OPEN

Estimating Rates of Progression and Predicting Future Visual Fields in Glaucoma Using a Deep Variational Autoencoder

Samuel I. Berchuck[1,2], Sayan Mukherjee[3] & Felipe A. Medeiros[1]*

In this manuscript we develop a deep learning algorithm to improve estimation of rates of progression and prediction of future patterns of visual field loss in glaucoma. A generalized variational auto-encoder (VAE) was trained to learn a low-dimensional representation of standard automated perimetry (SAP) visual fields using 29,161 fields from 3,832 patients. The VAE was trained on a 90% sample of the data, with randomization at the patient level. Using the remaining 10%, rates of progression and predictions were generated, with comparisons to SAP mean deviation (MD) rates and point-wise (PW) regression predictions, respectively. The longitudinal rate of change through the VAE latent space (e.g., with eight dimensions) detected a significantly higher proportion of progression than MD at two (25% vs. 9%) and four (35% vs 15%) years from baseline. Early on, VAE improved prediction over PW, with significantly smaller mean absolute error in predicting the 4th, 6th and 8th visits from the first three (e.g., visit eight: VAE8: 5.14 dB vs. PW: 8.07 dB; P < 0.001). A deep VAE can be used for assessing both rates and trajectories of progression in glaucoma, with the additional benefit of being a generative technique capable of predicting future patterns of visual field damage.

Glaucoma is a progressive optic neuropathy that results in characteristic changes to the optic disc and retinal nerve fiber layer[1]. Although damage from glaucoma is irreversible, early treatment can usually prevent or slow down progression to functional damage and visual impairment[2]. Assessment of functional damage is essential for management of the disease and standard automated perimetry (SAP) remains the default method for monitoring functional changes in the disease[3].

Estimation of rates of functional deterioration by SAP is essential for determining patient prognosis and aggressiveness of therapy. However, the best method for estimating such rates is still a matter of controversy[4]. While rates of change can be estimated by global parameters, such as mean deviation (MD), these rates would potentially ignore fast localized losses occurring in an otherwise mostly stable field. Several previous studies have approached this problem by attempting to project the high-dimensional (i.e., multiple locations) visual field to a lower dimensional representation, thus reducing variability and removing collinearity in the visual field measurements. This has been implemented successfully using a variety of unsupervised dimension reduction techniques[5–8], that learn latent features derived from the original visual field data. Therefore, from the original 52-dimensional visual field on the SAP 24-2 test, for example, these techniques arrive at a much lower number of "meaningful" latent dimensions summarizing the original visual field.

Rates of progression can then be derived by analyzing the slopes of latent features across time. However, although previously used techniques may be able to learn a latent feature space; they cannot be used as realistic generators of visual field data, as they are designed explicitly for dimension reduction. Besides estimating how fast a patient is progressing (i.e., the current rate of change), it is also important to be able to generate, or predict, future visual field data from the currently available data for that patient. Such predictions would carry important prognostic value, by determining the likely areas of future damage over time which can help inform the impact of the disease on quality of life[9].

[1]Duke Eye Center and Department of Ophthalmology, Duke University, Durham, North Carolina, USA. [2]Department of Statistical Science and Forge, Duke University, Durham, North Carolina, USA. [3]Departments of Statistical Science, Mathematics, Computer Science, Biostatistics & Bioinformatics, Duke University, Durham, North Carolina, USA. *email: felipe.medeiros@duke.edu
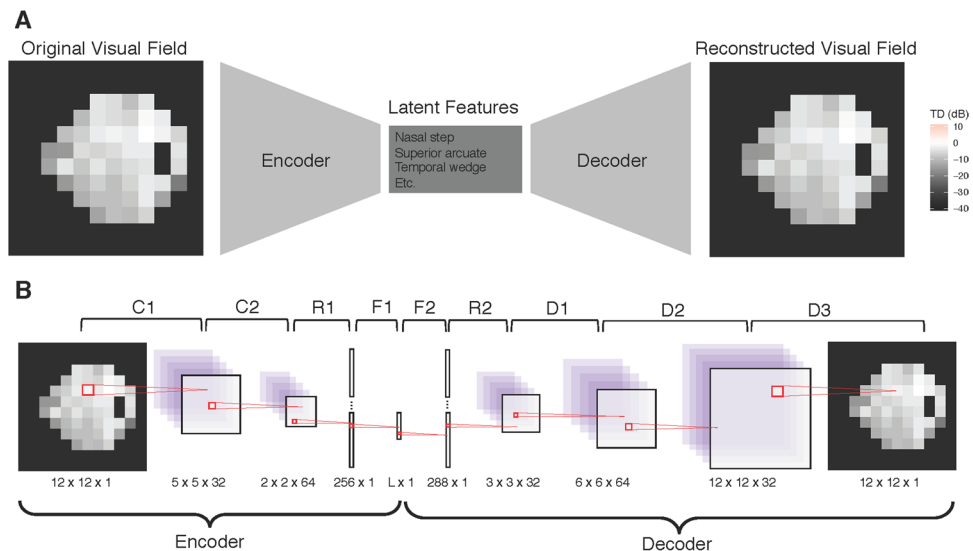
Figure 1. Illustrating the VAE for visual fields. (**A**) The top frame demonstrates the structure of the VAE, which is a dual-mapping from the original visual field to the latent features and then back to a reconstructed visual field. (**B**) The bottom frame provides details on the network architecture. There are four types of layers: convolutional (C), resizing (R), fully connected (F), and de-convolutional (D). The sizes at the bottom of each image reflect the transformed dimensions after each layer. For each layer, the kernel is 3×3 with a stride of 2. The activation function for all layers is the rectified non-linear unit, except for F1 and D3, which have the identity and sigmoid activations, respectively. The latent dimension, L, is a user specified parameter.

In contrast to previously used models, a generative model learns a data generating mechanism (i.e., distribution), thus allowing for predictions of future visual fields. Historically, the task of learning a distribution of a high-dimensional data object was untenable without making simplifying assumptions, leading to the use of techniques like point-wise (PW) regression for visual fields[10]. Modeling high-dimensional processes has become a reality with the explosion of deep learning. Deep learning has been applied effectively in health application, including the field of ophthalmology[11–13], and in particular for visual fields[14,15]. The most common generative models are the generative adversarial network and variational auto-encoder (VAE)[16,17]. These methods can learn the complex distributions of high-dimensional biological data by first learning a projection to a lower-dimensional latent space, and then a mapping back to the original image. The latent space of these generative models has previously been shown to reveal novel biological patterns[18,19]. In this manuscript, we implement a general form of the VAE to learn a clinically motivated latent representation which can be used to determine rates of disease progression, and consequently, a generative process for visual field data.

Methods

Visual fields included in this analysis came from participants enrolled in a prospective longitudinal study designed to evaluate functional impairment in glaucoma. Written informed consent was obtained from all participants and the institutional review board and human subjects committee at Duke University approved all methods. All methods adhered to the tenets of the Declaration of Helsinki for research involving human subjects and the study was conducted in accordance with the regulations of the Health Insurance Portability and Accountability Act.

During follow-up, patients underwent comprehensive ophthalmologic examinations, including review of medical history, visual acuity, slit-lamp biomicroscopy, intraocular pressure measurement, gonioscopy, dilated funduscopic examination, stereoscopic optic disc photography, and SAP using 24-2 Swedish interactive threshold algorithm standard (Carl Zeiss Meditec, Inc, Dublin, California, USA). Visual fields were excluded in the presence of eyelid or rim artifacts, fatigue effects, or evidence that the visual field results were caused by a disease other than glaucoma. Visual fields were also excluded if they had more than 33% fixation losses or more than 15% false-positive errors.

The VAE is an unsupervised technique, however in order to visualize the latent feature space of the VAE, we defined disease status as normal, suspect, and glaucoma. Glaucoma was defined by the presence of two or more repeatable glaucomatous visual field defects at baseline, defined as a pattern standard deviation with $P < 0.05$, or a Glaucoma Hemifield Test result outside normal limits, and corresponding optic nerve damage. Eyes were considered normal if they were recruited from the general population and had no visual field defects. The remaining eyes were considered suspects and had history of high intraocular pressure or suspicious glaucomatous appearance of the optic nerve, but in the absence of confirmed visual field defects.

Visual field preparation. The 24-2 SAP produces visual fields with 52 informative data observations that together constitute an eye's field of vision. The shape of a visual field is not rectangular, as is typically required for deep learning models, so square images were created by padding the visual field with zeros. As a result, the visual field images in our analysis were transformed to 12×12 squares (see Fig. 1A for an example visual field used in

	Total	Machine Learning Datasets		
		Training	Validation	Test
Number of Visual Fields	29,161	23,744 (81%)	2,577 (9%)	2,840 (9%)
Number of Patients	3,832	3,108 (81%)	363 (9%)	361 (9%)
Normal (n, %)	667 (17%)	547 (18%)	58 (16%)	62 (17%)
Suspect (n, %)	2,221 (58%)	1,793 (58%)	222 (61%)	206 (57%)
Glaucoma (n, %)	944 (25%)	768 (25%)	83 (23%)	93 (26%)
Number of visits	7.61 (7.35)	7.64 (7.36)	7.10 (7.24)	7.87 (7.29)
Follow-up period (years)	4.95 (5.25)	4.94 (5.24)	4.67 (5.36)	5.33 (5.24)
Age (years)	60.94 (15.66)	61.04 (15.77)	60.38 (15.9)	60.67 (14.47)
Baseline MD (dB)	−3.55 (5.71)	−3.57 (5.72)	−3.76 (6.12)	−3.23 (5.17)
Baseline PSD (dB)	3.42 (3.29)	3.42 (3.27)	3.49 (3.41)	3.42 (3.34)

Table 1. Demographic and clinical characteristics of subjects included in the study across training, validation, and test data sets. The training, validation, and test datasets are for one cross-validation fold. MD = mean deviation; PSD = Pattern standard deviation. Values are presented as mean (standard deviation), unless otherwise noted.

the analysis). To prepare the visual fields for the deep learning algorithm, all visual field images were converted to right eyes for uniformity. In this study, we represented functional loss using total deviation (TD) values, an age-adjusted measure of sensitivity loss, measured in decibels (dB). TD is a continuous measure, with large negative values indicating functional loss.

Prior to analysis, all visual fields were normalized to be in the range of zero and one. The training, validation, and test datasets were created by randomly sampling patients from the overall study population with 80%, 10%, and 10% probability, respectively. The randomization process was performed at the patient level, so that all images from a patient are included in at most one dataset. The demographics of the study datasets are illustrated in Table 1.

A deep generalized variational autoencoder for visual fields.   We utilized a VAE framework to create a generative process for visual field data. A VAE is an unsupervised learning method that transforms high-dimensional data to a small set of latent features[17]. Furthermore, the advantage of the VAE is that the latent features can be disentangled to generate synthetic visual fields[20]. The VAE was derived from variational Bayesian methods[21], and the model is optimized to have minimal image reconstruction loss, while maintaining a homogenous feature space based on a regularization prior. While the original VAE uses the Kullback-Leibler divergence for regularization[22], for visual field data, we use an alternate form of the VAE that uses the maximum mean discrepancy[23]. While this changes the interpretation of the VAE as a probabilistic model, it maintains all other properties. Going forward we refer to our model as a VAE to be consistent with machine learning literature.

VAEs consist of two components, the encoder and decoder. The encoder is a function that takes in a visual field and compresses it to a lower-dimensional space. The decoder (sometimes called the generator) is a probabilistic distribution that maps the latent features back to the original high-dimensional image space. Since the decoder is trained to reconstruct the original data using a small number of latent dimensions, those features end up representing key aspects of the data. In other words, the encoder-decoder process is forced to learn the most important features of the original data. Furthermore, because there is a regularization prior, the VAE learns features that are homogenous, and therefore can be interpreted clinically. The bottleneck structure of the VAE is presented in Fig. 1A.

Both the encoder and decoder used deep neural networks. The encoder is itself a network, but for the decoder, the network is the mean of a Gaussian distribution. Details of both networks can be found in Fig. 1B. The encoder of the VAE is comprised of two 2D convolutional layers, a reshaping layer and a fully-connected dense layer. The decoder begins with a fully-connected dense layer, followed by a reshaping layer, and then a sequence of de-convolutions, which up-sample until the original 12 × 12 dimension is reached. All layers use a 3 × 3 kernel size and a stride of two, and the activation is a rectified non-linear unit transformation, except for F1 and D3, which used the identity and sigmoid activations, respectively.

For each choice of latent dimension, the model was trained using the Adam optimizer, an extension of stochastic gradient descent, using 100 epochs and a batch size of 100[24]; we used a learning rate of 1e-4. The training epoch with the minimal validation loss was chosen as optimal. For robustness, we used cross-validation with five folds. The VAE was implemented using the deep learning library Keras (version 2.2)[25] with Tensorflow (version 1.9)[26] backend, all within RStudio (3.5.1)[27].

Dimension of the latent features.   The dimension of the latent space is a parameter that was specified to reflect the number of latent features that can adequately explain the high-dimensional visual fields. In this study, we explored the performance of the VAE over latent dimensions ranging from one to fifteen. Throughout the analysis all results based on the VAE are reported for all fifteen variations of the latent dimension in an attempt to determine the preferred latent space. To determine the optimal dimension of the latent space, we compared clinical performance metrics across dimensions.

Rates of visual field progression. We used the latent features of the VAE to determine rates of progression for glaucoma patients in the test dataset. For each patient, we obtained the latent features corresponding to each visual field using the encoder. Then, we studied the longitudinal trends of the latent features (instead of the complex 52-dimensional visual field). In particular, we quantified progression using both a linear and non-linear method. For the linear technique, progression was determined based on the global rate of change across all features from a model. In order to calculate this global rate, we performed ordinary least squares (OLS) regression with a zero-sum constraint on the design matrix; thus, allowing us to perform a hypothesis test on the mean rate of change across time. Progression was defined as a significant global rate of change based on a two-sided hypothesis test with type-1 error of 0.05. For the non-linear method, the same framework was used, with the addition of a quadratic term. Then, progression was defined as a significant global rate of change in either the linear or quadratic term using a joint F test, again using a two-sided hypothesis test with type-1 error of 0.05.

This rate of visual field progression can be interpreted as the speed of deterioration for a patient's visual field features, where features represent some lower dimensional representation of the visual field. This is reminiscent of SAP MD, which is a comparison of a current visual field to an age adjusted healthy baseline in one-dimension. As such, for comparison we used standard rates of visual field progression of MD using OLS linear regression across time, with progression defined as a significantly negative rate of change over time (alpha = 0.05).

Finally, in the absence of a gold standard of progression, we compared the progression methods by matching their specificities at 95%. Because all methods are p-values with a type-1 error of 0.05, this is automatically achieved with a cutoff of 0.05. Therefore, the method with the higher rate of detecting progression in the glaucoma patients in the test dataset has superior operating characteristics. Note that this is not a sensitivity, because not all patients are progressing, but rather a progression hit rate. For each method, we estimated hit percentage at two, four and six years from baseline visits and also for all patient follow-up. Only visits that occurred on or before the cutoff time are included in the analysis. This imitated a clinical setting where each metric is calculated at every visit and progression is diagnosed. For hypothesis testing, bootstrapped confidence intervals were presented[28,29] and all results are averaged across the cross-validation folds for robustness.

Predicting future visual fields. An advantage to the VAE modeling framework is the ability to generate visual fields from its latent features through the decoder. This motivated a method for predicting future visual fields through a two-stage procedure. First, for a longitudinal collection of visual fields, we obtained their corresponding latent features (obtained from the encoder) and modeled each dimension independently. This modeling was done using both the linear and non-linear regressions used for determining rates of change. Then, we used the predicted values of the latent features to generate future visual fields using the decoder. Note that because the decoder is a mean process, the generated visual fields are de-noised.

We assessed the prediction accuracy of this two-stage approach by predicting five consecutive visits, following the third, fifth, and seventh visits. Prediction results are presented for all patients, and only glaucoma patients, in the test dataset. Prediction accuracy was assessed using mean absolute error (MAE) for only the 52 informative locations (i.e., not the full 12×12 image). All results are compared to the established technique of visual field prediction, PW linear regression. One-sided Wilcoxon signed rank tests are presented to formally compare if the MAE from the VAE is smaller than from PW, using a Bonferroni corrected type-1 error, $0.05/900 \approx 0.00006$. All results are averaged across the cross-validation folds.

Results

This study included 29,701 visual fields; however, 540 visual fields were excluded due to false-positives greater than 15% or fixation error greater than 33%. This yielded 29,161 usable visual fields from 3,832 eyes. These patient eyes had an average follow-up of 4.95 years with a mean of 7.61 visits. Mean age at baseline was 60.94, mean baseline MD was −3.55 dB and mean baseline PSD was 3.42. From the usable visual fields, we created training, validation, and test datasets, by randomly sampling eyes with probability 80%, 10%, and 10%, respectively. Population characteristics for are presented in Table 1, where the training, validation, and test datasets represent summaries for a particular cross-validation fold.

The models were trained for 100 epochs across five cross-validation folds and the optimal losses are presented in Fig. 2 for the training and validation datasets across models with varying latent dimensions. On average, the minimal validation loss across trainings occurred at epoch 35. From the figure it is clear that the reconstruction losses were nearly zero for all models and that the regularization loss decreased with the number of dimensions after three dimensions, indicating that the latent space becomes more homogenous as the dimensions increase, while maintaining similar reconstruction. To determine the optimal number of latent dimensions, we will study the clinical results related to rates of progression and prediction.

We investigated the rates of visual progression determined from the latent space of the VAE and SAP MD. To motivate using the VAE for detecting progression, we have presented the longitudinal follow-up for examples patients of each disease status in a two-dimensional latent space (Fig. 3A). The predicted trajectories are presented as arrows and the observed longitudinal follow-up are underlaid using muted colors (non-linear trajectories were excluded because they were indistinguishable). Furthermore, the location of a healthy visual field (blue square) is displayed, where healthy is defined as the mean visual field of all the healthy eyes. In Fig. 3B,C, the first and second latent feature are shown across time with OLS regression lines indicating rates of change for each feature. The linear VAE progression metric is the average of feature slopes, which is a calculation of the speed of movement through the latent space. The non-linear metric is similar, but also accounts for any non-linear movement.

In Fig. 4, the progression detection hit rates are presented for of each of the diagnostic measures. The vertical line in each frame is a representation of the 95th percentile of MD, and thus metrics with no overlap are significantly superior. At two years from baseline, the linear VAE models with seven (Hit Rate: 0.24), eight (0.25), and ten (0.35) dimensions were significantly superior at detecting progression over MD (0.09); while at four years
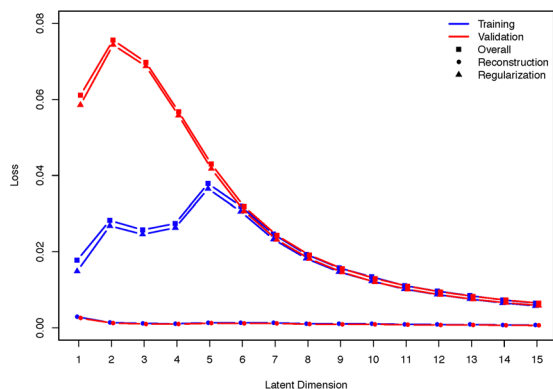
Figure 2. Training and validation loss over different latent dimensions. Overall loss is presented along with reconstruction and regularization loss.
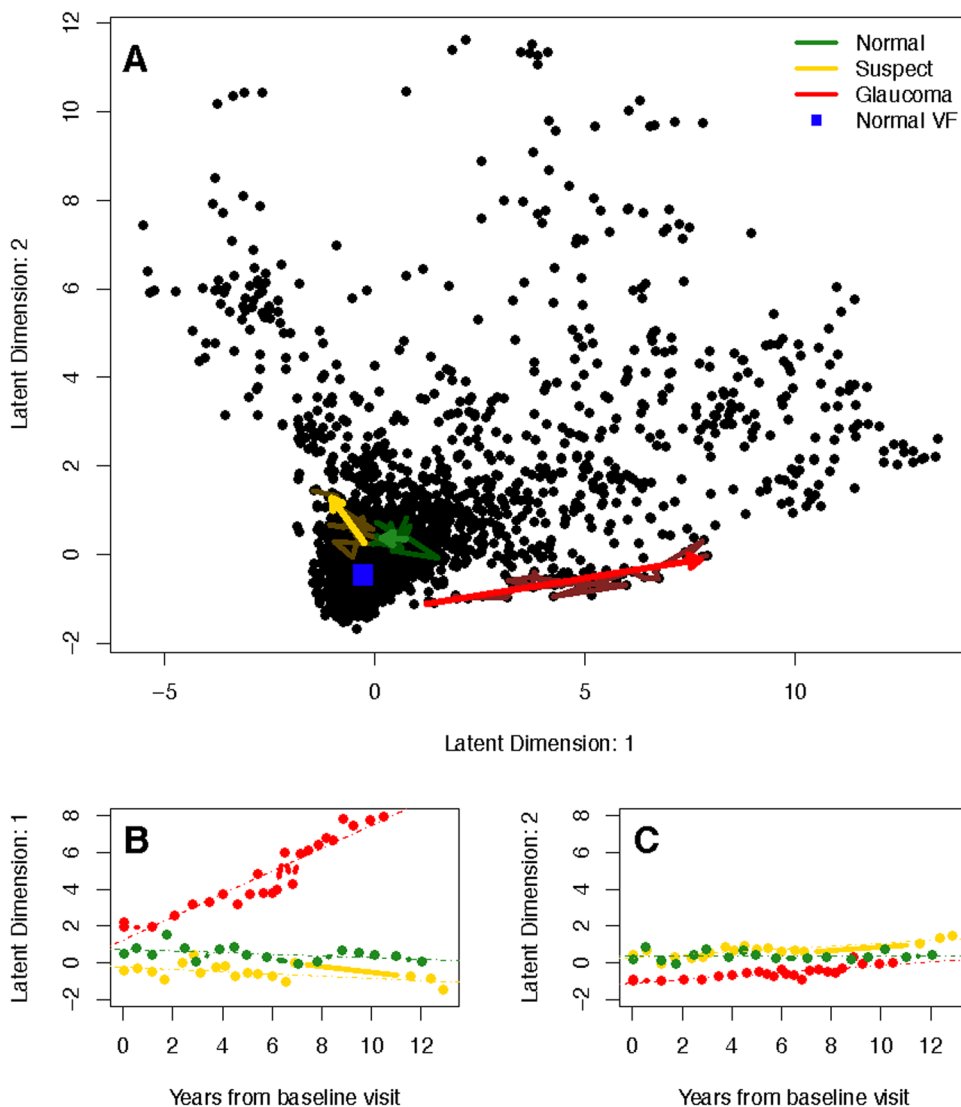

Figure 3. Visualizing the latent space of the VAE in two dimensions using a linear model. (**A**) The top frame shows predicted trajectories through the latent space, presented as arrows, are presented for example patients of each disease status. Underlaying the trajectories are the observed longitudinal follow-up for each patient using muted colors. The healthy visual field in latent space is a blue square near the origin. (**B**) The bottom left frame shows the longitudinal latent features from the first dimension for the example patients. OLS regression lines are presented, which indicate rates of change in each latent dimension. (**C**) The second dimension is shown in the bottom right.
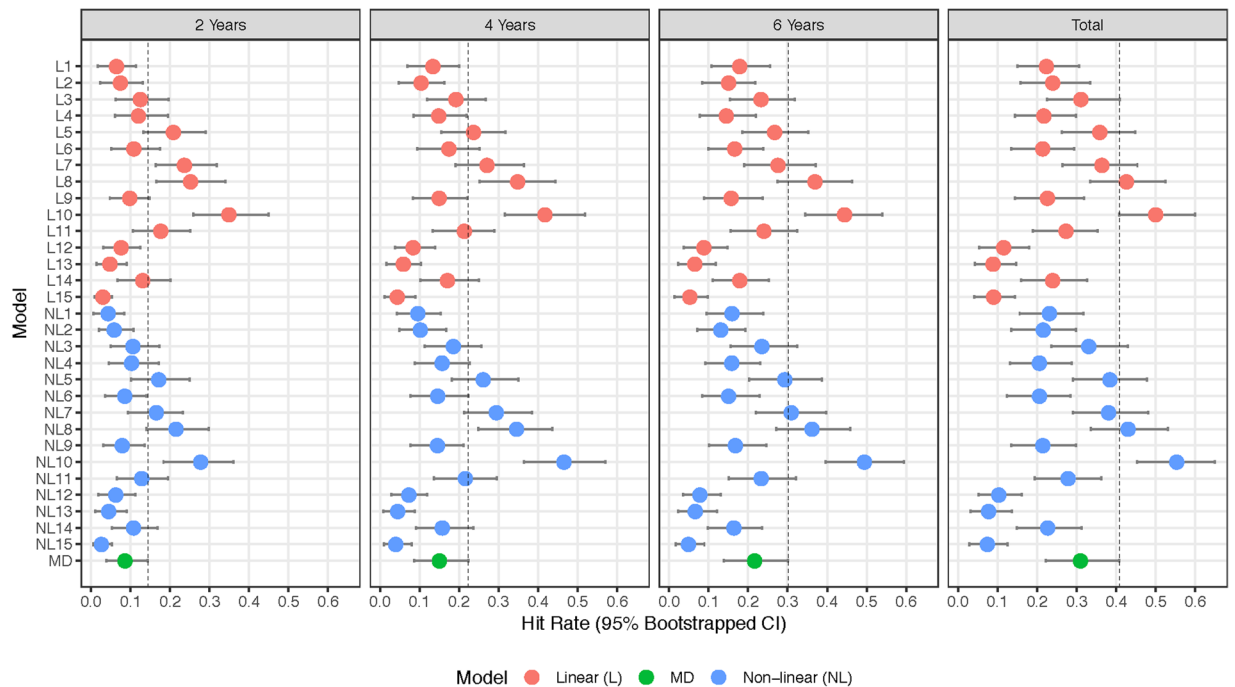
Figure 4. Visual field progression hit rate for diagnostic measures at two, four, and six years from baseline visit, and total follow-up, for glaucoma patients in the test dataset. Error bands represent 95% bootstrapped confidence intervals and are averaged over the five cross-validation datasets.

the models with eight (0.35) and ten (0.42) dimensions remained significant over MD (0.15), and at six years the model with ten dimensions (0.44) was significant over MD (0.22). The significance disappears when using all of the data in the linear model. A similar trend held for the non-linear VAE, and in particular the model with ten latent dimensions (2 Years: 0.28, 4 Years: 0.47, 6 Years: 0.49, Total: 0.55) was significant at all four time periods over MD (2 Years: 0.09, 4 Years: 0.15, 6 Years: 0.22, Total: 0.31). The summaries of the hit rates can be found in Supplementary Table S1.

To verify that the trained models effectively learned the generative process we assessed the predictive capacity and the results are presented in Figs. 5 and 6. Figure 5 shows boxplots of the MAE for all patients in the test dataset across all prediction scenarios and Fig. 6 shows results for only glaucoma patients. The y-axis of the boxplots has been truncated for visualization purposes, however all significance tests account for all data. Predictions for the PW model are given in blue, and all models with significantly smaller prediction errors, in relation to the PW model, are presented in red. Each column corresponds to the number of visits used for prediction (3: left, 5: middle, 7: right), and each row represents the number of visits predicted into the future (1–5). In general, the linear VAE performs better than PW prediction earlier on in follow-up (i.e., fewer visits used for prediction) and when predicting further into the future. The non-linear VAE method performs poorly across all settings. A similar pattern holds for glaucoma patients, with superiority of the linear VAE prediction becoming less striking overall. All prediction summaries are presented in the Supplementary Tables S2–7.

The ability of the VAE to predict future visual fields is further illustrated in Fig. 7, where predictions are presented of an example glaucoma patient with severe disease. The predictions, presented using the Humphrey Field Analyzer printout, are for the eighth visit and were based on the patient's initial three visits. Visual fields are presented for the true visual field, PW and the linear VAE models with eight and ten dimensions. These printouts are meant to be interpreted visually, not through the text, which is small due to the size of the printouts.

To further demonstrate the clinical utility of the VAE, we visualized it with a two-dimensional latent space across clinical measures. Figure 8 shows the relationship between the VAE latent feature space and clinical variables, disease status, age, MD, and PSD. Finally, in Fig. 9, we presented the generative distribution derived from the VAE for two latent dimensions, with the predicted trajectories of each of the example patients included. Overlaying each path is a blue line that indicates the distance traveled in the first year of follow-up, which is constant across time. These figures are presented utilizing one fold of the cross-validation.

Discussion

In this manuscript, we developed and validated a novel deep learning algorithm that produced a unified framework for learning the generative process of visual fields and detecting rates of glaucoma progression. The VAE used a low-dimensional latent space representation of the more complex and high-dimensional visual field image to produce a clinically relevant latent space. This was achieved through the model structure of the VAE, which learned the latent features through both a compression (i.e., encoder) and generative (i.e., decoder) process. Standard techniques for learning latent features, such as probabilistic principle components, factor analysis or independent components analysis, only include the compression component and, consequently, the latent space
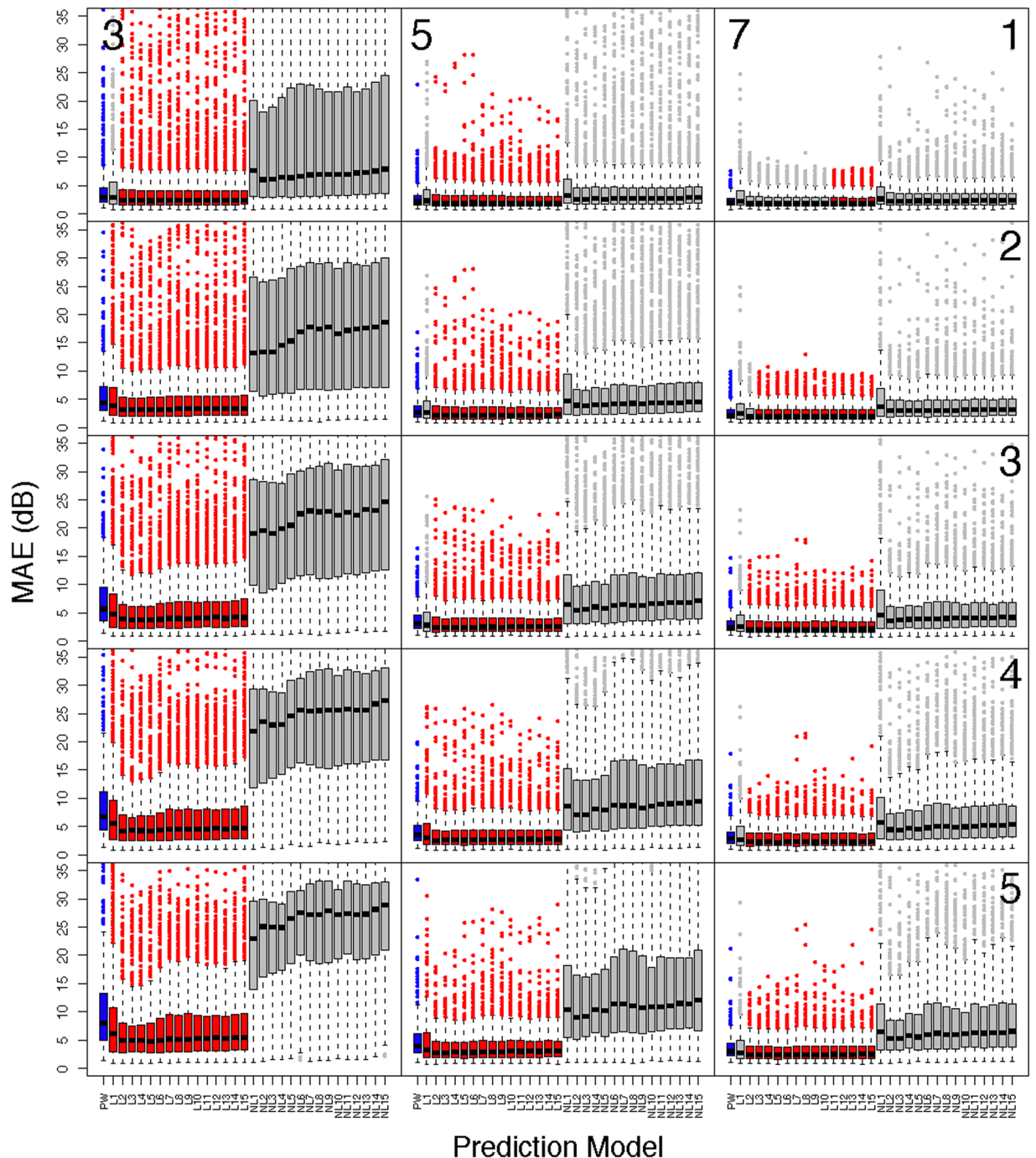
Figure 5. Comparing the prediction accuracy of the linear and non-linear VAE and PW methods in **all** patients from the test dataset. Boxplots present summaries of prediction using MAE (dB) over the varying number of latent dimensions. Each column indicates the number of visits used in prediction (3: left, 5: middle, 7:right), and the rows indicate how many visits into the future are predicted (1–5). The PW prediction is given in blue, and any model that is significantly smaller than PW is given in red. Significance is based on a one-sided Wilcoxon signed rank tests, using a Bonferroni corrected type-1 error, $0.05/900 \approx 0.00006$. All predictions are averaged across cross-validation folds.

can often be less realistic[7,30]. Furthermore, these standard techniques assume the latent features are a linear combination of the original image, while the VAE uses deep learning to allow for arbitrarily non-linear mappings.

Previous studies have used machine learning approaches to assess glaucomatous damage from visual fields[13,14,31,32]. In these studies, the glaucomatous status of the eyes was determined either by clinical measures or by clinical expertise, making the machine learning performance dependent on the chosen gold standard. This type of technique can be limited, because the end-point of interest is often inaccurate or different from the true clinical end-point. We circumvented this issue in our study by using an unsupervised method. Instead of training
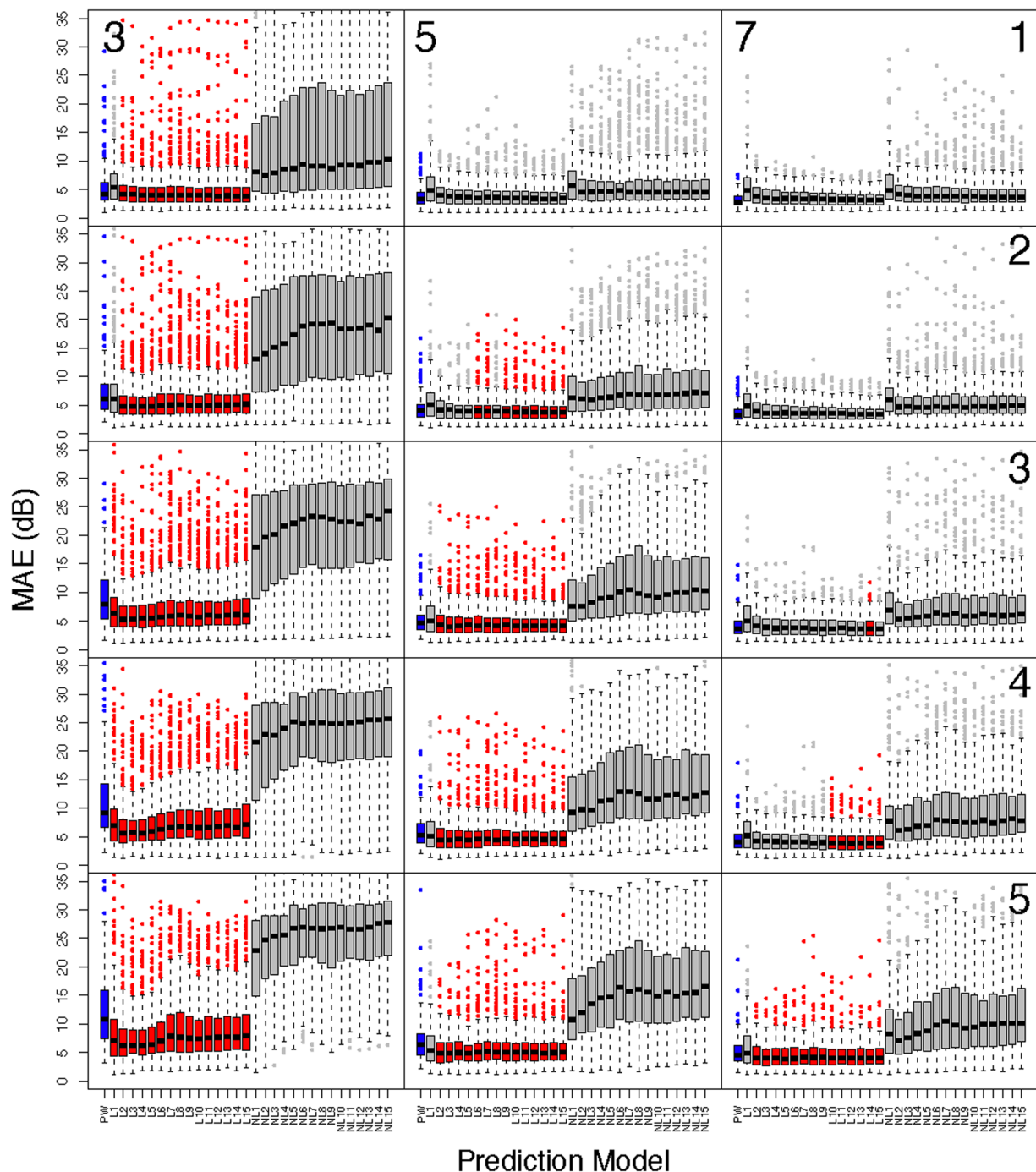
Figure 6. Comparing the prediction accuracy of the linear and non-linear VAE and PW methods in **glaucoma** patients from the test dataset. Boxplots present summaries of prediction using MAE (dB) over the varying number of latent dimensions. Each column indicates the number of visits used in prediction (3: left, 5: middle, 7:right), and the rows indicate how many visits into the future are predicted (1–5). The PW prediction is given in blue, and any model that is significantly smaller than PW is given in red. Significance is based on a one-sided Wilcoxon signed rank tests, using a Bonferroni corrected type-1 error, $0.05/900 \approx 0.00006$. All predictions are averaged across cross-validation folds.

against a noisy gold standard, the VAE trains a visual field against itself, in the process developing a generative model through clinically interesting latent features.

To visualize the latent features for visual fields, we presented the latent space in two-dimensions for clinical measures, glaucoma status, age, MD, and PSD (Fig. 8). Through inspection of these figures the latent space comes to life. Across all clinical measures it is clear that patients with no evidence of glaucoma reside near the origin, as this is the location of the normal disease patients, and young patients with MD and PSD values near zero. As the
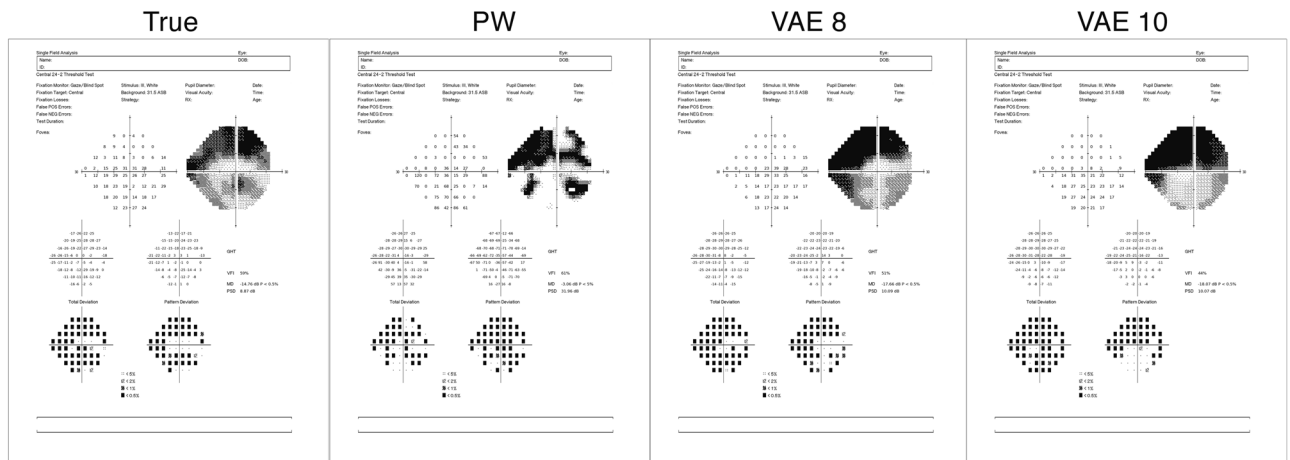
Figure 7. Demonstrating prediction capability through Zeiss printouts from an example glaucoma patient in the test dataset. Predictions are of the eighth visit using the first three. The linear VAE model presented is with eight and ten latent dimensions. These printouts are meant to be interpreted visually, not through the text, which is small due to the size of the printouts.
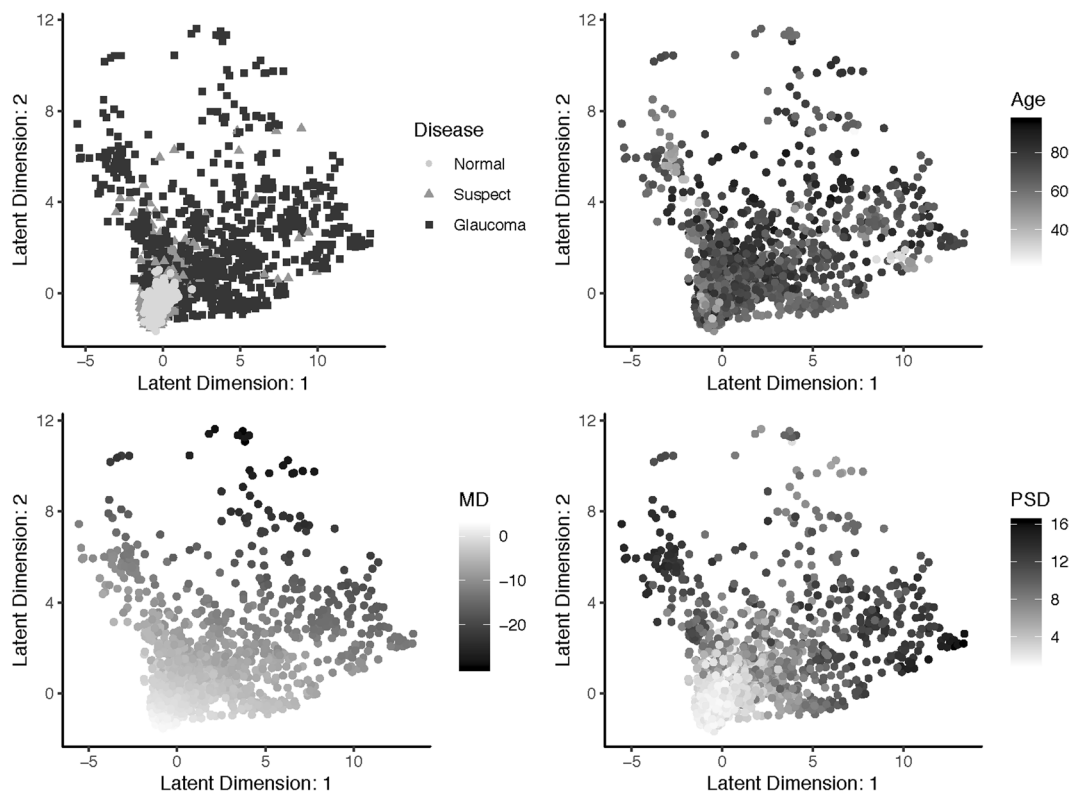

Figure 8. Visualization of the relationship between the VAE latent feature space in two dimensions and clinical variables, disease status, age, MD, and PSD. Visualizations are from one cross-validation fold.

latent features increase in either dimension the disease severity increases. In particular, both age and MD values appear to become more severe (i.e., older age and more negative MD) with an increase in either of the features or a combination of both. In contrast, PSD only demonstrates more severity (i.e., larger values) with an increase in either of the latent features, but not both. This illuminates a clinical phenomenon, that MD and age are linear risk factors, while PSD identifies local defects, which are typically not evident in the late stages of disease.

The latent space can also be used to visualize the generative process for visual fields. In Fig. 9, the distribution of visual fields across the two-dimensional latent space was displayed. This presentation provides more clinical context to the latent space, as the generative process is a smooth representation of the variability in visual fields. In particular, we can see the patterns of disease severity, as an increase in the second latent feature appears to indicate
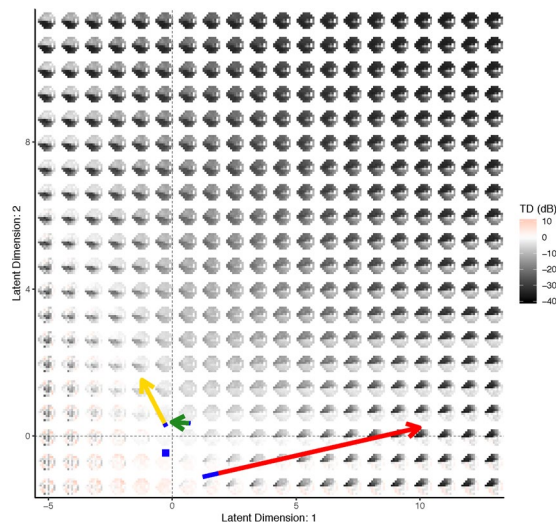
Figure 9. Visualizing the VAE generative distribution for visual fields with two latent dimensions. Included are the paths of the example patients from Fig. 3. The length of the blue line overlaying each path indicates the speed of change across time and is the length traveled in the first year. The average visual field from all of the normal patients is represented as a blue square. Visualizations are from one cross-validation fold.

a global worsening in the functional vision, while the first feature dictates worsening patterns in the superior (or inferior) hemisphere when moving to the right (or left) of the origin. Because the VAE produced a latent feature space that was clinically informative, it had utility for determining rates of progression.

In particular, we leveraged the clinical interpretation of the latent space to calculate progression rates by measuring the rate of a patient's movement through the generative distribution. We formally defined the progression rate based on the average of the rates across features, both in a linear and non-linear manner. This process was exemplified through the visual fields of a normal, suspect, and glaucoma patient in Fig. 3. The movement of each of the patients through latent space was displayed, along with the linear trajectories of the features across time. The glaucoma patient has the highest rate, indicated by the large positive slope in the first feature. In Fig. 9, the trajectory of each of the example patients overlays the generative process, with an indicator of the rate of movement through latent space (blue line). This novel presentation of the trajectories through the distribution of visual fields provided a literal clinical road map for making treatment decisions.

We provided evidence that the rate of change through latent space is superior to MD (Fig. 4) across the range of follow-up for a number of VAE models with both linear and non-linear rates, in particular with ten dimensions. This is not surprising, as MD can be thought of as a one-dimensional latent space representation of a visual field[33]. Therefore, it follows that the performance of the VAE would improve upon MD. The performance of both the linear and non-linear VAE method appeared to peak when the VAE was trained with ten latent dimensions. This indicates that choosing a VAE with ten dimensions is optimal when modeling visual fields. These results are promising, because the VAE significantly outperformed MD in the early years from follow-up and in the long term, making the results clinically meaningful[34].

Furthermore, the VAE can be used to predict the future location of a patient's latent features (Figs. 5, 6 and 7, and Supplementary Tables S2–7). This allows for accurate predictions of future visual fields, which can help illuminate patterns and severity of progression. Such predictions of whole visual fields are obviously not possible with a single global metric such as MD. Even if one uses PW regression, the VAE was shown to have superior prediction capabilities, when using a linear method. The non-linear technique was not shown to be effective for prediction purposes. Since the progression detection rates were comparable across linear and non-linear VAE methods and the non-linear predictions were poor, this motivates using the linear VAE in clinical practice. The predictions using Humphrey Field Analyzer printouts in Fig. 7 demonstrated the benefits of using the VAE as compared to PW regression. In particular, while the PW method was highly susceptible to local variability, the VAE produced stable (i.e., smooth) predictions. This resulted in predictions that were robust to the variability of individual entries and consequently, the true disease pattern visible is better illuminated.

The prediction capability of the VAE model, including patients with patterns in their visual fields, is reassuring, as global methods, like MD, typically fail at detecting localized defects[34]. The flexibility of the VAE, to not only improve rates of detecting progression over the global MD method, but also maintain superior predictions over PW regression, makes it clinically useful. In particular, because the VAE has the ability to detect progression at higher rates in the early years from baseline, clinicians will need fewer visits to obtain accurate detection of progression; thus, limiting the burden on patients.

A potential limitation, or motivation for future work, is that when training the deep learning model, we treated each of the visual fields as independent images. We overcame this limitation by modeling a patient's longitudinal visual fields series in latent space, which we showed to be an effective method for assessing rates of progression. In the future, we could make improvements by learning a generative model for longitudinal series of visual fields,

instead of a singular image. This would be a powerful technique for clustering patients, not by their visual fields, but by the characteristics of their progression.

Another extension to the method presented in this manuscript is to generate synthetic glaucomatous visual fields to be used as a benchmark dataset to validate new methods for glaucoma research. Existing methods attempt to generate longitudinal visual fields by modelling PW relationships across time with potentially non-linear fits[35,36], however they typically ignore spatial correlations in the visual field or can only predict stable fields. The decoder from the VAE can be used to simulate glaucomatous visual field series that accounts for spatial dependencies and the highly-nonlinear PW trends.

In conclusion, this manuscript showed the potential use of the VAE latent space for assessing rates and trajectories of glaucoma progression. The rates of progression can be considered a multi-dimensional extension of MD with improved abilities to detect progression and the additional benefit of a generative technique to predict future patterns and severity of visual fields.

Data availability
The data from this study are not available for public release at this time.

References
1. Weinreb, R. N., Aung, T. & Medeiros, F. A. The pathophysiology and treatment of glaucoma: a review. *JAMA* **311**, 1901–1911, https://doi.org/10.1001/jama.2014.3192 (2014).
2. Vianna, J. R. & Chauhan, B. C. How to detect progression in glaucoma. *Prog Brain Res* **221**, 135–158, https://doi.org/10.1016/bs.pbr.2015.04.011 (2015).
3. Chauhan, B. C. *et al.* Practical recommendations for measuring rates of visual field change in glaucoma. *Br J Ophthalmol* **92**, 569–573, https://doi.org/10.1136/bjo.2007.135012 (2008).
4. Gardiner, S. K., Mansberger, S. L. & Demirel, S. Detection of Functional Change Using Cluster Trend Analysis in Glaucoma. *Invest Ophthalmol Vis Sci* **58**, BIO180–BIO190, https://doi.org/10.1167/iovs.17-21562 (2017).
5. Goldbaum, M. H. *et al.* Progression of patterns (POP): a machine classifier algorithm to identify glaucoma progression in visual fields. *Invest Ophthalmol Vis Sci* **53**, 6557–6567, https://doi.org/10.1167/iovs.11-8363 (2012).
6. Goldbaum, M. H. *et al.* Using unsupervised learning with independent component analysis to identify patterns of glaucomatous visual field defects. *Invest Ophthalmol Vis Sci* **46**, 3676–3683, https://doi.org/10.1167/iovs.04-1167 (2005).
7. Sample, P. A. *et al.* Unsupervised machine learning with independent component analysis to identify areas of progression in glaucomatous visual fields. *Invest Ophthalmol Vis Sci* **46**, 3684–3692, https://doi.org/10.1167/iovs.04-1168 (2005).
8. Yousefi, S. *et al.* Learning from data: recognizing glaucomatous defect patterns and detecting progression from visual field measurements. *IEEE Trans Biomed Eng* **61**, 2112–2124, https://doi.org/10.1109/TBME.2014.2314714 (2014).
9. Abe, R. Y. *et al.* The Impact of Location of Progressive Visual Field Loss on Longitudinal Changes in Quality of Life of Patients with Glaucoma. *Ophthalmology* **123**, 552–557, https://doi.org/10.1016/j.ophtha.2015.10.046 (2016).
10. Fitzke, F. W., Hitchings, R. A., Poinoosawmy, D., McNaught, A. I. & Crabb, D. P. Analysis of visual field progression in glaucoma. *British Journal of Ophthalmology* **80**, 40–48, https://doi.org/10.1136/bjo.80.1.40 (1996).
11. Diaz-Pinto, A. *et al.* Retinal Image Synthesis and Semi-supervised Learning for Glaucoma Assessment. *IEEE Trans Med Imaging.* https://doi.org/10.1109/TMI.2019.2903434 (2019).
12. Medeiros, F. A., Jammal, A. A. & Thompson, A. C. From Machine to Machine: An OCT-Trained Deep Learning Algorithm for Objective Quantification of Glaucomatous Damage in Fundus Photographs. *Ophthalmology*, https://doi.org/10.1016/j.ophtha.2018.12.033 (2018).
13. Li, Z. *et al.* Efficacy of a Deep Learning System for Detecting Glaucomatous Optic Neuropathy Based on Color Fundus Photographs. *Ophthalmology* **125**, 1199–1206, https://doi.org/10.1016/j.ophtha.2018.01.023 (2018).
14. Yousefi, S. *et al.* Detection of Longitudinal Visual Field Progression in Glaucoma Using Machine Learning. *Am J Ophthalmol* **193**, 71–79, https://doi.org/10.1016/j.ajo.2018.06.007 (2018).
15. Wen, J. C. *et al.* Forecasting Future Humphrey Visual Fields Using Deep Learning. *arXiv preprint arXiv:1804*.04543, doi: arXiv:1804.04543 (2018).
16. Goodfellow, I. *et al.* Generative adversarial nets. *In Advances in neural information processing systems* (2014).
17. Kingma, D. P. & Welling, M. Auto-encoding variational Bayes. *arXiv preprint arXiv:1312*.6114, doi:arXiv:1312.6114 (2013).
18. Way, G. P. & Greene, C. S. Extracting a biologically relevant latent space from cancer transcriptomes with variational autoencoders. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing* **23**, 80–91 (2018).
19. Choi, H., Kang, H. & Lee, D. S., Alzheimer's Disease Neuroimaging, I. Predicting Aging of Brain Metabolic Topography Using Variational Autoencoder. *Front Aging Neurosci* **10**, 212, https://doi.org/10.3389/fnagi.2018.00212 (2018).
20. Doersch, C. Tutorial on variational autoencoders. *arXiv preprint arXiv:1606*.05908, doi:arXiv:1606.05908 (2016).
21. Blei, D. M., Kucukelbir, A. & McAuliffe, J. D. Variational Inference: A Review for Statisticians. *Journal of the American Statistical Association* **112**, 859–877, https://doi.org/10.1080/01621459.2017.1285773 (2017).
22. Kullback, S. & Leibler, R. A. On Information and Sufficiency. *The Annals of Mathematical Statistics* **22**, 79–86, https://doi.org/10.1214/aoms/1177729694 (1951).
23. Zhao, S., Song, J. & Ermon, S. InfoVAE: Information maximizing variational autoencoders. *arXiv preprint arXiv:1706*.02262, doi:arXiv:1706.02262 (2017).
24. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412*.6980 (2014).
25. Chollet, F. & others. *Keras*, https://github.com/fchollet/keras (2015).
26. Abadi, M. *et al.* TensorFlow: Large-scale machine learning on heterogeneous systems. *Software available from tensorflow.org* (2015).
27. Chollet, F., Allaire, J. & others. *R Interface to Keras* (2017).
28. Garway-Heath, D. F., Poinoosawmy, D., Fitzke, F. W. & Hitchings, R. A. Mapping the visual field to the optic disc in normal tension glaucoma eyes11The authors have no proprietary interest in the development or marketing of any product or instrument mentioned in this article. *Ophthalmology* **107**, 1809–1815, https://doi.org/10.1016/s0161-6420(00)00284-0 (2000).
29. Efron, B. & Tibshirani, R. J. *An Introduction to the Bootstrap.* (CRC press, 1994).
30. Wetzel, S. J. Unsupervised learning of phase transitions: From principal component analysis to variational autoencoders. *Phys Rev E* **96**, 022140, https://doi.org/10.1103/PhysRevE.96.022140 (2017).
31. Asaoka, R., Murata, H., Iwase, A. & Araie, M. Detecting Preperimetric Glaucoma with Standard Automated Perimetry Using a Deep Learning Classifier. *Ophthalmology* **123**, 1974–1980, https://doi.org/10.1016/j.ophtha.2016.05.029 (2016).
32. Kim, S. J., Cho, K. J. & Oh, S. Development of machine learning models for diagnosis of glaucoma. *PLoS One* **12**, e0177726, https://doi.org/10.1371/journal.pone.0177726 (2017).

33. Leske, M. C. *et al*. Predictors of long-term progression in the early manifest glaucoma trial. *Ophthalmology* **114**, 1965–1972, https://doi.org/10.1016/j.ophtha.2007.03.016 (2007).
34. Bengtsson, B. & Heijl, A. A visual field index for calculation of glaucoma rate of progression. *Am J Ophthalmol* **145**, 343–353, https://doi.org/10.1016/j.ajo.2007.09.038 (2008).
35. Russell, R. A., Garway-Heath, D. F. & Crabb, D. P. New insights into measurement variability in glaucomatous visual fields from computer modelling. *PLoS One* **8**, e83595, https://doi.org/10.1371/journal.pone.0083595 (2013).
36. Wu, Z. & Medeiros, F. A. Development of a Visual Field Simulation Model of Longitudinal Point-Wise Sensitivity Changes From a Clinical Glaucoma Cohort. *Transl Vis Sci Technol* **7**, 22, https://doi.org/10.1167/tvst.7.3.22 (2018).

Acknowledgements

Author contributions

S.I.B., S.M. and F.A.M. developed the primary questions, analytic and overall plan for this study. S.I.B. analyzed the data, contributed new analytic tools, and wrote the first draft of the paper. S.M. and F.A.M. provided feedback and key revisions to the initial draft. All authors read and approved the final manuscript.

Competing interests

Dr. Medeiros has the following declarations: Alcon Laboratories (C, F, R), Allergan (C, F), Bausch&Lomb (F), Carl Zeiss Meditec (C, F, R), Heidelberg Engineering (F), Merck (F), nGoggle Inc. (F), Sensimed (C), Topcon (C), Reichert (C, R). Drs. Berchuck and Mukherjee declare no potential conflict of interest.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-019-54653-6.

Correspondence and requests for materials should be addressed to F.A.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.