

Classification of normal and abnormal images of lung cancer

Divyesh Bhatnagar, Amit Kumar Tiwari, Vijayarajan V and Krishnamoorthy A

School of Computer Science and Engineering, VIT University, Vellore-632014, India

E-mail: vijayarajan.v@vit.ac.in

Abstract. To find the exact symptoms of lung cancer is difficult, because of the formation of the most cancers tissues, wherein large structure of tissues is intersect in a different way. This problem can be evaluated with the help of digital images. In this strategy images will be examined with basic operation of PCA Algorithm. In this paper, GLCM method is used for pre-processing of the snap shots and function extraction system and to test the level of diseases of a patient in its premature stage get to know it is regular or unusual. With the help of result stage of cancer will be evaluated. With the help of dataset and result survival rate of cancer patient can be estimated. Result is based totally on the precise and wrong arrangement of the patterns of tissues.

1. Introduction

Lung cancer detection in early stage is the important part and key to cure. Generally, it can be diagnosed with the help of X-rays, Chest films, CT scan, MRI, and with many more techniques. Among all these, there is a vital measure called neurotic determination that analyses examples of needle biopsies that will be acquired from the bodies of subject that is going to diagnosed. In last decades it was very difficult to diagnosed Lung cancer at early stages with the help of image processing and pattern recognition. But with the time new hybrid Lung cancer detection method come into existence and accuracy of diagnosis also get improved. Many methods were proposed that were based on the subtraction between two serial mass chest radiography. This was proposed to detect new lung nodules. Some methods are used to extract and analysed features of the pulmonary blood vessel region and lung and then diagnosed with different rules. Result obtained from diagnosis was used for the detection of tumour. Until now there is rigorous theory indicate that gives error free approximation. Artificial neural network have already been widely exploited in this research but now it get improved with the generalization ability which include training of several artificial neural networks and combining their predictions. Lung cancer diagnosis done with the help of X-ray chest film gives more effective results. Here LCDS is used to analysed or



handled with those cases that could not be diagnosed clearly by X-ray chest films. With the help of filters image processing can be done better. That results colour and contrast of the image get change.

2. Related Works

The work which improved the image quality and assessment of lung cancer and breast cancer images are found in [1]. The method [2] called high resolution computed tomography has changed the diagnostic evaluation of patients with ILD (interstitial lung disease). Using this image analysis approach, the doctor can easily diagnose idiopathic pulmonary brosis problem. The work achieved the correlation between the image frames over longer length of period is found in [3]. Also, it compared its efficiency with traditional video compression techniques with better accuracy. The better categorization of x-ray chest films was investigated using association rules are found in [4]. The main advantage of this work is that it preserves all data in the multimedia format and that would help the doctor to retrieve and analyse the data easily. The research done with an interface of combining coordinate protein and tumour cooperation system is found in [5]. The work helped to identify the point of interest of tying locals. The information extracted can be stored as a distributed system using cloud environment is found in [6]. The doctor as well as the patient can retrieve their information from this cloud for the future purposes. The classification of images and an retrieval system found in work [7] could be used for building a new natural language based query retrieval system for lung cancer record maintenance.

3. Methodology

We have implemented our system in MATLAB and testing is done with CT scan images.

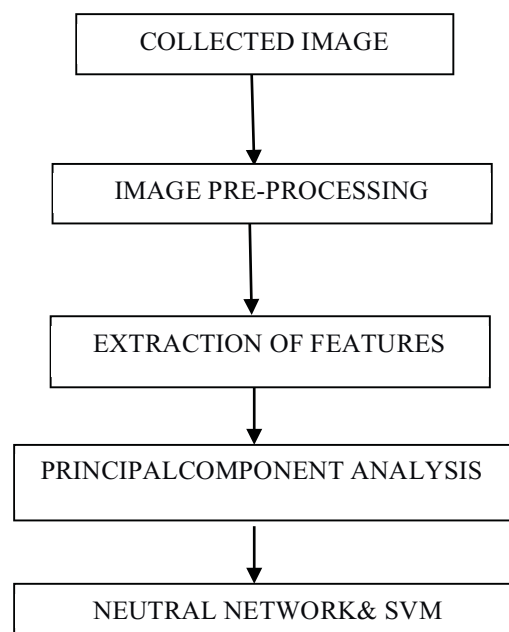


Figure 1. The work flow architecture

With the help of computed tomography images of lung cancer Pre-processing of Images can be done as shown in Figure 1. Image quality is affected by noise, motions so pre-processing is done basically for the better quality images. In this paper Histogram Equalization is to improve the quality of images and to reduce the redundancy present in images as shown in Figure 2.

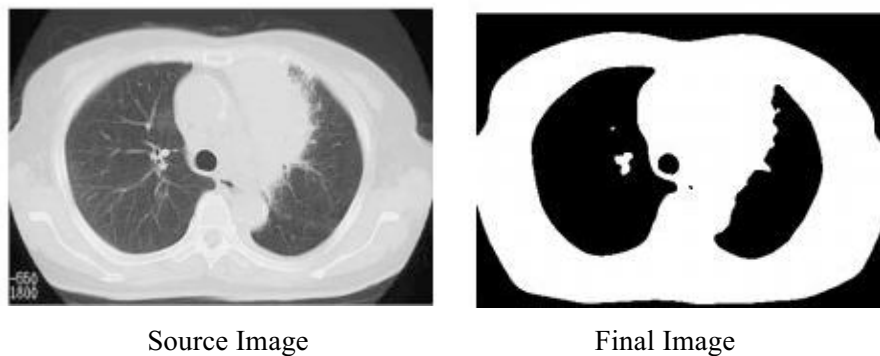


Figure2. After applying Histogram Equalization on CT scan image.

Operators of morphology include morphological operations. Morphological operations are applied on binary images. Binary image is the image with only two colours black and white. They are used to get the information about shape, convex hull /region of images. Next is extraction of features. Which is very important step and that is used to detect and separate the shape of given images. To predict the probability of lung cancer presence, we are using: binarization and GLCM, these two methods works on anatomy of lung and works on computing tomography. Grey Level Co-Occurrence Method is a method which is used to show the information in tabular form, it is used to show how different combinations of pixel value present in an image. We will use the grey co-matrix function in MATLAB for creating the grey-level co-occurrence matrix from image after that normalization of GLCM will be done with the help of proposed equation

$$P_{i,j} = V_{i,j} / \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} V_{i,j}$$

i represent number of rows and j represents number of column. From GLCM method different feature are extracted such as Probability, shade of cluster, Prominence of cluster. The MATLAB code for GLCM

```
p1=imread("img.jpg");
v1=rgb2gray(p1);
c1=imresize(v1,[128 128]);
t1=graycomatrix(c1);
disp(t1);
g1=imhist(c1);
```

Entropy

It tells the amount of image needed for compression

$$\sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} -P_{ij} * \log P_{ij}$$

Correlation

It measures the dependency of grey levels neighbouring pixels

$$\sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} (i,j)p(i,j) - \mu_x \mu_y$$

1.1. Binarization approach

Binarization check method works as shown in flow diagram which is discussed below

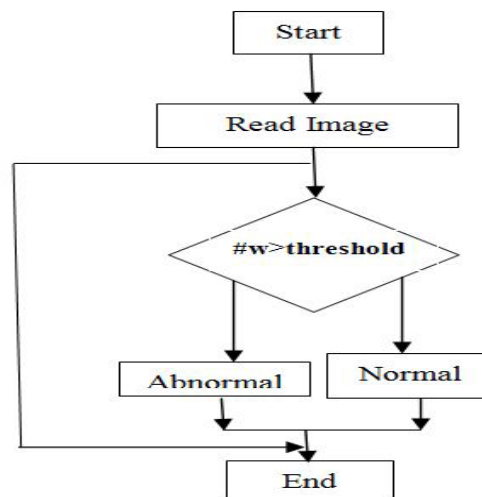


Figure 3. Binarization check method

In Binarization check method as shown in Figure 3, we have used the following steps-

1. Initially we will extract the number of white pixels.
2. Compare the white pixels with threshold value. If it is less than threshold value then it is normal else abnormal.
3. After combining the result of GLCM and Binarization approach we can conclude that whether the case is of normal or abnormal.

2.2 PCA (Principal component Analysis)

PCA is a method which uses an orthogonal transformation to convert values of set of possibly M correlated images into set of value of K uncorrelated variables called Eigen face. Value of Eigen faces is always equal or less than to the number of original face images i.e. $K < M$.

For better classification of images we will pass the extracted features through the PCA

- With the help of MATLAB calculate the mean and standard deviation.
- From every perception, subtract the specimen mean after that separation with test standard deviation. Calculate the coefficients and variance of PCA.
- Matrix will contain coefficients of all components and their respective variance will be stored in diagonal elements.
- The maximum variance in the result is required for better classifications.

2.3 NNC & SVM (Neural Network Classifier)

In this paper as a classifier tool we will use the supervised feed-forward back-propagation neural network. Neural network classifier is different from the other classifier like Bayesian and k – nearest neighbor classifiers. One important difference between the neural network classifier and traditional classifier is linearity of data. Neural network classifier can work on nonlinear data while traditional classifier requires linear data. To works with the higher accuracy neural network require wide range of input data for training. There are three layers in the architecture of the neural network initial layer, second is hidden layer and outer layer. The node of the initial layer is combined to the node of the second hidden layer and similarly the node of the second hidden layers is connected to the node of the outer layers. The outer layer consists at least one response variable.

Steps which are performed in neural network classifier –

1. First of all maintain a propagation feed-forward back network.
2. After creation of network train it with help of operations and sample data which are defined for it.
3. For the test sample, the input image is the extracted MAWH standardized data and to model the network to verify the input giving result patient is having symptom of cancer or not.
4. Classification rate is calculated from the result of neural network and sample data using some mathematical formula as shown in Figure 4.

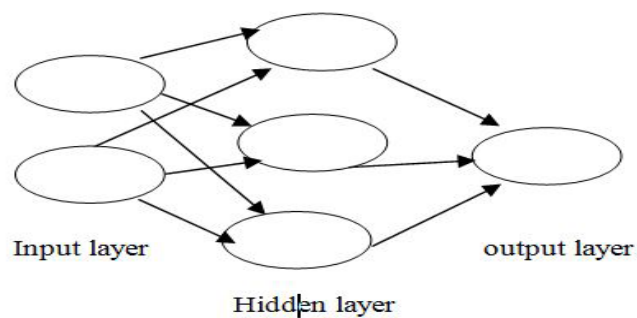


Figure 4. The neural net model

Survival rate and year

Survival rate of lung cancer is calculated on the basis of survival rate of people who are still alive having symptom of cancer. For example, a 4-year's rate of survival is 60% for a condition that would mean that 60% of survival rate will be there, after 4 years. Other people may or may not come under this category. The overall architecture is discussed in Figure 5 and 6.

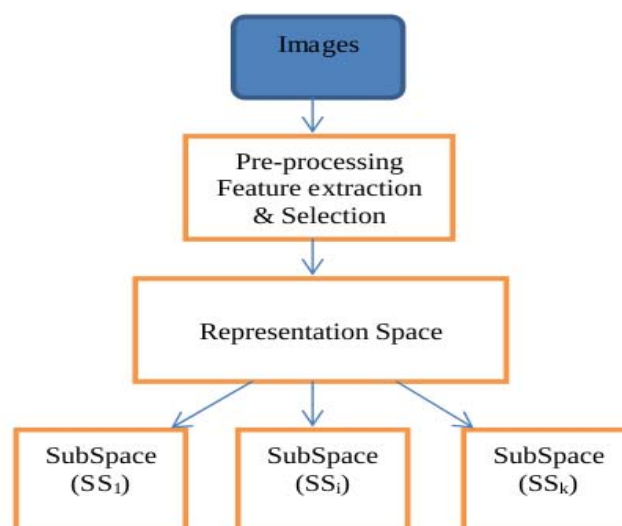


Figure 5. A Novel combination model (ANN_SVM)

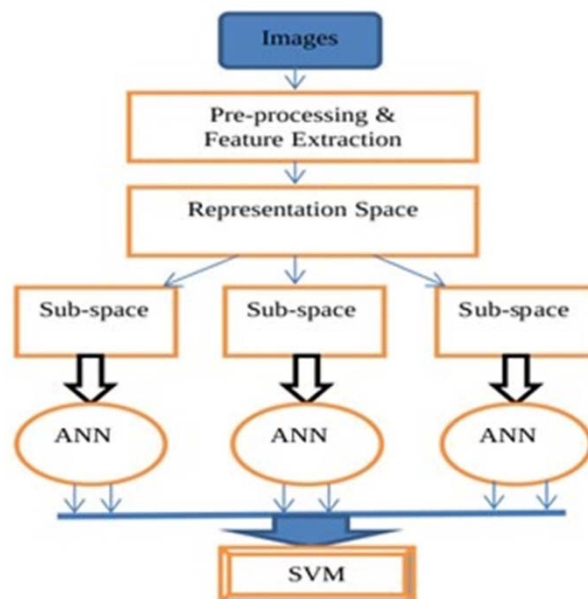


Figure 6. Using SVM to aggregate the classify results of all sub images

4. RESULTS AND DISCUSSION

After image processing some feature has been extracted and lung CT-SCAN has been collected then classification is carried out with the help of neural network to verify whether the image is normal or abnormal. The neural net training, the abnormal images and normal images are shown in Figure 7, 8 and 9 respectively.

Neural network classifier training

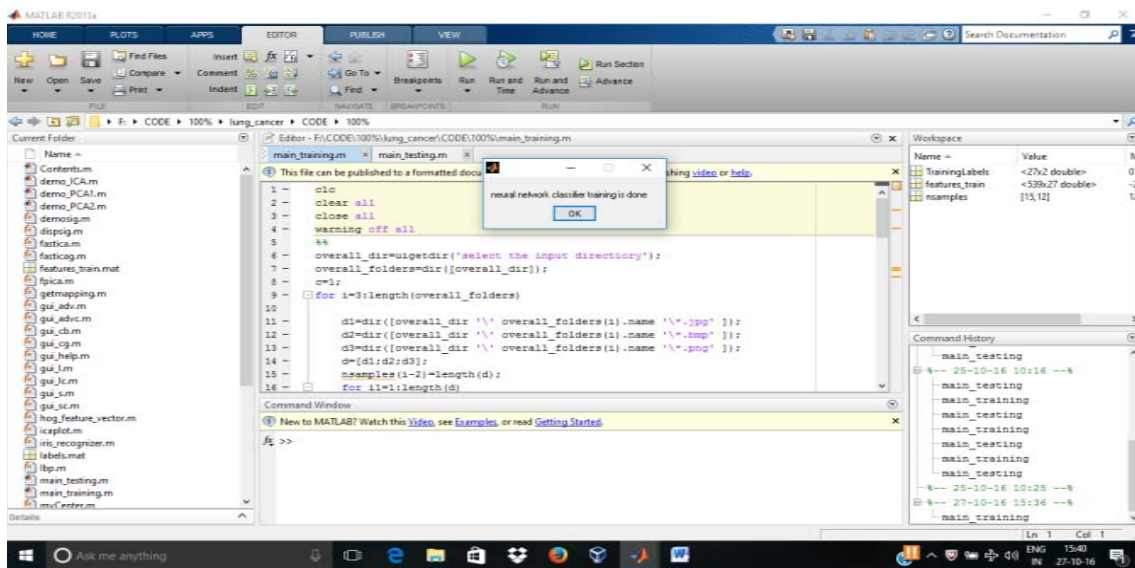


Figure 7. The neural net training

For abnormal lungs

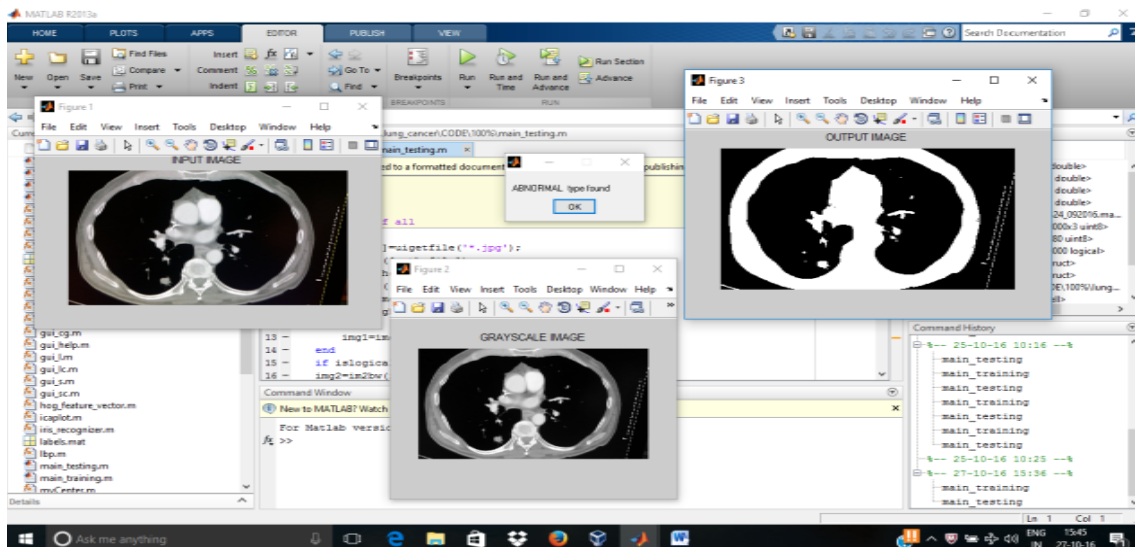


Figure 8. The abnormal lungs

First, load the database after that perform the training of all images and test all trained images. After that we will give that image and then calculate the feature extracted. After the feature extraction our next step is classification of the image which will be done with the help of the neural network. The algorithm comparisons are shown in Figure 10-13.

For normal lungs

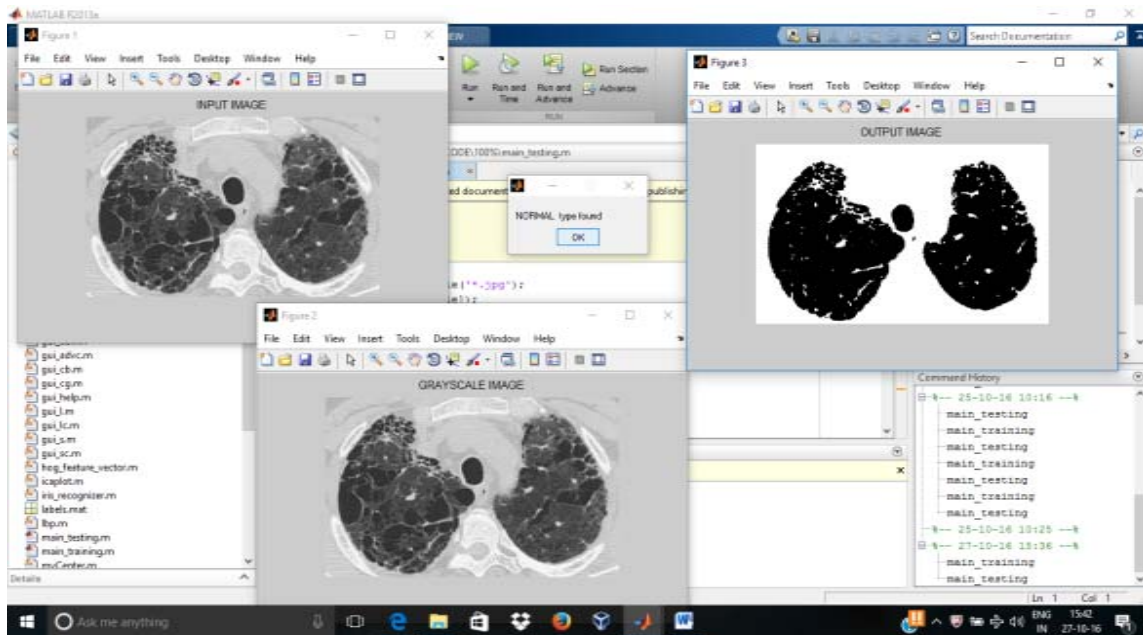


Figure 9. The normal lungs

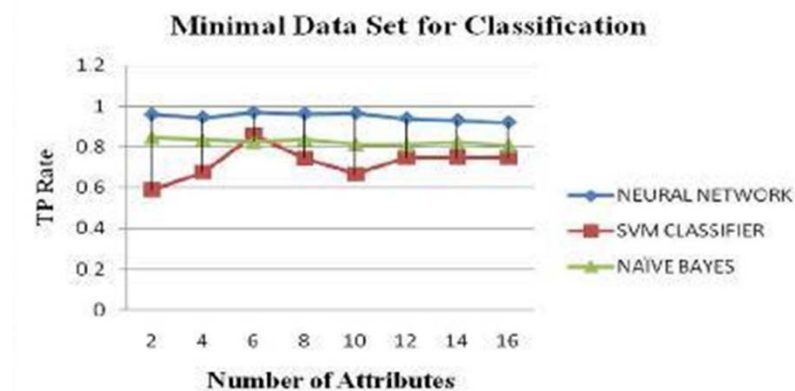
*COMPARISION BETWEEN ALGORITHMS:**A. USING TP RATE*

Figure 10. The TP Rate

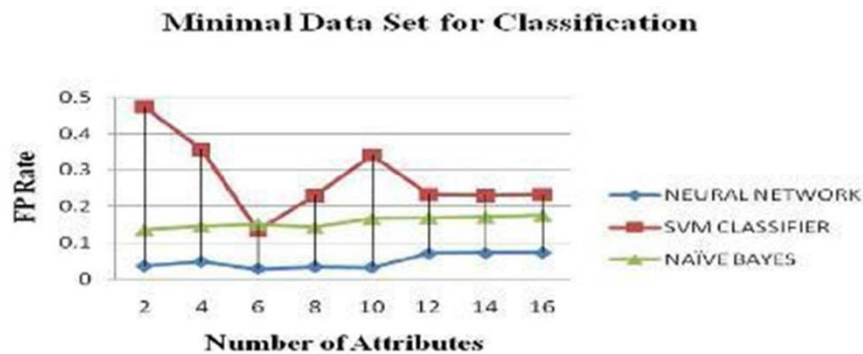
B. USING FP RATE

Figure 11. The FP Rate

C. ROOT MEAN SQUARE ERROR

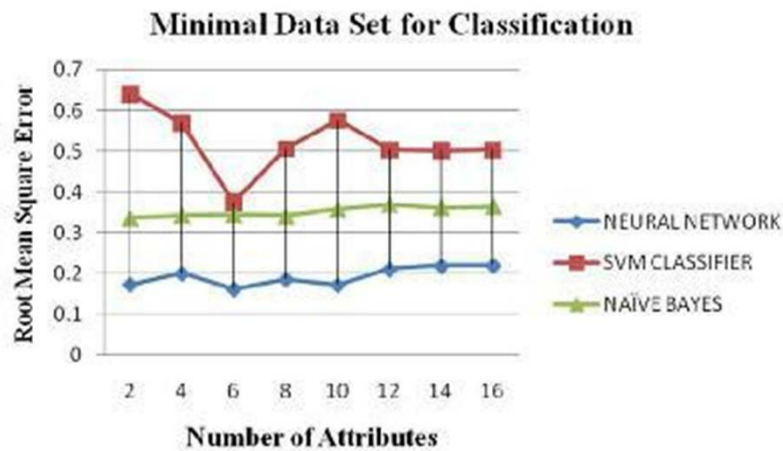


Figure 12. The root means square error

COMPARISON OF CLASSIFIERS AT 12 FEATURES: To find out which algorithm classifies maximum number of instances correctly. 30% of attributes has been taken in training phase and remaining attributes during the testing phase.

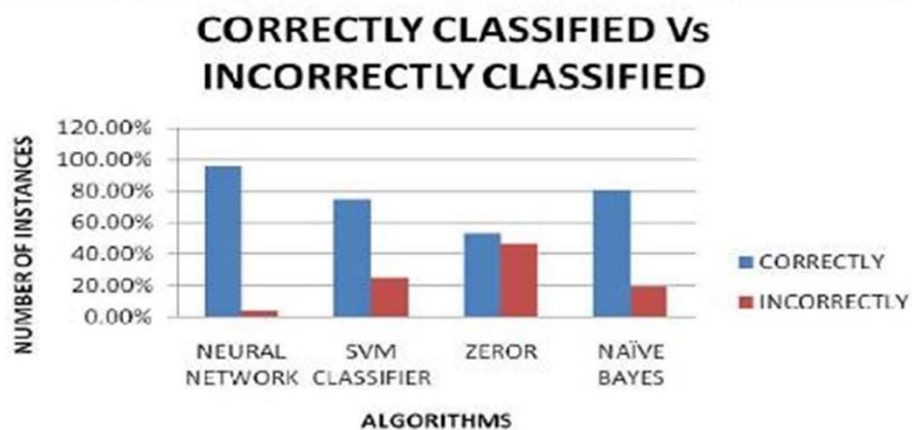


Figure 13. Comparison of classifiers at 12 features

5. CONCLUSION:

Using MATLAB and its tools Neural Network algorithm is implemented. Idea of execution is compared with other grouping calculations. Histogram equalization done to improve the quality of image and to reduce the redundancy present in image. With Binarization approach result can be concluded that whether the case is of normal or abnormal.

References

[1] AL-TARAWNEH, M. S. (2012). Lung growth location utilizing picture preparing strategies.

Leonardo Electronic Journal of Practices and Technologies, 11(21), 147-58.

[2] Albregtsen, F. (2008). Factual surface measures figured from dark level co-occurrence networks. Picture preparing research center, bureau of informatics, college of oslo, 5.

[3] Yang J Zhang, Frangi, and Yang, J.Y (2004). 2D PCA: another way to deal with appearance-dependent face representation and acknowledgment. 26(1), 131-137.

[4] Zubi, Z. S., and Saad, R. A. (2011, February). Utilizing a few information digging methods for early finding of Lung malignancy. In Proceedings of the tenth WSEAS worldwide gathering on Artificial Insight Knowledge Engineering and Data Bases (pp. 32-37).

[5] Ganeshan, Abaleke, Young, Chatwin, C. R., and Miles (2010). Surface investigation of non-little cell lung growth. Tumor imaging, 10(1), 137-143.

[6] Kannadasan, R., Sivashanmugam, G., Vijayarajan, V., Prabakaran, N., Krishnamoorthy, A., & Naresh, K (2016). An Integrated Approach for Configuring Hadoop Clusters by Ambari on Horton Sandbox.

[7] Vijayarajan, V., Dinakaran, M., Tejaswin, P., & Lohani, M. (2016). A generic framework for ontology-based information retrieval and image retrieval in web data. *Human-centric Computing and Information Sciences*, 6(1), 18.