IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 6, June Month 2021

DOI 10.17148/IJARCCE.2021.106128

# LUNG CANCER DETECTION USING ARTIFICIAL NEURAL NETWORKS

**Mrs. Kavitha B.C [1], Pooja [2], Roopa B. P[3], Vanitha M. R[4], Veena A.H[5]**

Assistant Professor, Department of Electronics and Communication Engineering, BGS Institute of Technology, Mandya Karnataka, India[1]

BE Students, Department of Electronics and Communication Engineering, BGS Institute of Technology, Mandya, Karnataka, India[2,3,4,5]

**Abstract:** Lung cancer is cancer that start in the lungs. Cancer is a disease where cancerous cells grow out of manage, taking over normal cells and organs in the body. The early detection of lung cancer is the most useful way to decrease the mortality rate. In this document we contrast two methods, a modified Hopfield Neural Network (HNN) and a Fuzzy C-Mean (FCM) Clustering Algorithm, used in segmenting sputum color metaphors. The segmentation grades will be used as a base for a Computer Aided Diagnosis (CAD) system for early detection of lung cancer. The manual analysis of the sputum samples is time overriding, inaccurate and requires intensive qualified person to avoid diagnostic errors. Both methods are designed to classify the image of N pixels along with M classes or regions. Due to intensity variations in the background of the raw images, a pre-segmentation process is developed to normalize the segmentation process. In this learn, we used 1000sputum color metaphors to test both methods, and HNN has shown a better classification result than FCM; though the latter was quicker in converging.

**Keywords:** Lung Cancer recognition, Sputum Cells, Thresholding Technique, Image Segmentation, Hopfield Neural Network, Fuzzy C-Mean Clustering

## I. INTRODUCTION

Lung cancer is considered to be as the main reason of cancer death worldwide, and it is difficult to detect in its early stages because symptoms appear only at superior stages causing the mortality rate to be the maximum among all other types of cancer. More people pass on because of lung cancer than any other types of cancer such as: breast, colon, and prostate cancers. There is significant evidence indicating that the early detection of lung cancer will reduce the mortality rate. The most recent estimates according to the latest data provided by world health organization indicates that around 7.6 million deaths global each year because of this type of cancer. Furthermore, mortality from cancer are expected to continue rising, to become around 17 million worldwide in 2030.There are many techniques to diagnosis lung cancer, such as Chest Radiograph (x-ray), Magnetic Resonance Imaging (MRI scan) and Sputum Cytology. though, most of these techniques are costly and time consuming. In other words, most of these techniques are detecting the lung cancer in its higher stages, where the patient's chance of survival is very low. Therefore, there is a great need for a new skill to diagnose the lung cancer in its early stages. Image processing techniques provide a high-quality quality tool for improving the manual analysis. A number of medical researchers utilized the study of sputum cells for early detection of lung cancer, most current research relay on quantitative information, such as the size,shape and the ratio of the affected cells. For this reason we effort to use automatic diagnostic system for detecting lung cancer in its early stages based on the analysis of the sputum colour images. In order to formulate a law we have developed a thresholding technique for unsupervised segmentation of the sputum colour image to divide the images into some meaningful sub regions. Image segmentation has been used as the primary step in image classification and clustering. There are many algorithms which have been projected in other articles for medical image segmentation, such as histogram analysis, regional growth, edge detection and Adaptive Thresholding. A reconsider of such image segmentation techniques can be found in. Other authors have deliberate the use of colour information as the key susceptible factor for cell segmentation for lung cancer diagnosis. The testing of sputum images have been used in. for detecting tuberculosis; it consists of analyse sputum images for detecting bacilli. They used analysis techniques and feature extraction for the improvement of the images, such as edge detection, heuristic knowledge, region labelling and removing

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 6, June Month 2021

DOI 10.17148/IJARCCE.2021.106128

## II.METHODOLOGY

**Data Collection:**

Data is the most significance part when you work on prediction systems. It plays a very vital role your whole project i.e., you system depends on that data. So selection of data is the first and the critical step which should be performed properly, for our project we got the data from the kaggle website. No additional in a row on the CT scan metaphors are obtained, thus the medical eminence and the background of the images are left as the task to be analysed in order to choose the suitable images to be used for the project. The dataset we choose wad particular based on the various factors and constraints we were going to take under the deliberation for our prediction system.

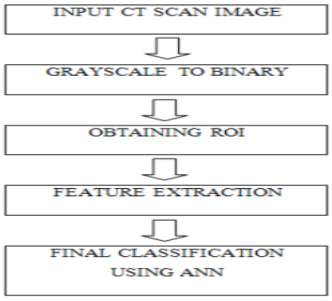



**Figure 1:** Block diagram of lung cancer Detection using fuzzy clustering

Image pre-processing

The main rationale of pre-processing is to lessen the amount of noise with minimum mean square worth for the image. Image pre-processing is a scheme for enhancing the quality of the original image where after enhancing the image quality is far better when compared with that of the inventive image. The obtained original CT image is obtained and it is pre-processed using wiener riddle as it is better for noise exclusion compared to other filters.

- Noise Removal: Many times image is corrupted by noise as the image is collected from many other sources thus initially noise is removed by making use of median filter which contains highest signal to noise ratio

Image Enhancement: The quality of the image can be enhanced in order to increase the intensity and thus can be proper for feature extraction.

- Grey scale to Binary conversion: Image binarization is a subclass of image segmentation as it divides an image into segments based on the value of pixels to a threshold value. The simplest of all the thresholding techniques is partition of the image using a single global threshold. With the use of this threshold value the upper half intensities is converted in full white and lower half into black.
- Removal of erroneous part :Sometimes some of the erroneous part is present which is not required in further process and thus need to be removed

Image Segmentation

The image segmentation is obtained by form a gray level co-occurrence matrix of the figure this template is used for attribute withdrawal The gray co matrix function creates a gray-level co-occurrence matrix by calculating how often a pixel with the intensity i.e. gray-level value i occurs in a comprehensive spatial affiliation to a pixel with the value j. The spatial relationship is defined as the pixel of attention and the pixel to accurate i.e. horizontally nearest, you can specify other spatial relationships between the two pixel. Each constituent i.e. i,j in the significant is simply the sum of the number of times that the pixel with value I occurred in the specified spatial affiliation to a pixel with rate j in the contribution image.

Clustering is a purpose of gathering set of objects in the same group is similar to each other than those set of objects in other group. Clustering can be done in two ways: solid clustering and fuzzy clustering. solid clustering type of data set belongs to one cluster. So it has some disadvantages like finite low disparity spatial resolution, noise, overlapping intensities, and non-uniform intensity which reduce the efficiency of assorted hard clustering methods in segmentation process. while in fuzzy clustering method each data can belong to more than one cluster. Clustering is a procedure of allotting information focuses to groups with the end goal that things in same group will be as analogous as could reasonably be expected and things in various group will be as exclusive as could be allowed.

**The architecture of an artificial neural network:**

To understand the concept of the architecture of an artificial neural network, we have to recognize what a neural network consists of. In order to identify a neural network that consists of a large number of artificial neurons, which are termed units arranged in a order of layers. Lets us glance at various types of layers available in an artificial neural network.

IJARCCE

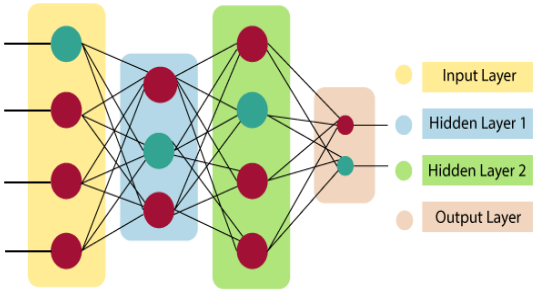
ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 6, June Month 2021

DOI 10.17148/IJARCCE.2021.106128

Artificial Neural Network mainly consists of three layers:




**Input Layer:**
As the name suggests, it accepts inputs in several unlike configures provided by the programmer.
**Hidden Layer:**
The hidden layer presents in-between input and productivity layers. It performs all the calculations to find hidden skin tone and patterns.
**Output Layer:**
The input goes from side to side a series of transformations using the hidden layer, which finally outcome in output that is conveyed using this layer.
The artificial neural network takes input and computes the slanted sum of the inputs and includes a bias. This computation is represented in the form of a convey function.

$$\sum_{i=1}^{n} Wi * Xi + b$$

It determines weighted total is accepted as an input to an activation function to produce the output. commencement functions choose whether a node should fire or not. Only those who are fired make it to the productivity layer. There are distinctive activation functions accessible that can be applied upon the sort of task we are performing.

## FUZZY CLUSTERING

Clustering is the process of dividing the data into homogenous regions based on the comparison of objects; information that is logically similar physically is stored together, in order to increase the effectiveness in the database system and to minimize the number of disk access. The progression of clustering is to assign the q feature vectors into K clusters, for each kth cluster Ck is its center. Fuzzy Clustering has been used in many fields like pattern gratitude and Fuzzy identification. A variety of Fuzzy clustering methods have been projected and most of them are based upon distance criteria. The most widely used algorithm is the Fuzzy C-Mean algorithm (FCM), it uses mutual reserve to compute fuzzy weights. This algorithm has as contribution a predefined number of clusters, which is the k from its name. Means stands for an average location of all the members of exacting cluster and the output is a partitioning of k cluster on a set of objects. The objective of the FCM cluster is to condense the total weighted mean square error:

$$J = (W^{q^k}, C^{(k)}) = \sum_{(q=1,Q)} \sum_{(k=1,K)} (W_{qk})^p \parallel x^{(q)} - c^{(k)} \parallel^2$$

The FCM allows each feature vector to belong to numerous clusters with various fuzzy membership values. Then the final cataloguing will be according to the highest weight of the feature vector over all clusters. The complete algorithm: Input: Vectors of objects, each object represent s magnitude, where v = {v1,v2,……,vn} in our case it will be an image pixels, each pixel has three dimensions RGB, K = number of clusters. Output = a set of K clusters which reduce the sum of distance error. Algorithm steps:
1. Initialize random weight for each pixel, it uses fuzzy weighting with positive weights {W qk } stuck between [0,1]
2. Standardize the initial weights for each qth feature vector in excess of all K clusters via

$$W_{qk} / \sum_{r=1,K} W_{qr}$$

3. homogenize the weights over k = 1,…,K for each q to obtain W qk , via

$$W_{qk} = W_{qk} / \sum_{(r=1,Q)} W_{rk}, q = 1,\ldots,Q$$

4.calculate new centroids C(k) , k = 1,….,K via

$$C^{(k)} = \sum_{(q=1,Q)} W_{qk} X^{(q)}, k = 1,\ldots,K$$

5.Update the weights {W qk } via

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 6, June Month 2021

DOI 10.17148/IJARCCE.2021.106128

$$W_{qk}=(1/\|x^q-c^k\|^2)^{1/(p-1)}/\sum_{(r=1,K)}(1/\|x^q-c^r\|^2)^{1/(p-1)},k=1,..K,q=1,...Q$$

6. If there is change in the input, replicate from step 3, else terminate.

7. Assign each pixel to a cluster based on the highest weight.

We applied the FCM clustering algorithm with the requirement mentioned above to one thousand sputum color images and maintain the result for further dealing out in the following steps. Our algorithm segments the images into nuclei, cytoplasm regions and clear background, however, the FCM is not susceptible to intensity variation, therefore, the cytoplasm regions are detected as one cluster when we rigid the cluster figure to three, four, five and six. Moreover, FCM failed in detecting the nuclei; it detected only part of it. By testing, the FCM algorithm takes fewer than 50 iterations to accomplish the obligatory results in 10 seconds on average.

## V. RESULTS

After applying the Fuzzy logic, extracted and filtered input images is mapped with trained images. It is collection of series of fixed size filters as known as kernels, which are used to perform the operation on images data yield the feature maps. Then trained images are moved on to the classification part where each image is classified as either normal or malignant. Output of each image has been displayed with more accuracy as whether cell image is normal or malignant.

**Non- Cancerous Lungs Image**

**Cancerous Lungs Image**

## VI. CONCLUSION

By using ANNs to slow task of manually extracting features can be eliminated. We will design a computer aided diagnosis (CAD) system to detect cancerous cell in CT Scan images. This CAD system comprises four stages: pre-processing, evaluating region of significance, feature taking out, final categorization using ANN. The pre-processing stage consists of various image enhancement techniques (contrast enhancement, thresholding, noise removal) to improve the visibility of tumors in CT scan images. We had taken out the ROI and GLCM matrix is formed from the enhanced image feature is extracted and give them asinput to the next stage. We will use artificial neural networks in the classification stage. After training is completed testing will be done in which the same steps will be applied to the testing image and the parameters of it will be compared to the matrix formed after training. A distance will be calculated with each and every row of the matrix this distance will give us the difference and the minimum distance will be the minimum difference and considered as the best match. And final result will be showed that whether the image is cancerous or not.

## VII.REFERENCES

1. Disha Sharma, Gangadeep Jindal, “Identifying Lung Cancer Using Image Processing Technique”, International Conference of Computational Techniques and Artificial Intelligence, 2011
2. AtiyehHashemi, Abdol Hamid Pilevar, Reza Rafeh, “Mass Detection in Lung CT Images Using Region Growing Segmentation and Decision Making Based on Fuzzy Inference System and Artificial Neural Network”, I.J. Image, Graphics and Signal Processing, 6, pp: 16-24, 2013
3. Anam Tariq, M. Usman Akram and M. YounusJaved, “Lung Nodule Detection in CT Images using Neuro Fuzzy Classifier”, Fourth International Workshop on Computational Intelligence in Medical Imaging (CIMI), pp:49-53, 2013
4. FatmaTaher, NaoufelWerghi and Hussain Al-Ahmad, “Bayesian Classification and Artificial Neural Network Methods for Lung Cancer Early Diagnosis”, IEEE, pp: 773-776, 2012
5. Varalakshmi. K,” Clasification of Lung Cancer Nodules using a Hybrid Aproach,” Proceeding in the Journal of Emerging Trends in Computing and Information Sciences, Vol. 4, No. 1 ISSN 2079-8407, Jan 2013
6. S.K. VijaiAnand, “Segmentation coupled Textural Feature Classification for Lung Tumor Prediction”, ICCCCT, pp: 518-524, 2010.
7. JIA Tong, ZHAO Da-Zhe, YANG Jin-Zhu, WANG Xu, “Automated Detection of Pulmonary Nodules in HRCT Images”, IEEE, 2007.