International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2019 IJSRCSEIT | Volume 5 | Issue 2 | ISSN : 2456-3307
DOI : https://doi.org/10.32628/CSEIT195287

Classification of malignant and Benign Lung Using Probabilistic Neural Network

Asmitha Shree R[1], Sajitha M[2], Subha S[2]

[1]Assistant Professor, Department of Computer Science and Engineering, Sri Krishna College of Technology Coimbatore, Tamil Nadu, India

[2]Department of Computer Science and Engineering, Sri Krishna College of Technology Coimbatore, Tamil Nadu, India

ABSTRACT

Lung Cancer is considered as one of the deadliest diseases among other lung disorders and cancer types and is the leading cause of cancer deaths worldwide. Lung cancer is a curable disease if detected in its early stages that makes up 13% of all cancer diagnoses and 27% of all cancer deaths. The objective of this paper is mainly focused on categorizing the patients Computed Tomography (CT) lung images as normal or abnormal. The images are subjected to segmentation to focus on detecting the cancerous region to classify. Effective feature selection and feature extraction is made by applying Watershed Transform and Principal Component Analysis. The emphasis is on the feature extraction stage to yield a better classification performance. The classification of CT images as benign or malignant is done using Machine Learning based Neural Network.

Keywords : Watershed transform, Gray Level Co-Occurrence Matrix, Probabilistic neural network.

I. INTRODUCTION

Cancer is disease that begins in cells which are the basic building blocks of a body. Though there are various types of cancers, all starts with the cells growing out of control. The type and stage of cancer can be found by determining where the abnormal growth occurs in the body and the type of cells that start to grow abnormally. The type of cancers often found are endometrial cancer, skin cancer, lung cancer, ovarian cancer and breast cancer. Among both men and women lung cancer by far is the leading cause of cancer deaths. The abnormal cells in the lung cluster together to form a tumor, unlike normal cells cancer cells grow uncontrollably, destroying the healthy lung cells around them. When the tumor cells in the lung starts to spread, it will stop the body parts from functioning properly. Generally Lung cancer doesn't cause signs and symptoms in its initial stages. Signs and symptoms of lung cancer occur when the cancer has spread to all other parts of the body. At this point, treating lung cancer is harder and complicated. Lung cancer often found in people who smoke..Also non-smokers can be caused by exposure to radon, air pollution, or other factors. The major types of lung cancer are non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC). About 10% to 15% of people with lung cancer will have small cell lung cancer (SCLC). At the initial stage the tumor cells aggressive and quickly starts to spread all over the body. Small cell lung cancer occurs only in heavy smokers. Non-Small-Cell Lung Cancer (NSCLC) is the most common type of cancer occurs for 80 to 85% of people. Lung cancer is common, compared to other cancers. Signs and symptoms caught at a very early

stage, it often can be treated effectively with surgery and even cured. Digital image processing has been deployed to computed tomography (CT) for improving accuracy in diagnosis. This provides a way in designing the computer-aided detection systems that can detect the benign and malignant tumors of the lungs. They are proved to be an important tool for medical science. For lung cancer detection one of the most important step is screening. Screening is the process used for identification of the nodules which is a white color spot present on lungs that is visible on an X-ray or Computed Tomography (CT) scans Images. Lung nodules will appear in X-ray or CT scan image only if its diameter is about 1 cm. A nodule may be either a benign or a mass. A nodule that is 3 cm or less in diameter is called Pulmonary or benign nodule which are non-cancerous. Another type of nodule whose size is larger than 3 cm is in diameter is called a lung mass which is more likely to be cancerous and needs to be detected as early as possible.

Detection of lung nodules in CT images using clustering techniques was performed earlier which used clustering techniques as a non-supervised learning step, and later data classification was performed from the cluster features obtained. K-Means clustering is used to categorize the items into groups from the given data set with certain features, and values[1]. Various clustering algorithms achieved different accuracy like the C-Means Clustering (FPCM) which achieved 92.50% accuracy, Possibility C-Means Clustering Algorithm (PCM) achieved 91.00% accuracy, Modified FCM achieved 89.50% accuracy and standard FCM showed least accuracy of 86.03%.

The random forest (RF) algorithm form a family of classification method which rely on the combination of several decision trees .The tree-based components are grown from a certain amount of randomness and the limitation of the this algorithm is that a large number of trees may make the algorithm slow for real-time prediction.

Support vector machines are supervised learning models with corresponding learning algorithms that perform data analysis and pattern recognition, used for classification was used for classifying the tumor as Benign or Malignant in the lung CT images. The basic SVM takes a collection of input data and for each input, it predicts which of the two classes forms the input, making it a non-probabilistic binary linear classifier.

The proposed system reads JPEG converted Dicom Format CT scan images of lungs and determines the presence of any abnormality through image processing techniques. Later it calculates certain features of the abnormality and feeds them into a Neural Network trained system to detect if the abnormality is cancerous.

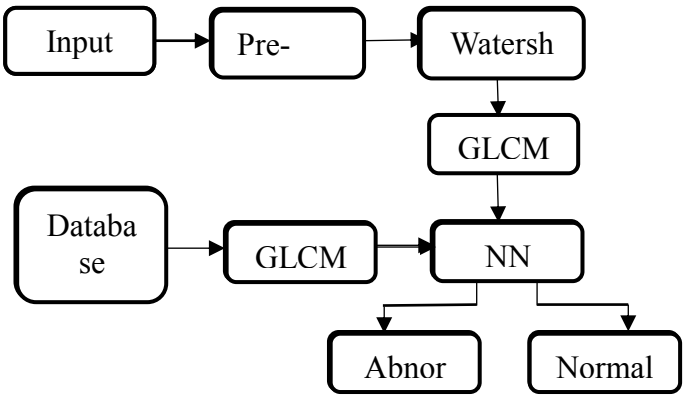

Fig.1 : A Flowchart Showing the proposed method of Cancer Cell Detection.

II. SYSTEM MODEL

Image Database

The development of computer-aided diagnostic (CAD) systems for lung nodule detection, classification, and quantitative assessment is facilitated through a well-characterized repository of computed tomography (CT) scans. The Lung Image Database Consortium (LIDC) completed such a database, establishing a publicly available reference

for the medical imaging research community with easy access. The medical images are in the dicom formats which have to be converted into readable format.

Image Pre-processing

Extraction of images is performed accurately when the image is in binary form rather than in their default Red Green Blue (RGB) form. Because RGB images are logically represented as a combination of 3 matrices, each having a dimension of 1 × n, where n is the number of pixels. These data are complex to process and therefore should be converted into a grayscale image which are logically a matrix with each pixel represented as a discrete level out of 0 to 255 [8 bits]. The dimension of the matrix in the gray scale image matches the dimension of the image represented in terms of pixels[8]. Further Image Smoothing, image resizing, image enhancement, image conversion into grayscale are the techniques are used in pre-processing. In proposed model for smoothing of image the median filter is used to remove the noise from the given image. Noise in image suppresses the important information and hence it should be removed. Median filter give the best result because it removes the noise without blurring the image whereas other techniques remove the noise and adds blurriness to the output image [4][12].

Fig.2: A Grayscale converted image.

Image Segmentation

Segmentation divides the input image into many segments. The main purpose behind the segmentation is simplifying the representation into more meaningful and easy for understanding. Watershed transform segmentation separates the foreground and background objects and separates them with different color for better visibility. The gradient magnitude is used to pre process a gray-scale image prior to using the watershed transform for segmentation. The gradient magnitude image has high pixel values along the object edges and low pixel values everywhere else in the image. Watershed transform would result in watershed ridge lines along the object edges. The topological gradient provides a global analysis of the image which removes unwanted contours caused due to the noise added to a image that can be significantly reduced by this approach[5][6][7].

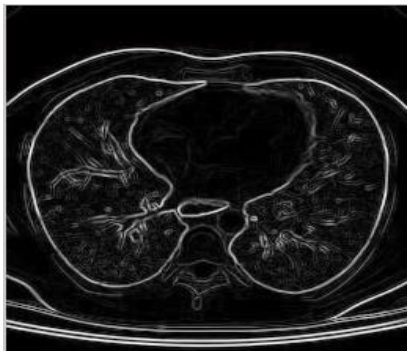

Fig.3: Gradient magnitude of Fig.2

Feature extraction

Gray Level Co-Occurrence Matrix (GLCM) is a common feature extraction method to recognize the shape and size of nodule present in image. In GLCM, first step is computing the co-occurrence matrix and the second step is calculating texture feature based on the co-occurrence matrix. GLCM texture considers the relation between two neighbouring pixels and use the contents of the GLCM to give a measure of the variation in intensity at the pixel of interest[13]. The main features to be extracted are Energy, Entropy, contrast, Homogeneity and correlation of the input image. Energy defines the measure of the extent of pixel pair repetitions or uniformity of an image. Entropy is the measure of randomness that is used to characterize the texture of the image and its value

will be maximum when all the elements of the co-occurrence matrix are the same. The contrast is a measure of intensity of a pixel and its neighbour over the image and is determined by the difference in the colour and brightness of the objects within the same field of view. Homogeneity obtains the measures of the closeness of the distribution of the GLCM elements to the GLCM diagonal. The Correlation texture measures the linear dependency of grey levels on neighbouring pixels[14]. The range of the values of these parameters are used to train the neural network and in turn for decision making. However, feeding more number of features into the neural network decreases efficiency. So features in the image are selected on interest. The value of these features are computed and fed into the neural network.

Classification

Initially nodules in lungs are small in size which then grows and is harmful for human body. The criteria for determining the cancer is nodule size >3cm in diameter. PNN consists of an input layer, which represents the feature vector. The input layer is interconnected with the hidden layer, which consists of the example vectors (the training set for the PNN). The example vector serves as the weights as applied to the input layer. Finally, an output layer represents each of the possible classes for which the input image can be classified.[13] After training, the model is used to make decision on the test CT scan image as malignant or benign.

III. RESULT

PNN is much faster than multilayer perceptron networks and gives more accurate result than multilayer perceptron networks. The performance of the proposed method is analysed by measuring sensitivity, specificity and accuracy and the values obtained are 100%, 75% and 96.67% respectively.

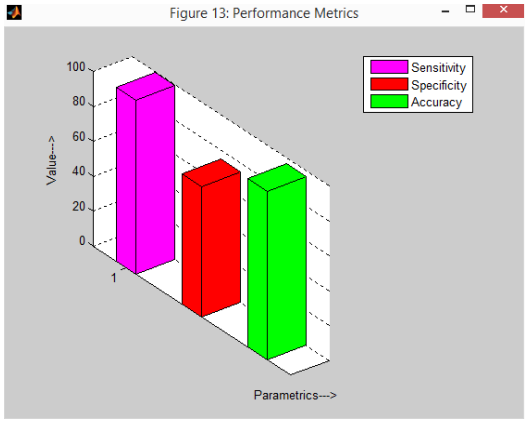

Fig.4: Performance metrics of Proposed System

IV. CONCLUSION

Cancer has been a potential threat to human eco system. The solution is early prediction of cancer and also its probabilities for an individual in future which will help to avoid cancer and decrease the death rate. Image processing techniques play a significant role for image improvement in earlier detection and treatment stages in many medical areas, where the time factor is vital to discover the abnormality issues in CT images. The CT image is processed and cancer cell is identified for that the high boost filter and watershed segmentation gives best results for pre-processing stage. From the region features are extracted to identify the size of the cancer cell and in turn the stage of the cancer is identified.

V. REFERENCES

[1] ANIKETBOMBALE, C.G.PATIL," Segmentation of Lung Nodule in CT Data Using K-Mean Clustering" International Journal Of Electrical, Electronics And Data Communication, ISSN: 2320-2084 Volume-5, Issue-2, Feb-2017.

[2] P.B.Sangamithraa, S.Govindaraju, "Lung Tumour Detection and Classification using EK-Mean Clustering," IEEE WiSPNET 2016 conference.

[3] Netto SMB, Bandeira Diniz JO, Silva AC, de Paiva AC, Nunes RA, Gattass M,"Modified Quality Threshold Clustering for Temporal Analysis and Classification of Lung Lesions",IEEE Trans Image

Process 2019 Apr;28(4):1813-1823. doi: 10.1109/TIP.2018.2878954. Epub 2018 Oct .

[4] Jue Jiang, Yu-chi Hu1, Chia-Ju Liu, Darragh Halpenny, Matthew D. Hellmann,"Connected Feature Streams For Automatic Lung Tumor Segmentation From CT Images", DOI 10.1109/TMI.2018.2857800, IEEE Transactions on Medical Imaging.

[5] G.Vijaya,A.Suhasini,"An Adaptive Preprocessing of Lung CT Images with Various Filters for Better Enhancement", DOI: 10.5829/idosi.ajcr.2014.7.3.84231, ISSN 1995-8943,IDOSI Publications, 2014.

[6] Niket Amoda, Ramesh K Kulkarni,"Image Segmentation and Detection using Watershed Transform and Region Based Image" International Journal of Emerging Trends & Technology in Computer Science,ISSN 2278-6856 Volume 2, Issue 2, March – April 2013.

[7] Pinaki Pratim Acharjya, Dibyendu Ghoshal,"An Overview on Watershed Transform and Its Consequences" International Journal of Engineering and Innovative Technology,ISSN: 2277-3754 Volume 1, Issue 5, May.

[8] Lamia Jaafar Belaid,Walid Mourou2 "Image Segmentation:A Watershed Transformation Algorithm ", Image Anal Stereol 2009;28:93-102, March 27, 2009.

[9] C. Saravanan,"Color Image to Grayscale Image Conversion" International Conference on Computer Engineering and Applications,DOI 10.1109/ICCEA.2010.192, IEEE Trans 2010.

[10] Niket Amoda, Ramesh K Kulkarni , "Efficient Image Segmentation Using Watershed Transform" ISSN : 0976-8491 IJCST Vol. 4, Issue 2, April - June 2013.

[11] Pinaki Pratim Acharjya, Dibyendu Ghoshal "An Overview on Watershed Transform and Its Consequences" International Journal of Engineering and Innovative Technology (IJEIT),ISSN: 2277-3754 Volume 1, Issue 5, May 2012 .

[12] Mokhled S. AL-TARAWNEH, "Lung Cancer Detection Using Image Processing Techniques," Leonardo Electronic Journal of Practices and Technologies, January-June 2012.

[13] G. Vijaya,A. Suhasini "An Adaptive Preprocessing of Lung CT Images with Various Filters for Better Enhancement" Academic Journal of Cancer Research,ISSN 1995-8943 IDOSI Publications, 2014.

[14] P. Mohanaiah, P. Sathyanarayana,L. GuruKumar"Image Texture Feature Extraction Using GLCM Approach",International Journal of Scientific and Research Publications, ISSN 2250-3153 ,Volume 3, Issue 5, May 2013.

[15] Anuja Kadam, Shirish Kulkarni,Nikita Naik "Segmentation of Lung Tumor Using GLCM Technique"International Journal of Engineering Research and Development,Volume 12, Issue 7 (July 2016).

[16] DONALD F. SPECHT, "Probabilistic Neural Networks", Neural Networks, Vol. 3. pp. 109 118, 1990.

[17] R.Pushpalatha, I.Anette Regina, " Cancer Detection through CT scan Images Using Gabor Filter Method and Classification Using Pnn", International Journal Of Merging Technology And Advanced Research In Computing , ISSN: 2320-1363.

Cite this article as :

Asmitha Shree R, Sajitha M, Subha S, "Classification of malignant and Benign Lung Using Probabilistic Neural Network", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 5 Issue 2, pp. 307-311, March-April 2019. Available at doi :
https://doi.org/10.32628/CSEIT195287
Journal URL : http://ijsrcseit.com/CSEIT195287