

International Journal of Research and Analytical Reviews

UGC Approved Research Journal

Periodicity - Quarterly



Atman Publishing Academy



International Journal of Research and Analytical Reviews

Atman Publishing Academy

2061-C/2/B, Nr. Adhyatma Vidya Mandir, Sanskar Mandal, Bhavnagar-364002.

Contact : 9427903033 E mail : editorsijrar@gmail.com, ijrar1@gmail.com



International Journal of Research and Analytical Reviews

ijrar.com

©IJRAR -All rights reserved. Reproduction in any form is strictly prohibited.

This work is licenced under Creative Commons International Licence Attribution 4.0 E-version.

Rs. 900

Subscription	1 year	2 years	5 years
Individual	3,000	6,000	15,000
Institutional	4,000	8,000	40,000
Advertisement	1000/Black	3000/colour	Per Page

Send your Paper(s)/Article(s) and Contact us on any one of following

E mail: (1) editorsijrar@gmail.com (2) ijrar1@gmail.com (3) drbjoshi@ijrar.com

Contact No.: +91 9427903033

-
1. Thoughts, language vision and example in published research paper are entirely of author of research paper. It is not necessary that both editor and editorial board are satisfied by the research paper. The responsibility of the matter of research paper/article is entirely of author.
 2. Editing of the IJRAR is processed without any remittance. The selection and publication is done after recommendations of at least two subject expert referees.
 3. In any condition if any National/International University denies accepting the research paper/article published in IJRAR, than it is not the responsibility of Editor, Publisher and Management.
 4. Only the first author is entitle to receive the copies of all co-author.
 5. Before re-use of published research paper in any manner, it is compulsory to take written permission from the Editor – IJRAR, unless it will be assumed as disobedience of copyright rules.
 6. All the legal undertakings related to IJRAR is subject to Bhavnagar Jurisdiction.

Editor

International Journal of Research and Analytical Reviews

Atman Publishing Academy

2061-C/2/B, Nr. Adhyatma Vidya Mandir, Sanskar Mandal, Bhavnagar-364002.

Contact : 9427903033 E mail : editorsijrar@gmail.com, ijrar1@gmail.com

Editor in chief**Dr. R. B. Joshi****Senior Advisory Board**

Dr. H. O. Joshi Retd. Prof. & Head, Department of Education, Saurashtra University, Rajkot, Gujarat.	Dr. Bhavesh Joshi Associate Professor College of Food Processing Technology & Bioenergy, Agricultural University, Anand – 388110, Gujarat	Vasantkumar Pathak Director, Pathak Group of Schools & College, Rajkot.
---	---	---

Editorial Board

Prof. (Dr.) Ami Upadhyay Director, Department of Humanities And Social Sciences, Dr. Babasaheb Ambedkar Uni. A'Bad.	Dr. Awa Shukla Asst. Professor & Director, Social Sciences Dept. Babasaheb Ambedkar Open University, Ahmedabad.	Dr. Dushyant Nimavat Associate Professor Department of English, Gujarat University, Gujarat, India
Dr. A. Heidari Faculty of Chemistry California South University (CSU) Irvine, California, U. S. A.	Dr. Bharat Ramanuj Professor & Head, Department of Education, Saurashtra University, Rajkot.	Dr. Nahla Mohammed Abd El-Aziz Assistant professor - Entomology Department, Faculty of Science Cairo University, Egypt.
Dr. Manahar Thaker Principal G. H. Sanghavi college of Education, Bhavnagar, Gujarat.	Dr. K. S. Meenakshisundaram Director, C. A. A., Great Lakes Institute of Management, Chennai	Dr. J. D. Dave I/c Principal P.D. Malviya Graduate Teachers' College, Rajkot, Gujarat.
Dr. M. B. Gaijan Associate Professor, Shamaldas Arts College, Bhavnagar.	Dr. A. K. Lodi H.O.D. Faculty of Education, Integral University, Lucknow(UP)	Dr. Trupti Pathak Assistant Vice President(Tech.) Clarise life Sciences, Ahmedabad. Gujarat.
Dr. K. Ramadevi Associate Professor Department of Civil Engineering Kumaraguru College of Technology, Coimbatore, Tamilnadu.	Dr. Jayant Vyas Professor & Head, Department of Education, M. K. Bhavnagar University, Bhavnagar	Dr. Dilip D. Bhatt Associate Prof. & Head, Department of English, V. D. K. Arts college, Savarkundla, Gujarat.
K. S. Dave Lecturer J. H. Bhalodia Women's College Rajkot, Gujarat.	Dr. Anil Ambasana Retd. Prof. & Head, Department of Education, Saurashtra University, Rajkot. Gujarat.	Dr. Sandeep R. Sirsat Associate Professor & Head, Department of Computer Science, Shri Shivaji Science & Arts College, Chikhli, Dist: Buldana (M.S.-India)

Review Committee

Editor & Head of Review Committee

Dr. S. Chelliah

Professor & Head,

Dept. of English and Comparative Literature,
Madurai Kamraj University, Madurai-21, **India.**

<p>Mr. Zeeshan Shah Senior Lecturer, Department of Multimedia and Communication, University College of Bahrain, Kingdom of Bahrain.</p>	<p>Dr. Samira Shahbazi Plant Protection & Biotechnology Research Group, Nuclear Agricultural Research School, Nuclear Science & Technology Research Institute (NSTRI), Iran</p>	<p>Dr. Belal Mahmoud Al-Wadi Lecturer, University of Dammam (Saudi Arabia), Founder & Vice President of the Jordanian Society for Business Entrepreneurship (Jordan)</p>
<p>Harish Mahuvakar Associate Professor & Head, Dept. of English, Sir P. P. Institute of Science, Bhavnagar, Gujarat, India.</p>	<p>Dr. Mainu Devi Assistant Professor (Sr. Grade) in Zoology, Diphu Govt. college, Karbi Anglong – Assam India.</p>	<p>Asim Gokhan YETGIN Assistant Professor, Faculty of Engineering, Dumlupinar University, Kutahya, Turkey.</p>
<p>Dr. A. Kusuma Assistant Professor, Department of Social Work, Vikramasimhapuri University, Nellore.(AP)</p>	<p>Prof. Rajeshkumar N. Joshi I/C Dean, Faculty of Arts & Humanities, C. U. Shah University, Gujarat, India.</p>	<p>Sunita. B. Nimavat Assistant Professor of English, N.P.College of Computer & Mgt., Kadi (North Gujarat).</p>
<p>Nahla Mohammed Abdelazez Assistant Professor Faculty of Science, Cairo University, Giza Governorate, Egypt.</p>	<p>Dr. Riyad Awad Associate professor, Structural Engineering, An - Najah National University, Nablus, Palestine.</p>	<p>Dr. Amer A. Taqa Professor Dept. of Dental Basic Science, College of Dentistry, Mosul University, Masul, Iraq.</p>



THADOMAL SHAHANI ENGINEERING COLLEGE

Organizes

INTERNATIONAL CONFERENCE ON EFFICACY OF SOFTWARE TOOLS FOR MATHEMATICAL MODELLING (ICESTMM'19)

19th & 20th April 2019

In association with



University of Mumbai



Indian Society for
Technical Education

ABOUT THE CONFERENCE

The ICESTMM' 19 is the first most comprehensive conference focused on the various aspects of advances in applications and efficiency of software tools for mathematical modeling in the field of Computer, Electronics, Information Technology, Humanities and Management. Fundamental research is heavily relying on abstraction and modeling of physical phenomena/ processes and artifacts. Enabling ICT market places are flooded with numerous software products for modeling and further post processing. It is observed that specific software products are best suited for a segment of problem domains to be demystified. People with experience can tell us which one fits well with a specific situation and context of problems to be solved. It is our earnest effort to invite all experts/and researchers working in this area to contribute to the cause of simulation and modelling using software best suited for the same.

About the TSEC

Established in 1983, Thadomal Shahani Engineering College is founded by the Hyderabad (Sind) National Collegiate Board (HSNC). Thadomal Shahani Engineering College (TSEC) is recognized by the Government of Maharashtra and is affiliated to the University of Mumbai. It offers a 4-year Bachelor of Engineering (B.E.) Degree program in Information Technology, Computer, Electronics & Telecommunication, Biomedical, Biotechnology and Chemical Engineering. It has been accredited by NBA till June 2019 for the former four under graduate programmes and is graded as an 'A' grade educational institution. It also offers Ph.D. program in the subject of Information Technology, Computer Engineering, and Electronics & Telecommunication. The college is centrally located in Mumbai at Bandra, the queen of suburbs; and is very closely connected and accessible from all the major local train stations and expressways.

About H(S)NCB

Hyderabad (Sind) National Collegiate Board, Mumbai (Estd 1949) is the brain child of Vidyasagar Principal K M Kundnani and Barrister H G Advani. Today the cooperative management of the Board is in the capable hands of Trustees, Mr. Anil Harish (President), Dr Niranjana Hiranandani (Immediate Past President), Mr. Kishu Mansukhani (Past President), Ms. Maya Shahani and Mr. Lal Chellaram. The Board was established in the wake of Partition to provide quality educational facilities to the displaced Sindhi population and the local community and generate employment opportunities for them. Since the past 70 yrs. the board has been, thus an active partner in nation building through education and professional prospects for the citizens. The Board has traversed a long way from 1 institution in 1949, to 14 institutions affiliated to University of Mumbai. Over the years the Board has also established 9 schools, 1 self-financed management Institute and Polytechnic. The H(s)NCB Institutes are spread over Churchgate, Colaba, Cuffe Parade, Worli, Bandra and Ulhasnagar.

Conference Objectives

- To provide a platform for Academicians, Researchers, Industry Professionals and Students to present their views and findings on Advancements in software tools and mathematical modeling.
 - To enhance and develop opportunities for further research in the field of Mathematical Modeling.
 - To bridge the gap between Academia and Industry.
-

Sub-Thematic Areas

In particular, conference addresses the Advancements, Trends and Innovation in the field of Mathematical Modeling using Software Tools. Some of the thematic areas which the conference aims to focus are:

- Computational Techniques
- Heat flow and mass transfer in Porous media
- Ecological and Environmental Modeling
- Numerical Analysis and Methods
- Computational aided Simulation and Modeling
- Information Theory and Coding
- Artificial Intelligence
- Neural Networks and Fuzzy Logic
- Advanced Computing Architectures and New Programming Models
- Statistical Methods in technical, economic and medical Sciences
- Nonlinear, Adaptive and Complex Systems High Performance Computing and Visualization

The above mentioned list of sub themes is not an exhaustive one, research in any other relevant theme is highly appreciated.

Conference Committee

Chief Patrons

Mr. Anil Harish (President, HSNCB)
Ms. Maya Shahani (Trustee, HSNCB)
Mr. Lal Chellaram (Trustee, HSNCB)

Patron

Dr. Dinesh Panjwani (Principal, RD & SH National College)

Convener

Dr. G. T. Thampi (Principal, TSEC)

Organizing Secretary

Dr. Mansha Kotwani (In-Charge, Mathematics)
Ms. Poonam Soni (Asst Prof., Mathematics)

Organizing Committee Members

Mr. Gaurav Mittal (Assistant Prof., TSEC)
Dr. Arti Deshpande (Assistant Prof., TSEC)
Dr. Shachi Natu (Assistant Prof., TSEC)
Dr. Darshan Ingle (Assistant Prof., TSEC)
Ms. Uttara Bhatt (Assistant Prof., TSEC)
Ms. Nagaveni Hebbar (Assistant Prof., TSEC)
Ms. Monica Tolani (Assistant Prof., TSEC)
Ms. Sanobar Shaikh (Assistant Prof., TSEC)

Advisory Board

- Dr. Amit Dutta (WRO, AICTE)
Dr. U V Kulkarni (Professor, SGSCES)
Dr. S K Ukarande (Principal, KHSIEIT & Dean Technology)
Dr. S M Khot(Principal, FCRIT,Vashi)
Dr. J W Bakal (Principal. SSJCE)
Dr. K R Pardasani (Professor,MANIT, BHOPAL)
Dr. B V Rathish Kumar (Professor, IIT, Kanpur)
Dr. Ashu Rani (Professor, University of Kota)
Dr. Subhash Shinde (Prof., LTCE, Navi Mumbai)
Dr. Hari Vasudevan (Principal, DJSCE)
Dr. Nazneen Ansari(SFIT, MUMBAI)
Dr. Gresha Bhatia (VESIT, MUMBAI)
Dr. Amol Adamuthe (Professor,RIT,Sangli)
Dr. P H Sawant (Principal, SPCF, MUMBAI)
Dr. Neeru Adlakha (Associate Prof, NIT,SURAT)
Dr. Arun Kumar(Associate.Prof,Govt College, Kota)
Dr. Safique Ahmad (Associate Prof, IIT Indore)
Dr. D B Gurung (Prof, Kathmandu University, Nepal)
Dr. Robin Antoine(Mathematics &Statistics, University of Westindies)
Dr. Adela Mihai (Mathematics &Computer, Bucharest)
Dr. Rajesh Buktar (Prof. SPCOE)
Dr. Dhananjay Kalbande(Prof. SPIT)
-

Contents

01	Intelligent IoT based Fire Detection and Alerting System in Moderately dense and Open Forest Ramkrishna V. Vhatkar, Ashish P. Paikrao, Sanket D. Shinde, Amol C. Adamuthe	01 – 09
02	Text Based Context Identification in Content-Based Image Retrieval Nikita Ahire, Jyotsana Chavan, Radhika Deshpande, Priyanka Ilhe, Prof. Shilpa Mene	10 – 12
03	Data Deduplication Based on Hadoop Mehul Savaliya, Kushal Trivedi, Keval Shah, Sanjay Vidhani, Era Johri	13 – 15
04	Stress Analysis of Type IISub Trochanteric Fracture Femur Bone Implantation with Prosthetic Plates and Screws Using Finite Element Method. Mr. Karan Bhojwani & Prof. Chandan Chaudhari	16 – 24
05	Analysis and Prediction of Crime Against Women Asmi Patel, Dhvani Shah, Namrata Poojary, Priya Mishra, Alvina Alphonso	25 – 30
06	Prevalence of Anemia Among Women in Maharashtra - A Statistical Analysis Siddhi Pawar, Ruchira Patil, Prof. Poonam Soni, Dr. Mansha Kotwani	31 - 35
07	Automated Jurisdiction System Naje Aqeel, Sardar Sobiya, Sayed Rehan, Adhikari Shahzeen, Prof. Apeksha Gopale	36 – 39
08	Capacity Planning for Public Toilet Dr. Ashfaq Shaikh, Arun Pandey, Shushant Jha, Indrajit Wagmare, Fatima Patel	40 – 44
09	ConnectU: A Local Business Platform Saly M. Lopes & Melbourne B. Lopes	45 – 50
10	Copy-Move Forgery Detection Using Deep Learning VARUN PANDE, ABHISHEK AGARWAL, DEEPANSHU NEGI, ANKIT KUMAR, MANE VANITA	51 – 55
11	Approximate Analytical Solution of Counter-Current Imbibition Phenomenon Using Homotopy Perturbation Method Dhara T. Patel & Amit K. Parikh	56 – 61
12	Digital Game-Based Learning (DGBL): Design and Development of an Educational Game : Design of Digital Games as Pedagogical Tools Dr. Nazneen Ansari, Siddhesh Yadav, Abhijeet Prabhu, Vidhi Shah, Anuj Potdar	62 – 69
13	SQL Query Generator Sanobar Shaikh (Guide), Yogita Motwani, Sakshi Pandita, Simran Lund, Sanjana Rohra	70 – 74
14	Dr Device- Service Provider for all Devices Dr. Ashfaq Shaikh, Madiha Baig, Rubina Siddique, Hafiah Ansari	75 – 79
15	Exploring Elliptic Curve Cryptography Arithmetic using the Sage Math Computer Algebra Software Santoshi Pote & B.K. Lande & Sanchali Kshirsagar	80 – 86

- 16 **Fake News Detection** 87 – 91
Swati Modi, Disha Modi, Pooja Patel, Noella Thomas, Dr. Vaishali Jadhav
- 17 **Fault Detection and Diagnosis of Merchant Ship After Experiencing Collision** 92 - 102
MainakBhaumik, Dr. Arun Kumar Lokhande, Jagdish Chahande, Suraj Marale, Shivam Shukla, Chaitanya Rothe, Mahesh Pisal, Mahesh Patil, Someshwar Haibathpure, Rekha Rathore, Nilesh Bodkhe
- 18 **Performance Evaluation of Association Rule Mining Using Weka Tool** 103 – 110
Bina Trivedi, Priti Nandankar, Anjali Singh, Rachel Mascarenhas, Dr. Nazneen Ansari
- 19 **A Review of Mathematical Multi-Criteria Decision Models with A case study** 111 – 124
Udaykumar S. Kashid, Archana U. Kashid, Saurabh N. Mehta
- 20 **Recent trends in Advance Coding Techniques** 125 – 129
Prof. Sandip J. Dawda & Dr. Saurabh Mehta
- 21 **Analysis of Frequent Amino Acid Association Patterns in Peptide Sequences of Dengue Virus Type 1** 130 – 134
Dr. Amita Jain & Ms. Janhavi Vadke
- 22 **Automatic Electoral Identity Generator with OTP Authentication** 135 – 140
Nirali Shah, Shirish Nair, Bhavin Oza, Russel D'britto, Ms. Purnima Kubde
- 23 **Architect: Augmented Reality application for architect's portfolio** 141 – 145
Sahil Raut, Shraddha Pawar, Remon Pereira, Monisha Murdeshwar, Sonali Vaidya
- 24 **Interactive Smart Mirror Screen** 146 – 150
Er Kashif Kaularikar, Ali Mehdi Hemani, Danish Ayub Khan, Nehal Contractor
- 25 **Application of Statistical Methods in Medical Sciences** 151 – 154
Manjiri P. Vartak & Daksha P. Sawant
- 26 **Similarity Analysis of an Unsteady Flow of Non-Newtonian Fluid past through a Porous Surface via Multiparameter Group Technique** 155 – 160
Jayshree Gangani & M. G. Timol
- 27 **Virtual Clothing Try-On Using Augmented Reality** 161 – 164
Neha Sharma, Nihal Shetty, Shreyash Shetty, Riken Shah, Sonali Vaidya
- 28 **Implementation of Embedding Algorithm for Data Hiding using Steganography Technique** 165 – 171
Dinaz R. Adajania & Prof. Sanjay Vidhani
- 29 **Managing Road Junctions Based on Traffic Flow Estimations & Road Conditions** 172 – 178
Dr. Riyazoddin Siddiqui, Shubham Parab, Ali Electricwala, Mustakim Merchant
- 30 **Model of Calcium Dynamics to Study the Role of Amyloid Beta in Alzheimer's Disease** 179 - 184
TEJAL KANGANE, SAIF SAYED, DR. MANSHA KOTWANI
- 31 **Numerical Estimation of Calcium Concentration in Liver Cells Involving Influx, out Flux and Buffers** 185 - 191
Yogita D. Jagtap & Neeru Adlakha
-

- 32 **Generation of 3d Retinal Model from Optical Coherence Tomography Images** 192 – 197
Ms.Mrinmoyee Mukherjee, Jerome Judith, Tabassum Chohan, Prasad Chavan, Camilla Dabre
- 33 **Android Based Bus Tracking Application** 198 – 203
Anagha Koroth, Tanmay Gajbhiye, Vishal Shrivastava, Sarvesh Pandey
- 34 **An Improved Pagerank Algorithm Based on Latent Semantic Model and Integration with N-Grams** 204 – 212
Bharti Khemani, Rushikesh Gajmal, Viral Patel, Deval Ghevariya, Daanyaal Kapadia
- 35 **Design and Implementation of Gas Sensor** 213 – 216
Varindra Kumar & Anita Kumari
- 36 **Parking Management System** 217 – 220
Dr. Nazneen Ansari, Nirantak Singh Raghav, Het Shah, Manan Sheth
- 37 **Placement Data Anatomization Using Tableau** 221 – 225
Sahil Rajendrakumar Raut, Shraddha Somnath Pawar, Nazneen Ansari, Vivian Brian Lobo, Vaibhav Kala, Snehal Dmello, Elizabeth George
- 38 **Proposed Research for Detecting Poachers** 226 – 230
Shreya Shivaji Gaikwad, Cyril Silvester Dabre, Pranit Mohan Mhatre, Dr. Nazneen Ansari
- 39 **Public Toilet Health and Hygiene monitoring using IOT** 231 – 237
Dr. Ashfaq Shaikh, Mohammed Amir Sayed, Riyaz Shaikh, Rushikesh Naik
- 40 **Rentkarde- Mobile Based Android Application.** 238 – 242
Er. Farida Attar, Mulla Mohammed Sufiyan, Khan Danish, Ghadi Ashish.
- 41 **Representation of finite transforms and its relations** 243 – 251
V. R. Lakshmi Gorty
- 42 **Research and Study on Domain-specific Search Engine using Naïve Bayes classifier with HITS algorithm** 252 – 259
Aliya Kazi, Amruta Dabholkar, Bhavana Choudhary, Shifa Shaikh, Bharti Khemani,
- 43 **Solution of Groundwater Infiltration Phenomenon in horizontal direction By DTM and RDTM Method** 260 – 268
Krisha Shah & Amit K. Parikh
- 44 **Statistical Analysis Of Air Pollution, Diseases and Mortality: A Case Study of Delhi** 269 – 276
Deepakshi Kasat, Bhairavi Savur, Poonam Soni
- 45 **The Address App** 277 – 281
Er. Kashif Kaularikar, Aves Cutleriwala, Abdul Wahab Machiwala, Mohd Maaz Khan, Khan Umair
- 46 **Classical Music Sub-Genre Classification Using Machine Learning** 282 – 286
Venitto Chettiar, Dexson Dsouza, Freya Dsouza, Feni Martina Fernando, Prajyoti Dsilva
- 47 **Emerging Trends in Cloud Computing: A Survey** 287 – 292
Apurva Saxena & Dr. Pratima Gautam
-

- 48 **Water Quality Monitoring & Notification System Using IoT** 293 - 297
Junaid D Momin, Raheen N Shaikh, Arkam Ansari, Mubashir Khan
- 49 **Dual-Phase-Lagfractional Thermoelastic Investigation of a Spherical Cavity with Finite Speed of Thermal Wave Propagation** 298 - 308
Gaurav Mittal¹, Vinayak. S. Kulkarni^{2*}
-



Intelligent IoT based Fire Detection and Alerting System in Moderately dense and Open Forest

¹Ramkrishna V. Vhatkar,²Ashish P. Paikrao,³Sanket D. Shinde, ⁴Amol C. Adamuthe

^{1,2,3}B. Tech Student,⁴Assistant Professor

^{1,2,3,4}Department of Information Technology,

^{1,2,3,4}Rajarambapu Institute of Technology, Rajaramnagar, MS, India

ABSTRACT: *In India and world, forest fires is one of the reasons behind the deforestation. The main objective of this project is to develop a system to detect forest fire in moderately dense and open forest. We designed Intelligent IoT based fire detection and alerting system by integrating multiple sensors. Proposed system detects the forest fire using flame, smoke, temperature and humidity sensors. Fuzzy logic is used to determine the possibility of fire based on integrated results of multiple sensors. Proposed fuzzy logic improves the early detection of fire and minimizes errors in the false alarm notification.*

Keywords: *Fire detection, sensor network, fuzzy logic.*

I. Introduction

The forest fire is an uncontrolled fire occurring in nature. Once the fire starts ignited it rapidly spreads all over the area in the forest. In India, there is by 46% increase in the number of forest fires in the last 16 years i.e. from 2003 to 2017. In just two years 2015 to 2017, the number is increased from 15,937 to 35,888. In 2017, the maximum number of different forest fires reported in Madhya Pradesh (4,781) followed by Odisha (4,416) and Chhattisgarh (4,373). This requires a much efficient and very effective method in the forest fire because these fires resulted in the loss of wild lives and damage of forest property. Once the fire starts, it rapidly spread all over area in the forest. Based on data received from the Forest Survey of India (FSI), "more than 1,500 cases of forest fires from the entire Maharashtra state between February and March so far," said R.S. Yadav, additional principal chief conservator of forest (protection) according to (Hindustan Times, 2018). According to a 2014 report from the government's National Institute of Disaster Management (NIDM), the country is losing an average of a million hectares to fires each year. The losses, combined with those from surging illegal logging. According to the Environment Ministry's 2015 Forest Survey of India report, India has lost 2,511 square kilometers of dense and moderately dense forest since 2013.

In literature, researchers implemented fire detection systems for home, industries, and forests. Sharma et al. (2017) proposed a forest fire detection system by monitoring the values of CO₂ level and temperature. Imteaj et al. (2017) presented system based on temperature, flame and smoke sensors. Review on existing fire detector based on automated fire voice alerting system for alerting any fire incidents in industrial premises is presented by Sathishkumar et al. (2016). The main objective of this paper is to design a system for the forest officer to detect smoke, fire, temperature that indicates the wildfire in forest. In addition, it generates an alert message on a display device when a substantive quantity of smoke, fire, motion detection is sensed. We can detect fire from suddenly rise in temperature and a decrease in humidity. Once the sensor detects these things, alert message directly sends to forest officer's display and further action taken by the respective authority.

The next section describes related work. Section III describes the design and implementation of proposed IoT based system. Section IV is about the fuzzy logic algorithms, membership functions, if then else rules. At last, conclusions are stated in section V.

II. Related Work

A) Image/video processing techniques

In the forest, various types of accidents are found one of them are wildfire which is available at a big rate of the chances of this fire has increased because of several impacts on the climate change and in the weak ecosystems. Right now, there are several various systems are present for monitoring purpose and detecting wildfires but some of them are not well developed. Another main aspect is Farmers try to extend their fields into forest areas, and that is why it is much important to detect motions in the forest which results in the safety of forest and existing wild lives in the forest.

The fire detection system is designed based on video input data. Digital image processing and embedded vision concepts are investigated. Thresholds are set to activate the fire alarms. The proposed system integrates the color spatial, temporal and motion information (Vijayalakshmi & Muruganand, 2018).

The fire detection system is based on the identification of the color of fire in an image. The color of fire are of various types like orange, red, yellow and white. To overcome the problem with color is overcome by the integration of other techniques such as smoke detection, background segmentation etc. The brightness level is identified to set the thresholds for the system that is proposed. Basically, fire and non-fire like images are differentiated. Paper compared five vision based fire detection system. Systems are based on flame color detection combined with other features such as motion and area of frame (Pritam & Dewan, 2017).

Authors used a fire detection technique based on the YCbCr color model. They separate luminance from chrominance by comparing other color spaces that are RGB and RGB(normalized). Separation of fire flame pixels from high-temperature fire center pixels is done based on statistical parametric values of fire image. They have tried to improve the accuracy of fire detection and tried to make the outputs real time with the help of moving camera (Premal & Vinsley, 2014).

Paper presented a stereoscopic imaging method for the purpose of the detection of the fire. Authors used the camera (low-cost binocular CCD camera), to gain the images in a stereoscopic manner. However, the system is only limited to the fire caused due to gas. Authors developed an algorithm to segment the zone of fire in HSV. Location, size and propagation of fire are determined (Li et al., 2014). A hybrid algorithm based on a feed-forward neural network is designed to detect fire flame/smoke. Results showed the success of the algorithm. The ordinary camera is used to detect flame and smoke for the real-time detection of hazards by taking and process video data (Kandil et al., 2010). Paper presented a fire detection system based on digital image processing. Fire source is identified based on dynamic characteristics and color information of the desired region (Xiao et al., 2009). Authors used image full segmentation, subsequent isolation of segments and simplification in the presence and absence of fire or smoke (Asatryan & Hovsepian, 2015). The motion detection algorithm is used for fire and smoke pixels detection. Further, fire and smoke pixels are separated. A pixel in appropriate spaces are chosen specifically and in parallel pixels are also selected based on the dynamics of the area. This is done to avoid and overcome false detection (Morerio et al., 2012). Wildland forest fire smoke detection system is designed using Faster R-CNN. Improvement in the proposed work is suggested for improving the performance of the system in case of thin smoke (Zhang et al., 2018).

B) Sensor networks

In literature, many papers experimented the use of wireless the sensor networks (WSN) for forest fires (Doolin & Sitar, 2005; Son et al., 2006; Bernardo et al., 2007; Pripuzić et al., 2008; Antoine-Santoni et al., 2009; Sazak & Abdullah, 2009; Aslan et al., 2012). Proposed system using a wireless the sensor network is designed for early forest fire detection. The main aim is quick detection of fire with minimizing the energy consumption of the sensor nodes. The proposed framework is tested using simulators (Aslan et al., 2012). To predict the forest fire, 24-hour weather data is collected through a rechargeable wireless sensor network. The obtained data is processed using a fuzzy inference system and big data analytics (Lin et al., 2018). Early forest fire detection system experiments with a smoke sensor, two gas sensors and microwave sensors. Indoor and outdoor tests are conducted to validate the results (Krüll et al., 2012). Presented improved MODIS for detection of forest fire by adapting the HJ-infrared sensor (Wang et al., 2012).

C) Soft computing methods

In literature, different soft computing methods are investigated for forest fire monitoring (Anezakis et al., 2016; Mahdipour & Dadkhah, 2014; Olivas, 2003). Different soft computing techniques are used to predict forest fire (Artés et al., 2017; Bui et al., 2017; Denham et al., 2008). Comparison of five neural network techniques is investigated to identify the best suitable forest fire prediction technique (Al Janabi et al., 2018). In literature, CNN is investigated for the detection of forest fire (Hohberg, 2015; Frizzi et al., 2016; Zhang et al., 2016). The wireless sensor network is used to gather the required information for the detection of a forest fire. Authors generalized Mamdani's inference algorithm. Proposed overlapping functions and overlapping indices avoid bad results (Garcia-Jimenez et al., 2017).

III. System Design And Implementation

In the development of any system, there are various initial steps are there and that how we can analyze the functionalities and design of system and objective of our project is to design Intelligent IoT based fire detection and alerting system in moderately dense and open forest. Early detection of danger occurred in the forest and generate an automatic alert system in terms of messaging and it will help to protect the environment.

To control wildfire, there should be proper mechanisms to detect and identify the immediate and as soon as any danger is found without any presence of motion action this danger should get inform to responsible authority and as soon as message get that group of people there should be fast reaction from those respective people as quickly as possible. If there is any type of delay or failure in the system of respective authority then it may cause a very huge fire which can't be controlled that easily or we can say it becomes almost impossible to stop the fire and from irreversible damages which can cause deaths of wild animals from the forest.

Fig.1 and 2 shows the system architecture and flowchart diagram of the proposed system. Fig. 1 describe the raspberry pi 3 microcontroller converts all of the various analog values from each sensor used in this system and converts each of them into the digital values. The several IoT platforms are implemented in this prototype which is systematically fully managed, to ease and derive various value from IoT devices is sent to display device and it was used to record data from the various devices. The fire detection system and motion detection devices are connected to the display device by using the lightweight wires and secured by the MQTT protocol. Actually, this IoT various devices can create a lot of different data. In fact, to detect signals by sensor technology on smart devices already in the use of accounts by 30 percent varies by the world's data. Such as some of IoT technology like Bluemix generates a NoSQL database which is further used for purpose of storage for various devices and that data sent by the proposed system developed in this project. Once our connection of hardware configured with the Raspberry Pi, our code for decision is written in Noob/Raspbian that is in python language however we can customize the code to get any desired format of sensor data in software design. In the implementation of smoke, fire, DHT11 and PIR sensor are done using standard library codes. This program is used in python in raspberry pi and then uploaded on board then again it uploaded into raspberry pi and we can achieve output on the monitor screen. From various sensor, further data is collected and the result will be displayed on a desktop device.

IV. Results and Discussion

This section presents the experimental details, results obtained using sensor networks, fuzzy results and discussion.

The system is designed with MQ2 sensor for smoke sensing, DHT11 sensor for temperature and humidity sensing, SEN16 sensor for flame sensing, HC-SR501 sensor for motion detecting. Each sensor is integrated with the raspberry pi 3 microprocessor with minimum interfacing components. The desktop device is interfaced with the microprocessor for transmitting and receiving information to and from the base station.

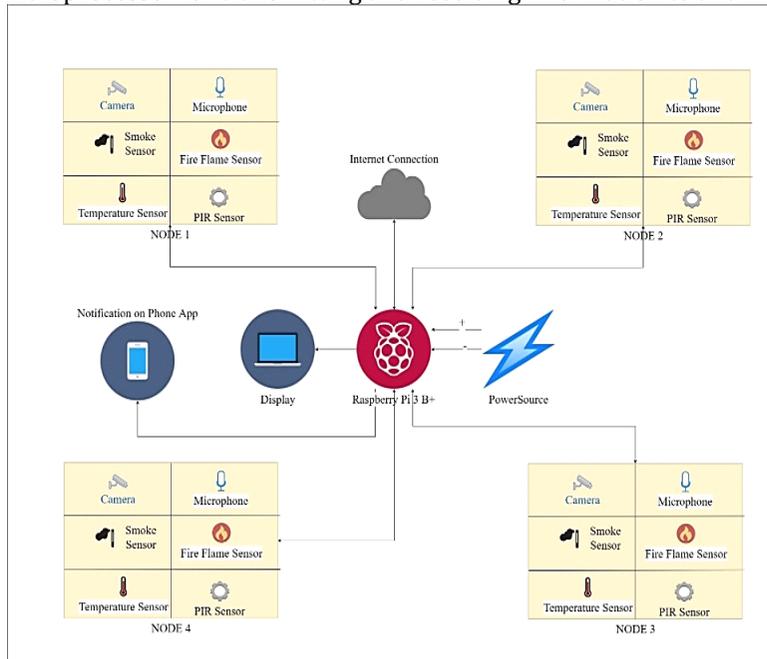


Fig. 1 System Architecture

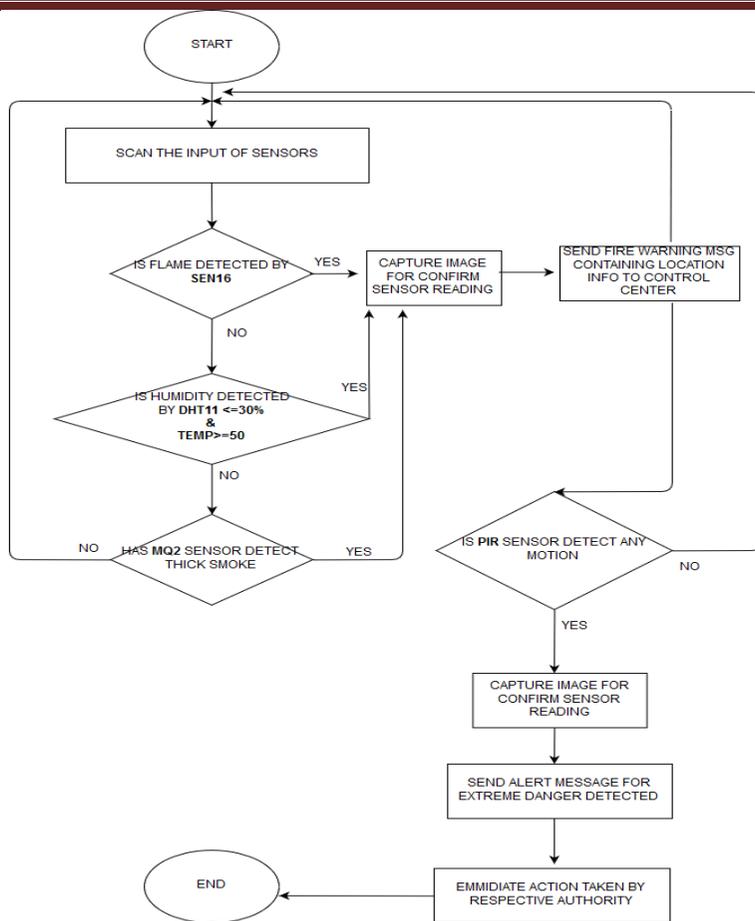


Fig. 2 Flowchart Diagram

With a few interfacing components, the sensor node is designed. Based on fuzzy logic the software is designed inside the node to analyze the fire risk level based on the smoke, flame, temperature, humidity and distance. The sensor node distance is calculated using the scale. It also considers the distance between the node and sensor. When smoke particles rise causes the rise in the threat of fire risk percentage. The temperature and humidity value are directly measured by the sensor and displayed in the display module. Once the node received its authenticated information from the forest, it sends the fire risk level to the base station to minimize wildfire. Using the display device in the base station the information is shared with users. Table 1 shows the results of flame and smoke detection using sensors.

Table 1 Hazard Detection

Sr. No.	Source	Sensor Used	Length (cm)	Detection
1	Match Box	SEN16 Flame	55	Flame detected
2	Cigarettes	MQ2 Smoke	4	Smoke detected
3	Lighter	SEN16 Flame	65	Flame detected
4	LPG gas	MQ2 Smoke	10	Smoke detected
5	Paper Burning	SEN16 Flame & MQ2 Smoke	87	Flame detected

The threat of fire risk in forest increases when temperature increases and humidity decreases. This threat level is formed in three different categories that are low, medium and high. The separate range is defined for each input and number of various variables for each one of the membership function is predefined as follows:

The range of temperature values for indicating at which level fire is present described below.

- Temp < 30°C - Low
- 30°C < Temp < 45°C -Medium
- Temp > 45°C - High

Fuzzy equation 1 is about rise in temperature. We have taken values of a, b as 0 and 45 respectively. We can detect the actual probability of wildfire using this membership functions. The threat of fire risk in forest increases with rise in temperature and decrease in humidity. Based on this fuzzy logic concept we draw various graphs for each different scenario. Fig. 3 describe the threat of fire risk increases in probability from 0 to 1 when rise in the temperature from 0° C to 45° C and above 45° C fire graph is constant i.e. probability of fire is 1 and graph is draw in MATLAB using fuzzy logic equation shown in Equation 1. Fire risk increases in probability from 0 to 1 when rise in temperature from 30° C to 45° C this high temperature generally measures in forest only because the normal temperature of the forest is less than 8° C normal temperature of the city.

$$\mu(x, a, b) = \left\{ \begin{array}{ll} 0; & x \leq a \\ \frac{1}{b-x}; & a \leq x \leq b \\ 1; & x \geq b \end{array} \right\} \tag{1}$$

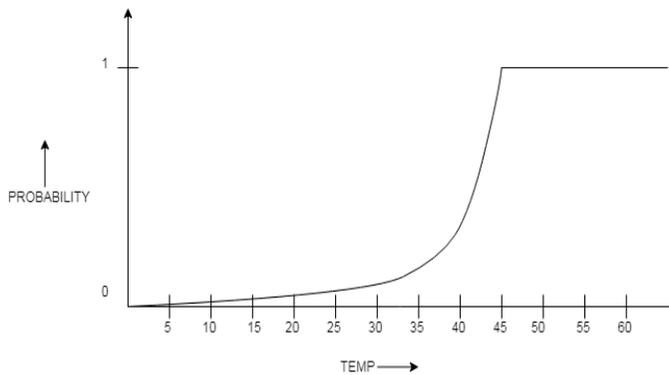


Fig. 3 Member Function of temperature input

In this system, we used two membership function one is for temperature and another is for humidity. Equation 2 shows membership function using fuzzy logic in humidity experiments. We have taken values of p, q is 3 and 6 respectively and y is humidity of the environment.

The range of humidity values for indicating at which level fire is present described below.

- Humidity > 60% - Low
- 30% < Humidity < 60% - Medium
- Humidity < 30% - High

```

Algorithm 1: Identification of sudden rise in temperature
for x in range (9):
    array[x]=array[x+1]
    array [9] =result. Temperature

count = 0
for x in range [8]
    y= x+1
    if array[y]>array[x]:
        count=count + 1

if count >= 8:
    print "increasing temperature detected, rapidly needs attention"
    
```

$$\mu(y, p, q) = \left\{ \begin{array}{l} 0; y \geq q \\ \frac{1}{y-2}; p \leq y \leq q \\ 1; y \leq p \end{array} \right\} \quad (2)$$

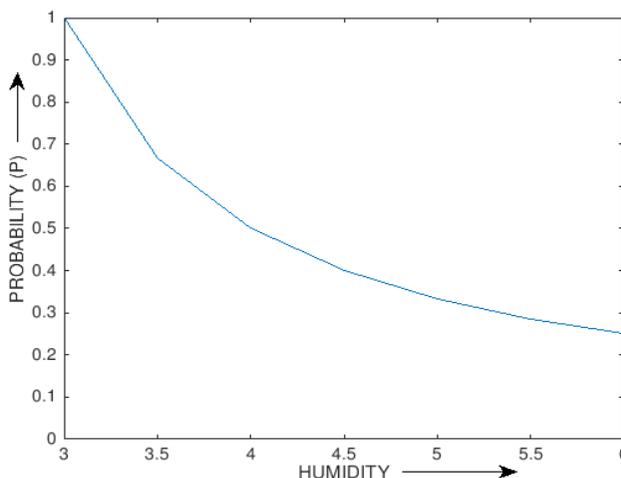


Fig. 4 Member Function of humidity

Fig. 4 describe the threat of fire risk increases in probability from 0 to 1 when describing in humidity from 60% to 30% and the graph is drawn in MATLAB using the fuzzy logic equation.

In table 2 shows the results of the fuzzy logic based system of fire detection. We get results of the probability of fire even at low temperature by using the following Equation 3.

$$P = \left(\left(\frac{1}{\text{MAX}-T} \right) \cdot \left(\frac{1}{H-2} \right) \right) \quad (3)$$

Where p stands for the probability of fire, MAX is the maximum temperature limit which we can change as per program, T and H represent temperature and humidity respectively.

Table 2 shows all results of combine sensors used in the project and integrate with raspberry pi to detect multiple results in one circuit connection for the safety of the forest. The table 2 consists of temperature, humidity, smoke, flame, distance and the output i.e. danger. The output of the smoke and flame sensor are considered Boolean. While the output of temperature and humidity sensors are categorized in Low, Medium and High. Table 2 presents the wildfire results based on following rules. Algorithm 2 presents the combined logic logic rules for fire detection. In fuzzy logic there are various conditions for if then else are generated which all are as follows:

- If (Temp is low && Humidity is high or (Smoke is not detected or Flame is not detected)) then no danger.
- If (Temp is low && Humidity is high or (Smoke is detected or Flame is not detected)) then fire is present.
- If (Temp is low && Humidity is high or (Smoke is not detected or Flame is detected)) then danger.
- If (Temp is medium && Humidity is medium or (Smoke is not detected or Flame is not detected)) then no danger.
- If (Temp is medium && Humidity is high or (Smoke is detected or Flame is not detected)) then danger.
- If (Temp is low && Humidity is low or (Smoke is not detected or Flame is not detected)) then no fire.
- If (Temp is high && Humidity is medium or (Smoke is not detected or Flame is detected)) then danger.
- If (Temp is high && Humidity is low or (Smoke is detected or Flame is detected)) then danger.
- If (Temp is high && Humidity is low or (Smoke is detected or Flame is not detected)) then danger.

Algorithm 2: Fuzzy logic for fire detection

```

import dht11
def fuzzy (temp, humidity):
    humidity=humidity/10
    global state
    max_temp=45
    float(temp)
    if max_temp > temp and humidity >3:
        state = (1/ (max - float(temp))) *(1/(humidity-2))
        print (state)
    else:
        state=10.0
        condition for very low fire alert message
        if state>=0 and state< 0.05
            print (Very low)
        condition for medium fire alert message
        if state>=0.2 and state< 0.1
            print (Medium)
        condition for high fire alert message
        if state>=0.3 and state< 0.2
            print (High)
        condition for extreme fire alert message
        if state>=1 and state< 0.5
            print (Extreme)
    else:
        print (Danger)
    
```

Table 2 Fuzzy results

Sr. No	Temperature	Humidity	Smoke	Flame	Distance	Fire Danger
1	Low	High	No	No	Far	No
2	Low	High	Yes	No	Average	Yes
3	Low	High	No	Yes	Close	Yes
4	Medium	Medium	No	No	Far	No
5	Medium	High	Yes	No	Average	Yes
6	Low	Low	No	No	Close	No
7	High	Medium	No	Yes	Far	Yes
8	High	Low	Yes	Yes	Average	Yes
9	High	Low	Yes	No	Close	Yes

V. Conclusions

The contribution of this proposed work is to develop an intelligent IoT system for forest fire detection. It performs tasks of monitoring forest, fire detection and alarm. In this project, we developed a system in which all sensor nodes identify accidental data such as fire, smoke, suddenly rise in temperature and decrease in humidity. Fuzzy logic is used to get early detection of danger i.e. detects fire danger at an early stage with high accuracy level from sensors data. We got fire alarms with a low, medium, high and extreme on forest officer's display device.

VI. Acknowledgment

This work was supported by Research Sensitization Scheme for College Students by Shivaji University Kolhapur, MS, India. The authors gratefully acknowledge all of these supports.

References

1. Anezakis, V. D., Demertzis, K., Iliadis, L., & Spartalis, S. (2016, September). A hybrid soft computing approach producing robust forest fire risk indices. In IFIP International Conference on Artificial Intelligence Applications and Innovations (pp. 191-203). Springer, Cham.

2. Al Janabi, S., Al Shourbaji, I., & Salman, M. A. 2018. Assessing the suitability of soft computing approaches for forest fires prediction. *Applied computing and informatics*, 14(2): 214-224.
3. Antoine-Santoni, T., Santucci, J. F., De Gentili, E., Silvani, X., & Morandini, F. 2009. Performance of a protected wireless sensor network in a fire. *Analysis of fire spread and data transmission. Sensors*, 9(8): 5878-5893.
4. Artés, T., Cencerrado, A., Cortés, A., & Margalef, T. 2017. Time aware genetic algorithm for forest fire propagation prediction: exploiting multi-core platforms. *Concurrency and Computation: Practice and Experience*, 29(9): e3837.
5. Asatryan, D., & Hovsepian, S. 2015. Method for fire and smoke detection in monitored forest areas. In *2015 Computer Science and Information Technologies (CSIT)* (pp. 77-81). IEEE.
6. Aslan, Y. E., Korpeoglu, I., & Ulusoy, Ö. 2012. A framework for use of wireless sensor networks in forest fire detection and monitoring. *Computers, Environment and Urban Systems*, 36(6): 614-625.
7. Bernardo, L., Oliveira, R., Tiago, R., & Pinto, P. 2007. A fire monitoring application for scattered wireless sensor networks. In *Proceedings of the International Conference on Wireless Information Networks and Systems*, Barcelona, Spain (Vol. 2831, p. 2831).
8. Bui, D. T., Bui, Q. T., Nguyen, Q. P., Pradhan, B., Nampak, H., & Trinh, P. T. 2017. A hybrid artificial intelligence approach using GIS-based neural-fuzzy inference system and particle swarm optimization for forest fire susceptibility modeling at a tropical area. *Agricultural and forest meteorology*, 233: 32-44.
9. Denham, M., Cortés, A., Margalef, T., & Luque, E. 2008. Applying a dynamic data driven genetic algorithm to improve forest fire spread prediction. In *International Conference on Computational Science* (pp. 36-45). Springer, Berlin, Heidelberg.
10. Doolin, D. M., & Sitar, N. 2005. Wireless sensors for wildfire monitoring. In *Smart Structures and Materials 2005: Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems*, 5765: 477-485. International Society for Optics and Photonics.
11. Frizzi, S., Kaabi, R., Bouchouicha, M., Ginoux, J. M., Moreau, E., & Fnaiech, F. 2016. Convolutional neural network for video fire and smoke detection. In *Industrial Electronics Society, IECON 2016-42nd Annual Conference of the IEEE* (pp. 877-882). IEEE.
12. Garcia-Jimenez, S., Jurio, A., Pagola, M., De Miguel, L., Barrenechea, E., & Bustince, H. 2017. Forest fire detection: A fuzzy system approach based on overlap indices. *Applied Soft Computing*, 52: 834-842.
13. Hindustan Times. <https://www.downtoearth.org.in/news/forest-fires-in-india-increased-by-125-per-cent-in-last-two-years-60349> (Accessed on 28/8/2018)
14. Hohberg, S. P. 2015. Wildfire smoke detection using convolutional neural networks (Vol. 2). Technical report, Freie Universitt Berlin, Berlin, Germany.
15. Imteaj, A., Rahman, T., Hossain, M. K., Alam, M. S., & Rahat, S. A. 2017. An IoT based fire alarming and authentication system for workhouse using Raspberry Pi 3. In *Electrical, Computer and Communication Engineering (ECCE), International Conference on* (pp. 899-904). IEEE.
16. Kandil, M., Salama, M., & Rashad, S. 2010. Fire detection using a dynamically developed neural network. In *ELMAR, 2010 PROCEEDINGS* (pp. 97-100). IEEE.
17. Krüll, W., Tobera, R., Willms, I., Essen, H., & von Wahl, N. 2012. Early forest fire detection and verification using optical smoke, gas and microwave sensors. *Procedia Engineering*, 45: 584-594.
18. Li, G., Lu, G., & Yan, Y. 2014. Fire detection using stereoscopic imaging and image processing techniques. In *Imaging Systems and Techniques (IST), 2014 IEEE International Conference on* (pp. 28-32). IEEE.
19. Lin, H., Liu, X., Wang, X., & Liu, Y. 2018. A fuzzy inference and big data analysis algorithm for the prediction of forest fire based on rechargeable wireless sensor networks. *Sustainable Computing: Informatics and Systems*, 18: 101-111.
20. Mahdipour, E., & Dadkhah, C. 2014. Automatic fire detection based on soft computing techniques: review from 2000 to 2010. *Artificial intelligence review*, 42(4): 895-934.
21. Morerio, P., Marcenaro, L., Regazzoni, C. S., & Gera, G. 2012. Early fire and smoke detection based on colour features and motion analysis. In *Image Processing (ICIP), 2012 19th IEEE International Conference on* (pp. 1041-1044). IEEE.
22. Olivas, J. A. (2003, August). Forest fire prediction and management using soft computing. In *Industrial Informatics, 2003. INDIN 2003. Proceedings. IEEE International Conference on* (pp. 338-344). IEEE.
23. Premal, C. E., & Vinsley, S. S. 2014. Image processing based forest fire detection using YCbCr colour model. In *Circuit, Power and Computing Technologies (ICCPCT), 2014 International Conference on* (pp. 1229-1237). IEEE.
24. Pripuzić, K., Belani, H., & Vuković, M. 2008. Early forest fire detection with sensor networks: Sliding window skylines approach. In *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems* (pp. 725-732). Springer, Berlin, Heidelberg.
25. Pritam, D., & Dewan, J. H. 2017. Detection of fire using image processing techniques with LUV color space. In *Convergence in Technology (I2CT), 2017 2nd International Conference for* (pp. 1158-1162). IEEE.
26. Sathishkumar, R., Vinothkumar, M., Devaraj Varatharaj, S., & Gowthaman, S. M. 2016. Design And Development Of Automatic Fire Detection Using Sms And Voice Alert System. *International Journal of Scientific & Engineering Research*, 7(5): 114-117.

27. Sazak, N., & Abdullah, H. 2009. The importance of using wireless sensor networks for forest fire sensing and detection in turkey. In 5th International Advanced Technologies Symposium (IATS'09) (pp. 13-15).
28. Sharma, A.K., Ansari, M.F.R., Siddiqui, M.F. and Baig, M.A. 2017. IOT enabled forest fire detection and online monitoring system. *Int J Curr Trends Eng Res (IJCTER)*, 3(5): 50-54.
29. Son, B., Her, Y. S., & Kim, J. G. 2006. A design and implementation of forest-fires surveillance system based on wireless sensor networks for South Korea mountains. *International Journal of Computer Science and Network Security (IJCSNS)*, 6(9): 124-130.
30. Vijayalakshmi, S. R., & Muruganand, S. 2018. Fire alarm based on spatial temporal analysis of fire in video. In 2018 2nd International Conference on Inventive Systems and Control (ICISC) (pp. 104-109). IEEE.
31. Wang, S. D., Miao, L. L., & Peng, G. X. 2012. An improved algorithm for forest fire detection using HJ data. *Procedia Environmental Sciences*, 13: 140-150
32. Xiao, J., Li, J., & Zhang, J. 2009. The identification of forest fire based on digital image processing. In *Image and Signal Processing, 2009. CISP'09. 2nd International Congress on* (pp. 1-5). IEEE.
33. Zhang, Q. X., Lin, G. H., Zhang, Y. M., Xu, G., & Wang, J. J. 2018. Wildland Forest Fire Smoke Detection Based on Faster R-CNN using Synthetic Smoke Images. *Procedia engineering*, 211: 441-446.
34. Zhang, Q., Xu, J., Xu, L., & Guo, H. 2016. Deep convolutional neural networks for forest fire detection. In *Proceedings of the 2016 International Forum on Management, Education and Information Technology Application*. Atlantis Press.

TEXT BASED CONTEXT IDENTIFICATION IN CONTENT-BASED IMAGE RETRIEVAL

¹Nikita Ahire, ²Jyotsana Chavan, ³Radhika Deshpande, ⁴Priyanka Ilhe, ⁵Prof. Shilpa Mene
Department of Information Technology, KKWIEER, Nashik, India.

ABSTRACT: Search engines are designed to find specific information on the basis of specific criteria entered by the user. The list of items that matches the criteria specified by the query is typically sorted as ranked. Search engine needs an input as text to search. To make this searching effective we can give input as text and image also. To search relevant images related to query, Content Based Image Retrieval (CBIR) algorithm can be used. And if the image contains text then Optical Character Recognition (OCR) algorithm is used to extract that text and to show related images about it. The system can detect objects and search their relevance. Further, the image is retrieved from the dataset. To find a desired image from a large collection is very difficult task. In CBIR, features related to visual content such as shape, color and texture are first extracted from a query image. The similarity between the set of features of the query image and that of each target image in a database is then computed, and the most relevant image is shown to the user.

Keywords: Text detection, OCR, CBIR, Image Retrieval

I. Introduction

Search engine provide an interface to a group of items that enables users to specify criteria about an item of interest and have the engine to find the matching items. The list of items that meets the criteria specified by the query is typically sorted as ranked. Search engine needs an input as text to search. It doesn't take image as input. The proposed system can take input as text and image also. To search relevant images related to query, Content Based Image Retrieval (CBIR) algorithm is used. And if the image contains text then Optical Character Recognition (OCR) algorithm is used to extract that text and to show related images about it. To find a desired image from a large collection is very difficult task. In CBIR, features related to visual content such as shape, color and texture are first extracted from a query image. The similarity between the set of features of the query image and that of each target image in a database is then computed, the target images are next retrieved which are most similar to the query image.

II. RELATED WORK

Text detection methods aim at automatically detecting and generating bounding boxes of words in natural scene images. All these methods extract geometric, structural and appearance features from candidate regions to verify if a region contains text or not. It is difficult to have one global parameter setting which would accommodate for all possible text variations in natural images [7]. Therefore, it is necessary to tune these parameters for every new alphabet, text style and size. In contrast, our approach focuses more on background connectivity rather than text regions. The proposed method does not extract text specific features. Therefore, it does not require any tuning for varying text size, style and orientation. Existing methods for text detection using visual saliency [6] mainly focus on bottom-up information such as edges, corners, color distinctiveness and lines of symmetry. Top-down models are task dependent and use saliency and context to steer the search for objects in images. In our approach, background information is used for text detection.

III. PROPOSED SYSTEM

The proposed system based on two main phases. First for training the dataset or storing the data into the database and second is used for accepting the user query and producing the search results. Therefore the entire system is described in three phases- Query interfacing, feature extraction and finally the results listing. In first phase, query will be accepted in the form of image or text. If the given input is image then its feature (color, shape, textures) will get extracted on the basis of CBIR technique and further it get compared with the trained dataset. And if the image contains text then text detection and text extraction methods helps to relate the text with the database through OCR technique. Once the input gets mapped with the dataset, the

most relevant result will be shown to the user by the system. System will still respond to the query if it is in simple text form.

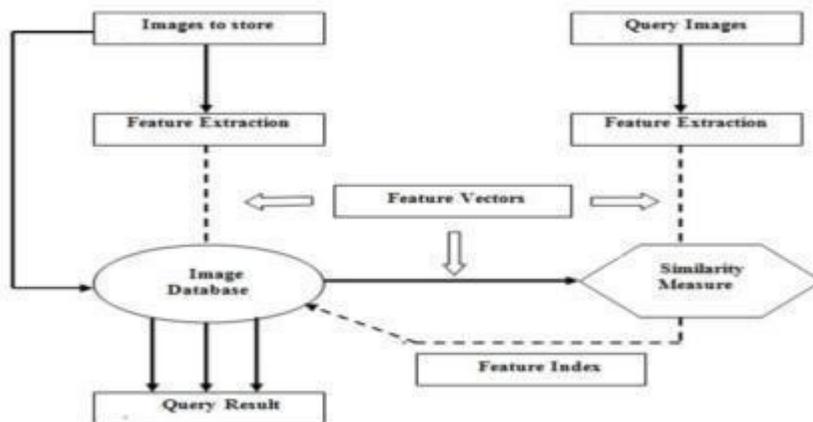


figure 1: flow of the system

3.1 Feature extraction

The content based images are retrieved by their image properties such as image objects edges, color distributions and the image textures.

3.2 Text detection

Text detection methods aim at automatically detecting and generating bounding boxes of words in images. Text detection methods can be categorized into two classes based on how they search character regions: a connected component and a sliding window approach.

3.3 Text extraction

This method extracts geometric, structural and appearance features from candidate regions to verify if a region contains text or not.

IV. ALGORITHM

4.1 OCR

OCR (optical character recognition) is the recognition of printed or written text characters by a computer. This involves photo scanning of the text character-by-character, analysis of the scanned-in image, and then translation of the character image into character codes, such as ASCII, commonly used in data processing. In OCR processing, the scanned-in image or bitmap is analyzed for light and dark areas in order to identify each alphabetic letter or numeric digit. When a character is recognized, it is converted into an ASCII code. Its working is as follows:

- 1) Loading the image file.
- 2) Improving image quality and orientation.
- 3) Removing lines.
- 4) Analyzing the page.
- 5) Detecting words and lines of text.
- 6) Analyzing and fixing of broken or merged characters.
- 7) Recognizing characters.
- 8) Saving the file.

4.2 CBIR

Content-based image retrieval (CBIR), also known as query by image content and content-based visual information retrieval (CBVIR) is the application computer vision techniques to the image retrieval problem, that is, the problem of searching for digital image in large database. Content-based image retrieval is opposed to traditional concept-based approaches. "Content-based" means that the search analyses the contents of the image rather than the metadata such as keywords, tags, or descriptions associated with the image. The term "content" in this context might refer to colors, shapes, textures, or any other information that can be derived from the image itself. CBIR is desirable because searches that rely purely on metadata are dependent on annotation quality and completeness.

As number of images are stored in image databases. Its working is as follows:

- 1) Extract features as Feature space.
- 2) Store feature space in feature database.
- 3) When a query comes in either the form of query-by example or query-by-keywords, its feature space will be compared with those in the feature database one-by-one.
- 4) And the relevant images with the smallest feature distance will be retrieved.

CONCLUSION

In this paper, the text is detected and extracted from images. The system also displays the information about that text. Further it will be classified on the basis of text-based and content-based algorithm which is Optical Character Recognition (OCR) and also uses the Content-Based Image Retrieval (CBIR) methodology. The system will be capable of detecting and extracting text from given input. Input can be in the form of text or image. The system will show similar images which are relevant to input.

References

- [1] T. Liu et al. 2011. Learning to detect a salient object. IEEE Trans. Pattern Anal. Mach. Intel., vol. 33, no. 2: 353–367.
- [2] S. Karaoglu, J. C. van Gemert, and T. Gevers. 2012. Object reading: Text recognition for object recognition. Proc.ECCV Workshops: 456–465.
- [3] S. Karaoglu, J. C. van Gemert, and T. Gevers. 2013. Con-text: Text detection using background connectivity for fine-grained object classification. Proc. ACM-MM: 757–760.
- [4] Mr. S.S. Mukati, Mr. Nitesh Rastogi.2016.A Survey on Image retrieval techniques with feature extraction. IRJET, Volume: 03.
- [5] R. Gnanarajan B. Jagadishkumar, S.T. Premkumar, B. Sunil Kumar. 2014.Content Based Image Retrieval (CBIR) Using Segmentation Process. ICETS, Volume 3.
- [6] A. Shahab, F. Shafait, A. Dengel, and S. Uchida. 2012. How salient is scene text?.Proc. IAPR Int. Workshop Document Anal. Syst.: 317–321.
- [7] C. Yao, X. Bai, W. Liu, Y. Ma, and Z. Tu. 2012. Detecting texts of arbitrary orientations in natural images. Proc. CVPR: 1083–1090.
- [8] B. Epshtein, E. Ofek and Y. Wexler. 2010. Detecting text in natural scenes with stroke width transform. Proc. CVPR: 2963–2970.
- [9] K. Wang, B. Babenko, and S. Belongie. 2011. End-to-end scene text recognition. Proc. ICCV: 1457–1464.

DATA DEDUPLICATION BASED ON HADOOP

¹Mehul Savaliya, ²Kushal Trivedi, ³Keval Shah, ⁴Sanjay Vidhani, ⁵Era Johri

¹ UG Student, ² UG Student, ³ UG Student, ⁴Assistant Professor, ⁵ Assistant Professor

^{1,2,3,4,5}Dept.ofInformation Technology

^{1,2,3,4,5}K.J.S.C.E, Vidyavihar, Mumbai, India

ABSTRACT: Data generated by user on social media and by various companies is increasing day by day. It is heavy challenge to copy these multiform of data in real time. To avoid data duplicates and increase the data reliability Hadoop distributed file system is designed to deal with duplicate data. MapReduce and HBase along with the new standard of Secure Hash Algorithm-3(Keccak) to speed up the deduplication procedure. The files with duplicate content will share a single copy which improves the utilization of the cloud storage space effectively.

Keywords: Data deduplication, HDFS, Secure Hash Algorithm-3

Introduction

Nowadays the amount of data generated from various source is tremendous. Typical real-world examples are Facebook, YouTube and Twitter etc., can be easily found in social applications, where file duplication occurs frequently. However, the critical challenge faced by providers is duplicate copies of file contents which consume more storage than needed. Duplicate data is being accumulated in large amount [2]. For example, user shares a multiple file with multiple users which consumes more cloud storage for a single file. Duplicate files need to be handled in a dataset; deduplication compares files without looking at the content of the file for faster run time. If data sets contain same files rather than identical files, deduplication process should identify the duplicate data by the content of the file to check which part of the content is same as previously stored data for better storage space savings. Data deduplication significantly decreases storage capacity because it eliminates excessive copies of data. For example, an email server which contains 10 instances of same file of 1 MB. It will require 10 MB storage, if everyone backup the email inbox without data deduplication. If data deduplication is implemented, then only one instance of the attachment is stored and each subsequent instance is just referenced back to the one saved copy [4].

RelatedWork

Paper: Dynamic Deduplication Decision in a Hadoop Distributed File System

Author:Ruay-Shiung Chang, ChihShan Liao, Kuo-Zheng Fan, and Chia-MingWu

Publisher: Hindawi Publishing Corporation International Journal of Distributed Sensor

Year:2014

Synopsis:

Data pre-processing can consume lot of time of the user which can prevent user from using the system. Data which needs to be stored is first pre-processed. Different pre-processing types are Normalization, Data cleaning, Data hiding, structuring of data, etc. which takes lot of time to implement. The process of splitting the data is called chunking which is the second step. The chunking is fixed to obtain more number of duplicates as much as possible. Hashing is the next step after chunking. It is essential part as the hashing values will be used to identify the duplicate files. For Hashing Secure Hash Algorithm-2 is used. Hash value for same data will be same, so it is accurate and efficient way to identify the duplicates. Data structure such as tree array or list is used to store the calculated hash value. To identify the duplicates, we need to compare value of the stored data with data to be uploaded. The amount of time taken to compare the hash value is called lookup time. For better performance the look up time should be less as much as possible. When user want to access the file the chunks are combined together to form the original data.

ProposedWork

The proposed system can be explained in 5 steps

1. File Aggregation

Hadoop is unable to deal with large number of small files. So, we need to assemble the small files into a large Hadoop file. Parallel Hadoop Aggregation file procedure is used. Initializing the Map function with key

parameter set to the name of the file and value set to the content of the file. The output is formatted in the Hadoop Sequence File. The output which is aggregated is collected by using MapReduce job.

2. Chunking

When data is stored in HDFS, it breaks the information down into separate blocks and distributes them to different nodes in a cluster.

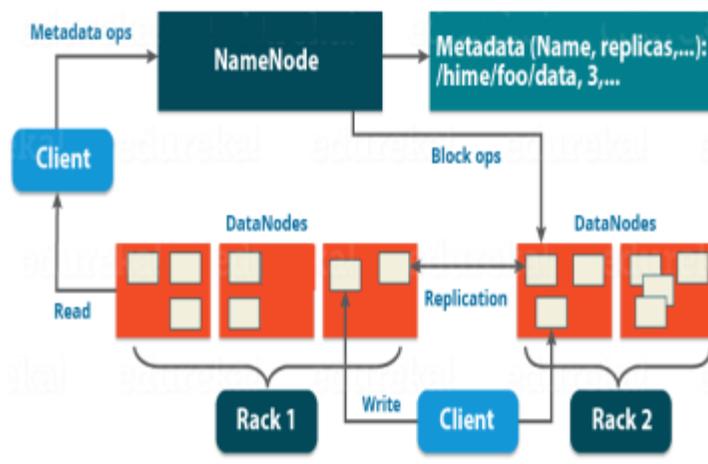


Figure 1.HDFS Architecture

HDFS consist of Namenode and Datanode. Namenode maintains and manages the Data Nodes and it records the metadata of all the files stored in the cluster. Actual data is stored on DataNodes

3. Hash Computation

Input for Hashing will be the output of the definite size partitioning function to compute the hashing of each chunk. For hashing the new Secure Hash Algorithm-3(SHA-3) will be used for security reasons. Hashing will be used to identify for the presence of the new chunk and existing chunks. Secure Hash Algorithm-3 is preferred over Secure Hash Algorithm-2 due to the time cost.

4. MapReduce and HBase

HBase is the Hadoop's distributed database which provides faster read and write operation in a large dataset. For future file access we store the chunk information in HBase which refer to duplicated files. MapReduce function is used to process large amount of data in-parallel. It is used to divide the input data into chunks which are independent of each other. The output of the map function is used as an input for reduce task. Both the input and output of the job is stored in the file system. It operates on key and value pair.

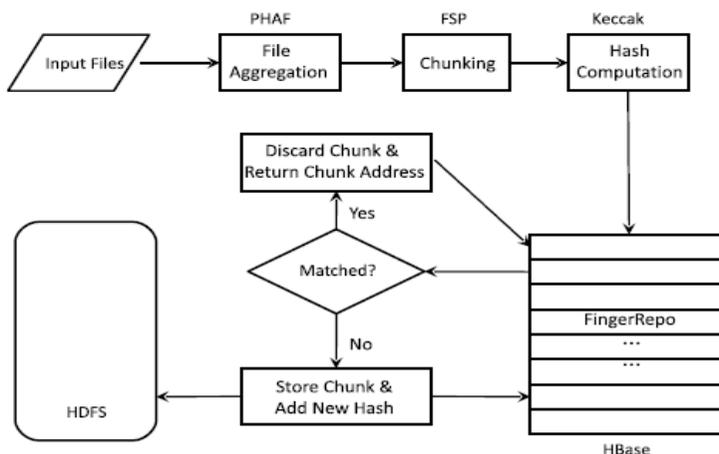


Figure 2. Flowchart of Deduplication

5. File Access

After applying the deduplication the file which contains the duplicate chunks will share a single copy.

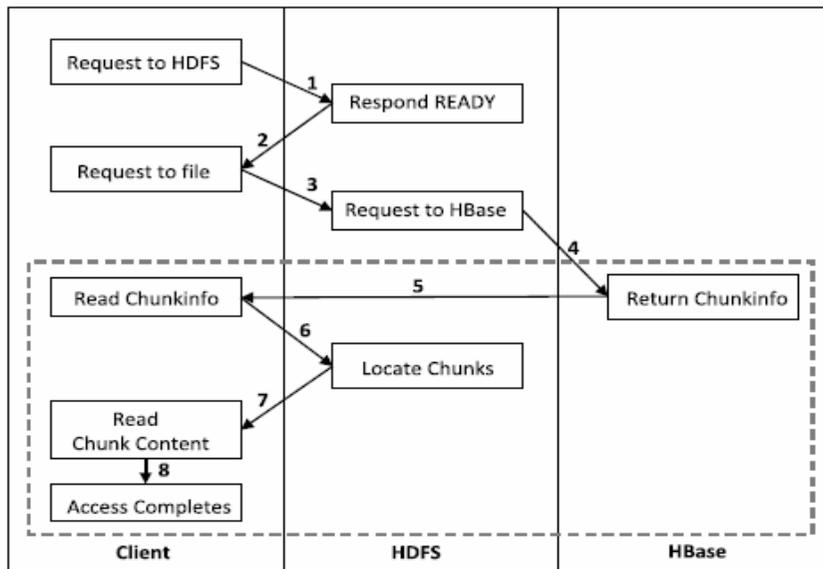


Figure 3.Steps of File Access after Deduplication

When user wants to access the file from the server, request is sent to HDFS and request is sent to HBase for the information of the chunks, HDFS locate the chunks of the file and all the chunks are together to form the file.

Implementation

Front-end of the system is developed using HTML,CSS and PHP.Connected the php and Hadoop file system using java for uploading, downloading and viewing the Hadoop directory. When the file is stored in HDFS, it chunks the file into blocks of 1MB. Hashing has been done using Secure Hash Algorithm-3.

Conclusion

In this technique we will utilize multiple clusters using parallel processing of the system. Data will be chunked into definite size partitions of 4k-bytes. The output of this chunk will be the input for the hashing. For hashing, Secure Hash Algorithm-3 is used over the Secure Hash Algorithm-1 and Secure Hash Algorithm-2 as it provides approximately 5 times better performance when same resources were used. The hashed chunks will be compared to the HBase for existing chunks and appropriate operations will be executed.

References

1. Dongzhan Zhang, Chengfa Liao, Wenjing Yan, Ran Tao and Wei Zhe, "Data Deduplication based on Hadoop", 2017 Fifth International Conference on Advanced Cloud and Big Data
2. Lars Kolb, Andreas Thor, Erhard Rahm, "Dedup: Efficient Deduplication with Hadoop", The Proceedings of the VLDB Endowment (PVLDB), 2012
3. Ruay-Shiung Chang, Chih-Shan Liao, Kuo-Zheng Fan, and Chia-Ming Wu, "Dynamic Deduplication Decision in a Hadoop Distributed File System", Hindawi Publishing Corporation International Journal of Distributed Sensor Networks Volume 2014
4. Apache Hadoop MapReduce.<https://hadoop.apache.org/docs/stable/hadoop-mapreduce-client/hadoop-mapreduce-client-core/MapReduceTutorial.html>

Stress analysis of type II sub trochanteric fracture femur bone implantation with prosthetic plates and screws using finite element method.

¹Mr.Karan Bhojwani & ²Prof.Chandan Chaudhari

¹UGStudent, ²AssistantProfessor

¹Department of Mechanical Engineering, SIES Graduate School of Technology, Navi Mumbai, India.

ABSTRACT: The use of finite element method (FEM) has grown to a more importance in the field of biomedical engineering. Although increased computational performance allow us to create more complex geometries in the field of orthopedic surgery. These computational methods help us to explore effect of various bio-materials on the sub-trochanteric fracture femur bone. In this study the type II sub trochanteric femur bone were modeled based on the dimensions. The femur bone is longest and strongest bone in the human body. It needs to support maximum weight of the body in between hip joint and knee joint. The femur bone contained linear elastic, isotropic, homogeneous material of calcium phosphate. The main objective of this study is to rectify the behavior of sub-trochanteric fracture femur bone for a man of age 48 years with prosthetic bone plate and screws under compression loading condition. Results were analyzed on the prosthetic plates and screws for Type II sub-trochanteric fracture femur bone i.e. (The fracture is within 2.5 cm of the lesser trochanter) during static loading condition with different bio materials. The comparison was made with various screws and plates made by various bio-materials which concluded in the implementation of the materials having minimum Von-misses stress, deformation of the fractured femur bone. On the basis of the results obtained, an alternative grade is examined and comparison has been made with the existing grade of titanium.

Keywords: Femur bone; Prosthetic; Biomaterials; Static structural; Stresses.

1. Introduction:

According to the World Health Organization [1], proximal femoral fracture or hip in simple terms in elderly people or with osteoporotic bone can occur frequently following a simple fall. The risk of hip fractures is approximately between 14%-20%, in the European nations. According to the reports 90% of hip fractures occur amongst people of 50 years of age and 90% occur in women. There are basically two types of hip fractures such as Intracapsular (Femoral Neck Fracture) and extra capsular (Trochanteric Fractures). Trochanteric fractures are generally osteoporotic and the risk of hip fractures generally increases with an increase in age and even with difference in sex [2]. On an assumption 9 million osteoporotic fractures were estimated out of which 1.6 million were hip fractures and the remaining forearm and clinical vertebral fractures. In research study it was also found that, the choice of internal fixation yield more robust fixation; especially in the regions surrounding the fastener points of the implant. The purpose of internal fixation is to keep the contact stability between the bone fractures. These techniques widely used in the field of orthopedics so as to repair the fracture bone with the aid of prosthetic bone plates and screws

Nomenclature:

FEM	Finite element modelling
FEA	Finite element analysis
CT	Computer tomography
PMMA	Poly methyl methacrylate
SS	Stainless steel

1.1 Femur Bone:

The femur is the longest and strongest bone in the human body, extending from the hip to knee. Femur bone is classified into three such as femur head, lateral epicondyles, and greater and lesser trochanter. The average length of a male femur bone is 48 cm i.e. 18.9 inch and 2.34 cm i.e. 0.92 inch in diameter [4]. The research study also claims that femur bone can able to support 30 times the weight of the adult person. Moreover, the femur bone supports all of the body weight during various activities such as jumping, walking, standing and running. The following figure 1 refers to anatomy of the femur bone. The femur bone is made of two primary biomaterials termed as cancellous bone and cortical bone. Cancellous bone is also termed as trabecular bone since this bone is very spongy in nature and found in core of the all the bone

structure. In contrast, cortical bone is the stronger part of the bone, and is located on the outer surface surrounding the cancellous bone.

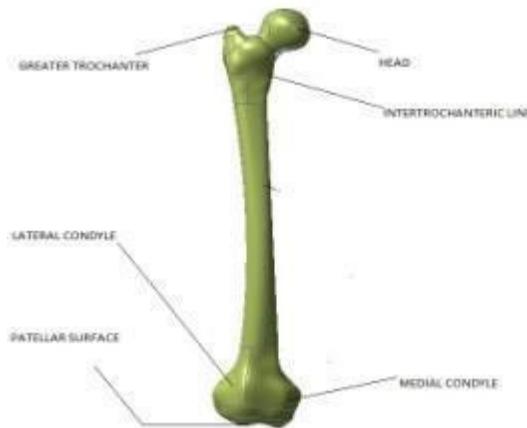


Figure1: femur bone anatomy

1.2 Biomaterials of fracture fixation plate:

Fixation plates are surgical tools, which are widely used to heal the fracture bones. This fixation technique provides a frictional force between the plate and bone to counteract the external forces experienced at the fracture region. A fixation mechanism is designed in such way that it eliminates the screw toggle and ensures less compression at the fracture region. Research study shows variety of Biomaterials implemented in orthopedic applications are listed in following table 1.

Table.1 mechanical properties of biomaterials for femur, plate and screws [5], [6],[7]

Bone and plate materials	Density (kg ⁻³)	Young's Modulus E (GPa)	Poisson Ratio	Ultimate Tensile Strength (MPa)	Ultimate Compressive Strength(MPa)
Femur cortical bone	1750	16.7	0.3	43.44	115.29
Femur cancellous bone	0.9	2.7e-4	0.26	0.19	0.17
Ti-6Al-4V	4500	120	0.32	993	1086
Alumina Al ₂ O ₃	8500	240	0.31	210	1920
Nylon 6/6	3720	300	0.21	65	16
PMMA	1180	220	0.2	47-79	83-124
SS316L	7750	193	0.31	485	570
Ti-29Nb-13Ta-4.6Zr	4500	80	0.33	990	1056

In this research study we used standard fixation plate with 150*15*5 mm dimensions having 8 holes. The fixation plates is made of various bio-materials along with 5 countersunk screws of same materials having 7.3 mm diameter of neck and 10 mm for head and 3 Counter screws with 5 mm diameter of neck and 8 mm for head It has been also noticed that the rigidity of an entire assembly that is generally sufficient with three screws in “normal” bone shows practically no increase beyond four screws per diaphyseal fragment. However, we have also seen that the quality of the fixation increases as the implant length increases. In practice, a sufficiently long implant is used so that one out of two holes can be left free in diaphyseal fixation

2. Structural Methodology:

First step in the structural methodology is to selection of suitable material of fixation bone plate and screw because this material is selected based on the following mentioned criteria.

- i]. Lower density
- ii]. Lower cost
- iii]. Recyclable and Biodegradable
- iv]. Higher strength & Stiffness which will ensure firm stability.
- v]. finally, material should also possess anti corrosive property along with homogeneous composition of fixation plate and screws preferred, because variations in elasticity in metallic implant may cause loosening of the implant.

2.1 Problem Identification:

The main requirement for the selection of biomaterial is based on the acceptability by the human body. The present study the algorithm followed would be starting with making a CAD model for the femur bone, followed by which a fixture parallel to trochanteric region was formed that could be easily attached ,considering the aim of the study being in that particular direction. We have considered various areas of contact and one type of compression loading. The analysis is based on the allowable stress and deformation level of the bone and the fixture system. The considered hypothesis are that the bone to be a model valid under the study of solid mechanics. It has been considered to be a two layer model with pores and with isotropic elastic mechanical properties. The bone is considered to be at the healing stage and the bone's bonded contacts are taken into account. The study is limited to type II sub trochanteric fracture only [8].

2.2 Geometry Modeling:

The three dimensional geometry model drawn from CT scan image based on the data received from a man of age 48 years as shown in figure 2. The femur bone consists with cancellous solid from the interior and cortical at the exterior as shown in figure 3. In orthopedic surgery various fixture plates stretching from greater trochanter to lesser trochanter were used. All the screws and plates were made according to the area of application taking into account various cross sectional geometry of the femur bone.

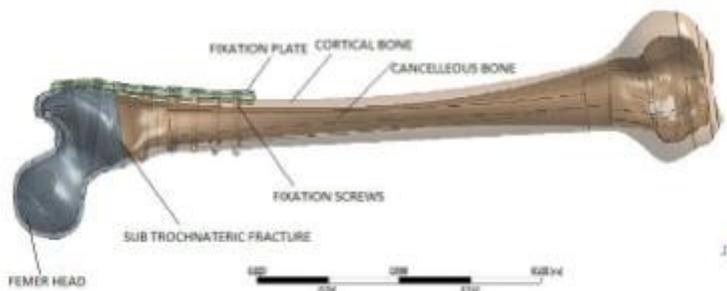


Figure2: threedimensional assemblies offemurbone, plate and screw

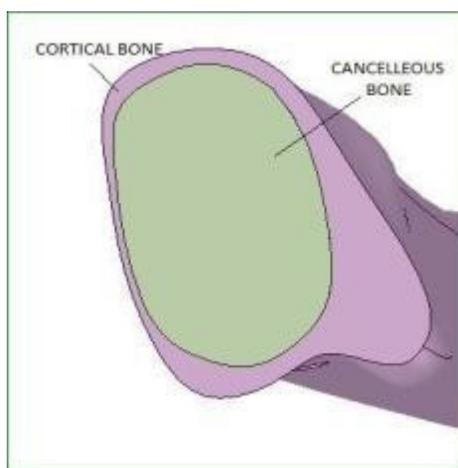


Figure3: cross SectionalAreaof FemurBone

2.3 Discretization/ Mesh Generation:

Meshing is the process of dividing the given domain into small number of nodes and elements. In this study conformal mesh is used, results were compared with the convergence criteria, resulting mesh model consists of tetrahedral elements with number of elements = 306714 and number of nodes= 412713 as

shown in the following figure 4. The refinement mesh is preferred along the fixation plate and femur head region so as to distribute uniform intensity of compressive load. The mesh statistics been evaluated as with numerous iterations.



Figure4: mesh design offemur, plates and screws.

2.4 Boundary Conditions:

Imposing the suitable boundary condition to the given study, the parameters such as deformation, equivalent stress i.e. Von misses stress on the fractured femur bone can be computed. The various biomaterials as shown in following Table.1 are used in the solver, the results were compared with different configuration of fixation plates and screws arrangements as seen in figure 5. In the present study the geometry is constrained in lateral condyle region and the static loading is performed under the linear elastic control with 250N, 500N, 750N, 1000N, 1250N and 1500N. All the measures considered so as protecting the radial nerves, currently orthopedic surgeons preferred titanium alloy during the orthopedic surgery. One of the advantages of titanium plate is that there is no need to remove them, because of their tolerance, thus avoiding a second surgical procedure.

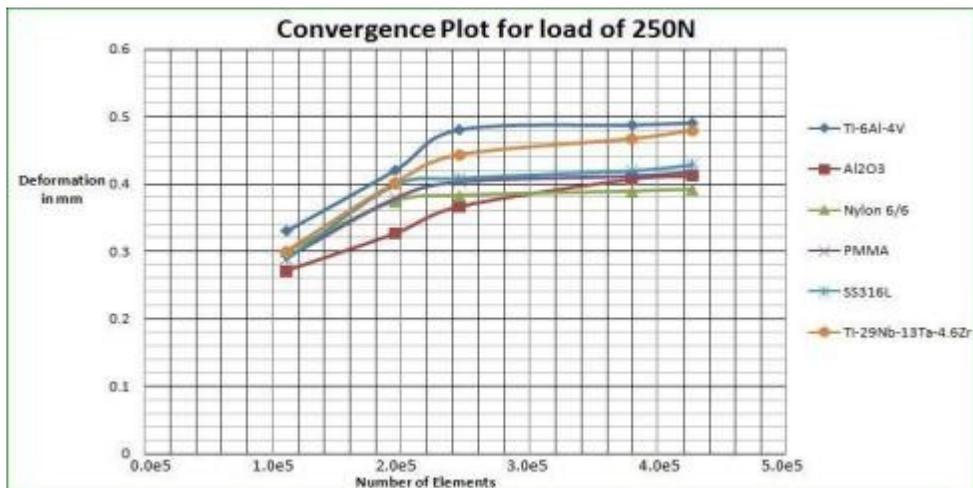


Figure 5: convergence plot for a load of 250N.

3. Results:

According to the concept of structural mechanics, at equilibrium condition the forces need to be balanced i.e. the reactions around lateral condyle region must be balanced with the applied compressive load. The results were assessed based on the convergence criteria. In general practice orthopedic surgeons use titanium alloy i.e. (Ti-6Al-4V) due to its strength, load sustainability, corrosion resistance and cost effective. Apart from the recent biomaterials focus should be drawn on composite materials listed in table 2 and table 3, which will ensure less equivalent stress and deformation respectively. Tables 4, 5, 6, 7, 8, 9 give us a graphical representation of equivalent stress on the bone at different loading conditions for different bio-materials. Similarly fig 7 and fig 8 would give the graphical representation of the same.

Table2: equivalent Stress on various bio-materials at different loading conditions

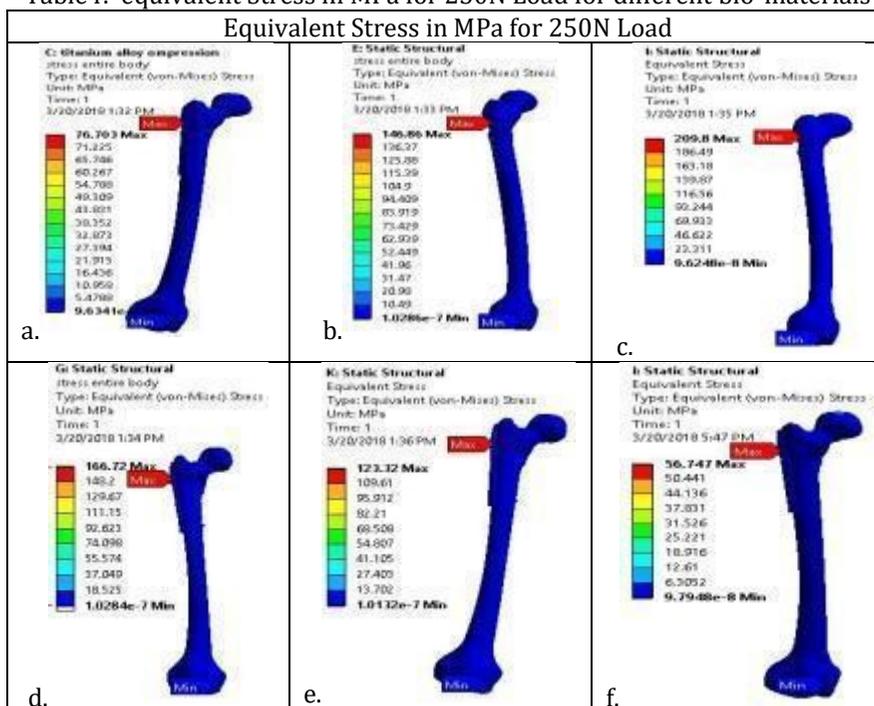
MATERIALS	LOAD(N)					
	250	500	750	1000	1250	1500
Ti-6Al-4V	76.703	153.41	230.11	306.81	383.52	460.22
AluminaAl2O3	146.86	293.72	440.57	587.43	734.29	881.15
Nylon 6/6	209.8	419.6	629.4	839.2	1049	1258.8

PMMA	166.72	333.44	500.16	666.89	833.61	1000.3
SS316L	123.32	246.63	369.95	493.26	616.57	739.89
Ti-29Nb-13Ta-4.6Zr	56.747	113.49	170.24	226.99	283.73	340.48

Table3: deformation of different bio-materials at different loading conditions

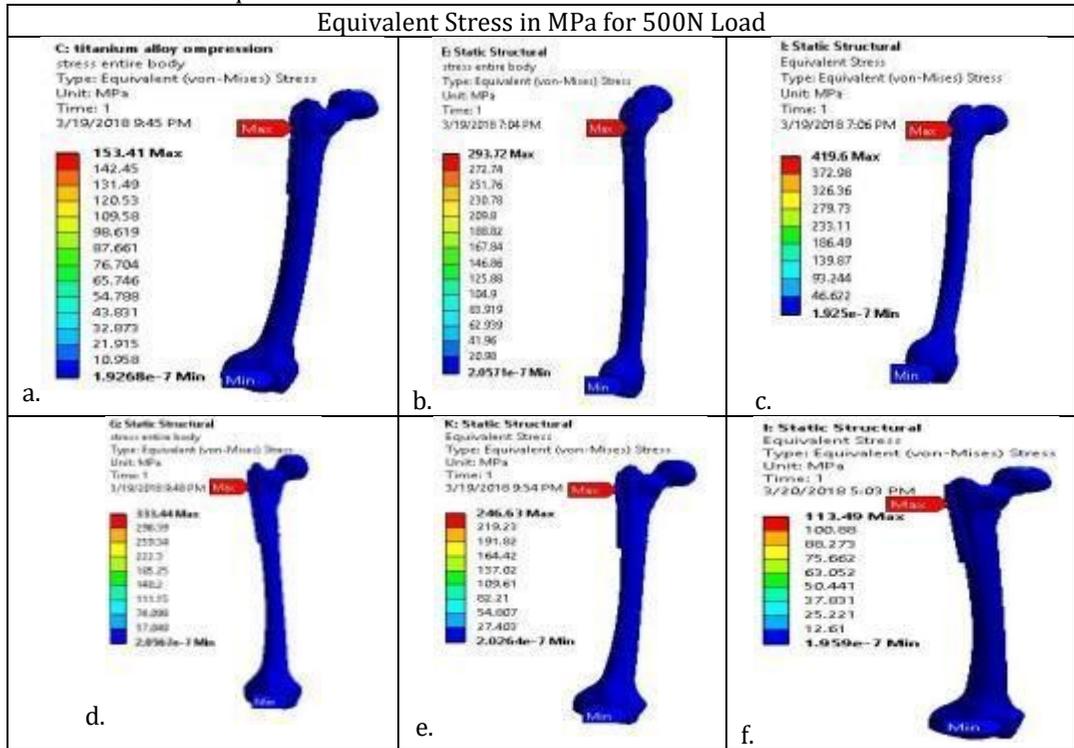
MATERIALS	DEFORMATION (mm)					
	250	500	750	1000	1250	1500
Ti-6Al-4V	0.49197	0.92395	1.3859	1.8479	2.0399	2.7718
AluminaAl2O3	0.41189	0.82372	1.2357	1.6476	2.0594	2.55
Nylon 6/6	0.39178	0.78356	1.1753	1.5671	1.9589	2.498
PMMA	0.41783	0.83567	1.2535	1.6713	2.0892	2.507
SS316L	0.42878	0.85755	1.2863	1.7151	2.1439	2.5727
Ti-29Nb-13Ta-4.6Zr	0.47925	0.95849	1.4377	1.917	2.3962	2.687

Table4: equivalent Stress in MPa for 250N Load for different bio-materials



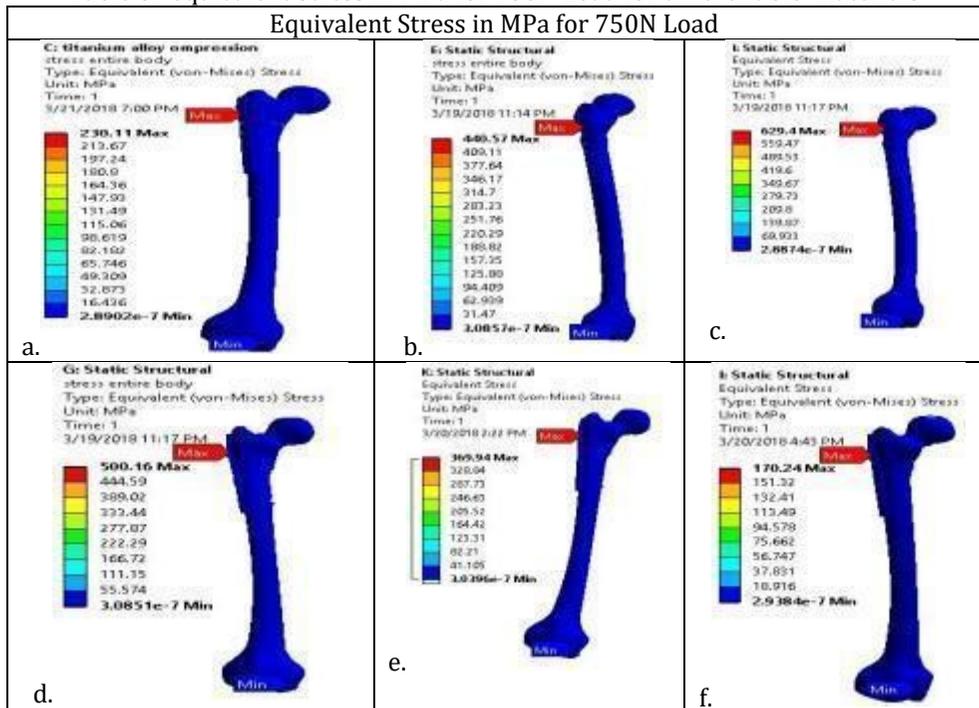
a: Ti-6Al-4V b. Alumina Al2O3 c. Nylon 6/6 d. PMMA e. SS316L f. Ti-29Nb-13Ta-4.6Zr

Table5: equivalent Stress in MPa for 500N Load for different bio-materials



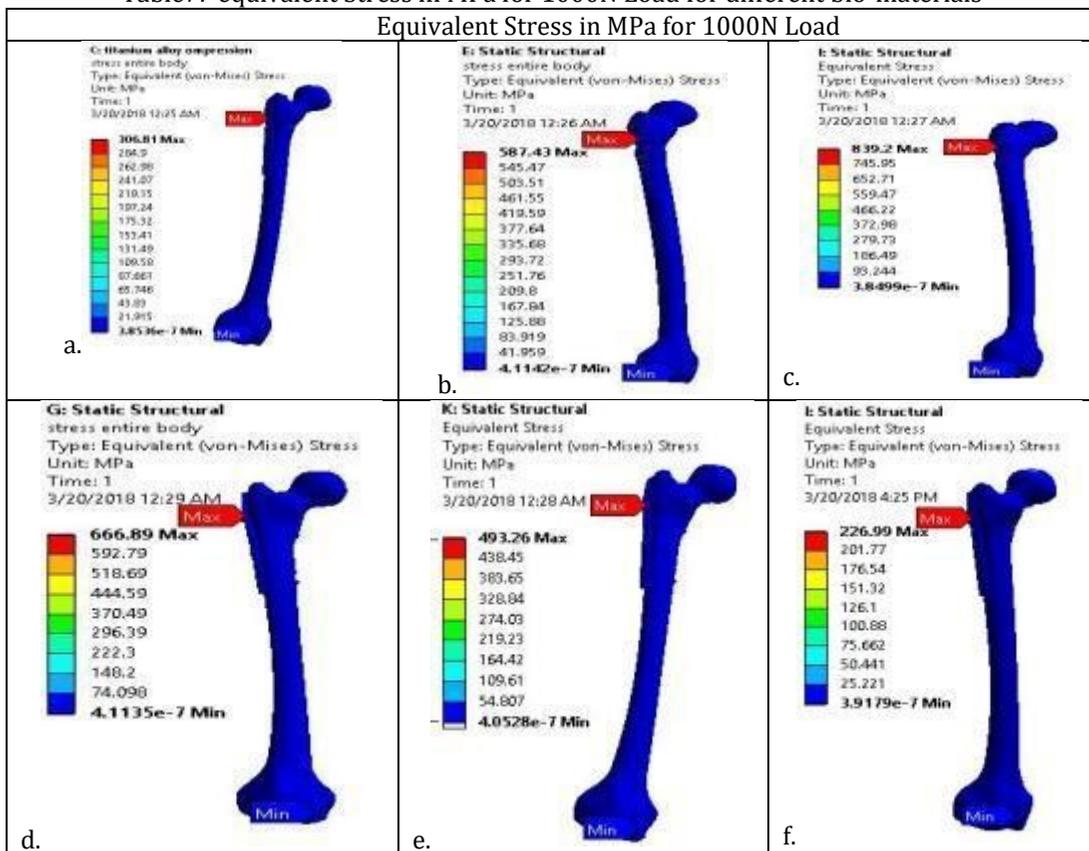
a: Ti-6Al-4V b. Alumina Al₂O₃ c. Nylon 6/6 d. PMMA e. SS316L f. Ti-29Nb-13Ta-4.6Zr

Table 6: equivalent Stress in MPa for 750N Load for different bio-materials



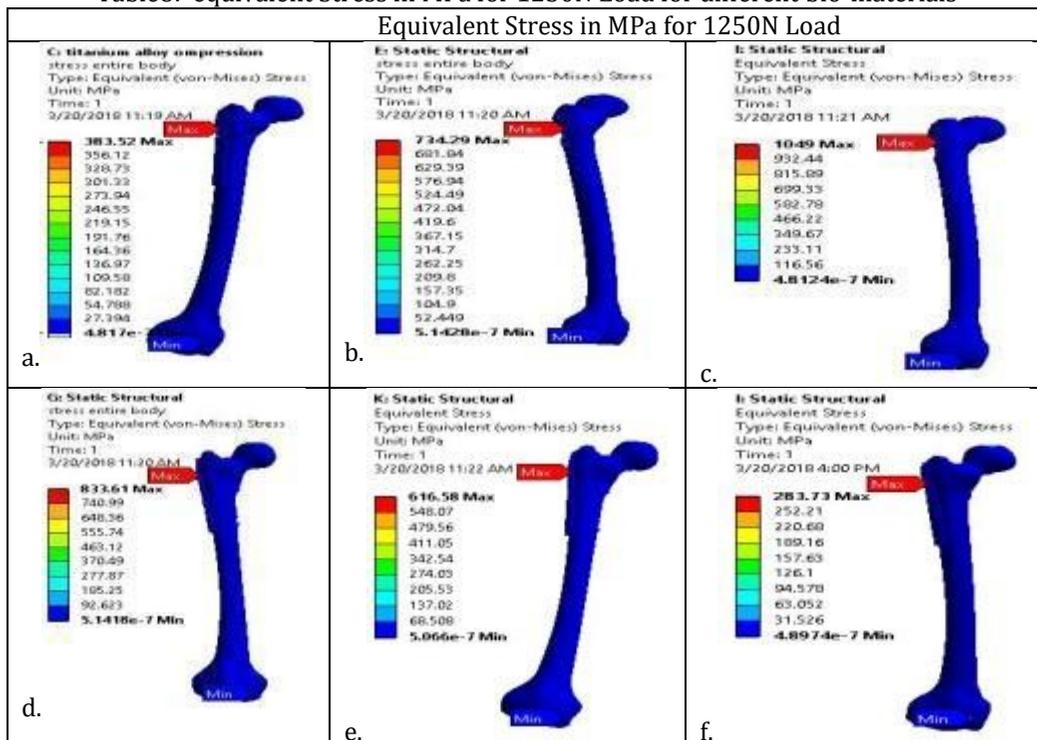
a: Ti-6Al-4V b. Alumina Al₂O₃ c. Nylon 6/6 d. PMMA e. SS316L f. Ti-29Nb-13Ta-4.6Zr

Table7: equivalent Stress in MPa for 1000N Load for different bio-materials



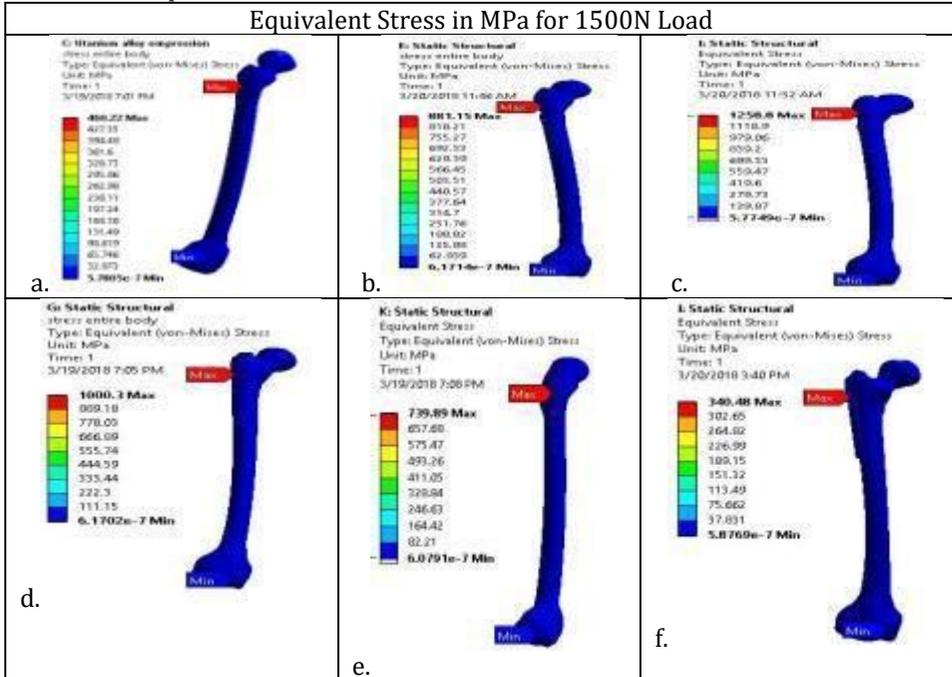
a: Ti-6Al-4V b. Alumina Al2O3 c. Nylon 6/6 d. PMMA e. SS316L f. Ti-29Nb-13Ta-4.6Zr

Table8: equivalent Stress in MPa for 1250N Load for different bio-materials



a: Ti-6Al-4V b. Alumina Al2O3 c. Nylon 6/6 d. PMMA e. SS316L f. Ti-29Nb-13Ta-4.6Zr

Table9: equivalent Stress in MPa for 1500N Load for different bio-materials



a: Ti-6Al-4V b. Alumina Al₂O₃ c. Nylon 6/6 d. PMMA e. SS316L f. Ti-29Nb-13Ta-4.6Zr

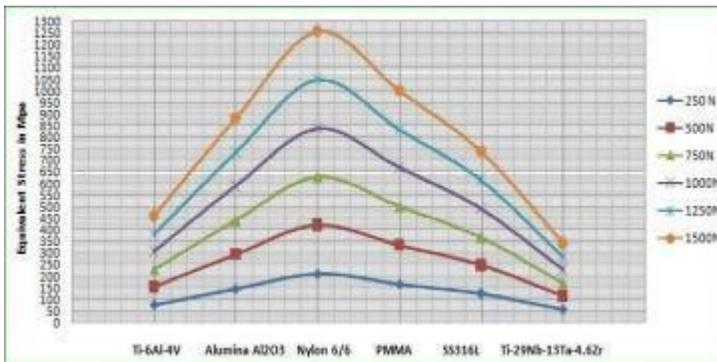


Figure7:graph of Equivalent Stress on various bio-materials at different loading conditions

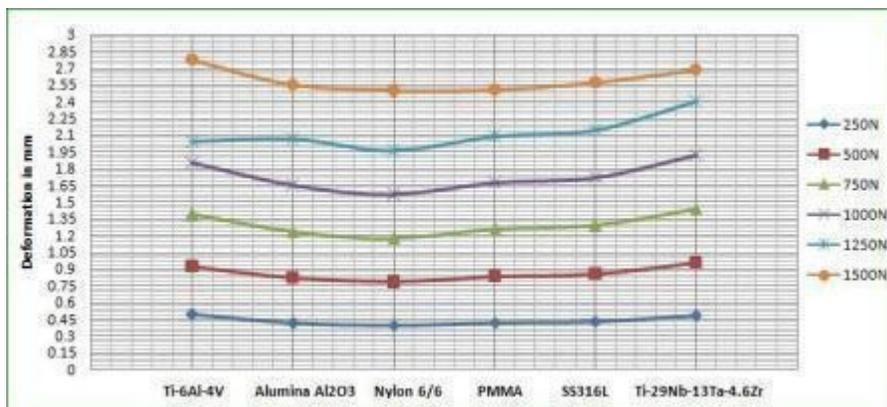


Figure8: deformation of different bio-materials at different loading conditions

4. Discussion:

In this research study, the geometry of fractured femur bone with fixation plates and screws is modeled and discretized successfully. Finite element method is used to investigate the condition under static loading

criteria. The aim of this study is to propose the substitute biomaterial for fixation plates and screws which can be used as alternative based on the convergence study, we can conclude that titanium alloy i.e. (Ti-29Nb-13Ta-4.6Zr) found to alternative material. This material can prove to be a great boon in orthopedic surgeries in the near future in terms of reducing the cost of performance and improving the human functionality.

References:

1. Prevention and Management of Osteoporosis, report of WHO scientific group. World Health Organization, Geneva 2003. ISSN 0512-3054.
2. Copper C, Campion G, Melton LJ 3rd Hip fracture in the elderly: a worldwide projection's. *Osteoporos Int.* 1992 Nov; 2(6):285-9. (PMID: 1421796)
3. Replacement arthroplasty versus internal fixation for extra capsular hip fractures in adults (Review) Parker MJ, Handoll HHG. *Cochrane Database Syst Rev.* 2006 Apr 19;(2):CD000086 (PMID:16625528).
4. A. Naderi-pour. Orthopedic One, The orthopedic Knowledge Expert. <https://www.orthopaedicsone.com/display/Review/Femur/> accessed 2010 April 05.
5. Ajay Dhanopia and Dr. Manish Bhargava, "Finite Element Analysis of Human Fractured Femur Bone Implantation with PMMA Thermoplastic Prosthetic Plate", 11th International Symposium on Plasticity and Impact Mechanics, Implast 2016. *Procedia Engineering* 173 (2017) 1685- 1665.
6. Reham B. Osman and Michael V. Swain. A critical Review of Dental Implant Materials with an Emphasis on Titanium versus Zirconia. *Materials (Basel).* 2015 Mar; 8(3): 932-958.
7. Garth Hastings, Jonathan Black, William Murphy. *Handbook of Biomaterial Properties.* Business Media -2nd edition, New York, cop.2016.
8. Seinsheimer, F. "Subtrochanteric fractures of the femur." *The Journal of Bone and Joint Surgery-American* Volume 60.3 (1978), 300-306.

Analysis and Prediction of Crime Against Women

¹Asmi Patel, ²Dhwani Shah, ³Namrata Poojary, ⁴Priya Mishra, ⁵Alvina Alphonso

^{1,2,3,4}Student, ⁵Assistant Professor

Information Technology, St. Francis Institute of Technology, Mumbai, India

ABSTRACT: The main purpose of this project is to analyze the crimes occurring district wise and predict the further crime trends using the past twelve years of crime records. The data is released by the National Crime Records Bureau (NCRB) and is made available on the Open Government Data Platform India (OGD). This data can be used to gain various insights about the different crime trends occurring in the country. The types of crimes in the dataset are molestation, rapes, dowry, exploitation, kidnapping, cruelty by relatives etc. K-means clustering data mining approach will be done to obtain crime trends and linear regression will be done to predict crime rates.

Keywords: K-means, Linear Regression, Types of crimes, Crime Analysis, Prediction, Visualization, scatter plot, correlation matrix

I. Introduction

India is a vast country with diverse societies. Position of women has been of great importance since ancient times in Indian culture. Unfortunately current scenario depicts a different story. According to National Crime Records Bureau, crime against women has significantly increased in recent years such as Dowry, Kidnapping, Insult to modesty, Rape, Assault, Importation of girls, Cruelty by relatives etc.

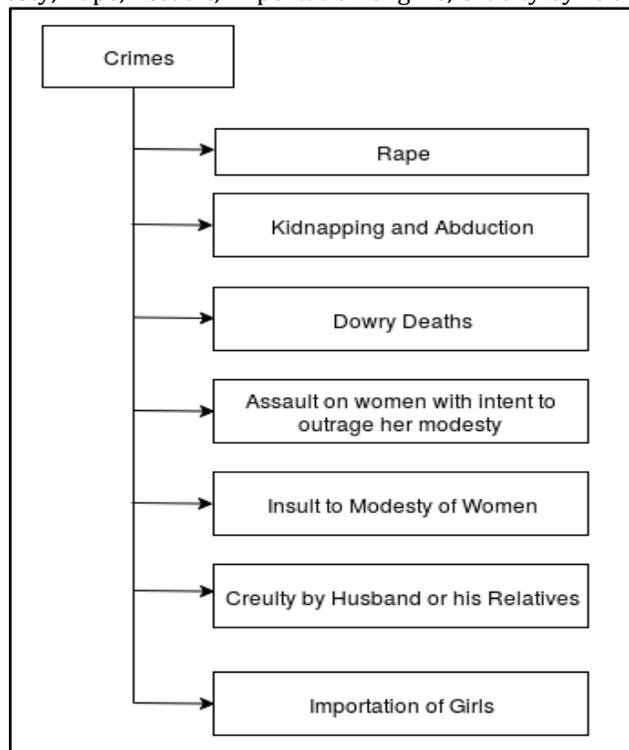


fig 1. crime hierarchy

It has become the topmost priority to the administration to enforce law and order to reduce this increasing rate of the crime against women. In country like India crimes against women has been always a serious issue leading to various legal norms and measures against it. With each passing year crime reports are generated leading to huge amount of data. Such data can be used to generate analysis and statistics which will help government and non-government organizations to initiate certain schemes and policies

accordingly. In 2012, Uttar Pradesh reported the highest cognisable crime rate of 455.8 among States of India, while Nagaland recorded lowest rates 47.7. The rates were calculated by National Crime Records Bureau as the number of incidents per 100,000 of the population. The crime rate differ in each and every state, and hence the measures to be taken also has to be different. This paper primarily focuses on analysis of crime against women which will help in order to generate predictions as an outcome this may help to generate valuable information from the existing data. The initial phase includes data acquisition provided by the government and processing it further for analysis and prediction.

II. Literature Review

Priyanka das et al. has proposed InfoMap algorithm which is used to detect communities, that displays changing trends of various crimes in the different states in India. This paper proposes a graph based approach to find crime similarity between states and to predict the crime patterns of the states. Graph based clustering is used to create a weighted graph and Cosine Similarity is used to calculate the crime similarities among the states. The prediction of the overall crime trend is done by combining the graphs and applying InfoMap algorithm.

Anant Joshi et al. proposed crime analysis using k-means algorithm. This paper includes the process of correlation discovery and discovery of patterns among fields. In this paper, the Rapid Miner software tool is used to carry out cluster analysis. The purpose of this paper is to analyze the crime which consists of theft, homicide and various drug offences which also include suspicious activities, noise complaints and burglar alarm. K-means clustering data mining approach has been used on a crime dataset from New South Wales region of Australia to generate clusters of homogenous elements. Analysis of these clusters is done using k-means where identification of k is done using silhouette measuring.

Sunil Yadav et al. proposed k-means algorithm to create clusters according to high low values. Patterns, forecast trends, determining relationships, mapping crime trends and possible suspects are discovered with the help of Weka tool and R tool. Result of the k-means is an input to the Apriori algorithm. The association among a number of attributes is discovered with the help of Apriori algorithm. The association between the suspect arrested during the year and the person acquitted in the same year is shown in the result. The data acquisition and data staging were the two biggest hurdle in the project. Various measures can be carried out in order to reduce the crimes.

Shiju Sathyadevan et al. proposed Apriori algorithm which is used to identify the trends and patterns in crime. The algorithm determines association rules which results in highlighting general crime trends in the database. The crime pattern for a particular place is the output of this phase. This paper has also proposed the Naïve Bayes algorithm. This algorithm classifies new article into a crime type to which it fits the best. The prediction has been implemented using the decision tree concept. The crime prone areas are graphically represented using a heat map, which indicates level of activity, usually darker colors to indicate low activity and brighter colors to indicate high activity.

III. Methodology

Modules that are included in the project are:

1. Data Pre-processing
2. Clustering
3. Linear Regression
4. Data visualization

K-means clustering technique is used for crime analysis since it is suitable for clustering large data sets, due to computational complexity. The clustered results makes it easy to identify crime trend over years and it can be used to design precaution methods for future.

Following is the System flow diagram:

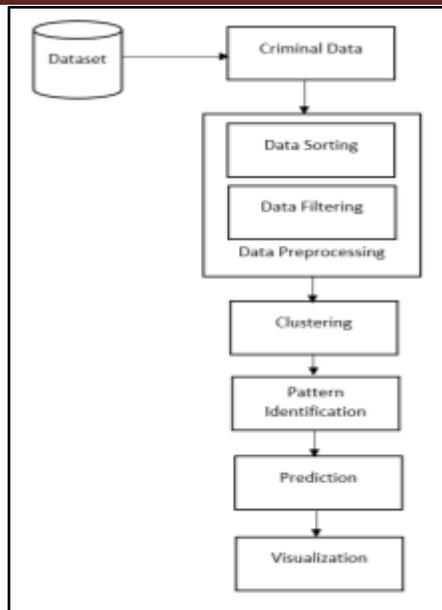


fig 2. system block diagram

DATA PRE-PROCESSING: Data Pre-processing is done in order to pre-process the data which consist of incomplete attributes and null values. The dataset is acquired from the government website www.data.gov.in. The data is provided by the National Crime Records Bureau . The dataset contains crime information of all 27 states and union territories from the year 2001-2012. Crimes such as Dowry, Kidnapping, Insult to modesty, Rape, Assault, Importation of girls, Cruelty by husband or his relatives are included in the dataset. In Order to manage data successfully missing values must be either removed or filled with the mean value. In Machine Learning models are based on Mathematical equation hence every categorical value must be transformed to numeric value. Further, the dataset is segregated into training and testing set into 80:20 ratio.

K-MEANS CLUSTERING: For clustering, k-means clustering algorithm is used which is unsupervised technique. The K-means clustering algorithm uses iterative refinement to produce a final result. The algorithm inputs are the number of clusters and the data set. The dataset is a collection of features for each data point. The dataset consist of States and union territories along with its districts, seven different crimes such as rape, dowry deaths, assault, insult to modesty, importation of girls, kidnapping, Cruelty by husband and his relatives. The algorithms starts with initial estimates for the centroids, which can either be randomly generated or randomly selected from the data set. So here the number of clusters, $k=3$. Three different clusters are formed based on the number of crimes in that particular place. The districts with highest number of crime rate will be grouped together in a cluster, similarly the one with lowest and average number of crimes. The algorithm then iterates between two steps:

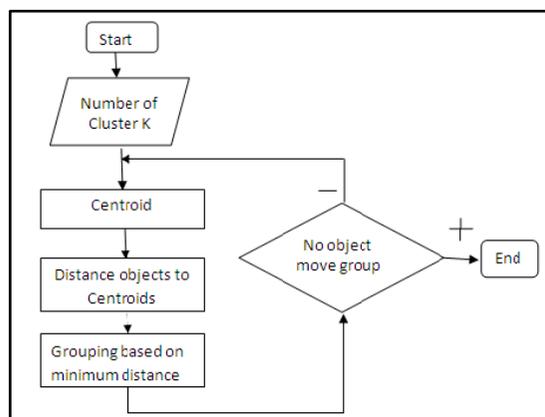


fig 3. k-means algorithm [8]

(a) Data assignment step: Each centroid defines one of the clusters. In this step, each data point is assigned to its nearest centroid, based on the squared Euclidean distance. Since the data is present in the form of 1d array the y coordinate becomes zero and the distance can be calculated by just finding the difference between the centroids and the dataset values.

(b) Centroid update step: In this step, the centroids are recomputed. This is done by taking the mean of all data points assigned to that centroid's cluster. The algorithm iterates between steps one and two until a stopping criteria is met (i.e., no data points change clusters, the sum of the distances is minimized, or some maximum number of iterations is reached). Here the maximum number of iteration is kept one. This algorithm is guaranteed to converge to a result. The result may be a local optimum (i.e. not necessarily the best possible outcome), meaning that assessing more than one run of the algorithm with randomized starting centroids may give a better outcome.

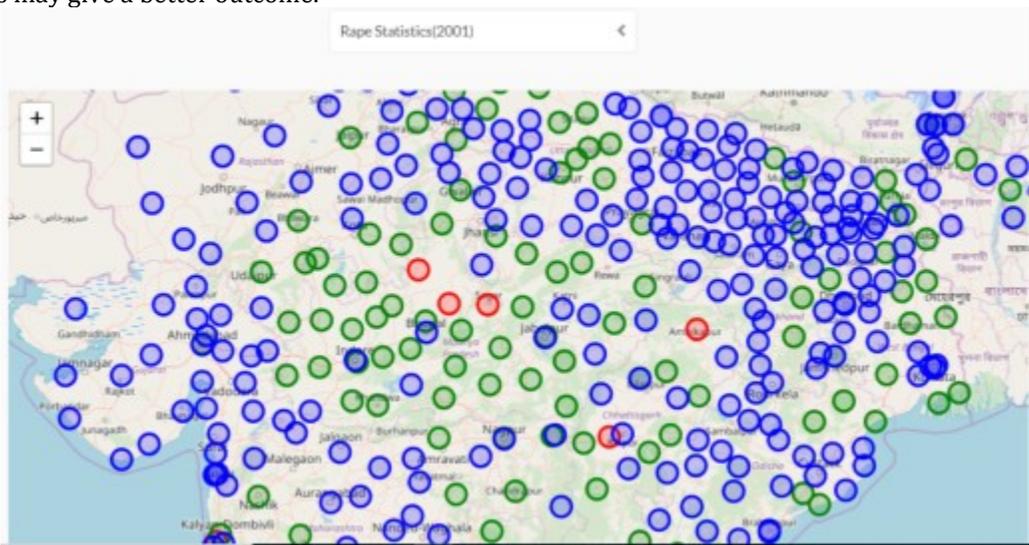


fig 4. k-means clusters

LINEAR REGRESSION: In order to forecast the crime rate for future years, linear regression technique is being used. This technique consists of a dependent and an independent variable. The linear regression line has an equation of the form $y = mX + c$, where m is the slope of the line, c is the coefficient of the line, X is the independent variable and y is the dependent variable. Here, the independent variable (X) is Year and the dependent variable (Y) will be the rate of specific crime from the dataset. The core idea is obtaining a line that best fits the data which can be used to predict any new feature value. The best fit line is the one for which total prediction error (all data points) are as small as possible. This best fitting line is called the regression line.

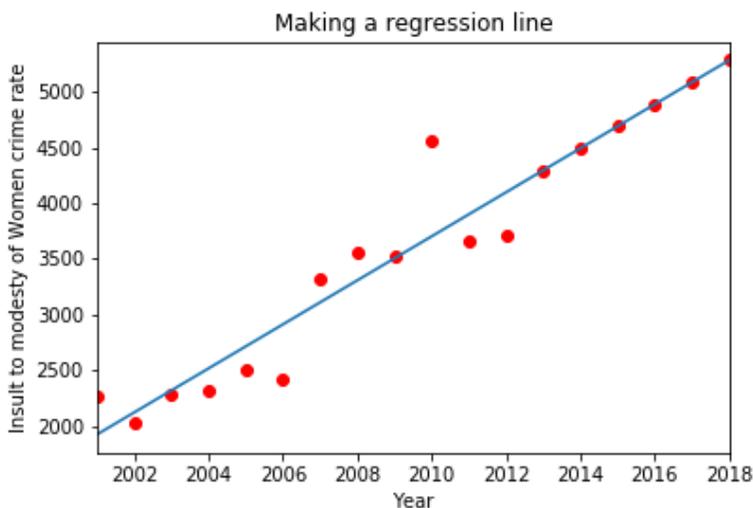


fig 5. linear regression

DATA VISUALIZATION:Data visualization is essential for representing insights from data in a graphical manner. With the large amount of data in dataset, one of the greatest challenge is to easily communicate the hidden patterns and findings in an easy and understandable manner. To visualise the data, there are many visualisation techniques available. Some of techniques that we are utilizing for the project are line chart, bar chart, correlation matrix, scatterplot.

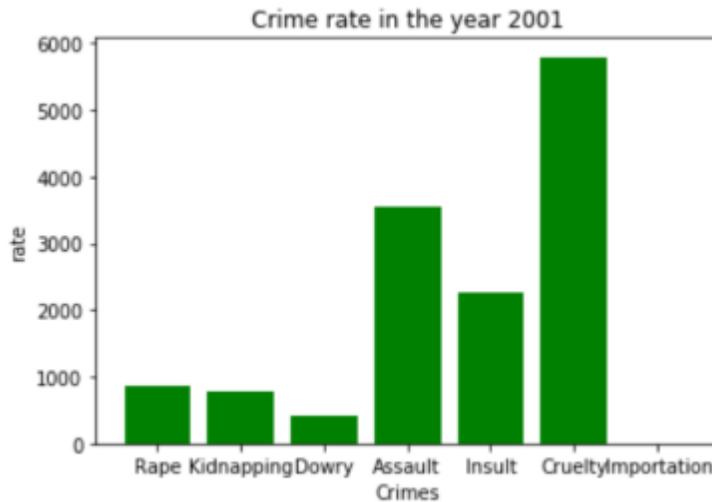


fig6. data visualization using bar chart

	Rape	Kidnapping and Abduction	Dowry Deaths	Assault on women with intent to outrage her modesty	Insult to modesty of Women	Cruelty by Husband or his Relatives	Importation of Girls
Rape	1.000000	0.548460	0.373808	0.650259	0.160520	0.610098	0.039514
Kidnapping and Abduction	0.548460	1.000000	0.486531	0.327858	0.158379	0.599610	0.011755
Dowry Deaths	0.373808	0.486531	1.000000	0.260285	0.201940	0.345819	0.101380
Assault on women with intent to outrage her modesty	0.650259	0.327858	0.260285	1.000000	0.330311	0.480329	-0.004923
Insult to modesty of Women	0.160520	0.158379	0.201940	0.330311	1.000000	0.269082	-0.005557
Cruelty by Husband or his Relatives	0.610098	0.599610	0.345819	0.480329	0.269082	1.000000	0.017730
Importation of Girls	0.039514	0.011755	0.101380	-0.004923	-0.005557	0.017730	1.000000

fig 7. correlation matrix

IV. Conclusion

Patterns related to the crimes in specific regions can be identified which can help concerned authorities to take preventive measures. Mapping by area in relation to the type of violence helps to prepare better strategies to prevent specific violence. The trends observed will help in better decision making and reducing the future crime against women. Analysis and prediction of crime by clustering and regression will help to identify the regions where specific crimes are more frequent as well as the crime rate, which in turn can be used by concerned authorities to focus on those areas.

V. Future Work

The proposed project mainly focuses on emphasizing the crime rates around a particular country (India). So as per the future vision other part of the world can be taken into consideration. Right now it is specifically bound to a single gender (Women). Other genders can be included further. There are certain crimes and cases which are unheard and unregistered around the globe and if they are taken into consideration the accuracy of the crime rates can be improved. Currently the dataset consist of only seven crimes but this can expanded to include many more crimes in the future.

VI. Reference

1. Priyanka Das, Asit Kumar Das. 2017. Behavioral Analysis of Crime against Women using a Graph Based Clustering Approach. IEEE International Conference on Computer Communication and Informatics (ICCCI): 1-5.
2. Anant Joshi, A. Sai Sabitha, Tanupriya Choudhury. 2017. Crime Analysis using k-means Clustering. Third International Conference on Computational Intelligence and Networks (CINE): 33-39
3. Sunil Yadav, Meet Timbadia, Ajit Yadav, Rohit Vishwakarma, Nikhilesh Yadav. 2017. Crime Pattern Detection, Analysis & Prediction. IEEE International Conference on Electronics, Communication and Aerospace Technology (ICECA): 225-230.
4. S. Sathyadevan, D. M. S, and S. G. S. 2014. Crime analysis and prediction using data mining. First International Conference on Networks Soft Computing (ICNSC): 406-412.
5. U. Thongsatapornwatana. 2016. A survey of data mining techniques for analyzing crime patterns. Second Asian Conference on Defence Technology (ACDT): 123-128.
6. Chhaya Chauhan, Smriti Sehgal. 2017. A review: Crime Analysis using data mining techniques and algorithms. International Conference on Computing, Communication and Automation (ICCCA): 21-25.
7. Javed Anjum Sheikh, Iqra Shafique, Madiha Sharif, Syeda Ambreen Zahra, Tuba Farid. 2017. IST: Role of GIS in Crime Mapping and Analysis. International Conference on Communication Technologies (ComTech): 126-131.
8. X. Yang, Y. Wang, D. Wu, A. Ma. 2010. K-Means based clustering on mobile usage for social network analysis purpose. Sixth International Conference on Advanced Information Management and Service (IMS): 223-228
9. S. Sivaranjani, Dr. S. Sivakumari, Aasha. M. 2016. Crime Prediction and Forecasting in Tamilnadu using Clustering Approaches. International Conference on Emerging Technological Trends (ICETT): 1-6.
10. Zhang Ying. 2016. Analysis of Crime Factors Correlation based on data mining techniques. International Conference on Robots & Intelligent System (ICRIS).
11. A. Bansal. 2015. Performance Comparison of Data Mining Techniques to Analyse Crime against Women. International Journal Of Scientific Research And Education (IJSARE): 4494-4512.
12. A. Babakura, Md N. Sulaiman and M. A. Yusuf. 2014. Improved Method of Classification Algorithms for Crime Prediction. International Symposium on Biometrics and Security Technologies (ISBAST): 250-255.
13. Chung-Hsien Yu, Max W. Ward, Melissa Morabito, and Wei Ding. 2017. Crime Forecasting Using Data Mining Techniques. 11th IEEE International Conference on Data Mining Workshops: 779-786.
14. Aman Kumar, Nikhil Tiwari, Prakhar Gupta, Dr. S. N. Rajan. 2017. Women Crime Prediction. International Research Journal of Engineering and Technology (IRJET).
15. National Crime Bureau Record. [Online]. Available: <https://data.gov.in>

PREVALENCE OF ANEMIA AMONG WOMEN IN MAHARASHTRA

A Statistical Analysis

¹Siddhi Pawar, ²Ruchira Patil, ³Prof. Poonam Soni, ⁴Dr. Mansha Kotwani
^{1,2}Student, Department of Biotechnology, ^{3,4}Assistant Professor,
Department of Mathematics, Thadomal Shahani Engineering College, Bandra, India.

ABSTRACT: This study has been undertaken to examine the prevalence of anemia among women in Maharashtra, India and the influence of various independent factors. The data was taken from the National Family Health Survey, India (NFHS)-4 for year 2015-16, conducted by International Institute of Population Sciences (IIPS). Data analysis was done using a set of statistical tools such as frequency tables and percentage. Additionally, the Chi-Square technique was used to test whether there is a relation or not between anemia and various independent factors such as age, marital status and maternity status and urban/rural location. The measure of association was checked using Cramer's V and Contingency coefficient. The results showed us that 48% of women in Maharashtra suffer from some type of anemia (Hb<12.0g/dL). Furthermore, there was a distinct relation between anemia and age, marital status, maternity status and urban/rural location. Also, anemia may be caused by other factors such as hypertension, diet intake, iron deficiency, vitamin B12 and vitamin C deficiency, folic acid deficiency and more, but they are not considered in the scope of this study. Further research needs to be done to check the influence of these factors on anemia.

Keywords: Anemia, Women, Maharashtra, Chi-Square, Contingency Coefficient, Cramer's V

I. INTRODUCTION

Anemia is a condition in which the number of red blood cells and their oxygen carrying capacity is insufficient to meet the physiological needs, which vary by age, sex, pregnancy status, smoking, etc. Anemia develops when the blood lacks enough healthy blood cells or hemoglobin. The World Health Organization (WHO) defines anemia as hemoglobin concentration below 12g/dL in women and 13g/dL in men. Relevance of anemia in South Asia is among the highest in the world, mirroring overall high rates of malnutrition. (WHO, 2011)

Iron deficiency is thought to be the most common cause of anemia globally, although other conditions such as micronutrient deficiency (folate, Vitamin B12 and Vitamin A), chronic inflammation, parasitic infection (malaria, cancer, tuberculosis and HIV), inherited disorders (e.g. haemoglobinopathies) can all cause anemia. (WHO, 2011)

Women are at risk of iron deficiency due to blood loss from menstruation especially during the reproductive years (Dutta, 2005). Iron deficiency anemia is detrimental to public health and results are retarded infant development, increased maternal and fetal mortality and morbidity, and reduced work performance (Park, 2007). In the world (excluding China), (DeMayer and Adiels, 1985) women among 15-49 years, 288.4 million (35% of non-pregnant and 43.9 million [51%] of pregnant) women are anemic. In India, 20-40% of maternal deaths are due to anemia (Royston, 1989). More than 50% of women do not have adequate iron stores for pregnancy (Perry et al, 1995).

The consequences of anemia for women increased risk of low birthweight or prematurity, prenatal and neonatal mortality, inadequate iron stores for the newborn, increased risk of maternal morbidity and mortality, and lowered physical activity, mental concentration, and productivity. (Gillespie & Johnston, 1998; Stoltzfus, 1997; Allen, 1997). Women with even mild anemia may experience fatigue and have reduced work capacity (Gillespie & Johnston, 1998).

In Maharashtra, recent survey done by International Institute of Population Sciences (IIPS) called as National Family Health Survey (NFHS-4) in the year 2015-16 reflects the magnitude of problem faced by women in the state. This research is aimed as investigating the prevalence of anemia and the factors on which it is dependent among the women in Maharashtra.

II. STATISTICAL ANALYSIS OF THE PROBLEM

Data

The data was taken from National Family Health Survey 2015-16 (IIPS and ICF, 2018) for the state of

Maharashtra which was the fourth in its series. It provides information on population, health, and nutrition for India and each state and union territory. The survey was funded by the United States Agency for International Development (USAID), the United Kingdom Department for International Development (DFID), the Bill and Melinda Gates Foundation (BMGF), UNICEF, UNFPA, the MacArthur Foundation, and the Government of India. Technical assistance for NFHS-4 was provided by ICF, Maryland, USA.

A total sample size of approximately 572,000 households was taken which was based on the size needed to produce reliable indicator estimates for each district and for urban and rural areas in districts in which the urban population accounted for 30-70 percent of the total district population. NFHS-4 fieldwork for Maharashtra was conducted in all 35 districts of the state from 1 April to 25 September

2015 by GFK Mode Private Limited. Information was collected from 26,890 households, 29,460 women age 15-49. Survey response rates were 96 percent for households and 94 percent for women. (IIPS and ICF, 2018)

Data analysis

The analysis of the data was done by using statistical tools such as frequency and percentage calculation to group the data. Primary outcome variables were created for hemoglobin measurements. The women were classified as mildly, moderately and severely anemic or normal based on the survey data. A hemoglobin concentration of less than 7 g/dl (gram per deciliter) was used to define severe anemia, 7.0–9.9 g/dl for moderate anemia, and 10.0–10.9g/dl to correspond to mild anemia in pregnant women and 10.0–11.9 g/dl for non-pregnant women. Women with hemoglobin levels greater than 12.0 g/dl were considered to have normal level/be non-anemic. Further they were classified based on independent variable like age (15-49), marital status, maternity status and urban/rural location. Table

1(a) gives the breakdown of women according to their age and anemic status, table 1(b) gives the breakdown of women according to their marital status and anemic status, table 1(c) gives the breakdown of women according to their maternity status and anemic status and table 1(d) gives the breakdown of women according to their location (Urban/Rural) and anemic status.

Table 1: frequency distribution and percentages of respondents (a) according to Age, (b) according to Marital status, (c) according to Maternity status and (d) according to Location (urban/rural)

(a)

Age Group in years	Mild (Percentage)	Moderate (Percentage)	Normal (Percentage)	Severe (Percentage)	Total
15-19	1753 (39.5)	426 (9.6)	2232 (50.3)	27 (0.6)	4438
20-29	3638 (37.1)	1010 (10.3)	5090 (51.9)	69 (0.7)	9807
30-39	2870 (36.6)	855 (10.9)	4061 (51.8)	55 (0.7)	7841
40-49	2255 (35.3)	639 (10)	3429 (53.7)	64 (1)	6387
Total	10516	2930	14814	215	28473

(b)

Marital Status	Mild (Percentage)	Moderate (Percentage)	Normal (Percentage)	Severe (Percentage)	Total
Not Married	2313 (38.6)	593 (9.9)	3038 (50.7)	48 (0.8)	5992
Currently Married	7567 (36.3)	2126 (10.2)	11008 (52.8)	146 (0.7)	20847
Widowed/Divorced/ Separated/Deserted	620 (37.9)	209 (12.8)	790 (48.3)	16 (1)	1635
Total	10500	2928	14836	210	28474

(c)

Maternity Status	Mild (Percentage)	Moderate (Percentage)	Normal (Percentage)	Severe (Percentage)	Total
Pregnancy	270 (26.2)	231 (22.5)	522 (51)	6 (0.6)	1029
Breastfeeding	1386 (41.2)	293 (8.7)	1667 (49.9)	17 (0.5)	3363
Neither	8862 (36.8)	2408 (10)	12619 (52.4)	193 (0.8)	24082
Total	10578	2932	14808	216	28474

(d)

Location	Mild (Percentage)	Moderate (Percentage)	Normal (Percentage)	Severe (Percentage)	Total
Urban	5253 (37.7)	1394 (10.0)	7218 (51.8)	70 (0.5)	13935
Rural	5263 (36.2)	1541 (10.6)	7590 (52.3)	145 (1)	14539
Total	10516	2935	14808	215	28474

Also, the data was evaluated using Chi-square test, Cramer's V test and contingency coefficient test analysis at 0.05 alpha levels. So, we have set the null and alternate hypothesis as follows:

H₀: There is no statistically significant association between anemia and independent variables among women in Maharashtra.

H_a: There is a statistically significant association between anemia and independent variables among women in Maharashtra.

In the first case of this hypothesis the independent variable is age, in the second case it is marital status, in the third case it is maternity status and in the fourth case it is location (urban/rural). Thus, we use Chi-square test of association to examine the effect of the variables on anemia. Set alpha to 0.05 and reject H₀ only if the calculated Chi-square value is greater than the value of Chi-square alpha and if p value < 0.05. Tables 1 and 2 displays observed and expected values and 100 percent of frequencies are above 5 in all the four cases.

Table 2: Expected values of (a) Age, (b) Marital status, (c) Maternity status and (d) Location (urban/rural)

(a)

Age Group in years	Mild	Moderate	Normal	Severe
15-19	1639.096	456.690	2308.701	33.511
20-29	3622.042	1009.184	5101.720	74.052
30-39	2895.934	806.874	4078.983	59.207
40-49	2358.925	657.251	3822.595	48.228

D.O.F = 9, $\chi^2 (\alpha) = 16.919$, χ^2 (calculated) = 31.369, Cramer's V = 0.019, Contingency Coefficient = 0.03

(b)

Marital Status	Mild	Moderate	Normal	Severe
Not Married	2209.594	616.161	3122.052	43.675
Currently Married	7687.486	2143.7160	10862.052	153.749
Widowed/Divorced	602.918	168.128	851.895	12.058

D.O.F = 6, $\chi^2 (\alpha) = 12.592$, χ^2 (calculated) = 28.992, Cramer's V = 0.018, Contingency Coefficient = 0.080

(c)

Maternity Status	Mild	Moderate	Normal	Severe
Pregnancy	380.101	105.957	535.135	7.805
Breastfeeding	1242.257	346.291	1748.939	25.311
Neither	8895.640	2479.750	12523.925	182.582

D.O.F = 6, $\chi^2 (\alpha) = 12.592$, χ^2 (calculated) = 215.046, Cramer's V = 0.061, Contingency Coefficient = 0.085

(d)

Location	Mild	Moderate	Normal	Severe
Urban	5146.465	1436.371	7246.943	105.220
Rural	5369.534	1498.629	7561.056	109.78

D.O.F = 3, $\chi^2 (\alpha) = 7.815$, χ^2 (calculated) = 80.082, Cramer's V = 0.032, Contingency Coefficient = 0.032

III. RESULT AND DISCUSSION

The data analysis showed us that 48% of women in Maharashtra suffer from some type of anemia. The survey was conducted on 28474 women. It was seen that nearly 36.9% women suffer from mild anemia,

10.3% women suffer from moderate anemia and 0.7% women suffer from severe anemia in the state of Maharashtra. (IIPS and ICF, 2018)

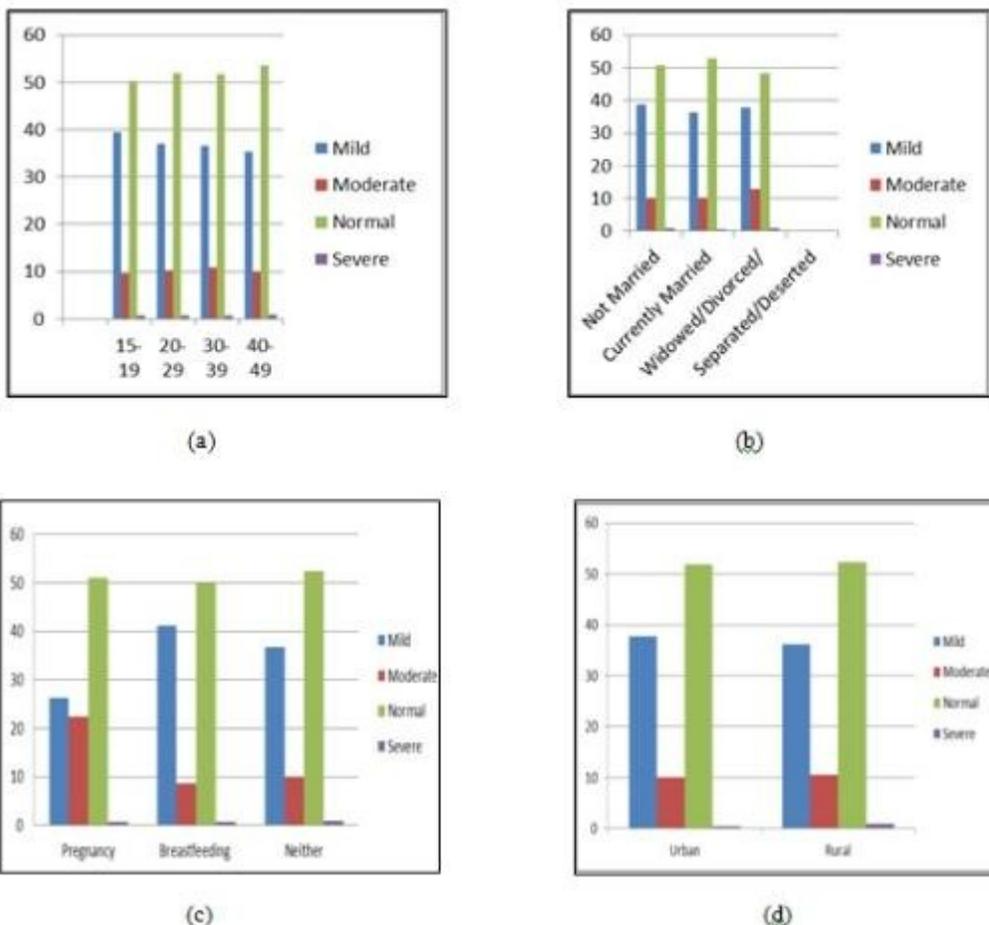


Figure 1: (a) age (b) marital status (c) maternity status and (d) location (urban/rural) as percentages of sample size for the four anemic categories mild, moderate, normal and severe.

Calculated Chi-square value of 31.369 was significant as p is 0.000257 ($p < 0.05$) at 9 Degrees of Freedom. Results of Chi-square test show that H_0 is rejected for independent variable age and that there may be a probable relationship between anemia and age of the women. This shows us that anemia is related to the age of women. Result of Cramer's V and contingency coefficient are 0.019 and 0.033 respectively. The measures of association show that correlation coefficient between age and anemia is 0.019 with a p value of 0.000257 ($p < 0.05$) for a sample size of 28473. This shows that the relation between anemia and age is weak but it may also be influenced by the large sample size.

Calculated Chi-square value of 28.992 was significant as p is 0.000061 ($p < 0.05$) at 6 Degrees of Freedom. Results of Chi-square test show that H_0 is rejected for independent variable marital status and that there may be a probable relationship between anemia and marital status of the women. This shows us that anemia is related to the marital status of women. Result of Cramer's V and contingency coefficient are 0.018 and 0.030 respectively. The measures of association show that correlation coefficient between marital status and anemia is 0.018 (Cramer's V) with a p value of 0.000061 ($p < 0.05$) for a sample size of 28474. This shows that the relation between anemia and marital status is weak but it may also be influenced by the large sample size.

Calculated Chi-square value of 215.046 was significant as p is 0.00001 ($p < 0.05$) at 6 Degrees of Freedom. Results of Chi-square test show that H_0 is rejected for independent variable maternity status and that there may be a probable relationship between anemia and maternity status of the women. This shows us that anemia is related to the maternity status of women. Result of Cramer's V and contingency coefficient are 0.061 and 0.085 respectively. The measures of association show that correlation coefficient between maternity status and anemia is 0.061 (Cramer's V) with a p value of -

0.00001 ($p < 0.05$) for a sample size of 28474. This shows that the relation between anemia and maternity status is weak but it may also be influenced by the large sample size.

Calculated Chi-square value of 80.082 was significant as p is 0.00001 ($p < 0.05$) at 3 Degrees of Freedom. Results of Chi-square test show that H_0 is rejected for independent variable location (urban/rural) and that there may be a probable relationship between anemia and location (urban/rural) of the women. This shows us that anemia is related to the location (urban/rural) of women. Result of Cramer's V and contingency coefficient are 0.032. The measures of association show that correlation coefficient between location and anemia is 0.032 (Cramer's V) with a p value of 0.00001 ($p < 0.05$) for a sample size of 28474. This shows that the relation between anemia and location (urban/rural) is weak but it may also be influenced by the large sample size.

IV. CONCLUSION

In the present study an attempt has been made to study the relation between anemia and independent variable age, marital status, maternity status and location (urban/rural). Data analysis revealed that there is an association between anemia and independent variable age, marital status, maternity status and location (urban/rural).

This study was population based using data from National Family Health Survey (NFHS-4) in the year 2015-16 done by IIPS by simple random sampling method, which ensures generalizability of results to the entire population. It is inferred from this study that sociodemographic, age, marital and obstetric factors were associated with anemia in reproductive age group. Hence, preventive measures improving the standard of living and imparting health education with emphasis on adequate dietary intake of iron should be undertaken. Ministry of Health and Family Welfare has revised the guidelines on IFA supplementation related to the National Nutritional anemia Prophylaxis programme. However, anemia may be caused by other factors such as hypertension, diet intake, iron deficiency, vitamin B12 and vitamin C deficiency, folic acid deficiency and more, but they are not considered in the present study. Further research and data analysis need to be done to check the influence of these factors on anemia.

V. REFERENCES

1. DeMayer E, Adiels-Tegman M. The prevalence of anaemia in the world. *World Health Stat Q.* 1985; 38:302-16.
2. Dutta DC. *Text Book of Gynaecology.* 4th ed. Calcutta: New Central Publication; 2005. p. 74-80, 175-86.
3. Gillespie, S. & Johnston, J.L. (1998): *The Prevalence of Anemia in the World*
4. International Institute of Population Sciences (IIPS) - National Family Health Survey (NFHS-4) Maharashtra State Factsheet http://rchiips.org/NFHS/pdf/NFHS4/MH_FactSheet.pdf
5. International Institute of Population Sciences (IIPS) - National Family Health Survey (NFHS-4) Maharashtra State Report <http://rchiips.org/NFHS/NFHS-4Reports/Maharashtra.pdf>
6. Park K. *Nutrition and health.* Park Text Book of Preventive and Social Medicine. 19th ed. India: M/S Banarsidas Bhanot publishers; 2007. p. 480-533.
7. Perry GS, Yip R, Zyrkowski C. Nutritional risk factors among low-income pregnant US women: The Centers for Disease Control and Prevention (CDC) pregnancy nutrition surveillance system, 1979 through 1993. *Semin Perinatol.* 1995;19(3):211-21.
8. Royston E. The prevalence of nutritional anemia in women in developing countries: A critical review of available information. *World Health Stat Q.* 1982;35(2):52-91.
9. World Health Organization: Global Prevalence of Anemia in 2011 http://apps.who.int/iris/bitstream/10665/177094/1/9789241564960_eng.pdf

AUTOMATED JURISDICTION SYSTEM

Naje Aqeel, Sardar Sobiya, Sayed Rehan ,Adhikari Shahzeen, Prof.Apeksha Gopale

Thane, Uran, Mumbai, Panvel, Mumbai

Department:Computer Engineering

AIKTC – School Of Engineering And Technology, Panvel, India

ABSTRACT: Every work of Judiciary System is done manually .From filing FIR to the last verdict of judge too much time is consumed. We are proposing a system that is trying to automate Indian judiciary system, so as to minimize the wastage of time of people involved in the case. The FIR and the charge-sheet will store online in our system which will make sure they are secure, also the system will predict act and section on the basis of keyword while creating charge-sheet. Judge and lawyers of both the parties will have complete details of case in the form of web or mobile software. In this paper we are going to discuss system's working scenario.

Keywords: E-judiciary, acts, section, predict sections, android, website,FIR.

I. Introduction

The Indian judicial system, a part of world's largest democracy, is very old to follow. Right from the monarchical rule to the British era and the modern system of the independent India.

More than 3 crore cases pending in different courts of India. Many of these cases are pending for more than 10 years. Below is the table that Shows approximate number of cases pending in Supreme Court, High Courts and District and Subordinate courts in India.

Table 1.1 Data about pending cases

Court	Pending Cases
Supreme Court	Around 60 Thousand
High Courts	Around 42 Lakh
District and Sub-ordinate Court	Around 2.7 Crore

This is one of the primary issues of Indian jurisdiction system. The main reason of this is delay in hearing and old manual work. Like the other pillars of democracy, the judiciary too has been found to engage in corruption. there are higher chances that evidence are intentionally or accidentally miss placed.

If there is some solution that can solve these kinds of problems then it is beneficial for judiciary. CourtCal (Automated Judiciary System) is one of the solution for these problems.

A. Objectives:

The main objectives of our system are:

- ✓ To reduce the load of jurisdiction system.
- ✓ To provide an interface that can predict sections at the time of FIR and Chargesheet registration.
- ✓ To decrease study load of lawyers by providing similar case details.
- ✓ To provide case history details to judge and lawyers.
- ✓ Also system will notify about about case hearing dates to manage and schedule case.

B. Scope

A website that will register police and stenotypist in our system. In this website stenotypist can type entire conversation. Police can submit FIR and chargesheet online. Lawyers and judge will have android app which will contain all the details of case. Case assignment and scheduling will be done through app.

II. EXISTING JURISDICTION SYSTEM

Our product is intended to be used as a system which will be an android application and website. To put the paper work of this government section to whole new technical level, this project is going to be the most efficient for judicial purpose. The data about all previous cases and current cases are preserved.

There are many problems which are being faced while working in jurisdiction:

- ✓ While putting charges against the criminal it is time consuming to look into charges and put the

right charge against him.

- ✓ The information like FIR, Charge-sheet can be lost or intestinally misplaced by some higher power for their own profit.
- ✓ It is possible that while transferring files/papers to some another place it misplaced.
- ✓ For lawyers, they need to dig into most of the previous similar cases to find the similarity to increase their winning chances or to straightens their argument. This process can be very time consuming and frustrating.
- ✓ They need to learn old cases study it's method and what actions were taken on to them, so that he can make precautions and identify cons and pros of the case.
- ✓ There are chances that judge has been changed and new one is sitting to give the decision, so new one doesn't know anything about the current case and he will be expecting some reports documents for he can study about the case and make his decision.
- ✓ It is possible that judge is not available on date, this can be problem because the peoples came by leaving their work, so they need proper notification about this kind of information.
- ✓ In India presenting a witness to court is the toughest work to do, like while travelling to court he can be killed or kidnapped or tourtured to change his statement, and how they know who is the witness? By the witness application which is being submitted hours ago.

III. METHODOLOGY

For the sake of understanding, let's take an example:

A person went to a police station to file FIR about a fight that happened across the street. first the police will file FIR, listen to the witness and take act against the suspect. At the same time another person will be documenting FIR online in our system. This page will be accessible after particular login conditions of an authorised police officer. This will be sent to the server where all the information about the case are stored. It will be preserved for future references. This information will be sent to jurisdiction and should be received by them within 12 hours. They will have access to read the information about the case.

Our web will also have a feature that create charge sheets. Based on crime system will predict charges. Now it will be upon the police officer which charge he/she is going to put on the accessed. This helps them by saving a lot of time that is used for researching and thinking of charges. All this done, the only thing left is to make sure the accused is present at the hearing. This application will have all the information about the case. The lawyers of both the parties will have this application which will update them. If lawyer want to learn about similar previous cases, he will have it just at the tip of his fingers.

The query will be performed in the background and search for keywords similar to the case.

The app will give him enough information to summarize the laws, acts and sections. It also stores and display the debates between the defender and prosecutor transcribed usually by court report or stenographer If the lawyer wants to produce the witness, he only need to fill an application of witness, which will notify the judge, the lawyer wants to produce a witness. If the judge wants to look at all the facts, he has it all in the application without being tempered.

When the judges set another date or adjourns the hearing or whatever reason, a court reporter will give an update on the case to both parties.

A. Use-case:

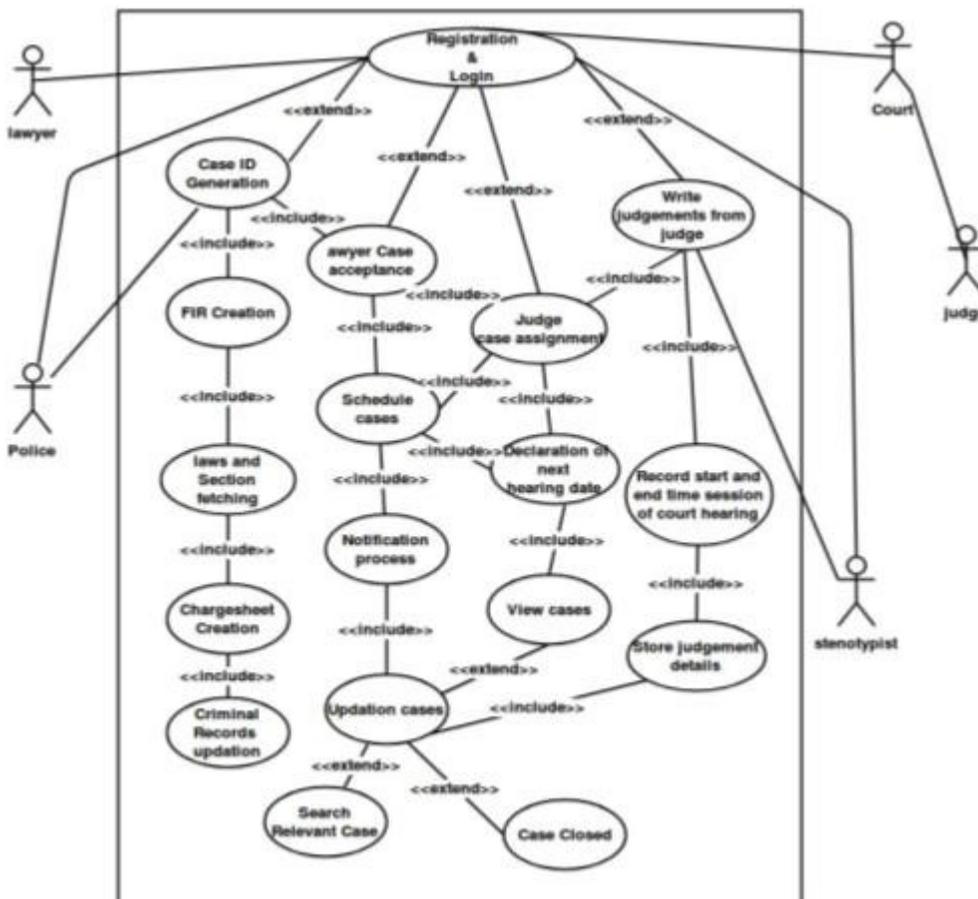


FIG 3.1. USE CASE DIAGRAM

Description : There are 4 roles; Police, lawyer, judge and stenotypist. User will register on our system and after log in based on role process will start.

- **Police** start typing FIR and charge sheet on our website and store it online. In police module, prediction of sections will be done at the time of FIR and chargesheet creation using multi label classification of supervised learning. When case comes in court, it will be assigned to judge and case scheduling will be done.
- **Lawyer** will schedule all the cases he/she is fighting, get details of his/her own cases and also get previous similar case details in android app. Scheduling and assigning of cases will be done through system. Lawyer will get notification from our system for case hearing.
- **Judge** will have details of all cases in app. Notification will be sent to judge for hearing. Judge will notify if he/she will be absent on schedule date.
- **Stenotypist** type entire conversation/ discussion at the time of hearing on our website. Discussion along with hearing start and end time will be store online.

IV. CONCLUSION

The proposed System will be able to minimize manual work and load of jurisdiction system. Case file and data related to case will be stored online in database server. System will predict sections at the time of FIR and chargesheet creation. Lawyers will get detail of similar cases. System will provide an interface for stenotypist to write conversation going at the timing of hearing, and scheduling of case will be provided by system

References

1. E-Court: Technology diffusion in court management- Proceedings of the Nineteenth Americas Conference on Information Systems, Chicago, Illinois, August 15-17, 2013.
2. ICT in Indian Court; Challenges & Solution-International Journal of Internet Computing (IJIC), ISSN No: 2231 – 6965, Volume-1, Issue-2, 2011
3. A secure E-Court System for India- October 2012
4. Machine learning algorithm for automatic Persian text classification.Iran Telecommunication Research Center
5. <https://lawrato.com/indian-kanoon/ipc>
6. WordNet and Cosine Similarity based Classifier of Exam Questions using Bloom's Taxonomy ijET – Volume 11, Issue 4, 2016

CAPACITY PLANNING FOR PUBLIC TOILET

¹Dr. Ashfaq Shaikh, Arun Pandey, Shushant Jha, Indrajit Waghmare, Fatima Patel

Assistant Professor, Student, Student, Student, Student Department of Information Technology

¹M.H.Saboo Siddik of College Engineering, Mumbai, India

ABSTRACT: The topic was undertaken as a small initiative towards restraining the problem of “Open Defecation” in urban slums of Mumbai. The area under study was slums of ‘A WARD’ in Mumbai, Our solution to this problem is a website with NGOs and Government bodies as its end user. The website has two powerful algorithm that can accurately predict the number of extra toilets required in a particular slum and also where it should be built exactly, which would be directly recommended on the map and via statistics. Through this website the concerned authorities can take necessary measures to tackle open defecation problem in slums and redirect the authorities to build the toilets required. This website will also generate statistics as per the required number of toilets for male and female and also urinals for gents separately. This website can effectively curb the grave problem of open defecation reducing health related problems and will be a small step toward Swachh Bharat.

Keywords: Capacity Planning, Public Toilets, Open defecation, A-Ward, Recommendation, Slums, Mumbai.

I. INTRODUCTION

The overall objective of the project is to break the age old behavior of defecating in the open along with making a clear view for the people on the importance of sanitation and toilet usage. Around 17% of the total estimated population lives in urban slums in India. In Mumbai currently there are 11,170 public toilets available i.e. 1 toilet is available for 750 peoples. BMC aims to have an average of 1 toilet per 50 people according to which there is a need to build 1.3 lakh public toilets and the fact is that the public toilet which are available for male: female is in the ratio 3:1. Capacity planning for public toilet is a website where a user can login or register to the website. As soon as the user goes through our website the user can select any ward from all the A wards so to recommend toilets as per requirements. After selecting any one ward we get some statistical information (total number of houses, total population, total number of public toilets) related to any area in that particular ward by that we can get the ratio to build toilet as per the units recommended. We can also recommend the toilets by entering manually the statistical data. This website will provide the solution by a leaflet map to come up with a logical and rational way to recommend the locations for constructions of toilets on a scale of Normal, Medium and Critical. By using this website we can even predict the future scope of planning the public toilets according to the census. This website also provides a complete documented report having all the information regarding the toilets construction with a better UI design.

II. LITERATURE REVIEW

RECOMMENDATION SYSTEM: A Recommendation system is a subclass of information filtering system that seeks to predict the “rating” or “preference” a user would give to an item. A user just has to provide a user input and then the system then runs some algorithm in the backend and shows the best recommended output. Recommender systems are utilized in a variety of areas including movies, music, news, books, research articles, search queries, social tags, and products in general. There are also recommender systems for experts, collaborators, jokes, restaurants, garments, financial services, life insurance, romantic partners, and Twitter pages.

CAPACITY PLANNING: Capacity planning is the process of determining the production capacity needed by an organization to meet changing demands for its products. In the context of capacity planning, design capacity is the maximum amount of work that an organization is capable of completing in a given period. Effective capacity is the maximum amount of work that an organization is capable of completing in a given period due to constraintssuch as quality problems, delays, material handling, etc. In our context Capacity Planning refers to determine the number of public toilets available in the particular area mapping with the population living in that area. And if the number of public toilets are deficient build the specified number of toilets so as to achieve its efficient utilization.

1. Paper Title: Online Facility for Location of WaSH services (IRJET)Year:

March 2018 Authors: Alisha Narnavere, Anoushka Patil, Ayushi Gaikwad, WaSH, web application, contains the information of washrooms,garbage disposal/dumpyards ,and drinking waterboothsindifferenttabswhichtheusercanaccessasrequired.Along with the complete information it will help you in finding thenearestone. Itisafastandsimplegarbagewashrooms,disposal/dump yards, and drinking water booths finder. The database of tens of hundreds of public toilets, dump yards, and drinking water booths is accessed via internet for fast and access. Search for a location on the map and it will show the user the nearest service. Tap on the map to get directions provided by Google Maps. Or if user is only concerned with the existence of service in an area, they can tap the directionsbuttonanditwillgivedirectionstotheclosestservice.The dynamicnatureofthedatabaseallowsittokeep Making updatation such as adding or modifying existing data. Furthermore, the user will have an access to upcoming news updates about WaSH sector and reforms that are constantly established in this sector on national and international scale.

2. Paper Title: Place Recommendation System (IRJET)Year:

March2017Authors:SunilSingh,RajwantGupta,SheenaPanjabi.Thisprojectfocusesondevelopinganandroidappl icationwhich can be used for recommending places to the user based on his currentGPSlocation.Forthisweuseuserbasedcollaborativefiltering algorithm which evaluates places based on user's current location, check-in and preferences. Collaborative filtering algorithm is used to recommend nearest and best places to the users from the device. For the purpose of recommendation, we use a self-driven database for information of places which incorporates real-world check-in spots. Our objective is to extract information about a place and recommend user their place of interest based on their ratings and systemalgorithm.

3. Paper Title: Efficient Recommender System using Collaborative Filtering Technique and Distributed Framework (IRJET)Year:

September 2016 Authors: Sonali B. Ghodake & Ratnamala S.Paswan. Data in the form of reviews, opinions, feedback, remarks, and complaint treated as Big Data cannot be used directly for recommendation system. This data first need to be filter/transform as per requirement. Thus, through this paper overview of need of recommendation using distributed framework is done. Proposed system makes use of user based collaborative filtering technique for recommendation, naive bayes classifier for classification and Hadoop is used as distributed framework. Apache Mahout framework provide flexibility in using pre-existing algorithms. As it is built on the Hadoop framework it solves the problem of scalability. Because of Hadoop, system is highly scalable, fault tolerant and it can handle data set of million ratings on single node. The proposed implementation is platform independent. Experimental results demonstrate that proposed system significantly improves the performance and scalability of recommender system over existing approach. Recommendation based on popularity. Architecture of implementation of Recommender system using Hadoop.

4. Paper Title: Implementing Collaborative Filtering On Large Scale Data Using Hadoop And Mahout (IRJET)Year:

July2015Author:SwatiSharma &ManojSethi.In thissystema combinedapproachofuser-userCFanditem-itemCFhasbeen presented to generate recommendations on Hadoop cluster using Apache Mahout, a library for machine learning algorithms. By using combined approach, accuracy of recommendation has improved.This approachhas scaledwellwiththehadoopplatform.Timeneededtosolvethetheproblemhas reduced.Mahoutisable to handle big data but it still lack some algorithms. There commendation for singleuserneedtobeimprovedforbetterresults.New computing platforms like Apache Spark are getting prominent in the field of Big Data analysis. Recommendation algorithms can be performed on such platforms for fasterperformance.

III. PROPOSEDSYSTEM

The website will mainly perform two functions i.e. recommending the numbers of extra toilets required in a area and secondly the area where these toilets will be built. The recommendation system has a smooth workflow with access to administrator and users separately. The main functionality lies in the Area tab where a dropdown list will be displayed and after selecting the A Ward option, On the left side of the screen an Open Street Map satellite view will be displayed with highlighted A ward area of Mumbai and on right hand side slums of A ward will be displayed along with statistics.

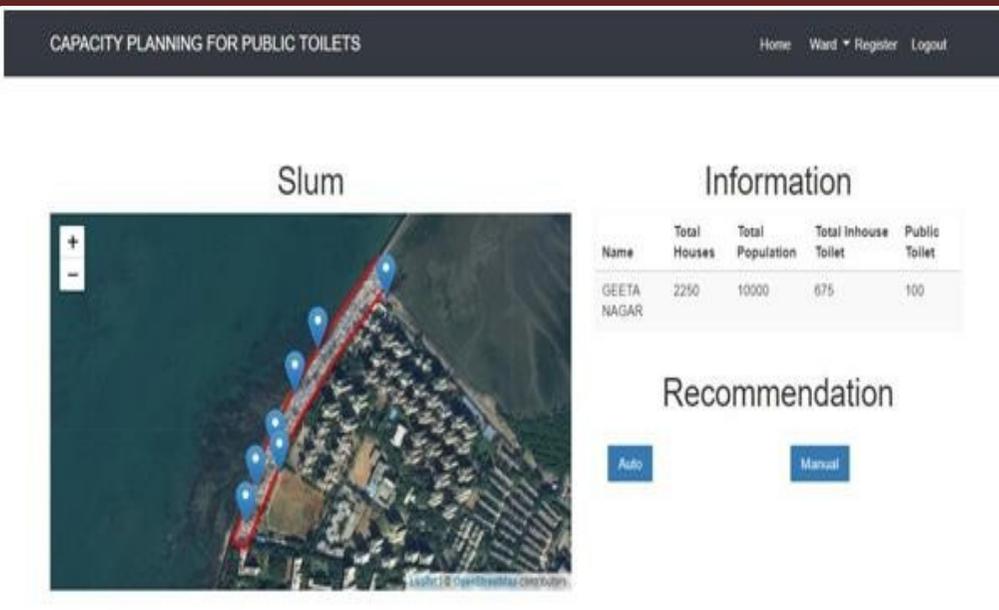


Figure 1: Statistics related to a particular selected slum

Now when we click on a particular slum then map of that slum is highlighted and table containing information of that slum is displayed with two buttons on the bottom i.e. Auto button and a Manual button. Auto recommends numbers based on our studies and input already provided to the database but since the name itself is capacity therefore future growth should be insight and hence manual button, which displays a form and by feeding asked information on the form it will automatically recommend the numbers. Since the BMC recommend that there should be at least 1 toilet per 50 person the recommended numbers are broadly classified as:

- 0-64 (In Green colour) - showing area with less requirement.
- 65-100 (In Blue colour) - showing area with moderate requirement.
- 101+ (In Red colour) - showing area with immediate requirement.

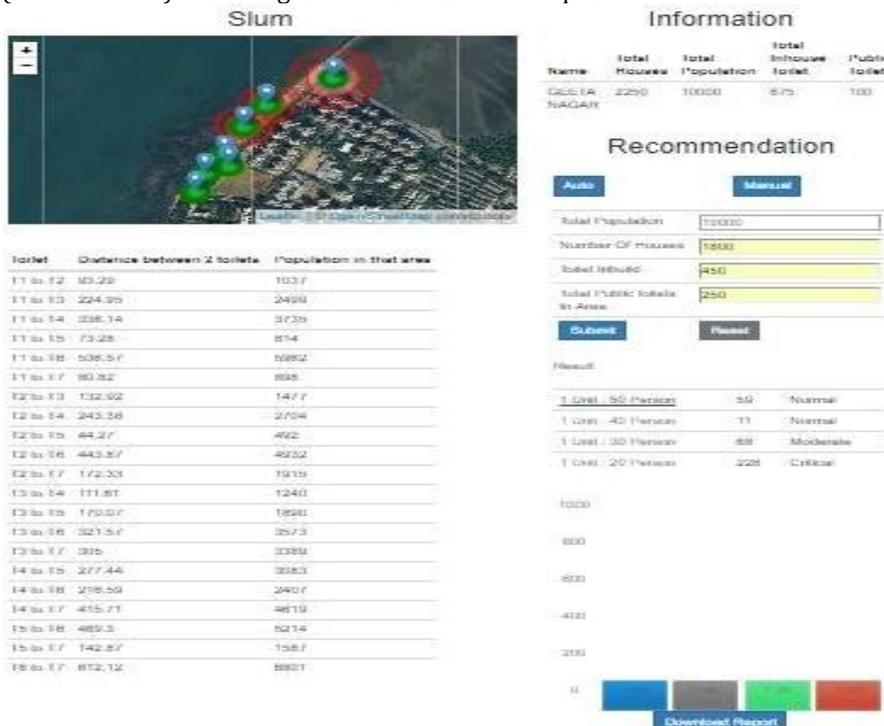


Figure 2: Recommended number of toilets to be built is shown.

The most important part in this system is to recommend the area where toilet should be constructed because we cannot pin point a particular coordinate on a map since there can be many reasons and disputes associated with that place so here we proposed a generic idea where it should be constructed which of course will be possible only after taking survey in that recommended region.

IV. METHODOLOGY

The parameters considered here are-

- Total number of population in aslum
- Total number of houses in aslum
- Total number of houses having inbuilttoilets
- Total number of public toilet available in thatarea

The entire process of tracking, monitoring and generating alerts involves following two steps:

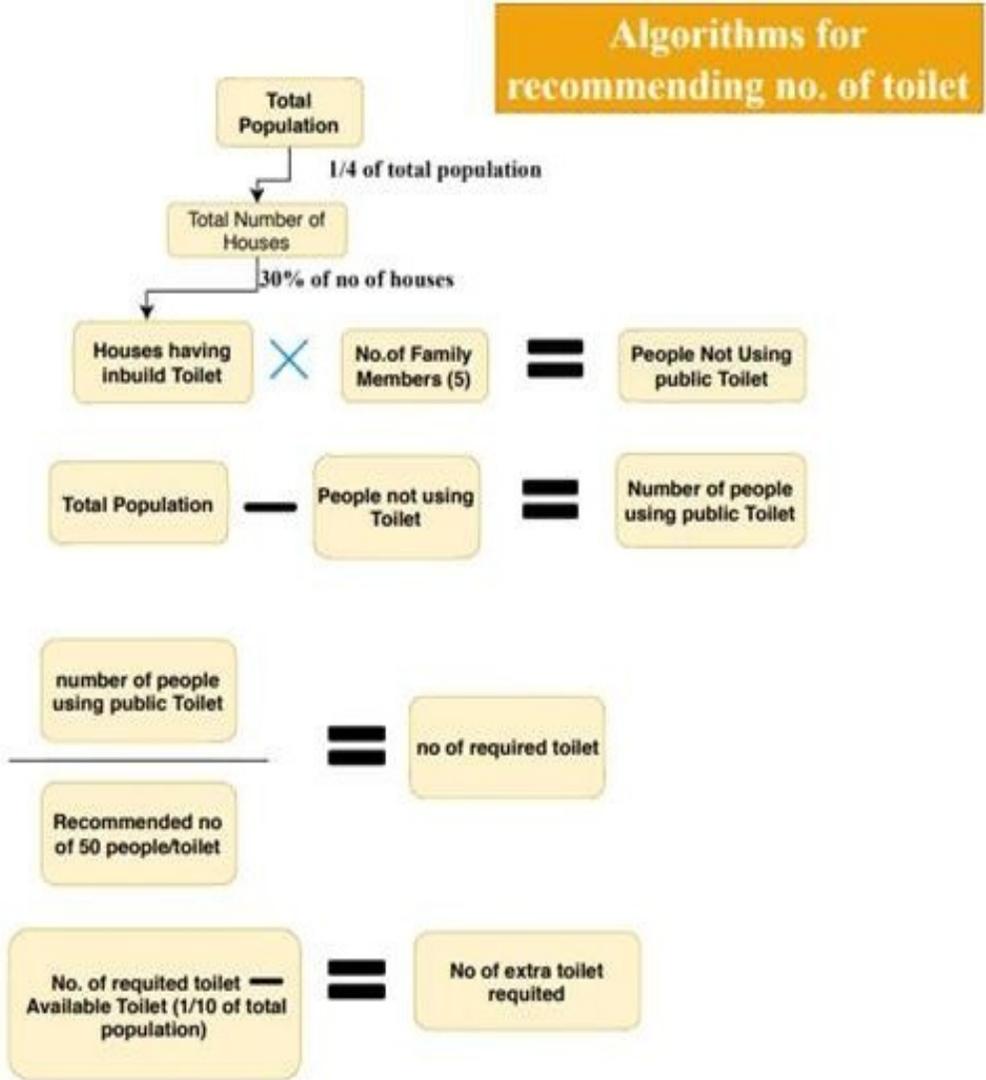


Figure 3: Algorithm for recommending the number of toilets to be built.

Here when the distance between a toilet and its nearest toilet is considered simultaneously and they are matched with the categories mentioned above and two circles with radius given as per above is drawn and the area which we recommend to build toilet is the region between two circles. The rule to follow here is Greater the distance between two nearest toilet, more will be the number of toilet to be constructed from the recommended number.

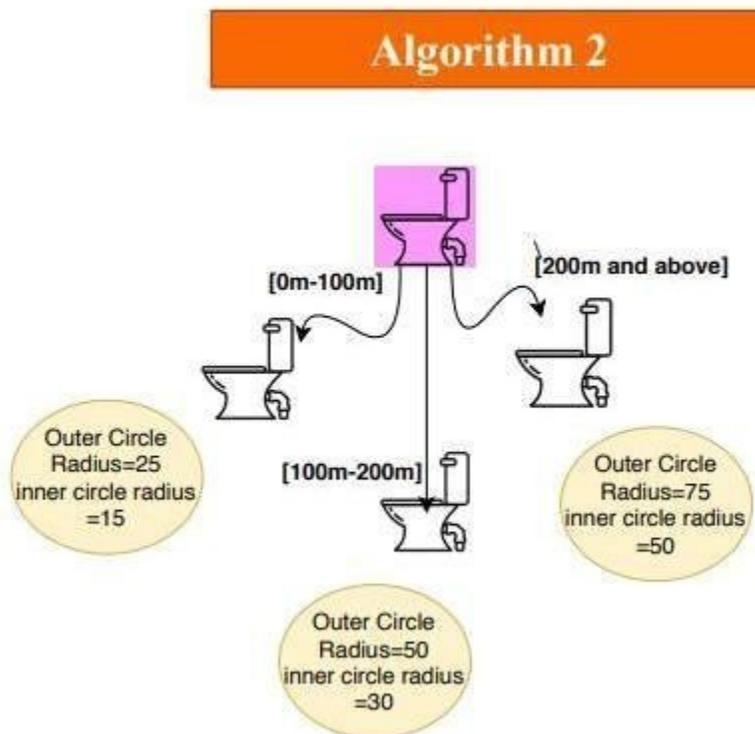


Figure 4: Algorithm for recommending the location where the toilets are needed to be built.

V. FUTURESCOPE

The website can help to collect information and maintain records and will also be helpful for eradicating the problem of open defecation from 4041 towns of India. New technologies such as AI and Machine learning can be used to enhance the system in the future after acquiring lots of data. An Android application can be built in which the government authorities can perform the same function as the website does and also using the GPS location can reach the recommended location.

VI. CONCLUSION

As per the proposed idea, a prototype system is built through which we march one step closer to clean and Swachh India thereby improving the standard of living and health of people of India.

VII. REFERENCES

- "Alisha Narnavere, Anoushka Patil, Ayushi Gaikwad, Prajakta Wasnik", "Online Facility for Location of WaSH services", International Research Journal of Engineering and Technology (IRJET), Volume: 05, Issue: 03, March-2018.
- "Sonali B. Ghodake, Ratnamala S. Paswan", "Efficient Recommender System using Collaborative Filtering Technique and Distributed Framework", International Research Journal Of Engineering and Technology (IRJET) Volume: 03, Issue: 09, September-2016.
- "Swati Sharma, Manoj Sethi", "Implementing Collaborative Filtering On Large Scale Data Using Hadoop And Mahout" International Research Journal Of Engineering and Technology (IRJET), Volume: 02, Issue: 04, July-2015.

ConnectU: A Local Business Platform

¹Saly M. Lopes & ²Melbourne B. Lopes

¹Assistant Professor, ²Android Consultant

¹Department of Computer Engineering,

¹St. John's College of Engineering and Management, Palghar, India

ABSTRACT: Today, enormous work has been done in the field of mobile technology. So is the way emerging that leads to serving services and products electronically. And as we know, development in technology has given a boost to many domains such as entertainment, social networking, artificial intelligence, e-commerce, marketing, tourism, web application/portals and so on. ConnectU is the platform that allows all kind of local professionals and businesses to showcase their skills to the intended consumers. ConnectU pins out the rural business and gives access to the service, for the use by locals. On ConnectU, one can find 360+ nearby services that are required on daily basis such as plumber, electricians, home service, car service, e-gadget service, grocery service, etc. Once a consumer locates his service, he can connect with the concerned service provider through the telephonic communication. A map visualization helps the consumer to search for the nearby service providers based on his geographic location. ConnectU application is a search engine of local businesses and is built on android technology using table index method. Currently, the application serves within the demographic area boundary of Vasai Taluka, a small town in India.

Keywords: Android technology, location based services, map-based visualization, search engine, service provider, and table indexing search.

I. Introduction

Nowadays, the demand for small businesses has increased, and so is the need of making them available over the online systems to be used by the consumers. This leads to the installation of a system that will bring up the rural enterprises. The setup of any e-commerce system in wireless environment is called mobile commerce (m-commerce). In m-commerce [1], the process of dealing of goods and services is done through wireless devices such as cell phones, personal digital assistants (PDAs).

This paper represents the mobile application named ConnectU, where the main agenda is to locate the nearby services which could be small, medium or micro-enterprises, locate the skilled workers and establish a connection with them through the smartphone/telephone. The connection established is between the service seeker and service provider who are the two main roles that interact and are benefited through this system. Service seeker is the user or consumer and service provider is the one with the business of its kind. The actual services provided through this application are local businesses owned by electricians, plumbers, painters, carpenters, consultant agents, real estate agents/brokers, photographers, fashion designers, caterers and many more. The services of this application can be availed within the geographical boundaries of Vasai taluka (District- Palghar, Maharashtra, India) only.

ConnectU is implemented on Android operating system and is available on Google Play Store and its downloadable link is given and can be used to install it (<https://play.google.com/store/apps/details?id=com.project.connectu>). The web services are hosted on a secure superior server. The map-based interaction provides reliable and fast service search. The consumer can view his location on the map and once he makes the search for a particular service, he can view the locations of all service providers for the search made within the area boundary. Alternatively, a distance wise list of searched services is displayed. The consumer, based on some criteria like budget, reviews, distance etc. will select the service provider that best fit the criteria and makes a telephone call for further assistance. The service provider is notified by SMS and a connection is recorded. Another interesting feature of this application is locating all the nearby services on one click within some range of distance from user location or within a predefined radius. The Google assisted voice-based search of the application involves locating local services, serving local people. Such features are beneficial where, due to lack of knowledge and unreachability of technology in rural regions the skill workers like the carpenter, plumbers, electricians, etc. sell their skills at a much cheaper cost and to small extent. Also, enterprises other than targeted can find this application beneficiary, so that the ideology can be expanded to the other regions of developing countries like India.

II. USER CHARACTERISTICS AND CHALLENGES WITH RURAL ENTERPRISES

Digital India is trending and making its way into the rural parts of the country, then too there is a starvation of internet, mode of digital communication and technology among majority of population. Illiteracy, poverty, lack of basic needs, no access to emerging technology, poor infrastructure, no connectivity of network, inaccessible internet, less social awareness, communication languages, non-availability of credit/debit cards are some of the basic problems faced by rural people (Standing Craig et al.,2006). Also, acquiring the trust of rural people to perform any online transaction is much difficult. They are afraid of investing in a large amount, and being getting cheated. Introducing a technological platform for rural enterprises in the simplest way could be a beneficial job. For example, in the traditional shopping method, firstly, shops are to be purchased on their own or on a lease. Secondly, the location of the shop has to be in the market. Thirdly, capital investment for the materials required in and on the shop. Whereas in e-commerce trading all the prior steps of traditional shopping are eradicated and one can store the materials in the small storeroom/godown, located anywhere geographically and provide the service upon electronic request. Mobile communication gives the independence to carry out the business under such situations. Most of the professionals, local business holders might not have the smartphone. Because of which they might fail to expand their services.

Hence, local economy digitization and empowerment becomes the main objective. The Internet and Mobile Association of India states that, as of December 2017, there were 291 million urban internet users and 187 million rural internet users estimated. Also, there are about 650 million mobile phone users out of which 300 million have smart phone, according to technology consultancy Counterpoint report. Where more than 50% people in India don't have access to the smartphone. And the majority of them are rural locals. Even though, the small enterprises can gain business by simply connecting their phone number (it could not be a smartphone) to the smart application to get noticed by the consumer.

III.BACKGROUND AND RELATED WORK

Currently, service-based application is a high-end trend with the intent of spreading worldwide, in the near future. The leading startups in India, compete among themselves to better cover the lifestyle of people and serve them the required services. Many applications such as UrbanClap, Justdial, Askme, Quikr, Housejoy are in lead with their own motive to serve services through a digital platform. UrbanClap [4] tends to serve as an e-marketplace [2] for local services and has made their mark up to major cities across India. With Justdial [5], the lead is sold to many service providers where either of them will have a chance to get the actual business. The 18% GST tax system, cost more for the online services rather than offline market. Likewise, Facebook's marketplace feature elaborates the social networking to e-commerce. It provides the electronic marketplace to buy and sell the items not only at the business level but also for the sellers [6]. Because it revolves around the buy/sell feature it lacks in providing filters to select an item under a particular category. And the items appear and disappear very quickly based on the item's availability. However, trust is the important factor which holds the intended user and businesses to a particular platform [2]. Thus, increased trust equals decreased risk.

ConnectU app serves both B2C as a primary and B2B as a secondary service. This platform serves its users "relevant data" based on user's current physical location. Here, peer-to-peer lead is generated and the user may contact the service provider of his choice. Vasai town is the 5th largest satellite city in Maharashtra, India. The 2011 census count put the estimated population of Vasai as 3.2 million people and this number is projected to grow up to 5 million by 2021. Knight Frank, a research agency, has estimated that growth rate in Vasai is much faster than that of Greater Mumbai, hence arises the need of all kind of services alike household, cleaning, e-gadgets etc. Vasai being the native place of both authors and with all the opportunistic estimates, make this area highly lucrative for setting up a venture like ConnectU. The DMIC (Delhi Mumbai Industrial Corridor – Indian Government Ambitious Projects) runs through Vasai town and will definitely nurture the growth of SMEs in this region. And will strive to the endless list of the service provider community.

Currently, around 350+ services are supported by ConnectU and will continue to add more as they come in picture. As of now 1500+ service providers, professional, skill workers and rural enterprises from all over Vasai-Virar town have enrolled themselves on ConnectU. Fig. 1 shows the demographic potential of the pilot area of operation.

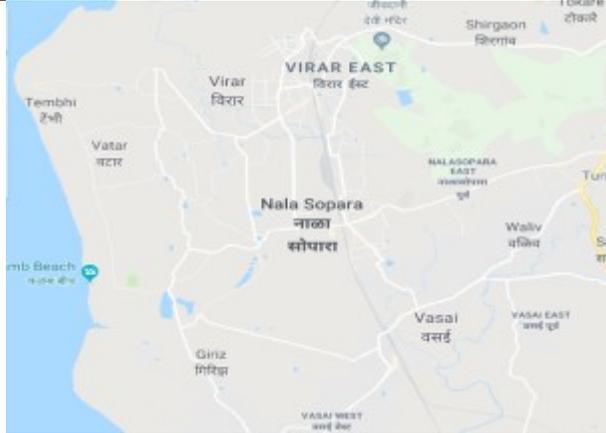


Fig. 1. Geographical location of Vasai-Virar town[7]

IV. METHODOLOGY AND DATA ANALYSIS

ConnectU is basically an intermediary form of platform, where the market place is available for public and owned independently by different businesses, but managed by the third party. Such type of public market increases the level of transparency. The third party involved here, has no control over any service provider's business. Only the record of which service is provided and by whom, is maintained in the application side. And the same record is provided for the consumers electronically. The entire service provider data is to be verified and validated upon its registration by the telecaller of ConnectU. And to maintain security and authenticity each user of ConnectU app is maintained by separate login account.

Figure 2. shows the process diagram of how service provider and service seeker will be connected to each other. The service provider and seeker will be connected to each other via cloud and communication between them will be carried out wirelessly. The consumer has to register on the ConnectU app to avail the services. And businesses have to be registered either through the application or through the website (Registration link- <http://spregistration.connectUapp.com>) before they are made available. The mobile application is the intermediary party that handles the storage, retrieval, and provision of data in the form of services.

Initially, data of service providers was collected by locating the business and visiting door to door in maximum region possible. Followed by advertising based marketing method was implemented to collect data. Also, there is the flexibility of language selection for data entry on the registration website. While collecting data, there was two major concerns of the people: trust and investment. The fields of data entries filled by the service provider are the name, description of the business, address, mobile number, and the geographic location. Geographic location is captured in the form of latitude and longitude of the service provider.

The consumer has a map visualization of all the services. The consumer can locate nearby services which are calculated based on latitude and longitude and can also locate services, searched through keyword which gives distance wise list of services as shown in fig. 3. ConnectU mobile application is a search engine that works on table indexing search [8].

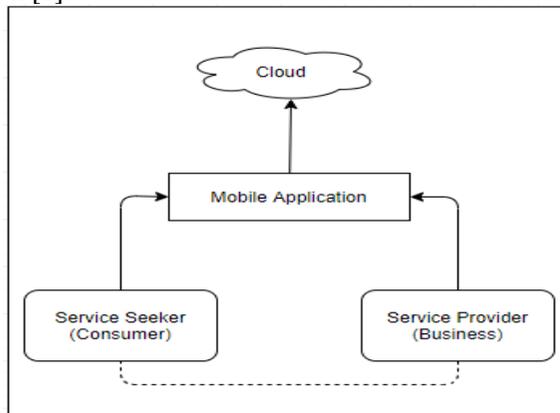


Fig. 2. Process diagram for connectU application

```
"SELECT DISTINCT l.lead_id, l.name, l.mobile1, l.landline, l.business_level,  
l.latitude, l.longitude, l.address, l.photo, l.ratings,  
l.total_ratings, ls.category, ls.sub_category, c.icon_name  
FROM lead_services ls  
INNER JOIN leads l ON ls.lead_id = l.lead_id  
INNER JOIN category c ON ls.category = c.category_id  
WHERE (l.latitude BETWEEN :southWestLat AND :northEastLat)  
AND (l.longitude BETWEEN :southWestLng AND :northEastLng)  
AND l.verify_status=1 AND ls.category !=1 GROUP BY l.lead_id"
```

(a)

```
"SELECT DISTINCT l.lead_id, l.name, l.mobile1, l.landline, l.business_level,  
l.latitude, l.longitude, l.address, l.photo, l.ratings,  
l.total_ratings, ls.category, ls.sub_category, c.icon_name  
FROM leads l  
INNER JOIN lead_services ls ON l.lead_id = ls.lead_id  
INNER JOIN category c ON ls.category = c.category_id  
WHERE l.name LIKE '%$user_keyword%' OR l.description  
LIKE '%$user_keyword%'  
AND l.verify_status=1 GROUP BY l.lead_id"
```

(b)

Fig. 3. Code to extract the latitude and longitude to calculate the: (a) nearby services and (b) list the services distance wise for keyword-based search.

V. PROPOSED SYSTEM

As mentioned in the previous sections, ConnectU is an application developed to serve as a virtual intermediary. The practical and physical modules are implemented successfully to be in use by the locals.

3.1 System Architecture of ConnectU Application

System diagram shown in fig. 4 represents the system modules and interaction among the modules. The mobile number verification module accepts the data of all the service providers and consumers who are going to operate through this application. In application-based registration, user has to enter his mobile number for registration and service provider has to register through web-based registration via the link mentioned earlier. The verification will be done by sending a one-time password (OTP). The SMS gateway used for sending OTP is maharecharge [9] service, one of the leading SMS vendors in India. Thereafter, the consumer can log in to the system. The details of the service provider are initially stored in a temporary database as they have to be verified before making those entries live. The data stored in panel storage is verified by the telecaller, who would personally make a call to the service provider to assure the data and move the entry on a live database.

The consumer, upon login into the system can see the map visualization of the current location and start a search. The keyword-based search is supported by table indexing search method [8] and voice-based search is driven from Google Assistant's audio search. Once the services are listed, the consumer can connect by making a call to anyone that matches the requirements. The intended service provider is notified through SMS. And entire transaction is stored on the live database in the form of which user contacted which service provider, the status of SMS notification and mobile numbers of service provider contacted.

The web services are developed using PHP and are hosted on Amazon Web Service (AWS) server [10], which provides fast, efficient and secure communication between applications and servers. The server uses a MySQL database to store its records of user details, business details, and transaction history.

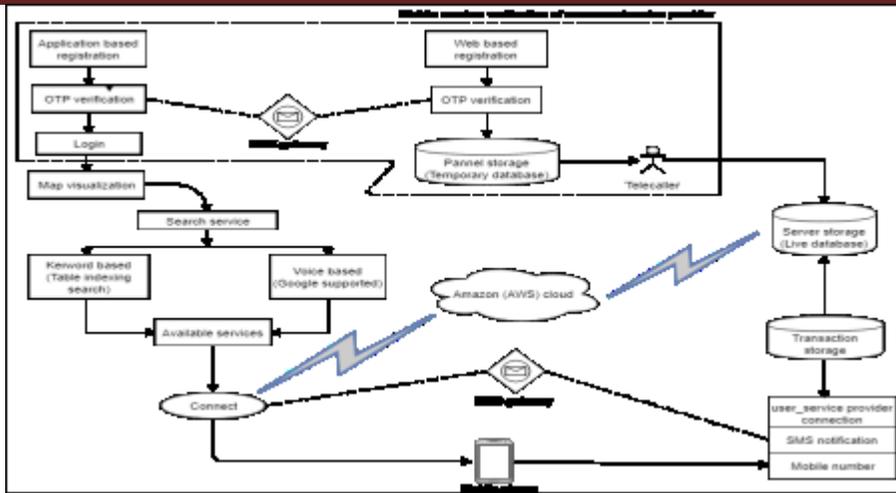


Fig. 4. System diagram for ConnectU application

3.2 Flowchart

In fig. 5, the entire flow works on the electronic catalogues of the service provider. User's request is searched across the catalogues that are stored in the database. The flowchart shows that ConnectU is a search engine for local services and businesses.

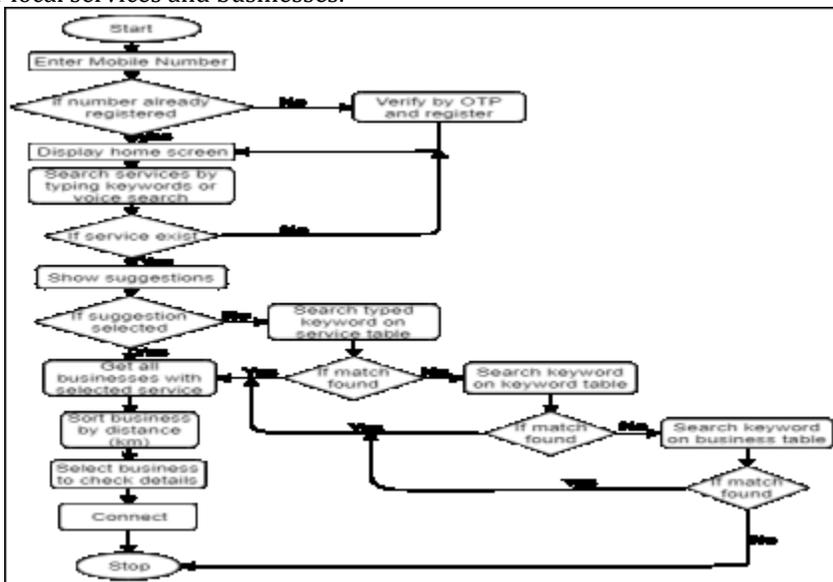


Fig. 5. A flowchart depicting the process of the ConnectU application

3.3 Local Advertisements

Displaying the catalogue of local businesses on one platform, leads to competition among the service providers. And the service with the nearest location gets served first based on the request for the service provider. Advertisement concept gives the lead to the one who comes into the public's eye. There are three different advertising mechanisms deployed:

3.3.1 Profile-based Advertisement

Displaying one's profile in the suggestions, when the user is on a similar profile. This strategy increases the visibility of advertising profile by the consumers.

3.3.2 Media based Advertisement

Displaying the video of business or product in one's profile. There is also a provision for the service provider to share the website link.

3.3.3 Notification based Advertisement

Sending the banner or text of a particular business as a notification, to the consumers. Also, the service provider can notify regarding the offers and discounts at the earliest.

3.4 Experimental results

Fig. 6 shows the graph of active users over the span of few months. The number of registration is categorized based on daily, weekly and monthly basis. The count of active users has gone up to 600+ over the period of three months. The launch of ConnectU application was on 26th January 2016.

Fig. 7 shows that a crash with respect to the mobile application is the point where activity terminates and no functioning of the application is entertained. Thus the unwanted termination of functioning creates a bad view among users for using the application.

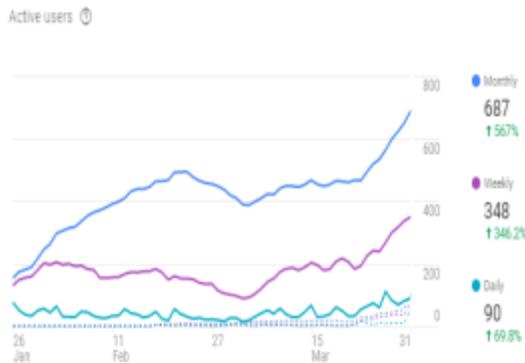


Fig. 6. Total number of active users for a span of three months.



Fig.7. The failure rate due to crashes on end-user devices.

3.5 CONCLUSION

The ConnectU application approaches the basic needs of people in local areas, of serving the local business. Sometimes, many of the services are not delivered to rural areas due to the challenges discussed. The main aim is to target such places and support them with a technological platform area wise. The service provider, on the other hand, could make some more business and reach a wide range of consumers. Raising such a platform requires an enormous number of data of several businesses. Also, attracting consumer about this beneficiary application is a challenging task. The strategy followed for the collection of service provider data is on the door to door basis. Hence, to conclude, more the data more the geographic area will be covered by the application on the field.

This local business directory has a beneficiary impact on both B2C and B2B categories. Data forecasts up to the turn of next decade, 90% of the world's population over the age of six years will have a mobile phone. As an advantageous opportunity in the m-commerce, ConnectU tends to make a success with all its integrated features in serving local, search-based services.

References

1. GeorgeAneet Shilpa, and Tanya Singh. 2015. M-commerce: Evaluation of the technology drawbacks.IEEE 4th International Conferencein Reliability, Infocom Technologies and Optimization (ICRITO)(Trends and Future Directions,1-6.
2. Standing C., Love P. E., Stockdale R., & Gengatharen D. 2006. Examining the relationship between electronic marketplace strategy and structure. IEEE Transactions on Engineering Management, 53(2): 297-311.
3. KhanM., Xu X., Dou W., & Yu S. 2016. OSaaS: Online Shopping as a Service to Escalate E-Commerce in DevelopingCountries. In 2016 IEEE 18th International Conference on High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS), 1402-1409.
4. <https://www.urbanclap.com/about>
5. <https://www.justdial.com/cms/overview>
6. Wilken, Rowan.2014. Places nearby: Facebook as a location-based social media platform. New Media & Society, 16(7): 1087-1103.
7. <https://www.google.com/maps/@19.417565,72.7967145,12.75z>
8. Zhou, Y., Xie X., Wang C., Gong Y., & Ma W. Y. 2005. Hybrid index structures for location-based web search. In Proceedings of the 14th ACM international conference on Information and knowledge management, 155-162.
9. www.maharecharge.com
10. Đorđević, B. S., Jovanović, S. P., & Timčenko, V. V. 2014. Cloud Computing in Amazon and Microsoft Azureplatforms: Performance and service comparison. IEEE 22nd In Telecommunications Forum Telfor (TELFOR), 931-934.

COPY-MOVE FORGERY DETECTION USING DEEP LEARNING

¹VARUN PANDE,²ABHISHEK AGARWAL,³DEEPANSHU NEGI, ⁴ANKIT KUMAR,⁵MANE VANITA

¹Student,²Student,³Student, ⁴Student,⁵Associate Professor

¹Computer Engineering,

¹RAIT, Nerul, Navi Mumbai

ABSTRACT: Nowadays anyone can edit images due to various advanced digital image processing software and tools. Detecting forgery in images is getting difficult day by day. A Copy-Move forgery is when some part of an image is copy and moved in another part. Copy move forgery detection is one of the most widely spread area for research within the blind image forensics field. The traditional methods used for detecting copy-move forgery were complex and time-consuming. Deep learning can be used to detect forgery faster. A Convolutional neural network method has been trained by using the ELA of an image. After then it gave the fast result for detecting forgery in an image.

Keywords: Forgery, ELA (Error Level Analysis), CNN, Pre-processing.

I. Introduction

Image is a great source of information in our day to day life. It is not only a source of information but also a source of validation of certain things. But in recent years due to the advancement of Digital image processing software and tools, manipulation in images has been increased significantly. Forgery in the image can cause false information or evidence and can influence people in a bad way. It is important to detect such forgery and prevent it from happening or spreading in society.

There is various type of forgery like copy-move, digital watermarking, image splicing, image retouching, etc. Among these, the Copy-Move forgery is hard to detect and identify. This is because in Copy-Move part image of the image is copied and moved in another part. Hence there is less change in intensity and background of the image. Also, this method of forgery is easy to implement [1]. Hence it is widely used.

Detecting forgery using a traditional method like Block-based and Key -Point takes requires a lot of time for a single image [2]. Also, as the quality of the image increases the time required to detect it also increases. Deep Learning can be used to train the model for detecting forgery. It will not only save time but its accuracy increases as the dataset increases. Convolutional neural network methods have been used in image classification, image forensics, and image hashing retrieval and many more. It can also be used to train such a model for forgery detection.

II. LITERATURE SURVEY

Table 1 Comparison of different forgery detection method.

Research Paper	Authors	Methodology	Advantages	Disadvantages	Year
Copy-Move Forgery detection using block-based method. [1]	Hajar Moradi-Gharghani, Mehdi Nasri	Block division PCA DCT	Achieves good accuracy.	Fails to detect forged regions at boundary areas. High computation cost.	2016
Copy-Move Forgery detection using adaptive over segmentation. [2]	Sevinc Beyram, Husrev Taha, N. Menon	Overlapping block division. PCA DCT	Able to detect forgery at boundary regions.	High Computation cost.	2016
Copy-Move Forgery detection using key point-based method.[3]	Jinchang Ren, Heng Yang		Achieves better accuracy than block-based method.	Inefficient on smooth surfaces.	2016
Copy-Move Forgery detection using block based and key point-based methodology [4]	Irene Amerini Lamberto Ballan Roberto Casadelli	SIFT	Scale invariant. Rotation invariant.	Not able to mirrored forged regions.	2017
Copy-Move Forgery detection using block based and key point-based methodology [5]	Xiaojie Guo, Xiachun Cao, Jiawan Zhang	MIFT	Mirror reflection invariant.	Not able to detect combination of mirrored and rotational forgery.	2017
Copy-Move Forgery using Deep learning [6]	Junlin Ouyang, Yizhi Liu, Miao Liao	CNN	Low computational cost.	Not able to detect multiple attacks.	2017

Over the past many years research has been done on detecting move-copy forgery from images. Above Table is the comparison of a few papers we have studied with its methodology used, its advantages and disadvantages.

III. PROPOSED SYSTEM

Our System consist of Deep learning model to detect Copy-Move Forgery. Convolution Neural Network has been used for this purpose. It consists of 3 parts, Data Preparation, Training Model, Detecting Forgery.

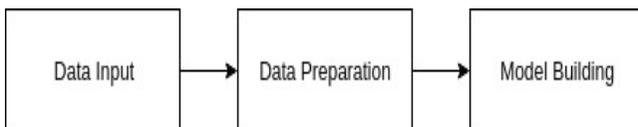


Figure 2 Model Flow

3.1 Data Preparation

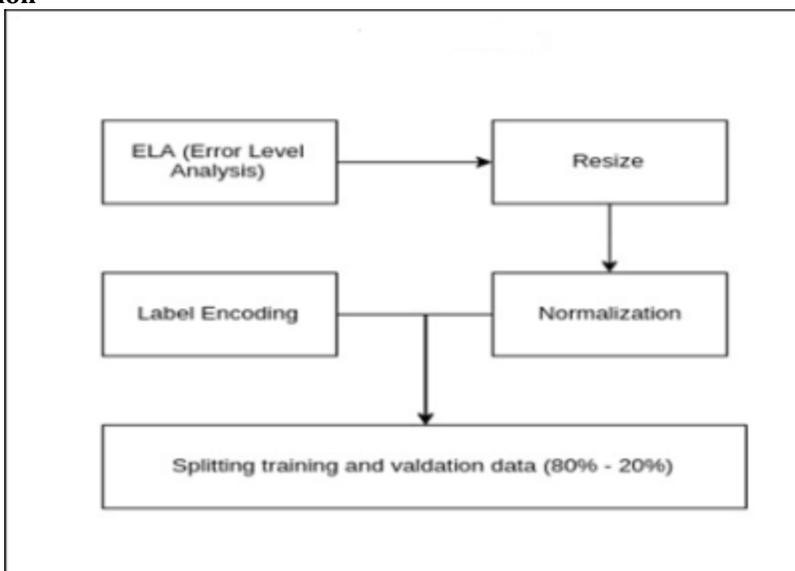


Figure 2 Data Preparation

In Data Preparation we have used Casia Dataset. First Error Level Analysis (ELA) for those datasets is generated. ELA detects areas of the image that is indifferent compression level. In a portion of an image have a significantly different error level then it can consist forgery [7]. For generating ELA first the image is re-saved at a 90% error rate. Then the pixel around its being compared. If there is virtually no change, then the cell has reached its local minima for error at that quality level. However, if there is a large amount of change, then the pixels are not at their local minima and are effectively original. After generating ELA of the datasets, it has been split into two parts for Training and Testing.

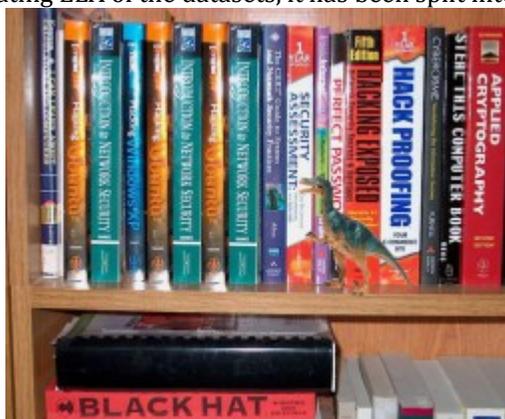


Figure 3.1 Forged Image

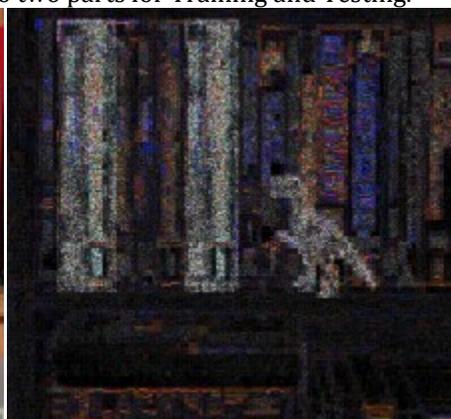


Figure 3.3 ELA of forged image

3.2 Training Model

For training the model custom CNN architecture is used. It consists of 2 convolution layer, 1 Max pool layer, 1 Flattening and 1 Full connection and lastly a SoftMax layer [8].

First the Data sets we had created in previous step is used as an input for the CNN. Convolution Layer is the first layer to extract features from the input image. It is a mathematical operation that takes two inputs such as image matrix and a filter. Image matrix ($h \times w \times d$) Filter matrix ($fh \times fw \times d$) Outputs a volume dimension $(h - fh + 1) \times (w - fw + 1)$ This is called Feature map. There are different filters available such as Identity, edge detection, box blur, Gaussian blur.

After convolution padding is done. It is required because sometime filter does not perfectly fit a image. At that time, we can either pad the picture with zeroes so that it fits or drop the part of the image where the filter does not fit. In pooling the dimension is reduced but the important feature is kept as it is [9]. The flattening step is needed so that you can make use of fully connected layers after some convolutional layers. In Fully connected layers every neuron is connected in one layer to every other neuron in another layer. It is in principle the same as the traditional multi-layer perceptron neural network (MLP). The flattend matrix goes through a fully connected layer to classify the images. For CNN TensorFlow is used.

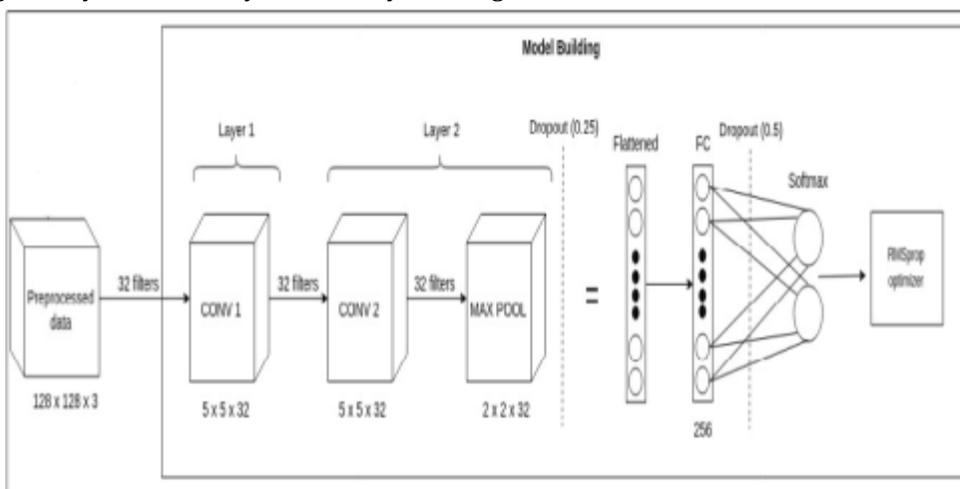


Figure 4 CNN Framework

3.3 Detecting forgery

After training the CNN model weights have been Generated. These weights are been used to detect forgery of an image.

IV. RESULTS AND DISCUSSION

We have tested of data On Casia Dataset 1 and 2. The data dets contain 14331 images including fake and real ones.

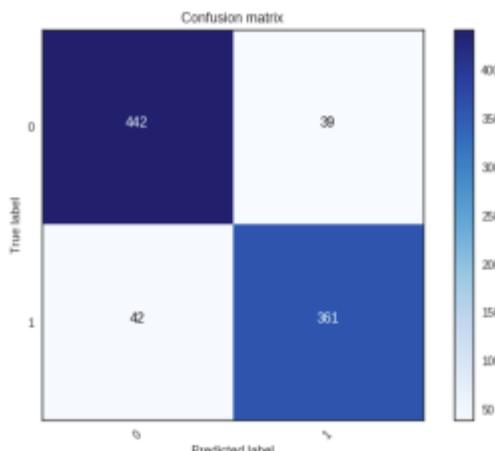


Figure 5 Confusion Matrix

The above diagram is the confusion matrix of the result. Here class 0 represents the Real Images and class 1 represents the Fake Images. The Cell 1,1 of the matrix shows Number of real images it correctly predicted. The cell 1,2 represents number of incorrectly predicted Real images. The Cell 2,1 represents incorrectly predicted fake images. The Cell 2,2 represents correctly predicted fake images.

Below is the Calculations of recall and Precision of the Results.

$$Recall = \frac{TP}{TP + FN}$$

$$Recall = 91.32 \%$$

$$Precision = \frac{TP}{TP + FP}$$

$$Precision = 91.89 \%$$

Table 3 Accuracy Table

Method	Accuracy	Recall	Precision
Previous	90.87%	80.33%	81.42%
Using Deep learning	98.07%	91.32%	91.89%

IV. OUTPUT



Figure 6.1 Real Image



Figure 6.2 ELA of Real Image

The above figure 6.1 is a real image and figure 6.2 is the ELA of that real image. It can be shown from the figure that there is no significant change in the ELA. No portion of ELA contain a larger change in the intensity. Hence in real images the ELA doesn't have a significant drop or rise.



Figure 6.3 Forged Image



Figure 6.4 ELA of Forged Image

The above figure 6.3 is forged image of figure 6.1 and figure 6.4 is the ELA of that forged image. In part that is forged in the original image it can be seen a significant change in intensity in the ELA to the forged images. When a move copy occurs the intensity of the object doesn't match to the new background. Hence it can detect forged images easily.

V. CONCLUSION

By using Convolutional neural network, image forgery detection is more accurate as well as this method is efficient in terms of complexity, cost and time compared to existing traditional methods. We will get better results using CNN which detects whether the input image is copy-move forged or not. The proposed system will be given quickly to detect some image using ELA and metadata to save CNN time.

References

- [1] Irene Amerini, Lamberto Ballan, Roberto Caldelli, Alberto Del Bimbo and Giuseppe Serra, SIFT-based Forensic method for Copy-Move attack Detection and transformation recovery, 2011.
- [2] Xiaojie Guo, Xiaochun Cao, Jiawan Zhang and Xuewei Li, MIFT: A mirror reflection invariant feature Descriptor, 2017.
- [3] Jjiangbin Zheng, Yanan Liu, Jinchang Ren, Tingge Zhu, Yijun Yan and Heng Yang, Fusion of block and key points-based approaches for effective copy move forgery detection, 2016Jiangbin Zheng, Yanan Liu, Jinchang Ren, Tingge Zhu, Yijun Yan and Heng Yang, Fusion of block and key points-based approaches for effective copy move forgery detection, 2016.
- [4] Junlin Ouyang, Yizhi Liu and Miao Liao, Copy-move forgery detection based on deep learning, 2017.
- [5] Maria Venitta, V. Sheerja Kumari, Image forgery detection using feature point matching, 2016.
- [6] Seema Sultana and Sunanda Dikshit, Image forgery detection using feature point matching and adaptive over segmentation, 2016.
- [7] Nor Bakiah Abd Warif, Mohd. Yamani Idna Idris, Ainuddin Wahid Abdul Wahab, Rosli Salleh, An evaluation of Error Level Analysis in image forensics, 2015
- [8] Yuan Rao and Jiangqun Ni, A deep learning approach to detection of splicing and copymove forgeries in images, 2016
- [9] Jaseela S and Nishada S. G, Copy move Image forgery detection technique, 2016.

Dataset Used:

1. <https://www.kaggle.com/sophatvathana/casia-dataset>

APPROXIMATE ANALYTICAL SOLUTION OF COUNTER-CURRENT IMBIBITION PHENOMENON USING HOMOTOPY PERTURBATION METHOD

¹Dhara T. Patel & ²Amit K. Parikh

¹Research Scholar, ²Principal, Mehsana Urban Science Institute of sciences, Ganpat University

^{1,2}Mathematics Department

^{1,2}Ganpat University, Mehsana, Gujarat, India

ABSTRACT: In this paper, we have discussed counter current imbibition phenomenon arising in inclined homogeneous porous medium during secondary oil recovery process. The approximate analytical solution of governing nonlinear partial differential equation is obtained in the form of infinite series using Homotopy Perturbation Method. The numerical and graphical solution have been obtained using MATLAB.

Keywords: Counter current imbibition phenomenon, Inclined homogeneous porous medium, Homotopy Perturbation method.

I. INTRODUCTION

Imbibition processes play an important role in fluid movement in plants and soil and one of the most important phenomenon to extract oil during secondary oil recovery process in reservoir.

When a porous medium filled with some fluid, is brought in contact with another fluid which preferentially wets the medium then there is a spontaneous flow of the wetting fluid into the medium and a counter flow of the native fluid from the medium. This phenomenon arising near to the common interface due to difference in wetting abilities of the fluids flowing in the medium is called counter current imbibition phenomenon.

Imbibition phenomenon has been investigated by various authors from different viewpoints analytically and numerically such as Brownscombe and Dyes (1952), Graham and Richardson (1959), Scheidegger (1960), Blair (1964), Verma (1969), Mehta and Verma (1977). P. M. Blair (1964) has derived numerical solution of the imbibition of water and the counter current flow of oil in porous rocks. J. Bourblaux, F. J. Kalaydjian (1990) have talked through experimental study of co-current and counter current flows in natural porous media. M. N. Mehta (1977) has described analytically the phenomenon of imbibition in porous media using a singular perturbation method. M. Pooladi-Darvish, A. Firoozabadi (2000) have discussed the similarities and differences of co-current and counter current imbibition and pointed out the consequences for practical applications. S. Yadav, M. N. Mehta (2009) have derived the mathematical model and similarity solution of counter current imbibition phenomenon in banded porous media. M. S. Joshi, N. B. Desai, M. N. Mehta (2011) have discussed an analytical solution of the counter current imbibition phenomenon in homogeneous porous media. A. K. Parikh, M. N. Mehta, V. H. Pradhan (2013) have described the counter current imbibition in vertical downward homogeneous porous media.

In this paper, we have discussed physical phenomenon of counter current imbibition in two phase (oil-water) flow of immiscible fluids in a homogeneous porous media with inclination. During secondary oil recovery process, when water is injected in inclined oil formatted homogeneous porous media, at the common interface ($x = 0$) the counter current imbibition occurs. The objective of this paper is to measure the saturation of water at any distance x and any time t in homogeneous porous media at particular angle of inclination θ . The nonlinear partial differential equation has been obtained for the counter current imbibition phenomenon, and solved by Homotopy Perturbation Method using appropriate initial condition.

II. STATEMENT OF PROBLEM

For the sake of mathematical study, we consider one-dimensional counter current imbibition phenomenon for which we choose a piece of cylindrical porous matrix from large natural field area and take vertical cross-section of this small finite incline cylindrical porous matrix as a rectangle which is inclined at small angle with x-axis and its three sides are impermeable. Since the water is injected in oil formatted area

and water and oil are flowing through porous matrix then it will satisfy Darcy's law for low Reynolds number.

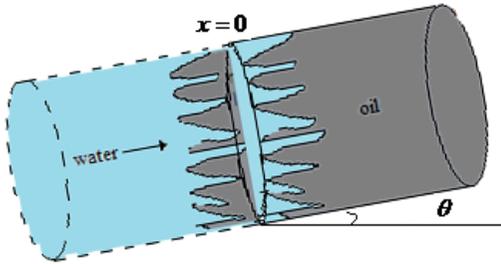


Figure-1

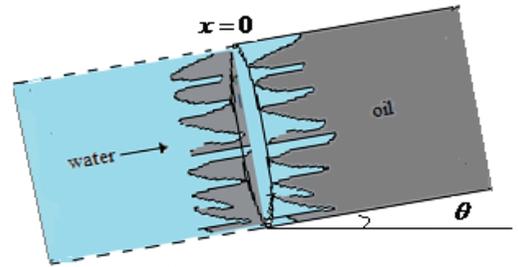


Figure-2

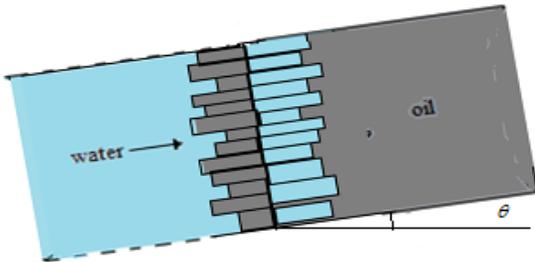


Figure-3

III. MATHEMATICAL FORMULATION

According to generalized Darcy's law (J. Bear. 1972) the seepage velocities of injected water V_w and native oil V_o are expressed as

$$V_w = -K \left(\frac{K_w}{\mu_w} \right) \left(\frac{\partial P_w}{\partial x} + \rho_w g \sin \theta \right) \quad (1)$$

$$V_o = -K \left(\frac{K_o}{\mu_o} \right) \left(\frac{\partial P_o}{\partial x} - \rho_o g \sin \theta \right) \quad (2)$$

In above equations K_w and K_o are the relative perm abilities which are function of water saturation .The permeability K is constant, the porous medium being homogeneous. P_w, μ_w, ρ_w are the pressure, viscosity and density of water respectively whereas P_o, μ_o, ρ_o are the pressure, viscosity and density of oil respectively. θ is the angle of inclination of the considered homogeneous porous medium.

The equation of continuity of injected water is

$$P \left(\frac{\partial S_w}{\partial t} \right) + \left(\frac{\partial V_w}{\partial x} \right) = 0 \quad (3)$$

Where P is porosity of the homogeneous porous medium.

It is well known fact that sum of saturation of water and oil is fully saturated (Unity). Hence we can define for fluid saturation of water and oil as

$$S_w + S_o = 1 \quad (4)$$

The fluid can flow through interconnected capillaries only due to capillary pressure which is defined as the pressure difference of the flowing fluid across their common interface. It may be written as

$$P_c = P_o - P_w \quad (5)$$

M.N.Mehta(1977) expressed the linear relationship between capillary pressure (P_c) and saturation of injected water (S_w) as

$$P_c = -\beta S_w \tag{6}$$

We assume that standard relationship due to A.E.Scheidegger, E.F. Johnson(1961), between phase saturation and relative permeability as

$$K_w = S_w, K_o = 1 - \alpha S_w \quad (\alpha = 1.11) \tag{7}$$

For more definiteness, we choose $\alpha \approx 1$ then

$$K_o \approx 1 - S_w = S_o \quad (\because S_w + S_o = 1) \tag{8}$$

According to Scheidegger A. E(1960).In countercurrent imbibition phenomenon, the sum of the velocities of injected water and native oil is zero. that is

$$V_w + V_o = 0 \tag{9}$$

Using equation (1) and (2) into equation (9),

$$\left(\frac{K_w}{\mu_w} K\right) \left(\frac{\partial P_w}{\partial x} + \rho_w g \sin \theta\right) + \left(\frac{K_o}{\mu_o} K\right) \left(\frac{\partial P_o}{\partial x} - \rho_o g \sin \theta\right) = 0 \tag{10}$$

From equations (10) and (5)

$$\left(\frac{K_w}{\mu_w} + \frac{K_o}{\mu_o}\right) \frac{\partial P_w}{\partial x} + \frac{K_o}{\mu_o} \frac{\partial P_c}{\partial x} = -\left(\frac{K_w}{\mu_w} \rho_w - \frac{K_o}{\mu_o} \rho_o\right) g \sin \theta \tag{11}$$

Using equation (11) into equation (1), we get

$$V_w = -\frac{\frac{K_w}{\mu_w} \frac{K_o}{\mu_o}}{\frac{K_w}{\mu_w} + \frac{K_o}{\mu_o}} K \left[(\rho_o + \rho_w) g \sin \theta - \frac{\partial P_c}{\partial x} \right] \tag{12}$$

On substituting the value of V_w from equation (12) into continuity equation (3), we get

$$P \left(\frac{\partial S_w}{\partial t}\right) - \frac{\partial}{\partial x} \left[\frac{\frac{K_w}{\mu_w} \frac{K_o}{\mu_o}}{\frac{K_w}{\mu_w} + \frac{K_o}{\mu_o}} K (\rho_o + \rho_w) g \sin \theta \right] + \frac{\partial}{\partial x} \left[K \frac{\frac{K_w}{\mu_w} \frac{K_o}{\mu_o}}{\frac{K_w}{\mu_w} + \frac{K_o}{\mu_o}} \frac{\partial P_c}{\partial x} \right] = 0 \tag{13}$$

For the investigation flow system involves water and viscous oil, therefore according to Scheidegger A. E (1960). we have

$$\frac{\frac{K_w}{\mu_w} \frac{K_o}{\mu_o}}{\frac{K_w}{\mu_w} + \frac{K_o}{\mu_o}} \approx \frac{K_o}{\mu_o} \tag{14}$$

On substituting values from equation (14), (8) and (6) into equation (13), we get

$$P \left(\frac{\partial S_w}{\partial t}\right) = \frac{K(\rho_w + \rho_o) g \sin \theta}{\mu_o} \frac{\partial(1 - S_w)}{\partial x} + \frac{K\beta}{\mu_o} \frac{\partial}{\partial x} \left[(1 - S_w) \frac{\partial S_w}{\partial x} \right] \tag{15}$$

Which is non-linear partial differential equation for the saturation of injected water during counter-current imbibition phenomena in incline homogeneous porous media.

Using dimensionless variables

$$T = \frac{K\beta}{L^2 P \mu_o} t, \quad X = \frac{x}{L}, \quad 0 \leq X \leq 1, \quad 0 \leq T \leq 1$$

And replacing $1 - S_w$ by S_o

$$\frac{\partial S_o}{\partial T} = S_o \frac{\partial^2 S}{\partial X^2} + \left(\frac{\partial S_o}{\partial X}\right)^2 - A \left(\frac{\partial S_o}{\partial X}\right) \tag{16}$$

Where $A = \frac{L(\rho_w + \rho_o) g \sin \theta}{\beta}$ is constant.

IV. APPROXIMATE ANALYTICAL SOLUTION

We will obtain the approximate analytical solution of (16) subject to initial condition $S_o(X,0) = 1 - Be^X$ by Homotopy perturbation method. First we construct homotopy of equation (16) as

$$\phi(u, p) = (1 - p) \left[\frac{\partial u}{\partial T} - \frac{\partial v_0}{\partial T} \right] + p \left[\frac{\partial u}{\partial T} - u \frac{\partial^2 u}{\partial X^2} - \left(\frac{\partial u}{\partial X}\right)^2 + A \frac{\partial u}{\partial X} \right] = 0 \tag{17}$$

where $p \in [0,1]$ is an embedding parameter. Choose $v_0(X, T) = 0$

So the equation (17) is written as,

$$\frac{\partial u}{\partial T} = p \left[u \frac{\partial^2 u}{\partial X^2} + \left(\frac{\partial u}{\partial X}\right)^2 - A \frac{\partial u}{\partial X} \right] = 0 \tag{18}$$

According to J.H.He (1999), solution of equation (18) can be expanded into infinite series as

$$u = u_0 + pu_1 + p^2u_2 + \dots \tag{19}$$

$$\frac{\partial(u_0 + pu_1 + p^2u_2 + \dots)}{\partial T} = p \left[\frac{\partial^2(u_0 + pu_1 + p^2u_2 + \dots)}{\partial X^2} + \left(\frac{\partial(u_0 + pu_1 + p^2u_2 + \dots)}{\partial X}\right)^2 - A \frac{\partial(u_0 + pu_1 + p^2u_2 + \dots)}{\partial X} \right] = 0$$

Equating coefficient of powers of p on both sides we get,

$$p^0 : \frac{\partial u_0}{\partial T} = 0, \quad u_0(X,0) = 1 - Be^X$$

$$p^1 : \frac{\partial u_1}{\partial T} = u_0 \frac{\partial^2 u_0}{\partial X^2} + \left(\frac{\partial u_0}{\partial X}\right)^2 - A \frac{\partial u_0}{\partial X}, \quad u_1(X,0) = 0$$

$$p^2 : \frac{\partial u_2}{\partial T} = u_0 \frac{\partial^2 u_1}{\partial X^2} + u_1 \frac{\partial^2 u_0}{\partial X^2} + 2 \left(\frac{\partial u_0}{\partial X}\right) \left(\frac{\partial u_1}{\partial X}\right) - A \frac{\partial u_1}{\partial X}, \quad u_2(X,0) = 0$$

And so on

On solving above equations, we can get

$$u_0 = 1 - Be^X$$

$$u_1 = (2B^2e^{2X} - Be^X + ABe^X)T$$

$$u_2 = (12B^2e^{2X} - Be^X + 2ABe^X - 18B^3e^{2X} - 8AB^2e^{2X} - A^2Be^X) \left(\frac{T^2}{2!}\right)$$

Substituting all these values into equation (19), and taking $p = 1$, we get

$$\begin{aligned} u &= u_0 + u_1 + u_2 + \dots \\ &= 1 - Be^X + (2B^2e^{2X} - Be^X + ABe^X)T \\ &\quad + (12B^2e^{2X} - Be^X + 2ABe^X - 18B^3e^{2X} - 8AB^2e^{2X} - A^2Be^X) \left(\frac{T^2}{2!}\right) + \dots \end{aligned} \tag{20}$$

Using equation (20), we can get solution of equation (16) as

$$\begin{aligned}
 S_w(X, T) &= 1 - S_o(X, T) \\
 &= Be^X - (2B^2e^{2X} - Be^X + ABe^X) \\
 &\quad - (12B^2e^{2X} - Be^X + 2ABe^X - 18B^3e^{2X} - 8AB^2e^{2X} - A^2Be^X) \left(\frac{T^2}{2!}\right) - \dots
 \end{aligned}
 \tag{21}$$

This is the required solution of governing equation (13) of the counter-current imbibition phenomenon which represents the saturation of injected water occupied by the schematic fingers of average length at any distance X and for any fixed time T .

V. RESULT AND DISCUSSION

Here for numerical calculation we consider the following values for constants which is taken from standard literature as follow

$$L = 1, g = 9.8, \rho_w = 0.3, \rho_o = 0.1, \beta = 0.1, \theta = 5^\circ \Rightarrow A \approx 0.034 \text{ and for } \theta = 45^\circ \Rightarrow A \approx 0.28$$

We select $B=0.05$ into equation (21).

Numerical and graphical presentation of solution (21) have been obtained by using MATLAB coding.

Figure (4) shows the graphs of S_w Vs. X for fixed time $T = 0.1, 0.2, 0.3, 0.4, \dots, 0.9$ and $\theta = 5^\circ$.

Table 1: Saturation of injected water (S_w) for different X for fixed time T at angle $\theta = 5^\circ$

X	T=0	T=0.1	T=0.2	T=0.3	T=0.4	T=0.5	T=0.6	T=0.7	T=0.8	T=0.9
X=0	0.05	0.0544	0.0591	0.0639	0.0689	0.0741	0.0795	0.0851	0.0909	0.0969
X=0.1	0.0553	0.0601	0.0651	0.0703	0.0757	0.0812	0.087	0.093	0.0991	0.1054
X=0.2	0.0611	0.0663	0.0717	0.0773	0.0831	0.089	0.0951	0.1014	0.1078	0.1145
X=0.3	0.0675	0.0732	0.079	0.085	0.0911	0.0974	0.1039	0.1105	0.1172	0.1241
X=0.4	0.0746	0.0807	0.087	0.0934	0.0999	0.1065	0.1133	0.1201	0.1271	0.1342
X=0.5	0.0824	0.0891	0.0958	0.1026	0.1094	0.1164	0.1234	0.1305	0.1376	0.1448
X=0.6	0.0911	0.0983	0.1054	0.1126	0.1198	0.127	0.1342	0.1414	0.1486	0.1559
X=0.7	0.1007	0.1084	0.116	0.1235	0.131	0.1384	0.1457	0.153	0.1602	0.1673
X=0.8	0.1113	0.1195	0.1275	0.1353	0.143	0.1506	0.1579	0.1651	0.1722	0.179
X=0.9	0.123	0.1317	0.1401	0.1482	0.1561	0.1636	0.1709	0.1778	0.1845	0.1909
X=0.10	0.1359	0.1451	0.1539	0.1622	0.1701	0.1775	0.1845	0.191	0.1971	0.2028

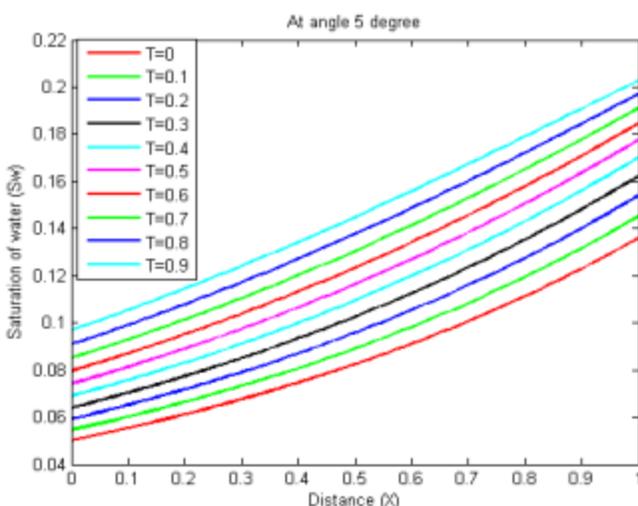


Figure (4): Graph of Saturation of water (S_w) Vs. Distance X for a fixed time T at angle $\theta = 5^\circ$

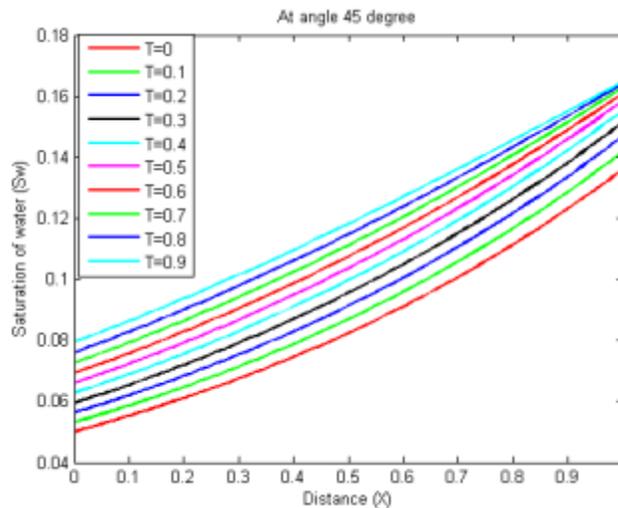


Figure (5): Graph of Saturation of water (S_w) Vs. Distance X for a fixed time T at angle $\theta = 45^\circ$

VI. CONCLUSION

In present paper the governing non-linear equation representing counter current imbibition phenomenon in homogeneous porous media with inclination is solved using Homotopy Perturbation method. It was observed that the saturation of water increases with time and increases with space variable in inclined homogeneous porous media. As value of angle θ is increase, saturation of water will decrease.

VII. REFERENCES

1. Scheidegger A. E. 1960. The physics of flow through Porous Media, University of Toronto Press.
2. P. M. Blair. 1964. Calculation of oil displacement by countercurrent water imbibition, Society of Petroleum Engineers Journal, Vol. 4(3), 195–202.
3. A. P. Verma. 1969. Statistical behaviour of fingering in a displacement process in heterogeneous porous medium with capillary pressure, Canadian J. Physics, Vol. 47, Issue 3, 319–324.
4. E.R. Brownscombe, A.B. Dyes. 1952. A Water imbibition Displacement - A Possibility for the Spraberry, Drilling and Production Practice, American Petroleum Institute, 383–390.
5. J.W. Graham, J.G. Richardson. 1959. Theory and application of imbibition phenomena in recovery of oil, Journal of Petroleum Technology, Society of Petroleum Engineers, Vol. 11(2), 65–69.
6. M. N. Mehta, A. P. Verma. 1977. A Singular Perturbation Solution of Double Phase Flow due to differential wettability, Indian J. Pure Appl. Math., Vol. 8(5), 523–526.
7. J. Bourblaux, F. J. Kalaydjian. 1990. Experimental Study of Cocurrent and Countercurrent Flows in Naturally Porous Media, SPE Reservoir Engineering, Society of Petroleum Engineers, Vol. 5(3), 361–368.
8. M. N. Mehta. 1977. Asymptotic expansions of fluid flow through porous media, Ph.D. Thesis, South Gujarat University, Surat.
9. M. Pooladi-Darvish, A. Firoozabadi. 2000. Cocurrent and countercurrent imbibition in a water wet matrix block, Society of Petroleum Engineers Journal, Vol. 5(1), 3–11.
10. S. Yadav, M. N. Mehta. 2009. Mathematical model and similarity solution of Countercurrent imbibition phenomenon in banded Porous matrix, Int. J. of Appl. Math. And Mech., Vol. 5(5), 76–86.
11. M. S. Joshi, N. B. Desai, M. N. Mehta. 2011. An analytical solution of Countercurrent imbibition phenomenon arising in fluid flow through homogeneous porous media, Int. J. of Appl. Math. and Mech., Vol. 7 (11), 21–32.
12. A. K. Parikh, M. N. Mehta, V. H. Pradhan. 2013. Mathematical model and analysis of countercurrent imbibition in vertical downward homogeneous porous media, British J. of Math. and Comp. Sci., Vol. 3(4), 478–489.
13. Darcy H. 1856. Les Fontaines publiques de la ville de Dijon, Dalmont, Paris.
14. J. Bear. 1972. Dynamics of fluids in porous media, American Elsevier Publishing Company, Inc.
15. A. E. Scheidegger. 1958. The Physics of flow through porous media, Soil Science, Vol. 86.
16. M. A. Patel, N. B. Desai. 2016. Homotopy analysis solution of counter-current imbibition phenomenon in inclined homogeneous porous media.
17. A. E. Scheidegger, E. F. Johnson. 1961. The statistically behavior of instabilities in displacement process in porous media, Canadian J. Physics, Vol. 39, Issue 2, 326–334.
18. J. H. He. 1999. Homotopy perturbation technique, comp. Meth. Appl. Mech. Eng., 178, 257–262.

Digital Game-Based Learning (DGBL): Design and Development of an Educational Game

Design of Digital Games as Pedagogical Tools

¹Dr. Nazneen Ansari, ²Siddhesh Yadav, ³Abhijeet Prabhu, ⁴Vidhi Shah, ⁵Anuj Potdar

¹Associate Professor, ^{2,3,4,5}B.E. Student

^{1,2,3,4,5}Department of Information Technology

^{1,2,3,4,5}St. Francis Institute of Technology, Mumbai, India

ABSTRACT: Digital game-based learning (DGBL) assimilates educational content into video games with the goal of engaging learners. This paper examines the limitations of the educational system towards the accomplishment of learning goals and emphasizes the need of a tool that engages and educates the learner. It describes the game design model and aims to measure its efficacy and effectiveness in achieving its goals in terms of the enthusiasm and active participation the players feel during gameplay, as well as the post-gameplay satisfaction with newly acquired knowledge and skills. This model based on flow theory and game design, stresses the importance of providing the player with immediate feedback, clear objectives and challenges. This paper will introduce the concept of such a game and the various learning processes that take place during each phase. The flow theory framework is used in order to maximize the impact of player's experience.

Keywords: Educational Games, Digital Game Based Learning, Game Design, Flow theory

I. Introduction

Traditionally, most classrooms make use of textbooks, charts, pictures and diagrams for the purpose of conveying information to the students. This has resulted in the emergence of a 'rote' learning culture [1]. Subsequently, the students memorize and reproduce the material in the examinations and are graded accordingly, regardless of their comprehension or understand. In this manner, they 'cheat' the system by not putting in long-term efforts, and suffer the consequences due to an inability to apply the knowledge they never truly attempted to learn [1, 2]. This culture is most prevalent in subjects like History, wherein remembering key facts (in form of dates and personalities) is of primary importance. The understanding of the significance and the impact of these events on the modern world events is considered optional and secondary, as the student is never tested on those.

This problem can be solved by using an unorthodox approach to education. Students do not find the earlier mentioned methods interesting enough and do not understand the purpose behind the inclusion of a specific topic in their curriculum. Grades alone as an incentive are insufficient to motivate them when there is the more expedient alternative of procrastinating, memorizing it a fortnight before the examinations and still managing to get good grades.

This paper recognizes the tremendous potential games have and presents a design model which utilizes the popularity of these games to exhibit educational content and provide an optimal learning experience for its players using the Digital Game Based Learning (DGBL) paradigm [3]. Section II explores the instructional method to be used within the game to optimize the learning process and how each phase of the process will be implemented. Section III discusses the flow theory, and the factors which can induce such a state for the players to make the most of the game. Section IV discusses the included in-game elements that encourage gameplay. Section V proposes a game to be built on the existing model. Section VI identifies the scope of the game with respect to its players, its content and its implementation. Section VII discusses the work done on the game, and subsequent future work to be carried out.

II. Instruction for Optimal Learning

Before designing the model, it is important to understand the proper instruction techniques that elicit the best performance from the students. According to Gagne (1996), the best sequence of events most instructions should follow to elicit optimal performance is.

Table 2.1: Gagne’s Events of Instruction and the internal learning process that it supports [4].

Internal Process	Instructional Event	Actions taken in Game
Reception	1.Gaining attention	Use of graphics & gameplay
Expectancy	2.Inform objectives	Inform players the game objectives & challenges
Retrieval to memory	3.Stimulate recall of prior learning	Recall learned skills from past levels.
Selective perception	4.Present content	Animations demonstrating educational content
Semantic encoding	5.Provide learning guidance	Game elements that guide the player.
Responding	6.Elicit performance	Ask learner to perform objectives
Reinforcement	7.Provide Feedback	Give informative feedback as per gameplay
Reinforcement and retrieval	8.Assess performance	Additional Feedback on performance at end of level.
Retrieval & Generalization	9.Enhance retention and transfer knowledge	Try the level again with varied approach

III. Flow Theory

Csikszentmihalyi M. (1991) has defined the ‘flow’ state as the one in which the individual experiences complete absorption and total involvement in their undertaken activity. It refers to the optimal psychological state where the goal driven activity is the only thing that matters to the person involved [5]. Past research indicates the positive effects of flow state on learning [6] and this should be taken into account while designing games that may act as learning materials. The game should be designed in such a manner, that the flow state is sustained, so that the player has a satisfying experience throughout gameplay.

Finneran and Zhang (2003) have asserted that the activities performed in computer-generated environments need to be decomposed into the main task and the artifact used to accomplish those activities [7]. They have also proposed a person-artifact-task (PAT) model that encapsulates the major components of a person involved in any computer-related activities. According to the model, the likelihood of ‘flow’experience is dependent on the interplay between the person, task and artifact [7].

In studies, the distinctive stages associated with attaining this state are: flow antecedents, flow experience and flow consequences [8]. The antecedents of flow include focused attention [8], a clear set of goals, immediate and appropriate feedback [8], potential control [7], a perception of challenges that are matched to the person’s skills [8], playfulness [6], speed and ease of use. The flow experience itself comprises a merging of action and awareness, concentration, a sense of control over activity [8], time distortion, and telepresence [7].The flow experience leads to increased learning, increased exploratory behavior [6], positive effect and perceived behavioral control.

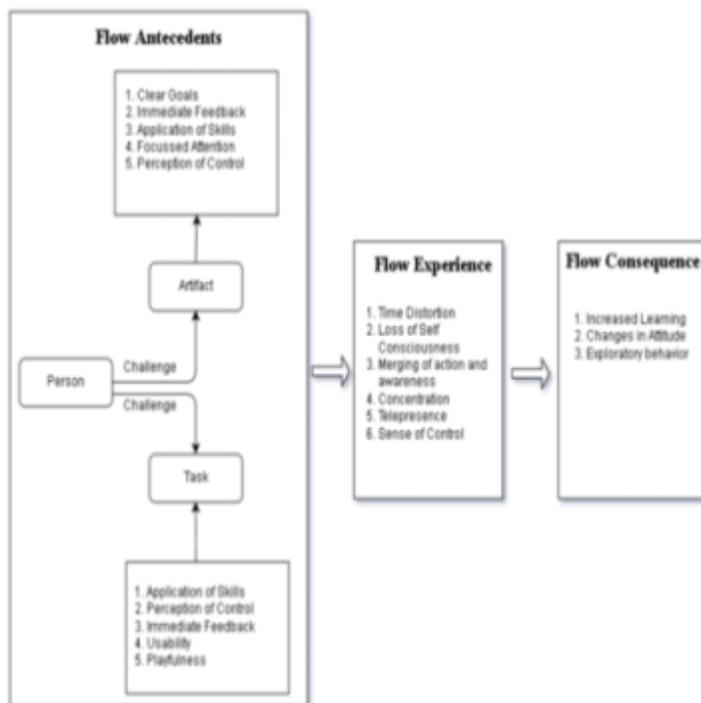


Figure 1: Framework of Flow in Computer Generated Environments.

In Fig. 1, the flow framework is presented. The framework comprises of the factors related to each stage and its respective components of the PAT model. The antecedents, speed and ease of use, are combined as a usability factor in the framework.

IV. Game Elements

Game elements are the components of a game that make it entertaining and interesting.

Basic elements of a game:

1. Identity: Player assumes the identity of a person or a thing who is acting out a task defined in the game. The player can control the character to either move around or make use of the actuators defined for the identity.
2. Space: Every game has the world in which it exists which is called its space. This is the foundation of every game as every rule or constraint is levied on the space [9]. It includes elements like Visual Space, Barriers, Colors, Sounds, and Lighting etc.
3. Goal: The goal of every game can be different depending on how the player will win the game and what will mark the end of the current level or the scenario in which the player is playing [9].
4. Components: The characters and the objects in the game are the components of the game. They exist within the space of the game [9].
5. Mechanics: These are the actions that characters and other components can perform, or have done to them, throughout or during a certain part of the game. This element shows all of the constraints put on the characters and helps create a difficulty level for the game [9].
6. Rules: Rules guide a player through the game and it tells them what they are supposed to do in order to win or what actions determine failure. Along with providing the information of how to play the game they also have a huge upper-hand in determining the over-all gameplay experience of the player [9].

Elements that enhance the gameplay experience:

1. Multiplayer: Multiplayer means that two or more players can play the same running game instance such that the actions of one player influence the space of the others as well. This can be done either on a single machine (offline) or two or more machines connected to a common network (online). The two broad multiplayer categories include competitive and collaborative gaming [3].

2. Collectibles: Collectibles encourage the players to spend more time at game exploration. A collection of a certain number of items resulting in a reward is the most prominent technique used in modern gaming [3].
3. Easter-eggs: Easter-eggs mean providing a reference to either a real world component or a component that does not belong to that space or that game. Their sole purpose is to arouse curiosity in entertainment genre and must be avoided in educational games [3].
4. Exploration/Discovery: Free-roam is one of the most appreciated features of modern games where players can explore the space of the game without the constraint of a single goal but with an intention to discover something new within the game. This is also very useful from a developer point of view to increase the scope of the game without digressing from main topic [3].
5. Strategy: A strategy is a complete algorithm for playing the game, telling a player what is to be done for every possible situation throughout the game. The strategy undertaken will determine the action the player takes at any stage of the game [10].
6. Score: The number of points, goals, runs, etc. achieved in a game by a team or an individual [11]. These points are usually based on Experience points (XP) system wherein the more the player plays the game, the bigger and more enticing the rewards he/she receives in terms of XP.
7. Achievements: Achievements being a great way to track player activity and encourage increased gameplay represents a quantitative goal or accomplishment. When the player meets or exceeds the goal, the achievement is considered earned, and the player is rewarded. The game rules define the goal and the game mechanics that describe how the achievement is earned. An achievement earns a player XP points. A player can see the total number of points that can potentially be earned in the game as well as how many points he or she has currently earned. The Game Center app allows players to compare their achievements with those of friends. The game can be designed in a way that the game mechanics allow the player to make progress towards completing the achievement [11].
8. Leaderboards: Many games offer scoring systems that measure how well a player does in the game. Scores are not just a way for players to measure their own progress; they also provide a way for players to compare their skills with those of other players. In Game Center, a leaderboard is a database of score data. Your game posts score to a leaderboard so that the player can later view those scores. When you add leaderboards to your game, you define what a score means in your game and how score data is formatted for display [11].

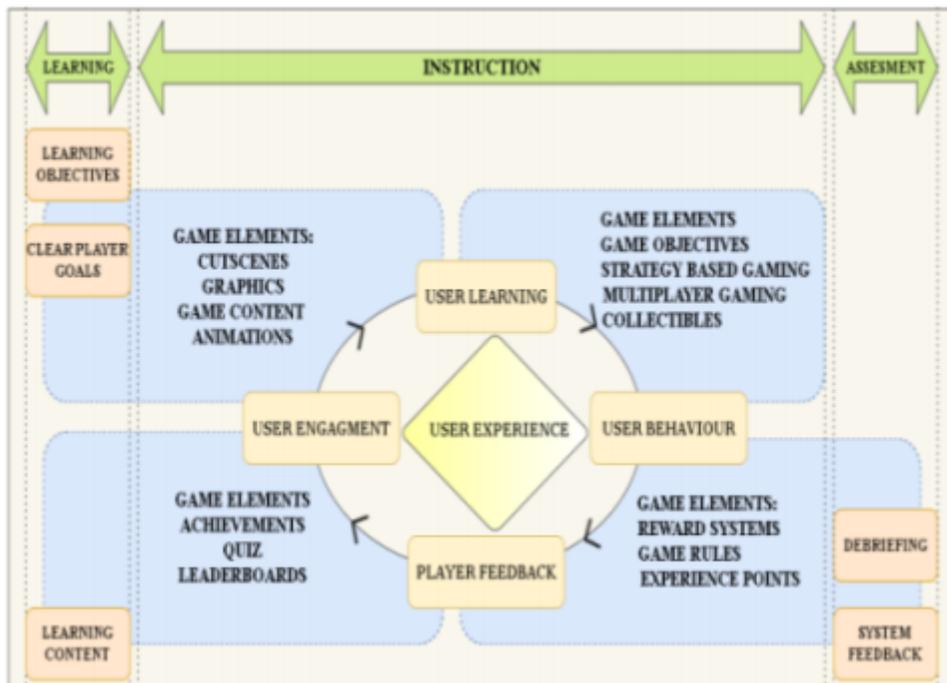


Figure 2: Proposed Model

V. Proposed Work

Based on this game design model, we have commenced game development in order to measure the effectiveness with which the model achieves its learning goals. These goals will be measured in terms of the enthusiasm and active participation the players feel during gameplay, as well as the post-gameplay satisfaction with their newly acquired knowledge and skills.

We decided to create a game depicting significant events of Indian History and test its effectiveness on students learning that event in their syllabus. The game follows a skill and drill format where the player will practice and repeat skills in order to better comprehend and retain momentous historical events. Every level shall be divided into three phases as explained below:

- Phase 1: Cutscenes

The level begins with a cut-scene that will depict the actual historic incident using various animations. The 3D models in the cutscene will help the visual learner and the sound and music will benefit the auditory learner by grabbing their attention and acquainting them with the storyline of the level.

- Phase 2: Gameplay

Now that the player is acquainted with the historic context of the level, he/she will play the lead character in the episode and defeat its opponents. The opponent and its army will be controlled using AI. The player must follow all the instructions accurately and must complete the level within the specified time, failing which he must replay the level to proceed further. The scoring system will record the players.

- Phase 3: Quiz

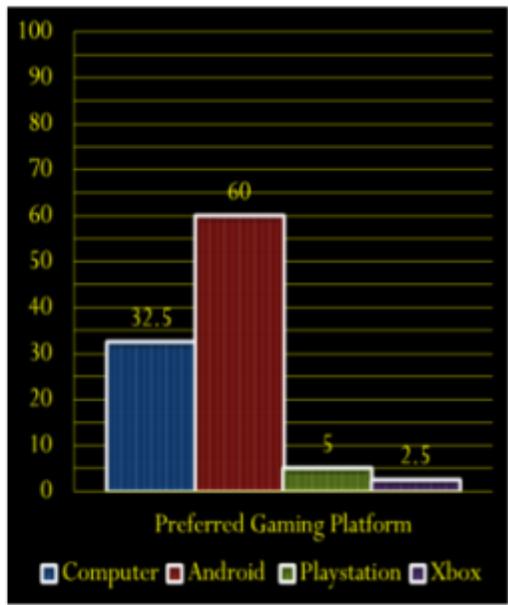
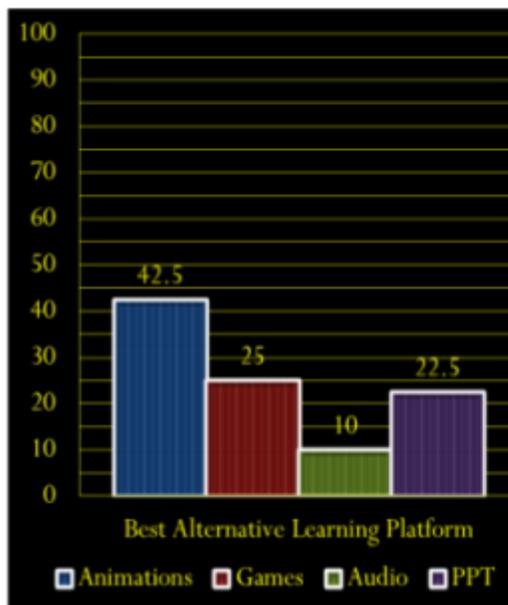
In this phase, the student must take the quiz based on earlier gameplay and cutscenes which tests comprehension and retention. Student must correctly answer at least 8 out of 10 questions to unlock the next level, failing which the cutscenes are replayed for the student to learn from his/her mistakes and try again. This tests the students understanding and acts as feedback to how effective the game has been as a learning tool.

VI. Scope of Proposed Work

In order to decide the scope of this game, we conducted surveys with diverse age groups to decide upon the most convenient and accessible platform for comfortable gameplay and seamless propagation of education content. We surveyed the students of St Francis D’Assisi High School belonging to age group 10 to 15 and collected the following data on their preferred gaming and learning platforms.

The questionnaire for it included following two questions:

1. Amongst which of the following ways do you prefer for learning history?
2. Which of the following is your most used device for gaming?



The results of the survey indicated that majority of student learners preferred Animations (42.5%) and Games (25%) as an alternative learning tool on Windows (32.5%) & Android platforms (60%).

VII. RESULTS AND FUTURE WORK

A Unity3d cross-platform Third Person Perspective game named 'Legends of Golden Sparrow' has been developed in order to demonstrate significant Indian Historical events. The meeting of Maratha King Shivaji and Adilshahi General Afzal Khan and the subsequent Battle of Pratapgarh has been dramatized. Once we finish developing this game, it will be tested for its effectiveness on the students we surveyed.



Figure 3: GUI



Figure 4: Shivaji and troops fighting Afzal Khan.



Figure 5: Fort where the event took place.

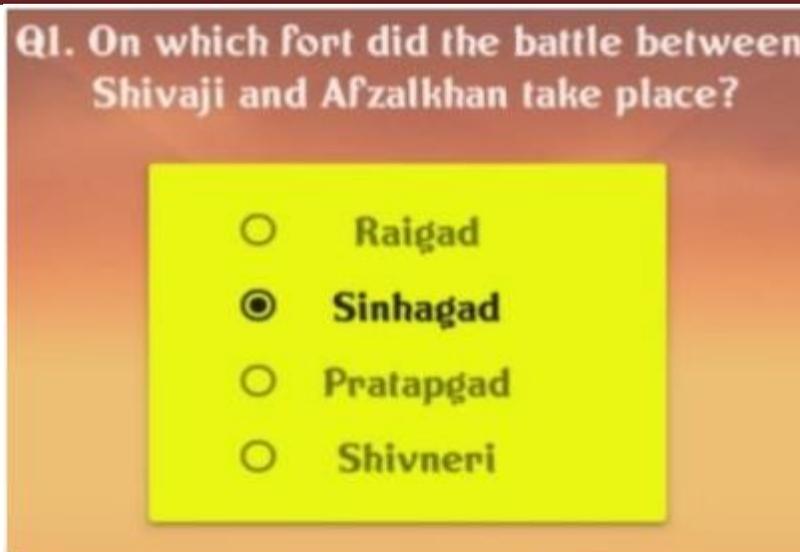


Figure 6: Quiz at end of level.

Figure 3 is a screenshot of the GUI of the game, the first screen the player encounters while running the game, the model on the right is Shivaji Maharaj the protagonist of the first level. An introductory cutscene will be shown which will introduce both the main character and the villain and the event about to take place. Figure 4 is a live action gameplay screenshot, where the player playing the role of Shivaji is aided by his troops in fighting Afzal Khan within the fort where the events took place. The fort can be explored to find collectibles and easter-eggs which will reward the player; the fort is shown in Figure 5. Figure 6 indicates the quiz at the end of the game which will test player understanding and retention.

This game will be a representation of the model that was developed, the actions within the game will be consistent with Gagne's instruction model, and the gameplay will be designed keeping the flow framework in mind making it easier for the players to have the flow experience until game completion. Future work will include addition of functionality, increased player interaction, increased gameplay content and testing the game on a bigger audience.

VIII. Conclusion

In this paper, we have discussed the problems that have plagued the educational system due to the rise of the 'rote' learning culture. We have utilized the popularity of games in order to address this problem and come up with a solution. In order to design such a model that provides the best experience in terms of education and gameplay, we first studied the instruction techniques used in educational settings that enable students to perform at their best; after that we understood the state of 'flow' and how it can be induced in players which would enable the players to perform their best in the game and derive the most out of its educational content.

We also incorporated various modern gameplay elements within the model to strike a healthy balance between an exciting game and a compelling immersive educational experience. Based on this model, we have also undertaken development of a game whose educational content, gameplay and platform were determined primarily by the players who would be playing the game. The structure of the game; its educational content was determined by the instruction techniques discussed earlier, while gameplay elements and to be used were decided by the data collected earlier. In the future, further development on this game will be carried out and the game will be available as a finished product to its targeted audience. And the feedback elicited by its players will be studied and reported in the future as a validation to the design model.

References

1. Rote learning an evil in education system, national survey reveals - Times of India, The Times of India, 2016. [Online]. Available <http://timesofindia.indiatimes.com/home/education/news/Rote-learning-an-evil-in-education-system-national-survey-reveals/articleshow/17564012.cms>. [Accessed: 28- Dec- 2016].
2. Sommer, W. G. (1990). Procrastination and cramming: How adept students ace the system. *Journal of American College Health*, 39(1), 5-10.

3. Kiili, K. (2005). Digital game-based learning: Towards an experiential gaming model. *The Internet and higher education*, 8(1), 13-24.
4. Gagné, R. M., & Medsker, K. L. (1996). *The conditions of learning: Training applications*.
5. Mihaly Csikszentmihalyi, *Flow: The psychology of optimal experience*, New York : Harper & Row, 1990.
6. Webster, J., Trevino, L. K., & Ryan, L. (1993). The dimensionality and correlates of flow in human-computer interactions. *Computers in human behavior*, 9(4), 411-426.
7. Finneran, C. M., & Zhang, P. (2003). A person-artefact-task (PAT) model of flow antecedents in computer-mediated environments. *International Journal of Human-Computer Studies*, 59(4), 475-496.
8. Chen, H., Wigand, R. T., & Nilan, M. S. (1999). Optimal experience of web activities. *Computers in human behavior*, 15(5), 585-608.
9. Hoffman, D. L., & Novak, T. P. (1996). Marketing in hypermedia computer-mediated environments: Conceptual foundations. *Journal of marketing*, 60(3), 50-68.
10. Aumann, R. J. (1985). What is game theory trying to accomplish?. In *Frontiers of Economics*, edited by K. Arrow and S. Honkapohja.
11. "Leaderboards, Achievements, and Challenges", Developer.apple.com, 2016.[Online].Available: https://developer.apple.com/library/content/documentation/NetworkingInternet/Conceptual/GameKit_Guide/Achievements/Achievements.html. [Accessed: 27- Dec- 2016].

SQL QUERY GENERATOR

¹Sanober Shaikh (Guide), ²Yogita Motwani, ³Sakshi Pandita, ⁴Simran Lund, ⁵Sanjana Rohra

¹Assistant Professor, ²student, ³student, ⁴student, ⁵student

¹I.T, ThadomalShahani Engineering College, Mumbai, Maharashtra, India.

²I.T, ThadomalShahani Engineering College, Mumbai, Maharashtra, India.

³I.T, ThadomalShahani Engineering College, Mumbai, Maharashtra, India.

⁴I.T, ThadomalShahani Engineering College, Mumbai, Maharashtra, India.

⁵I.T, ThadomalShahani Engineering College, Mumbai, Maharashtra, India.

ABSTRACT: SQL stands for Structured Query Language. It is a standardized programming language, deals with relational database and performs various operations with data such as update, delete, query, insert. All the RDBMS(Relational Database Management Systems) such as Oracle, MySQL use SQL as their standard database language.

Further, the SQL language is divided into clauses, expressions, queries, statements etc. SQL is a declarative, not an imperative programming language. It allows users to define and describe the data, create and drop tables. SQL has many more different functions.

SQL is important because it helps us to find the needed information or data easily, it is a query language, not a programming language, It is used to store and retrieve data from the database quickly. It is used for the query, insert, collect and manage data from the database.

Almost every database system will need SQL for further processing.

This paper provides an overview of a system which can be used for generating the SQL queries without having to manually write them. It can be used for training purpose as well as by anyone who wants to perform actions on their databases (anyone who don't have much knowledge of database queries). SQL queries are used for database manipulation and to perform different operation. There are different complex query syntax which needs to remember which is not an easy task.

The proposed system provides the GUI for query generation where user just needs to click on icons/buttons to generate the query. Here, no need to remember the syntax for different operations. User just need to click on particular button and then enter the required information.

Keywords: Database, SQL queries, query generator, tables, columns, tuples, application.

I. INTRODUCTION:

I.T. in Education (2019)², Query languages are used to generate the queries in a database, and Structured Query Language (SQL) is the standard which is used for it. Under the SQL query languages, there are different types of languages which is used for generating queries, including MySQL, Oracle SQL. Query languages for other databases, such as NOSQL and graph databases consists of Cassandra Query Language (CQL), Data Mining Extensions (DMX) and XQuery.

, SQL stands for Structured Query Language. SQL is used to communicate with a database and is the standard language for relational database management systems, according to ANSI. SQL commands are used to perform operations such as update data or retrieve data from a database.

The dynamic websites on the World Wide Web consists of large data and these can be handled by the databases. The SQL was developed in 1970s that is System R, new databases software in the IBM laboratories. And to manage the data in System R, the SQL language was used. At first it was called SEQUEL, but was later renamed to just SQL.

Mahesh Verma, VastuaAgarwal&TapanAcharaya (2017)³, In 1979, a company Relational Software also called as Oracle, created the modified version of SQL called as Oracle V2.

SQL provides great flexibility to users by providing support to distributed systems that is databases that can run on several computers at a time. It is certified by ANSI and ISO that SQL has become the standard for database query language and provides the basis for variety of databases applications on the Internet. It is used for both industry and academic levels and is used for both individual computers and corporate servers. With the invention in the database technology the SQL applications have become affordable for regular user.

This is due to the invention of various open-source SQL database applications such as MySQL, PostgreSQL, SQLite and many more.

Addison Wesley,(1997)⁴, SQL commands consists of DDL, DML, DCL and TCL commands:

DDL stands for Data Definition Language, and is used for creating, modifying and deleting databases, tables, columns etc. It includes commands such as CREATE, ALTER, RENAME, DROP etc.

DML stands for Data Manipulation Language, and it is used to create or manipulate the actual data. The CRUD operations of SQL can be performed by INSERT, SELECT, UPDATE and DELETE commands that comes in DML. *CRUD(create, read, update and delete) operations are the four basic functions of any persistent storage mechanisms.*

DCL stands for Data Control Language and includes commands such as GRANT, REVOKE etc.

TCL stands for Transaction Control Language and includes commands such as COMMIT, SAVEPOINT, ROLLBACK, etc.

Also other commands such as order by, having, group by, aggregate functions, relational algebra operators, etc can be used to get the required amount of data from the current databases.

II. REVIEW OF LITERATURE:

DataQuest (2018)¹, Edgar Frank "Ted" Codd^(superscript 1) (19 August 1923 – 18 April 2003) was an English computerscientist who, while working for IBM, invented the relational model for database management.

SQL was initially developed at IBM by Donald D. Chamberlin and Raymond F. Boyce after learning about the relational model from Ted Codd in the early 1970s

By 1986, ANSI and ISO standard groups officially adopted the standard "Database Language SQL" language definition. New versions of the standard were published in 1989, 1992, 1996, 1999, 2003, 2006, 2008, 2011 and most recently, 2016.

2.1 OBJECTIVES OF THE SYSTEM:

Provide the GUI for SQL query generation.

Quickly and easily generate SQL queries.

Help the developer in solving complex problems related to database.

Helps the developer in developing the backend of the application/system quickly/easily.

Helps the developer to know the result of their action on database without modifying actual database.

Helps in understanding/learning SQL queries at the basic level.

2.2 METHODOLOGY:

The data and information has been collected from secondary sources like, reputed journals,magazines, reference books and websites. Further, a survey was conducted in relation with database queries.

III. EXISTING SYSTEM:

3.1 RazorSQL:

*(Razor SQL)*⁵RazorSQL provides a SQL tool, SQL editor, and database administration tool for different operating systems.RazorSQL has been tested on various databases and can be connected to various databases via either JDBC or ODBC driver.

RazorSQL helps to make use of the relational database engine that is present inside it or not and does not requires end user administration. Some of the main functions that RazorSQL provides are tools for creating, editing, altering, dropping, and viewing database objects and is used for importing and exporting data. It provides programming editor which supports different programming languages.

As it provides support for various programming languages so provides limited commands for database manipulation. Also the GUI is not much user friendly and we cannot generate queries without connecting to databases.

3.2 GUI:

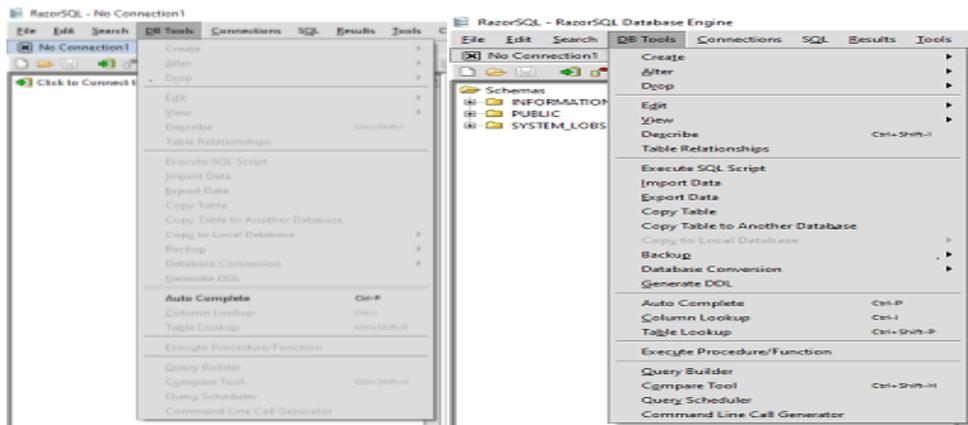


Fig. 1

Fig.1 shows, user needs to connect to database for generating queries whereas our system allows user to generate queries without having to connect to database and the operations available for manipulation are very less and basic whereas our system facilitates generation of many complex queries.

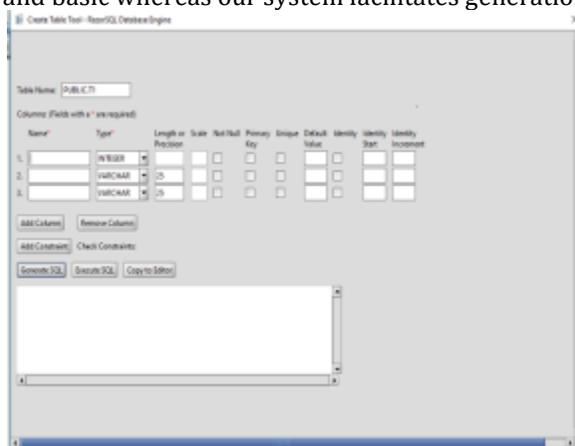


Fig. 2

Fig. 2 shows, options to select the values for creating tables.

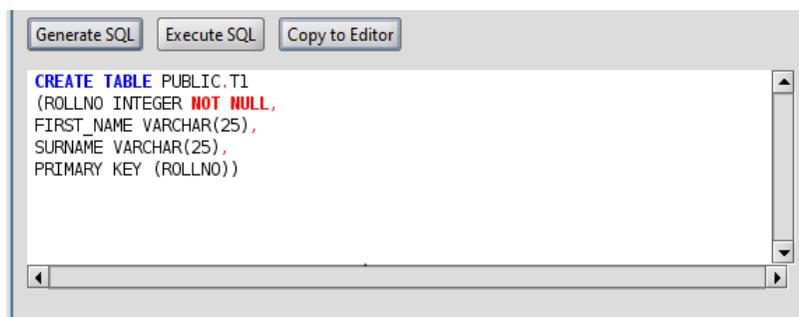


Fig. 3

Fig. 3 shows, there is no option of placing queries into the application or to text document.

IV. IMPORTANCE OF THE PROPOSED SYSTEM:

SQL Queries are important for development of any application or system. It is used at backend for data manipulation. Learning syntax of SQL queries may not make you a great developer or may not build skill to learn complex queries.

It is important for developer to have the knowledge of queries rather than the knowledge of syntax. The main advantages of our proposed system are as follows:-

4.1 GUI

It provides user friendly GUI which helps user to generate queries quickly.

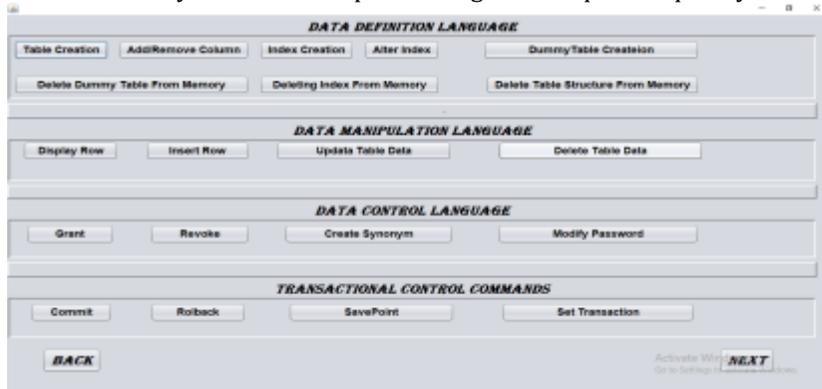


Fig. 4

Fig. 4 shows, the GUI which user gets for query generation. PL/SQL function generation is also possible with our system.

4.2 Data manipulation without connecting to database:

If for any security purpose user doesn't want to disclose his/her database then user can go with the first option that is QUERIES.

This section will allow the user to generate the queries without connecting to any database.



Fig. 5

Fig. 5 shows, System can also be used for learning purpose

If user doesn't wants to connect to their database then he/she can go with the first option. QUERIES section can be used for learning purpose as well.

4.3 Storing the generated queries:



Fig. 6

Fig. 6 shows, options to store the queries.

Here after the query has been generated the user gets 3 options:-

1. Copy query to plain document:

Here the query is placed in text document. The document may either already exists or it can be created at that time by that user depends on the user.

So that, if user wants to paste the query into the main application only after seeing the status of the database after performing different commands.

2. Copy query to my application:

Here the query is directly placed in the main application after seeing the output.

Here user after clicking on the option, user needs to browse its application from the window that will appear.

After the application has selected the query is placed at the end of the file.

3. Dismiss:

Here selecting this option will cancel the query that is no action is taken after generation of query.

V. CONCLUSION:

Our system provides GUI for SQL query generation and also allows us to operate on dummy databases, which helps the user to get to know the effects of various operations on the databases before applying them to their application.

Our system can also be used for learning purpose as well.

It can be used by trainers to train their trainees.

VI. ACKNOWLEDGMENT:

The authors would like to acknowledge our teacher/guide Ms. SanoberShaikh in order to provide the support and helps to carry out this research work in order to develop such type of the system. We are very thankful for their understanding and vision

REFERENCES:

1. DataQuest(2018),Decoding Digital:A cybermedia Publication vol XXXVI No 07.
2. I.T. in Education(2019):Express Computer Volume NO:30.
3. Mahesh Verma, VastuaAgarwal&TapanAcharaya (2017), Apps World.
4. "C. J. Date with Hugh Darwen: A Guide to the SQL standard : a user guide to the standard database language SQL, 4thed., Addison Wesley, USA 1997, ISBN 978-0-201-96426-4".
5. www.razorsql.com,"Richardson Software, LLC 543 Helene St. Gaithersburg, MD 20878."Razor SQL".

DR DEVICE- SERVICE PROVIDER FOR ALL DEVICES

Dr. Ashfaq Shaikh, Madiha Baig, Rubina Siddique, Hafia Ansari
Asst. Prof. IT Dept.at M. H SabooSiddik College, Student , Student , Student
Department of Information Technology
M. H. SabooSiddik of College Engineering, Mumbai, India

ABSTRACT: DR. DEVICE is an android application that provides repairing services to the customer of different mobile, tablet and laptops brand. The customer can request the services from their devices the request will be stored on server, the user will get the notifications from time to time once request is posted, once the problem is resolved then user will get the device on his place so our app provides the unique feature where user will not required to visit service centre. The server will get the number of request from various users and collect the devices with There will be two applications one for the customer who will request to take the service to repair his/her phone. And the other one for the vendor who will authenticate the service request from the user for repairing and will look after the number of days taken to repair the app and will look after all the details about the customer and will generate the cost. This app will basically reduce customers effort to go the service centre and repair their devices because the vendor will provide pick up and drop services from the given address with warranty of few months. It will also provide the customer with an option of replacement of its device given for repair for the number of days it will take to repair. With this feature the customer will not have to keep his important work at hold and can continue working. Also, there will be a number of accessories, mobiles, laptops and tablets of different brands for the user to purchase. There will also be live tracking when the device is being given to repair till it is returned to the customer.

Keywords: Android, Dr. Device, Smart phones, Receivers, Vendors, Service centers.

I. INTRODUCTION

Internet tends to be the backbone of all the technologies. The Doctor Device-Service Provider for All Devices is a progressive step in the field of service centres. Any android user can make use of such app to locate and communicate with the service centres in the vicinity. The proposed system can be used by any android user, in addition of getting services from the service centre this mobile application will also include online shopping for the various accessories or devices provided by different mobile or laptop companies. The very first task as the user install the app is the registration process, once the registration is completed the user has the liberty of using all the features. If the user is already registered, then he/she just have to login. After that the user have to book the service as per the need which means that whatever repair is to be done will be selected along with the device that the user has (wide range of devices). Most people are so busy in their tight schedule that they don't get the time to get their devices repaired. The targeted audience for this application is the corporate sector as well as the people who have their businesses solely dependent on technology. If their devices are hampered it may even lead to a lot of financial loss. This strategy of repairing service will get the increase profit in Business sector.

This Application will provide repairing facilities to customers who want their devices to get repaired and even get a backup device till the repair is done and then return it back to the service provider. Application provides a solution to this problem by making platform to repair and sell items through the Company Dr. Device. Our system is highly interactive and based on Android, so it is more responsive to user. There be different activities depending on the type of person that is using the services. Wide array of devices that are open for repairing and troubleshooting with easy UI design.

II. LITERATURE REVIEW

The rapid improvements and developments day by day in the field of Mobile Computing and medical sciences have created an outstanding scenario for the developments and enhancements of electronic-Health and mobile-Health, especially after the introduction of mobile phones and Smartphone in the market. Early mobile application solutions required enterprises to choose between cradle-based synchronization through a wire line network or a pure online wireless solution (PENTA group, 2010) [2].

However, Mobile applications are rapidly rising segments of the international mobile market, which consist the software runs on a mobile device and performs certain tasks for the clients. Because of the various

functions including the user interface for basic telephone and messaging service as well as advanced services, mobile Apps are widely used by customs. Also, mobile Apps are a large and continuously growing market and served by an increasing number of mobile App developers, publishers, and providers (the Mobile Marketing Association group, 2008) [1]. An increasing proportion of services are now electronic services delivered over the Internet. The survey regarding this application includes information gathering from various sources. These sources include various related web sites and similar projects developed previously. The different projects with the similar idea of our proposed system are given below:

- i. **Micromax Care** is your personal device assistant to help manage your Micromax product, in India. This app is a smart extension of your Micromax ownership experience. The features and services provided on this app are recommended and endorsed by Micromax. This app has three main important features
 - Doorstep pickup and drop service for repair requests
 - Completely paperless claims process
 - Completely transparent process with real-time updates on repair requests, through app and web
- ii. **Yaantra- Gadgetwood** in India's 1st professionally managed smartphones and tablets repair service. We can help you repair all brands of smartphones and tablets you own i.e. Accidental damages like broken screen, broken touch and display, damages in body, damages due to water, dead smartphone, broken USB or charging jack, non-responsive devices, dead devices, not only this you can also take help on most common software problems i.e. software update, hanging problem, slow performance, application download, removing unresponsive apps, data backup and restore, email setup and with everything that you may come across.[2]
- iii. **Helpbit** is your one stop solution to book services from trusted service providers in the **UAE** and **KSA** such as Electronics Repair (Mobile, Laptop, Tablet, etc), Home Services, House Cleaning, Moving, Home Appliances Repair, Plumbing, Automotive, Pet Care, Beauty Home Services and much more. In **Electronics**: Repair all kind of electronics at Helpbit repair centre from Mobiles, Laptops, Tablets, Game Consoles, TVs, Printers, Home Appliances, CCTV Installation. in addition to Windows & Mac Support, WIFI Support, Data Transfer and Brand Warranty. You can get your mobile repaired at your doorstep with Helpbit.
- iv. **One Plus Care** is also a mobile service provider app only for one plus users. It has following hassle free services like
 - Convenient service modes to get your device repaired such as:
 - Getting your OnePlus device picked for repair from an address selected by you
 - Jumping the queue by pre-booking a visit to an authorized Service Centre near you as identified by the OnePlus Care App
 - Track your repair journey from within the App
 - View complete the history of your services and transactions with us
- v. **Zed secure** is a mobile Protection Plan by Solvy Tech Solution. It covers Accidental Damage and Liquid Damage for your mobile. Coverage begins from the date of purchase of the device and ends on completion of 365 days from the date of purchase of the device. Key Highlights of Zed Secure Mobile Protection Plan are:
 - Accidental damage protection
 - Liquid damage protection
 - Repair through Brand's authorized Service Centre
 - Paperless Documentation process
 - Free pick up and drop

III. PROPOSED SYSTEM

DR. Device is an Android based application that will be a service provider for devices. There would be one actor playing their part, which is the consumer who wants to use the services and get their devices repaired. The consumer first must select which type of services they want and whether they want a pick and drop service for their device or not. The system will check the authenticity of the user by either registering them or logging in if they already have an account. The users can also log in through other third-party applications. After which the consumer can book their service and select whether he/she wants a standby device for that period when their device is getting repaired. After all this the user is all set and the technician will visit for the pickup, as soon as the diagnostic of the device/ devices is done the user will then get their

timeline of when the device will be delivered back. There will also be an on-app mobile accessory store for selling items along with refurbished devices.

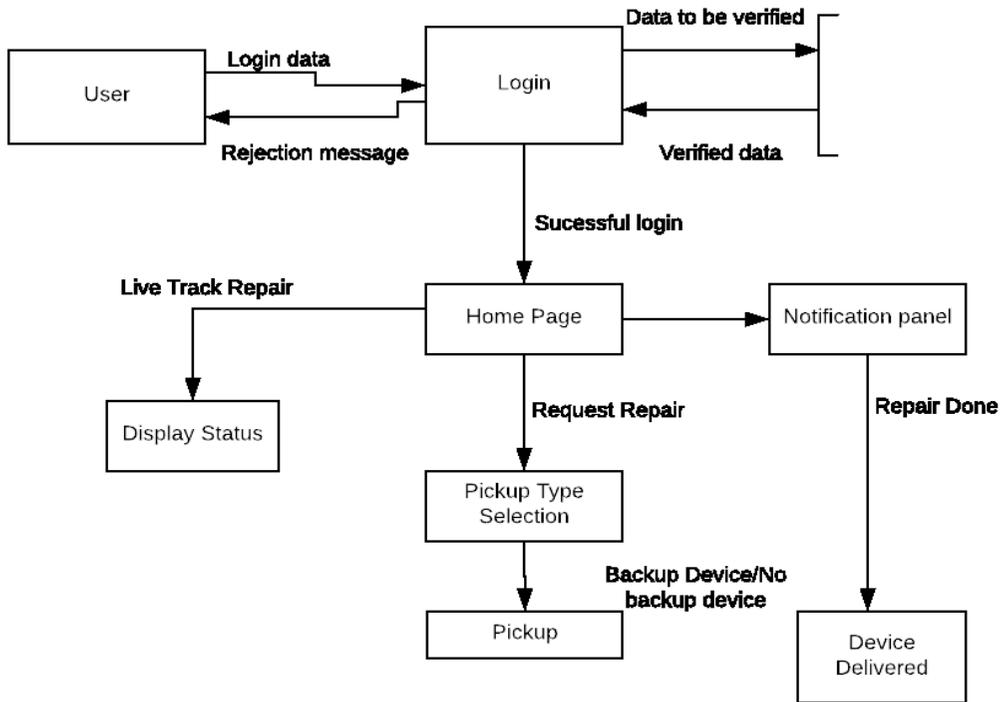


Fig 1: System Flow

The purpose of this project is to provide android user services more efficiently than the existing system. There are some disadvantages of the existing systems like it is either a service provider for mobile or for laptops and not for both at the same time or it is only for specific mobile brands or specific country. These disadvantages can be overcome by **Dr Device-Service Provider for All Services** and it can be made handy available to every android user.

The Screenshot shows how the user can select the timeslots for pickup as per the preference, there are multiple operating time available which the user can use. The pickup time can also be selected a day prior after which the app will take you to the backup device option and whether the user wants data along with the phone.

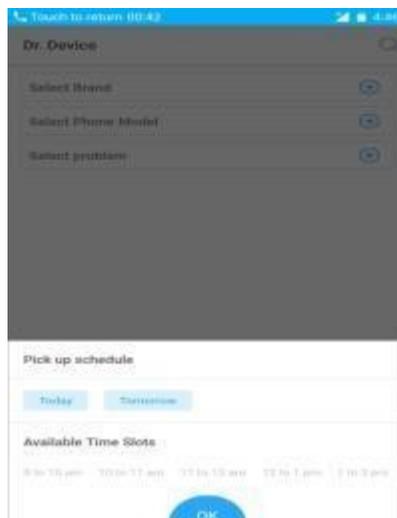


Fig 2: Screenshot of Pickup schedule interface

This shows the categorization of device and how user can select the brand of device, select the phone model. After which the user must state the device issue and then the pickup can be done based on the availability of services.



Fig 2.1: Screenshot of category interface

Previously people could not get help or locate the service centres conveniently in case of mobile, laptop damages or any other emergencies. Thus, our app is proposed to provide doorstep services to the user to assist people and fulfil their requirements easily. [1]

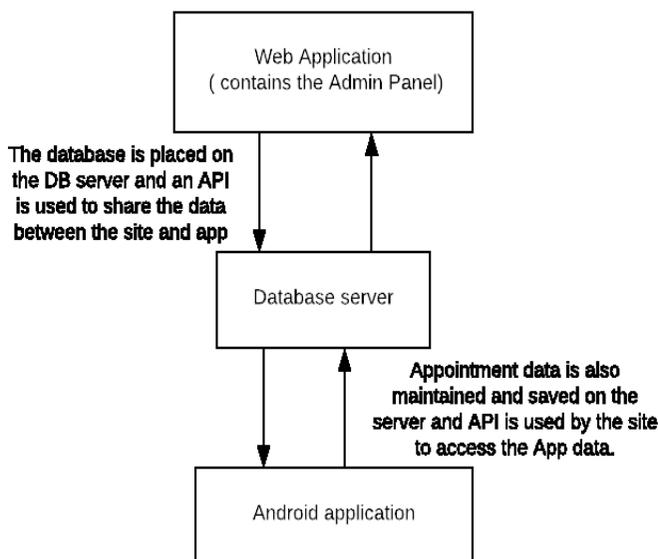


Fig 3: Block Diagram of Appointment Data Management

Laravel is the open source PHP framework majorly used for the development of web applications, it follows model view controller architectural pattern. Due to these features our admin panel uses this framework particularly for managing the contents like the types of devices and their variable models. It is going to be a web based application with consists of login, categories of devices and models. Admin can efficiently add

and remove the device or their specific model from the list these changes are going to be prominent on the application. We are using MySQL for this database driven web application which has multiple options of managing data in an efficient manner, login information of the users can be saved here along with the system details. MySQL is particularly going to be a middle application database service between the app and the web application

VI. FUTURE SCOPE

The major goal of this application is to provide efficiency to users majorly targeting the people at corporate level. DR. DEVICE is composed of two main components: a client-side application which runs on Android handsets, and a server-side application which will handle and interact with various client-side features [1]. The system is designed to provide features of mobile phones and laptops, feasible services provided by the service centres, locations of all the service centres in the vicinity, live tracking, 24/7 customer services and extended warranty period, etc. The system has a unique feature of online shopping where the users can buy mobile phones and laptops of various brands.[3] The above-proposed model is easy to implement considering the ASP model. The model is simple, secure and reliable.

This application can be improved in the future by adding the following functionalities:

- Extending to Apple iOS.
- Biometrics Registration and Login.
- Providing services for a wider range of devices.

VII. CONCLUSION

The proposed paper shows the flow, structure, and working of the DR. DEVICE application. DR. DEVICE is user-friendly i.e handy for users. It is freely available on play store. Thus, it is a time-saving application vitally for the people at the corporate level as well as cost-efficient application. So, we can conclude that the proposed system can be used to reduce human efforts by providing services at the doorstep and the feature of online shopping, hand in hand, with the modern technology. [1]

IX. REFERENCE

1. "Developing an Android based learning application for mobile devices", Telematics and Information Systems (EATIS), 2012 6th Euro American Conference, de Clunie, G.T.Fac. de Ing. de Sist., Computacionales, Univ. Tecnol. de Panama, Panama City, Panama Serrao, T. ; Monteiro Braz, J.R.- . Serr o, T. Rangel, N. Castillo, A. Gomez, B. Rodriguez, .de Barraza, Riley, J
2. "Implementation of Cloud Messaging System Based on GCM Service".Computational and Information Sciences (ICCIS), 2013 Fifth International Conference. Penghui Li Transp. Manage. Coll., Dalian Maritime Univ., Dalian, China Yan Chen ;Taoying Li ; Renyuan Wang ; Junxiong Sun 2"A public safety application of GPS-enabled smartphones and the android operating system" Systems, Man and Cybernetics, 2009. SMC 2009. IEEE International Conference-Whipple, J.Inf. Syst. Eng. Dept., Southwest Res. Inst., San Antonio, TX, USA Arensman, W. ;Boler, M.S.
3. "Unified platform for the delivery of notifications to smartphones notification" Carpathian Control Conference (ICCC), 2012 13th International. Mojziso, A. Inst. of Control & Informatization of Production Processes, Tech. Univ. of Kosice, Kosice, Slovakia Mojžiš, M.

Exploring Elliptic Curve Cryptography Arithmetic using the Sage Math Computer Algebra Software

¹Santoshi Pote & ²B.K. Lande & ³Sanchali Kshirsagar

¹Associate Professor, ²Professor, ³Assistant Professor

^{1,3}Electronics and Communication,

^{1,3}Usha Mittal Institute of Technology (SNDT women's university), Mumbai, India

²Mukesh Patel School of Technology Management of Engineering (NMIMS), Mumbai, India

ABSTRACT: The aim of this paper is to provide an introduction which emphasizes the mathematical and algorithmic components and building blocks suitable for mathematics students, while liberally illustrating the theory with examples. Most textbooks for a mathematics audience limit themselves to pen and paper calculations, which fails to give the student a sense of either the asymptotic complexity for the algorithms or access to a practical range for cryptographic study. Textbooks which take a computational view usually miss the conceptual framework of the mathematics and are either tied to a particular commercial software package or emphasize low-level computations in C or Java which requires a stronger computer science background. We choose to use the computer algebra software Sage for experimental exploration since this package is both freely available and designed for intuitive interactive use. We hope that this paper will fill a niche by emphasizing a mathematical presentation of structures in elliptic curve cryptography, without sacrificing the explicit exploration of the field. This paper contributes to the field of elliptic curve cryptography pedagogy by discussing and implementing small-scale prime field and extension field example.

Keywords: Cryptography, elliptic curve arithmetic, Sage.

I. Introduction

This paper deals with the implementation of elliptic curve arithmetic using Sage open source software over prime field and extension field. The implementation embrace arithmetic over both the fields, key generation, encryption and decryption. Elliptic curve cryptography is an attractive option in case of smart cards, internet security and cryptocurrency. The abundant mentioned cryptocurrency Bitcoin uses elliptic curve *secp256k1*. Wherever "sec" is standards for efficient cryptography and p is prime field [4]. The safety of elliptic curve cryptography is relies on numerous finite field operation. The computation of those field operation is extremely complicated in MATLAB and in C language. Sage is open source software and it is well documented so that we can easily implement different application such as encryption, decryption and digital signature on realistic elliptic curves. For example the realistic elliptic curves domain parameter which are defined in SEC 2: Recommended Elliptic Curve Domain Parameters Certicom Research Version 1.0 [4]. The layout of the paper in the following manner section (2) we have discussed theory of elliptic curve and arithmetic of elliptic curve, section (3) describe Sage software, in section (4) Sage functions which are used for elliptic curve prime field, section (5) discusses Sage function for elliptic curve over binary field. Finally section (6) is conclusion and future scope. The Sage software is beneficial to implement most of the cryptographic algorithm but our focus in this paper is on elliptic curve cryptography.

II. Elliptic Curves

An elliptic curve is a mathematical object which has certain specific properties. We can perform computations using the points that lie on an elliptic curve, and we can define problems using them that are computationally difficult. This computation can be used to obtain secure cryptosystems. Elliptic curve cryptography can be spilt in two broad areas. The first broad area concerns the mathematical "environment" in which we work, that is the elliptic curves themselves. The second broad area concerns the algorithms and systems using these structures to perform some interesting cryptographic operation, for example signing or key agreement [1][2].

Elliptic curves are the most essential part of "ECC" A generalized Weierstrass equation of an elliptic curve is

$$y^2 + a_1 xy + a_3 y = x^3 + a_2 x^2 + a_4 x + a_6 \quad (1)$$

From Eq.1 the simplified Weierstrass equation is

$$y^2 = x^3 + Ax + B \quad (2)$$

Where A and B are curve parameters that have constraints, and Eq.2 could be either defined on complex, real, integers, or any kind of finite field element. Elliptic curve over finite fields are provide more unpredictability which ensure secure cryptosystem. The finite field is often the field of integers mod a prime number.

That means the set of points on the curve are also an integer coordinate. It is possible to define point addition and multiplication on these elliptic curve points. In this paper we will be using this Eq.2 throughout our example of elliptic curves, because it is simpler to use and is very helpful to the study of ECC over prime field.

III. Elliptic curve

3.1 Elliptic curve over finite field (F_p)

Let F_p be a prime finite field so that p is an odd prime number, and let $A, B \in F_p$ satisfy $4A^3 + 27B^2 \neq 0 \pmod{p}$. Then an elliptic curve $E(F_p)$ over F_p defined by the parameters $A, B \in F_p$ defined over prime field modulo p is cryptographically good if the curve is nonsingular. This happen when the $\Delta_E = 4A^3 + 27B^2 \neq 0$. That means the polynomial $x^3 + Ax + B$ do not have multiple roots. Now consider two distinct points $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$ from $E(F)$ and define how to add those points that is determine $P_3 = P_1 + P_2 \cap E(F)$.

When $P_1 \neq P_2$, slope is $l = \frac{y_2 - y_1}{x_2 - x_1}$ and when $P_1 = P_2$ $l = \frac{3x_1^2 + A}{2y_1}$

The coordinates of P_3 are given by

$$x_3 = l^2 - x_1 - x_2 \quad (3)$$

$$y_3 = l(x_1 - x_3) - y_1 \quad (4)$$

It can be shown that addition law is associative law is associate, that is $(P + Q) + R = P(Q + R)$. It is also commutative $P + Q = Q + P$.

3.2 Elliptic Curves over extension field (F_{2^m})

Let F_{2^m} be a characteristics 2 finite field and let $A, B \in F_{2^m}$ satisfy $b \neq 0$ in F_{2^m} . Then the elliptic curve $E(F_{2^m})$ over F_{2^m} defined by the parameters $A, B \in F_{2^m}$ consists of the set of solutions of points $P=(x,y)$ for $x, y \in F_{2^m}$ to the equation

$$y^2 + xy = x^3 + Ax^2 + B \quad (5)$$

Two distinct points $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$ from $E(F_{2^m})$ and define how to add those points that is determine $P_3 = P_1 + P_2 \cap E(F)$.

When $P_1 \neq P_2$, slope is $l = \frac{y_2 + y_1}{x_2 + x_1}$ and when $P_1 = P_2$ $l = x_1 + \frac{y_1}{x_1}$

The coordinates of P_3 for point addition are given by

$$x_3 = l^2 + l + x_1 + x_2 \quad (6)$$

$$y_3 = l(x_1 + x_3) + x_3 + y_1 \quad (7)$$

And for point doubling

$$x_3 = l^2 + l + A \quad (8)$$

$$y_3 = l^2(l + 1) + x_3 \quad (9)$$

The set of points on $E(F_{2^m})$ forms an abelian group under this operation.

IV. SAGE software

Sage (System for Algebra and Geometry Experimentation) is mathematical software with options covering several aspects of mathematics, together with algebra, combinatorics, numerical mathematics, number theory, cryptography and calculus. The first version of Sage was released on 24th February 2005 as free open source software under the terms of GNU General Public License, with the initial goals of creating an "open source alternative to Magma, Maple, Mathematica and "MATLAB". The originator and leader of the Sage project, William Stein, is a mathematician at the University of Washington. Sage uses the Python Programming language, support procedural, functional and object-oriented construct [5][6][7][8].

4.1 Features of Sage

- a) A browser-based notebook for review and re-use of previous inputs and outputs, including graphics and text annotations. It is compatible with Firefox, Opera, Google Chrome and Safari.
- b) A text-based command-line interface using IPython.
- c) Support for parallel processing using multi-core processor, multiprocessor, or distributed processing
- d) Numerical linear algebra
- e) Matrix manipulation including sparse array.
- f) Libraries of number theory functions.
- g) 2D and 3D graphs of symbolic functions and numerical data.

4.2 Advantages of Sage

- a) Sage intended for audience in mathematics student (from high to graduate school), teachers, and research mathematicians.
- b) Sage uses highly optimized mature software like GMP, PARI, Gap, and NTL, and so very fast at certain operations.
- c) The source code must be freely available and readable, so users can understand what the system is really doing and more easily extend it.
- d) It is well documented
- e) It is user friendly.

4.3 Functions of Sage over Elliptic Curve

a) Defining an Elliptic curve

By using the command `EllipticCurve([A,B])` we can define an elliptic curve $y^2=x^3+2x+7$ over rational field where a,b are the coordinates of the elliptic curve.

```
E=EllipticCurve([2,7])

print E
```

b) Defining an Elliptic Curve over Finite Field.

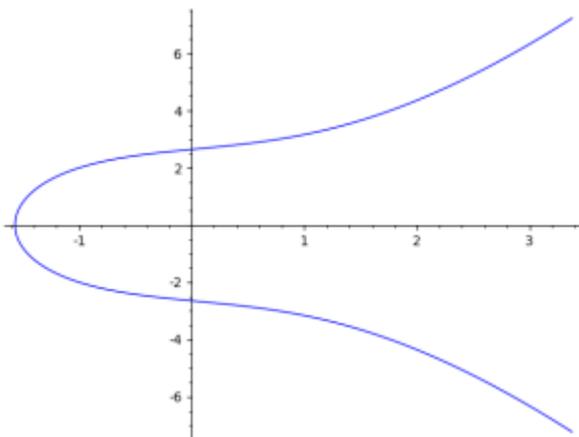
Finite field of size 1489 can be defined using the in-built command `FiniteField(1489)` and then using the command for elliptic curve we can define the curve $y^2=x^3+2x+7 \pmod{1489}$ over the finite field of size 1489.

```
F=FiniteField(1489)

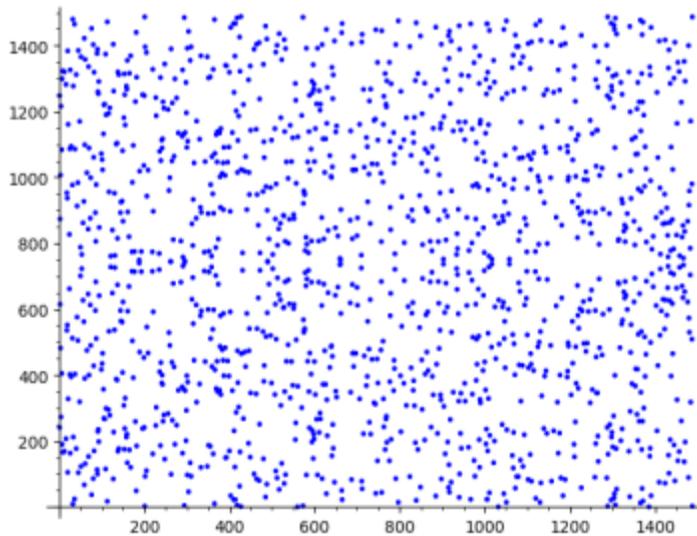
E=EllipticCurve(F,[2,7])

print E
```

c) Plotting of Elliptic Curve $y^2=x^3+2x+7$ over the rational field



d) Plotting of Elliptic Curve $y^2=x^3+Ax+B$ over the Finite Field of size 1489



The example described here is of manual computation of elliptic curve point. Let E be given by $y^2 = x^3 + 4x + 4 \pmod{5}$. First step is to compute all the points on the curve by substituting $x=0,1,2,3,4$ and solve for y . Substitute each of these into the equation(2) and find the different values of y .

$$\begin{aligned} x \equiv 0 &\Rightarrow y^2 \equiv 4 \Rightarrow y \equiv 2,3 \pmod{5} (0,2)(0,3) \\ x \equiv 1 &\Rightarrow y^2 \equiv 9 \Rightarrow y \equiv 3,2 \pmod{5} (1,2)(1,3) \\ x \equiv 2 &\Rightarrow y^2 \equiv 0 \Rightarrow y \equiv 0 \pmod{5} (2,0) \\ x \equiv 3 &\Rightarrow y^2 \equiv 4 \Rightarrow \text{no solution} \\ x \equiv 4 &\Rightarrow y^2 \equiv 4 \Rightarrow y \equiv 2,3 \pmod{5} (4,2)(4,3) \end{aligned}$$

Therefore, following are the points on the elliptic curve along with point at infinity $(0,2), (0,3), (1,2), (1,3), (2,0), (4,2), (4,3)$

For large prime field it is a tedious job to find all the points on the curve manually as shown in the above example. Sage computes these points more quickly than doing the calculations by hand over any large number of prime field or binary field.

e) Total number of points /orders of elliptic curve computed using following function. The total number of points on the curve are 1420, and following function can generate the points as per the requirement. Different variation of functions are available to print the points. The following command generate only six points.

```
F=FiniteField(1489)
E=EllipticCurve(F,[2,7])
E.order()
```

The first point represent point at infinity and the remaining points which are laying on the curve.

f) The arithmetic of elliptic curve over finite fields is based on addition of two points and doubling of point. The arithmetic of finite field is involved different modular operations such as addition, multiplication, subtraction which is obtained by additive inverse and division that is multiplicative inverse. These field operations are very difficult when solving it manually for large field.

Example1: (point addition) Suppose E is defined by $y^2 = x^3 + 2x + 7 \pmod{1489}$. Now add the point (852,1043) and point (410,22). The slope $l \equiv \frac{22-1043}{410-852} \equiv 1050 \pmod{1489}$. Then we have the following equation to compute the third point on the curve.

$$x_3 = l^2 - x_1 - x_2 \equiv 1050^2 - 852 - 410 \equiv 867 \pmod{1489}$$

$$y_3 = l(x_1 - x_3) - y_1 \equiv 1050(852 - 867) - 1043 \equiv 1075 \pmod{1489}$$

This means that (852,1043)+(410,22)=(867,1075) which is also on the curve and easily computed by using following Sage command

```
p=1489,
A=2
B=7
EC=EllipticCurve(GF(p),[A,B]);#EC
P=EC.random_point();#P
Q=EC.random_point();#Q
x1 = P[0]; y1 = P[1];x2 = Q[0];y2 = Q[1]
print "Point1=",P
print " Point2=",Q
R=P+Q;R
print"Addition of two Points=",R
```

```
Output:
Point1= (852 : 1043 : 1)
Point2= (410 : 22 : 1)
Addition of two Points= (867 : 1075 : 1)
```

Example2: (Point Doubling) suppose E is defined by $y^2 = x^3 + 2x + 7 \pmod{1489}$. Now double the point (879,1149). The slope $l \equiv \frac{3*879^2+2}{2*1149} \equiv 1332 \pmod{1489}$. We have the following equation to compute the third point on the curve.

$$x_3 = l^2 - 2 * x_1 \equiv 1332^2 - 2 * 879 \equiv 556 \pmod{1489}$$

$$y_3 = l(x_1 - x_3) - y_1 \equiv 1332(879 - 556) - 1149 \equiv 255 \pmod{1489}$$

This means that P+P=(879,1149) + (879,1149) = (556,255) which is also on the curve and easily computed by using following Sage command

```
p=1489;A=2;B=7
F=GF(p)
EC=EllipticCurve(GF(p),[A,B]);# EC
P=EC.random_point()
x1 = P[0]; y1 = P[1]
print "Point1=",P
S=2*P # Doubling Point
output:
Elliptic Curve defined by y^2 = x^3 + 2*x + 7 over Finite Field of size 1489
Point1= (879 : 1149 : 1)
Doubling of point= (556 : 255 : 1)
```

g) Elliptic curve over binary field or Extension field is generated by following functions. Consider the binary field $F(2^m)$ where $m=3$. This field is called characteristics two field having only two components $\{0,1\}$. To implement the field $F(2^3)$ which consist of eight field element the we required irreducible polynomial having degree 3. For $F(2^3)$ the irreducible polynomial is $f(\theta) = \theta^3 + \theta + 1$ where θ is the root of the polynomial. So, the field consist of element $\{0,1, \theta, \theta^2, \theta^3, \theta^4, \theta^5, \theta^6\}$. The task of selecting irreducible polynomial and computing field element is accomplished by Sage easily.

```
F.<a> = GF(2^3, F.modulus);#Finite Field in a of size 2^3
E=EllipticCurve(GF(2^3,'theta'),[1, 1, 0, 0, theta]);# Elliptic Curve defined by y^2 + xy = x^3 + x^2 + theta over
Finite Field in a of size 2^3
E.points() #[(0 : 1 : 0), (0 : theta^2 + theta : 1), (1 : theta^2 : 1), (1 : theta^2 + 1 : 1), (theta : theta^2 : 1), (theta : theta^2 + theta : 1), (theta + 1 : theta^2 + 1 : 1),
(theta + 1 : theta^2 + theta : 1), (theta^2 + theta : 1 : 1), (theta^2 + theta : theta^2 + theta + 1 : 1)]
P=E(1,theta^2);# Random Point selected on Elliptic curve
Q=E(theta + 1,theta^2 + 1); #Random Point selected on Elliptic curve
R=P+Q; $ Addition of two points
Output: (1 : theta^2 + 1 : 1)
```

V. Elliptic curve in application

Specifically, each elliptic curve define with domain parameters over field are sextuple: $T=(p, A,B, G, n, h)$ where p specifying the field, element A,B are the field element, G is the base point on elliptic curve and prime n is the order of G .

The curve used in Bitcoin cryptocurrency is koblitz curve *secp256k1* having following parameter:[4]

1. Equation $y^2 = x^3 + 7$ ($A = 0, B = 7$)
2. Prime Field (p) = $2^{256} - 2^{32} - 977$
3. Base point (G) =
 $(79BE667EF9DCBBAC55A06295CE870B07029BFCDB2DCE28D959F2815B16F81798,$
 $483ADA7726A3C4655DA4FBFC0E1108A8FD17B448A68554199C47D08FFB10D4B8)$
4. Order (n) = $FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFEBAAEDCE6AF48A03BBFD25E8CD0364141$

The curve's name is *secp256k1*, where SEC stands for Standards for Efficient Cryptography and 256 is the number of bits in the prime field. To begin working with elliptic curves, let's confirm that the base point (G) is on the curve ($y^2 = x^3 + 7$)

$G = (79BE667EF9DCBBAC55A06295CE870B07029BFCDB2DCE28D959F2815B16F81798,$
 $483ADA7726A3C4655DA4FBFC0E1108A8FD17B448A68554199C47D08FFB10D4B8)$

5.1 Utilizing ECC for Public Key Cryptography

Private keys are the scalar, denoted with "s". The public key is the resulting point of the scalar multiplication or sG , which is usually denoted with "P". $P(x,y)$ is actually a point on the curve. Here's how you can derive the public key from the private key using Sage function.

```
p = 2^256 - 2^32 - 977; #prime field of 256 bit
A=0; B=7# Elliptic curve parameter
F = GF(p); E = EllipticCurve( F, [A,B] );#Generate Elliptic curve
G=E(0x79BE667EF9DCBBAC55A06295CE870B07029BFCDB2DCE28D959F2815B16F81798,
0x483ADA7726A3C4655DA4FBFC0E1108A8FD17B448A68554199C47D08FFB10D4B8);#
Base point G of elliptic curves
s=999; # scalar is a private key
P=s*G; # Public Key generated from scalar multiplication
```

Output:

Prime field

256bit: 115792089237316195423570985008687907853269984665640564039457584007908834671663

Elliptic curve equation:

Elliptic Curve defined by $y^2 = x^3 + 7$ over Finite Field

of size 115792089237316195423570985008687907853269984665640564039457584007908834671663

Public key generated by scalar multiplication.

(68073332634951771983379810576461367518612914887001815873609470706324019594353 :

100265349306802633132698962749090572900634375768193967970766835776642803928888 : 1)

The above key generation example shows that how we can easily generate keys for public key cryptography using Sage function on a large number of prime field curve.

VI.CONCLUSION

Sage strives to provide a consistent and uniform access to features in a wide area of mathematics ranging from group theory to numerical analysis and beyond visualization in two and three dimensions, animation, networking, databases. The arithmetic of elliptic curve is a central of discussion where it plays a very important role in cryptography. In this paper, a brief and yet a quite comprehensive exploration of elliptic curve cryptography using Sage software have been illustrated. With the help of Sage, we can solve the arithmetic of elliptic curve very fast and accurate for realistic elliptic curve (secp112r1, secp128r1, secp160r1, secp192r1, secp256r1).

REFERENCES

1. Lawrence C. Washington. Elliptic Curves Number Theory and Cryptography. Chapman & Hall CRC press, 2008.
2. J.H.Silverman. The arithmetic of Elliptic Curves, Graduate Texts in Mathematics, Vol.106, Springer, 1986.
3. Miller V.S. Use of elliptic curves in cryptography, Advances in Cryptology- Crypto 85, LNCS Vol. 218. Springer (1985).
4. Certicom Research. Standards for efficient cryptography2: Recommended elliptic curve domain parameters Certicom.2000.
5. Paul Zimmermann, AlexandreCasamayou ,Nathann Cohen, Guillaume Connan. Computational Mathematics with SageMath.
6. Prof. Bernhard Esslinger.The CrypTool Book: Learning and Experiencing Cryptography with CrypTool and SageMath.
7. The Sage development team, Sage Tutorial Release 8.6. January 21 2019.
8. Ted Kosan, Sage for newbies, Copyright © 2007.

Fake News Detection

¹Swati Modi, ²Disha Modi, ³Pooja Patel, ⁴Noella Thomas, ⁵Dr. Vaishali Jadhav

^{1,2,3,4}Student, ⁵Professor

Information Technology

St. Francis Institute of Technology, Mumbai (400103), India

ABSTRACT: Due to reduced internet costs and ease of internet access, people are more exposed to satire, clickbait, hoaxes and several other misleading contents which is commonly referred to as Fake News. Fake news leads to a lot of misunderstandings like creating a bias in elections, chaos, and confusion regarding the current events. Hence it is very significant to distinguish between fake and real news and restrict the widespread of fake news. To tackle this problem, a hybrid machine learning approach of stance detection and document similarity based on the acquired news content was used in this paper. Experimental Results shows that the model reaches an accuracy of 85.5% which indicates that the method proposed can be effective enough and applicable in the real-world scenario.

Keywords: Fake News Detection, Stance Detection, Document Similarity, Content Based classification, Support Vector Machines (SVM), Naive Bayes, Passive-Aggressive, Gradient boosting, Long Short-term memory(LSTM).

I. Introduction

Social Media and online news websites enable people to access news very easily. Most of these news articles are fake and spread more quickly with no verification or security and a lot of assumptions. Some examples of fake news that were widely spread during demonetization period were 'GPS chips in new notes to detect black money', '10 coins are not genuine and invalid'.

The biggest challenge is the lack of an efficient way to tell the difference between real news or a fake one; even humans are often unable to tell the difference between them. Hence, Artificial Intelligence especially Machine Learning and Natural Language Processing (NLP) has proved its capability in Fake News Detection. Human language is very complex due to its ambiguity, syntactic, semantics as well as referential complexity hence it is very difficult to teach machines to understand how we communicate which makes it very interesting to approach the problem via machine learning perspective. Machine learning techniques can help to automate the task of detecting fake news.

While there are already existing studies to detecting fake news using machine learning, they are mostly comparative study or just researches made on the topic. Most of the implementations are not very accessible to the common audience and do not consider proof checking with internet search results which can improve the accuracy and reliability to a great extent. In this paper, we aim to efficiently classify fake news using a combination of machine learning approaches to Stance detection and Document Similarity. To make this project accessible to the common audience we are developing a chrome extension that once installed, can detect whether a news article is fake or not in just one click.

II. Literature Review

Several kinds of research and implementations to detect fake news using machine learning were taken into consideration for this paper. Most of the papers emphasize the capability and importance of stance detection in the process of detecting fake news. The insights drawn from different papers are discussed as follows:

Razan Masood et al. [1] discusses detecting Fake News using stance detection through traditional machine learning approaches where final score obtained was 82.1 percent using Craig Silverman's emergent dataset involving 2595 articles where 71159 labeled pairs and 49972 pairs for training and testing respectively were used in this process. Training dataset shows that 73.13 percent were unrelated, 17.83 percent were discussed, 7.36 percent agreed and 1.68 percent disagreed where 0.25 percent score is added to each one of them and after classification, it adds 0.75 percent score to agree, disagree and discuss. In this, the entire article is divided into the first 5 sentences (heading) where four sentences from tail & 10 percent of the middle sentence (i.e. min 2 sentences) give the best performance.

Samir Bajaj [2] proposes a method for detecting Fake news using deep learning based on only its content by building a classifier where the problem is approached via pure NLP perspective. Dataset was taken from www.kaggle.com and signal media news respectively. Multiple Models like Logistic Regression, Two-layer Feedforward Network, Recurrent Neural Network, Gated Recurrent Units, Long Short-Term Memories, Bidirectional RNN with LSTMs, Convolutional Neural Network with Max-Pooling and Attention-Augmented Convolutional Neural Network were used. Results show that CNN with max pooling and attention has the maximum precision of 0.97 which concluded that features like the source of news, associated URLs, topic, country and publication year can further improve the accuracy.

Another method proposed by Xiaowei Wu et al. [3] is by using machine learning to detect the stance of newspaper headlines on their bodies, which can serve as an important factor in detecting fake news. Various machine learning models including support vector machines (SVM), softmax, multinomial Naive Bayes, and multilayer perceptron classifier (MLP) were implemented. The Dataset used was provided by Fake News Challenge (FNC-1).40350 instances were randomly selected as the training set. Each stance is one of the {unrelated, discuss, agree, disagree}. It is found that the “similarity” and “bow” features better describe the stances of the headlines towards the bodies, than the “word sentiments” and “polarity features”. The addition of both “similarity” and “bow” features to improve the performance best than adding a single type of features. Using scikit learn, these models were implemented to learn from the training data using k-fold (k=10) cross-validation. The accuracy rate for all models in the test set clearly shows that MLP Classifier yields the best performance.

Qi Zeng et al. [4] method for detecting fake news is by using handcrafted feature engineering and Neural Network. The Dataset used was provided by Fake News Challenge (FNC-1) consisting of 49972 headline-body pairs for the task of stance detection. In the handcrafted feature, engineering features were concatenated and fed into a Gradient Boosting classifier. In the second approach, feature extractor and classifier are the two components where feature extractors are Bilateral Multiple Perspective Matching, Attentive Reader with full attention, Attentive Reader with simple attention and Bidirectional Encoder which are all RNN based are used. Results after evaluation show that attention reader with full attention had a maximum accuracy of 86.5 percent compared to all others which conclude to improve precision on agree category first and then on disagree category for good accuracy.

III Proposed Solution

The proposed solution is to make use of a hybrid approach of stance detection and document similarity to detect whether a news article is fake or not. A chrome extension will be developed which will allow users to check if the article they are reading is fake or real.

Whenever a user is reading an online news article and wants to know if it is fake or not, the chrome extension makes a request and sends the current URL and news article to a flask server where a Machine learning model is deployed. The flask server forwards the call to the model which returns a prediction whether the news article is fake or real. The result is displayed on the user's screen.

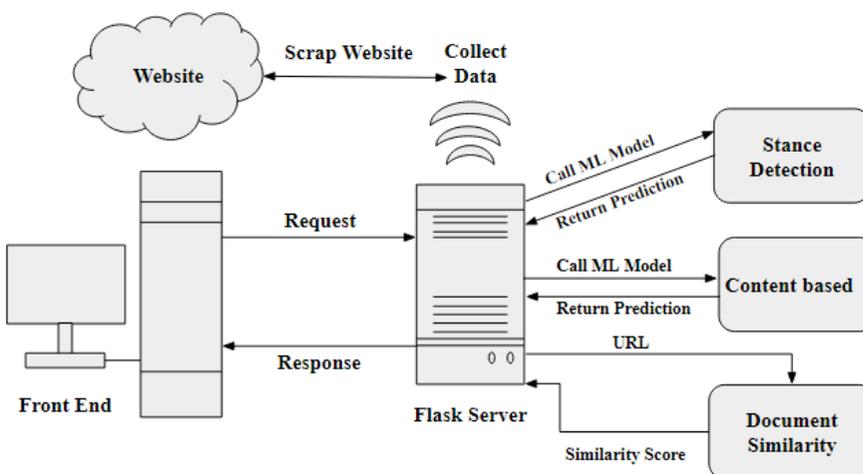


Fig.1. System Block Diagram

The model has 3 major modules - Stance detection, Document similarity, and Content-based classification. Results from each module are taken into consideration to calculate the final score which decides whether the news article is fake or not. To further reduce the widespread of fake news articles online, the extension has a "flag as fake" option which will allow users to mark unreliable news articles as fake.

IV. Implementation

4.1 Dataset

Data used for this project was collected from various sources in public domains.

- The dataset from Fake News Challenge-1 [10] contains the headline, body, and stance. The dataset is provided as two CSVs. Train bodies.csv contains the body text with the corresponding id and Train stances.csv contains the labeled stances and body id.
- The dataset from Fake News Corpus [11] contains millions of news articles.
- The final collection has a total of 51,000 articles where each article is labelled as "Fake" or "Real". This dataset is shuffled and split into 80% training set and 20% test set.

4.2 Data Pre-processing

For In this phase, we reduce the size of original data and represent the data using unigram and vector-based model. The data need to be in a clean format, which is done by stop-word removal, tokenization, lower casing, sentence segmentation, and punctuation removal. The last step is the classification process which is to train the classifier.

4.2.1 Stop Word Removal

Stop-words are unimportant words that are removed such as, a, about, an, are, as, at, be, by, for, from, how, in, is, of, on, or, that, the etc which were removed from each document, and the processed documents were stored.

4.2.2 Stemming

After tokenizing the data, the next step is stemming. Stemming is a process where words are reduced to a root by removing inflection through dropping unnecessary characters, usually a suffix.

4.2.3 Lemmatization

Lemmatization normally aims to remove inflectional endings only and to return the base or dictionary form of a word.

4.2.4 TF-IDF for feature extraction

TF-IDF, which stands for term frequency—inverse document frequency, is a scoring measure widely used in information retrieval (IR) or summarization. TF-IDF is intended to reflect how relevant a term is in a given document.

4.3 Stance Detection

In Stance Detection, we find the relation between the body of a news article and its title. Specifically, the body text may agree, disagree, discuss or be unrelated to the title. According to the literature review, Credible news sources show a positive stance whereas if the stance is unrelated, the credibility of the news source is considered to be low.

In this paper, for implementing stance detection, many algorithms ideal for classification like Support Vector Machine, Naive Bayes, Gradient Boosting, Passive aggressive classifier are used. The preprocessed data is fed to all the models. The following sections describe each in detail.

4.3.1 Support Vector machine

A Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyperplane. In other words, given labeled training data (supervised learning), the algorithm outputs an optimal hyperplane which categorizes new examples. The accuracy achieved by using SVM is 73%

4.3.2 Naive Bayes

Naive Bayes is a supervised learning classification technique which assumes the presence of a particular feature in a class is unrelated to the presence of any other feature. Working on the principles of conditional probability, it derives the Bayes theorem as follows:

$$P(C|X) = [P(X|C) * P(C)] / P(X)$$

Where,

$P(C|X)$ is called our posterior probability.

$P(X|C)$ is the conditional probability or likelihood.

$P(C)$ is the prior probability of the outcome

$P(X)$ is the probability of the predictor variables

The accuracy achieved by using Naive Bayes is 67%

4.3.3 Passive-Aggressive Classifier

Passive-Aggressive is a family of online algorithms for classification and regression problems that learn from massive streams of data. The accuracy achieved by using Passive-aggressive is 79%.

4.3.4 Gradient boosting

Gradient boosting is a machine learning technique for regression and classification problems, which produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees. The accuracy achieved by using Gradient Boosting is 69%.

4.4 Content Based Classification

In Content Based Classification, we classify the article based on the linguistic features and writing style of the article. For Content Based Classification, we extracted frequency and tf-idf features and implemented Recurrent Neural Network-Long Short term memory (LSTM) and achieved an accuracy of 92%.

4.4.1 Long Short term Memory

LSTM is a recurrent neural network (RNN) architecture that remembers values over arbitrary intervals. LSTM is well-suited to classify, process and predict time series given time lags of unknown duration. Relative insensitivity to gap length gives an advantage to LSTM over alternative RNNs, hidden Markov models and other sequence learning methods.

4.5 Document Similarity

Another strategy used is Document Similarity which is an Instance-based learning approach. In document similarity, the content of a news article is compared with the content of other related web pages on the internet. The related web pages are determined by querying the headline to search engines and considering the top n search results. The similarity between the articles will be calculated using cosine similarity with TF-IDF. We also search for refuting words in the articles and then compute and output a score for the classification of the news.

Results obtained from each of the 3 modules - Stance detection, Document similarity, and Content based classification will be taken into consideration for calculating the final score. Based on this score, the news will be classified as "Fake" or "Real".

V. Results

The accuracy of each model that was employed for stance detection model and content based classification model, are summarized in table 1 and table 2

5.1 Results of Stance Detection model

Table 5.1: Accuracy of different models for stance detection

Algorithm	Accuracy
Support Vector Machines	73%
Naïve Bayes	67%
Passive-Aggressive classifier	79%
Gradient Boosting	69%

5.2 Results of Content Based Classification model

Table 5.2: Accuracy of LSTM model for content-based classification

Algorithm	Accuracy
Long Short-Term Memory (LSTM) + Frequency Features	89%
Long Short-Term Memory (LSTM) + TF-IDF Features	92%

It has been observed that Passive-Aggressive for stance detection has the highest accuracy as compared to other algorithms for detection of fake news. The accuracy of LSTM model for content-based classification with TF-IDF features is 92%.

VI. Conclusion

In this paper, we addressed the issue of detecting fake news using a machine learning approach. We implemented three major techniques - Stance detection, Content based classification, and Document similarity. Stance detection with Passive-aggressive classifier reaches an accuracy of 79% while LSTM model of Content based classification reaches 92%.

We have developed a chrome extension which will make the system accessible to all online news readers. Our performing model reaches an accuracy that is considerable and over humans' ability to spot fake news.

The overall accuracy of the model is 85.5% which indicates that the method proposed can be effective enough and applicable in the real-world scenario.

VI. Future Work

Although the accuracy of the model is 85.5%, it can further be improved considering the following factors

- A Lot of fake news is shared on social media sites especially on platforms like WhatsApp and Facebook so fake news detection can be implemented on such platforms.
- Currently the model classifies news as “Fake” or “Real” based on its content and URL. Considering other metadata like source of article, publishing medium, geographical region, and domain can improve the efficiency and performance.
- Using more data for training from different sources can improve the accuracy and performance of the model.

VII. References

1. R. Masood and A. Aker. The Fake News Challenge: Stance Detection using Traditional Machine Learning Approaches. Proceedings of the 10th International Joint Conference on Knowledge Discovery. Knowledge Engineering and Knowledge Management. 2018.
2. BasuSamir Bajaj. The Pope Has a New Baby! Fake News Detection Using Deep Learning. Stanford University CS 224N - Winter 2017.
3. Xiaowei Wu . Sizhu Cheng . Zixian Chai . Fake News Stance Detection. <http://cs229.stanford.edu/proj2017/final-reports/5244160.pdf> .
4. Qi Zeng. Quan Zhou. Shanshan Xu. Neural Stance detectors for Fake News Challenge. Department of Physics. Stanford University <https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1174/reports/261936.pdf>
5. S. Gilda. Evaluating machine learning algorithms for fake news detection. 2017 IEEE 15th Student Conference on Research and Development (SCORED). 2017.
6. M. Granik and V. Mesyura. Fake news detection using naive Bayes classifier. 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON). 2017.
7. Manisha Gahirwal. Sanjana Moghe. Tanvi Kulkarni. Devansh Khakhar. Jayesh Bhatia. Fake News Detection. International Journal of Advance Research. Ideas and Innovations in Technology. 2018.
8. S. Pal. T. K. S. Kumar. and S. Pal. Applying Machine Learning to Detect Fake News. Indian Journal of Computer Science. vol. 4. no. 1. p. 7. 2019.
9. H. Ahmed. I. Traore. and S. Saad. Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques. Lecture Notes in Computer Science Intelligent. Secure. and Dependable Systems in Distributed and Cloud Environments. pp. 127–138. 2017.
10. Fake News Challenge 1. <http://fakenewschallenge.org/>
11. Fake News Corpus. <https://github.com/several27/FakeNewsCorpus>

FAULT DETECTION AND DIAGNOSIS OF MERCHANT SHIP AFTER EXPERIENCING COLLISION

Mainak Bhaumik^{1*}, Dr. Arun Kumar Lokhande², Jagdish Chahande³, Suraj Marale⁴, Shivam Shukla⁵, Chaitanya Rothe⁶, Mahesh Pisal⁷, Mahesh Patil⁸, Someshwar Haibathpure⁹, Rekha Rathore¹⁰, Nilesh Bodkhe¹¹

^{1*3 4 5 6 7 8 11}Assistant Professor in Mechanical Engineering Department in MGM's CET, Kamothe, New Mumbai, India -410209

²Professor in Chemical Engineering Department, and HOD in Mechanical Engineering Department, in MGM's CET, Kamothe, New Mumbai, India -410209

⁹Assistant Professor in Civil Engineering Department, in MGM's CET, Kamothe, New Mumbai, India -410209

¹⁰Assistant Professor in Applied Science Department, in MGM's CET, Kamothe, New Mumbai, India - 410209

ABSTRACT: *The study and analysis involves optimal damage identification, estimation of work and rectification of oil tanker named as M.T.Nautical Global XVI (Ex-Gulf Petroleum I), which was collided with M.V. Pac Signius, a cargo container ship. The collision occurred at the early morning wee hours of 0500hrs when both the ships were running and a fatal damage occurred to M.T.Nautical. After collision M.T.Nautical put anchor at the same place of outer anchorage to get it rectified as further movement in damage condition was not advisable without conducting the repairs. The accident occurred in the Indian Arabian water in outer anchorage of Kandla. The abroad ship owner company M/s Ajman, UAE, Gulf Shipping Services FZZ allotted the work to the Indian company M/s Tangar Ship Management Pvt. Ltd. to get it rectified at collision occurred area which is outer anchorage of Kandla. The ship affected area was starboard side and poop deck area. The marine grade steel plate of hot rolled graded as LR-A of thickness 16 mm has been replaced successfully by cutting, grinding and welding process. The finished final work was leak tested by spraying water at a pressure of 3kg/cm², under the supervision of Indian Register of Shipping (IRS) surveyor and Mercantile Marine Department (MMD) surveyor jointly. The steel replacement by welding technique as per welding procedure specification (WPS) was shielded metal arc welding (SMAW) and fusion core arc welding (FCAW). A satisfactory and healthy repair works conducted, ship trial taken and certification of repair and vessel fitness for voyage also done after the replacement of steel plates, primer coating on plates and ship color painting on the plates. All the certifications were done by the IRS class, MMD surveyor, vessel owner and Vendor Company.*

Keywords: *Ship collation, oil tanker, MT Nautical Global XVI, LR Gr A, poop deck, starboard, aft, bulkhead, steel plate, steel angle, plate replacement, oxy acetylene cutting, arc welding, fusion welding, clit support, ship class surveyor, ship trial, Kandla port.*

I. Introduction

The merchant ships are for various purposes in the sea and ocean water to perform multiple operations. In offshore operations the offshore supply vessels are performing various supporting works to the offshore owner company like piles, pipelines, cargo, ballast supply and inspection and maintenance works offshore platforms etc. The oil tankers and cargo carriers are heavy in weight, bulky in size. The approximate average length of the oil tanker and cargo carriers are 150mts to 300mts based on gross tonnage (GRT), load carrying capacity and design of the vessel. During the undergoing of ship movement and voyage in favorable or adverse situations of high tide conditions the ship experiences rolling, pitching, damage and collision occurs with nearby ship or in the port area or with ice barge. Here is the storey of collation of MT Nautical Global XVI with MV PaccSegnius. MT Nautical Global XVI carries International Maritime Organization (IMO) No. as 9113135. The type of vessel is oil tanker. The vessel gross tonnage (GRT) is 7134tons, overall length 127.5m, breadth as 21m, depth 10.04m, total height from keel to mast 34.859m, BHP 4118kw year of build is 1995 by Pan united shipping, Singapore. The owner and manger name is Prince shipping services – Ajman, United Arab Emirates (UAE). The other vessel with which MT Nautical collided named as MV PaccSegnius. It is having IMO No. as 9443358, vessel type is general cargo vessel, GRT is 21094tons, overall length 179m, breadth as 27m, year of build is 2012, ship builder named as Zheijang Ship industry, China, the ship owner and manager is Pacc Ship manger.

Sometimes due to automatic mode of operation of ship at the wee hours the collision and accident works. In case of collision the damage is very much fatal towards the ship body and steel of the ship point of view. The ships are required to do periodical survey and maintenance like annual survey, three surveys or five yearly surveys. If any major problem occurs in between, the ships are required to conduct maintenance and survey by the ship class surveyor and MMD surveyor. Based on the severity of the damage the repair and maintenance can be conducted. It can be done at the offshore zone area also if it is minor and accepted by the surveyors. In most of the cases she is required to the anchorage or in the port or in dry dock location. The main hull problem identification purpose underwater surveys can be carried out by the divers and based on the severity decision can be taken. Hull underwater excess marine growth can be cleaned by the drives and she can resume in operation. If underwater or overwater steel plate damage occurs it requires to replace by the new steel plate. In most of the cases the dry docking is costly and it is not advisable also to carry out such operation unless and until it is mandatory. The over water repair, inspection and maintenance can be carried out at anchorage area. In deep sea anchorage area the rolling and pitching of sea influences the works, but it's quite economical to perform the operations. The Fig. 1 illustrates MT Nautical Global XVI Oil tanker before collation undergoing voyage operation with cargo oil. Whereas the Fig. 2 shows the vessel MV PaccSeginus with which MT Nautical Global XVI collided.



Fig. 1: MT Nautical Global XVI Oil tanker before collation undergoing voyage operation.



Fig. 2: MV PaccSeginus general cargo vessel undergoing voyage operation before collation with MT Nautical Global XVI.

II. PROBLEM DEFINITION

The assigned work is about the collision work rectification of vessel MT Nautical Marine Global XVI which was carrying crude oil from Dubai to Kandla for unloading purpose. Major damages occurred at the starboard side and aft side poop deck area. At the starboard side the damages occurred at the above water

line side steel plates, onboard steel plates and inside bulk head straight structure. Similarly at the aft side poop deck area also same damages occurred at the side plate, overboard area and bulk head structure. The selected material as per Lloyds standard as hot rolled carbon steel of Lloyds shipping grade steel as LR Grade A. The standard density of LR Gr A steel material, from the steel data handbook as 8000 kg/m³. It has been ordered and taken onboard steel plates with structure angle as 15 tons of materials with consideration of cutting and clut allowances. The electrodes used for structural steel fabrication is E7018 and E6013. The collision cause and damage reason of MT Nautical Global XVI was due to vessel movement at around 15 nautical miles with loaded oil cargo. The other vessel MV PaccSeginus was carrying general cargo. Due to automatic Pilot mode movement of the vessel and negligence of the crew and onboard Pilot the sever dashing occurred with higher momentum. First of all it got hit at the starboard side of the vessel. Due to unable of reducing speed and of unable to take immediate turn the ship became uncontrollable at Sevier impact of collision, jerk and momentum it took turn and dashing occurred at the aft side poop deck are of the vessel. It was resulted as Sevier fatal damage to her and unbalancing of vessel MT Nautical. Hence the vessel was unable to move further. It was required to put anchor at same spot and was informed for assistance and rectification her damaged body plates and structure. The Fig. 3 shows about the effected steel plate of MT Nautical Global XVI. The forward main deckstarboardaffected area due to collision has been marked in the ship drawing which has been illustrated in the Fig. 4. The affected area has been marked by red ink. The estimated replaceable new steel plate area (A1) is 100m², which is bearing length (L1) 25m and breadth (B1) 4m. The Fig. 5 is about forward starboard ship side in drawing, marked by red ink. It has estimated replaceable area (A2) as125m². It is having height length (L2) 25m and height (H1) as 5m. The Fig.6 illustrates aft starboard side in the ship drawing marked by red ink. It is having estimated replaceable steel area (A3) as 77.55m². This area consists of length (L3) as 14.10m and height (H2) as 5.5m. The Fig.7 shows the poop deck starboard side drawing with marked affected area by red ink. It has been estimated replaceable area (A4) as 20m². The height (H3) is 5m and breadth (B2) as 4m. The estimated total replaceable area (A) is 322.55m². The steel plate thickness (t) is 12mm. hence estimated total volume (V) is 3.87m³. From the steel data hand book for LR Gr A steel, the density (ρ) taken as 7,880kg/m³. The Fig.8 to 16 illustrates the ship affected photographs.

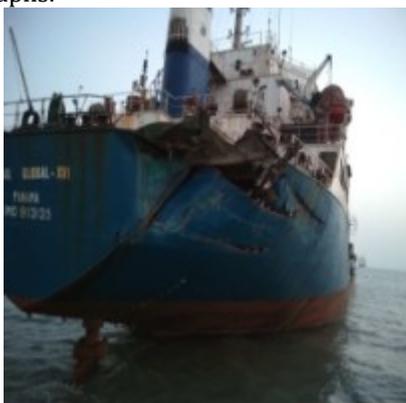


Fig.3: Effected MT Nautical Global XVI after collision with MV PaccSeginus.

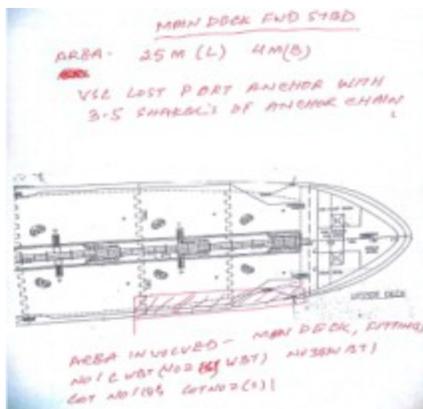


Fig.4: MT Nautical Global XVI main deckforward starboard affected area marked by red ink.

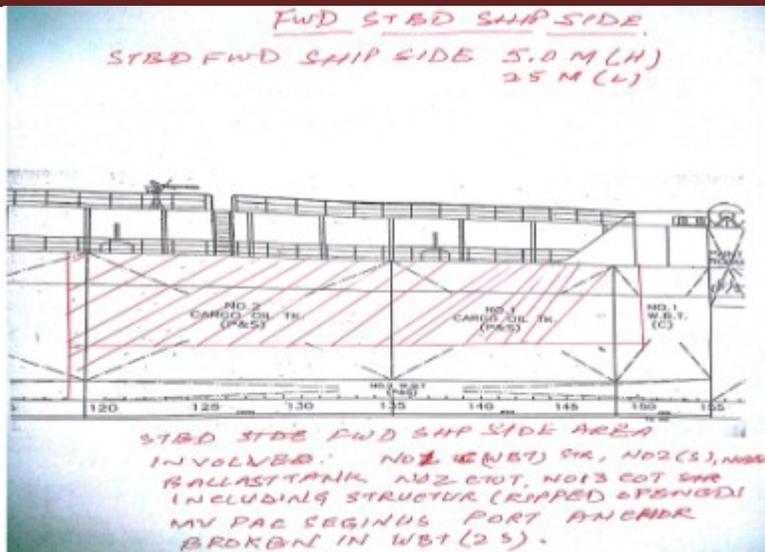


Fig.5: MT Nautical Global XVI forward Starboard ship side affected area marked by red ink.

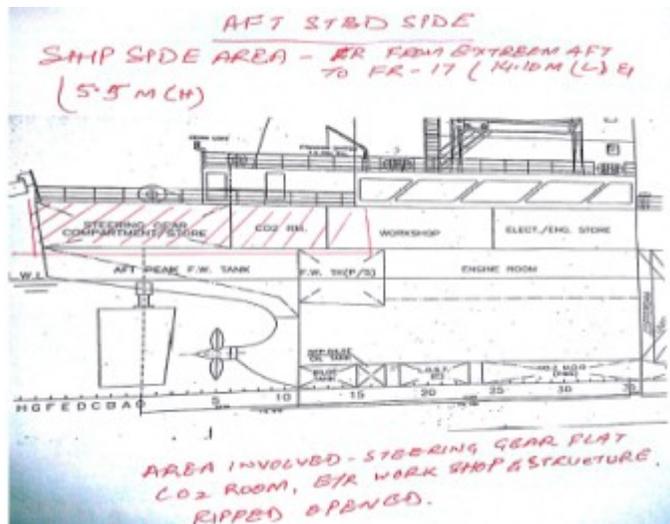


Fig.6: MT Nautical Global XVI aft Starboard side affected area marked by red ink.

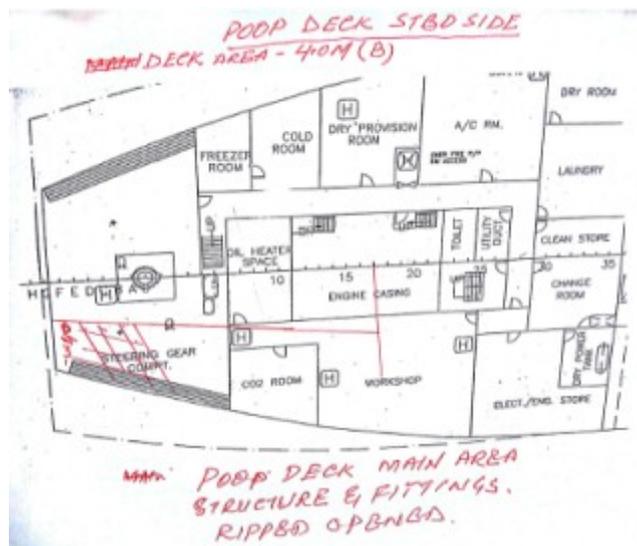


Fig.7: MT Nautical Global XVI poop deck Starboard side affected area marked by red ink.



Fig.8: MT Nautical Global XVI starboard side affected location in the ship photograph.



Fig.9: MT Nautical Global XVI starboard side affected location in the ship photograph.

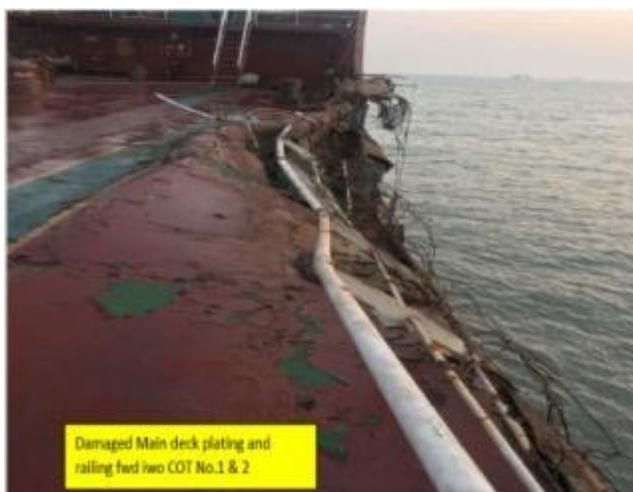


Fig.10: MT Nautical Global XVI main deck affected location in the ship photograph.

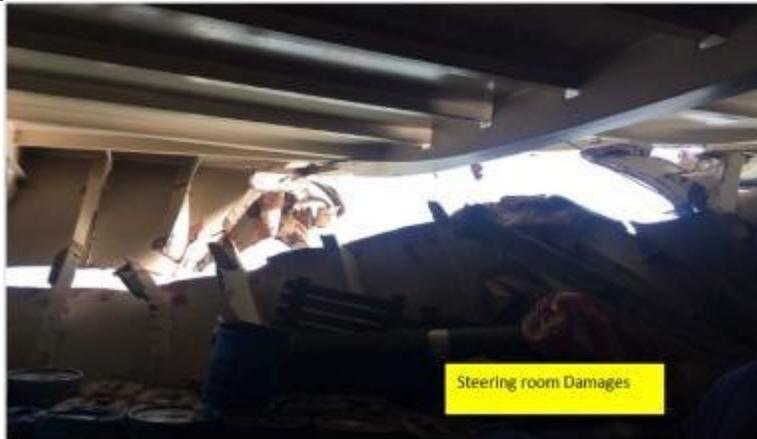


Fig.11: MT Nautical Global XVI steering room affected location in the ship photograph.



Fig.12: MT Nautical Global XVI steering room side shell plating affected location in the ship photograph.

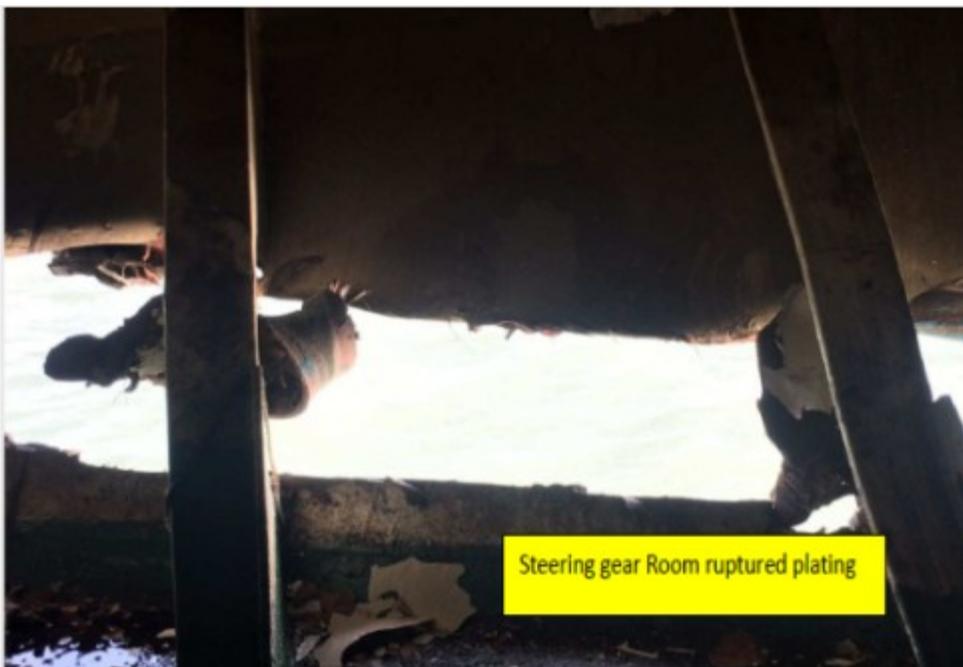


Fig.13: MT Nautical Global XVI steering gear room ruptured plating affected location in the ship photograph.



Fig.14: MT Nautical Global XVI buckled side vertical angles affected location in the ship photograph.

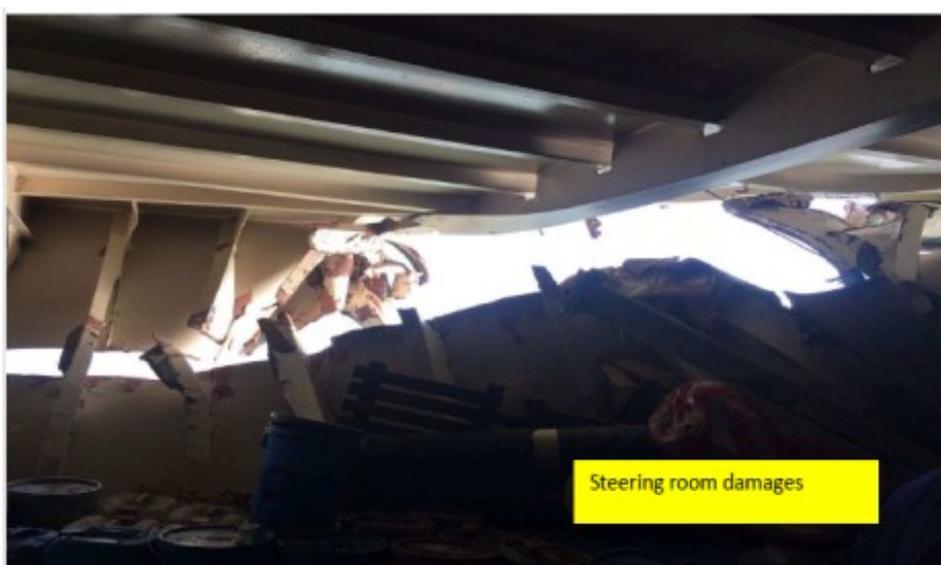


Fig.15: MT Nautical Global XVI steering room damaged location in the ship photograph.



Fig.16: MT Nautical Global XVI damaged bollard in the ship photograph.

III. METHODOLOGY

The replacement of over water steel plate, onboard steel plate bulk head structure is the main focus of the assigned job by the ship owner party in time to avoid revenue losses required which is required to be earning by the mother vessel MT Nautical Global XVI. After finalization of vendor it is required to be onboard steel materials, welding machine, power hand grinder, oxy acetylene cylinders, chipping tools, wire brush and other tools and tackles, welding machine cables, welding rod preheating oven. M/s Tangar ship technical Superintendent, welding team, fitters, helpers and supervisors to be onboard at the ship anchorage location for 24 hours until the job gets completed and handover of vessel being done to the onboard Captain of the MT Nautical ship owner. The port clearances and other formalities has to be attended by the ship agent on priority with shore supporting vehicles and crew boat to onboard the ship materials and personal movement as per requirement. The empty oxy acetylene cylinders refilling and onboard works attended by the crew boat from the Kandla port Jetty to the outer anchorage area where the vessel MT Nautical Global XVI has been anchored. The technical works involves visual check up of the vessel by the Tangar technical team and marking of the damage plates required to be crop. After marking the oxy acetylene has cutting had been attended for couple of days. Temporary scaffolding work carried out with the strong wooden planks on the scaffolding to perform cutting, fitting and welding operations at the above water side steel plate and bulk head structure replacement. One team of fitters cut down the specific required steel at particular location with clit preparation. The plate edge preparation also carried out by hand power grinding machine. The selected plates at the main affected starboard area alignment had been carried out with clit supports. After cutting bulk head angles got replaced by arc welding. Next work simultaneously attended as above water straight vertical plates fitting and onboard flat horizontal plates fitting. The selected WPS is shielded metal arc welding upto plate butt root gap of 1.2 to 2mm. In critical bending plate welding area where the aligning of butt steel joint gap over raised more than 2mm to 6mm has been redefined welding joining technique as FCAR. After finishing first root run the dry penetrant (DP) testing done. The results showed to the surveyor and technical superintendent and onboard Captain. Full run of welding done by day and night operation. At specific area gouging also carried out as per superintendent's instructions. Finally hydro testing of the cooled plates had been carried out by water jet at a pressure of 3kg/cm². Survey work closed and sea trail taken.

IV. RESULTS AND DISCUSSIONS

The rectification work of damaged steel and structure of MT Nautical Global was allotted to the M/S Tangar Ship Management. Ltd. The steel plates as per Lloyds class specification has been selected as LR Gr A. The steel plate material specification was cross checked by the M/s Metal analysis and services Pvt. Ltd. It was found steel major constituent responsible for tensile strength and weld ability that is carbon (C) content in between 0.14 to 0.16. The Fig. 17 illustrates the steel materials metallurgical certificate. The same also had been certified by the Lloyds inspector. The repairable steel mass was estimated as 31tons. Considering the other factors of steel plate cutting, clit preparation, scrap loss etc. it was estimated the requirement of steel plate, angles and pipelines as 42.894tons. To execute the project total 45tons of steels was been procured. During the execution of repair work 45tons of steel plates, angles and pipelines were placed onboard in the MT Nautical Global XVI. The Fig. 18 shows the initiating of MT Nautical Global XVI damaged steel plate gas cutting. The gas cutting of damaged plate and the shape of new plate was carried out simultaneously parallel by different team of works. After steel cutting, steel fitment and welding was initiated. The steel plate fitment at the starboard vertical side, initiating has been shown in the Fig. 19. The steel plate fitment by clit support has been illustrated in the Fig. 20. The welded plate at the starboard side of MT Nautical Global XVI has been illustrated in the Fig. 21.

METAL ANALYSIS AND SERVICES PRIVATE LIMITED S.No. 048
 975, Basse (Old) Wharf, S. T. Road, Basse (N), Mumbai - 400 012. Page 1 of 1
 Tel: (022) 24242329/24242331 Fax: 24242334
 Email: metalanal@rediffmail.com

TEST RECORD - CHEMICAL ANALYSIS RESULT SHEET

Report No. _____ Description: LR CHS Sec. 2 Cr. B

Client: Jaguar Ship Management Test Method: Wt: IS 228-PV-1, 2, 3, 5, 6, 7, 8, 9, ASTM E 399-0910, ASTM E 399-0912, MAGCHEMIST

Ref: _____ Spectro: E 4357 --- 40 8979 --- E 4357 --- 18 8811, E 1264, MAGCSTO

Sample Received on: 5-5-12 Equipment Used: 376-3490 / Spectro Apparatus / TIME CTO - 10 (Exhausted) / 976

Date of completion of Test: 5-5-12 Spectroscopic Apparatus: 204501163001, 021 / Conductivity meter Type 870

Tested By: WAC / SM / SV / SM / LM / HV Spectral: SM / SM / SM / SV / HV / SM

Sl. No.	Lab No.	Sample Identification / Description	Elements in Percentage					
			C	Mn	P	S	CP	
1	10445	10 mm Thk. m.s. Plate	0.15	0.24	0.010	0.024	0.026	0.30
2		10 - 83						
3	10446	10 mm Thk. m.s. Plate	0.14	0.24	0.010	0.023	0.021	
4		10 - 86						
5	10447	10 mm Thk. m.s. Plate	0.14	0.18	0.010	0.022	0.022	0.29
6		10 - 87						
7	10448	10 mm Thk. m.s. Plate	0.14	0.18	0.010	0.023	0.015	0.28
8		10 - 82						
9								
10								

Received: _____ MS: _____ MS: _____ MS: _____ MS: _____ MS: _____

* NON HAZ. PARAMETER Ctd by: P. K. K. For Metal Analysis & Services Pvt. Ltd.
 ** TOLERANCE APPLIED Signature & Name of Tech. Manager (Chemist)

Fig. 17: The steel materials metallurgical testing certificate.

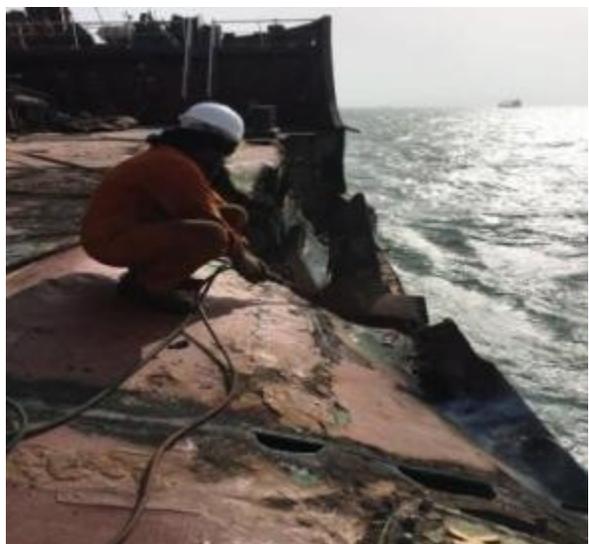


Fig. 18: MT Nautical Global XVI damaged steel plate cutting initiating.



Fig.19: Workers involved in steel plate fitment at the starboard side of MT Nautical Global XVI.

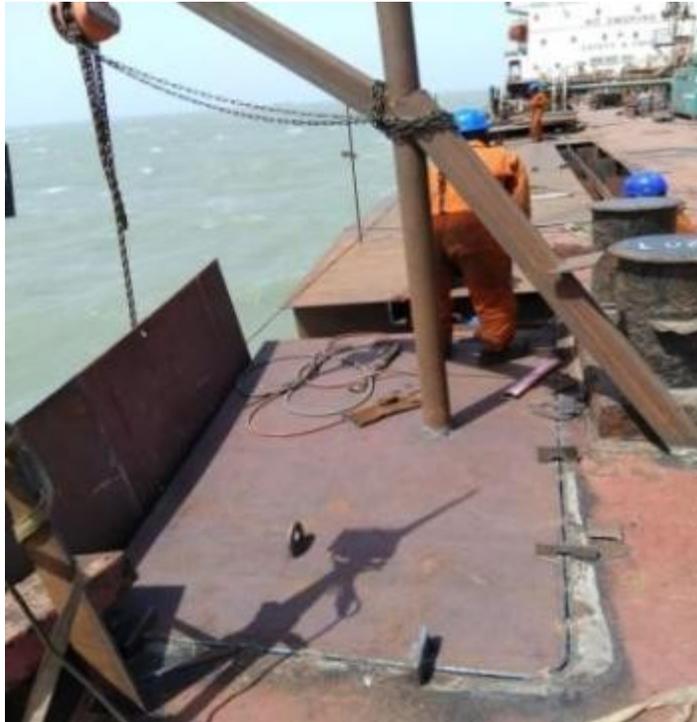


Fig.20: The steel plate fitment by using clit at the starboard onboard area.



Fig.21: MT Nautical Global XVI welded steel plate at the starboard side.

V. CONCLUSION

The heavily damaged oil tanker MT Nautical Global XVI checked out. It was identified the ship steel zones for the replacement by new steel as per class and MMD requirements. The requirements of steel were estimated. Approximately 44tons of steels were replaced by new one. Almost 2tons of steel spend for cutting allowances; scrap etc.The job was initiated and closed within the span of twenty days. The next survey of steel plates and machineries would be initiated after of three years during the dry docking as per class and MMD norms.

VI.. Acknowledgment

The authors are grateful to M/s Tangar Ship management to carry out the study project and analysis of the realistic fault detection and rectification with testing of Oil tanker in Indian sea water. Also the authors are

being grateful to MGM's College of Engineering authorities to allow us to carry out such study and analysis works and publication of same work in the journal. Not but the least, we are grateful to our family members, colleagues, friends and students for continuous moral support and encouragement.

References

1. IvanaGledić, JoškoParunov, PeroPrebeg, MaroĆorak, "Low-cycle fatigue of ship hull damaged in collision", *Engineering Failure Analysis*, vol.96, (2019), pp. 436-454.
2. AshkanBabazadeh, Mohammad Reza Khedmati, "Ultimate strength of cracked ship structural elements and systems: A review", *Engineering Failure Analysis*, vol. 89, (2018), pp. 242-257.
3. Y. Garbatov, F. Sisci, M. Ventura, "Risk-based framework for ship and structural design accounting formaintenance planning", *Ocean Engineering*, vol. 166, (2018), pp.12-25.
4. Kujala, F. Goerlandt, B. Way, D. Smith, M. Yang, F. Khan, B. Veitch, "Review of risk-based design for ice-class ships", *Marine Structures*, vol. 63, (2019), pp. 181-195.
5. Yan Liu, Dan M. Frangopol, "Probabilistic risk, sustainability, and utility associated with ship grounding hazard", *Ocean Engineering*, vol.154, (2018), pp. 311-321.
6. Bin Liu, Y. Garbatov, Ling Zhu, C. GuedesSoares, "Numerical assessment of the structural crashworthiness of corroded ship Hulls in stranding", *OceanEngineering*, vol.170, (2018), pp. 276-285.
7. Yan Liu, Dan M. Frangopol, Minghui Cheng, "Risk-informed structural repair decision making for service life Extension of aging naval ships", *Marine Structures*, vol.64, (2019), pp. 305-321.
8. Young JoonSeo, Min Ho Ha, Zaili Yang, Syamantak Bhattacharya, "The Ship Management firm Selection : The case of South Korea", *The Asian Journal of Shipping and Logistics*, vol. 34, (2018), pp. 258-268.
9. YanyanSha, Jorgen Amdalh, "Numerical investigations of a pontoon wall subjected to Ship collision loads", *Ocean Engineering*, vol. 172, (2019), pp. 234-244.
10. Isaias Vicente-Cera, Javier Moreno-Andres, David Amaya-Vias, Miriam Biel-Maeso, Marina G. Pintado-Herrerab, Pablo A. Lara-Martinb, Asuncion Acevedo-Merinoa, Juan Antonio Lopez-Ramireza, Enrique Nebota, "Chemical and microbiological characterization of cruise vessel wastewater discharges under repair conditions" vol. 169, (2019), pp. 68-75.
11. Rob Williams, Scott Veirs, Val Veirs, Erin Ashe, Natalie MastickOceans, "Approaches to reduce noise from ships operating in important killer whale habitats, vol. xxx (xxxx), pp. xxx-xxx.

Performance Evaluation of Association Rule Mining Using Weka Tool

¹Bina Trivedi, ²Priti Nandankar, ³Anjali Singh, ⁴Rachel Mascarenhas, ⁵Dr. Nazneen Ansari

^{1,2,3,4} Under Graduate Student⁵Associate Professor

^{1,2,3,4,5}Department of Information Technology,

^{1,2,3,4,5}St. Francis Institute of Technology, Mumbai, India

ABSTRACT: In recent times, there have been many advancements in the flourishing field of data mining. Association rule mining is one of the most popular and well researched technique in this field for determining frequent patterns or associations between data items in voluminous collection of datasets. It aims to discover correlations in data by finding rules that enable us to predict the occurrence of a particular item based on the occurrence of other items in the dataset. In this paper, apriori and FP growth L algorithm are selected for analysis. Their performance is evaluated using Weka 3.8 which is a java based machine learning tool. We have considered the parameters like number of rules supported by the algorithm, grouping of attributes, generation of data, number of scans etc.

Keywords: Data mining, Association Rule, Evaluation, Weka tool, Apriori, FP-growth

I. INTRODUCTION

Data mining is from a large number of incomplete, noisy, fuzzy and random data extract implicit useful information and knowledge in which people do not know in advance but is potential[1]. Data Mining, is defined as a large amount of Data in Data warehouse, or other information obtained potentially useful and ultimately understandable patterns of nontrivial process[2]. Although data mining is used in various field of medical science, it has not still been well known and its advantages has not been adequately discussed. The huge amount of different medical records of each need to stored in databases and data warehouse. The advanced system also include the status of patient regularly. The data which is stored in such a system that may contain valuable knowledge hidden in medical records [3].

The problem is of prediction in any other system like banking sector, which can be divided into two parts : The first one is learning part and second part is decision making. In the learning part, a huge data set is converted into a smaller data set. Number of features and objects in this new set are simplified than the original set in different ways [4]. This learning part is important to make accurate results. This Data sets is used for prediction when unknown results occur with different algorithm. Using modern technology, information can be easily found and also secure. No one can access this information without permission.

The study presented a technique of discovering association rules of diabetes mellitus (DM) with complication or linkage among groups of data[5]. The efficiency of data mining algorithm method is an important in data mining in knowledge discovery. The main objective of the research is to evaluate the data mining algorithm using association rule algorithm on bank data using weka tool. The following steps of data mining need to be carried out in order to apply data mining [6]:

A. Data Integration:

First of all the data are collected and integrated from all the different sources. Data integration involves combination of different data source from various data set, which get stored by using different technologies. The old data get merged with new data.

B. Data Selection:

This is an important step because the results comes from selection of appropriate and high quality data. So in this step we select only those data which is useful for data mining.

C. Data Cleaning:

The database may contain irrelevant information which need to be removed. The data which we collected may contain various errors, missing values, noisy or inconsistent data. Noisy and inconsistent records also need to be removed. So we need to apply different techniques to remove such inconsistency present in data.

D. Data Transformation:

The data even after cleaning are not ready for mining as we need to transform from one format to appropriate forms for mining. The techniques which are used for transforming the data are smoothing, aggregation, normalization etc.

E. Data Mining:

After data transformation ,it is ready to apply data mining techniques on the data to discover the interesting patterns. Techniques like clustering and association analysis are among the many different techniques used for data mining.

F. Pattern Evaluation and Knowledge Presentation

This step involves visualization, transformation, removing redundant patterns etc from the patterns we generated. These kinds of results are hypothesis at this stage.

G. Knowledge Testing and Evaluation

This step helps user to make use of the knowledge acquired to take better decisions and evaluated on the principle of training and test data sets.

This paper consist of following sections: Section II contain Literature Review for evaluating most commonly used data mining algorithm in medical support system. Section III contain Implementation of apriori and FP tree using WEKA tool. Section IV contain Experimental Result in which Evaluation is done using bank data set on weka tool. Section V contains Comparative Analysis and Section VI contain Conclusion.

II. LITERATURE REVIEW

Lan Yu et al.[7] , In this paper researchers are focusing on a data mining task on test data of PHS by using Microsoft Association Rules algorithm in the database SQL Server 2005. The grades of vital capacity, grip strength, standing long jump and step test of a student are used for input attributes, and total score of the student is used for evaluation attribute. By using pattern recognition, statistical and mathematical techniques to sift workstation. Data mining helps to predict relationships or comparison, trends, patterns, exceptions and anomalies that might otherwise go unnoticed [8].

Dr. Mukesh Kumar [9] compared two association rule mining algorithms i.e apriori and predictive apriori algorithms .He applied the two algorithms on 15 different datasets and compared them using different measures of accuracy .He calculated various statistical measures using mat lab. Based on the results, he concluded that predictive apriori performs better as compared to apriori.

Chenlu Li et al. [10], in this paper, researchers focused on the medical system which has gradually become an important part of modern medical information system. The importance of information in the medical and healthcare information is focused, in which association rules mining is more widely used in medical applications. This paper will mine positive and negative association rules on medical and healthcare data. Because of huge databases, the traditional Apriori algorithm of association rules mining are difficult to give accurate and good outputs. These rules are mainly focusing on positive association rules, which can give only partial output. Where as ,the combination of negative association rules give more valuable data.

P. Kasemthaweesab et al.[11], this paper has presented a basic model for discovering an association of diabetes mellitus disease with states by applying various attributes like gender, age and occupation factors and tested it for finding the relationship between diagnosis data. This paper aims a brief theory of creating an association rule of diagnostic data of diabetes mellitus (DM). Data mining method via association rule was processed by WEKA tool . Understand the types of diabetes mellitus (DM) causing ophthalmic, renal and neurological, etc.

Cornelia Gyorodi et al. [12], they present a comparative study of association rules mining. A large number of data instances is considered. It proposes an evaluation of three association rule algorithms ie. Apriori, FP-growth and DynFP-growth algorithms .Different datasets is been examined and implemented on java . The platform uses Pentium 4 1.7GHz processor, with 256 MBRAM, Windows 2000 for evaluation of algorithms .10,000 to 5,00,000 instances and support factor between the range of 5% to 40%was used for evaluation .For the evaluation different number of instances is considered each time of similar data set and compares the execution time for all the three algorithms ,and other factor that was considered is support factor .The experiment concluded that DynFp-Growth performs better than FP-growth and also the performance of FP-growth decreases with support factor and performance of FP-growth increases with increases with support factor.

Lijun et al.[13], they have proposed study based on the examination of association rule algorithm. they consider the algorithm like AIS, SETM, Apriori , AprioriTid , AprioriHybrid , DIC, Partition and FP-growth algorithms. Analysis is done based on the advantages and disadvantages of algorithm. The outcome of this system is that setm is unsuitable in all the algorithms but it is very appropriate when added with the database management system and examined that the FP-growth is faster than the apriori algorithm.

Mukesh Sharma et al. [14] compared two association rule mining algorithms i.e apriori and predictive apriori algorithms .He applied the two algorithms on 15 different datasets and compared them using

different measures of accuracy .He calculated various statistical measures using matlab.Based on the results, he concluded that predictive apriori performs better as compared to apriori.

Tanna and Ghodasara [15] applied apriori algorithm using WEKA for mining frequent patterns from the transaction. The paper described that apriori algorithm is a simple to understand and effective algorithm. The paper presented the implementation of apriori algorithm using WEKA for association rule mining.

III. IMPLEMENTATION OF APRIORI AND FP TREE USING WEKA TOOL

Weka tool version 3.8.3 is been used for evaluation of algorithms. Following steps are executed for evaluation of association rule mining algorithm:

- The data is been loaded in weka tool and preprocessed.
- Apriori algorithm is been applied on the preprocessed data.
- By setting the values of number of rules and support factor we concluded the evaluation of apriori algorithm.
- FP-Growth algorithm is applied on same preprocessed data.
- The attributes of preprocessed data is changed from nominal to binary and then numeric to binary to start FP-Growth algorithm.
- By setting the values of number of rules and support factor we concluded the evaluation of FP-Growth algorithm.
- Finally, we compare the results displayed on weka explorer.

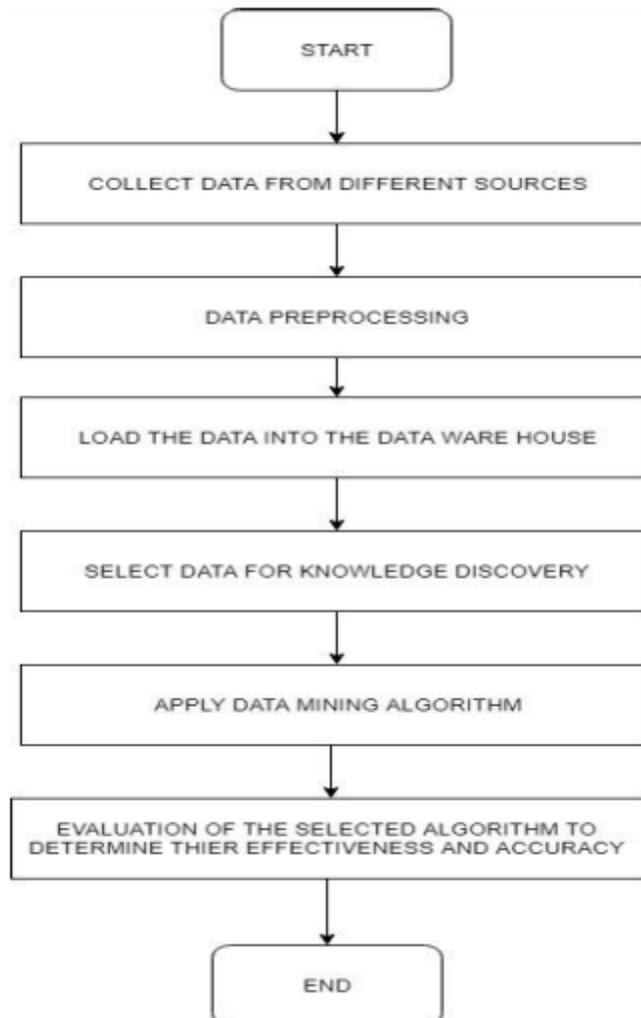


Fig. 1.Flow Chart for Implementation of algorithm

IV. EXPERIMENTAL RESULTS

Evaluation is done using bank data set on weka tool .The data set are preprocessed and then evaluation is done on data mining association rule algorithm.

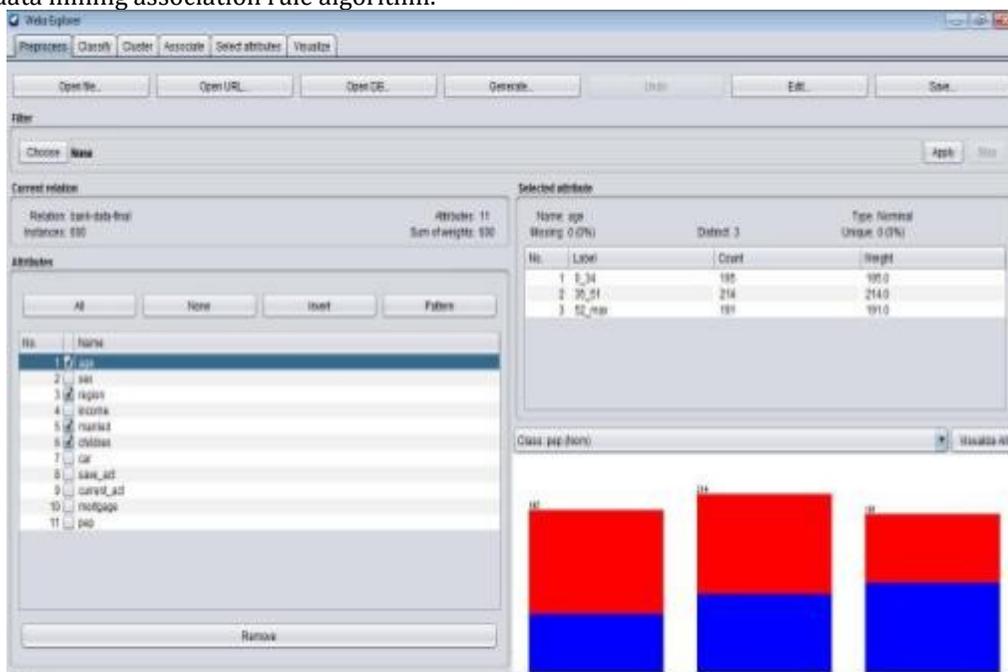


Fig. 2. Data after preprocessing

The apriori algorithm is applied on the preprocessed dataset by keeping support factor 0.1 and number of rules are 10 where set for start the execution of algorithm .The output of apriori algorithm as follow:

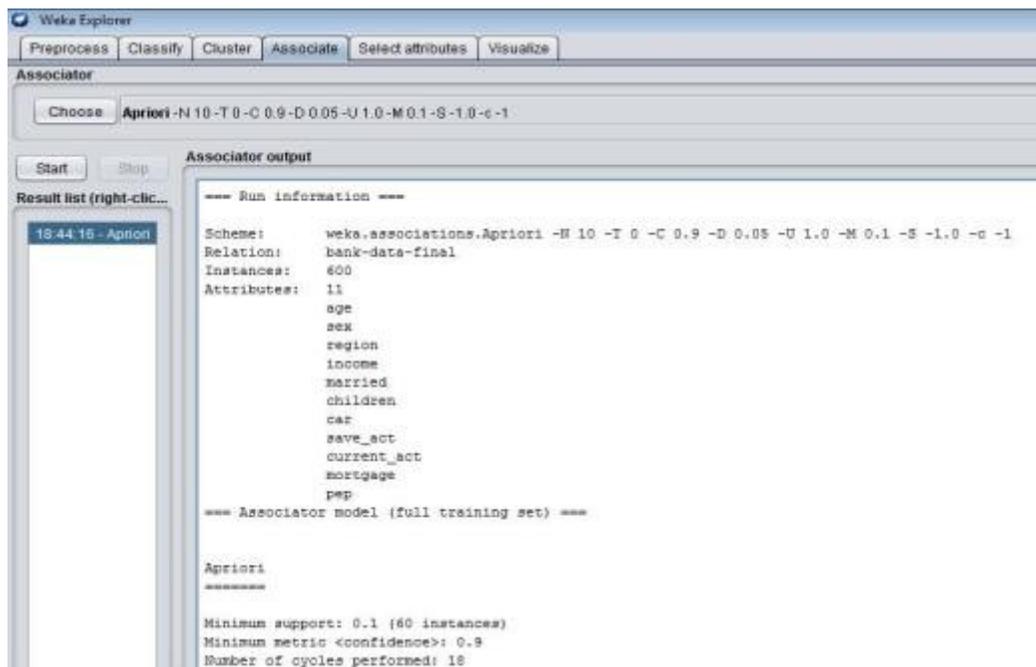


Fig.3 Apriori algorithm output

```

Generated sets of large itemsets:

Size of set of large itemsets L(1): 28

Size of set of large itemsets L(2): 232

Size of set of large itemsets L(3): 524

Size of set of large itemsets L(4): 277

Size of set of large itemsets L(5): 33

Best rules found:

1. income=43759_max 80 ==> save_act=YES 80 <conf:(1)> lift:(1.45) lev:(0.04) [24] conv:(24.8)
2. age=52_max income=43759_max 76 ==> save_act=YES 76 <conf:(1)> lift:(1.45) lev:(0.04) [23] conv:(23.56)
3. income=43759_max current_act=YES 63 ==> save_act=YES 63 <conf:(1)> lift:(1.45) lev:(0.03) [19] conv:(19.53)
4. age=52_max income=43759_max current_act=YES 61 ==> save_act=YES 61 <conf:(1)> lift:(1.45) lev:(0.03) [18] conv:(18.91)
5. children=0 save_act=YES mortgage=NO pep=NO 74 ==> married=YES 73 <conf:(0.99)> lift:(1.49) lev:(0.04) [24] conv:(12.58)
6. sex=FEMALE children=0 mortgage=NO pep=NO 64 ==> married=YES 63 <conf:(0.98)> lift:(1.49) lev:(0.03) [20] conv:(10.88)
7. children=0 current_act=YES mortgage=NO pep=NO 82 ==> married=YES 80 <conf:(0.98)> lift:(1.48) lev:(0.04) [25] conv:(9.29)
8. children=0 mortgage=NO pep=NO 107 ==> married=YES 104 <conf:(0.97)> lift:(1.47) lev:(0.06) [33] conv:(9.1)
9. income=43759_max current_act=YES 63 ==> age=52_max 61 <conf:(0.97)> lift:(3.04) lev:(0.07) [40] conv:(14.31)
10. income=43759_max save_act=YES current_act=YES 63 ==> age=52_max 61 <conf:(0.97)> lift:(3.04) lev:(0.07) [40] conv:(14.31)
    
```

Fig.4 Best rules generated for apriori algorithm

The FP-Growth algorithm is applied on the preprocessed dataset, but before applying the algorithm on dataset we need to convert the attribute type from nominal to binary and binary to numeric and then FP-Growth algorithm will execute on the data set.

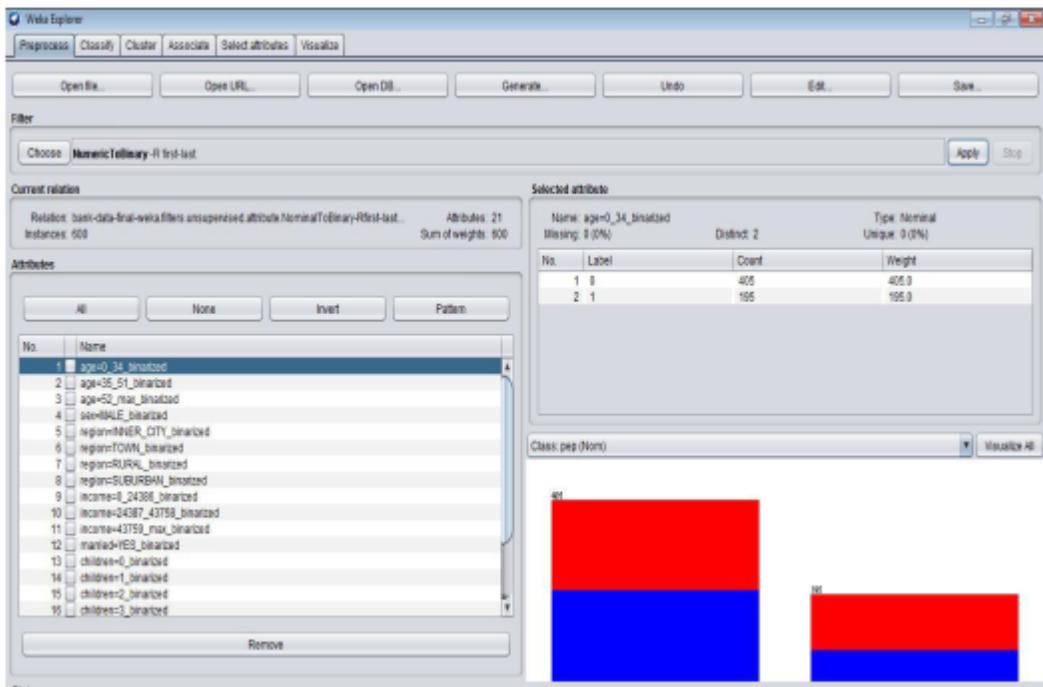


Fig.5 Dataset after changing attribute type

By keeping support factor 0.1 and number of rules are 10 where set for start the execution of algorithm. The output of FP-Growth algorithm is as follows:

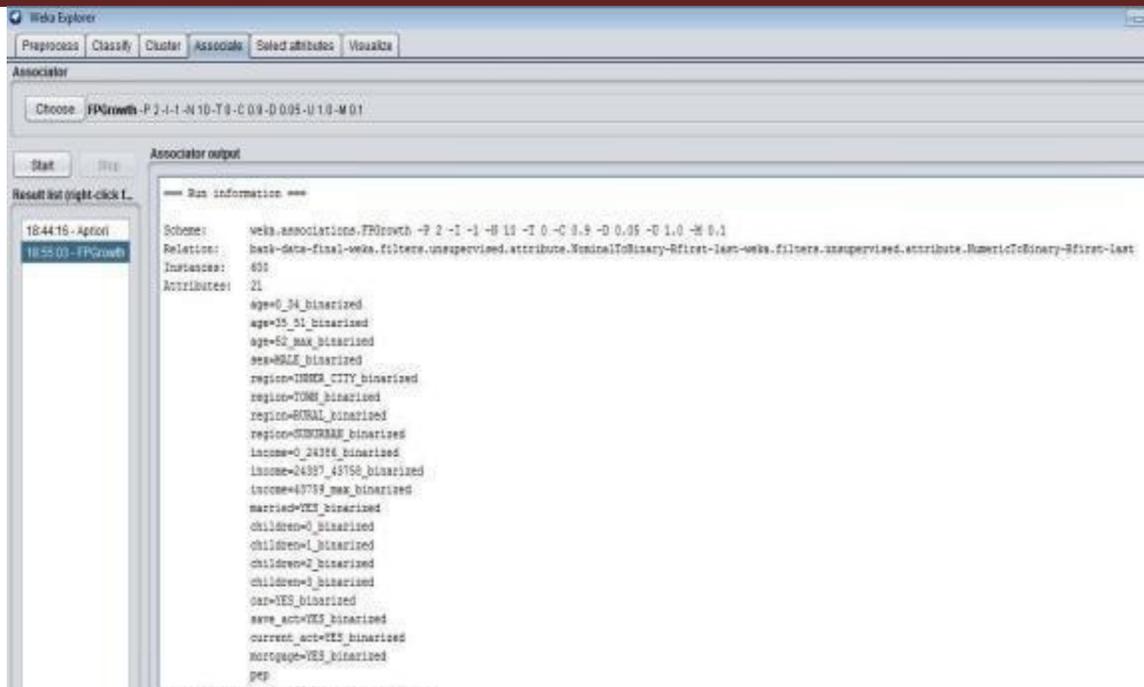


Fig.6 FP-Growth output

```

=== Associator model (full training set) ===

FPGrowth found 21 rules (displaying top 10)

1. [income=43759_max_binarized=1]: 80 ==> [save_act=YES_binarized=1]: 80 <conf:(1)> lift:(1.45) lev:(0.04) conv:(24.8)
2. [current_act=YES_binarized=1, income=43759_max_binarized=1]: 63 ==> [save_act=YES_binarized=1]: 63 <conf:(1)> lift:(1.45) lev:(0.03) conv:(19.53)
3. [age=52_max_binarized=1, income=43759_max_binarized=1]: 76 ==> [save_act=YES_binarized=1]: 76 <conf:(1)> lift:(1.45) lev:(0.04) conv:(23.56)
4. [current_act=YES_binarized=1, age=52_max_binarized=1, income=43759_max_binarized=1]: 61 ==> [save_act=YES_binarized=1]: 61 <conf:(1)> lift:(1.45) lev:(0.03) co
5. [current_act=YES_binarized=1, income=43759_max_binarized=1]: 63 ==> [age=52_max_binarized=1]: 61 <conf:(0.97)> lift:(3.04) lev:(0.07) conv:(14.31)
6. [current_act=YES_binarized=1, income=43759_max_binarized=1]: 63 ==> [save_act=YES_binarized=1, age=52_max_binarized=1]: 61 <conf:(0.97)> lift:(3.85) lev:(0.06)
7. [current_act=YES_binarized=1, save_act=YES_binarized=1, income=43759_max_binarized=1]: 63 ==> [age=52_max_binarized=1]: 61 <conf:(0.97)> lift:(3.04) lev:(0.07)
8. [income=43759_max_binarized=1]: 80 ==> [age=52_max_binarized=1]: 76 <conf:(0.95)> lift:(2.98) lev:(0.08) conv:(10.91)
9. [income=43759_max_binarized=1]: 80 ==> [save_act=YES_binarized=1, age=52_max_binarized=1]: 76 <conf:(0.95)> lift:(3.77) lev:(0.09) conv:(11.97)
10. [save_act=YES_binarized=1, income=43759_max_binarized=1]: 80 ==> [age=52_max_binarized=1]: 76 <conf:(0.95)> lift:(2.98) lev:(0.08) conv:(10.91)
    
```

Fig.7 Best Rule Generated for FP-Growth algorithm

By considering the parameters like memory utilization, execution time, understandability of output and visualization of output, we have done the evaluation. By considering all these parameters, we evaluate that the FP-Growth algorithm is better in performance, but in terms of understandability, the Apriori algorithm is easy to understand. FP-Growth Algorithm is better because it shows the best rules among all the rules found, and shown in Figure 7, the memory utilization is less in this algorithm because only best rules are generated and in case of apriori more rules are generated.

V. COMPARATIVE ANALYSIS

Sr No.	Parameters	Apriori	FP-growth
1.	Technique	It uses apriori property to joins and trims	It constructs pattern tree and pattern base from database.

		property.	
2.	Memory usage	Since more rules are generated large	Since only best rules are generated less memory is required.
		memory is required..	
3.	Execution time	Execution time is more as more rules are	Execution time is less as only best rules are
		generated.	generated ,depending on number of attribute
4.	Understandability	It has better understandability	It has less understandability as compared to apriori algorithm
5.	Visualization	One can view the actual values of	One can view the lowest and maximum
		attributes easily	value of the attribute
6.	Search method	Breadth-first search	Depth-first search
7.	No. of iterations	More number of iterations	Less of number of iterations

VI. CONCLUSION

It is concluded that Association Rule Mining is used to identify strong rules discovered in databases using some measures of uniqueness. The experimental results support the above mentioned statement. The algorithms used are apriori and FP-growth algorithm. It is inferred that from the experiment performed, FP-growth is better than apriori algorithm in terms of performance. It is possible to further extend this research by applying classification and clustering algorithms for evaluation.

VII. ACKNOWLEDGMENTS

We would like to thank our mentor, Prof. Nazneen Ansari for her great support and encouragement.

REFERENCES

- [1] Han, J., & Kambr, M. (2001). The concept and technology of data mining and. Xiao-f eng Meng translation.
- [2] Xiao-Li, M., & Xiao-Lei, Z. (2015, June). The application of data mining technology in computer network security. In 2015 Seventh International Conference on Measuring Technology and Mechatronics Automation (pp. 1126-1129). IEEE.
- [3] Dewan, A., & Sharma, M. (2015, March). Prediction of heart disease using a hybrid technique in data mining classification. In 2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom) (pp. 704-706). IEEE.
- [4] Ranganatha, S., Raj, H. P., Anusha, C., & Vinay, S. K. (2013). Medical data mining and analysis for heart disease dataset using classification techniques.
- [5] Kasemthaweesab, P., & Kurutach, W. (2012, July). Association analysis of diabetes mellitus (DM) with complication states based on association rules. In 2012 7th IEEE Conference on Industrial Electronics and Applications (ICIEA) (pp. 1453-1457). IEEE.
- [6] M. Kantardzie and A. Srivastava (2009), Data Mining: concepts, models, methods, and algorithms, Journal of Computing and Information Science in Engineering, vol. 5, pp. 394-395.
- [7] Yu, L. (2009, April). Association rules based data mining on test data of physical health standard. In 2009 International Joint Conference on Computational Sciences and Optimization (Vol. 2, pp. 322-324). IEEE.
- [8] Tong, Q., Zhou, Y. C., Wu, K. C., & Yan, B. P. (2007). Scientific Data Mining Grid Service Architecture. Jisuanji Yingyong Yanjiu/ Application Research of Computers, 24(6), 25-29.
- [9] Patel, S., & Patel, H. (2016). Survey of data mining techniques used in healthcare domain. International Journal of Information, 6(1/2), 53-60.
- [10] Li, C., Hao, F., Zhao, L., Song, L., & Dong, X. (2017, July). Analysis of medical and healthcare data based on positive and negative association rules. In 2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD) (pp. 1559-1564). IEEE.
- [11] Kasemthaweesab, P., & Kurutach, W. (2012, July). Association analysis of diabetes mellitus (DM) with complication states based on association rules. In 2012 7th IEEE Conference on Industrial Electronics and Applications (ICIEA) (pp. 1453-1457). IEEE.
- [12] Györfödi, C., Györfödi, R., & Holban, S. (2004). A comparative study of association rules mining algorithms. In SACI 2004, 1st Romanian-Hungarian Joint Symposium on Applied Computational Intelligence (pp. 213-222).

- [13] Lijun, H., Linghua, L., Xiaoniu, L., & Degao, W. (2009, April). Comparison and Analysis of algorithms for Association Rules. In 2009 First International Workshop on Database Technology and Applications (pp. 196-198). IEEE.
- [14] Sharma, M., Choudhary, J., & Sharma, G. (2012). Evaluating the performance of apriori and predictive apriori algorithm to find new association rules based on the statistical measures of datasets. International Journal of Engineering Research & Technology (IJERT), 1(6).
- [15] Tanna, P., & Ghodasara, Y. (2014). Using Apriori with WEKA for frequent pattern mining. arXiv preprint arXiv:1406.7371.

A Review of Mathematical Multi-Criteria Decision Models with A case study

¹Udaykumar S. Kashid, ²Archana U. Kashid, ³Saurabh N. Mehta,

¹Research Scholar, Department of Mathematics, Shri J.J.T. University, Jhunjhunu, Rajasthan, India,

¹Assistant Professor, Vidyalkar Institute of Technology, Mumbai, India

²Assistant Professor, Bharati Vidyapeeth's College of Engineering, Navi Mumbai, India,

³Research Guide, Shri J.J.T. University, Jhunjhunu, Rajasthan, India,

³Professor, Vidyalkar Institute of Technology, Wadala, Mumbai, India.

ABSTRACT: Multi-Criteria Decision Making/Analysis (MCDM/MCDA) is a Core of Decision Theory, as an important field of Operational Research (OR) in Mathematics. Multi-Criteria Decision model is a technique that permits Decision Makers (DM) to make the best decision in the presence of multiple, potentially conflicting criteria. These MCDM methods have attracted much attention from Decision makers, academics, researchers and practitioners. Multi-Criteria Decision model is used to solve areas problems such as sciences, engineering, technology, economics, military strategies, business, supplier selection, Sports entity and many more. This MCDM is a Method of evaluating conflict real world situations, based on different quantitative, qualitative, criteria under risky, certain, uncertain environments. In this paper we have discussed the some well noted and discussed the results of Multi-Criteria Decision Making (MCDM) techniques with their classifications and characteristics. Moreover, we have compared few MCDM methods with a case study on IPL 2018. The main objective of this paper is multifold: First, to finds the applicability of MCDM methods in different situations by evaluating their relative strengths and weaknesses. Second, comparisons of MCDM methods using a case study. Our study concluded with an observation that the different Multi-Criteria Decision methods present different ranking for the same problem with same multiple attributes and same multiple alternatives although same decision maker. However, this paper does not claim that any method is better than other methods across all possible circumstances, but rather it emphasizes the importance of investigating different Multi-Criteria Decision-Making techniques to rank the decisions of each method and the importance of finding the most appropriate method for ranking the decisions in consideration of the decision-making conditions.

Keywords: Operations Research, Decision theory, Multi-Criteria Decision Making (MCDM), Decision matrix, WSM, WPM, Analytical Hierarchy Process (AHP), Pairwise Comparisons.

I. Introduction

I.1. Motivation:

Operation Research is a relatively new Branch of Applied Mathematics discipline. Decision theory is the main tool, technique of Operation Research where Multi-Criteria Decision Making/Analysis (MCDM/MCDA) is a Core of Decision Theory. Multi-Criteria Decision Making (MCDM) models has incredible grown as a part of Operations Research, concerned with developing computational and Mathematical tools for supporting the subjective evaluation and selection of performance criteria by decision-makers. In our real life the main activity of a manager is the decision making. Many times, we make the decisions even without noticing them. The decisions are taken simply by common sense with imprecise knowledge of problem, judgment and expertise without using any mathematical or any other model in simple situations. But the decision we are concerned here for MCDM in with are complex and heavily responsible environment.

Selection of best Alternatives in a conflict environment is always a difficult task for Decision manager, because many qualitative and quantitative attributes in the best Alternatives selection process need to be considered. Although team management/Decision manager uses a variety of assessments such as present performance, past performance, and manager's judgments etc. depends on situations. To select and evaluate best Alternatives, quantitative approaches could greatly benefit the decision makers or team management and be significantly useful. To avoid the traditionally vague and subjective practice of best alternative selection, quantitative approaches could provide more objective results in a shorter time-frame. To accomplish these activities of performance analysis and optimal selection of alternatives, proper decisions require to be made to determine what actions need to be performed and how they are carried out so that the desired goal can be achieved. As a result, decision making becomes an essential part of the problem-solving procedure and a systematic approach such as mathematical modelling or the Multi-Criteria Decision-Making/Analysis (MCDM/MCDA) approach is essential.

I.2.MCDM and MCDA: BACKGROUND

There are two schools of decision analysis methods: Multi-Criteria Decision Making (MCDM) developed by the American school (C. Zopounidis and P. Pardalos. Handbook of Multi-Criteria Analysis. Springer, 2005), and Multi-Criteria Decision Analysis/Aid (MCDA) created by the European school (B. Roy. Decision-aid and decision-making. European Journal of Operational Research,45:324-331,1990). Most researchers use MCDM and MCDA interchangeably. In this research paper, the American school (MCDM) is followed.

Despite the criticism that multi-dimensional Techniques have received, some of them are widely used in Decision process. The Weighted Sum Model (WSM) is the earliest and probably the foremost widely used technique. The Weighted Product Model (WPM) can be considered as a modification of the WSM and has been developed to overcome some of its limitations. The Analytic Hierarchy Process (AHP), as proposed by Saaty [22,23] and is a later development 1994 has recently become increasingly Applicable in Decision theory. Professors Belton and Gear [1983], [6] suggested a modification to the AHP that appears to be more powerful than the original approach. Some other widely used methods are the ELECTRE [Benayoun, et al., 1966] and TOPSIS [Hwang and Yoon, 1981] and so on.

The MCDM techniques are usually categorized into two classes: Multi-Attribute Decision Making (MADM) and Multi-Objective Decision-Making (MODM) as shown in figure 1. The MODM methods are used to deal with continuous MCDM problems, and the MADM methods are utilized to handle discrete MCDM problems (Rezaei, 2015). In this MCDM class, we are usually confronted with a decision-matrix which contains information about several alternatives with respect to many criteria or attributes. The aim of the MCDM methods is to evaluate the alternatives and/or criteria.

In literature, many terms have been used for MCDM and these terms are given as below:

- Multi-Criteria Decision Analysis (MCDA)
- Multi-Objective Decision Making (MODM)
- Multi-Attributes Decision Making (MADM)
- Multi-Dimensions Decision-Making (MDDM)

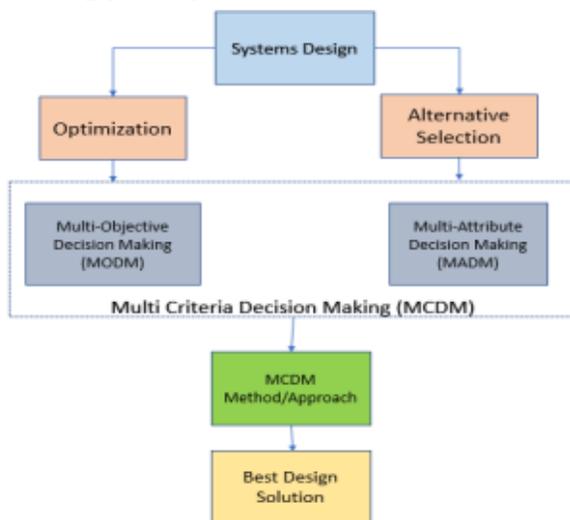


Fig.1. Multicriteria Decision Making types

Although MCDM as a subdiscipline of Operational Research (OR) has a relatively short history of about 48 years, over 70 MCDM techniques have been developed for facilitating the decision-making process (C. Zopounidis and P. Pardalos. Springer, 2005). There are many MCDM Techniques available in the literature. Each Technique has its own characteristics. There are many ways one can classify MCDM Technique. One way is to classify them according to the type of the data they use and Another way of classifying MCDM Technique is according to the number of decision makers involved in the decision process. Hence, we have single decision maker MCDM methods for involvement of one decision taker and group decision making MCDM for involvement of multiple decision takers. (for more information refer the journal of **Group Decision Making**). They are also classified according to the type of information and the salient features of the information as shown in figure 2.

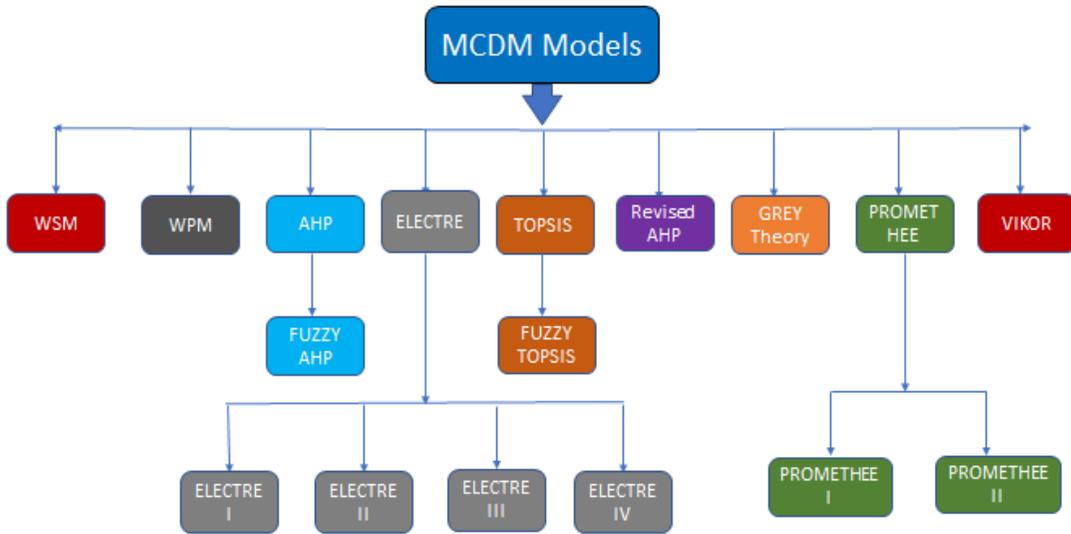


Fig.2 Hierarchical structure of MCDM Models

II. LITERATURE REVIEW:

Many MCDM techniques have been proposed during the past decades. Here, we briefly review few important MCDM techniques which are used for criteria weight determination and alternative evaluation based on the information of the decision-matrix. These are few methods mentioned here as, **SAW** (Simple Additive Weighting), **WASPAS** (Weighted Aggregated Sum Product ASsessment), **COPRAS** (COmplex PROportional ASsessment), **TOPSIS** (Technique for Order of Preference by Similarity to Ideal Solution), **VIKOR** (from Serbian: ViseKriterijumska Optimizacija I Kompromisno Resenje), **ELECTRE** (in French: ELimination Et Choix Traduisant la REalite), **BWM** (Best–Worst Method) and **EDAS** (Evaluation based on Distance from Average Solution), which are used for the evaluation of alternatives.

Besides the MCDM methods used for evaluation of alternatives, there are some other methods for determination of objective weights of criteria. The Standard Deviation (**SD**), **Entropy**, Analytical Hierarchy Process (**AHP**), and Artificial Neural Network (**ANN**) are few common methods which determine the objective weights based on the within-criterion variation information and assign smaller weights to a criterion if it's similar values across alternatives. **CRITIC** (CRiteria Importance Through Inter-criteria Correlation) is another method which considers within- and between-criterion variation information (Diakoulaki et al., 1995). The correlation between criteria is used to measure variations between criteria. In this CRITIC method, lower values of correlation have a positive effect on the weight of each criterion. SAW and TOPSIS had statistically similar performances. ELECTRE wasn't desirable in providing full, sorted ranks among the alternatives. VIKOR allotted identical ranks for many alternatives; once full, sorted ranks are needed, VIKOR is unfavorable. Notably, no ranking methodology was considerably sensitive to uncertainty levels once uncertainty modified symmetrically. The range of the MCDM methods is wider than the mentioned methods. However, these methods are reviewed because we use them in the research study. The details of noted work/results are summarized in Table 1.

Table.1. MCDM methods with its merits and demerits

Sr.no.	Method	Merits	Demerits	Description
1	The Weighted Sum Model (WSM) [8], [11], [20]	1.Additive utility hypothesis 2.Simple Mathematical calculations 3. Relative order of magnitude of the standardized scores remains same	1.Difficulty emerges on multi-dimensional problems 2.Single Dimension 3.ranges across criteria 4.Same unit	useful for evaluating several alternatives in accordance to the different criteria which are expressed in the same unit.
2	The Weighted Product Model (WPM) [11], [20], [29]	1. Multiplicative utility hypothesis 2.Like WSM, easy Calculation 3.Relative values are used rather than actual ones.	1. Subjective selection of criteria weight 2.If no of alternatives is large then it's lengthier and more difficult to solve. 3.No solution with equal weight of decision matrices	Comparison of with the other by the weights and ratio of one for each criterion.
3	Analytic Hierarchy Process (AHP)	1.Flexible, intuitive and checks	1.Very difficult to obtain	Includes pair wise

	[6], [7], [10], [22], [24]	inconsistencies 2. Problem is constructed into a hierarchical structure 3. No bias in decision making 4. not dataintensive.	Perfect consistency 2.Irregularities in ranking. 3.Doesn't consider theuncertainty, Incomplete Information 4. Rankreversal 5. Dependence on human judgment	comparison of different alternatives for different criterion and develops a linear additive model
4	Data Envelopment Analysis (DEA) [14], [27]	1. Relation between inputs and outputs are not necessary 2.Capable of handlingmultipleinputs and outputs 3.Inputs and outputs can have different units	1. Absolute efficiency cannot be measured 2.Statistical tests are not applicable 3. Does not deal with imprecise data	Used to find the efficiency of combination of multi inputs and multi outputs of the problem
5	Analytic Network Process(ANP)[19], [28], [29], [30]	1.Independence among criteria is not required. 2.Prediction is accurate due to priorities are improved by feedback	1. Time consuming, complexity in calculation 2.Difficult to convince decision making 3.Uncertainty – doesn't consider	AHP develops the decision problem from planning of differentobjectives, criteria and alternatives and pair wise comparison of the criteria to obtain the best alternative
6	Goal Programming (GP) [14], [27], [30]	1.Handles large scales of variables, constraints and objectives. 2. Can produce infinite alternatives.	1.Needs to be combine with other MCDM techniques forcriteria weight. 2.Solutions are may be not efficient	It has more than one objective which conflicts with one other, and by composing the objectives or target must be achieved by minimizing the irrelevant data.
7	Grey Analysis (GA) [14], [20]	1 Complete and precise information has a unique solution.	1.Optimal solution Does not exist	Methods deal with all incomplete and imprecise data
8	ELECTRE (in French: ELimination Et Choix Traduisant la REalité)[16], [27], [29]	1. Deals outranking relation by using pairwise comparisons 2. Handles complete or incomplete data	1.concordance index required 2.sometimes unable to identify the preferred alternative.	Used to select the best alternative with maximum advantage and least conflict in the various criteria
9	VIKOR (from Serbian: VlseKriterijumskaOptimizacija I KompromisnoResenje) [17], [25]	1. Like ELECTRE method 2.The best alternative is finding by minimizing regret group and maximizing utility group theory. 3.compromise solution with an advantage rate	1.The performance rating is quantified as crisp values 2. Doesn't consider imprecise or ambiguous data	Used to solve alternative preference in highly complex environment
10	TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) [5], [17], [25]	1.Numberof iterations remains the sameregardless of the number ofattributes. 2. Ability to measure the relative criteria of alternative in Mathematical form 3.Use of Euclidean Distance property	1. Rank reversal 2.Doesn't consider uncertainty in weightings. 3.Its use of Euclidean Distance doesnot consider the correlation ofattributes. 4.Difficult to understand because itconsists of many algorithms.	By Euclidean Distance finds the shortest distance from the positive ideal(S ⁺) solution and the farthest distance from the negative ideal solution(S ⁻) of alternatives

Apart from the MCDM methods which are listed, many other MCDM methods are available which have been listed as above with its purpose, merits and demerits. The suitability of each method and Description in which it can be applied has been described in Table1.

III. An Overview of Selected MCDM Methods:

For this paper, several methods have been reviewed, and eventually, WSM, WPM, and AHP methods are the ones which are used mostly in practice today are identified to described in this section. The main reasons for selecting these MCDM methods are 1) WSM is a proportional linear transformation of the raw data. It means that the relative order of magnitude of the standardized scores remains same. 2)WPM is modification

of the WSM model and more efficient than other methods in problem solving of MCDM [Atmojo, Cahyani, Lie, 2014]. 3) AHP is a compensatory method which assumes complete aggregation among criteria and develops a linear additive model. It's ability to produce accurate ratio scale measurements and combine them across many multiple criteria decision methods.[Kasperczyk, N., Knickel, K. (2004)].

Before we present more details on selected MCDM methods, we would like to present a brief description about decision matrix for readers understanding.

Decision Matrix:

An MCDM Problems are often simply expressed in matrix format. A decision matrix A is associate (m × n) matrix within which element a_{ij} indicates the performance of alternative A_i when it is evaluated in terms of decision criterion C_j, (for i = 1,2,3..., m, and j = 1,2,3..., n). It is additionally assumed that the decision maker has determined the weights of relative performance of the criteria (denoted as W_j, for j = 1,2,3, ..., n) as represented in [22].

$$\begin{array}{c}
 \text{Criteria } C_j \\
 \begin{array}{cccc}
 C_1 & C_2 & C_3 \dots\dots & C_n \\
 W_1 & W_2 & W_3 \dots\dots & W_n
 \end{array} \\
 \text{Alt. } A_i \\
 \left[\begin{array}{cccc}
 a_{11} & a_{12} & a_{13} & a_{1n} \\
 a_{21} & a_{22} & a_{23} & a_{2n} \\
 a_{31} & a_{32} & a_{33} & a_{3n} \\
 \vdots & \vdots & \vdots & \vdots \\
 a_{m1} & a_{m2} & a_{m3} & a_{mn}
 \end{array} \right]
 \end{array} \tag{1}$$

III.1. The Weighted Sum Model (WSM):

The weighted sum model (WSM) is probably the most frequently used approach, especially in singledimensional problems. If there are 'm' alternatives and 'n' criteria then, the best alternative is the one that satisfies (inthe maximization case) the following expression [Fishburn, 1967]:In the background of this technique, the additive utility hypothesis is applied, which means that the general price of each different is like the products' total sum.In issues with identical units' ranges across criteria, WSM is applicable; however, once the units' ranges vary, as an example once qualitative and quantitative attributes are employed, the matter becomes more difficult to handle, because the aforesaid hypothesis is desecrated, and hence, normalisation schemes should be applied as presented in [27].

$$A_i^{WSM} = \sum_{j=1}^n a_{ij} w_j, \text{ Where } i = 1,2,3, \dots, m, \text{ and } j = 1,2,3, \dots, n \tag{2}$$

$$\text{Where } \sum_{j=1}^n W_j = 1 \tag{3}$$

Where A_i^{WSM} is weighted sum value of A_i alternative, 'n' is the total number of criteria, a_{ij} is the actual value of the i^{th} alternative in terms of the j^{th} criterion, and w_j is the weight of importance of the j^{th} criterion.

Then best alternative is

$$A^* = \max\{A_i^{WSM}\}, i = 1,2,3, \dots, m \tag{4}$$

In single-dimensional cases, within which all the unit's area unit thesame (e.g., Radian, speed, rupees, seconds), the WSM will be used without any difficulty.Difficulty with this method comes whenit is applied to multi-dimensional decision-making problems. Then, in combining different dimensions, andconsequently totally different units, the additive utility assumption is desecrated, and the result is like "adding apples to bananas".

Case Study of IPL 2018:

As a case study, in any sport, formation of a good team is vital to its success. Player selection for a team in most sports is a subjective issue, commonly based on the coaches/team manager/Franchises' notions of what is needed to form the best team. Selection of players in a team is always a difficult task for Franchisee, because many qualitative and quantitative attributes in the player selection process need to be considered. Although team management uses a variety of assessments such as testingpractice matches, coaches'

judgments to select and evaluate players, quantitative approaches could greatly benefit the team management and be significantly useful to form the best team. For selection of best player Innings, runs, average, balls faced, strike rate, not out, for the bowler, bowl Economy i.e. average runs conceded by the player per over bowled, bowl Strike Rate, wickets etc. are the some of the decision criteria for evaluation and selection of best player as shown in Table.2.

In the case study, we are interested in determining the best alternative (i.e., best player). The primary advantage of WSM, WPM and AHP are their simplicity and transparency in forming an interactive selection Mathematical decision model to help decision makers in evaluating best alternative. we present a case study of cricket player’s performance analysis of batting in Indian premier league (IPL)2018 to illustrate the effectiveness and applicability of identified MCDM models. For this analysis we consider only Six randomly batsmen’s (alternatives) and six criteria for batting performance evaluation and ranking of players. Same procedure could be extending to applied for ‘m’ alternatives and ‘n’ criteria. Result may be will vary due to impact of additional criteria and its weight on additional alternatives. As IPL Twenty20 league is of limited overs tournament hence, strike rate of batsmen is the most important criteria followed by runs, average, innings, Balls faced and Not out innings. All these criteria are benefits attributes that is the higher value of any attributes signifies better performance of batsmen. These Batsmen criteria for performance evaluation are described in the Table.2.

Table 2 Batsmen Attribute and its description

Criteria	Description
Innings (C_1)	No. of innings played by a batsman in a series of tournament
Runs (C_2)	No. of runs scored by a batsman in a series of tournament
Average (C_3)	No. of Runs scored / (Total no. of Innings-Not out)
Balls faced (C_4)	No. of balls faced by a batsman in a series of tournament
Strike rate (C_5)	Total no. of Runs scored / Total no. of Balls faced by a batsman.
Not out (C_6)	No. of innings in which a batsman is not out

Batting Statistics of the selected players for our case study is taken from secondary source data as IPL 2018 (<https://www.iplt20.com/stats/2018/most-runs>). Note that as per the statistical data from IPL 2018, Kane Williamson (Sunrisers Hyderabad) was highest run scorer and Rishabh Pant (Delhi Daredevils) was second position.

Table-3. Actual data sheet

Sr.no	Name of Batsman	Innings C_1	Runs C_2	Average C_3	Balls faced C_4	Strike rate C_5	Not out C_6
P1	Rohit Sharma	14	286	23.83	215	133.02	2
P2	MS Dhoni	15	455	75.83	302	150.66	9
P3	Virat Kohli	14	530	48.18	381	139.10	3
P4	Kane Williamson	17	735	52.50	516	142.44	3
P5	Rishabh Pant	14	684	52.61	394	173.60	1
P6	AB de Villiers	11	480	53.33	275	174.54	2

Weight of selected criteria are obtained by subjective judgement of decision manager (team manager) as per his requirement of the team and his expertise in field of T 20 cricket with some beliefs as shown in Table 11.

Table.4 (Weight of Six criteria)

Criteria	Innings C_1	Runs C_2	Average C_3	Balls faced C_4	Strike rate C_5	Not out C_6
weight	0.157	0.180	0.166	0.098	0.311	0.089

Based on equation (3) we have formulated our $\sum_{j=1}^6 C_j = 1$

Decision matrix will be formed by equation (1) and the corresponding a_{ij} values are to be formed as follows. (from Table.3,4)

Table.5 Decision matrix:

Sr.no	Name of Batsman	Innings C_1	Runs C_2	Average C_3	Balls faced C_4	Strike rate C_5	Not out C_6
	Weight	0.157	0.180	0.166	0.098	0.311	0.089
P1	Rohit Sharma	14	286	23.83	215	133.02	2
P2	MS Dhoni	15	455	75.83	302	150.66	9
P3	Virat Kohli	14	530	48.18	381	139.10	3
P4	Kane Williamson	17	735	52.50	516	142.44	3
P5	Rishabh Pant	14	684	52.61	394	173.60	1
P6	AB de Villiers	11	480	53.33	275	174.54	2
	Sum	85.00	3,170.00	306.28	2,083.00	913.36	20.00

6x6

By equation (2) the scores of the six alternatives are:

$$A_i^{wsm} = \sum_{j=1}^n a_{ij} w_j$$

From Table.5, The overall score of each individual alternative is obtained through the weighted sum model (WSM) as follows.

as shown in Table.6 and Fig.3.

$$P_1^{wsm} = (14 \times 0.157) + (286 \times 0.180) + (23.83 \times 0.166) + (215 \times 0.098) + (133.02 \times 0.311) + (2 \times 0.089) = 120.19$$

$$P_2^{wsm} = (15 \times 0.157) + (455 \times 0.180) + (75.83 \times 0.166) + (302 \times 0.098) + (150.66 \times 0.311) + (9 \times 0.089) = 174.03$$

$$P_3^{wsm} = (14 \times 0.157) + (530 \times 0.180) + (48.18 \times 0.166) + (381 \times 0.098) + (139.10 \times 0.311) + (3 \times 0.089) = 186.40$$

$$P_4^{wsm} = (17 \times 0.157) + (735 \times 0.180) + (52.50 \times 0.166) + (516 \times 0.098) + (142.44 \times 0.311) + (3 \times 0.089) = 238.77$$

$$P_5^{wsm} = (14 \times 0.157) + (684 \times 0.180) + (52.61 \times 0.166) + (394 \times 0.098) + (173.60 \times 0.311) + (1 \times 0.089) = 226.71$$

$$P_6^{wsm} = (11 \times 0.157) + (480 \times 0.180) + (53.33 \times 0.166) + (275 \times 0.098) + (174.54 \times 0.311) + (2 \times 0.089) = 178.34$$

Hence our result is $P_4^{wsm} > P_5^{wsm} > P_3^{wsm} > P_6^{wsm} > P_2^{wsm} > P_1^{wsm}$ (where ">" stands for "better than").

Table 6 WSM score and ranking

Sr.no	Name of Batsman	WSM	Rank	Performance Index %
P1	Rohit Sharma	120.19	6	50.34
P2	MS Dhoni	174.03	5	72.89
P3	Virat Kohli	186.40	3	78.07
P4	Kane Williamson	238.77	1	100.00
P5	Rishabh Pant	226.71	2	94.95
P6	AB de Villiers	178.34	4	74.69



Fig.3 WSM Ranking

As shown in figure 3 our numerical analysis predicts the ranking of a player based on WSM relative weights and performance Index As we can see Kane Williamson stands number 1 player because his number of runs is high and predominating on other criteria. Although, we have set high weight for strike rate criteria because WSM is a directly proportional to linear transformation of the raw data.

III.2. The Weighted Product Model (WPM):

The Weighted Product Model (WPM) is very similar to the Weighted Sum Model (WSM). The main difference being multiplication instead of addition in the model. Each alternative is compared with the others by multiplying of ratios, one for each criterion. Each ratio is raised to the power like the relative weight of the respective criterion. In general, to compare the alternatives A_k and A_l , the following product (Bridgman[1922] and Miller and Starr [1969]) as presented in [20] must be calculated:

$$R\left(\frac{A_k}{A_l}\right) = \prod_{j=1}^n \left(\frac{a_{kj}}{a_{lj}}\right)^{w_j} \quad \text{where } i = 1,2,3,\dots, k, l, \dots, m, \text{ and } j = 1,2,3,\dots, n \quad (5)$$

where: 'n' is the number of criteria, a_{kj} and a_{lj} are the actual value of the k^{th} and l^{th} alternative for the j^{th} criterion, and w_j is the weight of the j^{th} criterion. If $R\left(\frac{A_k}{A_l}\right) > 1$, then alternative A_k is preferred than alternative A_l (in the maximization case), if $R\left(\frac{A_k}{A_l}\right) < 1$, then alternative A_l is preferred than alternative A_k . The best alternative is the one that is better than or at least equal to all the other alternatives. The WPM is also known as dimensionless analysis because it eliminates any units of measure. Thus, the WPM can be applied in single-and multi-dimensional decision-making problems. An advantage of this model is that instead of the actual values it can use relative values.

Consider the problem presented in the previous example as a case study in Table 5.

$$R\left(\frac{P_1}{P_2}\right) = \prod_{j=1}^n \left(\frac{a_{1j}}{a_{2j}}\right)^{w_j}$$

$$R\left(\frac{P_1}{P_2}\right) = \left(\frac{14}{15}\right)^{0.157} \times \left(\frac{286}{455}\right)^{0.180} \times \left(\frac{23.83}{75.83}\right)^{0.166} \times \left(\frac{215}{302}\right)^{0.098} \times \left(\frac{133.02}{150.66}\right)^{0.311} \times \left(\frac{2}{9}\right)^{0.089} = 0.6113$$

It means that $P_2 > P_1$, similarly, $R\left(\frac{P_1}{P_3}\right) = 0.7160 < 1$ implies that $P_3 > P_1$, ...and so on. Similarly, we required to compare each P_1 to P_4, P_5, P_6 . Then similarly we will compare P_2 to P_3, P_4, P_5, P_6 and so on. It means that we must take total 15 comparison ratios. $\left\{ \text{For } n, \text{ total compare ratios are } \frac{n(n-1)}{2} \right\}$. If number of alternatives is large then it's lengthier and more difficult to solve. This is one of the drawbacks of WPM.

Hence it was modified by normalising each a_{ij} By dividing $\sum_{i=1}^m a_{ij}$ for the j^{th} criterion and the normalized score R_{ij} was determined using Eq. (6) as follows as shown in Table 7.

$$R_{ij} = \frac{a_{ij}}{\sum_{i=1}^m a_{ij}} \tag{6}$$

The overall score of players (P_i) was determined from R_{ij} value using Eq. (7) as presented in [29]

$$P_i = \prod_{j=1}^n (R_{ij})^{w_j} \tag{7}$$

The overall score defines the final scores of each (player) alternative. The final scores of each player are shown in Table 8 and Fig.4.

Table.7. Normalized Matrix: R_{ij}

Sr. no	Name of Batsman	Innings C_1	Runs C_2	Average C_3	Balls faced C_4	Strike rate C_5	Not out C_6
P1	Rohit Sharma	0.165	0.090	0.078	0.103	0.146	0.100
P2	MS Dhoni	0.176	0.144	0.248	0.145	0.165	0.450
P3	Virat Kohli	0.165	0.167	0.157	0.183	0.152	0.150
P4	Kane Williamson	0.200	0.232	0.171	0.248	0.156	0.150
P5	Rishabh Pant	0.165	0.216	0.172	0.189	0.190	0.050
P6	AB de Villiers	0.129	0.151	0.174	0.132	0.191	0.100
	checksum	1.000	1.000	1.000	1.000	1.000	1.000

Table 8. Scores obtained using WPM and ranking

Sr.no	Name of Batsman	WPM	Rank	Performance Index %
P1	Rohit Sharma	0.1146	6	61.11
P2	MS Dhoni	0.1874	1	100.00
P3	Virat Kohli	0.1600	4	85.34
P4	Kane Williamson	0.1841	2	98.24
P5	Rishabh Pant	0.1657	3	88.38
P6	AB de Villiers	0.1543	5	82.30



Fig.4. WPM Ranking

As shown in figure 4 our numerical analysis predicts the ranking of a player based on WPM relative weights and performance Index. As we can see MS Dhoni is 1 position and Kane Williamson stands number 2 player although his number of runs is high and predominating on other criteria. Because, we have set high weight for strike rate criteria and WPM is directly proportional to linear transformation of the raw data with multiplication hypothesis instead of additive hypothesis like WSM.

III.3. Analytical Hierarchy Process (AHP):

One of the most applicable MCDM techniques The Analytic Hierarchy Process (AHP) for formulation and analysing complex problems in conflict situation, based on Mathematics and Science. It was developed by Thomas L.Saaty (1980) as presented in [22] and has been extensively studied and modified since then. Later, T.L. Saaty (1996) done study on the further development of the Analytic Network Process (ANP) Technique. The AHP has attracted the interest of many scientists and researchers mainly due to the nice

mathematical properties with proofs of the method and the important fact that the required input data for decision matrix are easy to obtain. The AHP is a multi-criteria decision theory support tool which can be used to solve complex decision problems. It uses a multi-level hierarchical structure of objectives, alternatives, criteria, and its sub criteria, the pertinent data are derived by using a set of pairwise comparisons. These pairwise comparisons are used to obtain the weights of the decision criteria, and the relative performance measures of the alternatives in terms of each individual decision criterion. If the comparisons matrix is inconsistent, then AHP provides a mechanism for improving consistency.

AHP is a study of relative measurement through pairwise comparisons of criteria and depends on the judgments of decision makers to derive priority scales (T.L.Saaty, 2008) as presented in [24].With the help of Saaty’s 9-point linear scale and the detailed steps of AHP is described as follows.

a) Pairwise Comparisons of Criteria:

An approach based on pairwise comparisons which was first proposed by T.L. Saaty (1980) has long attracted the interest of many researchers. Pairwise comparisons are used to determine the relative importance of each alternative in terms of each individual criterion. In this approach the decision-maker (DM) must express his judgement about the value of one single pairwise comparison at a time. Usually, the DM must select his answer among Saaty’s 9-point linear scale.(see also Table 9). Each choice is a linguistic form. Some examples of such linguistic phrases are: "player P_1 batting performance is very poor than P_2 ", or " P_1 batting performance is of the same poor as P_2 ", or " P_1 is a little more good than P_2 ", and so on. Since pairwise comparisons are the key factor of these decision-making processes, correctly quantifying them is the most crucial step in Multi-Criteria Decision-Making methods which use qualitative data. The DM are unable to make selections from an infinite set. For example, DM cannot justify between two very close values of importance, say 5.00 and 4.98. Psychological experiments have also shown that individual’s person cannot simultaneously compare more than seven objects (plus or minus two) (Miller, 1956). This is the main reasoning used by T.L. Saaty to establish 9 as the upper limit of his scale, 1 as the lower limit and a unit difference between successive scale values. The values of the pairwise comparisons in the AHP are determined according to the scale introduced by T.L. Saaty (1980). According to these 9 scales, the available

values for the pairwise comparisons are members of the set: $\{9, 8, 7, 6, 5, 4, 3, 2, 1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \frac{1}{6}, \frac{1}{7}, \frac{1}{8}, \frac{1}{9}\}$ i.e.

$C_{ij} = \frac{1}{C_{ji}}$ (see also Table 10).

Table.9.(Saaty,1980)9-point preference scale

Scale	Compare factor of i and j
1	Equally Important
3	Weakly Important
5	Strongly Important
7	Very Strongly Important
9	Extremely Important
2,4,6,8	Intermediate value between adjacent

For our case study example as shown in Table 10 Decision Makers define with brainstorming a pair-wise comparison matrix for criteria using Saaty’s 9 scale, so, criteria weights are computed with subjective judgement of DM is as shown in Table11.

Table 10. Pairwise comparison criteria matrix

Criteria	Innings C_1	Runs C_2	Average C_3	Balls faced C_4	Strike rate C_5	Not out C_6
Innings C_1	1.000	1.000	1.000	2.000	0.333	2.000
Runs C_2	1.000	1.000	1.000	3.003	0.500	2.000
Average C_3	1.000	1.000	1.000	2.000	0.500	2.000
Balls faced C_4	0.500	0.333	0.500	1.000	0.333	2.000
Strike rate C_5	3.000	2.000	2.000	3.000	1.000	2.000
Not out C_6	0.500	0.500	0.500	0.500	0.500	1.000
sum	7.000	5.833	6.000	11.503	3.167	11.000

Normalise each a_{ij} by dividing $\sum_{i=1}^m a_{ij}$ for $j = 1, 2, \dots, n$ by Eq. (6) as shown in Table 11.

$$R_{ij} = \frac{a_{ij}}{\sum_{i=1}^m a_{ij}}$$

Table.11 R_{ij}

	Innings C_1	Runs C_2	Average C_3	Balls faced C_4	Strike rate C_5	Not out C_6	Weights	Total	CM
Innings C_1	0.143	0.171	0.167	0.174	0.105	0.182	0.157	0.942	6.239
Runs C_2	0.143	0.171	0.167	0.261	0.158	0.182	0.180	1.082	6.263
Average C_3	0.143	0.171	0.167	0.174	0.158	0.182	0.166	0.995	6.221
Balls faced C_4	0.071	0.057	0.083	0.087	0.105	0.182	0.098	0.586	6.146
Strike rate C_5	0.429	0.343	0.333	0.261	0.316	0.182	0.311	1.863	6.261
Not out C_6	0.071	0.086	0.083	0.043	0.158	0.091	0.089	0.533	6.131
sum	1.000	1.000	1.000	1.000	1.000	1.000	1.000	λ Max	6.210

One of the most practical issues in the AHP methodology is that it allows for slightly non-consistent pairwise comparisons. If all the comparisons are perfectly consistent, then the following relation should always be true for any combination of comparisons taken from the judgment matrix: $a_{ij} = a_{ik} a_{kj}$. However, perfect consistency rarely occurs in practice. In the AHP the pairwise comparisons in a judgment matrix are adequately consistent if the corresponding consistency ratio (CR) is less than 10% (Saaty, 1980) as presented in [22]. First the consistency index (CI) needs to be estimated. This is done by adding the columns in the judgment matrix and multiply the resulting vector by the vector of priorities (i.e., the approximate eigenvector) obtained earlier as consistency measure (CM) in table 11. This yields an approximation of the maximum eigenvalue, denoted by λ max. Then, the CI value is calculated by using the formula as presented in [14]

$$CI = \frac{(\lambda_{max} - n)}{(n-1)} \tag{8}$$

Next the consistency ratio CR is obtained by dividing the CI value by the Random Consistency Index (RCI) as given in Table 12.

The random Consistency index (RCI) values for various sizes of matrices are suggested by many researchers. The present study used the values as suggested by J.A. Alonso, M.T. Lamata, (2006) as presented in [19]. In this case the order of the matrix is 6 and thus the value of RCI is considered as 1.24.

Table.12. Random Consistency index(RCI)

n	1	2	3	4	5	6	7	8	9
RCI	0	0	0.58	0.90	1.12	1.24	1.32	1.41	1.45

where λ_{max} is the maximum eigenvalue and n is the size of the crisp comparison Matrix. In the present case from Table.11. $\lambda_{Max} = 6.210$, $n = 6$, hence by Eq. (8)

$$CI = \frac{(6.210 - 6)}{(6 - 1)} = 0.040$$

For consistency ratio (CR), T.L. Saaty (1980) as presented in [22].

$$CR = \frac{CI}{RCI} = \frac{0.040}{1.24} = 0.0339 < 0.10 \tag{9}$$

It turns out that a reciprocal matrix A with positive entries is consistent if and only if eigenvalue $\lambda_{Max} = n$. In practice, $CR \leq 0.10$ is considered acceptable, otherwise Any higher value at any level indicate that the judgements warrant to re-examination of pair wise comparison matrix. Our Eq. (9) indicates that the pair-

wise comparison matrix is not suffering from inconsistency and can be used for relative weight calculation as shown in Table 11.

$$C_i = \frac{\sum_{j=1}^n C_{ij}}{n} \quad i = 1,2,3, \dots m$$

$$C_1 = \frac{\sum_{j=1}^6 C_{1j}}{6} = \frac{0.942}{6} = 0.157, \text{ similarly, } C_2 = 0.180, C_3 = 0.166, C_4 = 0.098, C_5 = 0.311, C_6 = 0.089$$

Ranking of best alternative is calculated as shown in Table.13 and Fig.5.

Table.13. AHP Best Alternatives

Player	SCORES:	Rank	Performance Index %
Rohit Sharma	0.11920	6	59.61
MS Dhoni	0.19996	1	100.00
Virat Kohli	0.16055	4	80.29
Kane Williamson	0.18755	2	93.79
Rishabh Pant	0.17516	3	87.60
AB de Villiers	0.15759	5	78.81
checksum	1.000		



Fig.5. AHP ranking

As shown in figure 5 our numerical analysis predicts the ranking of a player based on AHP Model relative weights and performance Index. As we can see MS Dhoni is again 1 position and Kane Williamson stands number 2 player although his number of runs is high and predominating on other criteria. Because, we have set high weight for strike rate criteria in pair wise comparison matrix as shown in table 10 .Also, the Main reason is that AHP helps to derive ratio scale priorities from approximate pairwise verbal comparisons relative to any attribute, quantitative or qualitative and established Consistency Index i.e. $CI = \frac{(\lambda_{max} - n)}{(n-1)}$ as a measure of the consistency or reliability of information as presented in T.L.Satty1977.

IV. Over all Comparison of Ranking by WSM, WPM and AHP Models:

Table 14. Comparison of Ranking by SM,WPM,AHP

Sr.no	Name of Batsman	WSM	WPM	AHP
		Rank	Rank	Rank
P1	Rohit Sharma	6	6	6
P2	MS Dhoni	5	1	1
P3	Virat Kohli	3	4	4
P4	Kane Williamson	1	2	2
P5	Rishabh Pant	2	3	3
P6	AB de Villiers	4	5	5



Fig.6 WSP,WPM,AHP Comparison ranking

We have seen that from WPM and AHP model ,the performance of MS Dhoni is very decent and consistent and after that the performance of Kane Williamson is good and followed by Rishabh Pant, Virat Kohli, AB de Villiers and Rohit Sharma. The overall Performance Index of MS Dhoni is good and due to his consistent contribution in all selected criteria we can say that it could be one of the major reasons for Chennai Super

Kings to win the IPL cup in 2018. Sunrisers Hyderabad still not win the title IPL 2018, but we have found that by WSM, WPM and AHP methods as shown in Table 14, the Performance Index of Kane Williamson player is not consistent in all selected criteria as compared to MS Dhoni. The main reason is that special emphasis is required on the integration of human judgments into decisions and on the measurement of the consistency of DM's judgments.

V. CONCLUSIONS:

This paper provided a review of MCDM Models used for Mathematical analysis of Alternatives with case study of IPL 2018. It was found that there are various issues in evaluations of alternatives and criteria. In our case study WSM, WPM and AHP having different ranks of the players, especially at first Rank (Table 6, 8, 13). As per the statistical data from IPL 2018, Kane Williamson (Sunrisers Hyderabad) was highest run scorer, Rishabh Pant (Delhi Daredevils) was on second position, but they are on second and third position where MS Dhoni (Chennai Super Kings) is on First position in both WPM and AHP models as shown in Table 14. Although the mathematical methods for processing the pertinent data are rather simple, the real challenge is in quantifying these data with incomplete and Imprecise Information in highly conflict situation. The main issue is that to identify proper optimal model for the selection of best alternative in MCDM techniques. Operations research provides a systematic framework for dealing with such problems. There are several techniques under MCDM but each having its own characteristic and scope of performance. However, WSP, WPM and AHP are unable to handle the inherent uncertainties and imprecisions in mathematical evaluation of criteria and alternatives. Hence, other MCDM methods should be developed to address this limitation. We hope this paper will help researchers and decision makers in solving MCDM problems effectively.

VI. ACKNOWLEDGMENT:

This work is partially supported by Vidyalankar Institute of Technology, Wadala, Mumbai, India, under the scheme of "Higher Studies Sponsorship Policy", Year 2018/19.

VII. REFERENCES:

- [1] Abbas Mardani, Ahmad Jusoh, Khalil MD Nor, Zainab Khalifah, Norhayati Zakwan, Alireza Valipour, (2015). "Multiple criteria decision-making techniques and their applications – a review of the literature from 2000 to 2014", *Economic Research-Ekonomska Istraživanja*, 28:1, 516-571, DOI: 10.1080/1331677X.2015.1075139
- [2] Ali Emrouznejad, Marianna Marra, (2017). "The state-of-the-art development of AHP (1979–2017): a literature review with a social network analysis", *International Journal of Production Research*, 55:22, 6653-6675, DOI: 10.1080/00207543.2017.1334976.
- [3] Arbel, A., Vargas, L. (2007). "Interval judgments and Euclidean centers", *Mathematical and Computer Modelling* 46(7), 976–984.
- [4] Bana e Costa, C.A., Vansnick, J.C, (2008). "A critical analysis of the eigenvalue method used to derive priorities in AHP", *European Journal of Operational Research* 187(3), 1422–1428.
- [5] Behzadian M, Khanmohammadi Otagh Sara S, Yazdani M, Ignatius J, (2012). "A state-of-the-art survey of TOPSIS applications". *Expert Systems with Applications*; 39:13051–69. doi: 10.1016/j.eswa.2012.05.056.
- [6] Belton, V., Gear, T. (1983) "On a short-coming of Saaty's method of analytic hierarchies". *Omega* 11(3), 228–230.
- [7] Benítez, J., X. Delgado-Galván, J. Izquierdo, R. Pérez-García, (2012). "Improving Consistency in AHP Decision-making Processes." *Applied Mathematics and Computation* 219: 2432–2441.
- [8] Bernard ROY, (1990). "Decision-aid and decision-making" *European Journal of Operational Research* 45 August, pp.324-331.
- [9] Boz'oki, S, (2008). "Solution of the least squares method problem of pairwise comparison matrices", *Central European Journal of Operations Research* 16(4), 345–358.
- [10] Brunelli, Matteo. (2015). "Introduction to the Analytic Hierarchy Process". Springer Briefs in Operations Research. P. 83. 978-3-319-12502-2 (electronic). <http://dx.doi.org/10.1007/978-3-319-12502-2>
- [11] C.L. Hwang, K. Yoon, (1981) "Multiple Attribute Decision Making Methods and Applications", Springer, Berlin Heidelberg.
- [12] Cavallo, B., D'Apuzzo, L., (2009). "A general unified framework for pairwise comparison matrices in multicriteria methods" *International Journal of Intelligent Systems* 24(4), 377–398.
- [13] Edmondas Kazimieras Zavadskas, Zenonas Turskis, Simona Kildienė, (2014). "State of art surveys of overviews on MCDM/MADM methods", *Technological and Economic Development of Economy*, 20:1, 165-179.

- [14] Evangelos Triantaphyllou, Stuart H. Mann, (1995). "Using the Analytic Hierarchy Process for Decision Making in engineering applications: some challenges", *International Journal of Industrial Engineering: Applications and Practice*, Vol. 2, No. 1, pp. 35-44.
- [15] Fedrizzi, M., S. Giove. (2007). "Incomplete Pairwise Comparison and Consistency Optimization", *European Journal of Operational Research* 183: 303-313.
- [16] Govindan K, Jepsen MB, (2016). "ELECTRE: A comprehensive literature review on methodologies and applications", *European Journal of Operational Research*; 250:1-29. doi: 10.1016/j.ejor.2015.07.019.
- [17] Hwang, C.L., Yoon, K. (1981) "Multiple Attribute Decision Making: Methods and Applications". Springer-Verlag, New York. <http://dx.doi.org/10.1007/978-3-642-48318-9>
- [18] I. L. Tomashevskii, (2018). "Optimization methods to estimate alternatives in AHP: The classification with respect to the dependence of irrelevant alternatives", *Journal of the Operational Research Society*, 69:7, 1114-1124, DOI: 10.1080/01605682.2017.1390533
- [19] J.A. Alonso, M.T. Lamata, (2006). "Consistency in the analytic hierarchy process: a new approach". *Int. J. Uncertain. Fuzziness Knowledge Based Syst.* 14(4), 445-459.
- [20] Martin Aruldoss, T. Miranda Lakshmi, V. Prasanna Venkatesan, (2013). "A Survey on Multi Criteria Decision Making Methods and Its Applications." *American Journal of Information Systems* 1, no. 1: 31-43. doi: 10.12691/ajis-1-1-5.
- [21] Pawel Tadeusz Kazibudzki, (2012). "Note on some revelations in prioritization, theory of choice and decision-making support methodology", *African Journal of Business Management* Vol. 6(48), pp. 11762-11770.
- [22] Saaty, T. L., (1980). "The Analytic Hierarchy Process", McGraw-Hill, New York.
- [23] Saaty, T. L., 1983, "Priority Setting in Complex Problems," *IEEE Transactions on Engineering Management*, Vol.30, No.3, pp.140-155.
- [24] Saaty, T.L. (2008). "Decision making with the analytic hierarchy process". *International Journal of Services Sciences*, 1(1), pp. 83-98.
- [25] Serafim Opricovic, Gwo-Hshiung Tzeng, (2004). "Compromise solution by MCDM methods: A comparative analysis of VIKOR and TOPSIS", *European Journal of Operational Research* 156, pp. 445-455.
- [26] Tapas Kumar Biswasa, Pritha Sahab, (2019). "Selection of commercially available scooters by new MCDM method", *International Journal of Data and Network Science* 3, pp. 137-144.
- [27] Triantaphyllou E. (2000). "Multi-criteria Decision-Making Methods: A Comparative study", *Applied Optimization*, vol.44. Springer, Boston, MA. https://doi.org/10.1007/978-1-4757-3157-6_2
- [28] Tseng, G.H. dan Huang, J.J. (2011). "Multiple Attribute Decision Making, Methods and Applications", CRC Press. Boca Raton.
- [29] Velasquez, Hester, (2013). "An Analysis of Multi-Criteria Decision-Making Methods", *International Journal of Operations Research*, Vol. 10, No. 2, pp. 56-66.
- [30] Wann-Ming Wey, Kuei-Yang Wu, (2007). "Using ANP priorities with goal programming in resource allocation in transportation", *Mathematical and Computer Modelling* 46, pp. 985-1000.
- [31] William Ho, Xiaowei Xu, Prasanta K. Dey, (2010). "Multi-criteria decision-making approaches for supplier evaluation and selection: A literature review", *European Journal of Operational Research*, Vol. 202, Issue 1, pp. 16-24.
- [32] Zavadskas, E.K.; Podvezko, V, (2016). "Integrated determination of objective criteria weights in MCDM", *International Journal of Information Technology & Decision Making*, 15, pp. 267-283. [CrossRef]

Recent trends in Advance Coding Techniques

¹Prof. Sandip J. Dawda & ²Dr. Saurabh Mehta

¹Assistant Professor, ²Professor

¹Electronics and Communication Department

¹Government Engineering College, Sector-28, Gandhinagar, Gujarat, India

²Electronics and Telecommunication Department

²Vidyalankar Campus, Vidyalankar College Marg, Wadala (East), Mumbai, Maharashtra, India.

ABSTRACT: *In the recent trends in wireless communication, beyond 4G, some of the important objectives need to be discussed like increased capacity, improved data rate, decreased latency, and better quality of service. Particularly if we consider the scenario based on channel coding techniques, In wireless communication, the high bit error rate requires the use of new advanced channel coding techniques which are at the core of the paper. In this detailed, we are enlist various advance coding techniques. Here we presents propose classification of advance coding techniques. Each code merely gives effective throughput compare to traditional coding technique. All such coding techniques can really applicable for upcoming wireless communication which is really helpful to fulfill the requirement of current data usage.*

Keywords: *Luby Transform Code, Tornado Code, Polar Code, Raptor code, Fountain code*

I. Introduction

In many ways wireless based networks require to enhance its capacity according to the requirement of current data usage. Of the current options, Cellular pre-3G technologies don't provide any useful rates of data. Using 3G technologies becoming widely available and emerging 4G technologies, wireless broadband access will be upgraded. Recent technology constituent like Mobile WiMax, long-term evolution (LTE), LTE advance and 5G are coming in the market as a segment of the advancement of current wireless based technologies. Each technology has been designed with different standards. For new development in standards, parameters and strategies will be assigned like modulation techniques, Accessing techniques, and Coding techniques. Specimen of all such new standard, one of the major parts is channel coding techniques. With the help of advance research in coding, we can able to achieve better performance compare to traditional coding techniques. One can achieve good performance compared to previous version techniques. Here various advance codes like Turbo codes, LDPC codes, Rate compatible LDPC and turbo codes, Fountain codes(Tornado code, Raptor Code R10 and RQ, Luby Transform codes, Online codes etc.) and Polar code. Many of this code have been continued and will be adopted into a number of standards (David Declerq, M. F. 2014). In this paper, we will compare each code with their application, complexity and comparison based on their performance. Remaining paper is organized as follows: In Section II will discuss brief history and its evolution. Section III will give you brief idea about each code and their application along with their decoding algorithm. Section IV will compare the performance for advance forward error correcting codes for next generation wireless communication. Section V will compare decoding cost/complexity for different codes.

II. Background

Error correction coding is one of the most important elements of a modern communication system. Redundant information (parity) is added to a message in error correction to make the message more resilient to noise-induced transmission errors. A paper published by Shannon (Shannon, C. E., July and October 1948) was used as the field of error correcting code. Shannon forecast that reliable communication will be possible by adding redundancy across a memory less channel, so long as the communication rate is not greater than the channel's ability. In order to achieve that, Shannon had never proposed a coding system. Efforts have been made since 1948 in the ongoing search for codes capable of achieving these theoretical limits. The codes which have been developed over the years can mainly be classified in three types: block codes, convolutional codes and concatenated codes. Hamming(Hamming, R. W., 1950), Bose-Chaudhuri-Hocquenghem (BCH)(Hocquenghem, A., September 1959, Ray-Chaudhari, R. B. a. D., March

1960, Ray-Chaudhuri, R. B. a. D., September 1960), Reed-Solomon (RS)(Solomon, I. S. R. a. G., June 1960) and Low-Density Parity Checks(Gallager, R. G., January 1962, Gallager, R. G., 1963, Neal, D. J. C. M. a. R. M., August 1996) have been the main blocking codes developed during the past 60 years. Some of the major developments in decoding block code algorithms include Berlekamp-Massey hard decision to decode an algorithm for RS codes (Berlekamp, E. R., March 1968], building a block code-trellis scheme that allows Viterbi's soft decision to decode (Viterbi, A. J., April 1967) linear block codes (Wolf, J. K., January 1978), as well as developing BP (belief propagation) sparse graph code decoding (LDPC) (MacKay, 2005) (Gallager, R. G., 1963, Pearl, J., August 1982, MacKay, D. J. C., 2005). In 1955 Elias first introduced CC (Convolutional codes)(Elias, P., March 1955). In order to decode these codes, two important decoding algorithms were developed, namely Viterbi trellis decoding(Viterbi, A. J., April 1967) and the algorithm MAP (Maximum A-Posteriori)(L. R. Bahl, J. C. F. J. a. J. R., March 1974). Forney developed the first concatenated codes and Berrou et al developed many later Turbo codes (very good concatenated codes)(Forney, G. D., 1966, C. Berrou, A. G. a. P. T., May 1993). There are further two development in coding theory by combining modulation and coding as TCM (Trellis Coded Modulation)(Ungerboeck, G., January 1982) and other one using multiple antennas as STC (Space Time Codes) (Alamouti, S. M., October 1998). The Shannon target is now almost achieved with Turbo and LDPC codes. If you use large block (code) sizes, these codes can work arbitrarily close to Shannon when used in the AWGN channel and binary erasure channel(MacKay, D. J. C., 2005). In MacKay, Turbo and LDPC codes are discussed in terms of their representation of the factor graph and are therefore also part of a family of codes known as sparse graph codes, which also include repeat accumulate (RA) and fountain codes. Due to the low density figures representing it, these codes are called sparse. BP (originally created by Pearl and Gallager) is the principal decoding algorithm used for those codes(Gallager, R. G., January 1962, Gallager, R. G., 1963, Pearl, J., August 1982).

III. DIFFERENT CODES

Codes can be characterized and explain in briefly with respect to their invention and application. Figure 1 shows various advance coding techniques according to their discovery year Figure 2 presents different classification based on their coding rate whether rated or rateless code.

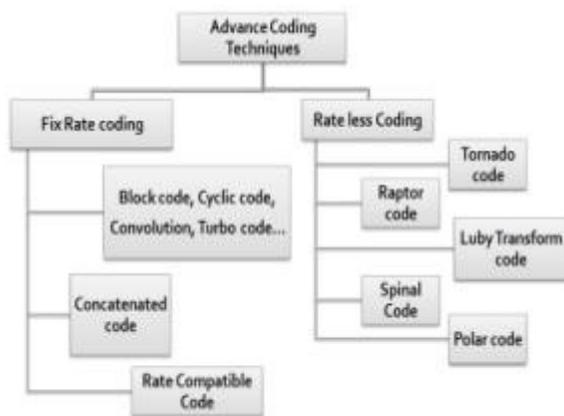


Fig. 1 Discovery of various recent codes

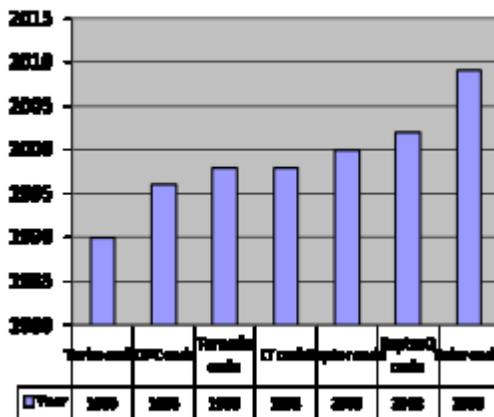


Fig. 2 Classification of Advance Coding Techniques

3.1 Turbo Codes

Turbo code was developed in the period 1990-91 (but first released in 1993) as a high-performance forward error correction (FEC) code. These were the first practical codes that approach the channel capacity very closely. Reliable communication for a theoretical maximum code rate is still possible at a specific noise level. Turbocodes apply to 3G/4 G mobile communication (e.g.

Table 1: Various Codes and their application in wireless communication

Coding Techniques	Code capability	Inventor	Development Year	Standard/Application	Preferable decoding Techniques	
Turbo Codes	Forward Error correction	Claude Berrou	1990-91	Wireless WAN(3G/4G) HSPA, EV-DO, IEEE 802.16 (WiMAX), and LTE. Satellite Communication DVB-RCS and DVB-RCS2	BCJR Algorithms (Bahl, Cocke, Jelinek and Raviv) Viterbi algorithm	
LDPC Code	Forward Error correction	Claude Berrou	1963 Re-discover 1996	Digital Video Broadcasting DVB-S2 / DVB-T2 / DVB-C2, DMB-T/H Wifi/Ethernet IEEE 802.11n-2009 IEEE 802.3an	Soft-in-soft-out (SISO) techniques such as SOVA, BCJR, MAP,	
Fountain codes	Tornado Code	Erasure correcting codes/ Forward Error Correction	Michael Luby	1998	---	---
	Luby Transform Code	erasure correcting codes/ Forward Error Correction	Michael Luby	1998	---	---
	Raptor Code	erasure correcting codes/ Forward Error Correction	Amin Shokrollahi	2000	DVB-H standards IETF RFC 5330	belief propagation decoding algorithm
	RaptorQ Code	erasure correcting codes/ Forward Error Correction	Amin Shokrollahi	2000	IETF RFC 6330	belief propagation decoding algorithm
Polar Code	Forward Error correction	Erdal Arkan	2009	---	Successive Cancellation list decoding algorithm	

UMTS and LTE) and to satellite communications as well as other applications where designers aim to secure trustworthy data transmission over the presence of corrupting sound via bandwidth or latency-constrained communication links. Today, turbo codes compete with performance - like LDPC codes.

3.2 LDPC Code

A low density Parity Check (LDPC) code, the method of transmitting a message via a reliable communication channel, is a linear error correction code (MacKay, D. J., 2003, Moon, T. K., 2005). A sparse bipartite graph is used to build the LDPC (Shokrollahi, A., 2003). LDPC codes can be considered as capacity - approach codes, which means that the noise threshold is set to the theoretical maximum (the Shannon limits) for a symmetrical memory-less channel very near (or even arbitrarily close to the BEC) and practical constructions. The noise limit defines an upper limit of the channel noise to the extent that the likelihood of lost information can be as small as desired. LDPC codes can be decoded in a time linear to their block length using iterative belief propagation technology. In applications where reliable information transfer over bandwidth or channel-constrained links in the presence of noise is required, LDPC codes are becoming increasingly widely used. The LDPC codes have been lagging behind those of other codes, in particular turbo codes. LDPC codes were also known as Gallager codes in honor of Mr G. Gallager, the LDPC concept developer at the Massachusetts Institute of Technology in 1960 (Dana Mackenzie, 2005).

3.3 Tornado Code

Tornado codes are a class of error correction erasure codes. Tornado codes need a constant C, which requires more redundant blocks than Reed-Solomon, but is far faster to generate and can fix erasures faster. In small length and around 10,000 times faster in bigger length, software-based implementation is approximately 100 times faster than the Reed-Solomon erasure codes (Mitzenmacher, M., 2004). Several other similar erasure codes, especially online codes, LT codes, and raptor codes have arisen since the introduction of Tornado codes. A layered approach is used by tornado codes. All layers, with the exception of the last one, use a LDPC error correction code, which is fast but may fail. The final layer uses a slower but optimum Reed-Solomon correction code for failure recovery. Tornado codes tell how many levels, how many blocks of recovery are at each level, and how much blocks are generated by the distribution for non-final levels.

3.4 Luby Transform Code

The first class of practical fountain code that is almost optimally erasure correction codes are Luby's transformation codes (LT codes). In 1998 Michael Luby invented them and in 2002 they were published (M.Luby, 2002). Like some other fountain codes, LT codes depend for encoding and decoding speed on sparse bipartite graphs to trade overhead reception. The distinctive feature of LT codes is that the message is encoded and decoded using a particularly simple algorithm based on the exclusive or operation).

LT coding is rate-less, as the encoding algorithm may generate an infinite number of message packages in principle (i.e., the proportion of packets that need to be received to decode the message may be arbitrarily small). These erasure codes are used to reliably transmit digital data on the erasure channel. Raptor codes (see, for instance IETF RFC 5053 or IETF RFC 6330) are next generation beyond LT codes and have linear time coding and decoding. Raptor codes are coded in two encoding stages, with an LT encoding in the second stage.

3.5 Raptor Code

Raptor codes are the first known class of fountain codes with linear time encoding and decoding, (rapid tornado; see Tornado codes)(Shokrollahi, A., 2006). In 2000/2001 they were invented by Amin Shokrollahi and published in an extended abstract for the first time in 2004. Raptor codes represent a significant improvement over LT codes that are the first practical class of fountain codes, theoretical and practical.

The raptor codes encode the message in a number of symbols k in a potentially endless sequence of coding symbols, as is usually the case for fountain codes, so that knowledge of any k or more coding symbols allows for a certain non-zero probability of recovering the message. With the number of symbols received above k , the likelihood that the message can be recovered increases to close to 1, once the number of symbols received is just slightly greater than k . For instance, RaptorQ codes for the latest generation are less than 1 percent, where k symbols have been received, and when $k+2$ symbols have been received, there is a possibility that $k+2$ symbols will fail to decode. A symbol may have any size of hundreds or thousands of bytes from a single Byte.

Raptor codes may be either systematic or not. In the system, in the set of coding symbols the symbols of the original message are included. An example of a systematic raptor code is the code defined in the third generation wireless mobile data transmission and multi-casting partnerships, and used by DVB-H IP data transmission standards for handheld devices (see external links). Raptor codes are also defined in IETF RFC 5053 in these standards. RaptorQ defined in IETF RFC 6330 is one of the most advanced versions.

3.6 Polar Code

Polar coding, which Erdal Arıkan recently invented (Arıkan, E., July 2009), is an encoding/decoding scheme that has the capability of symmetrical binary memory-less channels to achieve. Polar code can achieve symmetrical ability of discrete binary memory-less channels. Polar Codes are encoded and decoded to low complexities ($O(N \log N)$) and N is the code block length. In addition, the probability of the block error of Polar Codes decay to zero, like $O(2^{-\text{alternatives to } ON})$ for any rate below capacity. Since polar coding is a new technique, many questions about polar codes remain unanswered.

IV. Decoding Complexity

Here we summarized the various coding techniques in the table which includes code capability, invention, development year, their various application and preferable decoding techniques as mention in table 2. we summarize coding techniques with their decoding cost or complexity in table II which depicts that what kind of latency is going to provide by different kind of codes. Decoding cost means their complexity to decode code once it is received at receiver side. Complexity means how much time it takes to process. How much adder, Memories is going to utilize to decode the code? So more number of such modules is going to utilize it takes more time so it gives more latency even system is going to be more & more complex.

TABLE 2 Decoding Cost

Coding Techniques		Decoding Complexity
Turbo Codes		-----
LDPC Code		$O(k^2)$
Fountain codes	Tornado Code	$O(n \log 1/\epsilon)$
	Luby Transform Code	$O(k \log k)$
	Raptor Code	$O(k \log 1/\epsilon)$
	RaptorQ Code	$O(k^{1.5})$
Polar Code		$O(2 - \sqrt{N})$.

V. Conclusions

In this paper, brief survey of various advance coding techniques has been shown based on their discovery and application. Here we find various emerging codes which are very much helpful for recent technologies and future technologies. All such code has much more performance capability compared to the traditional coding techniques. We also come to know that how much latency is going to provide by different kind of codes.

References

- [1] David Declerq, M. F. 2014. Channel Coding : Theory, Algorithms and Application (First ed.). Academic Press.
- [2] Shannon, C. E., July and October 1948. A mathematical theory of communication. s.l., Bell System Technical, p.379-423 and 623-656.
- [3] Hamming, R. W., 1950. Error detecting and error correcting codes. Bell System Technical, April, Volume 26,2, p.147-160.
- [4] Hocquenghem, A., September 1959. Codes correcteurs d'erreurs , vol. 2. s.l., Chiffres, p. 147-156.
- [5] Ray-Chaudhari, R. B. a. D., March 1960. On a class of error correcting binary group. s.l., Information and Control, vol. 3, p. 68-79.
- [6] Ray-Chaudhari, R. B. a. D., September 1960. Further results on error correcting binary group. s.l., Information and Control, vol. 3, p. 279-290.
- [7] Solomon, I. S. R. a. G., June 1960. Polynomial codes over certain finite fields. The Society of Industrial and Applied Mathematics, vol. 8, no. 2, p. 300-304.
- [8] Gallager, R. G., January 1962. Low-density parity-check codes. s.l., IRE Transactions on Informa-tion Theory, vol. 8, no. 1, p. 21-28.
- [9] Gallager, R. G., 1963. Low-density parity-check codes, Cambridge, Massachusetts: Ph.D. thesis, M.I.T..
- [10] Neal, D. J. C. M. a. R. M., August 1996. Near Shannon limit performance of low density parity check codes. s.l., Electronics Letters, vol. 32, no. 18, p. 1645-1646.
- [11] Berlekamp, E. R., March 1968. Nonbinary BCH decoding. s.l., IEEE Transactions on Information Theory, vol. 14, no. 2, p. 242-242.
- [12] Viterbi, A. J., April 1967. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. s.l., IEEE Transactions on Information Theory, vol. 13, no. 2, p. 260-269.
- [13] Wolf, J. K., January 1978. Efficient maximum likelihood decoding of linear block codes using a trellis. s.l., IEEE Transactions on Information Theory, vol. 24, p. 76-80
- [14] Pearl, J., August 1982. Reverend bayes on inference engines: A distributed hierarchial approach. Pittsburg, USA, s.n., p. 133-146.
- [15] MacKay, D. J. C., 2005. Information Theory, Inference, and Learning Algorithms, 1st ed, Cambridge: UK: Cambridge University Press.
- [16] Elias, P., March 1955. Coding of noisy channels. s.l., IRE Convention Record, vol. 4, p. 37-47.
- [17] L. R. Bahl, J. C. F. J. a. J. R., March 1974. Optimal decoding of linear codes for minimizing symbol error rate. s.l., IEEE Transactions on Information Theory, vol. 20, no. 2, p. 284-287. error rate. s.l., IEEE Transactions on Info
- [18] Forney, G. D., 1966. Concatenated Codes, 1st ed, Cambridge, Massachuset: USA: MIT Press.
- [19] C. Berrou, A. G. a. P. T., May 1993. Near Shannon limit errorcorrecting coding and decoding: Turbo codes.
- [20] Ungerboeck, G., January 1982. Channel coding with multilevel/phase signals. s.l., IEEE Transactions on Information Theory, vol. 28, no. 1, p. 55-67.
- [21] Alamouti, S. M., October 1998. simple transmit diversity technique for wireless communications. s.l., IEEE Journal on Selected Areas in Communications, vol. 16, no. 8, p. 1451-1458.
- [22] MacKay, D. J., 2003. Information theory, Inference and Learning Algorithms, CUP, s.l.: ISBN 0-521-64298-1.
- [23] Moon, T. K., 2005. Error Correction Coding, Mathematical Methods and Algorithms. Wiley, s.l.: ISBN 0-471-64800-0.
- [24] Shokrollahi, A., 2003. LDPC Codes: An Introduction,
- [25] Dana Mackenzie , , Communication speed nears terminal velocity, NewScientist, 9 July 2005
- [26] Mitzenmacher, M., 2004. Digital Fountains: A Survey and Look Forward. s.l., Proc. 2004 IEEE Information Theory Workshop (ITW)..
- [27] M.Luby, 2002. LT Codes. s.l., The 43rd Annual IEEE Symposium on Foundations of Computer Science.
- [28] Shokrollahi, A., 2006. Raptor Codes. s.l., IEEE Transactions on Information Theory, vol. 52, pp. 2551-2567.
- [29] Arıkan, E. (July 2009). "Channel Polarization: A Method for Constructing Capacity-Achieving Codes for Symmetric Binary-Input Memoryless Channels". IEEE Transactions on Information Theory 55 (7): 3051-73

ANALYSIS OF FREQUENT AMINO ACID ASSOCIATION PATTERNS IN PEPTIDE SEQUENCES OF DENGUE VIRUS TYPE 1

¹Dr. Amita Jain & ²Ms. Janhavi Vadke

¹Assistant Professor, ²Assistant Professor

¹Information Technology

¹Vidyalankar School of Information Technology, Wadala, Mumbai, India

ABSTRACT: Association rule mining is the important research area in data mining and Frequent Itemset Mining technique is one of the prominent techniques in pattern mining. In this study frequent itemset mining is used to find the frequent amino acids patterns in Dengue Virus Type1 data set. The Apriori approach is used to identify the frequent patterns in Dengue Virus Type1 amino acids sequences. This study reveals the association between the amino acids. These associations have been used to obtain association rules, secondary structures and the physiochemical properties. At present there are no proper vaccines for diseases like Dengue. The association between dominating amino acids will be useful for the drug designers to develop the antibiotics for this virulent disease caused by dengue viruses. This proposed system uses protein dengue virus type1 datasets from NCBI (National Centre for Biotechnology Information). Such models will be of great use in developing signatures that will provide better insight into the structures, functions and interactions of proteins.

Keywords: Data mining, Association rule, support, confidence, fuzzy set etc.

I. Introduction

Analysis of biological data plays an important role in medical and bioinformatics industry. But uncertainty in this biological information is the most unavoidable challenge of this era. The existing algorithms for association rule mining are inadequate to address the issues of uncertainty in the molecular data. The variation in the length of the sequences leads to the variation in the degree of relationships among amino acids. Ignorance of the parameters leads to uncertainty due to the dependencies of the objects and their patterns on the parameters.

Despite extensive research, the underlying mechanisms causing severe dengue is still not well understood due to the lack of appropriate models of infection and disease. However, even though it is pretty clear that both viral and host factors play important roles in the course of infection, a fundamental knowledge gap still remains to be filled regarding host virus.

The plenty of molecular sequences of dengue virus are available in various biological databases. The associations of amino acids present in peptide sequences of dengue virus can be explored and analyzed to generate the information which will be of crucial importance in understanding the structures, functions, binding sites, biochemical properties, protein folding and interactions of these peptide sequences.

Agrawal and Srikant [1] proposed an algorithm for extracting association rules from large databases in 1994. A number of techniques and methodologies have been proposed by various researchers [2] [3] [4] [5] [6] [7] [8] for association rule mining. A number of theories like probability, fuzzy set theory [9], intuitionistic fuzzy sets [10, 11], vague sets [12] and rough set theory [13] are reported in the literature to deal with uncertainty. All these theories, however, are associated with an inherent limitation, which is the inadequacy of the parameterization tool associated with these theories. Various research workers have employed fuzzy set and soft set [14] approaches for finding patterns in molecular sequences of MTBC [15][16][17].

II Research Methodology:

The dataset of peptide sequences of dengue virus is collected from online databases like NCBI. In this data the redundant, incomplete and complete sequences are identified. The length of all the sequences is computed to determine the range of variation of length in the sequences as shown in Table 2.1, Table 2.2 and Table 2.3. These parameters like length, redundancy, missing characters, and incomplete sequences are identified as sources of uncertainties.

Table 2.1: Sequences, their amino acids and length of corresponding sequences

Sequence number (Transactions)	Amino Acids(Item Set)	Length of Sequence (Parameter)
S_1	$a_1 \ a_2 \ \dots \ a_{19} \ a_{20}$	l_1
S_2	$a_{2,1} \ a_{2,2} \ \dots \ a_{2,19} \ a_{2,20}$	l_2
\vdots	\vdots	\vdots
S_n	$a_{n,1} \ a_{n,2} \ \dots \ a_{n,19} \ a_{n,20}$	l_n

Table 2.2: Sequence number and membership of amino acids in each sequence

Sequence number (Transactions)	Membership of Amino Acids(Item Set)
S_1	$\mu(a_{1,1})\mu(a_{1,2})\dots\mu(a_{1,19})\mu(a_{1,20})$
S_2	$\mu(a_{2,1})\mu(a_{2,2})\dots\mu(a_{2,19})\mu(a_{2,20})$
\vdots	\vdots
S_n	$\mu(a_{n,1})\mu(a_{n,2})\dots\mu(a_{n,19})\mu(a_{n,20})$
Summation(Σ)	$\Sigma\mu(a_{1,1})\Sigma\mu(a_{2,2})\dots\Sigma\mu(a_{19,19})\Sigma\mu(a_{20,20})$

Table 2.3: Sequences, their amino acids and length of corresponding sequences in different Ranges (Low, Medium and high)

Sequence number	Amino Acids(Item Set)	Length of Sequence (Parameter)	Ranges		
			R _{LOW}	R _{MID}	R _{HIGH}
S_1	$a_1 \ a_2 \ \dots \ a_{19} \ a_{20}$	l_1	0	1	0
S_2	$a_{2,1} \ a_{2,2} \ \dots \ a_{2,19} \ a_{2,20}$	l_2	1	0	0
\vdots	\vdots	\vdots	0	0	1
S_n	$a_{n,1} \ a_{n,2} \ \dots \ a_{n,19} \ a_{n,20}$	l_n			

III RESULTS AND DISCUSSION:

The maximal amino acid association patterns and their support (> threshold=331) are computed by fuzzy set approach and presented in Table 3.1.

Table 3.1: Support of maximal association patterns of dengue virus type 1

Maximal association patterns	Support
EGALKTV	SUPPORT = 339.5075
EGLIKTV	SUPPORT = 337.01398
GALIKTV	SUPPORT = 332.94965

The maximal amino acid association patterns and their confidence (> 0.05) are computed by fuzzy set approach and presented in Table 3.2.

Table 3.2: Confidence of most Frequent Maximum Patterns

EGALKTV GALKTV->E CONFIDENCE 0.9600038 EALKTV->G CONFIDENCE 0.9987888 EGLKTV->A	EGLIKTV GLIKTV->E CONFIDENCE 0.9583555 ELIKTV->G CONFIDENCE 0.9907415 EGIKTV->L	GALIKTV ALIKTV->G CONFIDENCE 0.9988162 GLIKTV->A CONFIDENCE 0.9467979 GAIKTV->L CONFIDENCE
---	---	--

CONFIDENCE 0.9436287 EGAKTV->L CONFIDENCE 0.9961138 EGALTV->K CONFIDENCE 0.97449356 EGALKV->T CONFIDENCE 0.9966864 EGALKT->V CONFIDENCE 0.9918528	CONFIDENCE 0.99583346 EGLKTV->I CONFIDENCE 0.9366982 EGLITV->K CONFIDENCE 0.9920625 EGLIKV->T CONFIDENCE 0.99714667 EGLIKT->V CONFIDENCE 0.98953503	0.9972472 GALKTV->I CONFIDENCE 0.94146055 GALITV->K CONFIDENCE 0.9812261 GALIKV->T CONFIDENCE 0.9970958 GALIKT->V CONFIDENCE 0.97966015
---	---	---

Table 3.3 shows the number of redundant and non-redundant sequences and also number of frequent patterns obtained in different ranges.

Table 3.3: Analysis on the basis of different ranges of peptide sequences of dengue virus type 1:

	REDUNDANT PEPTIDE SEQUENCES	NON REDUNDANT SEQUENCES	RANGE 1 [NON REDUNDANT]	RANGE2 [NON REDUNDANT]	RANGE3 [NON REDUNDANT]
RANGE			0 - 1140	1141- 2280	2281- 3420
Total No Of Sequences	12581	6625	4220	24	2382
Threshold	629.05	331.25	211	1.2	119.1
Frquent Amino acids	9	9	9	9	10
Maximal F Pattern	6	7	5	9	10
Two Frequent	35	36	32	36	45
Three Frequent	66	81	50	84	120
Four Frequent	60	102	38	126	210
Five Frequent	28	64	11	126	252
Six Frequent	5	21	-	84	210
Seven Frequent	-	3	-	36	120
Eight Frequent			-	9	45
Nine Frequent			-	1	10
Ten Frequent			-	-	1

Figure 3.1 shows the different ranges (low, medium and high) for peptide sequences of dengue virus type 1. It is observed that number of sequences are higher in low range.

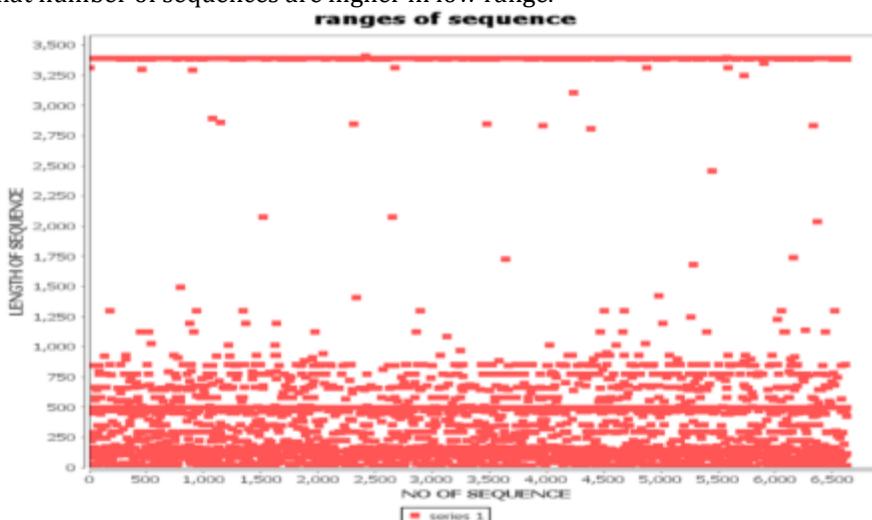


Fig 3.1:Dengue virus type 1 sequence in ranges

Table 3.4 depicts various physicochemical properties for non-redundant sequences in different ranges.

Table 3.4: Results generated from the peptide sequences of dengue virus type 1:

PARAMETERS	RESPONSIBLE AMINO ACID	COMPOSITION			
		NON REDUNDANT	RANGE1	RANGE2	RANGE3
hydrophobicity:	I V L F C M A G	46.53 %	47.03 %	46.23 %	45.65 %
aromaticity:	W Y F	7.98 %	8.02 %	7.91 %	7.92 %
protein stability:	R D C E K	23.46 %	22.95 %	23.81 %	24.38 %
c-beta branched :	V I T	21.63 %	22.28 %	21.12 %	20.48 %
polarity:	R N D C E Q H K S T G Y	58.04 %	58.46 %	57.86 %	57.30 %
salt bridged:	R N E K	21.30 %	20.69%	22.05 %	22.39 %
helix:	M A L E K R Q H	41.98 %	40.97 %	43.03 %	43.76 %
beta sheet:	V I T C W F Y	31.68 %	32.57 %	31.34 %	30.12 %
coil:	N D P S G	26.33 %	26.46 %	25.64 %	26.12 %
absorbance:	Y W C	6.70 %	6.74 %	6.10 %	6.62 %
aliphatic:	A V L I	28.37 %	28.05%	28.58 %	28.94 %

It is observed from Table 3.5 that Length of sequences is directly proportional to some physicochemical properties like protein stability, salt bridged formation, aliphaticity, positive charged, negative charged, and helix formation whereas it is inversely proportional to hydrophobicity, c-beta charged, polarity and beta sheet structure formation. Also it can be predicted that in range 3 (length of sequence is greater than 2837), **Arginine [R]** also occurs frequently.

Table 3.5: Analysis of above discussion

	INCREASES	DECREASES
Range	↑	
Length of sequence	↑	
Hydrophobicity		↓
Protein stability:	↑	
C-beta branched :		↓
Polarity:		↓
Salt bridged:	↑	
Aliphatic:	↑	
Positive charged:	↑	
Negative charged:	↑	
Structures:		
Helix:	↑	
Beta sheet:		↓

IV CONCLUSIONS

The fuzzy set approach is proposed and employed to mine amino acid association patterns in the peptide sequences of dengue virus type 1, discover association rules and predict secondary structures and physicochemical properties of peptide sequences. On the basis of results, it can be concluded that fuzzy set approach is quite effective in incorporating degree of relationship among amino acid association patterns in peptide sequences of dengue virus type 1. The generated association patterns will be able to give the insights of various relationships among amino acids, physicochemical properties and secondary structures of dengue virus type 1. These type of models can be developed to generate the information on molecular relationships and mechanisms involved in the disease which could be useful to bio medical scientists for development of methodology for diagnosis and treatment of diseases. The presence or absence of amino-acids can be used to effectively design drug targets. The use of biologics which are synthetic peptides used as drugs can be

very well benefitted by the knowledge of association rules framed in various proteins. Such type of tools can not only save time and money but also help to find out patterns which can be missed by normal analysis.

REFERENCES:

1. Agrawal, R., and Srikant, R.: 'Fast algorithms for mining association rules', Proceeding 20th International Conference Very Large Data Bases, 1994.
2. Jea, K.F., et al.: 'Discovering frequent itemsets by support approximation and itemset clustering', Data & Knowledge Engineering, , 2008, pp. 90-107.
3. Houtsma, M., and Swami, A. : 'Set-oriented mining of association rules', Research Report RJ 9567, IBM Almaden Research Center[C]. San Jose, 1993.
4. Agarwal, R.C., Agarwal, C.C. and Prasad, V. V. V.: 'A Tree Projection Algorithm for Generation of Frequent Item Sets', Journal of Parallel and Distributed Computing, 2001, 61, (3), pp. 350-371.
5. Panday, A., and Pardasani, K.R.: 'PPCI algorithm for mining temporal association rules in large database', Journal of Information & Knowledge Management 2009, 8, (4), pp. 345-352.
6. Panday, A., and Pardasani, K.R.: 'Rough Set Model for Discovering Multidimensional Association Rules', IJCSNS International Journal of Computer Science and Network Security, 2009, 9, (6), pp. 159-164.
7. Patel, R., Swami, D.K. and Pardasani, K.R.: 'Lattice Based Algorithm for Incremental Mining of Association rules', International Journal of Theoretical and Applied Computer Sciences, 2006, 1, (1), pp. 119-128.
8. Khare, N., Adlakha, N. and Pardasani, K.R.: 'Karnaugh Map Model for Mining Association Rules in Large Databases', IJCSNS International Journal of Computer and Network Security 2009, 1, (1), pp. 16-21.
9. Zadeh, L.A.: 'Fuzzy Sets', Information and Control, 1965, pp. 338-353.
10. Atanassov, K.: 'Intuitionistic fuzzy sets', Fuzzy Sets and Systems 1986, 20, pp. 87- 96.
11. Atanassov, K.: 'Operators over interval valued intuitionistic fuzzy sets', Fuzzy Sets and Systems 1994, 64, pp. 159-174.
12. Gau, W.L., and B., D.J. : 'Vague sets', IEEE Trans. System Man Cybernet, 1993, 23, (2), pp. 610-614.
13. Pawlak, Z.: 'Rough sets', Internat. J. Inform. Comput. Sci. , 1982, 11, pp. 341-356 .
14. Molodtsov, D.: ' Soft set theory-first results', Computers and Mathematics with Applications 1999, 37, pp. 19-31.
15. Jain, A., &Pardasani, K. R. : 'Mining fuzzy amino acid associations in peptide sequences of mycobacterium tuberculosis complex (MTBC).', Network Modeling Analysis in Health Informatics and Bioinformatics, 2015, 4, (1), pp. 1-14.
16. Jain, A., &Pardasani, K. R. : 'Soft fuzzy model for mining amino acid associations in peptide sequences of Mycobacterium tuberculosis complex.' CURRENT SCIENCE, 110(4), pp.603-618.
17. Jain, A., &Pardasani, K. R.: 'Fuzzy soft set model for mining amino acid associations in peptide sequences of Mycobacterium tuberculosis complex (MTBC).', Journal of Intelligent & Fuzzy Systems, 2016, 31, pp. 259-273.

Automatic Electoral Identity Generator with OTP Authentication

¹Nirali Shah, ²Shirish Nair, ³Bhavin Oza, ⁴Russel D'britto, ⁵Ms. Purnima Kubde

¹Student, ²Student, ³Student, ⁴Student. ⁵Assistant Professor

⁵Department of Information Technology,

⁵St. Francis Institute of Technology, Mumbai, India

ABSTRACT: Voting is a fundamental right of any citizen as per the constitution of India, possession of an Electoral Identity card (Voter ID) is mandatory to be eligible to cast a vote. In a conventional system, the user has to fill the form manually and upload documents which are then checked by a person at the government/municipal organization's side. This takes a lot of time at both, the applicant as well as the government/municipal organization's side. Thus, we introduce a system that will generate the Electoral ID faster by combining the Aadhar card (Unique id for every Indian Citizen) database with birth and death certificate database.

Keywords: Application development, Electoral Identity Card, Aadhar card, Birth Certificate, Death Certificate, EPIC.

I. Introduction

Generating an Electoral Identity Card or Voter Card is a manual process. Traditionally, it involves submitting original documents such as Birth Certificate and Aadhar card as proof at the nearest municipal office for verification. Former acts as a proof of nationality date of birth and the latter one as an address proof. Another way of applying for an Electoral Identity Card is through an online portal of the Election Commission of Government of India. This also includes uploading documents over the internet for verification and the Electoral Identity card will be issued if all the documents are found to be correct [1]Article on "Voter ID – How to apply, eligibility and documents required", Cleartax, Feb 21, 2019, 08:15 IST. Although these two mechanisms verify a person's identity, they do not solve other issues pertaining to issuing, delivery, multiple generation and revocation of Electoral Identity cards. Election Commission also does not maintain the data of deceased people and rely on data given by the Health Department. This causes an increase in chance of the names of deceased people to appear in the eligible voters list. Since this data is important from Election point of view, it has to be error free so as to get accurate and fair election results.

II. PROBLEM DEFINITION AND AREA

To develop an Automatic Electoral Identity Cards generation web application to help solve problems with the issuing, delivery, multiple generation, and revocation of Automatic Electoral Identity Cards without the need of going to the nearest municipal office.

This project will include areas like Database Management, Web Technologies, Application development and Computer Security.

III. objectives and expected outcomes

The main objective of this project is to solve all the problems faced by election commission in generation of Voter ID cards and reduce the delivery time to the minimum also to automate the process so that the applicant do not need to physically visit the nearest municipal office for voter ID.

Additionally, we aim to improve the system in following ways:

- Faster Generation of Electoral Identity Cards.
- Eliminate the overhead of time required for delivery.
- Easy revocation
- Mapping of Electoral Identity Card with Aadhar Card.

By deploying this system, we not only not only will be able to streamline the process but also help in efficient data storage and retrieval for the concerned government organization/municipality.

IV. APPLICATION AND RELEVANCE

Introduction of this system will prove to be a convenient option for the citizens but also will be beneficial at the municipality/government's side. The online mode of applying for the Electoral Identity Card will be

accessible to people from anywhere, 24x7. Similarly, at the municipality/government's side, the overhead of sitting and verifying the uploaded documents will be eliminated. This system is planned to be deployed in such a way that it can replace the existing system without having to changing the underlying hardware and software technologies. This would be beneficial as the downtime required while upgrading the old system to new one will be minimal. Hence, we aim to introduce the new system which is relevant and better than the existing system to ease the entire process. Moreover, since the Electoral Identity Cards generated in the new system will be linked with the Aadhar card, it would be easy to keep a track of how many Electoral Identity Cards a person has issued in his/her lifetime. This step will tackle the issue of duplicate Electoral Identity Card generation. The entire idea of generating Electoral Identity Cards automatically via online method promotes the idea of a Smarter and Digital India.

Abbreviations and Acronyms

4.1 Electoral Identity Card

A valid government identity card issued by the Election Commission of India, possession of which is mandatory to vote.

4.2 Aadhar Card

An important government identity card that has a 12-digit unique number for every citizen of India.

All major transactions and documents are meant to be linked with the 12-digit Aadhar number.

4.3 Birth Certificate

A proof of birth, issued by the hospital/municipality.

4.4 Death Certificate

A proof of death, issued by the hospital/municipality.

4.5 EPIC

A 10-digit unique number for each Electoral Identity Card.

V. RESEARCH METHODOLOGY

1.1 Research Approach

Since this system aims to solve the problems in the conventional system, the research work had to be started right from identifying the issues and drawbacks of the same. This included understanding how the current system works, identifying the alternative methods of Electoral Identity Card generation, finding out the total time involved right from submitting the application till the delivery of Electoral Identity Card at home via post. All this required meeting the officials of the Election Commission and other concerned government officials to know the exact operation of the current system.

Keeping in mind the immense population of this country, it was necessary to consider the amount of data about people that is involved here. A typical ward has no less than 10,000 people in a moderately sized city [2]Demographics on "Greater Mumbai", IndiKosh – All about India. Generating Electoral Identity Cards for every person meant that each person will have a unique number on his/her Electoral Identity Cards. If not, the number on the card should at least be unique within the ward. Research also included studying the contents on the Electoral Identity Card and the EPIC number. Electoral Photo Identity Card number is a digit number that is unique to every Electoral Identity Card. Methods of generating this number and identifying the 10-digit composition was equally necessary (see Figure 1). Along with this, possession of multiple Electoral Identity Cards has been reported several times in the news and is a crime according to the Constitution of India. Hence ways to uniquely identify each Card and map how many such Cards a person possess becomes necessary [3]Article on "Having two Voter ID cards punishable", The Times of India, Oct 31, 2013, 14:29 IST.

1.2 Research Outcomes

After much research to understand the working of the existing system and two visits to the municipal corporation of Mumbai, Brihanmumbai Municipal Corporation (BMC) it was noted that applying for an Electoral Identity Card can be done in two ways: Offline and Online. Offline method was largely preferred by the people since online method wasn't introduced until year 2015. The Offline method is largely time consuming and involves the applicant to physically visit the municipal or the election commission office for applying for an Electoral Identity Card. The applicant also has to carry his/her valid government IDs and a proof of nationality and birth. The Online method does not keep a track if people have applied for an Electoral Identity Card through the Offline method. In case of the Online method, the applicant has to upload the valid government IDs and they are stored by the Government of India in formats such as JPEG/PNG/PDF and rest of the data in form of text for other credentials. Saving data in picture formats consumes a lot of server space and takes time to load/process and generates an overhead. According to the BMC, Election

Commission relies on the data provided by the Health Department and there isn't much coordination between these two independent bodies. The Death records are not updated by the Health Department according to the Election commission and hence this results in discrepancies in the list of Eligible Voters list that is generated as the data provided by the Health Department lacks integrity. For generating an Electoral Identity Card, the most important documents and valid government IDs required are: Birth Certificate as a proof of birth and nationality and the Aadhar card as a proof for address (current place of residence). These two documents are possessed by the majority of the people and hence are used when applying for an Electoral Identity card in both, Offline and Online methods. The Health Department also maintains records of registered births and deaths digitally in a large a database. Election commission can use this data and generate Electoral Identity Cards automatically, which is our proposed system itself. The web portal of the Election commission also allows one to apply for an Electoral Identity Card only for a stipulated amount of time before the elections take place. This results in the system being unavailable.

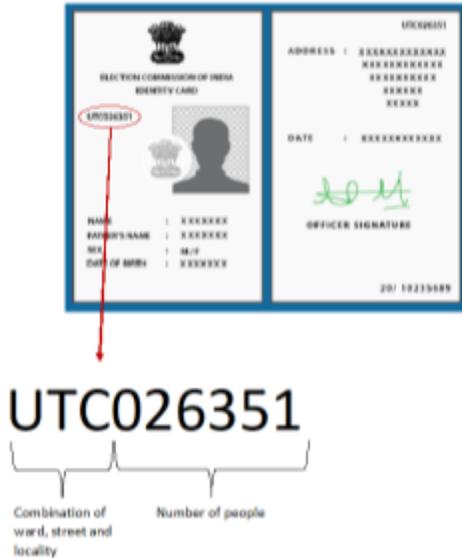


Figure 1: Showing EPIC number format

1.3 Technology and Methodology

Having mentioned the working of current system and its drawbacks, the proposed system will solve all of them with very less overhead of storing, processing and the entire time taken right from submitting the application to the delivery of the Electoral Identity Card at home.

The proposed system is so designed that it can be deployed in place of the current Online system without having to change much of the underlying technologies and hardware.

The proposed system will be available to the Municipal organizations or the Local Government/Election department of the concerned Municipality as a web application. This will allow it to be accessed from anywhere at any time. A central database will store all the data in tables and only photograph of the applicant in picture format. This will consume less space than storing documents in picture formats. Also storing images of pictures can be a threat to security in case the current system is hacked. Instead, all the data stored in the database will be encrypted using RSA/MD5 and hence will be secured [4]Article on "Why RSA encryption secure",RSA Encryption, Google Sites, "The simple mathematics of RSA", Nikos Drakos, Computer Based Learning Unit, University of Leeds, translated by Martin Ouwehand, Aug 19, 2001.

To applicants, this system will be accessible through a web browser. At the applicant's and government's side, lot of manual work will be saved as most of the processing will be done by the system itself, thus reducing the manual processing overhead.

The proposed system will use two existing government databases: Aadhar Card Database and Birth Certificate Database.

1.3.1 Working of the system

The system will use the existing Birth Certificate data and Aadhar Card details of the applicant from the respective databases. After a child is born, the birth certificate data will be stored in the database. Eventually as the child grows up, he/she will own an Aadhar card whose details will also be stored in the

database. After completing 18 years of age, the applicant will visit the government portal and enter his/her Aadhar card number. All the details required for the Electoral identity card will be mapped together with one common id. Once the applicant enters the Aadhar number, the details of the Aadhar card such as the name, date of birth, father’s name etc. will be verified with the birth certificate. If the records match, the Electoral Identity will be processed and its details will be stored in a different table. The system will check its database every day for applicants who are turning 18. For applicants who are turning 18 and those whose details from the birth certificate and Aadhar card match, the processed Electoral Identity card will be shown to the applicant for changes (if any). If the applicant has to update his/her latest Aadhar card details such as phone number and address so that the Electoral Identity can be delivered to the applicant at the mentioned address, he/she can do so and then upload his/her photograph through the front camera of the laptop/mobile phone. After the details are updated, an OTP will be sent to the mentioned phone number as an authentication of the applicant’s identity. If the OTP matches, the applicant will be given a preview of the Electoral Identity card. The Electoral Identity card will be delivered to the applicant after conformation of its correctness. After a person dies, the death certificate records will be updated in the database and the Electoral Identity card of that applicant will be revoked. Thus, it addresses problems pertaining to issuing, delivery, multiple generation and revocation of Electoral Identity cards.

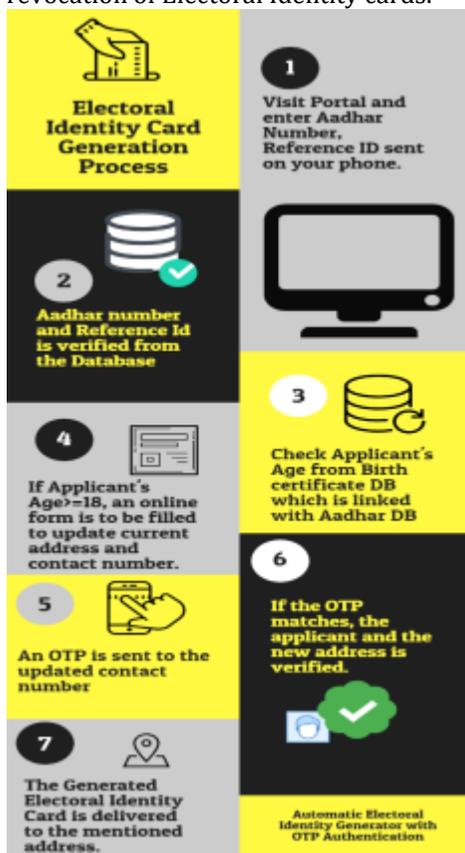


Figure 2: Showing general working of the proposed system.

1.4 Methods of Validation

To ensure that the Electoral Identity Card belongs to a legitimate application, there has to be a mean of authentication of the applicant which is absent in the current system. The current system does not guarantee that the person who is applying for an Electoral Identity Card will be the one to receive its delivery. This problem is solved in the proposed system by using OTP or One Time Password as a mean of authentication [5]Sudhakar K, Shrikant S, “Secured Mutual Authentication between two entities”, in 9th International IEEE conference 2015.OTP is will ensure that the applicant applied for an Electoral Identity Card is the same person to whom it belongs to. Moreover, the issue of delivery of the Electoral Identity Card is solved by using Digital Signature to validate the digital copy of newly Electoral Identity Card before downloading it.

VI. SURVEY

As mentioned in the Research Outcomes section, the municipal corporation of Mumbai, Brihanmumbai Municipal Corporation (BMC) was approached regarding the details of the working of current system. The Public Relations Officer (PRO) further guided as to who was to be approached and for what reason. The Assistant Health Officer at the Health Department of BMC at Parel provided enough information about the working of current system and the problems encountered in it. Further, employees at The Election Commission branch of BMC at Masjid Bundar, spoke about how the Electoral Identity Cards are generated and delivered

2.1 Literature Survey

Casting a vote is an important part in formation of a Democratic society. For establishment of any democratic society, more than half of the votes are necessary for choosing a ruling party. But since past few years, it has been observed that less than 60 percent of the total population actual cast their votes. There is a numerous reason as to why this happens. One reason why this happens is because a large portion of the population does not hold an Electoral Identity card. Automatic voter registration would not only help people leverage their full voter potential and support a more inclusive system, but it would also have a massive impact on young people and their access to the ballot. With voter registration among young people at a historic low—less than 50 percent—automatically registering students as soon as the turn 18 would not only encourage participation, but it would avoid many of the possible barriers young people face trying to register and eliminate the all too common excuses for not voting [6]Jazmin Kay, “How automatic Voter Registration Would Benefit Young People”, Generation Process, August 24, 2018. If the idea of voting online is added with generation of electoral identity, it will solve a great deal of problems in areas where the problem of booth capturing occurs [7]Vishal, Vibhu Chinmay, Rishabh Garg, “Online voting system linked with AADHAR,” in 3rd International IEEE conference 2016. Another idea of voting is to link the current voting system with Aadhar card. If Automatic Electoral identity cards and voting process is linked, it will lead to an efficient system for casting votes which will eliminate problems like bogus voting, booth capturing and other malpractices that are a threat to the idea of democracy [8]Himanshu Agarwal, G. N. Pandey, “Online voting system for India based on AADHAAR ID,” in Eleventh International IEEE conference 2013.

VII. RESULTS AND FINDINGS

After the Electoral Identity Card is generated, the resultant EPIC on the card will be unique for every applicant or at least unique in every ward. The EPIC number is an alphanumeric 10-digit number. First 6 digits of EPIC represent a combination of ward, street, locality and the next 4 digits represent number of applicants in that area. Number of applicants in an area can increase in an area beyond 1000. In that case, the last 4 digits can be shifted left in the first 6 digits.

VIII .CONCLUSION

The main objective of this proposed solution is to create an application for generating Electoral identity cards faster than conventional process and solve all the issues issuing, delivery, multiple generation, and revocation of Automatic Electoral Identity Cards. In conventional system, the user has to fill in the form manually and upload documents which are then checked by a person at the government/municipal organization's side. This takes a lot of time at both, the applicant as well as the government/municipal organization's side. By deploying this system, we not only will be able to streamline the process but also help in efficient data storage and retrieval for the concerned government organization/municipality. Thus, we aim to solve this problem by which the overhead on the sides will be reduced and the process of obtaining the Electoral ID will be a lot faster.

IX. ACKNOWLEDGEMENT

We thank Dr. Pranita Tipre, Asst. Health Officer at BMC's Health Department, Parel, Mumbai for her valuable guidance and insights.

References

1. Article on “Voter ID – How to apply, eligibility and documents required”, Cleartax, Feb 21, 2019, 08:15 IST.
2. Demographics on “Greater Mumbai”, IndiKosh – All about India.
3. Article on “Having two Voter ID cards punishable”, The Times of India, Oct 31, 2013, 14:29 IST.
4. Article on “Why RSA encryption secure”,RSA Encryption, Google Sites, “The simple mathematics of RSA”, Nikos Drakos, Computer Based Learning Unit, University of Leeds, translated by Martin Ouwehand, Aug 19, 2001.

5. Sudhakar K, Shrikant S, "Secured Mutual Authentication between two entities", in 9th International IEEE conference 2015 .
6. Jazmin Kay, "How automatic Voter Registration Would Benefit Young People", Generation Process, August 24, 2018.
7. Vishal, Vibhu Chinmay, Rishabh Garg, "Online voting system linked with AADHAR," in 3rd International IEEE conference 2016.
8. Himanshu Agarwal, G. N. Pandey, "Online voting system for India based on AADHAAR ID," in Eleventh International IEEE conference 2013.

ARchitect: Augmented Reality application for architect's portfolio

¹Sahil Raut, ²Shraddha Pawar, ³Remon Pereira, ⁴Monesha Murdeshwar, ⁵Sonali Vaidya

¹Student, ²Student, ³Student, ⁴Student, ⁵Assistant Professor

Department of Information Technology,
St. Francis Institute of Technology, Mumbai, India

ABSTRACT: Promotion of Architectural projects nowadays involves the use of non-portable 3D models or rendering them into images with improved aesthetics. With the recent emergence of better cameras and more accurate sensors in soon-to-be mainstream devices, Augmented Reality is transitioning from image or QR code based activations to marker-less Augmented Reality experiences. This paper presents an innovative way of marketing the architect's portfolio using a mobile Augmented Reality application named ARchitect. The Architects plans for their project are visualized in a portfolio that is used as target for displaying the specific 3D models. The proposed application includes storing of the models in online datasets and its retrieval on scanning target object to facilitate reduction of application size. The application helps improving the user experience by making the process of browsing the portfolio interactive, user-friendly and portable.

Keywords: Augmented Reality(AR), Marker based technology, Marker-less technology, 3D Modeling, Architect'sPortfolio

I. Introduction

For many years, Architecture agencies have focused primarily on the traditional method of building models manually using plaster of Paris, cardboard etc. This is a very tedious and cumbersome task. The drawback of this modus operandi is the immobility of the bulky 3D Models. To overcome this, architects market their projects using portfolios. A portfolio is a collection of creator's works intended to demonstrate a person's ability to a potential employer or customer. They can be in any form, printed such as brochures, newspaper, and digital such as websites, social media. These portfolios contain 2D images of the previously done work of the architect. But this resulted in losing the 3D aspect. To get the best of both worlds we are using Marker based/Marker-less augmented reality to develop an application that showcases a 3D model of architect project on user's own device. Augmented reality has the ability to enhance a user's experience of perceptions in real time which is demonstrated by mobile indoor AR like the ones related to interactive graphic AR applications and mobile outdoor platforms like location-based AR (Rattanarungrot, S., & White, M. 2014). With the help of AR 3D models, architects can create a lasting first impression and a more prolific networking experience. The project aims to develop a mobile application that showcases a novel way of advertising an architect's portfolio using augmented reality.

1.1 EXISTING SYSTEM

While researching on this application the following issues were observed in the existing system that needed to be tackled. Firstly, the concern of portability of 3D models and 3D simulation. In the prevailing system, architecture agencies make use of Computer-aided design (CAD) models in order to present their work as a simulation and also manually building models. But for this method to succeed either the customer has to go over to the company or the publicity team has to go over to the customer's house. In this methodology, it becomes tedious and time consuming for both parties. Another issue is the requirement of particular hardware. There are many features that can be implemented in mobile Virtual reality applications but they require a set of particular hardware that enables the device to map environments. As these devices are not economical, every user may not own one. Due to this drawback, some users may not be able to avail all the features of the respective mobile application. This also creates a challenge for the developers as they want most or all of the features to be available to their users. An additional issue is faced when applications use marker-based AR technology. The drawback of this is that it requires the user to place an image that acts as a reference to dynamically load the model. Hence the user has to continuously focus at the target to interact with the 3D model.

II. Literature Review

Extensive research was undertaken to ensure that ARchitect overcomes the downside of the existing system. Published research articles were studied to find solutions and also identify their limitations. The first paper presents a mobile augmented reality application that uses service-orientation for viewing cultural media objects in personalized interactive museum environments. The features that were included in this mobile application are media content acquisition from open service providers through a web service framework, photogrammetry service that allows users to obtain virtual 3D models of preferred cultural objects through image-based reconstruction techniques and the ability to integrate these acquired 3D reconstruction contents with existing associated contents in a personalized augmented reality museum environment so that they can be stored and saved for later viewing (Rattananarungrot, S., & White, M. 2014). This paper also introduces us to the concept of marker-less Augmented Reality and here the application was developed on iOS and native development platforms. For ARchitect, the application will be built on the Android platform and QR code will be provided for the same so that every android user will be able to download the application from google play store. The second study proposes a methodology of applying the AR technology to a construction project and a prototype system to test the applicability of the AR technology in rebar construction work (Hyeon-seung Kim, Chang-hak Kim, Hyoun-seok Moon, So-yeong Moon, Young-hwan Kim, Leen-seok Kang. 2013). In this paper Unifeye SDK (Software Development Kit) 3.5 was used to implement augmented reality. Instead of Unifeye SDK, modeling in Architect will be done using Blender as it is a free and open source 3D creation suite whereas many CAD tools are paid products and Unity 3D will be used to build an Android application which combines the real-time objects with virtual objects. Both software reduces the complexity of the modelling process and its deployment. A new method of building CARE (Contextual Augmented Reality Environments) environments is proposed in the paper. This method permits automatic generation of such environments through a semantic transformation of high-level domain ontologies. This enables AR environments to be created and maintained by domain-experts without experience in programming and graphic design (Krzysztof Walczak, Dariusz Rumiński, Jakub Flotyński. 2014). Many AR platforms were suggested in this paper but we found Vuforia as a better option for new developers who are working on storing 3D models in databases and desire user-friendly environment. For our last inspiration, we studied a research paper that it solves the issue of how to make human lips synchronization in Autodesk Maya Programming language of Maya – MEL (Maya Embedded Language) was used to associate with mouth positions to provide visemes for computer animation of speech (Antonio Moura; Ingrida Mazonaviciute; Joao Nunes; Justinas Grigaravicius. 2012). Autodesk-Maya is a relatively complex software to develop 3D models for ARchitect application. Hence to build 3D objects, Blender, which is user-friendly will be used to save time and cost of licensing.

III. Methodology

Taking into consideration the limitations and complexity of the systems researched in the above-stated paragraph, a new system is proposed with better efficiency, good user-experience and reduced complexity.

3.1 Proposed Solution

The system aims at developing an application that will aid the architects in advertising their portfolio in an innovative way using augmented reality. For user convenience, the application will be built on Android platform initially. This solves the problem of portability as the user can view the models sitting at home. The architectural models will be designed and built using CAD software. These models would be stored on a cloud database and retrieved whenever the target image is scanned. Instead of marker-based technology, marker-less will be used. marker-less AR does not need any pre-knowledge of a user's environment to overlay 3D data into reality and hold it to a fixed point in space. In ARchitect, this technology is used so that the object remains even if the marker is removed. Using this enables the user to only scan the image target once and move the device but still be able to interact with the 3D model.

The performance evaluation parameters are: Rendering speed, detailing of models and high accuracy. The project incorporates three main attributes:

- **Portfolio image** - On scanning the Architect's project portfolio image, regardless of the image being in any form i.e. physical or digital, the entire 3D model of the project will be displayed on the user's screen.
- **Architect's blueprint** - The user can view the 3D model of the floor plan by scanning the architect's blueprint.
- **Location-based information** - The real-life project will be scanned and its information like location, distance from landmark will be displayed using marker-less AR.

3.2 System Architecture

The development of the application comprises of 3 main components i.e. 3D modelling, cloud deployment and retrieval of the models, and building the Android application. The given System Flowchart (Figure 1) explains the entire Architectural Design of the application. It shows how exactly the data or commands will be processed starting from the user or the developer.

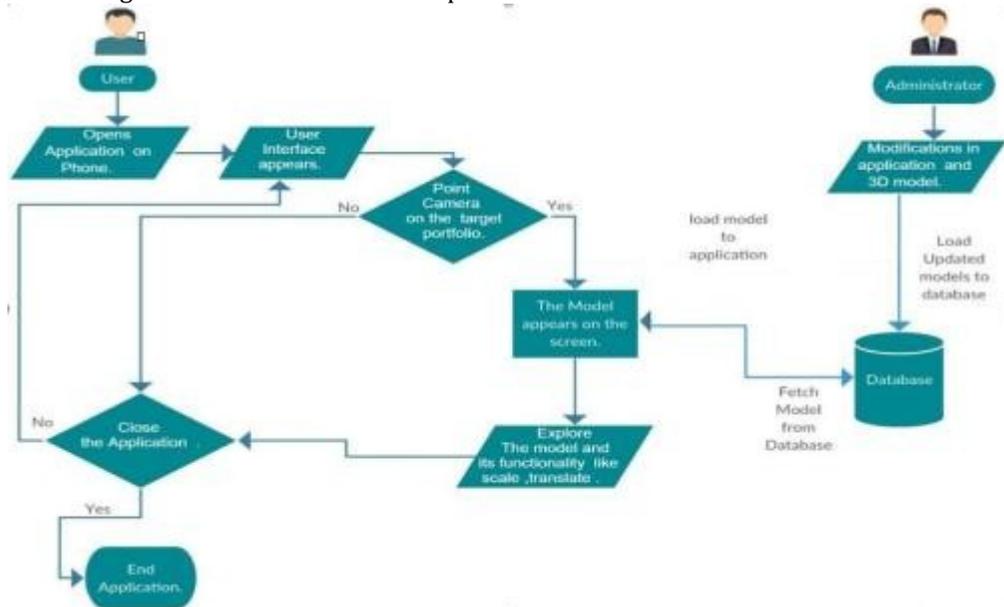


Fig.1:Architectural Design

The System is mainly divided into two parts that are the user side and administrator side. The user and the administrator have specific privileges in the system. Users are the individuals who make use of the basic functionalities of the application. Administrators are the developers of the system having privileged of the backend functionalities like modifications in the application and 3D models. In the following flowchart, the user accesses the interface of the application wherein the AR camera comes into play. The user has to point the camera towards a target Image set as a reference in the database or the user can point towards a plane surface where the user can use the functionality of the button to render an object in an open space. Once the object is rendered from the database connected to the application various tasks such as scale, translate, rotate can be executed.

3.3 Implementation

The system implementation will be divided into different stages which should be done in an incremental manner. The various steps are explained in detail below:

Step 1: Data Gathering

First, the architectural data is gathered for which the application is to be made. According to the data gathered the models are designed.

Step 2: Model Designing

As this project deals with architect designs, building structures and layouts are the basic design. To create 3D models Blender and unity are used. Blender v2.79 is an open source software which provides extensive functionalities in the field of modelling, animation, simulations, compositing etc.

Step 3: Importing models and setting up unity.

Once the model is developed it is imported to unity. To make the models visually appealing in the application various lightings and camera angles need to be set in unity.

Step 4: Uploading image targets and setting up the database is used to save the image targets.

Vuforia package is imported in unity which contains all the necessary prefabs. The license key of the database has to be entered in unity.

Step 5: Setting up functionalities

All the necessary functions are implemented using C# scripts.

Step 6: Building the application

Once Android is set up and all the settings are done the application is compiled and is ready to run.

Step 7: Testing

The targets are scanned via a smartphone/tablets camera and the object gets loaded. With the help of marker-less technology, the models should be placed manually and adjusted.

Step 8: Maintenance

As per the testing and user feedback, the application will be updated and bug fixes will be done along the way.

IV. Results

We have developed a basic version of ARchitect as a mobile AR application on Android platform. The application includes 3Dmodel database connectivity and tracking of the reference objects. The implementation of the application is illustrated below.

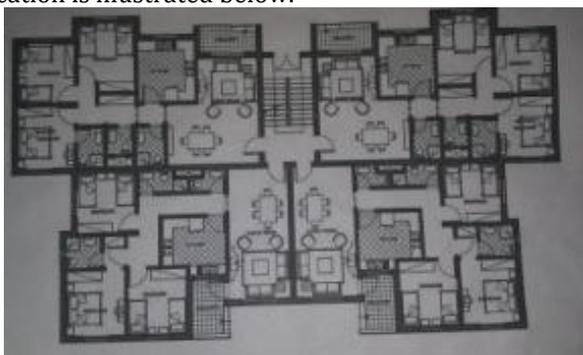


Fig.2: Blueprint of Model

Here Architects blueprint is used as reference target where the structure will render Figure 2 shows the blueprint of the building which should be scanned via the mobile device.

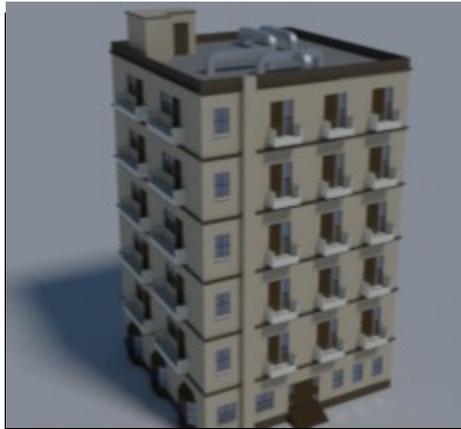


Fig.3: Model design in Blender

Figure 3 depicts the rendered image of the model of the outer structure of the architect’s portfolio building which was done using Blender



Fig.4: Implementation in Unity using Vuforia license

Figure 4 shows the model that was built using blender being implemented in augmented reality using Unity 3D and stored in Vuforia database.

V. Acknowledgement

We are highly grateful to our institute St. Francis Institute of Technology and the Information Technology Department for providing us with this learning opportunity with the required resources to accomplish our task. We wish to express our deep gratitude to our external guide Mrs. Purnima Kubde, Assistant Professor in the Department of Information Technology and Dr. Nazneen Ansari for all the advice, encouragement and constant supervision for the betterment of the project. This work would not have been possible without their intellectual and valuable suggestions. Our effort would have been incomplete without our Project coordinator Ms. Amrita Mathur.

VI. Conclusion

In this paper, we demonstrated an Augmented Reality based application which innovatively promoted the architect's portfolio. The application eases marketing in architectural field by automating the 3D models into an augmented reality portfolio application. The distinct features are the aptness of tracking the marker image (Portfolio images/Blueprints) and visualizing the 3D models on the users mobile or tablet screen. The architect's plans for their project are visualized in a portfolio that is used as a target for displaying the specific 3D models. Furthermore, the application supports dynamic changes to the 3D models by introducing various 3D transformation techniques which can be operated using the graphical user interface of the application. Online storing of 3D models into a database and retrieving them on scanning the target reduces memory requirement of the application on user's device. These features strongly elevate the user experience by making the process of browsing the portfolio interactive, user-friendly and portable.

References

1. Rattanaarungrot, S., & White, M. 2014. A service-oriented mobile augmented reality architecture for personalized museum environments. 2014 International Conference on Virtual Systems & Multimedia (VSMM). doi:10.1109/vsmm.2014.7136695.
2. Hyeon-seung Kim, Chang-hak Kim, Hyoun-seok Moon, So-yeong Moon, Young-hwan Kim, Leen-seok Kang. 2013. Application of Augmented Reality Object in Construction Project. Third World Congress on Information and Communication Technologies, Electronic ISBN: 978-1-4799-3230-6.
3. Krzysztof Walczak, Dariusz Rumiński, Jakub Flotyński. 2014. Building Contextual Augmented Reality Environments with semantics. International Conference on Virtual Systems and Multimedia (VSMM), DOI: 10.1109/VSM.2014.7136656.
4. Antonio Moura; Ingrida Mazonaviciute; Joao Nunes; Justinas Grigaravicius. 2012. Human lips synchronization in Autodesk Maya. International Conference on Information Management, Innovation Management, and Industrial Engineering.
5. M. Chu, J. Matthews and P. Love. 2017. Integrating mobile Building Information Modelling and Augmented Reality systems: An experimental study. Automation in Construction, vol.85, pp.305-316, 2018. Available: 10.1016/j.autcon.2017.10.032.
6. A. Clark, A. Dünser, and R. Grasset. 2011. An interactive augmented reality coloring book. SIGGRAPH Asia 2011 Emerg. Technol. - SA '11, pp.
7. L. Pombo and M. M. Marques. 2017. Marker-based augmented reality application for mobile learning in an urban park: Steps to make it real under the EduPARK project. International Symposium on Computers in Education (SIIE).
8. D.H. Shin, P.S. Dunston. 2008. Identification of Application Areas for Augmented Reality in Industrial Construction based on Technology Suitability. Automation in Construction, Elsevier Science B.V., Amsterdam, Vol. 17, No. 7, pp.882-894.
9. C. Lim, C. Kim, J.-I. Park, and H. Park. 2016. Mobile Augmented Reality Based on Invisible Marker. IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct).
10. M. Samir, A. Hanie, A. Aboulgheit, K. Hossam, A. Atia, and N. Elmasry. 2018. The Wanderer: Implementing markerless augmented reality with object position awareness. First International Workshop on Deep and Representation Learning (IWDRL).
11. Matthias Baldauf, Sebastian Zambanini, Peter Fröhlich, and Peter Reichl. 2011. Markerless visual fingertip detection for natural mobile device interaction. 13th International Conference on Human Computer Interaction with Mobile Devices and Services. ACM, 2011, pp. 539-544.
12. T.-W. Kao and H.-C. Shih. 2013. A study on the markerless augmented reality for picture books. IEEE International Symposium on Consumer Electronics (ISCE).
13. A. Ufkes and M. Fiala. 2013. A Markerless Augmented Reality System for Mobile Devices. International Conference on Computer and Robot Vision, 2013.

INTERACTIVE SMART MIRROR SCREEN

¹Er Kashif Kaularikar, ²Ali Mehdi Hemani, ³Danish Ayub Khan, ⁴Nehal Contractor

¹Assistant Professor, ²Student, ³Student, ⁴Student

¹Information Technology,

¹M. H. Saboo Siddik College of Engineering, Mumbai, India

ABSTRACT: *This paper describes the design and development of a smart mirror screen that represents an unobtrusive interface for the ambient home environment. The mirror provides a natural means of interaction through which the users can control the household smart appliances, check the weather, view schedule on a calendar and access personalized services. Emphasis is additionally given to confirm convenience in accessing these services with a minimum quantity of user intervention. Service-Oriented design has been adopted to develop and deploy the varied services, the appliances, and therefore the news and knowledge feeds all use internet service communication mechanisms. The smart mirror screen functionalities have been demonstrated by developing an easily extendable web widget. The integration of mirror with the smart device will allow increasing the usefulness of the mirror by utilizing the same space for added functionality and convenience to increase productivity and save users time. The smart screen makes it possible for the user to get information feeds, tasks, schedules, and even news while grooming and freshening up in the morning through the smart mirror screen. This mirror will provide added functionality to the same area making it more appealing to use it rather than a conventional mirror. It will digitalize the user's home making it futuristic and even act as decor making it more appealing.*

Keywords:

I. Introduction

Smart Screen promotes a paradigm where humans are surrounded by intelligent and natural interfaces offered by the interconnected heterogeneous computing devices embedded into everyday objects. The environment thus created is capable of recognizing and responding to the actions and presence of individuals (E. Aarts, H. Harwig, and M. Schuurmans. Ambient Intelligence, 2001). Smart screen can be seen as the driving force toward a more user-friendly and user-empowered smart environment for providing effective support to human interactions.

The smart screen is made aware of smart environments by using a variety of smart technologies. These technologies integrate sensing, processing, reasoning, and networking capabilities in addition to heterogeneous applications, services and digital contents. With all of these rich technologies involved, smart screen faces challenges on how to integrate them with the everyday objects, often unobtrusively, in order to provide computing intelligence in the surrounding environment.

The advanced technology available now plays a major role in providing intelligence about the home environment. We have smart door surveillance, smart wardrobe, smart dressing table, smart bed, smart pillow, smart mat, smart picture frame, and so on (S.H. Park, S. H. Won, J. B. Lee, and S. W. Kim, 2003). We aim to develop the smart screen as an interface to some of these smart devices that can be connected and controlled from a single interface.

The application of smart screen in the home environment may provide quality, convenience, efficiency, security, and safety to its residents (M. S. Raisinghani, A. Benoit, J. Ding, M. Gomez, K. Gupta, V. Gusila. D. Power, and O. Schmedding, 2004.). Aral for assisted living (F. Bomarius, M. Becker, and T. Kleinberger, 2006), especially for the elderly and the people with disabilities (P.L. Emiliani and C. Stephanidis, 2005) has already received much attention. Besides, the areas of home automation, communication and socialization, rest, refreshment, entertainment and sports, working, and learning at home (M. Friedewald, O. Da Costa, Y. Punie, P. Alahuhta, and S. Heinonen, 2005) will be influenced by the innovations of smart screen. Therefore, the smart systems in the house should not be only technology driven, but also, be able to blend in to the home surroundings and look natural. This provides comfort and convenience to individuals living in the home.

In this paper, we make the following contribution. First, we proposed and developed a futuristic smart mirror screen using off-the-shelf technologies that provide personalized data feeds, and other services in

addition to controlling the smart appliances in the household. The mirror can be used as a traditional mirror that essentially provides a sense of natural interaction with the surrounding environment. Second, we provide an easily extendable framework for integrating smart appliances and services with the mirror interface in order to automate the home environment.

The remainder of this paper is organized as follows. Section 2 briefly comments on some related works. This is followed by the description of the smart mirror screen including the design and architecture of the underlying home automation framework in section 3. The implementation detail, test cases and screen snapshots of the proposed system are provided in section 4. Conclusion and some thoughts on future work are presented in section 5.

II. Related Work

The proposed smart mirror screen represents a natural interface that facilitates access to personalized services and control of household smart appliances in the ambient home environment. This is an attempt to contribute to the design of a smart mirror-like interface as well as the smart environment in which the interface is used for interaction in the following, we briefly comment on some related research in this direction.

Among several projects, the work on creating an intelligent personal care environment uses an Interactive Mirror (Tatiana Lashina, 2004) in the bathroom to provide personalized services according to the user's preferences. For example, children can watch their favorite cartoon while brushing their teeth. The mirror can provide live TV feeds, monitor the latest weather, check the traffic information, and so on. The mirror is a combination of one or more LCD flat screen displays specifically combined with a mirrored surface and connected with a central processor to provide the intended services.

The AwareMirror (K. Fujinami, F. Kawsar, and T. Nakajima, 2005) is an augmented display that is placed in the bathroom for presenting personalized information to the user. It detects the position of a person in the bathroom using a proximity sensor and identifies her/him from the usage of a toothbrush. It provides useful information such as closest schedule, transportation information, and the weather forecast. The mirror is constructed by attaching an acrylic board in front of a monitor. A slider sensor on the mirror is used to navigate the information presented on the mirror. Although it attempts to provide an intuitive interface, it has some limitations that may restrict it from wider usage. For example, the state-of-use of a toothbrush for identifying a person might not provide accurate states to personalize information. Also, the use of magic mirror restricts dark color from going through it; and hence, requires special attention to the color of the contents to be displayed.

Another implementation (P.L. Emiliani and C. Stephanidis, 2005) can act as an assistant to elderly people by graphically showing the status of drug usage over a 24-hour period of time. It can keep track of all the drugs removed from the medicine cabinet and records it in a history log, in order to display the details of previous usage and to warn about possible lost and/or misplaced items. It may not be used as a traditional mirror and it is not suitable as an interface to access personalized services and to control household appliances, which are the two main goals we intend to achieve.

The work in (L. Ceccaroni and X. Verdaguer, 2004) proposes a Magical Mirror as an interface to provide basic services and enable control of Household appliances. The intended services to offer are interactive TV, weather data, news, music, reminders, and searches. Unlike our work, it promotes the use of ontology to personalize the services. However, conceptually, our work has similar objectivity to what the Magical Mirror intends to perform, except that we present a working prototype, whereas some of the functionalities in the Magical Mirror have been presented only by simulation such as the control of appliances. In addition, we use open standards like web services to communicate with the devices and customize various personalized services for the user, which is not present in the design of the Magical Mirror.

An earlier work (K. Ushida, Y. Tanaka, T. Naemura, and H. Harashima, 2002) presents i-mirror that attempts to include information services within the mirror interface as a natural way of providing interactive experiences to people. The i-mirror uses dedicated optical system including a video camera, magic mirrors, and a video projector to imitate a mirror. Its use of magic mirror to superimpose an image to the mirror is similar to the one in AwareMirror [12]. Three potential uses of the i-mirror have been explored such as the one with ability to show images in the dark; one with the capability of providing younger/older looks; and the one with memory.

In comparison to i-mirror and other works described above, our work is different in that we aim to develop a working system for providing services in the ambient home environment based on open standards and

off-the-shelf technology, where the smart mirror screen is the interface to access/control various data feeds, information services, and appliances in the environment.

III. Proposed Smart Mirror

Figure 1 and Figure 2 shows a schematic view of the proposed smart mirror screen. The smart mirror screen is eventually a technologically augmented interaction device through voice. The objective of designing the mirror is to provide a natural interface in the home environment for accessing various information services (news feeds, weather updates etc.) as well as controlling household smart appliances (light, camera, etc.) when required.

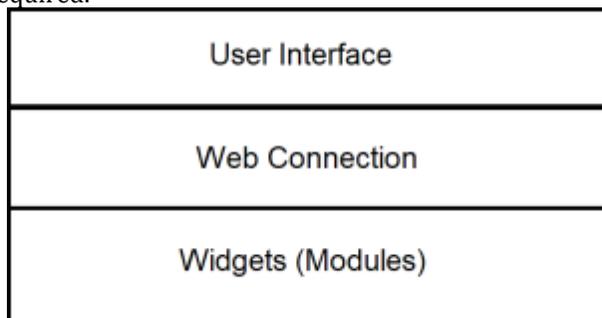


Figure 4

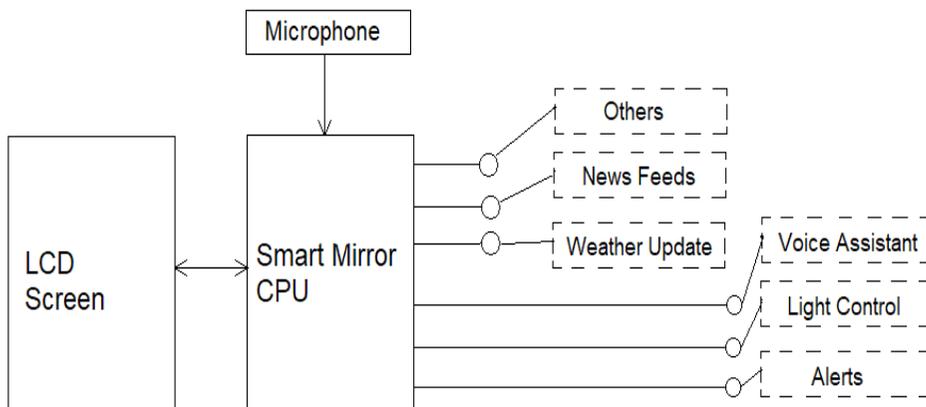


Figure 2

3.1 Functional overview

The proposed mirror is designed to perform several functionalities that can be summarized as follows:

- Mimic a natural mirror interface: A Two-Way Mirror is used to reflect the environment close to how a mirror reflects it.
- Remote control access to appliances: After the appliances such as lights, camera etc. has been added to the smart mirror screen, their functions become accessible from any terminal, in this case the smart mirror screen. All appliances will have standard features such as on and off, as well as customized ones when applicable, such as dimming the lights in different rooms of the house.
- Personalized information services: Users will have the ability to tap into various XML-based data such as Real Simple Syndication (RSS) feeds. This will allow them to obtain up-to-the-minute updates and converge on topics of interest, such as public news headlines, weather bulletins, stock reports, and consumer news updates on products of interest, etc.
- Home surveillance: The system provides access to the live feeds from the home security camera installed in strategic locations. This offers the user such conveniences as to be able to check who is at the door while doing another activity.

In the following section we describe the design and architecture of our system through which we have implemented the above functionalities.

3.2 Design and architecture

The design of the smart mirror screen can be seen from two perspectives. One is the mirror interface and the other is its underlying infrastructure based on which services are provided.

Figure 1 shows the high-level view of the mirror's user interface (UI). There are several aspects in this design. The mirror interface is decorated with several widgets. A widget is a simple window frame that contains an embedded browser. Unlike a window, widgets do not overlap nor do they contain complex interface elements. The mirror interface contains two types of widgets, one that enables remote device control (e.g. light on/off etc.), and the other enables access to various information services (e.g. news feeds, weather updates etc.).

Figure 3 shows the overall home automation system architecture. The mirror interface/screen described earlier stands as an interface for this system to access various services. The backbone of the system consists of Raspberry Pi 3. The devices like microphone, speaker are connected with the Raspberry Pi at strategic locations to minimise space and maximise throughput. This single design incorporates several functionalities, it is possible to cluster the services and use one or two Pi for the tasks to improve speeds. Also, in the architecture, the internet network leverages the latest standardized communication protocols such as web services and RSS feeds, thus allowing a heterogeneous operating environment to co-exist within the infrastructure. For example, the output on the display can be mirrored onto any device with a JavaScript enabled browsersystem. CPU would be able to expose its attached smart appliances via self-hosted web services, while consuming Real Simple Syndication (RSS) feeds streamed from the Internet gateway. The integration performed is seamless.

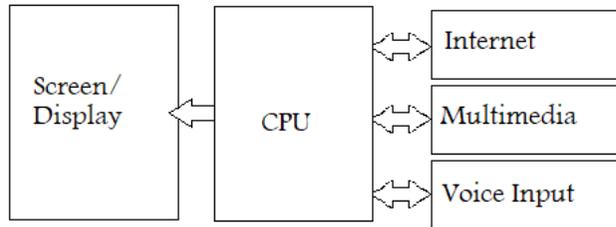


Figure 5

IV. PROTOTYPE IMPLEMENTATION

The system prototype is implemented using some keyenabling technologies. These are the web services and Electron JS. besides, the system uses a flat monitor, Two Way Mirror and a raspberry pi 3.

Using the aforementioned technologies, we developed our proposed system. As shown in figure 4, the system component diagram consists of five high-level components. These components are the User Interface, Hardware Abstraction, Communication. We describe each of these components in the following section.

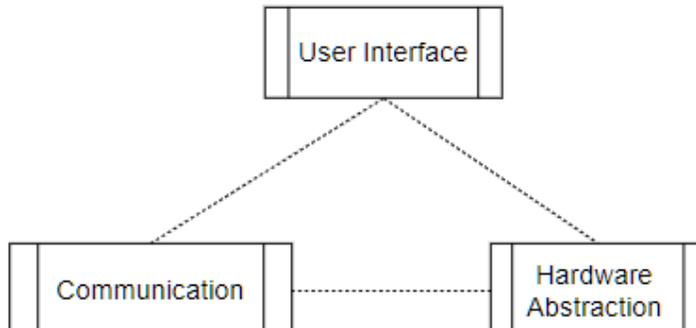


Figure 6

The User Interface (as presented in figure 2) is the application that runs on the mirror node, and is what the user comes directly in contact with. It consists of a collection of web widgets created in Electron JS that allows for another way of easily and quickly extending the system. Modules can be developed quickly, and are cross platform compatible. They are also based on free open source project maintained by GitHub and an active community of contributors. They are meant to provide the user easy access to all the necessary information at one glance. The responsibility of the User Interface component is to provide a framework engine for the widgets.

The Hardware Abstraction component is responsible for wrapping the various devices with device drivers that can be invoked from within a web service. This is where new devices should be added. Third party devices can be easily recognized and exposed to the smart mirror screen OS during runtime to control them remotely from the client. In our system Wi-Fi is used for control between the client device and the smart screen that the client is trying to control. Typically, the client computer is not physically near the device but uses the same Wi-Fi transceiver to send wireless signals to the smart screen.

The Communication component is responsible for taking care of messaging between various components. Web services are used to facilitate communication in our system, by allowing devices to expose services over the Internet in order to be accessed or called upon by other devices. Consumed by web widgets mentioned earlier, web services help separate functionalities from the user interface and provide a flexible way for smart devices to communicate. A single web service is inherently using the Server-Client pattern. The Communication component will work on any home network that supports TCP/IP. Overall, the use of web services allows us to create a system, which is based on an open standard. Systems developed with web services for ambient intelligence have been practiced by other researchers such as (V. Issamy, D. Sacchetti, F. Tartanoglu, F. Sailhan, R. Chibout, N. Levy, and A. Talamona, 2005).

V. Results

We have designed a futuristic smartmirror screen that provides voice interaction between users and the smart screen. The functionalities include the control of household appliances and access to personalized information services. The smart mirror screen may be a combination of 1 or a lot of display specifically combined with a reflected surface and connected with a mainframe to supply the meant services. We have developed a purposeful epitome to demonstrate our work. Overall, the prototype provides an easily extendable framework that can be utilized to provide even more functionality to the user. In our future work, we will investigate how the surrounding context of the user and the environment can be utilized in order to provide optimal service experiences in the home environment by controlling some of the connected devices.

References

1. Ushida, Y. Tanaka, T. Naemura, and H. Harashimaa2002, i-mirror: An interaction/information environment based on a mirror metaphor aiming to install into our life space. In Proceedings of the 12th International Conference on Artificial Reality and Telexistence (ICAT2002).New York, NY, USA. 113-118.
2. M. S. Raisinghani, A. Benoit, J. Ding. M. Gomez, K. Gupta, V. Gusila. D. Power, and O. Schmedding2004. Ambient intelligence: Changing forms of humancomputer interaction and their social implications. Journal of Digital Information, 5(4).
3. P.L. Emiliani and C. Stephanidis2005. Universal access to ambient intelligence environments: Opportunities and challenges for people with disabilities. IBM SystemsJournal, 44(3):605-619.
4. E. Aarts, H. Harwig, and M. Schuurmans. Ambient Intelligence2001, The Invisible Future. McGraw Hill, New York. 235-250.
5. F. Bomarius, M. Becker, and T. Kleinberger2006. Embedded intelligence for Ambient Embed- assisted living. ERCIM News, 67:19-20.
6. Ceccaroni and X. Verdaguer2004. Magical mirror: multimedia, interactive services in home automation. In Proceedings of the Workshop on Environments for Personalized Information Access - Working Conference on Advanced Visual Interfaces (AVI 2004), 10-21.
7. P.L. Emiliani and C. Stephanidis2005. Universal access to ambient intelligence environments: Opportunities and challenges for people with disabilities. IBM Systems Journal, 44(3):605-619.
8. M. Friedewald, O. Da Costa, Y. Punie, P. Alahuhta, and S. Heinonen2005. Perspectives of ambient intelligence in the home environment. Telematics and Informatics, 22(3):221-238.
9. Fujinami, F. Kawsar, and T. Nakajima2005. AwareMirror: A personalized display using a mirror. In Pervasive, 315-332.
10. V. Issamy, D. Sacchetti, F. Tartanoglu, F. Sailhan, R. Chibout, N. Levy, and A. Talamona2005. Developing ambient intelligence systems: A solution based on web services. Automated Software Engineering, 12(1):101-137.
11. Tatiana Lashina2004. Intelligent bathroom in European Symposium on Ambient Intelligence (EUSAI' 04), Eindhoven, Netherlands, 92-99.
12. S.H. Park, S. H. Won, J. B. Lee, and S. W. Kim2003. Smart home - digitally engineered domestic life. Personal Ubiquitous Computing, 7(3-4): 189-196.

APPLICATION OF STATISTICAL METHODS IN MEDICAL SCIENCES

¹Manjiri P. Vartak & ²Daksha P. Sawant

^{1,2}Assistant Professor, Statistics Department, Mithibai College, Mumbai, India

ABSTRACT: Nearly 52% of Indians aren't aware that they are suffering from high blood sugar. By 2030, India's diabetes numbers are expected to cross the 100 million mark.-International Diabetes Federation It's a serious major social health issue and hence we decided to analyze this using statistical techniques.

Major findings:

1. Factors which affect blood sugar level.
2. Various health disorders in type 2 Diabetes patients.
3. Ideal lifestyle for Type 2 Diabetes Patients.

Area of Conference: Social determinants of health

Keywords: Binary logistic regression, Graphical techniques, Odds ratio, SPSS, EXCEL.

I. Introduction

Diabetes is a group of metabolic diseases in which a person has high blood sugar, either because the body does not produce enough insulin, or because cells do not respond to the insulin that is produced. Blood glucose level is the amount of glucose (sugar) present in the blood of a human.

Blood glucose may be measured randomly (anytime), Fasting basis (collected after an 8 to 10 hour fast), Postprandial (after a meal).

Normal Range: Fasting: 70-110 mg/dl; Postprandial: 70-140 mg/dl. (mg/dl refers to Milligrams per Deciliter).

The diseases which are listed under Diabetes Mellitus are many with the most common being Type-1 diabetes and Type-2 diabetes. There is currently a worldwide epidemic of Type-2 diabetes which researchers believe is being driven in many areas because of sedentary life styles revolving around computers, video games, television and fast foods. Fast food restaurants are now in every corner of our world. That is why our project focuses on TYPE 2 DIABETES wherein we target the diabetic population and find out the underlying causes and risks of diabetes.

II. Objectives

- 1.To study factors which affects on controlling of blood sugar level.
- 2.To study various health disorders in type 2 Diabetes patients.
- 3.To suggest ideal lifestyle for Type 2 Diabetes Patients.

III. Methodology

i. Preparatory phase	ii. Designing the Questionnaire	iii. Pilot survey
iv. Modification of Questionnaire	v. Actual survey	vi. Data cleaning
vii. Coding of data	viii. Analysis using SPSS, EXCEL.	ix. Interpretation and conclusion

IV. Scope and Limitations

- a. **Target population: Type 2 Diabetes patients.**
 - b. Data collection sites: Various pathology laboratories and Diabetologist's clinics in Mumbai
 - c. Collected data of patient's lifestyle for a period of one month previous to the current report of blood glucose test by face-to- face interview.
 - d. Pilot Survey: Sample Size = 50 ; Main Survey: Sample Size = 610 ; Sample Size after cleaning = 589.
 - e. Sample consisted of **320 male** and **269 female** respondents.
4. **Data Interpretation**
 - i. Most affected age group: 50-60 years.
 - ii. In our sample 64% of the patients have Diabetes due to heredity.

- iii. Most of the people (46%) think that doing Exercise is most difficult task over diet, meditation, testing sugar level.
- iv. Most of the patients (70%) think that 'Diet' is a very significant factor which affect on blood sugar level.
- v. International Diabetes Federation (IDF) has accepted BMI value of >25kg/m² and 23kg/m² as the cut off for obesity for Indian men and women respectively, in our sample 71% male&75% female are obese.
- vi. In our sample 53% of patients use sugar substitute.
- vii. In our data 42% of patients tested their blood glucose level monthly.

5. Statistical Analysis: Model Building

Now we can fit the **BINARY LOGISTIC REGRESSION** model to the data with 9 independent variables to predict probability of blood sugar in control and to find factors which affects blood sugar level.

Our dependent variable y is binary

y=1 if blood sugar is in control and y=0 if blood sugar is not in control.

Binary logistic model: $\ln \frac{p}{1-p} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_q X_q$

Factors affecting blood sugar level

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
							Lower	Upper
BMI	-.128	.032	16.193	1	.000	.880	.826	.936
Rice freq.	-.126	.007	.007	1	.000	.890	.859	1.014
Yoga	.071	.008	82.530	1	.000	1.073	1.057	1.090
Walk(m/e)	.048	.007	51.348	1	.000	1.049	1.036	1.063
Stress	-.263	.049	28.486	1	.000	.769	.698	.847
Gender(1)	.089	.299	.089	1	.765	1.093	.608	1.965
Age	-.011	.021	.257	1	.003	.989	.950	1.031
First diagnose	.010	.023	.175	1	.676	1.010	.965	1.057
Exercise	.049	.015	10.569	1	.001	1.051	1.020	1.082
Constant	1.143	1.100	1.080	1	.299	3.136		

Variable(s) entered on step 1: BMI, Rice Freq., Yoga, Walk (morning/evening), Stress, Gender, Age, Age when diabetes is first time diagnose, Exercise.

Positive estimates indicates that presence of variable increases chance of blood sugar in control.

Negative estimates indicate that presence of variable decreases chance of blood sugar in control.

Same can be interpreted as follows:

- i. Increase in number of rice bowls multiplies the odds of having blood sugar level in control by 0.89. So the odds of having blood sugar level in control decreases with increase in number of rice bowls.
- ii. Increase in yoga duration multiplies the odds of having blood sugar level in control by 1.073. So the odds of having blood sugar level in control increases with increase in yoga duration.

As 86.8% of people were correctly classified, hence model is 86.8% accurate.

The addition of explanatory variables increases the percentage of correct classification significantly

CLASSIFICATION TABLE

	Observed	Predicted		
		BSL not control	BSL control	Percentage Correct
Step 1	Not control	416	38	91.6
	control	40	95	70.4
	Overall Percentage			86.8

The cut value is .500

THEORETICAL FRAMEWORK

Omnibus Tests of Model Coefficients

	Chi-square	df	Sig.
Step	318.120	9	.000
Step 1 Block	318.120	9	.000
Model	318.120	9	.000

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	14.299	8	.074

Model is good fit as p-value>0.05

Fitted model is

$$\ln\left(\frac{p}{1-p}\right) = 1.143 - 0.128\text{BMI} - 0.126 \text{ (RiceFreq.)} + 0.071(\text{Yoga}) + 0.048(\text{WALK(M/E)}) - 0.263\text{stress} + 0.089(\text{male}) - 0.011(\text{Age}) + 0.010(\text{First diagnose}) + 0.049(\text{Exercise}).$$

Conclusions:

As we have fulfilled all the objectives using several statistical techniques; based on that we have drawn conclusions.

Significant factor affecting blood sugar level

- i. BMI
- ii. Rice Frequency
- iii. Yoga
- iv. Morning Evening Walk
- v. Stress
- vi. Gender
- vii. Age
- viii. Age when diabetes is first time diagnose
- ix. Exercise

➤ **Other Health Disorders**

If high glucose levels retain in blood for a long time period then certain parts of the body begins to suffer. High glucose levels in blood damages the blood vessels & lead to thicker blood & other complications. There can be long-term complications of type 2 diabetes, especially if not treated properly.

Other Health disorders in Diabetes patients



Is there any connection between significant factor for type 2 Diabetes Other health disorders?

Odds Ratio measures of the strength of association or non-independence.

It approximates how much more likely it is for the outcome to be present among those with x=1 than among those with x=0.

		Independent Variable X	
		X=1	X=0
Outcome Variable Y	Y=1	a	b
	Y=0	c	d

Odds ratio = $\Psi = (a/b)/(c/d)$

	Odds Ratio
Heart Disease Vs. Non-veg	2.28
Heart Disease Vs. Obesity	1.90
Heart Disease Vs. Daily Physical Activity	0.50
Heart Disease Vs. Alcohol	2.23
Heart Disease Vs. Smoking	1.49
High BP Vs. Stress	1.62
Obesity Vs. Daily Physical Activity	0.47
Obesity Vs. Fried Food	1.87

Interpretation of odds Ratios

- **Heart disease** is **1.90** times more likely in people who are **obese** than who are not obese.
- **Heart disease** is **0.50** times less likely in people who do **daily physical activity** than who do not do

9.3 Recommendations for an ideal lifestyle for a diabetic based on our study

We sorted patients whose blood glucose level is in control & studied their life style & formulated the following day-plan for diabetes patients.

Factors	Plan
Smoking/alcohol/tobacco	Strictly prohibited
Exercise/morning or evening walk/yoga	30 mins+
Meditation	20 mins+
Sleeping hours	7 hrs per day
Lunch time	12:30 pm
Dinner time	8:30 pm
Rice frequency	1 bowl a day at afternoon
Tablets/insulin	As prescribed by the doctor
Balanced diet	Followed regularly

References

1. Johnson Richard A,1998. Applied Multivariate Statistical Analysis (4th Edition).
2. Gujarati, Damodar N., 1995. Basic Econometrics (McGraw-Hill Inc.) (Second Edition).
3. Cody, Ronald P. & Smith, Jeffrey K., 2006. Applied Statistics and the SAS Programming language (Elsevier Science Publishing Company) (Fifth Edition).
4. *Montgomery* and Peck, Introduction to Regression Analysis, Wiley. D.C.
5. Dielman Terry, Applied multiple regression
6. <http://www.statisticssolutions.com/methods-chapter/statistical-tests/the-multiple-linear-regression-analysis-in-spss/>
7. <http://www.stat.vcu.edu/help/SPSS/SPSS.Reggression.PC.pdf>
8. http://dss.princeton.edu/online_help/analysis/interpreting_regression.htm
9. www.kellogg.northwestern.edu/kis/tek/.../Introduction2SPSS.pdf

Similarity Analysis of an Unsteady Flow of Non-Newtonian Fluid past through a Porous Surface via Multiparameter Group Technique

¹Jayshree Gangani & ²M. G. Timol

¹Assitant professor,²professor

¹Department of Mathematics, ¹Uka Tarsadia University, Bardoli, Gujarat, India.

²Department of Mathematics, ²Veer Narmad South Gujarat University, Surat395007, Gujarat,India.

ABSTRACT: Multiparameter group technique is developed to find the similarity analysis of non-linear partial differential equations. The proposed method is applied to derive invariant solution for the system of non-linear partial differential equations governing the two-dimensional, unsteady squeezed flow over a porous surface for a power-law non-Newtonian fluid. In analysis the continuity, momentum and energy equations are written and cast into a non-dimensional form with the most general form of boundary conditions. Then partial differential systems with three independent variables are converted into an ordinary differential system, via single application using proposed multiparameter group method. The numerical solution of this system of ordinary differential equations is derived. Effects of flow behavior index and Prandtl number on the flow are also discussed.

Keywords: Similarity analysis, Group theoretic technique, Non-Newtonian Fluid, Power-law- fluid, Unsteady Flow

1. Introduction

While there is no existing general theory to find exact solution of nonlinear partial differential equations, many special cases have yielded to appropriate changes of variable. Transformations are the most powerful general analytic tool currently available in this area. Typically, these linearize the systems (for example the Kirchoff and hodograph transformations), reduce the equation to a nonlinear ordinary differential equation (for example the similarity transformation), transform the system to one already solved, or perform some other reduction of complexity. In general, these changes can be classified into three groups: Class I includes those, which are transformations only of the dependent variables, Class II, includes transformations only of the independent variables, and Class III consists of transformations of both dependent and independent variables. The mixed transformation in which both sets of variables are involved appears to be the most productive of the three general classes. Study of Non-Newtonian fluid flow has attracted much attention and importance than Newtonian fluids, due to their various technological and industrial applications. Boundary layers of Non-Newtonian fluids have attracted great interest because of their importance in engineering applications, such as metal extrusion, lubrication, food processing and heat exchangers, etc.

Similarity solutions of the laminar boundary-layer equations describing heat and flow in a quiescent fluid driven by a stretched surface subject to suction or injection obtained by Ali [1995]. The effect of thermal boundary layer on a stretched surface for a Non-Newtonian fluid was studied by Chamkha [1997], who obtained similarity solutions. The symmetry reductions of unsteady three-dimensional boundary layers of some non-Newtonian fluids were investigated by Yürüsoy and Pakdemirli[1997]. Tashtoush et al. [2001] investigated the heat transfer characteristics of non-Newtonian fluid on a power-law stretched surface of variable temperature with suction or injection. Aydın and Kaya [2005] presented the similarity solution and the numerical solution of laminar boundary layer flow over a horizontal permeable flat plate. Yürüsoy [2006] treated the unsteady boundary layer equations of power-law fluids over a stretching sheet.

In recent years, boundary layer studies have been focused on the chemical and biological detection systems, which involve extended surfaces called micro cantilevers. These extended surfaces are modeled as porous surfaces by many researchers working on Newtonian fluids. Khaled and Vafai [2004] analyzed the hydro magnetic effects on flow and heat transfer over a sensor surface. Mahmood et al. [2007] investigated flow and heat transfer over a permeable sensor surface placed in a squeezing channel. However, most chemical or biological liquids are non-Newtonian fluids.

In the present paper we have considered the two-dimensional, unsteady squeezed flow over a porous surface for a power-law non-Newtonian fluid is considered. Continuity, momentum and energy equations

are written and cast into a non-dimensional form. Boundary conditions are selected in a general form. Multiparameter group technique is applied to the equations. The partial differential system with three independent variables is converted into an ordinary differential system. The ordinary differential equations are solved numerically. Effects of flow parameters on the development of momentum and thermal boundary layers are discussed.

Nomenclature

u, v	Velocity component in the x and y direction	c_p	Specific heat conductivity of the fluid
U	Free stream velocity component	T	Temperature
ρ	Density of the fluid	n	Flow behavior index
k	Flow consistency index	x, y	Cartesian coordinates
k	Thermal conductivity of the fluid	T_s	Surface temperature
T_∞	Temperature of the ambient fluid	U_0	Average free stream velocity
L	Length of the horizontal surface	R_e	Reynolds number
P_r	Prandtl number	ψ	Stream function
A, B	Constants of fluid models	η	Transformed independent variable
F_1	Transformed dependent variable	F_2	Transformed dependent variable

2. Governing Equations (Mathematical Model)

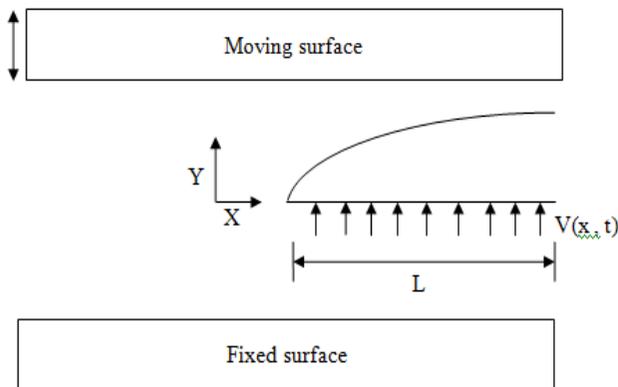


Figure -1 Geometrical picture

The flow geometry of the present problem is shown in Figure 1. Accordingly the basic governing equations of the two dimensional unsteady, incompressible laminar over a horizontal surface enclosed in a squeezing channel are given as [Akgil and Pakdemirli (2012)]

$$\frac{\partial \bar{u}}{\partial \bar{x}} + \frac{\partial \bar{v}}{\partial \bar{y}} = 0 \quad (1)$$

$$\rho \left(\bar{u} \frac{\partial \bar{u}}{\partial \bar{x}} + \bar{v} \frac{\partial \bar{u}}{\partial \bar{y}} + \frac{\partial \bar{u}}{\partial \bar{t}} \right) = nK \left| \frac{\partial \bar{u}}{\partial \bar{y}} \right|^{n-1} \frac{\partial^2 \bar{u}}{\partial \bar{y}^2} + \rho \left(\frac{\partial \bar{U}}{\partial \bar{t}} + \bar{U} \frac{\partial \bar{U}}{\partial \bar{x}} \right) \quad (2)$$

$$C_p \rho \left(\bar{u} \frac{\partial \bar{T}}{\partial \bar{x}} + \bar{v} \frac{\partial \bar{T}}{\partial \bar{y}} + \frac{\partial \bar{T}}{\partial \bar{t}} \right) = K \frac{\partial^2 \bar{T}}{\partial \bar{y}^2} \quad (3)$$

With the boundary conditions are

$$\left. \begin{aligned} \bar{u}(\bar{x}, 0, \bar{t}) &= \bar{s}(\bar{x}, \bar{t}), \quad \bar{v}(\bar{x}, 0, \bar{t}) = \bar{V}(\bar{x}, \bar{t}), \quad \bar{T}(\bar{x}, 0, \bar{t}) = T_s, \\ \bar{u}(\bar{x}, \infty, \bar{t}) &= \bar{U}(\bar{x}, \bar{t}), \quad \bar{T}(\bar{x}, \infty, \bar{t}) = T_\infty \end{aligned} \right\} \quad (4)$$

Where $\bar{V}(\bar{x}, \bar{t})$ is the suction or injection velocity of the permeable surface and $\bar{s}(\bar{x}, \bar{t})$ is the moving surface velocity. T_∞ is the temperature of the ambient fluid and T_s is the surface temperature.

3. Formulation of the problem

Now considering the dimensionless quantities in the equations (1) - (4). Where non-dimensional quantities used are (The non-dimensional parameters are related to their dimensional ones through the following relations):

$$u = \frac{\bar{u}}{U_0}, \quad v = \frac{\bar{v}}{U_0} R_e^{\frac{1}{n+1}}, \quad U = \frac{\bar{U}}{U_0}$$

$$x = \frac{\bar{x}}{L}, \quad T = \frac{\bar{T} - T_\infty}{T_s - T_\infty}, \quad y = \frac{\bar{y}}{L} R_e^{\frac{1}{n+1}}$$

$$t = \frac{U_0 \bar{t}}{L}, \quad R_e = \frac{\rho U_0^{2-n}}{K} L^n, \quad P_r = \frac{C_p \rho U_0 L}{K R_e^{\frac{n}{n+1}}}$$

$$s = \frac{\bar{s}}{U_0}, \quad F(x, t) = \frac{\partial U}{\partial t} + U \frac{\partial U}{\partial x}, \quad v = \frac{\bar{v}}{U_0} R_e^{\frac{1}{n+1}}$$

Where U_0, L, R_e and P_r are the average free stream velocity, length of the horizontal surface, Reynolds number and Prandtl number respectively. $F(x, t)$ is defined above from the free stream velocity for simplicity.

The non-dimensional form of equations (1) - (3) and boundary conditions (4) are as follows:

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0 \quad (5)$$

$$u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} + \frac{\partial u}{\partial t} = n \left| \frac{\partial u}{\partial y} \right|^{n-1} \frac{\partial^2 u}{\partial y^2} + F(x, t) \quad (6)$$

$$u \frac{\partial T}{\partial x} + v \frac{\partial T}{\partial y} + \frac{\partial T}{\partial t} = \frac{1}{Pr} \frac{\partial^2 T}{\partial y^2} \quad (7)$$

With the boundary conditions are

$$u(x, 0, t) = s(x, t), \quad v(x, 0, t) = V(x, t), \quad T(x, 0, t) = 1,$$

$$u(x, \infty, t) = U(x, t), \quad T(x, \infty, t) = 0 \quad (8)$$

$$\text{where } F(x, t) = \frac{\partial U}{\partial t} + U \frac{\partial U}{\partial x}$$

Introducing a non-dimensional stream function $\psi(x, y)$ such that

$$u = \frac{\partial \psi}{\partial y} \quad \text{and} \quad v = -\frac{\partial \psi}{\partial x} \quad (9)$$

which satisfies equation (5) identically.

Then equations (6) and (7) becomes

$$\frac{\partial^2 \psi}{\partial t \partial y} + \frac{\partial \psi}{\partial y} \frac{\partial^2 \psi}{\partial x \partial y} - \frac{\partial \psi}{\partial x} \frac{\partial^2 \psi}{\partial y^2} - \frac{\partial \psi}{\partial t} - U \frac{\partial \psi}{\partial x} = n \left| \frac{\partial^2 \psi}{\partial y^2} \right|^{n-1} \frac{\partial^3 \psi}{\partial y^3} \quad (10)$$

$$\frac{\partial T}{\partial t} + \frac{\partial \psi}{\partial y} \frac{\partial T}{\partial x} - \frac{\partial \psi}{\partial x} \frac{\partial T}{\partial y} = \frac{1}{Pr} \frac{\partial^2 T}{\partial y^2} \quad (11)$$

The boundary conditions (8) now become

$$\frac{\partial \psi}{\partial y}(x, 0, t) = S(x, t), \quad \frac{\partial \psi}{\partial x}(x, 0, t) = -V(x, t), \quad T(x, 0, t) = 1,$$

$$\frac{\partial \psi}{\partial y}(x, \infty, t) = U(x, t), \quad T(x, \infty, t) = 0 \quad (12)$$

Where t is time, $S(x, t)$ is surface temperature function, $U(x, t)$ is the suction or withdrawal and blowing velocity, $V(x, t)$ is the suction or withdrawal and blowing velocity of the porous surface.

The boundary value problem (10) and (11) are coupled non linear partial differential equations whose exact solution is indeed tough. We transform this set of non linear partial differential equations into ordinary differential equations by similarity technique so that it's numerical or closed form solution can be obtained. A multi parameter group technique is employed to find the proper similarity transformations. So similarity solution will exit.

4. METHODOLOGY OF THE PROBLEM

Our method of solution depends on the application of a two-parameter linear group of transformations. Group analysis is the only precise mathematical method to find all symmetries of a given differential equation and no adhoc assumptions or a prior knowledge of the equation under investigation is needed. The non-linear character of the partial differential equations (10), (11) subject to the boundary condition (12) of the motion of the fluid creates the difficulty in solving the equations. Commonly in this field of fluid mechanics, most of the researchers try to obtain the similarity solutions in such cases to the partial differential equations along with boundary conditions. Under this transformation the three independent variables will be reduced by one and the differential equations will transform into the ordinary differential equations.

4.1 The group systematic formulation of the problem

Introducing the group theoretic method, let us consider 2-parameter linear group transformation defined as

$$G_3: \begin{cases} t = A^{\alpha_1} \bar{t}; & x = B^{\beta_1} \bar{x}; & y = A^{\alpha_3} B^{\beta_3} \bar{y} & \text{(independent variables)} \\ \psi = A^{\alpha_4} B^{\beta_4} \bar{\psi}; & T = A^{\alpha_5} B^{\beta_5} \bar{T}; & U = A^{\alpha_6} B^{\beta_6} \bar{U} & \text{(dependent variables)} \end{cases} \quad (13)$$

Where α_i 's and β_i 's, A, B are constants we now seek relations among α_i 's and β_i 's such that the basic equations will be invariant under this group of transformation.

4.2 Invariance analysis

This can be achieved by substituting the group transformation G_3 into Equations (10), (11) and applying chain rule for transforming the derivative, the above Eqs. (10), (11) and boundary conditions (12) become

$$\begin{aligned} & A^{\alpha_4 - \alpha_1 - \alpha_3} B^{\beta_4 - \beta_3} \frac{\partial^2 \bar{\psi}}{\partial \bar{t} \partial \bar{y}} + A^{2\alpha_4 - 2\alpha_3} B^{2\beta_4 - 2\beta_3 - \beta_1} \frac{\partial \bar{\psi}}{\partial \bar{y}} \frac{\partial^2 \bar{\psi}}{\partial \bar{x} \partial \bar{y}} \\ & - A^{2\alpha_4 - 2\alpha_3} B^{2\beta_4 - 2\beta_3 - \beta_1} \frac{\partial \bar{\psi}}{\partial \bar{x}} \frac{\partial^2 \bar{\psi}}{\partial \bar{y}^2} - A^{\alpha_6 - \alpha_1} B^{\beta_6} \frac{\partial \bar{U}}{\partial \bar{t}} - A^{2\alpha_6} B^{2\beta_6 - \beta_1} \bar{U} \frac{\partial \bar{U}}{\partial \bar{x}} \\ & = n \left[A^{\alpha_4 - 2\alpha_3} B^{\beta_4 - 2\beta_3} \frac{\partial^2 \bar{\psi}}{\partial \bar{y}^2} \right]^{n-1} - A^{\alpha_4 - 3\alpha_3} B^{\beta_4 - 3\beta_3} \frac{\partial^3 \bar{\psi}}{\partial \bar{y}^3} \end{aligned} \quad (14)$$

$$\begin{aligned} & A^{\alpha_5 - \alpha_1} B^{\beta_5} \frac{\partial \bar{T}}{\partial \bar{t}} + A^{\alpha_4 + \alpha_5 - \alpha_3} B^{\beta_4 + \beta_5 - \beta_3 - \beta_1} \frac{\partial \bar{\psi}}{\partial \bar{y}} \frac{\partial \bar{T}}{\partial \bar{x}} \\ & - A^{\alpha_4 + \alpha_5 - \alpha_3} B^{\beta_4 + \beta_5 - \beta_1 - \beta_3} \frac{\partial \bar{\psi}}{\partial \bar{x}} \frac{\partial \bar{T}}{\partial \bar{y}} = \frac{1}{Pr} A^{\alpha_5 - 2\alpha_3} B^{\beta_5 - 2\beta_3} \frac{\partial^2 \bar{T}}{\partial \bar{y}^2} \end{aligned} \quad (15)$$

Subject to the boundary conditions:

$$\frac{\partial \psi}{\partial y}(x, 0, t) = s(x, t), -\frac{\partial \psi}{\partial x}(x, 0, t) = v(x, t), T(x, 0, t) = 1, \frac{\partial \psi}{\partial y}(x, \infty, t) = U(x, t), T(x, \infty, t) = 0 \quad (16)$$

Note that the basic equations remain invariant under the group G_3 of transformation if the powers of A and B in each term should be equal. Thus invariance of above equations under G_3 gives following relations among α_i 's and β_i 's:

$$\begin{aligned} \alpha_4 - \alpha_1 - \alpha_3 &= 2\alpha_4 - 2\alpha_3 = \alpha_6 - \alpha_1 = 2\alpha_6 = n\alpha_4 - 2n\alpha_3 - \alpha_3; \quad (17) \\ \alpha_5 - \alpha_1 &= \alpha_4 + \alpha_5 - \alpha_3 = \alpha_5 - 2\alpha_3; \\ \beta_4 - \beta_3 &= 2\beta_4 - 2\beta_3 - \beta_1 = \beta_6 = 2\beta_6 - \beta_1 = n\beta_4 - 2n\beta_3 - \beta_3; \\ \beta_5 &= \beta_4 + \beta_5 - \beta_3 - \beta_1 = \beta_5 - 2\beta_3; \quad (18) \end{aligned}$$

Solving equations (14) and (15) for α_i 's and β_i 's, we obtain the relations among these α_i 's and β_i 's provided that the basic equations are to be invariant under this group transformation for some arbitrary choice of P_1 and P_2

$$\begin{aligned} \frac{\alpha_3}{\alpha_1} &= \frac{2-n}{1+n}, \quad \frac{\alpha_4}{\alpha_1} = \frac{1-2n}{1+n}, \quad \frac{\alpha_5}{\alpha_1} = P_1, \quad \frac{\alpha_6}{\alpha_1} = -1, & (19) \\ \frac{\beta_3}{\beta_1} &= \frac{n-1}{n+1}, \quad \frac{\beta_4}{\beta_1} = \frac{2n}{1+n}, \quad \frac{\beta_5}{\beta_1} = P_2, \quad \frac{\beta_6}{\beta_1} = 1, \end{aligned} \quad (20)$$

It is to be noted that the proper form of P_1 and P_2 can be derived from the similarity requirement of equations (14)-(15).

4.3 The Absolute Invariants

The next step in this method is to find the so-called "absolute invariants" under the considered two parameter group of transformation. Absolute invariants are functions having the same form before and after the transformation. The absolute invariants [T. Y. Na] are:

$$\begin{aligned} \eta &= \frac{y}{t^{\frac{2-n}{1+n}} x^{\frac{n-1}{n+1}}}, \quad F_1(\eta) = \frac{\psi}{t^{\frac{1-2n}{1+n}} x^{\frac{2n}{1+n}}} \\ F_2(\eta) &= \frac{T}{t^{P_1} x^{P_2}}, \quad c = \frac{U}{t^{-1} x^1} \end{aligned}$$

Similarity solution exists if the main stream velocity is given by

$$U = c \frac{x}{t} \quad (21)$$

4.4 The Reduction to an ordinary differential equation

Substituting these in equations (10) and (11) as well as in boundary conditions (12) we obtain a set of ordinary differential equations as:

$$n \left| F_1 \right|^{n-1} F_1''' - \left(\frac{n-2}{n+1} \right) \eta F_1'' - (F_1' - 1) F_1' + \left(\frac{2n}{n+1} \right) F_1 F_1'' - c(1 - c) = 0 \quad (22)$$

$$\left(\frac{n-2}{n+1}\right)\eta F_2' - \left(\frac{2n}{n+1}\right)F_1F_2' + F_2(P_1 + P_2F_1') - F_1F_2' - \frac{1}{Pr}F_2'' = 0(23)$$

Now for Newtonian case, that is, for $n=1$, the equation (22) and (23) will be:

$$F_1''' + \left(\frac{1}{2}\right)\eta F_1'' - (F_1' - 1)F_1' + F_1F_1'' - c(1 - c) = 0(24)$$

$$\left(\frac{-1}{2}\right)\eta F_2' - F_1F_2' + F_2(P_1 + P_2F_1') - F_1F_2' - \frac{1}{Pr}F_2'' = 0(25)$$

Boundary conditions are transformed as

$$F_1(0) = 0, F_1'(0) = 0, F_2(0) = 1, F_1'(\infty) = 1, F_2(\infty) = 0(26)$$

System of equations (22)-(23) are system non-linear ordinary differential equations with boundary conditions (26). The Numerical solution of this system can be found by some suitable method.

5. Numerical solution

The partial differential system is transformed to an ordinary differential system via two parameter group transformation method are solved numerically by 4th order Runge-Kutta method. Boundary conditions are selected in a general form and the structures of functions appearing in the boundary conditions, for which similarity transformations are available, are determined. The resulting ordinary differential system is solved numerically.

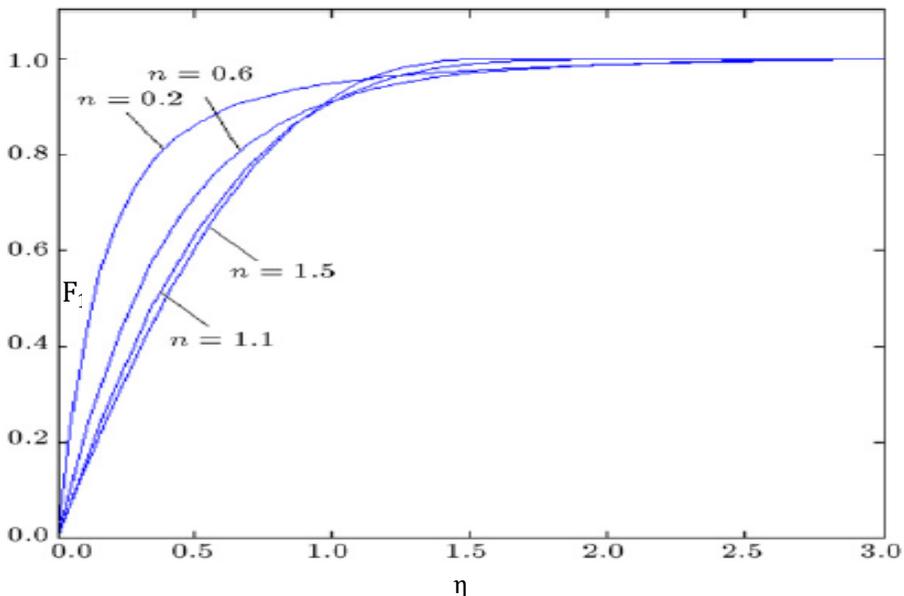


Figure-1: Velocity Variation for different values of flow behavior index n.

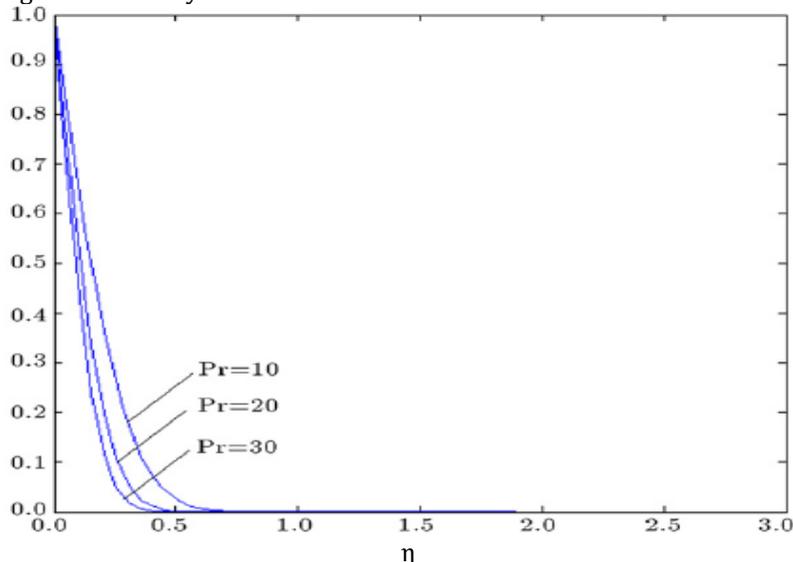


Figure-2: Variation in temperature for different values of flow behavior index Prandtl Number.

From Figure -1, it is observed that the effects of the flow behavior index, squeezing parameter, and Prandtl number on the velocity and temperature profiles are analyzed. Velocity profiles are qualitatively different for flow behavior indexes less than 1 and greater than 1. Increasing surface velocity parameters and squeezing parameters increases the function related to the x component of velocity. Finally, from Figure-2, it is clear that an increase in Prandtl number decreases the temperature profiles.

6. Conclusion

It is concluded that the application of multi parameter group theoretic method described in the beginning, the set of governing equations and the boundary conditions are reduced to highly nonlinear ordinary differential equations with appropriate boundary conditions in a single step. The outcome of this study is that, all possible conditions under which the similarity solutions for the present flow situation exists are automatically derived from the similarity requirements and thus the similarity solution is found in most general form. Further this similarity equation is solved numerically.

References:

1. Ali, M.E. "On thermal boundary layer on a power-law stretched surface with suction or injection", *Int. J. Heat Fluid Flow*, 16, pp. 280–290 (1995).
2. Chamkha, A.J. "Similarity solution for thermal boundary layer on a stretched surface of a non-Newtonian fluid", *Int. Commun. Heat Mass Transfer*, 24, pp. 643–652 (1997)
3. Yürüsoy, M. and Pakdemirli, M. "Symmetry reduction of unsteady three- dimensional boundary layers of some non-Newtonian fluids", *Internat. J. Engrg. Sci.*, 35, pp. 731–740 (1997).
4. Tashtoush, B., Kodah, Z. and Al-Ghasem, A. "On thermal boundary layer of a non-Newtonian fluid on a power-law stretched surface of variable temperature with suction or injection", *Heat Mass Transfer*, 37, pp. 459–465 (2001).
5. Khaled, A.-R.A. and Vafai, K. "Hydromagnetic squeezed flow and heat transfer over a sensor surface", *Internat. J. Engrg. Sci.*, 42, pp. 509–519 (2004).
6. Aydın, O. and Kaya, A. "Laminar boundary layer flow over a horizontal permeable flat plate", *Appl. Math. Computation*, 161, pp. 229–240 (2005).
7. Yürüsoy, M. "Unsteady boundary layer flow of power-law fluid on stretching sheet surface", *Internat. J. Engrg. Sci.*, 44, pp. 325–332 (2006).
8. Mahmood, M., Asghar, S. and Hossain, M.A. "Squeezed flow and heat transfer over a porous surface for viscous fluid", *Heat Mass Transfer*, 44, pp. 165–173 (2007).
9. Hansen, A. G. (1965): *Similarity analyses of boundary value problems in engineering*, Englewood Cliffs, Prentice Hall.
10. Akgül M.B., Pakdemirli M. (B 2012): Lie group analysis of a Non-Newtonian fluid over a porous surface, *Scientia Iranica J.vol. 19 issue-6*, pp1534–1540.

VIRTUAL CLOTHING TRY-ON USING AUGMENTED REALITY

¹Neha Sharma, ²Nihal Shetty, ³Shreyash Shetty, ⁴Riken Shah, ⁵Sonali Vaidya

¹Student, ²Student, ³Student, ⁴Student, ⁵Assistant Professor

Department of Information Technology,
St. Francis Institute of Technology, Mumbai, India.

ABSTRACT: *The ever-increasing use of internet has now made everything online. One of the industries to gain a lot of advancement is the e-commerce industry. However, major disadvantage lies in the logistics concerned with the return and exchange of products. One of the reasons for such inconvenience is inappropriate visual representation of clothes on one's body. According to the statistics, about 25%-30% of the total orders in a month are returned. Also, various known e-commerce sites fail to provide solution for the same. Due to this, Virtual try-on of clothes has received much attention in the recent times. The project focuses on Augmented Reality for clothing trial to offer real life experience to the end users. The user can virtually try clothes of their desired choices using the phone camera. This would probably reduce costs and the process of trying several clothes will be easier and faster for the consumers.*

Keywords: *augmented reality, online shopping, e-commerce, virtual, clothing, try-on, extended tracking.*

I. INTRODUCTION

Augmented Reality can be defined as the mixture of Digital information with user's environment and real time. In Augmented Reality, the new information is overlaid on the top of the existing environment. This gives the user and experience to live in real world in a more creative and innovative way. It provides 3D animated models of clothes and enables the user to view and manipulate desired clothes without having the pain of trying on for several times. Virtual try-on applications have gained a lot of popularity in recent times because they provide the ease with which a user can try various clothes without the effort of changing them physically. This helps to decide quickly and thereby improving the sales efficiency of retailers. Previous solutions included 3D reconstruction or modelling, motion capture, which is tedious and not robust for all body poses. This paper makes use of a method that evades the steps by integrating image-based renderings of the user and previously recorded garments. It transfers the appearance of a garment recorded from one user to another by matching input and recorded frames, image-based visual hull rendering, and online registration methods. Using images of real garments allows for a realistic rendering quality with high performance. This paper renders the cloth that was previously recorded on some other user which means that the cloth is not displayed on the actual user's body [1].

This paper presents a mixed reality system for 3D virtual clothes try-on that enables a user to see themselves wearing virtual clothes while looking at a mirror display without taking off their actual clothes. The user can choose virtual clothes for trying-on. The system physically simulates the selected virtual clothes on user and the user can see the virtual clothes them from various angles as they move. The major contribution of this paper is that they automatically customize an invisible (or partially visible) avatar based on the user's body size and the skin color. It does not focus on how user's image is captured and then clothing models can be rendered over them [2].

The goal of this paper is to study the possibilities of displaying objects that are larger than screen on mobile devices with real-time marker less tracking. Different approaches to marker less object and symbol identification will be investigated regarding their portability to mobile devices with sufficient performance. Depending on how promising these methods it will be implemented for further research. By improving and combining the methods, a marker less tracking technique for mobile devices will be developed. The main concern will be the performance of this technique and the visualization of larger-than-screen objects while the user moves around them [3].

II. PROPOSED SOLUTION

Considering the above-mentioned problems faced in online shopping and its consequences, augmented reality proves to be the best solution. Using this technology, the user can virtually try-on clothes on his/her body. This includes choosing the desired cloth and trying them virtually before buying. With regards to e-commerce it will reduce cost and increase sales also it will be easier, faster and efficient for the customers.

This involves gathering the clothes models. Blender v2.79 is an opensource software which provides extensive functionalities in the field of modelling, animation, simulations, composting etc. Once the model is developed, it is imported to Unity. In Unity a target image is set. Target image is to be held for the cloth to be overlaid on one's body. The user can now select various models one by one for trial. Extended tracking technology is used to retain the cloth on the body even if the target image is no longer available. This concept is explained in the section IV. Vuforia database is used to save the image target and the clothes database as well and hence it involves inculcating Vuforia plugin in Unity.

III. SYSTEM FLOW DIAGRAM

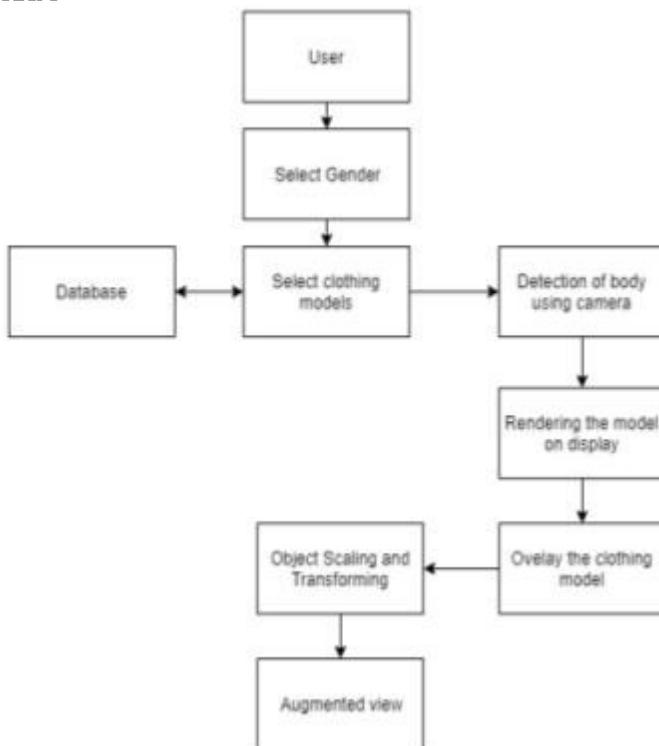


Fig. 1: System diagram

- Step1:**The user opens the application.
- Step2:**User selects the gender i.e. male or female followed by the selection of clothing model which will be directly retrieved from the database.
- Step3:**The clothing models are created using Blender and later imported into Unity3D and the target image is finalized using Vuforia. This is then later stored in the database.
- Step4:**Now the user gets himself detected using camera and clothing model gets rendered on the user's body.
- Step5:**Detector mainly performs two tasks-tracking and rendering which provides the body structured view to the overlay engine.
- Step6:**The overlay engine overlays the 3D model of cloth on the user's body.
- Step7:**The final augmented view is then provided to the user.

IV. EXTENDED TRACKING

Extended Tracking can be defined as a concept in which a target's pose information will be available even when the Target is no longer in the field of view of the camera and also when it cannot directly be tracked for other reasons. Extended Tracking makes use of a Device Tracker to enhance tracking performance and sustain tracking even when the target is no longer in view. Before Vuforia Engine 7.2, Extended Tracking was enabled on a per-target basis using the appropriate APIs. Starting with Vuforia Engine 7.2, Extended Tracking is automatically permitted for all targets when the Positional Device Tracker is enabled. Extended Tracking can intensify or enhance two kinds of user experience i.e. Game-like experiences with a large amount of dynamic content that requires the user to point the device away from the target as the user

follows the content and Visualizations of large objects like furniture, appliances, large home furnishings and even architectural models at the right scale and perspective.

The following figure depicts the impact of Device Tracking on an example Image Targets:



Fig. 1: Object on target image

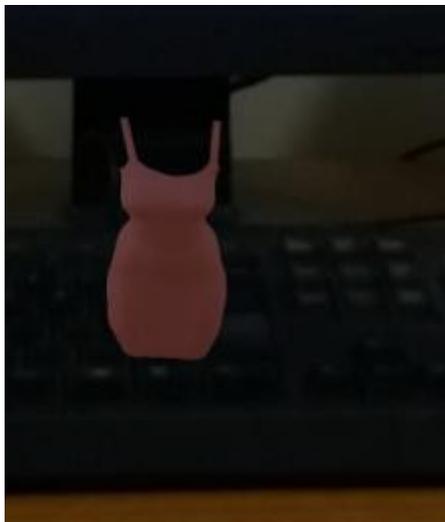


Fig. 2: Extended tracking technology used to retain the cloth even if the target image is no longer available.

V. SURVEY AND STATISTICS

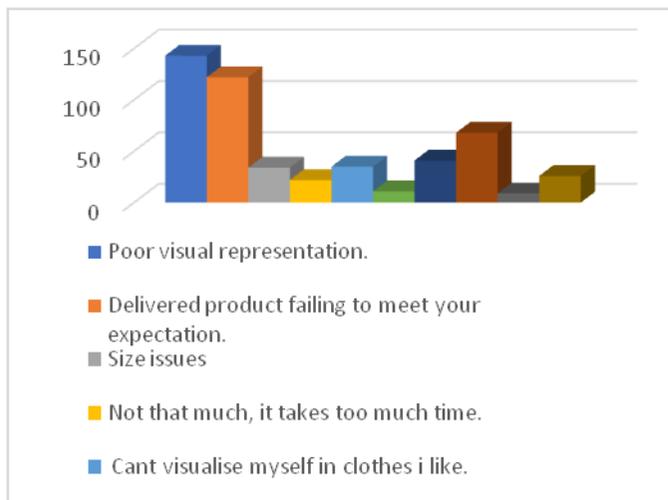


Fig. 1: Issues faced in online shopping

Following figure shows the statistics of various issues faced by the shoppers. Response was collected from 581 people. Maximum of them choose poor visualization to be one of the issues in online shopping.

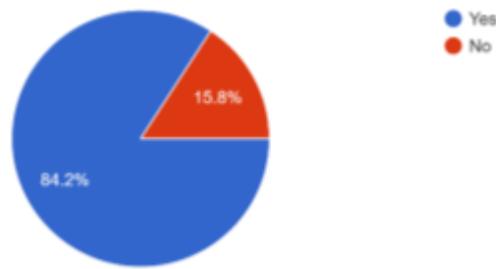


Fig. 2: Piechart

The survey conducted regarding the interest of the people in virtual clothing try-on was found to be 84.2%.

VI. ACKNOWLEDGEMENT

We are highly grateful to our institute St. Francis Institute of Technology and the Information Technology Department for providing us with this learning opportunity with the required resources to accomplish our task. We are highly indebted to our college for providing us with the infrastructure to work and also the SFIT Library for providing us with various books, newsletters, papers and project books. We wish to express our deep gratitude to our project guide Prof. Sonali Vaidya, Assistant Professor in the Department of Information Technology, for all the advice, encouragement and constant supervision for the betterment of the project. Also, our effort would have been incomplete without our Project coordinator Prof. Amrita Mathur.

VII. CONCLUSION

Today's trend is online shopping and with many e-commerce websites providing with all we want, people have switched to purchase all the products online. However, we wish we could try on some things such as clothes before purchasing them and the best combination is to buy the clothes online and get a chance to have them on the body before purchasing them. Virtual try-on applications have become popular because they allow users to watch themselves wearing different clothes without the effort of changing them physically. This helps users to make quick buying decisions and, thus, improves the sales efficiency of retailers. Considering this, we aim to build an application integrating Augmented Reality to make this combination a reality. The application is designed for Android smart phones and will be implemented in Unity in integration with Vuforia. The application will help users to select the models of clothes that they wish to try on themselves and get the real-time look by getting the clothes rendered on themselves.

REFERENCES

1. Stefan Hauswiesner, Matthias Straka, Gerhard Reitmayr, "Virtual Try-On through ImageBased Rendering", Visualization and Computer Graphics IEEE Transactions on, vol. 19, no. 9, pp. 1552 - 1565, 2013.
2. Miaolong Yuan, Ishtiaq Rasool Khan, Farzam Farbiz, Susu Yao, Arthur Niswar, Min-Hui Foo, "A Mixed Reality Virtual Clothes Try-On System", Multimedia IEEE Transactions on, vol. 15, no. 8, pp. 1958-1968, 2013.
3. Elisa Ziegler, "Real-time marker less tracking of objects on mobile devices", Computer Visualistics, 2013.
4. Unity User Manual (2018.3), <https://docs.unity3d.com/Manual/index.html>
5. Features of blender, <https://www.blender.org/>
6. Developers page, <https://www.vuforia.com/content/vuforia/en/augmented-reality-app-development-developer.html>

Implementation of Embedding Algorithm for Data Hiding using Steganography Technique

¹Dinaz R. Adajania & ²Prof. Sanjay Vidhani

¹Mtech in Information Security, ²Associate Professor

¹Department of Information Technology, ²Department of Information Technology

¹K J Somaiya College of Engineering, Mumbai, India, ²K J Somaiya College of Engineering, Mumbai, India

ABSTRACT: Steganography is the art and science of hiding information that incorporates hidden content in cover media so that suspicious activities are not provoked by an eavesdropper. Steganography is different from cryptography in that cryptography focuses on keeping data secret while steganography focuses on keeping data secret. The approach is to implement a new steganography embedding technique that is simple and will be used for all media file formats such as image, audio and video. This embedding algorithm embeds the master image file size, type of encryption algorithm (DES or AES), compression ratio of secret file size, data of secret file size which is also encrypted using a secret key. The Deflate compression technique is also used to compress secret file into the cover file. The proposed system maintains imperceptibility and data integrity property of steganography. The steganography tool is implemented and integrates types of steganography technique such as image, audio and video in one tool. The proposed method satisfies the reversibility property, imperceptibility, hidden capacity of data into cover image of steganography to reconstruct secret and cover images without any loss of information.

Keywords: Steganography, Cryptography, Deflate compression technique, cover file, stego file.

I. Introduction

The word Steganography is Greek and stands for covered writing. Steganography is the art of hiding data secretly in any cover medium such as image, audio and video. Steganography's main role is to conceal the information transmitted in some digitally covered media so that it does not provoke suspicious activity by eavesdroppers. Steganography is the art of hiding information that cannot be identified by its presence. Secret information is encoded in a way that conceals the existence of secret information. Steganography is a powerful security tool that offers a high level of security when combined with encryption [1].

Image steganography is defined by hiding secret data (image, text files) by taking cover object or media as an image. The image file acts here as a cover medium for hiding secret messages. The image selected for hiding is called the cover image, and the image obtained after steganography is called the stego image. The cover medium is an image of any extension (.jpeg, .gif, .png), the secret medium can be secret image, secret file, secret message.

Audio Steganography is defined by hiding secret data (image, text files, audio) by taking object or media as audio. The audio file acts here as a cover medium for hiding secret file or messages. The audio is selected for hiding is called cover audio and the audio media obtained after steganography is called the stego audio. The cover medium is an audio of any extension (.mp3, .mp4, .), the secret medium can be secret image, secret file, secret message, secret audio.

Video steganography is defined by hiding secret data (image, text files, audio, video) by taking cover object or media as a video. The video file acts here as a cover medium for hiding secret messages. The video selected for hiding is called the cover video, and the video obtained after steganography is called the stego video. The cover medium is a video of any extension (.avi), the secret medium can be secret image, secret file, secret message, secret audio and secret video.

This paper comprehends the following objectives:

- i. The embedding capacity of secret file can be same as size of cover image file but it is not less than cover image file.
- ii. Imperceptibility is maintained. The Deflate compression technique is used to compress secret file into the cover image. So the secret or stego file which is embedded in cover or master file is secured and maintains imperceptibility property of steganography.
- iii. Quality of file after embedding a secret file is not changed and also time required to embed files into cover image is very less

- iv. Data integrity is maintained and having less chance to decrypt secret file from cover or master file
- v. It should be encrypted using secret key is maintained so that there is less chance of obtaining secret file which is embedded in cover file.
- vi. The imperceptibility, embedding capacity, robustness, security and tamper resistance are the properties is achieved by implementing image steganography.
- vii. The image extension such as png, jpeg, gif, bmp files is embedded in to cover image.

II.Literature Survey

In 2017, the paper An overview and computer forensic challenges in image steganography, discusses the security flaws in steganography tools and give comparison for different steganography tools which is currently used and describes about the forensics challenges which means the steganography tool are not secured enough to hide secret file in cover file easily detected by steganalysis [1]. In 2017, the paper Enhancing the imperceptibility of image steganography for information hiding,[2] discusses about the main property of steganography which is imperceptibility which means in- visibility of steganographic algorithm because it is first and foremost requirement, since the strength of steganography lies in ability to be unnoticed by the human eye. In 2017, the paper Steganography methods on text, audio, video: A survey, which describes about the different methodologies for steganography and also give overview about each steganography techniques and relate to these methodologies are discussed [3]. In 2017,the paper Evaluation of best steganography tool using image features, [4] discusses about different new methodologies used for image steganography and also different tools for image steganography and give comparisons, tools merits and demerits are also described. In 2017, the paper A survey on image steganography[11] discusses about survey of papers related to image steganography techniques where all comparison techniques of image steganography with its merits and demerits are described. In this paper, there is literature survey of all image steganography technique is explained which comes under two main categories spatial domain steganography and partial domain steganography which consists of different embedding techniques. In 2016, the paper deflate compression algorithm, describes the step by step implementation of deflate compression algorithm which is lossless compression technique that uses a combination of LZ77 algorithm and Huffman coding [5].In 2015, the paper Steganography based information security with high embedding capacity discusses about how the two secret images can be embedded into single cover image without any loss in quality of cover image. Here, five parameters of steganography is achieved quality, embedding capacity, histogram analysis, computational cost and speed [14].In 2008, the paper Improvable Deflate algorithm, [10] describes theoretical implementation of Deflate compression technique and also how to implement LZ77 and Huffman encoding in deflate compression technique is discussed.In 2015, the paper Secure Data Transmission Using Video Steganography describes transmission of secure data within video using video steganography. In 2015, the paperAudio-Video steganography describes to hide text in both audio and video together using audio-video steganography. In 2016, the paperA Novel Technique for Embedding Audio inAudio to Ensure Secrecy describes about to embed different audio file types in any audio file using audio steganography. In 2015, the paperAn Imperceptible and Robust Audio Steganography Employing Bit Modification describes about the audio steganography enhanced the properties of steganography such as imperceptibility and robustness.

III. methodology

A. IMAGE STEGANOGRAPHY:

Image Steganography is defined as the hiding secret data (image, text files) by taking cover object or media as image.The image in which secret data is to be hided is called the cover image and the image obtained after steganography is called the stego image. The cover medium is image of any extension (.jpeg,.png,.gif,.bmp),the secret medium can be secret image, secret file, secret message[6][8].The below figure 1 shows the general block diagram of steganography which shows the cover image is embedded with a secret file or image combined with secret key to enhanced the security and gives output as embedded file or message, this is done by embedding process and retrieving process will take stego image as inputimage and gives output as retrieving message from stego or secret file. In this paper image steganography is enhanced in terms of embedding capacity, imperceptibility, reversible, without any loss of information, data integrity is maintained.

B. AUDIO STEGANOGRAPHY:

It is a technique used to transmit a hidden message within audio file. Embedding message in an audio file is much more difficult than hiding message in an image file.

C. VIDEO STEGANOGRAPHY:

A video file is used as a cover file to hide the secret message. It is less prone for steganalysis as a video file is a combination of text, image and audio. So it is difficult for an attacker to identify that which frame of video is containing the secret message.

D. ADDITION METHOD FOR IMAGE, VIDEO AND AUDIO STEGANOGRAPHY:

There are two types of images used in addition method, first sink file which is the file to be hidden in another image and secondly container file where the sink file is embedded[9][12][13]. The advantage of this method is solely due to its simplicity. Here, there is a partition of both container and sink files into two parts namely, structural part and data part. The structural part consists of header information and coloring palette, etc. while data part consists of actual image pixel values. Therefore, there are four different parts for container and sink images.

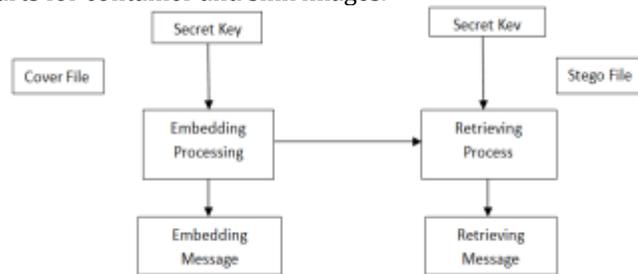


Fig.1 General block diagram of steganography

- StructuralPartofContainerImage
- DataPartofContainerImage
- StructuralpartofSinkImages
- DatapartofSinkImages

E. DEFLATE COMPRESSION TECHNIQUE FOR IMAGE, AUDIO AND VIDEO STEGANOGRAPHY:

Deflate is the new and efficient algorithm developed by Phillip W. Katz which uses both LZ77 and Huffman coding for compressing the data [10]. It was mainly used for .zip file format but Deflate are used for different file formats or softwares like GZIP, 7-zip etc. The algorithm for Deflate Compression is given below-

Step 1: The whole input stream is divided into series of blocks.

Step 2: LZ77 is implemented to find the strings repeated in each block and reference pointers are inserted with in the block

Step 3: Symbols are then replaced with the code-words based on their occurrences in the block using Huffman Encoding process

F. ALGORITHM FOR ENCRYPTION AND COMPRESSION IN IMAGE STEGANOGRAPHY:

Algorithm: Encryption and Compression in Image, Audio and Video Steganography

Input: Cover or Master file

Output: encrypted and compressed Stego file

Step 1: Embed header of cover file as it is into output stego image.

Step 2: Embed size of cover file in binary.

Step 3: Embed data of cover file in binary form.

Step 4: Embed version of algorithm used for encryption.

Step 5: Embed features that contain information about data

file compression and decryption. For example, if it is CEF then it is compressed file.

Step 6: Calculation of compression ratio and embed it into output stego image $CompressionRatio = (DataFilesize\ after\ compression / OriginalDataFilesize) * 100$ Embed the compression data into output stego image.

Step 7: Embed secret file size in binary format.

Step 8: Embed output stego file by tarray content to file i.e. master/cover file.

IV. experimental results

The experimental results show the comparative study of Image, Audio and Video steganography techniques are used previously and how in this project the disadvantages of all steganography techniques are overcome, comparative study of compression file size and time required for embedding and retrieving data file in master file for image, audio and video steganography are examined.

Type of steganography	Master Image File size	Data File size	Actual Steganography File size	Compressed File size	Time required for embedding file into master image file (in seconds)	Time required for retrieving steganography file (inseconds)
Text File is embedded into image file	D5C0191E.jpg(2990KB)	samp.txt(113 bytes)	3103KB	3063 KB	0.02	0.03
Message on clipboard of application is embedded into image	Cafe.jpg(15500 KB)	Message written in message box	CafeOutput.jpg (16550 KB)	CafeOutput.jpg (15700 KB)	0.02	0.03
Image is embedded into image	Del.jpg(4 KB)	Clear.jpg(10 KB)	DelOutput.jpg(14 KB)	DelOutput.jpg (8.24KB)	0.03	0.04
Gif is embedded into Gif	Th.gif(151 KB)	200w.gif(93 KB)	Outputth.gif(244 KB)	Outputth.gif(243 KB)	0.03	0.04
.jpg image is embedded into .png image	3.png(3.14KB)	About.jpg(1.43KB)	Output3.png(4.58)	Output3.png(4.57KB)	0.03	0.04
.png image is embedded into .png	2.png(4.41 KB)	3.png(3.14 KB)	2Output123.png(7.55 KB)	2Output123(7.43KB)	0.03	0.04
.png image is embedded into .jpg file	Box.jpg(3.09 KB)	4.jpg(3.71 KB)	BoxOutput.jpg(6.94 KB)	BoxOutput.jpg(6.80 KB)	0.03	0.04
.bmp image embedded into .bmp image	About.bmp(1.43 kB)	4.bmp(3 kB)	AboutOut.bmp(5.43 kB)	AboutOut.bmp(5 kB)	0.03	0.04

Table 1 Comparison of different extensions of image file of proposed image steganography tool

A. SCREENSHOTS OF IMAGE STEGANOGRAPHY TOOL:

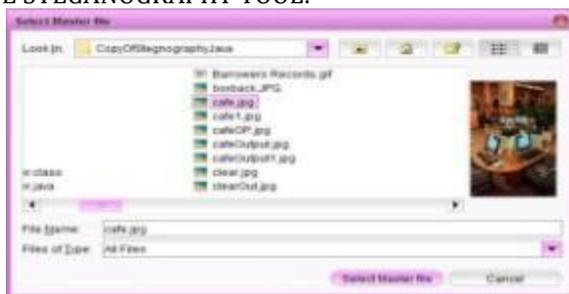


Fig. 2. Select master file

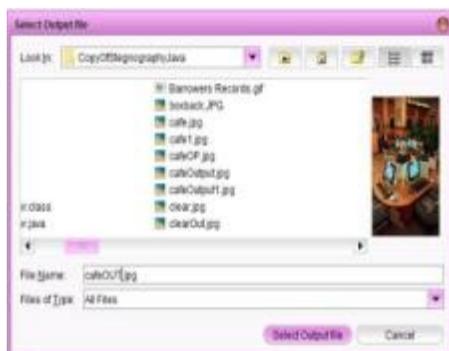


Fig. 3. Select output file as cafeOUT.jpeg

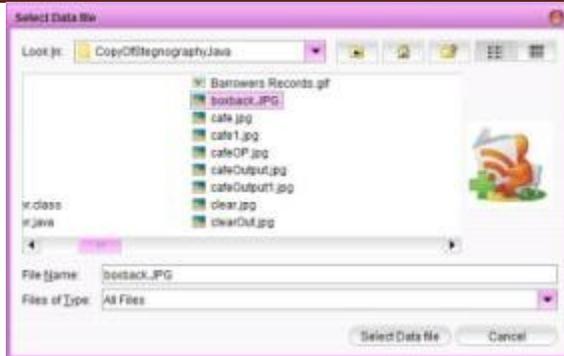


Fig. 4. Select Data file which to be embedded in cafeOUT.jpeg

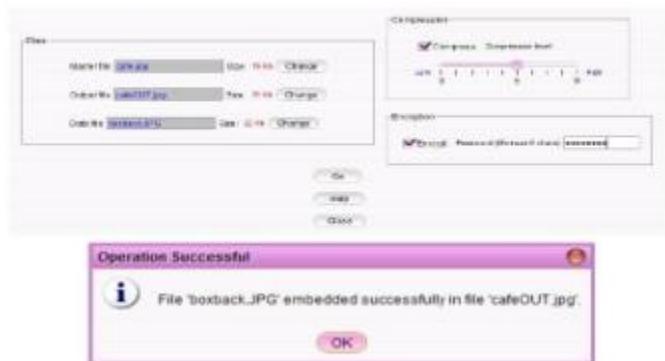


Fig. 5. Image is embedded in to the cafeOUT.jpg and successfully embedded

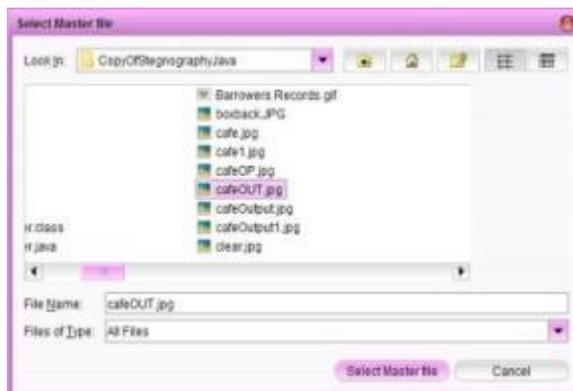


Fig. 6. Select Master file for retrieving secret image into cafeOUT.jpg



Fig. 7. Details of embedded file is retrieved



Fig. 8.password for embedded file



Fig.9 Secret image is retrieved

Type of steganography	Master Audio File size	Data File size	Actual Steganography File size	Compressed File size	Time required for embedding file into master audio file (in seconds)	Time required for retrieving steganography file (in seconds)
Image file is embed in video file	clearvobtool .avi(144KB)	Boxback.jpg(22 KB)	166KB	166 KB	0.02	0.03
Video file embed in video file	Profileviz1.a vi(255 KB)	Clervobtool.avi(144 KB)	Profileviz1OUT PUT1.avi(399 KB)	299 KB	0.02	0.03
Audio file embed in video file	Profileviz1.a vi(255 KB)	Hard_TON_-_07_-_Choc-ice_Dance.mp3(339 KB)	Profileviz2.avi(5 64 KB)	Profileviz2.av i(525 KB)	0.03	0.04
Image file embed in audio file	Hard_TON_-_07_-_Choc-ice_Dance. mp3(339	Café.jpg(15 KB)	Hard_TON_-_07_-_Choc-ice_DanceOut. mp3(354 KB)	Hard_TON_-_07_-_Choc-ice_DanceOuu t.mp3(354	0.03	0.04
Audio file embed in audio file	Hard_TON_-_07_-_Choc-ice_Dance. mp3(339 KB)	John_Wesley_C oleman_-_07_-_Tequila_10_Se conds.mp3(695 KB)	Hard_TON_-_07_-_Choc-ice_DanceOut1. mp3(1034 KB)	Hard_TON_-_07_-_Choc-ice_DanceOuu t1.mp3(1000 KB)	0.03	0.04
Video file embed in audio file	Hard_TON_-_07_-_Choc-ice_Dance. mp3(339 KB)	Clervobtool.avi(144 KB)	Hard_TON_-_07_-_Choc-ice_DanceOut2. mp3(483 KB)	Hard_TON_-_07_-_Choc-ice_DanceOuu t2.mp3(413 KB)	0.03	0.04

Table 3 Comparison table for audio and video steganography in a steganography tool

V. conclusion

In the proposed steganography tool, there are no lengthy separation for cover file and stego file. The Deflate compression algorithm is used for secret data is compressed to the cover file so the size of secret file is less. Therefore, there is less chance of identify the secret data in steganography tool. The embedding capacity of secret file can be same as size of cover image and the quality of file after embedding a

secret file is not changed and less time is required to embed secret file in to cover file. Data integrity is maintained and having less chance to decrypt secret file from cover image or master image file. There is no restriction of image, audio and video extension that means png, jpeg, gif, avi, mp3, mp4, and bmp files can be embedded in to the cover file and extension of file can be taken as cover file or master file. The image stego file is encrypted using secret key so that there is less chance of obtaining secret file from cover file. The imperceptibility, embedding capacity, robustness, security and tamper resistance are the main properties of steganography are achieved. The future scope of this steganography tool will be extracting audio from a video (containing audio) and later mixing both video and audio while decoding and retrieving the complete video.

References

1. Yari, Imrana Abdullahi, and Shahrzad Zargari. "An overview and computer forensic challenges in image steganography." *Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, 2017 IEEE International Conference on. IEEE, 2017.
2. Fouad, Mohamed M. "Enhancing the imperceptibility of image steganography for information hiding." *Computer Science and Information Systems (FedCSIS)*, 2017 Federated Conference on. IEEE, 2017.
3. Febryan, Aryfandy, Tito Waluyo Purboyo, and Randy Erfa Saputra. "Steganography Method on Text, Audio, Image and Video: A Survey." *International Journal of Applied Engineering Research* 12.21 (2017): 10485-10490.
4. Jensi Lakdawala, Jiya Rankawat "Evaluation of Best Steganography Tool using Image Features". *International Journal of Computer Trends and Technology (IJCTT)* V50(1):50-58, August 2017. ISSN:2231-2803. www.ijcttjournal.org. Published by Seventh Sense Research Group.
5. Oswal, Savan, Anjali Singh, and Kirthi Kumari. "Deflate Compression Algorithm." *International Journal of Engineering Research and General Science* 4.1 (2016): 430-436.
6. Sharma, Vipul. "Two new approaches for image steganography using cryptography." *Image Information Processing (ICIIP)*, 2015 Third International Conference on. IEEE, 2015.
7. Sirisha, B. Lakshmi, S. Srinivas Kumar, and B. Chandra Mohan. "Steganography based information security with high embedding capacity." *Recent Advances in Electronics & Computer Engineering (RAECE)*, 2015 National Conference on. IEEE, 2015.
8. Abdulzahra, Hayfaa, R. O. B. I. A. H. Ahmad, and Norliza Mohd Noor. "Combining cryptography and steganography for data hiding in images." *ACACOS, Applied Computational Science* (2014): 978-960.
9. Kour, Jasleen, and Deepankar Verma. "Steganography techniques—A review paper." *International Journal of Emerging Research in Management & Technology* ISSN (2014): 2278-9359.
10. Weimin, Wu, et al. "Improvable deflate algorithm." *Industrial Electronics and Applications, 2008. ICIEA 2008. 3rd IEEE Conference on. IEEE* 2008.
11. Menon, Nidhi. "A survey on image steganography." *2017 International Conference on Technological Advancements in Power and Energy (TAP Energy)*. IEEE, 2017.
12. Morkel, Tayana, Jan HP Eloff, and Martin S. Olivier. "An overview of image steganography." *ISSA*. 2005.
13. Abdulzahra, Hayfaa, R. O. B. I. A. H. Ahmad, and Norliza Mohd Noor. "Combining cryptography and steganography for data hiding in images." *ACACOS, Applied Computational Science* (2014): 978-960.
14. Sirisha, B. Lakshmi, S. Srinivas Kumar, and B. Chandra Mohan. "Steganography based information security with high embedding capacity." *Recent Advances in Electronics & Computer Engineering (RAECE)*, 2015 National Conference on. IEEE, 2015.
15. Gopalan, Kaliappan, and Jiajun Fu. "An imperceptible and robust audio steganography employing bit modification." *Industrial Technology (ICIT)*, 2015 IEEE International Conference on. IEEE, 2015.
16. Sahu, Chandan Kumar, and Prabira Kumar Sethy. "A novel technique for embedding audio in audio to ensure secrecy." *Communication and Signal Processing (ICCSP)*, 2016 International Conference on. IEEE, 2016.
17. Balaji, R., and Garewal Naveen. "Secure data transmission using Steganography." *Electro/Information Technology (EIT)*, 2011 IEEE International Conference on. IEEE, 2011.

MANAGING ROAD JUNCTIONS BASED ON TRAFFIC FLOW ESTIMATIONS & ROAD CONDITIONS

Dr. Riyazoddin Siddiqui, Shubham Parab, Ali Electricwala, Mustakim Merchant

Associate Professor, Department of Information Technology
M. H. Saboo Siddik College of Engineering, Mumbai, India .

ABSTRACT: Traffic optimisation has evidently proven to becoming one of the emerging areas for development. With exponential increase in surveillance devices and data analysis tools, the technology behind the traffic control system has improved. This paper presents a solution to the changing traffic density at intersections. The pre-timed traffic signals fail to tackle such situations, hence resulting in improper utilisation of the junction. Our system is much more deterministic in approach, with vehicle actuated traffic signals. It is more reliable and accurate than the traditional method. All the data crunching is based on the macroscopic view of the traffic. It calculates the traffic density considering existing road conditions. Further, it uses logistic function to calculate the chances of traffic jam, which will help analyze the situations better. Based on above features, the traffic is phased at junction accordingly.

Keywords: road, traffic, junctions, analysis, statistical.

I. INTRODUCTION:

Traffic Congestion is a major problem in major cities in India. In the past few decades, there is a huge migration of people into the cities. As buying two and four wheelers became easy and affordable these days. Every day 700 new vehicles are registered in cities like Mumbai, according to a recent study (Sen, 2017). The chart below shows the changing demography of vehicle population in India.

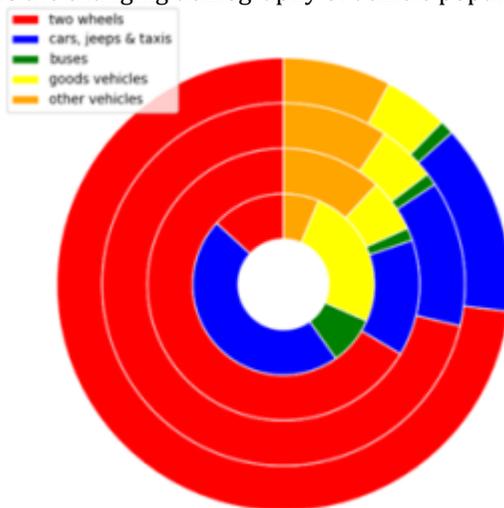


Fig 1: Changing vehicle population from 1961 to 2014 in India (In to Out)

As a result, due to lack of planning, cities face huge traffic overheads. This is an ever growing problem as it is estimated that 40% of India's population will be living in cities by 2030 (Sen S. , 2018). To tackle this, many have adopted idea of smart cities and tech parks where companies build their offices and residences annexed to the main land. While this helps accommodating to some extent. The road network grows and ramifies more.

Upon this, major cities are going through huge transformation with new building constructions and metro/monorail projects. Due to this, most of the roads are dug and narrowed. This brings major compromises in the organization and management. Other events affecting this can be anything causing road blockages like accidents, emergencies, potholes, etc. Also, an inevitable problem with old growing cities is they are not planned to operate on this scale. Hence, the road network in such area is surely to congest without proper guidance.

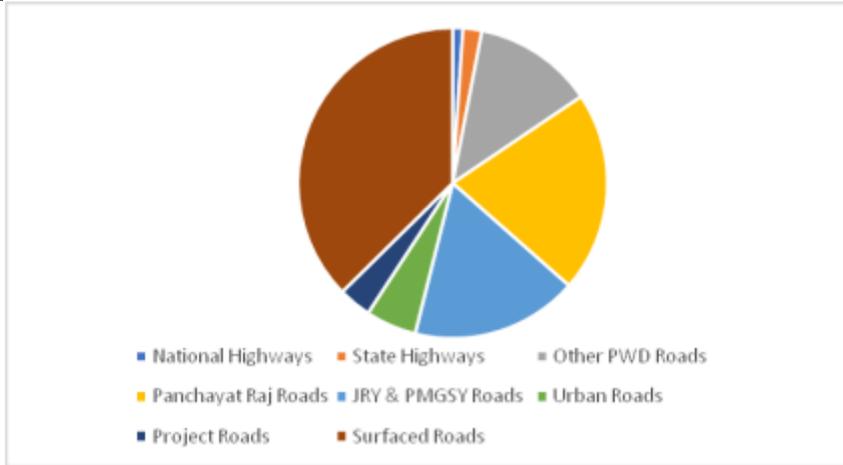


Fig 2: Total Road Percentage Share of Each Category of Road in 2014

The above diagram shows the percentage-wise road category in year 2014. In most cities, denser the networks makes it more difficult for officers and analyst to get insights from them. Also the work is divided into wards allocated different areas handled from control rooms. Hence, the analysis of different situations and there communication might prove to be a heavy task. Also, it might be impossible to sustain in critical situations where fast coordination and decision making is required.

The above requirements create the need to upgrade the current traffic control and surveillance system with augmented features by providing a centralized automated system(Sharma, 2012). Principle considerations of the undertaking are to include:

- Automated data collection and synchronisation from all the sources (CCTV, internet, etc). Data may contain direct surveillance data or JSON objects representing cars.
- Centralized data crunching and analysis.
- Visualization and interpretation of percept information (also simulation).
- Daily monitoring and storing the statistics.
- Accurate estimations and predictions of various events based on the data until then.
- Reinforced learning for better predictions.
- Intelligent suggestions for controlling and directing cars.

SYSTEM MODEL:

The system is based on an integrated, centralized system which directly depends on live CCTV footages from different landmark(Buxton & Gong, 1994). Then examining this data produce different verdicts, suggestions and warnings which the operator can look into and act. Further the system may even record any feedback on given suggestion to future use. This system can mainly be used in busy city’s road traffic control rooms.

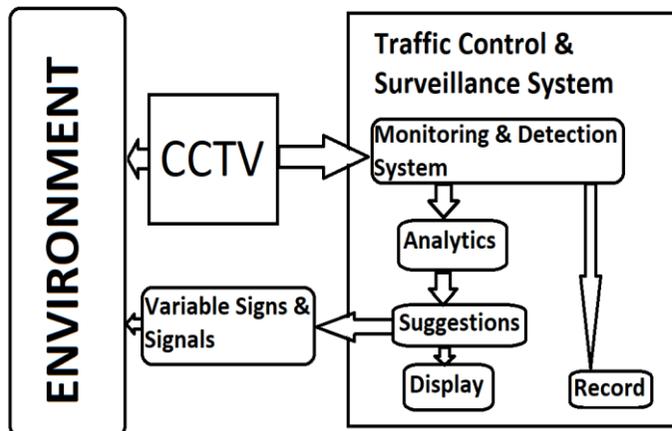


Fig 3: Architecture of Traffic Control & Surveillance System

1. Monitoring & Detection

The monitoring system captures CCTV footages and organizes them according to their area and connections with other nodes (junctions). Based on this, the system understands the traffic flow and density between junctions (nodes). Along with this, it also acts as a CCTV mirroring system to carryout analysis on each signal.

Detection system handles the vehicle detection and classification as well determining other characteristics like speed, model, etc. All this information is collected and organized according to relevance.

The sub-system also manages the feedback information related to the suggestion given by the system. The system can prompt here to look for any specific event to occur or check some conditions after applying any improvisation. This data will finally be recorded for future use.

2. Analytics

The sub-system handles the computation and decision making of the system. The data made avail by the above sub-system is only known here. This modular style of separating functionality makes future modification and improving easier. This can be subdivided into two parts: 1. Short term analysis 2. Long term analysis

The short term analysis manages the instantaneous movement of vehicles. This includes the live statistics of vehicles and their attributes. Based on these factors, the analytics dissects the conditions to extract some valuable insight on the current situation. This knowledge can be used to make quick decisions on the different scenarios like traffic congestion, accidents and emergency cases.

The long term analysis is based on the accumulated data till that time. This analysis uses the quantitative analysis and pattern tracking in the huge statistical data collected to tell about the usage and traffic behaviour. Long term decisions can be made to modify the road networks to allow better traffic flow while reducing the people’s waiting time in traffic.

3. Suggestion & Feedback

Being an automated control system, the system still retains the power to the people in-charge. The system suggests the optimisations and decisions to the user. In response the user handles the allowing/denying of the actions. This makes sure the user is aware of activities prior their execution.

The performed actions may/may not cause some effects on the statistical data from there on. Hence, to capture such results, the system monitor the environment for checking the effects caused. These results can then be analysed and recorded for future decision making.

III. Proposed Approach:

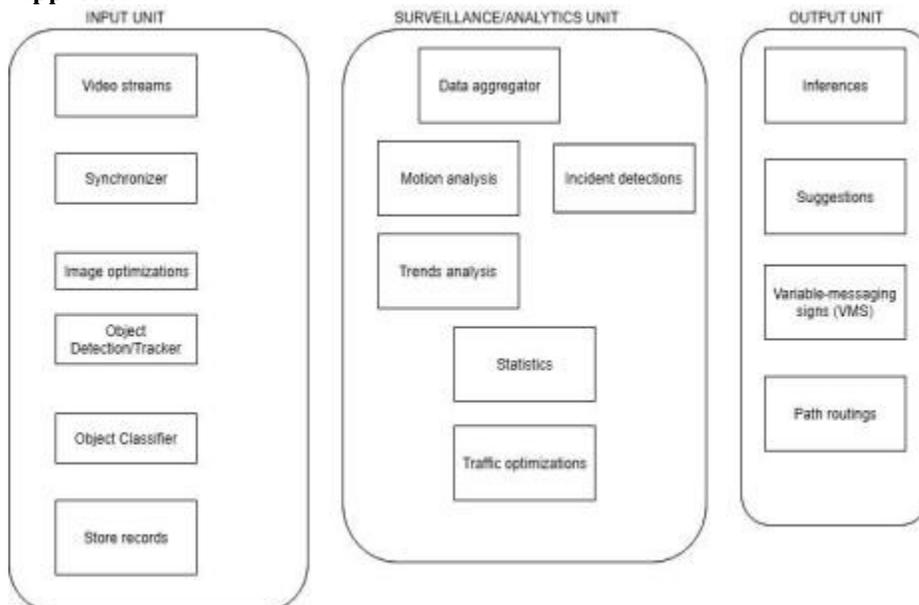


Fig 4: Block diagram of Traffic Control & Surveillance System

The system is supposed to be the build in modular fashion as shown above where each component has less or no dependency on another. This was each model can be developed further independently. Our main area of interest is surveillance unit in this paper while development of other units is equally important.

Our goal in this paper is to propose a deterministic model for managing the traffic at intersections based on kind of manoeuvre while accounting for different factors affecting. The current traffic management at an intersection divides the entire cycle time into equal pre-timed signals for each phase. According to our survey, the traffic signals work at 80 seconds per phase at any junction. While allows 20 seconds for pedestrians on normal intervals. Lesser time to cross the intersection also led to vehicle reckless speeding on denser routes. The constant halting of vehicles at signals also led to cascade traffic congestion on roads behind. The example below(a) shows primitive stage traffic jam due to pre-timed traffic signals.

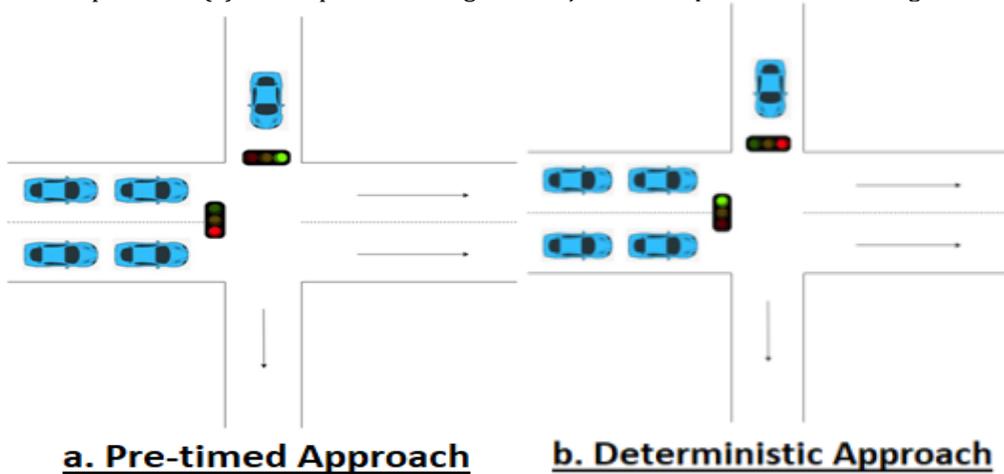


Fig 5: Comparison of two different approaches to handle traffic

This kind of system fails to support better mobility for highly travelled roads. Contrarily, our model will be able to asses such situation to optimize the traffic signals as shown above(b)(Ali, Kurokawa, & Shafie, 2013).

The system is based on two kinds of data sources for calculating all its predictions i.e statistical as well as real-time records. The prior records are used to train models while the newly perceived information is used to make inferences. Firstly, the intersections need to be efficiently divided into phases to maximize the carrying capacity of the junction. Traditionally, the roads are divided into 'n' number of phases for 'n' number of routes.

At junction, inflow = outflow

The traffic will always enter through one of the inflow roads and exit out of one of the outflow roads.

Combination (inflow, outflow) = inflow x outflow

Traffic density gives the approximate number of vehicles 1km of road contains(Geroliminis & Daganzo, 2008). It can be calculated as follows:

$$\rho(x) = \frac{cars}{nL - \sum_i^m a_i b_i} - \epsilon \tag{1}$$

Here, $\rho(x)$ = traffic density

$cars$ = number of cars

n = number of lanes

L = measured length of per lane (km)

$\sum_i^m a_i b_i$ = summation of all the road defects like potholes, construction, encroachment, etc

where, m - total number of defects

a_i = number of defect of i -th category

b_i = occupied road length per defect of i -th category

ϵ = error function (detection errors, accuracy, etc)

Preferably use gauss error function for normal distribution of density:

$$\text{erf}(x) = \frac{1}{\sqrt{\pi}} \int_{-x}^x e^{-t^2} dt$$

Or, with enough dataset we can use standard error of the mean for simplicity:

$$\text{erf}(x) = \frac{1}{N} \sum_i^n (\text{Observedoutput} - \text{Actualoutput})^2$$

The grouping of the routes into phases depends on the traffic density at each road. As the junction is shared by more than one road, it faces with different situations like merging, diverging and crossing of vehicles. We have to reduce the chances of these causing problems. The traffic density function depends on many independent random variables, hence their normalize sum tends towards normal distribution; hence the central limit theorem will save us. We will be using logistic function to find how probability of traffic breakdown(Kerner, 2011).

$$p = \frac{1}{1 + e^{-m(x-\hat{x})}} \quad (2)$$

Here, p = probability of traffic breakdown

e = Euler's constant

\hat{x} = mean traffic density for moderate traffic congestion (cars/km)

m = curve slope or steepness

x = current traffic density (cars/km)

Based on historical data, following are the rules that needs to be considered while building

- Diverging means same inflow road but different outflow roads.
- Crossing means different inflow and outflow roads at a time.
- Converging means different inflow roads but same outflow road.
- While diverging, $p(x=\text{traffic jam})$ for all inflow roads = [0.0, 0.33]
- While crossing, $p(x=\text{traffic jam})$ for all inflow roads = [0.0, 0.65].
- While converging, $p(x=\text{traffic jam})$ for all inflow roads = [0.0, 0.90].
- Else high traffic density roads need to flow separately.

ALGORITHM:

Step 1: START

Step 2: DECLARE VARIABLES n_{inflow} , $n_{outflow}$, ARRAY(road_label), ARRAY(angles), ARRAY(probabilities), ARRAY(routes), 2D-ARRAY(graph), ARRAY(colour)

Step 3: READ values n_{inflow} & $n_{outflow}$

Step 4: FOR $i = 0$ to $n_{inflow}+n_{outflow}$

4.1: READ label road_label[i]

4.2: READ value angles[i]

4.3: READ value probabilities[i]

Step 5: FOR $i = 0$ to n_{inflow}

5.1: FOR $j = 0$ to $n_{outflow}$

5.1.1: ADD (road_label[i], road_label[n_inflow+j]) to routes

Step 6: INITIALIZE graph[i][j] = 0 until $i, j < n_{inflow}*n_{outflow}$

Step 7: Looping through every route to find compatible routes

FOR $i = 0$ to $n_{inflow}*n_{outflow}$:

7.1: FOR $j = 0$ to $n_{inflow}*n_{outflow}$:

7.1.1: IF $i \neq j$ THEN:

7.1.1.1: SET $r11 = \text{routes}[i][0]$, $r12 = \text{routes}[i][1]$, $r21 = \text{routes}[j][0]$, $r22 = \text{routes}[j][1]$

7.1.1.2: SET $a11 = \text{angles}[r11]$, $a12 = \text{angles}[r12]$, $a21 = \text{angles}[r21]$, $a22 = \text{angles}[r22]$

7.1.1.3: SET $p11 = \text{probabilities}[r11]$, $p21 = \text{probabilities}[r21]$

7.1.1.4: IF $a11 == a21$ AND $a12 \neq a22$ AND $p11 \leq 0.9$ AND $p21 \leq 0.9$ THEN // DIVERGING ROUTES

7.1.1.4.1: graph[i][j] = 1

7.1.1.4.2: GOTO 7.1.2

7.1.1.5: IF $a11 \neq a21$ AND $a12 == a22$ AND $p11 \leq 0.33$ AND $p21 \leq 0.33$ THEN // CONVERGING

ROUTES

7.1.1.5.1: graph[i][j] = 1

7.1.1.5.2: GOTO 7.1.2

7.1.1.6: SET $LOW1 = \text{MIN}(a11, a12)$, $HIGH1 = \text{MAX}(a11, a12)$, $LOW2 = \text{MIN}(a21, a22)$, $HIGH2 = \text{MAX}(a21, a22)$

7.1.1.7: IF (($LOW1 \geq LOW2$ AND $LOW1 \leq HIGH2$) != ($HIGH1 \geq LOW2$ AND $HIGH1 \leq HIGH2$)) OR (($LOW2 \geq LOW1$ AND $LOW2 \leq HIGH1$) != ($HIGH2 \geq LOW1$ AND $HIGH2 \leq HIGH1$)) AND $p11 \leq 0.65$ AND $p21 \leq 0.65$ THEN // CROSSING ROUTES

7.1.1.7.1: graph[i][j] = 1

```

7.1.1.7.2: GOTO 7.1.2
7.1.1.8: IF NOT ((LOW1>=LOW2 AND LOW1<=HIGH2) != (HIGH1>=LOW2 AND HIGH1<=HIGH2)) OR
((LOW2>=LOW1 AND LOW1<=HIGH1) != (HIGH2>=LOW1 AND LOW1<=HIGH1)) THEN // NOT
CROSSING ROUTES
7.1.1.8.1: graph[i][j] = 1
7.1.1.8.2: GOTO 7.1.2
7.1.1.9: graph[i][j] = 0 // INCOMPATIBLE ROUTES
7.1.2: ENDIF
7.2: END LOOP
Step 8: STOP
    
```

IV. Performance Evaluation:

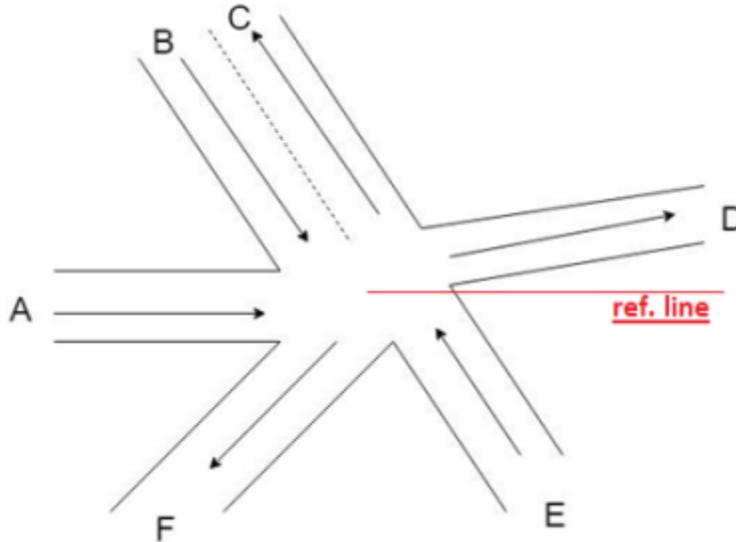


Fig 6: Sample junction representation

For instance, in above situation the traffic lights will allow each route to pass at once. Hence, n equals to 9. Here, we can count the number of roads based on type of traffic flow.

Above, inflow = A, B, E & outflow = C, D, F

Hence, possible routes = [AC, AD, AF, BC, BD, BF, EC, ED, EF]

At above junction, let's put density data to test using above formulae,

Road label	Density (cars/km)	Probability of Traffic Breakdown [0,1] ($\hat{x} = 97.95, m = 1/15$)
A	78.74	0.21743676353
B	212.7	0.999524182381
C	131.3	0.90232537093
D	52.32	0.0455641156375
E	32.23	0.0123541354768
F	80.42	0.237096075094

Following are some of the examples of crossing paths in above example:

Path 1	Angles (in degrees)	Path 2	Angles (in degrees)
AC	180-120	BD	120-13
BF	120-225	EA	300-180
EC	300-120	BD	120-13

Following are some of the examples of non-crossing paths in above example:

Path 1	Angles (in degrees)	Path 2	Angles (in degrees)
AF	180-225	BD	120-13
BF	120-225	ED	300-13
CD	120-13	AF	180-225

The diagram below (a) shows the undirected graph of all possible paths with links between vertices represents incompatible routes.

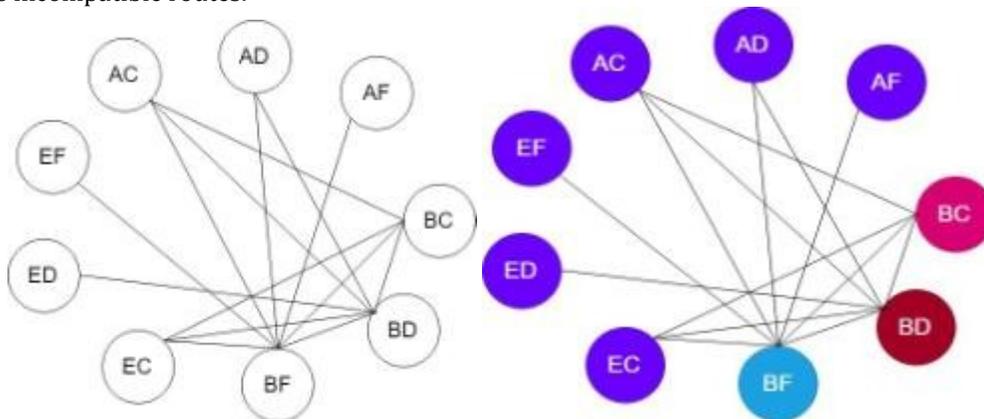


Fig. (a) Fig. (b)
Fig 7: Graph representation of compatible roads

Later, we can apply graph colouring algorithm to grouping vertices (routes) based on links with other routes (Batugedara & Lanel, 2017). Vertices with same colour are of same phase. Hence, we condensed the original 9 phases into 4 phases. It can be clearly understood from above diagram that due to high chances of traffic jam due to inflow from road B (i.e 99.95%), each route starting from B is given a separate phase.

V. Conclusion:

The proposed system improves the existing system in many features. While the grouping of road routes into phases based on type of manoeuvre reduced the waiting time of each vehicle, the overall maximum utilization of the intersection area made the junction more efficient. The system still relies on the statistical data for making some assumptions. Also, significant errors in detecting vehicles may lead to inaccurate results. The solution doesn't address the backward propagation of congestion as the outflow of vehicles is assumed to be congestion-free and dependent on the heading junction.

VI. Acknowledgment:

We thank our colleagues from M. H. Saboo Siddik College of Engineering who provided insight and expertise that greatly assisted the research.

We thank our team mentor Dr. Riyazoddin Siddiqui for assistance with statistical models & algorithms and other associated professors for comments that greatly improved the manuscript.

We would also like to show our gratitude to our maths professors for sharing their pearls of wisdom with us during the course of this research. We are also immensely grateful to Er. Mirza Zainab and Dr. Shaikh Ashfaq Amir for their comments on an earlier version of the manuscript, although any errors are our own and should not tarnish the reputations of these esteemed persons.

REFERENCES:

1. Ali, H., Kurokawa, S., & Shafie, A. (2013). Autonomous Road Surveillance System: A Proposed Model for Vehicle
2. Batugedara, B., & Lanel, G. (2017). Designing the traffic phases using graph colouring and book thickness to reduce the traffic for a selected junction in Colombo city.
3. Buxton, H., & Gong, S. (1994). Visual surveillance in a dynamic and uncertain world.
4. Dash, D. (2018). 'Traffic congestion costs four major Indian cities Rs. 15 lakhs crore a year'.
5. Geroliminis, N., & Daganzo, C. (2008). Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings.
6. Kerner, S. (2011). Optimum principle for a vehicular traffic network: minimum probability of congestion.
7. Khatri, D. (2018). Traffic violation: Police register 10,000 cases in Mumbai. September 28, 2018.
8. Sen, S. (2017). Every day, 700 new vehicles hit Mumbai roads.
9. Sen, S. (2018). From 20.3 lakh to 32 lakh, Mumbai's vehicle count up 56% in just 5 years.
10. Sharma, A. (2012). Traffic Congestion in Mumbai.

MODEL OF CALCIUM DYNAMICS TO STUDY THE ROLE OF AMYLOID BETA IN ALZHEIMER'S DISEASE

TEJAL KANGANE , SAIF SAYED , DR. MANSHA KOTWANI
THADOMAL SHAHANI ENGINEERING COLLEGE
P G KHER MARG T.P.S. III , BANDRA WEST
MUMBAI 400 050

ABSTRACT: Alzheimer's disease (AD) is a progressive neurological disorder in which the death of brain cells leads to memory loss and cognitive decline. AD is characterized by deposition of β -amyloid (A β) oligomers in the vicinity of neurons in the brain which causes dysfunction of neuronal calcium homeostasis. As A β induces the up-regulation of intracellular Ca²⁺ and the increased Ca²⁺ entry in the cytoplasm caused due to improper regulation of Ca²⁺ leads to neuronal cell death in Alzheimer's disease. This mathematical model highlights the effect of A β on calcium distribution in the presence of L type Calcium channel and PMCA pump in neuron cell this study depicts the relation between Ca²⁺ concentration and A β during neurodegenerative disorder.

Keywords: Alzheimer's disease, Amyloid beta, Calcium signaling, Calcium homeostasis

I. Introduction

Neurons are information messengers. They use electrical impulses and chemical signals to transmit information between different areas of the brain, and between the brain and the rest of the nervous system. They are born in the areas of the brain that are rich in precursor cells. Neurodegeneration is the progressive loss of structure or function of neurons, including death of neurons. Many neurodegenerative diseases – including amyotrophic lateral sclerosis, Parkinson's disease, Alzheimer's disease, and Huntington's disease – occur as a result of neurodegenerative processes. Such diseases are incurable, resulting in progressive degeneration and/or death of neuron cells.

Alzheimer's disease disrupts this communication among neurons, resulting in loss of function and cell death. AD is a chronic neurodegenerative disease which usually starts slowly and gradually and worsens overtime. It is recognized as the most common cause of dementia affecting the people aged 65 and above. AD is thought to be caused by the abnormal accumulation of proteins in and around the brain cells. One of the proteins involved is called **amyloid**, deposits of which form plaques around brain cells [1].

Calcium (Ca²⁺) ions have a major role in cellular signaling, as the involvement in the cytosol of cytoplasm exerts allosteric regulatory effects on many enzymes and proteins [6]. Calcium ion plays an important role in the structure and function of nerve cell circuit in brain. Ca²⁺ signaling helps in regulating neuronal growth, exocytosis, synaptic plasticity and cognitive function. The neuron normal function gets affected due to disturbances in Ca²⁺ homeostasis [7]. There are several parameters that affect calcium concentration in cytosol like buffers, pump leak, receptor etc. in normal and healthy neurons in brain. Many studies have shown the effect of endoplasmic receptors and calcium pump and various channel effect on calcium homeostasis in neuron and other body cells in normal condition [8]. But very few studies are available to study the role of calcium homeostasis in pathogenic condition which causes neurodegenerative disease. AD is one of the neurological disorder in which a protein called Amyloid Beta deposits in and around the brain cell which disrupts the calcium homeostasis in normal neuron cell.

Amyloid beta is a complex biological molecule that interacts with different receptors and forms insoluble assemblies and its abnormal depositions alter the normal neuronal conditions. A β is the major element of amyloid plaques formed due to the improper cleavage of amyloid precursor protein (APP). During AD, autophagy (mechanism for removing amyloids) is increased and the transfer of these autophagic vesicles to the lysosomes is blocked which may contribute to the deposition of amyloid around the cells.[1]

In order to understand the neurological disorder it is necessary to observe the role of protein which disrupts the normal function of neuron cell. Thus, we developed a mathematical model to study calcium homeostasis in the presence of A β deposition, L-type calcium channel and PMCA pump.

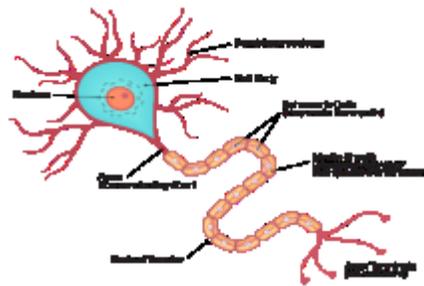


Figure 1 [4]

Those neurons where amyloid deposition is in excess have more Ca^{2+} level than the normal resting level. Calcium deposition is directly proportional to the deposition of amyloid around respective neurons. Amyloids perturb Ca^{2+} homeostasis. First, AICD affects the expression of Ca^{2+} channels and buffers. Second, $A\beta$'s perturb the balance between Ca^{2+} entry in and extrusion out of the cytoplasm. In healthy neurons, these processes equilibrate, leading to a basal Ca^{2+} level in the range 50–100nM.

II. Literature Survey

Ranjan et al. In this paper they built up a mathematical model for breaking down and understanding calcium signaling in presence and absence of β -amyloids statements. A hypotheses was tried worried to certainty that does β -amyloids have any impact in calcium signaling or not. The proposed model was a blend of three distinct models viz. The Fall Kiezer , Mitochondrial PTP model and Amyloid Deposition. They joined these three models to acquire a composite model which demonstrates the impact of amyloid fixation when there is presentation of calcium into the cell by thinking about its impact of expanding permeability. They plotted the activity potential without pathology and with beta amyloid misfolding and deposition. This model gives a total perspective on the sub-atomic procedures happening in the nerve cell during Alzheimer's disease . By consolidating the models they reproduced the opening of PTP utilizing amyloid deposition as a trigger.[1]

Kim N.Green et al. His review plates the proof for a role for calcium dyshomeostasis in the commencement of pathology, just as the proof for these pathologies themselves upsetting ordinary calcium homeostasis, which lead to synaptic and neuronal dysfunction, synaptotoxicity and neuronal loss, underlying the dementia associated with the disease. Underlying the phenotypic changes are the appearance of several hallmark pathologies. The first is the collection of the amyloid-peptide ($A\beta$) into extracellular plaques, where the second is the formation of neurofibrillary tangles (NFTs) made out of hyperphosphorylated π protein inside neurons.

Calcium may play a crucial job in any number of parts of AD – from regulating APP handling towards the amyloidogenic processing pathway, to fundamental memory and neuronal loss. In vivo evidence has demonstrated that $A\beta$ increments dendritic spine calcium levels, debilitates LTP and prompts learning and memory deficits, highlighting the synapse as a prime $A\beta$ target, although age related changes in calcium homeostasis may allow the buildup of pathologies associated with AD, through both direct and indirect mechanisms. Synaptic transmission is profoundly subordinate upon firmly directed calcium at both the pre- and post-synaptic membranes and it is at these membranes all the ion channels $A\beta$ has been accounted for to disturb are clustered, likely prompting synaptic dysfunction and more, while focusing on NMDA receptors with memantine is advantageous in AD patients, just as various models of the disease, setting calcium dyshomeostasis immovably at the centre of the disease.[2]

Joëlle De Caluwé et al. In their research paper ,they displayed the primary hypothetical model that formalizes a positive input circle that offers ascend to bistability. On account of the positive input circle among $A\beta$ and Ca^{2+} three unfaltering states exist together; low, stable one that relates to a solid circumstance, described by low dimensions of amyloids and Ca^{2+} . The halfway state is unsteady. The high enduring state compares to the obsessive circumstance where the dimensions of the two amyloids and Ca^{2+} turns out to be high which prompts neuronal cell passing. This model proposes that test evaluation requires a cautious ID of the condition of the neurons. The bistability is generally autonomous of the parameters of the numerical model and when just all Ca^{2+} or $A\beta$ controlled procedures.[3]

In the present study, the role of amyloid beta on the calcium dynamics has been discussed by including L-type calcium channel and PMCA Pump which is a transport protein in the plasma membrane of nerve cells which helps in expelling calcium Ca^{2+} from the cells. The consolidated model that incorporates the conduct

of amyloid beta, changes in penetrability, leak voltages and action for expelling calcium out of the cell brings about portraying the increasing calcium concentration into the cell. This may infer that gathering of amyloid beta oligomers pursued by expanded calcium signaling will cause neuronal cell death in Promotion.

III. Mathematical Formulation

In perspective on these preliminary observations, we propose the essential model schematized in Fig. 2. The 2 components of the model are the convergences of Aβ (without refinement among intra- and extracellular compartments, nor between the amyloid mixes of different lengths and in different oligomerization states) moreover, the intracellular Ca²⁺ fixation. In the equations, these fixation are implied by a and c, respectively. The Ca²⁺ fixation should represent the basal element of cytoplasmic Ca²⁺, or alternatively, the time typical centralization of this particle, whose value does not basically depend upon the short-lived Ca²⁺ peaks rising up out of the electrical movement of the neurons. Aβ is thought to be blended at a steady rate (V₁) and eliminated with a first-order kinetics, described by a rate constant k₁. Initiation of Aβ synthesis by Ca²⁺ is represented by a Hill term, with a maximal rate V_α, a half-saturation constant K_α and a Hill coefficient n. So also, Ca²⁺ enters the cytoplasm at a constant rate V₂ and is wiped out with a first-order kinetics, characterized by a rate constant k₂. [3]

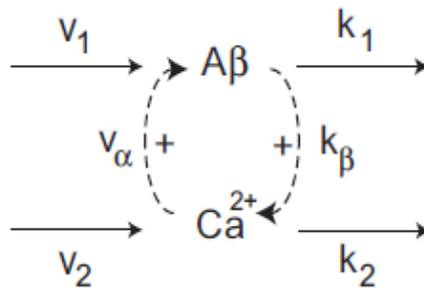


Fig. 1. Schematic representation of the minimal model describing the interplay between Ca²⁺ and Aβ during the onset of Alzheimer's disease. The positive feedback exerted by Ca²⁺ on the formation of Aβ and the fact that Aβ tend to increase intracellular Ca²⁺ create a positive loop that is responsible for the occurrence of bistability.

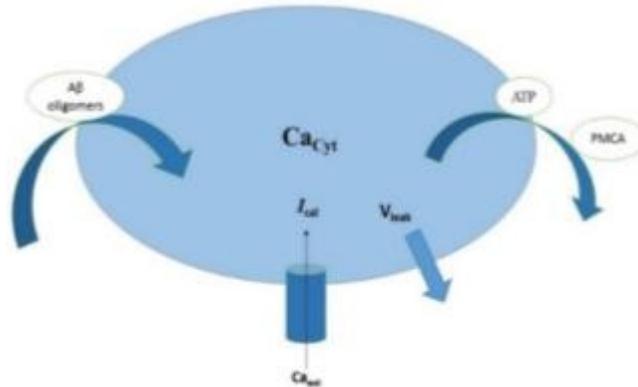


Fig.2 Calcium Dynamics in Neuron cell

Our mathematical model (discussed in Fig (1) and Fig (2)) of Calcium dynamics can in presence of amyloid beta in neuron cell can be written as

$$\frac{dCa}{dt} = k_{\beta} a^m + J_{in} - J_{out} \dots (1)$$

With Initial condition [Ca²⁺]_I = 0.05 at t=0

$$\frac{dCa}{dt} = k_{\beta} a^m + V_{leak} + I_l - J_{PMCA} \dots (1.1)$$

Also L type calcium channel already have been reported in neuron cell [9]. The I_L channel is responsible for the fast depolarization of the membrane at the initiation of an action potential and it has been modeled using Goldman-Hodgkin-Katz (GHK) equation.

$$I_l = \frac{P_{Ca} z_{Ca}^2 F^2 V_m}{RT} \frac{[Ca]_i - [Ca]_o e^{-\frac{zFV_m}{RT}}}{1 - e^{-\frac{zFV_m}{RT}}} \dots (1.2)$$

Where, P_{Ca} (m/s) is calcium permeability, z_{Ca} is valency of calcium, F (C/mole) is Faraday's constant, V_m (volts) is resting membrane potential, R (J/K mole) is universal gas constant, T (Kelvin) is absolute temperature, $[Ca]_i$ & $[Ca]_o$ are intracellular and extracellular $[Ca^{2+}]$ (μM) concentration respectively.

Now combining equation (1.1) & (1.2), We get

$$\frac{dCa}{dt} = k_{\beta} a^m + a_1 Ca + \frac{P_{Ca} z F V_m}{RT} \frac{Ca - [Ca]_o e^{-\frac{zFV_m}{RT}}}{1 - e^{-\frac{zFV_m}{RT}}} - \frac{V_{pm} Ca^n}{k_{pm}^n + Ca^n} \dots (1.3)$$

let 'u(t)' represent the calcium concentration in cytosol at time t and assume $(k/u) \ll 1$, We get

$$\frac{du}{dt} = k\beta a^m + a_1 u + \frac{T[u - 1800S]}{1 - S} - \frac{V_{pm} u}{k} \dots \dots (1.4)$$

Where

$$T = \frac{P_{Ca} z F V_m}{RT}, S = e^{-\frac{zFV_m}{RT}}$$

The above equation (1.4) calcium dynamics in neuron cell when amyloid beta is present in the cytosol. Equation (1.4) represent a linear differential equation with initial condition $u = 0.05$ when $t = 0$. We solve equation (5) using MATLAB for the given table (1) and the results hence been discussed in the form of figures. Numerical results have been discussed in details in the next section.

IV. Numerical Results and Discussion

The numerical values of biophysical constants which have been used to compute the numerical results are given in Table (1)

Table-1 List of Physiological Parameters

Parameters	Values	Biological Significance
V_2	4 nm/s	Constant rate of Ca^{2+} entry in cytoplasm
k_{β}	0.2 nm ³ /s	Rate constant of increased Ca^{2+} entry
k_2	0.1/s	Rate constant of Ca^{2+} elimination
m	4	Co-operativity coefficient
V_1	0.0065 nm/s	Constant rate of $A\beta$ synthesis
V_{α}	0.05 nm	Maximal rate of $A\beta$ synthesis
K_{α}	120 nm	Half-saturation constant
k_1	0.01/s	Rate constant of $A\beta$ elimination
K_{pm}	0.425 μM	Dissociation Constant of PMCA
n	1	Hill Co-efficient
V_{pm}	2.8 μM /s	Volume of PMCA pump
a_1	0.02/s	Leak Constant
z_{Ca}	2	Valency of calcium
F	96487 C/mole	Faraday's constant

R	8.314 J/Kmole	Universal gas constant
T	293K	Absolute Temperature
P_{Ca}	$5.4 \times 10^{-6} \text{m/s}$	Calcium Permeability
$[Ca]_o$	1.8mM	Concentration Ca^{+2} entry
V_m	-0.07 volts	Resting membrane potential
V_{cyt}	$523.6 \mu\text{m}^3$	Volume of Cytosol

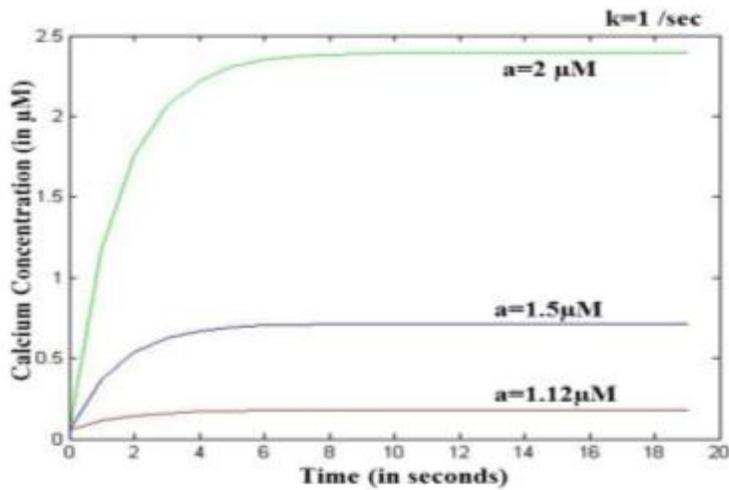


Fig.3 Temporal Calcium distribution in neuron cell for different values of amyloid beta for $k=1/\text{sec}$

Figure 3 shows the effect of amyloid beta concentration on calcium distribution in neuron cell. As amyloid beta concentration increases the calcium concentration increases. Thus, it can be stated that calcium concentration is directly proportional to amyloid beta concentration. Also the peak calcium concentration is $2.39 \mu\text{M}$ for $a=2 \mu\text{M}$, $0.71 \mu\text{M}$ for $a=1.5 \mu\text{M}$ and $0.174 \mu\text{M}$ for $a=1.12 \mu\text{M}$ at $t=5 \text{sec}$. Thus the amyloid beta concentration increases by 0.5 times calcium concentration increases by almost three times. This results discussed above in the figure are very important in order to study the role of amyloid beta in Alzheimer's disease.

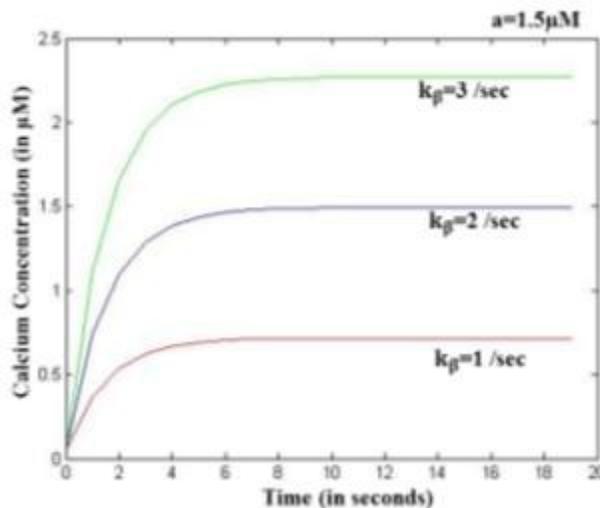


Fig.4 Temporal calcium concentration distribution in neuron cell for different values of amyloid beta binding affinity

Figure 4 shows variation in the calcium concentration as binding affinity of amyloid beta varies. Calcium concentration increases with the same rate as we change the value of amyloid beta affinity. It seems from the above figure initially calcium concentration increases with time rapidly and thereafter it shows a constant value after $t = 4$ sec for different amyloid beta affinity. This shows that amyloid beta disrupts the effect of PMCA pump which is responsible for expelling the excess calcium out of cell and it results the higher calcium concentration remains in cytosol which causes detrimental effects on cell function.

V. Conclusion

The model discussed above have been developed to study the calcium dynamics and role of amyloid beta during Alzheimer's disease. The results imply here are very important and showed that there is a significant effect of amyloid beta on calcium dynamics in neuron cell during Alzheimer diseases. The above model can be studied by adding more and more realistic parameters. Also, it can be extended to study various neurodegenerative diseases like the Parkinson's disease, and Huntington's disease. The result discussed above are in fully agreement with physiological parameters and can be of great use for further study in computational neuroscience and system biology.

References

1. Bobby Ranjan, KetHing Chong and JieZheng .Composite mathematical modeling of calcium signaling behind neuronal cell death in Alzheimer's disease. BMC Systems Biology 2018, 12(Suppl 1):10
2. Kim N. Green. Calcium in the initiation, progression and as an effector of Alzheimer's disease pathology. J. Cell. Mol. Med. Vol 13, No 9A, 2009 pp. 2787-2799
3. Joëlle De Caluwé, Geneviève Dupont . The progression towards Alzheimer's disease described as a bistable switch arising from the positive loop between amyloids and Ca^{2+} Journal of Theoretical Biology 331(2013)12-18
4. <https://toxtutor.nlm.nih.gov/14-004.html>
5. https://www.google.com/search?q=amyloid+beta&rlz=1C1CHBF_enIN729IN729&source=lnms&tbm=isch&sa=X&ved=0ahUKEw1q1fLMtNzhAhUt4nMBHWQbAeAQ_AUIDigB&biw=1366&bih=657#imgrc=joJ35QnjV_KmdM
6. Berridge, MJ. Bootman MD. Lipp, P. 1998. Calcium-a life and death signal. Nature, 395:645-648.
7. Jha, B. Joshi, H. Dave, D. D. 2016. Portraying the Effect of Calcium-Binding Proteins on Cytosolic Calcium Concentration Distribution Fractionally in Nerve. Interdisciplinary Sciences: Computational Life Sciences, 1-12.
8. Kotwani, M, Adlakha, N. Mehta, M N. 2017. Modeling of Endoplasmic Reticulum and Plasma Membrane Ca^{2+} Uptake and Release Fluxes with Excess Buffer Approximation (EBA) in Fibroblast Cell. International Journal of Computational Material Sciences and Engineering, 4(4):1750004-12.
9. Keener J. and Sneyd J. 1998. Mathematical Physiology, Vol. 8, Springer, 53-56.

NUMERICAL ESTIMATION OF CALCIUM CONCENTRATION IN LIVER CELLS INVOLVING INFLUX, OUT FLUX AND BUFFERS

¹Yogita D. Jagtap & ²Neeru Adlakha

¹Ph.D scholar,²Professor

¹Applied Mathematics and Humanity Department,

¹Sardar Vallabhabai National Institute of Technology, Surat, India

ABSTRACT: The specific calcium concentration patterns in the liver cells are responsible for various activities contributing to the function of the cell as well as liver as a whole. Thus, understanding of these calcium concentration patterns responsible for various processes in the liver cell is necessary for clinical and biomedical scientists for diagnosis and treatment of disorders of the liver. In this paper a mathematical model is proposed to investigate these calcium concentration patterns in a liver cell for one dimensional unsteady state case. The calcium buffering process, diffusion, out fluxes due to SERCA pump and influx due to source channel and leak has been incorporated in the model. The finite volume method has been employed to obtain numerical results and to study the relationship of various calcium concentration patterns with the processes of the cell like leak, buffering, influx and out flux etc. in the liver cell.

Keywords: Liver cell, influx, out flux, EGTA, BAPTA

I. Introduction

Liver is multi-functioning vital organ in human body. Almost all functions of liver are regulated by calcium level in it. The different chemicals like enzymes and proteins achieve its target role in presence of specific calcium concentration inside liver cell. The balanced chemical level in liver cell is regulated by cytosolic calcium. The endoplasmic reticulum (ER) is basic storage compartment in a liver cell (Dupont 2000). It bears clusters of channels releasing calcium. The influx of calcium through these channels and leakage through membrane of ER into cytosol increases calcium concentration in huge amount. The increase in calcium concentration level than normal range is toxic to cell for its survival. Therefore cell actively suck calcium back into ER with the help of SERCA pumps (Dupont, 2000 and Thomas, 1991). The maximum amount of free calcium bind with varieties of buffers (proteins) present in the liver cell (Fall, 2002).

In the past the number of attempts are noticed to numerical study of calcium dynamics in neuron cell (Tewari, 2010 and Jha 2015), astrocyte (Jha B.et. al. 2013, 2014), myocyte (Pathak, 2015, 2016), fibroblast (Kotwani, 2014, 2017), acinar cell (Manhas, 2014), oocyte (Pandey, 2013 and Naik, 2016), hepatocyte cell (Jagtap, 2018) etc. But no effort is reported in past to study calcium dynamics in liver cell involving influx, out flux, leak, pump and excess buffer. In this paper a finite volume model is proposed to study calcium dynamics in a liver cell involving influx through channel, leak, and out flux through SERCA pump and excess buffers like EGTA and BAPTA. The mathematical model is described in the following section.

II. MATHEMATICAL MODEL

The fluxes of calcium from endoplasmic reticulum (ER) into cytosol and vice versa in a liver cell can be written as,

Influx through calcium channel (I_{Ch}) and leakage (I_{Lk}) \leftrightarrow Out flux through SERCA pump (O_{SP}) and Out flux due to binding with buffer

The change in calcium concentration in cytosol (C_{Cy}) of liver cell w.r.t time due to diffusion, in and out fluxes and due to association and dissociation of buffer is expressed by following (Jafri, 1995),

$$\frac{\partial C_{cy}}{\partial t} = D_{Ca} \frac{\partial^2 C_{cy}}{\partial x^2} + V_R (I_{Ch} + I_{Lk}) - O_{SP} - k^+ [B]_{\infty} (C_{cy} - C_b) \text{ for, } 0 \leq x \leq 15; \quad t > 0. \quad (1)$$

Where, k^+ is buffer association rate constant and $[B]_{\infty} = \frac{K[B]_T}{K+C_{cy}}$. Where, $K = \frac{k^-}{k^+}$ is buffer dissociation constant.

$$I_{Ch} = P \alpha_{Chan} (C_{ER} - C_{Cy}) \quad (2)$$

$$I_{Lk} = \alpha_{leak} (C_{ER} - C_{Cy}) \quad (3)$$

Here, channel conductance is given by α_{Chan} (in sec^{-1}), and the open channel probability is represented by P. The C_{ER} represents calcium concentration in ER, and α_{leak} represents the leak flux constant,

By conservation law,

$$C_T = V_R C_{ER} + C_{Cy} \tag{4}$$

Using equation (4) in equation (2) and (3), and then rearranging we get,

$$I_{Ch} = P \frac{\alpha_{Chan}}{V_R} (1 + V_R) \left(\frac{C_T}{(1 + V_R)} - C_{Cy} \right) \tag{5}$$

$$I_{Lk} = \frac{\alpha_{leak}}{V_R} (1 + V_R) \left(\frac{C_T}{(1 + V_R)} - C_{Cy} \right) \tag{6}$$

The outward flux O_{SP} is given by following expression.

$$O_{SP} = P^{max} \frac{C_{cy}^2}{C_{cy}^2 + K_p^2} \tag{7}$$

Where, the maximal pump rate denoted by P^{max} and MichaelisMenten constant is denoted by K_p .

Substituting (5),(6),(7) in (1) gives,

$$\frac{\partial C_{cy}}{\partial t} = D_{Ca} \frac{\partial^2 C_{cy}}{\partial x^2} + (1 + V_R)(\alpha_{leak} + P \alpha_{Chan}) \left(\frac{C_T}{(1 + V_R)} - C_{Cy} \right) - P^{max} \frac{C_{cy}^2}{C_{cy}^2 + K_p^2} - k^+[B]_{\infty}(C_{cy} - C_b) \tag{8}$$

Table 1.Nomenclature and numerical values(JAFRI, 1995),

Symbol	Description	Numerical Value	Unit
D_{Ca}	Diffusion coefficient of calcium	200-300	$\mu m^2/s$
$[B]_T$	Total buffer concentration	50-100	μM
k^+	EGTA buffer association constant	1.5	$\mu M^{-1}S^{-1}$
K	EGTA buffer dissociation constant	0.2	μM
k^+	BAPTA buffer association constant	600	$\mu M^{-1}S^{-1}$
K	BAPTA buffer dissociation constant	0.17	μM
V_R	Volume ratio of Er with cytosol	0.185	
C_T	Total cytosolic calcium concentration	0.2	μM
α_{leak}	Ca^{2+} leak flux constant	0.11	S^{-1}
α_{Chan}	Maximum Ca^{2+} channel flux	0.6	S^{-1}
P^{max}	Maximum Ca^{2+} uptake through pump	0.9	$\mu M^{-1}S^{-1}$
K_p	Dissociation constant of Ca^{2+} to pump	0.1	μM

INITIAL CONDITIONS:The basal cytosolic calcium concentration initially before release of calcium from source channel on ER membrane is considered as **0.1 μM** (Dupont 2000).This gives,

$$(C_{cy})_{t=0} = 0.1 \mu M \tag{9}$$

BOUNDARY CONDITIONS:The left boundary condition at the node 1 located at $x=0$ where source channel is kept, is framed as given below,

$$\lim_{x \rightarrow 0} -D_{Ca} \frac{dC_{cy}}{dx} = C_s \tag{10}$$

Where, C_s represents the influx of Ca^{2+} from source channel.

The right boundary conditions can be framed by assuming the fact that, basal calcium concentration is **0.1 μM** away from source,

$$\lim_{x \rightarrow 15} C_{cy} = C_b = 0.1 \mu M \tag{11}$$

The liver cell is approximately $15 \mu m$ in length and C_{cy} approaches to the basal concentration of $0.1 \mu M$ as $x \rightarrow 15$. Therefore second boundary condition is taken at other side of liver cell.

III. Solution by finite volume method

To employ finite volume method the liver cell in one dimension is discretized into subintervals called as discrete control volumes as shown in "Fig.1". The general node Z is surrounded by control volume. The space between boundary nodes 1 and 32 is uniformly discretized by 30 nodes.(Versteeg, 1995).

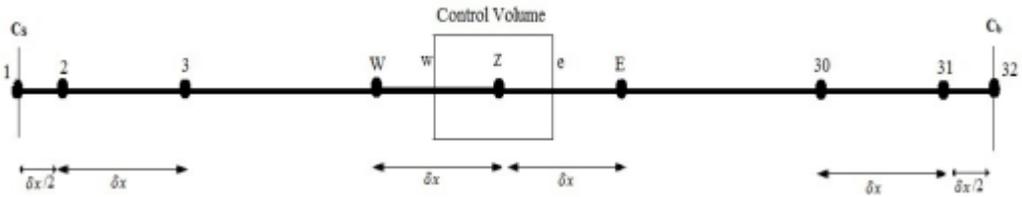


Figure1. Discretization of liver cell in one dimension

We express given nonlinear partial differential equation (8) into linear partial differential equation by assuming following two cases.

Case I) When $K_p \gg C_{cy}$: In this case $\frac{C_{cy}^2}{C_{cy}^2 + K_p^2} < \frac{C_{cy}^2}{K_p^2} < \frac{C_{cy}}{K_p}$, with this substitution equation (8) reduces to,

$$\frac{\partial C_{cy}}{\partial t} = D_{Ca} \frac{\partial^2 C_{cy}}{\partial x^2} + (1 + V_R)(\alpha_{leak} + P \alpha_{Chan}) \left(\frac{C_T}{(1 + V_R)} - C_{cy} \right) - P^{max} \frac{C_{cy}}{K_p} - k^+[B]_{\infty}(C_{cy} - C_b) \quad (12)$$

Rearranging equation (12) we get following form of equation,

$$\frac{1}{D_{Ca}} \frac{\partial C_{cy}}{\partial t} = \frac{\partial^2 C_{cy}}{\partial x^2} - l_1 C_{cy} + m_1 \quad (13)$$

Here, $l_1 = \frac{1}{D_{Ca}} \left[(1 + V_R)(\alpha_{leak} + P \alpha_{Chan}) + \frac{P^{max}}{K_p} + k^+[B]_{\infty} \right]$

$$m_1 = \frac{1}{D_{Ca}} [C_T(\alpha_{leak} + P \alpha_{Chan}) + k^+[B]_{\infty} C_b]$$

Case II) When $K_p \ll C_{cy}$: In this case put $K_p = \beta C$ for $0 < \beta < 1$ then,

$$\frac{C_{cy}^2}{C_{cy}^2 + K_p^2} = \frac{1}{\beta^2 + 1} \quad (14)$$

Using it in equation (8) gives,

$$\frac{\partial C_{cy}}{\partial t} = D_{Ca} \frac{\partial^2 C_{cy}}{\partial x^2} + (1 + V_R)(\alpha_{leak} + P \alpha_{Chan}) \left(\frac{C_T}{(1 + V_R)} - C_{cy} \right) - P^{max} \frac{1}{\beta^2 + 1} - k^+[B]_{\infty}(C_{cy} - C_b) \quad (15)$$

Rearranging equation (15) we get following form of equation,

$$\frac{1}{D_{Ca}} \frac{\partial C_{cy}}{\partial t} = \frac{\partial^2 C_{cy}}{\partial x^2} - l_2 C_{cy} + m_2 \quad (16)$$

Here, $l_2 = \frac{1}{D_{Ca}} [(1 + V_R)(\alpha_{leak} + P \alpha_{Chan}) + k^+[B]_{\infty}]$

$$m_2 = \frac{1}{D_{Ca}} \left[C_T(\alpha_{leak} + P \alpha_{Chan}) - \frac{P^{max}}{\beta^2 + 1} + k^+[B]_{\infty} C_b \right]$$

The equation (13) and (16) can be written together in the following form,

$$\frac{1}{D_{Ca}} \frac{\partial C_{cy}}{\partial t} = \frac{\partial^2 C_{cy}}{\partial x^2} - l C_{cy} + m \quad (17)$$

For sake of convenience we denote C_{cy} by C. Integrating equation (17) w.r.t time and space gives,

For discretization of equation (9), integrating it w.r.t t and x over control volume (CV) gives,

$$\int_t^{t+\Delta t} \int_{x_w}^{x_e} \frac{1}{D_{Ca}} \frac{\partial c}{\partial t} dx dt = \int_t^{t+\Delta t} \int_{x_w}^{x_e} \left(\frac{\partial^2 c}{\partial x^2} - lc + m \right) dx dt \quad (18)$$

Simplifying integration w.r.t x gives,

$$\frac{\delta x}{D_{Ca} \Delta t} [c_Z - c_Z^0] = \left\{ \varphi \left(\frac{c_E - c_Z}{\delta x} - \frac{c_Z - c_W}{\delta x} \right) + (1 - \varphi) \left(\frac{c_E^0 - c_Z^0}{\delta x} - \frac{c_Z^0 - c_W^0}{\delta x} \right) - l \delta x [\varphi c_Z + (1 - \varphi) c_Z^0] + m c_{\infty} \delta x \right\} \quad (19)$$

Where, $0 \leq \varphi \leq 1$. Rearranging the equation (19) we get,

$$\left[\frac{\delta x}{D_{Ca}\Delta t} + \frac{\varphi}{\delta x} + \frac{\varphi}{\delta x} + l\varphi\delta x \right] C_Z = \frac{\varphi c_E + (1-\varphi)c_E^0}{\delta x} + \frac{\varphi c_W + (1-\varphi)c_W^0}{\delta x} + \left[\frac{\delta x}{D_{Ca}\Delta t} - \frac{(1-\varphi)}{\delta x} - \frac{(1-\varphi)}{\delta x} - l(1-\varphi)\delta x \right] C_Z^0 + mc_\infty\delta x \tag{20}$$

For all nodes from 3 to 30, putting $\varphi = \frac{1}{2}$ in equation (20) by Crank Nicolson method we get,

$$\left[\frac{\delta x}{D_{Ca}\Delta t} + \frac{1}{2\delta x} + \frac{1}{2\delta x} + \frac{l\delta x}{2} \right] C_Z = \frac{c_E + c_E^0}{2\delta x} + \frac{c_W + c_W^0}{2\delta x} + \left[\frac{\delta x}{D_{Ca}\Delta t} - \frac{1}{2\delta x} - \frac{1}{2\delta x} - \frac{l\delta x}{2} \right] C_Z^0 + mc_\infty\delta x \tag{21}$$

It can be expressed as follow,

$$a_Z c_Z = a_E c_E + a_W c_W + a_Z^0 c_Z^0 + a_E c_E^0 + a_W c_W^0 + S_u \tag{13}$$

Where, $a_Z = \frac{\delta x}{D_{Ca}\Delta t} + \frac{1}{2\delta x} + \frac{1}{2\delta x} + \frac{l\delta x}{2}$; $a_Z^0 = \frac{\delta x}{D_{Ca}\Delta t} - \frac{1}{2\delta x} - \frac{1}{2\delta x} - \frac{l\delta x}{2}$; $a_E = \frac{1}{2\delta x}$; $a_W = \frac{1}{2\delta x}$;

$S_u = mc_\infty\delta x$

The left boundary condition at node 2 is kept at specified concentration $c_W = c_W^0 = c_s$ and $a_W = 0$

$$a_Z c_Z = a_E c_E + a_Z^0 c_Z^0 + a_E c_E^0 + S_u \tag{14}$$

Where, $a_Z = \frac{\delta x}{D_{Ca}\Delta t} + \frac{1}{2\delta x} + \frac{1}{\delta x} + \frac{l\delta x}{2}$; $a_Z^0 = \frac{\delta x}{D_{Ca}\Delta t} - \frac{1}{2\delta x} - \frac{1}{\delta x} - \frac{l\delta x}{2}$; $a_E = \frac{1}{2\delta x}$; $S_u = mc_\infty\delta x + \frac{2c_s}{\delta x}$

Similarly for node 31, east boundary of control volume is kept at specified concentration $c_E = c_E^0 = c_b$ and $a_E = 0$

$$a_Z c_Z = a_W c_W + a_Z^0 c_Z^0 + a_W c_W^0 + S_u \tag{15}$$

Where, $a_Z = \frac{\delta x}{D_{Ca}\Delta t} + \frac{1}{2\delta x} + \frac{1}{\delta x} + \frac{l\delta x}{2}$; $a_Z^0 = \frac{\delta x}{D_{Ca}\Delta t} - \frac{1}{2\delta x} - \frac{1}{\delta x} - \frac{l\delta x}{2}$; $a_E = \frac{1}{2\delta x}$; $S_u = mc_\infty\delta x + \frac{2c_b}{\delta x}$

The equations obtained for case-I and case-II solved separately by using Gauss elimination method. The time step taken for simulation is 0.001sec. The results are obtained by simulation on MATLAB to study the effect of various parameters on calcium concentration patterns.

IV. RESULTS AND DISCUSSION

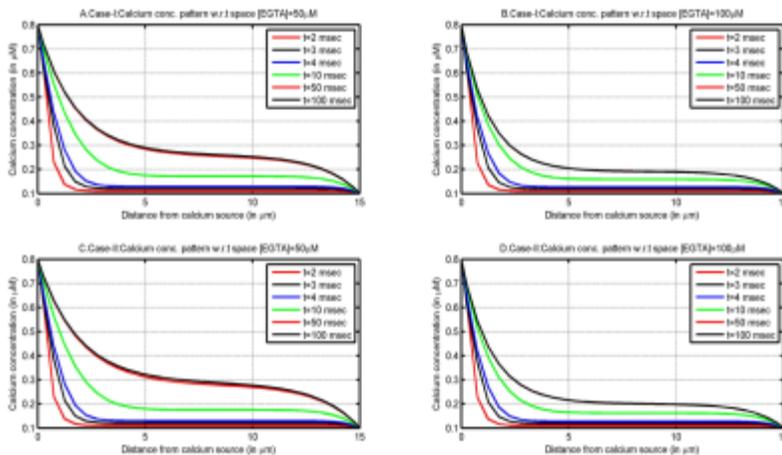


Figure2. Calcium concentration patterns w.r.t space in presence of EGTA buffer

A. Case-I [EGTA]= 50 μM, B. Case-I [EGTA]= 100 μM , C. Case-II [EGTA]= 50 μM, D. Case-II [EGTA]= 100 μM.

The effect of presence of EGTA buffer concentration on calcium concentration patterns w.r.t space at time $t=2,3,4,10,50,100$ msec is shown in figure 2. The calcium concentration is noticed maximum at node 1. It reduces sharply in the nearby regions, afterwards it attains basal level of calcium concentration towards other side of cell. From “**Fig.2 A and C**”, it can be seen that, the nodal calcium concentration is slightly maximum in case-II where $K_p \ll C_{cy}$ than in case-I where $K_p \gg C_{cy}$. This is due to the fact that, calcium concentration decreases with higher values of pump rate constant. The effect of increase in concentration of EGTA buffer from 50 μM to 100 μM is shown in “**Fig.2 A and B**”in case-I. At each time step nodal calcium concentration observed decreasing in presence of 100 μMEGTA buffer than 50 μM EGTA buffer. Same effect can be seen for case-II. Thus concentration of EGTA buffer plays important role in decreasing cytosolic calcium concentration.

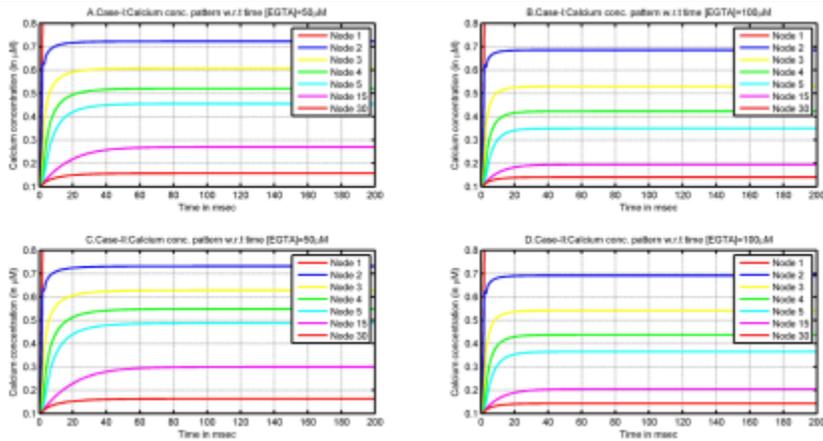


Figure3. Calcium concentration patterns w.r.t time in presence of EGTA buffer
 A. Case-I [EGTA]= 50 μM , B. Case-I [EGTA]= 100 μM , C. Case-II [EGTA]= 50 μM , D. Case-II [EGTA]= 100 μM

Figure 3 shows, calcium concentration patterns w.r.t time at node number 1, 2, 3, 4, 5, 15, 30 for case-I and case-II under consideration in occurrence of EGTA buffer. The calcium concentration elevate sharply in the initial span of time as soon as source channel opens. After that it almost attains steady state concentration at each node under consideration. The steady state concentrations are observed slightly maximum in case-II as observed in “Fig.3 A and C” respectively. The increase in concentration of EGTA from 50 μM to 100 μM causes decrease in calcium concentration as observed in case-I (In “Fig.3 A and B” respectively) and in case-II (In “Fig.3 C and D” respectively).

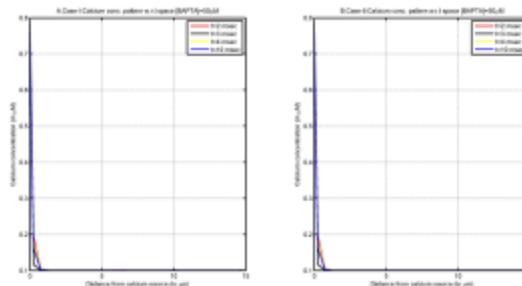


Figure4. Calcium concentration patterns w.r.t space in presence of 50 μM BAPTA buffer for A. Case-I, B. Case-II.

The occurrence of fast BAPTA buffer with concentration 50 μM affects the calcium concentration strongly. The BAPTA buffer binds with calcium within fraction of msec and reduces calcium concentration sharply as shown in “Fig.4 A and B” for case-I and case-II respectively. Due to high binding capacity of BAPTA buffer the calcium concentration remains tightly controlled. The pump rate constant has not shown any effect on calcium concentration in occurrence of BAPTA buffer.

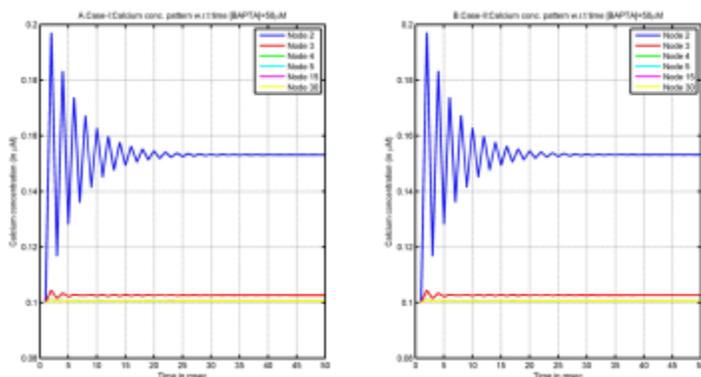


Figure5. Calcium concentration patterns w.r.t time in presence of 50 μM BAPTA buffer for A. Case-I, B. Case-II

The oscillations of calcium concentration patterns are observed in occurrence of BAPTA buffer in both cases under consideration as shown in “**Fig.5 A and B**”. The amplitude of oscillation is maximum at node 2, which is very close to source channel. This is because of the interplay between calcium and BAPTA buffer. As soon as calcium release in cytosol calcium concentration increases, soon after BAPTA binds with calcium and lowers its concentration. Therefore very few amount of calcium able to reach node 3 and further nodes.

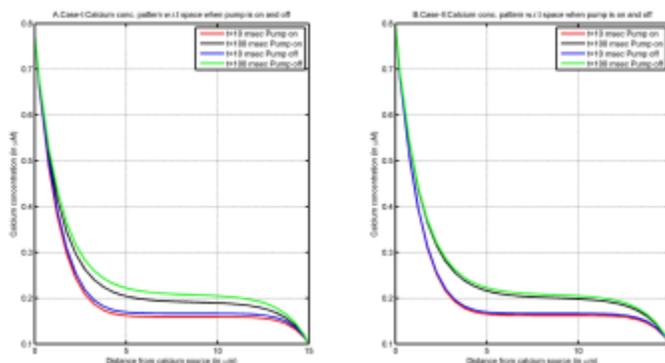


Figure6. Calcium concentration patterns w.r.t space when leak is on or off for A. Case-I, B. Case-II

The nodal calcium concentration patterns when leak is on or off is plotted at time $t=10$ and 100 msec for case-I and case-II in “**Fig.6 A and B**”. respectively. The calcium concentration increases slightly when leak is on in the region near to source. It remains almost constant away from source.

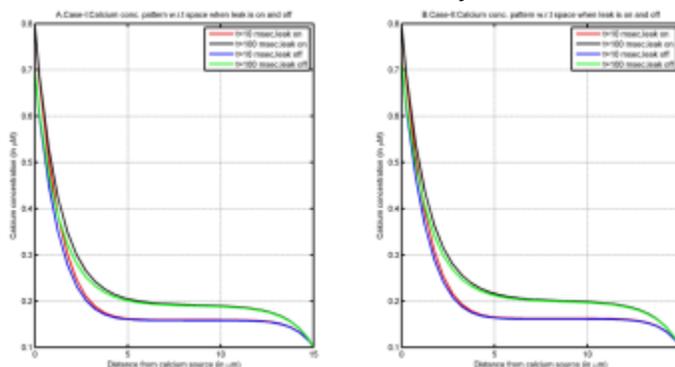


Figure2. Calcium concentration patterns w.r.t space when pump is on or off for A. Case-I, B. Case-II

The SERCA pumps actively collect calcium back into ER. The calcium concentration patterns when pump is on or off at time $t=10$ and 100 msec is plotted in “**Fig.7 A and B**”. respectively for case-I and case-II. When pump is not active, only buffers are working to lowers the calcium concentration, the calcium concentration is observed increasing in the region away from source. The gap in calcium concentration with and without pump is maximum at $t=100$ msec.

V. CONCLUSION

The one dimensional unsteady state case calcium dynamics in liver cell is successfully studied by employing finite volume method. The calcium concentration patterns are affected by occurrence of type of buffer and concentration of buffer, influx and out flux of calcium from source channel to cytosol. From the study it is concluded that, the buffers and pump lowers the calcium concentration in cytosol of liver cell and avoid increase in calcium level beyond the normal range which is hazardous to survival of cell. When the calcium concentration reaches basal level of concentration, source channels of calcium opens to supply the required amount of calcium to cell. Thus, cyclic movement of calcium from ER to cytosol and vice versa helps to maintain chemical homeostasis in liver cell, necessary for proper functioning of cell. Thus, such mathematical models can be developed to analyze the real life problems.

References

1. Dupont, G. et.al. 2000.Hierarchical organization of calcium signals in hepatocytes: from experiments to models.Biochimica et BiophysicaActa (BBA)-Molecular Cell Research,1498(2):134-152.
2. Amaya, M. and Nathanson, M.2013. Calcium signaling in the liver.Comprehensive Physiology.

3. Thomas, A.P, Renard, D.C and Rooney, T.A. 1991.Spatial and temporal organization of calcium signalling in hepatocytes. *Cell calcium*. 12(2-3):111-126.
4. Jafri, M.S. and Keizer, J. 1995. On the roles of Ca^{2+} diffusion, Ca^{2+} -buffers, and the endoplasmic reticulum in IP_3 -induced Ca^{2+} waves. *Biophysical journal*. 69(5): 2139-2153.
5. Fall, C. 2002. *Computational Cell Biology, Interdisciplinary Applied Mathematics*. V.
6. Tewari, S. and Pardasani, K.R. 2010. Finite element model to study two dimensional unsteady state cytosolic calcium diffusion in presence of excess buffers. *IAENG International Journal of Applied Mathematics*. 40(3):108-112.
7. Jha, A. and Adlakha, N. 2015. Two-dimensional finite element model to study unsteady state Ca^{2+} diffusion in neuron involving ER LEAK and SERCA. *International Journal of Biomathematics*. 8(1):1550002.
8. Jha, B, Adlakha, N. and Mehta, M. 2014. Two-dimensional finite element model to study calcium distribution in astrocytes in presence of excess buffer. *International Journal of Biomathematics*. 7(3):1450031.
9. Jha, B., Adlakha, N. and Mehta, M. 2013. Two-dimensional finite element model to study calcium distribution in astrocytes in presence of $vgcc$ and excess buffer. *International Journal of Modeling, Simulation, and Scientific Computing*. 4(2):1250030.
10. Pathak, K. and Adlakha, N. 2015. Finite element model to study calcium signalling in cardiac myocytes involving pump, leak and excess buffer. *Journal of Medical Imaging and Health Informatics*. 5(4):683-688.
11. Pathak, K. and Adlakha, N. 2016. Finite element model to study two dimensional unsteady state calcium distribution in cardiac myocytes. *Alexandria Journal of Medicine*. 52(3):261-268.
12. Kotwani, M., Adlakha, N. and Mehta, M.N. 2014. Finite element model to study the effect of buffers, source amplitude and source geometry on spatio-temporal calcium distribution in fibroblast cell. *Journal of Medical Imaging and Health Informatics*. 4(6):840-847.
13. Kotwani, M., and Adlakha, N. 2017. Modeling of endoplasmic reticulum and plasma membrane Ca^{2+} uptake and release fluxes with excess buffer approximation (EBA) in fibroblast cell. *International Journal of Computational Materials Science and Engineering*. 6(1):1750004.
14. Manhas, N. and Pardasani, K.R. 2014. Mathematical model to study IP_3 dynamics dependent calcium oscillations in pancreatic acinar cells. *Journal of Medical Imaging and Health Informatics*. 4(6):874-880.
15. Manhas, N., Sneyd, J. and Pardasani, K.R. 2014. Modelling the transition from simple to complex Ca^{2+} oscillations in pancreatic acinar cells. *Journal of biosciences*. 39(3):463-484.
16. Naik, P. and Pardasani, K.R. 2016. Finite element model to study calcium distribution in Oocytes involving voltage gated Ca^{2+} channel, ryanodine receptor and buffers. *Alexandria Journal of Medicine*. 52(1):43-49.
17. Panday, S. and Pardasani, K.R. 2013. Finite element model to study effect of advection diffusion and Na^+/Ca^{2+} exchanger on Ca^{2+} distribution in Oocytes. *Journal of medical imaging and health informatics*. 3(3):374-379.
18. Jagtap, Y. and Adlakha, N. 2018. Finite volume simulation of two dimensional calcium dynamics in a hepatocyte cell involving buffers and fluxes. *Communications in Mathematical Biology and Neuroscience*. 2018(15):1-16.
19. Jagtap, Y. and Adlakha, N. 2018. Simulation of buffered advection diffusion of calcium in a hepatocyte cell. *Mathematical Biology and Bioinformatics*. 13(2):609-619.
20. Versteeg, H.K. and Malalasekera, W. 1995. *An introduction to computational Fluid Dynamics, The finite volume control*. Longman, Londres.

GENERATION OF 3D RETINAL MODEL FROM OPTICAL COHERENCE TOMOGRAPHY IMAGES

¹Ms.Mrinmoyee Mukherjee, ²Jerome Judith, ³Tabassum Chohan, ⁴Prasad Chavan, ⁵Camilla Dabre

¹Assistant Professor, ²Student, ³Student, ⁴Student. ⁵Student

⁵Department of Information Technology,

⁵St. Francis Institute of Technology, Mumbai, India-401103

ABSTRACT: Human Eye is the most intricate sensory organ comprising of a multitude of substructures. Retina, a mere set of tissues and one third of a millimeter in width, consists 10 layers. The distortion in these retinal layers account for many diseases related to vision such as retinopathy, age related macular degeneration, eye cancer, retinal detachment, retinal degenerations and inherited. Two important type of imaging techniques which are being used for analysis of retina are and Optical coherence tomography (OCT) and digital fundus images. OCT scan means extracting the information about the inner retinal layers non-invasively. OCT scans provided a video with multiple frames each of which have a close similarity to a slice of eye (retina). Purpose of this project is to make a 3D model of retinal layers captured through OCT scans to anticipate a better picturing of retina to ophthalmologists. There are two main parts of project which include image processing module and graphics. The frames gathered from OCT scans are processed using image processing techniques to extract the details of the layers in the form of coordinates. Immediately after the coordinates of all the points are extracted a point cloud is formed. Using this point cloud, regeneration of a three dimensional model using these points is done.

Keywords: Medical Image Processing, Computational modeling, Retinal Prosthesis Modeling, OCT report, Patients report.

I. Introduction

Optical Coherence Tomography(OCT) is a non-invasive imaging test. OCT uses light waves to get cross-section images of your retina. OCT is a commercially available computer-assisted optical instrument that generates cross-sectional images (tomograms) of ocular structures with close to 10mm axial resolution three-dimensional images from within optical scattering media (biological tissue). In ophthalmology – Cross-sectional visualization is an extremely powerful tool in the identification and assessment of retina abnormalities. Resolving power of 10mm – (Time Domain) 5mm – (Spectral Domain) is able to provide detail for evaluating the vitreo-retinal interface, neurosensory retinal morphology, and the RPE-choroid complex. Grey-scale images are qualitatively superior to the color-scale images on high-resolution spectral OCT.[2]

The optical coherence tomography (OCT) imaging creates high-resolution cross-sectional scans of the human retina. This technique has been widely employed by clinicians for thickness measurements of different retinal layers to identify and assess the retinal abnormalities. OCT is also capable of creating 3D data sets of retinal structure based on acquiring images consecutively at different depths in the retina. In the analysis of the retinal OCT images, identification of the different retinal layers is important for assessment of retinal features or diagnosis of various pathologies.

Optical Coherence Tomography (OCT) technology is widely used in medical image applications. For example, the region between the Epithelium (EP) layer and the Lamina Propria(LP) layer is used to identify dysplastic cells in oral cancer OCT images.

II. PROBLEM DEFINITION AND AREA

Epiretinal Membranes (ERMs) most often occur in people over age fifty. It is estimated that at least 2% of people over 50 years old and 20% over age 75 have ERMs. Up to 20% of people with ERMs have them in both eyes, but symptoms and severity for each eye differ. So generation of a 3D retinal model using OCT images and developing a new OCT-based method is done for investigation of ERM pathology in human eyes. In India, there is a shortage of 1,27,000 eye doctors due to which 45% of patients suffer vision loss before diagnosis. [4] The number of ophthalmologists in India is 11,441 per million, whereas our population is 1.3

billion. Even when available, ophthalmologists in India are inconsistent. Hence a dire need for an automated system that would exhibit more precision in diagnosis has risen.

Layers of retina

2.1 Nerve Fiber Layer (NFL): The NFL layer is generated by the increase of the fibers of the optic nerve and its thickest near the optic disc that gradually diminishing toward the ora serrata.

2.2 Inner Plexiform Layer and Ganglion Cell Layer (IPL+GCL): consists of synaptic connections between the axons of bipolar cells and dendrites of ganglion cells.

2.3 Inner Nuclear Layer (INL): consists of the cell bodies of horizontal cells, bipolar cells, amacrine cells, interplexiform neurons, Müller cells, and sometimes displaced ganglion cells

2.4 Outer Plexiform Layer (OPL): It is a wide and also external band composed of inner fibers of rods and cones and a narrower inner band consisting of synapses between photoreceptor cells and cells from the inner nuclear layer.

2.5 Outer Nuclear Layer and Photoreceptor Inner Segments (ONL+PIS): one of the layers of the vertebrate retina, the light-detecting portion of the eye.

III. RESEARCH METHODOLOGY

3.1 Research Approach

The raw data were collected at the Chennai Eye Clinic, using a Stratus OCT (Carl Zeiss-Meditec, Inc.) instrument. Scanning protocols included a volumetric scan of the macula centered on the fovea with a scan area of 6.45×4.81 mm. The axial resolution of the images was $3 \mu\text{m}$, and the distance between the images was $\sim 140 \mu\text{m}$. These images were then used to create realistic in-silica meshes, on which retinal electrical propagation simulations were performed using finite element modeling (FEM). These anatomically correct computational scaffolds are constructed.

1.2 Working of the system

The system comprises of several steps such as segmentation, creating a nodal point cloud, creating a mesh and smoothing it.

3.2.1 Segmenting 2D OCT images to extract retinal strata.

The computational algorithm for automatic segmentation of the OCT retinal data sets. This algorithm was developed using the MATLAB programming platform and proved to be capable of accurately delineating the layered structure of the healthy retina.[3] Using the algorithm, it is possible for detecting seven different retinal strata which starting from the posterior of the eye and moving towards the inner vitreous surface, which starting from the posterior of the eye and moving towards the inner vitreous surface were as follows:

1. Choroid/Retinal Pigment Epithelium (RPE) layer.
2. RPE + Inner Segment (IS) + Outer Segment (OS) of Photoreceptors/Outer Nuclear Layer (ONL).
3. ONL/Outer Plexiform Layer (OPL).
4. OPL/Inner Nuclear Layer (INL).
5. INL/Inner Plexiform Layer (IPL) + Ganglion Cell Layer (GCL).
6. IPL + GCL/Nerve Fiber Layer (NFL).
7. Vitreous/Retina

3.2.2 Digitizing the retinal anatomical layers to in-plane nodal points.

A computational model of the retina under electrical stimulation could be created; a 3D representative mesh of the tissue in the form of finite elements was required. It aims at developing a mesh by defining a set of data points along the detected retinal borders and next fitting them to a predefined generic initial mesh. This fitting process ensures that the final constructed mesh reproduces the curvature of the retinal layers accurately. The set of data points are defined by selecting a group of pixels from the segmented continuous boundaries with constant spacing, that is down-sampling the detected borders by a constant rate.[4]

3.2.3 Creating an initial mesh based on extracted nodal points.

In the next step of creating the representative computational mesh of human retina, an initial model was created using tricubic Hermite basis functions based on the 3D data cloud. The cubic Hermite basis functions provide continuity of nodal values and nodal derivatives, making these is a smooth tissue such as the retina.

3.2.4 Smoothing the initial mesh based on acquired 3D data cloud.

The process of fitting the initial mesh to the data cloud consists of minimizing the least squares total distance between data points and the associated 2D retinal layer surface (i.e. minimizing the data projection errors), while the mesh is kept reasonably smoothed. Once the data projections were carried out for each data point, the nodal set was updated to progressively tailor the mesh to the digitized representation of the

clinically obtained data set. This process was carried out iteratively to reach the desired shape and accuracy.[8]

3.2.5 Simulating retinal electrical stimulation.

The above generated OCT-based 3D computational mesh as a scaffold helped to study the effects of two electrode array configurations on retinal electrical activation. Both configurations were located subretinally at the surface, and the methodology of their implementation has been detailed. Briefly, one bipolar electrode set-up and one hexapolar arrangement were implemented, using the demonstrated configuration.

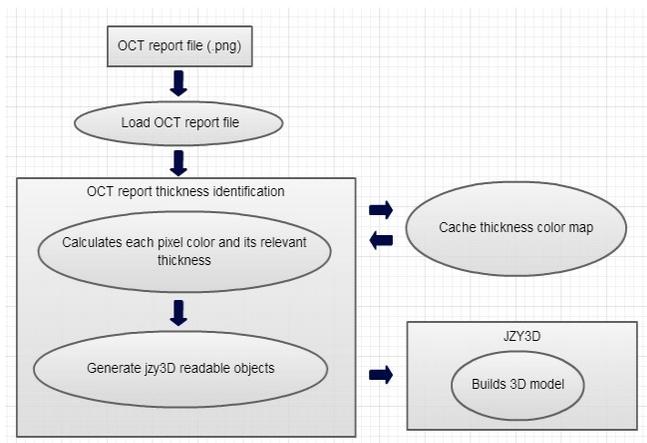


Figure 1: Flowchart for conversion of OCT images to 3D retinal model

Description: The flowchart shows that the proposed system converts the input i.e. OCT images to a 3D retinal model which is the output.

IV. SURVEY

As mentioned in the Research Approach section, the Chennai Eye Clinic was approached regarding the details of the working of current system. The Doctors at Chennai Eye Clinic provided enough information about the working of current system and the problems encountered in it.

4.1 Literature Survey

Sr no.	Title of the paper	Abstract	Positives	Negatives
1	A Comparative Study on Macular Optical Coherence Tomography Image Alignment	<ul style="list-style-type: none"> Macular OCT image alignment. 10 combinations of metric and transform in image registration. 	The combinations of rigid transform and SSD (Sum of Squared Differences) and NCC (Normalized Correlation Coefficient) metric achieve better results.	No shape prior is included in the alignment.
2.	Optical Coherence Tomography Image Segmentation	Segmentation of retinal layers in OCT images. It removes speckle noise, connects discontinuous layer boundaries, as well as enhances/sharpens weak image features	Capable of segmenting all the retinal layers that can be seen by human eyes.	
3.	Region segmentation in 3-D optical coherence tomography images	System uses mean value and a 3D filter-based-fuzzy-c-mean algorithm to cluster pixels in 3D OCT images and find the edge between different clustered regions. Segmented regions of	Uses 3D FBFCM algorithm to enhance 3-D spatial information for cluster processing. Allows low-	System fails to enhance spatial information when two conjunctive OCT images are far

		interest are reconstructed via AVIZO for 3D display purposes.	resolution OCT images to give better segmentation results.	apart during the scanning process.
4.	Advancement in OCT and Image Processing Techniques for Automated Ophthalmic Diagnosis	Discusses the basic functioning principle of OCT and various advancements that it has gone through starting from its invention by Huang et al. in 1991.	Discusses working of OCT system for retinal imaging and. An overview of its clinical use in macular disease diagnosis and monitoring the treatment is presented and various image analysis techniques developed.	De-noising techniques causes loss of information carrying data. Also, increased pre-processing steps tend to decrease computational efficiency.
5.	Retinal Thickness Measurements From Optical Coherence Tomography Using a Markov Boundary Model	Generated System for detecting retinal boundaries in optical coherence tomography (OCT) B-scan and quantitative measures of retinal thickness is made from OCT images.	To generate new system to automatically detect the retinal boundaries in optical coherence tomography (OCT) images of the eye.	If Blunders are detected then the edge-growing process for the outer boundary has picked the wrong edge near A-scan and thus reduces the effect of Markov Boundary Model.
6.	Statistical Modeling of Retinal Optical Coherent Tomography	Statistical model is based on nonlinear transform, called as Gaussianization, to convert the probability distribution function (pdf) of each OCT intra-retinal layer to a Gaussian distribution.	Contrast enhancement method and tested on 130 randomly selected slices from thirteen healthy 3D OCTs ,taken by Zeiss Cirrus HD-OCT.	

V. RESULTS AND FINDINGS

5.1 Segmentation

The segmentation process in image processing, aims at converting a digital image into multiple segments with sets of pixels which are also called as super-pixels. The aim of segmentation is to make things easier or change the representation of an image into something that is more meaningful and much easier to analyze. Segmentation of an image is normally used to locate objects and boundaries such as lines, curves in images. More precisely, image segmentation is the process of assigning a label to every pixel in an image in such a way that the pixels with the same label share certain features.

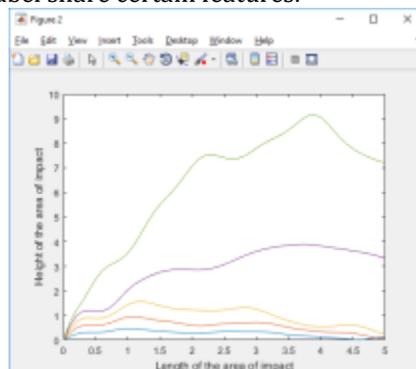


Figure 2: Segmentation & analysis graph

Description: The dataset of OCT images has been analyzed and the variation in the impacted layer has been compared. The parameters considered are the length and height of the area of impact in the OCT image dataset. [3] This segmentation process of the layers helps to find the thickness of individual layers and then the 3D scatter plot further helps in developing the 3D Model for the area of impact.

5.2 Scatter Plot

The 3D scatter plot shows trivariate points plotted in an X-Y-Z axis. It is particularly useful for examining the relationships among these variables. The influence of a categorical variable may be studied by using a different plotting symbol for each value of this variable. Hence, up to four variables (three numerical and one categorical) may be displayed on a single graph. The scatter plot procedure has the ability to rotate the data so that you can study the data from different angles and the ability to plot the multiple regression surface when retreating Y on X and Z using up to two-way multiple regression models.

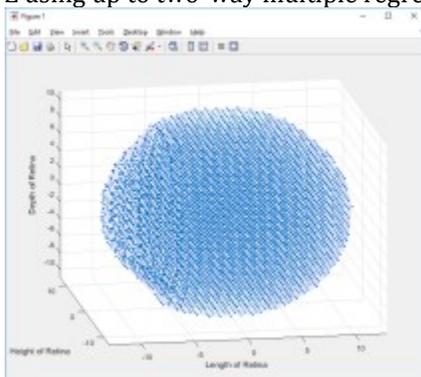
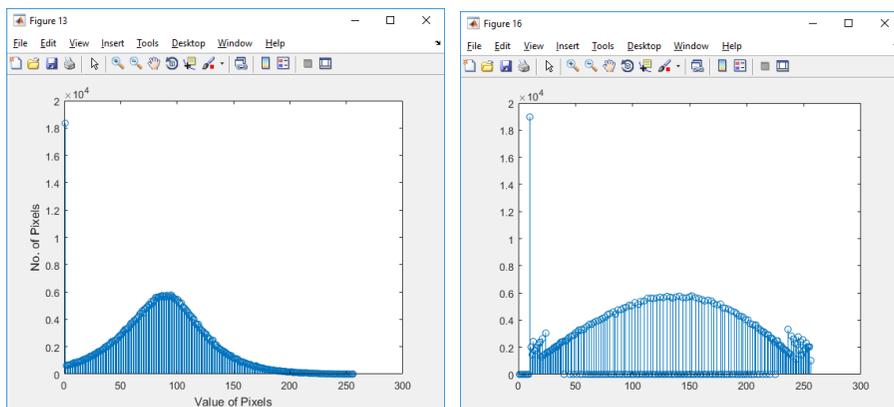


Figure3: Scatter Plot of the dataset

Description: The parameters here on X, Y and Z axis is the length, height and depth of retina and the 3D point cloud is formed to find the area of impact on the retinal layers. This analysis will further help in developing the 3D model of only the impacted retinal layers. The dataset used in this analysis are greyscale OCT images.[7]

5.3 Histogram

Histogram of an image in image processing refers to a histogram of the pixel intensity values. Histogram is widely used in thresholding. It aims at showing graphical representation of the pixel distribution in a digital image that plots the number of pixels for each value. In this, the x-axis represents the gray level intensities and the y-axis represents the frequency of these intensities. Histogram is used to equalize an image.



(Original image)

(Equalized image)

Figure 4: Histogram of the original and equalized image.

Description: The histogram of the first OCT image from the patient dataset shows that the less intensity areas of the layers are enhanced by doing the PDF (Probability Distribution Function) and CDF (Cumulative Distribution Function).

5.4 Image enhancement after histogram

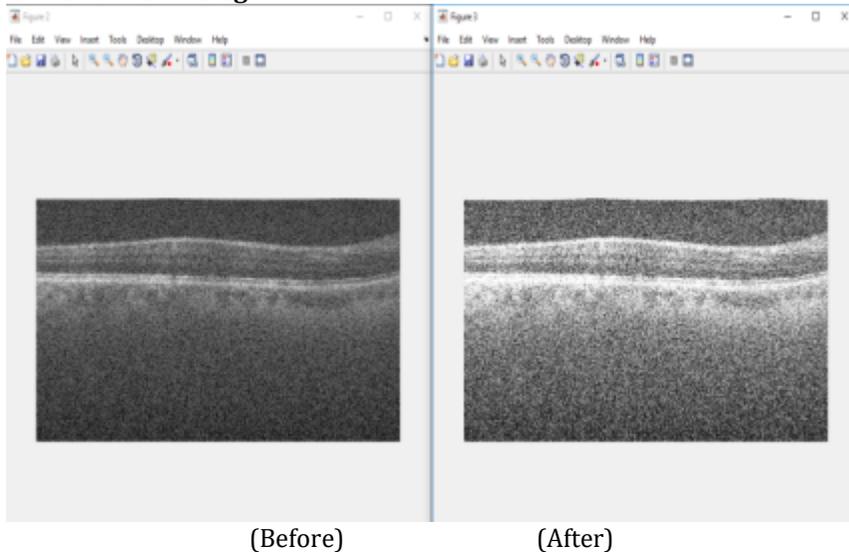


Figure 5: Image enhancement after histogram

Description: The Histogram, PDF and CDF functions have been done on the first set of OCT images and thus the pixel intensity of the retinal layers have increased thus enhancing the dataset. Thus image is clearer to find the distribution of pixels for determining the thickness. These layers can then be put together as a 3D Mesh in MATLAB will give the 3D retinal model. [6]

VI. CONCLUSION

In conclusion, the ophthalmologists will be able to diagnose patients' about their ocular ailments more effectively, and will be efficient in doing so. The model will use advanced image processing technologies to increase performance. Using MATLAB the 10 layers of the retina will be combined to form the 3D model. [2] As we have seen the retina has 10 layers which are scanned individually, and displayed in the Optical Coherence Tomography (OCT) images. The doctors have to just look at the model in order to find the underlying ailment. This will make the diagnosis quicker than the current technology.

VII. ACKNOWLEDGEMENT

We thank Dr. B. Ganesh, Chennai Eye Clinic, Tamil Nadu, for his valuable guidance and insights.

References

1. Rupsa Datta, Aditya, S., & Tibrewala, D. N. (2010). Advancement in OCT and image-processing techniques for automated ophthalmic diagnosis. 2010 IEEE Students Technology Symposium (TechSym). doi:10.1109/techsym.2010.5469183
2. Michael R. Hee, Joseph A. Izatt, Eric A. Swanson, David Huang, Joel S. Schuman, Charles P. Lin, Carmen A. Puliafito, James G. Fujimoto, "Optical Coherence Tomography of the Human Retina", *Arch Ophthalmol*, vol. 113, no. 3, pp. 325-332, 1995.
3. J. Duan, C. Tench, I. Gottlob et al., "Automated Segmentation of Retinal Layers from Optical Coherent Tomography Images Using Geodesic Distance", *Computer Vision and Pattern Recognition*, 2016.
4. A. Yazdanpanah, G. Hamarneh, B. Smith, M. Sarunic, "Intra-retinal Layer Segmentation in Optical Coherence Tomography Using an Active Contour Approach", *Proceedings of the 12th International Conference on Medical Image Computing and Computer-Assisted Intervention: Part II*, Springer-Verlag, vol. 5762, pp. 649-656, 2009.
5. Y. Zheng, R. Xiao, Y. Wang, C. G. James, "A Generative Model for OCT Retinal Layer Segmentation by Integrating Graph-Based Multi-surface Searching and Image Registration", *Proceedings of Medical Image Computing and Computer-Assisted Intervention*, vol. 8149, pp. 428-435, 2013.
6. Jazmin Kay, "How automatic Voter Registration Would Benefit Young People", *Generation Process*, August 24, 2018.
7. A. Hojjatoleslami, M. Avanaki, "Automated 3-D Intraretinal Layer Segmentation of Macular Spectral-Domain Optical Coherence Tomography Images", *IEEE Transactions on Medical Imaging*, vol. 28, no. 9, pp. 1436-1447, 2009.
8. J.J. Kaluzny, A. Szkulmowska, T. Bajraszewski et al., "Retinal imaging by spectral optical coherence tomography", *European Journal of Ophthalmology*, vol. 17, no. 2, pp. 238-245, 2007.

ANDROID BASED BUS TRACKING APPLICATION

Anagha Koroth, Tanmay Gajbhiye, Vishal Shrivastava, Sarvesh Pandey

BE student, BE student, BE student, BE student

Department of IT, Pillai College of Engineering, New Panvel, India

ABSTRACT: Millions of people travelling by public buses waste a lot of time waiting at bus stop and are not aware of the bus timings. We have proposed a Mobile(Android) based Bus Tracking System which overcomes the difficulty faced by the passengers by providing a real time tracking of public buses. This application focuses on tracking of various buses through GPS technology. It will help in better time management for the commuters. It features a one-click SOS button for emergency situations which would notify all the provided contacts of the user's location at the given time. It also includes an E-ticketing feature which make it easier to buy tickets online.

Keywords: Android, bus tracking, GPS;

I. INTRODUCTION

The system is an android application consisting of two applications one for the driver and the another for the user. The application in the driver's mobile phone will update the location every five minutes to the server connecting the user's application and driver's application. The system is basically meant for commuters who daily travel by buses to their destinations. The system includes methods for tracking the real time location of the bus, providing ETA prediction, emergency services for commuters. The ETA prediction module will be based on A* algorithm. In computer science, A* is a computer algorithm that is widely used in pathfinding and graph traversal, which is the process of finding a path between multiple points, called "nodes". It enjoys widespread use due to its performance and accuracy. The SOS service will store the contact number of recipients and send the location of the user to the entered recipients in case of an emergency. The E-ticketing will be based on QR code. The QR code will be linked to an e-wallet. The user needs to scan the QR code and the amount will be deducted from the wallet directly.

I. OBJECTIVE

The main objective of our project is to provide the real time location of public buses to commuters. To achieve this we will be using Google maps api and update the latitude and longitude of the bus with a delay of five minutes.

This will consists of three elements:

1. Driver's application that will provide the location to the server.
2. A server that will store the location of the bus
3. A user application that will display the location of the bus on the map.

II. METHODOLOGY

The Figure 1 shows the block diagram of the proposed system consists of three modules which are Live Location Alert and Payment . The brief descriptions of all the components are as follows:

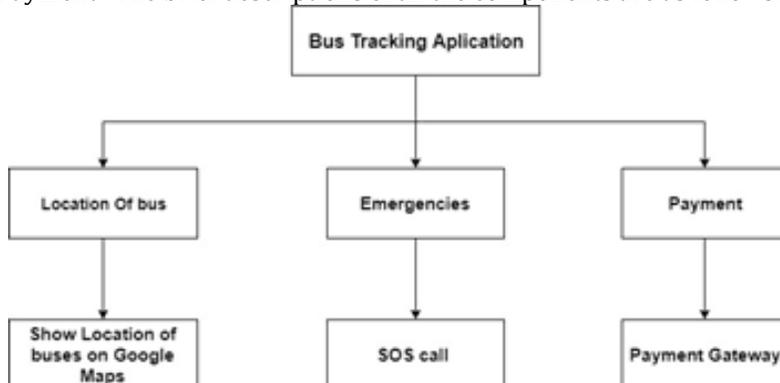


Figure1: Block diagram of proposed system

A. Driver's Application

The driver's application is an android application that sends the current location of the bus. The application works with the help of firebase. This application also allows new drivers and buses to be added. The application requires the name, email id and the number of the bus that is designated to that driver. As soon as the driver logs in through this application; the application starts sending location to the firebase.

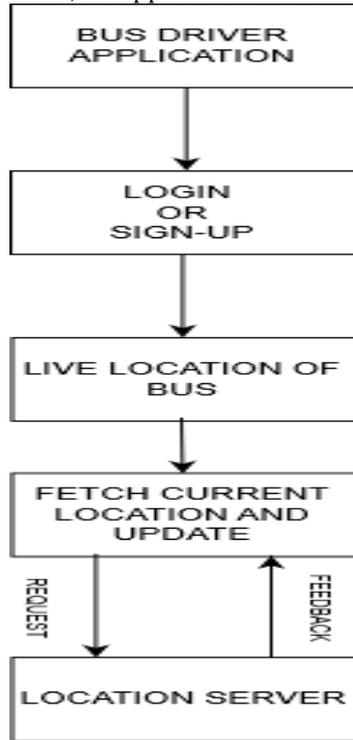


Figure 2: Drivers Application

B. User Application

The user application of the bus sent by the driver's application to firebase and displays it on the map. The user needs to enter the number of bus to be searched and the location of the bus will be displayed on the map

C. E-ticketing

The E-ticketing module allows the user to book tickets. This module requires the user to enter the source, destination and number of tickets. After the user has entered the source, destination and number of tickets the user is required to enter payment details and after validation the user is redirected to payment gateway. All the details and history of ticket booking is stored in this section of the application.

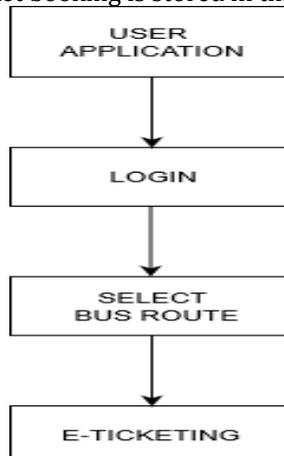


Figure 3 : Flowchart chart - E-Ticketing

D. Messages

The messages module displays the updates sent from the administration office .Any decision taken by administration regarding the buses will be displayed in the messages module.This module works in coordination with firebase.Any changes by administration is sent to the firebase and it is displayed in this module

E. Emergency

The Emergency module works on location services of the Android platform.The application requires the user to enter the emergency contact numbers which will be used to send the location of the user in case of emergency.This module retrieves the location coordinates of the user and sends it to the emergency contacts entered by the user.

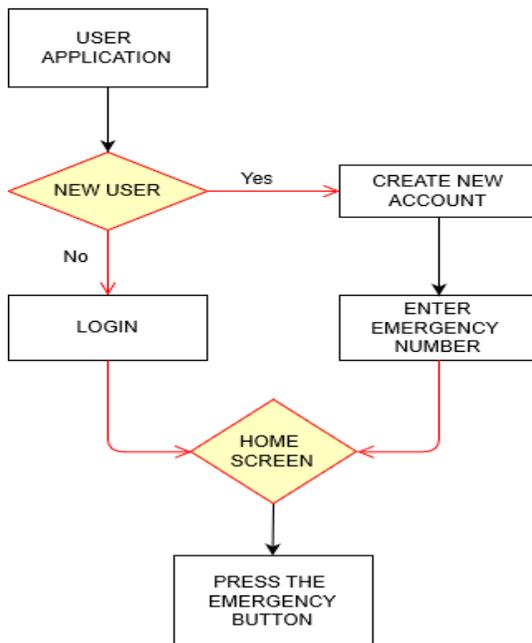


Figure 4 : Flowchart - Emergency module

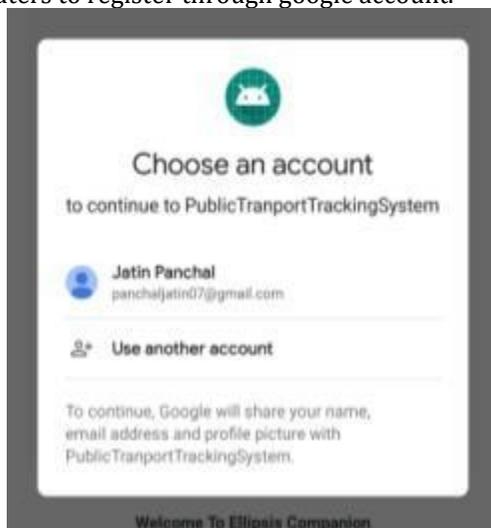
F. Firebase

The Firebase Realtime Database is a cloud-hosted NoSQL database that lets you store and sync data between your users in real time. Cloud Firestore enables you to store, sync and query app data at global scale.

III.RESULT

A. SIGNUP

This features allows new commuters to register through google account.



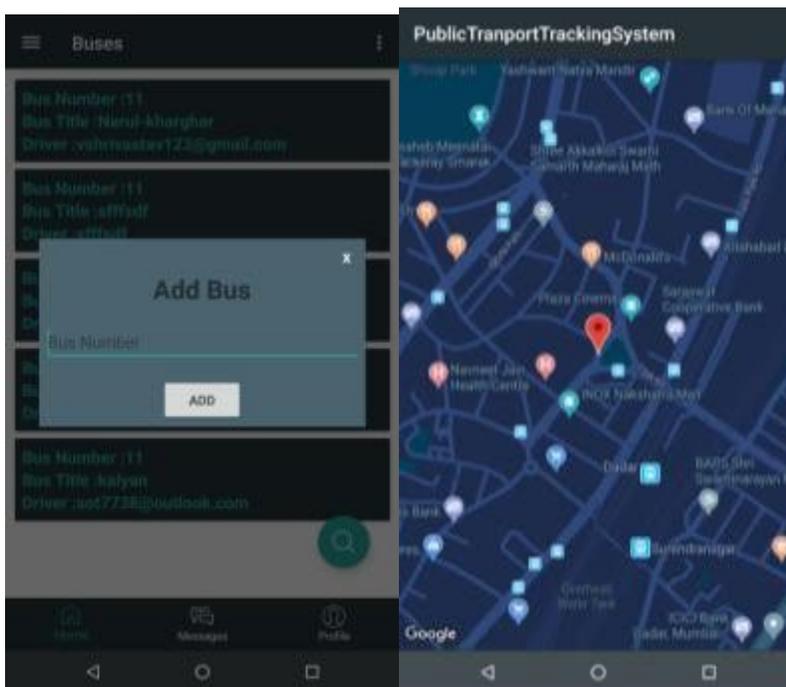
B. HOME

After logging it redirects to the Home screen where we get access to various other features of our app.



C. LOCATION

Provide bus number to the dialogue box appearing on the screen and then live location is shown on the map.



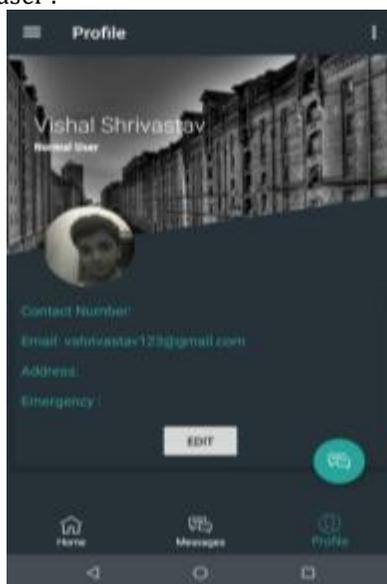
D.E-TICKETING

Provide source and destination and number of passengers .



E. PROFILE AND EMERGENCY

This screen shown the details of the user .



VI. CONCLUSION

The main goal of our android application is to ease the hassle of our daily commuters which they have to face on a regular basis while travelling through the local buses. With the help of our application, commuters can plan their travel accordingly as the location of their desired bus is known, thus saving a lot of time of the commuters which is lost while waiting for the bus on their respective bus stops. The Emergency Module can be of great help, especially to the women passengers travelling around in the local buses, hence providing safety to the commuters.

VII. FUTURE SCOPE

The application can be modified to provide an accurate Estimated Time of Arrival(ETA) of the desired bus searched by the commuter. With the help of accurate ETA prediction, commuters can get better

knowledge about the location and time the bus would take to reach their respective station, thus helping them plan their travel better.

VIII. REFERENCES

- [1] NushrathJahan,KamalHossen and Muhammad KamrulPatwary "Implementation of a Vehicle Tracking System Using Smartphone and SMS service " International Conference on Advances in Electrical Engineering(ICAE), 2017 8th IEEE International.
- [2] Li, Dongjiang, Cheng Cheng, and Bo Zhang."Vehicle Remote Monitoring System based on Android." Software Engineering and Service Science(ICSESS), 2016 7th IEEE International Conference on. IEEE, 2016.
- [3] C.V. Sonia, Dr. A.R. Aswatha, "An Android Application to Locate and Track Mobile Phone", International Journal of Engineering Trends and Technology, vol. 4, no. 5, May 2015.
- [4] Lee, SeokJu,GirmaTewolde, and JaerockKwon."Design and implementation of vehicle tracking system using GPS/GSM/GPRS technology and smartphone application."Internet of Things (WF-IoT), 2014 IEEE World Forum on. IEEE, 2016.
- [5] Alshamsi, Humaid, VetonKëpuska, and HazzaAlshamsi. "Real Time Vehicle Tracking" International Journal of Science and Technology 5.12 (2016).
- [6] P. Verma, J.S. Bhatia, "Design and Development of GPS-GSM Tracking System with Google Map Based Monitoring", International Journal of Computer Science, Engineering and Applications, vol. 3, no. 3, June 2015.
- [7] WiFi Sensing Based Real-time Bus Tracking and Arrival Time, 2018, Rui Zhang, WenpingLiu,YufuJia, Guoyin Jiang, Jing Xing, HongboJiang,Jiangchuan Liu.
- [8] Predicting Bus Arrival Time With Mobile Phone Based Participatory Sensing, 2014, Pengfei Zhou, YuanqingZheng, Mo Li.
- [9] Smart, Public Buses Information System , 2014, Rajesh KannanMegalingam, Nistu Raj, Amal Lehar Soman, Lakshmi Prakash, NivedhaSatheesh, Divya Vijay.

AN IMPROVED PAGERANK ALGORITHM BASED ON LATENT SEMANTIC MODEL AND INTEGRATION WITH N-GRAMS

¹Bharti Khemani, ²Rushikesh Gajmal, ³Viral Patel, ⁴Deval Ghevariya, ⁵ Daanyaal Kapadia

¹ Assistant Professor, ^{2,3,4,5} Student

¹ Computer Department

¹ Rizvi college of Engineering, Mumbai, India

ABSTRACT: The traditional PageRank (PR) just takes into account the web link structure, when distributing rank scores it treats all links equally, which results in topic drift. In this paper, Latent Semantic Model (LSM) and N-Grams is used to calculate the similarity between web pages, and the LSM PageRank (LPR) algorithm is introduced. Our analysis instead uses four gram data generated from an Internet corpus. In short, we validate LSA's dimension reduction optimality property but a full comparison was not possible due to computational constraints. In this algorithm we calculate the value of parent page is distributed to the child on the basis of page similarity between them. The experiment which combines with Nutch shows that the LSMPageRank algorithm performs better than the PageRank algorithm and retrieves better resultset.

Keywords: PageRank; LSMPageRank; Latent semantic model; Topic drift; N-Grams

I. Introduction

It's a great challenge for search engine to retrieve effective information from mass web pages. L. Page and S. Brin propose the PageRank algorithm [1,2] to calculate the importance of the pages. This algorithm has achieved success in Google's business application. But studies show that there are some problems in PageRank, for instance, calculating the PR value of mass pages costs a lot of time and space, and the algorithm results in topic drift.

Now some scholars have proposed effective optimization strategy, [3] proposes the Topic-Sensitive PageRank, this algorithm calculates PR value under different topic and sorts these query results when searching. While the algorithm is based on topic classification, it needs query context to judge the topic classification, without the context it may result in topic drift. [4] is probabilistic combination of link and content information, but the time complexity and space complexity of the algorithm are very large, which makes it difficult to be widely used. In [5], the author studies the link structure and proposes the Weighted PageRank algorithm, when distributing weights it depends on the importance of pages, so famous websites will get higher weights, which makes topic drift more serious.

Vector Space Model is used in [6], where the document is mapped to the vector space, by calculating the cosine value of the document vectors to calculate similarities, then distribute the PR value by weights. But the document is simply mapped to the vector, it ignores the deviation of synonyms and polyseme in different contexts, and the complexity of time and space are very large when the document vector is not on the case of dimension reduction. In Latent Semantic Model, Latent Semantic Indexing (LSI) is based on the principle that words that are used in the same contexts tend to have similar meanings. The important feature of LSI is its ability to extract the conceptual content of a body in text by making associations between that terms that occur in similar contexts [7]. Singular Value Decomposition (SVD) is used to reduce dimensionality, [8] presents a novel algorithm for clustering of web search results based on SVD. [9,10,11] suggest that LSM which contains LSI and SVD dimension reduction could calculate similarities between documents. In this paper, we propose an improved PageRank algorithm based on Latent Semantic Model.

In the next section, we give a brief review of PageRank. Section III presents LSM, and how we use it for calculating similarities. The improved PageRank algorithm based on LSM is described in Section IV. LSM experiment and result are given in Section V. Experiment results and analysis are given in Section VI. And the VII module consist of integration with N-gram. Conclusions and future works of the given in Section VIII.

II. REVIEW OF PAGE RANK

The PageRank algorithm is most widely used page ranking algorithms, states that if a page has important links to it, its links to other pages also become important[2, 12, 13, 14]. The Web-Graph model is considered as a directed graph G , and the pages are G 's vertices, links are G 's edges. The idea of PageRank is that if page u has a link to page v , then the author of u is implicitly conferring some importance to page v . The PageRank algorithm is defined as:

$$\forall v PR(v) = c \sum_{u \in I_v} PR(u) / O_u + (1-c) \tag{1}$$

Where O_u is the outdegree of page u , I_v represents the set of pages pointing to v , $PR(v)$ and $PR(u)$ are rank scores of page v and u . We construct the matrix A through graph G , then calculate all page rank scores through the iteration convergence of A [15]. In order to make the algorithm convergence, we must ensure that G is a strongly connected digraph, some pages don't have outgoing links or the users not follow the links to visit web pages, so we introduce a factor c , c is dampening factor that is usually set to 0.85[13].

III. LATENT SEMANTIC MODEL

In Latent Semantic Model, document set is mapped into a frequency matrix $M (M = (m_{ij})_{m \times n})$, which the rows represent terms and columns represent documents. Then there exists a factorization of the form:

$$M = U \Sigma V^T \tag{2}$$

Where U is in formula is an m -by- n unitary matrix where the matrix Σ is in formula is m -by- n diagonal matrix which nonnegative real numbers on the diagonal, and also V^T is in formula denotes the conjugate transpose of V , an n -by- n unitary matrix. Such a factorization is in formula called a singular-value decomposition of M . And we should reconstruct a matrix $M_k (rank(M_k) = k, k \leq rank(M))$, making the value of $\|M - M_k\|$ be minimum. Fig. 1 shows the construction process of M_k :

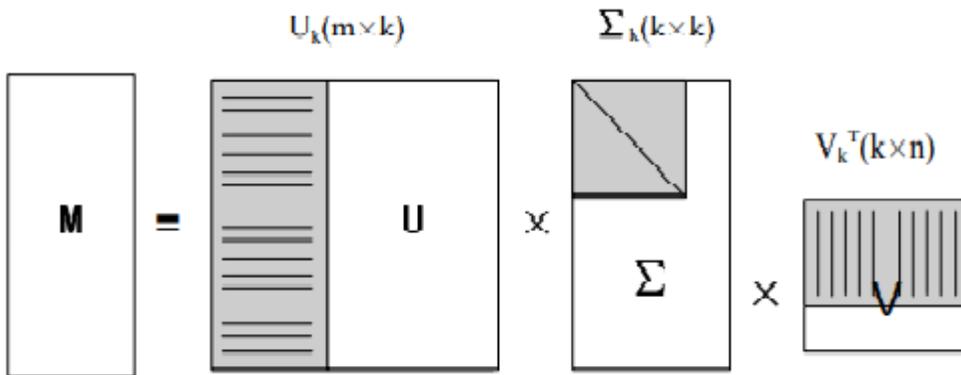


fig 1 . the construction process of M_k

So for LSM truncated SVD is used:

$$M = \begin{matrix} U & \Sigma & V^T \\ k & k & k \end{matrix} \tag{3}$$

where U_k is $m \times k$ matrix whose columns are first k left singular vectors of M , Σ_k is $k \times k$ diagonal matrix whose diagonal is formed by k leading singular value of M , V_k is $n \times k$ matrix whose columns are first k right singular vectors of M . And rows of U_k represent terms, rows of V_k represent documents. Because documents have been expressed on vectorization in Latent Semantic Model, we can calculate the similarity between two documents through the cosine-theorem[10].

IV. LSM PAGE RANK ALGORITHM

In PageRank, the rank scores of the parent page are shared by child pages equally, the pages get the same PR value, which result in unrelated pages. So how to retrieve better rank results and improve user satisfaction are the main problems we are facing. Based on the idea, documents are mapped to latent semantic space, and SVD is used to reduce dimensionality of semantic space. We calculate similarities between documents effectively through eliminating the deviation of synonyms and polyseme, besides, make use of web link structure to reassign the LPR value by document similarity weights. The LPR is defined as:

$$\forall v LPR(v) = c \sum_{u \in O_v} LPR(u) \cdot \alpha + (1-c) \tag{4}$$

Where LPR(u) represents the LPR value of u, α indicates the proportion of Sim(u,v) in sum of the similarities between page u and the pages which u links to. So we define α as:

$$\alpha = \frac{\sum_{v \in I_u} Sim(u,v)}{\sum_{p \in O_u} Sim(u,p)} \tag{5}$$

In (4) and (5), I_v represents the set of pages link to v, O_u represents the set of pages which u links to. Let $N = M_k$, and Sim(u,v) indicates the similarity between page u and v, Sim(u, v) is defined as follows:

$$Sim(u, v) = \frac{\sum_t n_{tu} n_{tv}}{\sum_t n_{tu}^2 \sum_t n_{tv}^2} \tag{6}$$

In semantic space model, the way to calculate the value of elements plays a vital role in LSMPageRank algorithm. In this paper, the classic IF-IDF algorithm is used to calculate the value, for example: $n_{ij} = TF_{ij} \times IDF_i$. And n_{ij} is defined as follows:

$$n_{ij} = TF_{ij} \times IDF_i = TF_{ij} \times \frac{|D|}{\sum_{i \in d} 1} \tag{7}$$

Where TF_{ij} is the term frequency, IDF_i is the inverse document frequency, and |D| represents the number of documents in D, d is a document of D, i represents term.

V. THE LSA EXPERIMENT AND RESULTS

The word totals were tabulated (words that appeared in only one context were discarded). Therefore for each context, we can generate a vector $x \cdot j \in N 60,768$. For ease of analysis, the counts in each entry were transformed via:

$$\begin{aligned} x'_{ij} &= \ln(1 + x_{ij}) \text{entropy}(x_{i \cdot}) \\ &= \ln(1 + x_{ij}) \sum p_l \\ &= 1 P(x_{il}) \log_2(P(x_{il})) \\ &= \ln(1 + x_{ij}) \sum p_l \\ &= 1 \sum x_{il} p_k \\ &= 1 x_{ik} \log_2(\sum x_{il} p_k = 1 x_{ik}) \end{aligned} \tag{8}$$

To represent all the data, we can concatenate each column together and we define the data matrix as $X = [x^1 \cdot 1 \cdot \dots \cdot x^p]$ where the $n = 60,768$ rows represent unique words and the $p = 30,473$ columns are the contexts. To find the salient dimensions of the data, the authors used singular value decomposition (SVD). See appendix A.1 for a review of the mathematics. Now, they “shrink” the space of experience following the method of least squares approximation which entails picking the first $d \ll p$ most salient dimensions. Denote the approximated matrix $X^{(d)}$.

There are many ways to compare vectors but they want to ensure that there is no effects due to the raw frequencies of the words (i.e. their magnitude). This is akin to representing the topic rather than how much is said about the topic. In information retrieval applications (as we have seen in class), the cosine measure works well, which was their choice. The best guess (g^*) to a synonym question would be to take the smallest angle between the vectors i.e. the largest cosine which can be denoted formally:

$$\begin{aligned}
 g^* &= \arg \min_{g \in \{a,b,c,d\}} \{ \theta(x^{(d)} T \cdot, x^{(d)} g \cdot) \} \\
 &= \arg \max_{g \in \{a,b,c,d\}} \{ \cos(\theta(x^{(d)} T \cdot, x^{(d)} g \cdot)) \} \\
 &= \arg \max_{g \in \{a,b,c,d\}} \langle x^{(d)} T \cdot, x^{(d)} g \cdot \rangle \sqrt{\langle x^{(d)} T \cdot, x^{(d)} T \cdot \rangle \langle x^{(d)} g \cdot, x^{(d)} g \cdot \rangle}
 \end{aligned}$$

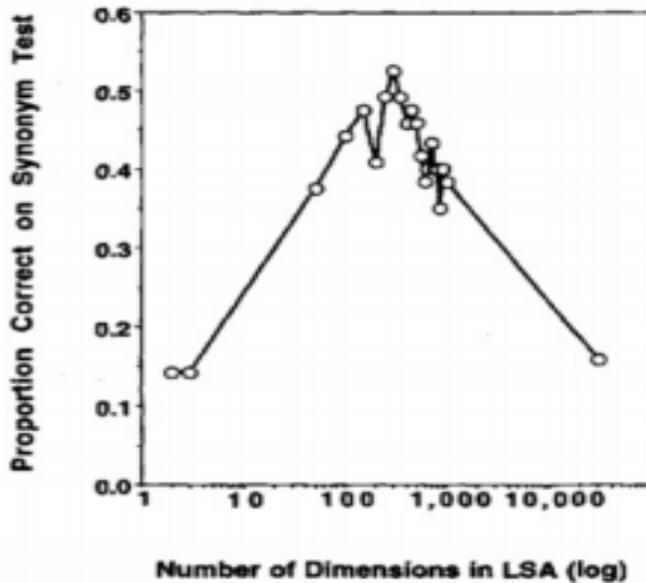


fig 2..proportion correct on synonym test versus number of dimension in lsa(log)

VI. EXPERIMENT RESULTS AND ANALYSIS

The aim of experiment is to compare LSMPageRank and PageRank. Nutch is an excellent open-source search engine[16], and the experiment is based on Nutch-0.9. The experiment includes four sections: Web data collection, Pretreatment, Algorithm and its Implement and Experiment results analysis. And the experiment steps are as follows:

- Web data collection: we crawled 500,000 pages from CNN for our web library named WebBase. When crawling these pages we created the web graph used to calculate the PR value and LPR value.
- Pretreatment: through parsing, filtering, getting rid of useless information, we created the term-document matrix.
- Calculate the LSMPageRank value and PageRank value: With the help of nutch 's plug-in mechanism, we implemented two algorithms and integrated them into nutch-0.9. Then we calculated the LPR value and PR value of every page in WebBase, created indexes for searching.
- Experiment results analysis: firstly, we observed the changes of PR value and LPR value of the same pages, secondly, compared the relevancy of the query results with PageRank and LSMPageRank.

A. Comparison to Previous Study

The crowning achievement of this crude duplication, is that we see a very clear effect of optimum dimensionality. In figure 3, we observe a local maximum at around 50-70 dimensions. This agrees with the result of the original LSA study where a dimension-reduction by a large factor yields optimal results. Note the precipitous drop in performance after $d = 110$. Most likely, if the knowledge space isn't sufficiently

condensed and the interrelations are not exploited, four word contexts offer very, very little information about the meaning of words

% correct on TOEFL synonym test by dimension

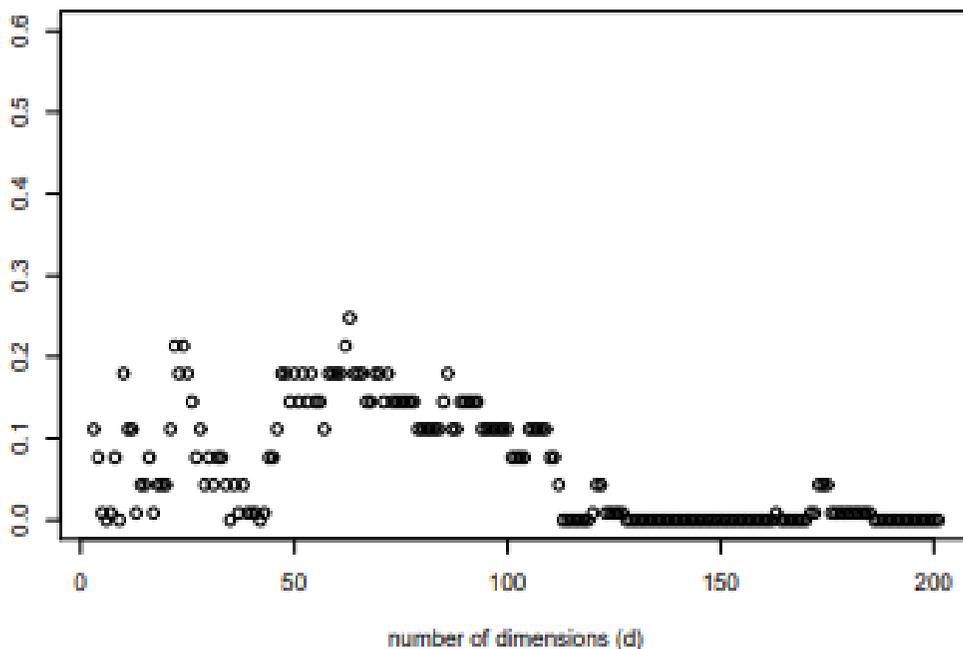


Figure 3: Our results

I still do not have any intuition on if many 4-grams could compete with encyclopedia articles. The more that are included, the better represented the knowledge space becomes. Since we were unable to increase p without computational issues, this was not able to be investigated.

B. Comparison between LPR Value and PR Value

We sorted query results by LPR value of the pages for the query “video”, and fetched the LPR value and PR value of top 50 pages to compare. Fig. 4 shows the result:

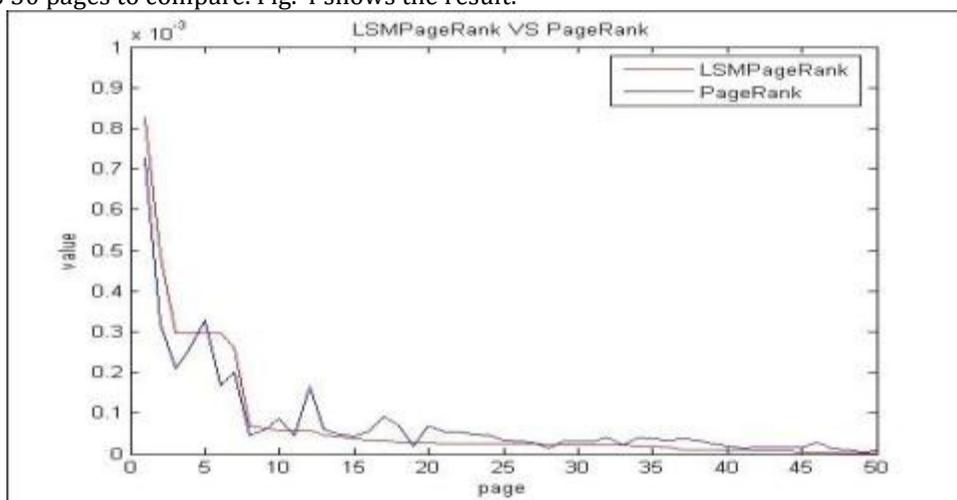


figure 4. the comparison of lpr value and pr value of the same pages

From Fig. 3 we can draw the conclusions:

- The LSMPageRank curve and PageRank curve are a downward trend, but LSMPageRank descends smoothly, PageRank curve appears peaks.

- Checking the peaks of PageRank curve, we found that these pages which have large outdegree and indegree but unrelated to query topic, the weights are relatively higher than LSMPageRank, which shows that LSMPageRank suppress topic drift more efficiency.
- We found there are only 12 of 50 results are related to query topic, the others have nothing to do with the topic. Fig. 4 shows that sum of LPR value of top 10 are higher than PR value, the PR value of others are higher than LPR value. So in LSMPageRank, if the page is related to query topic it will get higher weights, otherwise, it will get lower weights.

C. COMPARISON OF RELEVANCY

We got the results set named R when searching “Cricket”, to reduce the noise resultant from irrelevant pages, we categorized R into four classes based on their relevancy to the given query:

- C1: which contain very important information about the given query;
- C2: which have relevant but not important information about the given query;
- C3: which do not have relevant information about the given query even though they contain the keywords of the given query;
- C4: which include neither the keywords of the given query nor relevant information about it.

And w_p is the weight of page p. The value of w_p for an experiment could be decided through experimental studies. For our experiment, we set the weight and relevancy score as follows:

$$w_p = \begin{cases} 1, & p \in C_1 \\ 0.5, & p \in C_2 \\ 0.1, & p \in C_3 \\ 0, & p \in C_4 \end{cases}$$

$$S = \sum_{i \in R} (n+1-i) \times w_i \tag{10}$$

where i denotes the ith page in the result page-list, n represents the first n pages chosen from the result list, and S is relevancy score.

Table I shows the relevancy scores of the results produced by PageRank and LSMPageRank, and Fig. 5 is the histogram. From tableI and Fig. 5 we can see the LSMPagRank algorithm produces better results for the query “Cricket” than PageRank.

Table I. Relevancy Comparison of two Algorithms

R	C1 + C2		S	
	PageRank	LSMPageRank	PageRank	LSMPageRank
5	2	4	7.2	11.6
10	4	6	20.2	31.6
20	7	8	60.1	80.4
30	11	9	118	142.7
40	16	12	196.3	236.5

60	19	15	389.9	407.8
80	23	20	617.1	642.3
100	25	24	851.4	872.4
120	28	26	1137.9	1146.7

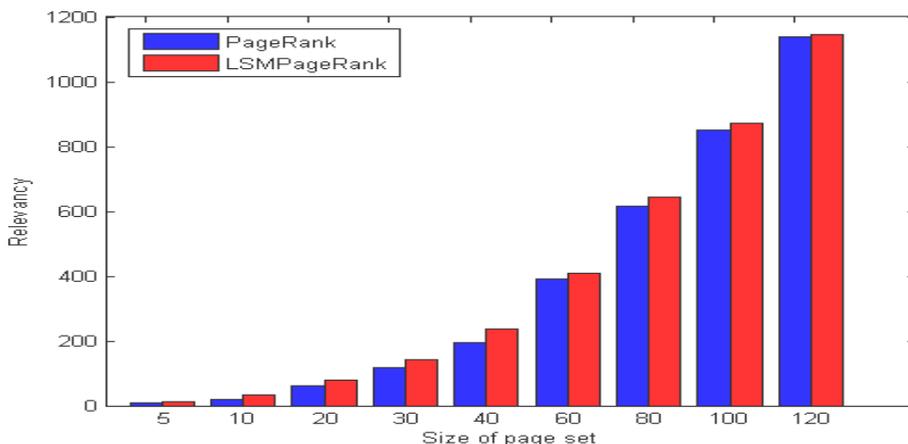


figure 5. the histogram of relevancy comparison of two algorithms

VII. INTEGRATION WITH N-GRAMS

The above provides ti way to handle some of the global constraints in the language. To obtain a multi-span language model, it remains to combine them with local constraints, such as provided by the n-gram paradigm. Obviously, the goal of the resulting integrated approach is to leverage the benefits of both. The integration can occur in a number of ways, such as straightforward interpolation, or within the maximum entropy framework [21]. In the following, we develop an alternative formulation for the combination of the n-gram and L!3A paradigms. The end result, in effect, is a modified n-gram language model incorporating large-span semantic information. To achieve this goal, we need t30 compute:

$$\Pr(w_q|H_{q-1}) = \Pr(w_q|H_{q-1}^{(n)}, H_{q-1}^{(l)}), \tag{11}$$

where the history H_{q-1} now comprises an n-gram component ($H_{q-1}^{(n)} = w_{q-1}w_{q-2} \dots w_{q-n+1}$) as well as an LSA component ($H_{q-1}^{(l)} = \&-I$). This expression can be rewritten as:where the history H_{q-1} now comprises an n-gram component ($H_{q-1}^{(n)} = w_{q-1}w_{q-2} \dots w_{q-n+1}$) as well as an LSA component ($H_{q-1}^{(l)} = \&-I$). This expression can be rewritten as:

$$\Pr(w_q|H_{q-1}) = \frac{\Pr(w_q, H_{q-1}^{(l)}|H_{q-1}^{(n)})}{\sum_{w_i \in \mathcal{V}} \Pr(w_i, H_{q-1}^{(l)}|H_{q-1}^{(n)})}, \tag{12}$$

where the summation in the denominator extends over all words in \mathcal{V} . Expanding and re-arranging, the numerator of (12) is seen to be:where the summation in the denominator extends over all words in \mathcal{V} . Expanding and re-arranging, the numerator of (12) is seen to be:

$$\begin{aligned} \Pr(w_q, H_{q-1}^{(l)}|H_{q-1}^{(n)}) &= \Pr(w_q|H_{q-1}^{(n)}) \Pr(H_{q-1}^{(l)}|w_q, H_{q-1}^{(n)}) \\ &= \Pr(w_q|w_{q-1}w_{q-2} \dots w_{q-n+1}) \\ &\quad \cdot \Pr(\vec{d}_{q-1}|w_qw_{q-1}w_{q-2} \dots w_{q-n+1}). \end{aligned} \tag{13}$$

Now we make the assumption that the probability of the document history given the current word is not affected by the immediate context preceding it. This reflects the fact that, for a given word, different syntactic constructs (immediate context) can be used to carry the same meaning (document history). This is obviously reasonable for content words, and probably does not matter very much for function words. As a result, the integrated probability becomes:

$$\Pr(w_q | H_{q-1}) = \frac{\Pr(w_q | w_{q-1} w_{q-2} \dots w_{q-n+1}) \Pr(\tilde{d}_{q-1} | w_q)}{\sum_{w_i \in V} \Pr(w_i | w_{q-1} w_{q-2} \dots w_{q-n+1}) \Pr(\tilde{d}_{q-1} | w_i)} \quad (14)$$

Note that, if $\Pr(w_i | w_{q-1} w_{q-2} \dots w_{q-n+1})$ is viewed as a prior probability on the current document history, then (14) simply translates the classical Bayesian estimation of the ngram (local) probability using a prior distribution obtained from (global) LSA.

VIII. CONCLUSION AND FUTUREWORK

On the basis of PageRank algorithm analyzing, we recognize that the average transmission strategy of PageRank will result in topic drift. So we combine Latent Semantic Model to propose an improved PageRank algorithm – LSMPageRank. Also Using the 4-grams from the Internet is a convenient way to test a variety of very short contexts. A problem I find to be of interest is optimizing over both length of context and dimensionality of context. This paper implements PageRank and LSMPageRank algorithm and integrates them into Nutch, then calculates the LPR value and PR value of crawled pages off-line. Integration of N-gram approach exploits the complementarity between n-grams, which inherently rely on syntactically-oriented, short-span relationships, and LSA language models, which tend to capture semantically oriented, large span relationships between words. The analyzed results indicate that LSMPageRank produce better query results than PageRank and suppress topic drift more effectively.

But we find that the time complexity of LSMPageRank is more larger than PageRank, so part of our future work is how to improve the algorithmic efficiency. Besides, we would like to consider the possibility of calculating the rank scores by using more than one level of reference pagelist.

REFERENCES

- [1] L. Page and S. Brin . “The anatomy of a large-scale hypertextual Web Search engine”. Proceedings of the Seventh International World Wide Web Conference.1998.
- [2] L. Page , S. Brin, R. Motwani and T. Winograd. “The PageRank citation ranking: Bringing order to the web”. Technical report, Stanford University, Stanford, CA, 1998.
- [3] N. Chomsky. Linguistics and Cognitive Science: Problems and Mysteries.” In A. Kasher (ed.), The Chomskyan Turn. Cambridge, MA: Basil Blackwell. 1991.
- [4] Web 1T 5-gram Corpus Version 1.1, 2006. Google, Inc.
- [5] Grolier. Grolier’s Academic American Encyclopedia. Arte Publishing, 1980.
- [6] Taher H. Haveliwa. “Topic—sensitive pageRank”. Proceedings of the 11th International Conference on World Wide WEB. Honolulu, Hawaii. ACM Press, 2002.
- [7] Matthew Richardson, Pedro Domingos. “The Intelligent Surfer Probabilistic Combination of Link and Content Information in PageRank”. Advances in Neural Information Processing Systems, 2002, 14, 1441-1448.
- [8] Wenpu Xing , Ali Ghorbani. “Weighted PageRank Algorithm”. Communication Networks and Services Research. 2004.
- [9] Gongwei Qian, Lin Ni, Yuan Miao, Rong Cao. “Extended PageRank algorithm based on Web link and content analysis”. Computer Engineering and Applications. 2007, 160-164. (in Chinese)
- [10] Deerwester, S., et al, “Improving Information Retrieval with Latent Semantic Indexing”, Proceedings of the 51st Annual Meeting of the American Society for Information Science 25, 1988, pp.36-40.
- [11] S Osinski, J Stefanowski, D Weiss. “Search Results Clustering Algorithm Based on Singular Value Decomposition”. Intelligent information processing and web mining. 2004.
- [12] Wei Song*, Soon Cheol Park. “Genetic algorithm for text clustering based on latent semantic indexing”. Computers & Mathematics with Applications. Volume 57, Issues 11-12, June 2009, Pages 1901-1907.
- [13] TK Landauer and ST Dumais. A solution to Platos problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. Psychological review, 1 (2):211-240, 1997.
- [14] Michael D. Lee, Brandon Pincombe, Matthew Welsh. “An empirical evaluation of models of text document similarity”. Proceedings of the 27th Annual Conference. 2005.

- [15] Cherukuri Aswani Kumar, Suripeddi Srinivas. "Latent semantic indexing using eigenvalue analysis for efficient information retrieval". Int. J. Appl. Math. Comput. Sci. 2006.
- [16] Amy N. Langville, Carl D. Meyer. "Deeper inside pagerank". Internet Mathematics 2004.
- [17] Arvind Arasu, Junghoo Cho, Hector Garcia-Molia,. "Searching the Web ". ACM transaction on internet technology. 2001.
- [18] Prasad Chebolu and Páll Melsted. "PageRank and The Random Surfer Model". Proceedings of the nineteenth annual ACM-SIAM symposium on Discrete algorithms. 2008.
- [19] Taher H. Haveliwala. "Efficient Computation of PageRank". Technical report, Stanford University, Stanford, CA, 1999.
- [20] R Khare, D Cutting, K Sitaker, A Rifkin. "Nutch: A flexible and scalable open-source web search engine". Oregon State University, 2004.
- [21] R. Rosenfeld, "A Maximum Entropy Approach to Adaptive Statistical Language Modeling," Computer Speech and Language, Vol. 10, London: Academic Press, pp. 187-228, July 1996.

DESIGN AND IMPLEMENTATION OF GAS SENSOR

¹Varindra Kumar & ²Anita Kumari

¹Postdoc Research Fellow, ²Professor and Head

¹Cambridge Graphene Centre, ¹Univ. of Cambridge, Cambridge, UK

²Chemical Engg. Dept. ²Thadomal Sahani Engineering College, Mumbai, India

ABSTRACT: A gas sensor is a rapidly emerging market with gas sensing solutions for CO, CO₂, O₂, NH₃, SO₂, NO, NO₂, H₂ and H₂S etc. where these gases are emitted from a variety of industrial components and are often a hazard to its surroundings. Gas sensors are used in industrial hygiene, safety solutions for its work force across many traditional industries, process control with clean room environment, clinical environment. The output from the sensor should provide a mechanism to tell the gaseous volume and type of gases available in its vicinity while the cost effectiveness, ease of its implementation, reliability plays a crucial role for its widespread application and usage. This paper reports about radio frequency (RF) based gas sensor design and implementation using a single core copper conductive wire in the form of a dual pole antenna. The variation in the relative dielectric constant of various gaseous media may provide a cost-effective mechanism to show the presence of a particular gas while the change in its resonance, radiation and other antenna parameters such as reflection coefficient, gain and efficiency may provide a method to show its volume within an enclosed chamber. The modeling and simulation result of the sensor has been obtained using a 3D field solver tool for various gases such as O₂, CO₂, NO₂ and NH₃ while the results have been plotted using Mat lab. The designed sensor resonates at 1.8 GHz for O₂ filled chamber while an introduction of other gases shifts its resonance to other frequencies. Using the obtained reflection coefficient, the conductivity (real and imaginary part) of the merged media can be obtained. A behavior of its conductivity along with the resonant frequency can be measured using any traditional low frequency measurement solution, providing a low-cost solution mechanism.

Keywords: Gas sensor, radio frequency, dipole, CO₂, O₂

I. INTRODUCTION

The effectiveness of a gas sensing, its recognition and quantification become an early indicator to improve air quality within environment across chemical process industries, automotive industries, smart cities and healthcare applications. Hence a research in chemical gas sensing is one of the most exciting field for its sensitivity, selectivity and repeatability. Although there exist some design solutions in the market with the application of low to high frequency components using traditional electronics design and micro electro mechanical system (MEMS) based gas sensor (Mamishv et al. 2004, Liu, et al. 2012, Azzarelli et al. 2014, Bailly et al. 2016) to detect and monitor various gases, the paper here talks about a high frequency-based gas sensing mechanism. The practical and effective performance of a gas sensing mechanism provides an effective value across a wide sector of market. In addition the gas sensing mechanism has been provided in the past using traditional solutions based mechanism on optical, mechanical, electrochemical, electronic and thermal properties (Barochi et al. 2011, Liu, et al. 2012) however these methods sometimes are either not very effective for its sensitivity, has higher cost (with additional high end instruments and other complex electronics based solution), require additional support mechanism and life expectancy (where the gas sensor lasts for a typical 5 - 10 years life time). With the application of high frequency-based solution and its extraction for impedance profile, the life expectancy of the gas sensor, sensitivity, cost effectiveness can be increased considerably. As the measurement of a gas sensing is highly dependent on the temperature and pressure of the gas itself, hence any parametric evaluation of the sensing mechanism needs to take pressure, temperature into account. The relative dielectric constant of a gas is highly dependent on its temperature and its gaseous state with vapor and liquid phase, hence the temperature dependency of its dielectric constant has also been obtained and shown here. Figs. 1 and 2 show a variation of the relative dielectric constant of CO₂, NO₂, O₂ and NH₃ gases (Itterbeek et al. 1948, Schmidt et al. 2003, Lee et al. 2011) with respect to its temperature. The relative dielectric constant at its ambient temperature 25 °C is used to calculate its resonance behavior. The effect of CO₂ gaseous volume within an enclosed chamber has been also obtained and shown in the paper.

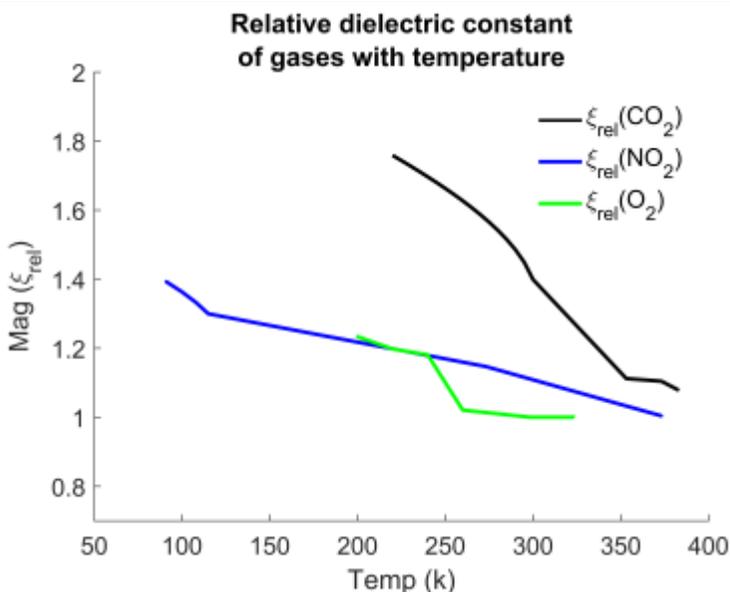


Figure 1: Relative dielectric constant for gases

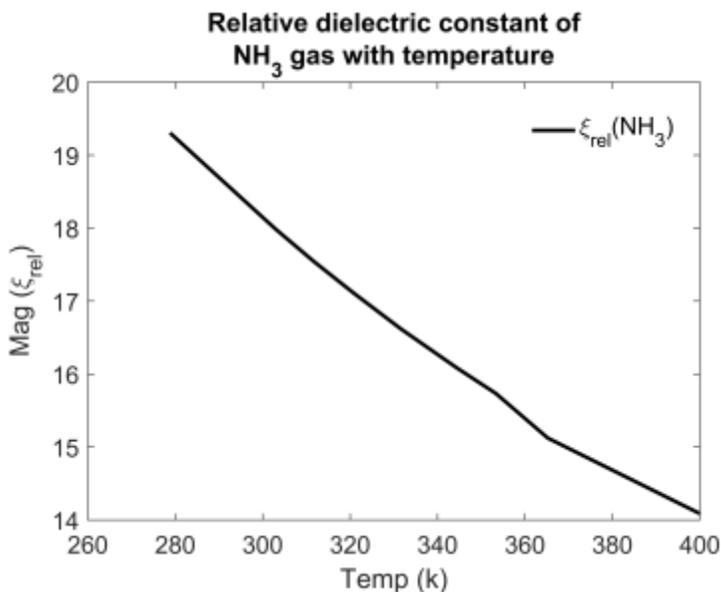


Figure 2: Relative dielectric constant for NH₃ gas

II. DESIGN METHODOLOGY

A copper conductive thin wire of radius 1.5 mm is used to construct the sensor, its other dimension is shown in Fig. 3.

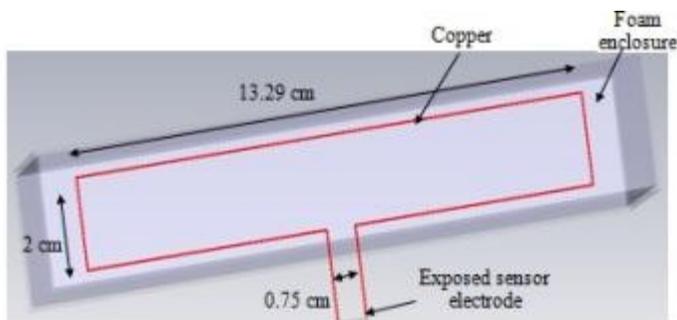


Figure 3: Dipole submerged in an enclosed chamber

The wire is enclosed within a foam chamber while there is a gas flow mechanism with a sealed entry and exit pipes connected to this enclosure. The sensor wire is connected to create a dipole antenna while two ends of the wire are exposed outside the enclosed chamber to excite the dipole within its intended frequency range. The dipole antenna although excites at 1.8 GHz for O₂ filled enclosure, providing a resonance within a thin layered enclosed chamber. The chamber is made of foam with relative dielectric constant of 1.0. After exciting the dipole with a signal frequency range from 1.0 GHz to 2.2 GHz, the resonance frequency shift is observed. As shown in the Fig. 4, the dipole antenna resonates at different frequency range based on the presence of gaseous media within its enclosure while providing a variation in its reflection coefficient.

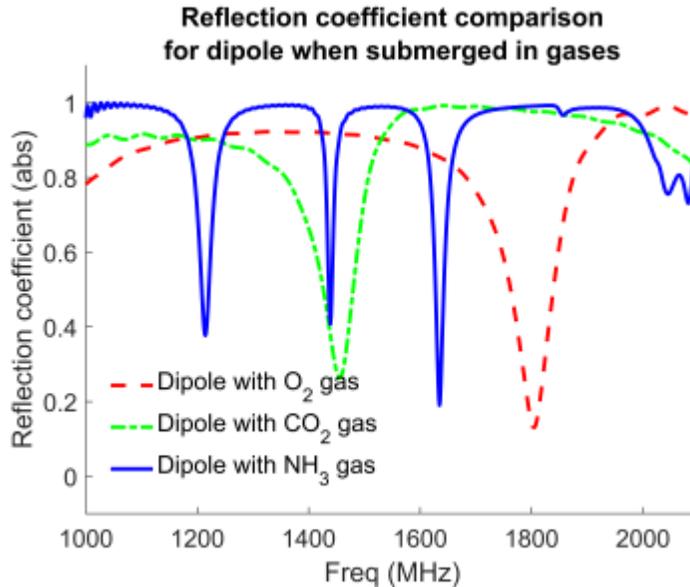


Figure 4: Reflection coefficient and resonance variation when exposed to O₂, CO₂ and NH₃ gases

The resonance and reflection coefficient parameters can be used to obtain the impedance profile of the structure within its gaseous media to provide a measurement value at low frequency. While a 3D field solver tool has been used to obtain its reflection parameter, the comparison has been plotted using Matlab. Similarly Fig.5 shows a comparison of the CO₂ gaseous media when the volume of CO₂ gas is varied within its enclosed chamber.

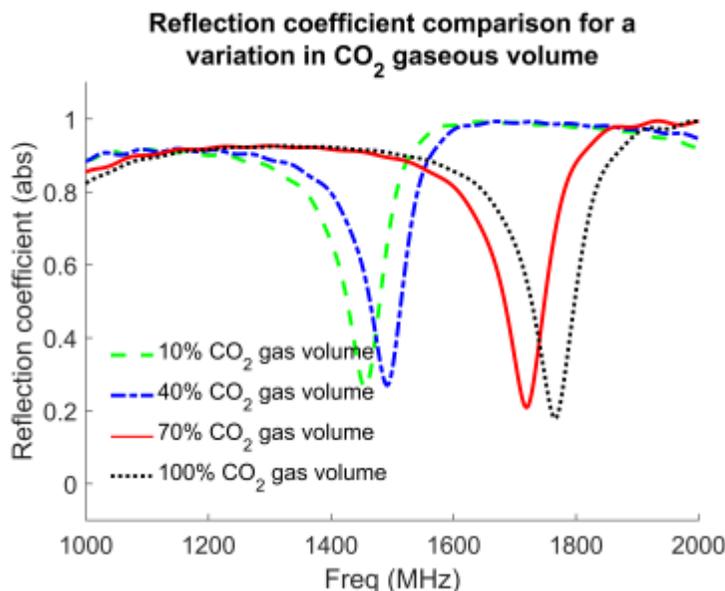


Figure 5: Reflection coefficient for CO₂ with gaseous volume change

III. RESULTS AND DISCUSSION

The designed Dipole antenna when enclosed within a foam enclosure filled with O₂ gas resonates at 1.8 GHz while a fixed volume of CO₂ and NH₃ gas changes its resonance with resonating frequency at 1.625 GHz and 1.43 GHz respectively. The change in its resonance along with the reflection coefficient become important parameters to show a change in impedance profile of the media. With the help of its impedance profile within the frequency range of interest, its resistive, capacitive and inductive profile can be obtained and a simple multi-meter at low frequency can be used to find out the presence of gases in its vicinity. The change in gaseous volume also affects the reflection coefficient and its resonance frequency, thus with the flow of particular gas, its volume in its vicinity can be obtained.

IV. CONCLUSIONS

A simple gas sensing mechanism using a dipole antenna has been designed for the detection of various gases along with the measurement of its volume. Due to its resonance and impedance profile, a low cost multi-meter can be used.

REFERENCES

- [1] Azzarelli, J.M., Mirica, K.A., Ravnsbæk, J.B. and Swager, T.M. 2014. Wireless gas detection with a smartphone via RF communication. *Proc. Natl. Acad. Sci.*, 111: 18162-18166.
- [2] Bailly, G., Harrabi, A., Rossignol, J., Stuerger, D. and Pribetich, P. 2016. Microwave gas sensing with a micro strip interdigital capacitor: Detection of NH₃ with TiO₂ nanoparticles. *Sens. Actuators B: Chemical*, 236: 554-564.
- [3] Barochi, G., Rossignol J. and Bouvet, M. 2011. Development of microwave gas sensors. *Sens. Actuators B*, 157: 374–379.
- [4] http://www.ddbst.com/en/EED/PCP/DEC_C1050.php
- [5] Itterbeek, A van and Clippeleir, K de. 1948. Measurements on the dielectric constant of gaseous ammonia, carbon oxide and hydrogen as a function of pressure and temperature. *Physica*, 14: 349-356.
- [6] Lee, H., Shaker, G., Naishadham, K., Song, X., McKinley, M., Wagner B. and Tentzeris, M. 2011. Carbon-nanotube loaded antenna-based ammonia gas sensor. *IEEE Transactions on Microwave Theory and Techniques*, 59: 2665-2673.
- [7] Liu, X., Cheng, S., Liu, H., Hu, S., Zhang D. and Ning, H. 2012. A Survey on gas sensing technology. *Sensors*, 12: 9635-9665.
- [8] Mamishev, A.V., Sundara-Rajan, K., Yang, F., Yanqing Du and Zahn, M. 2004. Interdigital sensors and transducers. *Proceedings of the IEEE*, 92(5): 808-845.
- [9] Schmidt, J.W. and Moldover, M.R. 2003. Dielectric permittivity of eight gases measured with cross capacitors. *International Journal of Thermophysics*, 24: 375-403.
- [10] Wu, B., Zhang, X., Huang, B., Zhao, Y., Cheng C. and Chen, H. 2017. High-performance wireless ammonia gas sensors based on reduced graphene oxide and nano-silver ink hybrid material loaded on a patch antenna. *Sensors*, 17: 2070.

Parking Management System

¹Dr. Nazneen Ansari, ²Nirantak Singh Raghav, ³Het Shah, ⁴Manan Sheth

¹Associate Professor, ^{2,3,4} B.E. Student

^{1,2,3,4}Department of Information Technology

^{1,2,3,4}St. Francis Institute of Technology, Mumbai, India

ABSTRACT: Traffic has always been a serious issue in metropolitan cities. Adding fuel to the fire, the improving standard of living, has led to people purchasing more private vehicles. The most frustrating thing while driving is finding an open parking space for your vehicle. Due to this, people park their vehicles along the roadside or in no parking zones, which lead to traffic and towing. As a solution to this, government has established public parking spots at various places. However most of these places are underutilized due to lack of awareness. The proposed solution will be a platform that aims to solve this problem by providing an accessible application to users, and ease of management for parking space providers. This platform allows you to book a parking spot by leveraging both public and private spaces, thus reducing roadside parking that causes congestion and also protects the vehicle from damage or theft. Hence, such an application which will be a win-win situation, as the government will be benefited by reduced peak hour traffic, the parking lot usage increases, and the vehicle owners are assured of the safety of their vehicle.

Keywords: Parking, Smart city services, Maps, Traffic, Mumbai, Metropolitan city, Vehicles

I. Introduction

In metropolitan cities like Mumbai, the most frustrating thing about having a vehicle is finding a parking spot. The average vehicular density per square meter here is 591, which is way higher than average national density is 300. Due to shortage of parking space, people here tend to park on the roadside or in no parking zones. While visiting South Mumbai, at night, one can observe that all the vehicles (especially cars) are parked along the roadside. There are trivial reasons for it but mainly due to the buildings which were not built considering the vehicles that each resident may have and hence a majority of buildings there don't have parking space. This not only consumes time but also leads to increase in traffic congestion. The drivers tend to drive at slow speeds when in hunt for a parking spot.

This problem is not critical when the road is broad but is elevated when the road becomes narrow, like single lane roads. This again leads to consumption of time, increase in traffic congestion and noise pollution due to honking by impatient drivers. There also is a chance of towing of the vehicle or getting damaged when parking along the roadside. Roadside parking in Mumbai is considered as the most recorded traffic violation every year.

According to TOI [1], in 2013-14, 1.12 lakh cases of unauthorized parking were recorded between April 1, 2013 and March 31, 2014. These are just the recorded ones but many of them go unrecorded as the owners tend to bribe the authorities. All this leads to one problem that is inaccessibility of parking spaces due to no knowledge about parking space nearby. Hence, there needs to be a solution that can maximize the utilization of available parking space which will lead to reduction in the above mentioned problem.

In the last decade there has been a rapid increase in private vehicle ownership, about 45% over the last five years in Mumbai [2], which has resulted to high traffic congestion in most city centers. As a corrective measure, the government has widened road spaces; however they have been completely taken over for parking of vehicles on many arterial roads in cities. The police have recently introduced the system of putting clamp locks on vehicles parked in unauthorized places to reduce illegal parking. As a result, people now prefer to park their vehicles in dedicated parking slots, however are not able to do so. The reason for this is a combination of factors like lack of knowledge of nearby parking space and their availability. If people can find parking spots near them right on their fingertips, they will always prefer to park their vehicles in trusted places. This will guarantee the safety of their vehicles and eradicate the fear of their vehicles getting towed or damaged.

Hence, such an application which shows and lets you book parking spaces will be a win-win situation, as the government will be benefited by reducing peak hour traffic, the parking lots by more customers, and the vehicle owners by assurance of safety.

II. Literature Survey

While there exist systems that recommend parking spaces [3, 4], they generally do so for public and indexed parking lots. A lot of approaches have been developed to solve this problem but they are insufficient either they are not sufficient to cope with the user needs or they are not able to communicate properly [5] with the external interfaces or they are too costly [6].

There is also the danger of accidentally parking in illegal parking spaces run by local parking mafia that may cause fines, robberies or damages to a vehicle. Available parking management applications simply make use of GPS provided location to find parking lots that may or may not be empty. This can easily be done without external applications, using Google Maps.

The level of precision of these third party applications is low in showing the availability of spaces. In India there is a large dis-proportionality between the number of cars on the road to the number of open parking spaces [7], thus there is a need to utilize private parking spaces as well. This application has more functionality that provides bidirectional usage, influenced by Airbnb [8], allowing users to both rent and occupy parking spaces.

Currently, there are no other platforms that have this functionality. This allows users having private parking to rent out the space to other users that may be nearby, thus utilizing the space that may remain empty during office hours and earn some money as the parking fee.

A paper by Donald Shoup [9] describes the increasing cost of parking with respect to the cost of building such structures. It clearly states that the high cost of parking is due to minimum cost restrictions enforced in parking slots. This cost may be reduced by removing minimum parking cost limits, which will make parking vehicles cost effective and hence relieve traffic caused by roadside parking.

The paper on ParkNet [10] provides details and implementation for drive by sensors attached on vehicles that detect parked cars and open slots available for legal roadside parking. A section in this paper describes prototypes built using ultrasonic sensors to detect nearby cars and collect all relevant data. This paper may be useful for analysis of parking data collected by vehicles fitted with the device and help enforce parking and traffic laws from the knowledge gained from this data, but this process requires hardware to be fitted on vehicles and thus may be inefficient on a large scale.

A similar paper on parking space discovery [11] explains the use of VANET i.e. Vehicular Ad-hoc Networks, an IEEE 802.11 Wireless LAN standard to collect spatio-temporal information on traffic and parking space availability. The algorithm user here involves vehicles broadcasting their information like - speed and location, this information is received by the server as well as other vehicles nearby. Information is then aggregated from all vehicles to determine a topology of vehicles, their velocities, which can be used to further derive the amount and average speed of traffic.

A recent paper in the IEEE journal [12] provides a generalized recommended architecture for car parking management based on Cyber Physical Systems. This paper defines a simple method based on a client-server architecture, where a centralized server stores information about all free spaces and informs users as requested. However, this system only works for free and publicly visible spaces, and does not provide an option for booking slots or online payment.

Our project aims to resolve some of these problems and add more functionality, by taking all these references into consideration. We conducted a survey to check the feasibility of such a system.

1. While going out, can you easily find a parking place near you?

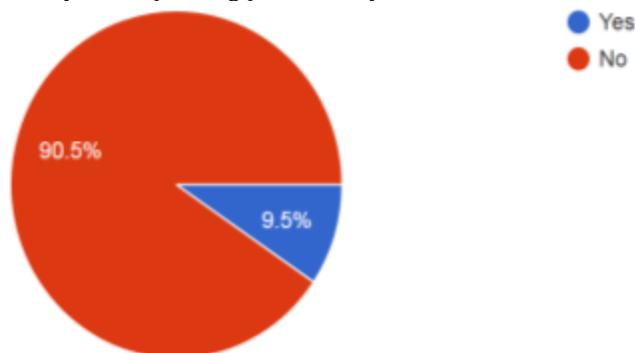


Figure 1: The parking problem

2. Would you rather park your car on the road, or in a parking lot for a minimal charge?

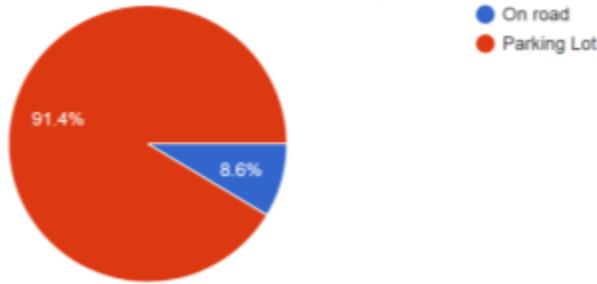


Figure 2: Feasibility study

III. Proposed Solution

To control the problems mentioned before, a system is to be developed that will act as a network of both public and private parking spaces available. The user once registered and logged in to the app can either rent out their private parking spot or may want to book a parking spot.

When booking an available parking spot, they need to select the destination location and the system will show all the available spaces nearby (within 500m-1000m). Registered users who want a parking space for some time can find open lots near their destination by dropping a marker on the map. Available spaces within a radius will be show as a red marker, and its information can be viewed by tapping the marker.

It will list out all the public/private parking grounds like government parking, airport parking or even any residential parking which can be rented out. The user can then select the amount of time and number of vehicles (when in groups). Once payment is done online or later via COD, the system will direct the user to that parking space. Location tracking will be done via Google Maps API, and available parking will be shown for a region of radius computed from latitude and longitude coordinates obtained from Google. When a unique code is shown at the entrance of parking, the user is then allowed to park their car.

A dashboard will be provided to the parking space owners to manage their parking space. They can set day and time when the parking space will be available and view all the users that have used the space previously. Users providing their parking may charge a fee rated by the hour, and limit its availability duration. All user interaction will be done on the mobile application, while the communication of data will take place in the background with a REST API server, as shown in Fig 3. Performance may be measured by the speed of fetching available parking based on user location, and the accuracy of this data.

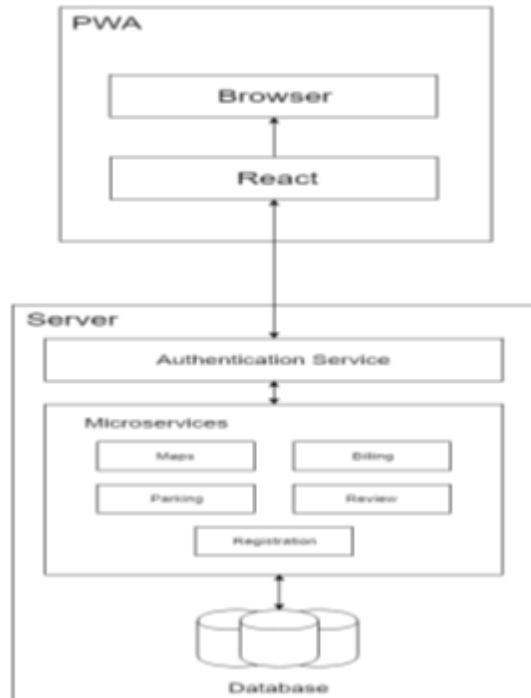


Figure 3: System Architecture

IV. Working

1. Users may install the app on their phone either via their app stores or from the website.
2. Users can then login or register via social media accounts like Google, Facebook, using OAuth2 or via email id and password.
3. A user's current location is first sent to the server, which then calculates parking spots within a radius of 500m and sends the results back.
4. The results are marked on the map with their respective prices and distances, and the user can view additional details and confirm their booking with their registered vehicle.
5. Users can register additional vehicles, add their own parking spots to rent out, and cancel bookings as required.

V. Conclusion

The primary goal of this project is to create a user friendly and accessible application through which users can easily book parking spaces without the hassle that makes their travel comfortable and payments easy. There are various benefits of this system such as increase in utilization of parking spaces that may be unknown to the public and reduction in traffic and congestion caused due to roadside parking. This application can also help users generate some passive income by renting out their parking spots, once they are willing to consider security and privacy concerns for parking in residential areas.

References

1. <https://timesofindia.indiatimes.com/city/mumbai/At-430/km-Mumbai-has-highest-car-density/articleshow/52825161.cms>
2. <https://timesofindia.indiatimes.com/city/mumbai/no-of-vehicles-in-city-up-50-in-5-yrs/articleshow/56302918.cms>
3. Chia-Ying Lin, Yi-Lung Lu, Meng-Hsun Tsai, Hui-Ling Chang, Utilization-based parking space suggestion in smart city, Consumer Communications & Networking Conference (CCNC) 2018 15th IEEE Annual, pp. 1-6, 2018.
4. Fei-Yue Wang, Liu-Qing Yang, Jian Yang, Yanli Zhang, Shuangshuang Han, Kai Zhao, Urban intelligent parking system based on the parallel theory, Computing Networking and Communications (ICNC) 2016 International Conference on, pp. 1-5, 2016.
5. Richard Arnott, John Rowse, Modeling Parking, Journal of Urban Economics, vol. 45, no. 1, Elsevier BV, January 1999, pp. 97-124.
6. Erik Verhoef, Peter Nijkamp, Piet Rietveld, The economics of regulatory parking policies: The (IM)possibilities of parking policies in traffic regulation, Transportation Research Part A: Policy and Practice, vol. 29, no. 2, Elsevier BV, March 1995, pp. 141-156.
7. Amihai Glazer, Esko Niskanen, Parking fees and congestion, Regional Science and Urban Economics, vol. 22, no. 1, Elsevier BV, March 1992, pp. 123-132.
8. Airbnb, <https://www.airbnb.com/>
9. Donald Shoup, The High Cost of Free Parking, ISBN:1351178067, 9781351178068, https://books.google.co.in/books?id=J_w6DwAAQBAJ Routledge, 2017
10. Mathur, Suhas and Jin, Tong and Kasturirangan, Nikhil and Chandrasekaran, Janani and Xue, Wenzhi and Gruteser, Marco and Trappe, Wade, ParkNet: Drive-by Sensing of Road-side Parking Statistics, Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services, MobiSys, pp. 123-136, 2010
11. Caliskan, Murat and Graupner, Daniel and Mauve, Martin, Decentralized Discovery of Free Parking Places, Proceedings of the 3rd International Workshop on Vehicular Ad Hoc Networks, VANET 2006, pp. 30-39.
12. Shafiq ur Rehman, Volker Gruhn, Recommended architecture for car parking management system based on cyber-physical system, Engineering & MIS (ICEMIS) 2017 International Conference on, pp. 1-6, 2017.

Placement Data Anatomization Using Tableau

¹Sahil Rajendrakumar Raut,²Shraddha Somnath Pawar,³Nazneen Ansari,⁴Vivian Brian Lobo, ⁵Vaibhav Kala, ⁶Snehal Dmello, ⁷Elizabeth George

^{1,2} B.E. Student,³Associate Professor, ^{4,5,6,7}Assistant Professor

^{1,2,3,4,5,6,7}Department of Information Technology

^{1,2,3,4,5,6,7}St. Francis Institute of Technology, Mumbai, India

ABSTRACT: *Placements for students in engineering have been of paramount significance. Providing the best amenities to students for better placement opportunities have always been the motto for educational institutes. Lack of knowledge and ignorance to the basic fundamentals of engineering, quants, verbal and logic have been the main reason for many failed interviews and test. Thus, exploring profound historical data for prediction creates a better chance of placement for students. Data visualization with data mining algorithm and prediction software helps in creating analysis for placement data and help institutes improve success rate in placement. Therefore, this study aims to use data visualization tool tableau along with k-means clustering technique for monitoring historic placement data to analyze future trends of placement interviews and aptitude.*

Keywords: *Classification, Data Mining, Placement Anatomization, Clustering, Tableau*

I. Introduction

Massive amount of placement data is accumulated by institutes each year in a bid to provide better facilities and placement opportunities to students. If this data is anatomized effectively and worked upon, it can yield cognizance and provide information about various data. Analysis of results from the data often becomes a burdensome task in the absence of right data mining technique, poor data preprocessing and visualizing software. Thus, we examined the placement data using industry standard software tableau for visualization purpose to gain insights into the placement data. We used real-time placement data of St. Francis Institute of Technology. A google survey form was distributed to the alumni members and the final year students currently giving placement interviews. We accessed various factors such as pre-placement tests, interview questions and important topics and correlated it to capitulate useful analyses and recommendations to improve placement process in place.

II. Literature Review

The significance of data mining is to discover knowledge and to apply knowledge through the use of existing data [1]. Data mining predicts future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions [2]. Tableau is a powerful data visualization and business intelligence tool. Data visualization with data mining algorithm and prediction software helps in creating analysis for placement data and help institutes improve success rate in placement. Lack of knowledge and ignorance to the basic fundamentals of engineering, quants, verbal and logic have been the main reason for many failed interviews and test. Thus, exploring the profound historical data for prediction creates a better chance of placement for students.

III. Proposed Solution

Data visualization along with visualization and prediction software tremendously boost the chances of better analysis and placement of students. Among various visualization tools, Tableau, i.e., the proprietary drag and drop tool, makes any visualization possible stand out among respective groups [3]. Figure 1 shows visualization and anatomization of data using tableau where it takes data from placement dataset and performs k-means clustering to represent in a graphical format. k-means is a numerical, unsupervised, nondeterministic, iterative method [4]. It is simple and very fast; so, in many practical applications, the method is proved to be an effective way that can produce good clustering results. Boundless analysis of data helps in providing answers to placement queries.

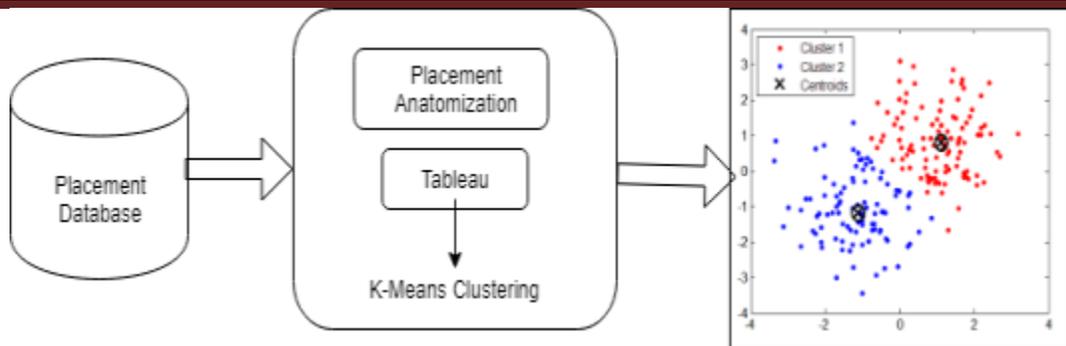


Figure 1. Visualization and anatomization of data using tableau for placement dataset followed by *k*-means clustering.

IV. Data Collection

The data collection process consists of the following steps:

1. Promulgating placement Google forms to final-year students and alumni members.
2. The form was prepared with a motto to explore the area of questions asked most often during any interview process.
3. The form comprised students name, 10th, 12th, engineering aggregate score, books referred as well as name of the companies the student has attended placement and the questions asked during the test in quants, verbal and logic.
4. Students who advanced to the personal interviews were asked additional questions about their technical and human resource interview.

V. Tableau

Data visualization is necessary for collecting plethora of knowledge and insights about the given data. Using the right data visualization tool for obtaining knowledge is significant for the development and progress of the system. Aesthetic elements play a major role in developing this perspicacity [5]. Tableau is one of the fastest evolving business intelligence (BI) and data visualization tool. Placement data collected and transformed into various data source such as excel, text, csv can be connected without any hassle to tableau workspace and performing preprocessing, analysis and visualization on the data.

VI. K-Means Clustering

Since clustering algorithms are used to find useful and unidentified classes of patterns, it helps in understanding the pattern of most likely asked questions to accumulate knowledge about the interview and testing system of placement companies in engineering colleges [6]. Among the various available clustering algorithms, *k*-means clustering algorithm is a simple, practical and efficient clustering algorithm [2]. The process of clustering in tableau follows the given pattern. The clustering technique that tableau uses is *k*-means clustering where it segregates data into *k* clusters for given number of clusters. The algorithm finds out the mean value of all points in the cluster also called as centroid through an iterative procedure that minimizes distances between individual points in a cluster and the cluster centre. Tableau uses Lloyd's algorithm with squared Euclidean distances to compute *k*-means clustering for each *k* [7]. Combined with the splitting procedure to determine the initial centres for each *k* > 1, the resulting clustering is deterministic with the result dependent only on the number of clusters. Tableau uses the Calinski-Harabasz criterion to assess cluster quality. The Calinski-Harabasz criterion is defined as follows:

$$\frac{SS_B}{SS_W} \times \frac{(N-k)}{(k-1)} \tag{6.1}$$

where SS_B is the overall between-cluster variance, SS_W is the overall within-cluster variance, *k* is the number of clusters, and *N* is the number of observations. The cohesiveness of the clusters depends on the value of this ratio where it increases with the increase in the value of the ratio.

VII. Implementation

7.1 Data Preprocessing

The data collected from alumni students through Google survey form is collectively stored in an excel sheet, which is needed to be cleaned before clustering. Therefore, unwanted rows from the dataset are

deleted and spelling mistakes are corrected manually. All the empty entries are replaced by null value. Figure 2 shows desired dataset, which is loaded in Tableau.

Form Responses 1 Name	Form Responses 1 Gender	Form Responses 1 Class	Form Responses 1 PID number	Form Responses 1 10Th Aggregate	Form Responses 1 12Th Aggregate	Form Responses 1 Engg aggregate	Form Responses 1 Companies that y...	Form Responses 1 Verbal Ability
Shivraj Bhaygude	Male	BE-IT-A	111016	84.9200	86.0000	69.2150	Tata Consultancy Ser...	Email writing
Shivraj bhaygude	Male	BE-IT-A	111016	84.9200	86.0000	69.2150	Tata Consultancy Ser...	Email writing
Pooje vasant sawant	Female	BE-IT-B	111103	89.8400	87.8300	78.1570	Tata Consultancy Ser...	Email writing
Selvin lopes	Male	BE-CMPN-A	112066	88.4600	71.2700	59.7200	Caggemini	null
Kajol	Female	BE-IT-A	111002	89.0700	79.0000	68.0000	Tata Consultancy Ser...	Email writing
Nitesh Kadam	Male	BE-IT-A	111059	85.0700	78.3300	61.9000	Persistent Systems L...	null
Ahilesh Nair	Male	BE-IT-A	121266	73.3800	69.6800	60.0000	Persistent Systems Ltd.	null
Mandar Kadam	Male	BE-CMPN-B	112056	90.0000	80.0000	59.0000	Larsen & Toubro Infot...	null
Kalpita Patekar	Female	BE-IT-A	121271	77.6500	76.7200	55.0600	Caggemini	null
Darryl Fernandes	Male	BE-IT-A	111035	85.3800	77.0000	62.5600	Tata Consultancy Ser...	Email writing
Sahil P. Dodti	Male	BE-IT-A	111029	90.4600	77.8300	65.7200	Persistent Systems L...	Synonyms
Raisa Ferreira	Female	BE-IT-A	111041	91.0000	76.0000	63.9300	Larsen & Toubro Infot...	null

Figure 2. Data loaded into tableau.

7.2 Clustering

On x-axis, 10th aggregate is added, and as per the engineering aggregate, k-means clustering is done on Tableau. Clusters are formed as per the company in which students have placed. Clusters formed in the shape of circle in Fig.3 depicts the different colors with list of colors indicating specific company on right-hand side. On hovering the cluster, detailed information is shown.

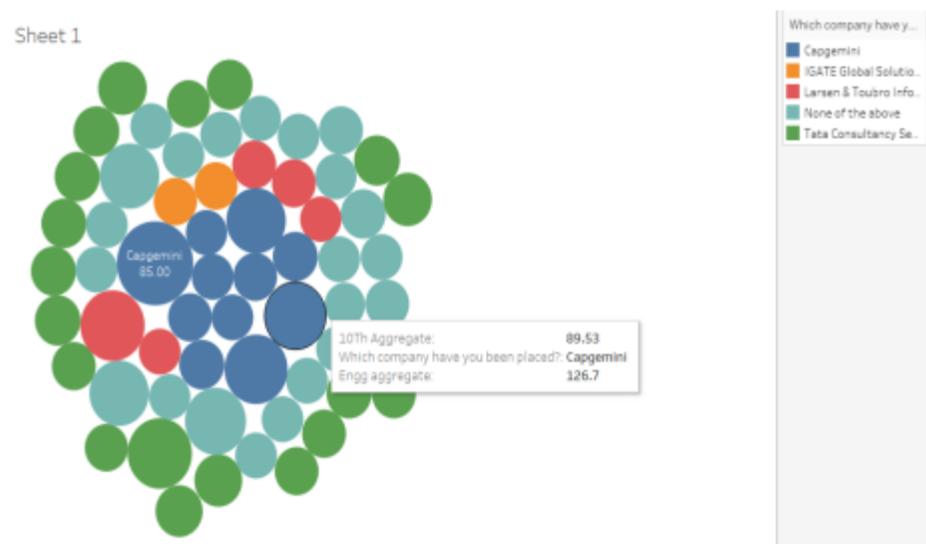


Figure 3. Clustering of dataset with respect to marks.

7.3 Data Anatomization

On x-axis, B.E. aggregate is added and on y-axis 10th aggregate is added and Figs.4 and 5 portray various questions asked during interview for a particular company, which is analyzed using Tableau graphical representation. Even which books are referred most often for interview preparation is analysed in Figure 6 so that it gets easier for future batches to prepare for interviews.

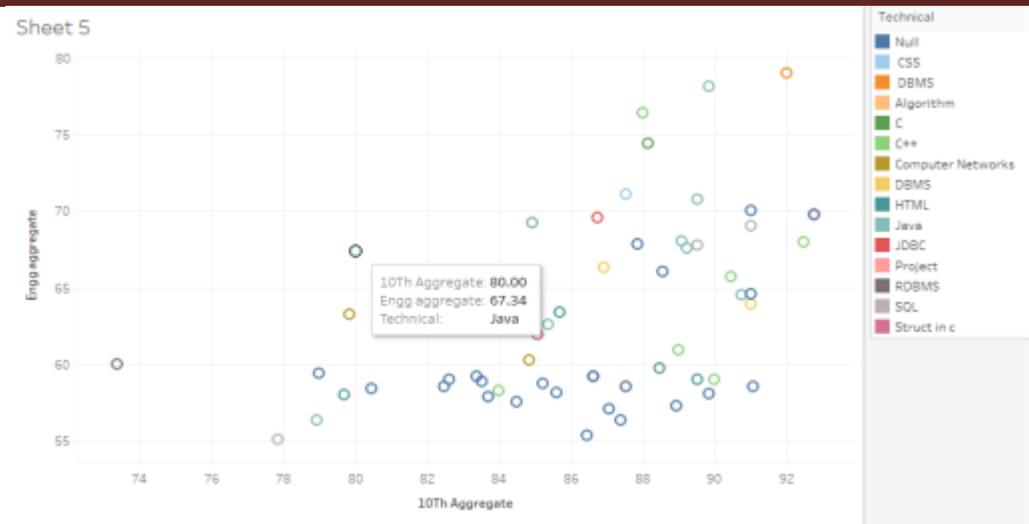


Figure 4. Technical questions asked by specific companies.

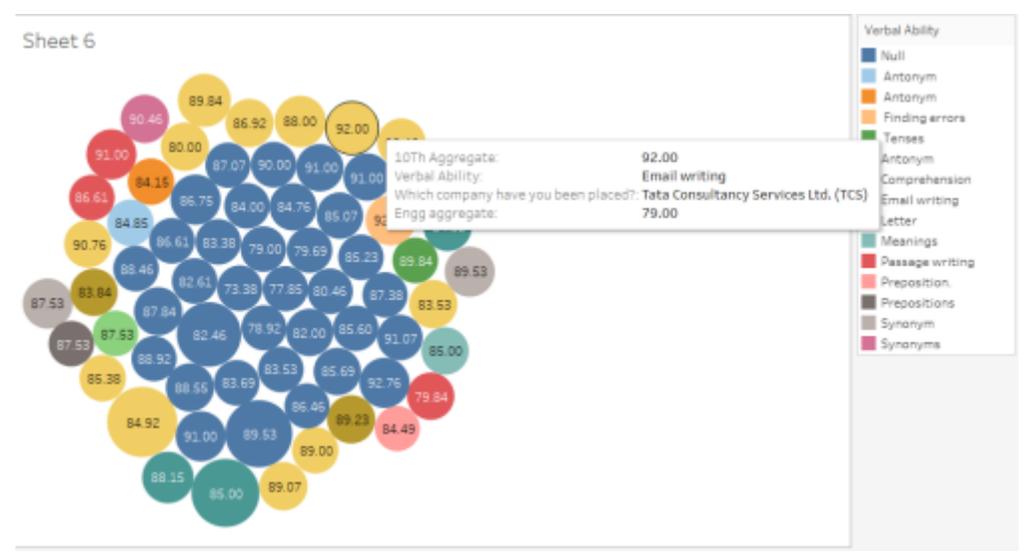


Figure 5. Verbal questions asked by specific companies.

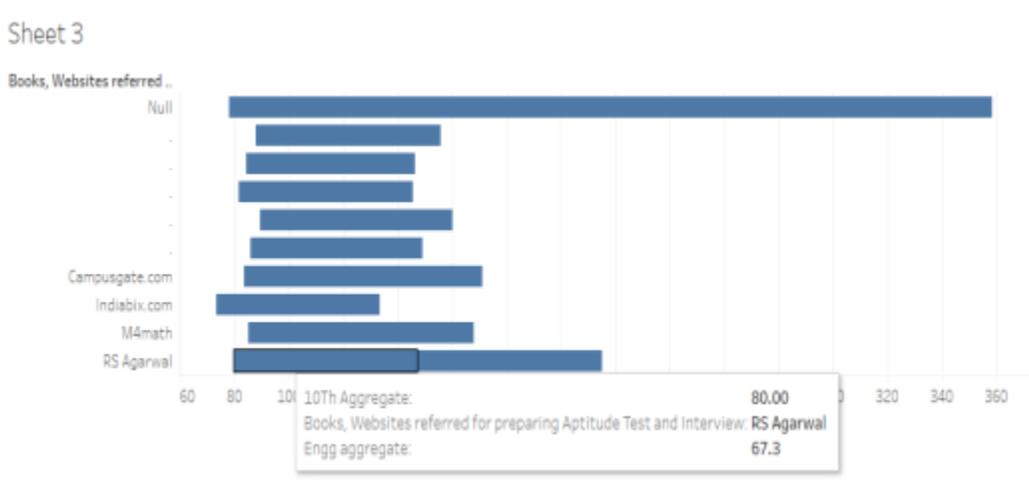


Figure 6. Books referred by students for interview preparation.

VIII. Results

Figure 3 depicts cluster formation of various companies in which *k*-means clustering algorithm is applied by taking engineering aggregate of students. The diagrammatic representation helps placement committees to analyze number of students placed in specific companies based on their aggregates. The anatomization carried out in Figure 4 on the placement data revealed the most frequent technical topics asked to the student during interviews based on their aggregate. This helps future batches to prepare interview questions as per their dream company. The analysis in Figure 5 shows clusters of verbal questions asked by particular companies along with student's marks. The histograms in Figure 6 shows various study material used by students to clear the interviews taken by the companies.

IX. Conclusion

This method of using tableau for clustering data intend to provide insights and cognizance of placement data for better future planning and increasing efficiency and improving progress of success rate in placements. The study explores placement data using tableau and helps in giving judgements on the most preferred topics to study, books to refer and most common questions asked by every company. The placement history and academic scores of students play an important role in generating factors and inferences that can help in increasing success rate of students sitting for placements and help them get their desired companies. *k*-means clustering in the data visualization tool tableau helps in achieving it.

References

1. Qiao, J., & Zhang, Y. (2015, November). Study on K-means method based on data-mining. In Chinese Automation Congress (CAC), 2015 (pp. 51-54). IEEE.
2. Tiwari, H. (2014). Data mining, warehousing and OLAP technology. *Discovery*, 24, 58-62.
3. Nair, L., Shetty, S., & Shetty, S. (2016). Interactive visual analytics on Big Data: Tableau vs D3. js. *Journal of e-Learning and Knowledge Society*, 12(4).
4. Na, S., Xumin, L., & Yong, G. (2010, April). Research on k-means clustering algorithm: An improved k-means clustering algorithm. In *Intelligent Information Technology and Security Informatics (IITSI), 2010 Third International Symposium on* (pp. 63-67). IEEE.
5. Data visualization through tableau software, EduPristine, [Online]. Available: <https://www.edupristine.com/blog/data-visualization-tableau-software> [Accessed on January 31, 2019].
6. Gulati, H., & Singh, P. K. (2015, March). Clustering techniques in data mining: A comparison. In *Computing for Sustainable Global Development (INDIACom), 2015 2nd International Conference on* (pp. 410-415). IEEE.
7. Find Clusters in Data, Refine Your Dashboard, [Online]. Available: <https://onlinehelp.tableau.com/current/pro/desktop/en-us/clustering.htm>. [Accessed on January 31, 2019]
8. Ojha, A., Pattnaik, U., & Sankar, S. R. (2017, August). Data analytics on placement data in a South Asian University. In *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)* (pp. 2438-2442). IEEE.

Proposed Research for Detecting Poachers

¹Shreya Shivaji Gaikwad, ²Cyril Silvester Dabre, ³Pranit Mohan Mhatre,
⁴Dr. Nazneen Ansari

^{1,2,3} Under Graduate Student ⁴Associate Professor
^{1,2,3,4}Department of Information Technology,
^{1,2,3,4}St. Francis Institute of Technology, Mumbai, India

ABSTRACT: Wildlife species like elephants, rhinos and tigers are getting killed ruthlessly by poachers, to plunder their claws, ivory, horn, skin, etc. The forest rangers conduct patrols with the aim of preventing poachers from carrying out illegal poaching activities. However, the sanctuaries and wildlife conservatories are spread over a substantial land and it is manually impossible for a team of rangers to patrol this widespread land day in and day out. Poaching activities bring a huge chunk of income which is persuading the poachers to carry out such activities risking their lives. Hence there is a need to automate the process of monitoring poaching activities. Hence, we propose a novel approach to track the poachers using deep learning algorithm like faster RCNN which can be demonstrated over a simulated environment AirSim built upon unreal engine using unmanned aerial vehicles or drones.

Keywords: Simulation, Unmanned Aerial Vehicle (UAV), AirSim, Wildlife Conservation, Anti-poaching

I. INTRODUCTION

Wildlife poaching is a significant threat to large-bodied animal species and it is one major reason of the population declines of key wildlife species such as elephants, tigers, and rhinos, which are crucial to the functioning of natural ecosystems. Many wildlife species are on the verge of extinction due to such poaching activities (F. Fang, 2017). Latest news says “87 Elephants Killed by Poachers in Africa’s Last Safe Haven” At least 87 elephants have been killed for their ivory in Botswana in recent months, according to the conservation non-profit Elephants without Borders, which discovered the carcasses (Rachael Bale, 2018). The forest rangers conduct patrols with the aim of preventing poachers from poaching animals, either by catching the poachers or by removing animal traps set by the poachers but it is not possible for forest rangers to patrol the entire area effectively and efficiently. Legally and illegally hunting has taken a great toll on India’s elephant (poaching facts n.d.) populations over the past two centuries.

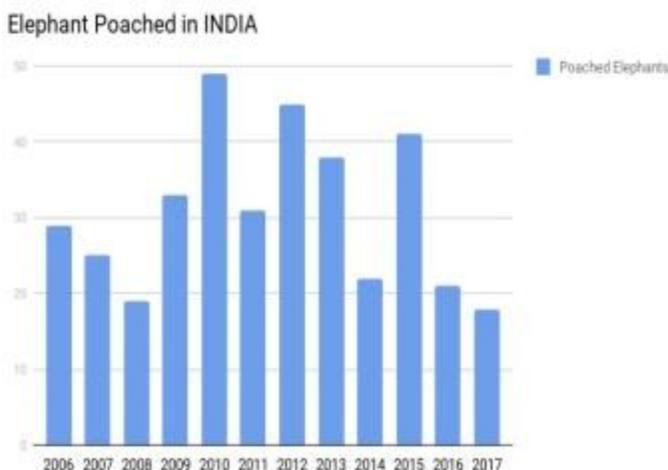


Fig. 1. Elephant Poaching in Select Regions Confirmed by WPSI in India (2006–2017) [3]

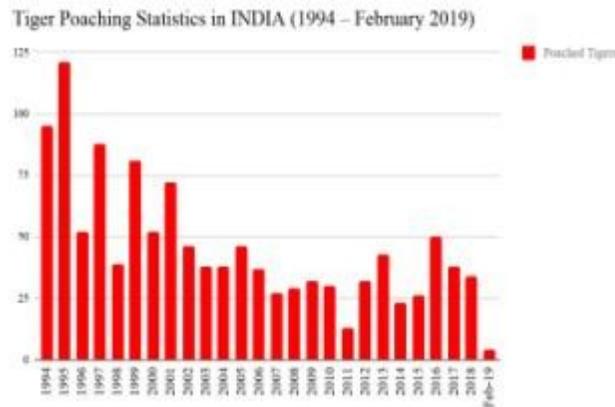


Fig. 2. Tiger Poaching in INDIA Statistics (1994–February 2019) [3]

II.LITERATURE REVIEW

M. Gor et al. (2017) investigated that wild animals straying out of wildlife sanctuaries and national parks can be tracked by auto generative location tracking and movement patterns which was implemented using Wireless Sensor Network, GPS with the accelerometer and the Wi-Fi shield where base station is monitored by using appropriate modeling that can visualize the motion path and specific location of the animal with predefined boundaries that the animal should not cross.

Carrio et al. (2017) have stated that Unmanned Aerial Vehicles (UAVs) are being widely used for several types of military as well as civilian applications from security, surveillance, and disaster rescue. The review was implemented on recently reported applications of deep learning for UAVs, including the most recent developments as well as their performances and limitations. Images acquired from UAVs are currently the existing type of information being exploited by deep learning, mainly due to the low cost, low weight, and low power consumption of image sensors in UAVs. This noticeable fact explained the dominance of CNNs among the deep learning algorithms used in UAV applications, given the excellent capabilities of CNNs in extracting useful information from images.

The Vision Heatmap system is used to map the human traffic which occurs at specific time at particular space (Djajadi et al., 2017). Use of cameras has been proposed to overcome the drawback of limited area that gets covered when motion sensors are used. The main purpose is to develop heatmap system by using camera vision which helps to find, track and record the movement of human (poachers) in a specific time and to inform the authorized person about it. The overall method for generating heatmap system can be implemented by carrying out the steps successively-preparing the development environment, taking images, clear the back image, calculate the image difference and detecting movement, use threshold to reduce/remove pixel noise, path making for busy area, path with color gradation, real time vision heat map overlaid on visual image and finally image stitching.

Railway Tracks are also laid in wild areas for better transport facility but these rail routes are dangerous for wild animals. Various animals have been killed from being overrun by trains. This paper (Mathur et al., 2014) addresses conservation of elephants by preventing them from being overrun by trains with the help of wireless sensor network.

The global communication network and advance animal tracking technology is used. For animal tracking units (Wall et al., 2014) four broad approaches are defined for wildlife monitoring in real time which includes proximity, geo fencing, movement rate and immobility are used to wildlife spatial activity and behavior which is impossible with conventional tracking systems. By using the Real Time Monitoring (RTM), environmental variable can be managed, processed, accessed and disseminated for the improvement of public safety and global monitoring for conservations. RTM used in animal tracking either by satellite based or by ground-based system for mobile communication technologies can be integrated into tracking units which makes possible to track animal and also for processing data in real time. RTM allows to observe the position and the movement of the animal within Geographic Information System.

Klein et al. (2015) stated that the technological advancement in sensors, big data processing and machine intelligence as well can provide affordable and effective measures of conservation of wildlife. This system can be used to monitor the endangered species and ecosystem. A strategic vision for how data-driven approaches to conservation can drive iterative improvements through better information and outcomes-

based funding has been presented. The approach to biodiversity monitoring involves periodically sending observers to a predetermined set of survey sites to collect data over relatively short survey windows. Logistical hurdles, personnel costs, and time constraints made it difficult to scale these traditional surveys to meet the increasing demands of global conservation.

Bondi et. al. (2018) have proposed a system to assist rangers with poaching activities in national park and conserved areas by detecting poaching incidents and predicting future incidents. The aerial vehicles had been used to create a dataset of 70 historical thermal infrared videos. It was done by flying the drones at night on a pre-programmed path. After labelling of the videos, individual frames had been used to train for animal and poacher detection. AirSim-W has been used which includes African savanna environment in Unreal Engine. Python API has been used to make the UAV fly in a zig-zag pattern and get back to object of interest such as poachers to make sure the object is in frame. The captured RGB images were converted to infrared images. The segmentation map created from RGB simulation specified the object captured in each image.

The paper (Shah et al., 2017) discusses SPOT (Symmetric Poacher Detector) which is an application that enhances the ability of the drones to detect poachers and animals by using state-of-the-art AI techniques to automatizing the detection process in near real time. SPOT is divided into two parts that is online and offline. The offline process includes preprocessing of training and testing videos, labelling training and testing videos, transferring the labels to faster RCNN format, training the faster RCNN. The online process includes capturing of live video stream, preprocessing, sending to Azure and then using Faster RCNN to find bounding boxes. Furthermore, the results of precision-recall have been compared with other promising application like EyeSpy.

Fang et al. (2017) have proposed an anti-poaching tool CAPTURE (Comprehensive Anti-Poaching tool with Temporal and observation Uncertainty Reasoning) which uses game-theoretic algorithm to predict the probable poaching areas and plan effective patrols. CAPTURE uses hierarchical behavioural model. Out of the two layers, the first layer incorporates how the poachers behave considering their activity in the past. The second layer is used to detect the signs of poaching. Logistic model has been used in both the models to aggregate the behavior without knowing the total number of poachers in both the layers.

III. PROPOSED SYSTEM

3.1 System Architecture

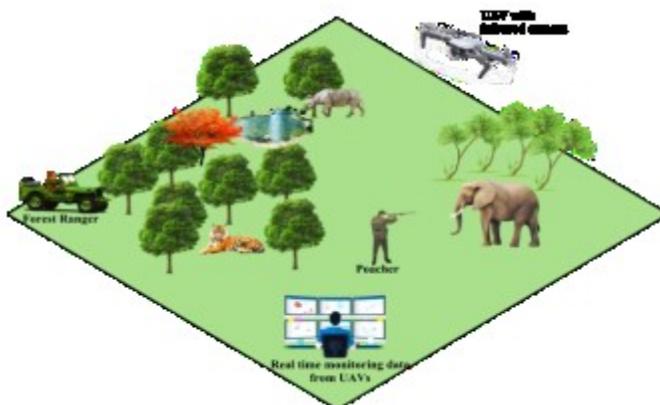


Fig.3. System Architecture

The system architecture is a simulation environment where forests, animals, poachers are simulated. Unmanned Aerial Vehicles are deployed in these regions along with infrared cameras attached to them. At night when it is difficult for the rangers to spot these poachers, they can be easily detected by these infrared cameras due to the heat that is emitted by them. As they are the warmer objects in the colder environment the heat emitted will be more and will appear brighter in comparison with the background. Knowing the unique thermal profile will help to differentiate the poachers from animals and other warm objects that are in the region. The video of the flights of the UAV is then recorded for training and testing purpose. The following flowchart shows the steps wise process to collect data for training and testing and how the poacher detection can be done on live video stream.

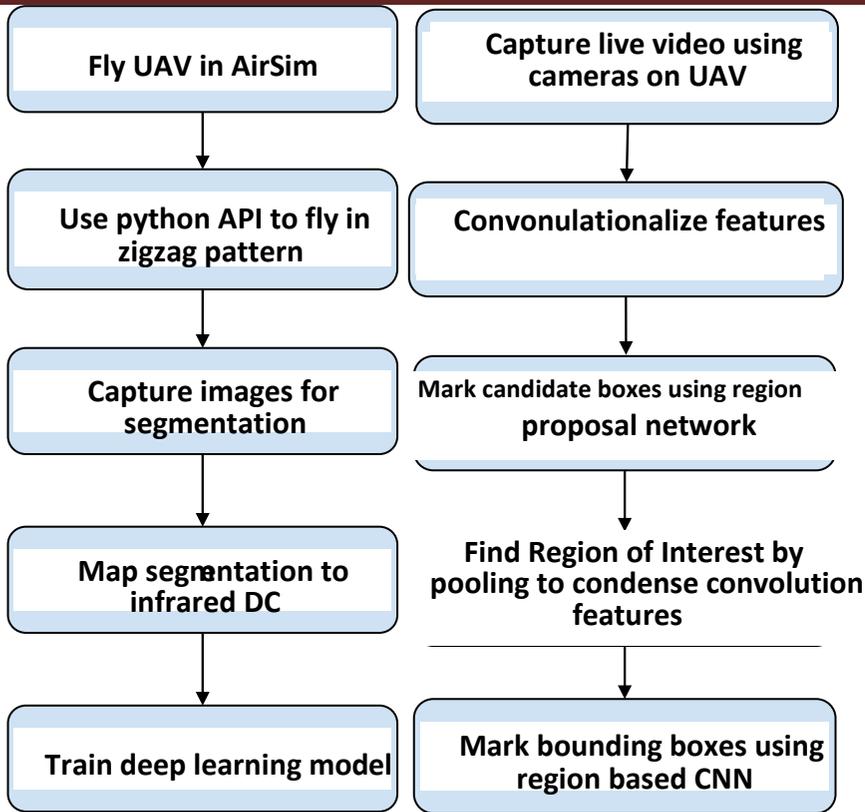


Fig.4. Flowchart to generate dataset training

Fig.5. Flowchart to carry out poacherfor detection using Faster RCNN

3.2 AirSim Simulation Environments

Airsim (Microsoft Airsim n.d.) is an open source simulator which is commonly used to simulate drones and cars in unreal engine. It can be simply developed using an unreal engine. Airsim is used to decrease the gap between the technical Simulation and the real world. It exposes API so you can interact with the developed tool in the simulation programmatically. The platform tries to positively influence development and testing of data-driven machine intelligence techniques such as reinforcement learning and deep learning. Airsim ensures that interaction with the simulation is possible. The API can be used to retrieve images, get state and control the vehicle as well. APIs are exposed through Remote Procedure Call, which can be easily accessed through various languages such as Python, C++, C# and java. API is also available as a part of separate, independent cross-platform library, so you can deploy them on a companion computer on your vehicle. There are two ways you can generate training data from AirSim for deep learning. First way is to just simply press the record. Second way is the better way as to generate the data exactly in the way you want by accessing the API, by using this it allows you to be in full control of How, What, Where, and When want to log the data and the Third way to use Airsim is to use “Computer vision mode” where you don’t have any kind of vehicle or physics. You can simply use the keyboard to move around or use the available API position in the cameras and collect the images as depth, disparity, images on surface, or of any object segmentation.

3.3 Unmanned Aerial Vehicle (UAV)

UAV (Aircraft System n.d.) is an abbreviation of Unmanned Aerial Vehicle which can be also called as drone. In some cases, the acronym of the UAV may be expanded to UAVS (Unmanned Aircraft Vehicle System) whereas the FAA has adopted the acronym as UAS (unmanned vehicle system) so that to reflect the fact that the complex uses the ground stations in it and other elements besides the actual vehicle. It is the component of Unmanned Aircraft System (UAS) which include UAV, a ground base controller and also the system for communication in it. UAV can be a remote-controlled aircraft which can be flown by pilot at ground control station or else it can be a pre-programmed flight plan or more complex dynamic automation system. These systems are commonly used in different missions, including reconnaissance and attack roles. UAV can also be defined as the vehicle which can be controlled, a vehicle system with sustained level flight and which is powered by jet or reciprocating engine. Control system used in UAV is different from manned aircraft. For

remote human control, a camera and video link almost always replace the cockpit windows; radio-transmitted digital commands replace physical cockpit controls. Autopilot software is used on both manned and unmanned aircraft, with varying feature sets.

IV.CONCLUSION

This paper encourages the reader to understand the need to espy methods for wildlife conservation. Animals are integral part of the ecosystem and their survival is important for maintaining the ecological balance. However, merciless poaching has endangered their existence. Hence it is important to reach out a larger community to carry out research to detect and prevent poaching. A simulation environment will provide a platform to involve more number of people in carrying out research on a crucial issue of detecting poaching activities that are carried out in unreachable lands. It is difficult to carry out real time monitoring and detection of poaching due to various factors like dense forests, environmental conditions, insufficient data, and terrain. Furthermore, use of aerial vehicle can aid easy monitoring of wildlife and poaching activities. Use of simulation environments can enhance ability and ease of training and testing of algorithms do develop a sound approach for real time monitoring and detection of poaching activities.

REFERENCES

- [1] Fang, F., Nguyen, T. H., Sinha, A., Gholami, S., Plumtre, A., Joppa, L., ... & Critchlow, R. (2017). Predicting poaching for wildlife protection. *IBM Journal of Research and Development*, 61(6), 3-1.
- [2] Rachael Bale.(n.d.).Retrieved from <https://www.nationalgeographic.com/animals/2018/09/wildlife-watch-news-botswana-elephants-poaching/>
- [3] Elephant Poaching Statistics (n.d.), Retrieved from <http://www.poachingfacts.com/poaching-statistics/>
- [4] Gor, M., Vora, J., Tanwar, S., Tyagi, S., Kumar, N., Obaidat, M. S., & Sadoun, B. (2017, July). GATA: GPS-Arduino based Tracking and Alarm system for protection of wildlife animals. In 2017 International Conference on Computer, Information and Telecommunication Systems (CITS) (pp. 166-170). IEEE.
- [5] Carrio, A., Sampedro, C., Rodriguez-Ramos, A., & Campoy, P. (2017). A review of deep learning methods and applications for unmanned aerial vehicles. *Journal of Sensors*, 2017.
- [6] Djajadi, A., Jaya, P. U., & Muhida, R. (2017). Human Behavior Tracking with Vision Heatmap Toward Smart Building System. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, 9(2-8), 159-163.
- [7] Mathur, P., Nielsen, R. H., Prasad, N. R., & Prasad, R. (2014, May). Wildlife conservation and rail track monitoring using wireless sensor networks. In 2014 4th International conference on wireless communications, vehicular technology, information theory and aerospace & electronic systems (VITAE) (pp. 1-4). IEEE.
- [8] Wall, J., Wittemyer, G., Klinkenberg, B., & Douglas-Hamilton, I. (2014). Novel opportunities for wildlife conservation and research with real-time monitoring. *Ecological Applications*, 24(4), 593-601.
- [9] Klein, D. J., McKown, M. W., & Tershy, B. R. (2015). Deep learning for large scale biodiversity monitoring. In Bloomberg Data for Good Exchange Conference.
- [10] Bondi, E., Dey, D., Kapoor, A., Plavis, J., Shah, S., Fang, F., ... & Tambe, M. (2018, June). Airsim-w: A simulation environment for wildlife conservation with UAVs. In Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies (p. 40). ACM.
- [11] Shah, S., Dey, D., Lovett, C., & Kapoor, A. (2018). Airsim: High-fidelity visual and physical simulation for autonomous vehicles. In *Field and service robotics* (pp. 621-635). Springer, Cham.
- [12] Bondi, E., Fang, F., Hamilton, M., Kar, D., Dmello, D., Choi, J., ... & Nevatia, R. (2018, April). Spot poachers in action: Augmenting conservation drones with automatic detection in near real time. In *Thirty-Second AAAI Conference on Artificial Intelligence*.
- [13] Sanchez-Lopez, J. L., Molina, M., Bavle, H., Sampedro, C., Fernández, R. A. S., & Campoy, P. (2017). A multi-layered component-based approach for the development of aerial robotic systems: The aerostack framework. *Journal of Intelligent & Robotic Systems*, 88(2-4), 683-709.
- [14] van Gemert, J. C., Verschoor, C. R., Mettes, P., Epema, K., Koh, L. P., & Wich, S. (2014, September). Nature conservation drones for automatic localization and counting of animals. In *European Conference on Computer Vision* (pp. 255-270). Springer, Cham.
- [15] Microsoft/AirSim(n.d.).Retrieved from <https://github.com/Microsoft/AirSim/>.
- [16] Gholami, S., Ford, B., Kar, D., Fang, F., Tambe, M., Plumtre, A., ... & Mabonga, J. (2018, June). Evaluation of Predictive Models for Wildlife Poaching Activity through Controlled Field Test in Uganda. In *Workshops at the Thirty-Second AAAI Conference on Artificial Intelligence*.
- [17] Liu, X., Yang, T., & Yan, B. (2015, November). Internet of Things for wildlife monitoring. In 2015 IEEE/CIC International Conference on Communications in China-Workshops (CIC/ICCC) (pp. 62-66). IEEE.
- [18] How Do You Protect Endangered Animals at Night? Ask an Astrophysicist (n.d.) Retrieved on <https://news.nationalgeographic.com/2018/04/thermal-infrared-cameras-drones-poaching-conservation-animals-spd/>.
- [19] Aircraft System(n.d.). Retrieved on <https://www.theuav.com/>.

Public Toilet Health and Hygiene monitoring using IOT

¹Dr.Ashfaq Shaikh, ²Mohammed Amir Sayed, ³Riyaz Shaikh, ⁴Rushikesh Naik

¹Assistant Professor, ²Student, ³Student, ⁴Student

¹ Department of IT

¹M. H. Saboo Siddik of College Engineering, Mumbai, India

ABSTRACT: *The problem in urban slums related to sanitation is three-fold namely, availability of toilets for the masses at the locations which need them the most, cleanliness and hygiene of the available public toilets, behavioural element of people moving away from defecating in open to use toilets. This paper deals with the second element which is to track cleanliness and hygiene of available public toilets to make sure they are clean and usable. This paper proposes the use of technology to ensure that public toilets which are existing and clean are well maintained, hygienic and usable. This is achieved by identifying factors or parameters which determine the cleanliness and hygiene of a public toilet. Keeping this in view, we have proposed a system that will automatically track and monitor above parameters through usage of sensors and other available low-cost mechanisms to effectively monitor the hygiene level and generate alerts. Our system is based on Internet of Things (IOT) which will consist of Arduino interfaced with multiple sensors to track the identified parameters and will send alert to web portal which will be further used by the centralized admin to notify the cleaners to clean the toilets. The system will also generate usability rating of individual toilets by analysing the sensor values and the cleanliness pattern.*

Keywords: *IOT, Arduino, sensors, monitoring, hygiene, web portal, alerts.*

I. Introduction

Swachh Bharat Abhiyan (SBA) or Swachh Bharat Mission (SBM) is a nation-wide campaign in [India](#) for the period 2014 to 2019 that aims to clean up the streets, roads and infrastructure of India's cities, towns, and rural areas. The overall objective of the Swachh Bharat Mission is to provide complete sanitation solutions for all of India's 4041 statutory towns. Eliminating open defecation, use of insanitary toilets and manual scavenging is all part of this elaborate scheme. Breaking the age-old behaviour of defecating in the open along with sensitizing people on the importance of sanitation and toilet usage are some of the themes that figure prominently in the blueprint. Provision of household and public toilets, solid waste management facilities, capacity building and awareness generation through information, education and communication (IEC) campaigns are the core components of the Mission. The urban component has been allotted Rs. 62,009 crores of which the Government of India would be pitching in Rs. 14,623 crores. The government has proposed to spend \$20 billion on the construction of toilets alone, the funds for which will be released annually to each of the states. So far, 8 million toilets have been built [1].

The problem related to sanitation are availability of toilets for the masses at the locations which need them the most, cleanliness and hygiene of the available public toilets, behavioural element of people moving away from defecating in open to use toilets. The main aim of this paper is the second element which is ensuring the cleanliness and hygiene of available public toilets to make sure they are clean and usable. Though the construction of public toilets has soared drastically within the previous decade, but the challenge of monitoring and maintenance of these toilets poses a huge problem to government authorities. Our main focus is to come up with a system that can track the cleanliness and hygiene of the available public toilets. The problem regarding the cleanliness and hygiene has two dimensions, namely identifying factors or parameters which determine the cleanliness and hygiene of a public toilet and automatic tracking and monitoring of these defined parameters through usage of sensors and other available low-cost mechanisms to effectively monitor the hygiene level and generate alerts

II. Existing System

IIT Kharagpur Invents Low-Cost Smart Hygiene Monitor to Alert Authorities About Cleanups

A smart device to monitor levels of dirt and send out alerts to stakeholders when things need attention for cleaning up, that is what a group of researchers of IIT Kharagpur have developed. The low-cost Smart Hygiene Monitor may soon be used to keep the public facilities clean and maintain hygiene standards at all

times. This low-cost dirt detector can be placed in public/community toilets, clinics, hospitals, operation theatres, kitchens and so on and is capable of recording the levels of ammonia, Sulphur dioxide, carbon monoxide, carbon dioxide, volatile organic compounds, and particulate matter with the help of sensors. In simple terms, the device can sense how dirty an area is and send online alerts to the authorities to clean it up. The research team behind the development of this smart device was headed by Aurobindo Routray, a professor in the Electrical Engineering department of the institute. Dr. Routray has been teaching in the college since 1991 and has been doing this research for the last one and half years. The prototype of the device was launched recently and it is yet to undergo field trials that will help evaluate its practical application and efficiency [2].

III. Proposed System

Our proposed system consists of Arduino along with multiple sensors that are deployed at toilet level which will monitor and track the status of the available public toilets based on the identified parameters that are responsible for the health and hygiene of the toilets and if the threshold values of these parameters are breached, then an alert message will be sent the sweeper responsible for that particular toilet and also the corresponding values will be sent to the web server that can be visualized with the help of a dashboard that can be used by the centralized admin to monitor the toilet. Figure 1 depicts the workflow of our proposed system.

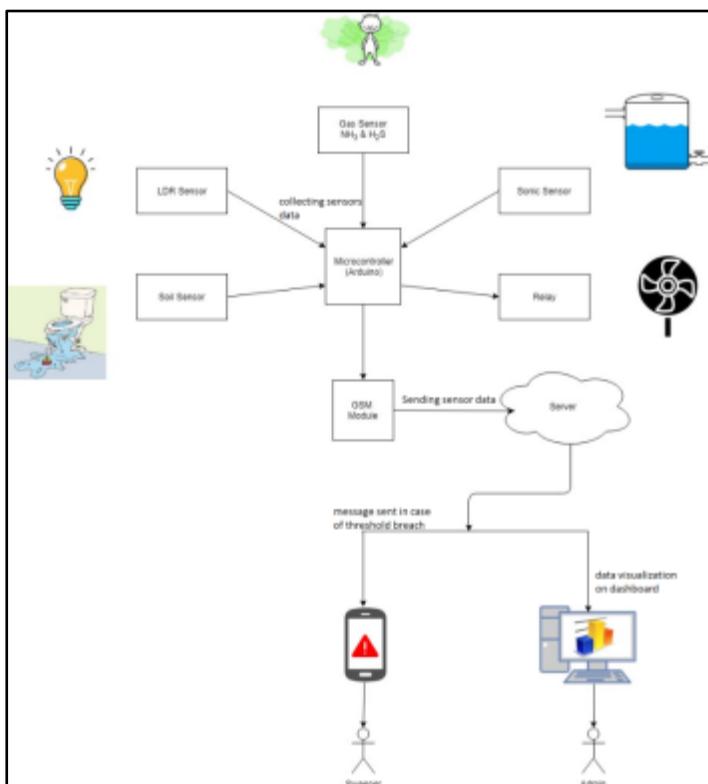


Figure 7: Workflow of proposed system

III. Methodology

The tracking and monitoring of the toilets are based on following defined parameters that contributes to the health and hygiene of the toilet:

Table 1: Parameters responsible for health and hygiene.

Parameters (Problems)	Description
Foul Smell	Most of the toilets have bad odour that makes them reluctant to visit
Shortage of Water	Water to wash after defecating and also for washing hands.
Blockage of Septic Tanks	Blockage that would make water rise through the commode.
Lights	No proper functioning of the lights.
Soap and Handwash	Inadequate quantity and non-availability.
Cleaners Activity	No proper monitoring of untidy toilets.

The above-mentioned parameters are tracked by the various sensors that are connected with Arduino and are as follows:

Arduino will have following sensors to tackle various parameters:

- The LDR sensor is used to detect the presence of any person in the toilet which can be also used as a counter to keep the count of people using the toilet, once a certain limit of usage is exceeded the sweeper will be notified to clean the toilet. Also, the count of the people visiting the toilet is used to estimate the of availability of the soap and alerts once the limit is exceeded.
- The MQ135 sensors are used to detect the NH_3 and H_2S gas which are present in the urine and the faeces which causes bad odour in the public toilet.
- The relay is used to control the fan whenever the NH_3 and H_2S gas level rises the fans would be turned on automatically.
- The sonic sensor is used to keep the track of the water level in the water tank and soil sensor is used to keep the track of the sludge level in the septic tank if they exceed the prescribed threshold value the sweeper will be alerted by the server with a message to his phone about the problem.

The entire process of tracking, monitoring and generating alerts involves following two steps:

A. Tracking and sending data from Arduino:

The Arduino is initialized which powers on all the sensors. These sensors monitor the parameters which they are defined for and works simultaneously.

- If the air quality of toilet is bad, then the MQ15 sensor, based on the defined threshold for distinguishing between good or bad air quality, triggers the alert once the threshold is breached.
- If the water level in tank falls below 50% in the tank, then the ultrasonic sensor which keeps track of this, triggers the alert.
- If the septic tank exceeds the safety level, then the soil moisture sensor will detect that and trigger the alert.
- If the lighting ambience level in the toilet falls below the predefined threshold, then an alert is triggered.
- If the quantity of handwash falls to empty based on the count of people using it, then an alert is triggered.

All these alerts which are triggered by various sensors when their threshold is breached are sent to a web server from the Arduino via GSM module. Figure 2 shows the prototype model of our proposed system.

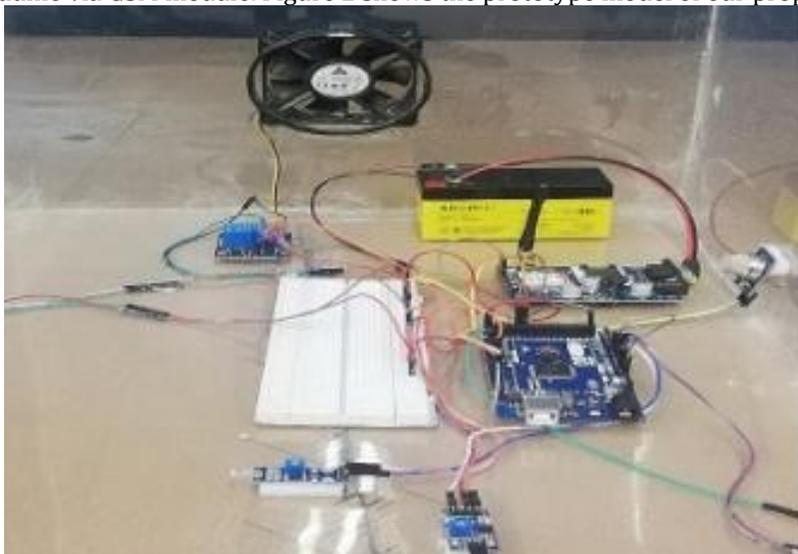


Figure 8: Prototype of proposed system.

B. Monitoring the toilet status on dashboard:

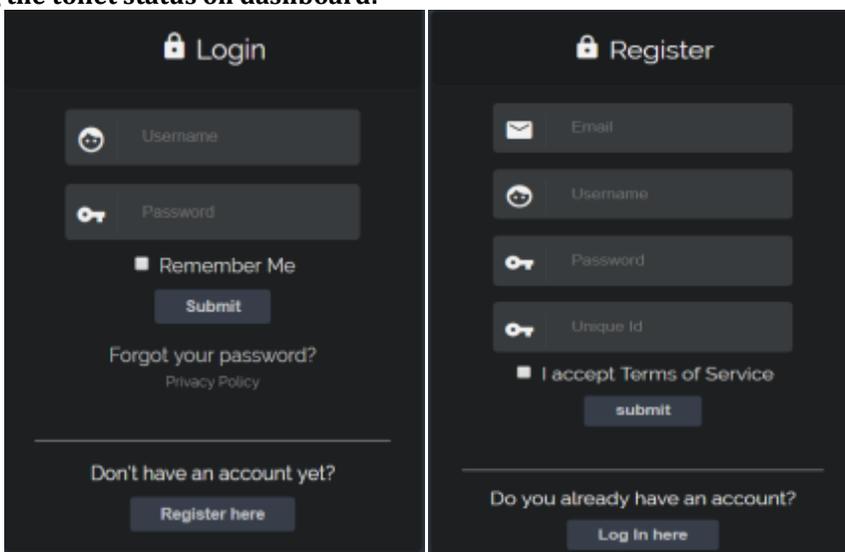


Figure 3: Admin login and register.

Figure 3 shows the admin registration and login and fig. 4 shows the dashboard functionality which contains the “Edit Toilet” option by which we can add a new toilet, delete the existing toilet from the dashboard, also we have the sweeper registration option to register the sweeper

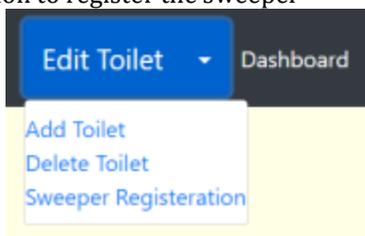


Figure 4: Modify toilet.

Figure 5 shows the admin dashboard which consists of current readings of the sensors for individual toilets. The parameters which are displayed on the dashboard home page turns red if the threshold value of that parameter is breached hence it remains green.

SR NO	NAME	TIMESTAMP	AIR QUALITY	NO OF PEOPLE	WATER TANK	SEPTIC TANK	LIGHT STATUS	HANDWASH STATUS	CURRENT STATUS	RATING
1	toilet1	2019-02-24 10:20:30	BAD	0	75.22 %	NOT-CLOG	WORKING	NOT-EMPTY	DIRTY	★★★★★
2	toilet2	2019-02-13 11:58:33	GOOD	45	56 %	CLOG	WORKING	EMPTY	DIRTY	★★★★★
3	toilet3	2019-02-13 08:23:25	GOOD	78	67 %	NOT-CLOG	WORKING	NOT-EMPTY	CLEAN	★★★★★
4	toilet4	2019-02-13 08:58:29	GOOD	86	35 %	CLOG	WORKING	NOT-EMPTY	DIRTY	★★★★★
5	toilet5	2019-02-11 08:59:11	BAD	43	29 %	CLOG	WORKING	NOT-EMPTY	DIRTY	★★★★★

Figure 5: Admin dashboard homepage.

$$\text{Toilet Rating} = \frac{\text{Total number of times the toilet current status is Dirty}}{\text{Total number of entries}}$$

Each toilet is assigned ratings based on the above-mentioned formula. It helps the admin to focus on the best and worst performing toilets.

On clicking the current status of any individual toilet, the user gets directed to the next page containing the detail information of that toilet as shown in fig. 6.

SR NO	TIMESTAMP	AIR QUALITY	COUNT	WATER TANK	SEPTIC TANK	LIGHT STATUS	HANDWASH STATUS
1	2019-02-01 21:15:13	GOOD	34	23%	CLOG	WORKING	EMPTY
2	2019-02-02 21:16:32	GOOD	45	100%	CLOG	WORKING	EMPTY
3	2019-02-03 21:16:51	GOOD	45	85%	NOT-CLOG	WORKING	EMPTY
4	2019-02-04 02:10:52	GOOD	45	55%	NOT-CLOG	WORKING	NOT-EMPTY

Figure 6: Detailed sensor data.

In this page, all the sensors data is from the start of the day is shown and also there is an option for the user to send the text messages to the sweeper of that toilet in case of non-responsiveness of sweeper. SMS feature is implemented using the **Text Local API**.

SR NO	TIMESTAMP	LDR COUNTER	SONIC SENSOR	MQ135 SENSOR	MOISTURE SENSOR	LDR LIGHT SENSOR
1	2019-02-01 16:47:39	WORKING	WORKING	WORKING	WORKING	WORKING
2	2019-02-01 17:07:15	WORKING	WORKING	WORKING	WORKING	WORKING
3	2019-02-02 17:09:29	NOT WORKING	WORKING	NOT WORKING	NOT WORKING	WORKING
4	2019-02-03 17:09:29	NOT WORKING	WORKING	NOT WORKING	NOT WORKING	WORKING

Figure 7: Sensor maintenance status.

Figure 7 displays the maintenance aspect of the system is unique. It shows whether all the sensors are working properly or not.

Sweeper	Description
Profile	
Name :	John snow (shai)
Address	North,Behind the wall,Thane(w) 400-607
Rating	★★★★★
Working Hours	0hrs
AreaUnder	North
ph-no	9879393042

Figure 8: Sweeper details.

$$\text{Sweeper Rating} = \frac{\text{Total number of early response} (\leq 15\text{min})}{\text{Total number of response}}$$

Figure 8 depicts sweeper details. The sweeper rating is also calculated based on the above-mentioned formula which keeps into account, the response time of the sweeper.

Figure 9 shows the response of the sweeper can be analyzed graphically as well for the better understanding of the admin.



Figure 9: Sweeper response analysis

SR NO	ALERT TIMESTAMP	RESPONSE TIMESTAMP	TIME TAKEN	TIME(MIN)	PERFORMANCE
1	2019-02-11 02:11:56	2019-02-11 02:13:36	0 d 0 hr 1 min	1.67 min	EXCELLENT 😊
2	2019-02-12 02:13:51	2019-02-12 02:14:08	0 d 0 hr 0 min	0.28 min	EXCELLENT 😊
3	2019-02-13 02:14:29	2019-02-13 02:16:32	0 d 0 hr 2 min	2.05 min	EXCELLENT 😊
4	2019-02-14 02:16:53	2019-02-14 02:17:32	0 d 0 hr 0 min	0.65 min	EXCELLENT 😊
5	2019-02-15 02:19:17	2019-02-15 02:25:58	0 d 0 hr 6 min	6.65 min	EXCELLENT 😊

Figure 10: Sweeper response and rating status.

An alert is generated whenever any event occurs which makes the toilet's current status dirty. Hence the timestamp of that alert is recorded and the timestamp at which the toilet's current status turns clean is recorded and the difference of this timestamps is used to measure the performance of the sweeper which is depicted by fig. 10 in detail.

$$Difference = \frac{Alert\ Timestamp}{Response\ Timestamp}$$

- If Difference ≤ 15 min then, performance is “EXCELLENT”
- If Difference >15 & ≤ 30 min then, performance is “GOOD”
- If Difference >30 & ≤ 45 min then, performance is “AVERAGE”
- If Difference >45 & ≤ 60 min then, performance is “POOR”

If the response timestamp is not recorded then it means the user has not responded yet.

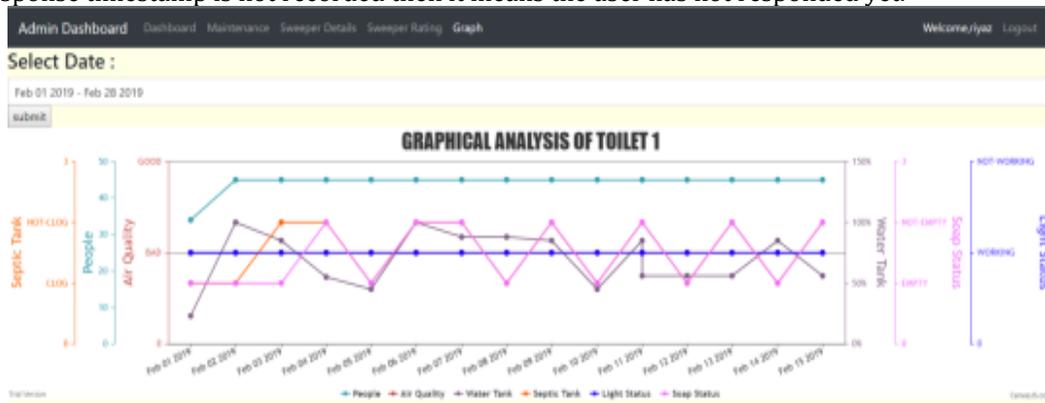


Figure 11: Graphical analysis of toilet.

The sensors data is also shown in the form of line graph in fig. 11 for better visualization of the overall status of the toilet

V. Future scope

We can implement features like automating the toilet lights which will save the electricity consumption. We can implement smart bin, in which an alert is sent to the sweeper when the dustbin is about to overflow. Smoking cigarettes and drinking alcohol in the toilets can be detected using specialized sensors. The necessity of additional toilets can also be recommended by tracking the record of the people using the toilet.

VI. conclusion

As per the proposed idea, a prototype system is built which automatically monitors the hygiene level of the toilets and raises an alert when the corresponding thresholds are breached. The system was tested with a pre-determined concentration of ammonia gas to imitate urine stench and so the water level in the tank was kept at above threshold. The lighting ambience and the count of people were also taken into consideration while testing the system. The system performed on an above average level thereby satisfying the overall aspect of the system goals.

VII. references

- [1] <https://www.pri.org/stories/2015-10-06/8-million-new-toilets-later-india-isnt-much-cleaner>.
- [2] <https://swachhindia.ndtv.com/iit-kharagpur-invents-low-cost-smart-hygiene-monitor-to-alert-authorities-about-clean-ups-11053/>
- [3] Osathanunkul, Kitisak & Hantrakul, Kittikorn & Pramokchon, Part & Khoenkaw, Paween & Tantitharanukul, Nasi. 2017. Configurable automatic smart urinal flusher based on MQTT protocol. 58-61. 10.1109/ICDAMT.2017.7904934.
- [4] Schlebusch T., Leonhardt S. 2011. Intelligent Toilet System for Health Screening. In: Hsu CH., Yang L.T., Ma J., Zhu C. (eds) Ubiquitous Intelligence and Computing. UIC 2011. Lecture Notes in Computer Science, vol 6905. Springer, Berlin, Heidelberg.
- [5] E. Elakiya, K. Elavarasi, R. P. Kaaviya Priya, 2018, Implementation of Smart Toilet (Swachh Shithouse) Using IOT Embedded Sensor Devices, International Journal of Technical Innovation in Modern Engineering & Science (IJTIMES), Volume 4, Issue 4, pp 65- 74.
- [6] K. Osathanunkul, K. Hantarkul, P. Pramokchon, P. Khoenkaw and N. Tantitharanukul. 2016. Design and Implementation of an Automatic Smart Urinal Flusher, International Computer Science and Engineering Conference (ICSEC2016), Chiang Mai, Thailand, pp 14-17.

RENTKARDE- MOBILE BASED ANDROID APPLICATION.

Er. Farida Attar, Mulla Mohammed Sufiyan, Khan Danish, Ghadi Ashish.

Asst. Prof. IT Dept. at M. H Saboo Siddik College of Engineering, Student, Student , Student

Department of Information Technology

M. H. Saboo Siddik of College Engineering, Mumbai, India

ABSTRACT: RENTKARDE is an Android based application that provides renting facilities. There would be two actors playing their part in this product, first the person Renter who rents his products which will be verified and further make it available to Purchaser to buy through its Mobile Application. The other actor will be Purchaser that will look through the items that are available on Application and wish to rent it for a particular time. Authenticity of Purchaser will be checked and provide him with the renting service. For the Renter he will upload the items that he wish to rent and provide a description about the product along with the amount of time he/she wants the product to be rented. The Purchaser will go through the products and choose the items mentioned in different Categories such as Clothing, Furniture, Electronics, Books, etc. Then he/she will provide the span of time they would like to rent it that would be in categories Daily, Weekly, Monthly. According to the time assigned by Purchaser the price would be decided and the product would be made off to rent. There will also be a Tracking of items both on Renter and Purchaser side.

Keywords: Android, RentKarde, Renting Facilities, Renters, Purchaser.

I. INTRODUCTION

The aim of the project is to develop an Android application that provides a platform for establishing the renting services. RENTKARDE is the company that will provide the platform in form of Android Application through which the Customers/User could allow their items to get rented or take items for renting purposes. There would be two actors playing their part in this product first the person Renter who rents his items that would be sent off to the Company and the Company will verify the Renter's and products Authenticity and further make it available to Purchaser to rent through its Web Application or Mobile Application. The other actor will be Purchaser that will look through the items that are on Company's Application and wish to rent it for a particular time. Company will check the Authenticity of Purchaser and provide him the renting service. For the Renter he will upload the items that he wish to rent and provide a description about the product along with the amount of time he/she wants the product to be rented. The Purchaser will go through the website and choose the items mentioned in different Categories such as Clothing, Furniture, Books, etc. Then he/she will provide the span of time they would like to rent it that would be in categories Hourly, Weekly, Monthly. According to the time assigned by Purchaser the price would be decided and the product would be made off to rent. There will also be a Tracking of items both on Renter and Purchaser side. The purpose of the applications to benefit the people who do not want an item to buy due to its high cost and also the reason that they would not use that item regular which will then be wasted of their price that they paid. It also benefits the person who is renting its item to the company by earning some money in return. Also a Business for the company through providing these renting services.

II. EXISTING SYSTEM

AMAZON-

Amazon provides means to rent books through Textbook Rental Store. You are allowed to rent a book through Textbook Rental Store for 30 days period after using the book you can return it. To return a textbook after the first 30 days of the rental period, simply go to Your Account on Amazon.com and select Your Textbook Rentals to view your rental library. Select the textbook you wish to return and then select "Return books" to print the return shipping label. It also have an application for books known as "Kindle" Amazon now loans you free books, as long as you're a Prime member. You can borrow one book a month by going to the "Kindle Owners' Lending Library" in the Kindle Store on your Kindle device. Books with the Prime icon can be borrowed for free for the month. To have advantage to this renting service you need to be a Prime member so this barely justify as a renting service. But they limit upto renting Books.

FURNITURE RENTING APPLICATION-

Furlenco and Fabrento are the well known application to rent a furniture. They provide renting service of furniture with good discounts and offers across India. They are in business when we talk about dealing with good furniture and interior designing. But the only limit upto renting furniture.

RENTGAMES:

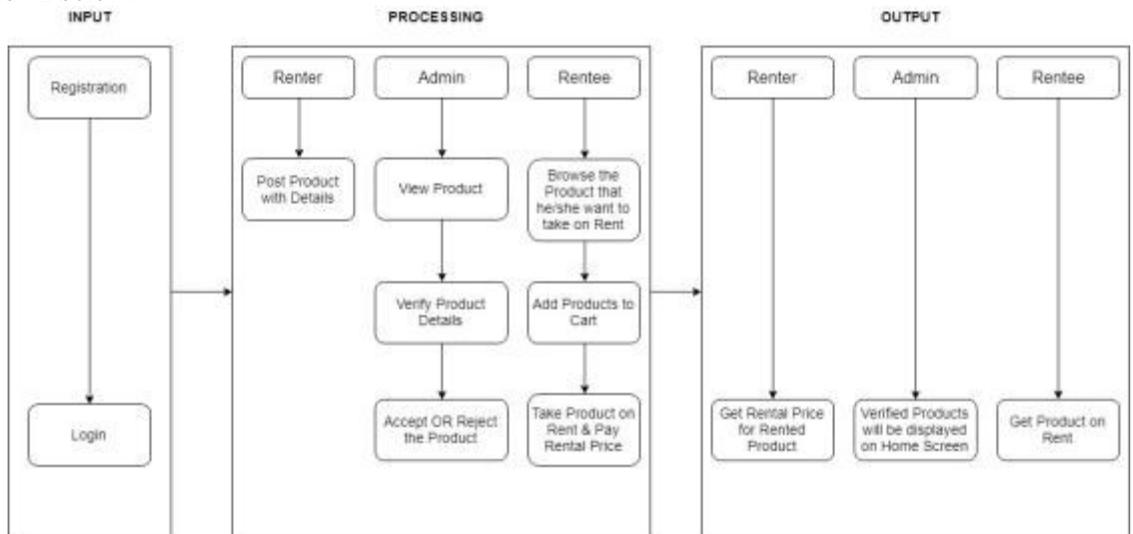
This application is well known for renting games both electronic and non electronic. To apply for renting service you need to be a member of the rent games community by registering on their website or having their Android application. You need to choose which type of membership plan that is available depending on the price limit assigned to each membership plan. They will provide one game at time. But this application limits upto renting only gaming console.

Advantages

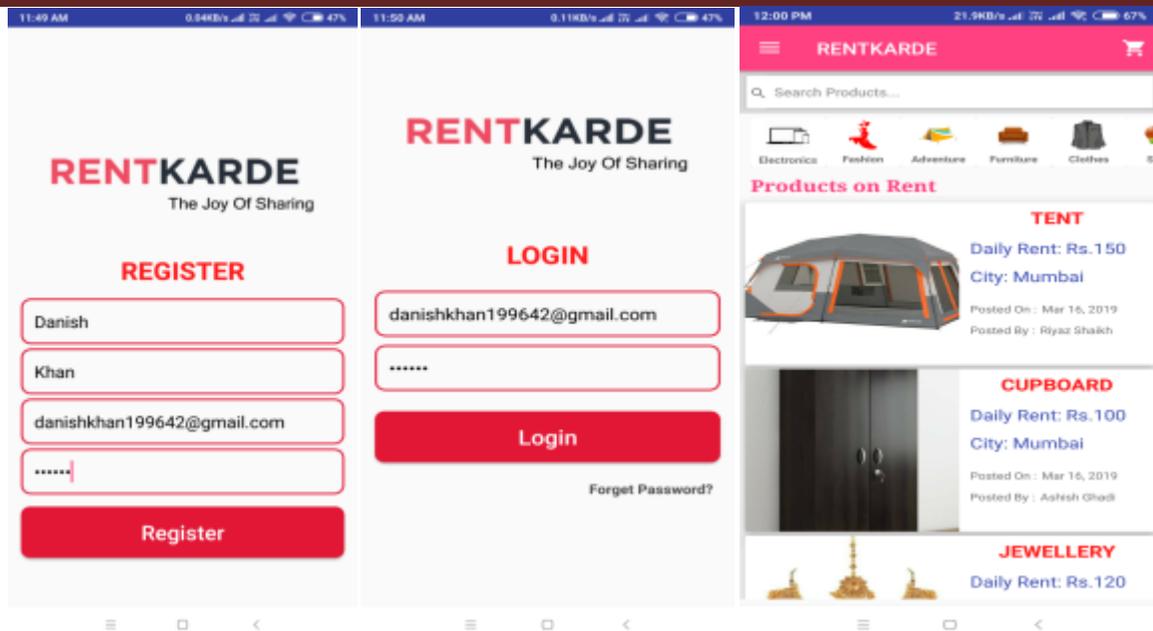
- (1) These renting applications are famous and well known in business market.
- (2) These renting applications have their own maintenance system and have their own stock that they send for renting.
- (3) They have good and customize application and websites that play an important role in their marketing strategies.

III. PROPOSED SYSTEM

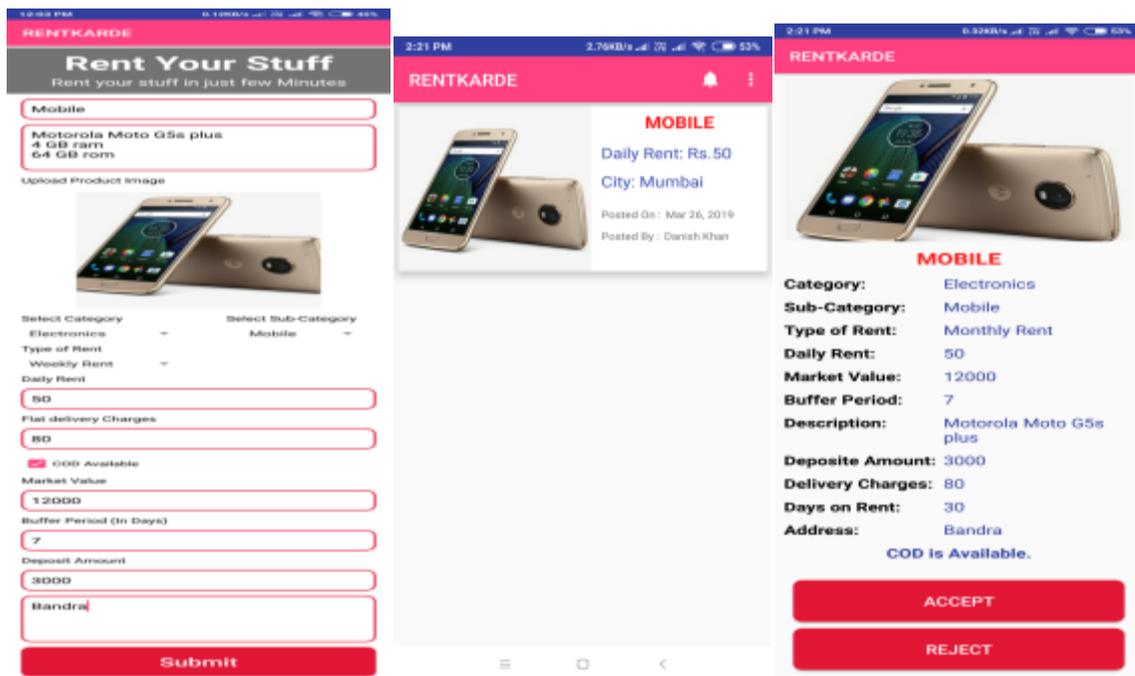
There are three important phases of the system Input, Processing, Output. In Input phase the system will take the Registration details and Login Credentials. Processing part will involve two parties Renter and the Rentee or Purchaser. Renter will add the product that they want to make available for renting along with its details and amount of time product is available for renting. Purchaser will browse and select the product they want to rent add the details like how long they want the product. Output phase will involve the Renter getting the money for rent and Rentee the product that he choose for themselves. The System Flow Diagram is shown below.



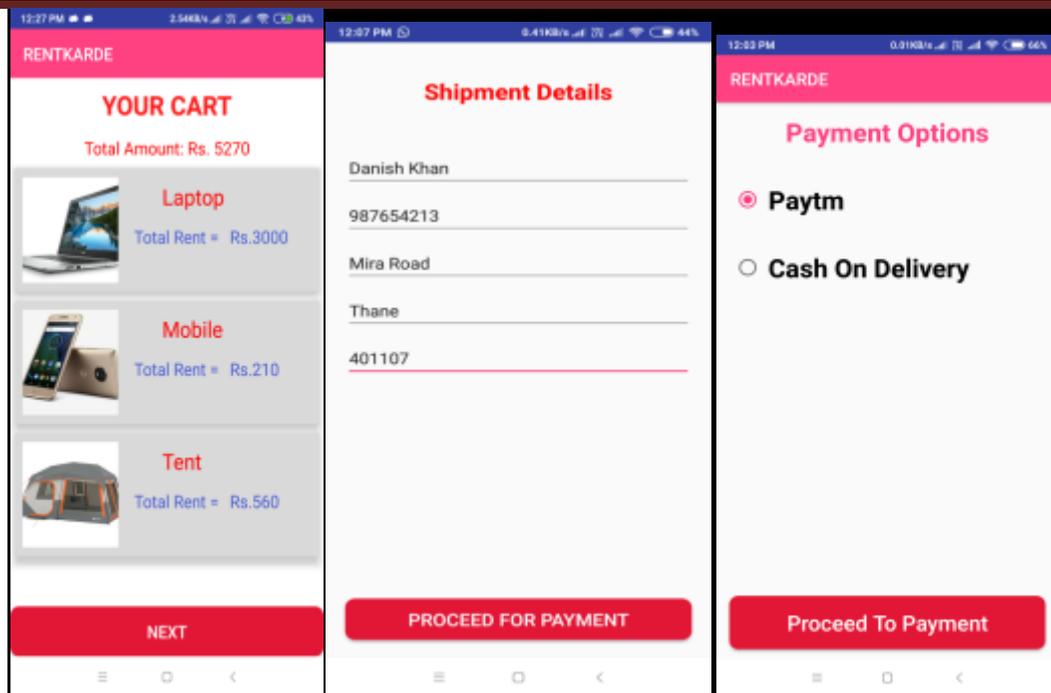
The system will begin with the user registering to the application where he/she enters its Name, Phone Number, Email ID and username and password. After registration user will login into the application with its credentials and then user can act as one of two parties the Renter and the Purchaser. The Home Screen will consist of the list of products available for renting by different users. The list of products can also be available for viewing in forms of categories such as Electronics, Books, Clothing, Furniture etc.



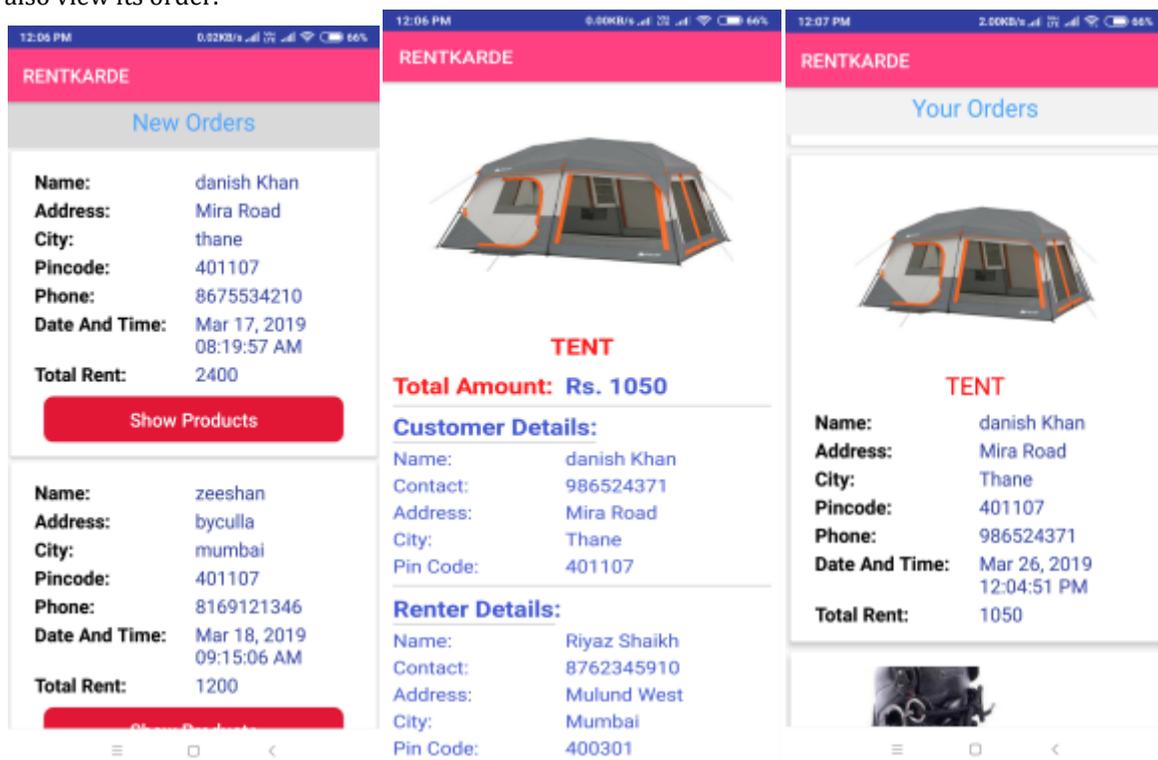
The application Drawer has Rent your stuff, Orders, Categories. By selecting renting your stuff new activity opens which has Title, Description, Uploading image, Selecting categories, type of rent, Daily rent, Delivery Charges, market value, Buffer Period, Deposit Amount, Address. After the product is submitted by the Renter the product goes to verification on Admin panel then the Admin verifies the product and according to that he choose to accept or reject the product. Then if the product is accepted it is displayed on the Home screen of the application.



The Purchaser or the rentee will add the product that they want to get for rent to the cart or directly place the order. They need to enter the shipment details and also choose the method of payment and then proceed the payment.



Once the order is placed, the Admin will get the notification of the order. The Admin will be able to see the details of the products including the address of the Renter and the Purchaser which will be displayed and the total Rent amount. The orders placed by the users will be seen in his Orders activity. The Purchaser can also view its order.



IV. FUTURE SCOPE

The major goal of this application is to provide platform to have renting facilities. RENTKARDE also has an API that during registration the User is asked to upload the scan Aadhar-card for proper authenticity of User/Customer. The privacy of the User/Customer are maintained. User/Customer can easily login by

entering their ID and Password which were set during the registration and most importantly no breaching of data will be done. The system has a unique feature of online shopping where the users can rent mobile phones and laptops, games, clothing, books and many other product. The above-proposed model is easy to implement considering the ASP model. The model is simple, secure and reliable.

This application can be improved in the future by adding the following functionalities:

- Extending to Apple iOS.
- Biometrics Registration and Login.
- Providing services for a wider range of devices.

V. CONCLUSION

Development of market and business domain application with integration of different features is a true form of challenge. Deployment of such systems may lead to operational bugs and fixes. The proposed system aims to provide a method of communication between two people for renting their products. Also the application will have monitoring capability towards both product and its user. The technologies suggested for implementation of the proposed system are also evidently easy to implement and error correcting. Bug fixing and error correction in such technologies is not a troublesome task. Providing the rating system to user for purpose of users sincerity and avoiding suspicious and many more functionalities the urged system proves instrumental in providing a good platform for renting the products.

VI. REFERENCE

1. <https://in.linkedin.com/company/rentkarde>
2. <https://www.justdail.com/mumbai/rentkarde>
3. <https://www.zaubacorp.com/company/RENTKARDE-SERVICES-PRIVATE-LIMITED/U72900MH2016PTC282834>
4. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.454.9931&rep=rep1>

Representation of finite transforms and its relations

V. R. Lakshmi Gorty

Professorin Mathematics, Basic Science and Humanities, SVKM's NMIMS University, MPSTME
Mumbai-400056, India.

ABSTRACT: In the present paper, the author considers several new finite integral transforms like finite Hankel transforms, finite Hankel Clifford transforms, finite Mellin, finite Fourier transforms and many more to develop relations between the transforms and its representations as twin transforms. Representations involving these finite transforms with new transforms are shown. Some examples are also given as illustrations of the results studied here.

Keywords: finite Hankel transforms, finite Laplace transforms, finite Fourier transforms, finite Mellin, relations.
AMS Classification: 44A10, 44A15, 44A20, 34A30.

I. Introduction

Finite Hilbert Transform was defined by [1] as:

$$S(f(x); y) = \int_0^y \frac{1}{y-x} f(x) dx. \quad (0.1)$$

Finite Widder-like transforms stated here as:

$$P(f(x); y) = \int_0^y \frac{2x}{y^2 - x^2} f(x) dx \quad (0.2)$$

for order 1.

$$P_{y-1}(f(x); y) = \int_0^y \frac{2x^{y-1}}{y^2 - x^2} f(x) dx \quad (0.3)$$

for the order $y-1$.

Finite Laplace transforms defined in [2] as:

$$L\{f(x); y\} = \int_0^1 e^{-xy} f(x) dx. \quad (0.4)$$

Sneddon in [2] defined finite Hankel transforms as given by:

$$h_{1,\nu}(f(u); y) = \int_0^1 u J_\nu(uy) f(u) du. \quad (0.5)$$

Finite Hankel-Clifford transforms is stated in [3] as:

$$H(f(x); u) = \int_0^1 u^\nu (ux)^{-\nu/2} J_\nu(2\sqrt{xu}) f(x) dx. \quad (0.6)$$

Integral representation of Kummer functions given by Abramowitz et al [pp. 558, 4] as

$$M(a, b, z) = \frac{\Gamma(b)}{\Gamma(a)\Gamma(b-a)} \int_0^1 e^{-zu} u^{a-1} (1-u)^{b-a-1} du. \quad (0.7)$$

Kummer transformation as related in [4] given by:

$$M(a, b, z) = e^z M(b-a, b, -z). \quad (0.8)$$

$$e^z M(b-a, b, -z) = \frac{\Gamma(b) e^z}{\Gamma(a)\Gamma(b-a)} \int_0^1 e^{-zu} u^{b-a-1} (1-u)^{a-1} du. \quad (0.9)$$

Finite Fourier cosine transform and finite Fourier cosine transform from [2] can be written as:

$$F_c(f(x); y) = \int_0^y \cos\left(\frac{u\pi x}{y}\right) f(x) dx. \tag{0.10}$$

$$F_s(f(x); y) = \int_0^y \sin\left(\frac{u\pi x}{y}\right) f(x) dx. \tag{0.11}$$

Beta function in [5] is given by:

$$\int_a^b (x-a)^m (b-x)^n dx = (b-a)^{m+n+1} \beta(m+1)(n+1). \tag{0.12}$$

The incomplete Beta function given by [2]:

$$\beta_x\{f(u); y\} = \int_0^x u^{c-y-1} (1-u)^{y-2} du. \tag{0.13}$$

Finite Mellin transforms is defined from [2] as:

$$M\{f(x); y\} = \int_0^1 (u^{-x-1} - u^{x-1}) f(x) dx. \tag{0.14}$$

New finite transforms are be stated as:

$$\varpi\{f(x); u\} = \int_0^1 x u^{x^2} f(x) dx. \tag{0.15}$$

$$N\{f(u); y\} = \int_0^y y u^{-y^2-1} f(u) du. \tag{0.16}$$

$$\Theta\{f(u); y\} = \int_0^y u^y f(u) du. \tag{0.17}$$

$$\diamond(f(x); y) = \int_0^1 M(a, b, y-x) f(x) dx. \tag{0.18}$$

$$\Xi(f(u); x) = \int_0^1 u^{\frac{1}{y^2-x^2}-1} du. \tag{0.19}$$

In this study the relations between finite transforms and its relations with other finite transforms have been established.

II. MAIN RESULTS

Lemma 2.1. The Kummer transforms of finite Laplace transforms from (1.4) can be related as

$$M\{L\{f(x); y\}; u\} = \int_0^1 e^{-(x-y)} \left(e^{(x-y)} M(b-a, b, -(x-y)) \right) f(x) dx. \tag{2.1}$$

Proof.

$$M\{L\{f(x); y\}; u\} = \frac{\Gamma(b)}{\Gamma(a)\Gamma(b-a)} \int_0^1 \int_0^1 e^{-(x-y)u} u^{a-1} (1-u)^{b-a-1} du f(x) dx.$$

Since

$$e^p M(b-a, b, -p) = \frac{\Gamma(b) e^p}{\Gamma(a)\Gamma(b-a)} \int_0^1 e^{-pu} u^{b-a-1} (1-u)^{a-1} du.$$

If $\text{Re}(b) > \text{Re}(a) > 0$, then Kummer's functions are related by Kummer's transformations as in (1.7). Thus gives

$$M(a, b, (x-y)) = e^{-xy} M(b-a, b, -(x-y)).$$

Lemma 2.2. The Kummer's transformations of finite Laplace transforms is represented by

$$M\{L\{f(x); u\}; y\} = \diamond(f(x); y) \tag{2.2}$$

where (1.4), (1.7) and (1.18) exists.

Proof. Here left hand side of (2.2) is represented as

$$\begin{aligned} M\{L\{f(x); u\}; y\} &= \int_0^1 \frac{\Gamma(b)}{\Gamma(a)\Gamma(b-a)} \left(\int_0^1 e^{-(x-y)u} u^{a-1} (1-u)^{b-a-1} du \right) f(x) dx. \\ &= \int_0^1 M(a, b, x-y) f(x) dx. \\ &= \diamond(f(x); y). \end{aligned}$$

Lemma 2.3. The representation for

$$N\{\varpi\{f(x); y\}; u\} = \frac{y}{2} P\{f(x); y\} \tag{2.3}$$

hold true provided each member of the assertions (1.15) and (1.16) exists.

Proof.

$$\begin{aligned} N\{\varpi\{f(x); y\}; u\} &= \int_0^y y u^{y^2} \left(\int_0^1 x u^{-x^2-1} f(x) dx \right) du \\ &= \int_0^y x y \left(\int_0^1 u^{y^2-x^2-1} du \right) f(x) dx \\ &= y \int_0^y \frac{x}{y^2-x^2} f(x) dx \\ &= \frac{y}{2} P\{f(x); y\}. \end{aligned}$$

Thus $N\{\varpi\{f(x); y\}; u\}$ can be represented as $\frac{y}{2}$ times the finite Widder-like transforms.

Lemma 2.4. The representation

$$\Theta\{M\{f(x); y\}; u\} = P(f(x); y) \tag{2.4}$$

exists considering (1.17) holds true, where $M\{f(x); y\}$ is finite Mellin transform of the first kind as in (1.14).

Proof.

$$\begin{aligned} \Theta\{M\{f(x); y\}; u\} &= \int_0^y \left(\int_0^1 u^{y-x-1} - u^{y+x-1} du \right) f(x) dx \\ &= \int_0^y \left(\frac{1}{y-x} - \frac{1}{y+x} \right) f(x) dx \\ &= P(f(x); y). \end{aligned}$$

Lemma 2.5. The incomplete beta-functions given in [6] is now extended in a new representation as

$$\int_0^x u^{c-a-1} (1-u)^{a-2} du = - \sum_{a=0}^{c-a} \frac{\Gamma(c) x^{c-a-1} (1-x)^{a-1}}{\Gamma(a)\Gamma(c-a)}. \tag{2.5}$$

$$\int_0^x u^{c-a-1} (1-u)^{a-2} du$$

$$= \left[\begin{aligned} & u^{c-a-1} \frac{(1-u)^{a-1}}{(-1)(a-1)} \\ & - (c-a-1) u^{c-a-2} \frac{(1-u)^a}{(-1)^2 (a-1)(a)} + \dots \\ & + (c-a-(c-a)) u^{c-a-(c-a)-1} \\ & \times \frac{(1-u)^c}{(-1)^{(c-a)-1} (a-1)(a) \dots (a+c-a)} \end{aligned} \right]_0^x$$

$$= - \sum_{a=0}^{c-a} \frac{\Gamma(c) x^{c-a-1} (1-x)^{a-1}}{\Gamma(a) \Gamma(c-a)}.$$

Thus $\int_0^x u^{c-a-1} (1-u)^{a-2} du = - \sum_{a=0}^{c-a} \frac{\Gamma(c) x^{c-a-1} (1-x)^{a-1}}{\Gamma(a) \Gamma(c-a)}$ can be is represented.

Lemma 2.6. The finite Laplace transforms of incomplete beta-functions can be represented as certain sum of Kummer transformations provided $c - y > 0$ and $y > 0$ and $0 \leq x \leq 1$ is given by:

$$e^z L \{ \beta_x \{ f(u); y \}; x \} = - \sum_{y=0}^{c-y} e^y M(c-y, c, -z). \tag{2.6}$$

Proof. From (1.3), the finite Laplace transforms is considered as:

$$L \{ \beta_x \{ f(u); y \}; x \} = \int_0^1 e^{-xz} \left(\int_0^x u^{c-y-1} (1-u)^{y-2} du \right) f(x) dx.$$

$$= - \int_0^1 e^{-xz} \left(\sum_{y=0}^{c-y} \frac{\Gamma(c) x^{c-y-1} (1-x)^{y-1}}{\Gamma(y) \Gamma(c-y)} \right) f(x) dx.$$

$$e^z L \{ \beta_x \{ f(u); y \}; x \} = - e^z \int_0^1 e^{-xz} \left(\sum_{y=0}^{c-y} \frac{\Gamma(c) x^{c-y-1} (1-x)^{y-1}}{\Gamma(y) \Gamma(c-y)} \right) f(x) dx.$$

$$= - \sum_{y=0}^{c-y} \frac{e^z \Gamma(c)}{\Gamma(y) \Gamma(c-y)} \left\{ \int_0^1 e^{-xz} x^{c-y-1} (1-x)^{y-1} f(x) dx \right\}.$$

$$= - \sum_{y=0}^{c-y} e^z M(c-y, c, -z).$$

Hence proved.

Corollary 2.7. : The finite Mellin transforms can be represented from Laplace transforms of infinite range. For a given Laplace transforms [2] of the form:

$$L[f(x); s] = \int_0^\infty e^{-sx} f(x) dx, \tag{2.7}$$

for s being a parameter.

Proof. Let $x = -\log\left(\frac{t}{a}\right); dx = -\frac{dt}{t}$.

For $x = 0; t = a$ and $x = \infty; t = 0$.

Then from (2.7), implies

$$L[g(t); s] = \int_0^a a^{-s} t^{s-1} g(t) dt,$$

denoted as new representation of finite Mellin transform as:

$$M\{g(t); y\} = \int_0^a a^{-y} t^{y-1} g(t) dt. \tag{2.8}$$

When $a = 1$, (2.8) can be written as:

$$M_F\{f(x); y\} = \int_0^1 x^{y-1} f(x) dx. \tag{2.9}$$

Lemma 2.8. The Finite Mellin transforms of (1.19) kind of expression is equal to $\frac{1}{2}$ of finite Widder-like transforms of order $(y-1)$. Then mathematically given by:

$$M_F\left(\left(\Xi(f(u)); x\right); y\right) = \frac{1}{2} P_{y-1}\{g(x); y\}. \tag{2.10}$$

Proof. The finite Mellin transforms is given by (1.14) and new transforms considered from (1.19) the lhs of (2.10) becomes:

$$\begin{aligned} M_F\left(\left(\Xi(f(u)); x\right); y\right) &= \int_0^y x^{y-1} g(x) \left(\int_0^1 u^{y^2-x^2-1} du \right) dx \\ &= \int_0^y \frac{x^{y-1}}{y^2-x^2} g(x) dx \\ &= \frac{1}{2} P_{y-1}\{g(x); y\}. \end{aligned}$$

Hence $M_F\left(\left(\Xi(f(u)); x\right); y\right) = \frac{1}{2} P_{y-1}\{g(x); y\}$ is proved.

Lemma 2.9. The Finite Hilbert Transform from (1.1) and finite Mellin transforms from (2.9) is represented as

$$i) S(y\pi M_F(f(x); y)u) = M_F(f(x); y). \tag{2.11}$$

$$ii) M_F\{S\{u f(u); x\}y\} = M_F\{f(u); y\}. \tag{2.12}$$

Proof.

i)

$$\begin{aligned} &S(yM_F(f(x); y)u) \\ &= \int_0^y \frac{1}{y-u} \left(\int_0^1 y x^{y-1} f(x) dx \right) du \\ &= \left(\int_0^1 x^{y-1} f(x) dx \right) \\ &= M_F(f(x); y). \end{aligned}$$

Thus it proved that Finite Hilbert transforms of Finite Mellin transforms is Finite Mellin transform itself.

ii) To prove $M_F \{S\{u f(u); x\}; y\} = M_F \{f(u); y\}$, the lhs of (2.12) is defined by:

$$\begin{aligned} &M_F \{S\{u f(u); x\}; y\} \\ &= \left(\int_0^1 x^{y-1} \right) \left(\int_0^u \frac{1}{u-x} f(u) du \right) dx \\ &= \int_0^1 \left(\int_0^u \frac{x^{y-1}}{u-x} dx \right) f(u) du \\ &= \int_0^1 u^{y-1} f(u) du \\ &= M_F \{f(u); y\}. \end{aligned}$$

It proved that finite Mellin transforms of finite Hilbert transforms is finite Mellin transform itself.

Lemma 2.10. Finite Hankel-Clifford transforms (1.6) of finite Fourier cosines transforms (1.10) is

$$\begin{aligned} &h_{1,\nu} \left\{ u^{\nu+1} F_c \left\{ (y^2 - x^2)^{\nu-\frac{1}{2}}; y \right\}; u \right\} \\ &= \frac{y^{2\nu} \pi \Gamma(2\nu)}{(2\pi)^\nu \Gamma(\nu)} h_{1,\nu}(f(u); \pi). \end{aligned}$$

Proof.

$$\begin{aligned} &h_{1,\nu} \left\{ u^{\nu+1} F_c \left\{ (y^2 - x^2)^{\nu-\frac{1}{2}}; y \right\}; u \right\} \\ &= \int_0^1 u^{\nu+1} \left(\int_0^y (y^2 - x^2)^{\nu-\frac{1}{2}} \cos\left(\frac{u\pi x}{y}\right) dx \right) f(u) du \\ &= \int_0^1 u^{\nu+1} \left[\sqrt{\pi} 2^{\nu-1} \Gamma\left(\nu + \frac{1}{2}\right) \left(\frac{y^2}{u\pi}\right)^\nu J_\nu(u\pi) \right] f(u) du. \\ &= \frac{y^{2\nu} 2^{\nu-1} \Gamma\left(\nu + \frac{1}{2}\right)}{\pi^{\nu-\frac{1}{2}}} \int_0^1 u J_\nu(u\pi) f(u) du. \\ &= \frac{y^{2\nu} \pi \Gamma(2\nu)}{(2\pi)^\nu \Gamma(\nu)} h_{1,\nu}(f(u); \pi). \end{aligned}$$

Hence proved.

Lemma 2.11. The finite Hankel-Clifford transformation of finite Widder-like transforms is finite Hankel-Clifford transforms i.e.

$$H_{2,\nu} \left(P(f(x); u); y \right) = H_{2,\nu} \left(f(x); y \right). \tag{2.13}$$

Proof. Here the finite Hankel-Clifford transforms as stated by Dorta Diaz [7] is considered as in (1.5).

$$\begin{aligned}
 & H_{2,\nu}(P(f(x);u);y) \\
 &= \int_0^1 x^\nu (xy)^{-\nu/2} J_\nu(2\sqrt{yx}) \left(\int_0^x \frac{2x}{x^2-u^2} du \right) f(x) dx \\
 &= \int_0^1 x^\nu (xy)^{-\nu/2} J_\nu(2\sqrt{yx}) f(x) dx \\
 &= H_{2,\nu}(f(x);y).
 \end{aligned}$$

Hence proved that finite Hankel-Clifford transformation of finite Widder-like transforms is finite Hankel-Clifford transforms. Hence proved that $H_{2,\nu}(P(f(x);u);y) = H_{2,\nu}(f(x);y)$.

Lemma 2.12. The representation of finite Hankel transforms (1.5) is considered over finite Fourier sine transforms (1.12) as:

$$h_{1,\nu} \left\{ u^{\nu+1} F_s \left\{ (yx-x^2)^{\nu-\frac{1}{2}}; y; u \right\} \right\} = \frac{y^{3\nu+1} \pi \Gamma(2\nu)}{(2\pi)^\nu \Gamma(\nu)} h_{1,\nu+1}(f(u); \pi). \tag{2.14}$$

Proof. From (1.12), the finite Hankel transforms is considered over finite Fourier sine transforms

$$\begin{aligned}
 & h_{1,\nu} \left\{ u^{\nu+1} F_s \left\{ (yx-x^2)^{\nu-\frac{1}{2}}; y; u \right\} \right\} \\
 &= \int_0^1 u^{\nu+1} \left(\int_0^y x (y^2-x^2)^{\nu-\frac{1}{2}} \sin\left(\frac{u\pi x}{y}\right) dx \right) f(u) du \\
 &= \int_0^1 u^{\nu+1} \left[\sqrt{\pi} 2^{\nu-1} y^{\nu+1} \Gamma\left(\nu+\frac{1}{2}\right) \left(\frac{y^2}{u\pi}\right)^\nu J_{\nu+1}(u\pi) \right] f(u) du. \\
 &= \frac{y^{3\nu+1} 2^{\nu-1} \Gamma\left(\nu+\frac{1}{2}\right)}{\pi^{\nu-\frac{1}{2}}} \int_0^1 u J_{\nu+1}(u\pi) f(u) du. \\
 &= \frac{y^{3\nu+1} \pi \Gamma(2\nu)}{(2\pi)^\nu \Gamma(\nu)} h_{1,\nu+1}(f(u); \pi).
 \end{aligned}$$

Thus proved.

Lemma 2.13. The finite Hilbert transforms on incomplete beta functions with weight function to it is given by the finite Hilbert transforms as:

$$P(x^{-m-n+2} \beta(f(u);x);y) = P(f(x);y). \tag{2.15}$$

Proof. From (1.12), Beta functions can be used as:

$$\begin{aligned}
 & \int_0^y \frac{2x}{y^2-x^2} \left(\int_0^x u^{m-1} (x-u)^{n-1} du \right) f(x) dx. \\
 & \int_0^y \frac{2x}{y^2-x^2} x^{-m-n+2} (x^{m+n-2} \beta(m,n)) f(x) dx. \\
 &= P(f(x);y) \beta(m,n).
 \end{aligned}$$

Hence proved (2.15).

Lemma 2.14. Finite Fourier cosine series over Finite Laplace transforms and Finite Fourier sine series over Finite Laplace transforms can be represented by

$$F_c \{xL (f(x);u);y\} = \int_0^y \frac{y^2}{y^2 - \pi^2} \left[1 - e^{-x} \left\{ \cos\left(\frac{\pi x}{y}\right) - \frac{\pi}{y} \sin\left(\frac{\pi x}{y}\right) \right\} \right] f(x) dx. \tag{2.16}$$

$$F_s (xL \{f(x);u\};y) = \int_0^y \frac{y^2}{(y^2 - \pi^2)} \left[e^{-x} \left\{ \frac{\pi}{y} \cos\left(\frac{\pi x}{y}\right) \right\} - \frac{\pi}{y} \right] f(x) dx. \tag{2.17}$$

Proof. Finite Fourier cosine series is represented by (1.10) and Finite Laplace transforms is given by (1.4).

The LHS of

$$F_c \{xL (f(x);u);y\} = \int_0^y x \operatorname{Re} \left\{ \int_0^1 e^{-ux} e^{i\frac{u\pi x}{y}} du \right\} f(x) dx = \int_0^y \frac{y^2}{y^2 + \pi^2} \left[1 - e^{-x} \left\{ \cos\left(\frac{\pi x}{y}\right) + \frac{\pi}{y} \sin\left(\frac{\pi x}{y}\right) \right\} \right] f(x) dx.$$

Hence

$$F_c \{xL (f(x);u);y\} = \int_0^y \frac{y^2}{y^2 - \pi^2} \left[1 - e^{-x} \left\{ \cos\left(\frac{\pi x}{y}\right) - \frac{\pi}{y} \sin\left(\frac{\pi x}{y}\right) \right\} \right] f(x) dx.$$

Also in the similar way the proof can be done for:

$$F_s (xL \{f(x);u\};y) = \int_0^y x \left[\int_0^1 e^{-xy} \sin\left(\frac{u\pi x}{y}\right) du \right] f(x) dx = \int_0^y \frac{e^{-x} \left\{ \frac{\pi}{y} \cos\left(\frac{\pi x}{y}\right) + \sin\left(\frac{\pi x}{y}\right) \right\} - \frac{\pi}{y}}{\left(1 + \left(\frac{\pi}{y}\right)^2 \right)} f(x) dx = \int_0^y \frac{y^2}{(y^2 + \pi^2)} \left[e^{-x} \left\{ \frac{\pi}{y} \cos\left(\frac{\pi x}{y}\right) \right\} - \frac{\pi}{y} \right] f(x) dx = \int_0^y x \operatorname{Im} \left[\int_0^1 e^{-xy} e^{i\left(\frac{u\pi x}{y}\right)} du \right] f(x) dx.$$

Lemma 2.15. The finite Laplace transforms of finite Fourier transforms can be given by

$$L \{x F_c(f(x); u); y\} = -\int_0^1 \frac{y^2 e^{-yx}}{\pi^2 + y^2} \left(\cos \pi x + \frac{\pi}{y} \sin \pi x + e^{yx} \right) f(x) dx. \quad (2.17)$$

Proof. The LHS of (2.17) is represented as

$$L \{x F_c(f(x); u); y\} = \int_0^1 \operatorname{Re} \left[\frac{-e^{-yx} (\cos \pi x + i \sin \pi x) \left(1 - i \left(\frac{\pi}{y} \right) \right)}{\left(1 - \left(i \frac{\pi}{y} \right)^2 \right)} + \frac{\left(1 - i \left(\frac{\pi}{y} \right) \right)}{\left(1 - \left(i \frac{\pi}{y} \right)^2 \right)} \right] f(x) dx$$

$$= -\int_0^1 \frac{y^2 e^{-yx}}{\pi^2 + y^2} \left(\cos \pi x + \frac{\pi}{y} \sin \pi x + e^{yx} \right) f(x) dx.$$

Hence established.

III. CONCLUSION

In this study several new finite integral transforms were introduced and developed relations between the transforms and represented them as twin transforms. Such developed twin finite transforms may be used when repeated transforms of different natures occur in an experiment/ study. The finite transforms will be proved useful in applications in Mathematical Physics.

IV. ACKNOWLEDGEMENT

I thank SVKM's NMIMS University, MPSTME for sponsoring this work.

References

- [1] Dang Vu Giang, Finite Hilbert Transforms Logarithmic Potentials and Singular Integral Equations, arXiv:1003.3070 [math.GM], May, 2018.
- [2] Sneddon I. N. (1979), The use of integral transforms, Tata McGraw-Hill, New-Delhi.
- [3] S. P. Malgonde and V. R. Lakshmi Gorty, Orthogonal series expansions of generalized functions and the finite generalized Hankel-Clifford transformation of distributions, Rev. Acad. Canaria, (2008), Cienc., XX (Nums.1-2), 49-61.
- [4] M. Abramowitz and I. A. Stegun (edited by), Handbook of mathematical functions with Formulas, Graphs and Mathematical tables, United States department of commerce, National Bureau of standards, 1970.
- [5] W. W. Bell, Special Functions for Scientists and Engineers, Dover Publications, 2004.
- [6] A. R. Di Donato and M. P. Jarnagin, The Efficient Calculation of the Incomplete Beta-Function Ratio for Half-Integer Values of the Parameters a, b, February 22, 1967, pp. 652-662. Link: <https://pdfs.semanticscholar.org/efc8/5198cc9c9fb2da29ee40ff4802498b47499a.pdf>.
- [7] Dorta Diaz, J. A and Méndez Perez, J. M. R. (1992): "Dini's series expansions and the Finite Hankel-Clifford transformations", Jour. Inst. Math. and Comp. Sci., (Math. Ser.) Vol.5, No.1, 1-17.

Research and Study on Domain-specific Search Engine using Naïve Bayes classifier with HITS algorithm

¹Aliya Kazi,²Amruta Dabholkar,³Bhavana Choudhary, ⁴Shifa Shaikh,⁵Bharti Khemani,

^{1,2,3,4}Students of Rizvi college of Engineering,⁵Assistant Professor of Rizvi college of Engineering, Mumbai, Maharashtra, India

ABSTRACT: Nowadays there are many effective tools to minimize the difficulties faced while acquiring information from the web, such as domain-specific search engines. We can build such a domain-specific web search engine, by matching the user entered input with contents of our dataset to obtain desired results. Traditional search engines return irrelevant results in case of short and ambiguous query. One solution to this problem is to organize the search results via categorization, in particular, the classification. The hits algorithm used traditionally can be combined with a classification algorithm namely Naïve Bayes classifier. This algorithm has been selected on the basis of comparison between different classification algorithms, KNN, decision tree.

Keywords: : domain-specific, Naive Bayes, decision tree, KNN algorithm, search engine, classification algorithm, HITS algorithm.

I. Introduction

The traditional search engines, for any user entered query, provide a list of results that are ranked according to their significance to the query, using HITS algorithm. Conversely, the ranking is independent of the subject of the document. So the results of different topics are not grouped together within the result output from a search engine. In this study we will compare different approaches namely, Naïve Bayes, decision tree and KNN algorithm. While comparing these, we will see which method gives more effective results given a common data set, training file and testing file. An effective tool for obtaining information about a certain topic or a certain type is a domain-specific search engine. For example:

- www.yummly.com recommends recipes personalized to the individual's tastes, semantic recipe search, a digital recipe box and shopping list
- www.jinni.com lets the user a search engine and recommendation engine for movies, TV shows and short films.
- www.imdb.com allows the user to search for reviews of movies. Type a movie title, and it provides links to relevant reviews from newspapers, magazines, and individuals from all over the world.
- www.nomadicmatt.com allows the user to search about travel, with special facilities for searching by activity, category and location.

The three different classification algorithms that we will be using for this comparative study are:

- Naïve Bayes algorithm
- Decision Tree algorithm
- KNN (K Nearest Neighbors) algorithm

To optimize the search criteria for domain-specific search engines we make use of machine learning techniques. These techniques allow search engines to be searched quickly with minimal effort and are suited for re-use across many domains. We will implement a simple classification program in "r studio" for each of these algorithms, to identify which executes faster in comparison to the other two.

This study shows us how Naïve Bayes performs better when compared to other similar algorithms like decision tree and KNN. Normal search engines make use of HITS algorithm to rank pages based on the search query. We suggest the use of Naïve Bayes algorithm for this process. The algorithm will iteratively search for the entered keyword in the webpage content, and based on the frequency of occurrence of that keyword, we rank the pages.

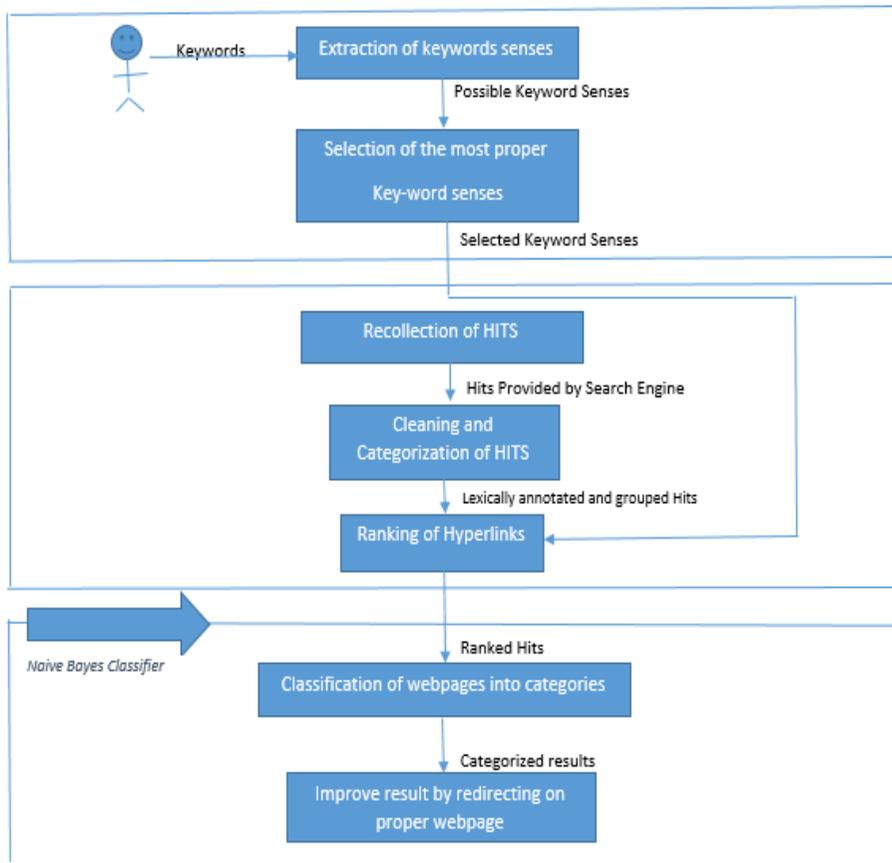


Fig.1

II. DEFINITION & BACKGROUND

▪ Naïve Bayes:

Naive Bayes algorithm is a probabilistic classifier that works on probability of features of an object belonging to a particular group.

It is called “naïve” based on its method of assuming the event of occurrence of every feature is independent of that of others and “Bayes” part, it refers to the statistician and philosopher, Thomas Bayes hence the name.

Assume that you are trying to identify an animal based on its color, shape and hair coat, then a brown colored, black eyed, long ear and four legged will most probably be a dog .Although these features are interdependent, all of these properties individually contribute to the probability that this animal is a dog and that is why it is known as “naïve.”

The Mathematics of the Naive Bayes Algorithm

As mentioned above the foundation of this algorithm is Bayes theorem. It gives us a method to calculate the conditional probability, i.e., the probability of an event based on previous knowledge available on the events.

More formally, Bayes’ Theorem is stated as the following equation:

$$P(A|B)=P(B|A)P(A)P(B)$$

Before moving onto the proof of the equation above, let us have a look at the terminologies used

- $P(A|B)$: Probability (conditional probability) of occurrence of event A given the event B is true
- $P(A)$ and $P(B)$: Probabilities of the occurrence of event A and B respectively
- $P(B|A)$: Probability of the occurrence of event B given the event A is true

The terminology in the Bayesian method of probability (more commonly used) is as follows:

- A is called the **proposition** and B is called the **evidence**.

- P(A) is called the **prior** probability of proposition and P(B) is called the **prior** probability of evidence.
- P(A|B) is called the **posterior**.
- P(B|A) is the **likelihood**

This sums the Bayes' theorem as

$$\text{Posterior} = (\text{Likelihood}) \cdot (\text{Proposition prior probability}) \cdot \text{Evidence prior probability}$$

Derivation of Bayes' Theorem

For a joint probability distribution of two events A and B, P(A∩B), the conditional probability[1],

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Similarly,

$$P(B|A) = \frac{P(B \cap A)}{P(A)}$$

Therefore,

Bayes' Theorem for Naive Bayes Algorithm

In a machine learning classification problem, there are multiple features and classes, say, C₁, C₂, ..., C_k. The main aim in the Naive Bayes algorithm is to calculate the conditional probability of an object with a feature vector x₁, x₂, ..., x_n belongs to a particular class C_i,

$$P(C_i | x_1, x_2, \dots, x_n) = P(x_1, x_2, \dots, x_n | C_i) \cdot P(C_i) \cdot P(x_1, x_2, \dots, x_n) \text{ for } 1 \leq i \leq k$$

Now, the numerator of the fraction on right-hand side of the equation above is P(x₁, x₂, ..., x_n | C_i). P(C_i) = P(x₁, x₂, ..., x_n, C_i)

$$\begin{aligned} P(x_1, x_2, \dots, x_n, C_i) &= P(x_1 | x_2, \dots, x_n, C_i) \cdot P(x_2, \dots, x_n, C_i) \\ &= P(x_1 | x_2, \dots, x_n, C_i) \cdot P(x_2 | x_3, \dots, x_n, C_i) \cdot P(x_3, \dots, x_n, C_i) \\ &= \dots \\ &= P(x_1 | x_2, \dots, x_n, C_i) \cdot P(x_2 | x_3, \dots, x_n, C_i) \dots P(x_{n-1} | x_n, C_i) \cdot P(x_n | C_i) \cdot P(C_i) \end{aligned}$$

The conditional probability term, P(x_j | x_{j+1}, ..., x_n, C_i) becomes P(x_j | C_i) because of the assumption that features are independent. From the calculation above and the independence assumption, the Bayes theorem boils down to the following easy expression:

$$P(C_i | x_1, x_2, \dots, x_n) = \left(\prod_{j=1}^n P(x_j | C_i) \right) \cdot P(C_i) \cdot P(x_1, x_2, \dots, x_n) \text{ for } 1 \leq i \leq k$$

The expression P(x₁, x₂, ..., x_n) is constant for all the classes, we can simply say that

$$P(C_i | x_1, x_2, \dots, x_n) \propto \left(\prod_{j=1}^n P(x_j | C_i) \right) \cdot P(C_i) \text{ for } 1 \leq i \leq k$$

▪ K nearest neighbour:

A k nearest neighbor classification rule will assign the class having the highest similarity score to a test point x_l, given a predefined similarity metric. This similarity metric can order the training point using their similarity to x_l. These scores are calculated by summing up the similarities of the k nearest neighbors in each class. The classification rule compares these scores and return the class having the highest, it is defined as with K the number of nearest neighbors otherwise[2]

$$\text{If } \phi(x_k) = y'$$

$$\phi(x') = \text{argmax}_{y' \in c} P(y') \prod_{j=1}^d P(x_j / y')$$

$$\delta(y', \phi(x_k)) = 1_i$$

else 0;

When KNN is used for classification, class with the highest frequency from the K-most similar instances can be used as basis for calculating the output. The class with the most votes is taken as the prediction with essence to the votes for each instance in the class. For a new data instance, the class probabilities can be calculated as the normalized frequency of samples belonging to each of those classes in the set of K most similar instances. For example, in a binary classification problem (class is 0 or 1):

$$p(\text{class}=0) = \frac{\text{count}(\text{class}=0)}{\text{count}(\text{class}=0) + \text{count}(\text{class}=1)}$$

It is more appropriate to choose a K value with an odd number to avoid a tie, if the value being used is K and we have an even number of classes (e.g. 2). Similarly we can use an even number for K when you have an odd number of classes. By looking at the class of the next most similar instance in the training dataset and expanding K by 1, we can break ties easily.

▪ Decision tree:

Another important algorithm belonging to the family of supervised learning algorithms is the decision tree algorithm. Unlike all the other supervised learning algorithms, even the decision tree algorithm can be used

for solving classification and regression problems. Decision Tree is used with the general motive of creating a training model which can be used to predict class or value of target variables by learning decision rules inferred from prior provided training data. Compared to other classification algorithms, the understanding level of Decision Trees algorithm is easier. The decision tree algorithm tries solving problems, by using tree-like representation. Here each leaf node corresponds to a class label and each internal node of the tree corresponds to an attribute[3,7].

Formulae:

$$E(S) = \sum_{i=1}^c -p_i \log_2 p_i$$

$$E(T, X) = \sum_{c \in X} P(c)E(c)$$

$$Gain(T, X) = Entropy(T) - Entropy(T, X)$$

III. traditional working of search engine (using hits algorithm):

Jon Kleinberg developed in 1998, a link analysis algorithm named Hypertext Induced Topic Search (HITS) or hubs and authorities, to rate Web pages. It is an iterative algorithm that quantifies every page of its value as an authority and as a hub. The foundation of the algorithm is that a web page serves two purposes: to provide information on a topic, and to provide links to other pages giving information on a topic. Based on above defined purposes, pages are categorized as:

Authority: page that provides vital and trustworthy information about the topic searched, and is pointed by many hubs.

- Hub: pages that contain links to “authorities“ that is pointing to many pages.

Ranking of the web page is decided by analysing their textual contents against the user entered search query. After collection of the web pages, the HITS algorithm neglects the textual contents of web pages and concentrates on the structure of the webpage only.

▪ Working:

When the user enters a query, HITS algorithm first finds a large amount of relevant pages that are then returned by the search engine. Then it produces ranking on the following pages, by assigning authority score and hub score to each page. The page having the highest authority score will be ranked highest

▪ Pseudocode:

1. G: = set of pages
2. for each page p in G do
3. p.auth = 1 // p.auth is the authority score of the page p
4. p.hub = 1 // p.hub is the hub score of the page p
5. function HubsAndAuthority(G)
6. for step from 1 to k do // run the algorithm for k steps
7. for each page p in G do // update all authority values first
8. for each page q in p.incomingNeighbors do // p.incomingNeighbors is the set of pages that link to p p.auth += q.hub
9. for each page p in G do // then update all hub values
for each page r in p.outgoingNeighbors do //p.outgoingNeighbors is the set of pages that p links to
p.hub += r.auth

IV. improved working of search engine using combined approach:

▪ **Problem Definition:**

Traditional search engines using hits algorithm return search query results based on the content and hyperlinks of a certain webpage. However an individual cannot sort through the thousands of results to search their relevant content. To ease the user’s efforts of finding only necessary information from

irrelevant web pages, we can make use of classification algorithms. Classification algorithms basically categorize data by identifying the class under which the data will fall.

We can improve the results returned by search engine to obtain only those web pages having data which is requested by the user by combining hits algorithm with a classification algorithm. The classification algorithm will be implemented on the result returned by hits algorithm to classify them into proper categories. Now the user can easily select what they want from these categorized results.

V. COMPARITIVE STUDY

b) Implementation of classification algorithm in R tools:

In this study we will compare different approaches namely, Naïve Bayes, decision tree and KNN algorithm. Of the three algorithms, Naïve Bayes was selected due to its easy and fast execution,observed during its implementation in R tools.

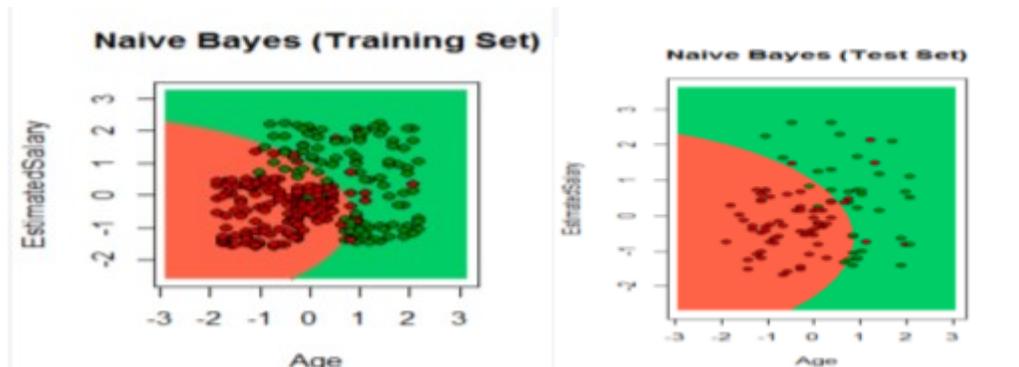


Fig2.Example of Naive Bayes in R tools

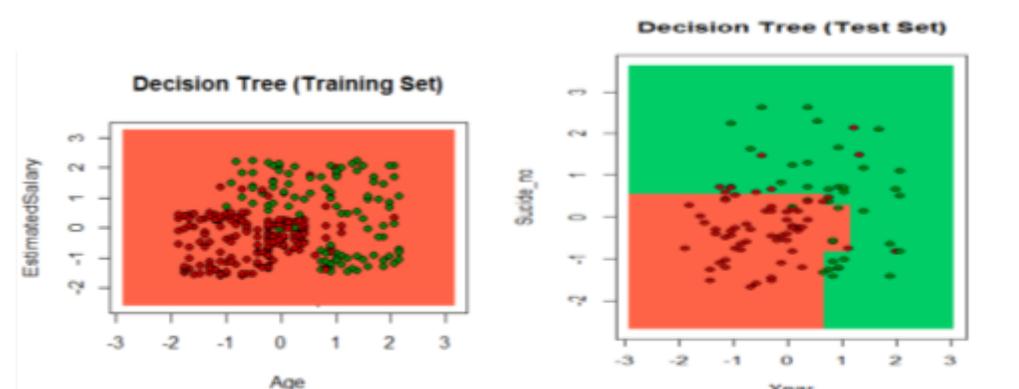


Fig3.Example of Decision tree in R tools.

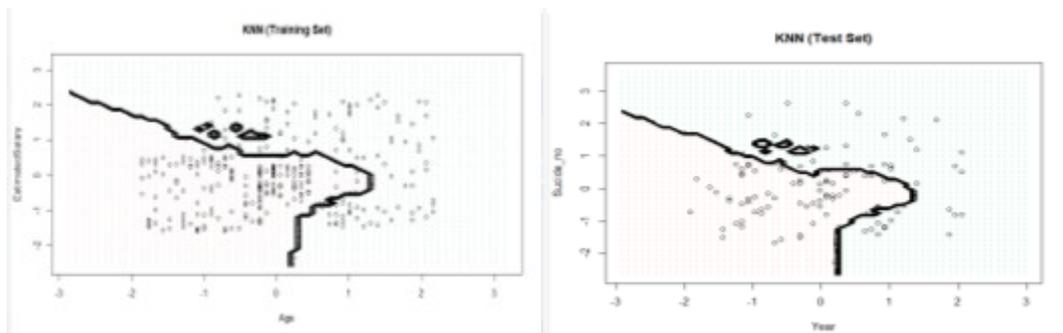


Fig4.Example of KNN in R tools

c) Complexities:

The time complexity of:

- i) Naïve Bayes is $O(Np)$, where N is the number of training examples and p is the number of features
- ii) Decision trees is $O(Nkd)$ which actually lies somewhere in between $O(Nk\log N)$ and $O(N^2k)$.
- iii) KNN is $O(nd + kn)$, where n is the cardinality of the training set and d the dimension of each sample.

d) Classification on the basis of Factor

Property	Decision Tree	KNN	Naïve Bayes Classifier
Accuracy	Good	Good	Excellent
Learning Speed	Very Good	Excellent	Good
Tolerance to missing values	Very Good	Average	Very Good
Tolerance to irrelevant attribute	Very Good	Good	Very Good
Tolerance to noise	Good	Average	Average
Attempts for incremental learning	Good	Excellent	Excellent
Explanation ability	Excellent	Good	Excellent
Support Multi Classification	Very Good	Good	Very Good
Speed of classification	Good	Average	Excellent

Table.1

e) Classification on the basis of computational time

Dataset	Size of Dataset	Time	KNN	Naïve Bayes	Decision Tree
Weather Nominal	Small (14 instances)	To Build Model	0 sec	0 sec	0.02 sec
		To Test Model	0.02 sec	0 sec	0 sec
Segment Challenge	Medium (1500 instances)	To Build Model	0 sec	0.08 sec	0.16 sec
		To Test Model	0.42 sec	0.31 sec	0.06 sec
Super Market	Large (4627 instances)	To Build Model	0.02 sec	0.06 sec	0.06 sec
		To Test Model	45.55 sec	0.28 sec	0.03 sec

Table.2

f) Implementation on different size of datasets:

Dataset	Size Of Dataset	KNN	Naïve Bayes	Decision Tree
Weather Nominal	Small (14 Instances)	100%	92.85%	99%
Segment Challenge	Medium (1500 Instances)	81.66%	100%	98%
Super Market	Large(4627 Instances)	89.42%	73.713%	63.7%

Table.3

g) Selection of classification algorithm

Domain specific search engines apply algorithms to automatically recognize the topic and content of a specific page. A web site organizes web content in a hierarchical structure, which is analogous to the traditional system, where each document belongs to one or more topics and subtopics. If the user wants to search for certain information related to a single topic, rather than searching through all available information, he/she only requires to seek the documents within the category they are interested in. However, they generally require humans to compile the list, and thus can only cover a small portion of the whole Web. On the web, users will typically express their information need with only three terms or less, and thus due to a gap in knowledge, the searcher is often not able to generate the correct query. Therefore, nowadays numerous search engines allow the users to specify their domain of interest or to describe their interests in a profile to assist the search. Such approaches are domain limited and non-universal. Also such information will not be willingly provided by every user.

The basic idea behind this is to classify these pages depending upon the text entered by user in the search engine. For this we will take a dataset of 1000's of keywords which relates and compares the texts with contents of dataset and then relatively redirect the user on a relevant web page. This method of redirecting

users can be made simple with the help of a few classification algorithms such as KNN, Decision tree and Naïve Bayes Algorithm.

A website can consist of any number of web pages with a wide variety of keywords. In a dataset of search engine, if we use any algorithm in light of SEO (Search Engine Optimization) then it should produce a better and more relevant output. Under the concept of Machine learning, the most commonly used algorithms are KNN, Decision tree and Naïve Bayes classifier. All three of them can produce an output for the same problem, but it is not necessary that all of them are equally efficient, as their complexities depend upon the method they use, sequence of steps and requirements. If we use KNN, we are required to provide the number of nodes in the network, it means number of webpages are to be mentioned first and only then it will classify them accordingly. Problem will arise if the user fails to update number of nodes, then the update numbered webpage will not be under consideration, and if the search engine receives input query matching content of that newly updated page then it may return a problem.

On the other hand if we use a decision tree algorithm for them same, it basically calculates the entropy of each attribute and accordingly calculates the gain. The attribute having the highest gain becomes the root node for our decision tree. This indicates that it has a greater probability to get searched by users and hence accordingly splits the data and constructs the tree further by similar iterative calculations. Entropy is an indicator of degree of randomness of your provided data. By using entropy, decision trees organize more than they classify the data. Also each tree has branches, if the user has entered by traversing on a particular branch, then if he/she wants to visit a node (webpage) in other branch then the user would require to backtrack the path. This process of traversing and backtracking becomes more complex and tedious if dataset is very large.

In Naïve Bayes classifier, the algorithm always checks for the probability of a given category in a hash table and accordingly lists the links or directly redirects onto the relevant pages. It generally provides better classification of webpages than the other two classification algorithms that we have discussed earlier.

As per the observation from the figures [Fig.2, Fig.3, Fig.4] produced by the three different algorithms and their respective complexities we have come to the conclusion that Naïve Bayes classifier algorithm has higher probability of accurate results for the classification. As it is less sensitive, fast, works great in manual practices, can be used for both multiclass and binary classification problems, Works with continuous and discrete data. Whereas Decision Trees sometimes suffer from over fitting, work less efficiently with Non-numerical data, their prediction accuracy is low for dataset, many class labels used in it leads to complications in calculations and KNN (K nearest neighbor) has skewed class distribution, means if during training phase particular class is very frequent then it will tend to prevail the major voting of new example, it does not work efficiently with high dimensional data, needs to define a value of K (Webpages).

VI. IMPLEMENTATION OF THE COMBINED APPROACH:

Up till now HITS algorithm uniquely provides the ranks to the webpages by comparing the contents of Web pages and gives the hyperlinks to which browser can redirect. But we are to handle a specific type of information request optimally, the search engine needs to be fine-tuned hence we are working with fine-tuning process as Domain-Specific Search Engines for which HITS algorithm does not provide accurate results hence combined approach is been used with best classification algorithm, Naïve Bayes Classification Algorithm.

▪ Working:

HITS:

- 1) Count the number and quality of links to a page to determine a rough estimate of how important the website is.
- 2) The underlying assumption is that more important websites are likely to receive more links from other websites.
- 3) Not all hyperlinks related to the conferral of authority. Find the pattern authoritative pages Using Naïve Bayes Classifier.

Naïve Bayes Classifier:

The Naïve Bayes classifier will be implemented on the result returned by hits algorithm to classify them into proper categories. Now the user can easily select what they want from these categorized results.

- 1) Categorize data by identifying the class under which the data will fall.
- 2) Improve the results returned by search engine to obtain only those web pages having data which is requested by the user by combining hits algorithm with a Naïve Bayes classifier.

VII. Future work:

Our future work will mainly focus on using unlabeled training data, since generally greater quantities of unlabeled data are available, and parameters may be more reliably estimated from larger amounts of data for which manual labelling is costly and error-prone.

Also we would like to propose the idea of creating a complete web search engine based on a classification algorithm using Naïve Bayes methodology rather than just a domain-specific one. This is due to the simplicity of the Naïve Bayes classifier and the accuracy it provides.

Another suggestion would be to optimize the domain specific search engine for image classification alongside text classification.

VIII. Conclusion

Recent years continue to see an exponential growth in the extent of information or collective data available on the Internet. Following this never settling trend, we argue that not only will the public need efficient tools or services to help them sort through this information, but also intelligent techniques to help build and maintain these services. Proposed study shows the combine approach works on selected results which have higher frequency of occurrence, so the modified HITS algorithm gives higher performance in terms of timing.

Through this paper we have shown that creation and maintenance of domain-specific search engines can be significantly aided by machine learning techniques. Much future work in each machine learning area has already been discussed. However, it cannot be denied that there exist many other areas where domain-specific search engines can be automated using machine learning techniques. For example, text classification can decide which documents on the Web are relevant to the domain.

IX. References

1. Wai Yuen Tang, Lam For Kwok, " Domain Specific Search Engine for Students", City University of Hong Kong, Hong Kong, 2004
2. Debajyoti Mukhopadhyay , Aritra Banik , Sreemoyee Mukherjee, Jhili Bhattacharya1 , "A Domain Specific Ontology Based Semantic Web Search Engine", West Bengal University of Technology, India,
3. The Curriculum Development Council (1991). An English-Chinese Glossary of Terms Commonly Used in the Teaching of Mathematics in Secondary Schools. HK: The Education Department.
4. McCallum, A., Nigam, K., Rennie, J. and Seymore, K. (1999). Building domain-specific search engines with machine learning techniques, Proc. AAAI-99 Spring Symposium on Intelligent Agents in Cyberspace. USA: The AAAI Press.
5. Yogendra Singh Rajput, Priyasaxena "A Combined Approach Used to find Domain Using Naïve Bayes Classifier", MP, India, 2016
6. Peiqian Liu, Jingjing Feng. An improved Naive Bayes Text Classification Algorithm. Microcomputer Information, No 27, 2010.
7. Qian Zhou, Mingsheng Zhao, Min Hu. Study on Feature Selection in Chinese Text Categorization. JOURNAL OF CHINESE INFORMATION PROCESSING, VOL 18, No 3, 2004.
8. Lei Dai, Weidong Ma, Lingnan Wang, Jianguo Ma. The Design and Implementation of Weight-based Naïve Bayes Classifier
9. Tom M. Mitchell. Machine Learning. China Machine Press, 2008
10. The Curriculum Development Council (1999). Syllabuses for Secondary School Mathematics Secondary 1-5. HK: The Education Department.

Solution of Groundwater Infiltration Phenomenon in horizontal direction By DTM and RDTM Method

Krishna Shah¹ & Amit K. Parikh²

Research Scholar¹, Principal²

Mehsana Urban Institutes of sciences,Ganpat University,Mehsana - 384012

ABSTRACT: In this paper, we apply Differential transform method and reduced-differential transform method to solve Groundwater Infiltration Phenomenon in horizontal direction. The definition and basic theorems in additive to new generalized transforms of DTM and RDTM are explored. Here we have done comparison of both the methods. The RDTM produces an approximate solution without any discretization or perturbation and can reduce the size of computations.

Keywords: : Differential transform method; Reduced differential transform method; Groundwater Infiltration Phenomenon in horizontal direction.

1. Groundwater Infiltration Phenomenon in horizontal direction:

1.1 STATEMENT OF THE PROBLEM:

The reservoir field with water of height $OA = h_{\max}$ = maximum height with impermeable bottom and surrounding of this reservoir is unsaturated homogeneous soil. Infiltration phenomenon is well demonstrated in the following figure 1 that shows a cross section of the reservoir surrounded by unsaturated porous medium.

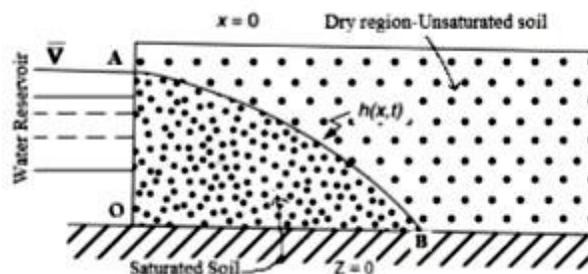


Figure 1: A scheme of groundwater Infiltration.

The height of the free surface is zero, when $OB = x = 1$, the dotted arc below the curve is saturated by infiltrated groundwater and above the curve is the dry region of unsaturated soil. The bottom is assumed impermeable, so water cannot flow in a downward direction. Infiltration is the process by which the groundwater of the reservoir has entered into the unsaturated soil through vertical permeable wall. The groundwater infiltrated in an unsaturated soil; its velocity decreases as soil becomes saturated. The infiltrated groundwater will enter in unsaturated soil then the infiltrated groundwater will develop a curve between saturated soil and unsaturated soil, which is called a water table or water mound. To understand this one-dimensional infiltration phenomenon and for the sake of its mathematical formulation, we have taken some assumptions. The governing nonlinear partial differential equation for the height of infiltrated water is known as Boussinesq's equation [3]. The purpose of the study of infiltration is to determine the effective height of the free surface as a measure of initial storage capacity of a porous stratum.

1.2 MATHEMATICAL FORMULATION

Consider the maximum height h_{\max} of the groundwater reservoir and adjacent formation of the reservoir is unsaturated porous media shown in the figure 11 by dotted region. The groundwater is infiltrated through the vertical adjacent side with the following simplifying assumptions:

1. The stratum has height h_{\max} and lies on top of a horizontal impervious bed, which we label as $z=0$;
2. We ignore the transversal variable y ;
3. The water mass which infiltrates the soil occupies a region described as

$$\alpha. = \{(x, z) \in R : z \leq h(x, t)\}$$

We are assuming that there is no region of partial saturation. This is an evolution model. Clearly, $0 \leq h(x, t) \leq h_{\max}$, h_{\max} maximum height of free surface and the free boundary function 'h' is also an unknown function for this problem. In this situation, we arrive at a system of three equations with unknowns the two velocity components 'u', 'w' and 'p' in a variable domain: one equation of mass conservation for an incompressible fluid and two equations for the conservation of momentum of the Navier-stokes type.

The resulting system is too complicated and can be simplified for the practical computation after introducing a suitable assumption, the hypothesis of almost horizontal flow, i.e. assuming that an almost horizontal flow where speed $u(u, 0)$, so that 'h' has small gradients. Hence vertical component of the momentum equation can be written as:

$$p \left(\frac{du_z}{dt} + u. \Delta u_z \right) = - \frac{\partial p}{\partial z} - \rho g \tag{1}$$

The inertial term has been neglected (the left-hand side). Integration equation (1) with respect to z gives for the first approximation, it gives

$$p + \rho g z = \text{constant} \tag{2}$$

We now calculate the constant on the free surface $z = h(x, t)$. If we impose continuity of the pressure across

the interface, we have $p=0$ (assuming constant atmospheric pressure in the air that fills the pores of the dry region $z > h(x, t)$). We then get

$$p = \rho g (h - z) \tag{3}$$

Hence the pressure can be established using of the hydrostatic approximation. We go now to the mass conservation law which will give us the equation. We proceed as follows: let us take a section

$S = (x, x + a) \times (0, C)$. Then

$$\phi \frac{\partial}{\partial t} \int_x^{x+a} \int_0^h dy dx = - \int_{\partial S} u. n dl \tag{4}$$

where ' ϕ ' is the porosity of the medium, i.e., the fraction of volume available for the flow circulation, and 'u' is the seepage velocity, which obeys Darcy's law in the form that includes gravity effects[4],

$$u = - \frac{k}{\mu} \Delta (p + \rho g z) \tag{5}$$

On the right-hand lateral surface, we have $u. n \approx (u, 0). (1, 0) = u$ i.e. $\left(\frac{k}{\mu}\right) p_x$.

while, on the left-hand side, we have the velocity as '-u'. Using the formula for 'p' and differentiating with respect to 'x', we get

$$\phi \frac{\partial h}{\partial t} = \frac{p g K}{\mu} \frac{\partial}{\partial x} \int_0^h \frac{\partial}{\partial x} h dz \tag{6}$$

Thus we obtained Boussinesq's Equation as [3],

$$\frac{\partial h}{\partial t} = \frac{p g K}{2 \mu \phi} \frac{\partial^2}{\partial x^2} (h^2) \tag{7}$$

With constant $\beta = \frac{p g K}{2 \mu \phi}$, this is fundamental equation in ground water infiltration which respects porous medium equation. On choosing new dimensionless variable $T = \frac{p g K}{\mu \phi} t$ and $X = \frac{x}{L}$ in Equation (7), it gives,

$$\frac{\partial h}{\partial T} = \left\{ h \frac{\partial^2 h}{\partial X^2} + \left(\frac{\partial h}{\partial X} \right)^2 \right\} \tag{8}$$

The equation (8) gives the height of the water mound with initial and boundary condition as,

$$h(0, T) = h_{\max} \quad X=0 \text{ and } T > 0 \tag{9}$$

which satisfied the initial condition

$$h(X, 0) = h_0 \quad X > 0 \text{ and } T = 0 \tag{10}$$

Equation (8) is a fundamental equation in groundwater infiltration.

2. Differential Transform Method

2.1 Introduction:

The classic differential transform method, introduced by Zhou[1], is based on the definition of the differential transform, which is a Taylor series. Thus, it requires a number of calculations to obtain the basic properties of the differential transforms. Some of DTM applications are mentioned in [2-7].

2.1.1 Two dimensional Differential Transform Method:

Consider a function of two variables $w(x, t)$ and suppose that it can be represented as product of two single-variable functions, i.e., $w(x, t) = f(x)g(t)$: Based on the properties of differential transform, function $w(x, t)$ can be represented as

$$w(x, t) = \sum_{i=0}^{\infty} F(i)x^i \sum_{j=0}^{\infty} G(j)t^j = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} W(i, j)x^i t^j, \tag{11}$$

Where $W(i, j) = F(i)G(j)$ is called the spectrum of $w(x, t)$.

2.1.2 Definition:

If function $w(x, t)$ is analytic and differentiated continuously differentiable with respect to time t and space x in the domain of interest, then let

$$W(m, n) = \frac{1}{m!n!} \left[\frac{\partial^{m+n}}{\partial x^m \partial t^n} w(x, t) \right]_{x=0, t=0} \tag{12}$$

where the t-dimensional spectrum function $W(m, n)$ is the transformed function which is called T-function in brief.

The differential inverse transform of $W(m, n)$ is defined as follows:

$$w(x, t) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} W(m, n)x^m t^n \tag{13}$$

Combining (12) and (13) gives the solution as

$$w(x, t) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \frac{1}{m!n!} \left[\frac{\partial^{m+n}}{\partial x^m \partial t^n} w(x, t) \right]_{x=0, t=0} x^m t^n \tag{14}$$

Function Form	Transformed Form
$w(x, t) = u(x, t) \pm v(x, t)$	$W(m, n) = U(m, n) \pm V(m, n)$
$w(x, t) = cu(x, t)$	$W(m, n) = cU(m, n)$
$w(x, t) = \frac{\partial}{\partial x} u(x, t)$	$W(m, n) = (m + 1)U(m + 1, n)$
$w(x, t) = \frac{\partial}{\partial t} u(x, t)$	$W(m, n) = (n + 1)U(m, n + 1)$
$w(x, t) = \frac{\partial^{r+s}}{\partial x^r \partial t^s} u(x, t)$	$W(m, n) = \frac{(m + r)! (n + s)!}{m!n!} U(m + r, n + s)$
$w(x, t) = u(x, t)v(x, t)$	$W(m, n) = \sum_{r=0}^m \sum_{s=0}^n U(r, n - s)V(m - r, s)$
$w(x, t) = x^\alpha t^\beta$	$W(m, n) = \delta(m - \alpha, n - \beta) = \begin{cases} 1 & m = \alpha, n = \beta \\ 0 & \text{otherwise} \end{cases}$

Table 1

2.2 Application of Differential transform Method:

According to the DTM and table 1, we can construct the following transformation of Equation (8) $\frac{\partial h}{\partial t} =$

$$\left\{ h \frac{\partial^2 h}{\partial x^2} + \left(\frac{\partial h}{\partial x} \right)^2 \right\}$$

With Initial condition $h(X, 0) = e^{-X}$ as:

$$\begin{aligned} (n + 1)H(m, n + 1) &= \sum_{r=0}^m \sum_{s=0}^n (r + 1)(m - r + 1)H(r + 1, n - s)H(m - r + 1, s) \\ &+ \sum_{r=0}^m \sum_{s=0}^n (m - r + 1)(m - r + 2)H(r, n - s)H(m - r + 2, s) \end{aligned} \tag{15}$$

By definition of DTM

$$H(m, n) = \frac{1}{m!n!} \left[\frac{\partial^{m+n}}{\partial X^m \partial T^n} h(X, T) \right]_{X=0, T=0} \tag{16}$$

To Apply initial condition in equation (16) putting $n=T=0$ we get,

$$H(m, 0) = \frac{1}{m!0!} \left[\frac{\partial^m}{\partial X^m} h(X, 0) \right]_{X=0} \tag{17}$$

Applying initial condition in equation (17) and putting different values of m , i.e m=0,1,2,3,4,5,... we get

$$H(0, 0) = \frac{1}{0!} \left[\frac{\partial^0}{\partial X^0} h(X, 0) \right]_{X=0}$$

$$H(0, 0) = \frac{1}{1} [e^{-X}]_{X=0}$$

So we get H(0,0) = 1

$$H(1, 0) = \frac{1}{1!} \left[\frac{\partial^1}{\partial X^1} h(X, 0) \right]_{X=0}$$

$$H(1, 0) = \frac{1}{1} \left[\frac{\partial^1}{\partial X^1} e^{-X} \right]_{X=0}$$

$$H(1, 0) = [-e^{-X}]_{X=0}$$

So we get H(1,0) = -1

$$H(2, 0) = \frac{1}{2!} \left[\frac{\partial^2}{\partial X^2} h(X, 0) \right]_{X=0}$$

$$H(2, 0) = \frac{1}{2} \left[\frac{\partial^2}{\partial X^2} e^{-X} \right]_{X=0}$$

$$H(2, 0) = \frac{1}{2} [e^{-X}]_{X=0}$$

So we get H(2,0) = $\frac{1}{2}$

In similar manner, we will get H(3,0) = $-\frac{1}{6}$

$$H(4,0) = \frac{1}{24}$$

$$H(5,0) = -\frac{1}{120}$$

Therefore, the values that we found are

$$H(0,0) = 1, H(1,0) = -1, H(2,0) = \frac{1}{2}, H(3,0) = -\frac{1}{6}, H(4,0) = \frac{1}{24}, H(5,0) = -\frac{1}{120}$$

(18)

Now putting n=0 and different values of m=0,1,2,3,4,... in Equation (15)

$$\begin{aligned} (n+1)H(m, n+1) &= \sum_{r=0}^m \sum_{s=0}^n (r+1)(m-r+1)H(r+1, n-s)H(m-r+1, s) \\ &+ \sum_{r=0}^m \sum_{s=0}^n (m-r+1)(m-r+2)H(r, n-s)H(m-r+2, s) \end{aligned}$$

n=0, m=0, we get

$$\begin{aligned} (0+1)H(0,0+1) &= \sum_{r=0}^0 \sum_{s=0}^0 (r+1)(0-r+1)H(r+1, 0-s)H(0-r+1, s) \\ &+ \sum_{r=0}^0 \sum_{s=0}^0 (0-r+1)(0-r+2)H(r, 0-s)H(0-r+2, s) \end{aligned}$$

So putting r=s=0

$$H(0,1) = (1)(1)H(1,0)H(1,0) + (1)(2)H(0,0)H(2,0)$$

$$= (1)(1)(-1)(-1) + (1)(2)(1) \left(\frac{1}{2}\right)$$

$$= 1+1=2$$

n=0, m=1, we get

$$\begin{aligned} (0+1)H(1,0+1) &= \sum_{r=0}^1 \sum_{s=0}^0 (r+1)(1-r+1)H(r+1, 0-s)H(1-r+1, s) \\ &+ \sum_{r=0}^1 \sum_{s=0}^0 (1-r+1)(1-r+2)H(r, 0-s)H(1-r+2, s) \end{aligned}$$

Putting s=0

$$(1)H(1,1) = \sum_{r=0}^1 (r+1)(2-r)H(r+1,0)H(2-r,0) + \sum_{r=0}^1 (2-r)(3-r)H(r,0)H(3-r,0)$$

$$H(1,1) = (0+1)(2-0)H(1,0)H(2,0) + (1+1)(2-1)H(2,0)H(1,0) + (2-0)(3-0)H(0,0)H(3,0) + (2-1)(3-1)H(0,0)H(2,0)$$

$$H(1,1) = (1)(2)(-1)(0.5) + (2)(1)(0.5)(-1) + (2)(3)(1)\left(-\frac{1}{6}\right) + (1)(2)(-1)(0.5)$$

$$H(1,1) = -1 - 1 - 1 - 1$$

Therefore $H(1,1) = -4$

$n=0, m=2$, we get

$$(0+1)H(2,0+1)$$

$$= \sum_{r=0}^2 \sum_{s=0}^0 (r+1)(2-r+1)H(r+1,0-s)H(2-r+1,s) + \sum_{r=0}^2 \sum_{s=0}^0 (2-r+1)(2-r+2)H(r,0-s)H(2-r+2,s)$$

$$H(2,1) = \sum_{r=0}^2 (r+1)(3-r)H(r+1,0)H(3-r,0) + \sum_{r=0}^2 (3-r)(4-r)H(r,0)H(4-r,0)$$

$$H(2,1) = (1)(3)H(1,0)H(3,0) + (2)(2)H(2,0)H(2,0) + (3)(1)H(3,0)H(1,0) + (3)(4)H(0,0)H(4,0) + (2)(3)H(1,0)H(3,0) + (1)(2)H(2,0)H(2,0)$$

$$H(2,1) = (1)(3)(-1)\left(-\frac{1}{6}\right) + (2)(2)(0.5)(0.5) + (3)(1)\left(-\frac{1}{6}\right)(-1) + (3)(4)(1)\left(\frac{1}{24}\right) + (2)(3)(-1)\left(-\frac{1}{6}\right) + (1)(2)(0.5)(0.5)$$

$$H(2,1) = \frac{1}{2} + 1 + \frac{1}{2} + \frac{1}{2} + 1 + \frac{1}{2}$$

$$H(2,1) = 4$$

Similarly $H(3,1) = -\frac{8}{3}$

Now putting $n=1$ and different values of $m=0,1,2,3,4,\dots$ in Equation (15)

$$(n+1)H(m, n+1)$$

$$= \sum_{r=0}^m \sum_{s=0}^n (r+1)(m-r+1)H(r+1, n-s)H(m-r+1, s) + \sum_{r=0}^m \sum_{s=0}^n (m-r+1)(m-r+2)H(r, n-s)H(m-r+2, s)$$

$n=1, m=0$

$$(1+1)H(0,1+1)$$

$$= \sum_{r=0}^0 \sum_{s=0}^1 (r+1)(0-r+1)H(r+1,1-s)H(0-r+1,s) + \sum_{r=0}^0 \sum_{s=0}^1 (0-r+1)(0-r+2)H(r,1-s)H(0-r+2,s)$$

$$(1+1)H(0,1+1) = \sum_{s=0}^1 (1)(1)H(1,1-s)H(1,s) + \sum_{s=0}^1 (1)(2)H(0,1-s)H(2,s)$$

$$2H(0,2) = (1)(1)H(1,1)H(1,0) + (1)(1)H(1,0)H(1,1) + (1)(2)H(0,1)H(2,0) + (1)(2)H(0,0)H(2,1)$$

$$2H(0,2) = (1)(1)(-4)(-1) + (1)(1)(-1)(-4) + (1)(2)(2)\left(\frac{1}{2}\right) + (1)(2)(1)(4)$$

$$2H(0,2) = 4+4+2+8$$

$$2H(0,2) = 18$$

Therefore $H(0,2) = 9$.

Putting different values of m and n in equation (15) we got different $H(x,t)$ which are as follows:

$H(0,0) = 1$	$H(1,0) = -1$	$H(2,0) = \frac{1}{2}$	$H(3,0) = -\frac{1}{6}$
$H(0,1) = 2$	$H(1,1) = -4$	$H(2,1) = 4$	$H(3,1) = \frac{8}{3}$
$H(0,2) = 9$	$H(1,2) = -27$	$H(2,2) = \frac{81}{4}$	$H(3,2) = -\frac{81}{2}$
$H(0,3) = \frac{271}{6}$	$H(1,3) = -\frac{542}{3}$	$H(2,3) = \frac{1084}{3}$	$H(3,3) = \frac{8672}{18}$

And so on...

Now by equation(13) we get,

$$h(X, T) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} H(m, n) X^m T^n$$

so, $h(X, T) = H(0,0)X^0T^0 + H(1,0)X^1T^0 + H(2,0)X^2T^0 + H(3,0)X^3T^0 + H(0,1)X^0T^1 + H(1,1)X^1T^1 + H(2,1)X^2T^1 + H(3,1)X^3T^1 + H(0,2)X^0T^2 + H(1,2)X^1T^2 + H(2,2)X^2T^2 + H(3,2)X^3T^2 + H(0,3)X^0T^3 + H(1,3)X^1T^3 + H(2,3)X^2T^3 + H(3,3)X^3T^3 + \dots$

Putting the value from the above table, we get,

$$h(X, T) = 1 - X + \frac{X^2}{2!} - \frac{X^3}{3!} + 2T - 4XT + 4X^2T - \frac{8}{3}X^3T + 9T^2 - 27XT^2 + \frac{81}{4}X^2T^2 - \frac{81}{2}X^3T^2 + \frac{271}{6}T^3 - \frac{542}{3}XT^3 + \frac{1084}{3}X^2T^3 - \frac{8672}{18}X^3T^3 + \dots$$

$$h(X, T) = 1 - X + \frac{X^2}{2!} - \frac{X^3}{3!} + 2T \left[1 - 2X + 2X^2 + \frac{4X^3}{3} \right] + 9T^2 \left[1 - 3X + \frac{9}{4}X^2 - \frac{9}{2}X^3 \right] + \frac{271}{6}T^3 \left[1 - 4X + 8X^2 - \frac{32}{3}X^3 \right]$$

But

$$e^{-X} = 1 - X + \frac{X^2}{2!} - \frac{X^3}{3!}$$

$$e^{-2X} = 1 - 2X + 2X^2 + \frac{4X^3}{3}$$

$$e^{-3X} = 1 - 3X + \frac{9}{4}X^2 - \frac{9}{2}X^3$$

$$e^{-4X} = 1 - 4X + 8X^2 - \frac{32}{3}X^3$$

Therefore $h(X, T) = e^{-X} + 2e^{-X}T + 9e^{-X}T^2 + 45.167e^{-X}T^3 + \dots$ (19)

3.Reduced differential transform method

3.1 Introduction:

Consider a function of two variables $u(x, t)$ and suppose that it can be represented as product of two single-variable functions, i.e., $u(x, t) = f(x)g(t)$: Based on the properties of differential transform, function $u(x, t)$ can be represented as

$$u(x, t) = \sum_{k=0}^{\infty} F_k x^k \sum_{j=0}^{\infty} G_j t^j = \sum_{k=0}^{\infty} U_k(x) t^k, \tag{20}$$

Where $U_k(x)$ is called t-dimensional spectrum function of $u(x, t)$.

3.1.1 Definition:

If function $u(x, t)$ is analytic and differentiated continuously with respect to time t and space x in the domain of interest, then let

$$U_k(x) = \frac{1}{k!} \left[\frac{\partial^k}{\partial t^k} u(x, t) \right]_{t=0} \tag{21}$$

where the t-dimensional spectrum function $U_k(x)$ is the transformed function which is called T-function in brief.

The differential inverse transform of $U_k(x)$ is defined as follows:

$$u(x, t) = \sum_{k=0}^{\infty} U_k(x) t^k \tag{22}$$

Combining (12) and (13) gives the solution as

$$u(x, t) = \sum_{k=0}^{\infty} \frac{1}{k!} \left[\frac{\partial^k}{\partial t^k} u(x, t) \right]_{t=0} t^k \tag{23}$$

For more illustration, consider the general nonlinear partial differential equation:

$$Lu(x, t) + Ru(x, t) + Nu(x, t) = g(x, t),$$

Function Form	Transformed Form
---------------	------------------

$u(x,t)$	$U_k(x) = \frac{1}{k!} \left[\frac{\partial^k}{\partial t^k} u(x,t) \right]_{t=0}$
$u(x,t) = v(x,t) + w(x,t)$	$U_k(x) = V_k(x) + W_k(x)$
$u(x,t) = cv(x,t)$	$U_k(x) = cV_k(x)$ (c is constant)
$u(x,t) = v(x,t) w(x,t)$	$U_k(x) = \sum_{k_1=0}^k v_{k_1}(x) W_{k-k_1}(x)$
$u(x,t) = \frac{\partial}{\partial t} v(x,t)$	$U_k(x) = (k+1)V_{k+1}(x)$
$u(x,t) = \frac{\partial^m}{\partial x^m} v(x,t)$	$U_k(x) = \frac{\partial^m}{\partial x^m} V_k(x)$
$u(x,t) = v^m(x,t) \frac{\partial}{\partial t} w(x,t)$	$U_k(x) = \sum_{k_m=0}^k \sum_{k_{m-1}=0}^{k_m} \dots \sum_{k_2=0}^{k_3} \sum_{k_1=0}^{k_2} \frac{\partial}{\partial x} W_{k_1} V_{k_2-k_1} V_{k_3-k_2} \dots V_{k_{m-1}-k_{m-2}} V_{k-k_{m-1}}$

Table 2

with initial condition

$$u(x, 0) = f(x), \tag{24}$$

where $L = \frac{\partial}{\partial t}$, R is a linear operator which has partial derivatives, $Nu(x,t)$ is a nonlinear term and $g(x, t)$ is an inhomogeneous term. According to the RDTM and Table 1, we can construct

$$(k + 1)U_{k+1}(x) = G_k(x) - RU_k(x) - NU_k(x),$$

where $U_k(x)$, $RU_k(x)$, $NU_k(x)$ and $G_k(x)$ are the transformation of the functions $Lu(x, t)$, $Ru(x, t)$, $Nu(x, t)$ and $g(x,t)$ respectively.

In real application, by consideration of $U_0(x) = f(x)$ as transformation of initial condition (24), the function $u(x,t)$ can be written by a finite series of equation (23) as

$$\widetilde{u}_n(x, t) = \sum_{k=0}^n U_k(x)t^k$$

where n is order of approximation solution.

The exact solution is given by

$$u(x, t) = \lim_{n \rightarrow \infty} \widetilde{u}_n(x, t).$$

3.2 Application of Reduced differential transform Method:

According to the RDTM and table 2, we can construct the following transformation of Equation (8)

$$\frac{\partial h}{\partial T} = \left\{ h \frac{\partial^2 h}{\partial X^2} + \left(\frac{\partial h}{\partial X} \right)^2 \right\}$$

With Initial condition $h(X, 0) = e^{-X}$ as:

$$(k + 1)H_{k+1}(X) = \sum_{k_1=0}^k H_{k_1} \frac{\partial^2}{\partial X^2} H_{k-k_1}(X) + \sum_{k_1=0}^k \frac{\partial}{\partial X} H_{k_1}(X) \frac{\partial}{\partial X} H_{k-k_1}(X) \tag{25}$$

By definition of RDTM

$$H_k(x) = \frac{1}{k!} \left[\frac{\partial^k}{\partial T^k} u(X, T) \right]_{T=0} \tag{26}$$

To apply initial condition putting $T=0$ in the above equation we get

$$H_k(x) = \frac{1}{k!} [u(X, 0)]$$

Putting $k=0$ we get,

$$H_0(x) = e^{-X}$$

Now putting $k=0$ in equation (25) we get,

$$(0 + 1)H_{0+1}(X) = \sum_{k_1=0}^0 H_{k_1} \frac{\partial^2}{\partial X^2} H_{0-k_1}(X) + \sum_{k_1=0}^0 \frac{\partial}{\partial X} H_{k_1}(X) \frac{\partial}{\partial X} H_{0-k_1}(X)$$

$$(1)H_1(X) = H_0(X) \frac{\partial^2}{\partial X^2} H_0(X) + \frac{\partial}{\partial X} H_0(X) \frac{\partial}{\partial X} H_0(X)$$

$$(1)H_1(X) = e^{-X} \frac{\partial^2}{\partial X^2} e^{-X} + \frac{\partial}{\partial X} e^{-X} \frac{\partial}{\partial X} e^{-X}$$

$$H_1(X) = e^{-X} e^{-X} + (-e^{-X})(-e^{-X})$$

$$H_1(X) = 2e^{-2X}$$

Now putting $k=1$ in equation (25) we get,

$$(1 + 1)H_{1+1}(X) = \sum_{k_1=0}^1 H_{k_1} \frac{\partial^2}{\partial X^2} H_{1-k_1}(X) + \sum_{k_1=0}^1 \frac{\partial}{\partial X} H_{k_1}(X) \frac{\partial}{\partial X} H_{1-k_1}(X)$$

$$2H_2(X) = H_0(X) \frac{\partial^2}{\partial X^2} H_1(X) + H_1(X) \frac{\partial^2}{\partial X^2} H_0(X) + \frac{\partial}{\partial X} H_0(X) \frac{\partial}{\partial X} H_1(X) + \frac{\partial}{\partial X} H_1(X) \frac{\partial}{\partial X} H_0(X)$$

$$2H_2(X) = e^{-X} 2(-2)(-2)e^{-2X} + e^{-2X} e^{-X} + (-e^{-X})(-4e^{-2X}) + (-4e^{-2X})(-e^{-X})$$

$$H_2(X) = 8.5e^{-3X}$$

Now putting $k=2$ in equation (25) we get,

$$(2 + 1)H_{2+1}(X) = \sum_{k_1=0}^2 H_{k_1} \frac{\partial^2}{\partial X^2} H_{2-k_1}(X) + \sum_{k_1=0}^2 \frac{\partial}{\partial X} H_{k_1}(X) \frac{\partial}{\partial X} H_{2-k_1}(X)$$

$$3H_3(X) = H_0(X) \frac{\partial^2}{\partial X^2} H_2(X) + H_1(X) \frac{\partial^2}{\partial X^2} H_1(X) + H_2(X) \frac{\partial^2}{\partial X^2} H_0(X) + \frac{\partial}{\partial X} H_0(X) \frac{\partial}{\partial X} H_2(X)$$

$$+ \frac{\partial}{\partial X} H_1(X) \frac{\partial}{\partial X} H_1(X) + \frac{\partial}{\partial X} H_2(X) \frac{\partial}{\partial X} H_0(X)$$

$$3H_3(X) = 143$$

So $H_3(X) = 47.67$

So we got $H_0(X) = e^{-X}$, $H_1(X) = 2e^{-2X}$, $H_2(X) = 8.5e^{-3X}$, $H_3(X) = 47.67$

And so on...

By equation (22)

$$h(X, T) = \sum_{k=0}^{\infty} H_k(X) T^k$$

$$h(X, T) = H_0(X) + H_1(X)T + H_2(X)T^2 + H_3(X)T^3 + \dots$$

$$h(X, T) = e^{-X} + 2e^{-X}T + 8.5e^{-X}T^2 + 47.67e^{-X}T^3 + \dots \quad (27)$$

4. Conclusion : From initial conditions and equation (19) and (27) we can see that with same initial conditions we got similar answer and RDTM is easier to apply than DTM.

References:

1. J.K.Zhou, Differential Transformation and Its Application for Electrical Circuits, Huazhong University Press, Wuhan, China, 1986.
2. R.Abazari and A.Borhanifar, " Numerical study of the solution of the Burgers and coupled Burgers equations by differential transformation method," Computers & Mathematics with Applications, vol.59, no.8, pp.2711-2722, 2010.
3. A.Borhanifar and R.Abazari, "Exact solutions for nonlinear Schrodinger equations by differential transformation method," Journal of Applied Mathematics and Computing, vol.35, no.1-2, pp.37-51, 2011.
4. A.Borhanifar and R.Abazari, "Numerical study of nonlinear Schrodinger and coupled Schrodinger equations by differential transformation method," Optics Communications, vol.283, no.10, pp.2026-2031, 2010.
5. R.Abazari and M.Abazari, "Numerical simulation of generalized Hirota-Satsuma coupled KdV equation by RDTM and comparison with DTM," Communications in Nonlinear Science and Numerical Simulation, vol.17, no.2, pp.619-629, 2012.
6. R.Abazari, "Solution of Riccati type matrix differential equations using matrix differential transform method," Journal of Applied Mathematics & Informatics, vol.27, pp.1133-1143, 2009.
7. R.Abazari and R.Abazari, "Numerical study of some coupled PDEs by using differential transformation method," World Academy of Science, Engineering and Technology, vol.66, pp.52-59, 2010.

8. S. Pamuk, Solution of the porous media equation by Adomians decomposition method, Phys. Lett. A 344 (2005), 184-188.
9. H. Saberi nik, M. S. Zahedi, R. Buzhabadi, S. E ati, Homotopy perturbation method and Hes polynomials for solving the Porous media equation, Comp. Math. Model., 24(2) (2013) 279-292.
10. E. Abdolmaleki, S.A. Youse , Application of VIM method for nonlinear Porous media equations, Am. J. Numer. Anal., 2(1) (2014) 11-13
11. M. Sari, Solution of the porous media equation by a compact nite di erence method, Math. Probs. Eng., 2009 (2009) Article ID 912541.
12. Afshin Babaei, Alireza Mohammadpour, Analytical approximations of the porous medium equations by reduced differential transform method 2015
13. Amit Parikh, LAP LAMBERT Academic publicatuion, solution of one dimentional groundwater flow, problems, 2015
14. Bear J. and Bachmat Y., Introduction to Modeling of Transport Phenomena in Porous Media, Kluwer, Academic Publishers, Dordrecht, The Netherlands, 1990.
15. Bear J., Dynamics of Fluids in Porous Media, American Elsevier, New York, 1972.
16. Boussinesq J., Comptes Rendus Acad. science Journal, Math. Pures Appl., 10, 5-78, 1903.
17. Darcy H., Les Fontaines publiques de loville de Dijon, Dalmont, Paris, 1856.
18. Desai N. B. and Mehta M. N., A Solution of Seepage of ground water in soil: Homotopy Perturbation method. Int. J. of Appl. Math and Mech. 6 (2): 85-92, 2010.

STATISTICAL ANALYSIS OF AIR POLLUTION, DISEASES AND MORTALITY: A CASE STUDY OF DELHI

¹Deepakshi Kasat, ²Bhairavi Savur, ³Poonam Soni

^{1,2}Biotechnology Engineering, ³Department of Mathematics
Thadomal Shahani Engineering College, Mumbai, India

ABSTRACT: *The alarming rate at which the air quality of cities worldwide is decreasing, has become a major cause for concern. In India, especially Delhi, the Central Pollution Control Board has categorized its air quality index (AQI) as 'severe'. There have been various evidences linking the pollution levels to deteriorating health conditions. This paper presents the significance of considering air pollution as a serious health hazard by analyzing the relationship between air pollution, climate change and diseases. Levels of NO₂, SO₂ and particulate matter 10 in the atmosphere of Delhi over a period of 16 years are considered. Average annual temperature and humidity are used as climate change parameters for a period of 7 years.*

The diseases studied are Chronic Obstructive Pulmonary Diseases (COPD), Ischemic Heart Diseases (IHD), and Ischemic Stroke (IS), based on their high degree of correlation with considered atmospheric parameters. Pearson's correlation coefficient for each disease with respect to the environmental factors are calculated. PM₁₀ and NO₂ show strong correlation with each disease over the 16-year period, however PM₁₀ correlation weakened when considered for only the recent years. Temperature too showed a strong correlation with all diseases consistently, whereas humidity seemed to have no major effect. Hence, it is inferred that PM₁₀, NO₂ and temperature change play a significant role in controlling the disability adjusted life year (DALYs), years lived with disability (YLDs) and deaths caused by COPD, ISH and IS.

Keywords: : Pearson's Correlation, COPD, PM₁₀, air pollution, DALYs

I. INTRODUCTION

Pollution refers to the contamination of the earth's environment with materials that interfere with human health, quality of life or the natural functioning of the ecosystems (Rizwan et al, 2013). Environmental problems in Delhi, India, are a threat to the well-being of the city's and area's inhabitants as well as the flora and fauna. Delhi, the sixth-most populated metropolis in the world (include all of NCR and it becomes second most populous metropolitan area), is one of the most heavily polluted cities in India, having for instance one of the country's highest volumes of particulate matter pollution. In May 2014 the World Health Organization announced New Delhi as the most polluted city in the world (Paradkar et al, 2016). Even though there has been acceptance of this fact, there is yet more place for research regarding the extent of pollution, the effect of pollution on health, the degree of this effect, the outcome of policies made, and what further actions can be taken to lessen it.

The prime air pollutants have been broadly classified as outdoor and indoor pollutants. Outdoor pollutants include remains of fossil fuel, carbon particles and metallic particles in the atmosphere from industrial and automobile emissions, toxic gases i.e., nitrogen dioxide, carbon monoxide, sulfur dioxide etc. and ozone, tobacco smoke etc. On the otherhand, indoor pollutants include toxic gases produced from kitchen fuels, building materials i.e., asbestos, lead, tobacco smoke (Ghosh and Parida, 2015). This paper studies the level of pollution by considering air pollutants- NO₂, SO₂ and PM₁₀, as well as climate change (temperature and humidity), and their effects on the city's health and disease statistic.

Air Pollutants

NO₂

NO₂ is a yellow brown gas. According to a report by the Greenpeace organization (Myllyvirta, 2018), India has three of the worst 50 nitrogen dioxide emissions sources. Oxides of nitrogen are formed by burning of fuel containing nitrogen at high temperatures. The main sources of NO₂ at the hotspots from highest emissions to lowest are public electricity and heat production, manufacturing industries and construction, road transportation, residential and other sectors, agricultural waste burning and other energy industries. The report states that emissions in New Delhi are a mix of multiple contributing sources including transport, coal power plants and manufacturing.

SO₂

SO₂ is an air pollutant that causes acid rain, haze and many health-related problems and is emitted when coal is burned to produce electricity. On burning coal, the sulfur that it naturally contains combines with oxygen to form sulfur oxides. According to a study led by University of Maryland, US, India is becoming the biggest emitter of SO₂ (Li et al, 2017). Since 2007 India’s sulfur emission has increased 50%. The largest emission sources are not in the most densely populated areas of India and thus SO₂ emissions are not causing as many health problems, but as the demand for electricity grows the impact may worsen, researchers said.

PM₁₀

PM stands for particulate matter and 10 is the diameter measurement- 10 micrometers or less. There is no single source for this mixture of particles. Some of the main sources are traffic, industry, domestic fuel burning, natural sources like sea salt and pollen (Karagulian et al, 2015). According to the data released by WHO (World Health Organization) 14 Indian cities, including Delhi were in a list of worlds’ most polluted in terms of PM₁₀ (WHO, 2015).

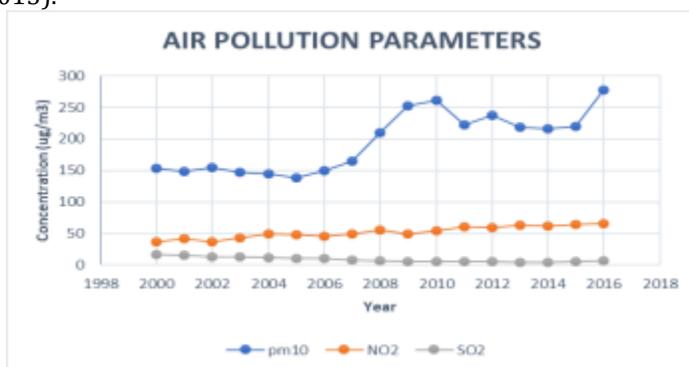


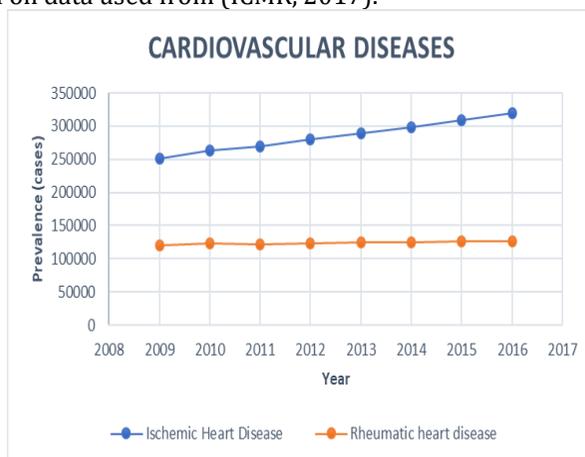
Fig.1 Air Pollution Parameters

Health and Diseases

Air pollution has been on a rise for over two decades now, and its effect on health has started to become prominent since a few years. The effects of deteriorating air quality have been studied on the prevalence of the following diseases/disorders:

- Breathing and respiratory disorders
- Cardiovascular diseases
- Stroke
- Neonatal disorders
- Infertility (male and female)

The graphs below are based on data used from (ICMR, 2017).



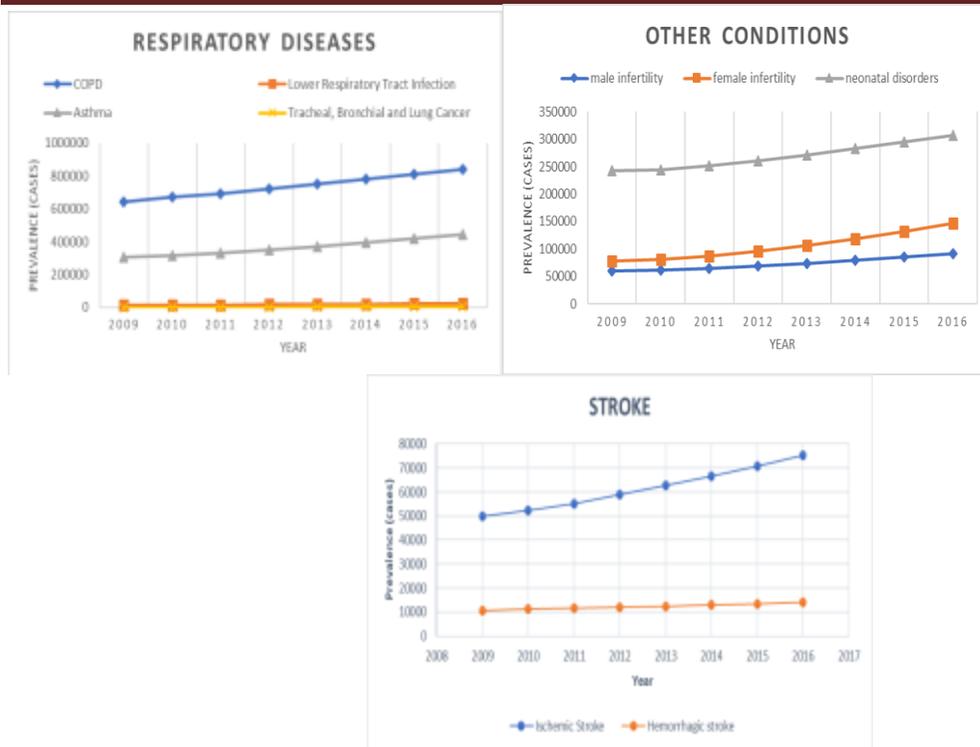


Fig.2 Respiratory Diseases, Cardiovascular Diseases, Stroke, Other Conditions

1. Respiratory Diseases

The presence of pollutants in the air damage the cells of the lung tissue. Exposure may worsen previously existing respiratory diseases or cause diseases like COPD, asthma, tracheal, bronchial and lung cancer and lower respiratory tract infection (Lockwood et al, 2009).

Chronic Obstructive Pulmonary Disease (COPD)

The highest prevalence has been shown by COPD since 2009. It is an umbrella term used to describe progressive lung disease including emphysema and chronic bronchitis. The major cause is cigarette smoking and other types of tobacco smoke, air pollution, chemical fumes or dusts from the environment (Kopsaftis et al, 2018).

2. Cardiovascular

Exposure to air pollution mainly particulate matter may cause short term as well as long term cardiovascular effects. It may cause blocking of the arteries resulting in strokes, myocardial infarction and cardiac failure.

Ischemic Heart Disease (IHD)

A condition in which the blood flow is restricted or reduced to the heart muscle. Some of the known causes include smoking, high cholesterol and blood pressure. It is also dependent on other factors like age, family history and sex to some extent. However, having shown the greater prevalence of the two diseases considered, and with evidence of its increasing trend with air pollution's, it forms an interesting disease to be studied further (Science Daily).

3. Stroke

The presence of particulate matter increases the risk for strokes by blocking the arteries.

Ischemic Stroke (IS)

Ischemic Stroke occurs when an artery to the brain is blocked, which may cause the brain cells to die. This happens due to the narrowing of blood vessels as a result of fatty deposits or plaque build-up. The risk factors include high blood pressure, smoking, high cholesterol and use of illegal drugs like cocaine, heroin, ecstasy etc. Other factors include age, gender and heredity (Ringer and Jimenez, 2018).

4. Other:

Infertility and neonatal disorders trends have also shown considerable similarity to the air pollution parameters' increasing graph. Hence, based on prevalence over the past 8 years, chronic obstructive pulmonary disease, ischemic heart disease and ischemic stroke have been top most victim diseases of the

ones considered. Statistical analysis and mathematical models(Soni et al, 2019) are important tools to measure the degree of effect and predict influence of environmental factors on diseases.

Measures

1. YLDS-Years lost due to Disability.It measures the amount of time of living with a disability.
2. DALYs-Disability-adjusted Life year (Havelaar et al, 2014). WHO defines DALY as one lost year of healthy life.

DALYs= YLLs (years of life lost) + YLDs

3. Death-Number of deaths occurring due to the disease.

OBJECTIVE AND SCOPE

The paper studies the Air Pollution and Climate Change trend of Delhi and identify its correlation with health and diseases. The diseases with the highest correlation with air pollution and climate change are identified and further categorized based on age and gender. The environmental effects on each category are identified.

The study attempts to understand the differences in correlation of each air pollution factor and diseases with time.However, the present paper covers 3 diseases in particular (Chronic Obstructive Pulmonary Disease, Ischemic Heart Disease and Ischemic Stroke). It also takes into consideration annual average air pollution parameters of Delhi from 2000 to 2016.

STATISTICAL ANALYSIS

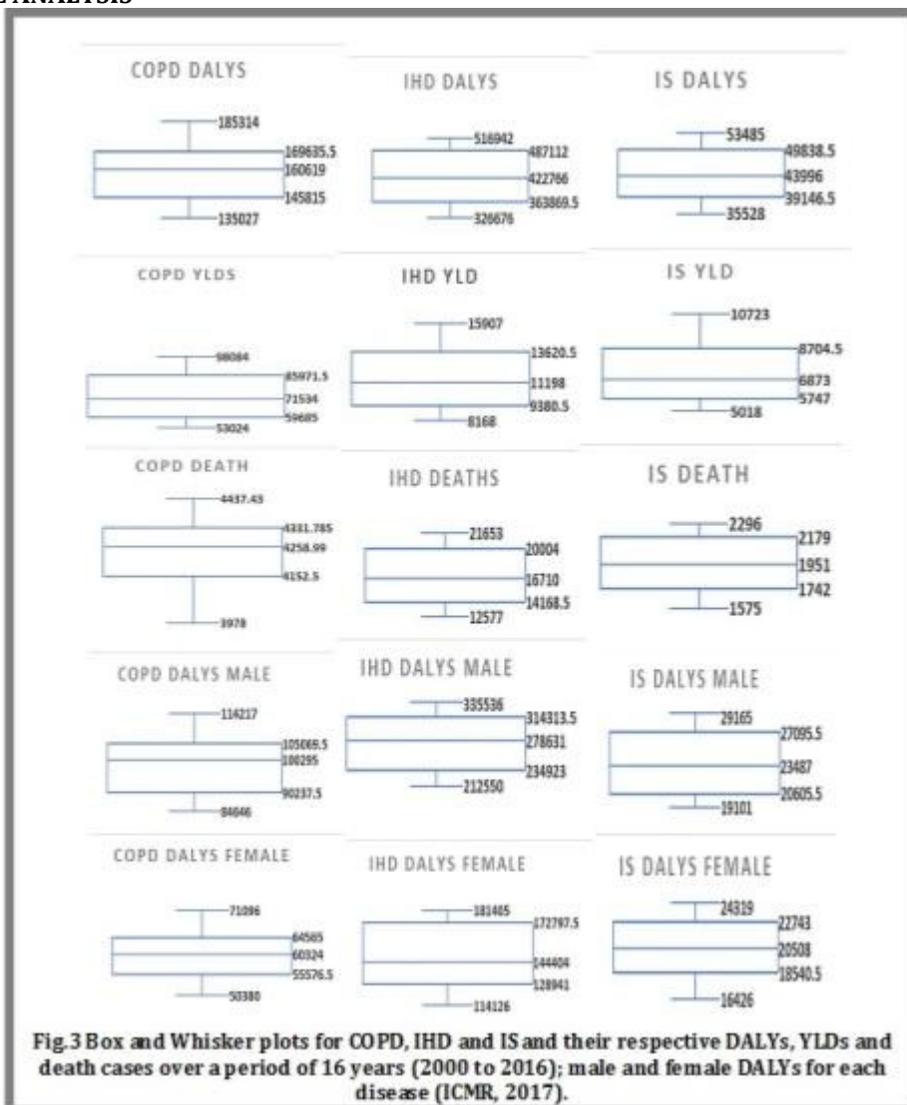


Fig.3 Box and Whisker plots for COPD, IHD and IS and their respective DALYs, YLDs and death cases over a period of 16 years (2000 to 2016); male and female DALYs for each disease (ICMR, 2017).

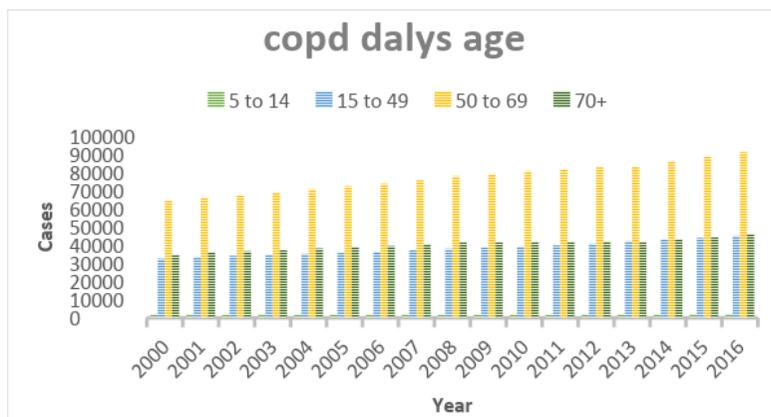
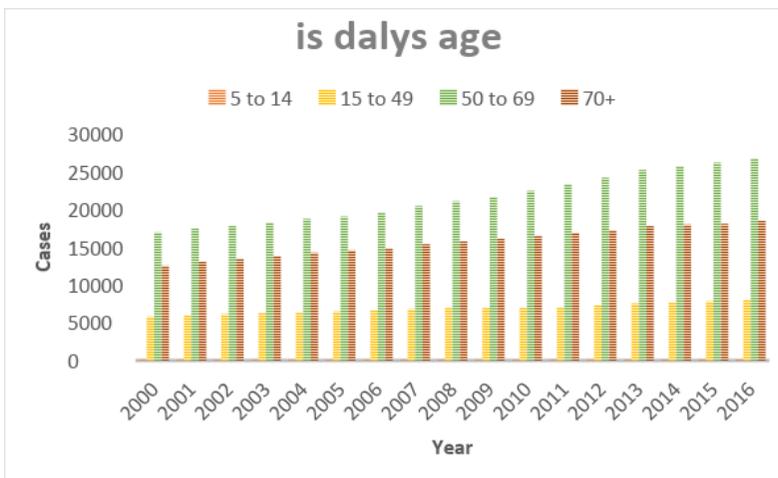
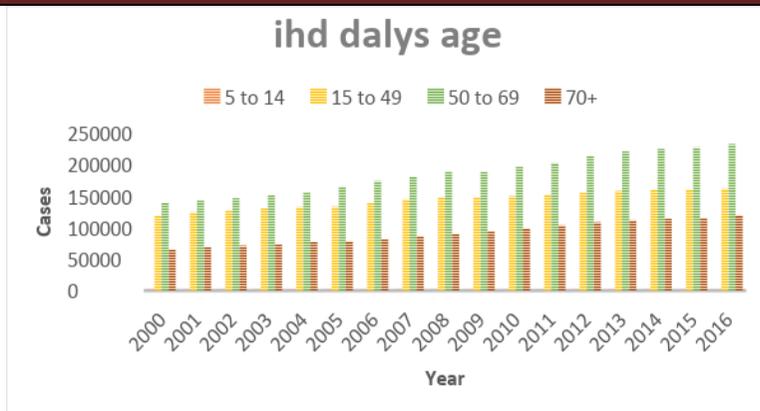


Fig.4 Graphs for age wise DALYs of COPD, IHD and IS (ICMR, 2017).

Based on the above data collected(CPCB, 2016)(ICMR, 2017), Pearson’s Correlation Matrices are shown below and correlation coefficients more than 0.5 are considered to be “strong”.

	2000-2008			2009-2016			2000-2016		
	PM ₁₀	NO ₂	SO ₂	PM ₁₀	NO ₂	SO ₂	PM ₁₀	NO ₂	SO ₂
COPD DALY	0.5872	0.9097	-0.9871	0.1099	0.8332	0.3056	0.8391	0.9580	-0.8937
COPD YLD	0.6211	0.8971	-0.9804	-0.0447	0.9025	0.1339	0.8357	0.9563	-0.8713
COPD	0.5211	0.9181	-0.9889	0.7724	-0.1445	0.8904	0.4790	0.5075	-0.5851

DEATH									
IHD DALY	0.6428	0.8990	-0.9758	-0.1810	0.9358	0.0043	0.8436	0.9637	-0.9118
IHD YLD	0.5828	0.9007	-0.9853	-0.0282	0.8875	0.1552	0.8224	0.9533	-0.8583
IHD DEATH	0.6382	0.9020	-0.9773	-0.1523	0.9360	0.0314	0.8419	0.9620	-0.8961
IS DALY	0.5946	0.9083	-0.9862	-0.1484	0.9251	0.0372	0.8244	0.9654	-0.8948
IS YLD	0.5984	0.9077	-0.9826	-0.0201	0.8845	0.1733	0.7952	0.9394	-0.8084
IS DEATH	0.5849	0.9102	-0.9873	-0.1770	0.9344	0.0022	0.8332	0.9674	-0.9188
COPD DALYS M	0.6559	0.9013	-0.9763	0.1640	0.8118	0.3500	0.8624	0.9542	-0.9005
COPD DALYS F	0.4733	0.9055	-0.9852	0.0435	0.8552	0.2503	0.7996	0.9584	-0.8786
IHD DALYS M	0.6911	0.8917	-0.9638	-0.1238	0.9286	0.0492	0.8587	0.9617	-0.9176
IHD DALYS F	0.5390	0.9031	-0.9862	-0.2644	0.9381	-0.0625	0.8132	0.9633	-0.8972
IS DALYS M	0.6804	0.8984	-0.9690	-0.1300	0.9319	0.0410	0.8433	0.9594	-0.8834
IS DALYS F	0.4965	0.9038	-0.9880	-0.1720	0.9126	0.0322	0.7931	0.9674	-0.9047
COPD DALYS 5-14	-0.3728	0.2226	-0.2300	-0.0160	0.8054	0.0844	0.4209	0.6825	-0.4300
COPD DALYS 15-49	0.5907	0.8940	-0.9871	-0.0363	0.8837	0.1364	0.8163	0.9553	-0.8664
COPD DALYS 50-69	0.5887	0.9124	-0.9837	0.0983	0.8444	0.3038	0.8455	0.9600	-0.8992
COPD DALYS 70+	0.5810	0.9117	-0.9897	0.3649	0.6555	0.5579	0.8353	0.9324	-0.8985
IHD DALYS 5-14	-0.8455	-0.5890	0.4892	-0.2830	-0.8032	-0.3830	-0.9130	-0.8993	0.7955
IHD DALYS 15-49	0.6429	0.8891	-0.9748	-0.1878	0.9063	0.0042	0.8351	0.9552	-0.9518
IHD DALYS 50-69	0.6155	0.8801	-0.9693	-0.1759	0.9226	-0.0307	0.8219	0.9634	-0.9110
IHD DALYS 70+	0.6190	0.9033	-0.9815	-0.1418	0.9365	0.0388	0.8423	0.9624	-0.8986
IS DALYS 5-14	0.3416	0.8015	-0.9139	0.0194	0.8290	0.2082	0.5935	0.8015	-0.5681
IS DALYS 15-49	0.5741	0.9041	-0.9895	-0.0735	0.8663	0.1081	0.7911	0.9590	-0.8831
IS DALYS 50-69	0.6379	0.9056	-0.9778	-0.1647	0.9308	0.0249	0.8243	0.9601	-0.8747
IS DALYS 70+	0.5438	0.9077	-0.9898	-0.1477	0.9330	0.0301	0.8292	0.9677	-0.9259

	T	H		T	H		T	H
COPD DALYS	0.6526	-0.71	COPD DALYS M	0.65	-0.71	COPD DALYS 70+	0.73	-0.69
COPD YLD	0.5794	-0.68	COPD DALYS F	0.65	-0.71	IHD DALYS 5-14	-0.63	0.52
COPD DEATH	0.1442	0.11	IHD DALYS M	0.46	-0.64	IHD DALYS 15-49	0.47	-0.68
IHD DALYS	0.4358	-0.62	IHD DALYS F	0.40	-0.61	IHD DALYS 50-69	0.47	-0.65
IHD YLDS	0.6107	-0.71	IS DALYS M	0.47	-0.64	IHD DALYS 70+	0.44	-0.62
IHD DEATH	0.4420	-0.63	IS DALYS F	0.56	-0.69	IS DALYS 5-14	0.71	-0.79
IS DALYS	0.4420	-0.66	COPD DALYS 5-14	0.74	-0.82	IS DALYS 15-49	0.65	-0.76
IS YLDS	0.5095	-0.71	COPD DALYS 15-49	0.63	-0.72	IS DALYS 50-69	0.49	-0.65
IS DEATH	0.6145	-0.63	COPD DALYS 50-69	0.61	-0.67	IS DALYS 70+	0.47	-0.63

(T- Temperature, H- Humidity) (worldweatheronline)

RESULT AND DISCUSSION

From the above analysis, the following correlations can be established:

- i. From 2000-2008, PM₁₀ and all considered disease parameters show a “strong” relationship (exceptions: COPD DALYS Female, IS DALYS Female, age 5 to 14 for all three diseases)
- ii. From 2000-2016, PM₁₀ shows a “very strong” correlation with all considered disease parameters (exceptions: PM₁₀- COPD death, COPD and IHD 5 to 14 age)
- iii. However, when PM₁₀ is considered from 2009 to 2016, it showed a very poor relationship with all disease parameters (except COPD death).
- iv. NO₂ shows a “very strong” relationship throughout the 16-year period (exception: IHD 5 to 14 age)
- v. SO₂ consistently showed “poor” to “very poor” relationship with all considered diseases throughout. (exceptions: COPD death, COPD age 70+, IHD age 5 to 14)
- vi. Temperature showed “fair” to “strong” relationship overall, whereas humidity showed majorly “poor” throughout.

CONCLUSION:

The paper makes an attempt to analyze the correlation between diseases COPD, Ischemic Heart Disease and Ischemic Stroke and different pollution parameters like SO₂, NO₂, PM₁₀, temperature and humidity taking the city of Delhi as a case study. The correlation matrix of all three diseases shows a strong correlation is observed between NO₂ and PM₁₀ whereas temperature and humidity show fair and poor correlation respectively. Climate change (temperature) therefore, can be considered as an important contributor to health issues in the near future with respect to COPD, IHD, IS.

Considering that, air pollution factors show similar trends like cardiovascular and respiratory diseases and ischemic stroke, it is safe to statistically call it a health hazard. It may be concluded that the incidence of these diseases (COPD, IHD AND IS) is a function of pollution parameters (NO₂ and PM₁₀) and therefore in order to address the health issues associated with these diseases the discussed pollution parameters showing highest correlation must be brought under safe limits. SO₂ shows a poor correlation as the release of SO₂ in air is more prominent outside the cities, in factory setups. The paper does not consider seasonal changes in environmental factors as well as factors like vehicle density, road dust and industrial pollution separately. Therefore, the paper is amenable to inclusion of statistical analysis of data on these factors.

REFERENCES

1. Ghosh, D. and Parida, P. 2015. Air Pollution and India: Current Scenario. International Journal of Current Research. Vol.7, Issue 11.
2. Havelaar, A., Devleeschauwer, B., Torgerson, P., Duchateau, L. and Oyen, H. 2014. DALY Calculation in Practice: A Stepwise Approach. International Journal of Public Health.
3. Karagulian, F., Belis, C., Dora, C., Prüss-Ustün, A., Bonjour, S., Adair-Rohani, H. and Amann, M. 2015. “Contributions to Cities’ Ambient Particulate Matter (PM): A Systematic Review of Local Source Contributions at Global Level.” Atmospheric Environment, Elsevier.
4. Kopsaftis, Z., Wood-Baker, R. and Poole, P. 2018. Influenza vaccine for people with chronic obstructive pulmonary disease (COPD), Cochrane Database of Systematic Reviews, Issue 6. Art. No.: CD002733.
5. Li, C., McLinden, C., Fioletov, V., Krotkov, N., Carn, S., Joiner, J., Streets, D., He, H., Ren, X., Li, Z. and Dickerson, R. 2017. India Is Overtaking China as the World’s Largest Emitter of Anthropogenic Sulfur Dioxide. Nature search Journal.
6. Lockwood, A., Welker-Hood, K., Rauch, M. and Gottlieb, B. 2009. Coal’s Assault on Human Health. A Report from Physicians for Social Responsibility.

7. Myllyvirta, L. 2018. New satellite data reveals world's largest NO₂ air pollution emission hotspots-Greenpeace Media Briefing. Retrieved from <https://storage.googleapis.com/planet4-international-stateless/2018/10/07426a79-no2-air-pollution-analysis-greenpeace.pdf>.
8. Paradkar, A., Nanivadekar, R. and Gharad, P. 2016. Research Report on Delhi Air Pollution. International Journal of Science Technology and Management. Vol.5, Issue 3.
9. Ringer, A. and Jimenez, L. 2018. Stroke (brain attack). Mayfield Brain and Spine, Mayfield Clinic.
10. Rizwan, S., Nongkynrih, B. and Gupta, S. 2013. Air Pollution in Delhi: Its Magnitude and Effects on Health. Indian Journal of Community Medicine. 38. 4-8.
11. Soni, P., Kumar, A. and Rani, A. 2019. Fluoride Ion Leaching Kinetics in Soil: A Mechanistic Model, International Journal of Emerging Technologies and Innovative Research, Vol.6, Issue 3, pp.135-137.
12. Air Pollution in Delhi, An Analysis. Central Pollution Control Board. ENVIS Centre on Control of Pollution Water, Air and Noise. Retrieved from http://cpcbenvs.nic.in/envis_newsletter/Air%20pollution%20in%20Delhi.pdf (2016)
13. Database on Source Apportionment Studies for Particulate Matter in the Air (PM₁₀ and PM_{2.5}). World Health Organization. Retrieved from www.who.int/quantifying_ehimpacts/global/source_apport/en/ (2015).
14. Indian Council of Medical Research, Public Health Foundation of India, and Institute for Health Metrics and Evaluation. 2017 GBD India Compare Data Visualization. New Delhi: ICMR, PHFI, and IHME.
15. Ischemic Heart Disease. ScienceDaily. Retrieved from www.sciencedaily.com/terms/ischaemic_heart_disease.htm
16. New Delhi Monthly Climate Averages. Retrieved from www.worldweatheronline.com/lang/en-in/new-delhi-weather-averages/delhi/in.aspx

THE ADDRESS APP

**Er. Kashif Kaularikar, Aves Cutleriwala, Abdul Wahab Machiwala, Mohd Maaz Khan,
Khan Umair**

Assistant Professor, Student, Student, Student, Student
Information Technology.
M. H. Saboo Siddik College of Engineering, Mumbai, India.

ABSTRACT: *"The Address App" is an Online Real Estate Management System which is Android and Web based related to online application of estate. The biggest advantage of this is, it provides an online tool to communicate between Buyer and Seller through Agents. The main objective of this project is to give a platform to the agents to establish themselves which are very much underrated in our country. The Seller can sell its property by contacting the suited Agent and post an Ad on the Android Application and on Website, interested Buyer can check the Ad and they can contact the Agent which holds the property, for rent, lease or purchase. Our concept is to give best and quick result to the user.*

Keywords: : Agents, seller, buyer, app, website.

I. Introduction

"The Address App" is a startup thought which will be by and large contacting individuals who will buy/sell their property at a viable cost. This thought will give a help to the Agents which are underrated in this nation. Agents will assume an extremely fundamental job in this undertaking thought. For the most part, when the buying/selling of a property is done, the Agents ought to get 1% from the Buyer side and 1% from the Seller side that implies aggregate of 2% individual of the rate of property sold. Nonetheless, India is where Agents are given less significance in this field. Developing such an Android Application and a Website which will guarantee the Agents that they'll get the complete 2% from both sides (Buyer and Seller). Aim of the project is to make an android application and a website which systematically handles and manages the users who are willing to buy/sell properties at an effective price through the Agents. Users will have the capacity to surf the properties on Android Application as well as on the Website.

II. Existing System

Like those of different enterprises, organizations in land have overwhelmingly digitized their procedures to expand economy. In any case, for some organizations, those procedures are altogether housed and oversaw on various, contradictory stages. Bringing together information and making it effectively open to workers who require it encourages organizations to receive the best rewards of their advanced relocation. Determining these advantages is particularly imperative to land organizations, as they handle monstrous measures of printed material for every property with which they are included. The real estate fundamentally incorporates a client, a dealer, and an agent. The purchaser scans for the property and contacts the regarded agent as and when required. The impact in any real estate is on the end client. The application can likewise be utilized as a stage to offer the properties, the agent again goes about as a middle person and posts the property for the dealer.

- *Magic bricks*

Magicbricks.com is a property portal that serves to the global market with its services and many online features. It was launched in the year 2006 by Times Group, Magicbricks has quickly risen to the No. 1 spot in the Property Portal in India. The Magicbricks design is based on rigorous research, unique product developments, and innovative initiative which has been accepted by users all over the world. To serve the users, features on the Magicbricks portal are constantly invented, evaluated and updated. In addition to this, magicbricks.com is synonymous with internationally acclaimed property fairs. It displays top-notch properties of major Indian metros are showcased for audiences in India and abroad. Till date, magicbricks.com has successfully hosted 75 domestic and over 15 international property shows around the globe.

- *99acres*

99Acres is an Indian real estate database website founded in the year 2005. Its parent company is Info Edge, established by Sanjeev Bikhchandani. Its headquarters are in Noida, Uttar Pradesh. Info Edge also owns jobs portal Naukri.com. Since 2011, 99acres users can browse and search for different properties through a mobile phone.

- *Realtor*

Realtor.com is operated by the real estate network Move, Inc., It is owned by News Corporation. Ryan O'Hara is the chief executive officer (CEO) of both realtor.com and also Move. The website is licensed to operate by the National Association of Realtors (NAR) which is the real estate industry's largest trade association. The company's business model focuses on selling advertising and leads to agents and brokers around the world. Parent office located in San Jose, California, the company moved to Santa Clara by 2016. The new headquarters was designed by the Gensler which is an architectural firm to resemble a "deconstructed house".

- *Common floor*

Common floor provides users with the following products and processes

This ecosystem innovatively combines at a single stop:

1. Exhaustive search options for renting and buying property.
2. Easy to use search analytics.
3. Apartment and locality reviews by other users.
4. A comprehensive and verified data base of apartments across India.

Apartment community management tools.

III. Proposed System

Our framework essentially plans to focus more on the agents. The present sprinters in the market impact more on the clients than the agents. The issue with the real estate management is that the agent isn't paid legitimately therefore prompting wrongdoing and remuneration. We have proposed to relegate just a single property to one agent, no two agents will be dole out to a similar property in this manner defeating the issue of repetition. Existing systems in the market does not have an overseen domain regarding agents. The agents will post the photographs of the property through the website, the end users and the sellers will deal with from the android application as well as on the Website and the Agents will deal with the activities through the website. The properties and data posted about any property will be genuine and no falsified data will be given. Another component in our proposed framework will be the User perspective of the property, we will utilize a map view, to make the information easy to use and to enhance the User association. We have additionally added an option to store the information and utilize it, the end user can save the properties that they like.

Selecting agents

Each and every agent working in address app will be authorized and there will be a systematic approach to recruit Agents.

The approach:

- The company will pick up agents based on the RERA license.
- On the account of the language spoken by the agent.
- The past experiences of the agent based on the work he has done
- Finally, an interview process will be done by the HR of the company.

In the event that some Agent has officially posted the property then no other Agent will post the equivalent property thereby decreasing the redundancy. The commission of the Agent is just liable on the seller side when leasing a property. No arbitrary individual can transfer the property, in contrast to competitors, just the approved Agent can post the data about the property, in this way diminishing the property misrepresentation.

Interaction through app

- Registration module: Users have to register and after successful registration user get their login id and password which they will require to log into the app
- Login module: login module will help in authentication of user accounts. User who has valid login id and password can only login to system.
- Search for their required property: Users must fill the complete details whatever mentioned and filter their type of property.

- Contacting the agent: User will get the information of the respected agent holding the respected property and then they will further negotiate

Interaction through website

- Login module: Valid Users and Agents can log into the system.
- View any requests from customers: Agents can view the list of customers who have requested for negotiation.
- Update: Agent can edit \ update house details.
- Responding the call from the seller: The agent will respond to the seller if he/she wantsto sell their property.
- Posting photos: Agents will be the one who will be posting photos of the properties that are going to be sold.
- Search for their required property: Users must fill the complete details whatever mentioned and filter their type of property.
- Contacting the agent: User will get the information of the respected agent holding the respected property and then they will further negotiate.

Favorite option

The app also offers an option which enables user to mark their favorite properties. This enables the user to minimize time consumption and can easily reach out to the selected properties.

Realistic Viewing

Another component in our proposed framework will be the User perspective of the property, we will utilize a map view, to make the information easy to use and to enhance the User association. We have additionally added an option to store the information and utilize it, the end user can save the properties that they like. It gives users a realistic view of the properties they are willing to post or buy.

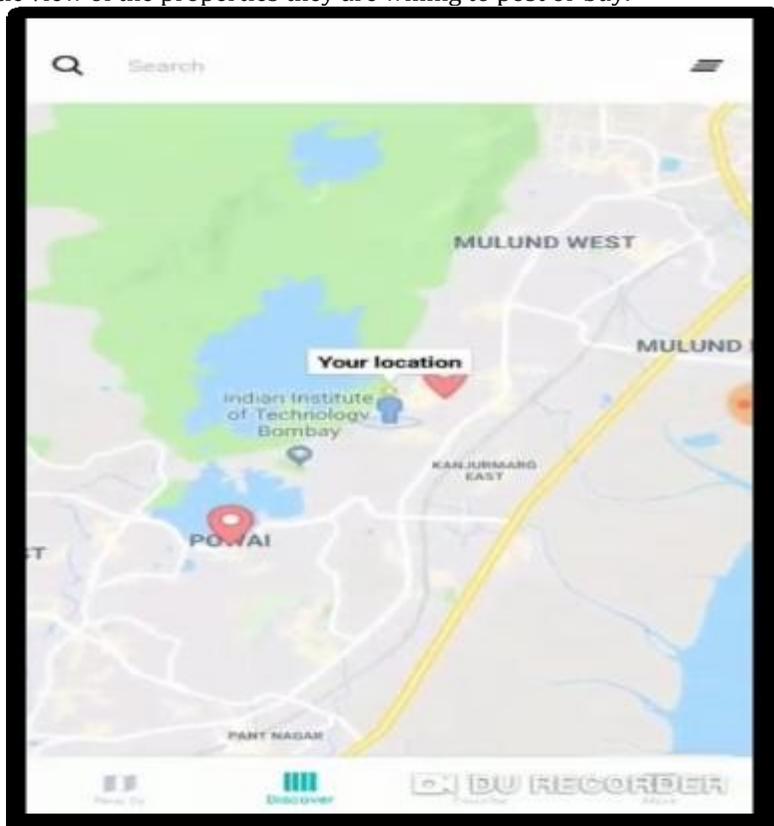


Figure 1. Map View

Sequence diagram

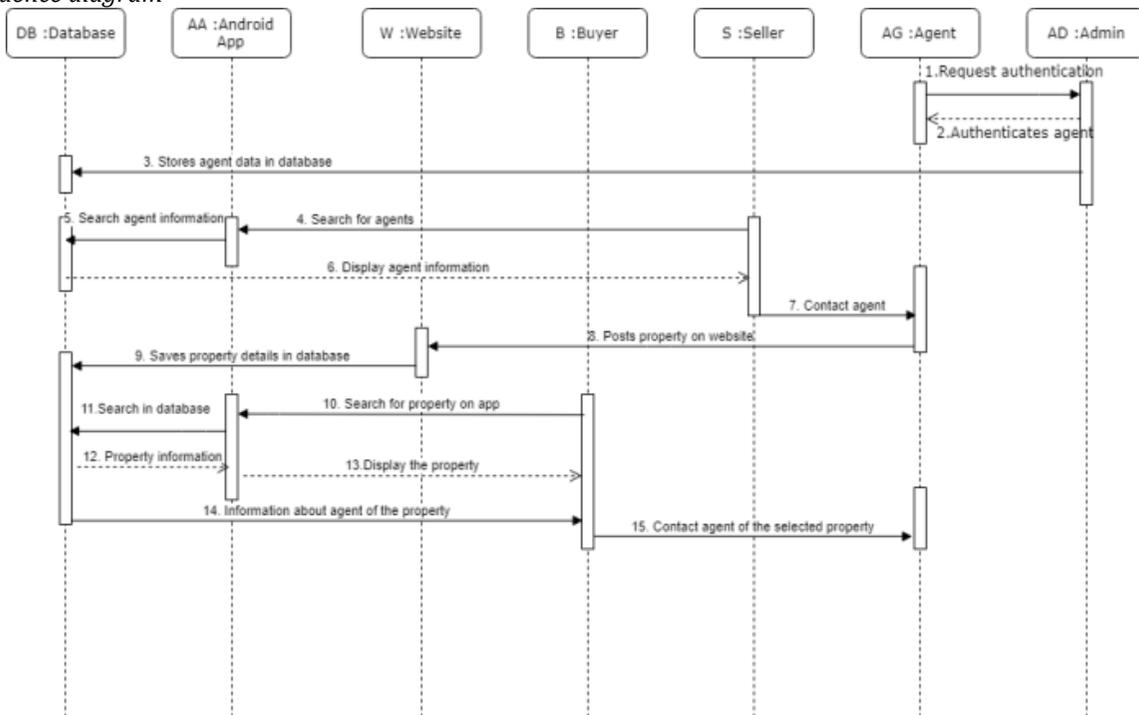


Figure 2. Sequence diagram

System Architecture

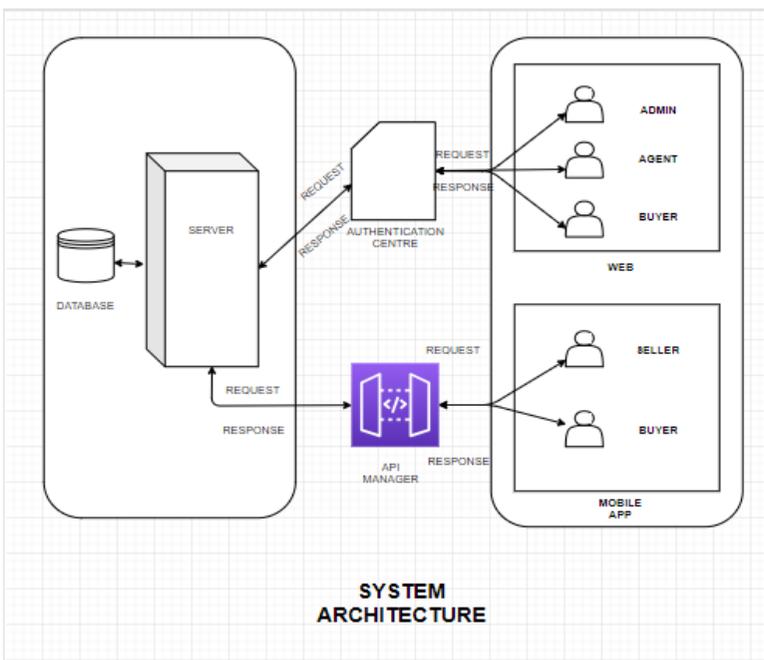


Figure 9. System Architecture

IV. RESULTS AND CONCLUSION

This is an Android and Website, which will be used to buy/sell your properties at an effective price through Agents. This system includes easy UI for every module that makes it efficient for further processing of the application. The project emphasizes more on the agents as the mediators between the user and the seller. It gives a platform to the agents to establish themselves and will be hired for the same.

The idea of assigning only one property to only one agent makes this project more realistic and flexible. The data redundancy and property theft will be reduced immensely courtesy of this function of the project. No two agents will be dole out to a similar property in this manner defeating the issue of repetition.

V. Future Scope

Data Mining systems can be introduced for better managing of data, finding patterns, help in making better decisions. It can help in predetermining what a user desires and to show the required properties based on the previous results. Artificial Intelligence can help in creating more efficient search by using different search techniques

VI. References

1. Peter Ekman, Randle D.Raggio & Steven Thompson, "Developing Smart Commercial Real Estate".
2. Carlos D. Barranco, Jesus R. Campana and Juan M. Medina, "A Real Estate Management System based on Soft Computing".
3. Ting Kien Hwa and Mona Isa, "The Significance of Information Technology on Real Estate Agency Firm Earnings".
4. Yao Fengge, Han Jing, "The Empirical Study on Fractal Characteristics of Real Estate Market", 2001.8-2010.9.
5. Yi Bin, Liu Weijun, "Research on Prediction Methods of Residential Real Estate Price based on improved BPNN".
6. Xiaojin Zhong, Haizang Kuai, "Research on the problems and solutions of real estate internet advertising".
7. Zhangxi Lin, G.D. Anderson, T.J. Anderson, "Web based services for real estate: model and implementation".
8. Jing liu, Song-zheng zhao, "Analysis and design of real estate property right registration information system for whole life cycle".
9. Zhuo-jun Hang, Hui qiang Li, "Multiple project management system of real estate enterprises".
10. Bogdan Walek, Jiří Bartoš, "Web application for searching in real estate web portals".

CLASSICAL MUSIC SUB-GENRE CLASSIFICATION USING MACHINE LEARNING

Venitto Chettiar, Dexson Dsouza, Freya Dsouza, Feni Martina Fernando, Prajyoti Dsilva

Department of Information Technology,
St. Francis Institute of Technology, Mumbai, India

ABSTRACT: Music plays an important role in the life of people. Not only does it help in reducing stress and affecting the emotions of a person but also it helps in enhancing the performance both physically and mentally. Music can be of various genres like Classical, Rock, Metal, Pop, Jazz, Blues, Hip-hop, etc. This paper focuses mainly on the Classical genre of music. Classical music has more benefits on a person than any other genre. Few of the benefits include: it makes the brain work better, reduces stress and thus helping in a person sleep better, it also helps in expressing emotions better and thus reduces pain both emotional and physical. This paper classifies the classical music into its four major sub-genres i.e. Opera, Renaissance, Orchestral and Baroque. The dataset is divided into training set and testing set. The training set is trained using algorithms like K-Nearest Neighbor, Random Forest, Support Vector Machine, Multi-Layer Perceptron, Recurrent Neural Network. The genres are classified using features like dynamic, rhythmic, tonal and spectral. By using different classifying algorithms as mentioned above, the generated results are expected to give an accuracy of approximately 80%.

Keywords: : Music, Classifiers, Machine Learning, Classical

I. Introduction

Music is an integral part of the society. Music lowers a person's stress level and improves health. It also reduces depression level in a person and strengthens the learning process and helps develop a good memory. Thus, it can be said that music makes a person happy. There are multiple genres of music like Rock, Metal, Hip-Hop, Jazz, Classical, Pop, Classical, Country, R and B, etc. Though there are numerous genres of music, it is seen that a person usually likes listening to a particular genre only.

Music player apps provide you with playlists according to a particular genre but it is usually seen that these playlists are curated according to the artist history and music similarity.

It is easier for experts to classify songs into different genres and even into but sub-genres. A layman can classify

a few songs into its genres but will find it difficult to classify it into its sub-genres. Sometimes, even experts find it difficult to classify a song into its sub-genre if that song lies in the cusp of two sub-genres or if the song is composed to belong to multiple genres.

Classifiers like k-Nearest Neighbor (kNN), Support Vector Machine (SVM), Random Forest and Multilayer Perceptron (MLP) classify the songs according to the features extracted and thus make it easier to classify a song into its sub-genre. The aim of this paper is to classify a Classical Music into its sub-genre like Renaissance, Baroque, Orchestral and Opera.

A. Renaissance

Renaissance music dates back to medieval Europe. Music historians believe that it started around 1400 and went on till around 1600, with the commencement of the Baroque music. The music of the Renaissance period was heavily influenced by the developments of that era which mainly signified the rise of the modern world. It also included recomposed works of the ancient Greece and Rome. The music of that era was about the innovations and discoveries happening during that era, increase in commercial activities and the rise of a new class- the bourgeois class. Composers like Palestrina, Lassus, Victoria and William Byrd were the pioneers of the Renaissance music. Though there are no recorded versions of the music of that era, artists in the 20th recreated and recorded much of the music of that era.

B. Baroque

Baroque refers to a generation that started around 1600 and ended around 1750, and included composers like Bach, Vivaldi and Handel, who pioneered new styles like the concerto and the sonata. Baroque has been derived from the Portuguese word *barroco*, or "oddly shaped pearl," "baroque" has been widely used since the nineteenth century to describe the period in Western European art music from about 1600 to 1750.

C. *Orchestral*

Orchestral music originated in the late 1600 stage spectacles and 1700 opera orchestras. 'Orchestra', this word came into common use in the 1800, when music of the Baroque and Classical eras often was composed for specific occasions. The success of pieces like John Adams' "A Short Ride in a Fast Machine," shows that the orchestra remains a vital medium, even in an age of high technology.

D. *Opera*

Opera music was first written around 1600. Opera uses texts that are dramatic and music to express their different stories. The word *opera* is an abbreviation of the Italian phrase *opera in musica* ("work in music"). Section II of this paper provides a gist information of the previous works in this field. Section III gives the working of the Classical Music Classification system and Section IV talks about the results generated by this system.

II. Research Methodology

Authors A. Goulart, R. Guido, C. Maciel in the paper titled "Exploring different approaches for music genre classification" have used the algorithm SVM (Support System Vector) for classification of music. It achieved accuracy up to 65.06%. Features used are entropies and fractal dimensions. The music sample used is Digital Music files. The issues faced are ambiguity and subjectivity in the classification and the dynamism of music styles (A. Goulart et al, 2012).

Authors R. Quinto, R. Atienza, and N. Tiglaio in the paper titled "Music Sub-Genre Classification Using Deep Learning" have used LSTM (Long Short Term Memory) and MLP (Multi-Layer Perceptron) algorithms for classifying music into sub-genre. They achieved a maximum accuracy of 89.824%. The feature used is Mel-Frequency Cepstral Components (MFCC) and music Sample is Jazz dataset. The issues faced by them are unsuccessful results by using MLP,SVM and KNN (R. J. M. Quinto et al,2017).

In the paper published by S. Chapaneri, R. Lopes, D. Jayaswal titled "Evaluation of Music Features for PUK Kernel based Genre Classification" have used the SVM algorithm to classify genre. It is used for the implementation of SMO (sequential minimal optimization). They have also used WEKA toolkit. They achieved an accuracy of 82%. The features used by them are 4 broad categories i.e. Dynamic, Rhythm, Tonal and Spectral. Cross validation is performed on the samples. The music sample used is Bench Mark Database. The issues faced in this work is that some genres suffer from confusion due to unclear meaning and definition(S. Chapaneri et al,2015).

In the paper titled "Statistical Evaluation of Music Information Retrieval Experiments" by A. Flexer, the algorithm is the Gaussian Mixture Model classifier. He achieved an accuracy of 93%. The feature used is Mel-Frequency Cepstral Components (MFCC). He used GTZAN dataset. The issues faced are by removing certain genres from classification deteriorates the results. Some genres are similar and thus affects the result(A. Flexer,2006).

III. Proposed Work

The classical music classification system is as shown in Fig. 1. The dataset consisted of 480 classical songs. In the 480 songs, there were 115 Baroque songs, 130 Opera songs, 105 Orchestra songs and 130 Renaissance. Each song is 30 seconds segment. Total of 37 features were extracted for each song in the dataset. For the complete dataset, a total of $37 \text{ feature} * 480 \text{ songs} = 17,760$ features were extracted. These features were stored in a CSV file.

The next step is Feature Extraction. In this phase various features like Zero Crossing rate (ZCR), Mel-Frequency Cepstral Coefficient (MFCC), Spectral Bandwidth, Spectral Rolloff, Spectral Contrast, Spectral Centroid, etc. are extracted from the dataset. These features help in classifying a song into its sub-genre as the features of a song belonging to a particular sub-genre are similar.

In classification phase the dataset was split randomly into training set and testing set in the ratio of 80:20. This ratio ensures that the system trains itself well and has enough data to produce a result for the testing. The classification is done using algorithms like k-Nearest Neighbour (kNN), Support Vector Machine (SVM), Random Forest and Multi-Layer Perceptron (MLP).

A. *K-Nearest Neighbor*

In k-Nearest Neighbour (kNN) performance was tested for different values of k ranging from 1 to 10. The KNN makes cluster of music samples with similar features. It classifies music according to the clusters made. Thus the songs that are not similar according to the features would lie as an outlier and will not be classified.

B. Random Forest Algorithm

In Random Forest the performance was tested for different no. of estimators ranging from 1 to 20. The random forest algorithm would divide the songs based on the classes Baroque, Opera, Renaissance and Orchestral. The class would depend on the feature. Then after the first iteration it will further divide the tree on other features.

C. Support Vector Machine

In SVM different kernels were used such as RBF, Polynomial and Linear. Penalty parameter of SVM was set to 100. Accuracy of RBF kernel was found out to be highest out of the three kernels.

D. Multilayer Perceptron

For MLP classification, the number of layers and hiddenUnits per layer were varied for different tests. Dropout layers are added after every perceptron layer, and 0.8 was chosen as the dropout rate. This was implemented using Keras on Tensor flow. The maximum no. of iterations were set to 400. Hidden Layer size was set to 40x40, this indicates the no. of neurons to be used for learning.

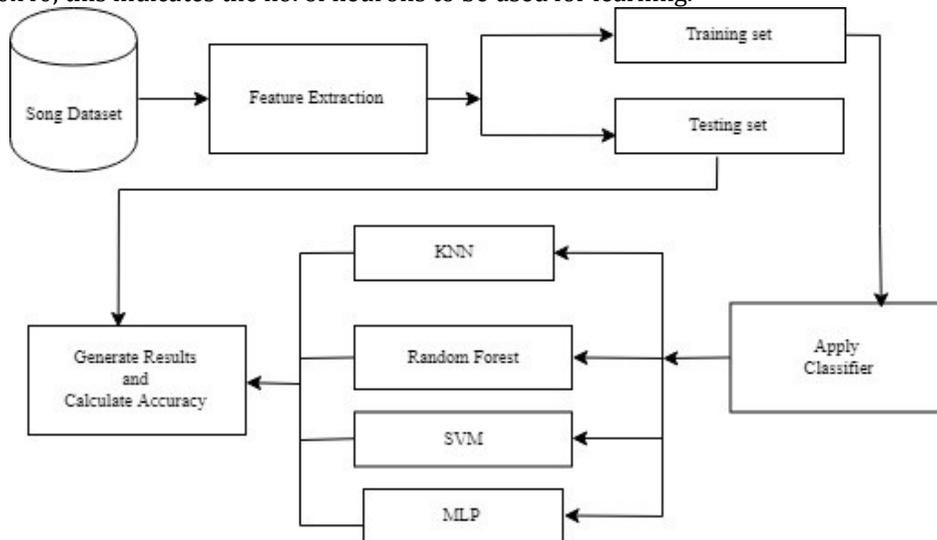


Fig. 1. Flow Diagram of the Classical Music Classification System

IV. Results And Discussion

The output of the classifier changes according to the algorithm used and also on the training and testing data. For the purpose of this paper the dataset of 480 songs is divided in a ratio of 80:20 i.e. 80% of the songs is used for training the classifier and 20% of the songs is used to test the classifier. This number is kept constant for all the algorithms. Accuracy of each algorithm was computed with respect to different parameters and parameters for the best accuracy were recorded. Performance of each algorithm along with its optimal parameters were shown in the form of a confusion matrix.

The k-Nearest Neighbor algorithm an accuracy of 86.5% is acquired. The confusion matrix is shown in Fig. 2. According to the Confusion Matrix, sub-genre Baroque songs are classified with ease followed by Opera, Renaissance and Orchestra respectively

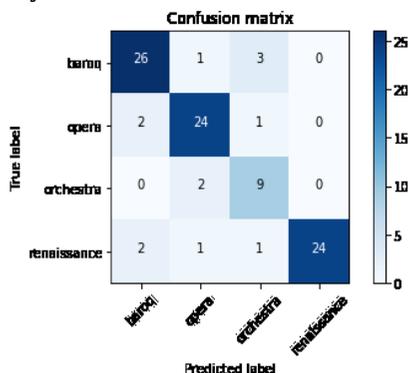


Fig. 2. Confusion matrix of the results of k-Nearest Neighbor Algorithm

The Random-Forest algorithm gives an accuracy of 85.4%. Fig. 3 shows the confusion matrix. The Confusion Matrix clearly shows that sub-genre Baroque has the most number of songs classified correctly followed by Renaissance, Opera and Orchestra.

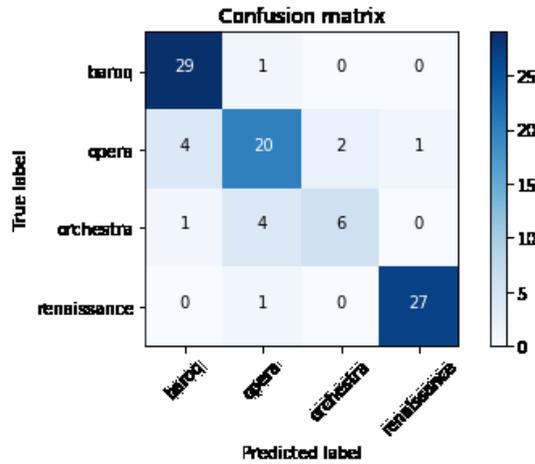


Fig. 3. Confusion Matrix for the results of Random-Forest Algorithm.

The Support Vector Machine algorithm gives an accuracy of 84.4%. The confusion matrix for SVM is shown in Fig. 4. Again, the sub-genre Baroque has the most number of songs classified correctly followed by Opera, Renaissance and Orchestra respectively.

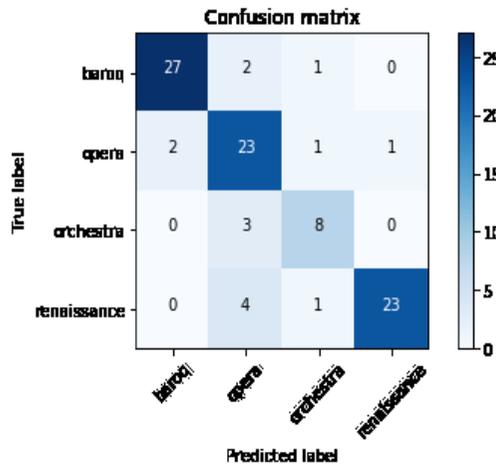


Fig. 4. Confusion matrix of the results of SVM algorithm.

The Multi-Layer Perceptron algorithm gives an accuracy of 79.2%. The confusion matrix for MLP is as shown in Fig. 5. The Confusion Matrix clearly shows that sub-genre Baroque has the most number of songs classified correctly followed by Opera, Renaissance and Orchestra.

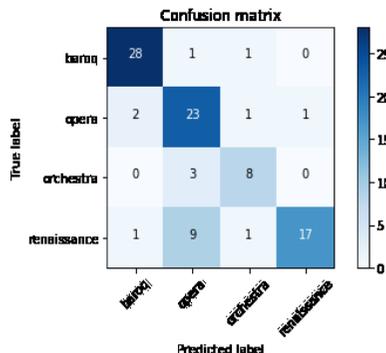


Fig. 5. Confusion Matrix for the results of MLP algorithm.

TABLE I: COMPARATIVE STUDY OF THE ALGORITHMS

Algorithm	KNN	Random Forest	Support Vector Machine	Multilayer Perceptron
Accuracy	86.5%	85.4%	84.4%	79.2%

V. Conclusion And Future Scope

From the TABLE I, it is clear that kNN gives the highest accuracy and MLP the lowest accuracy. In all the algorithms, it is seen that the songs of sub-genre Baroque is classified correctly the most number of time. Opera and Renaissance follow next and then Orchestra.

The Classical Music Classification can be further improved if additional songs are added to the dataset. Also, the system can be used to classify additional sub-genres like Avant-Garde, Modern Classical, Romantic, Chant, Choral, Minimalism, etc. The system can also be implemented for other genres like Rock, Jazz, Hip-Hop, R and B, etc.

References

1. Goulart, R. Guido, C. Maciel, "Exploring different approaches for music-genre classification", Egyptian Infomatics Journal, April, 2012. [Online]. Available:<https://www.sciencedirect.com/science/article/pii/S1110866512000151>
2. R. J. M. Quinto, R.O. Atienza, N.M.C. Tiglao, "Jazz music sub-genre using deep learning," Proc. of the 2017 IEEE Region Conference (TENCON), Malaysia, November 5-8, 2017
3. S. Chapaneri, R. Lopes, D. Jayaswal, "Evaluation of music features for PUK kernel based genre classification," ICACTA,2015. [Online].Available:<https://www.sciencedirect.com/science/article/pii/S1877050915003555>
4. Flexer, "Statistical Evaluation of music information retrieval information," Journal of new music research, June, 2006.[Online].Available:<https://www.sciencedirect.com/science/article/pii/S1877050915003555>
5. "Simple K - Nearest Neighbor Algorithm",D.umn.edu, 2018.
6. Analytics, T. model and T. Srivastava, "Tuning Random Forest model — Machine Learning — Predictive modeling", Analytics Vidhya, 2018.
7. Institute, "Building Predictive Model using SVM and R - DnI Institute", Dni-institute.in, 2018.
8. "A hypothetical example of Multilayer Perceptron Network.", ResearchGate, 2018.
9. L. Santos, "Recurrent Neural Networks · Artificial Inteligence", Leonardoaraujosan- tos.gitbooks.io, 2018

EMERGING TRENDS IN CLOUD COMPUTING: A SURVEY

¹Apurva Saxena & ²Dr.Pratima Gautam

¹Research Scholar (Computer Science Engineering), ² Dean of Computer Science and Application,
Rabindranath Tagore University, Rabindranath Tagore University, Bhopal, India,

ABSTRACT: *Cloud computing an emerging technology provides various services to the users like infrastructure, hardware, software, storage etc. For working cloud in data security, it is necessary that cloud computing network should always be free from outside attack/ threats. Here in this paper we focused on promises, issues, applications and challenges of cloud computing and we tried to propose some of the best solutions for issues generated in proving data security at various levels.*

Keywords: : *Music, Classifiers, Machine Learning, Classical*

I. INTRODUCTION

Cloud computing is a centralized controlling system in which minimal resources are offered by the providers, due to which intruder gain easy access to the resources and breach the security. Cloud computing allows the user to take benefit of the technologies. It is used for delivery of its services like-servers, storage, databases, (Joshi et.al, 2011) networking, software, analytics and many more. It is agile for the organization in improving the services provided to the user. One of the foremost characteristics of the cloud computing by which it provides flexibility to the user through the reduction of cost. Other one cloud computing services is speed which is provided to self service and on demand large amount of computing resources transfer within a minute without any pressure of capacity planning. Here cloud knew at which time how much amount of power, bandwidth and storage is required according to the geographic conditions (Ajeet Gautam, et.al, 2012). By using on-time data centers in computing, productivity gets increases. It becomes more reliable as it secures data, disaster recovery. Types of cloud services are IaaS (Infrastructure as a Service) in this customer can get infrastructure and virtual machine, networks on rent. PaaS (Platform as a Service) it is designed to make a platform for the user to develop a web or mobile apps without applying setup behind it (Apurva, et.al, 2018). SaaS (Software as a Service) it is designed to develop a software on demand for the user. IBM cloud works on all the services of the cloud. Cloud provider host and manages the application, underlying infrastructure and handle maintenance too. Cloud computing is a model for enabling omnipresent, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction (Apurva, et.al, 2019) . There are four deployment models to deploy cloud computing:

[1] **Public Cloud:** In simple terms, public cloud services are characterized as being accessible to clients from a third party service provider through the Internet. The term “public” does not always mean free, even though it can be free or fairly inexpensive to use. A public cloud does not mean that a user’s data is publically visible; public cloud vendors typically provide an access control mechanism for their users. Public clouds provide an elastic, cost effective means to arrange solutions. Examples are AWS (Amazon) cloud, Microsoft Azure cloud.

[2] **Private Cloud:** A private cloud offers many of the benefits of a public cloud computing environment, such as being elastic and service based. The difference between a private cloud and a public cloud is that in a private cloud-based service, data and processes are managed within the organization without the restrictions of network bandwidth, security exposures and legal requirements that using public cloud services might required. In addition, private cloud services offer the provider and the user greater control of the cloud infrastructure, improving security and resiliency because user access and the networks used are restricted and designated.

[3] **Community Cloud:** A community cloud is controlled and used by a group of organizations that have shared interests, such as specific security requirements or a common mission (Microsoft Azure 2019) . It works as private cloud also. The members of the community share access to the data and applications in the cloud (Zurück zum Seitenanfang 2019).

[4] **Hybrid Cloud:** A hybrid cloud is a combination of a public and private cloud that interoperates. In this model users typically outsource non-business critical information and processing to the public cloud, while keeping business-critical services and data in their control. Example Google cloud provider.

Virtualization is the new technology of the cloud computing. This technology divided physical computers into virtual devices by which they can easily manage the task. Generally cloud provider is using traditional security system to avoid authorized access of the resources. Virtualization is a key point in the cloud system that provides multiple virtual instance of a physical resource and if a single instance of a resource susceptible then connected clients are affected (Hwan-Seok Yang, 2013).

The remainder of this paper is organized as follows. Section II presents a discussion on the promises of cloud. Section III highlights the cloud computing issues and challenges. Section IV considers possible solutions of cloud computing issues and challenges. Section V defines the applications of cloud computing. Section VI concludes this paper.

II. PROMISES OF CLOUD

- A. **Speedy Elasticity:** Elasticity is defined as the ability to extent resources both up and down as needed. According to the consumer, the cloud appears to be infinite, and the consumer can purchase as per the need. This is one of the essential characteristics of cloud computing (Zurück zum Seitenanfang 2019).
- B. **Measured Service:** The cloud service are controlled and monitored by the cloud provider. This is vital for billing, access control, resource optimization, capacity planning and other tasks.
- C. **On-Demand Self-Service:** Here consumer can use cloud services as needed without any human interaction with the cloud provider.
- D. **Ubiquitous Network Access:** In this the cloud provider's capabilities are available over the network and can be accessed.
- E. **Resource Pooling:** Resource pooling allows a cloud provider to serve up its consumers via a multi-tenant model. Physical and virtual resources are assigned and reassigned according to client demand.
- F. **Cost efficiency** – The major reason behind shifting to cloud computing is that it cost effectiveness than an on-premise technology. Here Cloud offers unlimited storage space, saving money and assets of the companies.
- G. **High Speed** – Cloud computing gives you to deploy any service quickly in few seconds.
- H. **Excellent accessibility** – Storing the information in cloud allows portability. For client it is highly accessible and flexible technology (Zurück zum Seitenanfang 2019) .
- I. **Back-up and restore data** – Once the data is stored in Cloud, it is easier to get the back-up and recovery of that, in minimum time.
- J. **Manageability** – Here maintenance and management is free. Cloud provider provides the service of cloud timely; ensure seamless working of all the services (Zurück zum Seitenanfang 2019) .
- K. **Strategic edge** – Cloud computing provides any company or organization a competitive boundary over the competitors when it comes to accessing the latest and mission critical applications.

III. CLOUD COMPUTING ISSUES AND CHALLENGES

In cloud computing security is a wide topic to discuss (Microsoft Azure 2019). It is combination of technologies and policies to protect the data, services and infrastructure. This mixture is an objective of all possible attacks. There is plethora of issues and challenges on cloud computing, one most important is reliability on new technology, dependency on cloud services, selecting perfect cloud setup, hacking of brand, recovery of lost data, transparency of service provider and many more (Microsoft Azure 2019) . Some issues facing by the cloud computing are as the following:

- a. Data security
- b. Malicious Insider
- c. Denial of Service
- d. Service Hijacking
- e. Data Location

Cloud computing challenges are as follows:

1. Security
2. Managing cloud Spending

3. Lack of resources
4. Migration
5. Vendor lock-in

IV. PROPOSED SOLUTIONS FOR ISSUES AND CHALLENGES

But here we have done research and are working on some issues provided possible solutions which can be implemented to get better outcomes. On some issues we are trying to get the desired outcome (Zurück zum Seitenanfang 2019) . Furthermore new possible solutions are required in cloud computing which can protect the data in an efficient manner. In data security customer exposes their data in the cloud platform which raises the risk of hacking their data through intruders. The solution is that tokenization used in Kerberos also applies inside the cloud which added the security to the cloud. In malicious insider act damages the whole process of the cloud providers and users sensitive data. A suitable solution can be implemented by the customer itself and managed by them inside any cloud services. In denial of service an attacker tries to interrupt the service by sending inappropriate data traffic on the cloud platform and tries to affect those services. An attacker tries to interrupt the services by sending inappropriate data traffic on the cloud platform and tries to affect those services. Service Hijacking in which cloud provider user login their account using credentials. It gets hacked or stolen or miss used by attacker for deleting or stealing or altering it. MFA (Multi Factor Association) in this after activating this service in cloud provider and mange it through Google authenticator application (Zurück zum Seitenanfang 2019). After pass the login credential from any cloud provider a user will get a code generated on this application after applying this code only the account will be open. Last phase is also an application is used for password management for user inside cloud provider to secure the service from the hacker. In data location data can be upload and retrieve data anytime from anywhere on the web. SSL (Secure Socket Layer) can securely transfer the data within the cloud provider and between multi clouds provider (Zurück zum Seitenanfang 2019).

Some of the challenges are generated in cloud computing such as security is on higher risk when traffic gets increase from malicious attacker through DDoS attack which target the network. Cloud flare provides better service on internet for security. When users get increased or traffic becomes higher than cost increases, unable to govern the services, performance and security get affected. There are many software companies such as cloud health and many more which provides cloud computing services related to cost management, governance, automation, performance and security (Salam Ismaeel, et.al 2016) . For lack of resources staff may need to acquire new skills and create or update foundation process. Only one solution is to higher the skilled workers to make the resources acceptable. Because of migration some problem arises in terms of data security. SMS (Server Migration Service) is an agent less service which makes easier and faster for user to migrate on-premises workloads to any cloud providers. Vendor lock-in many cloud providers have many facilities and flexibility to attract the customer. The image of an instance which is launch on cloud provider can be downloaded and import that image from that cloud provider and export to other cloud provider. Here in table1 it is tried to show the combine explanation of cloud issues and challenges with their appropriate solutions using cloud services by the clients to get proper outcome.

S.No.	ISSUES	CHALLENGES	SOLUTIONS
1.	Security	<ul style="list-style-type: none"> • Storage and computing • Virtualization • Internet and services • Network problem • Access 	<ul style="list-style-type: none"> ✓ Cryptography ✓ Monitoring Virtual machines ✓ Authentication Process ✓ Multi Factor Authentication (MFA)
2.	Unwanted Access	<ul style="list-style-type: none"> • Malicious Attacker • Provider Chains • Data Theft 	<ul style="list-style-type: none"> ✓ Authorized User ✓ Security level agreement(SLA)
3.	Lack of Resources	<ul style="list-style-type: none"> • Increases business agility 	<ul style="list-style-type: none"> ✓ Higher skilled workers ✓ Provide training

		<ul style="list-style-type: none"> Flexibility Innovative skills 	
4.	Migration	<ul style="list-style-type: none"> Inter cloud migration On-premise to cloud 	<ul style="list-style-type: none"> ✓ EBS(Elastic Block Storage) ✓ SMS(Server Migration Service)
5.	Vendor Lock-in	<ul style="list-style-type: none"> No Communication[21] Difficult to migrate Export of data 	<ul style="list-style-type: none"> ✓ Cloud provider service can be use ✓ Image can be made through instance ✓ Use of image data can import from one cloud to another

Table1: Cloud Computing issues, challenges with these solutions.

V. APPLICATIONS OF CLOUD COMPUTING

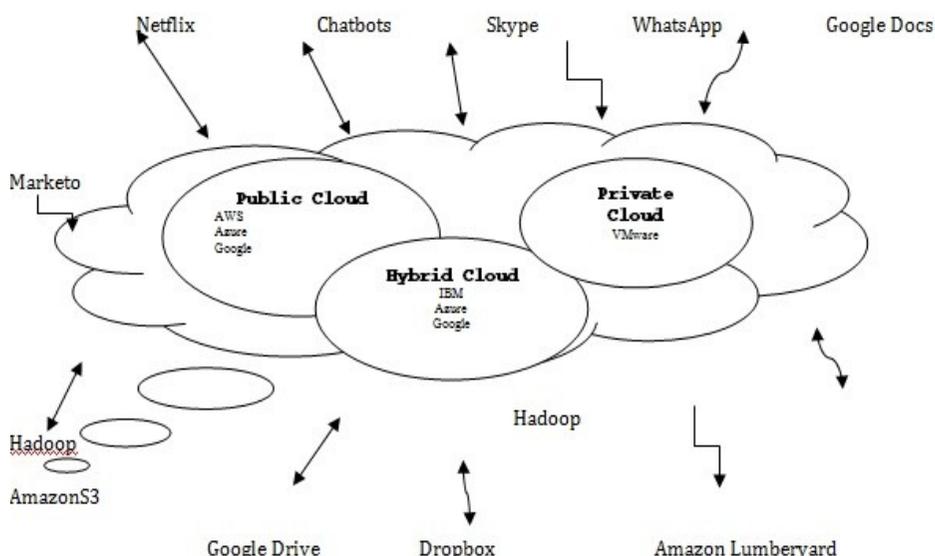


Figure1: Different types of cloud model with cloud providers and its applications.

Here in this figure1 it is tried to show the club of different types of cloud models such as public cloud which consist of different cloud providers like Amazon AWS, Microsoft Azure and GCP (Google Cloud provider), private has VMware cloud provider and Hybrid cloud consist of different cloud provider which uses these cloud for their self and open for public use also like IBM, AWS and Azure (Zurück zum Seitenanfang 2019) . If we use machine learning with cloud computing then there will be more applications (Anubha Dubey, 2016). The cloud computing has many applications in different services of cloud like it offers:-

- It provides scalable resources. Due to its on-demand streaming service, it faces server load at peak times. The move to migrate from in-house data centers to cloud without having to spend in setup and maintenance of costly infrastructure.
- Chatbots can provide customize solution, messages through cloud provider. Siri, Alexa and Google Assistant are all cloud-based natural-language intelligent bots (Zurück zum Seitenanfang 2019) .
- For communication cloud uses different application like skype, whatsapp. In regard of productivity concern, office tools like Microsoft Office 365 are used.
- For business management applications like customer relationship management (CRM) and enterprise resource planning (ERP) are based on a cloud service provider. Salesforce, Hubspot, Marketo are all applications of cloud.
- In Social Media the most popular and overlooked application of cloud computing are Facebook, LinkedIn, MySpace, Twitter, and many other sites use cloud computing.

- Cloud computing enables data scientists to beat into any organizational data to analyze it for patterns and insights, find forecast future crisis and help in data backed decision making. Hadoop, Cassandra, is one of the tools used in cloud computing (Zurück zum Seitenanfang 2019).
- The cloud can provide an environment to cut expenses and launch your apps in the market quicker. LoadStorm and Blaze Meter are popular testing tools.
- The cloud also provides more flexibility in the sense that client can enjoy large storage and on-demand backups. Recovery is also performed faster in the cloud because the data is stored over a network of physical servers. Drop box, Google Drive and Amazon S3 are trendy examples of cloud backup solutions.
- This can speed up and simplify the development process. Amazon Lumberyard is a popular mobile game development tool used in the cloud.

VI. CONCLUSION

Cloud computing provides easy computing and access to high performance computing, networking, storage and infrastructure through internet. Cloud computing have potential to provide high efficiency and cost savings. As cloud computing is the development trend in the future which provides infinite computing and capability, but still security, storage and privacy is a big challenge for cloud computing. So, in this paper we have tried to show the various benefits of cloud computing with all its security features. But in near future we will work on artificial intelligence, machine learning, and internet of things for providing more security to the available data. So that we can increase the efficiency and accuracy of the cloud features.

VI. ACKNOWLEDGEMENT

The author's would like to thank Dr.Anubha Dubey (Independent Researcher and analyst, Bioinformatics, Bhopal) for her expert advice and encouragement for this manuscript.

REFERENCES

1. Ajeet Kumar Gautam, et.al, 2012 An Improved Hybrid Intrusion Detection System in Cloud Computing. International Journal of Computer Applications (0975 – 8887) Volume 53– No.6, September.
2. Al Awadhi, E,et.al, 2013 Assessing the security of the cloud environment. 17-20 Nov. IEEE Conference, Page(s) 251 – 256.
3. Anubha Dubey, 2016 Applications of Machine Learning: Cutting Edge Technology in HIV Diagnosis, Treatment and Further Research, Computational Molecular Biology, 6 (3), 1-6.
4. Apurva Saxena, et.al, 2018 Honeypot based new algorithm for data security, Journal of Engineering Research and Application Vol.8, Dec.
5. Apurva Saxena, et.al, 2019 Kerberos based Data security in Research & Production Honey Pot. EPH - International Journal of Science and Engineering Volume-5, Issue-1, Jan.
6. Dhamija, et.al, 2000 *A user study using images for authentication*. in Proc. 9th USINEX Security Symp., Denver, CO, Aug., pp. 45– 58.
7. G. E. Blonder, 1996 Graphical password. U.S. Patent 5 559 961, Sep. 24.
8. Gagan,et.al, 2016 Dynamic Cluster based Privacy-Preserving Multi-Keyword Search over Encrypted Cloud Data IEEE 6th International Conference - Cloud System and Big Data Engineering (Confluence)
9. Hwan-Seok Yang, 2013 A study on attack information collection using virtualization technology. 74:8791–8799 DOI 10.1007/s11042-013-1487-8, Springer Science + Business Media New York.
10. Joshi Ashay Mukundrao, et.al, 2011 Enhancing Security in Cloud Computing Information and Knowledge Management. ISSN 2224-5758 (Paper) ISSN 2224-896X,Vol 1, No.1.
11. Kumar Shridhar, et.al, 2014 A Prevention of DDos Attacks in Cloud Using Honeypot. International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 Impact Factor (2012): 3.358 Volume 3 Issue11, November .
12. Liangxuan Zhang, et.al, 2017 Privacy-Preserving Attribute-Based Encryption Supporting Expressive Access Structures. IEEE Second International Conference on Data Science in Cyberspace.
13. Microsoft Azure 2019 <https://azure.microsoft.com/en-in/overview/what-is-cloud-computing/>
14. Muhammet Baykara, et.al, 2015 A Survey on Potential Applications of Honeypot Technology in Intrusion Detection Systems. International Journal of Computer Networks and Applications (IJCNA) Volume 2, Issue 5, September – October .
15. Paul. A.J, et.al, 2007 A Fast and Secure Encryption Algorithm For Message Communication. IET-UK International Conference on Information and Communication Technology in Electrical Sciences (ICTES 2007), Dr. M.G.R. University, Chennai, Tamil Nadu, India. pp. 629-634.
16. Salam Ismaeel,et.al 2016 Cloud Security: Basic Principles at research gate publication
17. Sultan Aldossary, et.al, 2016 Data Security, Privacy, Availability and Integrity in Cloud Computing: Issues and Current Solutions. International Journal of Advanced Computer Science and Applications, Vol. 7, No. 4.

18. X. Suo, et.al, 2005 Graphical passwords: A survey. in Proc. 21stAnnu.Comput. Security Appl.Conf., Dec. 5-9, , pp. 463-472.
19. Yunfei CI, et.al, 2017 Design and Implementation of the Components of the Symmetric Cryptographic Algorithm. IEEE Second International Conference on Data Science in Cyberspace.
20. Zurück zum Seitenanfang 2019, Amazon Web Services, Inc. oder Tochterfirmen. Alle Rechte vorbehalten.

Online Resources:-

- <https://www.monitis.com/blog/3-types-of-cloud-computing-services>
- http://www.cloudcouncil.org/Cloud_Computing_Use_Cases_Whitepaper-4_0.pdf

Water Quality Monitoring & Notification System Using IoT

¹Junaid D Momin, ²Raheen N Shaikh, ³Arkam Ansari, ⁴Mubashir Khan

¹B.E Student, ²B.E Student, ³B.E Student, ⁴Assistant Professor

Department of Computer Engineering,

Anjuman-I-Islam Kalsekar Technical Campus, Navi Mumbai, India

ABSTRACT: In overall monitoring of water, standard takes a lot of time and it appears to be a little complex as it undergoes several laboratory testing, which results in time-consuming and for this a real time monitoring of water goodness by using IoT has been proposed. IoT together with sensors is quite effective. Some other technologies had certain drawbacks which it couldn't reveal it. So to remove this short coming and providing an efficient and socio-economic solution has been the main focus of this project. Here we are using sensors which will monitor water on the bases of its level, conductivity, ph value, flow control, and the most important to detect if any presence of micro-organisms is found. The accessed data are controlled by the use of the micro-controller. By using IoT, the information can be collected and the water pollution can be inquired to the public or the person in charge of it. This can make atmosphere to adapt the good quality if water.

Keywords: : IoT, Sensors, socio-economic, ph-value, micro-organisms, bacteriophage.

I. Introduction

Water is the important source of fuel for all life that lives on earth without which no lives can survive on this planet. But due to major issues related to water that are arising nowadays impacts human life. Majority of which is an overflow of water, over storage of water which leads to contamination, leakage in the pipeline which results in impure water and many more.

It is a trendy work to check water quality regularly using agile technologies and in fact, we common people cannot detect that. Mostly it is found that water tanks are controlled manually to check the flow of water. If the person is found absent, the water keeps on flowing despite the loss of water. So this problem can be controlled by using solenoid valve which can be controlled by the user from anywhere on getting the indication of water level through the ultra-sonic sensor.

This wireless sensor technologies are been developed for data acquisition, building a better environment, to enhance the livelihood, to improve one's better life. In other projects, they have used sensors which controls the flow of water or either detects the temperature of water or maybe to check the level of chlorine present in it.

In this project, the user can not only detect the flow of the water but can also detect the quality of water based on its ph value, present of any micro-organisms. On getting such indication the user can decide to stop the flow of water. Because the water quality that is present in the tank tend to be contaminated with either of these contaminations.

Ensuring the safety of water is a challenge due to the excessive sources of pollutants, most of which are man-made. Water quality is affected by both point non-point sources of pollutants, which include entry of soil through leak in pipeline, over stored of water which leads to rise in micro-organisms, over rate of salinity in water, etc. Poor quality of water spreads disease; causes death hampers scion economic progress.

Water quality monitoring is defined as the collection of information at set locations at regular interval of time in order to provide data which may be used to define current condition of water. Main objective of water quality monitoring include measurement of critical water quality parameters such as microbial, physical chemical properties provide early warning identification of hazards.

II. APPROACH TO WATER QUALITY:

Design is the very first step in the development phase for any techniques and principles for the purpose of defining advice, a process or a system insufficient detail to permit its physical realization. Once the software requirements have been analyzed and specified the software design, coding, implementation and testing that are required to build and verify the software the crop.

A use case model is basically something use for business analyzing and for presentation of the steps that mostly defines the interactions between a user and a system. It helps in understanding the detail of the

interactions and also helps to set the expectations of how the user will work with the system. For the proposed system the working and the flow of use case is as follows:

Working : The sensors which are used in this system will monitor the parameters of water. This parameters are then calculated and compared. After the analyzing the value, a report is generated based on which a notification to the user is send. Arduino- with overall water monitoring mechanism with specific regard of monitoring the quality of water as seen by primary and secondary actors.

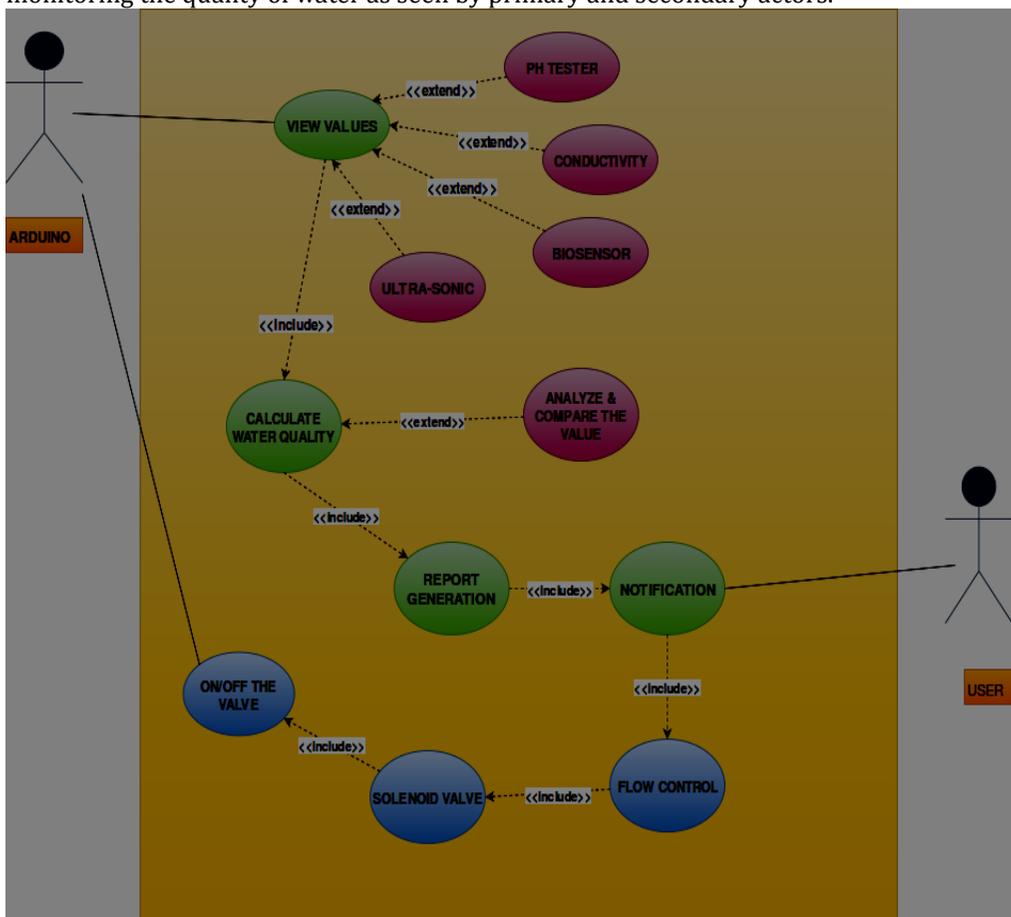


Figure 1: Use Case diagram for Water Quality Monitoring & Notification System using IoT.

Success scenario:

1. The value obtained by different sensors which are induced inside the water and calculated.
2. These values are analyzed and compared with the values obtained.
3. A report is generated on the basis of comparison of the value.
4. If the value lies above threshold value a notification is send to the user.
5. The user can decide the flow control of the water based on the report obtained.
6. This control is made under solenoid valve, which can be operated to control the flow of water by switching into ON/OFF.
7. This control data is traced back to Arduino.

III. SENSORS:

In the current system, it acquires the working based on different types of sensors for fetching the quality of the water. Sensors such as Ultra-sonic, Conductivity, pH tester, Bio-sensor and Solenoid valve, will be used to sense the quality of the water and provide the reading to the micro-controller. Micro-controller that we shall use is Arduino, which will upload the data on the cloud as well as check the quality of water with the help of sensors connected to it.

3.1 Ultra-Sonic Sensor:

Ultra sonic sensor is the device which measures the distance between the two objects or any material or substances. Ultra-sonic sensor in water reads the level of water in the tank. When this sensor is

placed in water it starts generating ultra-sonic waves which bombards or touches the surfaces of water present in tank which result in measuring the level of water present in tank as compared to the total height or the capacity of the tank.

This sensor consists of speaker and the mike which is transmit the wave to measure the distance. This results in monitoring of level of water in the tank and since there is no direct contact of the sensor with water.

3.2 Conductivity sensor

The word conductivity defines the ionic strength present in the solution. More the concentration of ions higher is the conductivity and if less is the ions lower is the conductivity of the solution. After all, conductivity is directly linked to the level of salinity present in any solution. Conductivity of water can be marked with this sensor since more the salinity higher are the risk of water getting turbid. So to avoid this, conductivity has the probe that test the level of ions present in the water. This can preserve the normal water getting contaminated by salty water if it found to be saline and thus, can be avoided by the user ob being getting a notification.

3.3 pH Tester Sensor

pH tester tests the acidic & basic nature of any solution. It has the pH value ranging from 0-14, which is further divided into 3 sub-ranges, mainly 0-6 is the pH value for acidic nature, 8-14 is the pH value for alkaline nature and 6 is the pH value for neutral nature which means the solution is pure i.e. Water.

Hence, pH tester in water can detect the pH of water so as to check its purity. Higher the pH value, higher are the chance of solubility of any substances or elements which can make the water toxic and can result in fatal for health. This tester can solve the acidic and basic purity present in water.

3.4 Detection of Bacterias:

Bio-sensor is the device which works on biology symptoms. This sensor can detect the presence of any microorganisms in water. Mostly the water borne organisms are not visible to naked eyes and hence, they easily enter into body and causes many harmful diseased to the living being. So, this sensor when placed inside the water it can detect the presence or the rise of any tiny microorganisms, which grows or are found when water is stored for long time and is not been used and this becomes a situation where water borne organisms grow.

With the help of bio-sensor one can detect and avoid getting infected by contaminated water. Mostly microorganisms are found into two categories, pathogenic and non-pathogenic organisms, i.e. harmful and non-harmful organisms. So this sensor can detect the presence of any such organisms and pass the data on cloud with which a user can get notified for the same and decide. After being getting the notification a user can decide the flow of water by controlling the solenoid valve depending on the purity of water.

IV. ACTUATORS :

Solenoid valve is one kind of sensor which can control the flow of water depending on user. When the result from ultrasonic sensor is retrieve and stored and when the user is informed about the level of water in the tank, the user can control its flow by just controlling the solenoid valve. These result in reduction of wastage of water.

There is no requirement of any other person to operate the motor manually or it is not necessary for the person to remain present at the time of filling the tank. Since it is mostly found that the person remains absent when the tanks seems to overflow and there results in wastage of water and hence, this can reduce the chance of wasting the water.

V. MONITORING AND CONTROL BLOCK :

This system will work with using micro-controller (Arduino-UNO) along with different sensors which are pH tester, Conductivity, Bio-sensor, Ultra-sonic, Solenoid valve and also a wifi module to connect the micro-controller to the cloud.

Wifi module (ESP8266) is the complete wifi network which connect everything as serving wifi adapter, wireless based design on its simple connectivity. The process used to send or receive data online or regularly.

All the Parameter sensors are connected to the Arduino Board which monitors the value of parameters placed inside the water. After getting the values, they are analyzed and compared with the threshold values. This generates the report and this report with the help of wifi-module will be stored on

cloud and a notification will be delivered to the user.

VI. STORAGE OF DATA :

Cloud computing is the most widely used internet-based computing service in which huge number of groups of remote servers are networked to allow centralized data storage and so one can easily access data from online to access computer services or resources. Cloud Service Models are well designed in Infrastructure as a Service (IaaS), Software as a Service (SaaS) and Platform as a Service (PaaS). With the help of Infrastructure as a Service (IaaS), we can virtualize the computing resources. Software as a Service (SaaS) is a mechanism for delivering software that provides remote access to software as a web-based service. Platform as a Service (PaaS) allocates a platform and an environment which allows the developers to build applications and services through the internet.

VI. COMMUNICATION :

There are Arduino shields, which can fix on board. We can use them externally also through ESP8266, GSM module and Sigsbee. Sigsbee devices can transmit data over long distances by passing data through a mesh network of intermediate devices to reach more distant ones. Signee has a defined rate of 250 kbit/s, best suited for intermittent data transmissions from a sensor or input device.

GSM modules are used to send a message or miscall over a network. It's a transceiver that can be added to any existing microcontroller based setups via UART (serial link) to enable the system to communicate over the Internet via Wi-Fi. Commands will be sent from an android application through the internet and to Wi-Fi network to ESP8266 with Arduino Mega. For data storing we can save it on Google drive.

In the proposed system, it consist of an micro-controller that is Arduino which is connected with different kinds of sensors such as pH-tester, conductivity, bio-sensor, etc, which are placed or induced inside the water to monitor its quality. This system uses this commercial hardware to allow the acquisition of data.

After the data acquisition, with a defined sampling period this data is calculated and then analyzed and compared with the stored data. When the comparison is completed it generates the reports based on the historical acquired data. This report is then forwarded to the user as a notification.

Based on the notification and the report of water quality obtained, the user can decide to control the flow of the water, if the value obtained is beyond the threshold value. This flow control of water is made with the help of solenoid valve which is a sensor which controls the flow of water .

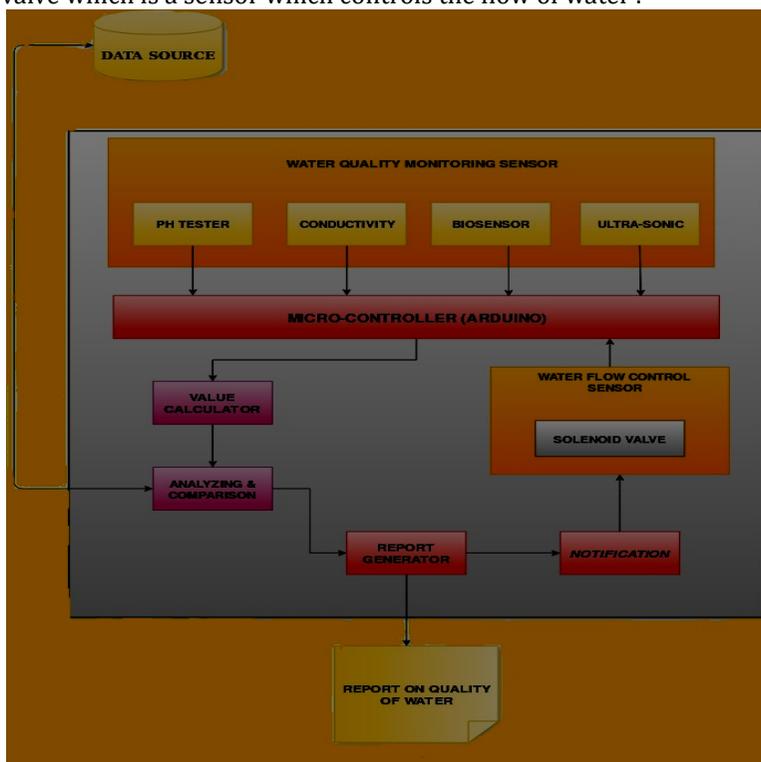


Figure 2 System Architecture of Water Quality Monitoring & Notification System using IoT.

VII. CONCLUSION :

This system is based on real-time water quality analyzing which has low cost maintenance and easy to use. With the help of sensors many harmful effects to society can be avoided and this can be well organized if it is governed properly, through which people can be notified or alarmed for the water quality they shall use which can prevent them from getting infected by various harmful chemicals, metals, pollutants and bacteria.

Since it is based on IoT, it has wide range of uses and application which can be taken into demonstration which can result in change in human lifestyle in various aspects, one of them is this system which tries to overcome the problem that arises through water by detecting the quality of water and informing the same to the user. On prior notification a user can decide whether the water must be taken in to use or not depending on the quality it has been mentioned. For this system to be operated a user need not have to be well trained or instructed. It is easy to use and operate.

assumption of the APT is that there should not be arbitrage in the market and the investors can earn only normal profit. Jarquebera test is used to test the normality of data.

References

1. SayaliWadekar, VinayakVakare, RamratanPrajapati, ShivamYadav, VijaypalYadavSmart Water Management Using IOT, 2016.
2. Tha. Sugapriyaa, S. Rakshaya, K. Ramyadevi, M. Ramya,P.G. Rashmi SMART WATER QUALITY MONITORING SYSTEM FOR REAL TIME APPLICATIONS 2017.
3. A.N.Prasad, K. A. Mamun, F. R. Islam, H. Haqva Smart Water Quality Monitoring System 2015

DUAL-PHASE-LAG FRACTIONAL THERMOELASTIC INVESTIGATION OF A SPHERICAL CAVITY WITH FINITE SPEED OF THERMAL WAVE PROPAGATION

Gaurav Mittal¹, Vinayak. S. Kulkarni^{2*}

Assistant Professor¹, Professor^{2*}

¹Department of Mathematics, Thadomal Shahani Engineering College, Mumbai-400050, Maharashtra, India.

e-mail: gmsre.1979@gmail.com

^{2*}Department of Mathematics, University of Mumbai, Mumbai-400098, Maharashtra, India.

ABSTRACT: A fractional thermoelastic dual-phase-lag model associated with one relaxation time is presented in this article. This model is prepared for a perfectly thermoelastic spherical cavity of finite annular radius whose inner and outer surfaces are assumed to be traction free. The dual-phase-lag heat conduction model of Tzou (1995) that contains the temperature gradient, thermal displacement and two different translations of heat flux vector are used to formulate the heat conduction model. Due to the involvement of relaxation time $\tau_s > 0$ and delay time translations $\tau_Q > \tau_T > 0$, the finite speed of thermal wave propagation has been achieved and the microstructural interactions and corresponding thermal variations have been studied. The analytical results for distribution of temperature, displacement and thermal stresses are obtained in the Laplace domain. These results are inverted in the time domain by the Gaver-Stehfest algorithm. The convergence of infinite series solutions has been discussed by Kuznetsov convergence theorem. The numerical results obtained are compared with coupled and generalized thermoelasticity theories.

Index Terms- Spherical Cavity, Dual-Phase-Lag, Fractional Thermoelasticity, Relaxation Time, Delay-Time Translations.

AMS Classification- 74F05; 34K28; 34N99; 34Q74; 35Q79; 35R11

I. INTRODUCTION

Categorization of various conducting materials as per their ability to conduct the heat energy subjected to high-temperature environments give rise to an important problem of material science and engineering. From last two decades, fractional calculus has been used successfully to reinvestigate the existing models of classical and generalized thermoelasticity and attains its popularity among research community to categorize the materials with the help of fractional order applied to the classical thermoelasticity theories. Povstenko (2005) successfully replaced integer orders derivatives with respect to time variable by Caputo (1967) time fractional derivative in classical Fourier law of heat conduction and fractional thermoelasticity theory was initiated. Sherief and Latief (2010,2013) further deduced the limiting cases for weak and strong conductivity cases of fractional thermoelasticity theory and applied these limitations to a one-dimensional thermoelastic fractional heat conduction problem. Youssef (2010) designed another thermoelastic model with suitable modifications using fractional theory and supports the uniqueness theorem proposed by Sherief et al. (2010). Validation of theoretical results was made through the solution of a fractional thermoelastic problem for half-space subjected to ramp-type heat flux where variations of change in temperature and thermal stresses were computed numerically and compared to classical and generalized thermoelasticity theories. Youssef and Elsibai (2011) designed a new mathematical model of fractional thermoelasticity based on the non-Fourier and coupling effect between the temperature and strain.

The generalized solution for thermoelastic vibration of a bounded nano-beam resonator induced by ramp-type heating was developed as an application to material science. Hamza et al. (2014), (2016) derived the fractional theory of thermoelasticity associated with two relaxation times and proved the uniqueness theorem for same, later designed one-dimensional fractional thermoelastic heat conduction problems in Cartesian and Spherical domain to support the theory derived. The results for temperature change, thermal displacement and thermal stresses were calculated using state space approach and Laplace transforms and then compared with earlier developed classical models of thermoelasticity for both models. Abbas (2015) considered a thermoelastic infinite body with a spherical cavity in the context of the theory of fractional thermoelasticity assuming the corresponding inner surfaces are traction free and subjected to thermal shock. Analytical solutions were iterated numerically and plotted graphically, compared with various models. Waleed (2016) applied the fractional order theory of thermoelasticity to a one-dimensional

problem of distribution of thermal stresses and temperature in thermoelastic spherical shell subjected to sudden change in the temperature of its external boundary, numerical results were obtained and shown graphically.

Sherief and Hussein (2017) obtained the fundamental solution of thermoelasticity with two relaxation times for an infinite space which is acted upon by a spherically symmetric instantaneous point source of heat. The operational method and the Laplace transform technique are used to derive the solution in the form of a series of functions. Sherief and Raslan (2017) studied the effect of an axisymmetric temperature distribution when a solid sphere composed of a thermoviscoelastic material is subjected to a distributed heat source that acts inside its inner sphere. This problem was solved analytically by Laplace transform technique and corresponding results were depicted graphically.

Sherief and Allam (2017) solved a two-dimensional problem for a sphere in the theory of thermoelasticity without energy dissipation. The spherical surface was taken to be traction free and subjected to an axisymmetric temperature distribution that is harmonic in time. Numerical results were represented graphically and discussed. Ezzat and Bary (2018) applied fractional magneto-thermoelasticity to one-dimensional problems of a thermoelectric spherical shell subjected to an arbitrary thermal loading of its external boundary in the presence of a uniform magnetic field. Some comparisons were shown in the figures to estimate the effects of the fractional order and ramping time parameters. Yong and Peng (2017) considered a spherical cavity in an infinitely extendable thermoelastic medium, subjected to thermal stress loading was investigated for change in temperature, thermal displacement and associated thermal stresses. The governing equations for the problem were formulated and then solved by Laplace transform together with its numerical inversion. The non-dimensional results were obtained and illustrated graphically. Hendy et al. (2018) constructed a new mathematical model of magneto electro-thermoelasticity in the context of a new consideration of fractional Green-Naghdi heat conduction law without energy dissipation. The governing coupled equations are applied to one-dimensional problem for a perfectly conducting spherical cavity subjected to an arbitrary thermal shock in the presence of a uniform magnetic field.

Theory of fractional thermoelasticity associated with relaxation times and phase-lags is not limited to theoretical developments however due to various applicable features the scientists and engineers are applying it to various engineering problems like to design heat exchangers, to demagnetize ferromagnetic metallic thin films (2000) and to synthesize the metals and semiconductor nanomaterials (2016) boiler tubes etc., where temperature induced elastic deformation occurs. Following the text books and research articles given in the reference section, one can explore that the theory discussed in the manuscript can also be applicable to controlling of dynamical systems, hyperbolic theories in biology like radio-frequency heating of human tissue and flash burn of the human skin. The presented theory can also be used by chemical engineers to observe the fundamental process of chemical reaction dynamics such as bond formation and break.

Following the recent literature, it can be found that so far very few research work is carried out in fractional thermoelasticity in finite thermoelastic medium especially in case of thermoelastic bodies having a spherical cavity. In the present article, an attempt is made to investigate the distribution of temperature, thermal displacement and associated thermal stresses, when suddenly an external heat source is applied to the outer boundary of the perfectly thermoelastic solid body having a spherical cavity whose surfaces are traction free. In this article, a dual-phase-lag problem is formulated in the context of the fractional order generalized theory of thermoelasticity associated with one relaxation time. The analytical results are obtained in the Laplace domain and corresponding inversions has been carried out using Gaver-Stehfest algorithm (1970a, 1970b, 1966).

The numerically computed results are shown graphically for a particular value of time variable. Some comparisons are made describing the limitations of generalized and classical theories of thermoelasticity for weak and strong conductivity cases. To our knowledge, no one has developed a dual-phase-lag model for the fractional thermoelastic problem for a solid body having a spherical cavity in a finite domain associated with relaxation time. This is the new and novel contribution to this field.

II. THE MATHEMATICAL MODEL

2.1 The Fractional Dual-Phase-Lag Heat Conduction Equation

Application of classical Fourier's law of heat conduction to thermal models leads to the parabolic differential equation for the temperature field. In order to achieve the finite speed of thermal wave Maxwell-Cattaneo modify Fourier's law as

$$\left(1 + \tau \frac{\partial}{\partial t}\right) \vec{Q} = -\kappa \Delta T, \tag{1}$$

where \vec{Q} is the thermal flux vector, ΔT is temperature gradient, κ denotes the thermal conductivity and $\tau > 0$ represents relaxation time.

Later, Sherief et al. (2010) modified the theory of generalized thermoelasticity in view of definition of Caputo fractional derivative (1967) and reconstructed the Maxwell-Cattaneo law of heat conduction as

$$\left(1 + \tau \frac{\partial^\alpha}{\partial t^\alpha}\right) Q_i = -\kappa_{ij} T_{,j} \quad 0 < \alpha \leq 1, \tag{2}$$

where Caputo time fractional derivative of order $\alpha \in (0,1]$ of an absolutely continuous function $\Psi(t)$ is given by

$$\frac{d^\alpha}{dt^\alpha} \Psi(t) = I^{1-\alpha} \left\{ \frac{d}{dt} [\Psi(t)] \right\}, \tag{3}$$

where I^ξ is Riemann-Liouville fractional integral of the function $\Psi(t)$ of order ξ which is given as

$$I^\xi \Psi(t) = \int_0^t \frac{(t-s)^{\xi-1}}{\Gamma(\xi)} \Psi(s) ds, \tag{4}$$

where $\Psi(t)$ is a Lebesgue integrable function and $\xi > 0$.

In order to study the microstructural interactions within solid heat conductor at microscopic level Tzou (1995) considers the non-Fourier effect of heat conduction and proposes a dual-phase-lag model in the following form

$$\vec{Q}(P, t + \tau_Q) = -\kappa \Delta T(P, t + \tau_T), \tag{5}$$

where the translations τ_Q, τ_T are the phase-lags in heat flux vector and temperature gradient respectively. The translations τ_Q, τ_T are responsible to capture the thermal wave behaviour and effect of phonon-electron interactions respectively. These lags are small positive quantities and well known as intrinsic properties of the medium.

Applying divergence operator on both the sides of equation (5), one will have the following vector equality

$$\Delta \cdot \vec{Q}(P, t + \tau_Q) = -\kappa \Delta^2 T(P, t + \tau_T). \tag{6}$$

Following Mittal and Kulkarni (2018) and applying Taylor’s series expansion of time fractional order α to both the sides of equation (6) and after retaining terms up to 2α order in τ_T and τ_Q one has governing time fractional heat conduction equation in two different translations of heat flux vector τ_Q and temperature gradient τ_T as

$$\left\{ \frac{\partial}{\partial t} + \frac{\tau_Q^\alpha}{\alpha!} \frac{\partial^{\alpha+1}}{\partial t^{\alpha+1}} + \frac{\tau_Q^{2\alpha}}{(2\alpha)!} \frac{\partial^{2\alpha+1}}{\partial t^{2\alpha+1}} \right\} (\rho c_p T + \gamma T_0 e) = \left\{ 1 + \frac{\tau_T^\alpha}{\alpha!} \frac{\partial^\alpha}{\partial t^\alpha} + \frac{\tau_T^{2\alpha}}{(2\alpha)!} \frac{\partial^{2\alpha}}{\partial t^{2\alpha}} \right\} \kappa \Delta^2 T, \alpha \in (0,1], \tag{7}$$

where T is the absolute temperature and T_0 is the reference temperature of the medium such that, $\left| \frac{T-T_0}{T_0} \right| \ll 1$, and $\alpha \in (0,1]$ is fractional order applied to the time variable partial derivatives and Δ^2 is the Laplacian operator given as

$$\Delta^2 \equiv \left(\frac{\partial^2}{\partial r^2} + \frac{2}{r} \right) = \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial}{\partial r} \right), \tag{8}$$

the positive constants c_p, ρ are the specific heat capacity and density respectively. The quantity γ is given by $\gamma = \alpha_t (3\lambda + 2\mu)$, where α_t is the coefficient of linear thermal expansion of the material and λ and μ are Lamè constants.

2.2 Mathematical Formulation

A solid heat conductor having spherical cavity of finite annular radius $a_1 \leq r \leq a_2$ in an isotropic thermoelastic medium occupying the region $S \subset \mathbb{R}^3$ is defined by

$$S = \{(r, \theta, \phi) \in \mathbb{R}^3 | a_1 \leq r \leq a_2, 0 \leq \theta \leq 2\pi, 0 \leq \phi \leq \pi\} \text{ where } 0 < a_1 < a_2. \tag{9}$$

The initial temperature of spherical cavity is zero. The inner and outer spherical surfaces of the cavity are traction free. The inner and outer cavity surfaces are subjected to time dependent temperature $\Psi(r, t)$, $\forall t > 0$ which is described by a continuous integrable function $\Psi: D \times (0, \infty) \rightarrow \mathbb{R}$. Under these more realistic boundary conditions, the thermal variations in the spherical cavity are required to be determined.

Following Hamza et al. (2016) because of spherical symmetry, the radial displacement component is given by

$$u_r = u(r, t). \tag{10}$$

The strain tensor has the following components

$$e_{rr} = \frac{\partial u}{\partial r}, \tag{11}$$

$$e_{\theta\theta} = e_{\phi\phi} = \frac{u}{r}, \tag{12}$$

$$e_{\theta\phi} = e_{r\phi} = e_{r\theta} = 0. \tag{13}$$

The cubical dilatation e is given by

$$e = \frac{\partial u}{\partial r} + 2 \left(\frac{u}{r} \right) = \frac{1}{r^2} \frac{\partial}{\partial r} (r^2 u). \tag{14}$$

The equation of motion in the absence of body forces is given by

$$\rho \frac{\partial^2 u}{\partial t^2} = (\lambda + 2\mu) \frac{\partial e}{\partial r} - \gamma \frac{\partial}{\partial r} \left(1 + \frac{\tau_s^\alpha}{\alpha!} \frac{\partial^\alpha}{\partial t^\alpha} \right) T. \tag{15}$$

The corresponding heat energy equation is given as

$$\kappa \left\{ 1 + \frac{\tau_T^\alpha}{\alpha!} \frac{\partial^\alpha}{\partial t^\alpha} + \frac{\tau_r^{2\alpha}}{(2\alpha)!} \frac{\partial^{2\alpha}}{\partial t^{2\alpha}} \right\} \Delta^2 T = \left\{ \frac{\partial}{\partial t} + \frac{\tau_Q^\alpha}{\alpha!} \frac{\partial^\alpha}{\partial t^\alpha} + \frac{\tau_Q^{2\alpha}}{(2\alpha)!} \frac{\partial^{2\alpha}}{\partial t^{2\alpha}} \right\} (\rho c_p T + \gamma T_0 e), \tag{16}$$

where ρ is the material density, c_p is the specific heat of the material, T_0 is the reference temperature of the medium, γ is the material constant given by $\gamma = \alpha_t(3\lambda + 2\mu)$. α_t is the coefficient of linear thermal expansion and λ, μ are Lamé constants.

The constitutive equations for the stress components σ_{ij} can be written as

$$\sigma_{rr} = \left\{ 2\mu \frac{\partial u}{\partial r} + \lambda e - \gamma \left(1 + \frac{\tau_s^\alpha}{\alpha!} \frac{\partial^\alpha}{\partial t^\alpha} \right) (T - T_0) \right\}, \tag{17}$$

$$\sigma_{\phi\phi} = \sigma_{\theta\theta} = \left\{ 2\mu \frac{u}{r} + \lambda e - \gamma \left(1 + \frac{\tau_s^\alpha}{\alpha!} \frac{\partial^\alpha}{\partial t^\alpha} \right) (T - T_0) \right\}, \tag{18}$$

$$\sigma_{\theta\phi} = \sigma_{r\phi} = \sigma_{r\theta} = 0. \tag{19}$$

Equations (9) – (19) constitutes the mathematical formulation of the problem.

2.3 Dimensionless Form

To convert the dimensionless system of equations one can make use of the following non-dimensional variables

$$(r', u') = \varepsilon \zeta (r, u), (t', \tau'_T, \tau'_Q, \tau'_s) = \varepsilon^2 \zeta (t, \tau_T, \tau_Q, \tau_s), \sigma'_{ij} = \frac{\sigma_{ij}}{\mu}, \Theta' = \frac{T - T_0}{T_0}. \tag{20}$$

Dropping the primes for simplicity following equations are the non-dimensional conversions of constitutive equations (dropping primes for simplicity)

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial e}{\partial r} - \omega \frac{\partial}{\partial r} \left(\Theta + \frac{\tau_s^\alpha}{\alpha!} \frac{\partial^\alpha \Theta}{\partial t^\alpha} \right), \tag{21}$$

$$\kappa \left\{ 1 + \frac{\tau_T^\alpha}{\alpha!} \frac{\partial^\alpha}{\partial t^\alpha} + \frac{\tau_r^{2\alpha}}{(2\alpha)!} \frac{\partial^{2\alpha}}{\partial t^{2\alpha}} \right\} \Delta^2 \Theta = \left\{ \frac{\partial}{\partial t} + \frac{\tau_Q^\alpha}{\alpha!} \frac{\partial^\alpha}{\partial t^\alpha} + \frac{\tau_Q^{2\alpha}}{(2\alpha)!} \frac{\partial^{2\alpha}}{\partial t^{2\alpha}} \right\} (\Theta + \varepsilon e), \tag{22}$$

$$\sigma_{rr} = (\beta^2 - 2)e + 2 \frac{\partial u}{\partial r} - \nu \left(\Theta + \frac{\tau_s^\alpha}{\alpha!} \frac{\partial^\alpha \Theta}{\partial t^\alpha} \right), \tag{23}$$

$$\sigma_{\theta\theta} = \sigma_{\phi\phi} = (\beta^2 - 2)e + 2 \frac{u}{r} - \nu \left(\Theta + \frac{\tau_s^\alpha}{\alpha!} \frac{\partial^\alpha \Theta}{\partial t^\alpha} \right), \tag{24}$$

where

$$\varepsilon = \sqrt{\frac{\lambda + 2\mu}{\rho}}, \zeta = \frac{\rho c_p}{\kappa}, \beta = \sqrt{\frac{\lambda + 2\mu}{\mu}}, \nu = \frac{\gamma}{\rho c_p}, \omega = \frac{\gamma T_0}{\lambda + 2\mu}, \nu = \frac{\gamma T_0}{\mu}. \tag{25}$$

Equations (20) – (25) constitutes non-dimensional conversion of governing equations (9) – (19).

2.4 The Boundary Conditions

Assuming quiescent initial state, one can apply the following boundary conditions on the spherical cavity as

$$\Psi(r, t) = \begin{cases} H(t)\Theta_{a_1}, & \text{for } r = a_1 \\ H(t)\Theta_{a_2}, & \text{for } r = a_2' \end{cases} \tag{26}$$

where $\Theta_{a_1}, \Theta_{a_2}$ are constant strengths of the temperature applied at inner and outer boundary of the cavity respectively and $H(t)$ is the Heaviside unit step function.

The boundaries of the cavity are traction free, physically there is no mechanical load is applied to the inner and outer boundary surfaces, this means

$$\sigma_{rr}(r, t)|_{r=r_1} = 0, \tag{27}$$

$$\sigma_{rr}(r, t)|_{r=r_2} = 0. \tag{28}$$

Assuming quiescent state, one will have the following equations describing initial conditions

$$\Theta(r, 0) = \dot{\Theta}(r, 0) = 0, \tag{29}$$

$$\sigma_{rr}(r, 0) = \dot{\sigma}_{rr}(r, 0) = 0, \tag{30}$$

$$u(r, 0) = \dot{u}(r, 0) = 0. \tag{31}$$

The equations (26) – (31) describes the imposed initial and boundary conditions of the problem.

III. THE SOLUTION

The solution of the governing partial differential equations along with assumed initial and boundary conditions of the model has been attained on removal of time variable t by applying Laplace integral transforms. The resultant ordinary differential equations in space variable r are solved in context of the Modified Bessel’s functions. Finally, the inversion of the Laplace transform is carried out numerically using Gaver-Stehfest algorithm to obtain the time domain solution for a particular value of time t .

3.1 Transformation in Laplace Domain

Theorem : Following Liang et al. (2015), if $\alpha > 0, n = [\alpha] + 1$, and functions $f(t), f'(t), f''(t), \dots, f^{(n-1)}(t)$ are continuous in $[0, \infty)$ and of exponential order, while ${}^C D_0^\alpha f(t)$ with order α is piecewise continuous on $[0, \infty)$, then Laplace transform of Caputo fractional derivative of $f(t)$ is defined as follows

$$\mathcal{L}({}^C D_0^\alpha f(t)) = s^\alpha \mathcal{L}(f(t)) - \sum_{k=0}^{n-1} s^{\alpha-k-1} f^{(k)}(0), \tag{32}$$

Due to homogeneous initial conditions, in view of above theorem, one can assume following equality

$$f(x, 0) = \frac{\partial}{\partial t} f(x, 0) = \frac{\partial^2}{\partial t^2} f(x, 0) = \dots = 0, \tag{33}$$

In view of formula given in equation, (32), the non-dimensional governing equations, (21) – (24) are transformed in Laplace domain as

$$s^2 \bar{u} = \frac{d\bar{e}}{dr} - \omega \left(1 + \frac{\tau_s^\alpha}{\alpha!} s^\alpha \right) \frac{d\bar{\theta}}{dr}, \tag{34}$$

$$\left\{ 1 + \frac{\tau_r^\alpha}{\alpha!} s^\alpha + \frac{\tau_r^{2\alpha}}{(2\alpha)!} s^{2\alpha} \right\} \Delta^2 \bar{\theta} = \left\{ s + \frac{\tau_Q^\alpha}{\alpha!} s^{\alpha+1} + \frac{\tau_Q^{2\alpha}}{(2\alpha)!} s^{2\alpha+1} \right\} (\bar{\theta} + \epsilon \bar{e}), \tag{35}$$

$$\bar{\sigma}_{rr} = (\beta^2 - 2)\bar{e} + 2 \frac{d\bar{u}}{dr} - \nu \left(1 + \frac{\tau_s^\alpha}{\alpha!} s^\alpha \right) \bar{\theta}, \tag{36}$$

$$\bar{\sigma}_{\theta\theta} = \bar{\sigma}_{\phi\phi} = (\beta^2 - 2)\bar{e} + 2 \frac{\bar{u}}{r} - \nu \left(1 + \frac{\tau_s^\alpha}{\alpha!} s^\alpha \right) \bar{\theta}. \tag{37}$$

Boundary conditions (26) – (28) transformed in Laplace domain as

$$\bar{\Psi}(r, s) = \begin{cases} \frac{\theta_{a1}}{s}, & \text{for } r = a_1 \\ \frac{\theta_{a2}}{s}, & \text{for } r = a_2 \end{cases}, \tag{38}$$

$$\bar{\sigma}_{rr}(r, s)|_{r=a1} = 0, \tag{39}$$

$$\bar{\sigma}_{rr}(r, s)|_{r=a2} = 0. \tag{40}$$

Applying the operator $\Delta^2 \equiv \frac{1}{r^2} \frac{\partial}{\partial r} \left(\frac{\partial}{\partial r} r^2 \right)$ to both sides of equation (34) and simplifying, one obtains

$$(\Delta^2 - s^2)\bar{e} = \omega \left(1 + \frac{\tau_s^\alpha}{\alpha!} s^\alpha \right) \Delta^2 \bar{\theta}. \tag{41}$$

For the sake of simplicity, one can assume following representation of coefficients involved in dual-phase-lag heat conduction and stresses equations given by (42) – (44)

$$\varrho = \left\{ s + \frac{\tau_Q^\alpha}{\alpha!} s^{\alpha+1} + \frac{\tau_Q^{2\alpha}}{(2\alpha)!} s^{2\alpha+1} \right\}, \tag{42}$$

$$\vartheta = \left\{ 1 + \frac{\tau_r^\alpha}{\alpha!} s^\alpha + \frac{\tau_r^{2\alpha}}{(2\alpha)!} s^{2\alpha} \right\}, \tag{43}$$

$$\varphi = 1 + \frac{\tau_s^\alpha}{\alpha!} s^\alpha, \tag{44}$$

where $\varrho, \vartheta, \varphi$ are the functions of Laplace parameter s .

Rewriting equations (34) and (35) using equations (42) – (44), one have following equations

$$\vartheta \Delta^2 \bar{\theta} = \varrho (\bar{\theta} + \epsilon \bar{e}), \tag{45}$$

$$(\Delta^2 - s^2)\bar{e} = \omega \varphi \Delta^2 \bar{\theta}. \tag{46}$$

Assuming the following substitution for simplicity

$$\psi = \frac{\vartheta}{\varrho}, \tag{47}$$

the variable ψ must be a function of Laplace parameter s .

Eliminating $\Delta^2 \bar{e}$ between equations (45) – (46) and simplifying for $\bar{\theta}$, one will have

$$[\Delta^4 - \Delta^2 \{s^2 + \psi(1 + \omega \epsilon \varphi)\} + s^2 \psi] \bar{\theta} = 0. \tag{48}$$

equation (48) can be factorized as

$$(\Delta^2 - p_1^2)(\Delta^2 - p_2^2)\bar{\theta} = 0, \tag{49}$$

where, p_1^2, p_2^2 are the roots with positive real parts of the following characteristic equation

$$p^4 - p^2 \{s^2 + \psi(1 + \omega \epsilon \varphi)\} + s^2 \psi = 0, \tag{50}$$

are given below

$$p_1^2, p_2^2 = \frac{(s^2 + \psi(1 + \omega\epsilon\varphi)) \pm \sqrt{(s^2 + \psi(1 + \omega\epsilon\varphi))^2 - 4s^2\psi}}{2} \tag{51}$$

The solution of equation (49) is given as

$$\bar{\theta} = \bar{\theta}_1 + \bar{\theta}_2, \tag{52}$$

where

$$(\Delta^2 - p_1^2)\bar{\theta}_1 = 0 \text{ and } (\Delta^2 - p_2^2)\bar{\theta}_2 = 0. \tag{53}$$

Equations (34) – (53) represents the conversion of non- dimensional governing equations (21) – (24) in Laplace domain.

3.2 Solution in Laplace Domain

Solving equation (49) for $\bar{\theta}$ and then substituting in equations (34) – (37), one will have the following results for distribution of temperature $\bar{\theta}$, displacement \bar{u} and thermal stresses $\bar{\sigma}_{rr}, \bar{\sigma}_{\phi\phi}, \bar{\sigma}_{\theta\theta}$ bounded at infinity in Laplace domain as

$$\bar{\theta}(r, s) = \frac{1}{\sqrt{r}} \sum_{i=1}^{i=2} A_i (p_i^2 - s^2) K_{1/2}(p_i r), \tag{54}$$

$$\bar{e}(r, s) = \frac{\omega\varphi}{\sqrt{r}} \sum_{i=1}^{i=2} A_i p_i^2 K_{1/2}(p_i r), \tag{55}$$

$$\bar{u}(r, s) = -\frac{\omega\varphi}{\sqrt{r}} \sum_{i=1}^{i=2} A_i p_i K_{3/2}(p_i r), \tag{56}$$

$$\bar{\sigma}_{rr}(r, s) = \frac{\omega\varphi}{\sqrt{r}} \sum_{i=1}^{i=2} A_i \left\{ \beta^2 s^2 K_{1/2}(p_i r) + \frac{4p_i}{r} K_{3/2}(p_i r) \right\}, \tag{57}$$

$$\bar{\sigma}_{\phi\phi} = \bar{\sigma}_{\theta\theta} = \frac{\omega\varphi}{\sqrt{r}} \sum_{i=1}^{i=2} A_i \left\{ \beta^2 s^2 - 2p_i^2 - \frac{2p_i}{r} \right\} K_{1/2}(p_i r), \tag{58}$$

where $K_{\frac{1}{2}}(\cdot), K_{\frac{3}{2}}(\cdot)$, are modified Bessel functions of second kind of order 1/2 and 3/2 respectively and A_1, A_2 are unknown constants depending upon s . Applying the transformed boundary conditions (38) – (40) to the results presented by equation (54), (57), one gets the following simultaneous equations in unknown constants $A_1(s)$ and $A_2(s)$ as follows

$$A_1(p_1^2 - s^2)K_{1/2}(p_1 a_1) + A_2(p_2^2 - s^2)K_{1/2}(p_2 a_1) = \frac{\theta_{a_1} \sqrt{a_1}}{s}, \tag{59}$$

$$A_1(p_1^2 - s^2)K_{1/2}(p_1 a_2) + A_2(p_2^2 - s^2)K_{1/2}(p_2 a_2) = \frac{\theta_{a_2} \sqrt{a_2}}{s}. \tag{60}$$

Solving equations (59) – (60) simultaneously, one will have following $A_1(s), A_2(s)$

$$A_1(s) = \frac{\theta_{a_1} \sqrt{a_1} K_{\frac{1}{2}}(p_2 a_2) - \theta_{a_2} \sqrt{a_2} K_{\frac{1}{2}}(p_2 a_1)}{s(p_1^2 - s^2) \left\{ K_{\frac{1}{2}}(p_1 a_1) K_{\frac{1}{2}}(p_2 a_2) - K_{\frac{1}{2}}(p_1 a_2) K_{\frac{1}{2}}(p_2 a_1) \right\}}, \tag{61}$$

$$A_2(s) = \frac{\theta_{a_2} \sqrt{a_2} K_{\frac{1}{2}}(p_1 a_1) - \theta_{a_1} \sqrt{a_1} K_{\frac{1}{2}}(p_1 a_2)}{s(p_1^2 - s^2) \left\{ K_{\frac{1}{2}}(p_1 a_1) K_{\frac{1}{2}}(p_2 a_2) - K_{\frac{1}{2}}(p_1 a_2) K_{\frac{1}{2}}(p_2 a_1) \right\}}. \tag{62}$$

Equations (54) – (62) constitutes the analytical solution of the problem in Laplace domain.

3.3 Gaver-Stehfest Algorithm

Applying Gaver-Stehfest algorithm (1970a), (1970b), (1966) to obtain the inverse Laplace transforms to the equations (54) – (62) to approximate the time domain solution using the equation (63) given below

$$f(t) \approx f_M(t) = \frac{\ln(2)}{t} \sum_{k=1}^{2M} \left[(-1)^{M+k} \left\{ \sum_{j=\lfloor \frac{k+1}{2} \rfloor}^{\min(k, M)} \frac{j^{M+1}}{M!} \binom{M}{j} \binom{2j}{j} \binom{j}{k-j} \right\} \cdot \bar{f} \left(\frac{k \ln(2)}{t} \right) \right], \tag{63}$$

and $\lfloor x \rfloor$ is the flooring function and $2M$ is an even integer whose value depends on the word length of the computer used.

3.4 The Convergence Criterion

If $f: (0, \infty) \rightarrow \mathbb{R}$ is a locally integrable function such that its Laplace transform $\bar{f}(s)$ exists for all $s > 0$ and the sequence $f_M(t)$ is defined by equation (63), then the convergence of sequence $f_M(t)$ depends on the values of the function f in the neighbourhood of t . If the function f is of bounded variation in the neighbourhood of t then the sequence $f_M(t) \rightarrow \frac{f(t+0) + f(t-0)}{2}$ as $M \rightarrow \infty$.

IV. NUMERICAL SCHEME

Following fractional theory of thermoelasticity by Povstenko (2015) the thermal variations are studied by numerical values of thermal properties of spherical cavity by considering various order (α) of fractional differential coefficient in time variable. As a special case for coupled thermoelasticity ($\alpha \rightarrow 0$), Cattaneo approach to generalized fractional thermoelasticity ($\alpha = 0.2, 0.5, 0.7$) and generalized theory of thermoelasticity ($\alpha = 1.0$) are discussed numerically.

4.1 Dimensions

Following cavity dimensions for numerical computations.

- The inner radius of spherical cavity $a_1 = 0.1 \text{ m}$.
- The outer radius of spherical cavity $a_2 = 0.2 \text{ m}$.

4.2 The Material Properties

Following Hamza et al. (2016) the numerical scheme has been applied for copper material with physical properties (SI-units)

$$\alpha_t = 1.78 \times 10^{-5} K^{-1}, \beta^2 = 4, \epsilon = 0.0168 N \cdot m \cdot J^{-1}, \kappa = 386 J \cdot K^{-1} \cdot m^{-1} \cdot s^{-1}, \lambda = 7.76 \times 10^{10} N \cdot m^{-2}, \mu = 3.86 \times 10^{10} N \cdot m^{-2},$$
$$\rho = 8954 \text{ kg} \cdot m^{-3}, \tau_T = 0.015 \text{ ps}, \tau_Q = 0.021 \text{ ps}, \tau_s = 0.031 \text{ s}, c_p = 383.1 J \cdot K g^{-1} \cdot K^{-1}, \theta_{a_1} = 400 K, \theta_{a_2} = 500 K, T_0 = 293 K.$$

Gaver-Stehfest algorithm is applied to the equations (54) – (62) to obtain the non-dimensional variations of temperature Θ , displacement u

and thermal stresses $\sigma_{rr}, [\sigma_{\phi\phi}, \sigma_{\theta\theta}]$ for fixed time $t = 0.1 \text{ s}$. Numerically computed results are plotted for the finite cavity space ranging from $0.10 \leq r \leq 0.20$ in the radial direction. Figures 1-4 are describing the change of temperature, thermal displacement, radial and hoop stresses respectively. The resultwise observations are summarized below:

V. RESULTS AND DISCUSSION

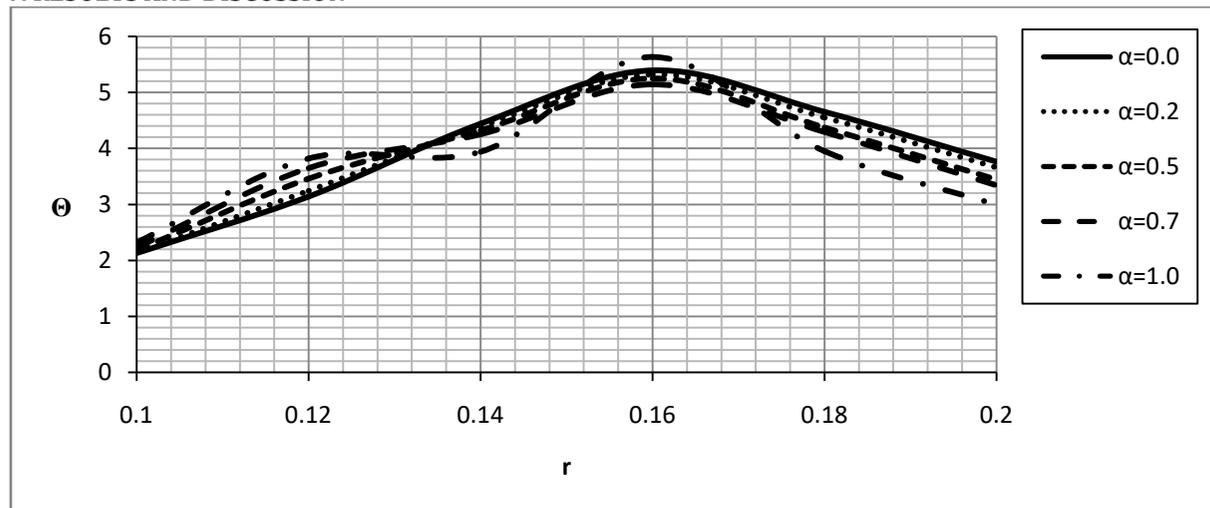


Figure 1: Distribution of temperature $\Theta(r, t)$ with in the cavity space

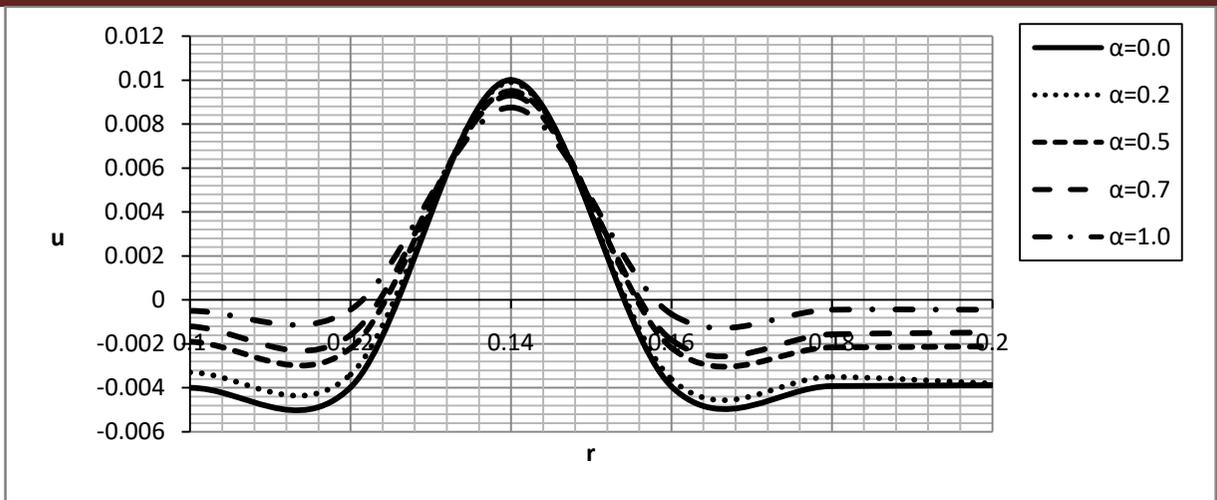


Figure 2: Distribution of thermal displacement $u(r, t)$ with in the cavity space

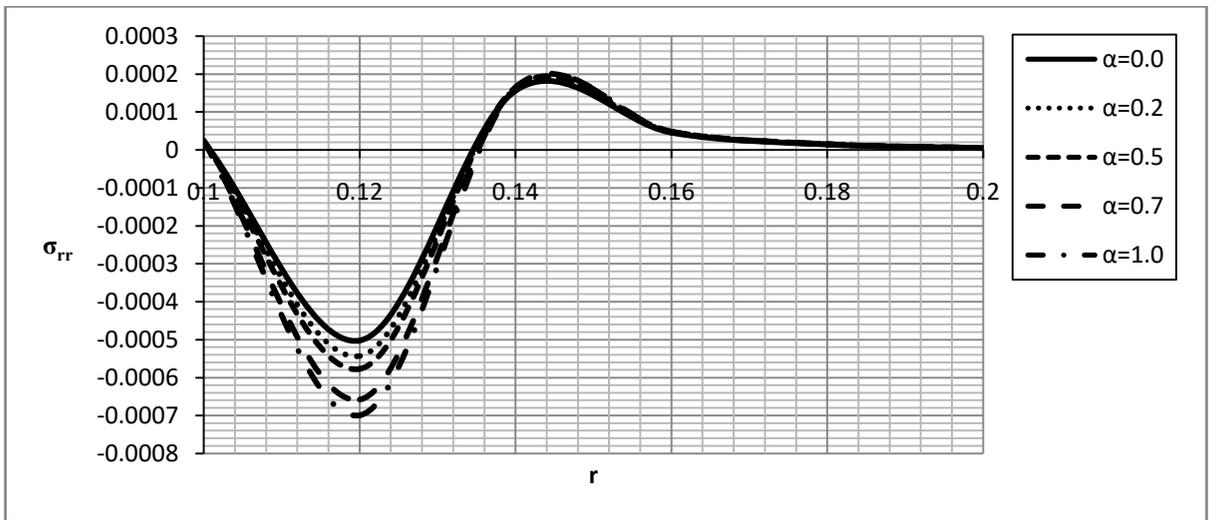


Figure 3: Distribution of hoop stress $\sigma_{rr}(r, t)$ with in the cavity space

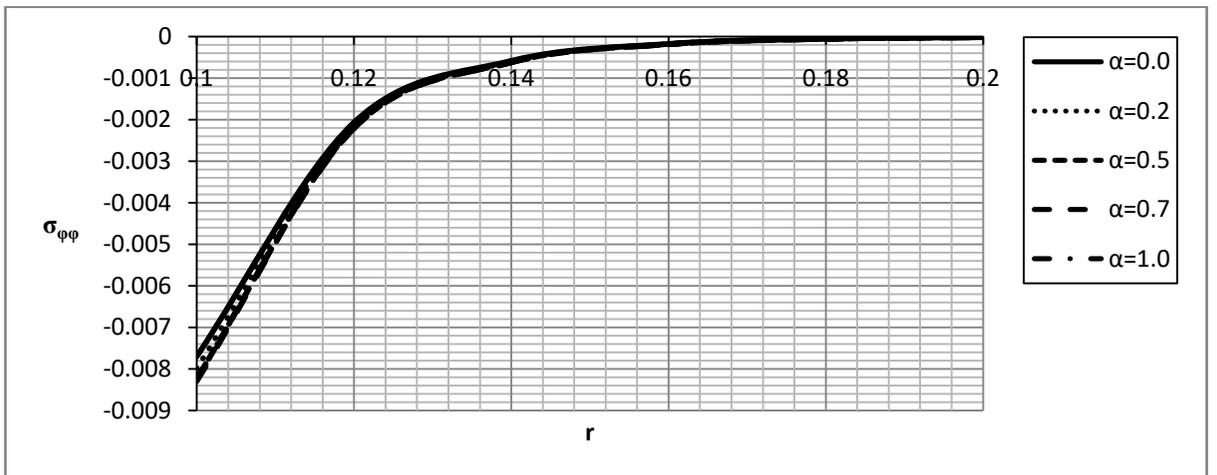


Figure 4: Distribution of hoop stress $\sigma_{\phi\phi}(r, t)$ with in the cavity space

Figure 1: Illustrates the non-dimensional change in temperature Θ inside the elastic body with the spherical cavity in radial direction. The distribution of temperature attains maxima for all the fractional orders

around $r = 0.16$ inside the cavity. As further cavity space increases the temperature decreases. It can be seen that the distribution of temperature is inversely proportional to the fractional order applied.

Figure 2: Exhibits the non-dimensional values of thermal displacement u inside the thermoelastic body with the spherical cavity in the radial direction. The displacement curves for all the fractional orders are symmetrically placed around $r = 0.14$. One can note that as and when the values of fractional order α increases the peakedness of displacement curves increases rapidly. It has been observed that the change in displacement is directly proportional to the fractional order α .

Figure 3: Depicts the non-dimensional variations of radial thermal stresses σ_{rr} . The radial stress variations around the inner cavity are compressive and tensile in nature around $0.10 \leq r < 0.14$ and $0.14 \leq r \leq 0.16$ respectively and very close for all fractional orders that are achieving zero around $r = 0.18$ onwards. Observing the results around $0.10 < r < 0.14$ it has found that the radial stresses σ_{rr} are inversely proportional to the fractional order α . The numerical values of radial thermal stresses σ_{rr} are zero at both traction free spherical boundaries $r = 0.1$ and $r = 0.2$.

Figure 4: Shows the plots of the non-dimensional variations of hoop stresses $\sigma_{\phi\phi}$ and $\sigma_{\theta\theta}$. The patterns of hoop stresses shows tensile behaviour. Hoop stress variations are low around the inner cavity boundary and very close for all fractional orders are negligible around $r = 0.16$ onwards. All of the computed results for thermal stresses show smooth variations and reaches to zero at the outer cavity boundary $r = 0.2$. Hence satisfies the limiting boundary conditions proposed for extremities of the occupied cavity region.

Noticing results graphically it can be found that while $\alpha \rightarrow 0$ the fractional theory is quite close to the classical coupled theory of thermoelasticity, similarly as and when $\alpha \rightarrow 1$ from $\alpha = 0.7$ results are thoroughly neighbouring hence fractional theory of thermoelasticity tends towards Lord-Shulman generalized theory (1967) of thermoelasticity. It has been found that the results are compatible with classical coupled theory and generalized theory for weak ($\alpha \rightarrow 0$) and normal conductivity ($\alpha = 1.0$) thermoelastic problems respectively. Hence the fractional theory of thermoelasticity used in this model is found to be applicable for fractional orders $\alpha \in (0,1)$ proposed by Sherief et al. (2010). The numerical results for change in temperature Θ , displacement u and thermal stresses $\sigma_{rr}, \sigma_{\phi\phi} = \sigma_{\theta\theta}$ are computed by applying various fractional orders $\alpha = 0.0, 0.2, 0.5, 0.7, 1.0$ inside the various points thermoelastic body with spherical cavity in radial sense for time $t = 0.1s$.

VI. THE CONCLUDING REMARKS

This is an attempt to investigate the thermal behaviour of a spherical cavity in a bounded domain. A dual-phase-lag model consists of relaxation time and two different translations have been framed in the context of fractional order thermoelasticity and studied the microstructural interactions and thermal variations within the spherical cavity. The significant outputs of this model are summarised as follows:

- The finite speed of thermal wave propagation has been achieved by applying small values of relaxation time τ_s and phase-lags τ_q, τ_T to the fractional heat conduction equation and the associated stress relations proposed in this model considering various values of fractional order. Every thermal parameter has been considered with an involvement of relaxation time τ_s and phase-lags τ_T, τ_Q .
- The Classical coupled and generalized thermoelasticity theories are recovered by considering the various special cases for the different fractional orders (α) of derivatives and two different translations under consideration. As $\alpha \rightarrow 0$ the fractional theory is quite close to coupled theory of thermoelasticity and when $\alpha \rightarrow 1$ fractional theory of thermoelasticity tends towards generalized theory thermoelasticity.
- Every thermal parameter $\Theta(r, t), u(r, t), \sigma_{rr}(r, t), \sigma_{\phi\phi}(r, t), \sigma_{\theta\theta}(r, t)$ found is attaining the limiting values and satisfies the imposed initial and boundary conditions proposed in the model for all fractional orders considered.
- Numerical results computed in this manuscript for temperature change, displacement and thermal stresses are validated by the equilibrium and compatibility equations of generalized thermoelasticity theory. The results obtained are also compared with classical and generalized thermoelasticity theories.
- The convergence of infinite series solutions has been discussed using Kuznetsov convergence theorem (2013).
- The results found in this article lead new research ideas in the field of coupled fractional thermoelasticity, especially in the finite spherical domain. According to this work, the fractional order α becomes a new investigator to categorize various materials as per their ability of heat conduction.

• This model can be re-examined in context of three-phase-lag heat conduction models to explore various other situations involving microstructural heat conduction problems by changing the various types of heat sources.

REFERENCES

- [1] Abbas, I. A. 2015. Eigenvalue Approach to Fractional Order Thermoelasticity for an Infinite Body With a Spherical Cavity, JAAUBAS. 1-5.
- [2] Bigot, J. Y., Vomir, M. and Beaurepaire, E. 2009. Coherent Ultrafast Magnetism Induced by Femtosecond Laser Pulses, J. Nat. Phys. 5, 515-520.
- [3] Caputo, M. 1967. Linear Models of Dissipation Whose Q is Almost Frequency Independent-II, Geophys. J. Int. 13, 529-539.
- [4] Ezzat, M. A. and El-Bary, A. A. 2018. Thermoelectric Spherical Shell With Fractional Order Heat Transfer, J. Micro. Tech. 24, 891-899.
- [5] Ezzat M. A. , El-Karamany, A. S., El-Bary, A. A. and Fayik, M. A. 2014. Fractional Ultrafast Magnetism Laser Induced Magneto-Thermoelastic Behaviour in Perfect Conducting Metal Films, J. Elect. Waves and Appl. 28, 64-82.
- [6] Gaver, D. P. 1966. Observing Stochastic Processes and Approximate Transform Inversion, Oper. Res. 14, 444-459.
- [7] F. Hamza, F., Abdou, M. and Abd. El-Latif, A. M. 2014. Generalized Fractional Thermoelasticity Associated With Two Relaxation Times, J. Therm. Stresses. 37, 1080-1097.
- [8] Hamza F., Abdou, M. and Abd. El-Latif, A. M. 2016. 1D Applications on Fractional Generalized Thermoelasticity Associated With Two Relaxation Times, Mech. Adv. Mat. Struct. 23, 689-703.
- [9] Hendy M. H., Amin, M. M. and Ezzat M. A. 2018. Magneto-Electric Interactions Without Energy Dissipation for a Fractional Thermoelastic Spherical Cavity, J. Micro. Tech. 24, 2895-2903.
- [10] Kilbas, A. A., Srivastava, H. M. and Trujillo, J. J. 2006. Theory and Applications of Fractional Differential Equations, Elsevier, New-York.
- [11] Kuznetsov, A. 2013. On the Convergence of the Gaver-Stehfest Algorithm, SIAM. 51, 2984-2998.
- [12] Liang, S. Ranchao W. and Chen L. 2015. Laplace Transform of Fractional Order Differential Equations. EJDE. 139, 1-15.
- [13] Lord, H. W. and Shulman Y. 1967. A Generalized Dynamical Theory of Thermoelasticity, J. Mech. Phys. Solids. 15, 299-309.
- [14] Miller, K. and Ross B. 1993. An Introduction to the Fractional Calculus and Fractional Differential Equations, John Wiley and Sons, Inc. New York.
- [15] Mittal, G. and Kulkarni, V. S. 2018. Dual-Phase-Lag Thermoelastic Problem in Finite Cylindrical Domain with Relaxation Time, MMMS. 14(5), 837-856.
- [16] Pan, I. and Das, S. 2013. Intelligent Fractional Order Systems and Control: An Introduction, New-York, Springer.
- [17] Povstenko, Y. Z. 2005. Fractional Heat Conduction Equation and Associated Thermal Stress, J. Them. Stress. 28, 83-102.
- [18] Povstenko. Y. Z. 2015. Fractional Thermoelasticity, Solid Mechanics and Its Applications, Switzerland, Springer.
- [19] Sherief, H. H., El-Sayed, A. and El-Latif, A. 2010. Fractional Order Theory of Thermoelasticity, Int. J. Solids. Struct. 47, 269-275.
- [20] Sherief, H. H. and El-Latif, Abd A. M. 2013. Application of Fractional Order Theory of Thermoelasticity to a 1D Problem for a Half-Space, ZAMM. 2, 1-7.
- [21] Sherief, H. H. and Hussein, E. M. 2017. Fundamental Solution of Thermoelasticity With Two Relaxation Times for an Infinite Spherically Symmetric Space, J. Appl. Math. Phys. 68:50.
- [22] Sherief, H. H. and Raslan, W. E. 2017. A 2D Problem of Thermoelasticity Without Energy Dissipation for a Sphere Subjected to Axisymmetric Temperature Distribution, J. Therm. Stress. 40, 1461-1470.
- [23] Sherief, H. H. and Allam, A. A. 2017. 2D Axisymmetric Problem for a Sphere With Heat Sources in the Theory of Generalized Thermoelasticity, Intl. J. App. Mech. 9, 1-23.
- [24] Stehfest, H. 1970a. Algorithm 368, Numerical Inversion of Laplace Transforms, Comm. Assn. Comp. Mach. 13, 47-49.
- [25] Stehfest, H. 1970b. Remark on Algorithm 368, Numerical Inversion of Laplace Transforms, Comm. Assn. Comp. Mach. 13, 624-624.
- [26] Tzou, D. Y. 1995. A Unified Field Approach for Heat Conduction From Macro to Micro Scales, J. Heat Trans. 117, 8-16.
- [27] Waleed, E. R. 2016. Application of Fractional Order Theory of Thermoelasticity to a 1D Problem for a Spherical Shell, J. Theo. Appl. Mech. 54, 295-304.
- [28] Yalin, Lu., Dajani, R. and Knize, R. J. 2016. Ultrafast Laser Assisted Fabrication of ZnO Nanorod Arrays for Photon Detection Applications, App. Surf. Sc. 253, 7851-7854.
- [29] Yong-bin, M. and Peng, W. 2017. Dynamic Response of an Infinite Medium With a Spherical Cavity on Temperature-Dependent Properties Subjected to a Thermal Shock Under Fractional-Order Theory of Thermoelasticity, J. Therm. Stress. 41, 1-11.
- [30] Youssef, H. M. 2010. Theory of Fractional Order Generalized Thermoelasticity, J. Heat. Trans. 132, 1-7.

- [31] Youssef, H. M. and Eman Al-Lehaibi, A. 2010. Fractional Order Generalized Thermoelastic Half-Space Subjected to Ramp-Type Heating, Mech. Res. Commn. 37, 448-452.
- [32] Youssef, H. M. and Elsibai, K. A. 2011. Vibration of Nano Beam Induced by Ramp Type Heating, World J. Nano Sci. Engg. 1, 37-44.
- [33] Zewail, A. H. 2000. Femtochemistry : Atomic-Scale Dynamics of the Chemical Bond Using Ultrafast Lasers (nobel lecture), Ange. Chem. Intl. Ed. 39, 2586-2631.