Lung Nodule Detection and Classification based on Feature Merging and Genetic Algorithm

Shalini V. Wankhade
Phd Research Scholar, Sathyabama Institute of Science And technology,
Chennai, India
svwankhade.sae@sinhgad.edu

Dr. Vigneshwari S.
Associate Professor,
Sathybhama Institute of Science and Technology, Chennai
vigneshwari.cse@sathybama.ac.in

Abstract:
The purpose of this study is to increase the accuracy of the early detection of lung nodes through the design of Computer-Aided Detection (CAD). Comparative studies of the effectiveness of the most commonly used methods for extracting and classifying features were conducted, and methods that give the highest accuracy and the least false-positive results were identified. To develop an accurate automatic detection system for early detection of lung nodules, we investigated directly candidates for lung nodules without removing blood vessels. The feature is extracted by using feature extraction methods. To utilize the extracted features, the extracted features were connected using the feature merge technique, and the features were selected by the new hybrid feature vector. The genetic algorithm (GA) search based on the classification accuracy rate of the used classifier was also applied to the hybrid feature vector. To archive, high classification accuracy was selected three classifiers and performed a comparison of their performance.

Keywords: *Computer-Aided Detection, Genetic algorithm, Morphological technique, Computed tomography, Segmentation, Classification.*

1. INTRODUCTION

As a result, high air pollution in recent years and the spread of smoking, lung cancer poses a great threat to humanity due to high rates of infection and difficulty in treatment and rapid spread [1]. Because early detection increases the chances of survival of the patient during 5 years to 70%, and also increases the chances of successful treatment if diagnosed in the early stages, this has led to the increasing significance of the development of early detection methods [2], [3]. One of the most important methods used in the diagnosis of lung cancer is computed tomography (CT) of the chest of the patient. This is not surprising, because in this study it is in many sections that radiologists and doctors bring the number of large images that allow examining all parts of the lungs [4].

However, due to the increase in the number of images obtained as a result of this investigation, in addition to using a low dose of radiation during the study to protect patients at high risk of radiation exposure, all studies of these images by radiologists are a complex and acute task [5]. These systems are known as Computer-Aided Detection (CAD) systems [6], [7], which allow researchers to process and analyze these images and design a computer-based system that allows them to automatically determine the presence of pulmonary nodes. CAD consists of four main stages for automatic detection of pulmonary nodes: the pre-treatment stage for increasing image contrast and reducing noise. This is followed by an automatic segmentation stage that aims to separate the area of the human lungs, followed by a procedure for extracting lung nodules using a digital CT image and a final phase is a classification [8] shows in Fig. 1.

Feature extraction and classification for the detection of nodules are described in this paper. This paper aims at the development of a fully automatic CAD system for the early detection of lung cancer. First and second stages of the system, which is an image preprocessing step and an automatic segmentation system, [9] have been previously completed and published. The performance of the complete system is calculated by a classification accuracy and using Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) dataset. In general, the system

ISSN: 2233-7857 IJFGCN

CAD process of extraction of candidate pulmonary nodes is to remove blood vessels followed by the removal of the main traits of sought candidates. But the blood vessels can cause small lung nodules, attached to the blood vessels [10] [11]. To overcome the problem of possible loss of small lung nodules, the CAD system proposed in this study detects pulmonary nodules without removing blood vessels, and the main focus of extracted lung nodules is to extract the major key feature of lung candidates to increase the detection accuracy.

Recently, many studies are carried out to improve the accuracy of the classification system. The most important of these methods is the merge technology, which can be divided into three levels: data merge, feature-level merge, and solution-level merge. A data merging that combines data from different data sources to produce new data sets with higher information content [12]. Merging solutions where a classification decision of a group of classifiers is made, classifiers can be of the same type or of different types using the same set of features or different types of them to obtain accurate and objective classification results [13].

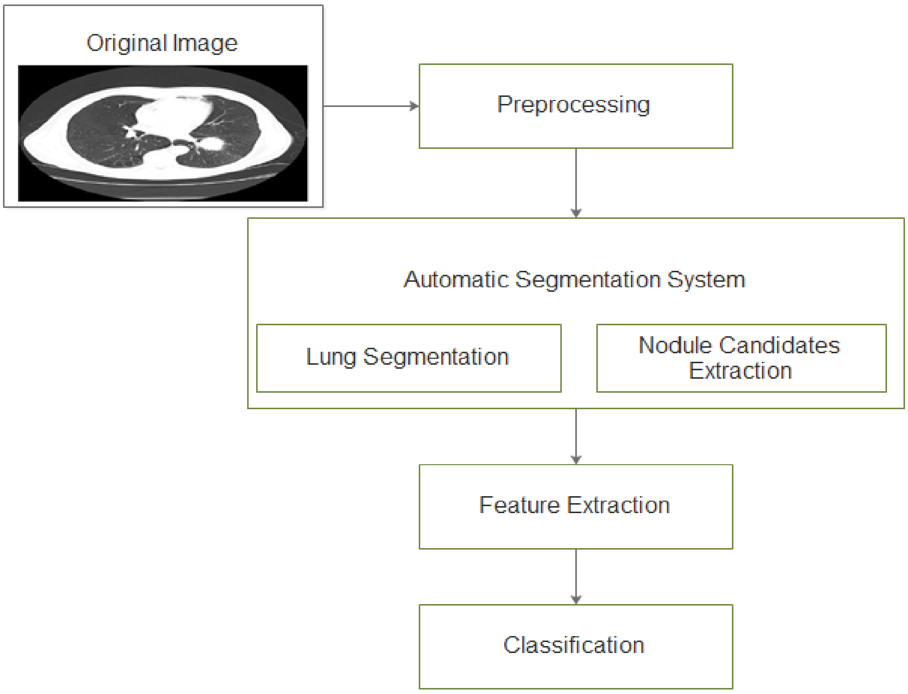

Figure 1: The stages of CAD system

Feature merging, in which different feature types are selected and combined to generate new features set with the removal of correlated and repeated information for improving accuracy. In this work, the features of merging will be applied [14]. In the classification stage, classifiers detect the nodules using extracted hybrid feature vector of lung node candidates as input to the classifier, and its output is to the input pattern is analyzed as nodules or non-nodules. Several classifiers are used such as CNN, NB, and SVM. The effectiveness of the proposed system is compared with existing classifiers.

2. METHODOLOGY

Mentioned in earlier, any CAD has four main stages: a preprocessing stage followed by an automatic segmentation procedure for lung extract and candidates for lung nodules, then a third stage-extraction of the main features of candidates for lung nodules, and finally a classification stage to identify lung nodes. Segmented candidates of pulmonary nodes are used as input to the third stage of CAD, which forms the stage of extraction and selection of signs. It is aimed at using advanced methods; mining function to load basic features segmented into the main candidates and selects the most important features. The resulting selected feature is a classification step that aims to detect lung nodules by classifying the extracted lung nodule candidates into nodules or non- nodules.

ISSN: 2233-7857 IJFGCN

Feature Extraction

An important step in the classification process is to extract features [15]. The purpose of the extraction of the function is to achieve a significant reduction in data and the determination of useful measures. This entails a reduction in dimensions, where the data of higher dimensional space is converted into a space of smaller dimensions. The resulting data (features) are designed to be informative, not superfluous, and facilitate the subsequent stages of learning and generalization.

We use four different techniques for extracting such as Gray-level co-occurrence matrix (GLCM), Value histogram (VH), Histogram of Oriented Gradients (HOG), and statistical features based on wavelet coefficients. We'll talk about these in more detail:

1) Histogram of Oriented Gradients (HOG): Texture of the image described by using a histogram of the features of the image. This type of feature was first adopted by Dalal and Triggs [16], where input images are subdivided into smaller areas called cells. For each cell, the Pixel gradient is calculated and the histogram of the gradient is constructed. These gradients are blocks in the classification of each neighboring cell pic, describing the texture of the image object, describing the outline of the object, and giving the dark shape and outline of the image object. For greater accuracy, cell histograms are normalized based on histograms of cells belonging to the same block. This normalization depends on the constancy of shading and lighting changes. HOG was calculated for each node candidate image. The resulting feature vector has large dimensions, reaching a maximum of 1806336 for each image. Feature vectors dimensions reduction and eliminate redundancy of information, it employs Principal Component Analysis (PCA), which makes it more efficient to manipulate and store data [17]. PCA is proposed by Hotelling and it is a mathematical method [18]. PCA is a conversion algorithm. The input signal for the PCA algorithm is correlation variables, and the output signal is linear correlation variables by applying an orthogonal transformation of these uncorrelated variables, called principal components. By applying PCA, there is an extra benefit next to riding uncorrelated parameters. The dimension of the output data from the PCA is smaller than the input data, after applying the main PCA obtained by organizing the eigenvalues calculated from the covariance matrix, the feature vector size of each ROIs reached 150 dimensions.
2) Gray Level CO-Occurrence Matrix (GLCM): Texture of GLCM based on wavelet transform coefficients features: Wavelet Transform is a spatial transformation, in addition to feature scales to obtain multi-resolution characteristics [19]. In this work, wavelet properties are calculated as follows; node candidate patterns decompose on two levels by applying the wavelet conversion 2-D Daubechies (DB2). Then, each horizontal, diagonal, vertical detail was obtained by wavelet decay structure decay levels one and two in addition to approximate coefficients second level. After receiving wavelength coefficients, the texture properties of GLCM were calculated from them. GLCM is a statistical dependency matrix based on relationships of adjacent or adjacent pixels to shows image texture [20]. First, the GLCM is assigned for each of the first and two-level sub bars to characterize node candidates. GLCM was used in many previous works to extract the texture properties of pulmonary node candidates due to their effectiveness [21], [22]. If the multiresolution analysis of the wavelet transform allows obtaining a characterization of the features around the lung nodule candidates at different levels and scales, in addition to the application in GLCM more each wavelet sup-bands is then calculated in the second-order statistical texture features of each GLCM more to get amore characterized and un- correlated texture features based on wavelet characteristics and pulmonary nodule candidates, increasing the classification accuracy [23]. The statistical features of the second order are entropy, contrast, energy, correlation, and homogeneity. There is a set of 5 features calculated for each wavelet sub-band, getting a total of 35 features.
3) Value Histogram: Value histogram (VH) function, which means the histogram distribution of CT image intensity values [24]. VH properties are extracted by constructing histogram distribution of the pixel intensity values of the CT image. The function vector size shows the

ISSN: 2233-7857 IJFGCN

number of extracted histogram bins, determined according to the number of bins, which gives the best classification accuracy rate. In this way, different VHs were built with a different number of bins [24].

4) Statistical Feature: In statistical calculations, the texture of the object in the image is described in an indirect way to non-deterministic properties that manage and describe the relationships between the values of Gray level intensity of the image. Two categories of statistical characteristics are uses: the first order and the second order. To determine the statistical characteristics of the first order, a histogram calculation of the gray level intensity image values are used. Statistical characteristics of the first order determined by this histogram include average Gray level, obliquity, variance, kurtosis, and entropy and foreground area pixels [25]. The statistical properties of the gray level Co-occurrence Matrix (GLCM), calculated for the gray level intensity image, are correlation, homogeneity, energy, contrast, and entropy. They are considered to be another statistical feature and was calculated using the Haralick transforming [25]. The resulting vector of statistical characteristics consists of 11 properties.

Feature Merging

Features are the key to the success of the classification process, and therefore the selection of the optimal feature set leads to the success of the classification process. In the process of merging features, after removing non- important and noisy features, a new feature set is created from a different set of features in different domains [14]. The merge process is a type of feature selection, and the feature selection process consists of excluding correlated, unimportant, or distorted features that are not included in the calculation. To improve the classification efficiency of the system, four different vectors were used separately, and simple concatenation was applied to combine them into a new hybrid feature vector. Next, the vector performance of each function set and the new hybrid function are compared.

After forming a hybrid function vector, the next step is to remove excess and correlated information, called feature selection. Removing redundant information improves classification accuracy, reduces the size of processing data, reduces the computational burden, and increases the speed of classification algorithms [26], [27]. The genetic algorithm (GA) was utilized as a selection step. The GA was originally developed by Holland [28]. It is built on three basic processes: the selection process based on the fitness function, it mutates, and each of the crossover. According to the terminology of the GA, the feature subsets are called individuals or chromosomes consisting of genes. Also, the individual qualities are measured by the fitness function. The fitness function selected as the classification accuracy rate. The individual size of the GA population is equal to the size of the hybrid feature vector. An individual in a population of GA occurs in the form of a bit string (these and Zeros), where 1 means that the corresponding function of the hybrid function vector is selected, and 0 means that the corresponding function is excluded. After the generation of the initial population, GA functions in a frequent manner to bring the population to an optimal point according to the fitness function. For each generation of GA, the fittest individuals are selected using a selection operation to act as parents to generate a new generation of GA through basic GA operations: reconnecting the chromosomal genes of each two selected parents (crossover), inverting the genes within an individual randomly from 1 to 0 or Vice versa (mutated), and individuals that are accepted as is without changes to the next generation of GA are determined by the selection operation [29].

In this paper, the GA algorithm was applied to the hybrid feature vector as a feature selection method using three classifiers: the Naïve Bayes (NB), the Convolutional Neural Network (CNN), and the Support vector machine (SVM).

Lung Nodule Detection and Classification

The final stage is the classification of the candidate into nodule and non-nodule. Three classifications were selected and carried out a comparison of their effectiveness. The NB, CNN, and SVM classifiers

ISSN: 2233-7857 IJFGCN

are used. The inputs to each classifier are feature vectors from feature extraction and merge stages. The classifier was trained and its performance was evaluated. For the training phase, each classifier uses 25 % data size and 75 % is used for the testing phase.

1) The CNN Classifier

CNN is a deep learning algorithm, which takes the input image, the highlight image, the different aspects/subjects, and can be distinguished from each other. CNN used to classify and recognize its high-precision images.

2) The NB Classifier

In statistics, the NB classifier is one of the simplest Bayesian network models of simple probabilistic classifiers" based on applying Bayes theorem with strong (naive) independence assumptions between features, even if the result of the kernel density estimates with high accuracy. NB assigns a class label to a problem instance represented as a vector of feature values drawn from a finite set of class labels.

3) The SVM Classifier

SVM is a supervised learning method, used the kernel function to process the input data. The result of the classification obtained through the final sum by applying the activation function. The hyperplane is used to classify input data into two classes (binomial classification). A vector that is very close to the boundary is called the reference vector, and the distance between the hyperplane and the reference vector is called the boundary. SVM uses several types of cores, such as linear, radial base functions (FBF), polynomials [29], [30].

By applying the GA technique is a function of the selection tool resulted in a decrease in the size of the feature vector as follows: in the case of applying CNN classification, the number of features was reduced from 180 to 168 functions, NB classification from 182 to 117 and in the case of SVM classification from 182 to 150 features.

3. EXPERIMENTAL RESULTS AND DISCUSSION

To calculate the performance of the system various classifiers classification accuracy, sensitivity, and specificity calculated are followed [29]:

$$Accuracy = [(TP + TN)/(TP + TN + FP + FN)] \times 100 \quad (1)$$

$$Sensitivity = [TP/(TP + FN)] \times 100 \quad (2)$$

$$Specificity = [TN/(TN + FP)] \times 100 \quad (3)$$

Where, FP, FN, TP, and TN represent false positives, false negatives, true positives, and true negatives, respectively [29]. The main goal of this study is to obtain the highest accuracy, sensitivity, and specificity of the CAD system designed to detect lung nodules in the early stage. To achieve this objective, feature fusion and feature selection techniques were applied to the extracted features. And, the classification accuracy rate, sensitivity (s) and specificity (SP) were calculated for each classifier using four types of extracted properties. Table 1, 2, 3 shows the accuracy, S, and SP obtained from 3 different classifiers. Fig. 2 shows graphical representation of classification accuracy using Future extraction techniques and proposed hybrid feature, it shows the use of feature fusion technology obtaining the highest classification results of the three classifiers.

Classifiers	**Features**				
	HOG	VH	Wavelet Features	Statistical Features	Hybrid Features
CNN	65 %	89 %	90%	88%	97%

ISSN: 2233-7857 IJFGCN

NB	67%	80 %	87%	75%	95%
SVM	70 %	87%	85%	80%	94%

TABLE 1: Classification accuracy using four Future extraction techniques and proposed hybrid feature

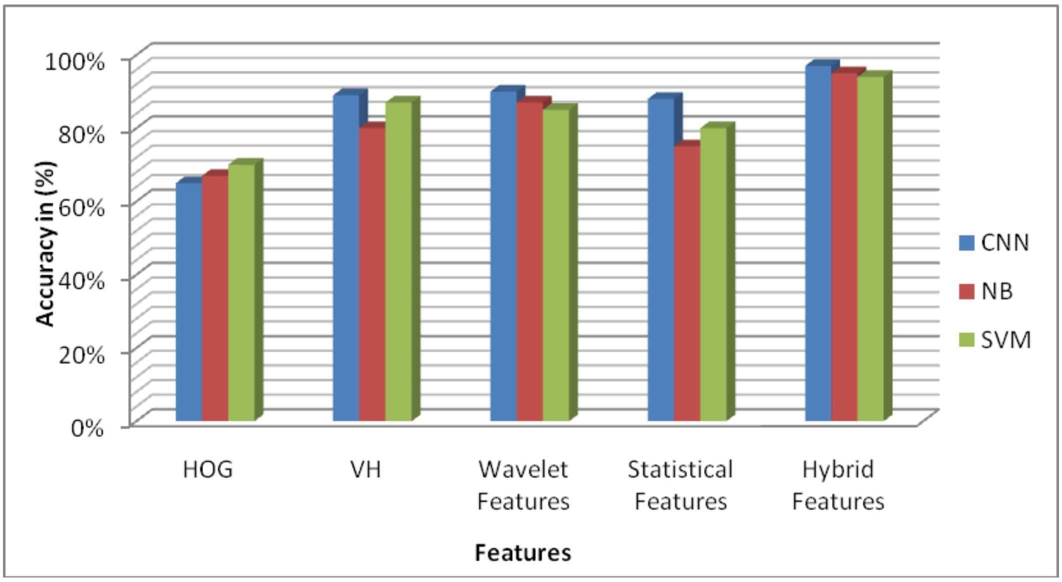

Figure 2: Classification accuracy graph

Fig. 3 shows graphical representation of sensitivity of classifier using Future extraction techniques and proposed hybrid feature, it shows the use of feature fusion technology obtaining the highest 100 % of sensitivity results of the two CNN and SVM classifiers.

Classifiers	**Features**				
	HOG	VH	Wavelet Features	Statistical Features	Hybrid Features
CNN	70 %	86 %	95 %	88 %	100 %
NB	67 %	85 %	86 %	75 %	99.1 %
SVM	69 %	80 %	88 %	80 %	100 %

TABLE 2: Sensitivity of classifiers using four Future extraction techniques and proposed hybrid feature

ISSN: 2233-7857 IJFGCN

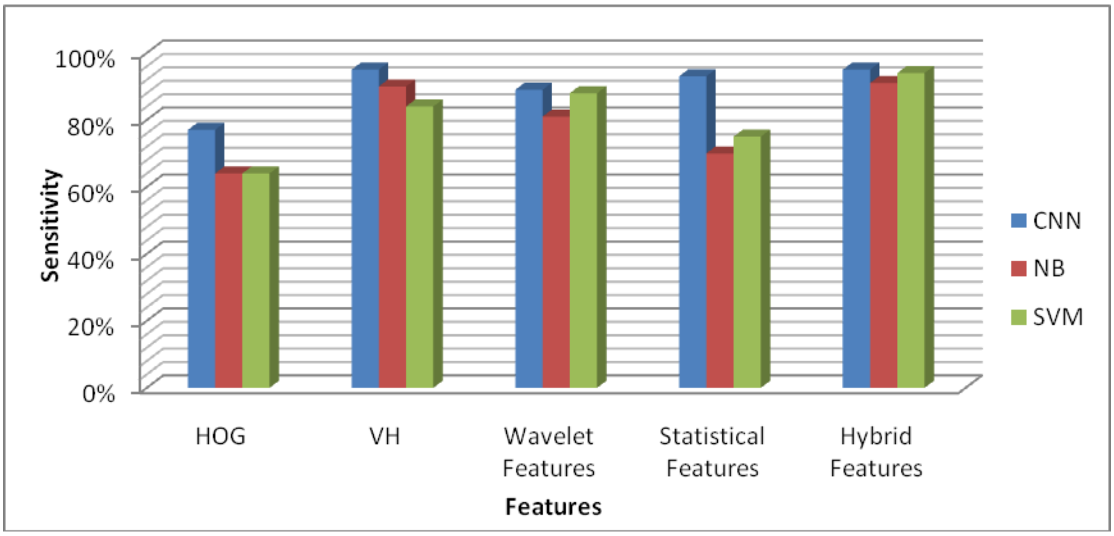

Figure 3: Sensitivity of classifiers graph

Fig. 4 shows specificity of classifier graph, it shows the use of feature fusion technology obtaining the highest 95 % of specificity results of the CNN classifiers.

Classifiers	**Features**				
	HOG	VH	Wavelet Features	Statistical Features	Hybrid Features
CNN	77 %	95 %	89 %	93 %	95%
NB	64 %	90 %	81%	70 %	91 %
SVM	64 %	84 %	88 %	75 %	94 %

TABLE 3: Specificity of classifiers using four Future extraction techniques and proposed hybrid feature

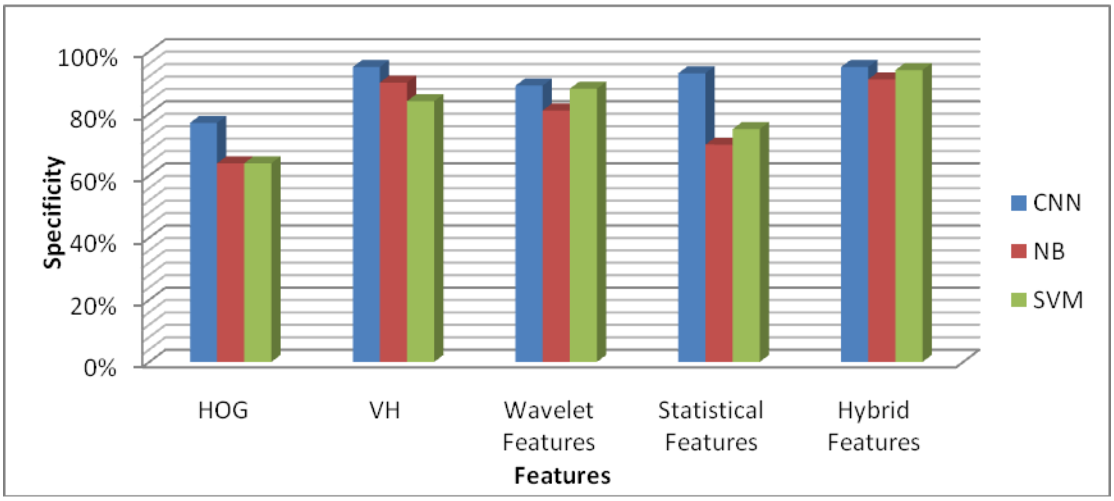

Figure 4: Specificity of classifiers graph

ISSN: 2233-7857 IJFGCN

International Journal of Future Generation Communication and Networking
Vol. 13, No. 4, (2020), pp. 4380–4389

Comparing the classification results, it is clear that the use of feature fusion technology obtaining the highest classification results of the three classifiers. Classification accuracy is reached to 97 %, 95 % and 94 %, Sensitivity reached to 100 %, 99.1 % and 100 %, and Specificity reached to 95 %, 91 % and 94 % with FP of
.06 and .1 and .057 and the FN is .008, 0 and 0 in the case of CNN, NB, and SVM. Therefore, the adoption of feature fusion gives a large effect on the improvement of the classification performance.

When the GA method is applied as a feature selection technique to a hybrid feature vector using three classifiers, feature vector size decreases. Table 4 shows the number of features before and after the decrease and the corresponding accuracy, S, and SP for each classifier.

	Classifiers		
	CNN	NB	SVM
Hybrid feature size	182	182	182
Selected feature size	168	117	150
Accuracy in (%)	99.6 %	99.2 %	99.6 %
Sensitivity in (%)	100 %	100 %	100 %
Specificity in (%)	99.2 %	94 %	99.2 %

TABLE 4: The accuracy, Specificity, and Sensitivity of classifiers and the feature vector size after using the GA technique

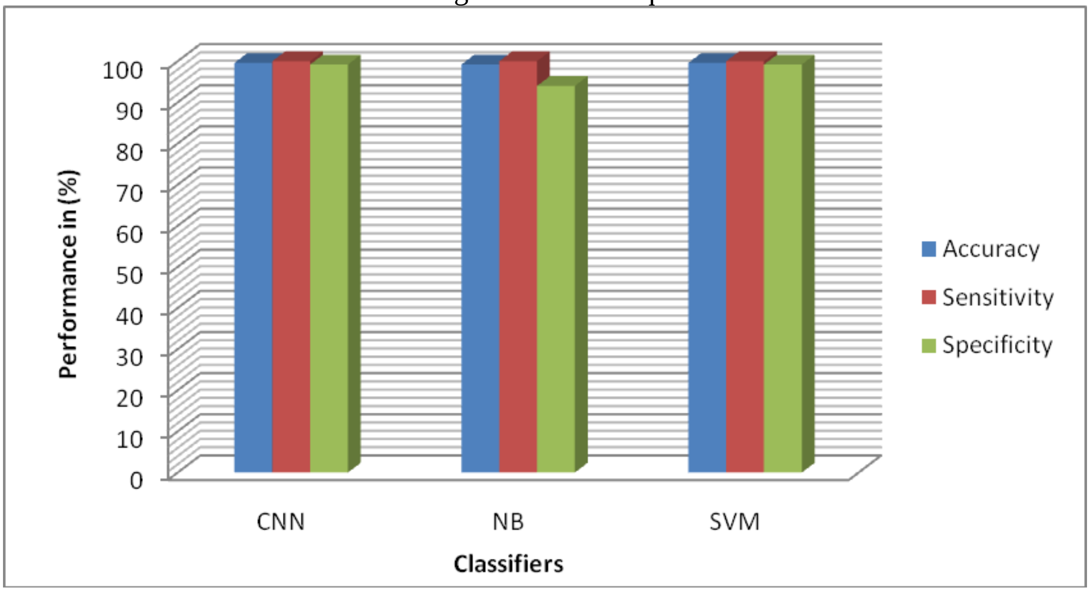

Figure 5: System performance using GA

As is clear from Table 4, using the GA feature selection, the technology has grown with the addition of classification accuracy, S and SP, and reduced feature vector size. Due to this improving the system performance and reduction of calculation time. The system performance is increased by using GA shows in Fig
5. Although the features are produced by the NB classifier are reduced, accuracy and specificity are lower than the other two classifiers, CNN and SVM. The classification result of CNN and SVM are

ISSN: 2233-7857 IJFGCN

the same but the feature of SVM is less than CNN.

4. CONCLUSION

The main objective of this study is to develop an automatic detection system (CAD) for the accurate detection of lung nodules on CT images. The main goal was to get the maximum accuracy of detecting lung nodules. The system contains four main stages. The processing to improve the quality of CT images, the automatic segmentation stage to extract candidates, the feature extraction and selection stage to obtain the most important features, and the classification stage for the final detection of lung nodules. In the proposed system, four feature extraction techniques are used. Feature merging steps used for four different sets of extracted features to generate a hybrid feature vector. These feature vectors are used to evaluate their performance. The classifiers are CNN, NB, and SVM. Using the feature vector the accuracy, sensitivity, and specificity of classifiers are calculated. Comparison shows, the hybrid feature gives the highest accuracy, sensitivity, and specificity. As a result, it was concluded that the feature merging method improved the detection accuracy of the lung nodule and improved system performance. A vector hybrid characterized as a genetic algorithm (GA) selective algorithm, this paper tries improvement of classification accuracy, improvement of system performance, shortening of calculation time.

REFERENCES

1. A. Jemal, T. Murray, E. Ward, A. Samuels, R. C. Tiwari, A. Ghafoor, E. J. Feuer, M.J. Thun, Cancer statistics, CA Cancer J. Clin. Vol. 55, pp. 10–30, 2005
2. A. Micheli, P. Baili, M. Quinn, E. Mugno, R. Capocaccia, P. Grosclaude, Life expectancy and Cancer Survival, The EUROCARE Working Group, in the EUROCARE-3 cancer registry areas, Ann. Oncol. V28– V40, 2003.
3. G.K. Singh, B.A. Miller, B.F. Hankey, Changing area socioeconomic patterns in U.S. cancer mortality, 1950–1988. Part II: lung and colorectal cancers, J. Natl. Cancer Inst. 94 916–925, 2002.
4. X. Ye, X. Lin, G. Beddoe, J. Dehmeshki, "Efficient computer-aided detection of ground-glass opacity nodules in thoracic CT images", in Proc.29th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., pp. 4449–4452, 2007.
5. T.W. Way, B. Sahiner, H.P. Chan, L. Hadjiisk, P.N .Cascade, A. Chughtai, N. Bogot, E. Kazerooni, "Computer-aided diagnosis of pulmonary nodules on CT scans: Improvement of classification performance with nodule surface features", Med. Phys., vol. 36, no. 7, pp. 3086–3098, 2009
6. S.C. Park, J. Tan, X. Wang, V. Lederman, J.K. Leader, S.H. Kim, B. Zheng, "Computer-aided detection of early interstitial lung diseases using low-dose CT images", Phys. Med. Biol., vol. 56, no. 4, pp. 1139–1153, 2011.
7. L. Song, X. Liu, A. Yang, K. Pang, C. Zhou, X. Zhao, Y. Zhao, "A novel approach of computer-aided detection of focal ground-glass opacity in 2D lung CT images", Proc. SPIE, vol. 8670, pp. 86702W, 1–6, 2013.
8. T. Manikandan, "Challenges in lung cancer detection using computer-aided diagnosis (CAD) systems – a key for survival of patients", Arch Gen Intern Med Volume 1 Issue 2, 2017
9. H.A. Elmikati, F.E. Abou-Chadi, Sh.S. Kishk, M.I. Obayya, H.M. Amer, "Automatic Detection of Human's Lung Nodules Using an Optimal Gray Level Threshold", International Journal of Scientific & Engineering Research Volume 8, Issue 5, ISSN 2229-5518, May-2017
10. Early Lung Cancer Action Program(ELCAP), available from: http://www.via.cornell.edu/lungdb.html.[Last cited on 2011 Dec 05].
11. W.J. Kostis, A.P. Reeves, D.F. Yankelevitz, and C.I. Henschke, “Threedimensional segmentation and growth-rate estimation of small pulmonary nodules in helical CT images,” IEEE Transactions on Medical Imaging, vol. 22, no. 10, pp. 1259–1274, 2003.
12. B. V. Dasarathy, "Decision Fusion", Computer Society Press, 1994.
13. L.I. Kuncheva, J.C. Bezdek, R.P.W. Duin, “Decision templates for multiple classifier fusion: an experimental comparison,” Pattern Recognition, vol. 34, pp. 299- 314, February 2001.

ISSN: 2233-7857 IJFGCN

14. U. G. Mangai, S. Samanta, S. Das, P.R. Chowdhury, " A Survey of Decision Fusion and Feature Fusion Strategies for Pattern Classification ", IETE technical review, vol 27, ISSUE 4,JUL-AUG, 2010.
15. V. Kuppusamy, R.C. Gopalakrishnan, "Feature Extraction Based Lung Nodule Detection in CT Images", International Journal of Applied Engineering Research, ISSN 0973-4562, vol. 11, no. 4, pp 2697-2700, 2016.
16. N. Dalal, B. Triggs, "Histograms of oriented gradients for human detection", in Proc. Comput. Vis. Pattern Recognit., pp. 886–893, 2005.
17. S.C. Park, B.E. Chapman, B. Zheng, "A multistage approach to improve performance of computer-aided detection of pulmonary embolism depicted on CT images: preliminary investigation", IEEE Trans. Biomed. Eng., vol. 58, no. 6, pp. 1519–1527, Jun. 2011
18. I.J.S. Jeya, S.N. Deepa, "Lung Cancer Classification Employing Proposed Real Coded Genetic Algorithm Based Radial Basis Function Neural Network Classifier", Hindawi Publishing Corporation Computational and Mathematical Methods in Medicine Volume 2016, Article ID 7493535, 2016.
19. Y. Tao, L. Lu, M. Dewan, A.Y. Chen., J. Corso, J. Xuan, M. Salganicoff, A. Krishnan, "Multi-level ground glass nodule detection and segmentation in CT lung images", in Proc. Med. Image Computer-Assisted Surg., Part II, LNCS 5762, pp. 715–723, 2009.
20. V.S. Thakare, N.N. Patil, "Classification of Texture Using Gray Level CoOccurrence Matrix and Self- Organizing Map", International Conference on Electronic Systems, Signal Processing and Computing Technologies, IEEE, 978-1-4799-2102-7/14, 2014
21. M. Firmino, G. Angelo, H. Morais, M.R. Dantas, R. Valentim, "Computer aided detection (CADe) and diagnosis (CADx) system for lung cancer with likelihood of malignancy", Firmino et al. Bio. Med. Eng. OnLine vol. 15, Issue 2, 2016.
22. H. Hotelling, "Analysis of a complex of statistical variables into principal components", Journal of Educational Psychology, vol. 24, Issue 6, pp. 417-441, 2016.
23. I. Buciu, A. Gacsadi, "Directional features for automatic tumor classification of mammogram images", Biomed Signal Process Control. Volume 6, Issue 4, pp. 370-378, October 2011.
24. X. Liu, L. Ma, L. Song, Y. Zhao, X. Zhao, C. Zhou, "Recognizing Common CT Imaging Signs of Lung Diseases Through a New Feature Selection Method Based on Fisher Criterion and Genetic Optimization", IEEE Trans. Biomedical and Health Informatics, vol. 19, no. 2, March 2015.
25. P.V. Ramaraju, S. Praveen," Classification of lung tumor Using Geometrical and Texture Features of Chest X-ray Images", International Journal for Research in Applied Science & Engineering Technology (IJRASET), Special Issue-3,ISSN: 2321-9653, November 2014
26. Y. Saeys, I. Inza, P. Larra˜naga, "A review of feature selection techniques in bioinformatics", Bioinformatics, vol. 23, no. 19, pp. 2507–2517, 2007
27. P. Arora, M. Singh," New Intensity Based Features for Classification of Mammograms", Conference Paper in Canadian Conference on Electrical and Computer Engineering, 978-1-4799-3010-9, 2014.
28. J.H. Holland, "Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence", U Michigan Press; 1975.
29. Sh. Aalaei, H. Shahraki, A. Rowhanimanesh, S. Eslami, "Feature selection using genetic algorithm for breast cancer diagnosis: experiment on three different datasets", Iranian Journal of Basic Medical Sciences, vol. 19, no. 5, pp. 476–482, MAR 2016
30. S. Sivakumar, C. Chandrasekar, "Lung Nodule Detection Using Fuzzy Clustering and Support Vector Machines", International Journal of Engineering and Technology (IJET), ISSN : 0975-4024, vol.5, no.1, Feb-Mar 2013.

ISSN: 2233-7857 IJFGCN