

# Longitudinal Alignment of Disease Progression in Fibrosing Interstitial Lung Disease

Wolf-Dieter Vogl<sup>1,3,\*</sup>, Helmut Prosch<sup>2</sup>, Christina Müller-Mang<sup>2</sup>,  
Ursula Schmidt-Erfurth<sup>3</sup>, and Georg Langs<sup>1</sup>

<sup>1</sup> Computational Imaging Research Lab, Department of Biomedical Imaging and Image-guided Therapy, Medical University Vienna, Austria

{wolf-dieter.vogl,georg.langs}@meduniwien.ac.at

<sup>2</sup> Department of Biomedical Imaging and Image-guided Therapy, Medical University Vienna, Austria

<sup>3</sup> Christian Doppler Laboratory for Ophthalmic Image Analysis, Department of Ophthalmology and Optometry, Medical University Vienna, Austria

**Abstract.** Generating disease progression models from longitudinal medical imaging data is a challenging task due to the varying and often unknown state and speed of disease progression at the time of data acquisition, the limited number of scans and varying scanning intervals. We propose a method for temporally aligning imaging data from multiple patients driven by disease appearance. It aligns follow-up series of different patients *in time*, and creates a cross-sectional spatio-temporal disease pattern distribution model. Similarities in the disease distribution guide an optimization process, regularized by temporal rigidity and disease volume terms. We demonstrate the benefit of longitudinal alignment by classifying instances of different fibrosing interstitial lung diseases. Classification results (AUC) of Usual Interstitial Pneumonia (UIP) versus non-UIP improve from AUC=0.71 to 0.78 following alignment, classification of UIP vs. Extrinsic Allergic Alveolitis (EAA) improves from 0.78 to 0.88.

## 1 Introduction

Observing the progression of a disease over time is an essential factor during treatment, staging and prognosis [1]. Longitudinal studies with images acquired from subjects at multiple timepoints capture subject-specific disease development trajectories [1]. To study and model disease progression across subjects, it is crucial to establish accurate correspondence of disease stages across individuals. Typically, choosing the initial examination as reference point does not reflect an accurate fit of disease stage, since the extend of disease at the time

---

\* This work has received funding from the Austrian Science Fund (FWF P22578-B19, PULMARCH), the European Union FP7 (KHRESMOI FP7-257528), the Austrian Federal Ministry of Science, Research and Economy and the National Foundation for Research, Technology and Development.

of first diagnosis can vary [2]. In this paper we propose a method for computational alignment of disease progression stages across subjects based on follow-up sequences of imaging data, taking varying progression speeds into account.

A group of diseases for which the progression can be determined via High-Resolution CT (HRCT) are fibrosing interstitial lung diseases (FILD) [3], including three of the most commonly encountered clinical entities: Idiopathic Pulmonary Fibrosis (IPF) with Usual Interstitial Pneumonia (UIP), Nonspecific Interstitial Pneumonia (NSIP), and chronic Extrinsic Allergic Alveolitis (EAA). Various forms of fibrosis and inflammation induce destructive changes to the lung. They are visible as distinctive patterns in HRCT data. Examples are parenchymal reticulation, centrilobular nodules, ground-glass attenuation, and honeycombing [2]. Identifying the extent, spatial distribution and severity of these patterns is essential in the diagnosis of FILD and the differential diagnosis of its entities [3]. To the authors' knowledge, current work features only one study dealing with quantification of disease progression of FILD in subsequent HRCT images [4], and no study exists with a population-wide image based FILD disease progression model.

*Related work.* The question of modelling temporal trajectories across individuals has been studied extensively in the context of shape development. Durrleman et al. [5] and Hart et al. [6] proposed spatio-temporal models from longitudinal data based on diffeomorphic shape changes. Hart et al. developed an interpolation scheme to average individual growth models at every timepoint, independently. Durrleman et al. computed the evolution of a mean template shape and a 4D deformation field of the individual growth models, which describes the variability in appearance and shape evolution. Dittrich et al. [7] found that differences in fetal brain growth speed can be reliably identified based on a non-parametric shape representation. In the case of FILD the shape changes due to pathology and disease progression are negligible. Major changes occur in appearance, texture and its global spatial distribution in the organ. Furthermore, in contrast to development, the timepoint origin within the population is not known in disease progression. Fonteijn et al. [8] proposed an event-based disease progression model where a time line of a disease is developed by finding an optimal ordering of a set of discrete events within a population using a MCMC algorithm.

*Contribution.* In this paper we propose a novel method to perform alignment of disease progression based on appearance. We assume that two acquisitions with similar spatial distribution of disease patterns are at a similar disease progression stage. We first identify appearance patterns that occur frequently across the population and can be identified with high stability. To ensure spatial consistency across the data, we map all data into a common reference space. Then, we calculate dissimilarity measures across the entire population based on the spatial distribution of these patterns. We perform alignment by minimizing a cost term that penalizes dissimilarity of cases with similar progression status. Alignment cannot permute the acquisition sequence of individual patients, but we allow

interval durations to change (*non-rigid longitudinal alignment*), reflecting differences in progression speed. We regularize alignment by inject assumptions about volume increase or decrease of a specific pattern. After alignment we construct a model of disease progression based on the aligned cases. We can fit new data to this model, and can compare the fit of individual data to multiple models representing different diseases. Results demonstrate that the comparison of model fit can serve as accurate classifier, and that the alignment substantially improves classification performance.

## 2 Identifying Appearance Classes in a Joint Reference Space

*Tissue classification.* We perform unsupervised texture classification on the segmented lungs across all cases [9]. A Gray-Level Co-occurrence Matrix (GLCM) is computed for each voxel with various window sizes ( $w$ ) and distances ( $d$ ). To obtain a orientation independent GLCM, it is summed up for all 13 distinct directions. In our case we computed a GLCM for  $w = 33$  and  $d = 9$ , for  $w = 17$  and  $d = \{1, 5, 9\}$ , as well as  $w = 5$  and  $d = 1$ , resulting in 5 GLCMs per voxel. From these GLCMs the Haralick features are computed, resulting in a feature vector of size 65 per voxel. The dimensionality of the feature vector is reduced to 15 by using PCA, and k-Means clustering with  $k=15$  clusters is applied across the entire population.

*Atlas building.* To map all voxels into a joint reference space we build a lung atlas by group-wise Large Deformation Diffeomorphic Metric Mapping (LDDMM) registration using ANTS [10] with a cross-correlation similarity measure. The atlas consists of a mean shape and mean intensity volume. Finally, an expert radiologist identified the two label clusters that are forming reticular and honeycombing patterns, which we then merged into a binary labeling (non-fibrotic/fibrotic tissue). These two clusters are used exemplarily in the further computation.

## 3 Longitudinal Alignment

Let  $\{\mathbf{L}_1, \dots, \mathbf{L}_N\}$  be a set of  $N$  atlas transformed binary label volumes with corresponding acquisition timepoints  $\mathbf{t}^0 = \langle t_1^0, \dots, t_N^0 \rangle$ , where the timepoints are the relative temporal distance in days to the first acquisition per subject. The assignment of an acquisition to a specific subject is defined by the binary variable  $c_{i,j}$ , where  $c_{i,j}$  is 1 if acquisition  $i$  and  $j$  are from the same patient, and 0 otherwise.  $\neg c$  is the negation of  $c$ . We apply a Gaussian filter to each  $\mathbf{L}_i$ , resulting in  $\mathbf{G}_i$ , and compute a dissimilarity matrix  $\mathbf{D} \in \mathbb{R}^{N \times N}$ , where  $\mathbf{D}_{i,j}$  is the Bray-Curtis Dissimilarity (BCD) [11]:  $\mathbf{D}_{i,j} = \sum_{x \in \mathbf{X}} |\mathbf{G}_i(x) - \mathbf{G}_j(x)| / \sum_{x \in \mathbf{X}} |\mathbf{G}_i(x) + \mathbf{G}_j(x)|$ . Furthermore, we define  $v_i$  as the relative volume of disease labeled tissue to the whole lung volume for acquisition  $i$ . Now, we perform longitudinal alignment by minimizing an objective function  $C(\mathbf{t})$  to obtain a mapping  $\mathbf{t}^0 \rightarrow \mathbf{t}$ . The objective function consists of three terms, (1) a tissue distribution similarity term, (2)

a temporal rigidity term that strengthens adherence to the temporal intervals within the series of acquisitions of the same patient, and (3) a regularization term that penalizes a decrease of volume of the diseased tissue. Hence, we minimize:

$$\mathbf{t} = \underset{\mathbf{t}}{\operatorname{argmin}}(C(\mathbf{t})) = \underset{\mathbf{t}}{\operatorname{argmin}}(S(\mathbf{t}) + \lambda R(\mathbf{t}) + \mu V(\mathbf{t})), \quad (1)$$

subject to  $t_{\min} < t_i < t_{\max}$ ,  $t_i \in \mathbf{t}$  and  $t_i < t_j \Leftrightarrow t_i^0 < t_j^0$ ,  $c_{i,j} = 1$ ,  $t_i, t_j \in \mathbf{t}$ . The first constraint restricts the timepoints being in an interval  $[t_{\min}, t_{\max}]$ . The second constraint keeps the order of the acquisitions within a subject.  $\lambda$  and  $\mu$  are weighting factors of the second and third term. The similarity term is:

$$S(\mathbf{t}) = \sum_{1 \leq i \leq N} \sum_{1 \leq j \leq N} w(t_i, t_j) \mathbf{D}_{i,j}, \quad w(t_i, t_j) = \exp\left(-\frac{(t_i - t_j)^2}{h^2}\right), \quad (2)$$

where  $w$  weights the costs of dissimilarity of spatial tissue distribution higher for pairs of acquisitions with similar time point, and lower for those with time points farther apart. The temporal rigidity term  $R(\mathbf{t})$  increases if the length of the period between acquisitions of the same subject differs from the actual period ( $|t_i - t_j|$ ). We can enforce temporal rigidity if we assume that the progression of disease occurs with the same speed across the population, or allow for deviations if this is not the case, i.e., if the disease can progress faster for some subjects.

$$R(\mathbf{t}) = \sum_{1 \leq i \leq N} \sum_{1 \leq j \leq N} |(t_i - t_j) - (t_i^0 - t_j^0)| \cdot c_{i,j} \quad (3)$$

The third term  $V(\mathbf{t})$  is the disease volume regularization term, which penalizes acquisitions of different patients with a higher disease volume to be sorted before acquisitions with a lower disease volume.

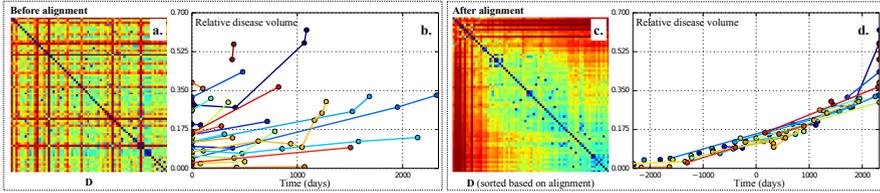
$$V(\mathbf{t}) = \sum_{1 \leq i \leq N} \sum_{1 \leq j \leq N} \max(0, -(v_i - v_j)(t_i - t_j)) \cdot \neg c_{i,j} \quad (4)$$

As constrained minimization algorithm we use Sequential Least Squares Programming. Fig. 1 illustrates the effect of longitudinal alignment of the acquisitions. The dissimilarity matrix is calculated once, and the optimization only manipulates the acquisition time vector  $\mathbf{t}$ . To illustrate the level of pattern distribution coherence across volumes, the right side of Fig. 1 shows the dissimilarity matrix after permuting the volumes corresponding to their optimized value  $t_i$ .

## 4 Disease Progression Model Building

After alignment we can estimate an average disease progression model from the aligned data. A cross-sectional spatial distribution  $\bar{\mathbf{G}}$  at an arbitrary timepoint  $\mathbf{t}$  is computed from the aligned label volumes as follows:

$$\bar{\mathbf{G}}(\mathbf{t}, \mathbf{t}) = \frac{1}{Z} \sum_{1 \leq i \leq N} w(\mathbf{t}, t_i) \cdot \mathbf{G}_i, \quad Z = \sum_{1 \leq i \leq N} w(\mathbf{t}, t_i) \quad (5)$$



**Fig. 1.** a) Dissimilarity matrix  $\mathbf{D}$  sorted by initial timepoints  $\mathbf{t}^0$ . b) Plot relative disease volume  $\mathbf{v}$  versus un-aligned timepoints  $\mathbf{t}^0$  per patient. c)  $\mathbf{D}$  sorted by aligned timepoints  $\mathbf{t}$ . d) Plot relative disease volume  $\mathbf{v}$  versus aligned timepoints  $\mathbf{t}$ .

The weighting factor function  $w$  is analogous to (2). Label distribution volumes which are close to the specified timepoint  $t$  are given a higher weight than volumes that are farther away. A time-parameterized spatiotemporal disease progression model  $\mathbf{M}(t)$  is obtained by computing  $\bar{\mathbf{G}}$  at any timepoint in the interval  $[t_{\min}, t_{\max}]$ . Fig. 2 illustrates the spatial distribution within the model at various timepoints.

#### 4.1 Fitting a New Subject to the Disease Progression Model

Let  $\{\mathbf{G}_1^s, \dots, \mathbf{G}_n^s\}$  be a set of  $n$  atlas-transformed smoothed label acquisitions of a new subject that was not part of the model building process, and  $\mathbf{t}^{0s}$  be the corresponding timepoints for this subject. We can fit the observed data of the new subject to the disease model built as described in the previous section, by minimizing an objective function similar to (1). The optimal timepoints  $\mathbf{t}^s$  for the new observations in the model  $\mathbf{M}(t)$  are obtained by finding a mapping  $\mathbf{t}_0^s \rightarrow \mathbf{t}^s$  so that the distances of the acquisition label distribution to the model label distribution at these timepoints are minimized. Hence, we minimize:

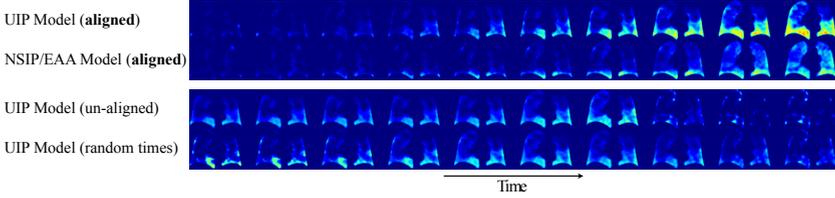
$$\mathbf{t}^s = \underset{\mathbf{t}^s}{\operatorname{argmin}} \sum_{1 \leq i \leq n} \left( D_{\text{KL-Sym}}(\mathbf{M}(t_i^s) || \mathbf{G}_i^s) \right) + \xi R(\mathbf{t}^s) \tag{6}$$

where  $D_{\text{KL-Sym}}$  is the symmetric Kullback–Leibler divergence.  $R$  is the same timepoint rigidness constraint as in (3), and  $\xi$  is the weighting factor of this term. The same time-constraints as in (1) are used for optimization.

As a result we obtain a cross-sectional spatiotemporal disease progression model from the temporally aligned acquisitions. Furthermore, new acquisitions or series of acquisitions can be fitted to the model.

## 5 Experiments and Results

The evaluation dataset contains 70 HRCT thorax acquisitions at maximum inspiration level from 31 patients. The axial resolution is  $512 \times 512$  voxels with a x,y-spacing of 0.55mm to 0.85mm and a slice thickness of 1mm. An expert radiologist determined the disease class based on the inspection of the HRCT



**Fig. 2.** Disease distribution progression models. 1<sup>st</sup> and 2<sup>nd</sup> row: comparison of aligned models UIP vs. non-UIP. 3<sup>rd</sup> and 4<sup>th</sup> row: UIP un-aligned and UIP with random starting-points. Note the differences in aligned UIP and non-UIP. The color indicates the cross-sectional spatial distribution of the fibrotic tissue.

**Table 1.** Summary of classification performance based on non-aligned and aligned data: UIP vs. non-UIP (UvN) resp. UIP vs. EAA (UvE) from models with non-rigid, rigid and un-aligned timepoints, resp. random starting points.

	UvN				UvE				
	Alignment	Sens.	Spec.	F1	AuC	Sens.	Spec.	F1	AuC
non-rigid		<b>0.875</b>	0.667	0.757	0.771	<b>0.875</b>	<b>0.875</b>	<b>0.875</b>	<b>0.875</b>
rigid		0.750	<b>0.800</b>	<b>0.774</b>	<b>0.775</b>	0.750	<b>0.875</b>	0.808	0.812
un-aligned		0.688	0.733	0.710	0.710	0.688	<b>0.875</b>	0.770	0.781
random		0.812	0.733	0.771	0.773	0.812	<b>0.875</b>	0.843	0.844

images, as well as on the histopathology report. UIP, possible UIP, NSIP, EAA, and ‘Other’ are the disease classes defined beforehand. The number of patients (acquisitions) having UIP are 11(18), possible UIP 5(11), NSIP 5(15), EAA 8(21), and other 2(3). The proposed spatiotemporal disease progression model  $\mathbf{M}$  was evaluated in terms of the accuracy during two classification tasks: UIP (UIP+possible UIP) vs. non-UIP and UIP vs EAA. We assume that the spatiotemporal distribution of the fibrosis is different in these two models, and that a new subject can be classified by determining which model fit better to the acquisitions. The evaluation was performed in a Leave-One-Out Cross-Validation setting per subject for both the model building and the classification. An UIP resp. non-UIP spatiotemporal model was computed from the time-adjusted training data for the left and right lung separately, since the speed of disease progression varies for each side of lung. The smoothing factor  $\sigma$  was set to 5.  $t_{\min}$ ,  $t_{\max}$  was set to  $\pm \max(\mathbf{t})$  and  $h$  was set to 730 (= 2 years). Each longitudinal alignment was performed once with the temporal rigidity constraint (Equation 3) and once without this term by setting the weights  $\lambda$  and  $\xi$  to  $5 \times 10^{-3}$  (rigid) resp. 0 (non-rigid) in (1) and (6). The weight  $\mu$  for the disease volume regularization term was set to  $5 \times 10^{-3}$ . As initial experiments showed, setting the weights  $\lambda$  and  $\mu$  to these values guides the optimization process without dominating it. To evaluate the effect of alignment, we also computed a model from the un-aligned data and a random model, where the timepoint of the first acquisition for each subject is shifted randomly.

In order to determine the disease class, the test subject acquisitions are fitted to the left and right UIP resp. non-UIP models as described in (6), where in the optimization process the KL-Divergence of both sides are summed up. Finally, the disease label is determined as the label to which model the KL-Divergence of the fitted acquisitions is smaller.

A summary of the classification performance is given in Table 1. In all experiments alignment improves classification performance. In case of UIP vs. non-UIP the Area under Curve (AUC) increases from 0.710 (unaligned) to 0.771 (non-rigid) and 0.775 (rigid). The sensitivity increases from 0.688 to 0.875 (non-rigid) and 0.750 (rigid), whereas the specificity increases only in the rigid case from 0.733 to 0.8, and for non-rigid alignment the specificity is reduced to 0.667. In the case of UIP vs. EAA the AuC increases from 0.781 (unaligned) to 0.875 (non-rigid) resp. 0.812 (rigid). Sensitivity increases from 0.688 to 0.750 (rigid) resp. 0.875 (non-rigid), whereas the specificity remains stable at 0.875.

## 6 Discussion

We propose to align disease progression across patients based on imaging data. The alignment establishes temporal correspondence across patients, and enables building progression models that reflect observations made across multiple patients. The alignment is based on appearance patterns that are identified by unsupervised learning in a study population. It allows fitting a new subject to an existing model, and thereby allows for evaluating the similarity between an individual and a population whose imaging data spans different disease stages.

We evaluated the longitudinal alignment in terms of classification performance, by differentiating between UIP and non-UIP, resp. UIP and EAA. After alignment the differences in disease pattern distribution in UIP and non-UIP become apparent (Fig. 2). In UIP the distribution is subpleural and basilar and in non-UIP it is more diffusely distributed across the entire lung. These findings are consistent with the literature [2,3], and the method now allows for a quantitative comparison across cases. Temporal non-rigidity during alignment improved the sensitivity of classification.

As distance measure we chose BCD instead of Euclidean distance. This measure is closely related to the Dice's coefficient (DSC), with  $BCD = 1 - DSC$ . In early disease stages, where the disease label voxels are the minority, the Euclidean distance is dominated by the zero-labeled non-disease voxels, whereas BCD measures the overlap of non-zero disease label voxels only. The disease volume regularization term makes an assumption regarding the increase or decrease of a pattern while the disease advances. This can be estimated from the individual cases. However, current work focuses on methods to allow for analysis of patterns where there is no fixed a priori assumption in this regard.

In the experiments we focused on the disease patterns of honeycombing and reticulation, which are the dominant factors in UIP. In principle the method works with any dissimilarity measure, and by incorporating other appearance patterns such as ground-glass opacities or centrilobular nodules, we expect an

enhancement of the model and further insight into the disease progression of ILDs. Furthermore, evaluation of the alignment and the model building on a larger lung dataset as well as on other progressing diseases with longitudinal image data available are ongoing.

## References

1. King Jr., T.E.: Clinical advances in the diagnosis and therapy of the interstitial lung diseases. *American Journal of Respiratory and Critical Care Medicine* 172(3), 268–279 (2005)
2. Raghu, G., Collard, H.R., Egan, J.J., Martinez, F.J., Behr, J., Brown, K.K., Colby, T.V., et al.: An official ats/ers/jrs/alat statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *American Journal of Respiratory and Critical Care Medicine* 183(6), 788–824 (2011)
3. Hodnett, P.A., Naidich, D.P.: Fibrosing interstitial lung disease. A practical high-resolution computed tomography-based approach to diagnosis and management and a review of the literature. *Am. J. Respir. Crit. Care Med.* 188(2), 141–149 (2013)
4. Arzhaeva, Y., Prokop, M., Murphy, K., van Rikxoort, E.M., de Jong, P.A., Gietema, H.A., Viergever, M.A., van Ginneken, B.: Automated estimation of progression of interstitial lung disease in ct images. *Medical Physics* 37(1), 63–73 (2009)
5. Durrleman, S., Pennec, X., Trouvé, A., Braga, J., Gerig, G., Ayache, N.: Toward a comprehensive framework for the spatiotemporal statistical analysis of longitudinal shape data. *International Journal of Computer Vision* 103(1), 22–59 (2013)
6. Hart, G., Shi, Y., Zhu, H., Sanchez, M., Styner, M., Niethammer, M.: Dti longitudinal atlas construction as an average of growth models. In: *MICCAI Workshop on Spatiotemporal Image Analysis for Longitudinal and Time-Series Image Data* (2010)
7. Dittrich, E., Riklin Raviv, T., Kasprian, G., Donner, R., Brugger, P.C., Prayer, D., Langs, G.: A spatio-temporal latent atlas for semi-supervised learning of fetal brain segmentations and morphological age estimation. *Med. Image Anal.* 18(1), 9–21 (2014)
8. Fonteijn, H.M., Clarkson, M.J., Modat, M., Barnes, J., Lehmann, M., Ourselin, S., Fox, N.C., Alexander, D.C.: An event-based disease progression model and its application to familial Alzheimer’s disease. In: Székely, G., Hahn, H.K. (eds.) *IPMI 2011*. LNCS, vol. 6801, pp. 748–759. Springer, Heidelberg (2011)
9. Mariolis, I., Korfiatis, P., Kalogeropoulou, C., Daoussis, D., Petsas, T., Costaridou, L.: Computer aided diagnosis of diffuse lung disease in multi-detector CT – selecting 3D texture features. In: Bamidis, P., Pallikaraki, N. (eds.) *XII Mediterranean Conference on Medical and Biological Engineering and Computing 2010*. IFMBE Proceedings, vol. 29, pp. 208–211. Springer, Heidelberg (2010)
10. Avants, B.B., Yushkevich, P., Pluta, J., Minkoff, D., Kordzykowski, M., Detre, J., Gee, J.C.: The optimal template effect in hippocampus studies of diseased populations. *Neuroimage* 49(3), 2457–2466 (2010)
11. Bray, J.R., Curtis, J.T.: An ordination of the upland forest communities of southern wisconsin. *Ecological Monographs* 27(4), 325–349 (1957)