Multi-scale Analysis of Imaging Features and Its Use in the Study of COPD Exacerbation Susceptible Phenotypes

Felix J.S. Bragman[1], Jamie R. McClelland[1], Marc Modat[1],
Sébastien Ourselin[1], John R. Hurst[2], and David J. Hawkes[1]

[1] Centre for Medical Image Computing, University College London, UK
[2] Centre for Inflammation and Tissue Repair, University College London, UK

Abstract. We propose a novel framework for exploring patterns of respiratory pathophysiology from paired breath-hold CT scans. This is designed to enable analysis of large datasets with the view of determining relationships between functional measures, disease state and the likelihood of disease progression. The framework is based on the local distribution of image features at various anatomical scales. Principal Component Analysis is used to visualise and quantify the multi-scale anatomical variation of features, whilst the distribution subspace can be exploited within a classification setting. This framework enables hypothesis testing related to the different phenotypes implicated in Chronic Obstructive Pulmonary Disease (COPD). We illustrate the potential of our method on initial results from a subset of patients from the COPDGene study, who are exacerbation susceptible and non-susceptible.

1 Introduction

Exacerbations of Chronic Obstructive Pulmonary Disease (COPD) are defined as a sudden worsening of symptoms, which accelerate the decline in lung function leading to an increased risk of mortality. Understanding their pathophysiology is critical for predicting the patients at greatest risk of hospitalisation. Recent work suggests that the frequency of exacerbations is a distinct phenotype [1]. This is described as an exacerbation susceptible phenotype, where a patient may exhibit distinct physiological patterns resulting in an intrinsic susceptibility.

Recent studies have suggested a potential link between changes in lung structure, function and exacerbations. A correlation between the progression of emphysema and the presence of exacerbations has been observed [2] whilst pulmonary arterial enlargement has been seen to be a related factor [3]. Further, regional ventilation defects have been observed prior to acute exacerbations [4]. These suggest a dependence between abnormalities in lung structure, the distribution of disease and exacerbations, which motivates our algorithm.

There is a growing interest in employing machine learning for the study and diagnosis of COPD. Classifiers are frequently trained with scalar values representing the whole lung [5] or individual lobes [6]. This ignores the spatial distribution of disease; which may be a signature of various COPD phenotypes.

We propose a novel framework for the analysis of lung pathophysiology. We hypothesise that the spatial distribution of disease is a discriminating factor in the presence of pathology. Our method is based on the measurement of image features representing the biomechanics and density of tissue, using a sliding box window at various anatomical scales. This is to deal with the bifurcating nature of the respiratory system. We apply it to the study of exacerbation susceptible and non-susceptible patients. The distributions measured at multiple scales are exploited to investigate differences between subtypes whilst classifying for the first time, those at greatest risk of further exacerbations.

2   Method

2.1   Non-rigid Registration

The NiftyReg registration platform[1] [7] is employed to find the spatial mapping between the lung at full inhalation (Ω^*) and end exhalation (Ω). This is performed using a stationary velocity field, parameterised through a cubic B-spline interpolation. The Local Normalised Cross Correlation (LNCC) drives the registration whilst the bending energy of the velocity field is used as the regularisation. The registration is performed by considering only the lungs, delineated by segmented masks. The background volume is set to 0 Hounsfield Units upon which the masks are diluted to include a 0 HU border within the lung volume.

2.2   Feature Extraction

The transformation $\varphi : \Omega \to \Omega^*$, resulting from the registration serves to map each coordinate $x \in \Omega$ to $x^* \in \Omega^*$, such that the position of voxels at expiration ($x \in \Omega$) is known within the inspiratory phase ($x^* \in \Omega^*$). Biomechanical and density-based feature sets are derived using the information embed within φ.

Biomechanical Feature Set. To quantify the transformation φ, we consider the deformation gradient tensor $\mathbf{F}$, which is defined as $\nabla_{x^*}\varphi(x)$. We derive 3 features from $\mathbf{F}$ to capture the respiratory process; the Jacobian determinant ($\det(\mathbf{F})$) and the first 2 moments of the distribution of the eigenvalues of the Lagrangian strain tensor ($\mathbf{E}$). The Jacobian determinant is defined as

$$\det(\mathbf{F}) = \det(\nabla_{x^*}\varphi(x)) \tag{1}$$

and measures the fractional volume change of voxels. The Lagrangian Strain Tensor $\mathbf{E}$ is derived from $\mathbf{F}$, by considering the Right Cauchy-Green Strain ($\mathbf{C}$)

$$\mathbf{C} = \mathbf{F}^\top\mathbf{F}, \;\; \mathbf{F} = \mathbf{R}\,\mathbf{U}, \;\; \mathbf{F}^\top\mathbf{F} = \mathbf{U}^\top\,\mathbf{R}^\top\,\mathbf{R}\,\mathbf{U} = \mathbf{U}^\top\mathbf{U} \; .$$

We are interested in analysing the stretches captured by $\mathbf{F}$. The tensor $\mathbf{C}$ results from a polar decomposition of $\mathbf{F}$, where the rotation component $\mathbf{R}$ is discarded

---

[1] http://sourceforge.net/projects/niftyreg

by considering its orthogonal properties. The tensor $\mathbf{C}$ is thus rotation free, solely containing information about the stretches $\mathbf{U}$. The computation of the Lagrangian Strain Tensor ($\mathbf{E}$) follows

$$\mathbf{E} = \frac{1}{2}\left(\mathbf{C} - \mathbf{I}\right) \ . \tag{2}$$

We derive the principal strains ($\lambda = \{\lambda_i \mid i = 1, 2, 3\}$) via an eigen-decomposition of $\mathbf{E}$. The trace ($\sum \lambda$), provides an overall measure of the magnitude of tissue strain whilst the variance ($\mathrm{Var}(\lambda)$) characterises anisotropy in the strain profile.

Density-Based Feature Set. The transformation φ allows us to compute corresponding measures of voxel density (HU) at inspiration ($\mathrm{I_{ins}}$) and expiration ($\mathrm{I_{exp}}$). We consider the distribution of HU in $\mathrm{I_{ins}}$ and $\mathrm{I_{exp}}$ and 2 scalar values; the percentage of emphysema ($\mathrm{\%LAA_{ins}} - 950\mathrm{HU}$) and gas trapping ($\mathrm{\%LAA_{exp}} - 856\mathrm{HU}$). The $\mathrm{\%LAA_{ins/exp}}$ metrics are computed as follows:

$$\mathrm{\%LAA_{ins}} - 950\mathrm{HU} = \frac{\sum_{x^* \in \Omega^*} (\mathrm{I_{ins}}(x^*) < -950)}{\sum_{x^* \in \Omega^*} x^*} \tag{3a}$$

and

$$\mathrm{\%LAA_{exp}} - 856\mathrm{HU} = \frac{\sum_{x^* \in \Omega'} ((\mathrm{I_{exp}}(x) \circ \varphi) < -856)}{\sum_{x^* \in \Omega'} x^*} \ . \tag{3b}$$

They are expressed as the percentage of voxels below -950 HU and -856 HU within $\mathrm{I_{ins}}$ and $\mathrm{I_{exp}}$. Within our framework, all features are calculated within local neighbourhoods across the lung, which is discussed below in Sect. 2.3.

2.3   Multi-scale Analysis of Imaging Features

Feature Distributions. We propose to sample the local variation of features (f_k) to quantify their distribution across the lung. This is performed by considering histograms ($h_i(f_k; x_j, \phi_i)$) of the local distributions of f_k. Each local feature distribution is centered at a voxel x_j ($j = 1 \cdots J$) within a neighbourhood ω governed by the scale ϕ_i, where $i = 1 \cdots n$ and j is the jth sampled neighbourhood. Thus, distributions at increasing scales of analysis (ϕ_i) can be computed (Fig. 1). The histograms are modelled by the first 4 statistical moments and the median. The feature f_k within ω centered at x_j is defined by:

$$H^j\left(f_k(x_j)\right) = \{\mu(h_1) \ \nu(h_1) \ \sigma(h_1) \ \gamma_1(h_1) \ \gamma_2(h_1) \ \cdots \\ \mu(h_n) \ \nu(h_n) \ \sigma(h_n) \ \gamma_1(h_n) \ \gamma_2(h_n)\} \tag{4}$$

where μ is the mean, ν the median, σ the variance, γ_1 the skewness and γ_2 is the kurtosis. A patient-specific matrix ($\mathbf{H}_p$, $p = 1 \cdots P$) is created such that

$$\mathbf{H}_p = \begin{bmatrix} H^1\left(f_1(x_j)\right) & \cdots & H^1\left(f_k(x_j)\right) & \mathrm{\%LAA^1_{ins/exp}}(x_j) \, \forall \phi \\ \vdots & \vdots & \vdots & \vdots \\ H^J\left(f_1(x_J)\right) & \cdots & H^J\left(f_k(x_J)\right) & \mathrm{\%LAA^J_{ins/exp}}(x_J) \, \forall \phi \end{bmatrix} \ . \tag{5}$$

The Jacobian determinant ($\det(\mathbf{F})$), the trace ($\sum \lambda$) and variance ($\mathrm{Var}(\lambda)$) of the strain eigenvalues and the voxel densities in $\mathrm{I_{ins}}$ and $\mathrm{I_{exp}}$ are modelled locally across the lung ($k = 5$). We incorporate the $\%\mathrm{LAA}-950\mathrm{HU}$ and $\%\mathrm{LAA}-856\mathrm{HU}$ for all ϕ_i, leading to $27n$ features per x_j. The number of sampled regions is determined by the sampling frequency of x_j at the finest scale (ϕ_1).

$$H^j\left(f_k = \sum \lambda; x_j\right) = \{\mu(h_1)\ \nu(h_1)\ \sigma(h_1)\ \gamma_1(h_1)\ \gamma_2(h_1)\ \cdots$$
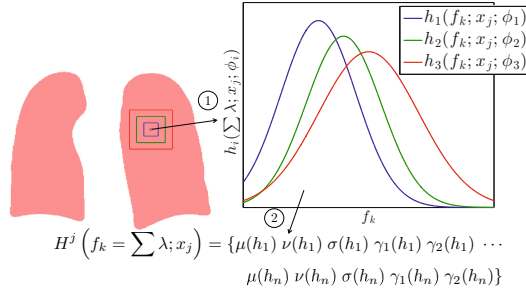$$\mu(h_n)\ \nu(h_n)\ \sigma(h_n)\ \gamma_1(h_n)\ \gamma_2(h_n)\}$$

Fig. 1. Illustration of the framework. 1) A feature f_k (e.g. $\sum \lambda$) at x_j is sampled at $n = 3$ scales, leading to 3 local histograms $h_i(f_k; x_j, \phi_i)$. 2) Statistical moments and the median of $h_i(f_k; x_j, \phi_i)$ are calculated for all ϕ_i, leading to the set $H^j(f_k(x_j))$.

Statistical Analysis of Features

Hypothesis testing using $\mathbf{H}_p$. Analysis of the distribution of values contained within each $\mathbf{H}_p$ allows hypotheses of changes in the global nature of local features to be made. For instance, consider the distribution of the variance of $\det(\mathbf{F})$ at all x_j. Each value demonstrates the local variation in volume change. The distribution of this measure across the lung will illustrate how the local variation is expressed, which may vary across subtypes. This facilitates a direct comparison of patient-specific distributions across phenotypes.

Principal Component Analysis of $\mathbf{X}$. We are interested in modelling the distribution of parameters across the studied population. We apply PCA on $\mathbf{X} = \left[\mathbf{H}_1^\top \cdots \mathbf{H}_P^\top\right]$. This seeks a low-dimensional projection ($d << 27n$) of $\mathbf{X}$, where the variance of the projected features is maximised. The entries of $\mathbf{X}$ are representative of the local histogram features measured at multiple scales. PCA of $\mathbf{X}$ allows one to compute the component scores within each neighbourhood defined by x_j. Thus, the computed scores can be projected to the image space to assess their distribution across the lung. Since the component scores are linear projections of the features measured at various scales, they will capture potential fractal properties in line with the nature of the lung anatomy. The distribution of the principal component scores can be analysed to model patient-specific distributions by computing their respective mean and variance. Thus, phenotype-specific distributions can be estimated to produce a clinically meaningful classifier. Importantly, classification in the PCA subspace prevents overfitting as PCA removes colinearity in the features.

3   Experiments and Results

3.1   Clinical Data

Inhale and exhale breath-hold CT images from the COPDGene study [8] were used. CT scans were acquired from multi-detector CT scanners, at full inspiration (200mAs) and at the end of normal expiration (50 mAs) with resolutions approximately equal to 0.66mm x 0.66mm x 0.73mm [8].

We tested our framework on $P = 20$ subjects with a GOLD 3 severity stage exhibiting $f = 0$ ($n = 10$) or $f \geq 6$ ($n = 10$) exacerbations per year. GOLD 3 patients were chosen due to their low variation in FEV_1. We chose two extreme sets ($f = 0$ and ≥ 6) to gauge the applicability of our framework in discriminating these phenotypes. The patient sets had a mean age of 60.2 and 67.5, a mean FEV_1%predicted of 42.1 and 40.5 and a mean FEV_1/FVC ratio of 42.4 and 47.2.

3.2   Algorithm Parameters

Prior to the registration, the masks were dilated with a sphere of 3 voxel radius. An analysis of the registration parameters was performed; demonstrating robustness in the registration to small parameter changes. The standard deviation of the LNCC Gaussian kernel was set to 3^3 voxels, whilst the weighting of the regularisation was 0.05% of the overall optimised cost function. The finest control point spacing of the B-spline grid was set to 5 voxels along each axis. After registration, the inhale lung mask was eroded by a spherical element with a 7 voxel radius. This was performed to ignore regions prone to discontinuities and which experience an extreme degree of motion. We performed the sampling using a cubic box window at scales 10, 20 and 30 mm^3 ($n = 3$), which is consistent with the size of the secondary pulmonary lobule. A sampling frequency of 10mm was used yielding approximately $7,500$ regions per lung. We ignored regions at all scales where 50% of the voxels fell outside the lung mask.

3.3   Multi-scale Analysis of Imaging Features

We investigated feature distributions at the 3 scales using $\mathbf{H}_p$. We calculated the mean and standard deviation of each feature within $\mathbf{H}_p$ for all 3 scales. This provided two patient-specific distributions of values for each feature. We performed a two-sample t-test for each subtype mean and standard deviation set to determine discriminating factors between both subtypes.

A significant difference in the mean of $\sigma\left(\det(\mathbf{F})\right)$ ($.12 \pm .01$ and $.21 \pm .02$) at all scales of analysis was found ($p < .03$). The feature $\sigma\left(\det(\mathbf{F})\right)$ illustrates the variation in local volume change. The lower variation seen by the exacerbation susceptible group suggested that they exhibit a more homogeneous pattern in their volume change. No significance was seen in the standard deviation of $\sigma\left(\det(\mathbf{F})\right)$ ($p < .20$). We observed a marked difference ($p < .05$) in the mean ($.15 \pm .02$ and $.27 \pm .04$) and standard deviation ($.12 \pm .01$ and $.22 \pm .02$) of $\sigma\left(\sum(\lambda)\right)$ at all scales. This insinuated that for the susceptible group, the anisotropy in the

magnitude of local tissue strain and its variation throughout the lung is more homogeneous compared to the non-susceptible patients. These suggested a possible distinction in physiological patterns, which were exploited in the classification.

Results from the PCA of matrix $\mathbf{X}$ corroborated the above, displaying evidence of distinct feature distributions across subtypes. (Figs. 2 and 3). Figure 2 illustrates 2 patient-specific principal component distributions for each subtype. These are characteristic of the phenotype distributions and are mostly consistent across each group. As the component scores are a linear projection of the features, Figure 2 suggests that there is a consistent physiological pattern per subtype. This is illustrated by a variation in the heterogeneity of the scores as observed in the analysis of $\mathbf{H}_p$. This reinforces the notion of phenotype-specific distributions and the discriminating power of the distribution of disease.
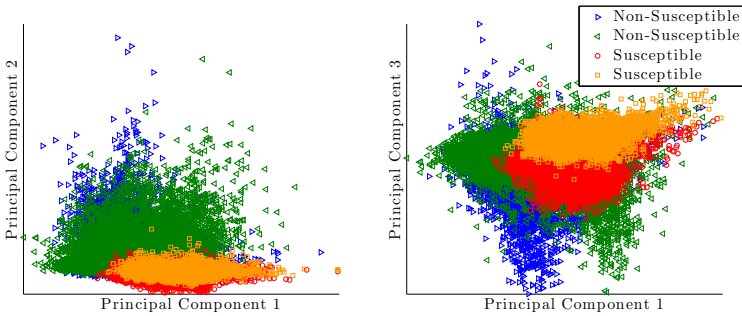

Fig. 2. An example of the multi-scale principal component distributions for each sampled neighbourhood x_j for 4 patients (2 susceptible and 2 non-susceptible patients). The first 3 principal components explain $\approx 55\%$ of the variance of $\mathbf{X}$.

As the principal component scores were computed per sampled region (x_j), we were able to couple them with their respective anatomical location (Fig. 3). This displays varying patterns in the physiology of the lung, consistently within and across subtypes. As the principal components aimed to fully explain the lung macrostructure and the deformation captured within $\mathbf{H}_p$, these maps display a novel way of viewing how lung physiology differs with the COPD phenotype and the frequency of exacerbations.

3.4   Classification of COPD exacerbation-susceptible patients

We aimed to classify exacerbation susceptible and non-susceptible patients based on the hypothesis that global and local patterns of disease differed across subtypes. This was shown in Fig. 2 and 3, where a rise in feature homogeneity coincided with exacerbation susceptibility. We performed the classification on the feature projections using the mean and the variance of the principal component scores as features. The set explaining 90% (17/81) of the variance of $\mathbf{X}$ was chosen.

(a) PC1   (b) PC2   (c) PC3
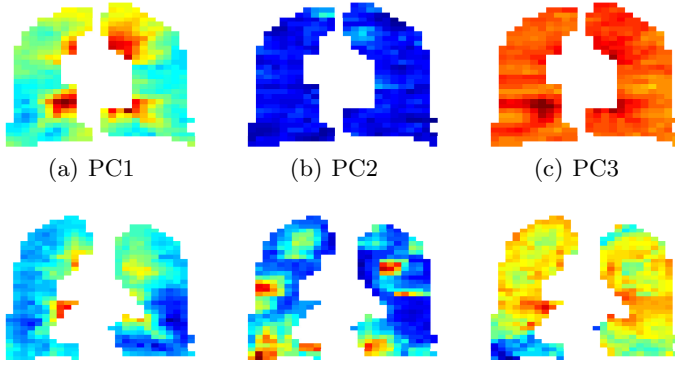
Fig. 3. Projection of the first 3 principal components of the multi-scale PCA into the image space. Coronal slice is at the mid-section. Top row: exacerbation susceptible phenotype. Bottom row: exacerbation non-susceptible phenotype.

A leave-one-out cross validation (LOOCV) was employed to test the classifier. LOOCV iteratively selects one patient ($\mathbf{H}_{unseen}$) as the testing data whilst the remaining are used for training. We assumed independence amongst each training set during the LOOCV to calculate accuracy and precision rates. For the classification, we projected $\mathbf{H}_{unseen}$ into the principal component space of $\mathbf{X}_{P-1}$ and used the mean and variance of the principal component scores as features. We used Support Vector Machines (SVM) as a classifier with a Gaussian radial basis function kernel $\sigma = 2.25$ and a soft-margin constant $C = 0.5$. Our framework has the unique ability to classify an unseen patient as either exacerbation susceptible ($f \geq 6$) or non-susceptible ($f = 0$) with a total accuracy of 75% (Table 1). This supports the applicability of our framework towards determining relationships between the distribution of disease with the clinical outcome.

Table 1. Classification results using Leave One-Out Cross Validation

	Susceptible	Non-Susceptible	Total
Classification accuracy (%)	80	70	75 ± 7.5

4   Conclusions

We have presented a novel framework for investigating global and local patterns of lung pathophysiology. The applicability of our framework in determining relationships between functional measures and the severity of disease has been shown, through an analysis of the exacerbation susceptible phenotype. Analysis of the local feature distributions displayed significant differences in the nature of lung function across subtypes. This translated to subtype-specific distributions

after dimensionality reduction, suggesting an intrinsic physiological behaviour attributed to both sets of patients. The main limitation of our work is due to the lack of patients analysed. We aim to include a larger population of patients to better demonstrate the clinical applicability of our work. This will allow us to correctly evaluate the performance of our classifier, and the consistency and utility of the derived feature distributions. Moreover, we intend to construct anatomical atlases to perform regional inter-patient statistics to investigate whether the spatial location of disease provides a further dimension to the analysis.

References

1. Hurst, J.R., Vestbo, J., Anzueto, A., Locantore, N., Müllerova, H., Tal-Singer, R., Miller, B., Lomas, D.A., Agusti, A., Macnee, W., Calverley, P., Rennard, S., Wouters, E.F.M., Wedzicha, J.A.: Susceptibility to exacerbation in chronic obstructive pulmonary disease. The New England Journal of Medicine 363(12), 1128–1138 (2010)
2. Tanabe, N., Muro, S., Hirai, T., Oguma, T., Terada, K., Marumo, S., Kinose, D., Ogawa, E., Hoshino, Y., Mishima, M.: Impact of exacerbations on emphysema progression in chronic obstructive pulmonary disease. American Journal of Respiratory and Critical Care Medicine 183(12), 1653–1659 (2011)
3. Wells, J.M., Washko, G.R., Han, M.K., Abbas, N., Nath, H., Mamary, A.J., Regan, E., Bailey, W.C., Martinez, F.J., Westfall, E., Beaty, T.H., Curran-Everett, D., Curtis, J.L., Hokanson, J.E., Lynch, D.A., Make, B.J., Crapo, J.D., Silverman, E.K., Bowler, R.P., Dransfield, M.T.: Pulmonary arterial enlargement and acute exacerbations of COPD. The New England Journal of Medicine 367(10), 913–921 (2012)
4. Kirby, M., Kanhere, N., Etemad-Rezai, R., McCormack, D.G., Parraga, G.: Hyperpolarized Helium-3 magnetic resonance imaging of chronic obstructive pulmonary disease exacerbation. Journal of Magnetic Resonance Imaging 37(5), 1223–1227 (2013)
5. Bodduluri, S., Newell, J.D., Hoffman, E.A., Reinhardt, J.M.: Registration-based lung mechanical analysis of chronic obstructive pulmonary disease (COPD) using a supervised machine learning framework. Academic Radiology 20(5), 527–536 (2013)
6. Murphy, K., Pluim, J.P.W., van Rikxoort, E.M., de Jong, P.A., de Hoop, B., Gietema, H.A., Mets, O., de Bruijne, M., Lo, P., Prokop, M., van Ginneken, B.: Toward automatic regional analysis of pulmonary function using inspiration and expiration thoracic CT. Medical Physics 39(3), 1650–1662 (2012)
7. Modat, M., Ridgway, G.R., Taylor, Z.A., Lehmann, M., Barnes, J., Hawkes, D.J., Fox, N.C., Ourselin, S.: Fast free-form deformation using graphics processing units. Computer Methods and Programs in Biomedicine 98(3), 278–284 (2010)
8. Regan, E.A., Hokanson, J.E., Murphy, J.R., Make, B., Lynch, D.A., Beaty, T.H., Curran-Everett, D., Silverman, E.K., Crapo, J.D.: Genetic epidemiology of COPD (COPDGene) study design. COPD 7(1), 32–43 (2010)