Spatio-Temporal Signatures to Predict Retinal Disease Recurrence

Wolf-Dieter Vogl[1,2(✉)], Sebastian M. Waldstein[2], Bianca S. Gerendas[2], Christian Simader[2], Ana-Maria Glodan[2], Dominika Podkowinski[2], Ursula Schmidt-Erfurth[2], and Georg Langs[1]

[1] Computational Imaging Research Lab, Department of Biomedical Imaging and Image-guided Therapy, Medical University Vienna, Vienna, Austria
{wolf-dieter.vogl,georg.langs}@meduniwien.ac.at
[2] Christian Doppler Laboratory for Ophthalmic Image Analysis, Vienna Reading Center, Department of Ophthalmology and Optometry, Medical University Vienna, Vienna, Austria

Abstract. We propose a method to predict treatment response patterns based on spatio-temporal disease signatures extracted from longitudinal spectral domain optical coherence tomography (SD-OCT) images. We extract spatio-temporal disease signatures describing the underlying retinal structure and pathology by transforming total retinal thickness maps into a joint reference coordinate system. We formulate the prediction as a multi-variate sparse generalized linear model regression based on the aligned signatures. The algorithm predicts if and when recurrence of the disease will occur in the future. Experiments demonstrate that the model identifies predictive and interpretable features in the spatio-temporal signature. In initial experiments recurrence vs. non-recurrence is predicted with a ROC AuC of 0.99. Based on observed longitudinal morphology changes and a time-to-event based Cox regression model we predict the time to recurrence with a mean absolute error (MAE) of 1.25 months, comparing favorably to elastic net regression (1.34 months), demonstrating the benefit of a spatio-temporal survival model.

1 Introduction

Biomarkers derived from medical imaging data are an essential tool for diagnosis, therapeutic decisions, and evaluation of treatment response. They provide valuable insight by quantifying informative changes in anatomic, physiological, biochemical or molecular processes [1]. In a clinical setting predictive biomarkers that estimate future disease development and treatment response are exceptionally beneficial, since they allow to personalize treatment, and to optimize its

W.-D. Vogl—The financial support by the Austrian Federal Ministry of Economy, Family and Youth and the National Foundation for Research, Technology and Development, the EU (FP7-ICT-2009-5/318068, VISCERAL), and OeNB (15929) is gratefully acknowledged.

S. Ourselin et al. (Eds.): IPMI 2015, LNCS 9123, pp. 152–163, 2015.
DOI: 10.1007/978-3-319-19992-4_12

effect. We propose a data-driven algorithm to identify spatio-temporal predictive imaging biomarkers or *signatures* in longitudinal medical imaging data. The algorithm localizes these markers in a joint reference space shared across individuals after anatomical alignment, and predicts treatment response using sparse learning methods. We demonstrate the effectiveness of the proposed method on a longitudinal ophthalmic imaging dataset of patients with macular edema secondary to Retinal Vein Occlusion (RVO). The algorithm predicts two variables: based on morphology observed in the early stage of treatment, it predicts if edema will recur in an individual patient after treatment. Secondly, it predicts the time-point of the first recurrence of edema.

Challenges in the Treatment of Retinal Diseases. Retinal vein occlusion is the second most common sight-threatening retinal vascular disorder after diabetic retinopathy with an estimated 16.4 million adults suffering from RVO worldwide [2]. In most cases a consequence of RVO is an exudation of fluid into the macula, a so-called macular edema. Left untreated, this edema inevitably leads to irreversible vision loss. An effective treatment for macular edema is an intraocular injection of anti-vascular endothelial growth factor (anti-VEGF) agents. However, without frequent treatment a recurrence of edema often occurs. The high treatment burden on a patient, the variable response to the anti-VEGF agent, the risk of complications as well as the very high cost of the drug ($2000 per injection) and the possible side-effects of frequent treatments make a solid pro re nata (PRN, as needed) regimen on an individualized basis with the smallest amount of anti-VEGF injections given while remaining effective, a vital necessity. In current clinical practice, PRN treatment is guided by the presence or absence of cystic retinal and subretinal fluid visible in Spectral-Domain Optical Coherence Tomography (SD-OCT) images acquired at monthly intervals. SD-OCT acquires 3-dimensional scans of the retina on a micrometer resolution and enables a visualization and quantification of microstructural changes in the eye. In Fig. 1 a SD-OCT reconstruction from a healthy retina and a retina with macular edema are shown, where in the diseased eye the retinal structure is disrupted by a spongy cystic fluid structure, causing a swelling of the retina, i.e. macular edema (Fig. 1c). The prediction of individual disease paths enables a personalized treatment with the potential reduction in the frequency of both monitoring visits and particularly injections, which provides the urgently needed relief of the current burden on both patients and healthcare systems in a high-frequency, high-cost therapy [2].

Contribution. In this paper we propose a method to predict treatment outcome based on observations made during early stages of treatment. We identify interpretable biomarkers in longitudinal imaging data for prediction. To compare features through follow-up examinations, and across patients, we obtain a joint coordinate reference space from retinal OCT images by intra-patient and inter-patient registration using fundus and OCT image landmarks. From pixel-wise features extracted from the initial three months within the reference frame coordinate system we predict the treatment response of *recurrence vs. non-recurrence*

(a) (b) (c)

Fig. 1. 3D SD-OCT reconstruction of a (a) healthy retina with the foveal pit in the center, and (b) a retina with edema, where the swelling of the retina is caused by cystic and subretinal fluid (c).

of macular edema in a 12 month follow-up using elastic net regularized generalized linear models. Furthermore, we predict the *time-point of the first recurrence of edema*. Cross-validation experiments demonstrate that the proposed method yields good prediction accuracy and interpretable results. Furthermore, we show that longitudinal features together with time-to-event survival statistics in the Cox proportional hazard model increase time-to-recurrence prediction accuracy.

Related Work. We pose the prediction of treatment response as a multivariate regression problem. In the following, we briefly review relevant related approaches. By treating each pixel/voxel or small groups of pixels as a single feature we are operating in a high-dimension-low-sample-size setting, where the feature dimension size p is several orders of magnitude larger than the number of patients n ($p \gg n$). Multivariate sparse linear regression methods, for instance Lasso [3] or elastic net [4] as well as non-linear regression methods such as Random Forests [5] are able to provide a prediction in such a setting [6,7]. In addition, they compute measures of feature-importance enabling insight into disease mechanisms. Such feature selection methods are used for instance in gene expression studies [8], fMRI network analysis [9] or predictions in structural neuroimages [10]. Most of these studies focus on prediction of the *present* condition from images, not dealing with the prediction of the *future*.

Sabuncu [11] proposed a sparse Bayesian multivariate prediction model combined with a survival model for studying longitudinal follow-up data. He demonstrated that an image based time-to-event prediction improves the result compared to a binary classification of converter/non-converter. Bogunović et al. [12] proposed a machine learning approach to predict the treatment response from retinal OCT in patients with age-related macular degeneration by classifying quantitative features extracted from the fovea center aligned image data.

To compensate for the anatomical variation in imaging studies the individual data is usually mapped to a common coordinate system or atlas [13]. Spatial variations in retinal images arise from different scanning positions and the varying anatomy of the retina across subjects. To ensure spatial consistency across the retinal data Abràmoff et al. [14] described a method to normalize eye fundus images by using visible landmarks such as the optic disc center, the fovea, and the main vessel arcades. Due to the smaller field of view of clinical OCT images compared to fundus images not all of these landmarks are available in OCT.

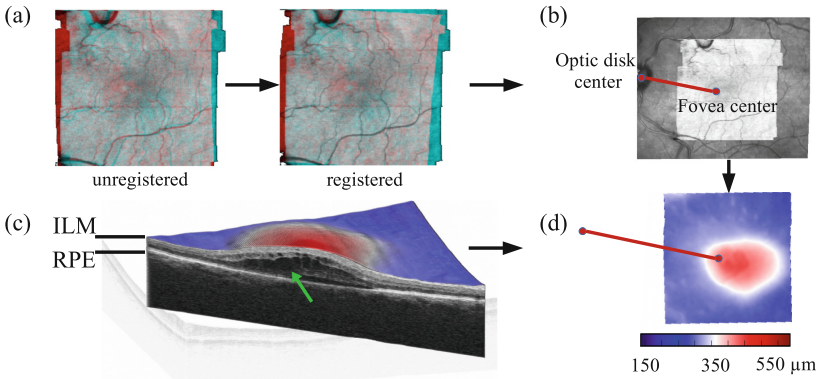

Fig. 2. Steps to get spatio-temporal signatures in a joint reference space: (a) Intra-patient registration via vessel structure. For illustration purpose 2D projections from OCTs of two time-points are overlayed and colored blue resp. red. The steps in the projections come from the motion-correction. (b) Inter-patient alignment via fovea center and optic disk center. (c) Disease features as total retinal thickness maps are obtained by measuring the distance between the ILM layer and RPE layer. The cut through the retina reveals the pathological cystic structure causing edema (green arrow). (d) Transformation of the thickness maps into the reference frame (Color figure online).

2 Spatio-Temporal Features in a Joint Reference Space

We normalize the retinal OCT images by transforming the data into a joint reference coordinate system using intra-patient follow-up registration based on the vessel structure, and inter-patient alignment via landmarks visible in OCT and fundus images.

First, we reduce motion artefacts introduced by patient movement during acquisition by the method described in [15]. To obtain landmarks for the intra-patient follow-up registration, automatic vessel segmentation is performed on the OCT images. Parameters for the affine registration are generated by applying Coherent Point Drift [16] to the segmented retinal vessel point sets, as described in [17] (Fig. 2a). The inter-patient affine registration is performed by aligning the fovea center and the optic disk center within patients (Fig. 2b). Due to the limited field of view of the macular centered OCT the optic disk is only visible in the corresponding fundus image. Hence, a registration of the fundus image to OCT is performed using the projection of the OCT image, registered rigidly to the fundus image by minimizing the normalized cross-correlation of the intensity values. The foveal center landmark (center of the foveal pit) (Fig. 1a), was set manually by expert readers. The optic disk center was identified by applying circular Hough transformation with varying radii on the binary threshold fundus image, and picking the circle showing the highest response. The optic disk center position is the center of the circle. Since the quality of the fundus images obtained by the OCT-scanner varies substantially, and the fovea center is sometimes hard to identify in cases with heavy pathology, we chose as reference scan the time-point

at which the foveal center as well as the optic disc center could be accurately determined. All other scans in the time-series were transformed to the selected one, and the reference scan was aligned with the other patients. Scans from the right eye were mirrored, in order to align their anatomy with scans from the left eye.

Spatio-temporal Signature of Disease. Total retinal thickness maps $\boldsymbol{M}$ are computed as the distance between the inner limiting membrane (ILM) surface, which is the boundary between the retina and the vitreous body, and the retinal pigment epithelium (RPE), which is the border between the retina and the choroid (Fig. 2c). These layers are identified using a graph-based surface segmentation algorithm [18]. These maps are transformed into the reference space using the obtained affine transformation (Fig. 2d).

Let $\boldsymbol{v}^{(m)} = (x_1, x_2, ..., x_k)$ be the vectorized pixel values from the transformed thickness map $\boldsymbol{M}$ for the month m. By concatenating the thickness map vectors and the change of thickness over time up to month m, a spatio-temporal signature vector is obtained for each individual: $\boldsymbol{x}^{(m)} = (\boldsymbol{v}^{(0)}, \boldsymbol{v}^{(1)}, ..., \boldsymbol{v}^{(m)}, \boldsymbol{v}^{(1)} - \boldsymbol{v}^{(0)}, \boldsymbol{v}^{(2)} - \boldsymbol{v}^{(1)}, ..., \boldsymbol{v}^{(m)} - \boldsymbol{v}^{(m-1)})$. These signatures are pooled in a matrix $\boldsymbol{X} \in \mathbb{R}^{n \times p}$, where each row represents a signature vector of a subject and each column is a distinct spatio-temporal anatomical position in the retina. $\boldsymbol{x}_i^{(m)} \in \mathbb{R}^p$ is the signature vector for subject i.

3 Prediction of Recurrence

We predict the treatment response based on data up to the initial loading phase of three monthly injections. Specifically, we predict for an individual based on the corresponding spatio-temporal signature $\boldsymbol{x}_i^{(m)}$ covering the three month loading phase, if a treated edema will recur in the future. In a second task we predict at which time-point the first recurrence happens for patients with recurring edema. For notational clarity, we drop the time-index $^{(m)}$ in the following explanation.

3.1 Prediction by Sparse Linear Regression

We assume that the continuous response variable y_i for an individual i is a weighted linear combination of the input variables $\boldsymbol{x}_i$: $y_i = w_0 + w_1 x_1 + ... + w_p x_p = \boldsymbol{x}_i \boldsymbol{w}$. The coefficients (or weights) $\boldsymbol{w}$ are estimated for a training-set $\boldsymbol{X}$ in a regularized way by minimizing the following objective function:

$$\underset{\boldsymbol{w}}{argmin} \frac{1}{2n} \|\boldsymbol{X}\boldsymbol{w} - \boldsymbol{y}\|_2^2 + \lambda P(\boldsymbol{w}) \tag{1}$$

Since in our case the system is strongly underdetermined ($p \gg n$), a regularization function P is necessary, where λ controls the amount of regularization. Recently, sparsity has been proposed as property of the coefficients [3,8], where only a few but relevant coefficients $\boldsymbol{w}$ are non-zero, highlighting predictive

anatomical regions and time-points. Sparsity can be obtained by applying the ℓ_1 norm on $\boldsymbol{w}$, known as Lasso regularization [3]. When features are strongly correlated, Lasso tends to pick only one of these features at random. Thus, Zou and Hastie proposed elastic net regularization [4] to overcome the limitation of Lasso regularization, by combining the ℓ_1 norm with a ridge regularization (ℓ_2 norm):

$$P(\boldsymbol{w}) = \rho\|\boldsymbol{w}\|_1 + \frac{(1-\rho)}{2}\|\boldsymbol{w}\|_2^2, \tag{2}$$

where ρ defines the ratio of the convex combination of ℓ_1 and ℓ_2 regularization.

The features x_i with non-zero coefficients w_i represent anatomical locations and time-points whose characteristics are informative for the prediction. They enable interpretation by ophthalmologists.

3.2 Categorical Variable Prediction

Using a generalized version of the sparse linear regression, where the outcome variable y is replaced by a function, categorical variables can be predicted using logistic regression. The probabilities of the binary outcomes are modeled as logit function (log of the odds): $\text{logit}(p_i) = ln(\frac{p_i}{1-p_i}) = \boldsymbol{x}_i\boldsymbol{w}$, where $p_i = \Pr(y_i = 1|\boldsymbol{x}_i)$. To obtain the coefficients Eq. 1 is modified such that:

$$\underset{\boldsymbol{w},c}{\text{argmin}} \sum_{i=1}^{n} (\log(\exp(y_i(\boldsymbol{x}_i\boldsymbol{w} + c)) + 1)) + \lambda P(\boldsymbol{w}) \tag{3}$$

The class probability p_i for a new case with covariates $\boldsymbol{x}$ can be predicted from the trained weights via the inverse logit: $p_i = \text{logit}^{-1}(\boldsymbol{x}\boldsymbol{w})$.

We compute the coefficients $\boldsymbol{w}$ by training an elastic net regularized logistic regression model using the spatio-temporal signature matrix $\boldsymbol{X}$ and the binary outcome labels $\boldsymbol{y}$ of non-recurring/recurring edema within 12 months. By applying the trained model on an unseen case we obtain the probability of recurrence p_i. Finally, we threshold the probability in order to obtain a binary outcome.

3.3 Temporal Variable Prediction

We predict the time to the first recurrence of edema using survival analysis, where the recurrence is modeled as a time duration T until an event happens. A common tool in survival analysis is the Cox proportional hazards (PH) model [19]. The model assumes a log-linear relationship of the covariates to a baseline hazard h_0, which describes how the risk of event changes over time based on baseline covariates. The time-parameterized hazard function is then:

$$h(t|\boldsymbol{x}_i) = h_0(t)\exp(\boldsymbol{x}_i\boldsymbol{w}) \tag{4}$$

A generalized linear model is obtained by formalizing the model as a hazard ratio: $\log\frac{h(t|\boldsymbol{X})}{h_0(t)} = \boldsymbol{X}\boldsymbol{w}$. Inference can be performed using partial likelihood [19]

$$L(\boldsymbol{w}) = \prod_{r\in D} \frac{\exp(\boldsymbol{x}_r\boldsymbol{w})}{\sum_{j\in R_i}\exp(\boldsymbol{x}_j\boldsymbol{w})} \tag{5}$$

where D is the set of indices of patient events and R_i is the set of indices of individuals which are at risk at time t_i. The coefficients are obtained by minimizing the regularized negative partial log-likelihood:

$$\underset{\boldsymbol{w}}{\operatorname{argmin}} - \sum_{r \in D} (\boldsymbol{x}_r \boldsymbol{w} - \log(\sum_{j \in R_i} \exp(\boldsymbol{x}_j \boldsymbol{w}))) + \lambda P(\boldsymbol{w}) \tag{6}$$

Predictions of hazards for a new individual with covariates $\boldsymbol{x}$ can be obtained by using Eq. 4 with the new covariates. In that case the baseline hazard h_0 has to be estimated by using the Breslow estimator:

$$h_0(t_i) = \frac{d_i}{\sum_{j \in R_i} \exp(\boldsymbol{x}_j \boldsymbol{w})} \tag{7}$$

where d_i is the number of events at t_i. Furthermore, we can estimate the individual survival function $S_i(t)$, which is defined as $S_i(t) = Pr(T > t)$, describing the probability that the time of event is later than some specified time t. The function can be estimated from the individual relative risk, the cumulative hazard Λ_0 and the baseline survival function $S_0(t)$ [20]:

$$S_i(t) = S_0(t) \exp(\boldsymbol{x}_i \boldsymbol{w}), \quad \text{with} \tag{8}$$

$$\Lambda_0(t) = \sum_{j:t_j \leq t} h_0(t_j), \quad \text{and} \quad S_0(t) = exp(-\Lambda_0(t)) \tag{9}$$

We obtain the coefficients $\boldsymbol{w}$ by training an elastic net regularized Cox PH model on the spatio-temporal signature matrix $\boldsymbol{X}$ and the time-points of events in month. For an unseen case we estimate from the covariates $\boldsymbol{x}$ the baseline hazard h_0 using Eq. 7 and the individual survival function $S_i(t)$ using Eq. 9. By computing the survival function for all time-points we determine the time-point of recurrence where the survival function drops below a given threshold.

A description of how to optimize the objective functions using coordinate descent can be found in [21] and [22]. We used the implementation from the package `glmnet` of the statistics software `R` for our computation.

4 Evaluation and Results

We evaluated the proposed method in two prediction experiments, (1) predicting recurrence vs. non-recurrence of edema within 12 months, and (2) predicting the time to the first recurrence of edema. Baseline SD-OCT scans and 12 monthly follow-up scans of 44 patients with central retinal vein occlusion (CRVO) were included. All patients received initial ranibizumab (anti-VEGF) injections for three months followed by a PRN regimen. Total retinal thickness maps were computed for all scans, and were transformed into the joint coordinate system as described in Sect. 2. All maps were smoothed with a Gaussian kernel with $\sigma = 1$. Figure 3 shows the aligned total retinal thickness maps for two patients with resp. without recurring edema.

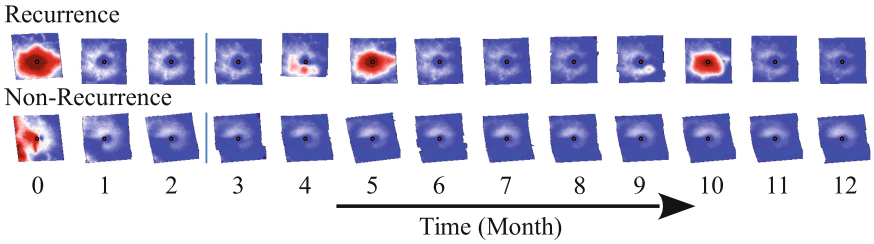

Fig. 3. 12 month follow-up series of aligned total retinal thickness maps for two patients with recurring edema at month 5 and 10 resp. without recurring edema.

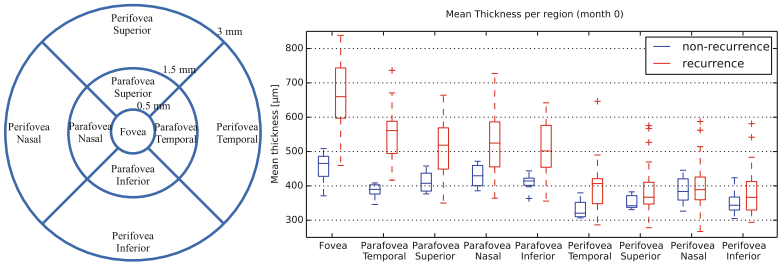

Fig. 4. (a) Division of macula in nine regions of interest defined by ETDRS [23]. (b) Mean retinal thickness of the baseline (month 0) per region stratified by patients with non-recurrence and recurrence of edema.

The time-point of a recurrence was determined algorithmically for each patient to serve as a standard-of-reference for evaluation. First, the total thickness maps were divided into nine circular regions of interest within three concentric circles with diameters of 1 mm, 3 mm and 6 mm, centered at the fovea (Fig. 4a) as defined by the Early Treatment Diabetic Retinopathy Study (ETDRS) design [23]. From each region the mean thickness was computed. A recurrence was defined as an increase of the mean thickness by 15 pixels ($= 29\,\mu m$) of two subsequent time-points in any region. 6 of the 44 patients showed no recurring edema within 12 months.

We used two-level nested five-fold cross-validation (CV) on patient level, where in the inner loop we conduct a grid search to tune the sparsity parameters ρ and λ, as well as the optimal threshold for the probability outcomes in terms of maximizing the F-score on the training set (harmonized mean of sensitivity and specificity). In the outer CV loop we measure the performance of the trained model on the test fold. All evaluations were repeated 20 times with random stratified CV partitioning, with at least one non-recurrence case in each fold.

To evaluate the benefit of using longitudinal data, we performed the training and testing on thickness maps up to month two ($\boldsymbol{x}_i^{(2)}$), one ($\boldsymbol{x}_i^{(1)}$), resp. the

Table 1. Evaluation results from the classification task non-recurrence vs. recurrence using total thickness maps up to month two, one and the baseline month, as well as up to two month with the thickness change over time information (2'). Classification results are non-recurrence vs. recurrence.

	Logistic Regression				Random Forest				Cox PH			
Month	0	1	2	2'	0	1	2	2'	0	1	2	2'
Sensitivity	**1.00**	**1.00**	**1.00**	**1.00**	0.88	0.85	0.78	0.81	0.83	0.83	0.83	0.45
Specificity	0.92	0.90	0.90	0.82	0.98	0.98	0.98	**1.00**	0.99	0.90	0.94	0.96
ROC AuC	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	0.91	0.93	0.96

baseline month only ($\boldsymbol{x}_i^{(0)}$). The tasks have been evaluated once with thickness difference between two subsequent time-points in the signature vector and once without this information, to evaluate the performance gain when utilizing disease changes over time.

Finally, models were computed from the whole dataset, from which the coefficients were mapped back to images to get interpretable results of the selected features.

Prediction of Non-recurrence vs. Recurrence. In the first experiment we evaluated the binary classification performance of recurrence vs. non-recurrence of edema within 12 months using sparse logistic regression in comparison to a Random Forest classification. As error measures we computed the sensitivity, specificity and the Receiver-Operating-Curve (ROC) Area-under-Curve (AuC), where the AuC was obtained from the predicted probabilities within the outer cross-validation loop, and a mean AuC has been computed from these. The classification results are listed in Table 1. In the classification task the baseline month is already enough to obtain predictive results with an AuC of 0.99 for logistic regression and Random Forest classification (note that at this point due to the small amount of patients with non-recurrence the confidence interval for sensitivity/specificity is large. For a sensitivity/specificity of 1.00/0.92 the 95 % confidence intervals are 0.52-1.00 / 0.78-0.98). When mapping the coefficients back to the reference frame (Fig. 5a) it can be observed that almost all features were selected from the fovea center and the parafovea temporal area at month zero. By comparing the mean thickness for each region between the two groups, as done in Fig. 4b, it can be observed that there is a (significant) difference of the mean thickness between the groups in these two areas, as well as almost no overlap of the mean thickness in the parafovea temporal area. The sparse feature selection correctly identified these areas and used them for prediction.

Prediction of time to Recurrence. In the second experiment we predicted the time to the first recurrence of edema for patients with recurring edema. We trained an elastic net regularized regression model, as well as Random Forest regression model with the thickness maps as input and the time-to-recurrence as outcome variables. To evaluate the benefit of using a survival model in comparison to a

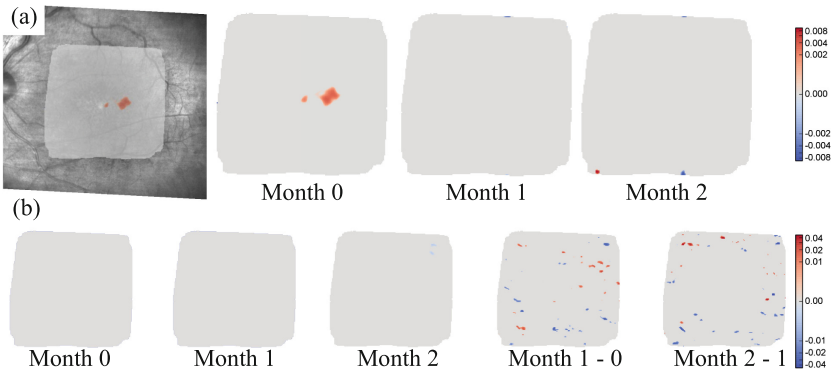

Fig. 5. Sparse coefficients mapped into the common reference coordinate system. (a): An exemplary fundus image mapped into the reference frame, overlaid by the retained coefficients from a single month and coefficients retained from the sparse logistic regression for month zero to two. (b) Coefficients retained from the Cox PH model for month zero, one and two as well as from the thickness differences between month one and zero resp. month two and one.

regression model, we furthermore trained a Cox PH model on the same dataset, where patients with no recurrence were set to censored after 12 months, since we do not know if edema recurred afterwards. For each patient we estimated the survival function $S_i(t)$ as described in Sect. 3.3. The time at which the survival function dropped below 0.5 has been set as the time at which an recurrence of edema is estimated. Patients with a survival function > 0.5 after twelve month got the non-recurrence label. Hence, the Cox PH model was used for both tasks, the binary classification and the time-point of recurrence estimation.

Mean absolute error (MAE) between predicted and true month of recurrence are reported in Table 2. The Cox PH model showed superior results compared to linear regression and Random Forest regression, with an MAE of 1.25 for Cox PH model resp. 1.34 for linear regression and 1.38 for Random Forest. Furthermore, the addition of thickness changes over time as features improved the performance for Cox and elastic net regression. By looking at the coefficients for the Cox PH model (Fig. 5b) it can be observed that almost only thickness difference features are picked. Features from the difference between month one

Table 2. Evaluation results of the time to recurrence prediction task. The mean absolute error (MAE) and the standard deviation (SD) is computed as the absolute difference between predicted and true time to recurrence in months.

	Elastic Net				Random Forest				Cox PH			
Month	0	1	2	2'	0	1	2	2'	0	1	2	2'
MAE [month]	1.37	1.40	1.43	1.34	1.39	1.38	1.54	1.46	1.30	1.26	1.29	**1.25**
SD	1.10	1.12	1.10	1.01	1.11	1.12	1.12	1.06	1.28	1.27	1.27	1.26

and zero are selected over the whole field of view, whereas from the month two to month one thickness difference only features from the perifoveal areas are selected, skipping the central and the parafoveal area. This indicates that the change of thickness between month zero and one in general as well as the change in the outer areas between month two and one are informative regarding the duration until recurrence.

5 Conclusion

In this paper we propose a method to extract spatio-temporal signatures from longitudinal retinal SD-OCT images transformed into a joint reference coordinate system, and use these features to predict the *future* development of disease under treatment. In particular we predicted two variables from image acquisitions during the initial three monthly treatments, (1) the non-recurrence vs. recurrence of edema within twelve months, and (2) the time to recurrence of edema. We demonstrated that sparse feature selection via elastic net in a multivariate generalized linear model setting yields accurate prediction and interpretable results. Furthermore, we showed that using survival models in terms of the Cox proportional hazards model increases the accuracy when predicting the temporal variable. The proposed methodology is an important step towards image-based individualization of patient treatment and disease management.

References

1. Smith, J.J., Sorensen, A.G., Thrall, J.H.: Biomarkers in imaging: Realizing radiology's future. Radiology **227**(3), 633–638 (2003)
2. Laouri, M., Chen, E., Looman, M., Gallagher, M.: The burden of disease of retinal vein occlusion: review of the literature. Eye **25**(8), 981–988 (2011)
3. Tibshirani, R.: Regression shrinkage and selection via the lasso. J. Roy. Stat. Soc. Ser. B (Methodol.) **58**, 267–288 (1996)
4. Zou, H., Hastie, T.: Regularization and variable selection via the elastic net. J. Roy. Stat. Soc. Ser. B (Methodol.) **67**(2), 301–320 (2005)
5. Breiman, L.: Random forests. Machine learning **45**(1), 5–32 (2001)
6. Hastie, T.J., Tibshirani, R.J., Friedman, J.H.: The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, Heidelberg (2011)
7. Rasmussen, P.M., Hansen, L.K., Madsen, K.H., Churchill, N.W., Strother, S.C.: Model sparsity and brain pattern interpretation of classification models in neuroimaging. Pattern Recogn. **45**(6), 2085–2100 (2012). Brain Decoding
8. Zou, H., Hastie, T.: Regression shrinkage and selection via the elastic net, with applications to microarrays. J. Roy. Stat. Soc. B. **67**, 301–320 (2003)
9. Langs, G., Menze, B.H., Lashkari, D., Golland, P.: Detecting stable distributed patterns of brain activation using gini contrast. NeuroImage **56**(2), 497–507 (2011)
10. Kandel, B.M., Wolk, D.A., Gee, J.C., Avants, B.: Predicting cognitive data from medical images using sparse linear regression. In: Gee, J.C., Joshi, S., Pohl, K.M., Wells, W.M., Zöllei, L. (eds.) IPMI 2013. LNCS, vol. 7917, pp. 86–97. Springer, Heidelberg (2013)

11. Sabuncu, M.R.: A Bayesian algorithm for image-based time-to-event prediction. In: Wu, G., Zhang, D., Shen, D., Yan, P., Suzuki, K., Wang, F. (eds.) MLMI 2013. LNCS, vol. 8184, pp. 74–81. Springer, Heidelberg (2013)
12. Bogunović, H., Abràmoff, M.D., Zhang, L., Sonka, M.: Prediction of treatment response from retinal oct in patients with exudative age-related macular degeneration. In: Chen, X., Garvin, M.K., L.J., (ed.) Proceedings of the Ophthalmic Medical Image Analysis First International Workshop, OMIA 2014, Held in Conjunction with MICCAI 2014, Boston, Massachusetts, September 14, 2014, pp. 129–136 Iowa Research Online (2014)
13. Reuter, M., Schmansky, N.J., Rosas, H.D., Fischl, B.: Within-subject template estimation for unbiased longitudinal image analysis. Neuroimage **61**(4), 1402–1418 (2012)
14. Abràmoff, M.D., Garvin, M.K., Sonka, M.: Retinal imaging and image analysis. IEEE Rev. Biomed. Eng. **3**, 169–208 (2010)
15. Montuoro, A., Wu, J., Waldstein, S., Gerendas, B., Langs, G., Simader, C., Schmidt-Erfurth, U.: Motion artefact correction in retinal optical coherence tomography using local symmetry. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) MICCAI 2014, Part II. LNCS, vol. 8674, pp. 130–137. Springer, Heidelberg (2014)
16. Myronenko, A., Song, X.: Point set registration: Coherent point drift. IEEE Trans. Pattern Anal. Mach. Intell. **32**(12), 2262–2275 (2010)
17. Wu, J., Gerendas, B.S., Waldstein, S.M., Langs, G., Simader, C., Schmidt-Erfurth, U.: Stable registration of pathological 3D-oct scans using retinal vessels. In: Chen, X., Garvin, M.K., Liu, J.J. (eds.) Proceedings of the Ophthalmic Medical Image Analysis First International Workshop, OMIA 2014, Held in Conjunction with MICCAI 2014, pp. 1–8. Iowa Research Online (2014)
18. Garvin, M.K., Abràmoff, M.D., Wu, X., Russell, S.R., Burns, T.L., Sonka, M.: Automated 3-D intraretinal layer segmentation of macular spectral-domain optical coherence tomography images. IEEE Trans. Med. Imaging **28**(9), 1436–1447 (2009)
19. Cox, D.: Regression models and life tables (with discussion). J. Roy. Stati. Soc. B **34**, 187–220 (1972)
20. Cox, D.R., Oakes, D.: Analysis of survival data. vol. 21. CRC Press (1984)
21. Friedman, J., Hastie, T., Tibshirani, R.: Regularization paths for generalized linear models via coordinate descent. J. Stat. Softw. **33**(1), 1 (2010)
22. Simon, N., Friedman, J., Hastie, T., Tibshirani, R.: Regularization paths for cox's proportional hazards model via coordinate descent. J. Stat. Softw. **39**(5), 1–13 (2011)
23. Chew, E.Y., Klein, M.L., Ferris, F.L., Remaley, N.A., Murphy, R.P., Chantry, K., Hoogwerf, B.J., Miller, D.: Association of elevated serum lipid levels with retinal hard exudate in diabetic retinopathy: Early treatment diabetic retinopathy study (etdrs) report 22. Arch. Ophthalmol. **114**(9), 1079–1084 (1996)