Classification of Lung Disease in HRCT Scans using Integral Geometry Measures and Functional Data Analysis

Dr Elke Thönnes
Department of Statistics and Centre for Scientific Computing
University of Warwick
Coventry, UK

Dr Abhir Bhalerao
Department of Computer Science
University of Warwick
Coventry, UK

Dr David Parr
Coventry and Warwickshire University Hospitals, Coventry
and Medical School, University of Warwick
Coventry, UK

Abstract

A framework for classification of chronic lung disease from high-resolution CT scans is presented. We use a set of features which measure the local morphology and topology of the 3D voxels within the lung parenchyma and apply *functional* data classification to the extracted features. We introduce the measures, Minkowski functionals, which derive from integral geometry and show results of classification on lungs containing various stages of chronic lung disease: emphysema, fibrosis and honey-combing. Once trained, the presented method is shown to be efficient and specific at characterising the distribution of disease in HRCT slices.

1 Introduction

Chronic obstructive lung diseases, such as emphysema and pulmonary fibrosis, are progressive respiratory diseases leading to a decline in lung function and, eventually, respiratory failure. High resolution computed tomography (HRCT) is currently the most accurate, non-invasive method of detecting and evaluating changes in lung parenchyma and so is used in clinical practice to diagnose and assess the severity of these diseases. The progress of disease manifests itself as textural changes of the imaged lung tissue and, in the case of emphysema, as a reduction of the mean lung density. In clinical practise, quantitative assessment of emphysema is usually based on summary statistics of the histogram of lung voxel values. Common statistics are percentile points and the voxel index, which is the proportion of lung voxels below a pre-defined threshold. However, such simple criteria can be unreliable, for example, when the pathological process is mixed such as in the presence of inflammation and fibrosis along with emphysema.

As chronic obstructive lung diseases express themselves as textural changes in HRCT scans, image texture analysis methods have been investigated as tools for a more robust quantification. The early work of Uppaluri *et al.* [10] used first and second order texture features together with fractal dimension to characterise emphysema. Chabat *et al.* [4] proposed the use of texture measures based on grey-level co-occurrence matrices (GLCM). They also included a number of shape features. The use of GLCM and the related shape and connectivity measures suggested by Chabet *et al.* are also used in more recent work by Xu *et al.* [11]. In Hoffman *et al.* [7] and Zavaletta *et al.* [12] spatial maps of lungs have been produced that graphically show the result of texture classification. The classifiers used in recent work vary from kNN classification [12], to Bayesian [11] and neural-network based methods [6]. In the following we explore the use of a systematic framework of morphological descriptors from integral geometry that integrates and extends current approaches based on histogram analysis.

2 Characterising Texture using Integral geometry

Integral geometry provides a family of measures, *intrinsic volumes* or *Minkowski functionals*, that characterise the morphology and topology of polyconvex sets. A set is polyconvex if it is a finite union of compact, convex sets. The foreground in a binary image is a finite union of voxels and therefore a polyconvex set.

In $\mathbb{R}^3$ there are four Minkowski functionals which are proportional to more commonly known quantities: volume V, surface area S, the mean breadth B (which is proportional to the integral of mean curvature) and the Euler-Poincaré characteristic χ. More formally, Steiner's formula can be used to define the Minkowski functionals for convex sets. In $\mathbb{R}^3$ Steiner formula shows that

$$V(K \oplus B_r) \quad = \quad V(K) + S(K)\, r + 2\pi B(K)\, r^2 + \frac{4}{3}\chi(K)\, r^3, \tag{1}$$

where $K \oplus B_r = \{x+y, x \in K \text{ and } y \in B_r\}$ is the dilation of the convex set K by a sphere of radius r. Hence we have

$$W_0(K) = V(K), \qquad W_1(k) = \frac{1}{3}S(K), \qquad W_2(K) = \frac{2\pi}{3}B(K), \qquad W_3(K) = \frac{4}{3}\chi(K). \tag{2}$$

In the following we will use the measures V, S, B and χ but refer to them collectively as Minkowski functionals.

Minkowski functionals have convenient mathematical properties. They are invariant to rigid motion, they are continuous in a certain sense and they are additive which allows for efficient computation of the functionals, see Section 3. Another important property is that Minkowski functionals are complete in the sense of Hadwiger's theorem: any image functional that is motion-invariant, continuous and additive is a linear combination of Minkowski functionals.

Having been shown to be effective in materials science for classification of two-phase media (see for example [1]), more recently, the functionals have also been applied to HRCT lung scans. In [2, 3] the measures are used to classify regions of interest of size 40x40x40 voxels into healthy, fibrotic and emphysematous tissue. The authors use an integrative filtering procedure to summarize the four integral geometry measures into a single quantity used for classification and provide evidence that these measures can outperform standard densitometric methods. In [9] the same measures are computed on sub-windows covering the whole lung and calibrated maps are produced illustrating the spatial distribution of emphysema and pulmonary fibrosis.

To apply the Minkowski functionals to HRCT scans, the images have to be binarized via thresholding. While [9] keep the threshold fixed, in this paper we extend the approach by letting the threshold vary and computing the Minkowski functionals for each threshold. This is a natural extension to the clinical densitometry approach as the zeroth order Minkowski functional (volume) as a function of attenuation is equivalent to the empirical cumulative distribution function of voxel values. In contrast to [2, 3] who examine large regions of interest (40x40x40 voxels) we consider small sub-windows (5x5x1 voxels). While this sacrifices some of the separation between Minkowski functionals for the different disease classes it allows us to localize the spatial distribution of disease. Also, rather than summarizing the measures into a single numerical quantity, we take account of the fact that we have functions of Minkowski functionals by using a classification technique from functional data analysis. This overcomes the statistical issues posed by having a large number of predictors relative to the sample size.

3 Computation of Measures and Functional Classification

Legland *et al.* [8] present an algorithm that computes Minkowski measures for 3D binary images. These Minkowski measures $m_i, i = 0,\ldots,3$, are local versions of Minkowski functionals $W_i, i = 0,\ldots,3$, and are computed on voxel level. The Minkowski functional of a set X restricted to a window Ω can then be computed by summing over the Minkowski values of the voxels in the window, that is $W_i(X \cap \Omega) = \sum_{x\in\Omega} m_i(x)$. The algorithm is implemented as a marching cubes type algorithm with hhe Minkowski measure for any 3x3x3 binary voxel configuration being pre-computed and kept in a look-up table. Linear filtering is then used to identify the voxel configuration for a given marching cube.

To apply the algorithm to HRCT lung scans we binarised the lung images for a range of different thresholds. We subdivided each binary image into a grid of windows of size 5x5x1 voxels and then computed the Minkowski functionals for each window. Thus we obtained Minkowski functionals as functions of attenuation value, see Figure 1 for examples. When increasing the window size the separation between the Minkowski functionals of different disease classification becomes more pronounced, however, spatial localisation of the classification into disease states becomes reduced. In our experiments we used the non-parametric supervised classification method by Ferraty and Vieu [5] to classify various diseased lung tissue. Our data represents functions which can pose problems for standard multivariate techniques such as classical linear discriminant analysis because we have a large number of predictors relative to the sample size. Functional data analysis, like the method by Ferraty and Vieu, can overcome these limitations. We give a short overview of the method, further details can be found in [5].

Let $\mathbf{w} = \{w_0, w_1, w_2, w_3\}$ be the observed Minkowski functionals which are functions of attenuation value and let C denote the categorical variable denoting the class. The classification method uses a Bayes rule as classification rule, that is it estimates the posterior probabilities

$$p_c(\mathbf{w}) = \mathbb{P}\Big(C = c \Big| \mathbf{W} = \mathbf{w}\Big) = \mathbb{E}\Big(\mathbf{1}_{[C=c]} \Big| \mathbf{W} = \mathbf{w}\Big) \tag{3}$$

and then assigns the class c with the highest estimated posterior probability to $\mathbf{w}$. The posterior probabilities are estimated using a kernel estimator which is based on a functional concept of proximity. In our experiments we used the L_2 metric applied to data smoothed via multivariate partial least squares regression.

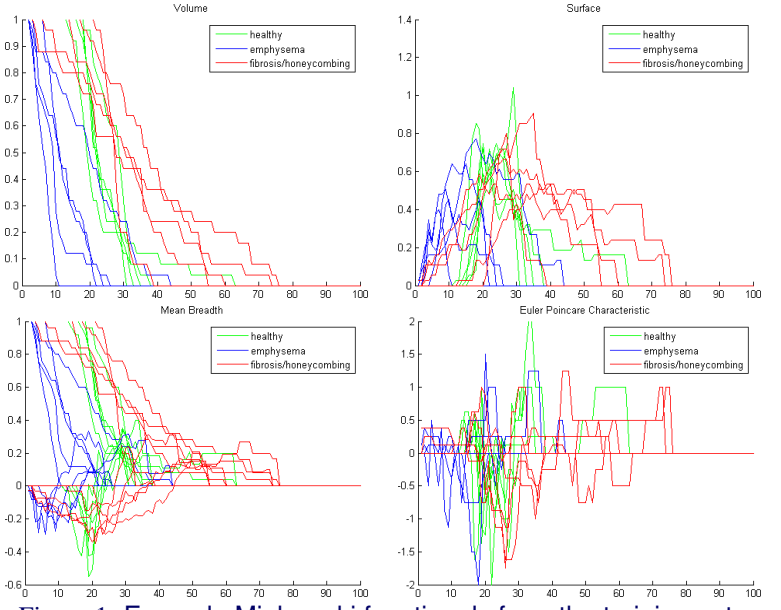

Figure 1: Example Minkowski functionals from the training set.

4 Results and Discussion

We manually selected 28 regions of size 30x30x5 voxels as a training set. 10 of these regions were examples of healthy lung, 10 fibrotic lung and 8 suffering from emphysema. (The number of emphysematous training regions is smaller as there were fewer such regions in the available HRCT scans). Both the training set and the lung tissue to be classified were subdivided into 5x5x1 voxel windows. While we choose the vertical size of classified regions to be just one voxel note that the Minkowski functional values each voxel take account of the configuration in the 3D neighbourhood. On the training set we achieved a misclassification rate of 7%.

Figure 2 illustrates some of our classification results. We should point out that the training set was classified by a non-expert and is likely to underestimate the variability of the disease patterns. Also, as the results are not based on a designed survey sample. Thus, at this stage, the results are indicative only but illustrate the promise of the presented method. The resolution of classification was chosen as a compromise between spatial localisation and appropriate separation in the training sample. However, the additivity of the Minkowski measures presents the opportunity to develop a hierarchical classification approach in which the resolution in a spatial region can be varied and is determined as part of the algorithm.

In summary, the classification method from functional data analysis has been shown to appropriately account of the fact that the integral geometry descriptors are functions of attenuation. Furthermore, while enhancing the assessment of emphysema the presented framework could readily be applied to any other lung disease that leads to structural changes and deformations of lung tissue, including pulmonary fibrosis and LAM. As well as further validation of the framework, we are investigating its wider application.

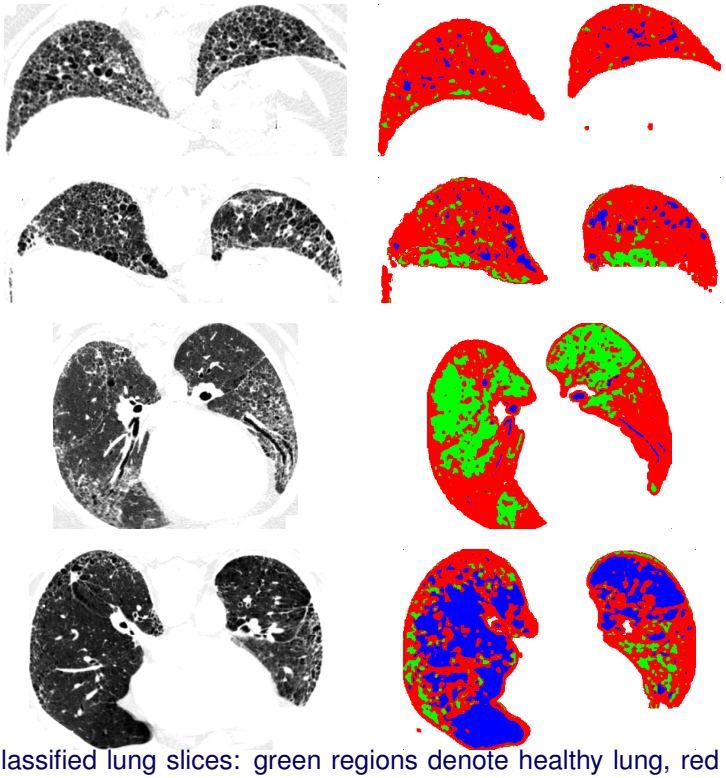

Figure 2: Classified lung slices: green regions denote healthy lung, red regions fibrotic, honeycombed tissue and blue regions are classified as suffering from emphysema.

References

[1] C.H. Arns, M.A. Knackstedt, and K.R. Mecke. *Characterising the Morphology of Disordered Materials*, pages 37–74. Lecture Notes in Physics 600. Springer, 2002.

[2] H.F. Boehm, C. Fink, U. Attenberger, C. Becker, J. Behr, and M. Reiser. Automated classification of normal and pathologic pulmonary tissue by topological texture features extracted from multidetector CT in 3D. *European Radiology*, 18, 2008. DOI: 10.1007/s00330-008-1082-y.

[3] H.F. Boem, C. Fink, C. Becker, and M. Reiser. Automated Characterization of Normal and Pathologic Lung Tissue by Topological Texture Analysis of Multi-Detector CT. In Giger and Karssemeijer, editors, *Medical Imaging 2007: Computer-Aided Diagnosis, Proceedings of SPIE Vol 6514*, page DOI: 10.117/12.702697, 2007.

[4] F. Chabat, G. Yang, and D. Hansell. Obstructive lung diseases: Texture classification for differentiation at CT. *Radiology*, 228:871–877, 2003.

[5] Frédéric Ferraty and Philippe Vieu. *Nonparametric Functional Data Analysis*. Statistics. Springer, New York, 2006.

[6] O. Friman, M. Borga, M. Lundberg, U. Tylén, and H. Knutsson. Recognizing Emphysema - A Neural Network Approach. In *Proceedings of 16th Int. Conf. on Pattern Recognition - ICPR'02*, page 10512, 2002.

[7] E. A. Hoffman, J. M. Reinhardt, M. Sonka, B. A. Simon, J. Guo, O. Saba, D. Chon, S. Samrah, J. Shikata, J. Tschirren, K. Palagyi, K. Beck, and G. McLennan. Characterization of the Interstitial Lung Diseases via Density-Based and Texture-Based Analysis of Computed Tomography Images of Lung Structure and Function. *Academic Radiology*, 10:1104–1118, 2003.

[8] D. Legland, K. Kiêu, and M.-F. Devaux. Computation of Minkowski measures on 2D and 3D binary images. *Image Analysis and Stereology*, 26:83–92, 2007.

[9] M. Charemza and E. Thönnes and A. Bhalerao and D. Parr. Integral Geometry Descriptors for Characterizing Emphysema and Lung Fibrosis in HRCT Images. In *First International Workshop on Pulmonary Image Processing (MICCAI 2008)*, pages 155–164, 2008.

[10] R. Uppaluri, M. Theophano, M. Sonka, E. A. Hoffman, and G. McLennan. Quantification of Pulmonary Emphysema from Lung Computed Tomography Images. *American Journal of Respiratory and Critical Care Medicine*, 156:248–254, 1997.

[11] Y. Xu, M. Sonka, G. McLennan, and E. A. Hoffman. MDCT-based 3-D texture classification of emphysema and early smoking related lung pathologies. *IEEE Transaction on Medical Imaging*, 25:464–475, 2006.

[12] V. A. Zavaletta, B. J. Bartholmai, and R. A. Robb. High resolution multidetector CT-aided tissue analysis and quantification of lung fibrosis. *Academic Radiology*, 7:769–771, 2007.