OPEN ACCESS INTERNATIONAL JOURNAL OF SCIENCE & ENGINEERING

MEDICAL IMAGE CLASSIFICATION FOR LUNG CANCER

Swastik Patro[1], Sachin Sahil[2], Deep Patel[3], Prof. Farhana Shaikh[4]

School of Computer Science and Technology MIT Academy Of Engineering, Alandi (D)Pune, India[1,2,3,4]

spatro@mitaoe.ac.in, ssahil@mitaoe.ac.in, dnpatel@mitaoe.ac.in, fjshaikh@mitaoe.ac.in

---

Abstract: Lung cancer takes more lives every year than the combined numbers of breast, colon, and prostate cancer. Diagnosis in the early stage increases the chance of survival of the patient. CT scan plays an important role in the detection of lung cancer and improved results than X-ray as CT scan is more powerful and sophisticated X-ray as it gives a 360-degree view of the organs. Due to this more number of CT scans are done and hence the radiologist has to evaluate all of these and this hampers the quality of diagnosis. Many SVM models have been used for nodule detection in Chest CT scan images but the accuracy has been more dependent on the preprocessing of the image for better accuracy, which consumes a lot of time. The purpose of this project is to address this aim and implement and train the SVM model to classify lung CT images into the nodule and non-nodule class.

Keywords—*Medical Image Classification, SVM , Lung Cancer*

I INTRODUCTION

Early detection of lung cancer increases the chances ofsurvival and can be cured easily, a CT scan of the lungs is one of the ways to reduce the risk of dying from lungcancer. As the CT scans are given to a radiologist and fora radiologist to manually check each CT scan it consumesa lot of time, as well as human error, is bound to happen. They need to segment the CT scans based on the nodule size and then forward the case to the doctors based on severity. It can be simplified using the classification of CT scan into different classes with the SVM(Support Vector Machines) andthis would reduce the time to analyse as well as human error. Medical image classification combines two fields i.e. imaging data and biomedical record. The images data areformed from the pixel values of the array, each body part of a human has a certain Hounsfield unit related to the pixel. In Computer-aided-diagnosis image classification plays a vital role to minimize the analysing time of an image. The challenges to image processing are methods and techniques used to exploit image processing results and objects here is usually not to increase accuracy but also decrease the false positives and to identify the nodule in the image (in case of lung cancer). Also understanding the features is important to image pre-processing and then find the region of interest as a complete represented feature of the segmented region has a significant issue for Image classification.

II RELATED WORK

This paper proposed an efficient lung cancer detection and prediction algorithm using multi-class SVM (Support Vector Machine) classifier. [1]Multi-stage classification was used for the detection of cancer. In every stage of classification image, enhancement and segmentation have been done separately in which binarization technique was used for foreseeing lung cancer. This system can also predict the probability of lung cancer. The algorithm can identify whether the input image contains a tumor cell or not and is ready to anticipate if there is any likelihood of growth. [2]In this paper, CT-Scan image is preprocessed and the ROI is separated in preparation for segmentation. Further, For segmentation Discrete Wavelet Transform is applied which is used in the next stage i.e. Feature Extraction. For feature extraction Gray Level Co- occurrence Matrix is used i.e. entropy, correlation, variance, dissimilarity, contrast and energy. The final model uses SVM for identifying cancerous nodules and non-cancerous nodules. The model provides better accuracy of nodule classification. In this project the clinical value of tumor heterogeneity is measured with F-FLT as a biomarker for lung cancer diagnosisand staging. [3] Support Vector Machine is used to train a vector of image features including heterogeneity extracted from PET image and CT texture features to improve diagnosisand staging of lung cancer. The SVM classification showed thebest performance of staging was achieved when all the image features are combined in the SVM training. [4]In this research paper , we learned about preprocessing of data in which we'll get the CT scan images as an input and preprocessing will be completed by such segments like segmentation , classifi cation on the basis of region of interest (ROI) , compression, transmission, reception, decompression. Then we'll perform further processes using a support vector machine (SVM). This research paper focuses on potential usage of feed forward back propagation neural networks as a judgement making for lung cancer. [5]The support vector machine is connected with the feedforward backpropagation neural network to create a hybrid algorithm that further reduces the computational complexity of the model. A set of 500 images was used, in which 75% of the data was used for training the model and the rest 25% was used for testing of the model. It follows a three step procedure for classification. The first step is preprocessing the dataset. The second step is feature extraction via SURF technique followed by genetic algorithm used for optimization. And the last step is a hybrid model of Feed Forward Back Propagation Neural Network and Support Vector Machine. [6]In this research paper, CLAHE Equalization is a technique used for preprocessing the CT scanned images and also increasing the contrast of images. Random walk segmentation is a method that is used for further segmentation. Under segmentation three processes happened, In the first process the portion containing lungs is segmented rom the CT scanned image ,the second is correction of border and the third is continuous pixel change is segmented. For feature extraction two techniques are used to foresee lung malignancy nearness i.e. binarization and GLCM. These two techniques depend on unequivocally identified with lung life systems and data of lung CT imaging. This paper discusses the use of image processing for the Lung cancer detection system. [7]This system accepts input in three formats CT, MRI and Ultrasound Images. The proposed model was developed using Genetic Optimization, PSO, and SVM algorithms for feature classification and selection. Canny filter is used for edge detection which further helps in detection of cancerous cells effectively from CT, MRI scans and Ultrasound images. For segmentation the technique used was super pixel segmentation and for de-noising the medical image, gabor filter is used. Matlab has been used for simulating the results of lung cancer detection systems. [8]In this paper, different machine learning classification algorithms are implemented for a comparative study of lung cancer detection. The dataset used for this dataset used in this research paper taken from UCI Machine Learning Repository and Data World. The algorithms used were Support Vector Machine (SVM), Naive Bayes, Decision Tree, and Logistic Regression. The highest accuracy achieved by the SVM model is 99.2%.

III.METHODOLOGY

A. Dataset

The dataset we have used is LIDC-IDRI (Lung Image Database Consortium and Image Database Resource Initiative) used by many of the research papers referred by us. It consists of around 1036 patients CT scan images along with a XML file containing the annotations of nodules present in the CT image by Radiologists. Each CT scan has been annotated by four radiologists .

B.Preprocessing

The images were in .dcm(Digital Imaging and Communications in Medicine) format , it was then processed using pydicom library and extracted metadata from the file . After reading the file , we have used the K-means algorithm (for K=2) to separate foreground (soft tissue / bone) and background (lung/air) . Then applied an erosion (of 3x3) to erode away finer elements and then diluted (of 8x8) to include some pixel values near the lung , so that the lung does not get clipped. and at the end we applied a larger dilation (of 10x10) to form the mask for lung . Also based on the metadata from the dcm files we have mapped the reading from xml files available(having the radiologist annotations) and trificated the images into three folders for further use .

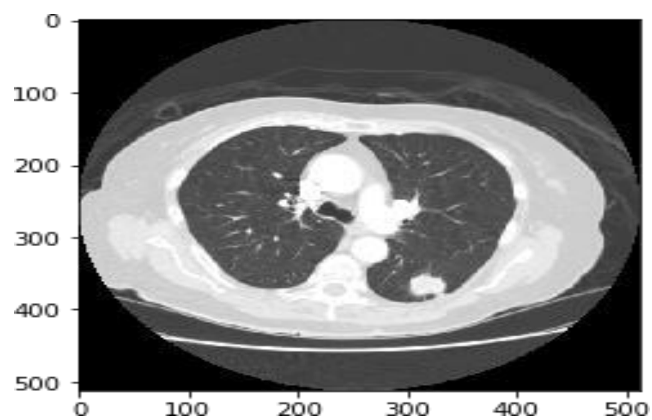

Fig. 1. Original image of CT scan (input to preprocessing module)

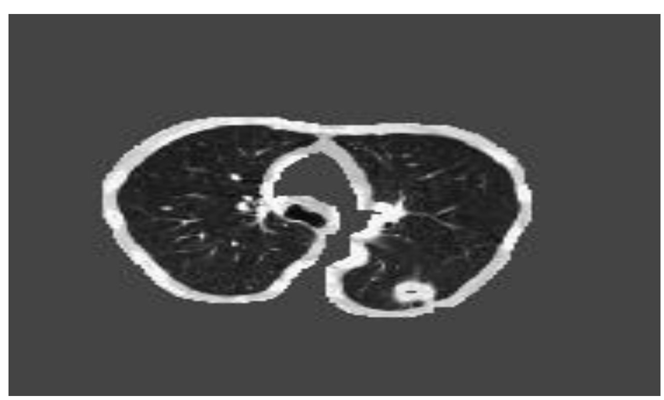

Fig. 2. Image after preprocessing and masking

C.SVM

In support vector machine algorithm the objective is to find a hyperplane in an N-dimensional space i.e. the number of features, the hyperplane helps in classifying the distinct points.There may be many possible hyperplanes for separation between two points. Thus, the maximum margin has to be chosen i.e the maximum distance between data points of both classes. This margin distance provides reinforcement so that future data points can be classified with more confidence. The required parameters for SVM are kernel, c and gamma.

D.Kernel

SVM Algorithms use a certain number of mathematical functions, these functions are known as Kernels. The basic function of the kernel is to accept data as inputs and transform it as per the required form. There are different types of kernel functions i.e. Linear, Nonlinear, Polynomial, Radial Basis Function (RBF), and Sigmoid. The kernel function used is Polynomial Function.

$$k(xi, xj) = (xi.xj + 1)^d \quad (1)$$

E.Converting images into Array

The jpg files need to be converted into a numpy array along with dividing the images into training , testing and validation folders ,the numpy array values are further flattened for it to be given as input in SVM .

F.Model Training

We have made use of the SVM algorithm. First, the model is trained to classify between nodules and non nodules using the polynomial kernel and taking image size as a flattened array of size 32x32 .Also a second model is trained to classify that the present nodule is cancerous or not .In the second model is trained to classify between cancerous nodules and non cancerous nodules using polynomial kernel and image size as 32x32 .

IV.RESULT

From the results obtained, our model performed better on some sets than others. The Accuracy of cancer classification is 85.71% and Nodule classification is 84.42% with the testing data which is comparable to previous work. Recall obtained from nodule classification is 86.24% and for cancer classifica- tion it is 88.81%. As better recall (True Positive Rate) would imply that the model is correctly able to classify the cancerous lung's CT scans and it would have greater significance in the field of medical science.. Precision of the nodule classification is 84.62% and for cancer classification it is 87.59%. The F-1 score for the nodule model is 0.85 and for cancer classification it is 0.88 . The AUC of the nodule classification is 0.84 and for cancer classification it is 0.85.

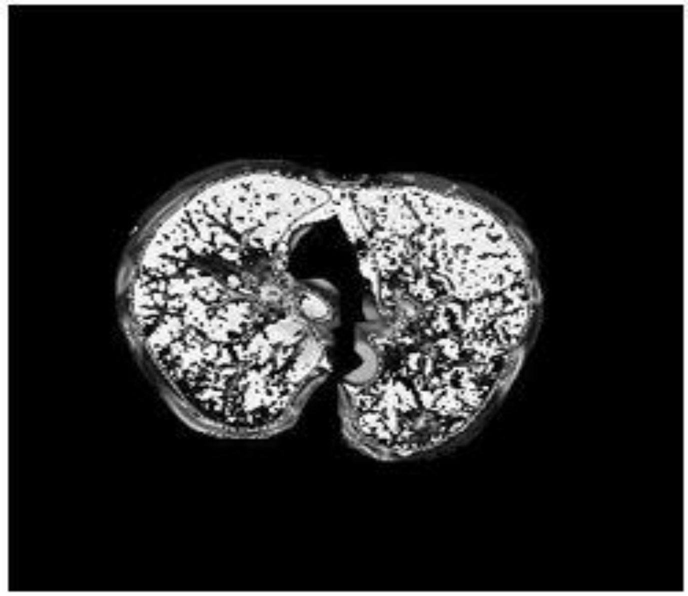

Fig. 3. Cancerous class image

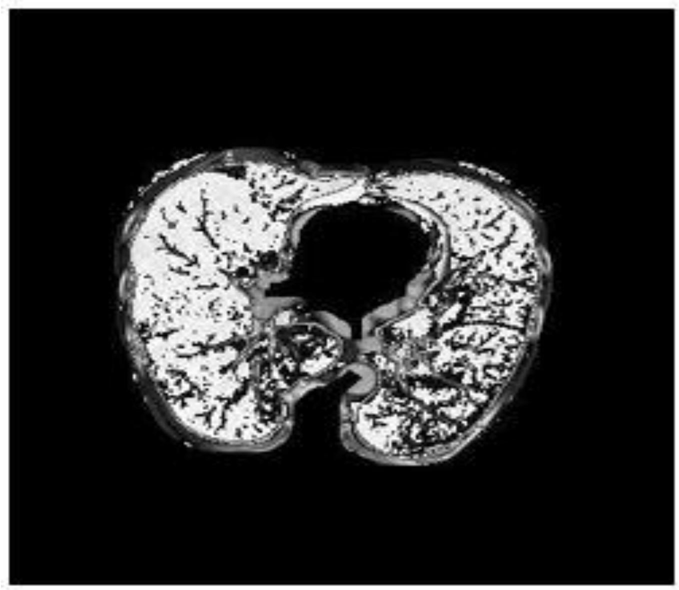

Fig. 4. Non-Cancerous class image

V.CONCLUSION

Medical image classification is an interesting research area, it combines the diagnosis problem and analysis purposes in the medical field. It has provided a detailed review of image classification techniques for the diagnosis of human body disease including imaging modalities used, each dataset and pros and cons for each technique. In the method we used was to reduce the training time without actually reducing the quality of the result , as using deep learning in this method could have given better results but the training time would have been more . More work can be done in this field with using different kernels for better results.

REFERENCES

[1] J. Alam, S. Alam and A. Hossan, ”Multi-Stage Lung Cancer De- tection and Prediction Using Multi-class SVM Classifier,” 2018 In- ternational Conference on Computer, Communication, Chemical, Ma- terial and Electronic Engineering (IC4ME2), 2018, pp. 1-4, doi: 10.1109/IC4ME2.2018.8465593.

[2] D. P. Kaucha, P. W. C. Prasad, A. Alsadoon, A. Elchouemi and S. Sreedharan, ”Early detection of lung cancer using SVM classifier in biomedical image processing,” 2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI), 2017, pp. 3143-3148, doi: 10.1109/ICPCSI.2017.8392305.

[3] N. Guo, R. Yen, G. E. Fakhri and Q. Li, ”SVM based lung cancer diagnosis using multiple image features in PET/CT,” 2015 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2015, pp. 1-4, doi: 10.1109/NSSMIC.2015.7582234.

[4] Q. Firdaus, R. Sigit, T. Harsono and A. Anwar, ”Lung Cancer Detection Based On CT-Scan Images With Detection Features Using Gray Level Co-Occurrence Matrix (GLCM) and Support Vector Machine (SVM) Methods,” 2020 International Electronics Symposium (IES), 2020, pp. 643-648, doi: 10.1109/IES50839.2020.9231663.

[5] Pankaj Nanglia, Sumit Kumar, Aparna N. Mahajan, Paramjit Singh, Davinder Rathee,A hybrid algorithm for lung cancer classification using SVM and Neural Networks,ICT Express,2020,ISSN 2405-9595,doi:10.1016/j.icte.2020.06.007.

[6] R. Sathishkumar, K. Kalaiarasan, A. Prabhakaran and M. Aravind, ”Detection of Lung Cancer using SVM Classifier and KNN Algo- rithm,” 2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN), 2019, pp. 1-7, doi: 10.1109/IC-SCAN.2019.8878774.

[7] Asuntha, A. , Brindha, A. , Indirani, S. and Andy, Srinivasan. (2016). Lung cancer detection using SVM algorithm and optimization tech- niques. 9. 3198-3203.

[8] R. P.R., R. A. S. Nair and V. G., ”A Comparative Study of Lung Cancer Detection using Machine Learning Algorithms,” 2019 IEEE Interna- tional Conference on Electrical, Computer and Communication Tech- nologies (ICECCT), 2019, pp. 1-4, doi: 10.1109/ICECCT.2019.8869001.