Manifold Learning of COPD

Felix Bragman[1], Jamie McClelland[1], Joseph Jacob[1,3], John Hurst[2] and David Hawkes[1]

[1]Centre for Medical Image Computing, University College London, UK
[2]UCL Respiratory, University College London, UK
[3]Department of Radiology in the Mayo Foundation for Medical Education and Research, Mayo Clinic, USA

Motivation and overview

- Chronic Obstructive Pulmonary Disease (COPD) is a heterogeneous disease with multiple pathological processes
- The evolution of local pulmonary damage can differ across patients with equal global values across the lung
- More accurate methods for quantifying disease spread [1] and efficiently computing pairwise similarities [2] are needed

Overview of our method

- Model lung disease progression and local biomechanics with **local disease and deformation probability distributions**
- Pairwise similarities between distributions computed using the **Earth Movers Distance**
- **Manifold learning and fusion** to embed the population into a lower-dimension manifold that parameterises various aspects of COPD progression

I. Quantifying disease and deformation in COPD

- Classification of voxels $\mathcal{Z}$ into emphysema and airway disease (fSAD) using Parametric Response Mapping [3]
- Non-rigid registration of paired breath-hold CT scans using NiftyReg [4] to calculate the Jacobian determinant map J
- Locally sample $\mathcal{Z}$ and J to create local disease and deformation distributions (Fig. 1 and Fig. 2)

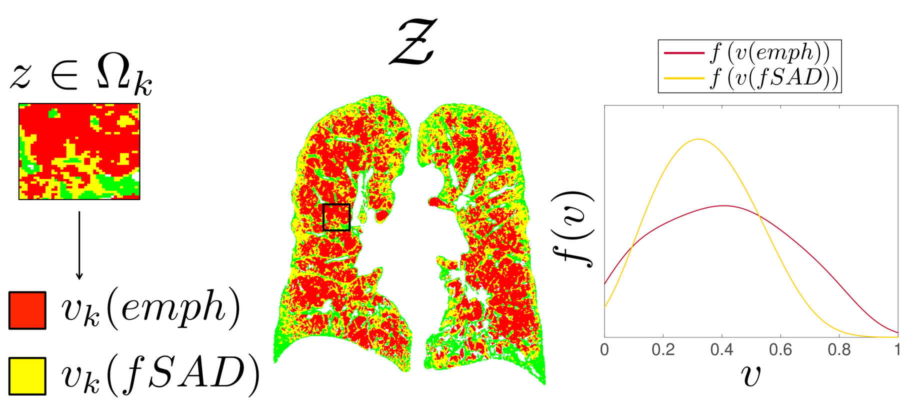

Fig. 1. The classification map $\mathcal{Z}$ is locally sampled to model two properties of disease spread: 1) diffuse or dense local destruction and 2) global homogeneity or heterogeneity

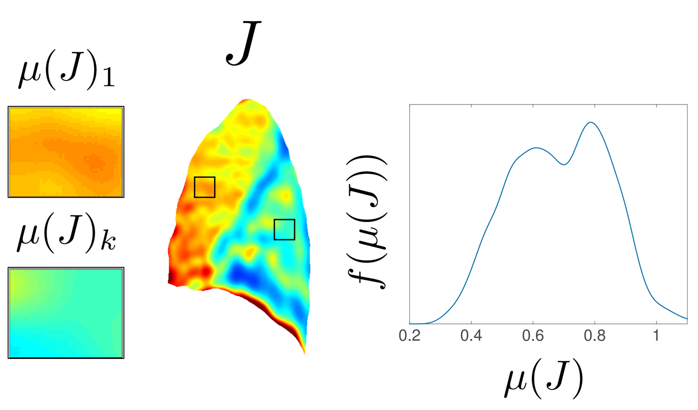

Fig. 2. The local mean of Jacobian map J Is sampled to model local biomechanics across the lung

II. Manifold learning and fusion of COPD

- Distances between the distributions of two patients *i* and *j* are computed with the Earth Movers Distance (EMD)
- A pairwise matrix $\mathcal{M}^{(\cdot)}$ is obtained by considering all pairwise distances in a population of *P = 743* COPD patients
- Separate embeddings for emphysema (y^e), fSAD (y^f) and lung deformation (y^J) are learned from pairwise matrices $\mathcal{M}^e$, $\mathcal{M}^f$ and $\mathcal{M}^J$ using Isomap [5]:

$$\min \sum_{i,j} \left(D_{i,j}^{(\cdot)} - ||y_i^{(\cdot)} - y_j^{(\cdot)}|| \right)^2$$

- The manifold fusion framework of Aljabar et al. [6] is used to combine the embeddings y^e, y^f and y^J into y^c (Fig. 3)

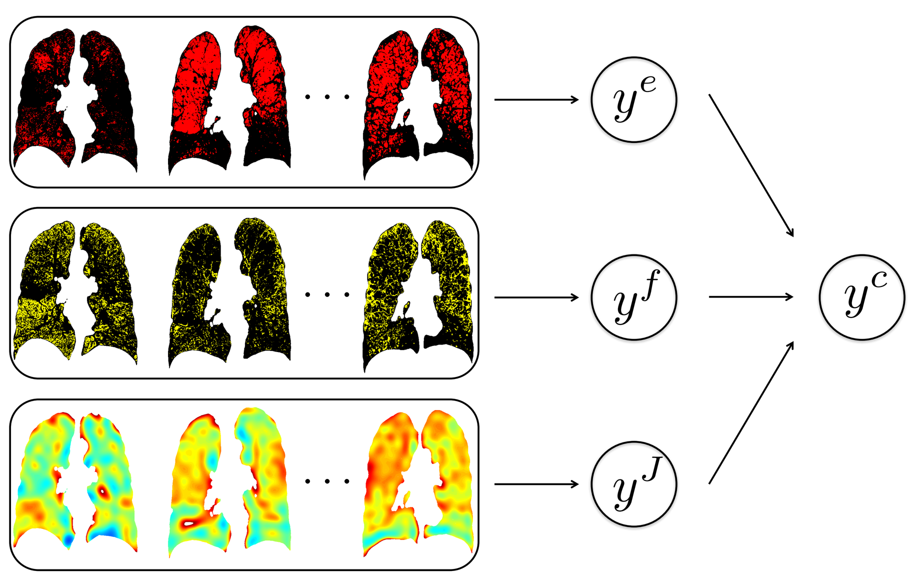

Fig. 3. Fusion of separate embeddings facilitates the construction of a low-dimensional representation of COPD that parameterizes several processes that drive its progression. Fusion is performed by applying Isomap on pairwise L_2 distances on the concatenated coordinates $Y = [s^e y^e, s^f y^f, s^J y^J]$ where $s^{(\cdot)}$ is scaling factor to yield unit-variance in the first component of $y^{(\cdot)}$

III. Prediction of COPD severity

- We considered two models of COPD to predict FEV_1% predicted
  A. y^{c1} : fused embeddings for emphysema and fSAD (Fig. 4A)
  B. y^{c2} : fused embeddings for emphysema, fSAD and Jacobian (Fig. 4B)
- Model B performed the best in predicting COPD severity in comparison to mean levels of emphysema, fSAD and Jacobian (Fig. 5)

- Manifold fusion was performed to create a joint model using mean pairwise distances: $y^{\mu(c_1)}$ and $y^{\mu(c_2)}$
- Correlation between the first component of $y^{(\cdot)}$ and FEV_1% predicted was calculated
- y^{c_1} and y^{c_2} had stronger correlations (0.67* and 0.70*) in comparison to $y^{\mu(c_1)}$ and $y^{\mu(c_2)}$ (0.60* and -0.65*) **p<0.001*

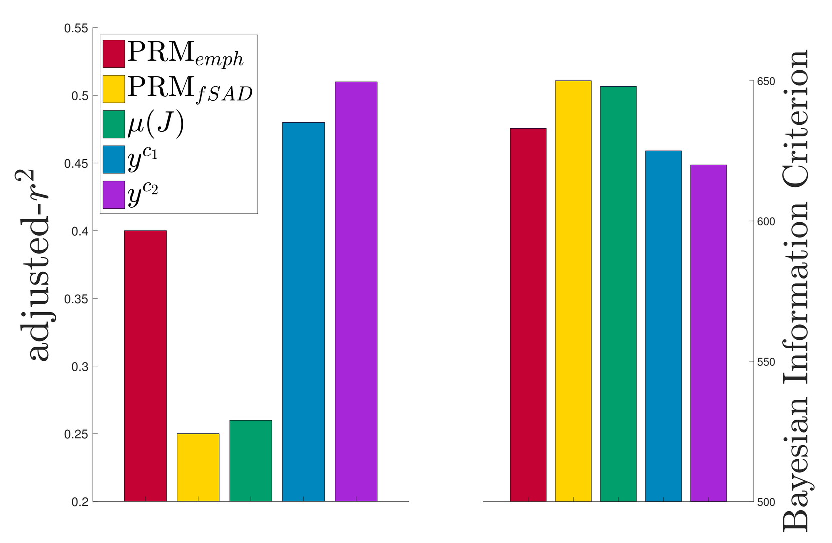

Fig. 5. Linear regression against FEV_1% predicted for each model. Coordinates from y^{c_2} performed best in the prediction of COPD severity.

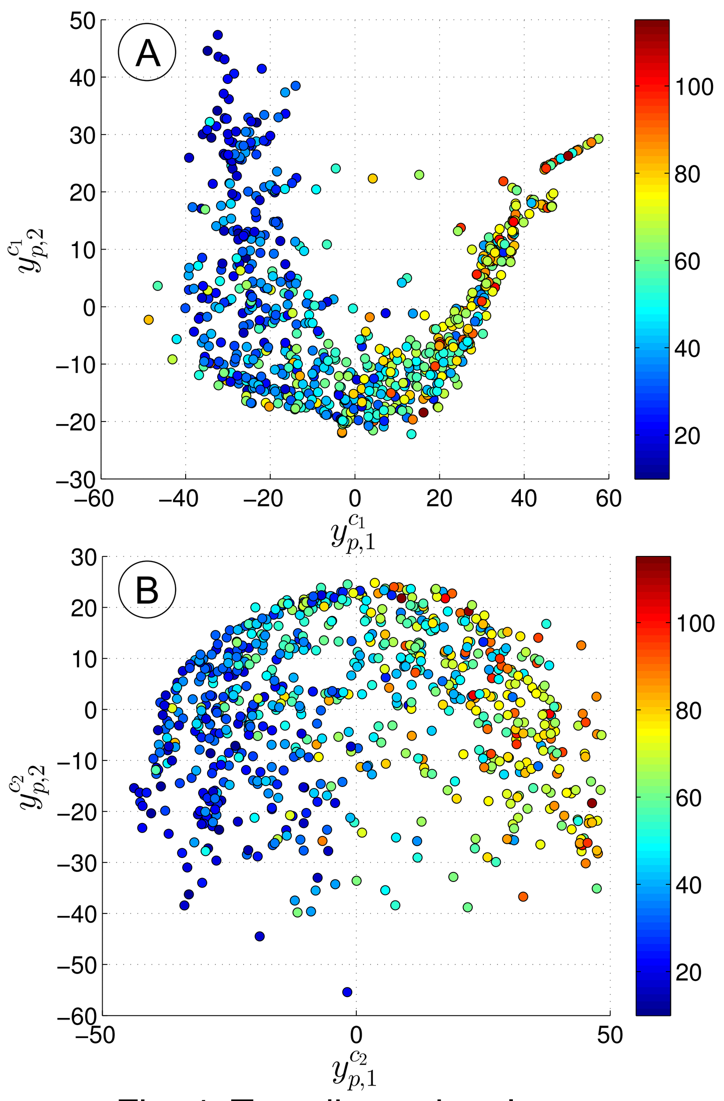

Fig. 4. Two-dimensional projection of the learned low-dimensional embeddings overlaid with FEV_1% predicted colour map.

IV. Trajectories of COPD progression

- Two trajectories of potential disease progression in the space of y^{c_1} were quantified by kernel regression:

$$y^c(l(\cdot)) = \frac{1}{v} \sum_i K(l_i - l) y_i^c$$

- The EMD between the disease distributions (Fig. 1) and idealised healthy distributions (peak at 0) for emphysema and fSAD were used as covariates (l_i)
- The trajectories correspond to potential subtypes of COPD where either emphysema or fSAD are the dominant mechanisms

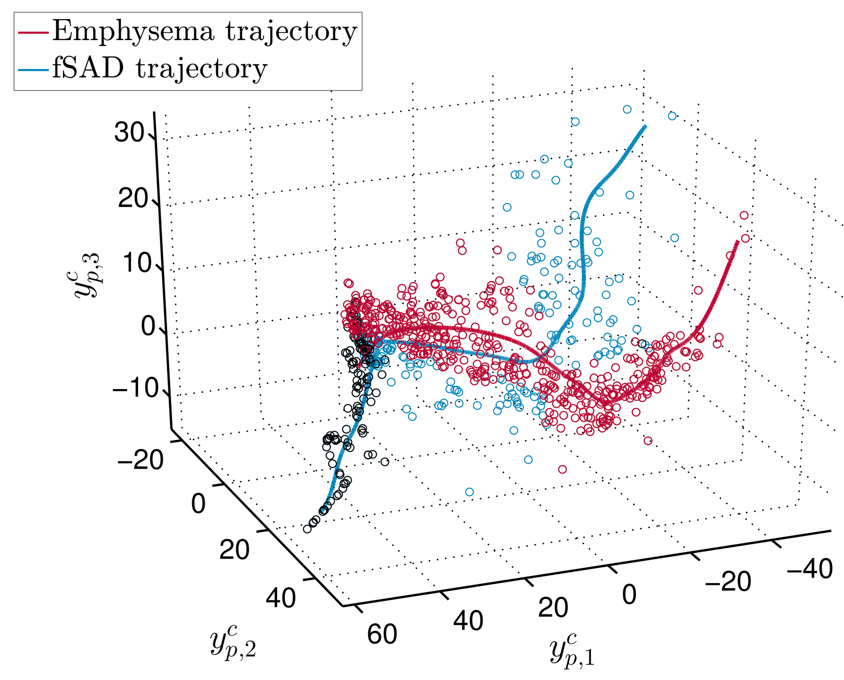

Fig. 6. Trajectories of COPD progression in the space of y^{c_1} parameterised by the disease distributions. Classification of patients into these potential subtypes improves the prediction of FEV_1% predicted with an adjusted-r^2 of 0.52 (emphysema) and 0.45 (fSAD).

V. Results and outlook

- The proposed disease and deformation distributions (Fig. 1 and 2) outperform conventional metrics that do not take into account local properties of COPD
- The position of a patient in the space of $y^{(\cdot)}$ may be critical for assessing COPD to inform therapeutic decisions based on the current COPD trajectory (Fig. 6)
- Complexity of the modelling can be improved by quantifying manifolds on a lobar basis or by considering additional textural measures

1. Harmouche et al. *Academic Radiology* 23(4), 2016
2. Batmanghelich et al. *MICCAI 2016 Workshop: MCV and BAMBI*
3. Galbán et al. *Nature Medicine* 18, 1711-1715 (2012)
4. Modat et al. *Computer Methods and Programs in Biomedicine* 98 (3), 2009
5. Tenenbaum et al. *Science* 290 (5500) , 2000
6. Aljabar et al. *IEEE Transactions on Medical Imaging* 30 (12), 2011

Acknowledgements. This work was supported by the EPSRC under Grant EP/H046410/1 and EP/K502959/1, and the UCLH NIHR RCF Senior Investigator Award under Grant RCF107/DH/2014. This research used data generated by the COPDGene study (phs000179.v3.p2), which was supported by NIH grants U01HL089856 and U01HL089897

EPSRC Pioneering research and skills
NHS National Institute for Health Research
COPDGene COPD Genetic Epidemiology