Published in final edited form as:
IEEE Trans Biomed Eng. 2014 April ; 61(4): 1143–1154. doi:10.1109/TBME.2013.2295605.

Glaucoma Progression Detection Using Structural Retinal Nerve Fiber Layer Measurements and Functional Visual Field Points

Siamak Yousefi, IEEE [Member],
Hamilton Glaucoma Center, Department of Ophthalmology, University of California, San Diego, CA 92093 USA

Michael H. Goldbaum,
Hamilton Glaucoma Center, Department of Ophthalmology, University of California, San Diego, CA 92093 USA

Madhusudhanan Balasubramanian,
Hamilton Glaucoma Center, Department of Ophthalmology, University of California, San Diego, CA 92093 USA

Tzyy-Ping Jung,
Institute for Neural Computation, Institute of Engineering in Medicine, University of California, San Diego, CA 92093 USA

Robert N. Weinreb,
Hamilton Glaucoma Center, Department of Ophthalmology, University of California, San Diego, CA 92093 USA

Felipe A. Medeiros,
Hamilton Glaucoma Center, Department of Ophthalmology, University of California, San Diego, CA 92093 USA

Linda M. Zangwill,
Hamilton Glaucoma Center, Department of Ophthalmology, University of California, San Diego, CA 92093 USA

Jeffrey M. Liebmann,
Department of Ophthalmology, New York University, New York, NY 10012 USA

Christopher A. Girkin, and
Department of Ophthalmology, University of Alabama, Birmingham, AL 35233 USA

Christopher Bowd
Hamilton Glaucoma Center, Department of Ophthalmology, University of California, San Diego, CA 92093 USA

Siamak Yousefi: syousefi@ucsd.edu; Michael H. Goldbaum: mgoldbaum@ucsd.edu; Madhusudhanan Balasubramanian: madhu@glaucoma.ucsd.edu; Tzyy-Ping Jung: jung@sccn.ucsd.edu; Robert N. Weinreb: rweinreb@ucsd.edu; Felipe A.

Correspondence to: Christopher Bowd, cbowd@ucsd.edu.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Medeiros: fmedeiros@ucsd.edu; Linda M. Zangwill: lzangwill@ucsd.edu; Jeffrey M. Liebmann: jml18@earthlink.net; Christopher A. Girkin: cgirkin@uab.edu; Christopher Bowd: cbowd@ucsd.edu

Abstract

Machine learning classifiers were employed to detect glaucomatous progression using longitudinal series of structural data extracted from retinal nerve fiber layer thickness measurements and visual functional data recorded from standard automated perimetry tests. Using the collected data, a longitudinal feature vector was created for each patient's eye by computing the norm 1 difference vector of the data at the baseline and at each follow-up visit. The longitudinal features from each patient's eye were then fed to the machine learning classifier to classify each eye as stable or progressed over time. This study was performed using several machine learning classifiers including Bayesian, Lazy, Meta, and Tree, composing different families. Combinations of structural and functional features were selected and ranked to determine the relative effectiveness of each feature. Finally, the outcomes of the classifiers were assessed by several performance metrics and the effectiveness of structural and functional features were analyzed.

Keywords

Biomedical engineering; biomedical signal processing; change detection; glaucoma progression; machine learning

I. Introduction

Glaucoma, a progressive optic neuropathy, is the second leading cause of blindness in the world [1]–[3]. It is estimated that there will be approximately 80 million people worldwide affected by glaucoma by 2020 [3]. Key aspects of glaucoma management are early detection and monitoring of disease progression (identifying the first evidence of disease related change and keeping track of it). Current gold standards for identifying glaucomatous progression are assessment of serial stereoscopic optic disc photographs (depicting anatomical structure) and assessment of serial psychophysical test results (tests of visual function, called visual field tests).

Stereoscopic disc photos provide a permanent record of the appearance of the optic nerve that is independent of specialized viewing instruments [4]–[7]. This technique remains the most widely used and accepted documentation of the condition of the optic nerve head [6]. Methods for detecting glaucoma and monitoring its progression over time using stereophotographs have been described previously [7]. Like other subjective assessments, subjective assessment of serial optic disc photographs is prone to expert errors. From the clinical point of view, many eye care professionals prefer to have access to more objective progression analyses.

Standard automated perimetry (SAP) is a psychophysical test that provides the clinicians an insight to the function of the visual field by reporting the retinal sensitivity to light stimuli [8]. SAP provides an objective assessment of the visual field in the form of sensitivity measurements at 52 different test points (for 24-2 stimuli) across the visual field. Recent improved testing algorithms like the Swedish interactive thresholding algorithm (SITA) and

the availability of progression detection software like guided progression analysis (GPA) and visual field index have solidified SAP as the preferred method for diagnosis and follow-up of functional visual field loss [8], [9].

Optical coherence tomography (OCT) is an objective tool for imaging the structure of the optic disc and the thickness of the surrounding retinal nerve fiber layer (RNFL, also damaged in glaucoma). OCT is a noncontact and noninvasive medical imaging technology that uses reflected light to produce detailed cross-sectional and 3-D images of the eye and provides a tool for visualizing structural maps of the RNFL [10]. Spectral domain OCT (SDOCT), also known as Fourier domain OCT, is a relatively recent advance in imaging that allows clinicians to detect retinal pathologies that might not be seen by the clinician or by using other ophthalmic instruments. Essentially, SDOCT identifies differences in optical reflection of the different retinal layers and illustrates them in a gray scale. Thus, SDOCT images depict reflective interfaces between retinal layers. SDOCT scanners generate high-resolution images in a relatively short time [10], [11].

Machine learning classifiers (MLCs) have been widely used in biomedical areas [12]–[14]. In an attempt to further increase the objective interpretation of optical imaging and visual function testing, MLCs that use a mathematical approach without human intervention are used to detect defect patterns and disease progression and to classify eyes as healthy or glaucomatous (see [15] for a review). The advantages of MLCs compared to traditional methods in glaucoma diagnosis are discussed in [16]. Recent advances in MLCs and ophthalmic imaging instruments suggest that more accurate prediction and detection of glaucoma progression is possible [17]–[19]. Relevance vector machine (RVM), which modifies support vector machines with Bayesian strategies to provide a probability of group membership, is an example of a MLC [20]–[22] that was successfully employed to separate healthy and glaucomatous eyes. RVM performs equal to, or better, than other statistical techniques [23]. In [24], the authors proposed spatial modeling of visual fields to enhance glaucoma progression detection accuracy. By properly modeling visual field dependencies, the authors reached a reasonable degree of accuracy detecting glaucomatous progression. To identify early glaucomatous progression from objective measurements, the current study sought to apply various classification methods to baseline and to follow up SDOCT measurements, SAP results and the combination of these measurements to determine which eyes remain stable and which glaucoma eyes show glaucomatous progression over serial follow-up. The chosen classification methods were examples of Bayesian, Lazy, Meta, and Tree classifiers. Several classifier types were evaluated to both determine which may be most useful for this classification problem and to keep our feature evaluation results independent of any particular classifier.

II. Methods

In this section, we first describe the technologies and instruments used for collecting structural and functional ophthalmic data of study participants. Subsequently, we describe the data acquisition and assessment of subjects and the methodology used to generate optimal features for analysis using MLCs. We have provided detailed descriptions of the MLCs evaluated in this study and their implementation details. Next, we describe a set of

performance metrics that we use to assess the performance of the MLCs. We then discuss how glaucomatous patterns are extracted by the MLCs by ranking features of structural and functional measurements of study participants. The outcome of the experiment and a thorough discussion is provided.

A. Instruments

Color photograph pairs were simultaneously obtained through maximally dilated pupils using a stereoscopic camera (Kowa Nonmyd WX3D, software version VK27E, Kowa Company Ltd., Tokyo, Japan). SAP-measured visual field sensitivity was tested at 54 points (of which 52 locations were used in feature sets) using the 24-2 SITA test strategy (Humphrey Field Analyzer II, Carl Zeiss Meditec, Dublin, CA, USA). SDOCT measurements of the RNFL thickness were obtained using the Spectralis RNFL circle scan (software version 1.5.2.0, Heidelberg Engineering, Heidelberg, Germany).

Fig. 1(top-left) shows an example photograph of the optic disk region and peripapillary retina. The retinal vessels and optic cup can be seen in this figure. Fig. 1(top-right) displays 24-2 SAP visual field absolute sensitivities in decibels at the available 52 test points that are uniquely specified by their angular location in superior, inferior, nasal, or temporal zones. In Fig. 1(bottom), a sample SDOCT image is shown. Retinal nerve fiber layer, ganglion cell complex, inner and outer plexiform layers, and other tissue layers can be seen from an OCT image.

B. Data Acquisition and Assessment

Each study participant underwent a comprehensive ophthalmic evaluation, including review of medical history, best corrected visual acuity, slit lamp biomicroscopy, intraocular pressure measurement with Goldmann applanation tonometry, gonioscopy, dilated slit lamp fundus examination, simultaneous stereoscopic optic disc photography, SAP visual field exam, and SDOCT imaging exam at each visit. All participant eyes were recruited from the University of California at San Diego (UCSD)-based diagnostic innovations in glaucoma study (DIGS) and the African Descent and Glaucoma Evaluation Study (ADAGES). The ADAGES is a multicenter study that includes UCSD, University of Alabama at Birmingham and New York Eye and Ear Infirmary. Both studies follow the tenets of the Declaration of Helsinki, Health Insurance Portability and Accountability Act guidelines and the Human Research Protection Program of each study site approved all methodology. Written informed consent was obtained from all study participants.

All classifiers (described below) were trained and tested on structural and functional measurements obtained of the same visit. To create the reference standard for classification assessment, all eyes evaluated were divided into two groups. The first group of eyes showed progression based on serial analysis of optic disc stereophotographs or based on visual field assessment using guided progression analysis (GPA, described below). The second group contained eyes that were stable using both progression detection techniques.

For progression assessment using stereophotographs, the baseline and each follow-up image were assessed for progressive glaucomatous optic neuropathy (PGON) by two observers using a stereoscopic viewing device to evaluate digitized paired images on a 21 inch or

larger computer monitor. PGON was defined as a decrease in the neuroretinal rim thickness, the appearance of a new RNFL defect or the enlargement of a preexisting RNFL defect. Observers were masked to patient identification and diagnosis. A third observer adjudicated any disagreement in assessment between the first two observers [25].

For progression assessment using SAP, progressive glaucoma was defined based on GPA analysis, a method for detecting changes available on the visual field instrument. GPA indicates change from baseline by evaluating all test points and indicates “likely progression” for the full field if change (greater than the variability observed in two baseline measurements) in three or more of the same points is repeatable in three consecutive exams [26], [27]. For this study, two consecutive “likely progression” results were required. Therefore, two groups of progressed glaucoma subjects were identified—progressed by PGON, and progressed by GPA. The PGON group represents glaucomatous progression based on structural evidence while the GPA group represents glaucomatous progression based on functional evidence. A total of 107 eyes from 100 subjects were identified as progressed by either PGON or GPA. A total of 632 OCT and SAP visual field measurements were collected from this group. The mean number of follow-up visits was 4.3, and the mean follow-up time was 2.2 years.

All stable eyes were selected from eyes that had been identified as glaucomatous at baseline with repeatable SAP defects (instrument software defined pattern standard deviation ≤5% of normal or Glaucoma Hemifield Test outside of normal limits) [28]. Stability was simulated by testing with SDOCT and SAP every week for five weeks, providing an average of 3.9 consecutive tests excluding baseline, for each eye. These eyes were considered stable (i.e., nonprogressing) because testing was completed within a mean of four weeks and detectable disease-related change in structure or function is not possible in this brief time (because glaucomatous change usually occurs over years). A total of 73 eyes from 39 subjects were included and a total of 358 OCT and SAP visual field measurements were collected.

Table I shows the demographic information of the subjects in the stable or progressing groups. The mean deviation (MD) and pattern standard deviation (PSD) of each group are listed in this table, as well. MD and PSD are global indices that indicate the deviation of a visual field from a mean of normal visual field and are descriptors of glaucoma related visual field defects. The lower the MD and higher the PSD, the more severe the defect.

C. Data Formats and Feature Sets for MLC

OCT RNFL thickness measurements and thresholds at each SAP test points were used to train and test the machine learning classifiers. MD and PSD values also were included resulting in a 54-D vector for SAP (threshold values at 52 test points, MD, and PSD).

RNFL thickness was measured using the Spectralis RNFL Circle Scan configuration, a single B-scan along a 3.4-mm circle centered on the optic disc composed of 1536 SDOCT A-scans. Spectralis software segments the image (isolating the RNFL) and calculates the average RNFL thickness for different sectors. In this study, RNFL thickness was obtained in six sectors; superior temporal, temporal, inferior temporal, inferior nasal, nasal, and superior nasal. Spectralis software also provides global RNFL thickness in addition to sectoral

measurements. The OCT data comprised these seven RNFL thickness measurements, resulting in a 7-D input vector. Quality of all SDOCT scans was reviewed according to standard protocol by the imaging data evaluation and analysis center staff, and only good quality scans were included in the analysis. All SAP results were reliable (fixation loss ≤33%, false positive results ≤15%) and were free of common tests artifacts. RNFL and SAP measurements for each visit were obtained within three months of each other. Average between-visit interval for progressing eyes was 0.6 years (SD = 0.3 years) and average between-visit interval for stable eyes was 8.3 days (SD = 2.0 days).

The RNFL and SAP data vectors were processed further to generate the feature vectors. For each eye, the difference between the baseline RNFL and SAP data vectors (obtained by the first exam date) and each follow-up RNFL and SAP data vector were calculated. This way, we obtained a longitudinal time series of features for each subject's eye. For instance, if the data are collected from a subject at baseline and at 4 follow-up visits, the longitudinal data set for this subject has four time points and each time point has a corresponding 7-D RNFL and a 54-D SAP (threshold values at 52 test points, MD, and PSD) feature vector. This is shown in Fig. 2 in more detail. Fig. 2(a) shows sample RNFL and SAP measurements and indicates how the data vectors are formed. Fig. 2(b) shows the longitudinal data vectors for a single subject. The longitudinal feature vectors, which are the norm 1 difference between the baseline and follow-up data vectors, are displayed in Fig. 2(c).

Different combinations of the RNFL and SAP features were fed to the machine learning classifiers to assess their effectiveness and power in detecting glaucoma progression patterns and separating stable from progressed eyes over time. Note that combining the features is done by concatenating two feature vectors. This is discussed in more details in the following sections.

D. Machine Learning Classifiers

To analyze the effectiveness of different classifiers and to assess the optimality of SAP and RNFL input features, we used classifiers from Bayesian, Instance-based, Meta, and Tree families of MLCs including Bayesian net, Lazy K Star, Meta classification using regression, Meta ensemble selection, alternating decision tree (AD tree), random forest tree, and simple classification and regression tree (CART) to detect glaucoma progression patterns from the longitudinal feature vectors, and to separate each eye into either the nonprogressed (i.e., stable) or progressed glaucoma. Eyes with at least 50% of follow-up exams classified as progressed by the MLC, or with two consecutive follow-up exams classified as progressed by the MLC, were assigned to the progressed glaucoma group; the remaining study eyes were assigned to the stable glaucoma group. Here, we briefly describe these classifiers.

Bayesian net employs factored representations of probability distributions that generalize the naive Bayesian classifier and explicitly represent statements about independence. In Lazy learning algorithms, the generalization beyond the training data is delayed until the arrival of a new observation. Lazy IB is actually a nearest neighbor classifier that assigns the nearest sample's class to the new instance [29]. Lazy K Star is another form of instance-based classifier that utilizes the entropic measures as the metric distance [30].

In Meta learning algorithms, multiple learners are integrated to compute a higher level learner. Meta classification using regression, first binaries each class and then a regression model is built for each class value to perform the classification [31]. In ensemble learning, multiple models are used to obtain a better predictive performance than any of the constituent models [32]–[34]. In ensemble selection, ensembles are constructed from libraries of thousands of models then forward stepwise methodology is used to augment the ensemble models in order to maximize their performance [34].

AD tree is a generalization of decision trees, voted decision trees, and voted decision stumps [35]. AD Tree is relatively easy to interpret and has strong connection to boosting. Random forest tree is an ensemble learning method for classification and regression. It constructs a forest of decision trees at the time of training and then outputs the class that is the mode of the classes output by individual trees [36], [37]. CART uses data to construct classification or regression decision trees that can then be used for classification of new observations.

E. Implementation

The machine learning classifiers were implemented in MATLAB (Mathworks, Natick, MA, USA) or Weka (The University of Waikato, New Zealand) to assess the effectiveness of structural and functional ophthalmic features. First, we used RNFL and SAP features separately and then we combined the SAP and RNFL features to assess whether the combined functional and structural data performed significantly better than either alone. This is a critical analysis to reveal the optimality of the SAP features for classifiers [38]. Several classification performance metrics outlined below were implemented using ten-fold cross validation (independent training and testing groups) to assess the machine learning classifier outcomes and in addition, independent feature ranking was performed to assess the discriminating power of the structural and functional features in detecting stable from progressing glaucoma eyes. We further implemented and analyzed the longitudinal evolution of the best features over time for subjects in each group, specifying the fastest glaucomatous progressed eye over time, and identifying and displaying the features corresponding to misclassified eyes.

The following default parameters and initializations were used, however, changing these default parameters did not change the outcomes significantly. For Bayesian net, a simple estimator was used for estimating the conditional probability tables of a Bayes network once the structure had been learned. The local hill climbing algorithm was used with the maximum number of parents set to 1. For lazy *k*-star, the global blending parameter was set to 20 and the missing mode was set to average entropy curves. In implementing Meta classification using regression, we used an improved version of M5 model tree [39] for the classification that was used along with regression. For Meta ensemble selection, a forward selection with greedy sort initialization was used, and the other parameters were set as follows: the hill climb iteration was set to 100, model ratio to 0.5, number of folds to 1, number of model bags to 10, and the validation ratio to 0.25. For AD tree, the number of boosting iterations parameter was set to 10. For Random Forest Tree we used unlimited maximum depth of the trees, and ten trees to be generated.

F. Performance Metrics

Performance metrics are mostly defined based on a confusion matrix. The confusion matrix is composed of four items: true positive (TP) which are positive instances correctly classified as positives; false positive (FP) which are negative instances incorrectly classified as positive; true negative (TN) which are negative instances correctly classified as negatives, and finally; False Negative (FN) which are positive instances incorrectly classified as negatives. A sample confusion matrix is shown in Table II.

True positive rate (TPR), false positive rate (FPR), F-Measure, Mathews correlation coefficient (MCC), area under receiver operator characteristic curve (AUROC), and area under precision recall curve (AUPRC) were calculated and compared for each method. The AUROC standard deviation and 95% confidence intervals were calculated based on the method reported by Hanley *et al*. [40]. The procedures for computing the performance metrics are explained below:

Specificity (TNR) is defined as the ratio of TNs and the total number of negatives

$$\text{Specificity}=\frac{\text{TN}}{\text{TN}+\text{FP}}.$$

Sensitivity (TPR) is defined as the ratio of TPs and the total number of positives

$$\text{TPR}=\frac{\text{TP}}{\text{TP}+\text{FN}}.$$

FPR is defined as the ratio of the number of FPs and the total number of negatives

$$\text{FPR}=\frac{\text{FP}}{\text{TN}+\text{FP}}.$$

F-Measure: The F-Measure is defined based on information retrieval concepts. It is a value between 0 and 1. A perfect classifier generates an F-Measure equal to 1. It is defined as follows:

$$F-\text{Measure}=\frac{2|\text{TP}|}{2|\text{TP}|+|\text{FP}|+|\text{FN}|}.$$

MCC: Indicates the quality of a classifier for binary classification, and is a better indication of quality when the classes are of very different sizes [41]. MCC ranges from −1 to 1, where −1 indicates adverse classification, zero corresponds to average, and 1 represents the perfect classification

$$\text{MCC}=\frac{\text{TP}*\text{TN}-\text{FP}*\text{FN}}{\sqrt{(\text{TP}+\text{FP})*(\text{TP}+\text{FN})*(\text{TN}+\text{FP})*(\text{TN}+\text{FN})}}.$$

AUROC: The ROC curve is used to characterize the tradeoff between true positive rate and false positive rates. AUROC, a widely used metric to measure classification model performance, is the area under ROC curve [42]. AUROC ranges from 0 to 1 and a perfect classifier results in an AUROC that is equal to 1.

AUPRC: Is a single-value measure based on information retrieval concepts. It ranges from 0 to 1 and depicts the tradeoff between recall and precision. An ideal classifier generates area under PRC equal to 1 [43].

G. Feature Selection

Independent feature ranking measures the discriminating power of each feature independent of the other features. Correlation-based feature ranking based on Pearson correlation coefficient and information gain feature ranking are two widely used independent feature ranking methods [44], [45] that can be used for feature ranking and selection. Evaluating the feature usefulness along with other features (i.e., dependent feature ranking) is another way of selecting an optimal subset of features. Both methods were performed in our analysis. For dependent feature selection, correlation-based feature subset selection (CSF) was used. CSF evaluates the discriminating power of a subset of features by considering the individual predictive ability of each feature, along with the degree of redundancy between them [46] using linear forward selection, to select a subset of the best performing features out of a pool of features [47]. Further explanation of this finding is provided in Section III

III. Results and Discussion

RNFL and SAP features from each study eye were fed to seven different MLCs. Different combinations of features (structural or functional) were selected and fed to classifiers to assess their effectiveness based on the performance metrics described earlier. The ROC curves for different classifiers and feature combinations are shown in Fig. 3. Fig. 3(a) displays the diagnostic accuracy of the MLCs when we used the combined set of all RNFL and all SAP features. Fig. 3(b) shows the diagnostic accuracy of the MLCs using all RNFL features only. Differences in other diagnostic performance measures of MLCs between RNFL only and RNFL-SAP combined features were not statistically significant (see Tables III and IV). Fig. 3(c) shows the diagnostic accuracy of the MLCs using all SAP features only. As can be seen from this figure, the performance decreases significantly compared to the performance of the classifiers using combined RNFL and SAP features and RNFL feature alone. Fig. 3(d) shows the diagnostic accuracy of the best performing MLC for the combined set of RNFL and SAP features, RNFL features only, and SAP features only.

The optimality of the RNFL features can be confirmed by looking at other performance metrics listed in Tables III–V. This general analysis suggests that, in this study population of early to moderate glaucoma, the SAP features do not add or improve the accuracy of MLCs in detecting glaucoma progression, while RNFL features provide more information than SAP features to MLCs for detecting progression. Note that the metrics listed in Tables III–V are the weighted average outcomes for the two classes.

We also were interested in comparing different classifiers when the input features were the same. As can be seen from Fig. 3(a) and (b) and Tables III and IV, there is no significant difference between the AUROC of the classifiers, indicating that adding SAP features does not improve the classifier accuracy. The performance of most MLCs (AUROC) was significantly lower when the input features were SAP only, based on Fig. 3(c) and Table V. Looking at other performance metrics, as can be seen from Fig. 3(a) and Table III, CART does not perform as well as the other classifiers at high specificities (90%). Figs. 3(a) and (b) and Tables III and IV suggest that for achieving a higher sensitivity, it is better to use Random Forest Tree or Lazy K Star. The same conclusion is true for MLCs when the features are RNFL alone (see Table III) and finally, when we use SAP features only (see Table V), the MLCs do not perform much better than chance at high specificities with the exception of Lazy *K* Star and Random Forest Tree. The results at 80% and 90% specificity for Tables III and V are very different indicating significant difference between the optimality of the SAP and RNFL features.

Using the above combinations of features provide us the usefulness and effectiveness of the entire set of RNFL or SAP features without providing insight into the effectiveness of each individual feature. To reveal the discriminating power of each single feature from the RNFL or SAP group, an independent feature ranking analysis was performed. The outcome of both independent correlation based and dependent CSF when RNFL and SAP features were combined (61-D feature set) are listed in Table VI, columns 2, and 3, respectively. These top ten combined features were selected out of 61 features based on their correlation ratio metric. As can be seen, both independent and dependent feature selection identify the same best four features, and seven features are selected in common by both methods.

As shown in Table VI, the four best features belong to RNFL feature set. This confirms the findings above that RNFL features alone provide enough information to MLCs for discriminating stable from progressing glaucoma eyes at this stage of disease. This finding led us to further analyze the effectiveness of features; therefore, we performed classification based on the first ten highly ranked features. Fig. 4 shows the ROC curves when using only the top ten ranked features and Table VII shows the AUROC measures of combined RNFL and SAP when using the best, best two, or top ten ranked features listed in Table VI. Note that when we fed a combination of all RNFL and all SAP features to the classifiers, Random Forest Tree classifier generated an AUROC of 0.88 (refer to Table III) and when we fed the selected top ten ranked features to the classifier, Random Forest Tree classifier generated an AUROC of 0.88, but improved sensitivity at higher specificities [compare the gray curve in Fig. 4 with Fig 3(a)]. As can be seen, this AUROC is much higher than the AUROC reported in [24] which is 0.69.

This analysis suggests that the MLCs performance using the top ten ranked features is sufficient to achieve the same diagnostic performance of the MLCs when using all RNFL and SAP features. This is supported by the fact that four top features belong to the RNFL feature set, and is in agreement with our findings described above.

Fig. 5 shows the longitudinal probability assignments for the first 12 eyes from both the stable and progressing group. Each color represents a subject’s eye, and the connected

circles show longitudinal data over time for that particular subject's eye. For instance, in Fig. 5(a), the first subject (in dark green connected circles) has two follow ups and the classifier has assigned this subject's eye to the stable group with probabilities 99%, and 99% at two follow up visits based on the recorded structural and functional data, respectively. Therefore, all two longitudinal features for this subject suggest that this subject has not progressed over time, which is a correct assignment. As can be seen from Fig. 5(a) and (b), the longitudinal features for each subject have been assigned correctly in most cases (points lie on the correct side of the 50% probability lines). Recall that we assigned an eye to the progressing glaucoma group if either more than 50% of the follow-up cases were classified as progressing or if two or more consecutive follow-up cases were classified as progressing. Note that the second rule is in agreement with visual field progressor software currently used in clinical applications and is a widely accepted objective criterion. The Bayesian Net MLC has been used in assigning the classes displayed in Fig. 5.

Fig. 6 displays the feature distribution of the two best RNFL features; Inferior-Nasal thickness and Inferior-Temporal thickness, to provide an insight into the features from stable and progressing groups. The Bayesian discriminating function for the corresponding features is displayed. The features inside the discriminating surface are assigned to the stable group, while the features outside this surface are assigned to progressing group. Note that we have displayed two features and the Bayesian discriminating surface corresponding to those features, instead of displaying all combined features in a 61-D space.

The distribution of the progression probability assignments is another objective way to assess the accuracy of the method and performance of the MLCs. Fig. 7 shows the distribution of the probability assignments to either stable or progressing groups, by the Bayesian net classifier. In most cases the classifier assigns the eyes to the expected class with a high confidence.

It is informative to see the evolution over time of the best feature, from Table VI, for some subjects from each group (stable and progressed). Fig. 8(a) and (b) shows the evolution of the Inferior-Nasal RNFL thickness feature for ten sample stable and ten progressing glaucoma eyes, respectively [at each follow up the feature is constructed by computing the norm 1 difference of measurement at baseline and measurement at follow ups—refer to Fig. 2(c)].

As can be seen, the features in stable glaucoma group [see Fig. 8(a)] are close to the zero line and tend to be mostly flat over time, while the features in the progressing glaucoma group [see Fig. 8(b)] have more negative values and tend to have negative slopes, indicating progression over time. This finding is in agreement with known disease-related decreases in RNFL thickness over time. Most of the features have this characteristic, however, some eyes were misclassified. As an example, for the last eye in Fig. 8(a) all the longitudinal features are negative and all indicate progression over time. In Fig. 8(b), the last eye has five follow ups and four of the follow-ups are positive, indicating no change (i.e., stability).

There are several possible explanations for the misclassification, described above. First, even though the features used for analysis are optimized, and the goal is to identify the

closest to ideal classifier, the MLCs are not perfect because they make some underlying assumptions that do not necessarily fit to the characteristics of the data. Second, measurements are collected from instruments that can be subject to measurement errors. For instance, SDOCT software segments the SDOCT images at apparent tissue boundaries and computes the thickness from the segmented images. In some cases, the segmentation algorithm exhibits localized failure that could result in inaccurate RNFL thickness measurements in a given sector (however, all scans were reviewed for segmentation failures and attempts were made to correct these failures, if identified). In addition, variation in image quality can affect RNFL thickness measurements, somewhat [48]. Visual function (i.e., SAP) measurements are subject to variability over time. This variability likely is the result, in part, of patient fatigue during testing, inattention, and increased variability that comes from disease-related intermittent cell firing and aging.

IV. Conclusion

Employing MLCs in discriminating stable from progressing glaucoma eyes using structural RNFL and functional visual sensitivity measurements is promising. In this population of early to moderate glaucoma eyes, the analyses suggest that the RNFL features alone provide similar diagnostic accuracy compared to classifiers that include both RNFL and SAP. More specifically, our findings suggest that the RNFL measurements in inferior nasal, inferior temporal, global, superior temporal, and temporal sectors of the RNFL provide the most discriminating power for separating stable from progressing glaucoma eyes. Using the ten best independent features, (the top four belonged to the RNFL feature set), resulted in the same outcome as using all 61 features. To obtain high diagnostic accuracy and higher sensitivity at high specificity, we suggest using a Random Forest Tree and not the simple CART classifier. Our experiments reveal that simple Bayesian Net classifiers perform as well as more complicated Meta or Tree-based classifiers when using RNFL features.

Acknowledgments

This work was supported by the Grants NIH R01EY022039, NIH K99EY020518, NIH R01EY008208, NIH R01EY011008, NIH R01EY019869, P30EY022589, an unrestricted grant from Research to Prevent Blindness (New York, NY), Eyesight Foundation of Alabama, Corinne Graber Research Fund of the NewYork Glaucoma Research Institute, David and Marilyn Dunn Fund, and participant incentive grants in the form of glaucoma medication at no cost from Alcon Laboratories, Allergan, and Pfizer.

References

1. Quigley HA, Broman AT. The number of people with glaucoma worldwide in 2010 and 2020. Brit. J. Ophthalmol. 2006 Mar.90:262–267. [PubMed: 16488940]
2. Weinreb RN, Khaw PT. Primary open-angle glaucoma. Lancet. 2004 May 22.363:1711–1720. [PubMed: 15158634]
3. Kingman S. Glaucoma is second leading cause of blindness globally. Bull World Health Organ. 2004 Nov.82:887–888. [PubMed: 15640929]
4. Jonas JB, Dichtl A. Evaluation of the retinal nerve fiber layer. Surv. Ophthalmol. 1996 Mar-Apr; 40:369–378. [PubMed: 8779083]
5. Bowd C, Weinreb RN, Zangwill LM. Evaluating the optic disc and retinal nerve fiber layer in glaucoma—Part I: Clinical examination and photographic methods. Semin. Ophthalmol. 2000 Dec. 15:194–205. [PubMed: 17585434]

6. Lim MC, Budenz DL, Gedde SJ, Rhee DJ, Feuer W. Digital stereoscopic photography with chevron drawings versus standard film photography: Estimates of cup to disc ratio measurements. Investigative Ophthalmol. Visual Sci. 2001 Mar.42:S131–S131.
7. Fingeret M, Medeiros FA, Susanna R Jr, Weinreb RN. Five rules to evaluate the optic disc and retinal nerve fiber layer for glaucoma. Optometry. 2005 Nov.76:661–668. [PubMed: 16298320]
8. Alencar LM, Medeiros FA. The role of standard automated perimetry and newer functional methods for glaucoma diagnosis and follow-up. Indian J. Ophthalmol. 2011 Jan.59:S53–S58. [PubMed: 21150035]
9. Bengtsson B, Heijl A. A visual field index for calculation of glaucoma rate of progression. Amer. J. Ophthalmol. 2008 Feb.145:343–353. [PubMed: 18078852]
10. Nassif N, Cense B, Park BH, Yun SH, Chen TC, Bouma BE, Tearney GJ, de Boer JF. In vivo human retinal imaging by ultrahigh-speed spectral domain optical coherence tomography. Opt. Lett. 2004 Mar.29:480–482. [PubMed: 15005199]
11. Yaqoob Z, Wu J, Yang C. Spectral domain optical coherence tomography: A better OCT imaging strategy. Biotechniques. 2005 Dec.39:S6–S13. [PubMed: 20158503]
12. Yousefi, S.; Kehtarnavaz, N.; Akins, M.; Luby-Phelps, K.; Mahendroo, M. Distinguishing different stages of mouse pregnancy using second harmonic generation images; Proc. 42nd Southeastern Symp. System Theory; 2010. p. 44-46.
13. Yousefi, S.; Kim, B.; Kehtarnavaz, N. Automating porosity features extraction from second harmonic generation images of cervical tissue; Proc. IASTED Int. Conf. Signal Image Process; 2011. p. 129-132.
14. Ham FM, Han S. Classification of cardiac arrhythmias using fuzzy ARTMAP. IEEE Trans. Biomed. Eng. 1996 Apr.43(4):425–430. [PubMed: 8626192]
15. Bowd C, Goldbaum MH. Machine learning classifiers in glaucoma. Optometry Vision Sci. 2008 Jun.85:396–405.
16. Chan K, Lee TW, Sample PA, Goldbaum MH, Weinreb RN, Sejnowski TJ. Comparison of machine learning and traditional classifiers in glaucoma diagnosis. IEEE Trans. Biomed. Eng. 2002 Sep.49(9):963–974. [PubMed: 12214886]
17. Acharya UR, Dua S, Du X, Sree SV, Chua CK. Automated diagnosis of glaucoma using texture and higher order spectra features. IEEE Trans. Inf. Technol. Biomed. 2011 May; 15(3):449–455. [PubMed: 21349793]
18. Hatanaka Y, Muramatsu C, Sawada A, Hara T, Yamamoto T, Fujita H. Glaucoma risk assessment based on clinical data and automated nerve fiber layer defects detection. Proc. IEEE Eng. Med. Biol. Soc. 2012; 2012:5963–5966.
19. Cheng J, Liu J, Xu Y, Yin F, Wong DW, Tan NM, Tao D, Cheng CY, Aung T, Wong TY. Superpixel classification based optic disc and optic cup segmentation for glaucoma screening. IEEE Trans. Med. Imag. 2013 Jun.32(6):1019–1032.
20. Bishop, CM.; Tipping, ME. Variational relevance vectormachines; Proc. 16th Conf. Uncertainty Artificial Intell; 2000. p. 46-53.
21. Tipping ME. Sparse Bayesian learning and the relevance vector machine. J. Mach. Learning Res. 2001; 1:211–244.
22. Racette L, Chiou CY, Hao JC, Bowd C, Goldbaum MH, Zangwill LM, Lee TW, Weinreb RN, Sample PA. Combining functional and structural tests improves the diagnostic accuracy of relevance vector machine classifiers. J. Glaucoma. 2010 Mar.19:167–175. [PubMed: 19528827]
23. Bowd C, Medeiros FA, Zhang ZH, Zangwill LM, Hao JC, Lee TW, Weinreb RN, Goldbaum MH. Relevance vectormachine and support vector machine classifier analysis of scanning laser polarimetry retinal nerve fiber layer measurements. Investigative Ophthalmol. Visual Sci. 2005 Apr.46:1322–1329.
24. Betz-Stablein BD, Morgan WH, House PH, Hazelton ML. Spatial modeling of visual field data for assessing glaucoma progression. Invest. Ophthalmol. Vis. Sci. 2013 Feb.54:1544–1553. [PubMed: 23341021]
25. Medeiros FA, Zangwill LM, Bowd C, Sample PA, Weinreb RN. Use of progressive glaucomatous optic disk change as the reference standard for evaluation of diagnostic tests in glaucoma. Amer. J. Ophthalmol. 2005 Jun.139:1010–1018. [PubMed: 15953430]

26. Heijl, A.; Lindgren, G.; Olsson, J. A package for the statistical analysis of visual fields. Proc. 7th Int. Visual Field Symp; The Netherlands; Amsterdam. 1987. p. 153-168.
27. Bengtsson B, Heijl A. A visual field index for calculation of glaucoma rate of progression. Amer. J. Ophthalmol. 2008 Feb.145:343–353. [PubMed: 18078852]
28. Johnson CA, Sample PA, Cioffi GA, Liebmann JR, Weinreb RN. Structure and function evaluation (SAFE)—Part I: Criteria for glaucomatous visual field loss using standard automated perimetry (SAP) and short wavelength automated perimetry (SWAP). Amer. J. Ophthalmol. 2002 Aug. 134:177–185. [PubMed: 12140023]
29. Aha DW, Kibler D, Albert MK. Instance-based learning algorithms. Mach. Learning. 1991 Jan. 6:37–66.
30. Cleary, J.; Trigg, L. K*: An instance-based learner using an entropic distance measure; Proc. 12th Int. Conf. Mach. Learning; 1995. p. 108-114.
31. Frank E, Wang Y, Inglis S, Holmes G, Witten IH. Technical note: Using model trees for classification. Mach. Learning. 1998 Jul.32:63–76.
32. Breiman L. Bagging predictors. Mach. Learning. 1996 Aug.24:123–140.
33. Freund, Y.; Schapire, RE. Experiments with a new boosting algorithm; Proc. 13th Int. Conf. Mach. Learning; Bari, Italy. 1996 Jul.. p. 148-156.
34. Caruana, R.; Niculescu-Mizil, A.; Crew, G.; Ksikes, A. Ensemble selection from libraries of models; presented at the 21st Int. Conf. Mach. Learning; Banff, AB, Canada. 2004.
35. Freund, Y.; Llew, M. The alternating decision tree learning algorithm; Proc. 16th Int. Conf. Mach. Learning; Bled, Slovenia. 1999 Jun.. p. 124-133.
36. Ho TK. The random subspace method for constructing decision forests. IEEE Trans. Pattern Anal. Mach. Intell. 1998 Aug.20(8):832–844.
37. Breiman L. Random forests. Mach. Learning. 2001 Oct.45:5–32.
38. Bengtsson B, Bizios D, Heijl A. Effects of input data on the performance of a neural network in distinguishing normal and glaucomatous visual fields. Invest. Ophthalmol. Vis. Sci. 2005 Oct. 46:3730–3736. [PubMed: 16186356]
39. Wang, Y.; Witten, IH. Inducing model trees for continuous classes; Proc. 9th Eur. Conf. Mach. Learning; 1997. p. 128-137.
40. Hanley JA, Mcneil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982; 143:29–36. [PubMed: 7063747]
41. Lund, O. Immunological Bioinformatics. Cambridge, MA, USA: MIT Press; 2005.
42. Fawcett T. An introduction to ROC analysis. Pattern Recog. Lett. 2006 Jun.27:861–874.
43. Davis, J.; Goadrich, M. The relationship between precision-recall and ROC curves; Proc. 23rd Int. Conf. Mach. Learning; 2006. p. 233-240.
44. Plackett RL. Pearson, Karl and the chi-squared test. Int. Statistical Rev. 1983; 51:59–72.
45. Quinlan JR. Induction of decision trees. Mach. Learning. 1986; 1:81–106.
46. Hall, MA. Correlation-based feature selection for machine learning. Proc. 17th Int. Conf. Machine Learning (ICML); Stanford University; Stanford, CA, USA. 2000 Jun-Jul. p. 359-366.
47. Gutlein, M.; Frank, E.; Hall, M.; Karwath, A. Large-scale attribute selection using wrappers; Proc. IEEE Symp. Comput. Intell. Data Mining; 2009. p. 332-339.
48. Balasubramanian M, Bowd C, Vizzeri G, Weinreb RN, Zangwill LM. Effect of image quality on tissue thickness measurements obtained with spectral domain-optical coherence tomography. Opt. Exp. 2009 Mar.17:4019–4036.

Biographies

Siamak Yousefi (S'10–M'13) received the Ph.D. degree from the University of Texas at Dallas.

He is a Postdoctoral Fellow at the Hamilton Glaucoma Center, University of California at San Diego, CA, USA where he is conducting research on ophthalmic image and data analysis. His research interests include biomedical image analysis, pattern recognition, and machine learning.

Dr. Yousefi is a member of the ARVO.

Michael H. Goldbaum, photograph and biography not available at the time of publication.

Madhusudhanan Balasubramanian received the Ph.D. degree from Louisiana State University, Baton Rouge.

He is currently Assistant Professor in the Department of Electrical and Computer Engineering, University of Memphis, Memphis, TN, USA. His research interest is at the intersection of computational science and engineering, biosolid mechanics and biofluid dynamics with emphasis in studying ocular structures, dynamics and the mechanism of vision loss in glaucoma.

Tzyy-Ping Jung received the Ph.D. degree from the Ohio State University, Columbus.

He is currently a Research Scientist and the Co-Director of the Center for Advanced Neurological Engineering, Institute of Engineering in Medicine, and the Director of the Swartz Center for Computational Neuroscience, Institute for Neural Computation at the University of California San Diego (UCSD), USA.

Robert N. Weinreb graduated from Harvard Medical School (MD) and MIT (electrical engineering).

He is a clinician, surgeon, educator, and scientist. He is the Distinguished Professor of Ophthalmology and Chairman of the Department of Ophthalmology, University of California, San Diego, USA.

Felipe A. Medeiros is Professor of Ophthalmology and Medical Director of the Hamilton Glaucoma Center, University of California San Diego, USA. He is also Director of Vision Function Research at the same institution.

Linda M. Zangwill photograph and biography not available at the time of publication.

Jeffrey M. Liebmann is presently Clinical Professor of Ophthalmology at New York University School of Medicine, NY, USA and Director of Glaucoma Services at Manhattan Eye, Ear, and Throat Hospital and New York University Langone Medical Center and Adjunct Professor of Clinical Ophthalmology at New York Medical College, Valhalla, New York.

Christopher A. Girkin, photograph and biography not available at the time of publication.

Christopher Bowd received the Ph.D. degree from Washington State University.

He is currently a Research Scientist at the University of California, San Diego, Hamilton Glaucoma Center. His current work involves early detection and monitoring of glaucoma

with structural imaging of the optic nerve, visual function and electrophysiological testing using standard and machine learning classifier-based analyses.

Fig. 1.

Top left: Sample optic disc photograph image, Top right: SAP visual points tested using the 24-2 system, and bottom: Sample OCT RNFL image and its scan type (circular optical section around the optic nerve).

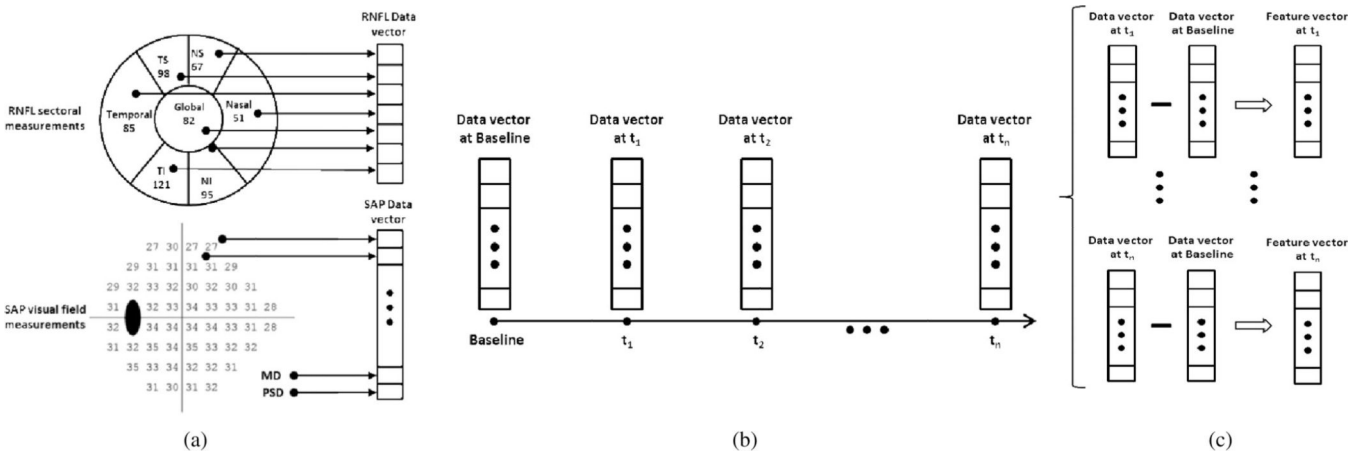

Fig. 2.

(a) Data vector formation, (b) longitudinal data formation for each subject's eye, (c) feature vector generation.

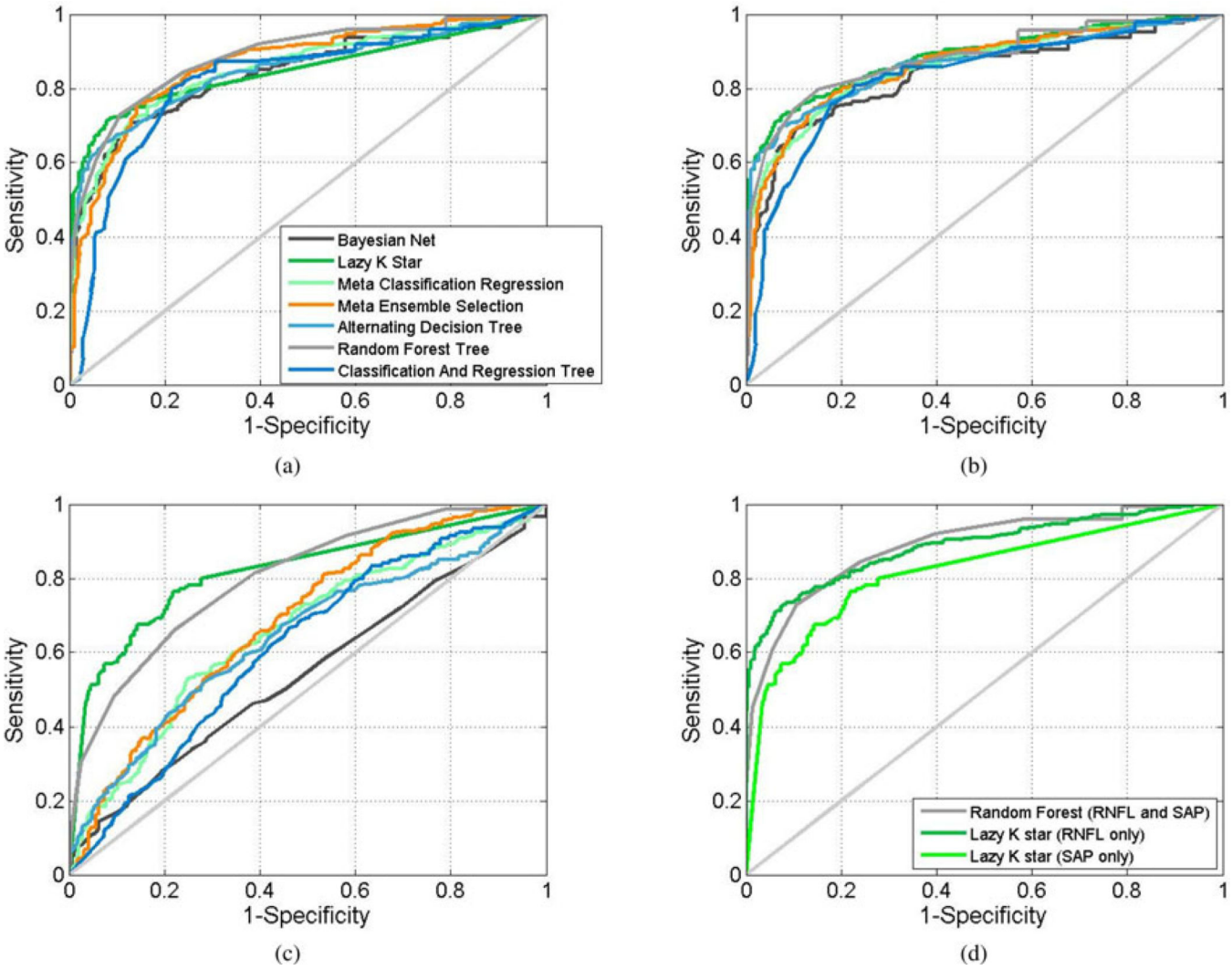

Fig. 3.

ROC curves for different feature sets; (a) Combined features from all RNFL and all SAP, (b) all RNFL features only, (c) all SAP features only, and (d) the best classifier (based on AUROC) for combined features from all RNFL and all SAP, all RNFL features only, and all SAP features only.

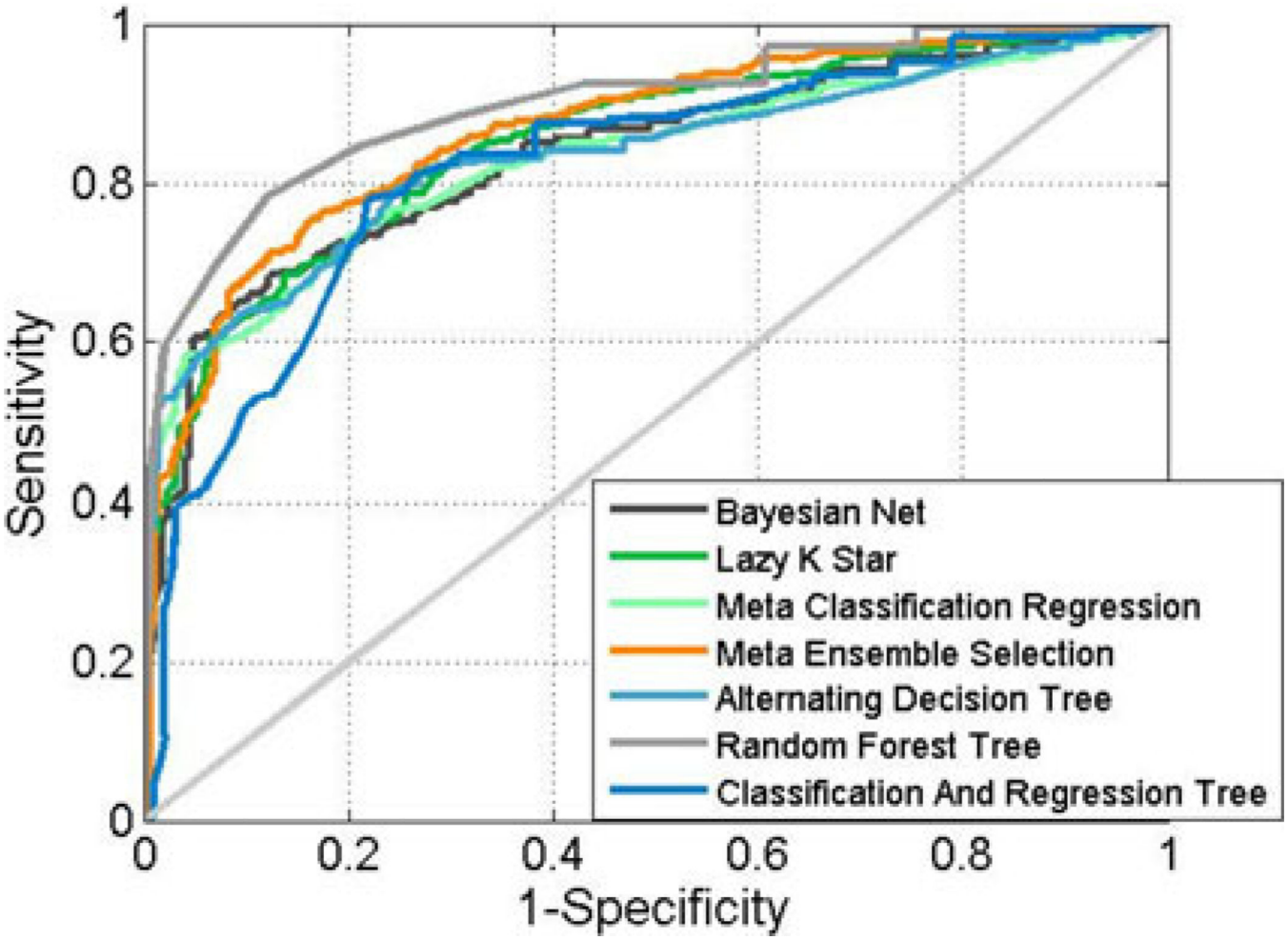

Fig. 4.
ROC curves for different classifiers when we use only top ten ranked features.

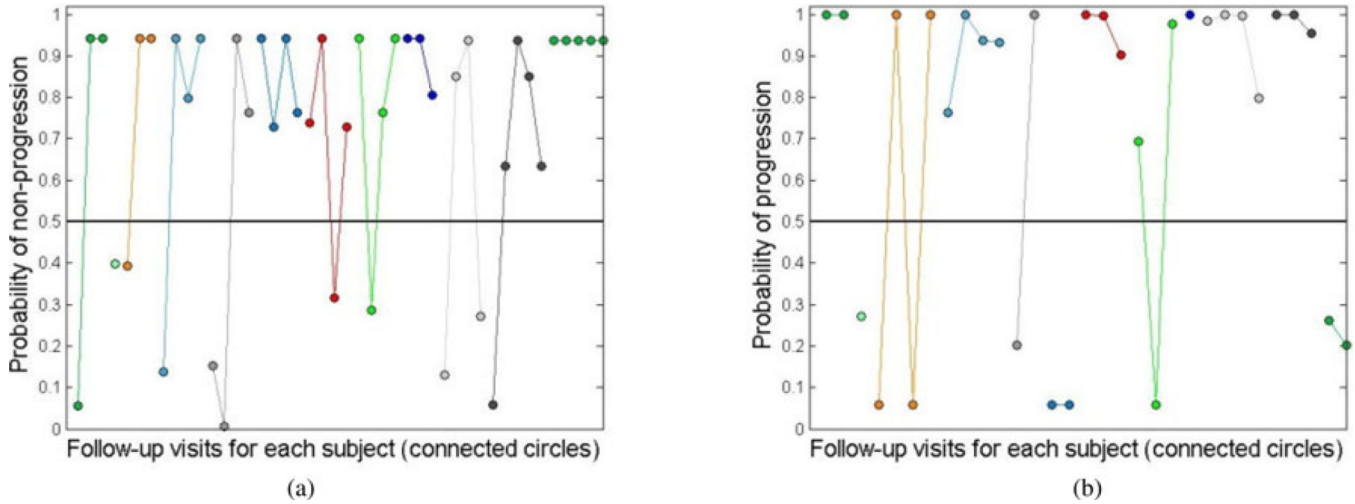

Fig. 5.

Probability assignment to longitudinal visit of each patient. (a) Stable group. (b) Progressing group. Connected circles in the same color indicate the classifier outcome for the same subject's eye longitudinal data.

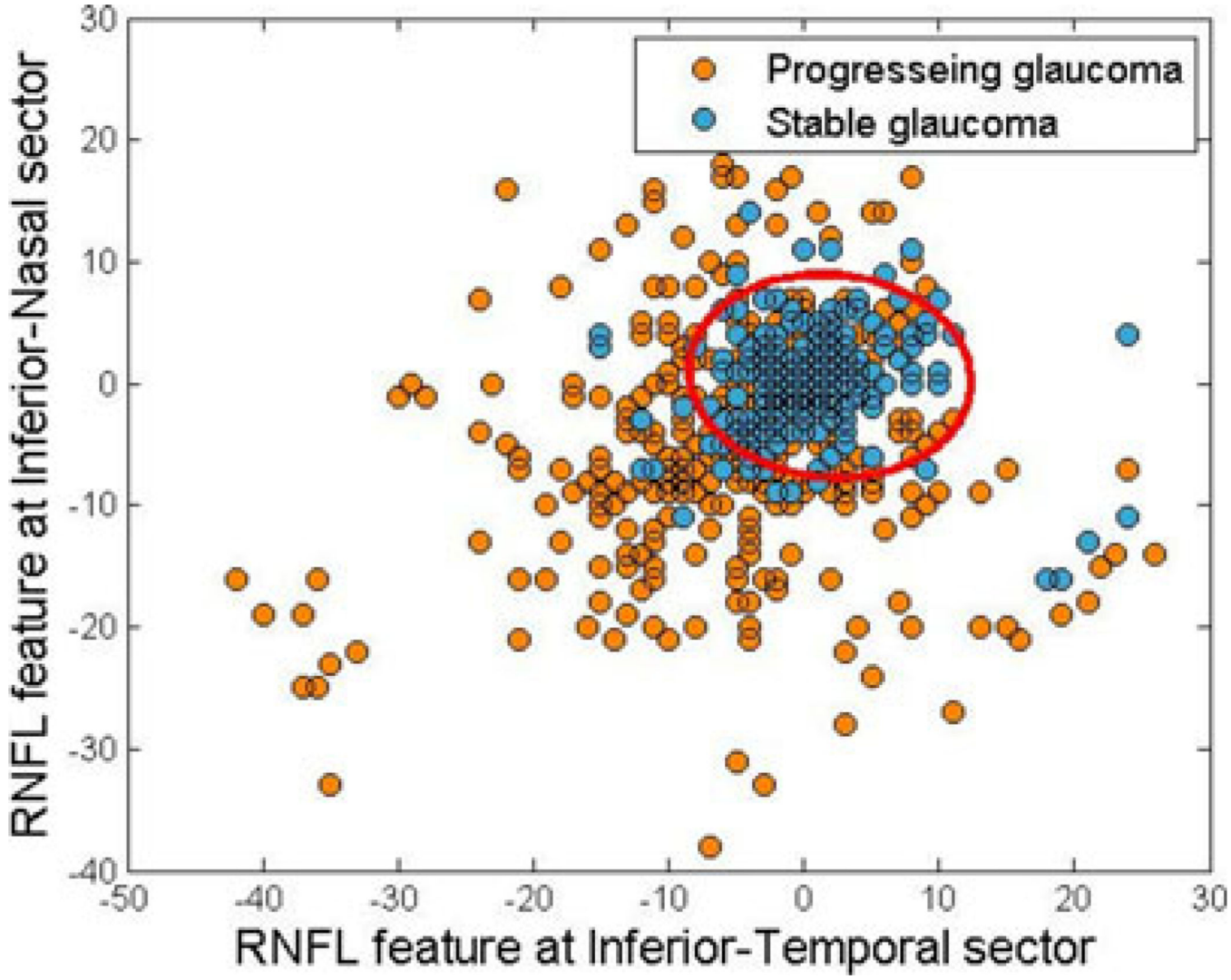

Fig. 6.

Feature distributions and Bayesian discriminating function plotted in 2-D for two best features. Each blue circle represents the RNFL inferior temporal versus inferior nasal feature (difference between RNFL thickness at baseline and follow-up in micron) at a follow-up of an eye that belongs to the stable group, and each orange circle represents RNFL inferior temporal versus inferior nasal feature of an eye that belongs to the progressing group.

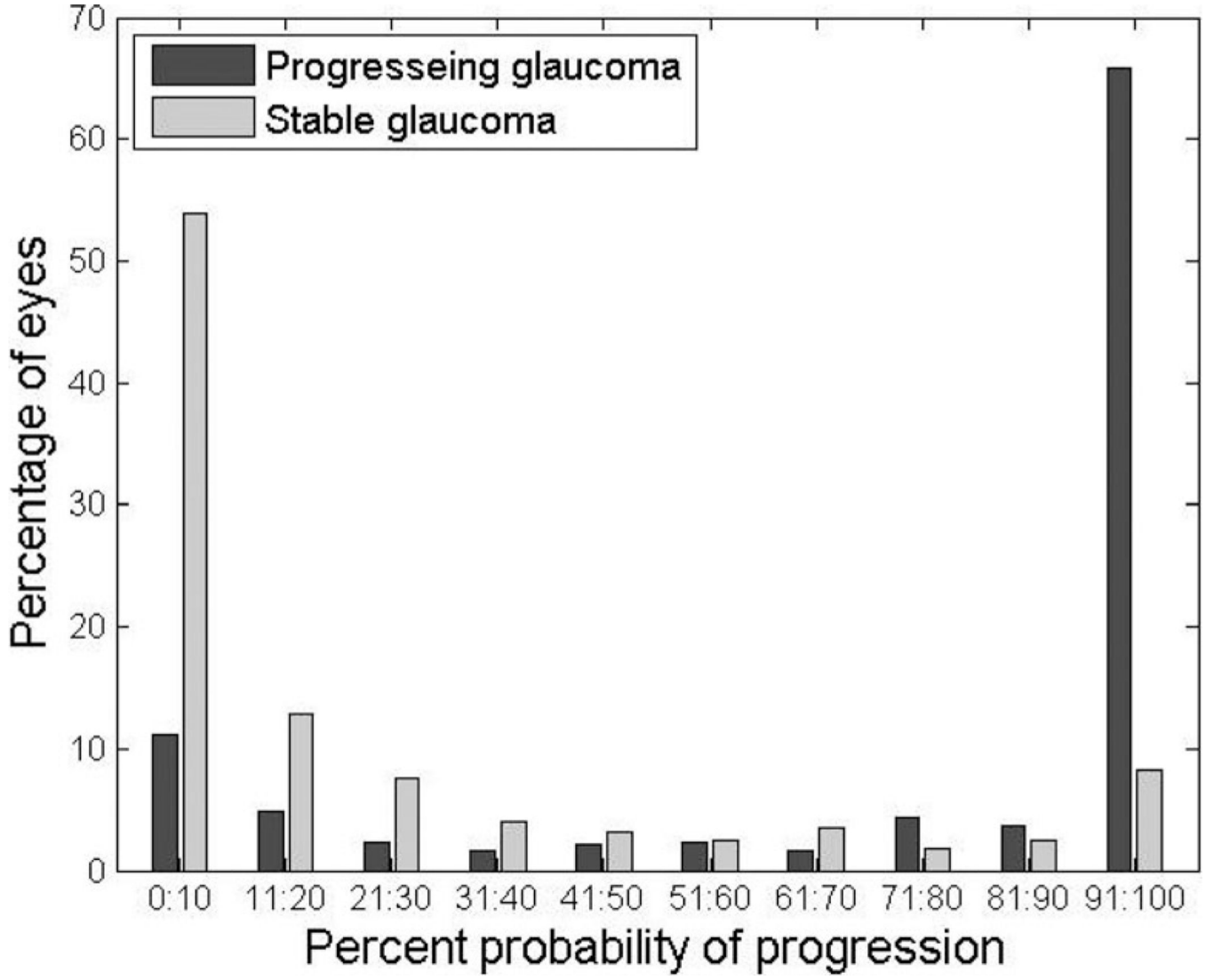

Fig. 7.
Percent probability of progression based on Bayesian Net classifier outcome.

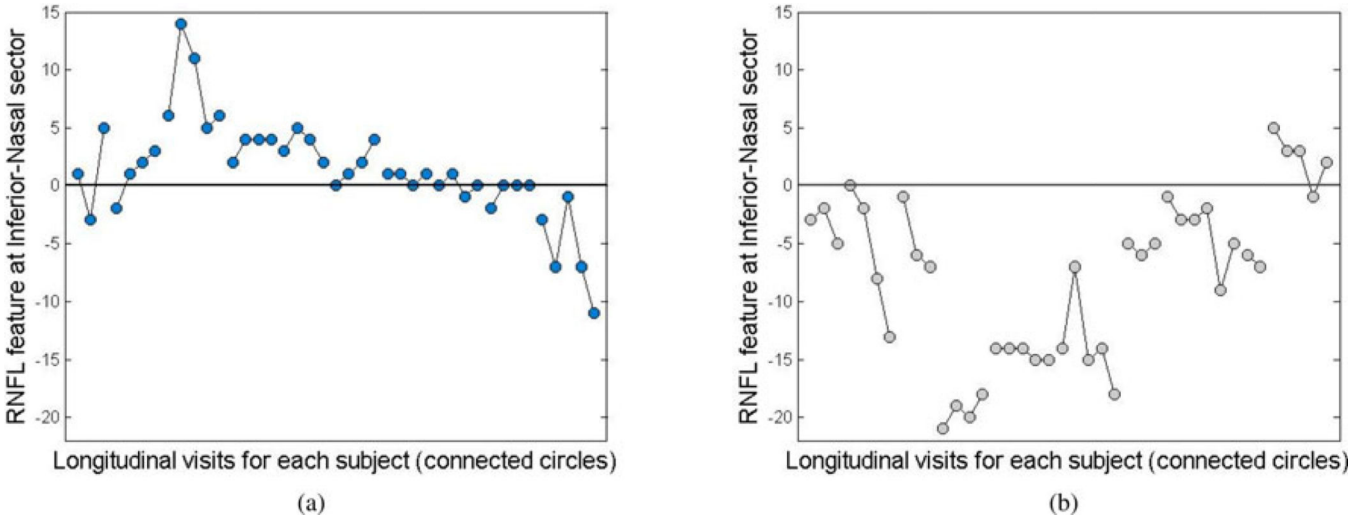

Fig. 8.

Evolution of features over time (visits); (a) RNFL feature at inferior nasal sector for ten stable eyes over the time course of follow ups, (b) RNFL feature at inferior nasal sector for ten progressing eyes over the time course of follow ups. (Connected circles belong to a single eye and each circle represents the feature value at that visit, sorted from first follow-up to the last follow-up from left to right).

TABLE I

Demographic Information From Subjects and Characteristics of Follow-Up Visits

Parameter	Stable	Progressed based on either photo or visual field or structure of function	p-value
Number of eyes	73	107	-
Number of subjects	39	100	-
Number of Follow-ups (SD)	3.9 (1)	4.3 (1.9)	<0.01
Length of Follow-up (SD)	32.2 days (11.9)	2.2 years (1.1)	-
Age at baseline in years (SD)	71.9 (9.0)	66.4 (12.3)	0.01
Gender (percent female)	44	56	0.20
Baseline SAP Mean Defect (MD) in dB (SD)	−5.6 (7.1)	−4.1 (4.7)	0.05
Baseline SAP Pattern Standard Deviation (PSD) in dB (SD)	5.5(4.0)	4.4 (3.8)	0.36

TABLE II

Confusion Matrix

	Actual positive	**Actual negative**
Predicted positive	TP	FP
Predicted negative	FN	TN

TABLE III

Performance Metrics (Weighted Average) Computed Using Combined All RNFL and All SAP Features for Different Classifiers

Classifier	TPR	FPR	F-Measure	MCC	AUPRC	Sensitivity (80% specificity)	Sensitivity (90% specificity)	AUROC (95% CI)
Bayesian Net	0.76	0.22	0.76	0.51	0.83	0.74	0.65	0.84 [0.80,0.87]
Lazy K Star	0.76	0.16	0.76	0.58	0.90	0.78	0.72	0.83 [0.80,0.86]
Meta Classification Regression	0.78	0.25	0.78	0.53	0.84	0.76	0.68	0.85 [0.82,0.88]
Meta Ensemble Selection	0.81	0.23	0.81	0.58	0.85	0.80	0.65	0.86 [0.83,0.89]
Alternating Decision Tree	0.78	0.25	0.77	0.52	0.84	0.75	0.68	0.85 [0.82,0.88]
Random Forest Tree	0.81	0.28	0.80	0.57	0.87	0.82	0.73	**0.88** [0.85,0.91]
Classification And Regression Tree (CART)	0.79	0.25	0.79	0.55	0.83	0.79	0.57	0.83 [0.80,0.82]

TABLE IV

Performance Metrics (Weighted Average) Computed Using Only All RNFL Features for Different Classifiers

Classifier	TPR	FPR	F-Measure	MCC	AUPRC	Sensitivity (80% specificity)	Sensitivity (90% specificity)	AUROC (95% CI)
Bayesian Net	0.77	0.21	0.78	0.54	0.82	0.76	0.68	0.84 [0.81,0.87]
Lazy K Star	0.80	0.21	0.80	0.58	0.88	0.80	0.73	**0.88** [0.86,0.91]
Meta Classification Regression	0.80	0.22	0.80	0.56	0.85	0.78	0.65	0.86 [0.83,0.89]
Meta Ensemble Selection	0.80	0.23	0.80	0.57	0.85	0.80	0.68	0.86 [0.83,0.89]
Alternating Decision Tree	0.79	0.24	0.79	0.55	0.85	0.77	0.71	0.87 [0.84,0.90]
Random Forest Tree	0.79	0.26	0.79	0.54	0.85	0.82	0.74	0.87 [0.84,0.90]
Classification And Regression Tree (CART)	0.79	0.23	0.79	0.55	0.81	0.77	0.59	0.84 [0.81,0.87]

TABLE V

Performance Metrics (Weighted Average) Computed Using Only All SAP Features for Different Classifiers

Classifier	TPR	FPR	F-Measure	MCC	AUPRC	Sensitivity (80% specificity)	Sensitivity (90% specificity)	AUROC (95% CI)
Bayesian Net	0.61	0.63	0.51	0.03	0.57	0.28	0.18	0.53 [0.49,0.58]
Lazy K Star	0.70	0.22	0.71	0.48	0.87	0.70	0.56	**0.82** [0.79,0.86]
Meta Classification Regression	0.65	0.43	0.64	0.22	0.65	0.39	0.24	0.65 [0.61,0.69]
Meta Ensemble Selection	0.69	0.45	0.67	0.28	0.69	0.41	0.25	0.68 [0.64,0.72]
Alternating Decision Tree	0.63	0.49	0.61	0.15	0.65	0.41	0.25	0.64 [0.60,0.68]
Random Forest Tree	0.73	0.40	0.71	0.40	0.80	0.63	0.50	0.80 [0.76,0.83]
Classification And Regression Tree (CART)	0.65	0.46	0.64	0.20	0.62	0.28	0.18	0.61 [0.57,0.65]

TABLE VI

Feature Ranking and Subset Feature Selection

Rank	Correlation based feature ranking	CSF feature selection
1	RNFL at Inferior-Nasal	RNFL at Inferior-Nasal
2	RNFL at Inferior-Temporal	RNFL at Inferior-Temporal
3	RNFL at Global	RNFL at Global
4	RNFL at Superior-Temporal	RNFL at Superior-Temporal
5	SAP point (9° nasal and 15° superior)	SAP point (21° temporal and 9° superior)
6	SAP point (3° nasal and 9° inferior)	SAP point (3° nasal and 9° inferior)
7	SAP point (21° nasal and 3° inferior)	SAP point (21° nasal and 3° inferior)
8	SAP point (9° temporal and 3° inferior)	SAP point (3° nasal and 3° inferior)
9	SAP point (21° temporal and 3° superior)	SAP point (21° temporal and 3° inferior)
10	RNFL at Temporal	RNFL at Temporal

TABLE VII

AUROC (95% CI) When Selecting 1, 2, and 10 Best Features Using Independent Feature Selection

Classifier	AUROC (95% CI) using 1 feature	AUROC (95% CI) using 2 features	AUROC (95% CI) using 10 features
Bayesian Net	0.74 [0.70 0.77]	0.81 [0.78 0.84]	0.83 [0.80,0.86]
Lazy K Star	0.80 [0.76 0.83]	0.83 [0.80 0.86]	0.84 [0.81,0.87]
Meta Classification Regression	0.78 [0.75 0.82]	0.83 [0.80 0.86]	0.83 [0.79,0.86]
Meta Ensemble Selection	0.78 [0.74 0.81]	0.83 [0.80 0.86]	0.86 [0.83,0.89]
Alternating Decision Tree	0.78 [0.74 0.82]	0.83 [0.80 0.86]	0.83 [0.80,0.87]
Random Forest Tree	0.77 [0.74 0.81]	0.78 [0.74 0.82]	0.88 [0.85,0.91]
Classification And Regression Tree (CART)	0.76 [0.71 0.79]	0.78 [0.74 0.81]	0.82 [0.79,0.86]