Lung Cancer Nodules Classification and Detection Using SVM and CNN Classifiers

Yashaswini S.L[1], K.V Prasad[2]

[1]M.Tech student, Dept of ECE, Bangalore Institute of Technology
[2]Professor & Head, Dept of ECE, Bangalore Institute of Technology

---

Abstract - *Cancer is a quite common and dangerous disease. The various methods of cancer exist in the worldwide. Lung cancer is the most typical variety of cancer. The beginning of treatment is started by diagnosing CT scan. The risk of death can be minimized by detecting the cancer very early. The cancer is diagnosed by computed tomography machine to process further. In this paper, the lung nodules are differentiated using the input CT images. The lung cancer nodules are classified using support vector machine classifier and the proposed method convolutional neural network classifier. Training and predictions using those classifiers are done. The Nodules which are grown in the lung cancer are tested as normal and tumor image. The testing of the CT images are done using SVM and CNN classifier. Deep learning is always given prominent place for the classification process in present years. Especially this type of learning is used in the execution of tensor Flow and convolutional neural network method using different deep learning libraries.*

Key Words: Lung cancer, deep learning, biomedical image classification, confusion matrix, microdicom.

1. INTRODUCTION

Lung cancer is recognized as the main reason behind the death caused due to cancer in the worldwide. And it is not easy to identify the cancer in its early stages since the symptoms doesn't emerge in the initial stages. It causes the mortality rate considered to be the highest among all other methods of cancer. The number of humans dies because of the dangerous lung cancer than other methods of cancer such as breast, colon, and prostate cancers. There exist enormous evidence indicating that the early detection of lung cancer will minimize mortality rate. Biomedical classification is growing day by day with respect to image. In this field deep Learning plays important role. The field of medical image classification has been attracting interest for several years. There are various strategies used to detect diseases. Disease detection is frequently performed by observant at tomography images. Early diagnosis must be done to detect the disease that is leading to death. One among the tools used to diagnose the disease is computerized tomography. Lung cancer takes a lot of victims than breast cancer, colon cancer and prostate cancer together. This can be a result of asymptomatic development of this cancer. The Chest computed tomography images are challenging in diagnostic imaging modality for the detection of nodules in lung cancer. Biomedical image classification includes the analysis of image, enhancement of image and display of images via CT scans, ultrasound, MRI. Nodules within the respiratory organ i.e. lung are classified as cancerous and non-cancerous. Malignant patches indicate that the affected person is cancerous, whereas benign patches indicate an affected person as a non- cancerous patient. This can be done using various classifiers.

2. EXISTING SYSTEM

Support Vector Machines is a method of machine learning approach taken for classifying the system. It examines and identifies the classes using the data. It is broadly used in medical field for diagnosing the disease. A support-vector machine builds a hyper plane in a very high or infinite-dimensional area, which can be utilized for classification, regression, or totally different operation like outliers detection.

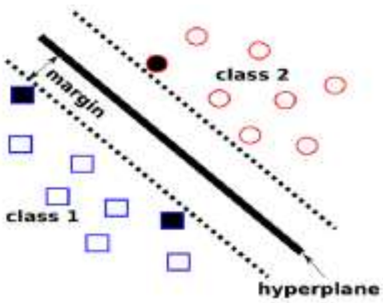


Fig -1: The SVM classifier representation.

Based on a good separation is obtained by the hyper plane in the SVM. After classification if the gap is large to the nearest training-data pictures of any class referred as functional margin, considering that in generally the larger the margin, the lesser the generalization error of the classifier. Fig-1 shows the support vector machine classifier that constructs a maximum margin decision hyper plane to separate two different categories. Support Vector Machine is a linear model applied for the classification and regression issues.

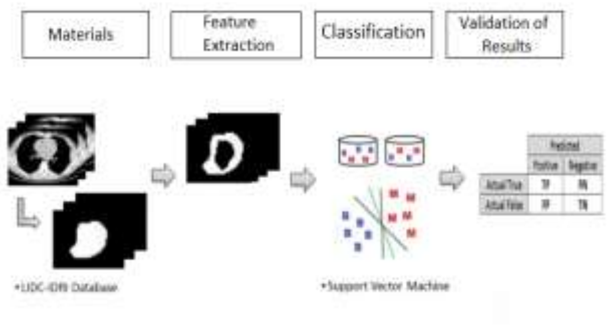


Fig -2: Training and prediction using SVM

SVM algorithm finds the points closest to the line from both. The classes of these points are referred as support vectors. The mixed data of tumor nodules and normal nodules are provided as input In SVM algorithm the input images given are trained and the results are predicted, tuning the various parameters. Fig-2 shows the training and prediction using SVM. Input images undergo feature extraction. At the training the various SVM parameters are tuned, and then the predictions are made using the hyper plane of SVM.

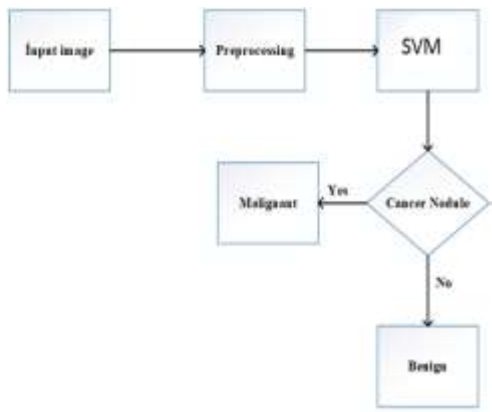


Fig -3: Testing phase of SVM

Fig-2 shows the training and prediction using SVM. Input images undergo feature extraction. At the training the various support vector parameters are tuned then the predictions are made using the hyper plane of SVM. In the testing phase the nodules in the lung cancer are classified as normal or tumor nodules. Fig 3 shows testing phase of SVM. Initially the input images are pre-processed. Later SVM operation takes place. The cancer nodules undergo testing processes. The CT scanned image undergoes median filtering and estimates whether the nodules are malignant or benign. Then the output will be shown as normal image or tumor image.

3. PROPOSED METHODOLOGY

Convolutional neural networks encompass of multiple layers in its structures. CNN could be feed forward and extremely tremendous approach especially in detection. Network structure is built easy; has less training parameters. A convolution neural network have multiple layers within the neural network, that consists of one or a lot of convolution layers and so succeeded by one or more fully connected layers as in a standard multiple layers in neural network. Convolution neural network architecture is typically employed collaboration with the convolution layer and pool layer. The pooling layer is seen between convolution layers. It confuses the features of the particular position. Since not all the location features are not important, it just needs other features and the position. The pooling layer operation consists of max pooling and means pooling. Mean pooling calculates the average neighbourhood inside the feature points, and max pooling calculates the neighbourhood inside a maximum of feature points.

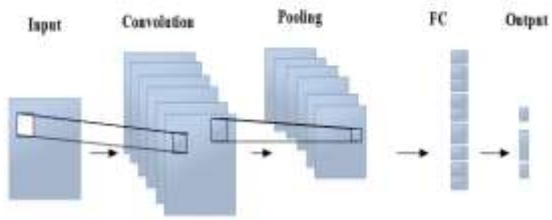


Fig -4: Structure of CNN model

A CNN uses the learned features with input and make use of 2D convolutional layers. This implies that this type of network is best for processing 2D images. Compared to other methods of image classification, the network uses very little pre-processing. This means that they can use the filters that have to be built by user in other algorithms. CNNs can be utilized in various applications from image and video recognition, image classification, and recommender systems to natural language processing and medical image analysis.

1. Input: This layer have the raw pixel values of image.

2. Convolutional Layer: This layer gets the results of the neuron layer that is connected to the input regions. We define the number of filters to be used in this layer. Each filters that slider over the input data and gets the pixel element with the utmost intensity as the output.

3. Rectified Linear Unit [ReLU] Layer: This layer applies an element wise activation function on the image data. We know that a CNN uses back propagation. Thus in order to retain the equivalent values of the pixels and not being modified by the back propagation, we apply the ReLU function.

4. Pooling Layer: This layer performs a down-sampling operation along the spatial dimensions are width and height, resulting in volume.

5. Fully Connected Layer: This layer is used to compute the score classes i.e. which class has the maximum score corresponding to the input digits.

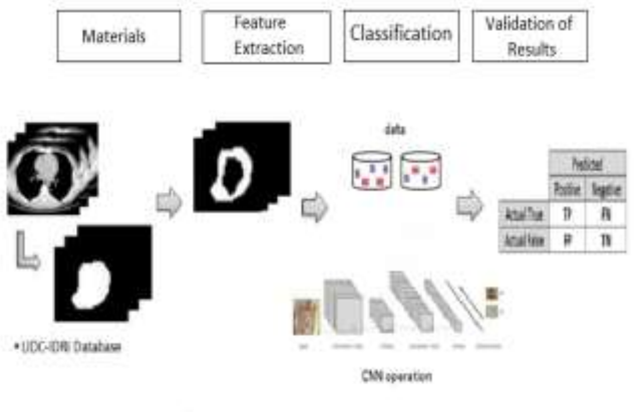


Fig -5: Training and prediction using CNN.

The datasets used are public databases that are utilized for diagnosis of carcinoma. The most frequently applied data from these datasets are LIDC-IDRI dataset. Dataset is having a carcinoma screening and computerized tomography scans, processed with the feature extraction.

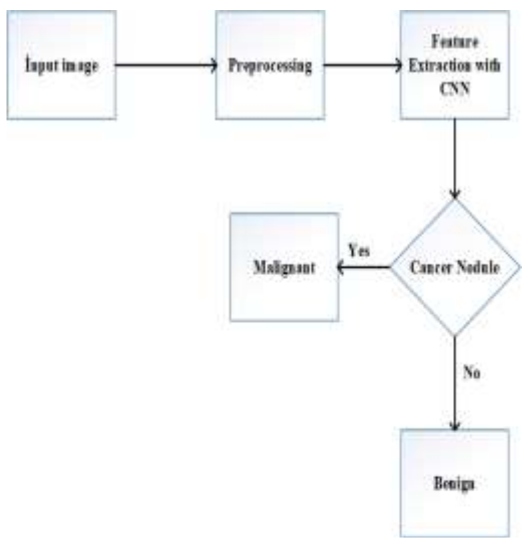


Fig -6: Testing phase of CNN

Then the step by step convolutional neural network operation takes place. The confusion matrix is predicted. Fig-6 shows the testing flow of CNN. The testing phase of CNN detects whether the cancer nodule is malignant or benign.

4. EXPERIMENTAL RESULTS

The dataset used in this paper is a collection of CT images of the carcinoma affected persons and also normal persons. Those images are of DICOM format, every individual image is having a multiple axial slices of the chest cavity. Those slices are displayed in the 2d form of slices. All the medical images are stored in microdicom format. The input image of dicom format is transformed by converting to .png, bmp and jpg format. The pydicom package which is available for spyder environment is used. The python language works good with all the dicom format images. At valuation several metrics are utilised. Using confusion matrix, the performance is calculated. The binary classification technique is also realized. Confusion matrix is the easily understandable metrics used to find the model's accuracy. The accuracy of the system is determined by looking at the TN, TP, FN, and FP. The Results for the SVM classifiers are shown as various parameters like confusion matrix, accuracy score, and reports are extracted. Then followed by receiver operating characteristic curve is obtained.

Table -1: Confusion matrix of the SVM

4	1
1	4

Accuracy score [0.8]

Table -2: Report for the SVM classifier

classes	Precision	Recall	f1-score	support
0.0	0.80	0.80	0.80	5
1.0	0.80	0.80	0.80	5
micro average	0.80	0.80	0.80	10
macro average	0.80	0.80	0.80	10
weighted average	0.80	0.80	0.80	10

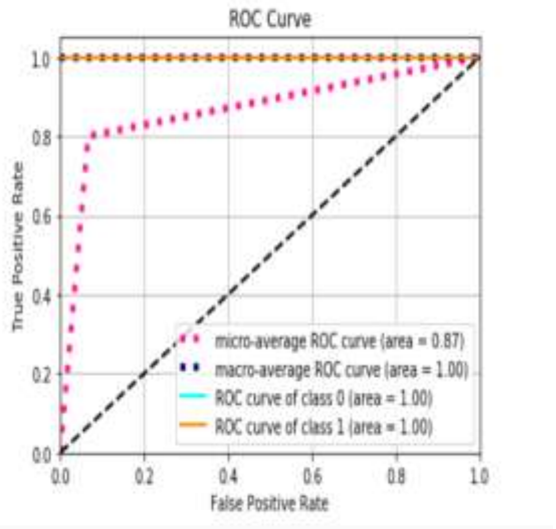


Chart -1: ROC curve for SVM

The Results for the CNN classifiers are as shown below various parameters like confusion matrix, accuracy score, and reports are extracted and followed by receiver operating characteristic curve. In the report the precision, recall,f1-score, support are obtained for both the classes.

The classes are both normal images and tumor images.

Table -3: Confusion matrix of the CNN

6	0
1	3

Accuracy score [0.9]

Table -4: Report for the CNN classifier

classes	Precision	Recall	f1-score	support
0.0	0.86	1.00	0.92	6
1.0	1.00	0.75	0.86	4
micro average	0.90	0.90	0.90	10
macro average	0.93	0.88	0.89	10
weighted average	0.91	0.90	0.90	10

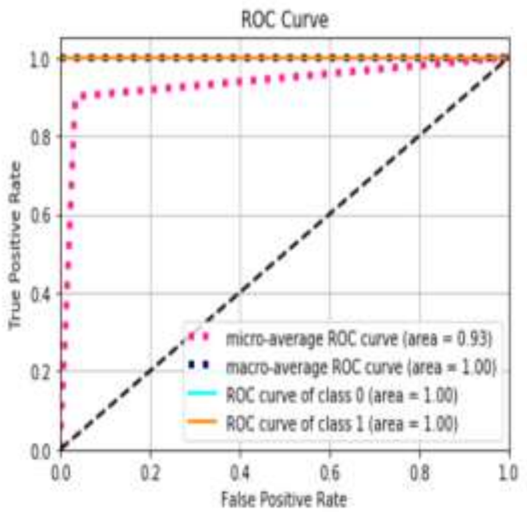


Chart -2: ROC curve for CNN

Comparison of SVM and CNN classifier is shown using the accuracy score and different parameters like precision, recall, f1-score,support parameters. The accuracy score with respect to two different classes and with the two different classifiers are obtained.

Table -5: Accuracy score

classifier	SVM	CNN
accuracy	0.8	0.9

Table -5: Comparison of various parameters of SVM and CNN

REPORT	SVM		CNN	
Classes	0.0	1.0	0.0	1.0
Precision	0.80	0.80	0.86	1.00
Recall	0.80	0.80	1.00	0.75
f1-score	0.80	0.80	0.92	0.86
Support	5	5	6	4

After testing, output is obtained using the SVM classifier and CNN classifiers to detect whether the nodules are malignancy or benign. The individual operation for SVM and CNN is obtained using the CT images as the input.

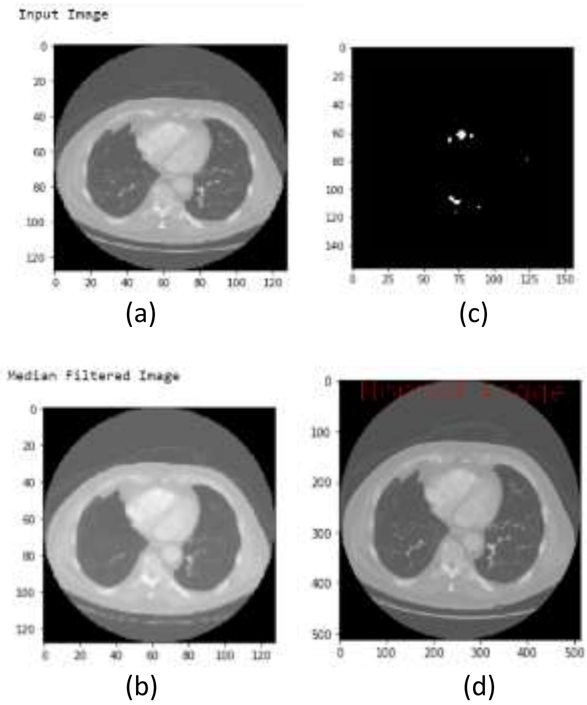
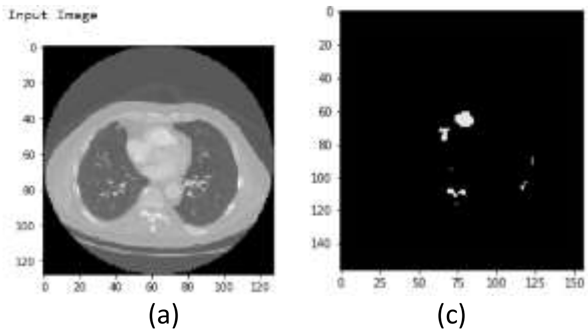
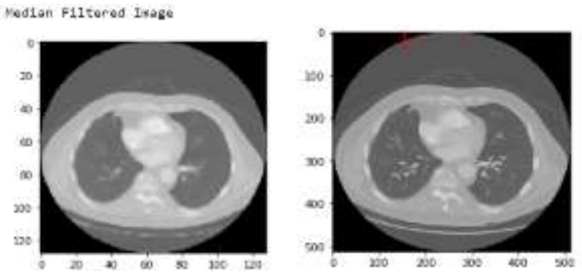


Fig -7: Lung cancer CT scans (a) Input image, (b) Median filtered image, (c) Nodules representation, (d) Detection of nodule as normal nodules.





(b) (d)

Fig -8: Lung cancer CT scans (a) Input image, (b) Median filtered image, (c) Nodules representation, (d) Detection of nodule as tumor nodules

5. CONCLUSION

This study draws attention to the diagnosis of lung cancer. Lung nodule classification is benign and malignant. The proposed method CNN architecture is specially regarded for its success in image classification compared to support vector machine. For biomedical image classification operation, it also obtains successful results. CNN architecture is used for classification in the study. Experimental results show that the proposed method is better than the support vector machine in terms of various parameters. The images in the data set used are rather small. In the future, the performance of the system can be improved with a larger dataset and an improved architecture. The proposed system is able to detect both benign and malignant tumors more correctly.

ACKNOWLEDGEMENT

We express our thanks to Department of Electronics and Communication, Bangalore Institute of Technology for providing their constant support and encouragement.

REFERENCES

[1] Anita Chaudhary, Sonit Sukhraj Singh “Lung cancer detection on CT images using image processing”, is computing sciences 2012 international conference, IEEE, 2012.

[2] G. Guo, S. Z. Li, and K. Chan, “Face recognition by support vector machines,” In Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition, Grenoble, France, pp. 196-201, March 2000.

[3] D. Nurtiyasari, and R. Dedi, "The application of Wavelet Recurrent Neural Network for lung cancer classification," Science and Technology-Computer (ICST), 2017 3rd International Conference on. IEEE, 2017.

[4] K. He, X. Zhang, S. Ren, and J. Sun, J. “Deepresidual learning for image recognition,” In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778). 2016.

[5] E. Dandıl, M. Çakiroğlu, Z. Ekşi, M. Özkan, Ö. K. Kurt, and A. Canan, “Artificial neural network-based classification system for lung nodules on computed tomography scans,” in 6th International Conference of Soft Computing and Pattern Recognition (soCPar), pp. 382–386, IEEE, 2014.

[6] J. Cabrera, D. Abigaile and S. Geoffrey, "Lung cancer classification tool using microarray data and support vector machines," Information, Intelligence, Systems and Applications (IISA), 2015 6th International Conference on. IEEE, 2015.

[7] H. M. Orozco and O. O. V. Villegas, “Lung nodule classification in CT thorax images using support vector machines,” in 12th Mexican International Conference on Artificial Intelligence, pp. 277–283, IEEE, 2013.

[8] H. Krewer, B. Geiger, L. O. Hall et al., “Effect of texture features in computer aided diagnosis of pulmonary nodules in low dose computed tomography,” in Proceedings of the IEEE InternationalnConference on Systems, Man, and Cybernetics (SMC),2013, pp. 3887–3891, IEEE, Manchester, United Kingdom,2013.

[9] M. Abadi, P. Barham, J. Chen, Z. Chen, A.Davis, J. Dean, M.cDevin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, “TensorFlow: A System for Large-Scale Machine Learning,” InOSDI 2016 Nov 2 (Vol. 16, pp. 265-283).

[10] Arvind Kumar Tiwari, “Prediction Of Lung Cancer Using Image Processing Techniques”, Advanced Computational Intelligence: An International Journal (Acii), Vol.3, No.1, January 2016.

[11] Raquel Kolitski Stasiu, Gerson Linck Bichinho, “Components Proposal for Medical Images and HIS”, Proceedings of the Fourteenth IEEE Symposium on Computer-Based Medical Systems, Page 73, 2001, ISSN:1063-7125.