A Lung Cancer Diagnosis using Fuzzy Logic

Miss. Minaj Chaugule, Dr. Kharat Govind U

***Abstract*— early detection of cancer can decrease mortality rate of cancer patients. In India there are large amount of lung cancer patients due to increase pollution and smoking. Hence we are targeting lung cancer for research. Various studies has proved that the overall 5-year survival rate for lung cancer patients increases from 14 to 49% if the disease is detected in time. Knowing this factor proposed paper is focused on early diagnosis of lung cancer. Here HRCT images of lung are used as an input image. Statistical parameters are principle components of proposed method. Mean square error, peak signal noise ratio, entropy, standard deviation, texture factor, Variance and VIF are some of statistical parameters. Fuzzy logic is used to get accurate results of diagnosis. The proposed method is implemented using MATLAB. Outcome of this project is accurate and fast detection of lung cancer without causing any kind of manual error.**

***Index Terms*—Lung cancer, Mean square error, peak signal to noise ratio, Variance.**

I. INTRODUCTION

[1] The lung is the essential respiration organ in air-breathing the two lungs are located near the backbone on either side of the heart. Their principal function is to transport oxygen from the atmosphere into the bloodstream, and to release carbon dioxide from the bloodstream into the atmosphere. A large surface area is needed for this exchange of gases which is accomplished by the mosaic of specialized cells that form millions of tiny, exceptionally thin-walled air sacs called alveoli.

[2] But there are various diseases which disturb the function of lung like cancer, ILD, CPOD, asthma and influenza. The diagnosis and treatment of disease is most important for proper functioning of human body.

[3] Lung cancer seems to be the common cause of death among people throughout the world. Early detection of lung cancer can increase the chance of survival among people. The overall 5-year survival rate for lung cancer patients increases from 14 to 49% if the disease is detected in time. Although Computed Tomography (CT) can be more efficient than X-ray. However, problem seemed to merge due to time constraint in detecting the present of lung cancer regarding on the several diagnosing method used. Hence, a lung cancer detection system using image processing is used to classify the present of lung cancer in a CT- images. In this study, MATLAB have been used through every procedures made. In image processing procedures, process such as image pre-processing, segmentation and feature extraction have been discussed in detail. We are aiming to get the more accurate and fast results. in next section we will review Papers published on the same topic.

II. PAPER SURVEY

In paper 'Sex and Smoking Status Effects on the Early Detection of Early Lung Cancer in High Risk Smokers using an Electronic Nose', published in IEEE Transactions on Biomedical Engineering, 2015. It states that "Objective: Volatile Organic Compounds (VOC) in exhaled breath as measured by electronic nose (e-nose) have utility as biomarkers to detect subjects at ris of having lung cancer in a screening setting. We hypothesize that breath analysis using an e-nose chemo-resistive sensor array could be used as a screening tool to discriminate patients diagnosed with lung cancer from high-risk smokers.

Methods: Breath samples from 191 subjects – 25 lung cancer patients and 166 high-risk smoker control subjects without cancer – were analyzed. For clinical relevancy, subjects in both groups were matched for age, sex, and smoking histories. Classification and Regression Trees and Discriminant Functions classifiers were used to recognize VOC patterns in e-nose data. Cross-validated results were used to assess classification accuracy. Repeatability and reproducibility of e-nose data were assessed by measuring subject-exhaled breath in parallel across two e-nose devices.

Results: E-nose measurements could distinguish lung cancer patients from high-risk control subjects, with a better than 80% classification accuracy. Subject sex and smoking status impacted classification as area under the curve results (ex-smoker males 0.846, ex-smoker female 0.816, current smoker male 0.745 and current smoker female 0.725) demonstrated. Two e-nose systems could be calibrated to give equivalent readings across subject-exhaled breath measured in parallel.

Conclusions: E-nose technology may have significant utility as a non-invasive screening tool for detecting individuals at increased risk for lung cancer.

Significance: The results presented further the case that VOC patterns could have real clinical utility to screen for lung cancer in the important growing ex-smoker population."[4]

In paper 'Study of Malignancy Associated Changes in Sputum Images as an Indicator of Lung Cancer', published in Proceedings of the 2016 IEEE Students' Technology Symposium .it states that "Lung cancer is one among the major causes of cancer related deaths. Fortunately, an early stage diagnosis can increase the survival rates of the patients. Sputum cytology is one of the easiest and cost-effective method for lung cancer diagnosis. Chances of misdiagnosis and sampling error related to sputum cytology led to the concept of malignancy associated changes. Malignancy associated changes (MAC) are the subtle changes that happens to the normal appearing cells near or distant from the

malignant cells. Literature suggests that these changes can be used as an indicator for lung cancer rather than using malignant cells which are very less in number compared to the normal appearing cells in sputum cytology images. The proposed work is intended to detect cells with MAC from sputum smear images. Analysis of nuclei texture features of sputum cell nuclei using Gray Level Co-occurrence Matrix and Gray Level Run Length Matrix from both normal and cancer patients revealed that both type of cells could be differentiated. Among 110 texture features calculated for each nuclei, a set of 35 features which clearly distinguishes normal cells and normal appearing cells were chosen. Support Vector Machine (SVM) classifier is used to classify the cells into two classes' i.e cells with MAC and cells without MAC. This study demonstrates that the presence of MAC cells in conventional microscopic sputum cytology
Images can be identified using image processing techniques and it can have some significance in the early detection of lung cancer."[5]

In paper 'Texture Analysis Based Feature Extraction and Classification of Lung Cancer', published in 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT). Authors states that "Lung cancer is most life-threatening disease, treatment of which must be the primary goal throughout scientific research. The early recognition of cancer can be helpful in curing disease entirely. There are numerous techniques found in literature for detection of lung cancer. Several investigators have contributed their facts for cancer prediction. These papers largely pact about prevailing lung cancer detection techniques that are obtainable in the literature. A numeral of methodologies has been originated in cancer detection methodologies to progress the efficiency of their detection. Diverse applications like as support vector machines, neural networks, image processing techniques are extensively used in for cancer detection which is elaborated in this work."[2]

In paper Multi-Stage Lung Cancer Detection and Prediction Using Multi-class SVM Classifier '' published in 2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2),authors says that "Recognition and prediction of lung cancer in the earliest reference point stage can be very useful to improve the survival rate of patients. But diagnosis of cancer is one the major challenging task for radiologist. For detecting, predicting and diagnosing lung cancer, an intelligent computer-aided diagnosis system can be very much useful for radiologist. This paper proposed an efficient lung cancer detection and prediction algorithm using multi-class SVM (Support Vector Machine) classifier. Multi-stage classification was used for the detection of cancer. This system can also predict the probability of lung cancer. In every stage of classification image enhancement and segmentation have been done separately. Image scaling, color space transformation and contrast enhancement have been used for image enhancement. Threshold and marker-controlled watershed based segmentation has been used for segmentation. For classification purpose, SVM binary classifier was used. Our proposed technique shows higher degree of accuracy in lung cancer detection and prediction". [1]

In paper 'Semi-Supervised Multi-Task Learning for Lung Cancer Diagnosis' published in IEEE 2018, it states that "Early detection of lung nodules is of great importance in lung cancer screening. Existing research recognizes the critical role played by CAD systems in early detection and diagnosis of lung nodules. However, many CAD systems, which are used as cancer detection tools, produce a lot of false positives (FP) and require a further FP reduction step. Furthermore, guidelines for early diagnosis and treatment of lung cancer are consist of different shape and volume measurements of abnormalities. Segmentation is at the heart of our understanding of nodules morphology making it a major area of interest within the field of computer aided diagnosis systems. This study set out to test the hypothesis that joint learning of false positive (FP) nodule reduction and nodule segmentation can improve the computer aided diagnosis (CAD) systems' performance on both tasks. To support this hypothesis we propose a 3D deep multi-task CNN to tackle these two problems jointly. We tested our system on LUNA16 dataset and achieved an average dice similarity coefficient (DSC) of 91% as segmentation accuracy and a score of nearly 92% for FP reduction. As a proof of our hypothesis, we showed improvements of segmentation and FP reduction tasks over two baselines. Our results support that joint training of these two tasks through a multi-task learning approach improves system performance on both. We also showed that a semi-supervised approach can be used to overcome the limitation of lack of labeled data for the 3D segmentation task."[3]

In next section we will review some mathematical parameters used in lung cancer diagnosis.

III. Principle Components

Statistical and Mathematical parameters are principle components of lung cancer diagnosis. In this section we will see those important principle components.

2.1 Statistical Analysis

a) Correlation: Correlation is an additionally a statistical procedure which demonstrates how factors are robustly related with each other. It extracts necessary information from the image. It is used to find the location in an image that is analogous to the reference image. Correlation is a measure of gray level linear dependence between the pixels at the specified positions relatively [10].

$$y(n) = \sum_{-n}^{n} x(v)h(n-v) \tag{1}$$

Where $x(v)$ - Image1 and $h(n-v)$ -Image 2(Shifted).

b) Entropy: It indicates average information of the image. The lowest value of entropy means no uncertainty of the image information. It is zero if the event is sure or impossible

$$E = -\sum_{x}^{m}\sum_{y}^{n} P[x,y]\log P[x,y] \tag{2}$$

$P[x,y]$ Is the probability difference between two adjacent pixels and log is the base2 logarithm. Deliberating Entropy E=0 if P=0 or 1. Entropy is supposed to be high throughout the image and is calculated by equation (1).

c) MSE (Mean Square Error) : The MSE represents the averaging of the squares of the errors between the two images. The error is the amount by which the values of the reference image differ from test image calculated form equation (3).

$$MSE = \sum_{0}^{m-1}\sum_{0}^{n-1}\left\| f(i,j) - g(i,j) \right\|^2 \quad (3)$$

$f(i,j)$ Represents the matrix data of original image and $g(i,j)$ represents the matrix data of test image. m represents the numbers of rows of pixels of the images and i represents the index of that row n represents the number of columns of pixels of the image and j represents the index of that column. MSE for the practical purpose allows comparing the true pixel values of original image to cancerous image.

c) Variance: It explains about the distribution of gray levels over the image. The value of the Variance is expected to be high, if the gray levels of the image are spread out extensively. It explains about the probability of distribution, describing how far the value lies from the mean that is anticipated value, which can also be defined as the moments of a distribution. Second moment (S=2) is given as by equation (5). The second moment is the Variance. μ_x is the average of x .Mean provides each pixel intensity for the whole image, whereas the variance gives each pixel variations form the neighboring pixels and is use to classify image into different regions or areas. It is the average of the squared differences from the mean. It is the variability around the value.

$$S^{th} = \sum (x_i - \mu_x)^2 \quad (4)$$

Steps to calculate variance of an image
1. Calculate Mean (simply average of numbers)
2. For each number, subtract the mean and the square of result (the squared difference)
3. Average of those squared differences

The mathematical expression for calculating Variance is given in equation (5), $N-1$ can be changed to N if the $\bar{x}$ is known prior rather than being estimated from the data

$$\text{var} = \sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x}) \quad (5)$$

d) Standard Deviation: If the value of standard deviation is less which means that the majority of the data is near the mean value and if it is more that means data is more spreaded over the image. The value of standard deviation is assigned to the center pixel of the image. All the steps are similar to calculate the standard deviation as the variance, only the last step is added as the square root, hence is the square root of the variance represented by equation (5)

$$SD = \sigma = \sqrt{\sigma^2} \quad (6)$$

d) Mean: It calculates the mean of the gray levels in the image. The Mean is supposed to be high, if the sum of the gray levels of the image is high. Mean depends on the first moment of the data. Technically, a moment is defined by a mathematical formula that just so happens to equal formulas for some measures in statistics. First moment **is** the mean which is represented as in equation (7) and (8)

$$S^{th} = \frac{(x_1^s + x_2^s + x_3^s + \dots. x_n^s)}{n} \quad (7)$$

First moment (S=1)

$$S^{th} = \frac{(x_1^1 + x_2^1 + x_3^1 + \dots. x_n^1)}{n} \quad (8)$$

This formula is identical to the formula to find the sample mean. Just add up all of the values and divide by the number of items in given data set. The mathematical expression of mean is given as in equation (9)

$$\mu = 1/N*M\sum_{x=0}^{M}\sum_{y=0}^{N}P[x,y] \quad (9)$$

Where $N*M$ (255*255) is the size of the image

IV. System development

The proposed methodology is presented in figure.1

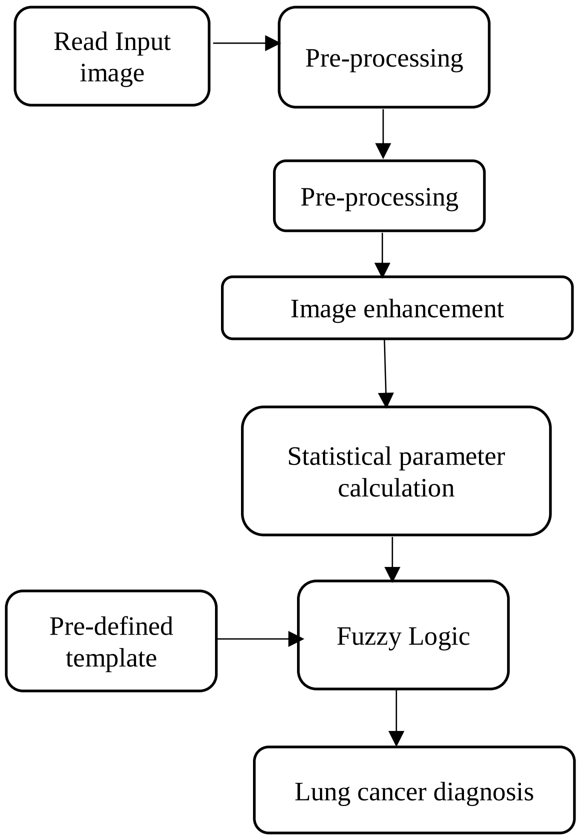

Figure. 1 Proposed methodology

The proposed methodology is based on principle component analysis of HRCT images. Statistical parameters are used as a principle component. Based on correlation of pre-defined template of principle component and real time principle component of a test image, fuzzy logic takes decision of lung cancer detection.

V. Results & Discussion

In the paper, the input image is biomedical image Input image is taken as gray level image. The median filter is a nonlinear digital filtering technique, often used to remove noise. To improve the quality of image we are using image enhancement algorithm. This algorithm enhances the image by focusing on parameters like contrast, brightness adjustment. For lung cancer diagnosis we have used correlation method. We analyzed standard HRCT images of cancerous and non-cancerous images. Fuzzy logic is used to avoid confusion between same statistical parameters in cancerous as well as non-cancerous images.

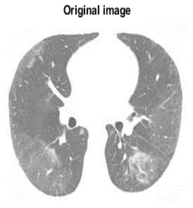

Fig. 2 Original Image

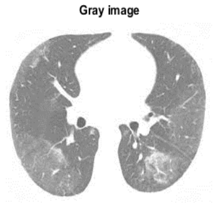

Fig. 3 Gray Image

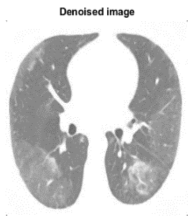

Fig. 4 Denoised Image

Fig. 5 Enhanced Image

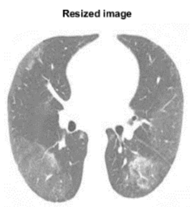

Fig. 6 Resized Image

VI. Conclusion

The proposed methodology is less time consuming and more accurate. It will be small step towards the research in early detection of lung cancer using image processing. The method will be more robust and cost effective as compared to existing methods. MATLAB allows us easy modification and upgradation of end user application. Handling of end user model in proposed implementation is very easy. It can be modified and updated easily as per end user requirement. Same methodology can be used with moderate changes for another type of cancer diagnosis.

Acknowledgement

I would like thank all technical and non-technical staff of Sharadchandra College of engineering and technology, Otur for their great support to implement this work. I am very much thankful to my guide Prof. Dr, Kharat Govind U for their immense support and guidance to implement this innovative system.

References

[1] Ojala, T., Pietikainen, M., & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence, 24(7), 971-987.*

[2] Radüntz, T., Scouten, J., Hochmuth, O., & Meffert, B. (2015). EEG artifact elimination by extraction of ICA-component features using image processing algorithms. *Journal of neuroscience methods, 243, 84-93.*

[3] Sairamya, N. J., Selvaraj, T. G., Ramasamy, B., Deivendran, N. P., & Subathra, M. S. P. (2018). Classification of EEG signals for detection of epileptic seizure activities based on feature extraction from brain maps using image processing algorithms. *IET Image Processing.*

[4] Sairamya, N. J., George, S. T., Subathra, M. S. P., & Kumar, N. M. (2019). Computer-Aided Diagnosis of Epilepsy Based on the Time-Frequency Texture Descriptors of EEG Signals Using Wavelet Packet Decomposition and Artificial Neural Network. *In Cognitive Informatics and Soft Computing (pp. 677-688). Springer, Singapore.*

[5] Sairamya, N. J., George, S. T., Ponraj, D. N., & Subathra, M. S. P. (2017, October). Automated Detection of Epileptic Seizure Using Histogram of Oriented Gradients for Analysing Time Frequency Images of EEG Signals. *In International Conference on Next Generation Computing Technologies (pp. 932-943). Springer, Singapore.*

[6] Kaya, Y., Uyar, M., Tekin, R., & Yıldırım, S. (2014). 1D-local binary pattern based feature extraction for classification of epileptic EEG signals. *Applied Mathematics and Computation, 243, 209-219.*

[7] Jaiswal, A. K., & Banka, H. (2017). Local pattern transformation based feature extraction techniques for classification of epileptic EEG signals. *Biomedical Signal Processing and Control, 34, 81-92.*

[8] Tiwari, A. K., Pachori, R. B., Kanhangad, V., & Panigrahi, B. K. (2017). Automated diagnosis of epilepsy using key-point based local binary pattern of EEG signals. *IEEE journal of biomedical and health informatics, 21(4), 888-896.*

[9] Sairamya, N. J., George, S. T., Ponraj, D. N., & Subathra,

M. S. P. (2018). Detection of Epileptic EEG Signal Using *Improved Local Pattern Transformation Methods. Circuits, Systems, and Signal Processing, 1-22.*

[10] Sairamya, N. J., George, S. T., Balakrishnan, R., & Subathra, M. S. P. (2018). An effective approach to classify epileptic EEG signal using local neighbor gradient pattern transformation methods. *Australasian physical & engineering sciences in medicine, 41(4), 1029-1046.*

First Author Minaj Chaugule has done her Bachelor of Engineering in Electronica and telecommunication. Now she is pursuing her Masters of Engineering in Embedded and VLSI.
Email: minajchaugule08@gmail.com

Second Author Prof. Kharat Govind U has done his Bachelor of Engineering in Electronics and Masters of Engineering in Power systems. His PhD topic was power Electronics Now he is working as an Principal at Sharadchandra college of engineering, Otur,Maharashtra,India.
Email: principalspcoe09@gmail.com