

SELECTION TECHNIQUES FOR EFFICIENT CLASSIFICATION AND ANALYSIS OF BIOMEDICAL IMAGES

Musibau Adekunle Ibrahim¹ and Oladotun Ojo²

¹Department of Information and Communication Technology, Osun State University, Osogbo, Nigeria

²Department of Physics, Osun State University, Osogbo, Nigeria

Abstract

Selection techniques to look for the relevant features that might have the best influence on the predictive accuracy are modified and applied during this paper. The selection search iteratively evaluates a subset of the feature, then modifies the subset and evaluates if the new subset is an improvement over the previous. The performances of the developed models are tested with some classifiers based on the feature variables selected by the proposed approach and the effects of some important parameters on the overall classification accuracy are analyzed. Our findings revealed that the proposed approach consistently improved the classification accuracy. The approach developed in this paper significantly improved the classification accuracy and compared favourably with the recently published work. The use of the feature selection search tool reduces the classification model complexity and produces a strong system with greater efficiency, and excellent results. The research results also prove that the amount of growing trees and the threshold values could affect the classification accuracy.

Keywords: Feature Selection, Multi-Fractal Descriptor, Classification Accuracy, Naïve Bayes, Bagged Decision Tree, Emphysema Patterns

1. Introduction

In machine learning, most classifier algorithms are presented with a set of training instances, where each instance, can be described as a feature vector or attribute values and a class label. For instance, in object recognition, the features might include the size, shape and height of the object, and the class label may be determined by different categories of this object. The first task is the selection of the appropriate classification algorithm that could be useful in classifying the feature sets. The classifier maps the space of feature values to the set of class values to formulate a predictive model [1-2]. The problem of the feature subset selection (FSS) in image classification of computed tomography (CT) images can be very challenging as there is need to select some important subset features upon which the algorithm can focus on. Selection of bad subsets features might eventually affect the performance accuracy of the classification system. The online emphysema database used for the experiments in this paper can be found in [3]. In this database, there are three different emphysema image classes: normal tissue (NT), centrilobular (CLE), and paraseptal (PSE). Recent paper by [4] also proposed multi-fractal based approach for the analysis and classification of emphysema images by extracting the self-similarity features of the images. In this technique, the Holder exponent for the power law approximation of intensity measures in pixel neighbourhoods has been used for the computation of multi-fractal spectrum for the classification of images [5-6]. Detailed analysis of the emphysema classification using the multi-fractal techniques can be found in [7-8]. There are four different multi-fractal intensity measures: the summation, maximum, Iso and inverse minimum, which can be used for the computation of the Holder exponent but only the summation measure is considered for the experimental analysis in this paper. The datasets used for the experiments in this paper are obtained from the multi-fractal and the alpha-histograms descriptors using the summation intensity measures for the Holder exponent computation. Further information on the computational analysis of the multi-fractal and alpha-histogram can be found in [4, 7-10]. In this paper, the multi-fractal data sets are generated by dividing the Holder exponents of each pixel $[\alpha_{\min}, \alpha_{\max}]$ of the emphysema image into 100 intervals and the corresponding fractal dimension of each pixel $f(\alpha_i)$ is calculated with the α values within the range $[\alpha_i, \alpha_{i+1}]$. $\{f(\alpha_i) / i = 0, 1, \dots, 99\}$ values for each emphysema class are used as the feature vectors for the multi-fractal data sets using the summation measure.

Correspondence Author: Musibau A.I., Email: kunle_ibrahim2001@yahoo.com, Tel: +2348134249446

Transactions of the Nigerian Association of Mathematical Physics Volume 12, (July – Sept., 2020), 117 –126

Similarly, the corresponding pixels count P_i with the alpha values within the range $[\alpha_i, \alpha_{i+1}]$ give the alpha-histogram, where $\{P_i / i = 0, 1, \dots, 99\}$ are directly used as the feature vectors for the alpha-histograms datasets. However, each data set consists of 100 feature values, but after removing the noisy outliers, only 50 feature values are used for the experiments. 24 images were randomly selected from each emphysema class for the training of the classification system while 6 images were used for testing the accuracy of the system. Each data set generated from the images consists of 90 observations and 50 predictor variables, thus the dimensionality of the data is 90×50 .

The presence of too many feature variables may sometimes reduce the accuracy of the classification system as some features may be redundant and non-informative. In addition, processing a high dimensional data requires large memory and may reduce the computational speed. This paper proposes to apply the feature selection (FS) technique to improve the accuracy of the classification system in CT emphysema images. The main research question that would be taken into account is "How does the FS approach affect the performances of the descriptors?"

2. Related Works

Different classification algorithms and techniques have been proposed and tested using various feature subsets. Some require extensive training while some need very little [5]. In the case of noisy data, different classifiers often provide different generalizations by using different decision boundaries. And different feature sets provide different representations of the input patterns containing different classified information of the input patterns [11-12]. Selecting the best classifier or the best feature sets is therefore very important as this may improve the performance of the classification models. This can be achieved by selecting a minimal set of features that has same or better predictive power as the original model.

FS algorithms can be broadly divided into two categories: The filter and Wrapper based approaches. A good example of the filter approach is the Relief and Focus algorithms, the Relief algorithm ranks each feature in the data set by assigning weights while the Focus is always searching for the minimal set of features that may be useful in classification [1-2, 13]. Correlation FS as discussed in [1] can be used to evaluate the predictive power of each feature and the degree of redundancy between them by selecting those subsets of features with low level of inter-correlation.

The Naive Bayes (NB) determines the class of a particular vector in the data by calculating its posterior probability. The posterior probability can be calculated using the Bayes theorem [14-15]. For instance, the probability of class c given feature vector V can be mathematically represented as $P(c|V)$, if V is a feature vector: $f_1, f_2, f_3, \dots, f_n|f$, represented by the set of classes $C = c_1, c_2, \dots, c_n|c$. The posterior probability of the likelihood and the prior probabilities according to Bayes theorem are given as $P(c_i | V) = (P(V/c_i) \times P(c_i)) / P(V)$, where $P(V/c_i)$ is the likelihood and is the probability of the occurrence of vector V given class c_i . $P(c_i)$ is called the prior probability and is the probability of class c_i . The likelihood $P(V/c_i) = p(V_1/C) \times p(V_2/C) \times \dots \times p(V_n/C)$, after calculating the probabilities for each class, the classifier would select the class with the highest probability [14-15]. Previous studies have shown that the performance of the classifier algorithms could be improved by using the FS [9].

Diagnoses and treatment of most medical images can be very difficult due to the ever increasing volume of clinical cases processed by the radiologist. The manual collections of data are therefore subject to high error rates, imprecision and uncertainty [6, 18]. In the experiment conducted by [19], the authors discovered NB to be superior to other classifiers even on data sets with many feature dependencies. Many studies in the literature have successfully applied NB for solving different medical problems [15, 20-21]; the results showed that the NB outperformed other algorithms. NB had also shown excellent results even with small data sets as proved by [22-23] and it can also be used to circumvent the problems of over fitting during classification.

The principal component analysis (PCA) is one of the methods that can be used for selecting features and reducing dimensionality [2, 6, 19]. In PCA, the data set is represented by a small number of uncorrelated features while retaining most of its information content. This can be done by removing correlated components that contribute only a small amount to the total variance in the data set. The methods of selecting relevant features using certain discriminating criteria could sometimes be achieved by the correlation coefficients like statistical tests such as t-test, F-test etc. [18]. In [24], the author proposed a minimum redundancy-maximum relevance criteria for FS and the results significantly improved the predictive power and generalization properties of the feature sets [25].

One of the approaches to dimensionality reductions is to transform a high dimensional feature space into a lower dimensional space [26-27] as this reduces the model complexities. In [28], the linear discriminant analysis (LDA) is combined with the PCA to obtain high discriminative patterns from a high dimensional feature space derived from the descriptors. The authors later applied Laplacian PCA (LPCA) to maximize the weighted local scatter instead of the global scatter of data as in the original PCA [28]. The results achieved significantly improved the classification accuracy. Support vector machines (SVMs) have demonstrated highly competitive performance in many real-world applications, such as bioinformatics, face recognition and image processing. In [2], SVM outperformed most of the previously proposed methods

in the diagnosis of cancer microarray data. The authors in [19] designed a biased maximum margin analysis and semi-supervised biased maximum margin analysis combined with the SVM to improve the performance of the traditional SVM as a relevant feedback for content based image retrieval (CBIR). In [14, 29], a novel algorithm for subspace learning technique was developed using SVM to exploit the user historical feedback log data for a CBIR.

The classifier algorithm constructed by the bagged decision tree (BT) can be used to create a classification model in a form of decision trees for predicting the class labels of the unknown data represented in a form of a matrix. This matrix contains predictors or feature vectors that could be used in constructing an ensemble of classification trees that can be used for predictions. Training of classification system involves random selection of features to improve the discriminative power of features by reducing the correlation between the trees in a classifier [15, 30]. Bootstrap aggregation as introduced by [6], is specifically based on constructing several training sets from the original set by resampling with replacement (bootstrapping) [25].

Based on the information provided by the literature, the most appropriate FS algorithm for improving the classification system in this paper can be classified under the filter approach. Among all the classifiers, NB and BT have been selected since they both have better characteristic features than other classifiers in terms of searching for the useful features and high level of dimensionality reduction in a data set.

Additionally, in this research, in order to select the most highly discriminating features from the feature vectors and the multi-fractal features, this paper proposes to calculate the column area under the curve (AUC) for all the features extracted and find the maximum mean AUC values for the best four columns in each data set. The four columns with the highest average AUC values would be selected as a four dimensional feature vector space for each data set during the classification process.

3. Materials and Methods

There are various types of features that can be extracted from the CT emphysema images for the multi-fractal analysis of emphysema patterns. In order to obtain feature descriptors with a very high discriminating power, this section considers the combinations of some of the important histogram features and the multi-fractal spectrum features for efficient classification of the emphysema patterns. The first histogram features are derived from the intensity histogram. An intensity histogram is a diagram in the form of a graph, plotting the number of pixels (fractional area) with a specific gray level versus the gray level value. It can be used for adjusting the brightness and contrast levels of the image. The shape of the histogram broadly describes the intensity distributions in the image. In some cases, a histogram may be scaled for adjusting the intensity levels or the contrast (e.g. histogram equalization). An example illustrating the intensity histogram of an emphysema CT image is presented in Figure 1.

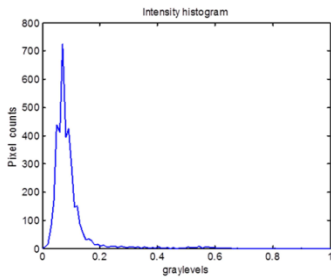


Figure 1. Intensity histogram of emphysema image

The useful features that can be derived from the intensity histogram are the minimum and maximum values on the x-axis, and the maximum of the histogram on the y-axis.

3.1. Histograms of Emphysema Image

An α image is a matrix of the same dimension $M \times N$ as the original image but filled with α -values, with one-one corresponding to image pixels. Further information on the computation of alpha image can be found in the previous work [4, 7-10, 32-33]. An alpha-histogram of an image is therefore constructed using the $\alpha(m, n)$ values of the image as the pixel intensity values. As an example, an alpha-histogram of an emphysema image class is presented in Figure 2. Alpha-histogram can also be used as a global descriptor of intensity values, just like the intensity histogram

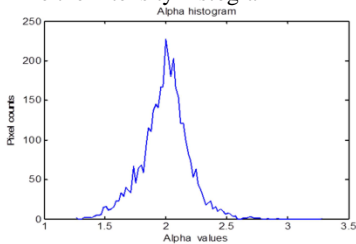


Figure 2. Illustrating an example of alpha-histogram of emphysema image

The features derived from the alpha-histogram would also provide some local characteristics of the images. However, multi-fractal spectrum provides several shapes that could give more useful information to describe the characteristic properties of the images than the alpha-histogram.

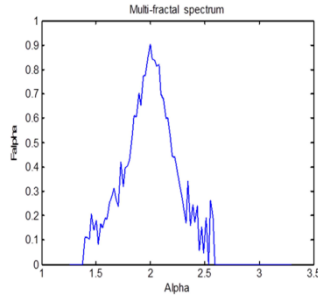


Figure 3. Multi-fractal spectrum of an emphysema image sample

Multi-fractal spectrum contains additional global information derived from the statistical self-similarity properties of the image at various scales to provide a global descriptor of the images. It generally has a higher discriminating power compared to intensity and the alpha-histograms. Furthermore, it can also be observed from the analyses that the combination of the features extracted from the histograms and the multi-fractal spectrum could generate a descriptor with a better discriminating power. The multi-fractal spectrum of the same emphysema image used in Figure 2 is presented in Figure 3.

There is the possibility of improving the classification accuracy of the emphysema images by cascading the results obtained from the alpha-histogram with the multi-fractal spectrum since the new descriptor will definitely provide solutions to some of the limitations of the alpha-histogram and the multi-fractal spectrum. The newly constructed descriptor would combine the characteristic features of the multi-fractal spectra and that of the histograms, which makes it more superior and discriminating for efficient and accurate classification. There are some features that are very useful in a multi-fractal spectrum but are lacking in the histograms and vice versa.

3.2. Feature Selection Approach

The system overview for the classification approach involving several stages is presented in Figure 4. The features extracted from the descriptors are used as the input features for the FS algorithm. The two popular classifiers used in this study are trained on the outputs obtained from the FS as presented in Figure 4.

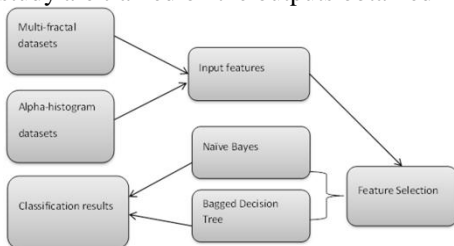


Figure 4. System overview of classification approach in emphysema images

FS is an important part of pre-processing data in machine learning as the selection of important features can make the training phase to be less time consuming. This can be done by reducing the dimensions of data and thereby making the classifier algorithms to operate faster [31]. FSS is a mapping from a m-dimensional feature space (input space) to n-dimensional feature space (output), which can be represented as follows:

$$Fss: RrXn \rightarrow RrXn, \tag{1}$$

where $m > n$, $RrXm$ is the matrix of the original data set with r instances or observations, $RrXn$ is the reduced feature set containing r observations in the subset selection. It is also a technique of selecting only the predictor variables, that provide the best predictive power by simplifying and improving the model interpretation.

Sequential forward selection (SFS) is used in this study to search for the relevant features that could yield optimum classification accuracy and increase the computational efficiency. SFS is one of the methods of selecting a subset of features in a data set by sequentially selecting the features until there is no further improvement in prediction accuracy. The important stages of the SFS algorithm are shown in Figure 5.

Evaluation of the subsets requires a scoring metric that grades a subset of features. In this study, a function handle is used to define a criterion to determine the relevant features to be selected. Dimensionality reduction is achieved by calculating an optimal subset of predictive features of the original data. The algorithm automatically stops when further selection of feature subset has no effect on the classification errors.

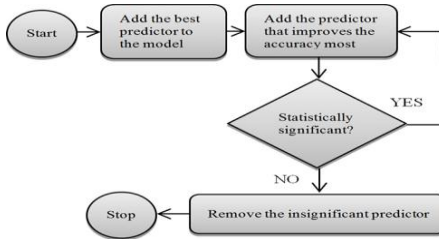


Figure 5. Sequential forward selection method

4. Results and Discussion

This section provides an outline of the experimental results obtained using images from the emphysema database [3], based on the implementation of the methods previously discussed. The feature vectors extracted from the multi-fractal spectra and the alpha-histograms have been used for classification and retrieving purposes. The histogram descriptors used for the classification experiment are constructed by dividing the range of α -values generated from the Holder exponent into 100 intervals. The alpha-histogram has been calculated for each alpha bin as the number of pixel counts with the α values within the α -range $[\alpha_i, \alpha_{i+1}]$. The average of the alpha-histogram for four randomly selected images has been calculated and used as the feature vectors. In the classification process of the NB classifier, the holdout partition method has been applied to divide the observations into training sets and test sets.

There is a scalar specifying the proportion of the number of observations to be randomly selected for validation. In order to achieve promising results since the accuracy of the classifiers depends on the training data; this scalar automatically selects 70 percent of the feature vectors for the training and 30 percent for testing. The performance of the classifiers is evaluated in the form of confusion matrix. A confusion matrix is a way of evaluating the performance of the classification model. Each entry in a matrix denotes the number of predictions made by the model where it classified the classes correctly or incorrectly. For instance, in a binary classification problem with two classes to classify, the metrics of the confusion matrix can be defined as follows:

True Positive (TP): This refers to the number of predictions where the classifier correctly predicts the positive class as positive.

True Negative (TN): This describes the number of predictions where the classifier correctly predicts the negative class as negative.

False Positive (FP): This defines the number of predictions where the classifier incorrectly predicts the negative class as positive.

False Negative (FN): It refers to the number of predictions where the classifier incorrectly predicts the positive class as negative.

Confusion matrix is a very efficient evaluation technique to measure performances in machine learning model. The overall accuracy for the samples that were correctly classified by the classifier can be achieved by the following mathematical equation:

$$(TP + TN)/(TP + TN + FP + FN) \tag{2}$$

While misclassification rate that determines those predictions incorrectly classified can be calculated using the following equation:

$$(FP + FN)/(TP + TN + FP + FN) \tag{3}$$

However, for a multi-class classification problem such as a 3-class classification problem as in this research work, precision, recall and F1-score measures can be calculated and combined to determine the performance of the model.

This matrix can therefore be represented as a matrix $M \in R^{k \times k}$, a square matrix whose diagonal elements represents the actual classification accuracy where k is the number of classes in the data set. The classification error of the classifiers can be calculated as follows:

$$Error = 1 - \frac{trace(Confusion\ Matrix)}{Sum(Confusion\ Matrix)} \tag{4}$$

where trace (.) is the sum of all the elements in the diagonal, and sum (.) is the sum of all the entries in the confusion matrix. The feature vectors from the data sets are also trained with the BT classifier and the performances of the classification algorithms are examined with different experimental settings. A dimensionality reduction step has been implemented due to the presence of large irrelevant features in the data sets. The classification results obtained before and after the feature reduction for the NB classifier are presented in Table 1 in the form of confusion matrices.

The feature techniques reduced the feature variables in each data set to a set of feature variables with the highest discriminating power, such that the classifiers can be trained with the newly selected features by the SFS method. The performance of the classification system is measured by the classification accuracy generated by the confusion matrices.

The results obtained by the SFS achieve better dimensionality reductions and increase in learning accuracy by simplifying the model complexity. In Table 1, the classification accuracy of the multi-fractal datasets using the NB algorithm, increases from 75.3% before the FS to 78.6% after the FS while in alpha-histogram, the classification accuracy increases from 75.6% before applying FS to 81.3% after applying FS. The classification results obtained by the BT algorithm are shown in Table 2. Additionally, the BT classifier outperformed the NB in multi-fractal datasets before and after the FS techniques, but the overall classification remains the same after the FS in the alpha-histogram data set.

Table 1. Naïve Bayes (NB) classification results with and without feature reductions

	Predicted											
	Multifractal datasets						Alpha_histogram datasets					
	NB without SFS			NB with SFS			NB without SFS			NB with SFS		
	NT	CLE	PSE	NT	CLE	PSE	NT	CLE	PSE	NT	CLE	PSE
NT	73	18	9	73	18	9	91	0	9	91	9	0
CLE	17	83	0	17	75	8	25	58	17	8	75	17
PSE	8	25	67	8	9	83	17	8	75	25	0	75

The reason for the improvement in the performance of the algorithms after the FS in the data sets can be traced to the reduction in the complexity of the models, as complex models sometimes over fit the data and generate additional errors. Simplifying the complex models that would include the feature variables that are uncorrelated with one another would always reduce the computational complexity, which might increase the accuracy. The difference in classification accuracy between the NB classifier and the BT over the multi-fractal datasets is 6.2% after the FS, while in alpha-histogram, the overall classification retains the same level even after the feature selection but a little bit higher before applying selection technique (Tables 1 and 2).

Table 2. Bagged decision tree classification results with and without feature reduction

	Predicted											
	Multifractal datasets						Alpha_histogram datasets					
	NB without SFS			NB with SFS			NB without SFS			NB with SFS		
	NT	CLE	PSE	N T	CLE	PSE	NT	CLE	PS E	NT	CLE	PSE
NT	82	9	9	91	0	9	81	19	0	91	9	0
CLE	8	75	17	25	67	8	0	75	25	0	75	25
PSE	0	17	83	8	0	92	0	25	75	17	8	75

The second approach is the concatenation of the relevant features obtained from the multi-fractal datasets and the alpha-histograms descriptors to generate new feature vectors. This is very easy since they both have the same number of rows and columns in dimension. Only the features selected by the FS technique have been used and the experimental results are presented in Table 3. The results in Table 3 reveal that this approach outperformed the results obtained in Tables 1 and 2. Significant improvements over the combined feature sets for the two classifier algorithms can be achieved after the FS. The difference in classification accuracy between the combined feature sets and the alpha-histogram datasets (Table 3) for the NB classifier is about 3%, while the accuracy over the multi-fractal for the BT improves by 6.1% after the selection technique.

Also in Table 3, the classification results of the combined features using the NB algorithm completely classified the normal emphysema images from the other pathological cases (CLE and PSE). The reasons for this improvement is due to the combination of the important feature variables with high discriminative power from both datasets since the irrelevant features have been filtered out by the FS methods. However, this approach consumes more processing time as the size of the dataset increases and thus reduces the computational speed.

Table 3. Classification results with and without feature selections for combined features

	Predicted											
	Multifractal datasets						Alpha_histogram datasets					
	NB without SFS			NB with SFS			NB without SFS			NB with SFS		
	NT	CLE	PSE	NT	CLE	PS E	NT	CLE	PSE	NT	CLE	PSE
NT	100	0	0	91	9	0	91	0	9	82	9	9
CLE	17	66	17	0	83	17	0	92	8	8	92	0
PSE	17	8	75	17	8	75	8	17	75	0	8	92

Furthermore, the pairwise t-test of the classification results before and after the feature selection were carried out in order to determine whether the differences in accuracy are statistically significant or not. The t-test results of the classification results of the combined features in Table 3 are shown in equation (5); $h = 0$ indicates a failure to reject the null hypothesis.

$$h = 0, p\text{-value} = 0.7165, ci = -27.0733, 20.4067 \tag{5}$$

This means the improvement achieved in the classification results is not statistically significant even at the 5% significant level. The same procedures are repeated for the previous results in Tables 1 and 2 before and after the feature selection in the multi-fractal and alpha-histogram data sets; the statistical results showed that the increase in the classification accuracy is not statistically significant since it failed to reject the null hypothesis at the 5% significant level. Another evidence to prove the statistical results is that the probability of observing a value of the test statistic, as indicated by the p values, is far greater than the α -value of 0.05. Additionally, the 95% confidence interval on the mean of the difference does contain zero in all the results as can be seen in equation (3). These evidences are enough reasons to conclude that none of the classification results presented in this section is statistically significant at the $\alpha = 0.05$ significant level.

An important parameter that can be used to evaluate the performance of the BT algorithm is the number of grown trees used in constructing the classification model. For the multi-fractal datasets, it can be observed in Figure 6a that, the BT performed well when the number of grown trees ranges from 37 to 49. In addition, the classification errors are at the minimum level and constantly stable (Figure 6a). The performances of the algorithm are not stable when the grown trees are less than 37 and hence generating more classification errors. However, for the alpha-histogram dataset (Figure 6b), the errors seem to be constant when the number of grown trees ranges from 20 to 40, and the performance of the algorithm is consistent. In order to fully optimize the performance of the classifier, the number of grown trees used for both datasets is therefore 40. Ensemble of decision trees, particularly the BT has a way of estimating the predictor importance. Measure of importance for each predictor variable can be achieved by evaluating the effect on the classification margin if the values of the variable are permuted across the out-of-bag observations.

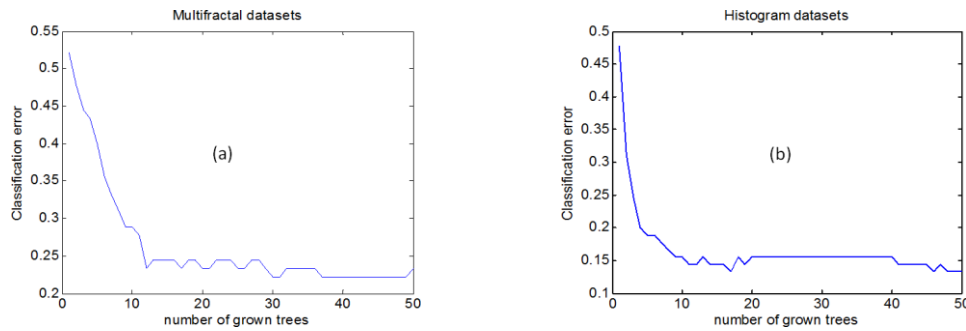


Figure 6. Illustrating the variations between the classification error and the grown trees for bagged decision. (a) Multi-fractal datasets (b) Alpha-histogram datasets

In other words, permuting a particular feature variable may either increase or decrease the classification accuracy. Figure 7 presents the results of the feature importance variables for the BT in each dataset. In the experiments, a threshold has been set to filter out those features whose ranking values are less than the required value; the features with the ranking values above this threshold value have been used for the classification process (Figure 7). In this case, the threshold has been reset to 0.42 in order to remove the unwanted features that could reduce the classification accuracy. The results after removing the features below the threshold level as in Figure 7 reflect that the chosen features have greater predictive power, which has led to further improvement in all features.

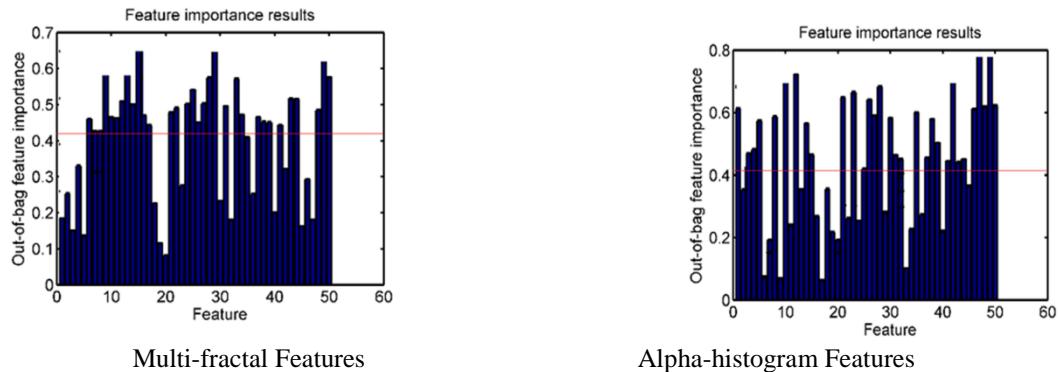


Figure 7. Showing the ranking values of each feature variable in the datasets for multi-fractal and alpha-histogram features. This improvement in the classification accuracy indicates that many features in the datasets are highly correlated and many are not strongly relevant. The FS ignored this set of data and only trained with the important features that would have significant impact on the overall accuracy.

The performances of different threshold levels are tested on the overall classification accuracies. It was verified experimentally that the larger the threshold value the more important the selected predictor variables (Figure 8). For instance, the threshold value of 0.42 over the combined feature sets gives the highest classification accuracy while the threshold value below this level reduces the classification accuracy (Figure8).

The reason is because, the smaller threshold values would allow more predictor variables that are not strongly relevant to be added to the selected variables trained with the classifier algorithms. This will eventually translate to a more complex model that could generate more classification errors and thus reduces the classification accuracy. However, further increase in the threshold values beyond 0.42 does not have any significant improvement in the accuracy. The plot of the test accuracy demonstrating the effect of the threshold level on the overall classification accuracy is presented in Figure8.

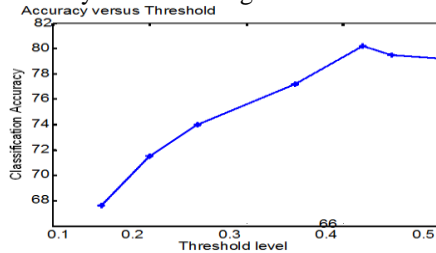


Figure 8. A graph of classification accuracy versus threshold level

5. Conclusion

This paper has presented a novel approach for improving the classification accuracy of emphysema images by employing the FS techniques. The FS approach has been implemented to remove irrelevant features in the emphysema CT images. The two machine learning algorithms considered in this study; the NB and BT performed well on the datasets used. The results achieved by the classifiers are compared, the performance of the BT has been slightly better than the NB algorithms.

The experimental results also confirmed that multi-fractal descriptors could be used for the analysis and classification of emphysema in CT images. The information from the alpha-histogram descriptors has been very useful as the combination of the relevant features in the form of hybrid for both descriptors improved the classification accuracy. During the implementation of the BT, some of the important parameters that could be used to evaluate the performance of the classification system are presented.

The experimental results proved that the number of growing trees and the threshold values could affect the classification accuracy. Overall, the performance of the classifiers after FS has been consistently higher than the results without FS. Further research work might be to cascade the two classifier algorithms together over the combined feature sets or other medical data sets. This cascaded technique can be used to construct a new descriptor with a very powerful feature to improve the existing results. The performance of the classifiers can also be improved by parallelizing the algorithms using the GPU parallel computing as this might improve the computational efficiency. In the future, other classification approach such as the local binary patterns (LBP) could also be implemented for further analysis of the emphysema images, and the results will be evaluated against the multi-fractal analysis.

Acknowledgements

The authors would like to sincerely thank Professor Mukundan, Department of Computer Science and Software Engineering at the University of Canterbury for his substantial contributions towards this study.

Reference

- [1] Han, Y., Park, K., & Lee, Y.K. (2011). Confident wrapper-type semi-supervised feature selection using an ensemble classifier. *2nd International Conference on Artificial Intelligence, Management Science and Electronic Commerce, AIMSEC – Proceedings.*; pp. 4581–4586, 2011. doi:10.1109/AIMSEC.2011.6010202.
- [2] Wang D., Nie F., &Huang H. (2015). Feature Selection via Global Redundancy Minimization, 4347 (c), pp. 1–14. doi:10.1109/TKDE.2015.2426703.
- [3] Sørensen, L., Shaker S.B., and De-Brujine, M. (2010). Quantitative Analysis of Pulmonary Emphysema using Local Binary Patterns. *IEEE transactions on medical imaging.* 29(2),pp.559–569.
- [4] Ibrahim, M. and Mukundan, R. (2014). Multi-fractal Techniques for Emphysema Classification in Lung Tissue Images. *Proceeding of International Conference on Environment, Chemistry and Biology*; pp. 115-119.
- [5] Duda, R., Hart P., and Stork D.,(2011). *Pattern classification and scene analysis*. New York.

- [6] Papadopoulos, A., Plissiti, M. and Fotiadis, D.(2005).Medical-image processing and analysis for CAD systems. *Journal of Applied Signal Processing*. 2(5), pp. 25–41.
- [7] Ibrahim, M. &Mukundan, R.(2015).Cascaded techniques for improving emphysema classification in computed tomography images. *Artificial Intelligence Research*; 4 (2), pp. 112–118.doi:10.5430/air.v4n2p112.
- [8] Ibrahim M.A., Ojo O.A., Oluwafisoye P.A.(2017). Analysis of Emphysema Patterns Computed Tomography Image, *International Journal of Scientific & Engineering Research*, 8(12), 2157-2168.
- [9] Ibrahim M.A., Ojo O.A, P.A. Oluwafisoye P.A. (2018). Identification of Emphysema Patterns in High Resolution Computed Tomography Images. *Journal of Biomedical Engineering and Informatics* (Canada, 2017), 4(1),16-24, 2018.
- [10] Ibrahim M.A., Ojo O.A.(2018) “New Multi-fractal Techniques for Pattern Analysis in Biomedical Images”, *SIA OmniScriptum Publishing, Latvia, European Union;LAMBERT Academic Publishing; ISBN:978613989296-9*.
- [11] Xu, J.W. &Suzuki, K.(2014). Max-AUC feature selection in computer-aided detection of polyps in CT colonography. *IEEE Journal of Biomedical and Health Informatics*. 18(2),pp. 585–593,doi:10.1109/JBHI.2013.2278023.
- [12] Zhang, H.(2009). A Novel Bayes Model: Hidden Naive Bayes. *IEEE Transactions on Knowledge and Data Engineering*, 21 (10), pp. 1361–1371.
- [13] Liu,Q., Shi, S., Zhu, H. &Xiao, J.(2014). A Mutual Information-Based Hybrid Feature Selection Method for Software Cost Estimation Using Feature Clustering. *IEEE 38th Annual Computer Software and Applications Conference*. pp. 27–32, 2014; doi:10.1109/COMPSAC.99.
- [14] Mangai, J.A., Nayak, J. and Kumar, V.S.(2013).A Novel Approach for Classifying Medical Images Using Data Mining Techniques. *International Journal of Computer Science and Electronics Engineering (IJCSSEE)*; 1,2.
- [15] Bhuvaneswari, R. &Kalaiselvi, K.(2012). Naive Bayesian Classification Approach in Healthcare Applications. *International Journal of Computer Science*. 3 (1), pp. 106–112.
- [16] Zhang, H., and Su, H.(2008). Naive Bayes for optimal ranking. *Journal of Experimental & Theoretical Artificial Intelligence*. 20(2), pp. 79–93.
- [17] Wang, L.M.,Li,X.L., Cao, C.H.andYuan, S.M. (2006).Combining decision tree and Naive Bayes for classification. *Knowledge-Based Systems*, 19(7), pp. 511–515.
- [18] Kohavi, R. and John, H.(2011). Artificial Intelligence Wrappers for feature subset selection. *Data Mining and Visualization Silicon Graphics*. 97, 273–324.
- [19] Domingos, P. (1997). On the Optimality of the Simple Bayesian Classifier under Zero-One Loss. 29, pp. 103–130.
- [20] Bluma, A.L. and Langley, P.(1997).Artificial Intelligence Selection of relevant features and examples in machine. pp. 245–271.
- [21] Liu, Y., Guo, J. &Lee, J. (2011). Halftone Image Classification Using LMS Algorithm and Naive Bayes. *IEEE Transactions on Image Processing*. 20 (10), pp. 2837–2847, doi:10.1109/TIP.2011.2136354.
- [22] Demšar, J., Leban, G. and Zupan, B.(2005). Freeviz-an intelligent visualization approach for class-labeled multidimensional data sets. *Proceedings of IDAMAP*. 1, pp. 13–18.
- [23] Supriyanto, C., Yusof N. &Nurhadiono, B.(2013).Two-Level Feature S election for Naive B ayes with Kernel Density, *Estimation in Question Classification based on Bloom's Cognitive Levels*. pp. 1–5.
- [24] Ladha,L. T. D.(2011). Feature Selection Methods and Algorithms. *International Journal of Computer Science and Engineering*. 3 (5), pp. 1787–1797.
- [25] Ruiz, V. and Nasuto, S.(2005). Biomedical-image classification methods and techniques. *Costaridou, L. (ed.)*.
- [26] Zhang, H.A. (2009). A Novel Bayes Model: Hidden Naive Bayes. *IEEE Transactions on Knowledge and Data Engineering*. 21(10), pp. 1361–1371.
- [27] Mikolajczyk, K. &Schmid, C. (2005). Performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27 (10), 1615–30; doi:10.1109/TPAMI.2005.188.
- [28] Zha G.&Matti, P.(2007). Local Binary Patterns with an Application to Facial Expressions. *Pattern Analysis and Machine Intelligence, IEEE Transactions*; 29 (6), pp.915–928.
- [29] Marsland, S.(2014).Machine Learning: An Algorithm Perspective. (2nd ed.). USA: CRC Press.
- [30] Kallergi, M.(2017). Evaluation strategies for medical-image analysis and processing methodologies. *International Journal of Image Analysis and Processing Methodologies*; 2(5), pp.75–91.

- [31] Randen, T. and Husoy, J.H. (2005).Filtering for texture classification: a comparative study. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 21(4), pp. 291–310.
- [32] Mukundan, R. and Hemsley, A.(2010).A.Tissue Image Classification Using Multi-Fractal Spectra. *International Journal of Multimedia Data Engineering and Management*. 1(2), pp. 62– 75.
- [33] Hemsley, A. and Mukundan, R. (2009). Multifractal Measures for Tissue Image Classification and Retrieval. *11th IEEE International Symposium on Multimedia*. pp. 618–623.