Predicting Macular Edema Recurrence from Spatio-Temporal Signatures in Optical Coherence Tomography Images

Wolf-Dieter Vogl, Sebastian M. Waldstein, Bianca S. Gerendas, Ursula Schmidt-Erfurth, and Georg Langs

***Abstract*—Prediction of treatment responses from available data is key to optimizing personalized treatment. Retinal diseases are treated over long periods and patients' response patterns differ substantially, ranging from a complete response to a recurrence of the disease and need for re-treatment at different intervals. Linking observable variables in high-dimensional observations to outcome is challenging. In this paper, we present and evaluate two different data-driven machine learning approaches operating in a high-dimensional feature space: sparse logistic regression and Random Forests based extra trees (ET). Both identify spatio-temporal signatures based on retinal thickness features measured in longitudinal spectral-domain optical coherence tomography (OCT) imaging data and predict individual patient outcome using these quantitative characteristics. We demonstrate on a dataset of monthly SD-OCT scans of 155 patients with central retinal vein occlusion (CRVO) and 92 patients with branch retinal vein occlusion (BRVO) followed over one year that we can predict from initial 3 observations if the treated disease will recurr within the covered interval. ET predicts the outcome on 5-fold cross-validation with an area under the receiver operating characteristic curve (AuC) of 0.83 for BRVO and 0.76 for CRVO. Logistic regression achieved an AuC of 0.78 and 0.79 respectively. At the same time the methods identified stable predictive signatures in the longitudinal imaging data that are the basis for accurate prediction. Furthermore, our results show that taking spatio-temporal features into account improves accuracy compared to features extracted at a single time-point. Our results demonstrate the feasibility of mining longitudinal data for predictive signatures, and building predictive models based on observed data.**

***Index Terms*—Predictive models, biomarkers, optical coherence tomography.**

I. Introduction

PRECISION medicine aims to deliver appropriate and timely treatment for each individual patient. Its diagnostic, prognostic, and therapeutic strategies enable optimal efficacy while reducing potential treatment-related morbidity [1]. The primary challenge in precision medicine is identifying predictive markers for future disease progression and treatment response. In this paper we demonstrate that such *spatio-temporal signatures* and corresponding *predictive models* can be learned from large-scale imaging data, without explicit patho-physiological models of disease mechanisms.

A. Background, Disease and Treatment Options

Identifying biomarkers able to differentiate patients into clinically meaningful subgroups of disease, outcome or treatment response is an important unmet medical need in precision medicine. However, the vast amount of biomedical data collected from molecular, genetic, and imaging sources renders extraction of relevant predictive biomarkers difficult [2]. Data-driven approaches based on machine-learning are promising avenues to identify and validate robust and sensitive biomarkers in big data [3].

Retinal vein occlusion (RVO), caused by obstruction of the retinal venous system, is the second-most common retinal vascular disease after diabetic retinopathy [4] with an estimated 16.4 million adults affected worldwide [5]. RVO is divided into central retinal vein occlusion (CRVO) and branch retinal vein occlusion (BRVO), depending whether the site of occlusion is at the central retinal vein or at a retinal branch vein. The increase in hydrostatic pressure due to the veinous obstruction results in fluid extravasation, swelling of the central retina (macular edema), hemorrhage, and frequently retinal ischemia. Untreated, this disease ultimately results in irreversible visual impairment [6].

Intra-ocular injections of anti-vascular endothelial growth factor antibodies (e.g., ranibizumab (Lucentis, Genentech, Inc., South San Francisco, CA)) are the current state-of-the-art therapy for macular edema in both CRVO [7] and BRVO [8]. They generally lead to a substantial reduction in macular edema and a restoration of visual acuity. However, regular clinic visits and continuous treatment with intra-ocular injections are required to prevent recurrent edema formation in most patients. Medication costs of $\sim$ \$2000 per injection and the high-frequency of visits put a substantial burden on patients and the health-care system [6]. In addition, intra-ocular injections are invasive and may confer serious side-effects such as the risk of devastating intra-ocular infection (endophthalmitis) [9]. Thus, a treatment paradigm based on precision medicine with the goal of reducing unnecessary

Manuscript received December 14, 2016; revised April 22, 2017; accepted April 26, 2017. The financial support by the Austrian Federal Ministry of Economy, Family and Youth and the National Foundation for Research, Technology and Development (OPTIMA), and by the Austrian Science Fund (FWF I2714-B31) is gratefully acknowledged.

W.-D. Vogl and G. Langs are with the Computational Imaging Research Lab, Department of Biomedical Imaging and Image-guided Therapy, Medical University Vienna, Vienna, Austria (email: wolf-dieter.vogl@meduniwien.ac.at, georg.langs@meduniwien.ac.at)

W.-D. Vogl, G. Langs, S.M. Waldstein, B.S. Gerendas, and U. Schmidt-Erfurth are with the Christian Doppler Laboratory for Ophthalmic Image Analysis, Vienna Reading Center, Department of Ophthalmology and Optometry, Medical University Vienna, Vienna, Austria.

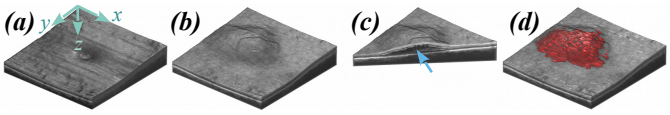

Fig. 1. Three-dimensional reconstruction of fovea centered optical coherence tomography (OCT) scans. (a) Healthy retina with the fovea pit in the center; (b) retinal swelling (macular edema) caused by central retinal vein occlusion; (c) cut through volumetric scan showing cystoid intraretinal and subretinal fluid (blue arrow); (d) distribution of edema within the retina.

monitoring visits and injections is desired to relieve the major socioeconomic and personal burdens. Predicting disease-free stability is an extremely difficult problem and thus patients are often seen infinitely to avoid the miss of a recurrence. Consequently, the reliable prediction of the individual future response to therapy is an essential step.

In current clinical practice, patients are monitored at monthly intervals using spectral domain OCT (SD-OCT) [10], which provides a 3-dimensional reconstruction of the retina at a micrometer resolution. SD-OCT reveals detailed in vivo information about pathologic changes in the retina, such as cystoid and subretinal fluid in macular edema secondary to RVO (Fig. 1). These micro-morphology changes within the retina are the candidate features for the proposed algorithm.

B. State-of-the-art

Identifying predictive features in a high dimensional pool of pixel-wise extracted candidates within a relatively small sample size is based on a family of regression methods that assume only a small number of the variables observed are linked to a prediction target. Methods such as least absolute shrinkage and selection operator (LASSO) [11] or elastic net [12] perform regularization inducing sparseness and produce predictive and interpretable results in multiple areas of medical imaging such as predictions from structural neuroimages [13], functional magnetic resonance imaging (fMRI) data [14] and microarray data [12]. Rasmussen et al. [15] discussed the influence of sparsity regularization parameters on interpretability of the fMRI classification model. They concluded that networks may be overseen when tuning sparsity on maximizing prediction accuracy only and proposed a more careful selection of regularization parameters for visualization. Hastie et al. [16] suggested the "one-standard error" rule should be used for model interpretation. Thereby, the most parsimonious model is chosen whose error is no more than one standard error above the minimum error determined in a cross validation setting.

Another strategy of obtaining a predictive model in the high dimension and small sample size environment is selecting important features and training a model using the random forests (RF) algorithm [17]. Such an approach has been applied successfully in, amongst many others, fMRI analysis [18], brain lesion segmentation [19], gene analysis [20] and glaucoma diagnosis [21]. Recently, machine learning algorithms have been applied on retinal OCT images for prediction. De Sisternes et al. used LASSO regularization in their statistical model to to predict the progression of age-related macular degeneration (AMD) from quantitative features extracted from the images [22]. Quantitative OCT features combined with a machine-learning algorithm were also used by Bogunović et al. [23] to predict the treatment response from retinal OCT in patients with AMD. Niu et al. [24] used a RF based approach to predict geographic atrophy progression for a future time-point from SD-OCT image features, and to identify the relevant features based on RF out-of-bag (OOB) feature importance measure.

C. Contribution

We propose a data-driven machine-learning approach that identifies predictive spatio-temporal imaging biomarkers or *predictive signatures* and corresponding *prediction models* in large-scale SD-OCT data of patients with CRVO or BRVO to predict whether a treated macular edema will recur within the observed interval from initial images. This proposal is a substantial extension of previous work [25]. First, we normalize all imaging data across scans and patients to a joint reference space using automatically detected and matched fundus and OCT image landmarks. In this space, we can compare local imaging features across different scanning protocols and scanner vendors, across patients and during follow-up examinations. Second, we extract pixel-wise feature maps covering the underlying retinal morphologic structure and changes over time from the aligned scans. This results in a high dimensional vector of feature candidates for each scan and each patient. Third, we treat prediction and signature identification as a sparse regression problem, and evaluate different techniques for prediction and prediction regularization while preventing overfitting to the training data. We evaluate the logistic regression generalized linear model (GLM) with sparsity regularization assuming that only a fraction of the features contain predictive information. We also evaluate "extremly randomized trees" or extra trees (ET), [26] based on the RF classifier of Breiman [17] that is able to select important features for prediction without overfitting to the data due to the randomness in the decision tree ensemble [26].

II. Method

The proposed method consists of three parts: first the transformation of imaging data into a joint reference space, second the extraction of individual spatio-temporal features from the images, and third the prediction of disease recurrence based on the features using a machine learning approach.

A. Joint Reference Space of Fovea-centered Retinal Images

As a first step, we normalize the longitudinal OCT scans from all patients by transforming them into a joint reference space to enable comparisons among the population. We propose an en face two-dimensional projection reference space, with the transformations obtained from a two-step intra- and inter-patient registration process (Fig. 2a). The retinal vessel structure segmented from the OCT projection image serves as the basis for stable corresponding landmarks in the intra-patient registration. The center of the fovea and of the optic nerve head (ONH) are the guiding landmarks in the inter-patient registration process.

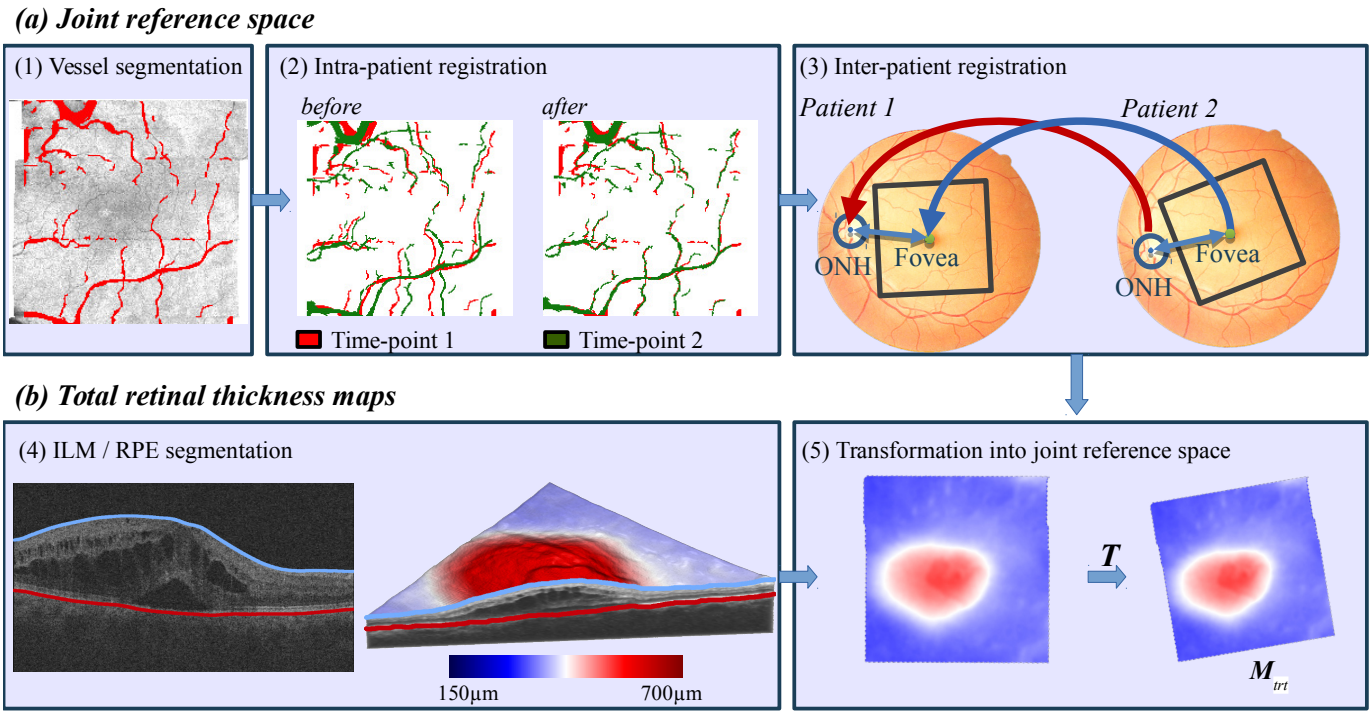

Fig. 2. Processing steps to obtain spatio-temporal disease signatures. (a) Intra-patient registration using vessel segmentation and inter-patient registration usint the fovea center and optic nerve head (ONH) landmarks, respectively, to obtain a joint reference space. (b) Automatic layer segmentation to obtain total retinal thickness (TRT) maps. Finally, TRT maps are transformed into the joint reference space.

The registration pipeline contains following steps. First, in a preprocessing step, we reduce motion artifacts in x-direction introduced by patient and eye movement during acquisition using the method described by Montuoro et al. [27]. The retinal layers are then segmented using a graph-based surface segmentation algorithm [28] with the uppermost and lowermost layers being the inner limiting membrane (ILM) and the retinal pigment epithelium (RPE), respectively (Fig. 2b). We segment the vessel structure in the OCT projection image for the intra-patient registration as proposed by Wu et al. [29]. Vessels, amongst other (pathologic) structures, cause shadows in the layers beneath them. We obtain a projection $\mathcal{P}$ of these structures on a 2D surface by averaging intensity values from the RPE surface to $20\,\mu\mathrm{m}$ towards the ILM. We segment vessel structures after denoising $\mathcal{P}$ using BM3D [30] and masking pathologic shadow structures, which in contrast to vessels are amorphous and not tube shaped, using the multiscale Frangi vesselness filter [31] with windows sizes 4×4, 2×2, 2×1, and 1×2. The final vessel segmentation is obtained by applying region growing on each filter image with pixels having the highest intensity values as seed points and combining them by using the intersection of all candidate segmentations. The intra-patient affine transformation is obtained by applying coherent point drift [32] to the segmented retinal vessel point sets.

The population-wide normalization is performed by aligning the fovea and ONH center across patients. Because the ONH is not visible in macular-centered OCT images, we use the corresponding confocal scanning laser ophthalmoscope (SLO) fundus image, which is registered to the OCT either by scanner software or by using a rigid registration minimizing the normalized cross-correlation of intensity values of the projection image $\mathcal{P}$ and the SLO image. The ONH center is detected automatically by thresholding the adaptive contrast enhanced SLO image and applying a RANSAC-based circle detection [33] which segments the optic disc. The circle center is considered to be the ONH center position. The fovea center landmark is obtained automatically by an algorithm described in [34]. The affine transformation, $\boldsymbol{T}$, that warps a scan into the reference frame consists of scaling, rotation and translation. The scaling parameter is obtained by normalizing the distance between ONH and fovea center to a population mean of $4.3\,\mathrm{mm}$ [35]. The translation parameter is obtained by shifting the fovea center location to the coordinate origin (0,0). The rotation parameter is obtained by rotating around the fovea center such that the angle between the fovea and the center of the optic disc versus the horizon is $5.6°$, which is the reported population mean [36]. The various image and pixel resolutions are unified finally by resampling $\mathcal{P}$ to a 250×250 image with an isotropic pixel resolution of $30\,\mu\mathrm{m}$ and the coordinate origin being at the image center using linear interpolation.

B. Individual Spatio-temporal Signature of Disease

The morphology of the retina and the underlying disease is modeled as a total retinal thickness (TRT) map, $\boldsymbol{M}_{trt}$. It represents the distance between the segmented ILM and RPE layer in $\mu\mathrm{m}$ units at each point (Fig. 2b). The map is transformed into the joint reference space by applying the affine transformation, $\boldsymbol{T}$, obtained previously and resampling

it to the isotropic resolution. Finally, the map is smoothed with a Gaussian of $\sigma = 1.0$ to reduce noise in the data.

An additional map is derived from $\boldsymbol{M}_{trt}$ containing the image gradient magnitude, $\boldsymbol{M}_{gm} := ||\nabla \boldsymbol{M}_{trt}||$, which is computed by filtering $\boldsymbol{M}_{trt}$ using a Gaussian derivate kernel with a σ of 3. The steepness of the retina at each spatial position is modeled with $\boldsymbol{M}_{gm}$. Fig. 3a illustrates the feature maps extracted for an individual patient. From these maps a spatio-temporal signature describing the morphologic status and changes in the retina are created in the following way: Let $\boldsymbol{v}^{(m)}$ be a row vector containing the vectorized values of the thickness map $\boldsymbol{M}_{trt}$ for month m, and $\boldsymbol{v}_G^{(m)}$ be the vectorized $\boldsymbol{M}_{gm}$ maps. We obtain a disease signature vector, $\boldsymbol{x}^{(m)}$ (Fig. 3b) describing the development of a disease up to month m by concatenating the vectorized maps and the change relative to baseline (= month 0), $\Delta\boldsymbol{M}$:

$$\begin{aligned}\boldsymbol{x}^{(m)} = (&\boldsymbol{v}^{(0)}, \boldsymbol{v}^{(1)}, \ldots, \boldsymbol{v}^{(m)}, \boldsymbol{v}^{(1)} - \boldsymbol{v}^{(0)}, \ldots, \boldsymbol{v}^{(m)} - \boldsymbol{v}^{(0)}, \\ &\boldsymbol{v}_G^{(0)}, \boldsymbol{v}_G^{(1)}, \ldots, \boldsymbol{v}_G^{(m)}, \boldsymbol{v}_G^{(1)} - \boldsymbol{v}_G^{(0)}, \ldots, \boldsymbol{v}_G^{(m)} - \boldsymbol{v}_G^{(0)})\end{aligned} \quad (1)$$

Further clinical variables and covariates (e.g. age) can be included by adding them to the feature vector.

Finally, a design matrix $\boldsymbol{X}^{(m)} \in \mathbb{R}^{n \times p}$ is created for training and evaluation of a machine learning algorithm by pooling the signature vectors with a total length of p from n patients. Each row represents a signature vector $\boldsymbol{x}_i^{(m)}$ for a patient, and each column is a distinct spatio-temporal feature from an anatomical region in the retina.

C. Prediction of Disease Recurrence

We predict treatment response in terms of a clinically relevant classification task based on the spatio-temporal signatures. In particular, we predict whether a patient will suffer from recurring edema within the period of 1 year after receiving initial treatment of 3 monthly injections (loading phase). The prediction is based on the signature vector describing the treatment response in this loading phase. Fig. 4 illustrates a time-series of the maps for patients with and without recurrent edema. Formally, we predict from the signature vector $\boldsymbol{x}^{(2)}$ the binary outcome variable, y encoding recurrence of edema within the observed period ($y = 1$) or non-recurrence ($y = 0$) using a predictive model based on the training data, $\boldsymbol{X}^{(2)}$. We use GLM with a sparseness regularization of the coefficients and ET. Both operate in a high-dimension and small sample size setting with feature selection embedded in the algorithms. In the following, we drop the index $^{(m)}$ for notational clarity.

1) Prediction by sparse logistic regression: The sparse logistic regression is based on the GLM, where a linear relation between the input covariates $\boldsymbol{x}$ and a response function $f(y)$ based on the outcome variable y is assumed. The relation is formulated as a weighted linear combination: $f(y) = w_0 + w_1x_1 + w_2x_2 + w_3x_3 + ... + w_px_p = \boldsymbol{xw}$. The probability P of belonging to a class $P = Pr(y = 1|\boldsymbol{x})$ or its log-odds is used for binary outcome variables in the GLM setting: $logit(P) = \log(\frac{P}{1-P}) = \boldsymbol{xw}$.

We imply that only a small subset of features is necessary for a predictive model and thus introduce sparsity of the coefficients as prior information. Tibshirani proposed the $\mathcal{L}_1$ norm on $\boldsymbol{w}$ as regularization, known as LASSO [11]. When features are strongly correlated, LASSO picks only one of them at random and in case of $p > n$ the solution has at most n non-zero weights [37]. Thus, Zou and Hastie proposed a combined $\mathcal{L}_1$ and $\mathcal{L}_2$ regularization to overcome these limitations, denoted as elastic net [38]. We use this approach, because we have a strong correlation of spatially adjacent features.

The weights $\boldsymbol{w}$ are estimated from the design matrix $\boldsymbol{X}$ using a maximum likelihood estimation (MLE) by minimizing following log-likelihood loss function:

$$\begin{aligned}\underset{\boldsymbol{w}}{\operatorname{argmin}} -\frac{1}{n}\sum_{i=1}^{n} y_i \boldsymbol{x}_i^T \boldsymbol{w} - \ln(1 + \exp(\boldsymbol{x}_i^T \boldsymbol{w})) + \\ +\lambda\left(\rho||\boldsymbol{w}||_1 + \frac{1-\rho}{2}||\boldsymbol{w}||_2^2\right)\end{aligned} \quad (2)$$

The non-negative parameter λ specifies the general amount of regularization, whereas ρ defines the ratio between LASSO ($\rho = 1$) and the ridge ($\rho = 0$) penalty. The class probability P is predicted using the trained weights $\boldsymbol{w}$ and the inverse logit for a new case with covariates $\boldsymbol{x}_k$: $p = logit^{-1}(\boldsymbol{x}_k\boldsymbol{w})$.

Features with non-zero coefficients $\boldsymbol{w}$ are characteristic spatio-temporal regions selected for prediction. We are able to identify and interpret these regions by mapping the weight vector back to the 2D images (Fig. 3c).

We used the "glmnet" package from the statistics software R for our computations, which optimizes (2) using a coordinate descent approach [39]. Changes in the regularization parameters, λ and ρ, have a major influence on the predictive performance of the model. Whereas an optimal λ is determined by the "glmnet" algorithm, ρ has to be set beforehand. We use a covariance matrix adaptation evolutionary strategy (CMA-ES) [40] for an efficient hyper-parameter search of ρ, implemented in the "Optunity" software package [41]. We use area under ROC curve (AuC), which is determined in a cross-validation (CV) setup, as the loss function in the CMA-ES optimization process.

2) Prediction by extremly randomized trees: In the ET algorithm, each decision tree is built from a bootstrap sample of the training set. Each split in the tree is chosen among the best split from a random subset of all features. In contrast to RF where the optimal split threshold at each node in the tree is determined, in ET random thresholds are drawn for each candidate feature and the one resulting in the best split is picked. The quality of a split is determined by an impurity criterion. ET uses the normalized information gain based on information entropy [26].

ET provides an "importance" score for each feature that is computed from all nodes where the feature was chosen as splitting criteria, and averaging the reduction of the impurity criterion at each of these nodes to its child nodes over all trees, weighted by the proportion of samples reaching the node. Features chosen at the top of the tree contribute more to the feature importance score due to the larger fraction of input samples. We obtain a feature importance vector, $\boldsymbol{f}_{imp}$, by computing the importance score for each feature. Analogous to the weights from the sparse logistic regression method, $\boldsymbol{f}_{imp}$

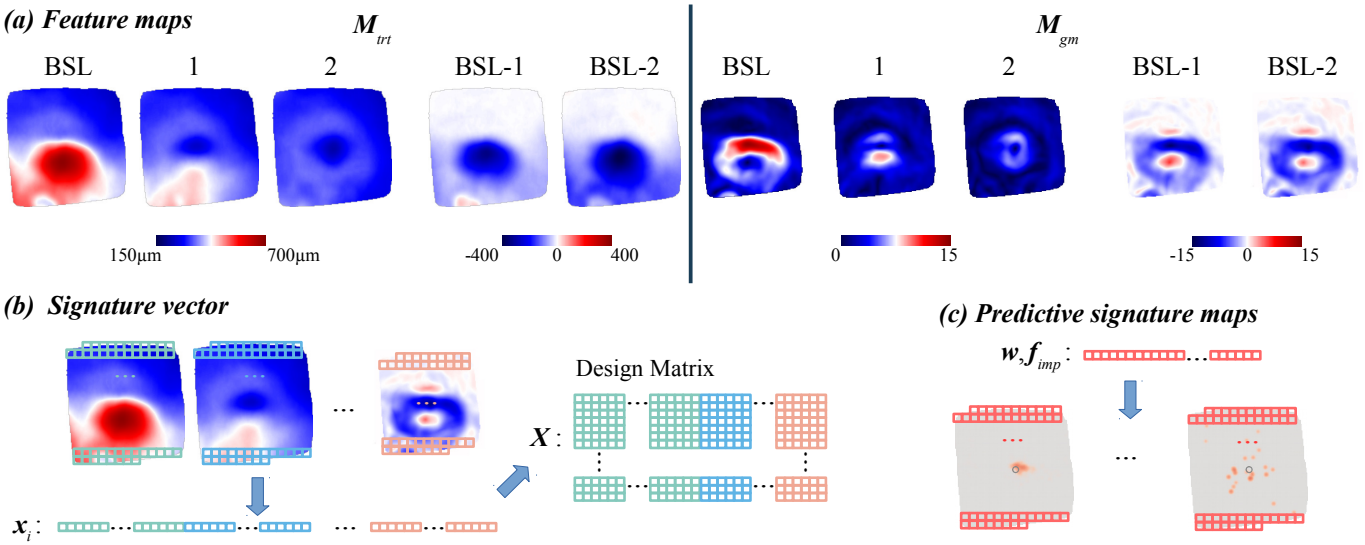

Fig. 3. Feature extraction and interpretation. (a) Feature maps in the joint reference space extracted from one patient's time series. (Left, $\boldsymbol{M}_{trt}$) Total retinal thicknesss (TRTs) for the first 3 months and the difference to baseline (BSL). (Right, $\boldsymbol{M}_{gm}$) Maps of the TRT gradient magnitudes and difference to baseline. (b) The feature maps are vectorized to obtain a signature vector $\boldsymbol{x}_i$. To train a model, the vectors are pooled into a design matrix $\boldsymbol{X}$. (c) Both prediction models provide a measurement of the predictive power of each feature $\boldsymbol{w}$ and $\boldsymbol{f}_{imp}$ for elastic net and extra trees, respectively. By remapping this vector we obtain predictive feature maps, which reveal anatomical regions that have influence on the outcome.

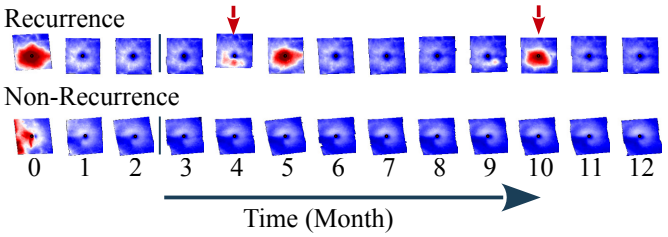

Fig. 4. 12 month follow-up series of total retinal thickness maps in the joint reference space for two patients, one with recurring edema (red arrows), and the other without recurrence. All patients were treated at the first three months (loading phase), followed by a individual treatment after recurring edema. The outcome (recurrence vs. non-recurrence) is predicted from treatment response in the loading phase, encoded as a signature vector.

are mapped back to a 2D image, and are used for interpretation of the model (Fig. 3c).

The main hyper-parameters are the number of trees (n_{trees}) and the number of features to consider when splitting a node (m_{feat}). Empirically, good default values for m_{feat} are in the range of $\sqrt{p}$ to $1/2p$, where p is the total number of features. In our experiments, we set $n_{trees} = 500$ and determined the optimal value of m_{feat} within the proposed range using the CMA-ES algorithm described above. The out-of-bag (OOB) error estimate, which is the error obtained by testing each tree on samples not included in the bootstrap set of that tree, is used as the loss function. In our experiments we used the ET implementation from the "scikit-learn" toolbox (Version 0.16.1) [42].

Using the method described, we predict the outcome for a new patient receiving treatment based on extracted imaging features that are transformed in a joint reference space combined with the prediction model trained on examples observed previously. The embedded feature selection in the prediction model serves as the basis for interpretation of predictive spatio-temporal features.

III. Experiments & Results

A. Patient Data

We evaluated the method on a dataset containing one eye each from patients with CRVO ($n = 155$) and with BRVO ($n = 92$) followed over 12 months. The study was conducted in compliance with the Declaration of Helsinki and all participants provided written informed consent before inclusion. Approval was obtained from the ethics committee at the Medical University of Vienna and at each study site where images were acquired. All retinal SD-OCT images were acquired from patients at monthly intervals. Patients received ranibizumab injections for the first 3 months (loading phase), followed by treatment based on individual need that depended on visual acuity response and OCT fluid variables.

6×6-mm macular-centered volumetric images were acquired at two different scanner types, Cirrus HD-OCT® (Carl Zeiss Meditec, Dublin, CA; $n_{CRVO} = 44, n_{BRVO} = 33$), having a resolution of $200 \times 200 \times 1024$ and a voxel spacing of approx. $30\,\mu\text{m} \times 30\,\mu\text{m} \times 1.96\,\mu\text{m}$, and Spectralis OCT® (Heidelberg Engineering, Dossenheim, GER, $n_{CRVO} = 111, n_{BRVO} = 59$), having a resolution of $512 \times 49 \times 496$ and a voxel spacing of $11\,\mu\text{m} \times 119\,\mu\text{m} \times 3.87\,\mu\text{m}$.

This dataset allowed us to evaluate and assess the heterogeneous characteristics of retinal vein occlusion diseases by including two disease subtypes. By pooling data from two scanner types and transforming them into the joint reference space, we were able to increase the total number of samples and to provide comparable time-series, which are independent of resolution and specific imaging properties originating from the different scanner types.

Total retinal thickness and gradient magnitude maps were computed and transformed into the joint reference space

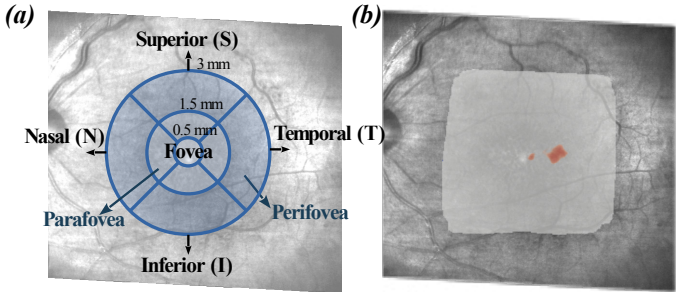

Fig. 5. Scanning laser ophthalmoscope (SLO) fundus image of a left eye with the optic disk located nasally. (a) Early treatment diabetic retinopathy study (ETDRS) grid dividing the macula into 9 subregions of interest as defined in [44]. Anatomic directions: nasal (N), temporal (T), superior (S) and inferior (I). (b) Overlay of a predictive signature map.

as described in Section II-A. Spectralis scanners provide a hardware motion correction via eye-tracking and an alignment of OCT follow-up scans via scanner software. Hence, motion correction and intra-patient registration were skipped for this scanner type. Due to varying field-of-views in the registered SD-OCT images not all scans cover the whole 3mm radius around the fovea. Thus, pixels in the TRT feature maps with more than 5% missing values were discarded and the other missing values were imputed using kNN imputation [43] using the 10 nearest TRT maps. Hence, the final feature map covered a 4.74 mm and 5.13 mm diameter in CRVO and in BRVO, respectively. We verified the correctness of intra-patient and inter-patient registration in all cases by validating the correct alignment of follow-up images by overlaying the en-face projections $\mathcal{P}$ and by validating the correct placement of fovea and ONH center positions. Furthermore, the correct segmentation of the layers have been approved by visual inspection of all TRT maps.

The time-point of recurring edema was determined algorithmically, serving as a standard-of-reference for evaluation. First, the aligned total retinal thickness maps, $\boldsymbol{M}_{trt}$, were divided into nine circular sections centered at the fovea, within three concentric circles of diameters 1, 3 and 6mm, as defined by the early treatment diabetic retinopathy study (ETDRS) design [44] (Fig. 5a). The mean thickness was computed for each section. An increase in mean thickness of more than $29\,\mu\mathrm{m}$ between two subsequent timepoints in any region was defined as recurrent edema. This threshold had been determined based on 46 cases, where time-point of recurrence was manually assigned. 28 patients ($= 18\%$) with CRVO and 20 ($= 22\%$) with BRVO did not show any recurrent edema within the 12 month follow-up.

B. Evaluation

The evaluation covered the two main aspects of the proposed technique that the models predict outcome accurately and provide interpretable results. First, we evaluated the algorithmic *prediction accuracy* for the two different regression approaches based on the known outcome at month 12 (edema recurrence). We varied the number of months and specific features taken into account. Second, we assessed the automatically identified predictive signatures observed in the imaging data and how hyper-parameters influenced signature interpretability.

1) Prediction accuracy: We measured the generalization of the proposed models using a nested five-fold stratified CV partitioning on the patient level to iteratively split the data into training and test sets. ET and GLM models were trained on each fold with an increasing number of months available, up to 3 months, to quantify the benefit of using longitudinal information as opposed to only baseline data. To assess the efficacy of the feature groups for prediction accuracy, we computed models with combinations from the features $\boldsymbol{M}_{trt}$, $\boldsymbol{M}_{gm}$ and $\Delta\boldsymbol{M}$. The performance measure we used was ROC AuC computed from the probabilistic outcomes of the classifiers on the test folds as well as sensitivity and specificity computed from a threshold of the probabilities. The optimal threshold was determined by maximizing the F_1-score on the training fold and applying it to the outcome of the test fold. F_1-score is the harmonic mean of sensitivity and precision. It is computed as: $2 \cdot TP/(2 \cdot TP + FP + FN)$, where TP, FP, and FN are the number of true positive, false positive and false negative cases, respectively. Due to the randomness in CV partitioning and the ET algorithm, we repeated the evaluation 20 times and computed the mean AuC by vertical averaging [45] as well as the sensitivity and specificity from the summed up confusion matrices. Furthermore, 95 % confidence intervals (CIs) for ROC AuC were obtained by stratified bootstrapping using the R package "pROC" [46]. CIs for sensitivity/specificity were computed from the mean values of the confusion matrices using Newcombe's efficient-score method [47]. The hyper-parameters ρ, λ and m_{feat} were tuned on the training folds with 100 iterations using the CMA-ES algorithm combined respectively with the AuC and OOB error loss function, as described in the Methods section.

2) Predictive signature maps: Because each feature of the predictive signature corresponds to a distinct spatio-temporal location, the vectors $\boldsymbol{w}$ and $\boldsymbol{f}_{imp}$ can be mapped back to the 2D retina projection, resulting in coefficient maps and feature importance maps, respectively (Fig. 5b). For clarity, we denote these maps subsequently as *predictive signature maps*. Our evaluation of the predictive signature maps was two-fold. First, we computed predictive signature maps from the best-performing models for each disease and assessed them from a clinical perspective. Second, we analyzed according to Rasmussen et al. [15] how the hyper-parameters λ and ρ, as well as m_{feat} and n_{trees} influence the feature selection and interpretability of the predictive signature maps, by varying each of these paramaters while keeping the other one fixed.

a) Parameters: We chose the best-performing models and hyper-parameters, i.e. those with the highest AuC score of the prediction accuracy evaluation, for the assessment of the predictive signature maps. These were the 3-month $\boldsymbol{M}_{trt}$ & $\boldsymbol{M}_{gm}$ & $\Delta\boldsymbol{M}$ feature-set for BRVO and the 3-month $\boldsymbol{M}_{trt}$ & $\Delta\boldsymbol{M}$ for CRVO. ρ was set to 0.1 and λ tuned by the glmnet algorithm such that the prediction error is minimized in a five-fold CV. We denote this value as min_{cv}. Alternatively, λ was set to be the most regularized version that is within 1 standard error (SE) to the min_{cv} model, as proposed in [48]. This so-called "one-standard error" rule results in a simpler but

still predictive model, and may lead to a better interpretability [16]. We denote the model as $1se_{cv}$. In the ET model, the parameters were set to the following values: $n_{trees} = 5000$, $m_{feat} = 0.5$. Whereas the highest performance was obtained already with fewer trees and features, the accuracy did not drop by extending the forest and number of features, although the predictive signature maps became less noisy with an increasing number of trees.

The influence of the hyper-parameters were evaluated by fixing one parameter and varying the other one when computing a model. For the logistic regression models, λ was fixed at min_{cv} and ρ varied over $\{0.5, 0.1, 0.01, 0.001\}$. For a fixed ρ of 0.01, we computed models for λ from $\{1.0, min_{cv}, 1se_{cv}, 15.0\}$. The ET models were computed for m_{feat} of $0.5 \times p$ and n_{trees} varied over $\{50, 100, 500, 5000\}$, where p is the overall number of features. Finally, for a fixed n_{trees} of 5000 we trained models for $m_{feat} = \{\log_2 p, \sqrt{p}, p/3, p/2\}$.

C. Results

1) Prediction accuracy: All classification results are listed in Table I and Table II for BRVO and CRVO, respectively. In BRVO, the recurrence of edema was predicted from the first 3 months with an AuC of 0.83 (Fig. 6a) and a sensitivity/specificity of 0.77/0.77 using the feature-set $\boldsymbol{M}_{trt}$ & $\boldsymbol{M}_{gm}$ & $\boldsymbol{\Delta M}$ and the ET model. Using features from the first month only resulted in a drop of AuC to 0.77. Better results were obtained in BRVO using extra trees than by logistic regression, which had a maximum AuC of 0.78. Prediction performance in CRVO was slightly inferior to BRVO with an AuC of 0.79 and a sensitivity/specificity of 0.67/0.79. Logistic regression outperformed extra trees (Fig. 6b). Accuracy using features from the first month only was already high with an AuC of 0.78 and 0.77 for logistic regression and extra trees, respectively. Prediction in BRVO also benefited from the derived features $\boldsymbol{M}_{gm}$ and $\boldsymbol{\Delta M}$ by improving the sensitivity. Although, upon adding $\boldsymbol{M}_{gm}$ to the CRVO model, the AuC and sensitivity declined in the logistic regression model, and, to a lesser degree, in the ET model. ET feature selection seems to be less sensitive to irrelevant features, and provides a more constant prediction accuracy when including or excluding $\boldsymbol{M}_{gm}$ features.

2) Predictive signature maps: Predictive signature maps for the models with the highest AuC values are presented in Fig. 7 (BRVO) and Fig. 8 (CRVO). Both methods selected distinct predictive regions for both diseases. There is an overlap between regions selected by ET and by min_{cv} logistic regression, with ET regions being more extended. The $\lambda = 1se_{cv}$ model is less noisy than $\lambda = min_{cv}$ and selected mostly regions from baseline foveal and parafoveal areas.

In Fig. 9, we illustrate on one example how changes in hyper-parameter values affect the predictive signature maps. Decreasing the amount of regularization by lowering λ resulted in additional predictive areas appearing. Increasing the ratio of $\mathcal{L}_2$ regularization by decreasing ρ caused mainly a dilation of the areas around already selected sparse regions. For the ET model, increasing n_{trees} or m_{feat} led to a sharper

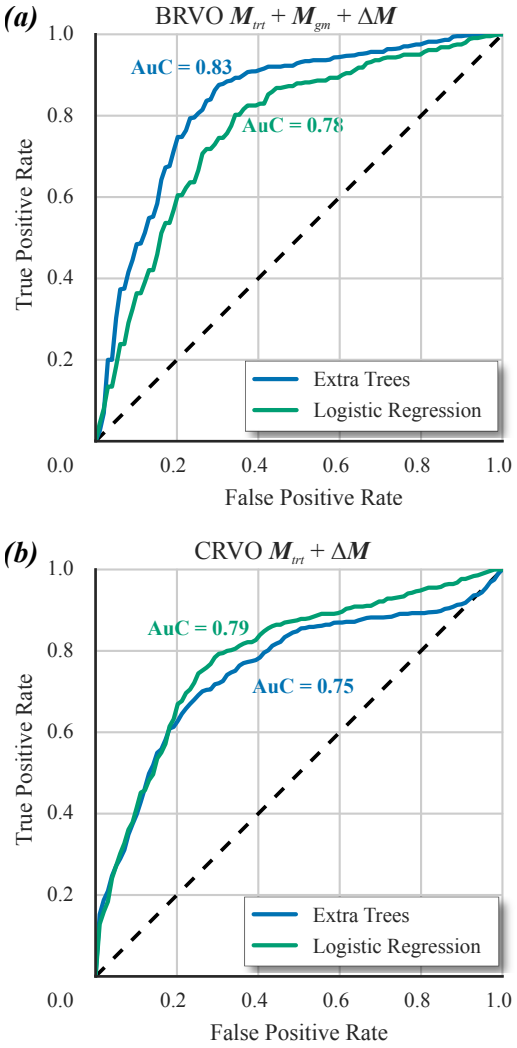

Fig. 6. Area under ROC curve (AuC) curve for (a) branch retinal vein occlusion (BRVO) and (b) central retinal vein occlusion (CRVO) with the best-performing feature set. The 3-month $\boldsymbol{M}_{trt}$ & $\boldsymbol{M}_{gm}$ & $\boldsymbol{\Delta M}$ feature set was used for BRVO and the 3-month $\boldsymbol{M}_{trt}$ & $\boldsymbol{\Delta M}$ for CRVO.

feature relevance map with more defined "important" areas, as illustrated in 9b. A reduced number of n_{trees} caused a noisier and speckled predictive signature map. A reduction in m_{feat} led to a more diffuse and blurred map.

IV. Discussion and Conclusion

In this paper, we presented a longitudinal data-driven method to predict the *future* treatment outcome in terms of recurring edema based on imaging features extracted from SD-OCT images. By transforming these imaging features into a joint reference space based on intra- and inter-patient registration we were able to compensate for anatomical variations and scanning positions, and to maintain vendors and spatial resolutions as invariant. We evaluated two different models operating with high dimensional data and low sample-size, which are elastic net regularized logistic regression and extra trees, a tree-based ensembling method. We demonstrated that both of the approaches analyzed yield predictive and

TABLE I
CLASSIFICATION PERFORMANCE ON BRANCH RETINAL VEIN OCCLUSION (BRVO) DATASET USING TOTAL RETINAL THICKNESS MAPS ($\boldsymbol{M}_{trt}$) AND GRADIENT MAGNITUDE ($\boldsymbol{M}_{gm}$) FEATURES, AS WELL AS CHANGE OF VALUES RELATIVE TO BASELINE $\Delta\boldsymbol{M}$. VALUES IN BRACKETS ARE THE 95 % CONFIDENCE INTERVALS (CIS).

		Extra Trees			Logistic Regression		
	# Months	AuC	Sensitivity	Specificity	AuC	Sensitivity	Specificity
$\boldsymbol{M}_{trt}$ & $\boldsymbol{M}_{gm}$	3	**0.83** (0.81/0.85)	0.74 (0.51/0.89)	0.78 (0.66/0.86)	0.78 (0.75/0.80)	0.67 (0.44/0.84)	0.76 (0.64/0.85)
$\boldsymbol{M}_{trt}$ & $\boldsymbol{M}_{gm}$ & $\Delta\boldsymbol{M}$	3	**0.83** (0.81/0.85)	**0.77** (0.54/0.91)	0.77 (0.65/0.86)	0.77 (0.73/0.79)	0.67 (0.44/0.84)	0.76 (0.64/0.85)
$\boldsymbol{M}_{trt}$	3	0.73 (0.70/0.76)	0.73 (0.50/0.89)	0.60 (0.48/0.71)	0.78 (0.75/0.80)	0.64 (0.41/0.82)	0.80 (0.68/0.88)
$\boldsymbol{M}_{trt}$ & $\Delta\boldsymbol{M}$	3	0.81 (0.78/0.83)	0.73 (0.50/0.88)	0.80 (0.68/0.88)	0.78 (0.76/0.80)	0.63 (0.41/0.82)	0.80 (0.68/0.88)
$\boldsymbol{M}_{trt}$ & $\boldsymbol{M}_{gm}$	2	0.79 (0.77/0.82)	0.72 (0.49/0.88)	0.75 (0.63/0.84)	0.73 (0.70/0.75)	0.60 (0.38/0.80)	0.76 (0.65/0.85)
$\boldsymbol{M}_{trt}$ & $\boldsymbol{M}_{gm}$ & $\Delta\boldsymbol{M}$	2	0.80 (0.77/0.82)	0.72 (0.49/0.88)	0.74 (0.62/0.83)	0.78 (0.75/0.80)	0.65 (0.42/0.83)	0.77 (0.65/0.86)
$\boldsymbol{M}_{trt}$	2	0.78 (0.75/0.80)	0.60 (0.38/0.79)	**0.82** (0.70/0.89)	0.76 (0.72/0.77)	0.62 (0.39/0.81)	0.79 (0.67/0.87)
$\boldsymbol{M}_{trt}$ & $\Delta\boldsymbol{M}$	2	0.78 (0.75/0.80)	0.64 (0.41/0.82)	0.79 (0.67/0.87)	0.77 (0.74/0.79)	0.62 (0.39/0.81)	0.80 (0.68/0.88)
$\boldsymbol{M}_{trt}$ & $\boldsymbol{M}_{gm}$	1	0.80 (0.77/0.82)	0.74 (0.51/0.89)	0.75 (0.63/0.84)	0.75 (0.72/0.77)	0.65 (0.42/0.83)	0.75 (0.63/0.84)
$\boldsymbol{M}_{trt}$	1	0.76 (0.74/0.79)	0.62 (0.40/0.81)	0.76 (0.64/0.85)	0.74 (0.71/0.76)	0.62 (0.40/0.81)	0.77 (0.65/0.85)

TABLE II
CLASSIFICATION PERFORMANCE ON CENTRAL RETINAL VEIN OCCLUSION (CRVO) DATASET USING TOTAL RETINAL THICKNESS MAPS ($\boldsymbol{M}_{trt}$) AND GRADIENT MAGNITUDE ($\boldsymbol{M}_{gm}$) FEATURES, AS WELL AS CHANGE OF VALUES RELATIVE TO BASELINE $\Delta\boldsymbol{M}$. VALUES IN BRACKETS ARE THE 95 % CONFIDENCE INTERVALS (CIS).

		Extra Trees			Logistic Regression		
	# Months	AuC	Sensitivity	Specificity	AuC	Sensitivity	Specificity
$\boldsymbol{M}_{trt}$ & $\boldsymbol{M}_{gm}$	3	0.75 (0.72/0.77)	0.61 (0.41/0.78)	0.80 (0.72/0.87)	0.71 (0.68/0.73)	0.54 (0.34/0.72)	0.80 (0.72/0.86)
$\boldsymbol{M}_{trt}$ & $\boldsymbol{M}_{gm}$ & $\Delta\boldsymbol{M}$	3	0.75 (0.72/0.77)	0.62 (0.42/0.79)	0.79 (0.71/0.86)	0.72 (0.69/0.74)	0.51 (0.32/0.70)	0.82 (0.73/0.88)
$\boldsymbol{M}_{trt}$	3	0.76 (0.74/0.79)	0.63 (0.43/0.80)	0.81 (0.73/0.87)	0.77 (0.75/0.79)	0.63 (0.43/0.80)	0.81 (0.72/0.87)
$\boldsymbol{M}_{trt}$ & $\Delta\boldsymbol{M}$	3	0.75 (0.73/0.78)	0.59 (0.39/0.76)	0.82 (0.74/0.88)	**0.79** (0.77/0.81)	**0.67** (0.47/0.83)	0.79 (0.71/0.86)
$\boldsymbol{M}_{trt}$ & $\boldsymbol{M}_{gm}$	2	0.73 (0.71/0.76)	0.59 (0.39/0.76)	0.79 (0.70/0.85)	0.70 (0.68/0.73)	0.53 (0.34/0.72)	0.81 (0.73/0.87)
$\boldsymbol{M}_{trt}$ & $\boldsymbol{M}_{gm}$ & $\Delta\boldsymbol{M}$	2	0.74 (0.71/0.76)	0.62 (0.42/0.79)	0.76 (0.68/0.83)	0.74 (0.72/0.76)	0.52 (0.33/0.71)	0.82 (0.74/0.88)
$\boldsymbol{M}_{trt}$	2	0.74 (0.72/0.77)	0.61 (0.41/0.78)	0.79 (0.71/0.86)	0.77 (0.75/0.79)	0.62 (0.42/0.79)	0.79 (0.71/0.86)
$\boldsymbol{M}_{trt}$ & $\Delta\boldsymbol{M}$	2	0.76 (0.73/0.78)	0.62 (0.42/0.79)	0.82 (0.74/0.88)	0.78 (0.75/0.80)	0.64 (0.44/0.81)	0.79 (0.71/0.86)
$\boldsymbol{M}_{trt}$ & $\boldsymbol{M}_{gm}$	1	0.76 (0.73/0.78)	0.61 (0.41/0.78)	0.80 (0.72/0.86)	0.77 (0.75/0.79)	0.56 (0.36/0.74)	0.83 (0.75/0.89)
$\boldsymbol{M}_{trt}$	1	0.77 (0.74/0.79)	0.63 (0.43/0.80)	**0.84** (0.76/0.89)	0.78 (0.76/0.80)	0.66 (0.46/0.82)	0.79 (0.70/0.85)

interpretable results in terms of predictive signatures from the initial treatment phase of 3 months.

1) Predictive signature maps: "Baseline only" already provided a high AuC in both diseases, and importantly, with a slight increase when adding additional months. This can be observed in the predictive signature maps as well, where the $\lambda = 1se_{cv}$ model selected features from baseline only because these were the most predictive ones. This observation is consistent with the findings of Rasmussen [15], who concluded that not necessarily the model with the highest accuracy results in the best interpretable model.

That the central total retinal thickness at baseline is an important predictor, corresponds well with the literature [49] for CRVO, where patients with a dry interval of more than 25 weeks at the last visit had a thinner central retinal thickness at baseline, often a complete resolution of macular edema after the first injection and were significantly younger. In addition, thickness change relative to baseline $\Delta\boldsymbol{M}$ got a high ranking, which indicates that a faster thinning is beneficial for not having recurrent edema. That central total retinal thickness is also important in BRVO is a new and unexpected finding, because edemas occur in BRVO in parafoveal and perifoveal areas in contrast to CRVO where edemas are central. It seems that if an edema extends over the foveal center the likelihood of a recurrence after treatment is greater. Considering the vascular anatomy of the central retina, the fovea itself is free from blood vessels to allow complete transparency. Thus, a foveal edema in BRVO may be a sign of a more severe leakage activity which is not confined to the occluded vascular bed itself but also spreads to adjacent, vessel-free retinal areas.

The ET model performed better than the logistic regression model in BRVO. The reason can be observed in the predictive feature map, where the ET model selected more regions for prediction than the logistic regression model, in particular from the $\boldsymbol{M}_{gm}$ changes relative to baseline and the $\boldsymbol{M}_{gm}$ features in the parafoveal nasal area.

2) Influence of hyper-parameters: In sparse logistic regression, the parameter ρ defines the ratio between $\mathcal{L}_1$ and $\mathcal{L}_2$. As seen in Fig. 9a, decreasing ρ mainly caused an extension around already selected sparse regions. Neighboring pixels were strongly correlated because the retina is inherently locally smooth and in addition a Gaussian filter was applied on the thickness maps, As ridge regularization tends to assign similar weights to strongly correlated features [16], a smaller ρ value up-regulates $\mathcal{L}_2$, which forces additional local correlated features to enter the model.

The parameter λ defines the general amount of regularization. Setting λ according to the "one-standard error" rule, led to less noisy feature relevance maps, highlighting only highly predictive areas (Fig. 9a). Lowering the regularization allowed other regions to enter the model. However, too little regularization causes overfitting to data and may lead to misinterpretation [15].

In the ET model, increasing m_{feat} resulted in sharper

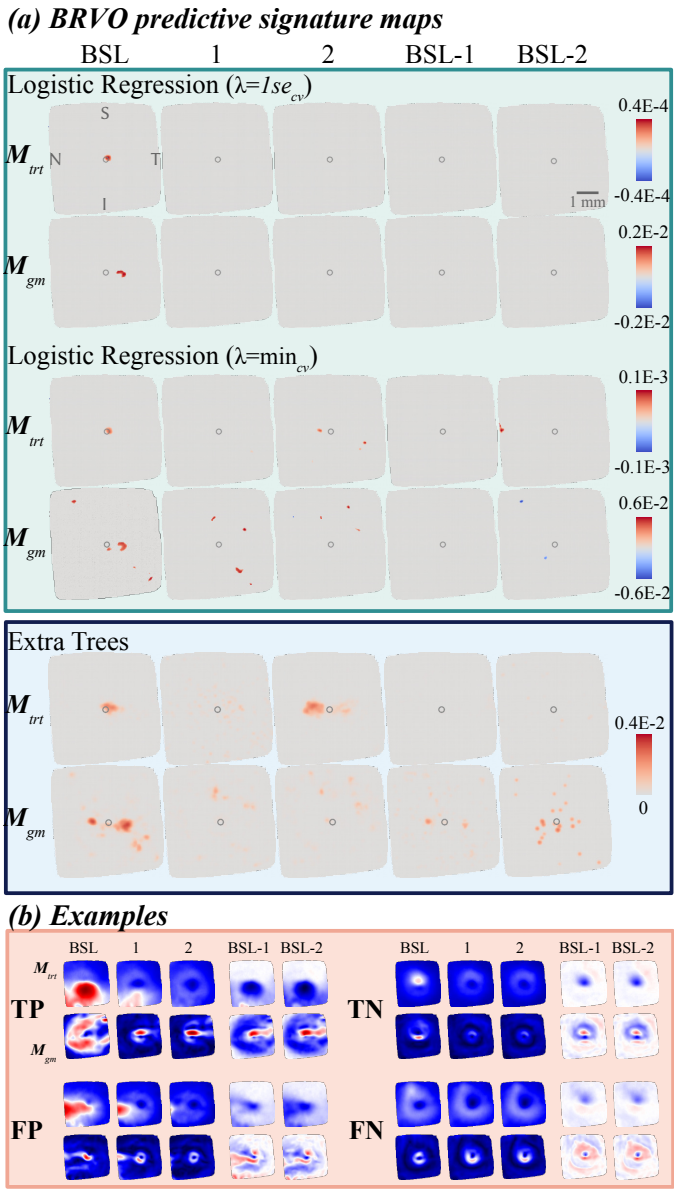

Fig. 7. (a) Predictive signature maps of $\boldsymbol{M}_{trt}$ and $\boldsymbol{M}_{gm}$ for branch retinal vein occlusion (BRVO) from baseline (BSL), month 1 and month 2 as well as the change relative to BSL, $\Delta\boldsymbol{M}$. The following parameter sets were used $\rho = 0.1$,$\lambda = 1se_{cv}$ and $\lambda = \min_{cv}$; $n_{trees} = 5000$, $m_{feat} = 0.5$. The gray circle at the map center determines the fovea position. Note that logistic regression selects $\boldsymbol{M}_{trt}$ features at the fovea center at BSL and a small parafoveal area at month 2, with relative low coefficient values. (b) Total retinal thickness and gradient magnitude feature maps of examples from true positive (TP), true negative (TN), false positive (FP) and false negative (FN) cases, respectively.

predictive signature maps. As Geurts et al. [26] stated, in the case of a high percentage of irrelevant features an increase in m_{feat} leads to a better chance of filtering out these variables and increases the probability of picking a feature with a higher amount of information gain at each split, resulting in a higher average information gain per split. Increasing the parameter n_{trees} reduced the noise in the maps. Because the total number of features is high and neighboring features are highly correlated, having a small number of trees results in a failure to select all predictive features, because in a split

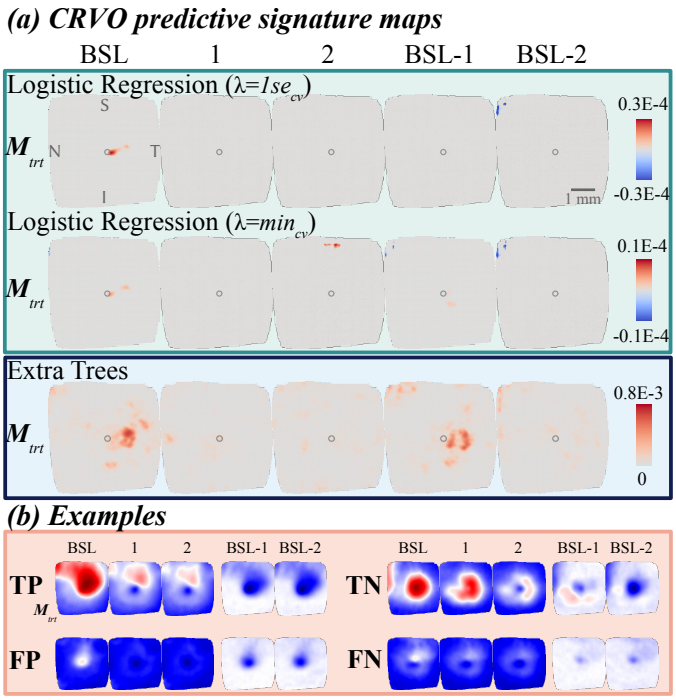

Fig. 8. (a) Predictive signature maps, $\boldsymbol{M}_{trt}$ for central retinal vein occlusion (CRVO) from the first 3 months and the change relative to baseline (BSL), $\Delta\boldsymbol{M}$. The following parameter sets were used $\rho = 0.1$,$\lambda = 1se_{cv}$ and $\lambda = \min_{cv}$; $n_{trees} = 5000$, $m_{feat} = 0.5$. The gray circle at the map center determines the fovea position. (b) Total retinal thickness feature maps of examples from true positive (TP), true negative (TN), false positive (FP) and false negative (FN) cases, respectively.

only one of the predictive correlated features is chosen. With a large number of trees, on average overall better performing features have a higher chance of being selected in a split.

3) Limitations: A limitation of this study is that we only analyzed imaging data, without considering other known factors that influence outcome such as age and blood pressure [49]. Clinical data was not available for this study. However, additional data can easily be incorporated into the model by extending the feature vector by these additional covariates. A further limitation of our models is that they cannot handle missing visits and the visit intervals must be fixed. However, outcome can be already predicted with a reasonable accuracy having the baseline scan only. Finally, the proposed algorithms are not capable of handling missing values at the borders of the feature maps, which occur due to off-centered image acquisitions. Though, as the predictive feature maps showed, most of the predictive regions are located close to the fovea center. Thus, a model based on a smaller field-of-view can be created, where SD-OCT images needs to be less accurately centered at the fovea.

4) Future work: Whereas we focused in this work on a particular treatment outcome in a specific retinal disease using total retinal thickness maps, the proposed methods could be easily extended to predict other treatment outcomes such as visual acuity in a broad spectrum of retinal diseases as for instance AMD or diabetic retinopathy. Furthermore, signatures identified in other pathologic structures segmented in the images such as fluid compartments or observable changes in

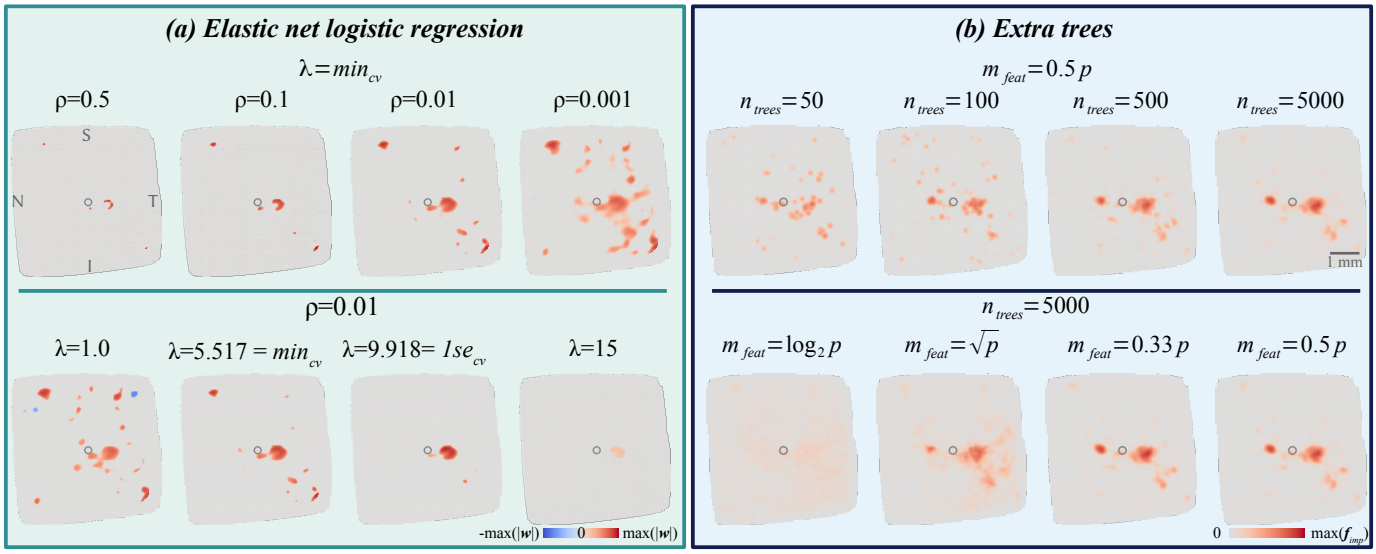

Fig. 9. Influence of the hyper-parameters ρ and λ on the regression coefficients $\boldsymbol{w}$ in logistic regression, as well as n_{trees} and m_{feat} on feature importance $\boldsymbol{f}_{imp}$ in extra trees (ET) models, where p is the total number of features. One of the parameters is fixed while the other is varied. The values min_{cv} and $1se_{cv}$ are the λ values with the highest area under ROC curve (AuC) score respectively the most regularized version where the error is within 1 standard error. Feature importance maps were smoothed with Gaussian $\sigma = 1$. In this illustration, predictive signature maps from branch retinal vein occlusion (BRVO) $\boldsymbol{M}_{gm}$ baseline features are used. The gray circle at the map center determines the fovea position.

photoreceptor cells may improve the predictive power of the models.

5) Conclusion: Our method shows potential for clinical application in the care of patients with RVO and potentially other indications. Findings based on the feature maps obtained may serve as robust and clinically meaningful imaging biomarkers applicable in physicians' daily practice. Machine learning in retinal imaging promises to introduce solid precision medicine standards into the care of patients with retinal vascular disease, one of the major vision-threatening diseases of modern times.

REFERENCES

[1] R. Mirnezami, J. Nicholson, and A. Darzi, "Preparing for precision medicine," *New England Journal of Medicine*, vol. 366, no. 6, pp. 489–491, Feb 2012.

[2] H. J. Aerts, E. R. Velazquez, R. T. Leijenaar, C. Parmar, P. Grossmann, S. Carvalho, J. Bussink, R. Monshouwer, B. Haibe-Kains, D. Rietveld *et al.*, "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nature communications*, vol. 5, 2014.

[3] U. Schmidt-Erfurth and S. M. Waldstein, "A paradigm shift in imaging biomarkers in neovascular age-related macular degeneration," *Progress in Retinal and Eye Research*, vol. 50, pp. 1–24, Jan 2016.

[4] R. Channa, M. Smith, and P. A. Campochiaro, "Treatment of macular edema due to retinal vein occlusions," *Clinical ophthalmology (Auckland, NZ)*, vol. 5, p. 705, 2011.

[5] S. Rogers, R. L. McIntosh, N. Cheung, L. Lim, J. J. Wang, P. Mitchell, J. W. Kowalski, H. Nguyen, T. Y. Wong, I. E. D. Consortium *et al.*, "The prevalence of retinal vein occlusion: pooled data from population studies from the united states, europe, asia, and australia," *Ophthalmology*, vol. 117, no. 2, pp. 313–319.e1, Feb 2010.

[6] M. Laouri, E. Chen, M. Looman, and M. Gallagher, "The burden of disease of retinal vein occlusion: review of the literature," *Eye*, vol. 25, no. 8, pp. 981–988, Aug 2011.

[7] P. A. Campochiaro, D. M. Brown, C. C. Awh, S. Y. Lee, S. Gray, N. Saroj, W. Y. Murahashi, and R. G. Rubio, "Sustained benefits from ranibizumab for macular edema following central retinal vein occlusion: Twelve-month outcomes of a phase iii study," *Ophthalmology*, vol. 118, no. 10, pp. 2041–2049, Oct 2011.

[8] D. M. Brown, P. A. Campochiaro, R. B. Bhisitkul, A. C. Ho, S. Gray, N. Saroj, A. P. Adamis, R. G. Rubio, and W. Y. Murahashi, "Sustained benefits from ranibizumab for macular edema following branch retinal vein occlusion: 12-month outcomes of a phase iii study," *Ophthalmology*, vol. 118, no. 8, pp. 1594–1602, Aug 2011.

[9] R. Merani and A. P. Hunyor, "Endophthalmitis following intravitreal anti-vascular endothelial growth factor (VEGF) injection: a comprehensive review," *International Journal of Retina and Vitreous*, vol. 1, no. 1, Dec 2015.

[10] A. Fercher, C. Hitzenberger, G. Kamp, and S. El-Zaiat, "Measurement of intraocular distances by backscattering spectral interferometry," *Optics Communications*, vol. 117, pp. 43–48, May 1995.

[11] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[12] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, vol. 67, no. 2, pp. 301–320, Apr 2005.

[13] B. M. Kandel, D. A. Wolk, J. C. Gee, and B. Avants, "Predicting cognitive data from medical images using sparse linear regression," in *Information Processing in Medical Imaging: 23rd International Conference, IPMI 2013, Asilomar, CA, USA, June 28–July 3, 2013. Proceedings*, J. C. Gee, S. Joshi, K. M. Pohl, W. M. Wells, and L. Zöllei, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 86–97.

[14] M. K. Carroll, G. A. Cecchi, I. Rish, R. Garg, and A. R. Rao, "Prediction and interpretation of distributed neural activity with sparse models," *NeuroImage*, vol. 44, no. 1, pp. 112–122, Jan 2009.

[15] P. M. Rasmussen, L. K. Hansen, K. H. Madsen, N. W. Churchill, and S. C. Strother, "Model sparsity and brain pattern interpretation of classification models in neuroimaging," *Pattern Recognition*, vol. 45, no. 6, pp. 2085–2100, Jun 2012.

[16] T. Hastie, R. Tibshirani, and M. Wainwright, *Statistical learning with sparsity: the lasso and generalizations*. CRC Press, 2015.

[17] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[18] G. Langs, B. H. Menze, D. Lashkari, and P. Golland, "Detecting stable distributed patterns of brain activation using gini contrast," *NeuroImage*, vol. 56, no. 2, pp. 497–507, May 2011.

[19] E. Geremia, D. Zikic, O. Clatz, B. H. Menze, B. Glocker, E. Konukoglu, J. Shotton, O. M. Thomas, S. J. Price, T. Das, R. Jena, N. Ayache, and A. Criminisi, *Decision Forests for Computer Vision and Medical Image Analysis*. London: Springer London, 2013, ch. Classification Forests for Semantic Segmentation of Brain Lesions in Multi-channel MRI, pp. 245–260.

[20] R. Díaz-Uriarte and S. Alvarez de Andrés, "Gene selection and classification of microarray data using random forest," *BMC Bioinformatics*, vol. 7, no. 1, pp. 1–13, 2006.

[21] U. R. Acharya, S. Dua, X. Du, V. Sree S, and C. K. Chua, "Automated diagnosis of glaucoma using texture and higher order spectra features," *IEEE Transactions on Information Technology in Biomedicine*, vol. 15, no. 3, pp. 449–455, May 2011.

[22] L. de Sisternes, N. Simon, R. Tibshirani, T. Leng, and D. L. Rubin, "Quantitative SD-OCT imaging biomarkers as indicators of age-related macular degeneration progressionpredicting amd progression using SD-OCT features," *Investigative ophthalmology & visual science*, vol. 55, no. 11, pp. 7093–7103, 2014.

[23] H. Bogunović, M. D. Abràmoff, L. Zhang, and M. Sonka, "Prediction of treatment response from retinal oct in patients with exudative age-related macular degeneration," in *Proceedings of the Ophthalmic Medical Image Analysis First International Workshop, OMIA 2014, Held in Conjunction with MICCAI 2014, Boston, Massachusetts, September 14, 2014*, L. J. Chen X, Garvin MK, Ed. Iowa Research Online, 2014, pp. 129–136.

[24] S. Niu, L. de Sisternes, Q. Chen, D. L. Rubin, and T. Leng, "Fully automated prediction of geographic atrophy growth using quantitative spectral-domain optical coherence tomography biomarkers," *Ophthalmology*, vol. 123, no. 8, pp. 1737–1750, Aug 2016.

[25] W.-D. Vogl, S. M. Waldstein, B. S. Gerendas, C. Simader, A.-M. Glodan, D. Podkowinski, U. Schmidt-Erfurth, and G. Langs, *Information Processing in Medical Imaging: 24th International Conference, IPMI 2015, Sabhal Mor Ostaig, Isle of Skye, UK, June 28 - July 3, 2015, Proceedings*. Cham: Springer International Publishing, 2015, ch. Spatio-Temporal Signatures to Predict Retinal Disease Recurrence, pp. 152–163.

[26] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Machine Learning*, vol. 63, no. 1, pp. 3–42, Apr 2006.

[27] A. Montuoro, J. Wu, S. Waldstein, B. Gerendas, G. Langs, C. Simader, and U. Schmidt-Erfurth, "Motion artefact correction in retinal optical coherence tomography using local symmetry," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014: 17th International Conference, Boston, MA, USA, September 14-18, 2014, Proceedings, Part II*, P. Golland, N. Hata, C. Barillot, J. Hornegger, and R. Howe, Eds. Cham: Springer International Publishing, 2014, pp. 130–137.

[28] M. K. Garvin, M. D. Abràmoff, X. Wu, S. R. Russell, T. L. Burns, and M. Sonka, "Automated 3-d intraretinal layer segmentation of macular spectral-domain optical coherence tomography images," *IEEE Transactions on Medical Imaging*, vol. 28, no. 9, pp. 1436–1447, Sep 2009.

[29] J. Wu, B. S. Gerendas, S. M. Waldstein, G. Langs, C. Simader, and U. Schmidt-Erfurth, "Stable registration of pathological 3d-oct scans using retinal vessels," in *Proceedings of the Ophthalmic Medical Image Analysis First International Workshop, OMIA 2014, Held in Conjunction with MICCAI 2014*, X. Chen, M. K. Garvin, and J. J. Liu, Eds. Iowa Research Online, 2014, pp. 1–8.

[30] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug 2007.

[31] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, "Multiscale vessel enhancement filtering," in *Medical Image Computing and Computer-Assisted Intervention — MICCAI'98: First International Conference Cambridge, MA, USA, October 11–13, 1998 Proceedings*, W. M. Wells, A. Colchester, and S. Delp, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1998, pp. 130–137.

[32] A. Myronenko and X. Song, "Point set registration: Coherent point drift," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 12, pp. 2262–2275, Dec 2010.

[33] R. C. Bolles and M. A. Fischler, "A ransac-based approach to model fitting and its application to finding cylinders in range data." in *IJCAI*, vol. 1981, 1981, pp. 637–643.

[34] J. Wu, S. M. Waldstein, A. Montuoro, B. S. Gerendas, G. Langs, and U. Schmidt-Erfurth, "Automated fovea detection in spectral domain optical coherence tomography scans of exudative macular disease," *International Journal of Biomedical Imaging*, vol. 2016, pp. 1–9, 2016.

[35] B. J. Lujan, F. Wang, G. Gregori, P. J. Rosenfeld, R. W. Knighton, C. A. Puliafito, R. P. Danis, L. D. Hubbard, R. T. Chang, D. L. Budenz *et al.*, "Calibration of fundus images using spectral domain optical coherence tomography," *Ophthalmic Surg Lasers Imaging*, vol. 39, no. 4 Suppl, pp. S15–20, 2008.

[36] K. Rohrschneider, "Determination of the location of the fovea on the fundus," *Investigative ophthalmology & visual science*, vol. 45, no. 9, pp. 3257–3258, 2004.

[37] M. R. Osborne, B. Presnell, and B. A. Turlach, "On the lasso and its dual," *J. Comput. Graph. Statist.*, vol. 9, pp. 319–337, 2000.

[38] H. Zou and T. Hastie, "Regression shrinkage and selection via the elastic net, with applications to microarrays," *Journal of the Royal Statistical Society: Series B. v67*, pp. 301–320, 2003.

[39] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," *Journal of Statistical Software*, vol. 33, no. 1, 2010.

[40] N. Hansen and A. Ostermeier, "Completely derandomized self-adaptation in evolution strategies," *Evolutionary computation*, vol. 9, no. 2, pp. 159–195, 2001.

[41] M. Claesen, J. Simm, D. Popovic, Y. Moreau, and B. De Moor, "Easy hyperparameter search using optunity," *arXiv preprint arXiv:1412.1114*, 2014. [Online]. Available: https://arxiv.org/abs/1412.1114

[42] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[43] O. Troyanskaya, M. Cantor, G. Sherlock, P. Brown, T. Hastie, R. Tibshirani, D. Botstein, and R. B. Altman, "Missing value estimation methods for DNA microarrays," *Bioinformatics*, vol. 17, pp. 520–525, 2001.

[44] E. Y. Chew, M. L. Klein, F. L. Ferris, N. A. Remaley, R. P. Murphy, K. Chantry, B. J. Hoogwerf, and D. Miller, "Association of elevated serum lipid levels with retinal hard exudate in diabetic retinopathy: Early treatment diabetic retinopathy study (etdrs) report 22," *Archives of Ophthalmology*, vol. 114, no. 9, p. 1079, Sep 1996.

[45] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, Jun 2006.

[46] X. Robin, N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J.-C. Sanchez, and M. Müller, "proc: an open-source package for r and s+ to analyze and compare roc curves," *BMC Bioinformatics*, vol. 12, p. 77, 2011.

[47] R. G. Newcombe, "Two-sided confidence intervals for the single proportion: comparison of seven methods," *Statistics in Medicine*, vol. 17, no. 8, pp. 857–872, Apr 1998.

[48] T. J. Hastie, R. J. Tibshirani, and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*. Springer, 2011.

[49] A. E. Hoeh, T. Ach, K. B. Schaal, A. F. Scheuerle, and S. Dithmar, "Long-term follow-up of oct-guided bevacizumab treatment of macular edema due to retinal vein occlusion," *Graefe's Archive for Clinical and Experimental Ophthalmology*, vol. 247, no. 12, pp. 1635–1641, Dec 2009.