ORIGINAL ARTICLE

A new fusion model for classification of the lung diseases using genetic algorithm

C. Bhuvaneswari [a,*], P. Aruna [a], D. Loganathan [b]

[a] Department of Computer Science and Engineering, Annamalai University, Chidambaram, Tamil Nadu, India
[b] Department of Computer Science and Engineering, Pondicherry Engineering College, Puducherry, India

Abstract Automatic classification of lung diseases in computed tomography (CT) images is an important diagnostic tool for computer-aided diagnosis system. In this study, we propose a new image based feature extraction technique for classification of lung CT images. A novel fusion based method was developed by combining the Gabor filter and Walsh Hadamard transform features using median absolute deviation (MAD) technique and hence, it possesses the advantages of both models. The proposed system comprises of three stages. In the first stage, the images are preprocessed and features are extracted by novel fusion based feature extraction technique, followed by second stage, in which extracted features are selected by applying genetic algorithm which selects the top ranked features. In the final stage, classifiers namely decision tree, K nearest neighbor (KNN), Multi layer perceptron Neural Networks (MLP-NN) are employed to perform classification of the lung diseases. A total of 400 datasets for the diseases bronchitis, emphysema, pleural effusion and normal lung were used for training and testing. The classification accuracy of above 90% is accomplished by multilayer perceptron neural network classifier. The system has been tested with a number of real Computed Tomography lung images and has achieved satisfactory results in classifying the lung diseases.

© 2014 Production and hosting by Elsevier B.V. on behalf of Faculty of Computers and Information, Cairo University.

1. Introduction

In the recent years, medical CT Images have been applied in clinical diagnosis widely. It assists physicians to detect and locate pathological changes with more accuracy. Computed tomography images can be distinguished for different tissues according to their different gray levels. Lung diseases can be caused by infection, an exposure at the workplace, medications and various disorders. X-ray chest radiography and computer tomography (CT) are two common anatomic imaging modalities that are routinely used in the detection and diagnosis of a variety of lung diseases.

* Corresponding author. Mobile: +91 9994249954.
E-mail address: bhuvana.csdept@gmail.com (C. Bhuvaneswari).

Medical images play a vital role in patient diagnosis, therapy, surgical, medical reference, and training. The Digital Imaging and Communications in Medicine (DICOM) standard [1,17] allows storing textual descriptions, known as metadata, along with the images. It was the most important breakthrough since the discovery of the X-rays, and CT has remained a cornerstone of diagnostic radiology throughout the years. An excellent overview of its technology and applications is given by Kalender [2]. CT imaging had evolved far enough in terms of speed and resolution to make it a valuable tool in the imaging of the lungs [3]. CT lung image classification with correction for perfusion gradient was suggested by Chabat et al., [4].

However, the large number of features actually represents a problem. It leads to the "dimensionality curse" problem [5], where the indexing structures degrade and the significance of each feature decreases, making the process of storing, indexing and retrieving extremely time consuming. Pathologies are clearly identified using automated CAD system [6]. It helps the radiologist in analyzing the digital images to bring out the possible outcomes of the diseases. The medical images are obtained from different imaging systems such as MRI scan, CT scan, Ultrasound scans.

A new process was proposed by automated method to quantification of affected lung parenchyma intensity distortion caused by gravity dependent perfusion gradient. Associated with the characteristics of exploitation and exploration, GAs can efficiently deal with large search spaces, and hence are less prone to get stuck into a local optimum solution when compared to other algorithms. This derives from the GAs ability to handle multiple concurrent solutions (individuals) in the search space and apply probabilistic genetic operators [7–9]. From the image processing point of view, it is important to gather as much features as possible to represent the images, yielding vectors with hundreds or even thousands of features to represent the images. Ginneken [10] has classified the lung regions extraction approaches into two different categories: either rule-based or pixel classification based category. Most of the proposed approaches belong to rule-based category [11,12], where a sequence of steps, tests and rules are used in the extraction process.

The computerized tomography has been found to be the most reliable method for early detection of tumors. Chiou et al. [12] proposed application of neural network based hybrid system for lung nodule detection, which based on artificial neural network architectures were developed for improving diagnostic accuracy and speed for lung cancerous pulmonary radiology. The configuration of the HLND system included the following processing phases; data acquisition and pre-processing, in order to reduce and to enhance the figure-background contrast, quick selection of nodule suspects based upon the most prominent feature of nodules, the disk shape and completed features pace determination and neural classification of nodules [13–15].

Genetic algorithms (GAs) are among the most used techniques to perform feature selection due to GAs ability to obtain either exact or approximate solutions in very large search spaces within tractable time. GA performs adaptive searching, following the standard concepts from natural genetics and evolution based on natural selection [16]. Owing to their potential, GA is employed in this work to perform feature selection.

The proposed work is organized as follows, the input of the system is CT lung images for which preprocessing is done to enhance the image, feature extraction is carried out by the proposed fusion median absolute deviation techniques which fuse the features of the gabor filter and Walsh Hadamard transform, feature selection is through genetic algorithm and the classification of the images is done through the classifiers and the performance measures, classification accuracy of each methods is discussed in this work.

2. Materials and methods

CT images of Lung diseases such as Emphysema, Bronchitis, Pleural effusion and normal lung are considered for classification in this work. The patient's age ranging from 15 to 50 comprising of both male and female are taken in this work. The images are obtained from the 16 slice Philips MX 16 evo CT scanner from Sri Manakula vinayagar medical college and hospital, Madagadipet, Puducherry and are anamolized by the radiologist. The original image is split into non-overlapping blocks of 16×16 windows and gray scale image is extracted.

3. Proposed work

The flow diagram of the proposed work is given in Fig. 1. The figure explains the block diagram for the proposed work that is represented in various steps as follows.

3.1. Preprocessing

Median filter and morphological smoothening filter techniques are applied to remove the noise from the images and enhance the image. The median filter will remove the salt and the pepper noises and produces the enhanced image. The dilation and erosion process is done for morphologically smoothening of the images. Erosion involves the alteration (removal) of pixels at the edges of regions, i.e., exchanging binary 1 value to 0, while dilation is the reverse process with regions growing out from their boundaries. The Erosion followed by dilation is known as Opening operation which suppresses the bright details smaller where dilation followed by erosion is closing operation which suppresses the dark details are computed. Herein this work the darker details are suppressed where the dilation is performed followed by erosion. The preprocessed image is obtained by applying the median filter and morphological smoothening.

3.2. Feature extraction

The feature extraction refers to creating a subset of new features by combination of the existing features. The purpose of feature extraction is to reduce original dataset by measuring certain features that distinguish one region of interest from another. Each subimage is taken from top left corner of the original image and the texture features and pixel coefficient values are extracted. The analysis and characterization of textures present in the medical images can be done by using statistical feature extraction method. Texture analysis is a quantitative method that can be used to quantify and detect structural abnormalities.
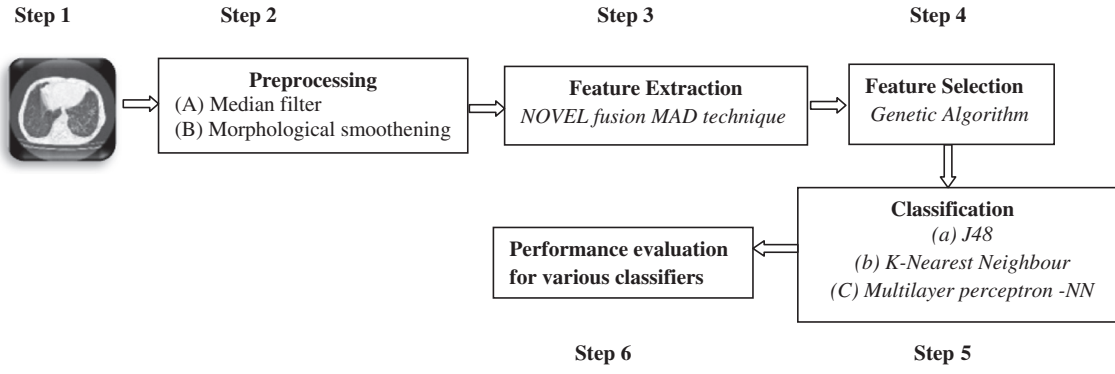
Step 1   Step 2   Step 3   Step 4

Preprocessing
(A) Median filter
(B) Morphological smoothening

Feature Extraction
NOVEL fusion MAD technique

Feature Selection
Genetic Algorithm

Classification
(a) J48
(b) K-Nearest Neighbour
(C) Multilayer perceptron -NN

Performance evaluation for various classifiers

Step 6   Step 5

Figure 1   Block diagram of the proposed work.

Gabor feature
(Texture Co-Efficient)

Walsh Hadamard Transform
(Pixel Co-Efficient)

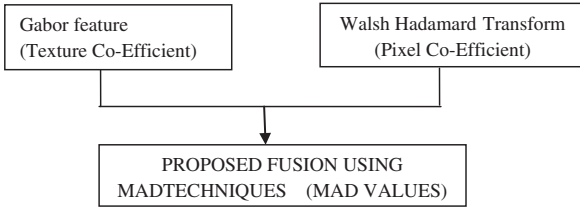PROPOSED FUSION USING
MADTECHNIQUES (MAD VALUES)

Figure 2   Block diagram of the proposed fusion technique.

Fig. 2 illustrates the proposed novel fusion feature extraction technique which combines the features of gabor filter and Walsh Hadamard transform.

3.2.1. Gabor filter

The analysis and characterization of textures present in the medical images can be done by using statistical feature extraction method. Gabor filters are band-pass filters with orientation selective and frequency-selective properties having optimal joint resolution in spatial and frequency domains. The 2D Gabor filter constitutes a sinusoidal plane of specific frequency and modulated Gaussian

$$G(x,y)\frac{1}{2\pi\sigma_x\sigma_y}\exp\left(-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)\right) \tag{1}$$

where σ is the spread of Gaussian window and x, y are co-ordinates.

Frequency and orientation representations of Gabor filters are found to be particularly appropriate for texture representation and discrimination. The Gabor filter will get the orientation as input, set with a given direction gives a strong response for locations of the target images that have structures in this given direction. So changing orientation will highlight the edges better. The image size of 256×256 pixels is divided into subimages of 12 blocks. The output of the gabor filter is the gabor co-efficients on the whole image. The features extracted from the Gabor filter are *sum, mean, min, max, median*.

3.2.2. Walsh–Hadamard transform

The Walsh–Hadamard transform (WHT) is a non-sinusoidal, orthogonal transformation technique decomposing a signal into a basic function set which are Walsh functions with rectangular or square waves with values of $+1$ or -1. The Walsh–

Hadamard transform returns sequency values. The Fast Walsh–Hadamard transform (FWHT) represents signals with sharp discontinuities accurately using fewer coefficients than Fast Fourier Transform. FWHT is a divide and conquer algorithm that recursively breaks a WHT of size N to two smaller WHTs of size $N/2$.

$$H_N = \frac{1}{\sqrt{2}}\begin{pmatrix} H_N - 1 & H_N - 1 \\ H_N - 1 & -H_N - 1 \end{pmatrix} \tag{2}$$

For a discrete series of N length the set of Walsh functions is N by N Hadamard matrix. The output of the WHT is the pixel co-efficients on the whole image .The features extracted from the WHT are sum, mean, min, max, median.

3.2.3. Proposed novel Fusion Median Absolute Deviation (MAD) technique

The Fusion of multiple samples extracted from the identical source similar to using multiple images from the lung CT scan is referred as multisampling. Fusion of multiple processing methods for individual samples is referred Multi-algorithm. In template level fusion, multiple templates combine to form a single template. Feature level fusion contains richer information than any other levels of fusion. The main advantages of feature fusion are that it derives the most discriminatory information from the original feature set and also eliminates redundant information from the feature set.

The features extracted from Walsh Hadamard Transform (WHT) and Gabor filter are fused using template level fusion. WHT extracts the features from the frequency domain and the Gabor captures the salient visual features corresponding to spatial localization, orientation selectivity, and spatial frequency. Features from Gabor and WHT are fused using Median Absolute Deviation techniques between the two features. MAD uses the median for the deviation scores, and it is more robust than the standard deviation as a measure of dispersion and is less susceptible to the influence of outliers when compared to the standard deviation. The advantage of the MAD is that it can be converted into values that approximate the standard deviation.

The median absolute deviation is calculated by taking the median of the absolute deviations from the median ($|X_1$-MED$|$, $|X_2$-MED$|$, …, $|X_n$-MED$|$) and the mean of the absolute deviations from the median is given by ($\sum (|X_n$-MED$|))/N$. The Gabor co-efficients are represented as $g_{i1}, g_{i2}, g_{i3}\ldots\ldots\ldots g_{in}$ and the Walsh Hadamard co-efficient

represented as $w_{i1}, w_{i2}, w_{i3} \ldots \ldots \ldots w_{in}$, fused feature vector $x_{i1}, x_{i2}, x_{i3} \ldots \ldots \ldots x_{in}$ is got by normalizing feature vector to obtain using MAD with the average of the same.

The features extracted from each block are combined for the whole image thus resulting in 60 features for each image. The extracted features are fed in for feature selection to obtain the top ranked features.

3.3. Feature selection

Feature Selection (FS) algorithm is a process of choosing a reduced relevant features that improves classification by searching for the best feature subset, from the fixed set of original features according to a given classification accuracy. It removes irrelevant or redundant features dealing with the dimensionality cruse, thus leading to reduce the computational and memory cost. The feature selection problem involves the selection of a subset of 'd' features from a total of 'D' features, based on a given optimization criterion. Genetic algorithm was utilized to search for the optimal solutions. The top 60 features were selected and their efficacy is investigated with various classifiers.

3.3.1. Genetic algorithm (GA)

Genetic algorithm [7] is a population based search methods and it moves from one set of points (population) to another set of points in a single iteration with likely improvement using set of control operators. GA is viewed as function optimizer, though problem ranges to which GA is applied are quite extensive features [8]. It improved fitness through evolution. A solution to a given problem is represented in the form of a string, called 'chromosome', consisting of a set of elements, called 'genes'', that hold a set of values for the optimization variables. Once an initial population is created, the individuals of this population are assessed through a fitness function, which informs the goodness of a chromosome in the solution of the optimization task. In addition, fitness is the function to be optimized by minimization or maximization in a genetic algorithm. Based on this fitness function, chromosomes are selected and some genetic operators like mutation, crossover is applied in the selected chromosomes. These chromosomes evolve, better individuals till it reaches the global optimum solution.

A. GA operators. The simplest form of GA involves three operators: selection, crossover (single point), and mutation. **Selection** selects chromosomes in population for reproduction. The fitter a chromosome, the more times it will be chosen to reproduce.

$$Ps(i) = \frac{f(i)}{\sum_{j=1}^{N} f(i)} \tag{3}$$

where $P_s(i)$ and $f(i)$ are probability of selection and fitness value for ith chromosome respectively [9].

The pseudocode and general framework for the genetic algorithm are represented in the Figs. 3 and 4 respectively. The procedure followed for the genetic algorithm is explained as follows.

In the initial population of the of the search space 30, these are the features extracted from the novel method of feature extraction and is combined via a crossover operator to

```
Begin;
Generate random population of P solutions
(chromosomes);
For each individual i∈P: calculate fitness (i);
For i=1 to number of generations;

Randomly select an operation (crossover or
mutation);
If crossover;
Select two parents at random ia and ib;
Generate on offspring ic = crossover (ia and ib);
Else If mutation;
Select one chromosome i at random;
Generate an offspring ic = mutate (i);
End if;
Calculate the fitness of the offspring ic;
If  ic is better than the worst chromosome then
replace the worst chromosome by ic;
Next i;
Check if termination = true;
End;
```

Figure 3   Pseudocode for genetic algorithm.

produce offspring, two point crossover with crossover rate of 0.7. The individuals in the population are then evaluated via a fitness function, the fitness function of the proposed work is the root mean square error (RMSE) and the process of crossover, evaluation, and selection is repeated for a predetermined number of generations or until a satisfactory solution has been found or condition fails. In the feature selection formulation of the genetic algorithm, individuals are composed of bit strings: a 1 in bit position indicates that feature should be selected; 0 indicates this feature should not be selected with the mutation rate is 0.01 and the stopping condition is Root mean square error threshold of 0.001 or 500 operations with generation of 500, fitness condition is RMSE and the crossover operator is two point. The demonstration of the genetic algorithm with the fitness function is represented in Fig. 5.

The figure represents the fitness function graph for various generations. The generations 7 and 27 are represented in the figure as a sample.

The datasets are ranked from the 1st column to the 59th column and the plot of the top ranked features is given as
y = Columns 1 through 27.
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2.
Columns 28 through 54
2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4 4 4
Columns 55 through 59
4 4 4 4 4

Current plot of k ranges from 1 to 60, the values in the column 1–4 represent the four class labels and 30 top ranked feature selected columns are
58, 32, 29, 15, 57, 60, 54,4,1, 8,53, 13,20,59,10, 35,42,37,7,16, 51, 3,56,6, 34, 9,2, 5, 17, 19
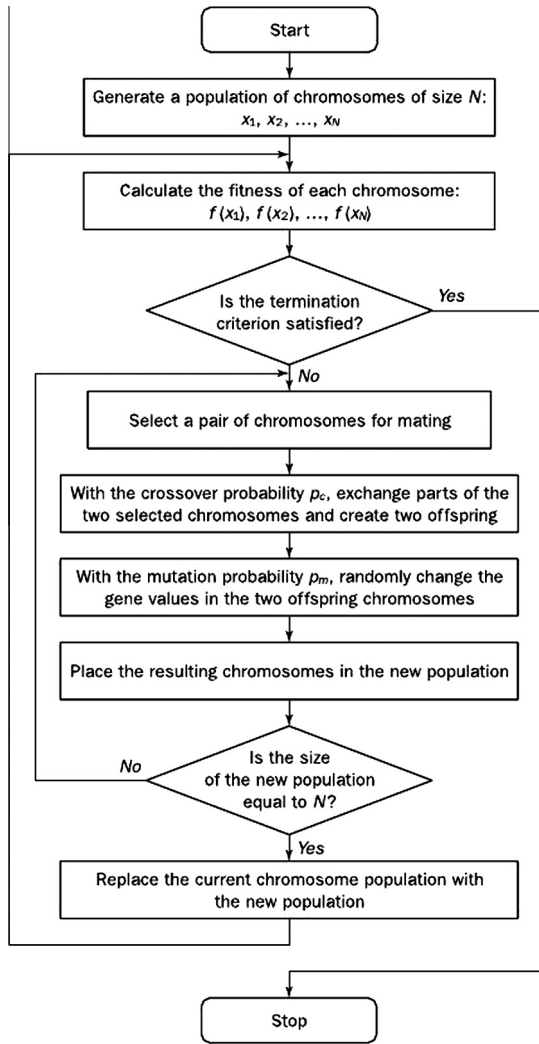
Figure 4 Framework for genetic Algorithm.

The above mentioned columns are represented as the top ranked features and the selected features are fed into the classifiers.

4. Classifiers

4.1. Decision trees

Decision tree learner is a tree structure where each non-leaf node represents a test on a feature, each branch denotes an outcome of the test, and each leaf node represents a class label. The decision tree classifier became popular due to the fact that the construction of a decision tree classifier does not require any domain knowledge, and the acquired knowledge in a tree form is easy to understand. In addition, the classification step of decision tree induction is simple and fast. Besides the splitting criterion, another interesting challenge of building a decision tree is to overcome the over-fitting of the data. J48 is the C4.5 algorithm's earlier version developed by J. Ross Quinlan. Decision trees represent information from machine learning algorithms, providing a quick way to express data structures [20]. C4.5 is a widely used decision tree based classifier, which is the implemented by C4.5 in this work to constructs pruned trees and subtree raising techniques.

4.2. K-nearest neighbor

K-NN is a typical instance-based prediction model. By K-NN, the class label of a new testing sample is decided by the majority class of its k closest neighbors based on their Euclidean distance The Nearest Neighbor Classification underlying intuitions are straightforward, and examples based on class of nearest neighbor are classified. More than one neighbor is considered and so the technique is usually called k-Nearest Neighbor (k-NN) where k nearest neighbors determines class. As training examples are needed at run-time, it is also called Memory-Based Classification. It is considered a Lazy Learning technique as induction is delayed by run time. As classification is training examples based it is also called Example-Based Classification or Case-Based Classification [21,22]. The formula for Minkowski distance is

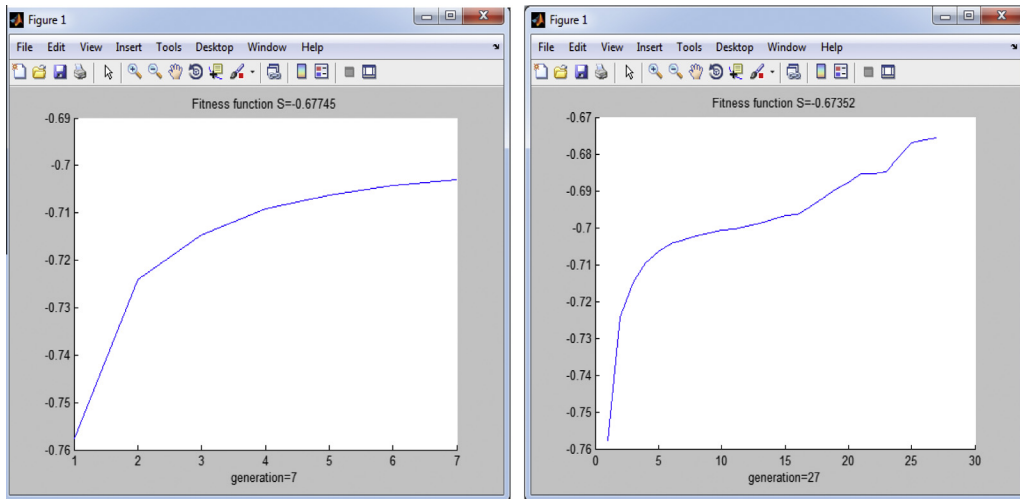$$MDp(q, xi) = \left(\sum_{f \in F} |Qf - Xif| \right)^{\frac{1}{p}} \tag{4}$$

Figure 5 Representation of the fitness function for various generations.
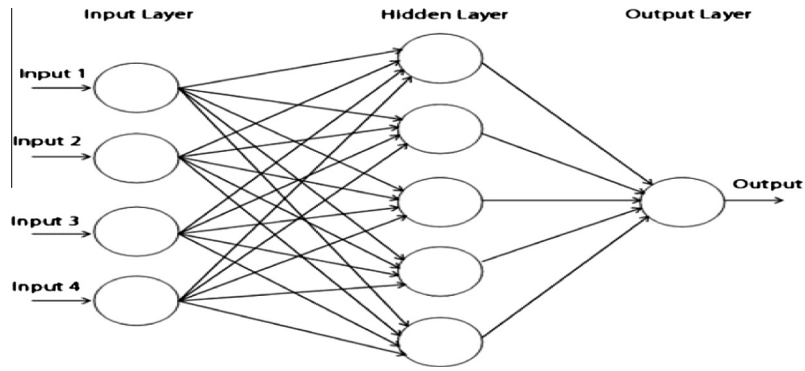
Figure 6   Block diagram of MLP-NN.

The KNN classifier ranks the images neighbors among the training set, and uses the class labels of k most similarity neighbors to predict the class of the input image.

4.3. Multilayer perceptron neural networks

A multilayer perceptron (MLP) has a hidden layer between the input and output layers. MLP can implement nonlinear discriminants (for classification) and nonlinear regression functions (for regression). The biological neuron's functions are modified by computing a differentiable nonlinear function (like a sigmoid) for MLP's each artificial neuron [23]. The input layer neurons as sensory units compute identity function, $y = x$. Each neuron in hidden and output layers computes *sigmoidal function* of input and weight values products sum of corre-

sponding connections. The output of jth neuron, Oj, in hidden or output layer is mathematically represented as

$$Ojth = \frac{1}{(1 + e^{-net_j})} \tag{5}$$

where $net_j = \sum W_{ji}o_i$ and $W_{ji} = the$ weight of the connection between ith and jth neuron.

It is a two-layer network because the input layer performs no computation. The hidden units must implement a nonlinear function i.e., sigmoid or else it is equivalent to a simple perceptron. Sigmoid can be seen as a continuous, differentiable version of thresholding. Extension of the perceptron learning algorithm to multiple layers by error back propagation from the outputs back to the inputs. Learning of hidden-to-output
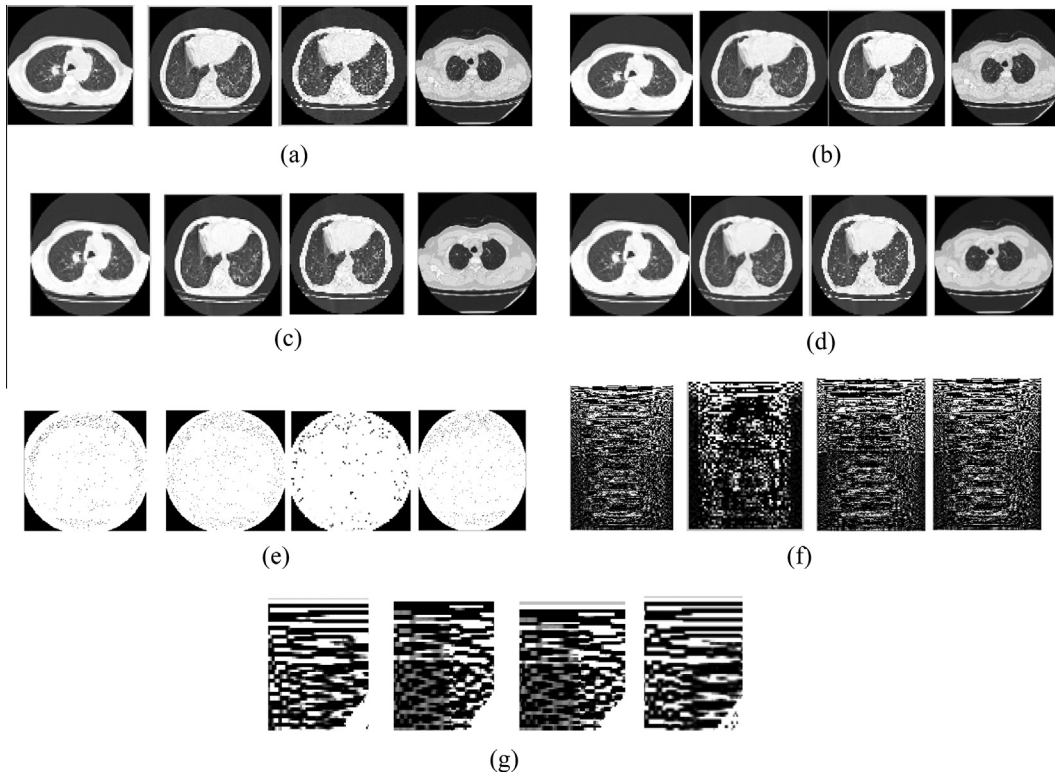

Figure 7   Output of various methods. (a) Input image, (b) enhanced image, (c) median filtered image, (d) morphologically smoothened image, (e) gabor filtered image, (f) Walsh Hadamard transform and (g) proposed fusion technique.

Table 1 Classification accuracy (in percentage).

Feature extraction and selection techniques	J48	KNN	MLP-NN
Gabor filter and genetic algorithm	77.19	72.88	85.96
Walsh Hadamard transform and genetic algorithm	72.88	74.58	77.19
Proposed fusion and genetic algorithm	83.05	86.44	91.53
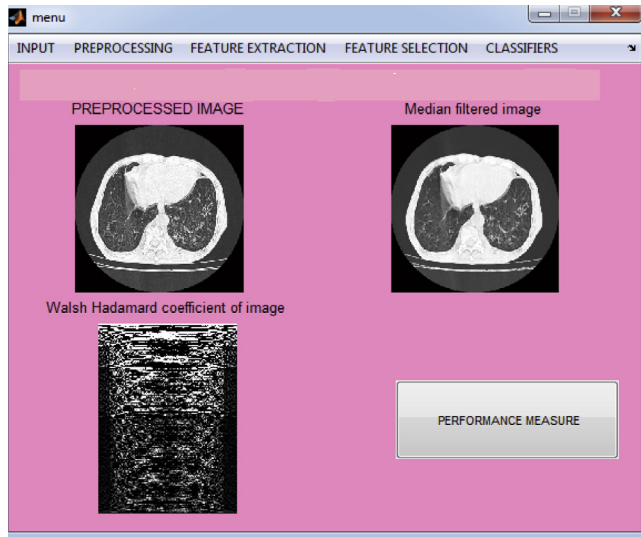
Figure 8 Output screen.
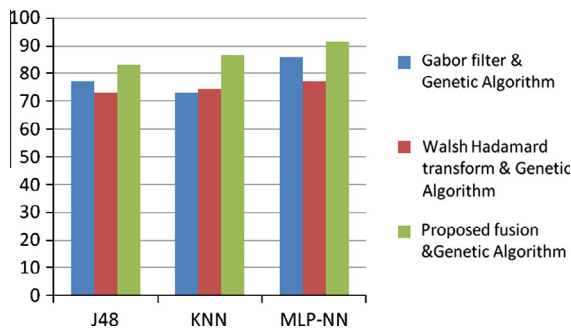
Figure 9 Classification accuracy (in percentage).

weights: like perceptron learning by treating the hidden units as inputs. Learning of the input-to-hidden weights: applying the chain rule to calculate the gradient.

Fig. 6 illustrates the block diagram of MLP-NN which represents the three layers. In the proposed work the input layer is represented as neurons where there is one hidden layer with the output layer having the possibility of the four class labels denoting the classification of the diseases.

5. Experiment and results

Our proposed method is implemented on lung diseases CT dataset based on proposed flow diagram as shown in Fig. 1. The input dataset consists of 400 images: 100 images are normal lung image and each diseases of bronchitis, emphysema, pleural effusion images each comprising of 100 images respectively. The texture features are extracted from the gabor filter and the pixel co-efficient is computed by the Walsh Hadamard transform and the features from Gabor and WHT are fused using Median Absolute Deviation. The fusion technique will combine the features extracted from the gabor and the WHT and the features overlapping these methods are eliminated and most appropriate features of mean, median, max, min of 15 windows of a image resulting in the 60 features are obtained in this proposed feature extraction method.

Feature selection is carried out using GA. The 60 extracted features are fed into selection process. The next step is to determine the relevance of each selected feature to the process of classifying the lung diseases. During the evaluation process by using GA, some features may be selected many times as the number of generation increases. If the feature was selected more times that feature was given as more important in the feature selection. The number of times the features selected was max, min, mean, and median. The parameter set for the GA algorithm is as follows: Population size is 30, Cross Over probability is two point crossover with crossover rate 0.7, mutation type is flip flop mutation rate is 0.01, and stopping condition is RMSE threshold of 0.001 or 500 operations with 500 generation fitness condition with two point crossover operator is employed and the fitness function is the Root mean square error value. Results show that, if the number of sample images increased, we get good classification accuracy for the tenfold cross validation method. A comparative study is performed for various classifiers using different datasets.

Fig. 7 represents the output achieved by various methods using matlab code for the four different dataset.

The output screen of the research work in the graphical user interface is given in Fig. 8.

The output of the various classification algorithms with the classification accuracy is depicted below for three types of classifiers in Fig. 9. The classification accuracy is tabulated in Table 1.

Table 2 presents the classification of datasets by their respective diseases for a given set of data. The performance

Table 2 Classification details disease wise.

Classifiers	Correctly classified as emphysema	Correctly classified as Bronchiestasis	Correctly classified as pleural effusion	Correctly classified as normal
GA_J48	87	85	82	84
GA_KNN	87	88	85	87
GA_NN	92	91	90	92

Table 3 Performance measures for various classifiers.

Techniques	J48			KNN			MLP-NN		
	Precision	Recall	F-Measure	Precision	Recall	F-Measure	Precision	Recall	F-Measure
Gabor filter and GA	0.75	0.78	0.78	0.72	0.71	0.77	0.84	0.86	0.85
Walsh Hadamard transform and GA	0.72	0.71	0.73	0.75	0.75	0.75	0.75	0.78	0.77
Proposed fusion and GA	0.83	0.83	0.72	0.88	0.86	0.87	0.92	0.91	0.92
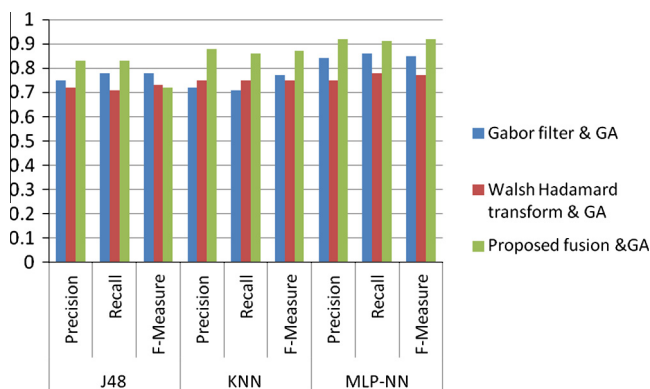
Figure 10 Performance measures.

measures such as precision, recall, F-Measure are presented in the Table 3.

The precision refers to the measurement deviation from true value and its scatter, recall refers to the measurement deviation from the false value. The F-Measure refers to a measure that combines precision and recall is the harmonic mean of precision and recall. The various values of the precision, recall, F-measure are calculated for the J48, KNN, MLP-NN classifiers.

The graphical representation for precision, recall and F-measure is presented in Fig. 10.

6. Conclusions

In this work a novel fusion based feature extraction is proposed and feature selection is done by genetic algorithm that selects the top ranked features and classification is done through J48, KNN, MLP NN classifiers to classify the lung CT dataset. The algorithm has been designed based on the concept of texture and pixel co-efficient features. This method effectively works well for the detection of lung diseases with high sensitivity, specificity and accuracy. The classification accuracy for the tenfold crossvalidation method shows 85.96 for gabor filter, 77.19 for Walsh Hadamard transform and 91.53 for the proposed feature extraction method. Results show that MLP NN classifier with median absolute deviation techniques and genetic algorithm for feature selection yields better results. This justifies the choice of using fusion with the genetic algorithm with the classifier. This approach has potential for further development because of this simplicity that will motivate to classify the types of lung diseases. The developed classification system is expected to provide valuable diagnosis for the physicians. The work can further be extended by including more feature extraction and selection methods for classifying more lung diseases.

References

[1] Graham R, Perriss R, Scarsbrook A. Dicom demystified: a review of digital file formats and their use in radiological practice. Clin Radiol 2005:1133–40.

[2] Kalender WA. Computed tomography: fundamentals, system technology, image quality, applications. 3rd ed. Erlangen: Publicis Publishing; 2011.

[3] Pugatch RD, Faling LJ. Computed tomography of the thorax: a status report. Chest 1981;80:618–26.

[4] Chabat, F, Hansell DM, Yang GZ. CT lung image classification with correction for perfusion gradient. In: 7th international conference on image processing and its applications; January 1999. p. 402–406.

[5] Mathew TV. Genetic algorithm. Bombay, Mumbai: Indian Institute of Technology; 1996.

[6] Sastry K, Goldberg D, Kendall G. Genetic algorithms. In: burke E, kendall G, editors. Search methodologies-introductory tutorials in optimization and decision support techniques. Berlin: Springer; 2005. p. 97–125 [Chapter 4].

[7] Golberg DE. Genetic algorithms in search, optimization and machine learning. Newyork: Addison Wesley; 1989.

[8] Haupt RL, Haupt SE. Practical genetic algorithms. second ed. New Jersey, United States: John Wiley & Sons; 2004.

[9] Melanie M. An introduction to genetic algorithm. Bradford Book. London, England: The MIT Press Cambridge, Massachusetts; 1999, Fifth printing.

[10] Malcok M, Aslandogan Y, Yesildirek A. Fractal dimension and similarity search in high-dimensional spatial databases. In: IEEE international conference on information reuse and integration, Waikoloa, Hawaii, USA; 2006. p. 380–384.

[11] Speckmann EJ, Elger CE. Introduction to the neurophysiological basis of the EEG and DC potentials. In: Neidermeyer E, Lopes da Silva F, editors, Electroencephalography, 5th ed., Lippincott Williams & Wilkins, New York, 2005 [Chapter 2].

[12] 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.

[13] Ginneken BV, Romeny BM, Viergever MA. Computer-aided diagnosis in chest radiography: a survey. IEEE Trans Med Imaging 2001;20(12).

[14] Kanazawa K, Kawata Y, Niki N, Satoh H, Ohmatsu H, Kakinuma R, Kaneko M, Moriyama N, Eguchi K. Computer-aided diagnosis for pulmonary nodules based on helical CT images. Comput Med Image Graph 2008;22(2):157–67.

[15] Lin D, Yan C. Lung nodules identification rules extraction with neural fuzzy network. Neural Inform Process IEEE 2002;4:2049–53.

[16] Golberg DE. Genetic algorithms in search, optimization and machine learning. Reading MA: Addison Wesley; 1989.

[17] Graham R, Perriss R, Scarsbrook A. Dicom demystified: a review of digital file formats and their use in radiological practice. Clin Radiol 2005;60:1133–40.

[20] Loh WY. Improving the precision of classification trees. Annal Appl Stat 2009;3(4):1710–37.

[21] Hassan MR, Hossain MM, Bailey J, Ramamohanarao K. Improving k-nearest neighbour classification with distance func-tions based on receiver operating characteristics. In: Machine learning and knowledge discovery in databases LNAI. Berlin Heidelberg: Springer; 2008. p. 489–504.

[22] Gil-Pita, Yao R. Evolving edited k-nearest neighbor classifiers. Int J Neural Syst Dec 2008;18(6):459–67.

[23] Roy K, Chaudhuri C, Kundu M, Nasipuri M, Basu DK. Comparison of the multilayer perceptron and the nearest neighbor classifier for handwritten digit recognition. J Inform Sci Eng 2005;21(6):1245–57.