*International Journal For Technological Research In Engineering*
*Volume 7, Issue 9, May-2020* *ISSN (Online): 2347 - 4718*

# ANALYSIS ON TECHNIQUES FOR LUNG CANCER CLASSIFICATION USING CT IMAGES

Keerti Shreedhar Hasyagar[1], Maya B S[2]
[1]B.E (Student), [2]Assistant Professor
Department of CS&E, BIT, Bangalore

***ABSTRACT: Lung cancer is one of the causes of cancer deaths. Lung cancer is a malignant lung tumor characterized by uncontrolled cell growth in tissues of the lung. Although, CT scan imaging is best imaging technique in medical field, it is difficult for doctors to interpret and identify the cancer from CT scan images. Therefore, computer aided diagnosis can be helpful for doctors to identify the cancerous cells accurately. Many computer aided techniques using image processing and machine learning has been researched and implemented. In this research, various lung cancer classification techniques are studied where the CT images are preprocessed using different filters like Median, Gabor, Gaussian etc., features are extracted and used for classification using the different classifiers such as SVM, CNN, Random Forest etc.***
***KEYWORDS: Lung Cancer, Computed Tomography (CT), Neural Networks, Feature Extraction, Classifiers***

## I. INTRODUCTION

Lung cancer, is additionally perceived as carcinoma of the lung, is a dangerous lung described by unlimited cell development in tissues of the lung. If not treated, this development can spread past the lung by procedure of metastasis into close-by tissue and different parts of the body. Most tumors that begin in the lung, known as essential lung malignancies, are carcinoma that got from epithelial cells. The significant sorts are little cell lung carcinoma (SCLC) and non-little cell lung carcinoma (NSCLC). The regular side effects are hack or hacking up blood, weight reduction, shortness of breath, and mid-section torments. The larger part (85–90%) of instances of lung disease are because of long haul introduction to tobacco smoke. Around 15–25% of cases happen in individuals who have absolutely not smoked. These cases are regularly created by a mix of hereditary components and presentation to asbestos and different types of air contamination, including second-hand smoke. Lung cancer can be dissected taking into account mid- section radio charts and registered tomography (CT) examines. The determination is checked by biopsy which is normally performed by bronchoscopy or with CT direction. Treatment results rely on upon the kind of cancer, the stage, and the individual's general wellbeing is measured by execution grade. Normal medications incorporate chemotherapy, and radiotherapy. NSCLC is here and there taking care of with surgery, though SCLC more often than not reacts better to chemotherapy and radiotherapy. By and large, 17% of individuals in the United States determined to have lung growth. According to American Association of Central Cancer Registries (NNACCR) report, 234,030 peoples are influenced by lung cancer in USA in the year 2018. Not only men (14%), 13% of women in USA also were influenced by lung cancer. Furthermore, worldwide analysis 154,050 out of 234,030 analyzed as fatal. Screening methodology has helped prediction of lung cancer, but an earlier detection of cancer and accuracy of cancer detection is difficult to maintain. So, computer aided automatic detection (CAD) process has to be applied to the clinical center for developing an effective cancer prediction system using an optimized and intelligent technique. The optimized lung image processing methods are utilized for the examination of inner details of body, restore the details, extract valuable information, and also create a knowledgeable system for the diagnosis of lung cancer.

## II. LITERATURE REVIEW

[1] Proposed a method to classify the lung cancer that first segments the region of interest (lung) and then analyses the separately obtained area for nodule detection in order to examine the disease. Initially, the segmentation of the CT images has been carried out by using K- Means clustering method. To the clustered result, EK-Mean clustering is applied. Further the features like entropy, Contrast, Correlation, Homogeneity and the area are extracted from the tumorous part of Fuzzy Ek- Means segmented Image. For feature extraction, statistic method called Gray Level Co-occurrence Matrix (GLCM). Classification is done by using the supervised neural network called the Back-Propagation Network (BPN). Results of the classification gives, whether the CT Image is a normal Image or cancerous. Back Propagation Neural network produces an accuracy of about 90.87%.
[2] Presents an approach which uses a Convolutional Neural Network (CNNs) to classify tumors seen in lung cancer screening computed tomography scans as malignant or benign. CNNs have special properties such as spatial invariance and allow for multiple feature extraction. When such layers are cascaded, leading to Deep CNNs, it has been shown widely that the accuracy of prediction increases dramatically. In this work, CNN is used for the analysis of CT scans with tumors, using domain knowledge from both medicine and neural networks. The results show that the accuracy of classification for this network performs better than both the traditional neural networks, and existing CNNs built for image classification purposes. The proposed CanNet a form of CNN achieves an accuracy of 76%.
[3] Aims to improve accuracy, sensitivity and specificity of early detection of lung cancer through a combination of image processing techniques and data mining. The Lung

  

Image Database Consortium dataset (LIDC) has been used for training and testing purpose for this study. The Computed Tomography (CT) scan image of the lungs is pre-processed and the Region of Interest (ROI) segmented, retained and compressed using a DWT (Discrete Waveform Transform) technique. The resulting ROI image is decomposed into four sub frequencies, bands LL, HL, LH, and HH. Again, the LL sub frequency is decomposed into four sub-bands, applying a 2- level DWT to the ROI based image. Further, features such as entropy, co-relation, energy, variance and homogeneity are extracted from the 2-level DWT images using a GLCM (Gray level Co-occurrence Matrix) with classification effected by means of an SVM (Support Vector Machine). Classification identifies whether the CT image is normal or cancerous. The classifier achieves an accuracy of 95.16%, sensitivity of 98.21% and specificity of 78.69%.

[4] Evaluates the various computer-aided techniques, analyzes the current best technique and finds out their limitation and drawbacks and finally proposes the new model with improvements in the current best model. The system makes use of Median filter and Gaussian filter in pre-processing stage. After preprocessing the processed image is segmented using watershed segmentation. This gives the image with cancer nodules marked. Features like area, perimeter, eccentricity, Centroid, Diameter and pixel Mean Intensity have been extracted in feature extraction stage for the detected cancer nodules. Classification of cancer nodule is performed using Support Vector Machine. Proposed model detects the cancer with 92% accuracy.

[5] Processes CT images using image processing techniques. The texture features along with statistical features are extracted and supplied various extracted features to classifiers. The seven different classifiers known as k-nearest neighbors classifier, support vector machine classifier, decision tree classifier, multinomial naive Bayes classifier, stochastic gradient descent classifier, random forest classifier, and multi-layer perceptron (MLP) classifier are used for classification and the results are compared. Dataset of 15750 clinical images consisting of both 6910 benign and 8840 malignant lung cancer related images is used to train and test these classifiers. Among different classifiers analyzed, kNN achieves an accuracy of 86.21%, SVM 57.47%, Decision Tree 81.77%, Multinomial Naïve Bayes 57%, Stochastic Gradient Descent 59.11%, Random Forest 83.17%, and Multi-Layer Perceptron which achieves the highest accuracy of 88.55%.

[6] Based on CT technology consists of several steps, such as image acquisition, preprocessing, feature extraction, and classification. In the preprocessing stage, RGB images are converted to grayscale images, the median filter and the Wiener filter are used to uproot noises, Otsu thresholding method is applied to convert CT images free from noise to binary images, and a MATLAB function is used to extract body region from binary images. In the feature extraction stage, features, like Contrast, Correlation, Energy, Homogeneity, are extracted through statistic method Gray Level Co-occurrence Matrix (GLCM). In the final stage, extracted features, together with Support Vector Machine (SVM) and Back Propagation Neural Network (BPNN), are used to identify lung cancer from CT images. The performance of the proposed system shows satisfactory results of 96.32% accuracy on SVM and 83.07% accuracy on BPNN respectively.

[7] Computed Tomography (CT Scans) of lungs of the patients from Lung Image Database Consortium (LIDC) is used as input data for image processing. A total of 910 images were taken from LIDC as the dataset for the implementation. 257 images were labeled normal, 331 as benign and 322 as malignant. In pre-processing stage conversion of RGB image to gray-scale image takes place because RGB images are too complex to process. Gray-scale image is further converted to Binary image. After Image Processing, the input images become more efficient and refined. These are input for the Convolution Neural Network. Convolution Filtering, Max Pooling filtering are steps in CNN which train the data to predict whether the lung image is cancerous (malignant) or non-cancerous (benign). An overall accuracy of 94.34% is obtained from proposed system.

[8] Proposes a region based Fuzzy C-Means Clustering (FCM) technique for segmenting the lung cancer region and the Support Vector Machine (SVM) based classification for diagnosing the cancer stage, which helps in clinical practice in significant way to increase the morality rate. Moreover, the proposed ECM-CSD (Efficient Classification Model for Cancer Stage Diagnosis) uses Computed Tomography (CT) lung images for processing, since it poses higher imaging sensitivity, resolution with good isotopic acquisition in lung nodule identification. With those images, the pre-processing has been made with Gaussian Filter for smoothing and Gabor Filter for enhancement. Following, based on the extracted image features, the effective segmentation of lung nodules is performed using the FCM based clustering. And the stages of cancer are identified based on the SVM classification technique. The proposed SVM classifier 93% accuracy is achieved for 70 numbers of training samples.

[9] Presents a computer-aided classification method in computerized tomography images of lungs. In the proposed system, MATLAB has been used for implementing all the procedures. The various stages involved include image acquisition, image preprocessing, segmentation, feature extraction and support vector machine (SVM) classification. First, the DICOM format lung CT image is passed as input which undergoes preprocessing. Then, a threshold value is calculated, and image is segmented into left lung and right lung. After that 33 features of each segmented lung are taken and passed as input to the SVM. Finally, the image is classified as cancerous or non-cancerous based on the training data. The SVM classifier achieves an accuracy of 86.25% for the given set of images.

[10] deals with improvement of the quality of lung image and diagnosis of lung cancer by reducing misclassification. The lung CT images are collected from Cancer imaging Archive (CIA) dataset, noise present in the images are eliminated by applying weighted mean histogram equalization approach which successfully removes noise from image, also enhancing the quality of the image, using improved profuse clustering technique (IPCT) for

  

segmenting the affected region. Various spectral features are derived from the affected region. These are examined by applying deep learning instantaneously trained neural network for predicting lung cancer. The system ensures that 98.42% of accuracy with minimum classification error 0.038.

Table 1: Related works comparison

| Sl No. | Authors | Techniques | Accuracy |
|---|---|---|---|
| 1 | P.B. Sangamithra and S.Govindaraju | Median Filter, Wiener Filter, EK-Mean Clustering and Backpropagation Network (BPN) | 90.87% |
| 2 | Prajwal Rao, Nishal Ancelette Pereira and Raghuram Srinivasan | CanNet (A form of CNN) | 76% |
| 3 | Deep Prakash Kaucha, P.W.C. Prasad, Abeer Alsadoon, A.Elchouemi and Sasikumaran Sreedharan | Discrete Wavelet Transform (DWT) and Support Vector Machine (SVM) | 95.16% |
| 4 | P.W.C. Prasad, Abeer Alsadoon, A.K. Singh and A. Elchouemi, | Gaussian Filter, Median Filter and Support Vector Machine (SVM) | 92% |
| 5 | Gur Amrit Pal Singh and P. K. Gupta | k-Nearest Neighbors (kNN), Support Vector Machine (SVM), Decision Tree classifier, Multinomial naive Bayes classifier, Stochastic Gradient Descent, Random Forest classifier, and Multi-Layer Perceptron (MLP) | 86.21% 57.47% 81.77% 57% 59.11% 83.17% 88.55% |
| 6 | Lingling Li, Yuan Wu, Yi Yang, Lian Li and Bin Wu | Median Filter, Wiener Filter, Support Vector Machine (SVM) and Backpropagation Neural Network (BPNN) | 96.32% 83.07% |
| 7 | Rohit Y. Bhalerao, Harsh P. Jani, Rachana K. Gaitonde and Vinit Raut | Convolution Neural Network (CNN) | 94.34% |
| 8 | M. S. Kavitha, J. Shanthini and R. Sabitha | Gabor Filter, Gaussian Filter, Fuzzy C Means, Support Vector Machine (SVM) | 93% |
| 9 | Chethan Dev, Kripa Kumar, Arjun Palathil, T. Anjali and Vinitha Panicker | Support Vector Machine (SVM) | 86.25% |
| 10 | P. Mohamed Shakeel, M.A. Burhanuddin and Mohamad Ishak Desa | Weighted Mean Histogram Normalization, Improved Profuse Clustering Technique (IPCT), Deep learning Instantaneously Trained Neural Network (DITNN) | 98.42% |

## III. CONCLUSION

Lung cancer is one of the causes of cancer deaths. It is difficult to detect because it arises and shows symptoms in final stage. With the development of various Machine Learning and Deep learning techniques computer Aided Diagnosis has become supplement and promising tool to support radiologists and pulmonary specialists. There have a been a lot research done on making the detection and classification of the lung cancer computer aided which is not prone to human errors. This study is an attempt to summarize some of the machine and deep learning techniques that are used in classification of lung cancer using the CT images. There are still many deep learning algorithms coming up almost every year on which the research should be carried on ensuring that the systems give the best accuracy to assist the doctors more effectively.

## REFERENCES

[1] P.B. Sangamithra, S.Govindaraju, "Lung Tumor Detection and Classification using EK-Mean Clustering", IEEE WiSPNET conference, 2016

[2] Prajwal Rao, Nishal Ancelette Pereira, Raghuram Srinivasan, "Convolutional Neural Networks for Lung Cancer Screening in Computed Tomography (CT) Scans", IEEE 2016

[3] Deep Prakash Kaucha, P.W.C. Prasad, Abeer Alsadoon, A.Elchouemi, Sasikumaran Sreedharan, "Early Detection of Lung Cancer using SVM Classifier in Biomedical Image Processing", IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI-2017)

[4] P.W.C. Prasad, Abeer Alsadoon, A.K. Singh, A. Elchouemi, "Lung Cancer Detection using CT Scan Images", Suren Makaju, 6th International Conference on Smart Computing and Communications, Elsevier, 2017

[5] Gur Amrit Pal Singh, P. K. Gupta, "Performance analysis of various machine learning-based approaches for detection and classification of lung cancer in humans",Springer, 2018

[6] Lingling Li, Yuan Wu, Yi Yang, Lian Li, Bin Wu, "A New Strategy to Detect Lung Cancer on CT Images", 3rd IEEE International Conference on Image, Vision and Computing, 2018

[7] Rohit Y. Bhalerao, Harsh P. Jani, Rachana K. Gaitonde, Vinit Raut, "A novel approach for detection of Lung Cancer using Digital Image Processing and Convolution Neural Networks", 5th International Conference on Advanced Computing & Communication Systems (ICACCS), IEEE, 2019

[8] M. S. Kavitha, J. Shanthini, R. Sabitha, "ECM-CSD: An Efficient Classification Model for Cancer Stage Diagnosis in CT Lung Images Using FCM and SVM Techniques", Journal of Medical Systems (2019) 43:73 Springer, 2019

[9] Chethan Dev, Kripa Kumar, Arjun Palathil, T. Anjali, Vinitha Panicker, "Machine Learning Based Approach for Detection of Lung Cancer in DICOM CT Image",Springer, 2019

[10] P. Mohamed Shakeel, M.A. Burhanuddin, Mohamad Ishak Desa, "Lung Cancer Detection from CT image using Improved Profuse Clustering and Deep Learning Instantaneously Trained Neural Networks", Measurement, Volume 145, Pages 702-712, October 2019

  