

# LUNG CANCER DETECTION USING IMAGE PROCESSING AND FCM

**Ayush Goel, Dhruv Takkar, Manpuneet Kaur**

Final Year Students, B. Tech Computer Science, MSIT

Received: May 08, 2019

Accepted: June 15, 2019

**ABSTRACT:** *The early detection of carcinoma may be a difficult downside, because of the structure of the cancer cells, wherever most of the cells square measure overlapped with one another. This project presents 2 phases that's image process and a Fuzzy cluster algorithmic rule, for segmenting humour colour pictures to sight the carcinoma in its early stages. The manual analysis of the humour samples is time intense, inaccurate and needs intensive trained person to avoid diagnostic errors. The segmentation results are used as a base for a computer aided identification (CAD) system for early detection of carcinoma which is able to improve the probabilities of survival for the patient. However, the intense variation within the greylevel and also the relative distinction among the pictures build the segmentation result less correct, therefore, we have a tendency to applied a histogram exploit, thresholding technique, dilation and have extraction as a pre-processing step altogether pictures to extract the nuclei and living substance regions, as a result of most of the quantitative procedures are supported the nuclear feature. The thresholding algorithmic rule succeeded in extracting the nuclei and living substance regions. Moreover, it succeeded in deciding the simplest vary of thresholding values. The FCM strategies is meant to classify the image of into bimodal categories. During this study, we have a tendency to used on hundred humour X-RAY and CT Scan pictures to predict the cancer identification.*

**Key Words:**

## I. INTRODUCTION

Lung cancer is that the most typical explanation for death in each men and girls throughout the planet. Statistics from the American Cancer Society calculable that in 2019 there'll be concerning 228,000 new cases of carcinoma within the U.S. occurred and over 142,000 deaths were because of the malady [1] some six.5% of men and girls within the U.S. are diagnosed with cancer of the respiratory organ at some purpose in their time period supported information from 2011-2013 that could be a serious issue to appear upon. In several developing countries, the incidence of carcinoma is starting to fall following public education concerning the risks of roll of tobacco smoking and also, the introduction of effective smoking-cessation programs. yet, carcinoma remains among the foremost common styles of cancers in each men and girls worldwide. This requires an efficient and time saving technique of cancer medical specialty. we have a tendency to propose through our analysis associate economical time-saving and reliable model of cancer medical specialty that uses image process and neural networks because the basis of the analysis.

### A. Input

For input, X-ray images and CT Scan images are used. The difference between X-ray image and CT scan image is illustrated below. However, it advised to feed a CT scan image as a source of input for better results and accuracy because of better pixel clarity.

Characteristic	X-Ray	CT Scan
1. Image dimension	2D	3D
2. Image clarity	Not much precise as compared to CT Scan	Very precise to observe internal organ injury.
3. Diagnostic Difference	Diagnosis of visible problems like pneumonia and widespread cancer	Diagnosis of soft tissues , muscles damage and precise organ damage clarity.
4. Availability	Easy	Not that abundant

### B. Segmentation

Segmentation could be a method that Permits to divide the image into numerous category elements. The class elements may be space of thought, the foreground space, the background space etc. The foreground space could overlap with the realm of thought therefore them distinctive category characteristics isn't determining nevertheless. Segmentation of medical pictures in 2nd, slice by slice has several helpful applications for the medical skilled such as: visual image and volume estimation of objects of interest, detection of abnormalities (e.g. tumours, polyps, etc.), tissue quantification and classification, and a lot of [1].

More exactly, image segmentation is that the method of distribution a label to each element in a picture specified pixel with an equivalent label share sure visual characteristics. All pixels in a very given region are similar with regard to some characteristic or computed property, like colour, intensity, or texture. Adjacent regions are considerably totally different with regard to an equivalent characteristic(s). Segmentation algorithms are supported one in all 2 basics properties of intensity values: separation and similarity. the primary class is to partition the image supported abrupt changes in intensity, like edges in a picture. The second class is predicated on partitioning the image into regions that are similar consistent with a predefined criterion. histogram thresholding approach falls beneath this class. Through the assistance of segmentation, pictures are divided to some regions that contents of every region have equivalent specifications. Dynamic the image illustration for easier rationalization is the main purpose of segmentation. The output of image segmentation could be a set of segments that together cowl the total image (edge detection). we have a tendency to are applying thresholding approach that is one in all the foremost powerful tools for image segmentation [10]. The metameric image obtained from thresholding has the benefits of smaller space for storing, quick process speed and ease in manipulation, compared with grey level image that sometimes contains 256 levels. Therefore, thresholding techniques have drawn heaps of attention throughout the past twenty years [2].

The simplest thresholding ways replace every element in a picture with a black element if the image intensity is a smaller amount than some fastened constant  $T$  (that is that the threshold value), or a white element if the image intensity is larger than that constant. within the example image on the correct, this ends up in the dark tree turning into fully black, and therefore the white snow turning into fully white [3]. On the opposite hand, in image process you have got way more data than grayscale distribution of a picture (I'm speaking particularly regarding the abstraction distribution of the pixels). Thus straightforward image segmentation ways, morphological operators, image derivatives etc. will typically offer quicker and higher results than sophisticated applied math analysis on histograms.

### C. Histogram Equalization

It is the first step of image processing operation. Histogram equalization is a statistical tool that is used to increase the global contrast of an image which allows the areas of lower local contrast to gain a higher contrast... Histogram equalization does this by dividing out the most occurred values. The method is useful in images with backgrounds and foregrounds that are both bright or both dark.[4] The function will not only enhance the contrast of lower contrast image but also highlight the under-exposed areas with better details. The advantage of this method is that it takes no time to compute and can be reverted (meaning to roll back the function applied on image to obtain the image as it is). Since we are considering bimodal classification and it is one of the handy methods to use for better results ,The histogram equalization function is defined as  $P(X[k])=nk/n$  where  $nk$  represents the number of times that the level  $X[k]$  appears in the input image compare to the total number of samples in the input image.

### D. Dilation

Dilation (usually represented by  $\oplus$ ) is one of the basic operators in mathematical morphology. It is a really successful morphological operation for binary images. The dilation operation uses a functioning element for targeting and expanding the shapes in the input image. The basic effect of the method on a binary image is to gradually zoom the boundaries of area of important information (i.e. white pixels, typically) which is achieved by growing the areas of foreground pixels in size while holes within those regions become smaller. In general cases structuring element is a  $3*3$  matrix as in our case and practical enlighten the foreground regions and holes in and balances out the background relatively. In simple words Dilation zooms in into an image for better clarity and precision of each pixel without affecting the image dynamics. With larger structuring elements, it is quite common to use an approximately disk-shaped structuring element, as opposed to a square one [5].

### E. Filtering Final Content

Filtering in image process could be a method that cleans up appearances and permits for selective lightness of specific data selection mathematical filters. - Image filtering is beneficial for several applications, as well as smoothing, sharpening, removing noise, and edge detection. Filters are often mounted within the photographic equipment to boost image quality and take away blur. A lot of advanced filters could also be able to reconstruct part broken pictures through averaging, using existing knowledge to estimate missing contents in a picture. we have a tendency to implement OTSU technique for this method. This algorithmic program returns one intensity threshold that separate pixels into 2 categories, foreground and background. Otsu's technique exhibits the comparatively smart performance if the bar graph is often assumed to possess bimodal distribution and

assumed to possess a deep and sharp depression between 2 peaks. One fashionable extension is that the two-dimensional Otsu's technique, that performs higher for the article segmentation task in clamorous pictures. Here, the intensity worth of a given component is compared with the common intensity of its immediate neighbourhood to boost segmentation results.

- **Feature Extraction**

After the fundamental image process is performed on respiratory organ region, the metameric nodules area unit used for feature extraction. Feature extraction is one among the foremost necessary steps during this system. A feature may be a vital piece of knowledge extracted from a picture that provides a lot of elaborated understanding of the image. A feature is outlined as a perform of 1 or a lot of measurements, the values of some quantitative property of associate degree object, computed in order that it quantifies some vital characteristics of the article. Feature extraction starts from associate degree initial set of measured knowledge and builds derived values (features) supposed to be informative and non-redundant, facilitating the next learning and generalization steps, and in some cases resulting in higher human interpretations. Feature extraction is expounded to spatiality reduction. Once the input file to associate degree formula is simply too giant to be processed and it's suspected to be redundant (e.g. identical measuring in each feet and meters, or the verbosity of pictures bestowed as pixels), then it is reworked into a reduced set of options (also named a feature vector). Determinant a set of the initial options is termed feature choice. the chosen options area unit expected to contain the relevant data from the input file, in order that the specified task is performed by victimisation this reduced illustration rather than the entire initial knowledge [8]. the data like value price, texture, Boundary space of the region targeting edge detection, distinction and intensity function the quantities of feature for feature extraction.

### G. Neuro Fuzzy Classification Final Content

Classification is that the method of dividing the info into homogenous regions supporting the similarity of objects; info that's logically similar physically is keep along, so as to extend the potency within the information system and to reduce the amount of access. The method of classification is to assign the letter of the alphabet feature vectors into K clusters, for every kth cluster  $C_k$  is its centre. Fuzzy cluster has been employed in several fields like pattern recognition and Fuzzy identification. A range of Fuzzy cluster strategies are planned and most of them are based mostly upon distance criteria. The foremost wide used rule is that the Fuzzy C-Mean rule (FCM), it uses reciprocal distance to figure fuzzy weights. This rule has as input a pre-defined variety of clusters, that is that the k from its name. Suggests that stands for a median location of all the members of specific cluster and therefore the output may be a partitioning of k cluster on a group of objects.

### [1] Scope of study

This paper focuses on the various image processing techniques to detect lung cancer using CT scans of the patients. The advantage of using CT scan images as compared to X-ray and MRI scans is less noise and less distortion. The CT scan images have been taken from NIH/NCI Lung Image Database Consortium dataset. The noise and unwanted pixels are removed using histogram equalization and various segmentation techniques which improve the efficiency of detection.

### TABLE I. Materials and methods

#### Pre-processing and FCM

This is done to make the image free from noise, distortion and corruption. It enhances the contrast of image for the further processing using the Otsu filter method. The sharp edges which has essential information is blurred during smoothing. Segmentation - It divides the noise free images into various segments which have the same specification. It is used to make things simple for analysing the images. It helps in locating the objects, boundaries and curves that exist in the image. It is used for assigning a label to every pixel in the image such that pixels with same label share similar visual characteristics. These characteristics include colour, intensity or texture. It is done using the thresholding method of segmentation. An optimal gray-level threshold value is used to separate objects from the target area. The pros of this method are it has fast processing [9]. Dilation - This is used to find image components which help in defining the shape or flux of the image which helps in its representation. Furthermore, operations like filtering, image filling, feature extraction and FCM classification further dig deep into the classification function to obtain the desired results.

#### LITERATURE SURVEY

Researchers have the aim to develop a system that detects and foresees cancer in its early stages. By exploitation segmentation, reprocessing, feature extraction and classification techniques of extracted databases, they have worked on the accuracy and improvement within the early prediction. Ada, Rajneet Kaur projected a machine procedure that divide the photographs into clusters in line with their

similarities. during this paper histogram equalization is employed for pre-processing of the photographs and have extraction method and neural network classifier to visualize the state of a patient in its early stage whether or not it's traditional or abnormal. In this paper Neural Network algorithmic rule is enforced exploitation open supply and its performance is compared to different classification algorithms. It shows the most effective results with highest TP Rate and lowest FP Rate and just in case of properly classification. It provides the 96.04% result as compare to different classifiers. Fatma Taher, Naoufel Werghi, Hussain Al-Ahmad, Rachid Sammouda [7] conferred 2 segmentation ways, Hopfield Neural Network (HNN) and a Fuzzy C-Mean (FCM) clump algorithmic rule, for segmenting mucous secretion color pictures to notice the carcinoma in its early stages. The HNN and FCM ways square measure designed to classify the image of N pixels among M categories. during this study, they used one thousand mucous secretion color pictures to check each way, and HNN has shown a far better classification result than FCM. The HNN succeeded in extracting the nuclei and protoplasm regions. during this paper authors uses a rule primarily based thresholding classifier as a pre-processing step. Overall, the thresholding classifier has achieved an honest accuracy of ninety-eight with high worth of sensitivity and specificity of eighty-three and ninety-nine severally. T. Sowmiya, M. Gopi, M. New Begin, L. Thomas Robinson explained many aspects of knowledge mining procedures that are used for carcinoma prediction for the patients. data processing ideas is beneficial in carcinoma classification. They conjointly reviewed the aspects of hymenopteran colony optimisation (ACO) technique in data processing. Hymenopteran colony optimisation helps in increasing or decreasing the malady prediction worth of the diseases. This analysis different data processing and hymenopteran colony optimisation techniques for applicable rule generation and classifications on diseases. in addition to that provides basic framework for more improvement in diagnosing on carcinoma.

## RESULTS:

### 1) Input Image

At this stage, noise and other instabilities being present in the image are being suppressed being equivalent to the high frequency suppressions in the frequency domain.



FIG1: CT-Scan Image

### 1) histogram equalisation

Histogram equalization is being used to provide a desirable shape to the histogram and has been found to be a powerful technique because of it being simple and fast.

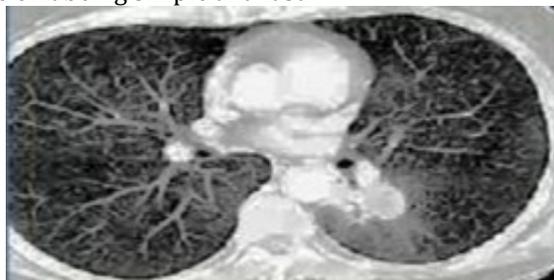


Fig 2

### 1) segmentation

With segmentation, images are being divided into regions having similar specifications such as colour, intensity and texture. Its output consists of segments which collectively covers the whole image.



Fig 3: High Contrast Image

**2) image filtering**

Image filtering is being used for smoothing, sharpening, removing noise, and edge detection



Figure 4: Less Distorted image

**3) dilation**

Involves the gradual enlargement of the boundary regions of the foreground pixel making them grow in size while making the holes present in the region smaller.



Figure 5: Image after dilation

**4) image Filling**

Flood-fill operation is being performed on binary and grayscale images by making use of in fill function.



Fig 6

**5) Feature extraction**

In order to make the input data less redundant, its data is being transformed into a reduced representation set of features and can be used to detect and isolate various desired portions or shapes of the image.



Fig 7

## 2) Neuro Fuzzy Classification

Fuzzy rule with some membership function is being used in order to diagnose if the cancer mass is being present inside the lung area. Adaptive neuro fuzzy classifier has been implemented by making use of various input parameters to produce final output. A combination of neural network and fuzzy logic is being used in adaptive neuro fuzzy inference system in order to get the advantages of both.

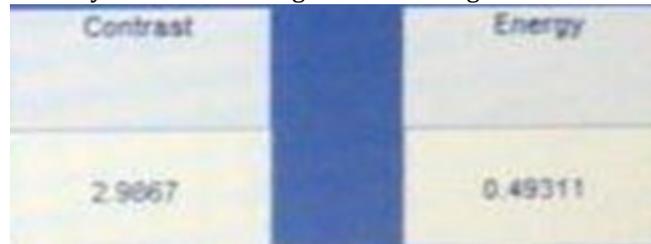


Fig 8

## CONCLUSION

In this paper, we have a tendency to propose to use image pre-process to discover the carcinoma symptoms at a really early stage. Some steps utilized in the method are histogram equalisation, Segmentation, Filtering, Dilation. By implementing these steps the nodules are detected then some valuable information is extracted. Then the get received are used for the classification of the affected stage. at that time, we have applied prediction model by applying that we have a tendency to predict from the obtained dataset from feature extraction to grasp how many people stricken by cancer or not. The cancer affected lungs region will be determined within the final output image. this method helps the radiologists and therefore the doctors by providing a lot of data and taking correct call for carcinoma patient in brief time with accuracy. Also, it finds its applications within the medical analysis moreover. Therefore, this technique costly pricey, less time overwhelming and simple to implement.

## REFERENCES

1. Cristobal G., Navarro. R., Space and frequency variant image enhancement based in Gabor representation, Pattern Recognition Letters, Elsevier, 1994.
2. Huang Q., Gao W., Cai W., Thresholding technique with adaptive window selection for uneven lighting image, Pattern Recognition Letters, Elsevier, 2004, 26, p. 801-808.
3. [https://en.wikipedia.org/wiki/Thresholding\\_\(image\\_processing\)](https://en.wikipedia.org/wiki/Thresholding_(image_processing)).
4. [en.wikipedia.org/wiki/Histogram equalization](https://en.wikipedia.org/wiki/Histogram_equalization)
5. R. Gonzalez and R. Woods *Digital Image Processing*, Addison-Wesley Publishing Company, 1992, pp 518 - 519, 549.
6. Anita Chaudhary, Sonit Sukhraj Singh, "Lung Cancer Detection on CT
7. Images Using Image Processing", International transaction on Computing Sciences.
8. Rachid Sammouda "Lung Cancer Detection Using Image Processing Techniques", Leonardo Electronic Journal of Practices and Technologies, (ISSN1583-1078), Issue 20, p. 147-158, January-June 2017.
9. Mr. Vijay A. Gajdhane, Deshpande, "Detection of Lung Cancer Stages on CT scan Images by Using Various Image Processing Techniques" Journal of Computer Engineering (IOSR-JCE), e-ISSN:2278-0661, p-ISSN: 2278-8727, Volume 16, Issue 5, Ver. III, Sep - Oct. 2014.
10. Ilya Levner, Hong Zhangm, "Classification driven Watershed segmentation ", IEEE TRANSACTIONS ON IMAGE PROCESSING VOL. 16, NO. 5, MAY 2007.
11. Nunes É.D.O., Pérez M.G., Medical Image Segmentation by Multilevel Thresholding Based on Histogram Difference, presented at 17th International Conference on Systems, Signals and Image Processing, 2010.