

Online Feature Selection for Classifying Emphysema in HRCT Images

M Prasad

{ mithunp@cse.unsw.edu.au }

Department of Electrical, Computer and Systems Engineering,
Rensselaer Polytechnic Institute,
110 8th Street, Troy, New York 12180-3590

Abstract— Feature subset selection, applied as a pre-processing step to machine learning, is valuable in dimensionality reduction, eliminating irrelevant data and improving classifier performance. In the classic formulation of the feature selection problem, it is assumed that all the features are available at the beginning. However, in many real world problems, there are scenarios where not all features are present initially and must be integrated as they become available. In such scenarios, online feature selection provides an efficient way to sort through a large space of features. It is in this context that we introduce online feature selection for the classification of emphysema, a smoking related disease that appears as low attenuation regions in High Resolution Computer Tomography (HRCT) images. The technique was successfully evaluated on 61 HRCT scans and compared with different online feature selection approaches, including hill climbing, best first search, grafting, and correlation-based feature selection. The results were also compared against “density mask”, a standard approach used for emphysema detection in medical image analysis.

Keywords— Feature subset selection, HRCT, emphysema.

I. INTRODUCTION

High Resolution Computer Tomography is a valuable imaging modality for assessing diffuse lung diseases and in particular, emphysema. Quantitative image analysis, a useful extension of visual evaluation of the CT scans, is of great assistance for radiologists performing diagnosis. The automated analysis of HRCT scans poses difficult problems, because the radiographic patterns observed are often varied and subtle. HRCT scans have high specificity for diagnosing emphysema and are the most accurate means of emphysema diagnosis in determining its type and extent. Emphysema is a common chronic respiratory disorder characterized by the destruction of lung tissue and is often reflected as areas of low attenuation in CT images [1] as can be seen in Figure 1. Visual evaluation by medical experts usually overestimates the percentage of damaged lung area. Further to this, reproducibility is poor both between examiners and between assessments made by the same examiner on different occasions [2]. Hence, there is a need for an objective and

accurate technique to detect and quantify emphysema that is useful to radiologists.

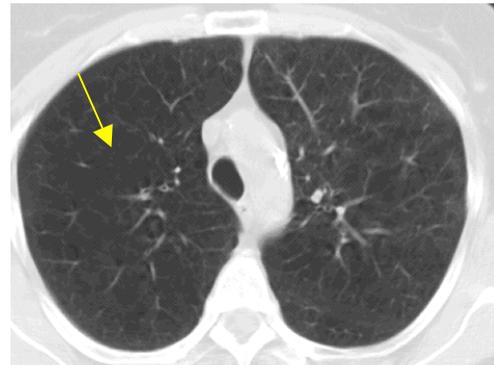


Figure 1: A typical HRCT scan containing emphysema. The regions pointed by arrows denote emphysema.

In many real world problems, feature selection is an essential part of data analysis. When applied as a pre-processing step to machine learning, feature selection is valuable in dimensionality reduction, isolating the most important information and thereby improving classifier performance. In recent years, datasets in many applications have become larger and contain more features. Correlated or non-informative features can behave like noise in the data, thereby degrading classifier performance, as both performance and cost are sensitive to the features used to construct the classifier. As a result, dimensionality reduction through feature selection plays an important role in classification tasks. The standard formulation of feature selection assumes that all candidate features are available from the beginning. However, in many real world problems, there are scenarios where not all features are present initially and must be integrated as they become available.

In the classic formulation of the feature selection problem, a learning system is presented with a training set D consisting of (x, y) pairs, where the x values are represented by fixed-length numeric feature vectors, and the y values are typically numeric scalars. The learner’s task is to select a subset of the elements of x that can be used to derive a mapping function f from x to y that is as “good as possible” according to some criterion C , and sparse with respect to x . If one cannot afford to wait until all features have arrived before learning begins, the problem is to derive an $x \rightarrow y$ mapping at each time step, that is as good

as possible using a subset of just the features seen so far. We call this scenario *online feature selection* (OFS). In OFS, the number of training examples is fixed, but the length of the feature vectors increases over time.

The aim of our research is to perform OFS for classification of emphysema in HRCT images where the features truly are generated online. In the HRCT domain, the potential feature space is enormous and only a few features can be held in memory at any given time. As a result, we show that this is a natural fit for OFS and is an approach that can outperform existing algorithms for emphysema classification. In this work, we deal with two main components: feature extraction and feature selection. We begin OFS for emphysema classification with training images and a corresponding ground truth that marks each pixel in the training image as either emphysema or non-emphysema. Our feature extractor, described in Section 3, generates combinations of image processing operators and applies them to the training images to produce features. A classifier is then produced from a linear combination of these features. The OFS algorithm, discussed in Section 4, selects a subset of those features and integrates them into the model as they arrive. The subset of features that was selected is stored and d more features are generated and added to the set. Then the online feature selection algorithm selects another subset. The last two steps are repeated until a stopping condition is met.

II. RELATED RESEARCH

A common approach to detecting emphysema is called “Density Mask” which is simply a thresholding technique [1]. Friman et al. [2] have combined image processing and neural networks into an emphysema detection system that produces high accuracy and reproducibility. Emphysema, along with other diffuse lung diseases, is a disease where textural features have been used widely for detection [1, 4]. Uppaluri et al. [5] has used a Bayesian classifier for recognizing several lung disorders, among them emphysema. Their approach partitions the lung into 31×31 blocks which subsequently are classified, which implies an immediate loss in accuracy. It has also been shown that Independent Component Analysis (ICA) can be used to perform feature subset selection for classification of emphysema using naive Bayes and C4.5 decision tree in HRCT images [6]. ICA is an iterative algorithm that is aimed at producing an entirely new co-ordinate system, with the first component being the “most non-gaussian”, the next being less non-gaussian than the first and so on. However, the use of ICA as a feature selection technique has been used in the traditional sense where it is assumed that all the features are available in the beginning. More recently, Singh [7] has explored Emphysema detection using JPEG compressed lung images with reasonable accuracy. Prasad et al. [8] have used semi-supervised learning techniques to classify different types of Emphysema.

Feature selection algorithms generally fall into three main categories: filters, wrappers and embedded methods [9]. In the filter model, feature selection algorithms can be classified into two groups, namely, feature weighting algorithms and subset

search algorithms. The Relief algorithm is a feature weighting algorithm that assigns a “relevance” weight to each feature [9]. It samples patterns randomly from the training set and updates the relevance values based on the difference between the selected pattern and the two nearest patterns of the same and opposite classes. Relief works by selecting those features whose scores exceed a user-defined threshold to form the final subset. For the subset search algorithms in the filter model, exhaustive search over all possible subsets of a feature set is not computationally feasible. Hence a number of authors have explored the use of heuristics, often in conjunction with branch and bound search [9, 10, 11]. Another school of thought argues that the bias of a particular classifier should be taken into account when selecting features. This technique is called the ‘wrapper’ [10]. It uses an induction algorithm together with a statistical resampling technique such as cross-validation to evaluate feature subsets. Wrappers often achieve better results than filters due to the fact that they are tuned to the specific interaction between an induction algorithm and its training data. However, they are usually much slower than filters as they repeatedly execute the induction algorithm and must be re-run when a different induction algorithm is used [11]. Embedded methods implement the same idea, but proceed more efficiently by directly optimizing a two-part objective function with a goodness-of-fit term and a penalty for a large number of variables.

While the feature selection literature is ample, the OFS literature is rather sparse and mostly involves example problems where all of the features are actually available beforehand [12]. Jiang et al. [13] have also explored online feature selection algorithm in a boosting manner to select the most representative features in the context of content-based image retrieval. Nurmi and Floreen [14] have also performed online feature selection using Gaussian kernels from time series data. The efficiency of the technique is evaluated for the purpose of activity recognition using toy data and real context data, gathered using a 3D accelerometer. Woodley et al. have proposed a tracking algorithm based on OFS in order to improve robustness to occlusions in the images [15]. Local features obtained from regions that overlap with outlier regions are excluded from the classifier. Thus, the classifier only considers features in regions consistent with the generative model. Recently, Glocer et al. [16] have also explored OFS in grayscale imagery but their work was limited to the problem of edge detection only.

III. FEATURE EXTRACTION

The lungs are initially located using software developed within the group [25] and features are extracted from the lungs identified. Feature extraction is an integral part of classifier construction. In image processing, a feature can be generated as the output of one of a series of operators applied to an image. The scalar value of a pixel in the output image is the value of the feature for that instance. Feature extraction encompasses not only the question of which operators are used but also the complexity with which they are combined. We study the effects of feature complexity by comparing simple

features that consist of a single operator to features whose structure is a tree of operators.

A. “Simple” Features

A “simple” feature is generated by running a single image processing operator on a raw input image. The operators used in this work are listed in Table 1. Some of these operators are neighborhood operators, where each output pixel is computed from a set of neighboring input pixels. For instance, each pixel in the output of the Gaussian smoothing operator is a weighted average of the pixels in its neighborhood in the input image. The weights are given by a Gaussian centered at that pixel. Similarly, the median filter provides the median of the pixels in its neighborhood. Binary thresholding converts a raw input image into a binary image by setting pixels below a certain threshold to zero and the remaining to one. The statistical operators perform various computations on the pixel values in a neighborhood of a given pixel and the result of those computations provide the pixel values for the output image.

A suite of morphological operations such as opening, closing and thresholding are included for feature extraction. In the HRCT setting, textural features are often used to characterize emphysema in the lung regions. Textural parameters for the classification of emphysema are calculated on a small neighbourhood surrounding each image point belonging to the lung region. The following methods are used for textural feature extraction in the context of emphysema classification:

1. moments of gray level histogram of a local area (Moments_Histogram)
2. gray level co-occurrence matrix method (GLCMM)
3. gray level run length matrix method (GLRLMM)
4. gray level difference method (GLDM)

The GLCMM, one of the well known texture analysis methods, estimates image properties related to second-order statistics. Each entry (i, j) in GLCM corresponds to the number of occurrences of the pair of gray levels i and j at a distance d apart at an angle θ in original image. The configurations of the co-occurrence matrix used in our experiments include $1 \leq d \leq 2$ and $0 \leq \theta \leq 90, \pm 45$ since these values are sufficient to cover uniformity of disease features. The GLRLMM is based on computing the number of gray level runs of various lengths in different directions. Each element of the GLRLM (i, j) specifies the estimated number of times a picture contains a run of length j , for gray level i , in the direction of angle θ . Three grey level run length matrices, where $0 \leq \theta \leq 360, \pm 45$, are used in our experiments. The full range of θ provides greater uniformity among the various disease features used in our experiments. GLDM is concerned with the spatial gray-level distribution and spatial dependence among the gray levels in a local area. The features extracted from the methods are displayed in Table 1; some features have multiple values, as discussed above.

Moments_Histogram	GLCMM	GLRLM	GLDM
-------------------	-------	-------	------

Mean	Energy	Short Emphasis	Mean
SD	Entropy	Long run emphasis	Contrast
Variance	Homogeneity	Gray level uniformity	Entropy
Energy	Contrast	Primitive length uniformity	SD
Entropy		Primitive percentage	Variance

Table 1: Textural Features

The final set of “simple” features are listed as categories in Table 2. A neighbouring window size of 12 x 12 is used for computing neighbourhood based features. It is reported by Prasad et al. [17] that window sizes less than 12x12 do not provide uniformity of disease patterns and window sizes larger than 12x12 are computationally expensive. These features form the foundation for constructing tree-structured features. Simple features used with a linear classifier restrict the hypothesis class to linear combinations of single operators. This is not rich enough to learn hypotheses that require combinations of operators. As a result, a richer set of features formed using the ‘simple’ features in the form of tree-structures are used to produce richer hypothesis class.

Feature	Category
Gaussian Smoothing	SMOOTH
Mean Filter	SMOOTH
Median Filter	SMOOTH
Moments_Histogram Features	STAT
GLCMM Features	STAT
GLRLM Features	STAT
GLRDM Features	STAT
Morphological Open, Close	MORPH
Binary thresholding	MORPH

Table 2: Operators used to produce simple features which also form the building block for all tree-structured features.

B. Complex Features

The richer set of complex features are constructed using tree-structures that have been demonstrated to work well in other situations [18]. Given a rich set of operators, these trees can represent a large variety of functions of the original input data. The operators for these features are the same as those used to generate simple features, but are combined only in ways that make sense given the nature of the operators. This constraint is provided by a context-sensitive grammar as they offer many advantages. Firstly, only sensible features are generated. Secondly, by restricting the way in which features can be combined, grammars greatly reduce the size of the search space and thirdly, they provide an efficient way to incorporate domain knowledge. The production of the grammar is described in Table 3. When using simple features, a linear classifier is restricted to linear combinations of single operators. This

also restricts the classifiers from learning sequential combinations of operators. The grammar is used to define the sequence of operations need to generate the complex features. For instance, an example of STAT (SMOOTH(x)) may correspond to the process of getting histogram related features from an image that has been smoothed using the Gaussian filter.

Production
STAT (SMOOTH (x))
MORPH (SMOOTH (x))
STAT (x)
STAT (STAT(X))
MORPH (STAT(X))
MORPH (STAT (SMOOTH(x))
MORPH (MORPH(x))

Table 3: Grammar used to generate features. Additionally, new features can be generated by adding, subtracting, multiplying or dividing any two features generated from these rules.

IV. ONLINE FEATURE SELECTION

Online feature selection (OFS) is used to provide a systematic way to search the large feature space at different time points. The OFS problem assumes that features arrive in stages but no new instances are added to the problem. At stage t , a new set f_t of features arrives. The set of all features at stage t is denoted by F_t . Thus, $F_t = \{f_t \cup F_{s,t-1}\}$, the union of features that have just arrived with the set of features that was selected at time $t-1$. At time t , after the arrival of f_t , a feature selection algorithm selects the subset $F_{s,t} \subseteq F_t$ based on some as yet unspecified criterion. In this way online feature selection can be viewed as adding a wrapper around a feature selection algorithm that is parameterized by the number of features added per stage, $d_t = |f_t|$, and by the feature selection algorithm it uses. At the core of the online feature selection framework is the criterion and the search that changes the choice of the features selected. In this work, two wrappers, a correlation based filter approach and grafting (an embedded technique) is used.

A. Hill Climbing using “empirical Bayes risk”

The hill climbing algorithm initializes a cache with d features. The initial features of the algorithm is the fitness of the initial cache. The cache is then mutated in one of three ways, each of which is equally probable: a randomly selected feature is removed from the cache, a randomly generated feature is added to the cache, or a randomly selected feature in the cache is replaced with a randomly generated feature. If this mutation improves the fitness, the mutation is kept. Forrest and Mitchell [19] refer to this as a random mutation hill climbing, or RMHC. In this work, the criterion used to evaluate the fitness of the cache is the “empirical Bayes risk” introduced by Green and Swets [20], which can be computed from the ROC curve by finding the point on the curve where the slope is 45 degrees.

B. Best First Search

It has been reported by Kovahi and John [21] that best first search is more robust than hill climbing. Two most common

variants of best first search are sequential forward selection and sequential backward elimination. Although the process can be slower, backward elimination is used in this work because it is not sensitive to the feature that is chosen first. The number of features per stage of the online feature selection algorithm is small enough that the computation time for either algorithm is much less than the time it takes to extract new features. Sequential backward elimination is initialized with a full set of features F_0 . The algorithm first removes each feature $f_i \in F_0$ from F_0 and trains an induction algorithm with the feature set $F_0 - \{f_i\}$. We use the Fisher Linear Discriminant as an induction algorithm in this work. The feature f_{max} whose removal minimizes the error in the resulting classifier is removed, and $F_1 = F_0 - \{f_{max}\}$. In the next round, for each of the remaining features $f_i \in F_1$, the algorithm tests the feature subset $F_1 - \{f_i\}$ and removes the feature f_{max} that minimizes the error of the resulting classifier to produce $F_2 = F_1 - \{f_{max}\}$. This process repeats until a local minimum of classifier error has been reached or some other stopping condition is met. Removing a feature may have only a small negative impact on performance, so backward elimination becomes a trade-off between marginally lower performance and a smaller set of features. This implies the need for some sort of regularization. Following Kovahi and John [19], we have used the same approach of adding a penalty $c=0.001$ per feature in order to force the algorithm to favour smaller subsets. Hence, the best first search algorithm minimizes the loss function given below where R_{BAYES} is the empirical Bayes risk, described earlier.

$$L = R_{BAYES} + c | \omega |_0 \quad (1)$$

C. Filter based Correlation Based Feature Selection

Comparisons were also performed with Correlation based Feature Selection (CFS) metric that has been commonly used in the filter framework for many machine learning problems [11]. The heuristic by which CFS measures the goodness of feature subsets takes into account the usefulness of individual features for predicting the class label along with the level of intercorrelation among them. The hypotheses on which the heuristic is based can be stated: *Good feature subsets contain features highly correlated (predictive of) with the class, yet uncorrelated with (not predictive of) each other.* The heuristic is formalized by the definition:

$$Merit_s = \frac{k \overline{r_{cf}}}{\sqrt{k + k(k-1)r_{ff}}} \quad (2)$$

where $Merit_s$ is the heuristic “merit” of a feature subset S containing k features, $\overline{r_{cf}}$ the average feature-class correlation, and r_{ff} the feature-feature intercorrelation. Equation 2 is nothing but Pearson’s correlation coefficient [22], where all variables have been standardized. Pearson’s correlation coefficient reflects the degree of the dependency between two variables. The numerator of the equation indicates how

strongly a feature group is correlated with the class and the denominator indicates how much redundancy is present among them. More details of the technique can be obtained in (Hall, 1999) where the performance of CFS using best first search has been shown to be superior to other search strategies [11].

D. Grafting

Grafting [12] uses the idea of regularization as an alternative to feature selection. In other words, the technique minimizes a regularized risk criterion of the form:

$$C = L(f(x)) + \lambda \sum_{j=1}^d |\omega_j| \quad (3)$$

The second term, the regularizer, is the l_1 norm of the weight vector. The free parameter, λ characterizes the trade-off between accuracy and complexity. The first term of the criterion function is the loss function, which is implemented as the binomial negative log likelihood (BNLL) loss described in Hastie et al. (2001):

$$L(f(x)) = \frac{1}{n} \sum_{i=1}^n \ln(1 + e^{-y_i f(x_i)}) \quad (4)$$

The grafting algorithm is based on the observation that the addition of feature i incurs a penalty of $\lambda |\omega_j|$. Thus adding the feature is only worthwhile if the reduction of the mean loss is greater than the increase in the penalty, and that this will only happen if $|\partial L / \partial \omega_j| > \lambda$. This gradient test is performed for each feature as it arrives, and it is faster than re-optimizing the classifier with respect to the new feature. If the weights do not pass the test, then they are discarded. On the other hand, if at least one weight passes, then the weight that maximizes the magnitude of the gradient is added and the model is optimized with respect to all of its parameters. Grafting differs from the other algorithms because it considers not only whether to add a new feature but also whether to drop currently selected features and even whether to discard features.

V. EXPERIMENTAL RESULTS AND DISCUSSION

The aim of performance evaluation is to compare the level of effort expended between algorithm-based classification versus manual classification of the same job. Accuracy is estimated using 10 fold cross-validation on a dataset consisting of 61 HRCT scans. The regions were labeled manually through interactive drawing of regions of interest in consultation with radiologists. They are shown as darker areas in Figure 2(a). With the different range of parameter values, the total number of simple features resulted in 49 features. Random mutation hill climbing was not used due to the small number of features. It is worth noting that the space of tree-structured features is much larger than the simple features space.

Because generating features is the time-limiting step, each algorithm was allowed to explore a total of 500 features. The hill climber was initialized to one. At each iteration, a feature was added to the cache, removed from the cache, or replaced with another feature. For best first search, the feature cache was initialized to five features. The backward elimination was run with a complexity penalty of $c=0.001$ for every feature. Grafting differs from the other algorithms in that whenever a new feature is added, it can check discarded features and potentially reincorporate them into the model. Ideally, the best result would happen by keeping all the features in memory, but this is not feasible. Instead, 20 iterations are run where 25 features are explored at every iteration. Evaluation was done using C4.5 decision trees [24]. C4.5 is an algorithm that represents the training data in the form of a decision tree and is known for their robustness and execution speed.

To illustrate the nature of these experiments, Figure 2 shows the results of emphysema classification using the features selected by the CFS technique, using the whole feature set along with “density mask” and the original image. “Density Mask” uses a simple thresholding technique where areas with attenuation of less than -910HU correlated closely with the pathologic assessment of emphysema. As can be seen in Figure 2(d), the output of the “density mask” algorithm shows that a lot of noise is picked up along with the emphysema regions. The mean accuracy comparison of the different techniques along with “density mask” is shown in Table 4 and it can be observed that using the feature selection techniques results in higher mean accuracy. CFS outperforms hill climbing, which in turn was slightly better than best first search and grafting.

Additionally, we also combined the different online feature selection techniques using a simple voting scheme. The features selected by the different techniques at each time point are chosen by a majority vote. However, the accuracy using the combined approach resulted in a very small improvement as can be seen in Table 4. The comparison of the different techniques was also made using statistical tests, and a significant difference between the online feature selection techniques and that of the original features was observed (all $p < 0.001$). While a significant difference was not observed among the online feature selection techniques, there was a difference between the techniques and “Density Mask” (all $p < 0.001$). Although no quantification was made, it was observed that the decision trees built by the online techniques were dependant on the order in which relevant features were seen.

Technique	Mean Accuracy
Hill Climbing	87.37
Best First Search	87.11
CFS	88.13
Grafting	86.97
Combined	89.33
“Density Mask”	79.36
Original Features	83.51

Table 4: Evaluation of different techniques for emphysema classification given by mean accuracy.

It was also observed that the tree-structured features outperformed the simple features for the same algorithm and hence, proving that there is an advantage to increasing the complexity of the feature structure. Table 5 shows comparison of the results for the different techniques when tree-structured features were used.

Technique	Simple features	Tree-structured features
Hill Climbing	85.75	87.63
Best First Search	82.17	85.19
CFS	85.37	88.31
Grafting	86.33	87.91

Table 5: Comparison of mean accuracy using different OFS techniques for simple alone and tree-structured features included.

VI. CONCLUSIONS

In this paper, we have presented a novel online feature selection algorithm for the detection of emphysema in HRCT images. In the HRCT domain, the computational bottleneck is the feature generator, not the classifier. Adding complexity to the structure of the features makes the feature space too large to search exhaustively but it increases the flexibility of the hypothesis class and can improve the performance of the resulting classifier. OFS provides a systematic way to search this larger feature space, and we compared classification using online feature selection as a pre-processing technique against classification in the original feature space along with the classical “density mask” approach. The OFS algorithms achieve higher accuracy comparatively and hence, we are able to provide better emphysema quantitative measures for medical experts. CFS in particular was found to be an efficient way to incorporate features in an online fashion. The OFS technique is more practical and widely applicable since many “real-world” computer vision applications are dominated by features that are not readily available in the beginning.

REFERENCES

- [1] M. Kinsella, N.L. Mueller, R.T. Abboud, N.J. Morrison, A. DyBuncio. *Quantification of emphysema by computed tomography using a density mask program and correlation with pulmonary function tests*, Chest 97 (1990) 315-321.
- [2] O. Friman, M. Borga, M. Lundberg, U. Tylén, H. Knutsson, *Recognizing emphysema - A Neural Network Approach*, Proceedings of 16th International Conference on Pattern Recognition, August 2002.
- [3] H.-U. Kauczor, K.R. Heitmann, C.P. Heussel, D. Marwede, T. Uthmann, M. Thelen. , *Automatic detection and quantification of ground glass opacities on HRCT scans of the lung: comparative study of a neural network, visual assessment and a density mask*, European Radiology, (1997) 1463-1472.
- [4] H. Mir, M. Hanmandlu and S. N. Tandon. *Texture analysis of CT images*, IEEE Engineering in Medicine and Biology, vol. 14, 1995.

- [5] R. Uppaluri, E. Hoffman, M. Sonka, P. Hartley, G. Hunninghake, and G. McLennan. *Computer recognition of regional lung disease patterns*. Am. J. Respir. Crit. Care Med., 160:648–654, 1999.
- [6] M. Prasad, A. Sowmya, I. Koch, *Feature Subset Selection using ICA for classifying Emphysema in HRCT images*, In proc. of international conference on intelligent sensors, sensor networks and information processing, Melbourne, Australia, 2004.
- [7] P. K. Singh, Emphysema detection in JPEG compressed HRCT lung images, Proceedings of the Eighth International Symposium on Signal Processing and Its Applications, 1, 195-198, 2005.
- [8] M. Prasad, A. Sowmya, P. Wilson, Multi-level classification of emphysema in HRCT lung images, Pattern Analysis & Applications, 2007 (in press, DOI: 10.1007/s10044-007-0093-7)
- [9] I. Guyon, A. Elissee, *An introduction to variable and feature selection*, Journal of Machine Learning Research, 3, 1157-1182, 2003.
- [10] R. Kohavi., G.H. John, *Wrappers for feature subset selection*. Artificial Intelligence 97, 273-324, 1997.
- [11] M. Hall, Correlation-based feature selection for discrete and numeric class machine learning. *Proceedings of the Seventeenth International Conference on Machine Learning*, pp. 359-366, 1999.
- [12] S. Perkins, J. Theiler, *Online feature selection using grafting*, Proceedings of the Twentieth International Conference on Machine Learning, pp. 592-599, 2003.
- [13] W. Jiang, G. Er, Q. Dai, L. Zhong and Y. Hou, “Relevance feedback learning with feature selection in region-based image retrieval”, Proc. of IEEE Int. Conf. Acoustics, Speech and Signal Processing, 2, 509-512, 2005.
- [14] P. Nurmi and P. Florén: “Online feature selection for contextual time series data (extended abstract).” Presentation at the PASCAL workshop on Subspace, Latent Structure and Feature Selection Techniques: Statistical and Optimisation Perspectives. Bohinj, Slovenia, 23–25, 2005.
- [15] T. Woodley and B. Stenger and R. Cipolla, *Tracking using online feature selection and a local generative model*. In Proceedings of the British Machine Vision Conference. British Machine Vision Association, pp. 790-799, 2007.
- [16] K. Glocer, D. Eads, J. Theiler: Online feature selection for pixel classification. *Proceedings of the 22nd international conference on Machine learning*, 249-256, 2005.

- [17] M. Prasad, A. Sowmya, P. Wilson, *Multi-level Classification of Emphysema in HRCT lung images*, Pattern Analysis and Applications Journal, 2007 (in press).
- [18] J. Koza, *On the Programming of Computers by Means of Natural Selection*. Cambridge: MIT Press, 1992.
- [19] S. Forrest, M. Mitchell, Relative building-block fitness and the building-block hypothesis. In L. D. Whitley (Ed.), *Foundations of Genetic Algorithms 2*, 109-126. San Mateo, CA: Morgan Kaufmann, 1993.
- [20] D. Green and J. Swets, *Signal Detection Theory and Psychophysics*, New York: Wiley, 1966.
- [21] R. Kohavi, G. H. John, *Wrappers for feature subset selection*, Artificial Intelligence, 97, 273-324, 1997.
- [22] J. S. Milton and J. C. Arnold, *Introduction to probability and statistics*, 3rd edition, McGraw-Hill, New York, New York, USA, 1995.
- [23] T. Hastie, R. Tibshirani and J. Friedman, *The Elements of Statistical Learning*. New York: Springer, 2001.
- [24] Tom Mitchell, *Machine Learning*, McGraw Hill, 1997.
- [25] P. T. Chiu and A. Sowmya. *Lung Boundary Detection and Low Level feature extraction and analysis from HRCT images*, VISIM: Information Retrieval and Exploration of Large Medical Image Collections, October 2001.

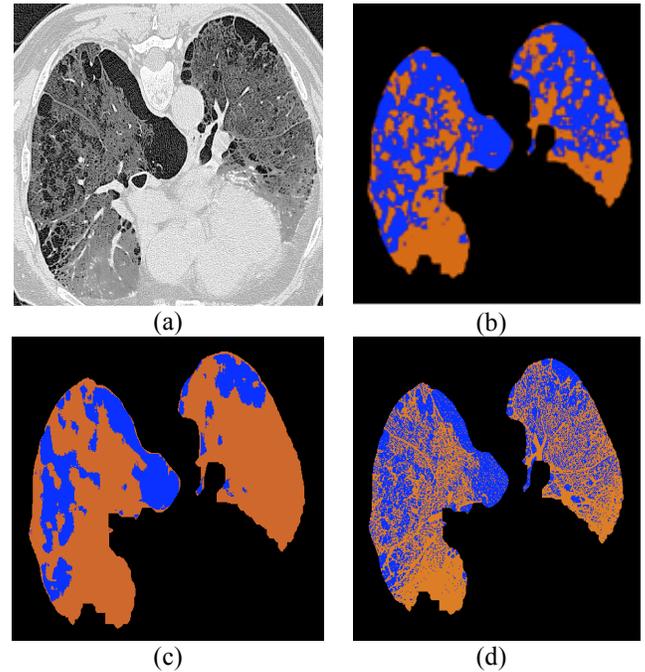


Figure 2: 2(a) contains the original image where the dark regions correspond to emphysema. 2(b) and 2(c) and 2(d) are the outputs of C4.5 in original feature space and in the feature space chosen by CFS respectively. The blue regions correspond to emphysema.