ENHANCED PREDICTION OF HEART DISEASE WITH OPTIMAL FEATURE SELECTION

N. D. Bhagwat[1] & Prof. J. R. Mankar [2]
[1] M.E. Student, [2]Assistant Professor
[1] Department of Computer Engineering,
[1] K. K. Wagh Institute of Engineering Education & Research, Nashik,
Savitribai Phule Pune University, Maharashtra, India.

Received: September 18, 2018 Accepted: November 02, 2018

ABSTRACT: *As the data is growing very rapidly in all domains, it is challenging task for data processing and analysis techniques to handle this data. So the traditional record maintenance strategy in the healthcare system is replaced with e-healthcare data efficiently. In the e-healthcare system, it is vital to develop a model that is able to handle the huge amount of data efficiently. This paper focuses on the challenging tasks of selecting critical features from the enormous set of available features and predicting heart disease. In classification problems, feature selection is one of the most widely used preprocessing steps. The existing system consists of a modified Differential Evolution (DE) algorithm with feed-forward neural network. This algorithm performs feature selection and optimization for cardiovascular disease. In proposed system, back-propagation model is used in which error minimization will be carried out in the prediction of heart disease to improve the performance in terms of accuracy, specificity and sensitivity. The performance of the system will be tested on dataset available in an online repository of the University of California, Irvine (UCI) which contains 303 data samples of individuals.*

Key Words: *Optimization, Feature selection, Differential evolution, Fuzzy AHP, Feed-forward neural network*

I. Introduction

Data analytics in a healthcare information system can extract valuable information due to the tremendous growth in the volume of healthcare data. As of late, healthcare organizations have been moving towards digitization of the huge volume of medicinal services information to use data analytics in healthcare to acknowledge broad advantages. The potential advantages incorporate identification of diseases at prior stages, cost decrease, customized treatment, enhanced patient experience, and superior care. A medical record is a document that contains information on patient identity and examinations, treatments, therapies and services which are given to the patient. The results of the digitized translation of medical records are known as electronic medical records (EMRs). EMRs are extremely complex and of huge size. It becomes difficult to manage such a huge volume of data with traditional data processing systems.

The EMRs contains various healthcare datasets which are generated from different medical sources such as diagnoses, procedures, medications, and lab results and are maintained under different attributes or features. The datasets with tens or hundreds or thousands of variables are available and for that the feature selection or attribute selection has become an important focus in many research applications.

The accuracy of the resulting classifier model can greatly be improved by feature selection. Moreover, it is important to find the relevant subset of predictive features. The process of obtaining the best possible values of decision variables based on the selected objective function is called optimization. For finding near-optimal solutions, evolutionary algorithms have been extensively used in various fields. In this paper, by selecting the critical features, the feature selection is carried out that form the root cause for the objective function of the problem under consideration, which is the prediction of cardio-vascular disease. The UCI data repository contains the heart disease dataset which is used for experimentation.

The differential evolution (DE) algorithm, which is a kind of evolutionary algorithm, has been adopted as an effective global optimizer. It is effective over continuous space because it is a powerful population-based stochastic technique. Differential evolution has been successfully implemented in many fields of engineering, such as mechanical engineering, pattern recognition, water resource management, and communication The output of the DE strategy which is the selected critical features has been analysed using fuzzy AHP and a feed-forward neural network, and heart disease prediction has been performed.

II. LITURATURE WORK

The concepts and algorithms of feature selection are introduced in [1]. It surveys existing feature selection algorithms for classification and clustering, groups and compares different algorithms with a categorizing framework based on search strategies, evaluation criteria, and data mining tasks, reveals unattempted combinations, and provides guidelines in selecting feature selection algorithms. An integrated system is built for intelligent feature selection. A self-adaptive DE (SaDE) algorithm was proposed in [2], within which each trial vector generation strategies and their associated control parameter values are gradually self-adapted by learning from their previous experiences in generating promising solutions. Differential evolution has been analyzed as a candidate global optimization method in [3] for feed-forward neural networks. Differential evolution seems not to provide any distinct advantage in terms of learning rate or solution quality in comparison to gradient based methods. The objective of the paper [4] is to match the performance of some totally different ways once employed by associate mechanicsform optimisation routine that styles vane shapes. A recently developed genetic algorithmic rule, Differential Evolution, outperforms additional ancient techniques. The DM techniques most frequently used in cardiology, the performance of DM models in cardiology and the comparisons of the performance of different DM models in cardiology are summarized and analyzed in [5].

To handle with high-dimensional dataset challenge and uncertainties, integration between rough sets based attribute reduction and IT2FLS was proposed in [6]. IT2FLS makes use of a hybrid learning process comprising fuzzy c-mean clustering algorithm and parameters tuning by chaos firefly and genetic hybrid algorithms. Fuzzy analytic hierarchy process (Fuzzy_AHP) technique was used in [7] to compute the global weights for the attributes based on their individual contribution. The attributes used was 13 commonly used heart failure attributes. The global weights were applied to train an ANN classifier for the prediction of HF risks in patients. The global weights represent the contributions of the attributes. A real-coded genetic algorithm (GA)-based system is proposed in [8] to select the critical clinical features essential to the heart diseases diagnosis. The heart disease database used in this study includes 352 cases, and 40 diagnostic features were recorded for each case. Using the proposed genetic algorithm, 24 critical features have been identified, and their corresponding diagnosis weights for each heart disease of interest have been determined. A weighted fuzzy rule-based clinical decision support system (CDSS) is presented in [9] for the diagnosis of heart disease which automatically obtains the knowledge from the patient's clinical data. An automated approach for generation of weighted fuzzy rules and developing a fuzzy rule-based decision support system, are the two phases of CDSS. The heart failure (HF) and left ventricular ejection fraction (LVEF) are measures for predicting heart disease in which pathophysiological differences are observed among patients diagnosed with HF and reduced LVEF compared with HF and preserved LVEF [10].

To predict more accurately the presence of cardiovascular disease in [11], the reduced number of attributes was used. For that, an intelligent system was investigated to generate feature subset with improvement in diagnostic performance. The hybrid forward selection technique was proposed for cardiovascular disease diagnosis which finds smaller subsets. Logistic regression (LR), multivariate adaptive regression splines (MARS), artificial neural network (ANN), and rough set (RS) techniques was hybridized in [12] to reduce the set of explanatory variables for effectively classifying heart disease. A decision support system is presented in [13] for heart disease classification based on support vector machine (SVM), which determines the support vectors in a fast and iterative manner, and integer-coded genetic algorithm (GA). Integer-coded genetic algorithm was used to maximize SVM's classification accuracy, selects the important and relevant features and discards the irrelevant and redundant ones.

In the analytic hierarchy process, a statistical criterion is presented for accepting/rejecting the pairwise reciprocal comparison matrices [14]. This system is capable of adapting the acceptance requirements to different scopes and consistency necessities which takes advantages in introducing the adaptability in the acceptance criterion and the simplicity of the index. The decision maker makes comparisons between pairs of attributes or alternatives in analytic hierarchy process (AHP). A robust regression technique was elaborated in [15] for eliminating the effect of an outlier in the comparison ratios, and compared with the eigenvector method and the logarithmic least squares regression. For representing the genotypes (a set of ten production rules of a Lindenmayer System with memory), a hybrid neuro-evolutive algorithm (NEA) was proposed in [16] that uses a compact indirect encoding scheme (IES), also has the ability to reuse the genotypes and automatically build modular, hierarchical and recurrent neural networks. A Lindenmayer System (L-System) is evolved in a genetic algorithm (GA) which is used to design the neural network's architecture.

Various data mining techniques like association rule mining, classification, clustering was implemented in [17] to analyze the different kinds of heart based problems. To develop an Intelligent System was main objective of [18] using data mining modeling technique, Naive Bayes. In this, user answers the predefined questions which were implemented as web based application. The hidden data was retrieved from stored database and compared the user values with trained data set. It assists healthcare practitioners to make intelligent clinical decisions by answering complex queries for diagnosing heart disease. The association rules in medical data were discovered in [19] to predict heart disease. A computational intelligence approach – Association rule mining is used to identify the factors that contribute to heart disease is considered along with the rule generation algorithm – Apriori. The data mining algorithms such as J48, Naïve Bayes, REPTREE, CART, and Bayes Net were applied in [20] for predicting heart attacks.

III. CONCLUSION

The Differential Evolution-based feature selection is adapted to select the optimal critical features from the actual available features. Fuzzy AHP with a feed-forward neural network is adapted to predict heart disease. The output of the modified Differential Evolution algorithm is given as input to the fuzzy AHP which computes local weights of alternatives and global weights of attributes and then feed forward neural network has been used to predict the heart disease.

IV. Acknowledgment

Authors would like to thank Prof. Dr. K. N. Nandurkar, Principal and Prof. Dr. S. S. Sane, Head of Department of Computer Engineering, K.K.W.I.E.E.R., Nashik for their kind support and suggestions. We would also like to extend our sincere thanks to all the faculty members of the department of computer engineering and colleagues for their help.

References

1. Huan Liu, Lei Yu, Toward Integrating Feature Selection Algorithms for Classification and Clustering, IEEE Transactions on Knowledge and Data Engineering, April 2005, pp. 491–502.
2. A.K. Qin, V.L. Huang, P.N. Suganthan, Differential evolution algorithm with strategy adaption for global numeric optimization, IEEE Trans. Evol. Comput. 13 (No.2) (April 2009).
3. J. Ilonen, Joni-Kristian kamarainen, J. Lampinen, Differential evolution training algorithm for form feed neural networks, Neural Process Lett. 17 (2003) 93–105.
4. T. Rogalsky, R.W. Derksen, S. Kocabiyik, Differential evolution in aerodynamic optimization, in: The Proceedings of 46th Annual Conference of Canada Aeronautical Space Institute., Montreal, QC, Canada, May 1999, pp. 29–36.
5. I. Kadi, A. Idri, J.L. Fernandez-Aleman, Knowledge discovery in cardiology: a systematic literature review, Int. J. Med. Inf. 97 (2017) 12–32.
6. Nguyen Cong Long, Phayung Meesad, Herwig Unger, A highly accurate firefly based algorithm for heart disease prediction, Expert Syst. Appl. 42 (2015) 8221–8231.
7. Oluwarotimi Williams Samuel, Grase Mojisola Asogbon, Arun Kumar Sangaiah, Peng Fang, Guanglin Li, An integrated decision support system based on ANN and Fuzzy AHP for heart failure risk prediction, Expert Syst. Appl. 68 (2017) 163–172.
8. Yan Hongmei, Zheng Jun, Jiang Yingtao, Peng Chenglin, Xiao Shouzhong, Selecting critical features for heart diseases diagnosis with a real-coded genetic algorithm, J. Appl. Soft Comput. 8 (2008) 1105–1111.
9. P.K. Anooj, Clinical decision support system: risk level prediction of heart disease using weighted fuzzy rules, J. Comput. Inf. Sci. Elsevier 24 (2012) 27–40.
10. A.L. Bui, T.B. Horwich, G.C. Fonarow, Epidemiology and risk profile of heart failure, Nat. Rev. Cardiol. 8 (2014) 30–41.
11. S. Shilaskar, A. Ghatol, Feature selection for medical diagnosis: evaluation for cardiovascular diseases, Expert Syst. Appl. 40 (10) (2013) 4146–4153.
12. Y.E. Shao, C.D. Hou, C.C. Chiu, Hybrid intelligent modelling schemes for heart disease classification, Appl. Soft Comput. 14 (Part A) (2014) 47–52.
13. Sumit Bhatia, Praveen Prakash, G.N. Pillai, SVM based decision support system for heart disease classification with integer-coded genetic algorithm to select critical features, in: Proceedings of the World Congress on Engineering and Computer Science (WCECS), San Francisco, USA October 22–24, 2008.
14. J. Alonso, M. Lamata, Consistency in the analytic hierarchy process: a new approach, Int. J. Uncertain. Fuzziness Knowledge-Based Syst. 14 (4) (2006) 445–459.
15. P. Laininen, R. H€am€al€ainen, Analyzing AHP matrices by regression, Eur. J. Oper. Res. 148 (3) (2003) 514–524.
16. M.L.C. Lídio, C.L.O. Roberto, R. Mauro, Optimization of neural net- works through grammatical evolution and a genetic algorithm, Expert Syst. Appl. Elsevier 56 (2016) 368–384.

17. Taneja, Abhishek, "Heart disease prediction system using data mining techniques," Oriental Journal of Computer science and technology 6.4 (2013): 457-466.
18. Pattekari, Shadab Adam, and Asma Parveen, "Prediction system for heart disease using Naïve Bayes." International Journal of Advanced Computer and Mathematical Sciences3.3 (2012): 290-294.
19. Ibrahim Umar Said, Abdullahi Haruna Adam, and Dr. Ahmed Baita Garko —Association Rule Mining on Medical Data to Predict Heart Disease, International Journal of Science Technology and Management Vol. No.4,Issue08, August 2015.
20. Masethe HD, Masethe MA, Prediction of heart disease using classification algorithms, In Proceedings of the world congress on Engineering and Computer Science 2014 Oct 22 (Vol. 2, pp. 22-24).