Pattern Anal Applic
DOI 10.1007/s10044-007-0093-7

THEORETICAL ADVANCES

Multi-level classification of emphysema in HRCT lung images

Mithun Prasad · Arcot Sowmya · Peter Wilson

Received: 8 January 2006 / Accepted: 24 September 2007

Abstract Emphysema is a common chronic respiratory disorder characterised by the destruction of lung tissue. It is a progressive disease where the early stages are characterised by a diffuse appearance of small air spaces, and later stages exhibit large air spaces called bullae. A bullous region is a sharply demarcated region of emphysema. In this paper, it is shown that an automated texture-based system based on co-training is capable of achieving multiple levels of emphysema extraction in high-resolution computed tomography (HRCT) images. Co-training is a semi-supervised technique used to improve classifiers that are trained with very few labelled examples using a large pool of unseen examples over two disjoint feature sets called views. It is also shown that examples labelled by experts can be incorporated within the system in an incremental manner. The results are also compared against "density mask", currently a standard approach used for emphysema detection in medical image analysis and other computerized techniques used for classification of emphysema in the literature. The new system can classify diffuse regions of emphysema starting from a bullous setting. The classifiers built at different iterations also appear to show an interesting correlation with different levels of emphysema, which deserves more exploration.

M. Prasad (✉) · A. Sowmya
School of Computer Science and Engineering,
University of New South Wales, Sydney, NSW 2052, Australia
e-mail: mithunp@cse.unsw.edu.au

A. Sowmya
e-mail: sowmya@cse.unsw.edu.au

P. Wilson
I-med Network, Sydney, NSW 2000, Australia
e-mail: pcwilson2@bigpond.com

1 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 diagnosis by radiologists is often based on visual recognition of imaging patterns augmented by anatomical knowledge. Emphysema is a common chronic respiratory disorder characterised by the destruction of lung tissue and is often reflected as areas of low attenuation in CT images [1] as shown in Fig. 1. Tobacco smoking is the main cause of emphysema, although there may be other contributing factors for development of the disease. Emphysema is among the top five diseases in the western world today in terms of rehabilitation and health care costs. From these perspectives, it is of vital importance to develop methods for diagnosing emphysema, both for clinical and research use [2].

Emphysema is a progressive disease, characterised by abnormal air spaces. These are typically small in the early stages, but become larger and involve the lung more diffusely over time. Large air spaces called bullae may develop, particularly in the later stages. Bullous emphysema is

Fig. 1 A typical HRCT scan containing emphysema. The regions outlined denote emphysema

histologically referred to as the presence of emphysematous areas with complete destruction of lung tissue. The classification of bullae is useful to evaluate patients as candidates for surgery. Figure 2 visually presents the appearance of bullous and diffuse regions of emphysema.

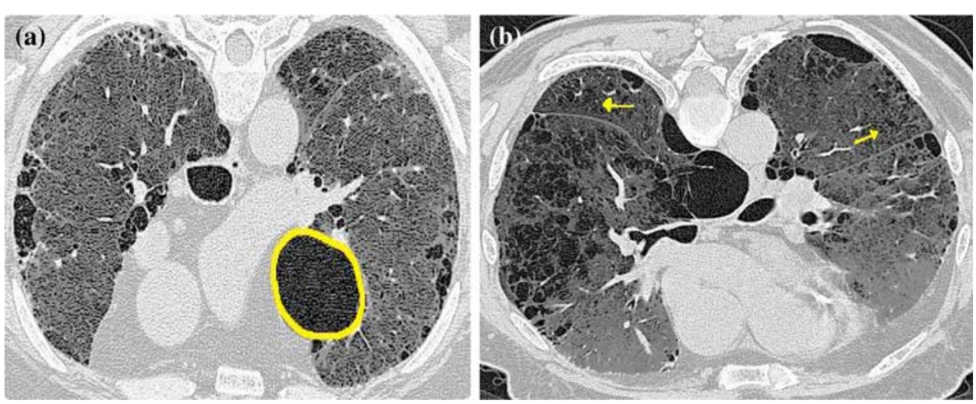
The techniques introduced in this paper are intended to automate the recognition process and assist radiologists in the diagnosis of emphysema by providing accurate measures of severity across each HRCT scan. This is achieved by using minimal anatomical knowledge. In automated emphysema detection in lung images, a common technique called "density mask'' is applied simply to threshold the image [1]. However, a fixed threshold yields unsatisfactory results when the degree of emphysema is low. Further, the issue of estimating the degree of emphysema has not been previously addressed. Computerised techniques for classifying emphysema have been explored [2–4] using texture and machine learning approaches with reports of reasonable accuracy. In a supervised learning framework, the system is given examples belonging to more than one class. All examples are labelled with respect to their membership in one of the classes. A machine learning system induces a general description of the classes from these examples. In the HRCT setting however, labelling is a very time consuming and expensive process because it requires expert effort. Therefore, it would pay to take as much advantage of the unlabelled data as possible.

The proposed system is called Inc_MVL. It uses Meta_MVL that is based on a multi-view framework where the domain features are partitioned into disjoint subsets known as views such that each view is sufficient to learn the target concept. Preliminary versions of Inc_MVL and Meta_MVL have been presented [7, 8]. Multi-View systems are based on the assumption that the views are both compatible and uncorrelated. If all examples are labelled identically by each view, the views are said to be compatible. Multi-view algorithms have been successfully developed for web-page classification [5, 6]. However, application of multi-view algorithms to vision problems has not been addressed to our knowledge. The overall system is shown in Fig. 3.

The contributions of this paper are the following:

1. The proposed system, called Inc_MVL, is an incremental version of Meta_MVL where provision has been made to incorporate images labelled by experts incrementally in order to improve the performance of the system. The approach taken in this work is based on active learning in the multi-view framework. Active learners aim to detect the most informative examples in the instance space and ask the experts to provide labels to such examples.
2. Meta_MVL is capable of classifying different levels of diagnosis automatically. The approach is based on multi-view learning. The levels range from the larger set of diffuse and bullous regions, to just bullous regions.
3. Inc_MVL is also capable of learning to detect diffuse regions from a bullous setting. Because labelling

Fig. 2 **a** A typical HRCT scan containing bullous emphysema. **b** A typical HRCT scan containing diffuse regions of emphysema



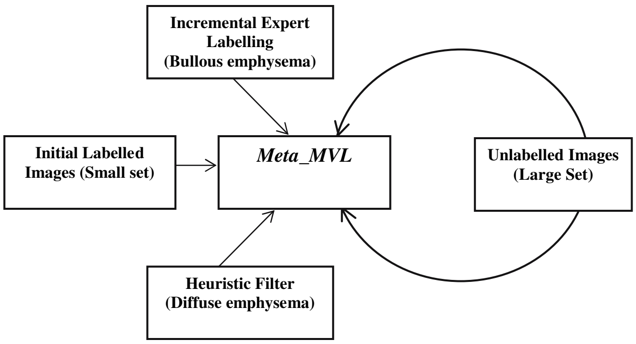


Fig. 3 Inc_MVL showing different modules

diffuse regions of emphysema is tedious and erroneous, a heuristic based filter is used to provide potential diffuse regions which are then used by the system as a labelled image used for training. The heuristic based filter is based on a suite of image processing techniques.

The paper is organised as follows: Related research is described in Sect. 2. Image pre-processing performed on the images along with the feature extraction is discussed in Sect. 3. A description of the view detection algorithm is provided in Sect. 4. Semi-supervised learning by reducing labelled data is introduced in Sect. 5. The multi-view learning system is described in Sect. 6. An outline of the procedure to incorporate labelled images by the experts incrementally in the multi-view system is provided in Sect. 7. Experimental results of classifiers are presented in Sect. 8. A description of the techniques used to detect diffuse emphysema in the multi-view framework is provided in Sect. 9 and conclusions are presented in Sect. 10.

2 Related research

A common approach to detecting emphysema is called "density mask" which is simply a thresholding technique [1]. Recently, Friman et al. [2] 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]. However, the issue of estimating the degree of emphysema has not been previously addressed. Uppaluri et al. [24] used a Bayesian classifier for recognizing several lung disorders, among them emphysema. Their approach partitions the lung into 31 $\times$ 31 blocks which are subsequently 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 in HRCT images [26]. 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. All of these systems are supervised and rely on the availability of labelled samples.

In a supervised learning framework, the system is provided with examples belonging to more than one class with the examples labelled with respect to their membership in one of the classes. A general description of the classes from these examples is induced by the machine learner. Recently there has been significant interest in supervised learning algorithms that combine labelled and unlabelled data [27]. However, a lot of the activity lies in text classification. The two main approaches used to exploit unlabelled data are to: (1) transform the input feature space using information in the unlabelled data and (2) iteratively label part of the unlabelled data [9]. Zelikovitz and Hirsh [10] apply the first approach where they use an unlabelled corpus to create a domain model that incorporates words that co-occur in the corpus. The second approach is more common than the first one and a number of different methods have been proposed [6–11].

Blum and Mitchell [6] introduced co-training, an approach based on semi-supervised learning in a muti-view framework. The goal of the technique is to improve classifiers trained with very few labelled examples using a large pool of unseen examples. Their problem was the classification of Universities' web pages as home pages of academic courses. The two sets of features used were text appearing on the document and the anchor text attached to hyperlinks pointing to the document. The co-trained classifiers outperformed the classifiers formed by supervised learning alone. In addition, Blum and Mitchell provide some theoretical insight into the co-training algorithm. They prove that any weak initial classifier can be boosted to arbitrarily high accuracy using unlabelled examples only if the feature sets representing the views are compatible and uncorrelated. Raskutti et al. [9] use unlabelled examples to enrich the original set of features with some additional, highly-informative attributes based on a clustering algorithm. The authors show that by using the enriched set of features, one can significantly improve the classification accuracy. Muslea et al. [11] propose a robust multi-view algorithm Co-EMT that combines semi-supervised and active learning. Co-EMT runs a semi-supervised learner in a multi-view framework and asks the user to label examples on which the combined prediction of the classifiers in each view is least certain.

In the multi-view setting, one assumes that each view is sufficient to learn the target concept and sufficiently uncorrelated. However, in practice, there are also views in which one can only learn a concept that is strictly *more general* or *more specific* than the concept of interest. The weaker view defines a class of concepts that is more general than the target concept and the stronger view defines concepts that are more specific. For example, in the case of emphysema detection, detection of bullous emphysema is more specific than detection of diffuse regions. Despite the limitations of the views, one can detect examples, which are very informative; that is, the unlabelled examples on which the views do not agree. In this case, at least one view is wrong and thus, obtaining the experts' answer to such query examples leads to faster convergence. The idea of combining strong and weak views appears in many applications. For example, Kushmerick et al. [12] describe the problem of classifying the various lines of text on a business card as a person's name, affiliation, address, phone number and fax. In this application, the strong view consists of the words that appear on each line. In the weak view, Kushmerick et al. [12] exploit the relative order of the lines on the card. By itself, the order of the text lines cannot be used to accurately classify the lines. However, when combined with the strong view, the ordering information leads to a classifier that clearly outperforms the stand-alone strong view. The DISCOTEX systems also use the idea of combining strong and weak views [13]. The system combines an information extraction system (strong view) with a text mining one (weak view). In this case, the task consists of extracting the relevant information from computer science job postings to the newsgroup austin.jobs; i.e., DISCOTEX must extract the job title, salary, location, progamming languages, development platforms and required degree. Nahm and Mooney [13] show that by combining strong and weak views, improved extraction accuracy will result. In this work, since classification is done at pixel level, it is tedious and time consuming for the user to label each pixel because of the large number of pixels in an image. Instead, where the views disagree (with respect to the labelled images), the regions labelled by an expert can be used in an active learning framework.

3 Image processing

Feature extraction is an integral part of classifier construction. In this work, textural features are used to characterize emphysema. The aim of the image-processing module is to extract textural parameters from lung regions in the HRCT image. There are essentially two main steps: automatic segmentation of the lungs and feature extraction.

In this step, the lungs in the image are located and extracted. A suite of classical image processing techniques is used to segment lung regions using in-house software [14]. This is quite a straightforward approach where the different morphological operations performed to segment lung regions include dilation, erosion and thresholding. The percentage area occupied by lung regions in the whole image is used to decide whether the image is of interest. A percentage value of less than 6 is considered unacceptable.

In the application, feature extraction is primarily based on texture as emphysema is a finding that can be well characterized by texture. Texture is a very commonly used term in computer vision. It is easy to recognise texture, but very difficult to define it precisely. A statistical approach is used to describe texture. A feature vector is defined as a set of textural parameters calculated on a small neighbourhood of 12×12 pixels surrounding each image point belonging to the lung region. Window sizes less than 12×12 do not provide uniformity of disease patterns and window sizes larger than 12×12 are computationally expensive. The textural parameters used in the experiments are based on the following methods:

1. moments of gray level histogram of a local area
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 [6]. 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.

4 Detection of views

The textural feature vector obtained from the various methods described in Sect. 3 is split into two uncorrelated sets using Pearson's correlation (Eq. 1). It is necessary to obtain two views in order to perform multi-view learning.

Table 1 Textural features

Moments of histogram	GLCMM	GLRLM	GLDM
Mean	Energy	Short run emphasis	Mean
SD	Entropy		Contrast
Variance	Homogeneity	Long run emphasis	Entropy
Energy	Contrast	Gray level uniformity	SD
Entropy			Variance
		Primitive length uniformity	
		Primitive percentage	

Pearson's correlation coefficient ρ reflects the degree of relationship between two variables [18]:

$$\rho = \frac{\mathrm{Cov}(X, Y)}{\sqrt{(\mathrm{Var}\ X)(\mathrm{Var}\ Y)}} \qquad (1)$$

The variables X and Y are the two feature values being tested for correlation. Pearson's correlation coefficient ranges from +1 to -1. A correlation coefficient of +1 means that there is a perfect positive linear relationship and a correlation coefficient of -1 indicates a perfect negative linear relationship between variables. A correlation coefficient of 0 indicates no correlation between the variables.

A correlation coefficient value between ± 0.3 is used to obtain two sets of features that are uncorrelated. The correlation coefficient between a feature and a set of features is computed by averaging the absolute value of the individual intercorrelation coefficients between the feature and all other features within the set. For example, let A and B be two sets of uncorrelated features initially empty. If two features f1 and f2 are uncorrelated, we assign f1 to set A and f2 to set B. For a new feature f3, we assign f3 to the set that is most correlated with f3. This process is repeated until all features are analysed. The uncorrelated sets of features are finally run through a combination of feature subset selection algorithms to retain features that have high discriminative power.

The Weka data-mining suite [17] was used for the experiments, within which all the feature subset selection (FSS) algorithms used in the experiments have been implemented. The various FSS algorithms used in the experiments are:

- *Information gain*: selects a subset of features by measuring the information gain with respect to the class.
- *Relief*: selects a subset of features by repeatedly sampling an instance and considering the given attribute value for the nearest instance of the same and different class.
- *Wrapper*: evaluates attribute sets by using a learning scheme.
- *Principal component analysis*: selects a subset of features by performing principal components analysis and transformation of the data.

Attribute values of all labelled samples are run through each of the feature subset selection algorithms and the feature subset resulting from each algorithm is obtained. The final set of features is obtained at the meta-level by choosing the most frequently occurring features in each of the resulting subsets. The two sets of features selected in each view are presented in Table 2.

5 Reducing labelled data

Labelling emphysematous lung images is a tedious and time-consuming process. Therefore, it is of paramount importance to learn a target concept based on as few labelled examples as possible. A common technique used to reduce the need for labelled training data is semi-supervised learning where the accuracy of the supervised learner is improved by introducing unlabelled examples in addition to the labelled ones.

5.1 Semi-supervised learning

Meta_MVL is based on semi-supervised learning performed in multiple views in order to achieve varying levels of emphysema diagnosis in HRCT images. Typically, a semi-supervised algorithm proceeds as follows: first, it uses the base learner L and the set of labelled examples L to learn an initial hypothesis h. Then h is applied to the unlabelled examples in U, and some or all of these examples, together with the label predicted by h, are added to L. The entire process is then repeated for a number of iterations.

The intuition behind semi-supervised learning is straightforward: even though the initial hypothesis h is

Table 2 Features used in each view for multi-view learning

View1	View 2
Mean (MOH)	Homogeneity (GLCMM $-$ d = 1, θ = 5)
Energy (MOH)	Homogeneity (GLCMM $-$ d = 1, θ = 90)
Sd (MOH)	Primitive percentage (GLRLM $-$ θ = 90)
	Short run emphasis (GLRLM $-$ θ = 90)
	Long run emphasis (GLRLM $-$ θ = 45)
	Grey level uniformity (GLRLM $-$ θ = 90)
	Primitive length uniformity (GLRLM $-$ θ = 45)

learned based on a small training set, its highest confidence predictions are likely to be correct. Thus, by adding the "most confident" examples in the unlabelled set to the training set, one enlarges the training set, based on which a more refined hypothesis can be learned. This hypothesis can be used further to label more examples accurately [15].

5.2 Active learning

The goal of active learners is to detect the most informative examples in the instance space and then ask the user to label them, thus reducing the amount of labelling in the instance space. The two standard approaches used in active learning are: query construction and selective sampling. Query construction asks the user about an example that is constructed by setting the value of each attribute so that the resulting query is as informative as possible. In contrast, selective sampling is a technique where the queries must be chosen from a given working set of unlabelled examples. Selective sampling is more popular because:

1. Most real-world problems consist of a large number of unlabelled examples.
2. Even though one may be able to artificially construct a query that is more informative than the examples in U, there is a high risk that the user may not be able to label such an example because it does not correspond to a real-world entity [16].

In this paper, Inc_MVL is based on selective sampling. However, instead of asking the user to label examples online, the labels are obtained from images labelled offline by experts.

6 Meta_MVL

Meta_MVL is a multi-view learning system based on co-training and semi-supervised learning together with a meta-learner. The co-training approach builds classifiers incrementally over each view. Using an initial set of labelled samples on hand, weak classifiers in each view are built. The intuition is that as long as h1 in view 1 is highly confident on an unseen example, it is very likely that the prediction is correct. This example is added as a labelled example so that the other view, regardless of its prediction will learn from it. The approach used in Meta_MVL combines meta-learning with co-training to provide a more robust technique (see Fig. 4). Meta learning refers to a single classification model derived by learning from multiple local classifiers. Naive Bayes and C4.5 are used as the base level learners whose output is combined at the Meta level.

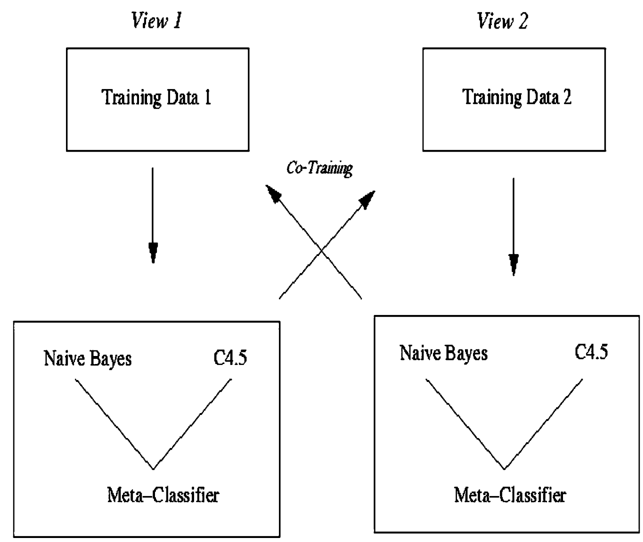


Fig. 4 Block diagram of Meta_MVL

Naive Bayes algorithms provide explicit probabilities for hypotheses and have been widely used in text classification tasks [19]. The Naive Bayes algorithm employs a simplified version of Bayes formula in order to classify each instance. The posterior probability of each possible class is calculated given the feature values present in the instance; the instance is assigned to the class with the highest probability. C4.5 is used to construct a decision tree with the training data [20]. C4.5 is a suitable classifier for this setting as it uses only high information-gain attributes to construct rules. The key idea is that two independent classifiers will be used rather than one in each view. For an unseen image, Naive Bayes and C4.5 predict the pixel labels independently. C4.5 is tweaked in order to predict the class label along with a probability measure. The Class distribution for an instance at the tree nodes is analyzed to obtain this value. An unseen example is classified taking into account not only the class distribution in the leaf but also the class distributions of the nodes in the path. The class distribution is computed taking into account the class distributions at the current node and at the predecessor of the current node. The pixels that have the same labels from the two classifiers in each view are used to obtain the most confidently predicted instances to add to the training set for each classifier. Finally, the predictions of the two classifiers in each view are combined by multiplying them together and then renormalizing their probability scores.

7 Learning through newly labelled examples incrementally

Meta_MVL is now extended to incorporate images labelled by medical experts, as they become available. The resulting

Given a new image, I_new consisting of labelled set <x_1,l_1,.....x_n,l_n>
For each x_i *in the labelled set,*
If $h1(x_i)$ *or* $h2(x_i)$ *disagree with* l_i,
Add x_i *to the training set with label* l_i *from the label set.*

Fig. 5 Inc_MVL-algorithm to incorporate examples labelled by the medical experts in the existing multi-view system

incremental system, called Inc_MVL, is described in Fig. 5:

The approach taken here is based on active learning used in multi-view learning [21]. Active learning algorithms try to detect the most informative examples in the instance space and ask the user to label them. Labelling each pixel or groups of pixels is a very tedious and time-consuming process for the medical experts. Hence, it is worthwhile to use images labelled by the experts to provide expert information to the existing multi-view system. When the classifiers in each view (i.e. *h1* and *h2*) label an example differently with respect to the label provided by the expert in the labelled image, that example is called a *contention point* as shown in Fig. 6d. It is the contention points that are most informative because when the classifiers disagree with the label provided by the expert, at least one of the views must be wrong. The correct label is the label assigned by the radiologist while labelling that image, and is provided to the view that mislabelled it. This also results in fewer examples required to co-train. Active Learning in the multi-view framework queries the user on examples where the two views disagree about its label. However, the case where both the views can be wrong without any disagreement is not currently handled in active learning systems. Inc_MVL also handles this case.

8 Experimental results

In recent years, more attention has been paid to developing standards for performance evaluation of computer vision algorithms [22]. The aim of performance evaluation is to compare the level of effort expended between algorithm-based classification versus manual classification of the same job. The measures used to quantify results are completeness and correctness, also known as recall and precision respectively. Recall is also the same as sensitivity. However, precision is slightly different from specificity, which is commonly used, in medical research. Completeness and correctness are used here because of their easy interpretation. They are percentage values given by:

$$\text{Correctness} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

$$\text{Completeness} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

where TP is true positives, FP false positives and FN false negatives.

High completeness means that the region segmented has covered the relevant disease pattern well, whereas high correctness implies that the region segmented does not contain many (incorrect) irrelevant disease patterns. Each classifier (Naive, C4.5 and Meta_MVL) was evaluated on a labelled test set comprising 60 HRCT scans, and the results are presented in Fig. 7. The classifier at iteration 0 is built using only 12 labelled samples (different from the test set). At each iteration, most confidently labelled positive and negative pixels from the unlabelled set of images are added to the training data (co-training). The most confidently labelled pixels from a group of 15 labelled images are also added to the initial labelled set at each iteration. This is done to ensure variability in the distribution of the textural features of each pixel.

Two sets of experiments were run. The first experiment incorporated labelled examples from the experts only at the last iteration (iteration 12) and the second experiment included labelled examples at early iterations (iterations 1–5). Figure 7a shows that the average correctness increases with iterations and Meta_MVL just outperforms

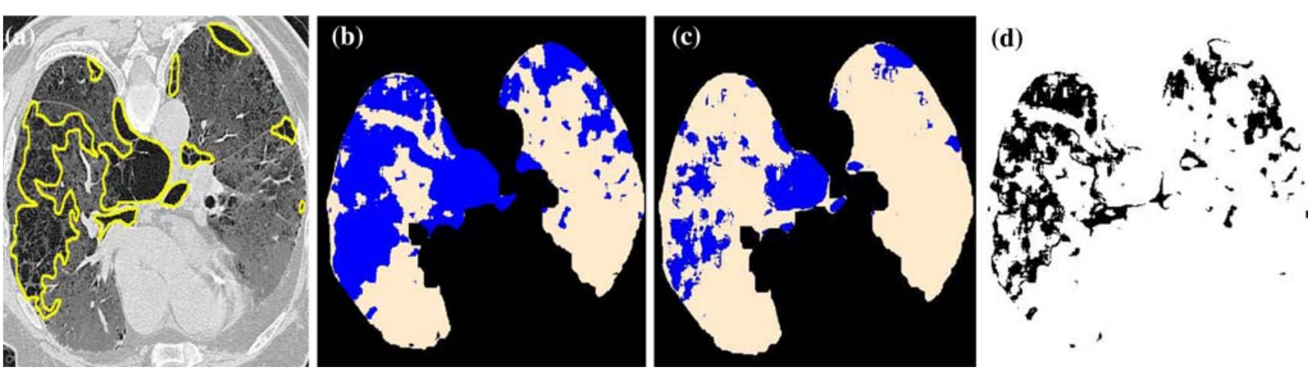
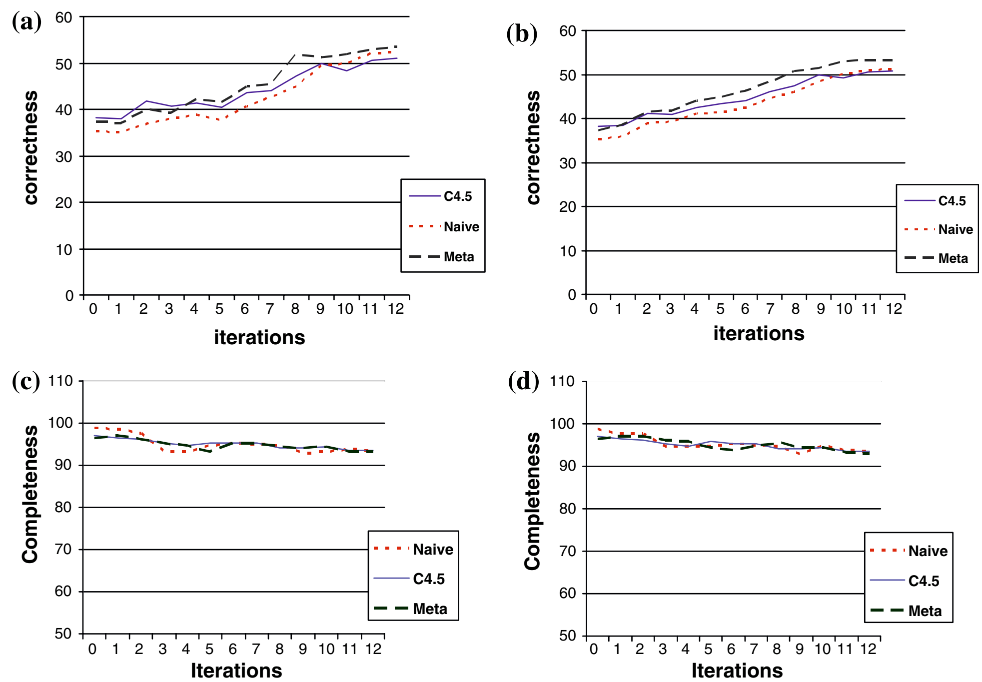


Fig. 6 **a** Shows the original image with the labelled regions marked in yellow. **b**, **c** Correspond to the output of classifier in view 1 and view 2, respectively, and **d** is the image containing contention points

Fig. 7 **a** Displays correctness versus iterations for Meta_MVL and **b** displays correctness versus iterations for Inc_MVL. **c**, **d** correspond to completeness versus iterations for Meta_MVL and Inc_MVL, respectively [images labelled by the experts were introduced at iteration 12 and at iterations (1–5)]



Naive Bayes and C4.5 classifiers although it starts badly for both experiments. However, when labelled images from the experts are incorporated into the system at early iterations, the system achieves higher correctness (Fig. 7b) than when labelled images are introduced at the 12th iteration. The starting completeness (average) values for all three classifiers are very high but drop marginally and remain steady through the iterations for both experiments as shown in Fig. 7c, d.

The initial classifiers in the two views were built using labelled regions from 12 HRCT scans. The regions were labelled manually by radiologists in the team through interactive drawing of regions of interest. 3 of the 12 regions contained emphysema and the others belonged to various other diseases. The system was then co-trained on 165 unlabelled and 15 labelled HRCT images selected from 13 patient studies. Note that each image consists of 2,62,144 pixels (512×512) and each labelled image contains 15000 labelled pixels on average. Classification is performed at pixel level, which results in tens of millions of instances for testing.

Meta_MVL was compared against the standard "density mask" approach to detect areas of emphysema on CT [1]. "density mask" uses a simple thresholding technique where areas with attenuation of less than −910HU correlated closely with the pathologic assessment of emphysema. Table 3 shows the comparison of "density mask" with the meta-classifier approach. Completeness and correctness measures were averaged for the two experiments. The meta-classifier outperforms "density mask" algorithm at the start and end of the co-training process for correctness. While completeness values are high for both algorithms, the "density mask" is slightly better because it covers a wider range of pixels that also includes misclassified instances. Figure 7a, b shows that Meta_MVL is only marginally better at the 12th iteration but outperforms C4.5 and Naive Bayes at early stages of iterations. It is worth noting in Fig. 7a that the meta-learner peaks early and then stabilises. In addition, the results of Meta_MVL were visually validated with the help of a radiologist in the team.

Figure 8a presents an original image where dark areas are emphysematous. The output of the "density mask"' algorithm (Fig. 8b) shows that a lot of noise is picked up along with the emphysema regions. When the degree of emphysema is low, the "density mask" algorithm provides unsatisfactory results, as is the case here. Figure 8c, d

Table 3 Comparison of completeness and correctness measures of the Meta_MVL with "density mask"

	Iteration	Average completeness (of both experiments)	Average correctness (of both experiments)
"Density mask"		98.33	26.33
Meta_MVL	0	95.63	37.00
Meta_MVL	12	93.39	53.40
C4.5	0	97.13	38.13
C4.5	12	93.67	51.00
Naive Bayes	0	96.97	34.96
Naive Bayes	12	93.53	52.30

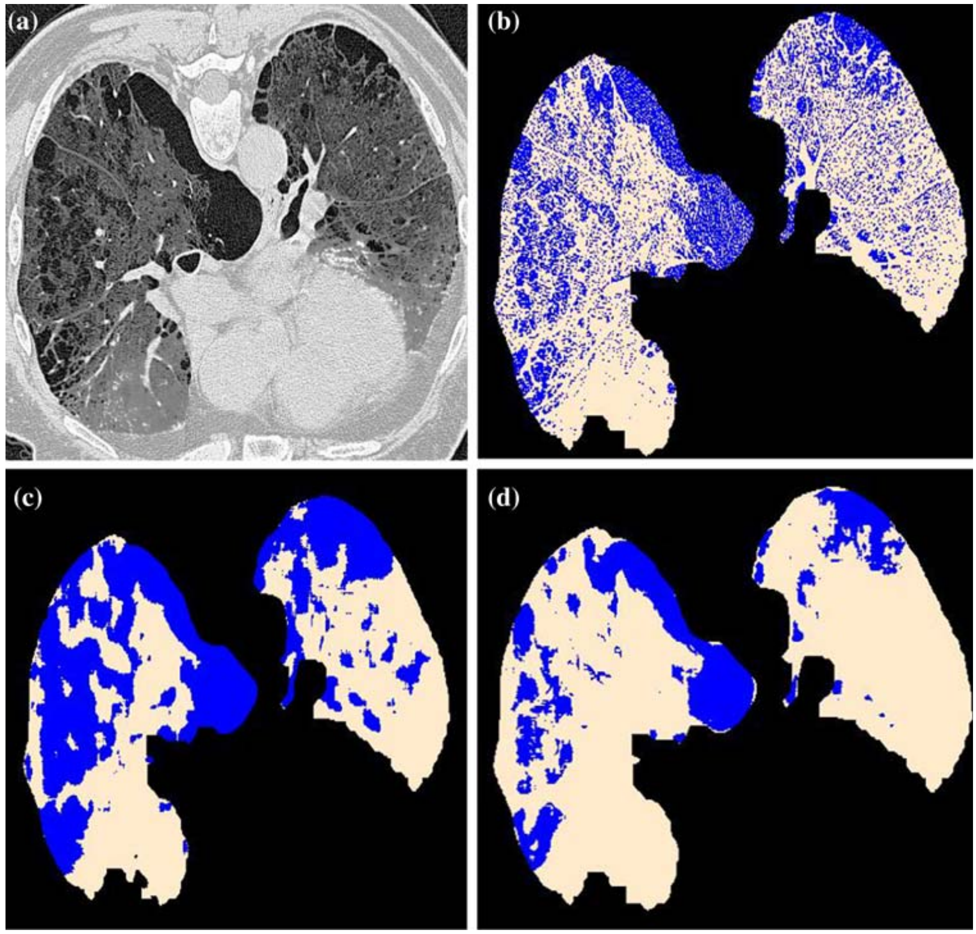


Fig. 8 **a** Contains the original image where the *dark* regions correspond to emphysema. **b** is the output of "density mask''. **c**, **d** correspond to the output of the Meta_MVL at iteration 0 and 12, respectively. The system after 12 iterations classifies only "more confident" regions whereas more marginal regions of emphysema can be classified at iteration 0

shows visual comparison of the output of the Meta classifier at iterations 0 and 12, respectively. The Meta-classifier displays emphysema regions after 12 iterations that could be termed "more confident", while the classifier at iteration 0 covers more emphysema regions, which are more marginal. This has been verified by an expert radiologist in the team who believes that classifiers at different iterations show an interesting correlation with different levels of diagnosis. These levels could vary from marginal emphysema to bullous regions of emphysema. In addition, the misclassified regions also diminish with more iterations as can be seen in Fig. 8d.

Comparison with other well-known techniques for emphysema detection was also performed. Two sets of experiments were carried out: (1) the training was done on the initial training data used for multi-view learning (i.e. 0th iteration) and (2) the training data comprised the initial training data (from the 0th iteration) and the labelled images from the co-training set used for multi-view learning. A variation of K-means clustering technique called seeded K-means was used to do the clustering which has been shown to be a good classifier for emphysema detection [25]. Seeded K-means is a semi-supervised clustering technique that uses labelled data to form initial cluster centers. The two seeds corresponded to emphysema and non-emphysema regions. Independent Component Analysis (ICA) as a feature selection algorithm for classification of emphysema has also been used to compare the performance of multi-view learning. The original features were reduced to 5 ICA components before performing classification using Naive Bayes and C4.5. Also, the technique proposed by Friman et al. [2] that combines

Table 4 Comparison of completeness and correctness measures of the Meta_MVL with "density mask"

	Average completeness (of both experiments)	Average correctness (of both experiments)
Seeded K-means	86.12	47.13
ICA-C4.5	81.27	49.78
ICA-Naive Bayes	82.55	52.51
Error Backpropagation	81.10	48.11
Support vector machine	85.90	49.90
Meta_MVL	93.39	53.40

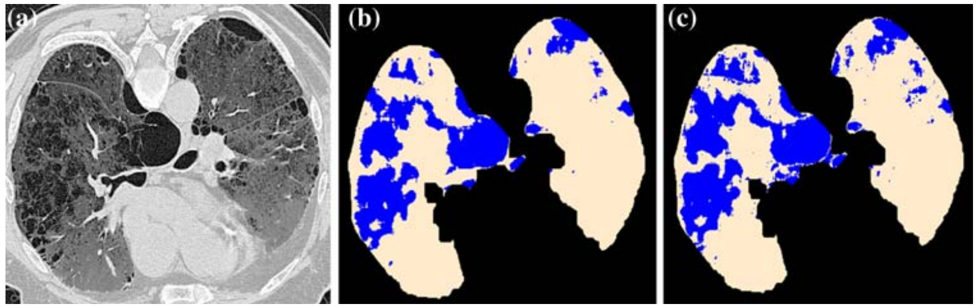


Fig. 9 **a** Contains the original image where the dark regions correspond to emphysema. **b** is the output of the Inc_MVL of our system at iteration 12. **c** Corresponds to the output of the Inc_MVL of the system at iteration 13 after learning from the heuristic based approach

image processing and neural networks also was used for comparison. The image processing technique is based on a filtering method called Normalized Convolution in order to smooth the image slightly. This is followed by a feature extraction where the features are based on texture, shape and global features and classification was performed using an ordinary error backpropagation neural network and support vector machine approach with Gaussian kernels. These techniques were implemented and compared in our work. Table 4 shows comparison of the different classification techniques with Meta_MVL. As can be seen from the table, the average completeness and correctness tends to be higher for Meta_MVL. Intuitively, this makes sense since the Meta_MVL technique is semi-supervised and is exposed to more data in the form of unlabelled examples. By learning from mistakes, the classification accuracy is improved compared to simple classification techniques.

9 Detection of diffuse regions of emphysema

Emphysema is a progressive disease, characterised by abnormal air spaces that are small in the early stages. The small air spaces typically form the diffuse appearance of emphysema. Smaller air sacs are very hard to pick by the human eye due to the diffuse and variable nature of the disease pattern. Hence, manual labelling becomes a very tedious and time-consuming process. To address this issue, a heuristic based filter is used to detect the larger subset of emphysema that includes both diffuse and bullous emphysema.[1]

9.1 Heuristic based filter

The heuristic based filter is used to provide examples that can be used to train the multi-view system incrementally to detect diffuse regions of emphysema. The goal of the heuristic based filter is to detect the larger set of emphysema that includes both marginal and bullous regions of emphysema. It uses a suite of morphology-based techniques along with heuristics applied on top of the existing Density Mask algorithm. The results of the heuristic based filter have been verified in consultation with a radiologist.

9.2 Learning to detect diffuse regions

The output of the Heuristic based filter was used as a labelled image to train Inc_MVL to detect diffuse regions of emphysema since it is difficult for radiologists to label diffuse regions of Emphysema. The classifier at iteration 12 was further trained to detect bullous regions using the heuristic based filter output. At this point, the labelled images from the heuristic based filter were fed incrementally. 12 unlabelled images were run through the heuristic based filter and fed to the meta-classifier in iteration 13. Figure 9a presents the original image where dark areas are emphysematous. The output of Inc_MVL at iteration 12 along with its output after learning from the heuristic approach is shown in Fig. 9b, c, respectively. This approach is useful to radiologists if they require viewing a bullous setting that is slightly more diffuse.

10 Conclusions

In this paper, an approach to perform multi-level diagnosis of emphysema detection based on co-training and active learning in an incremental setting has been presented. By incorporating images labelled by experts incrementally, it is shown that Inc_MVL can perform better in the existing multi-view framework. Informative examples are ones when the views disagree with respect to the label provided by the expert. For these, the correct labels are then provided by images labelled by experts. Results show that when these labelled images are incorporated into the system at early iterations, the performance of the system improves. Results have also been compared against

[1] The heuristic based filter was joint work with Mario Bou-Haidar. Refer to [23] for more details.

"density mask", a standard approach used for emphysema detection in medical image analysis. In general, the density mask method has been known to mark more pixels as emphysematous than warranted, and it has been speculated that many marked pixels do not represent true emphysema. The results of the method proposed here appear to confirm the speculation. Other well-known computerized techniques used for classification of emphysema have also been used for comparison and the results show that by learning from mistakes, classification accuracy can be improved. In addition, it is also shown that the system is able to recognise diffuse emphysema regions by training it with a heuristic based filter from a bullous setting. The heuristic based filter is used as a labelling technique to incorporate images incrementally. This approach has been taken since labelling diffuse regions of emphysema manually is a tedious task because of the variable nature of the radiographic patterns. A system that is capable of differentiating the appearance of emphysema regions has been successfully reported in this paper, which would help experts in the medical setting to analyse the progressive nature of the disease. Recognising and quantifying the different types of emphysema through the use of computer assisted techniques and anatomical knowledge is an area that still needs further exploration.

Acknowledgments This research was partially supported by the Australian Research Council through a Linkage grant (2002–2004), with Medical Imaging Australasia as clinical and Philips Medical Systems as industrial partners.

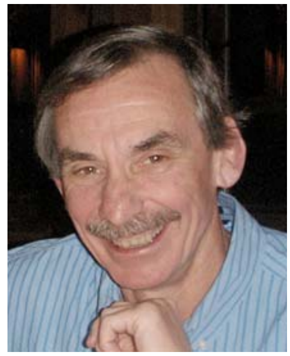
References

1. Kinsella M, Müller NL, Abboud RT, Morrison NJ, DyBuncio A (1990) Quantification of emphysema by computed tomography using a "density mask" program and correlation with pulmonary function tests. Chest 97:315–321
2. Friman O, Borga M, Lundberg M, Tylén U, Knutsson H (2002) Recognizing emphysema—a neural network approach. In: Proceedings of sixteenth international conference on pattern recognition, Quebec, Canada, pp 1–4
3. Heitmann KR, Kauczor H, Mildenberger P, Uthmann T, Perl J, Thelen M (1997) Automatic detection of ground glass opacities on lung HRCT using multiple neural networks. Eur Radiol 9:1463–1472
4. Mir H, Hanmandlu M, Tandon SN (1995) Texture analysis of CT images. IEEE Eng Med Biol Mag 14:781–786
5. Muslea I, Minton S, Knoblock CA (2002) Adaptive view validation: a first step towards automatic view detection. In: Proceedings of the nineteenth international conference on machine learning, Sydney, Australia, pp 443–450
6. Blum, Mitchell T (1998) Combining labelled and unlabelled data with co-training. In: Proceedings of the 1998 conference on computational learning theory, New York, pp 92–100
7. Prasad M, Sowmya A (2004) Multi-level emphysema diagnosis in HRCT lung images through robust multi-view and meta-learning. In: Asian Conference on Computer Vision, Jeju, S. Korea, pp 937–942
8. Prasad M, Sowmya A (2004) Multi-level emphysema diagnosis of HRCT lung images in an incremental framework. In: Proc SPIE Med Imaging, San Diego, USA, vol 5370, pp 42–50
9. Raskutti B, Ferra H, Kowalczyk A (2002) Combining clustering and co-training to enhance text classification using unlabelled data. In: Proceedings of the eighth international conference on knowledge discovery and data mining, Canada, pp 620–625
10. Zelikovitz S, Hirsh H (2001) Using LSI for text classification in the presence of background text. In: Proceedings of the tenth international conference on information and knowledge management. Atlanta, GA, pp 113–118
11. Muslea I, Minton S, Knoblock CA (2002) Active + semi-supervised learning = robust multi-view learning. In: Proceedings of the nineteenth international conference on machine learning, Sydney, Australia, pp 435–442
12. Kushmerick N, Johnston E, McGuinness S (2001) Information extraction by text classification. IJCAI-01 Workshop on Adaptive Text Extraction and Mining, Seattle, Washington
13. Nahm U, Mooney R (2000) A mutually beneficial integration of data mining and information extraction. In: Proc Seventeenth Nat Conf Artif Intell, Austin, TX, pp 627–632
14. Chiu PT, Sowmya A (2001) Lung boundary detection and low level feature extraction and analysis from HRCT images. VISIM: Information Retrieval and Exploration of Large Medical Image Collections, Utrecht, Netherlands
15. Muslea I (2002) Active learning with multiple views. Dissertation
16. Baum EB, Lang KE (1991) Constructing hidden units using examples and queries. In: Lippmann RP et al. (eds) Advances in neural information processing systems. Morgan Kaufmann, San Mateo, pp 904–910
17. Witten IH, Frank E (1999) Data mining: practical machine learning tools and techniques with Java implementations. Morgan Laufmann, San Fransisco
18. Milton JS, Arnold JC (1995) Introduction to probability and statistic*s*, principles and applications for engineering and the computing sciences, 3rd edn, McGraw-Hill International, New York
19. Mitchell T (1997) Machine learning. McGraw Hill, New York
20. Quinlan JR (1986) Induction of decision trees. Mach learn 1:81–106
21. Nigam K, McCallum A, Thrun S, Mitchell T (2000) Text classification from labelled and unlabelled documents using EM. Mach Learn 39:103–134
22. Bowyer K, Phillips P (1998) Overview of work in empirical evaluation of computer vision algorithms. In: Bowyer K et al. (eds) Empirical evaluation techniques in computer vision, IEEE Computer Society Press, CA, pp 1–11
23. Bou-Haidar M, Prasad M, Sowmya A (2004) Emphysema detection using a density mask. UNSW Computer Science and Engineering Technical Report no. UNSW-CSE-TR-0421
24. Uppaluri R, Hoffman E, Sonka M, Hartley P Hunninghake G, McLennan G (1999) Computer recognition of regional lung disease patterns. Am J Respir Crit Care Med 160:648–654
25. Prasad M, Sowmya A (2004) Multi-class unsupervised classification with label correction of HRCT lung images. In: Proceedings of international conference on intelligent sensing and information processing, Chennai, India, pp 51–55
26. Prasad M, Sowmya A, Koch I (2004) Feature subset selection using ICA for classifying emphysema in HRCT images. In: Proceedings of international conference on intelligent sensors, sensor networks and information processing, Melbourne, Australia, pp 515–519
27. Bertelli L, Cucchiara R, Paternostro G, Prati A (2006) A semi-automatic system for segmentation of cardiac M-mode images. Pattern Anal Appl 9:293–306

Author Biographies

Mithun Prasad received his PhD from the University of New South Wales, Sydney, Australia in 2006. He was a postdoctoral scholar at the University of California, Los Angeles and now a research associate at Rensselaer Polytechnic Institute, NY. His research interests are computer aided diagnosis, cell and tissue image analysis.

Arcot Sowmya is a Professor, School of Computer Science and Engineering, UNSW, Sydney. She holds a PhD degree in Computer Science from Indian Institute of Technology, Bombay, besides other degrees in Mathematics and Computer Science. Her areas of research include learning in vision as well as embedded system design. Her research has been applied to extraction of linear features in remotely sensed images as well as feature extraction, recognition and computer aided diagnosis in medical images.



Peter Wilson is a clinical Radiologist at Pittwater Radiology in Sydney. He was trained at Royal North Shore Hospital and taught Body Imaging at the University of Rochester, NY, prior to taking up his current position.