Nat. Volatiles & Essent. Oils, 2021; 8(4): 12264-12271

NVEO
Natural Volatiles &
Essential Oils

Lung Cancer Detection Based On Ct Images Using Feature Extraction And Classification Techniques

Dr. S G Balakrishnan[1] , Dr. P.Ramya[2] , Thallam Venkata Naga Hanumatha Rao[3] , Dr. G. Manikandan[4] , Hemavathi S[5] , R. Selvameena[6]

[1]Associate Professor, Dept of CSE, Mahendra Engineering College, Namakkal, Tamilnadu, India.

[2]Associate Professor, Dept of CSE, Mahendra Engineering College, Namakkal, Tamilnadu, India.

[3]Associate Professor, Dept of CSE,Hindu College PG Courses, Guntur, Andhra Pradesh, India.

[4]Assistant Professor, Dept of ECE, Dr.M.G.R Educational and Research Institute, Maduravoyal, Chennai, Tamilnadu, India.

[5a]Scientist, CSIR - Central Electrochemical Research Institute (CECRI), CSIR-Madras Complex, Chennai 600113, Tamilnadu, India.

[5b]Assistant Professor, Academy of Scientific & Innovative Research, Ghaziabad 201002, Uttar Pradesh, India.

[6]Assistant Professor, Dept of CSE, Dr.M.G.R Educational and Research Institute, Maduravoyal, Chennai, Tamilnadu, India.

Abstract

The image detection in medical field is a difficult task due to small lesions which are not identifiable in an early stage. The early medical diagnosis is important because it increases the mortality rate. Therefore prediction of lung cancer in its early stage is important. Here Linear Discriminant Analysis (LDA) classification is used to predict lung nodules and normal cells on MRI images. Dual Tree M-Band Wavelet Transform is used for feature extraction and LDA concentrates on the image classification. The proposed method undergoes three important stages includes in lung cancer detection such as pre-processing, feature extraction and classification. The accuracy results up to 95% of the images have been identified precisely.

Keywords: GLCM Feature Extraction, M-band wavelet co-efficient, Linear Determinant Analysis, Lung cancer detection.

1. Introduction

Abnormal growth of cells in any part of the body occupies additional space and leads to a disease called cancer. Therefore the growth of abnormal cells in lungs causes lung cancer [1]. Lung cancer might treated as a non-curable disease in the past period and this disease is considered to be 5th cause of death compared to all type of cancers that exists in the world [2]. It is the most common type of cancer and the death rate results nearly 1.59 million people per anum. Among all cancers the lung cancer presents the highest incidence for new cases of count 222,500 in both men and women with the death count of 155,870 in US in the year of 2017 [3]. Lung cancer be the second cause of death among men and women, it is the kind of prostate cancer for men and breast cancer for women [4].

Many lives can be saved through the earlier diagnosis of cancerous cells [5]. But accuracy detection of nodule cells by this earlier detection technique is difficult because of variations of morphological patterns in the lung tissues. Therefore classification method for lung nodule detection [6] was proposed to classify the lung nodules as benign or malicious. In correspondence with Lung Image Database Consortium (LIDC) the contour maps are extracted on basis of excel files [11]. The proposed cancer detection techniques are based on radiation method and every technique has own pros and cons. If radiation is too high then the image quality will get reduced also affects the patients health, hence an alternative method was developed through Computer Aided Diagnosis (CAD) in concern with human healthcare system. The main classification methods used nowadays are Artificial Neural Network (ANN) [7] and Support Vector Machine (SVM) [8]. The accuracy of the CAD system is improved by training the dataset and testing the dataset on basis of the system requirement. Results obtained for SVM classifier is 97% accuracy when compared with ANN classifier including significant reduction in false rate densities [9].

An appropriate method called Computed Tomography (CT) screening is considered to identify the lung cancerous cells [10] in the earlier stages. Lung carcinoma or lung nodule is defined as the malignant lung tumor caused in lung tissues because of hysterical cell growths. These cells can grow beyond the lungs into neighbour tissues and nearby parts of the body or other various parts of the body. This process is said to be metastasis. Benign nodules are referred to normal patients and malignant nodules are represented for the patients who suffering with cancerous cells [12].

2. Proposed Method

In this proposed method the lung nodule identification is carried out in three stages such as image pre-processing using median filter, feature extraction using GLCM features and classification of lung nodules using linear discriminant analysis classifier. Figure 1 gives the overall process of the proposed lung nodule classification method.

Figure 1: Nodule classification using LDA

The CT lung images collected from the dataset are usually consists of noise with low contrast because of image acquisition process. Hence the input image should undergo the process of image preprocessing in order to remove the impulse noise as well as to get better scan image contrast value. The preprocessing method not only improves the precision but also increases the nodule detection accuracy value. Here median filter is applied for removing the noise and this provides good smoothing consequences when compared with other noise removing filters.

The next stage is extracting the features and this feature extraction process plays a significant role in detection of true positive nodules precisely. Extraction of good features can efficiently differentiate the true nodules from the acquired candidate nodules. Intensity based features is used to extract the Dual Tree M-Band Wavelet (DTMBW) sub-band coefficients. The features such as skewness and kurtosis are used to decompose the DTMBW features. By taking the intensity values presented in each region of pixels the average intensity values for all the pixels in the region are calculated. The intensity value in the wide range of pixels is identified and interchanges with the overlapping of cells in an image. The pulmonary nodules are detected to be circular in shape and the blood vessels are identified with slender like structures. Skewness and variance features are defined in following equations 1 and 2.

$$Skewness_{Sk} = \frac{1}{\sigma_r^3}\sqrt{\sum_{r=0}^{R-1}(r-\bar{r})^3 s(r)} \quad (1)$$

$$Variance\ \sigma_r^2 = \sum_{r=0}^{R-1}(r-\bar{r})^2 s(r) \quad (2)$$

The skewness, variance, kurtosis and contrast measure are used to extract the lower and higher frequency components of DTMBT. These extracted features are stored in the feature database and used for the classification.

Once the intensity based features for each and every nodule candidate is calculated then the extracted features are analyzed with the suitable classifier. Here LDA classifier method is applied for the classification of benign (normal) cells and malignant (cancerous) lung nodules. The feature values are evaluated with the LDA classifier to discriminate between the malignant nodule and the benign cells, the extracted features is combined with the LDA. The function of linear discrimination analysis g(x) is defined using equation 3.

$$g(x) = W(X) \oplus \omega_0 \quad (3)$$

The term W(X) denotes the product weight vector and product feature vector. The term ω_o indicates the bias that is used to determine or locate the position of the decision surface.

3. Results and Discussions

The implementation process is done using the tool called MATLAB r2018 and the consequences are shown further. 130 images are taken for training process and 40 images are used for testing process. The size of the input image taken here is 256X256 and the classified output of nodules and non-cancerous cells and its intensity level is measured through accuracy calculation. The sample input image that taken in the dataset is given in figure 2.

Figure 2: Sample Input Image

Mainly 2 features are extracted from this feature extraction process which is carried out by applying GLCM feature reduction technique and fed as input to the LDA classifier. The image size after the feature extraction process reduced to 15X19.

Accuracy measurement is defined as the number of precise predictions made and it is termed as a different label classification. To evaluate performance of classification confusion matrix is used for evaluation. The confusion matrix in general comprises of four tables with the combination of predicted values and actual values. But the output consists of two or more classes based on the requirement. The accuracy is measured using the equation 4.

$$Accuracy\,(\%) = \frac{(TP + TN)}{(TP + TN + FP + FN)} \qquad (4)$$

Here P represents Positive i.e. resulted as positive, N represents Negative where the observation is resulted as not positive. TP stands for True Positive which implies the observation as positive therefore results in positive %. FN stands for False Negative that denotes the observation as positive and it results in negative %. TN stands for True Negative and observation seems to be negative as well as results in negative. FP stands for False Positive and observation is considered as negative and results as predicted positive. Histogram oriented gradient image representation for non-cancerous cells are shown in figure 3.

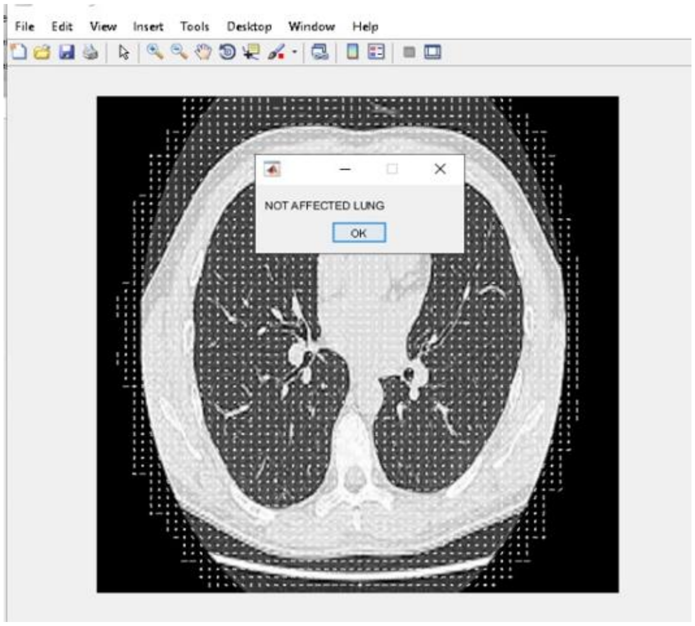

Figure 3: Histogram representation of non-cancerous lung

Figure 4: LDA classifier output

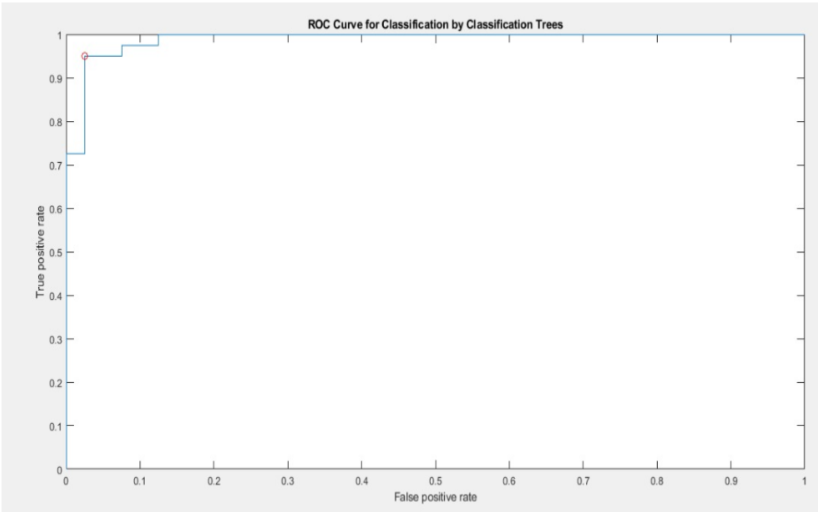

Figure 5: ROC curve for LDA Classifier

The confusion matrix result and the Region of Curve (ROC) which is applied for the calculation of accuracy of the LDA classifier are shown in figures 4 and 5 respectively. Therefore for our proposed model provides better accuracy of 99% compared to the conventional classifiers.

4. Conclusion

Detection of lung cancer in its early stage is significant since this leads to fatality. Therefore several digital imaging techniques had been proposed for the detection of cancerous cells but the precision accuracy is still being a challenging one. Hence a novel and precise lung cancer detection method is proposed. The proposed method undergoes three important stages includes in lung cancer detection such as pre-processing, feature extraction and classification. This method uses median filter for pre-processing the image, DTMBW is applied for feature extraction and LDA classifier is used for classification purpose to predict lung nodules and normal cells on MRI images. The accuracy results up to 99% of the images have been identified precisely.

References

1. Gajdhane, V. A., & Deshpande, L. M. (2014). Detection of lung cancer stages on CT scan images by using various image processing techniques. IOSR Journal of Computer Engineering (IOSR-JCE), 16(5), 28-35.

2. Vardell, E. (2020). Global Health Observatory Data Repository. Medical Reference Services Quarterly, 39(1), 67-74.

3. Siegel, R. L., Miller, K. D., Fedewa, S. A., Ahnen, D. J., Meester, R. G., Barzi, A., & Jemal, A. (2017). Colorectal cancer statistics, 2017. CA: a cancer journal for clinicians, 67(3), 177-193.

4. Moradi, P., & Jamzad, M. (2019). Detecting Lung Cancer Lesions in CT Images using 3D Convolutional Neural Networks. In 2019 4th International Conference on Pattern Recognition and Image Analysis (IPRIA) (pp. 114-118). IEEE.

5. Ravindranath, K., & Somashekar, K. (2017). Early detection of lung cancer by nodule extraction—A survey. In 2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT) (pp. 1-4). IEEE.

6. Tariq, A., Akram, M. U., & Javed, M. Y. (2013). Lung nodule detection in CT images using neuro fuzzy classifier. In 2013 Fourth International Workshop on Computational Intelligence in Medical Imaging (CIMI) (pp. 49-53). IEEE.

7. Nasser, I. M., & Abu-Naser, S. S. (2019). Lung cancer detection using artificial neural network. International Journal of Engineering and Information Systems (IJEAIS), 3(3), 17-23.

8. Asuntha, A., Brindha, A., Indirani, S., & Srinivasan, A. (2016). Lung cancer detection using SVM algorithm and optimization techniques. J. Chem. Pharm. Sci, 9(4), 3198-3203.

9. Taher, F., Werghi, N., & Al-Ahmad, H. (2015). Computer aided diagnosis system for early lung cancer detection. Algorithms, 8(4), 1088-1110.

10. Gopi, K., & Selvakumar, J. (2017). Lung tumor area recognition and classification using EK-mean clustering and SVM. In 2017 International Conference on Nextgen Electronic Technologies: Silicon to Software (ICNETS2) (pp. 97-100). IEEE.

11. Gupta, A., Märtens, O., Le Moullec, Y., & Saar, T. (2015). A tool for lung nodules analysis based on segmentation and morphological operation. In 2015 IEEE 9th International Symposium on Intelligent Signal Processing (WISP) Proceedings (pp. 1-5). IEEE.

12. Setio, A. A., Jacobs, C., Gelderblom, J., & van Ginneken, B. (2015). Automatic detection of large pulmonary solid nodules in thoracic CT images. Medical physics, 42(10), 5642-5653.