

RANKING AND CLASSIFICATION OF MONOTONIC EMPHYSEMA PATTERNS WITH A MULTI-CLASS HIERARCHICAL APPROACH

Sila Kurugol, George R. Washko, Raul San Jose Estepar,

Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA

ABSTRACT

Emphysema has distinct and well-defined visually apparent CT patterns called centrilobular and panlobular emphysema. Existing studies concentrated on the classification of these patterns but they have not looked at the complete evolution of this disease as the destruction of lung parenchyma progresses from normal lung tissue to mild, moderate, and severe disease with complete effacement of the lung architecture. In this paper, we discretize this continuous process into five classes of increasing disease severity and construct a training set of 1161 CT patches. We exploit three solutions to this monotonic multi-class classification problem: a global rankSVM for ranking, hierarchical SVM for classification and a combination of these two, which we call a hierarchical rankSVM. Results showed that both hierarchical approaches were computationally efficient. The classification accuracies were slightly better for hierarchical SVM. However, in addition to classification, ranking approaches also provided a ranking of patterns, which can be utilized as a continuous disease progression score. In terms of the classification accuracy and ratio of pair-wise constraints satisfied, hierarchical rankSVM outperformed the global rankSVM.

Index Terms— emphysema, COPD, multi-class classification, rankSVM

1. INTRODUCTION

Emphysema is the progressive destruction of the lung leading to a permanent dilation of the distal airspaces. A common CT based quantification technique of emphysema is dichotomization of lung parenchyma into emphysema and non-emphysema regions with a threshold in Hounsfield units. Despite the wide acceptance of this approach due to correlation with histopathology and clinical outcomes, it has multiple drawbacks like sensitivity to noise and imaging parameters.

The smallest physiologic subunit of the lung is the secondary pulmonary lobule (SPL) that includes airways, arteries, veins, lymphatics, and the lung interstitium. The pattern of tissue damage apparent in this structure is used by radiologists to classify disease type and rate its severity [1]. Em-

physema has two visually apparent patterns called centrilobular and panlobular emphysema. While panlobular disease can be described as complete effacement of the lung architecture, centrilobular emphysema is characterized by varying degrees of preservation of the structure of the SPL [1]. Existing techniques classify emphysema into one of these patterns often with a K Nearest Neighbor (KNN) or a Support Vector Machine (SVM) classifier. They use features such as vectors [2] or histograms [3, 4] of intensities, local binary patterns [5], wavelet transform [6], or texton signatures [7]. However, existing works have not studied the complete evolution of disease as it progresses from normal parenchyma to mild, moderate, severe centrilobular and severe panlobular emphysema. (Fig. 1). An automated approach for both emphysema classification and gradation will therefore be of great interest.



(a) Normal (b) Mild (c) Moderate (d) Severe (e) Panlob.

Fig. 1: Two sample patches per class are shown. The patch sizes are $24.18 \times 24.18 \text{ mm}^2$, the size of an SPL. CT intensity window was set to $[-1000; -500]$. Severity increases from (a) to (e) with increasing size of low attenuating regions.

This problem differs from a standard classification problem because of the inherent natural ordering of not only the classes, but also the patterns within a single class. It differs from a regression problem because it is difficult to assign a continuous score of disease severity for each pattern by visual evaluation. Clinician only assigns class labels to patterns. We therefore exploit three solutions: 1) A ranking approach (RankSVM [8, 9]) where the intrinsic class orderings of pairs of patterns are included as constraints into the SVM objective function, 2) hierarchical multi-class classification with binary SVM (H-SVM), 3) combination of 1) and 2), which we call hierarchical RankSVM (H-RankSVM). In the first approach, rankSVM provides a single global ordering of the patterns, while in the second approach, a hierarchical multi-class SVM classification is employed. The hierarchical approach considers the inherent class orderings when grouping the classes into two subsets at each level. This reduces the number of possible tree combinations and makes it computationally more efficient to select the optimal tree. The third approach combines the advantages of first two approaches. Compared to the global rankSVM, it provides a more accurate inter-class rank-

This work was funded by 1R01HL116931-01 and the COPDGene study NHLBI grants 2R01HL089897-06A1 and 2R01HL089856-06A1. Additional support provided by NIH grants K25 HL104085-04.

ing, since it works more locally concentrating on the boundary between two groups of classes at a time at each level of the hierarchy. Compared to H-SVM, it provides extra information: it provides both a classification and also a continuous ranking of patterns that can be utilized as a disease progression score. In this paper, we compare these three approaches.

We perform nested cross-validation experiments in a data set of 1161 emphysema image patches with the size of a secondary pulmonary lobule obtained from 267 COPD subjects. We compare the results of suggested approaches with each other and with standard multi-class classification techniques.

2. METHODS

We first describe the global rankSVM approach that provides not only a classification but also an ordering of the disease patterns. We then explain the hierarchical multi-class classification solution. We describe how we efficiently build the optimal binary classifier tree and how we train each binary SVM classifier in this tree while optimizing for the parameters with a nested cross validation approach. We finally explain our hierarchical rankSVM approach where binary SVMs at each node are replaced with rankSVMs.

2.1. RankSVM

Ranking was first used in literature to solve information retrieval tasks that requires ranking the relevancy of documents for a query [10]. Unlike classification, ranking requires classes (or rankings) to have an ordering. RankSVM defines a function $f(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x}$, such that $f(\mathbf{x}_i) > f(\mathbf{x}_j) \iff \mathbf{c}_i > \mathbf{c}_j$. Combining these equations, we get $\mathbf{w} \cdot \mathbf{x}_i > \mathbf{w} \cdot \mathbf{x}_j$. This is equivalent to standard classification where difference of pairs of data points ($\mathbf{x}_i - \mathbf{x}_j$) and the sign of difference of the labels of those points are fed into the classifier as data points and labels respectively. The vector \mathbf{w} is then learned using a standard SVM learning method. Since the inputs are pairs of data points, the problem size increases quadratically with the size of the training set. Efficient algorithms have recently been proposed to solve this problem [11].

We apply this efficient rankSVM formulation [11] to our problem in order to obtain a global ranking of patterns. We use the inherent ordering of classes from increasing severity levels as constraints when we train the ranking function $f(x)$. This function is then used to rank the test patterns. One approach to evaluate the performance is to calculate the ratio of the constraints satisfied in the test set. Another approach is converting the rankings to class labels and computing the classification error. We compute four thresholds over the ranking function to separate the ranked patterns into five classes. Thresholds are selected to minimize the classification error over the validation set.

2.2. Hierarchical Multi-Class Classification for Monotonic Classes

The binary classifiers such as SVM [12] can be extended to multi-class classification problems using techniques like

one-against-rest, one-against-one and hierarchical classification [13]. Hierarchical classification is efficient in terms of number of binary classifiers needed. Moreover, our problem has a special structure with monotonic classes from progressing disease levels (see Fig. 2(b)). This structure can be utilized to learn the optimal binary classification tree very efficiently. The binary classification tree (Fig. 2(a)) subdivides the set of classes into two subsets at each node. The division stops at the leaf nodes where each subset contains a single class. To classify a new pattern, a path is followed from the parent node to one of the leaf nodes of the tree according to the binary classifier decision at each node.

To build the optimal tree, it is necessary to decide how to partition the classes into two subsets at each node. This can be performed by comparing all the possible trees and choosing the optimal one according to a criterion. However, this approach is computationally expensive. A greedy approach that computes the best subsets at each node based solely on the criterion evaluated at that node can be used. This approach requires a small number of comparisons to build the tree, however it is not optimal. Instead, we use the special structure of the emphysema classification problem with monotonic class distributions to reduce the comparison at each node from $2^{k_i-1} - 1$ to $k_i - 1$ where k_i is the number of classes at that node i . For instance, at the first node there are five classes (1, 2, ..., 5) and 4 possible ways to subdivide these classes into two subsets. The location of these cut points can be only between class $k-1$ and k ($\{\{1\}, \{2, 3, 4, 5\}\}, \{\{1, 2\}, \{3, 4, 5\}\}, \{\{1, 2, 3\}, \{4, 5\}\}, \{\{1, 2, 3, 4\}, \{5\}\}$). With this approach, we limit the number of comparisons at each node. Therefore it is possible to evaluate all possible combinations of trees and select the optimal tree. For our problem with five classes, there are 14 possible trees. We train each possible classifier tree and select the optimal one based on the maximum accuracy criterion computed over the validation set.

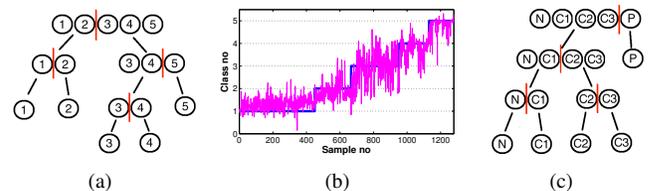


Fig. 2: (a) Multi-Class Hierarchical Classification Tree. (b) Regression plot showing the monotonicity of classes with increasing disease progression levels from 1 to 5, Normal to Panlobular. Blue line is the true class labels and the pink line is the output when we fitted a regression line to the features and the true class labels. (c) Optimal Multi-Class Hierarchical Classification Tree. The experiments resulted in the same optimal tree for H-RankSVM and H-SVM.

2.3. Classification at Each Node

Previous work on emphysema classification mostly used KNN classifier, however, KNN only provides a local decision irrespective of the global information in the training set. In-

stead we use either a rankSVM or a binary SVM classifier at each tree node.

Binary SVM: Binary SVM uses the training samples to learn the optimal separating hyperplane ($f(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x}$) with the orientation that maximizes the classifier margin ($\frac{2}{\|\mathbf{w}\|}$). For samples that are not linearly separable, kernel SVMs are used. We use kernel SVM with Radial Basis Function kernel ($k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma\|\mathbf{x}_i - \mathbf{x}_j\|^2)$) since it works well in our application and compare the results with linear SVM.

RankSVM: The alternative approach we propose to standard hierarchical SVM is hierarchical rank SVM where a binary rankSVM is applied at each node to rank the patterns. Unlike global rankSVM, hierarchical rankSVM provides a local ranking at each node. We then combine these piecewise linear local rankings into a final global ranking as illustrated in Fig. 3. To do so, we define a piecewise linear map between the median values of the rankings for each class computed over the training set.

2.4. Features

The classifiers work on features extracted from image patches of size 31×31 pixels ($24.18 \times 24.18 \text{ mm}^2$). The size of patches are selected as the average size of a secondary pulmonary lobule. Our feature set is the kernel density estimate (KDE) of intensity pdfs [4]. For KDE, a standard Gaussian kernel with bandwidth parameter σ was used and the parameter was estimated using the method in Botev et al. [14]. We extracted $d = 601$ features: The first 600 density values in the range $[-1050; -450]$ and the last feature computed as the sum of the density over all HU values larger than -450 .

2.5. Experimental Results

Data set: We utilized 1161 image patches labeled by an expert clinician in our experiments. The samples were selected from a group of 267 subjects. The number of samples for each class in the order of increasing progression levels was: NT=370, C1=287, C2=178, C3=178, P=148. The expert labeled four to six samples per patient at random based on prototypic expression of disease and without any prior spatial correlation.

Cross Validation: We used nested cross validation experiments. The data was first divided into training and test sets using 10-fold cross validation such that all patches from a single subject fell in either training or test set, but not both. The training set was then further divided into validation and training sets using 5-fold cross validation. The training, test and validation sets were all independent. We used a grid search over the validation set to find the optimal parameters that gave the best classification performance. F-score [15] was used to measure classification performance since it balances the classification errors from negative and positive classes.

In our experiments we computed the optimal tree hierarchy for the hierarchical binary SVM and rankSVM using

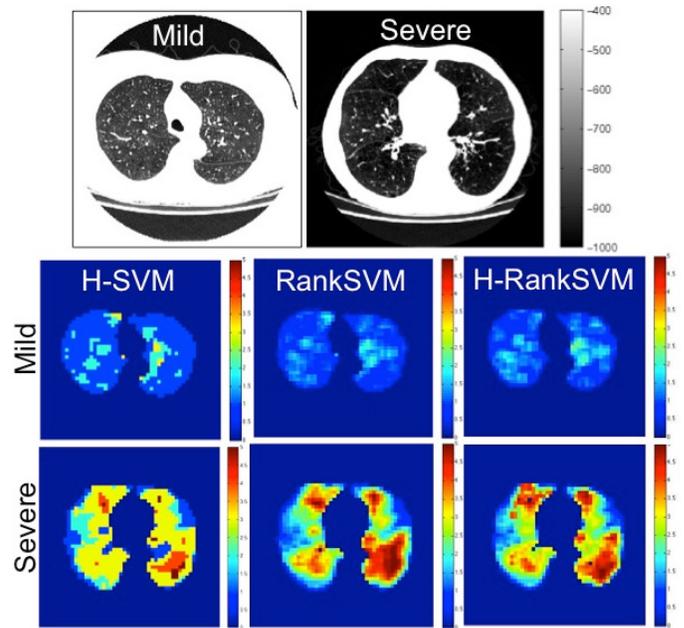


Fig. 3: First row shows two CT slices from two subjects. Left one has mild disease and right one has severe disease. Second and third rows show the results of the methods for mild and severe disease slices respectively. The columns show the results of H-SVM, RankSVM and H-RankSVM respectively. While H-SVM only provides discrete class labels from 1 (normal) to 5 (severe panlobular), rankSVM and H-rankSVM provides a continuous map in the same range.

training set success rate criterion. We used a brute-force search over all possible trees (14 trees). We obtained the same optimal tree for both H-SVM and H-RankSVM. For the rest of the experiments, we used the tree shown in Fig. 2 (c). Note that in the 10 fold cross validation experiments, most folds of the training set resulted in the same tree, which we use in the results we report.

We evaluated the performance of global rankSVM classifier, H-SVM classifier and H-RankSVM classifier and compared them against standard one-against-one, one-against-rest SVM, Naive Bayes and KNN classifiers that were used in previous work [4, 5]. In the KNN classifier, the number of nearest neighbors was set to the optimal value 5 reported in the previous work. For H-SVM classifiers, the results of both linear and kernel versions are reported in Table 1. The proposed kernel H-SVM method outperformed all classifiers and achieved comparable performance with one-against-one classifier. However, during testing, in one-against-one SVM, $n(n-1)/2$ binary classifiers are applied to each sample and the decision is made by majority voting. In hierarchical SVM, the number of classifiers that are needed to be applied to a sample is fewer. In our case, with five classes, one-against-one classifier applied 10 binary SVMs, while the hierarchical version only applied 4 while achieving the same performance. As expected, H-SVM achieved slightly better classification

accuracy compared to H-RankSVM. However, H-RankSVM additionally provides an intra class ranking of the patterns and a continuous disease progression map. We also computed the ratio of correct inter-class pair-wise orderings, and H-RankSVM outperformed global rankSVM, with values of 0.86 and 0.72 respectively.

Method	Mean		Classification Success Rate
	Sens.	Spec.	
RankSVM	0.598	0.874	0.64
H-RankSVM	0.665	0.896	0.69
H-SVM	0.669	0.905	0.70
H-SVM + Kernel	0.694	0.914	0.71
1-against-all	0.610	0.904	0.67
1-against-all + Kernel	0.650	0.904	0.68
1-against-1	0.646	0.898	0.68
1-against-1 + Kernel	0.694	0.916	0.71
N. Bayes	0.602	0.834	0.55
KNN	0.656	0.892	0.69

Table 1: Mean sensitivity and specificity metrics derived from the confusion matrices and averaged over five classes and classification success rates.

Fig. 3 compares the results of all three methods on CT images of two smokers, one with mild and the other with severe disease. H-SVM approach provides only discrete class labels from 1 (normal) to 5 (severe panlobular) for each patch. RankSVM and H-RankSVM provides a continuous map in the same range, which can be utilized as a disease progression score. Moreover, the expert visual evaluation of these slices agreed well with H-RankSVM results.

3. DISCUSSIONS AND CONCLUSIONS

In this paper, we presented one global ranking approach and two hierarchical approaches, one for classification, and one for ranking. These approaches took advantage of the progressive nature of the disease. RankSVM learned a global ranking function that satisfied the pairwise constraints from the training set, while H-RankSVM learned a local ranking function for each node of the hierarchy, which were later combined to provide a global ranking. H-SVM provided multi-class hierarchical classification. We used the monotonic relation between patterns that reflects the disease progression to limit the number of comparisons carried out when constructing the optimal trees. We compared the performance of these approaches with standard one-against-one and one-against-rest SVM approaches as well as with KNN classifier.

The H-SVM approach outperformed KNN and the other multi-class SVM approaches and had the same accuracy with one-against-one approach. However, the one-against-one approach required 14 binary classifications while the H-SVM required only 4 reducing the computational complexity.

H-SVM achieved slightly higher accuracies compared to

both rankSVM and H-RankSVM as expected since it optimizes this criterion. However, the ranking approaches additionally provides an intra class ranking for each pattern. The H-RankSVM approach outperformed global rankSVM in terms of both classification accuracy and the ratio of inter-class pairwise constraints satisfied. It also agreed well with visual expert evaluation.

4. REFERENCES

- [1] DM Hansell et. al, "Fleischner society: Glossary of terms for thoracic imaging," *Radiology*, vol. 246, pp. 697–722, 2008.
- [2] T. Zulueta-Coarasa, S Kurugol, JC Ross, G Washko, and R San Jose Estepar, "Emphysema classification based on embedded probabilistic PCA," in *EMBC, 2013*, 2013.
- [3] R Uppaluri et. al, "Quantification of pulmonary emphysema from lung CT images," *American journal of respiratory and critical care medicine*, vol. 156, no. 1, pp. 248–254, 1997.
- [4] CS Mendoza, GR Washko, JC Ross, AA Diaz, DA Lynch, JD Crapo, EK Silverman, B Acha, C Serrano, and R San José Estepar, "Emphysema quantification in a multi-scanner HRCT cohort using local intensity distributions," in *ISBI*, 2012.
- [5] L Sorensen, SB Shaker, and M De Bruijne, "Quantitative analysis of pulmonary emphysema using local binary patterns," *IEEE Trans. on Med. Imag.*, vol. 29, no. 2, pp. 559–569, 2010.
- [6] A Depeursinge, A Foncubierta-Rodriguez, D Van de Ville, and H Müller, "Multiscale lung texture signature learning using the riesz transform," *MICCAI*, pp. 517–524, 2012.
- [7] M Gangeh, L Sørensen, S Shaker, M Kamel, and M de Bruijne, "Multiple classifier systems in texton-based approach for the classification of CT images of lung," *Medical Computer Vision*, pp. 153–163, 2011.
- [8] T Joachims, "Optimizing search engines using clickthrough data," in *Proc. of ACM SIGKDD*, 2002, pp. 133–142.
- [9] F Pedregosa, Gramfort A, G Varoquaux, E Cauvet, C Pallier, and B Thirion, "Learning to rank from medical imaging data," *Machine Learning in Medical Imaging*, pp. 234–241, 2012.
- [10] Y Cao, J Xu, TY Liu, H Li, Y Huang, and HW Hon, "Adapting ranking SVM to document retrieval," in *Proc. of the ACM SIGIR conference*, 2006.
- [11] O Chapelle and SS Keerthi, "Efficient algorithms for ranking with SVMs," *Info. Retrieval*, vol. 13, pp. 201–215, 2010.
- [12] C Cortes and V Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [13] CW Hsu and CJ Lin, "A comparison of methods for multiclass support vector machines," *IEEE Trans. on Neural Networks*, vol. 13, no. 2, pp. 415–425, 2002.
- [14] ZI Botev, JF Grotowski, and DP Kroese, "Kernel density estimation via diffusion," *The Annals of Statistics*, vol. 38, no. 5, pp. 2916–2957, 2010.
- [15] M Sokolova, N Japkowicz, and S Szpakowicz, "Beyond accuracy, f-score and roc: a family of discriminant measures for performance evaluation," *Advances in Artificial Intelligence*, pp. 1015–1021, 2006.