

A New Framework for Early Detection and Diagnosis of Lung Lesion using Various Classifiers

Lim J Seelan, L. Padma Suresh

Abstract: Lung malignant is one of the most dangerous forms of cancer because it claims more than a million precious lives every year. So, lung nodule detection in chest Computed Tomography (CT) images becomes very necessary in the present clinical world. Thus the Computer Aided Diagnosis (CAD) arrangement is particularly important for early finding of lung cancer in this proposed CAD system. Initially the preprocessing technique is performed for enhancing, subsequently the lung extraction, lung border correction and lung segmentations are performed for finding the region of interest. After that, the feature values are calculated for the particular ROI. Finally, using the classifications techniques the overall performance of the proposed models is calculated.

Index Terms: curvelet transform, adaptive concave hull, optimized chan-veese algorithm, SVM, SVM naïve bayes, ANN.

I. INTRODUCTION

Cancer is the improvement of strangely and hysterical cells. It can harm the encompassing tissue extend to other pieces of the body. The end phase of the malignant growth development can prompt death; it could create from each cell kind in the human parts. Lung cancer is the most widely recognized explanation behind the malignant growth passing's, this sort of disease not just purpose the largest amount of male passing's, yet in addition motivation the female death. The degree of death rate may diminish if the development can be recognized and treated early, the danger of restoration will be high [1]. Easy distinguishing proof of lung disease will help to expand the survival rate and to improve the patient health. By utilizing computer tomography (CT) check the malignant growth is effectively identify (Computed Tomography). The figure of CT output will give the phase of lung malignant growth and furthermore various outcomes among typical and irregular lungs. Assessment system by along these lines requires gear that is to some degree intricate, costly and needs medicinal investigator in this procedure. Indeed, even restorative expert can likewise commit errors in separating ordinary and unusual lungs. Therefore, numerous analysts have given an assortment of substitute alternatives that can help specialists by using diverse image processing strategies [2].

Messay et al. [3] proposed Thresholding and rule based analysis to identify the lungs and detect nodules. Then a support vector machine and rule based analysis is used to

detect the number of false positives. Kuppusamy et al [4] used ant colony optimization to detect the edges and then the output was fed to a black circular neighbourhood algorithm. The output of this phase was the center of the detected nodules. Carvalho[5] explained the classification of lung diseases here used genetic algorithm to select the best model and features followed by a SVM for classification. Abduh et al [6] defined small windows for each ROI to calculate features, then a stepwise feature selection(SFS) algorithm was utilized to filter the best features were used as an input to KNN and SVM classifiers. Chiou et al. [7] proposed new technique for lung nodule detection by neural network based on hybrid system; this model is fully based on artificial neural network architectures using this technique the diagnostic accuracy and speed of the network were improved. P. Aruna et al. [8] suggested classification model based on decision tree, K nearest neighbor (KNN), Multi layer perceptron Neural Networks (MLP-NN) algorithm, based on the fusion based feature extraction technique the images are preprocessed and features are extracted, and next by applying genetic algorithm the features are selected. The authors explain various methods for detecting the nodules at an initial stage but all the method has some limitations. To overcome the problems the novel classifiers are proposed to detecting and classifying the tumor.

The novel classifier technique is based on the SVM, SVM naïve bayes and ANN classification model. This method is more helpful in the medical images for identifying the tumor and classifying the tumor cells. The resultant accuracy for various classifications is improved compared with the other modality image. The organization of the paper is as follows: Section 1 describes the introductory part of the proposed system and brief explanation about several researches papers, Section 2 emphasizes the proposed system, section 3 explains the experimental results and section 4 concludes the paper.

II. METHODOLOGY

Computer-aided detection (CAD) helps specialists in advanced image investigation and furthermore serves to radiologist for examine and assess the nodules cautiously in a brief span. Computer aided design frameworks help to check digital image and to mark possible disease. The general progression of proposed strategy is appeared in Fig.1. For the most part, CAD frameworks for lung nodules identification are made out of four fundamental stages: pre-processing, lung segmentation, feature extraction and classification.

Revised Manuscript Received on July 05, 2019.

Lim J Seelan, Research scholar, Department of EEE, Noorul Islam University, Thuckalay, Tamilnadu, India.

Dr. L. Padma Suresh, Principal, Baselios Mathews II College of Engineering, Sasthamkotta, Kerala-690521, India.

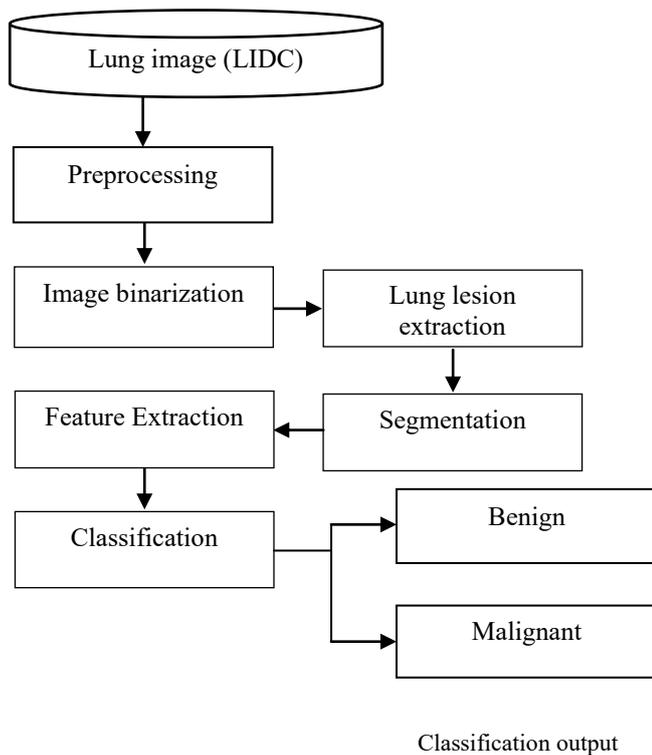


Figure 1: block diagram of proposed model

A. Image Preprocessing

In this medical image acquiring process Computed Tomography (CT) examine images are mostly used because it has low noise and improved transparency compared to another imaging technique images. The lung images are obtained from NIH/NCI Lung Image Database Consortium (LIDC) which is an on-line CT image dataset available in the “Cancer Imaging Archive”. Here the images are available in the DICOM format using image converters this images is convert into JPEG format. Normally, the size of the image is 512 x 512 and then it converted to 256 x 256 to obtain superior feature. In this proposed method there is 60 images is used for processing which consist of tumor and non-tumor images.

B. Image segmentation

Segmentation is basically utilized for segment of image. Image division is the strategy for isolating a picture into various parts. This is regularly used to identify object and edges of an image. It is a basic procedure for most image examination following strategies. There are different strategies are accessible for image segmentation. Here advanced chan-veese calculation is utilized for further processing. The optimized chan-veese algorithm is a combination of region based active contour and Hybrid BF-GWO algorithm.

C. Feature Extraction

The features of segmented images are dissected with the support of surface component of images. There are various numerical models accessible for ray level spatial distribution of images. Here Gray Level run length matrix features (GLRM) algorithm is helpful for concentrate the image features. Gray level run length matrix (GLRLM) is the

quantity of keeps running with pixels of gray levels i and run length j for a provided guidance.. The highlights properties are portrayed in table 1.

Table: 1. GLRLM Features

$SRLGE = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N \frac{P(i,j)}{i^2 j^2}$	$SRHGE = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N \frac{P(i,j) i^2}{j^2}$
$LRLGE = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N \frac{P(i,j) i^2 j^2}{1}$	$LRHGE = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N \frac{P(i,j) j^2}{i^2}$

D. Image classification

After the feature extraction model the result is given to the classification stage. These classification techniques are helpful for finding the stages and the performance of the classifier. In the proposed pipeline, we have used a Support vector machine (SVM), SVM naïve bayes and Artificial neural network (ANN) classifiers model to make automated decision about the nodule types (benign or malignant). This procedure is a powerful instrument for some classification execution and can easily make classifications. The result of feature extraction is given to every classifier and it gives accuracy, sensitivity, specificity and precision values. Support Vector Machine (SVM) is a supervised learning model. SVM approach is utilized to examine information and recognize a few examples, analysis and classification. It is broadly utilized in the bioinformatics because of its high accuracy. Naive Bayesian classifier is a basic characterization method, which estimates the class-conditional probability by assuming that the attributes are conditionally independent. For a wide range of applications Artificial Neural Networks (ANNs) have been developed. This model is mainly helpful for function approximation, feature extraction, optimization, and classification. The fed forward networks is one of the artificial networks for pattern recognition that are able to be trained for classifying net inputs according to the target classes.

III. EXPERIMENTAL RESULTS

In this model LIDC datasets are collected for processing. The dataset is accessible through the National Cancer Institute’s (NCI). The Cancer Image Archive (TCIA). This dataset includes 75 benign images and 75 malignant images these are in the DICOM format.



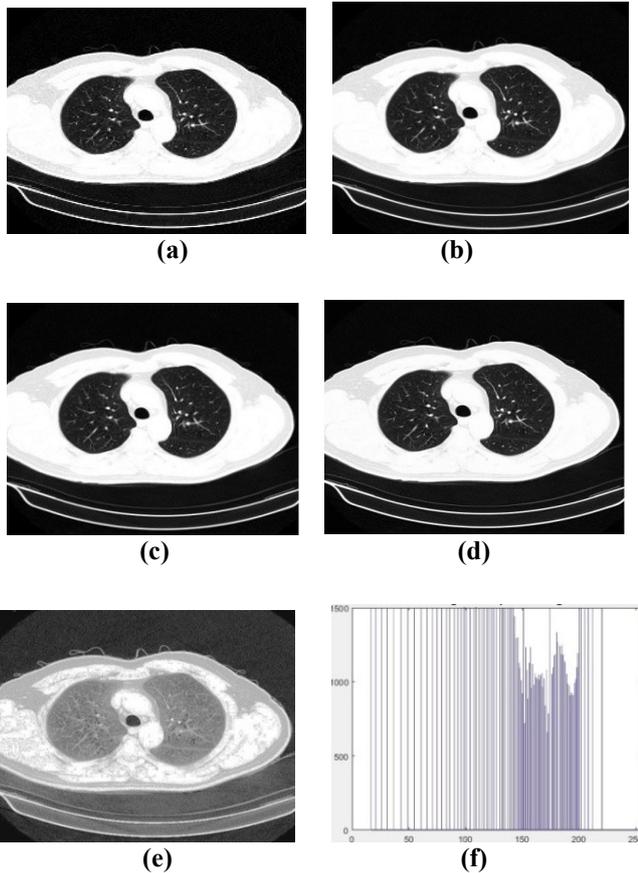


Figure 2: Figure: 2. (a) Input image (b) Gray scale image, Comparison of both (c) adaptive median filter image (d) Curvelet transformed image (e) Histogram equalized image, (f) Histogram equalized plot

Preprocessing methods are applied to the input image to eliminate the unwanted component and boost the feature of images. In this method initially the input image is altered into grayscale image. After alteration curvelet transform is used for de-noising. It effectively sharpens the edges and shapes. Then histogram equalization is performed for obtaining high contrast image. Figure 2 shows the preprocessed image, here 2(a) represents the input image of proposing method, 2(b) represents the grayscale conversion image. 2(c) & 2(d) shows the comparison of curvelet and adaptive filter image. In this comparison both techniques remove the noise, but curvelet transform effectively removes the noise because, if the edge in the images improves then image quality will be automatically improved. 2(e) & 2(f).shows the plots are generated for the resultant equalized images. Finally the performance measures such as PSNR, MSE, SNR, and SSIM are analyzed and the values are obtained. This is shown in table 2.

A fuzzy thresholding technique is helpful for CT image binarization. Here all the unrelated components are removed. Figure 3(a) and 3(b) shows the result of image binarization. Here the result of fuzzy output is compared with the Otsu thresholding technique. Compared with the Otsu thresholding technique the fuzzy thresholding techniques give better outputs. Next the pre-processing models the morphological operations and to identify the shapes and margins of the related components the canny edge detector is used. In morphological operation Dilation and erosion

operations are involved. Canny edge method uses two thresholds for detecting strong and weak boundaries, and finally includes only the weak edges in the output if they are linked to strong edges. The resultant output is displayed in 3(c) - 3(f).

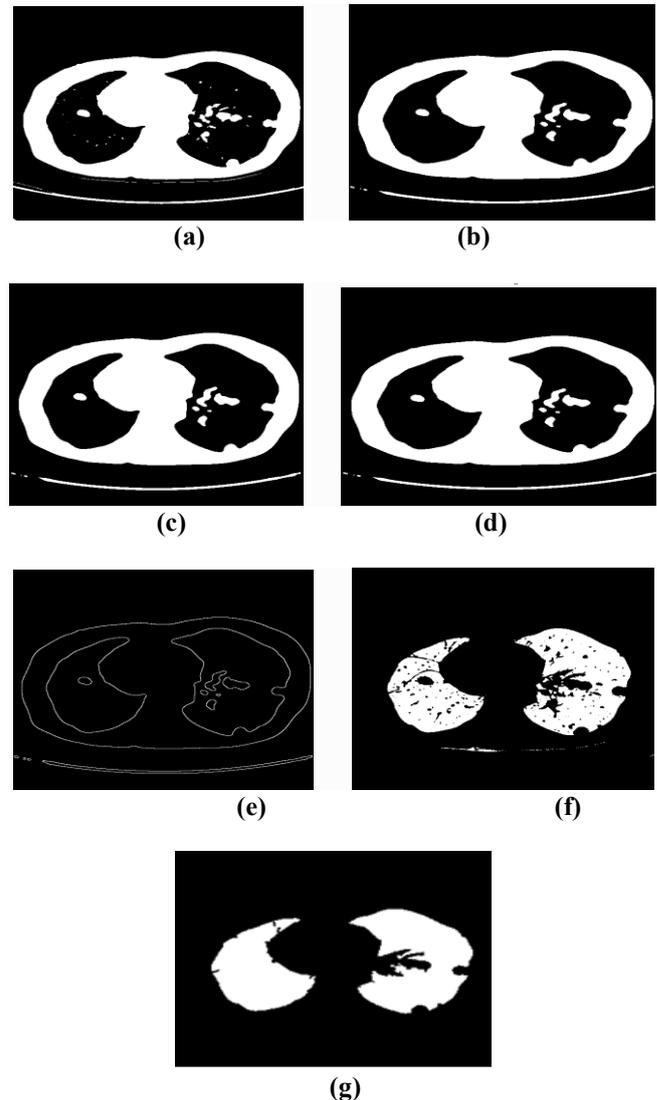


Figure: 3. comparison of image binarization result (a) fuzzy thresholding (b) Otsu thresholding (c) eroded image, (d) dilated image, (e) canny detector output, (f) border cleared image (g) adaptive concave hull output

A new adaptive concave hull algorithm is used to achieve exact original lung masks for a region based active contour model. This algorithm produces convex or non-convex polygons to characterize the lung area occupied by the given points. Output of concave hull algorithm is displayed in figure 3(g)

Table 2: Performance values of the database images

Metrics	MSE	PSNR	SNR	SSIM
Sample 1	59.1	35.66	30.41	0.9756
Sample 2	38.57	36.17	32.26	0.9735
Sample 3	69.76	35.45	29.69	0.9733
Sample 4	25.07	36.66	34.13	0.9766
Sample 5	43.17	36.04	31.77	0.9740
Sample 6	40.91	36.10	32.01	0.9742

After the preprocessing model the segmentation of tumor region from the enhanced image is performed. Here optimized chan-veese algorithms used for segmentation. This isolation will provide better results. The resultant segmented image is compared with the output of k-means clustering and region based active contour model. Both techniques provided same results and it missed some minute nodules, over segmentation problem occur in the k-means result. It is shown in figure 4(a) & 4(b). So to avoid this optimized technique is used.

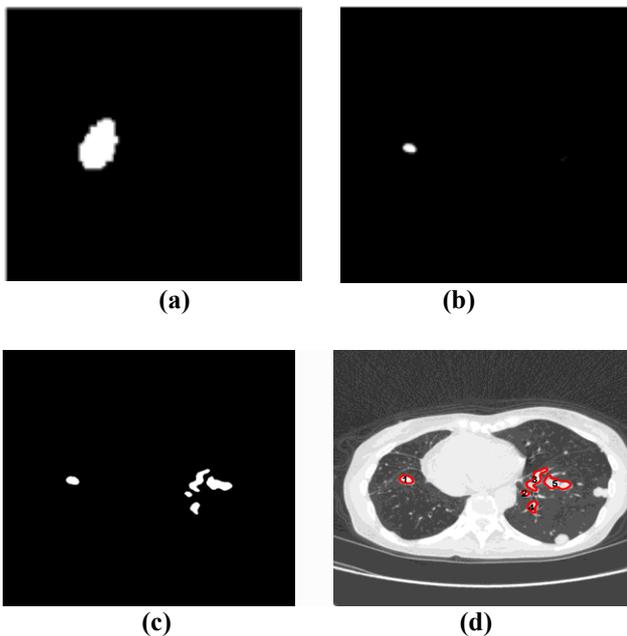


Figure 4: segmented images of
(a) k- means segmentation, (b) region based active contour model, (c) Optimized chan-veese, (d) labeled image of optimized chan-veese algorithm

It clearly separates the images with unclear boundaries. The major role of this work is to clearly marks the cancerous region. The result of the segmented image and the labeled image is shown in figure 4(c) & 4(d). This will mark the nodules efficiently. The area parameter centroid diameter of the nodule is calculated using this algorithm this is shown in table 3.

This strategy incorporates any sizes of the nodules and precisely smooth's the pulmonary vessels. For exact and powerful lung field segmentation with the nodule this proposed technique can be utilized. It doesn't require any client cooperation. It provides correct dominant points along convex and concave area on the lung boundary and reduces over segmentation and under segmentation.

Table 3: Parameter Measures

Region	Area (mm ²)	Perimeter (mm)	Centroid (mm)	Diameter (mm)
1	1680.0	148.7	125.5	46.2
2	60.0	25.7	206.6	8.7
3	56.0	24.4	320.6	8.4
4	42.0	22.1	243.8	7.7
5	20	12.1	199.3	5.9

The run length features values are given in table 4. Here the features like SRHGE, SRLGE, LRLGE, LRHGE are extracted.

Table 4: Run length Features

Features	GLRM			
	SRHGE	SRLGE	LRLGE	LRHGE
1	2170	321.0111	12.10332	2170
2	2474.04	851.4939	9.647325	2345
3	2474	561.4939	8.196796	2474
4	2474	851.4939	21.13615	2474.78
5	2344.09	851.4939	6.904325	2474
6	2574	7901.494	27.61273	2474.67
7	2564	971.4939	8.104962	2474.89
8	2474.56	851.4939	18.89785	2124
9	2234.75	451.4939	8.012987	2874
10	2474.09	231.4939	12.10332	2445

Based on the feature value at last SVM, SVM naïve bayes and ANN classifier classifies images and provides superior accuracy, sensitivity, specificity and precision values. The confusion plot of various classifiers is shown in figure 5. Here accuracy is calculated for every classifier which is given in table 5. From the figure, ANN classifier produces best results when compared to other classifier. Comparison shows that the classifiers used in the present work clearly produces better results than other classifiers. Figure 6 and 7, shows the comparison of proposed SVM, naïve bayes, and ANN.

Table 5: performance comparison

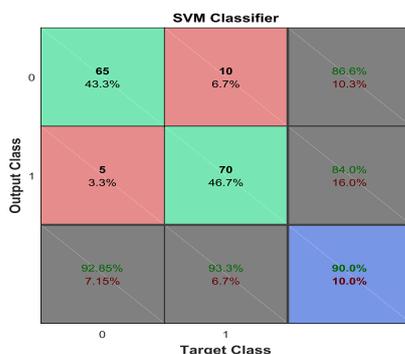
Classifier	SVM	Naive - Bayes	ANN
Accuracy	90%	92%	94.7%
Error	10%	6.7%	6.3%
Sensitivity	86.6%	93.3%	92%
Specificity	84%	84%	89.3%
Precision	93.3%	93.15%	92.4%

IV. CONCLUSION

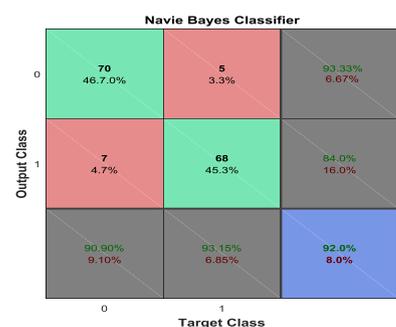
Lung disease is the most widely recognized reason for malignancy death on the world. premature identification of lung malignancy will significantly save the patient. The pre-processing for lung cancer is performed by curvelet transform and enhanced by histogram equalization method. Subsequently the lung extraction, lung border correction and lung segmentations are performed by various algorithms for finding the region of interest. After that, the feature values are calculated by GLRLM, for the particular ROI. Finally, using the classifications techniques the performances of the proposed models are calculated. The accomplished precision of proposed framework is 90%, 92% and 94.7% which is higher than other order strategies.

REFERENCES

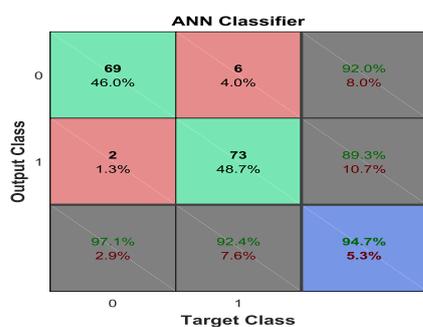
1. American Cancer Society, "What are the risk factors for non-small cell lungcancer? ", 2017, <http://www.cancer.org>
2. El-Baz, Ayman, Garth M. Beache, Georgy Gimel'farb, Kenji Suzuki, Kazunori Okada, Ahmed Elnakib, Ahmed Soliman, and Behnoush Abdollahi. "Computer-aided diagnosis systems for lung cancer: challenges and methodologies." *International journal of biomedical imaging 2013* .
3. T.Messay, R.C Hardie And S.K.Rogers "A New Computationally Efficient CAD System For Pulmonary Nodule Detection In CT Imagery," *Med.Image Anal.*, Vol.14,No3, Pp.390-406,2010
4. V.Kuppasamy, "Feature Extraction Based Lung Nodule Detection In CT Images," *IEE Trans. Biomed Eng.*, Vol. 56, No. 7, Pp.1810-1820, 2009
5. de Carvalho Filho, Antonio Oseas, Aristófanés Corrêa Silva, Anselmo Cardoso de Paiva, Rodolfo Acatauassú Nunes, and Marcelo Gattass. "Computer-aided diagnosis system for lung nodules based on computed tomography using shape analysis, a genetic algorithm, and SVM." *Medical & biological engineering & computing 55*, no. 8 (2017): 1129-1146..
6. Z.Abduh, M.A.Whed, And Y.M Kadah, "Robust Computer-Aided Detection Of Pulmonary Nodules From Chest Computed Tomography," *J. Med. Imaging Heal. Informatics*, Vol, 6,No. 3, Pp.693-699,2016
7. Chiou YS, Lure YF, Freedman MT, Fritz S. Application of neural network based hybrid system for lung nodule detection. In: *Computer-based medical systems, proceedings of sixth annual IEEE symposium*, IEEE; 1993. p. 211–216.
8. Bhuvaneswari, C., P. Aruna, and D. Loganathan. "A new fusion model for classification of the lung diseases using genetic algorithm." *Egyptian Informatics Journal 15*, no. 2 (2014): 69-77.
9. Lim J Seelan, Padma Suresh 'Lung Lesion Enhancement Using Adaptive Median Filter With Curvelet Transform' *Inernational Journal Of Engineering & Technology*.7 (2.33) (2018) 1323-1328.
10. Lim J Seelan, Padma Suresh 'An Effective Automatic Detection Of Lung Tumor Based On Novel Optimized Chan-Vese Algorithm' *International Journal of Recent Technology and Engineering* (2019).
11. Anifah, Lilik, Rina Harimurti, Zaimah Permatasari, Puput Wanarti Rusimanto, and Adam Ridianto Muhamad. "Cancer lungs detection on CT scan image using artificial neural network backpropagation based gray level cocurrence matrices feature." In *2017 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, pp. 327-332. IEEE, 2017..
12. Li, Xuanping, Xue Wang, and Yixiang Dai. "Adaptive energy weight based active contour model for robust medical image segmentation." *Journal of Signal Processing Systems 90*, no. 3 (2018): 449-465.



(a)



(b)



(c)

Figure 5: Confusion matrices of proposed models

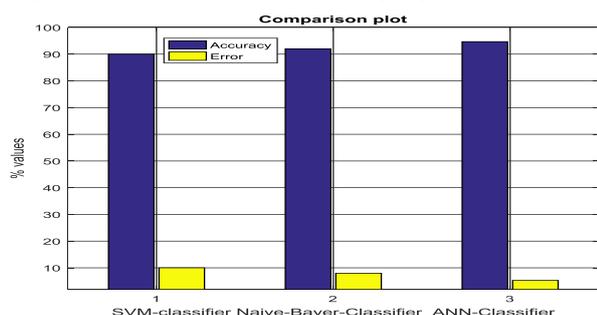


Figure 6: Accuracy and error plot of proposed models

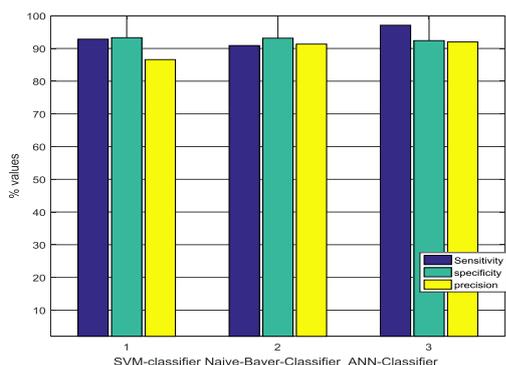


Figure 7: Sensitivity, specificity and precision plot of proposed models