IJCRT.ORG ISSN : 2320-2882
INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

Lung cancer detection and analysis using SVM binary classifier

Shabbir Pathan, Amita Dessai

Student, Assistant Professor

Electronics and telecommunication engineering,

Goa college of Engineering, Goa, India.

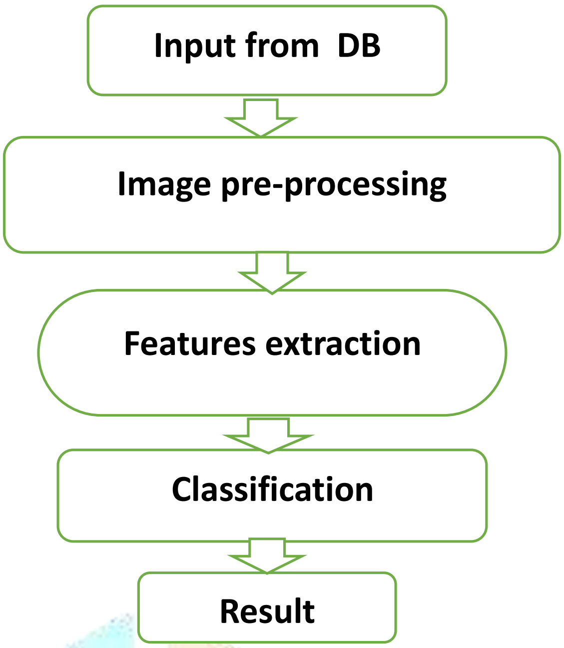
Abstract: Lung cancer has been responsible for the major deaths all over the world for the past few decades. The mortality rate of lung cancer has been at the peak in comparison with all other cancers worldwide. The most adequate way followed by radiologists for analysis of cancer in lungs is using computed tomography as it provides a detailed picture of cancer cells. The usual method followed is visual interpretation which can sometimes be wrong due to human error and is always time consuming. Therefore an efficient Lung cancer detection system is required by everyone all over the world. Thus this paper presents an automated approach for premature analysis of lung cancer using CT scan images. A numeral methodologies have been discovered to increase the efficiencies of the system for detecting the tumours in the lungs.Thus this algorithm is proposed using support vector machine, watershed segmentation and extraction of geometrical and textural features using gray level co-occurance matrix.

Keywords: SVM, GLCM, CT(Computed topography).

***I*. Introduction**: Lung cancer is the foremost cause of cancer deaths worldwide, with 1.8 million new cases being diagnosed each year. The most fatal cancer in women is breast cancer but lung cancer has excelled it over the years. On the other hand it has exceeded over prostate, pancreatic and colon cancer fused together in men. The sixth cause of deaths in United states is lung tumors without smoking.

Lung tumors arises in the tissues of the lungs first and slowly reaches the bronchi. The disease is capable of modifying the normal cells into cancerous cells. These cells than turn into tumors at a very farst rate.Thus the oxygenating all the parts of body becomes difficult as the breathing rate is affected.These cells can further multiply and spread to other body parts which eventually leads to most cancer deaths. Non-small cell lung cancer is most common, being responsible for 80 to 85 percent of cancers.In non-smokers ,non-small cell lung cancer is commonly found. Small cell lung cancer are accountable for 15 percent of lung cancers. These cancer is aggressive by nature and its hard detect it before it spreads fully. In this paper, an automated system is developed for the earliest detection of these cancer cells and thereby reduce the latency in detection by following the traditional methods.

***II*.Methodology:** The proposed system for lung cancer detection is carried out using five main steps as depicted in the flow chart below:



***III*.Data Collection** : The first step is to obtain the lungs dataset of cancer patients. For research work, the images have been downloaded from the Cancer Imaging Archive database. The images are converted from DICOM to JPEG format and stored. The image database contains Computed Tomography images of patients with and without lung cancer.The dataset consist of 218 JPEG images of patients.

***IV*.Image Pre-Processing:** The objective of image pre-processing stage is to quench of all the distortions in the image and to intensify some features useful for further processing. The salt and pepper noise mostly present in CT scan is filtered out using median filtering which ise quite effective technique in eliminating this impulse noise while preserving the edges. A median filter of size 3*3 was used and its contribution towards enhancement of the images.

***V*.Segmentation**:

The division of image in the nodes of the lung in the provided CT scan image is useful in obtaining improved resolution of the provided image. Transform such as Watershed have been used for segmentation. In Watershed Transform , the lines(watershed) are comparing edges in between indicator(marker) and is not subjected by edges having contrast that is lower and as a result the neighbourhood minima problem is solved. Segmentation of image is a crucial part as it performs the action of screening in biomedical image processing. The model is being divided into numerous segments. The vision and recognition system using computer therein the digitalized image is separated into numerous segments . Motive behind the segmentation process is to convert the provided CT scan images into more detailed image so that more information can be extracted easily. The image segmentation eliminates non vital details in the input image and certain details like curves, boundaries. The method that has been proposed, therein the process of segmentation further comprises of some stages. The input image provided is being transformed into image edge. Then image edge is converted into image that is dilated and filled image .In the final step the right and left lung both is segmented.

***VI*.Feature extraction**:

The output of watershed segmentation part is ued as an input to the feature extraction part. The coloured watershed matrix is superimposed on the original image and geometrical and textural features are extracted using GLCM(Gray level co-occurance matrix).The geometrical features extracted are Area, Eccentricity and Circularity. The textural features extracted are Contrast, Correlation, Energy, Homogeneity, Mean, Standard-Deviation, Entropy, RMS, Variance, Smoothness, Kurtosis and Skewness.
The Fifteen features extracted are as follows, where Ng is the number of gray levels, P is the normalized symmetric GLCM and p(i, j) is the (i, j)th element of the normalized GLCM

Table 6.1 Geometric features

Area	Number of pixels within the lungs
Eccentricity	$\frac{(\mu_{02}-\mu_{20})^2+4M_{11}}{A}$
Circularity	$\frac{4\Pi A}{P^2}$

Table 6.2 Texture features

Contrast	$F_1=\sum_{ij}[i-j]_p^2(i,j)$
Energy	$F_2=\sum_{j-1}[P(\mathrm{i},j)]^2$
Correlation	$F_3=\sum_{i-j}\frac{(i-\mu_i)(j-\mu_j)p(i,j)}{\sigma_i\sigma_j}$
Homogeneity	$F_4=\sum_{i',j=0}^{N-1}\frac{P_{ij}}{1+(i-j)^2}$
Mean	$F_5=\mu=\frac{1}{MN}\sum_{i=1}^{M}\sum_{j\neq 1}^{N}p(i,j)$
Standard deviation	$F_6=\sigma=\sqrt{\frac{1}{MN}\sum_{1-1}^{M}\sum_{j-1}^{N}[p(\mathrm{i},j)-\mu]^2}$
Entropy	$F_7=\sum_{i_1j=0}^{N-1}-\ln(P_{ij})P_{ij}$
Root mean square(rms)	$F_8=\frac{\mu}{\sqrt{2}}$
Variance	$F_9=\sigma^2=\frac{1}{MN}\sum_{i-1}^{M}\sum_{j-1}^{N}[P(i,j)-\mathrm{\mu}]^2$
Smoothness	$F_{10}=1-\frac{1}{1+\sigma}2$

Kurtosis	$F_{11} = \frac{1}{MN}\sum_{i-1}^{M}\sum_{j-1}^{N}\left[\frac{p_{(i,j)}-\mu}{\sigma}\right]^4$
Skewness	$F_{12} = \frac{1}{MN}\sum_{i-1}^{M}\sum_{j-1}^{N}\left[\frac{P(ij)-\mu}{\sigma}\right]^3$

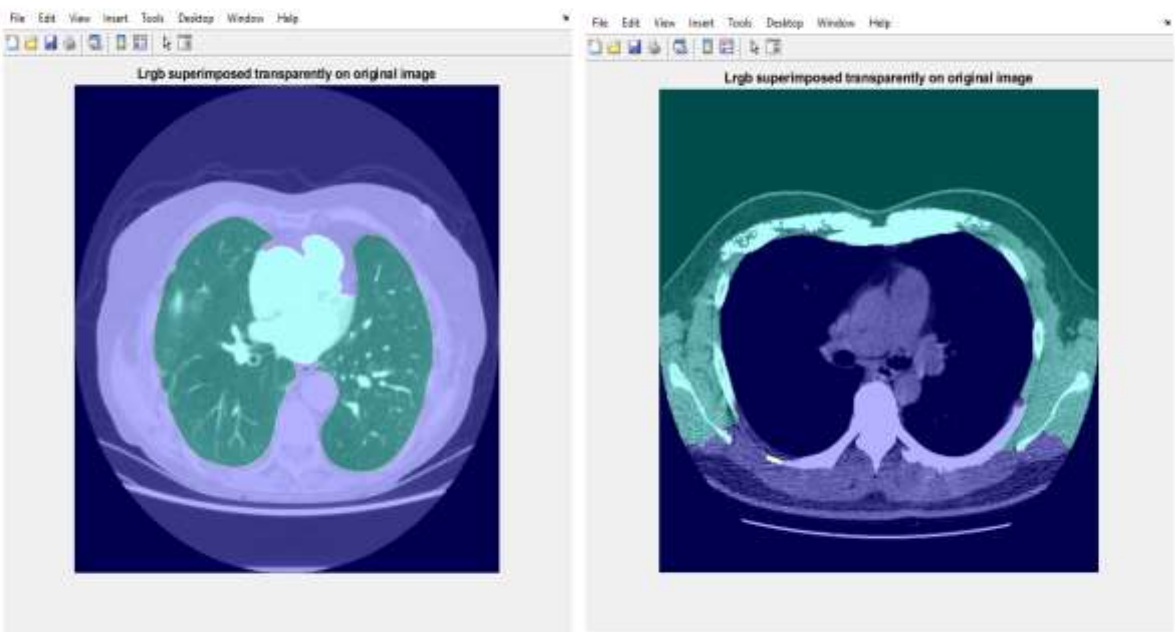
***VII*. Classification**:

The classification stage involves labelling of CT scan images as positive and negative images. For classification we have used binary SVM(Support vector machine) classifier.The extracted features of 160 images are feeded as input to SVM model which analysis all the inputs inputs into positive and negative classes and classifies them according to the machine learning patterns. The SVM algorithm thus finds the the best hyper plane that separates the two groups and thus classifies the lung CT images. For the best hyper plane data points of one class are separated from the other by largest margin between the two groups.

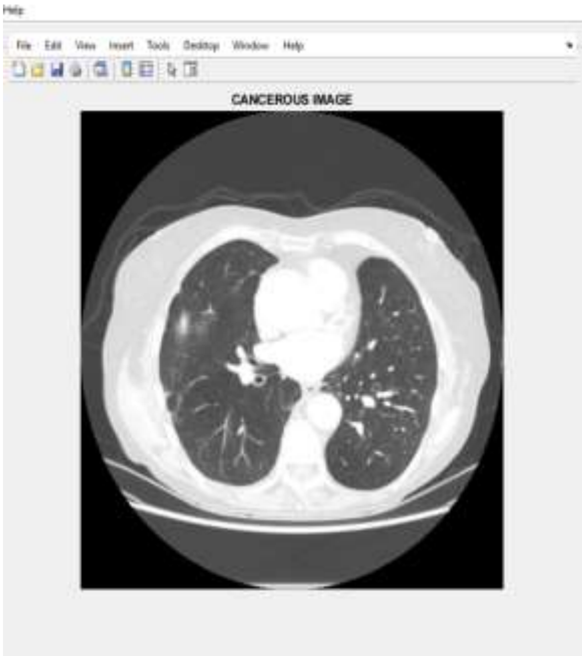
***VIII*. Results:**

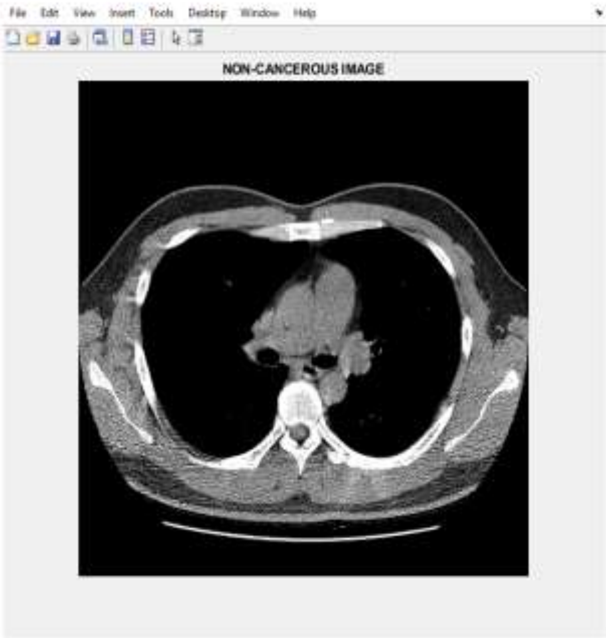
A dataset of 218 images were used, these dataset was divided into two parts training and testing.For training 160 JPEG images were used and for testing 58 JPEG images were used.

The images are converted to gray form and then all the images are advanced through the median filter. The left and right lungs are segmented out using watershed segmentation.The regional maxima(Lrgb image) is then superimposed on the original image. This provides a more valuable image through which information(features) can be easily and precisely extracted. The output as “Tested Objective Function Value” and optimized classifier is depicted in this figure below.



The 15 features are then extracted from this image and then used by the SVM classifier to classify the images as cancerous and Non-cancerous images.The final output is shown in the figure below and the message is displayed as”Patient is having cancer” or “Patient is not having cancer” .





IX .Conclusion:

In this paper, a system for spontaneous detection of lung cancer in CT images was successfully developed using image processing technique. The adopted methodology performs well in pre-processing, segmenting and extracting features from CT images. Median filtering technique was effective in eliminating impulse noise from the images without blurring the image. Mathematical morphological operations enable accurate segmentation of lung and tumor region. Three geometrical features i.e. area, cicularity and eccentricity and textural features i.e. Contrast, Correlation, Energy, Homogeneity, Mean, Standard-Deviation, Entropy, RMS, Variance, Smoothness, Kurtosis and Skewness were extracted from segmented region and fed to the input of the classifier for classification of lung CT images into cancerous or not. The proposed methodology detects the cancer cells efficiently with an accuracy of 91.37%.

References:

[1]JaneeAlam SabrinaAlam, Alamgir Hassan, "Multistage lung cancer detection and prediction using multiclass SVM classiffier. ",International journal of scientific and research publications,2018

[2]S.Kalavani,. Pramitchatterjee shikharJuyal, RishiGupta"Lung cancer detection using Digital image processing and artificial neural network"; International conference on electronics, communication and aerospace technology ICECA 2017.

[3]Silvia Moreno, Mario Bonfante, Eduardo Zurik, Hamerosan Juan," Study of medical image processing technique applied to lung cancer"; International journal of research and pulications,2019.

[4]Moffy Vas, Prof.AmitaDessai," Lung cancer detection system using lung CT image processing".International conference on computing communication control and automation.ICC0BEA 2017.

[5]RuchitaTekade, K.Rajeshwari," Lung cancer detection and classification using deep learning", International conference on computing communication control and automation, ICCUBEA 2018.

[6]Bhagvashri G.patil. Prof.sanjeev N Jain," Cancer cells detection using digital image processing methods", International Journal of latest Trends in engineering and technology,lJLTET

[7]S.Avinash; K.Manjunath; S.Senthil Kumar," An improved image processing analysis for detection of Lung cancer", International conference on inventive computation technologies ICIET 2016.

[8]S. Kalairam, Permit Chatterjee, Shikharjuyal, Rishigupta, "Lung cancer detection using digital image processing and artificial neural network", International conterence on electronics,communication and aerospace technology ICECA 2017.

[9] Kyamelia Roy, Ahana Ganguly, Rayna Banik, Sheli Sinha Chaudhury,

Chandrima Dutta, Madhurima Burman, Sayani Banik," A Comparative study of Lung Cancer detection using supervised neural network", 2019 International Conference on Opto-Electronics and Applied Optics (Optronix)

[10] Elmar Rendon-Gonzalez , Volodymyr Ponomaryov," Automatic Lung Nodule Segmentation and Classification in CT Images Based on SVM",2016 9th International Kharkiv Symposium on Physics and Engineering of Microwaves, Millimeter and Submillimeter Waves (MSMW)

[11] Sanjukta Rani Jena, Dr. Thomas George, Dr. Narain Ponraj;" Texture Analysis Based Feature Extraction and Classification of Lung Cancer", 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)

[12] Nidhi S. Nadkarni, Prof. Sangam Borkar," Detection of Lung Cancer in CT Images using Image Processing", 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)