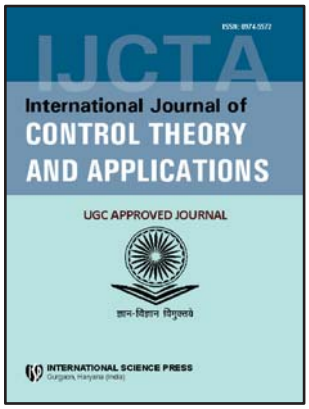

International Journal of Control Theory and Applications

ISSN : 0974-5572

Volume 10 • Number 33 • 2017

A Novel Optimization Based Ensemble Disease Prediction Model for COPD Prediction

Banda Srinivas Raja[a] and Tummala Ranga Babu[b]

[a]*Research Scholar, Department of ECE Acharya Nagarjuna University, Guntur, India*

[b]*Professor, RVR & JC College of Engineering, Acharya Nagarjuna University, Andhrapradesh, India*

Abstract: Medical image disease prediction is one of the essential tasks in both computational biology and computer vision is concerned. Image classification models on Chronic Obstructive Pulmonary Disease (COPD) data are growing rapidly for the prediction of disease level. COPD becomes a highly prevalent disease that impacts both patients and healthcare system. In the medical applications, image classification algorithms are used to predict the disease severity that can help in early diagnosis and decision-making process. Extensive amount of research have been carried out since many years to predict the disease patterns on medical image databases. However, most of the image classification models are failed to predict the disease patterns using the existing trained patches or templates. Accuracy, recall, precision, f-measure are the statistical measures which influence the performance of image classification models for disease prediction. Some of widely accepted image classification models such as SVM, FSVM, Random Forest, Naïve Bayes, KNN, GA,etc are used to evaluate the COPD disease prediction results. In our paper, we have designed and implemented a novel image classification model for COPD disease prediction. Experimental results proved that proposed model has high computational accuracy in terms of false positive rate, *f*-measure compared to traditional ensemble models.

Keywords : *COPD disease prediction, ensemble model, image classification.*

1. INTRODUCTION

Image processing plays a major role in medical diagnosis. Feature extraction is a significant phase of image classification. It is responsible for accurate representation of images. Many researchers proposed vast amount of feature extraction schemes for various classifiers. Both binary and multi-class classifications can affect performance level of feature extraction. Several researches have been carried out to develop a feature extraction scheme for optimal performance. The main objective of this scheme is to retrieve the relevant features of images can add it to a label. The main difficulty of image classification is to study properties of image features. The numerical features are assigned to classes. This classification is based on the image contents. Performance and accuracy rate of image classification models are determined by numerical properties, that signifies data. Among many feature extraction schemes, every individual approach has its pros and cons. Relevant features determine the strength of every approach. Below are three important features of image feature extraction.

Color Features: Color feature is a significant feature of image classification as well as retrieval. Color features are generally represented by color histograms. This feature does not depend on size, rotation and the zoom of images.

Texture Features: Texture features are considered for huge complex images sharing repetitive regions. It is robust in nature. Texture can be defined as a group of pixels with some common properties. It can be further categorized into two parts, those are:- spatial and spectral texture feature extraction.

Shape Features: These shape features are widely used in the area of objection recognition and shape description. Again shape features extraction scheme is split into further two sub-features, those are:- contour-based and region based schemes.

Region-based methods use the whole region while calculating the feature whereas, contour-based approaches evaluate the feature from its boundary value while discarding the inner regions.

Feature selection scheme can be categorized under dimensionality reduction approaches. It is mostly used in the area of data mining and knowledge discovery. It aims for deletion of irrelevant and duplicate features by keeping only the distinct ones. It limits data transmission rate and makes data mining more efficient. Some of other advantages are:- packet collisions, data rate and storage. The resulted outcome of feature selection enhances quality and performance. Feature selection algorithm has three basic steps, those are:- screening, ranking and selecting. In the screening phase, it discards unwanted records (like noisy and redundant image predictors). Ranking can be defined as a sorting scheme which adds ranks according to priority level. At last selecting can be described as extracting the required and important predictors while discarding the others. The results of this process contain screened, ranked and selected predictors which are essential.

Image classification approaches are implemented widely in educational and research area of biomedical domain. Numbers of image classification algorithms are developed to classify medical images. Classification process occurs in two phases, those are:- training and testing phase. Classification algorithm helps in constructing classification model by training dataset. In the testing phase, the overall performances are evaluated. Many classification models were proposed by various authors and researchers since years. Some of the popular classification models are briefly described below.

Support Vector Machine (SVM) is a specific type of construction-based classification model. It involves the concepts of statistical learning. This model shows better performance than that of other conventional approaches of image classification process. Decision boundaries are determined by hyperplanes. SVM also defines data points among different classes. Classification gives rise to both linear and non-linear problems which are overcome by SVM. The algorithm calls a mapping function for mapping purpose. Mapping is actually carried out by taking original data points from input space and transforms it to a high-dimensional feature space. All these operations are done through a specific kernel function.

ROC curve makes Naive Bayes model efficient and significant classification scheme. It is better as compared to logistic regression scheme, nearest neighbour, decision tree and neural network according. The NB model is constructed by NB classification algorithm and is mostly implemented in research domain. The model is a simple model having few parameters. It also achieves better performance than above classification approaches.

KNN is commonly known as K-Nearest Neighbour classification model. It comes under instance-based classification. The KNN algorithm computes approximation function and it is terminated when classification process is completed. Due to unavailability of training phase, the classification is termed as lazy learning. The classification process is completed only in the testing phase, as the training phase is absent. These training data are evaluated in testing phase to compute the performance. In case of large datasets, whole training data are split into smaller partial datasets. The most common application of KNN is implementation in density estimation process and parametric estimation process.

Like SVM, Hidden Markov Model (HMM) can be also categorized under statistical classification model. The whole system is considered as a Markov process. It also contains some unobserved or hidden states. This model is an example dynamic Bayesian model. Some applications of this model are:- Face recognition and detection process. The main disadvantage of this classification is, it discards state structures in super states.

Divide-and-conquer algorithm is applied in decision tree classification models for the construction of trees. The instances are integrated with various attributes. The tree consists of nodes and leaf nodes. Leaf nodes are descendent of nodes. The outcome of each node is either true (1) or false (0). Rules are constructed by the path from root to leaf nodes. Pre-conditions can be defined by using path and nodes The unnecessary and redundant pre-conditions are eliminated by tree pruning process.

As Random forest algorithm is responsible for classification of vast amount of data. Hence, it is considered as one of the efficient and best classification algorithms. In decision tree classification, classification algorithm generates classification models. The final result includes modal class formed by individual trees. It integrates weak learners to form strong learners, which is a strong concept of this classification.

Genetic algorithm (GA) helps to generate optimal solution for classifications. It is a type of evolutionary and stochastic algorithm. Candidate solutions are encoded by crossover and mutation operations. Fitness function is responsible for selection criteria for the generation of offspring. The initial population is randomly and haphazardly created. If the candidate solution gets minimum fitness score at each generation, then that satisfies the minimum criteria of fitness according to the genetic algorithm classification scheme.

In our paper, we have thoroughly studied and analyzed all pre-existing image feature extraction and classification models for disease prediction.

2. RELATED WORKS

M. Anthimopoulos et.al. introduced a new classification scheme for detection of Interstitial Lung Diseases (ILD) by merging the concepts of DCT and random forests [1]. They proposed this scheme in order to enhance the detection accuracy of ILD. The authors selected HRCT images for their work and studied various ILD patterns. They used a specific DCT-centred filter bank feature extraction. From this filter bank, local frequencies are evaluated. Final feature vector is computed by considering the graylevel histograms. Random Forest classification algorithm is used by the researchers for the classification. They validated their theory with experiments and showed that both the performance and efficiency of the technique outperforms other previously proposed research ideas. They got 89% accuracy rate through their method. Future research may be carried out to extend the 2-dimensional filter bank to 3-dimensions.

M. Anthimopoulos et.al. studied various image feature extraction schemes and identified the need of automated tissue characterization for prediction of ILD [2]. They presented a new Convolutional Neural Network (CNN) to classify ILD. They suggested a new method involving 2x2 kernel layers and LeakyReLU activators. It also contains a average pooling and three dense layers. They considered training datasets of 14696 images for computation of CNN collected from 120 CT scans out of huge sources. The researchers performed experiments and found that their method gives 85.5% accuracy rate, which is better than that of other developed approaches at that time. Future research may involve by extending CNN to 3D data and merging the suggested method with CAD in order to achieve better performance.

C. Tianrui, X. Gang et.al proposed a new approach for texture feature extraction in case of HRCT images [3]. ROI is considered as an effective measure for regional lung diseases like tumour. The researchers introduced an advanced segmentation method concentrating on tolerance granular space algorithm and region-growing approach. The outcome after implementing the approach focuses on extraction of texture parameters. The proposed algorithm considered mean value of ROI in order to choose seed points automatically and to enhance manual selection of algorithm. The proposed algorithm can perform well with lack of adjustment in threshold parameters. It is also responsible for improvement of tolerance relation system. HRCT is a perfect implementation of the presented scheme. The researchers proved that, their approach is more pertinence than that of conventional texture analysis.

J. K. Dash, et.al. analyzed and compared some of previously introduced rotation-invariant texture feature extraction in order to classify and retrieve ILD [4]. They used the basic concept of texture direction to retrieve texture features through DWT. The authors selected database with HRCT images having five types of lung tissue. They conducted tests and its outcomes proved that, the proposed technique has both good retrieval and classification strategy. Hence this scheme is responsible for automatic identification of disease. Performance rate is increased highly as compared to the other pre-existing schemes. The result included anisotropic texture of lung tissues.

J. K. Dash et.al. analyzed and identified the problem of block classification in case on lung tissue related diseases that is, presence of inaccurate and smooth boundaries [5]. They considered three automated image segmentation schemes and their effect on smooth boundaries out of numbers of lung tissue disorder patterns. The researchers selected HRCT images from a publicly available database which contains patients' data infected with ILD. They conducted experiments on these images. Different algorithmic concepts are merged together such as, Gaussian Mixture Model (GMM) ,Markov Random Field (MRF) and Mean Shift (MS). The authors used two-fold cross validation technique to choose best parameters and achieve best performance rate among all algorithms. They got best performance in case of Mean Shift algorithm.

J. D. Deng developed a new image categorization algorithm by feature schemes merged with feature analysis approach [6]. He used an edge descriptor which concentrates on canny detector and extended MPEG-7 features. Outcomes of the preliminary stage focus to enhance effectiveness. With the integration of semantic analysis, better categorization can be achieved. By the combination of some schemes though resulted low classification accuracy, it showed better performance in case of feature extraction. As dimension reduction is not an effective technique, there is need of merging different feature schemes for optimal performance. Further future work is needed to include low-level features along with semantic features which will resolve the categorization problems.

M. Kakar, et.al proposed an approach for extraction of fuzzy classification rules from texture-segmented high resolution HRCT lung images [7]. This approach can be applied on the field of diagnosis and high-precision radiation therapy. But it is still a challenging question in Non-Small Cell Lung Carcinoma (NSCLC) disease. For NSCLC disease, their presented classifier can extract fuzzy classification rules out of texture-segmented HRCT images. Fuzzy Information System (FIS) can be implemented on overlapping regions. They evaluated their technique over 138 regions retrieved out of CT scans. These CT scan images are collected from lung cancer patients. The researchers picked two variety of tissues C1(normal) and C2(diseased) as negative and positive respectively. Further future work can be done to implement classifier on 4DCT sequence for more accurate and updated result. Again the classifier can be extended to multi-classes in order to identify healthy tissues. This will optimize the treatment plan.

N. Kato, et.al. studied and analyzed different algorithms for classification of diffuse lung diseases, but none of them was able to classify heterogeneous texture distribution within a ROI. [8]. In reticular and honeycombing patterns, some features of ROI are recorded. The authors developed a new bag-of-feature method to resolve the above issue. Texture images can be presented as histograms, generated by clustering of images. Two techniques are used by them to extract the features, those are:- Scale Invariant Feature Transformation (SIFT) descriptor and intensity descriptor. These descriptors can be specified in a specific image class. The only flaw of local feature is, unavailability of translation and rotation. The researchers overcome this problem through sampling many local regions. They validated their work by experimenting on 5 image classes such as (ground, reticular, honeycombing, emphysema and normal) with 1109 ROIs and 211patients. They gain 92.8% classification accuracy, better than that of all other traditional approaches with GLCM feature.

B. Kaur and S. Jindal developed a new feature extraction scheme based on SURF method on OPEN CV platform [9]. They used C interface for application of different image processing algorithms. The authors combined feature extraction scheme with the novel model for their algorithms. They evaluated performance by analyzing execution period, rotation, accuracy, etc. They stated that, their method is best applicable to

classify complex medical images. They focused on feature extraction and selected dataset of medical images. Later extraction process is carried out. The OPEN CV platform presents a C++ interface for the execution of image processing algorithms. The authors concluded that, their work achieves better performance than that of traditional MATLAB routines used for image processing purpose.

Y. Lee, N. Kim, et.al tried to enhance efficiency through a cascade classification technique to diffuse ILD with HRCT images [10]. Classification takes place through all other classes. Order of classification accuracy was determined from highest to lowest accuracies. Classification followed by feature extraction was repeated for new images. Performance is evaluated by calculating computation cost and classification accuracy rate. To make this process automated, support vector machine approach can be applied. Ten-folding technique helps in cross validation. The researchers decreased computation cost with 46% and classification accuracy was increased upto 92.04%.

B. Li, W. Li and D. Zhao tried to overcome the issue of automatic image category annotation and multi-scale image classification with their newly developed approach [11]. They termed their method as multi-scale feature-based medical image classification. They retrieved numbers of features like gray-level, texture, shape features and frequency domain features. They presented a classification framework for this classification integrated with feature extraction. These features play a vital role in classification process and the resulted outputs are compared. The researchers gained significant accuracy as compared to the other existing techniques. The proposed technique is a comparative approach for multi-scale medical image classification. The developed technique results high accuracy as compared to all classifiers and highest accuracy was recorded when compared with the selected classifiers.

T. Nuzhnaya, et.al. introduced a new approach for classification of texture patterns in CT lung imaging [12]. They distinguished between normal and diseased tissue through vector quantization and image compression. Instead of considering the image of whole lung, they concentrated on retrieval of descriptors from each ROIs. These are determined by domain experts. Local optimal is computed after the evaluation of ROIs by using Generalized Lloyd algorithm. KNN, SVM and neural network classifiers are used for classification of normalized histograms. The authors gained 98% classification accuracy through their experiments. Texture descriptors can be implemented through searches, clustering, classification and other operations involving lung image databases.

3. ADVANCED ENSEMBLE CLASSIFICATION MODEL FOR DISEASE RISK PREDICTION

The present work introduces an automatic feature extraction and COPD severity classification for COPD CT images in the repository[2]. The overall workflow can be classified into two phases as shown in Figure . The ensemble classifier is usually considered to be more efficient and accurate than individual classifier. The simple majority voting is widely used method for voting best classifier among multiple base classifiers. We combine the weak classifiers by using the proposed feature selection and random forest methods instead of the class label.

1. Feature selection technique is proposed to extract the COPD relevant features for ensemble classification.
2. A novel ensemble feature selection based classification model for COPD classification among different severity classes.
3. Proposed ensemble classifier combines multiple weak classifiers for COPD severity classification.

This work have the following advantages: Gives better performance of classification accuracy, solves class imbalanced problem , best ensemble model in multi-level classification .It works better than other multiple classifier schemes which suffer from the problem of ensemble selection.

Preprocessing : Preprocessing of medical images is the primary step in disease classification process, to minimize noise and to optimize the image quality. Traditional adaptive median filter is used to enhance the image quality as well as remove the poison noise from the images. In the median filter, a window moves along the image and the computed median value of the window pixels becomes the output. It preserves the edges and reduces the noise in the image. Each pixel is replaced with the median value of the neighborhood of the input pixels. In our preprocessing model , we extended the traditional adaptive median filter to remove the noise in the training and test medical images.

Proposed Ensemble Disease Classification Model : Prior to COPD image classification, image features are extracted into database for training severity levels using the proposed ensemble classifier. These features are used to classify COPD into two classes, normal and emphysema disease severity. We implemented a novel ensemble image classifier using Naïve bayes, SVM , enhanced random forest tree and disease prediction algorithm as base classifiers on the training data.

Input : Featured Training data T_r , Unlabeled test Data T_e

Output : Test image class prediction.

Procedure: Initialize T_r and T_e

For each feature F(i = 1, n) in T_r and T_e

Do

If(F[i] is numeric)

Then $\text{INormalize}(\text{F}[i]) = (\text{F}[i] - \mu_{\text{F}[i]})/(\text{Max}(\text{F}[i]) - \text{Min}(\text{F}[i]))$

End if

If(F[i] is categorical)

Then

$\text{INormalize}(\text{F}[i]) = (\sum \text{Prob}(\text{F}[i]/\text{C}_m))/(\text{Max}(w(\text{F}[i])) - \text{Min}(w(\text{F}[i])))$

Where C_m represents the m-classes

End if

Done

Algorithm2 : Proposed Disease prediction algorithm

Input: Normalized data ND, Max iterations Max, bias λ , Weighted vector W, c is class label, I is instance, I[c] is nistance class, I[d] : dimensional instance value.

Output : Disease prediction class label.

Procedure:

Step 1: Initialize weights to all instances.

For each instance in ND

Do

$W_d \leftarrow 0$

Done

Step 3: Initialize bias to 0.

$\lambda \leftarrow 0$

Step 4: For each iteration i = 1,2,…Max do

For each instance I in D do

$$R \leftarrow \sum_{d=1}^{\#dimensions} \log(W_d).I_d + \lambda$$

Done

Done

Step 5: if R(log(W)I + bias) < 0

Then

For each dimension d do

$$W_d \leftarrow \log(W_d) + I_c.I_d$$
$$\lambda = \lambda + R$$

Done

//updata previous cached weight

$$W_{p,d} \leftarrow |\log(W_{p,d})| + R.\ C_t.\ I_d$$

//updata cached bias values

$$\lambda_p = \lambda_p + I[c].\ C_t$$

End if

$$C_t = C_t + 1$$

return $(W\text{-}\frac{W_{p,d}}{\sqrt{C_t}}, \lambda - \frac{\lambda_p}{\sqrt{C_t}})$

Positiveness checking:

$$R' \leftarrow \sum_{d=1}^{\#dimensions} (\log(W_d) + I_d) I_d + \lambda + 1$$

$$R' \leftarrow \sum_{d=1}^{\#dimensions} ((\log(W_d).I_d + I_d.I_d) + \lambda + 1$$

$$R' \leftarrow \sum_{d=1}^{\#dimensions} \log(W_d).I_d + \sum_{d=1}^{\#dimensions} I_d^2 + \lambda + 1$$

$$R' \leftarrow \sum_{d=1}^{\#dimensions} \log(W_d).I_d + \lambda + (\sum_{d=1}^{\#dimensions} I_d^2 + 1)$$

$$R' \leftarrow R + \sum_{d=1}^{\#dimensions} I_d^2 + 1 \qquad \text{---- Eq (1)}$$

From the Eq. (1) , as $\sum_{d=1}^{\#dimensions} I_d^2 > 0$, R' always greater than 1.

Gradient Descent optimization function using 2-Norm Regularization for decision boundary.

$$\ell(W_d, \lambda) = \sum_n e^{-I_c \sum_{d=1}^{\#dimensions} \log(W_d).I_d + \lambda} + (\lambda / 4) \| W_d \|^2$$

Here $\lambda / 4$ is used to optimize the gradient preprocessed w.r.t bias.

$$\frac{\partial \ell(W_d,\lambda)}{\partial \lambda} = \frac{\partial}{\partial \lambda}\sum_n e^{-I_c \sum_{d=1}^{\#dimensions} \log(W_d).I_d + \lambda} + (\lambda/4)\,||\,W_d\,||^2$$

$$\frac{\partial \ell(W_d,\lambda)}{\partial \lambda} = \frac{\partial}{\partial \lambda}\sum_n e^{-I_c \sum_{d=1}^{\#dimensions} \log(W_d).I_d + \lambda} + \frac{\partial}{\partial \lambda}(\lambda/4)\,||\,W_d\,||^2$$

$$\frac{\partial \ell(W_d,\lambda)}{\partial \lambda} = \sum_n \frac{\partial}{\partial \lambda}(-I_c \sum_{d=1}^{\#dimensions} \log(W_d).I_d + \lambda) e^{-I_c \sum_{d=1}^{\#dimensions} \log(W_d).I_d + \lambda} + \frac{\partial}{\partial \lambda}(\lambda/4)\,||\,W_d\,||^2$$

$$\frac{\partial \ell(W_d,\lambda)}{\partial \lambda} = \sum_n -I_c.e^{-I_c \sum_{d=1}^{\#dimensions} \log(W_d).I_d + \lambda} + (1/4)\,||\,W_d\,||^2$$

$$\text{as } \lambda \leftarrow \lambda - \eta.\frac{\partial \ell(W_d,\lambda)}{\partial \lambda}$$

$$\frac{\partial}{\partial W_d}\ell(W_d,\lambda) = \frac{\partial}{\partial W_d}\sum_n e^{-I_c \sum_{d=1}^{\#dimensions} \log(W_d).I_d + \lambda} + \frac{\partial}{\partial W_d}(\lambda/4)\,||\,W_d\,||^2$$

$$\frac{\partial}{\partial W_d}\ell(W_d,\lambda) = \frac{\partial}{\partial W_d}\sum_n e^{-I_c \sum_{d=1}^{\#dimensions} \log(W_d).I_d + \lambda} + \frac{\partial}{\partial W_d}(\lambda/4)\,||\,W_d\,||^2$$

$$\frac{\partial}{\partial W_d}\ell(W_d,\lambda) = \frac{\partial}{\partial W_d}(-I_c \sum_{d=1}^{\#dimensions} \log(W_d).I_d + \lambda)\sum_n e^{-I_c \sum_{d=1}^{\#dimensions} \log(W_d).I_d + \lambda} + (\lambda/4)2.\,||\,W_d\,||$$

$$\frac{\partial}{\partial W_d}\ell(W_d,\lambda) = (-I_c/W_d)\sum_n e^{-I_c \sum_{d=1}^{\#dimensions} \log(W_d).I_d + \lambda} + (\lambda/2)\,|\,W_d\,||$$

$$\text{as } W_d \leftarrow W_d - \eta.\frac{\partial}{\partial W_d}\ell(W_d,\lambda)$$

Algorithm 3: Ensemble Steps

$D_0 \leftarrow <1/D_1, 1/D_2, 1/D_3, ...1/D_N>$ N uniform data partitions

$C_1 \leftarrow$ call Alg orithm2

$C_2 \leftarrow$ Random forest

$C_3 \leftarrow$ NaiveBayes

$C_4 \leftarrow$ SVM

for each classifer *k* in C_i do

$C_i = \text{classifier}(D, D_{k-1})$ // Training phase

$\hat{P}_n \leftarrow C_i(x_n)$,$\forall n$ prediction on training instances

$\hat{E}_k \leftarrow \sum_n D_{k-1}[n] \;/\; P_n \neq \hat{P}_n$

$$\phi_k = \frac{|\,D\,|}{2.|\,D_{k-1}\,|}\log\left(\frac{1-\hat{E}_k}{\hat{E}_k}\right)$$

end for

return $f(\hat{m}) = \text{sgn}[\sum_k \phi_k.\; C_k(\hat{m})]$

4. EXPERIMENTAL RESULTS

To evaluate the novel ensemble COPD feature performanceon emphysema diasgonis , we used the online emphysema dataset from http://image.diku.dk/emphysema_database/.The database consists of 115 HRCT images of size 512x512 and these images belong to a group of 39 categories including smokers and non-smokers with COPD. Each image is labeled using the ROI pattern and severity by an experience pulmonologist and radiologist .The serverity patterns are classified as normal tissue(NT), paraseptal- emphysema (PSE) ,centrilobular- emphysema (CLE) and panlobular- emphysema (PLE). Also , the severity levels labeled for each slice is classified as no emphysema(0), minimal (1), mild (2) ,moderate (3) ,severe (4) and very severe (5). To evaluate the COPD severity level, we used different statistical metrics such as : Recall , Precision, false positive and True positive rates.

Sample COPD Feature Extraction Data in arff format :

@relation COPD

@attribute histogram real

@attribute LBP real

@attribute Similarity real

@attribute class {0,1,2,3,4,5}

@data

2.34375,90.03257751464844,7.437084032141644,0

1.953125,87.79067993164062,7.483955051587976,1

1.953125,87.33673095703125,8.126940046037946,0

2.734375,85.11886596679688,7.673354375930059,2

2.34375,84.82437133789062,7.727759225027902,0

2.734375,90.98129272460938,7.9947153727213545,0

2.34375,91.86859130859375,8.585700988769531,4

2.34375,91.84074401855469,8.160073416573661,0

2.34375,87.11166381835938,8.343639373779297,0

2.34375,86.01722717285156,8.483238220214844,5

1.953125,86.78131103515625,8.840520758377878,0

2.34375,93.05763244628906,8.116295224144345,0

2.34375,93.365478515625,7.547295611837636,3

1.953125,93.91098022460938,8.293260846819196,1

2.34375,88.34190368652344,8.179912567138672,0

2.34375,89.64805603027344,9.309344821506077,0

2.34375,90.06805419921875,7.86041986374628,1

2.34375,90.77301025390625,7.66626440960428,0

1.953125,90.70358276367188,8.528609502883185,0

2.34375,91.50962829589844,8.940238952636719,0

2.34375,87.98904418945312,7.33770287555197,3

1.953125,86.3433837890625,8.022798810686384,4

Table 1
Performance analysis of proposed classification model to the traditional models on No emphysema(0) and Minimal emphysema (1)

Algorithm	*TRUE*	*FALSE*	*F-measure*	*Accuracy*
Adaboost	0.612	0.365	0.618	0.632
SVM	0.577	0.418	0.599	0.651
FSVM	0.643	0.397	0.673	0.671
Naïve	0.641	0.362	0.661	0.701
DT + LBP	0.678	0.307	0.71	0.734
EnsemblModel	0.856	0.206	0.853	0.861
ProposedModel	0.8965	0.193	0.8854	0.9063

Table 1,describes the statistical comparison of proposed model with the traditional classification models on COPD data(No emphysema(0) and Minimal emphysema (1)). From the table ,it is observed that the true positive ,false postitive ,F-measure and accuracy improved on an average of 5% using the proposed ensemble classification model.

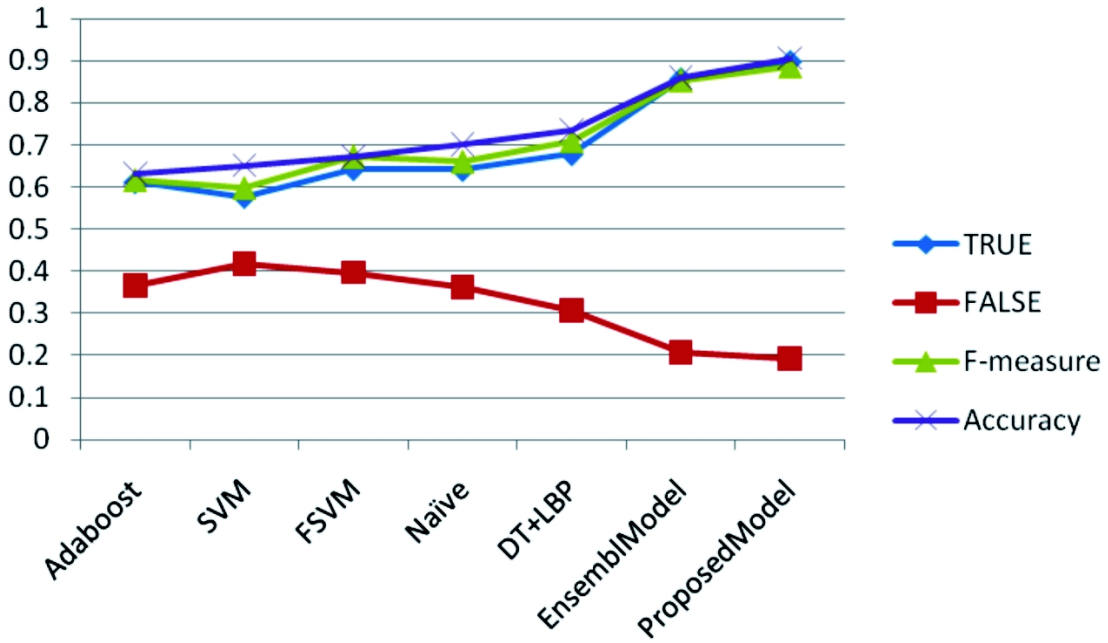

Figure 1: Performance analysis of proposed classification model to the traditional models on No emphysema(0) and Minimal emphysema (1)

Figure 1, describes the statistical comparison of proposed model with the traditional classification models on COPD data(No emphysema(0) and Minimal emphysema (1)). From the figure ,it is observed that the true positive ,false postitive ,F-measure and accuracy improve on an average of 15-20% when preprocessed with the proposed feature selection model and ensemble classification model.

Table 2
Performance analysis of proposed model to the traditional models on No emphysema(0) and Mild emphysema (2)

Algorithm	*True Positive*	*False Positive*	*F-measure*	*Accuracy*
Adaboost	0.631	0.351	0.642	0.672
SVM	0.519	0.398	0.642	0.548
FSVM	0.629	0.377	0.682	0.636
NaïveBayes	0.692	0.332	0.639	0.728
DT + LBP	0.658	0.347	0.734	0.726
Ensemble Model	0.931	0.156	0.926	0.936
Proposed Model	0.9654	0.106	0.964	0.959

Table 2, describes the statistical comparison of proposed model with the traditional classification models on COPD data(No emphysema(0) and Mild emphysema (2)). From the table ,it is observed that the true positive ,false postitive ,F-measure and accuracy improve on an average of 5-10% when preprocessed with the proposed feature selection model and ensemble classification model.

5. CONCLUSION

In this paper, a novel ensemble classification model for disease prediction has been implemented. The disease level of COPD has been successfully automated using the ensemble model. We have tested our model on COPD disease data to predict the level of disease. Some of widely accepted image classification models such as SVM, FSVM, Random Forest,Naïve Bayes, KNN, GA,etc are used to perform the experimental study on COPD data. Experimental results proved that proposed model has high computational accuracy (15%-20%) of improvement in terms of false positive rate, f-measure compared to traditional ensemble models.

REFERENCES

[1] M. Anthimopoulos, S. Christodoulidis, A. Christe and S. Mougiakakou, “Classification of Interstitial Lung Disease Patterns Using Local DCT Features and Random Forest”, “Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE”, pp.6040-6043, 2014.

[2] M. Anthimopoulos, S. Christodoulidis, L. Ebner,A. Christe and S. Mougiakakou, “Lung Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural Network”, “IEEE Transactions on Medical Imaging”, pp.1-10, 2015.

[3] C. Tianrui, X. Gang, W. Fang and Y. Chengdong, “Texture Features Extraction of Chest HRCT Image”, “International Conference on. IEEE”, pp.1-4, 2010.

[4] J. K. Dash, S. Mukhopadhyay, R. Dasgupta, M. K. Garg, N. Prabhakar and N. Khandelwal, “Wavelet-Based Rotation Invariant Texture Feature for Lung Tissue Classification and Retrieval”, “SPIE Medical Imaging. International Society for Optics and Photonics”, 2014.

[5] J. K. Dash, V. Madhavi, S. Mukhopadhyay, N. Khandelwal and P. Kumar, “Segmentation of Interstitial Lung Disease Patterns in HRCT

[6] Images”, “SPIE Medical Imaging. International Society for Optics and Photonics”, 2015.

[7] J. D. Deng, “Improving feature extraction for automatic medical image categorization”, “24th International Conference Image and Vision Computing New Zealand (IVCNZ 2009)”, pp.379-384, 2009.

[8] M. Kakar, A. Mencattini and M. Salmeri, “Extracting Fuzzy Classification Rules from Texture Segmented HRCT Lung Images”, “Journal of digital imaging 26.2 (2013)”, pp.227-238, 2013.

[9] N. Kato, M. Fukui and T. Isozaki, “Bag-of-features approach for improvement of lung tissue classification in diffuse lung disease”, “SPIE Medical Imaging. International Society for Optics and Photonics “,2009.

[10] B. Kaur and S. Jindal, "An implementation of Feature Extraction over medical images on OPEN CV Environment", "Devices, Circuits and Communications (ICDCCom), 2014 International Conference on. IEEE ", 2014.

[11] Y. Lee, N. Kim, J. B. Seo, S. O. Park, Y. K. Lee and S. H. Kan, "Improvement of Computational Efficiency Using a Cascade Classification Scheme for the Classification of Diffuse Infiltrative Lung Disease on HRCT", 2012.

[12] B. Li, W. Li and D. Zhao, "Multi-Scale Feature Based Medical Image Classification", "3rd International Conference on Computer Science and Network Technology", pp. 1182-1186, 2013.

[13] T. Nuzhnaya, V. Megalooikonomou, H. Ling, M. Kohn and R. Steiner, "Classification of Texture Patterns in CT Lung Imaging", "International Society for Optics and Photonics", 2011.