scientific reports

Check for updates

OPEN

Application of machine learning techniques in GlaucomAI system for glaucoma diagnosis and collaborative research support

Hubert Świerczyński[1,2]✉, Juliusz Pukacki[1], Szymon Szczęsny[2], Cezary Mazurek[1] & Robert Wasilewicz[3]

This paper proposes an architecture of the system that provides support for collaborative research focused on analysis of data acquired using Triggerfish contact lens sensor and devices for continuous monitoring of cardiovascular system properties. The system enables application of machine learning (ML) models for glaucoma diagnosis without direct intraocular pressure measurement and independently of complex imaging techniques used in clinical practice. We describe development of ML models based on sensor data and measurements of corneal biomechanical properties. Application scenarios involve collection, sharing and analysis of multi-sensor data. We give a view of issues concerning interpretability and evaluation of ML model predictions. We also refer to the problems related to personalized medicine and transdisciplinary research. The system can be a base for community-wide initiative including ophthalmologists, data scientists and machine learning experts that has the potential to leverage data acquired by the devices to understand glaucoma risk factors and the processes related to progression of the disease.

Keywords Machine learning, Personalized medicine, Software architecture, Glaucoma, Sensors

Glaucoma is a worldwide vision threatening disease that is a significant public health challenge. Substantial efforts have been made to improve non-invasive diagnostic methods based on retinal fundus images or optical coherence tomography (OCT)[1]. Advancements in deep learning algorithms and transfer learning enable objective quantification of the optic nerve head changes[2,3]. Convolutional neural networks can perform optic disc, cup and retinal vessel segmentation[4]. Alternative diagnostic options have become available in the recent years[5]. Eye and cardiac sensors can continuously record data during 24-hour session. Acquired sensor data is processed to build patient's diagnostic profile that provides insights independent of imaging techniques commonly used in ophthalmology[6].

Triggerfish (Sensimed) device is based on a contact lens sensor with embedded strain gauge[7]. It measures ocular volume changes during a whole day. Series of measurements in the units of millivolt equivalents is recorded every 5 minutes (each series called burst has 300 values). Triggerfish measurements are related to the changes of intraocular pressure (IOP). Properties of such relation were investigated in many studies[8,9]. IOP is one of the most important factors in the diagnosis and management of glaucoma[10]. Triggerfish contact lens sensor can also record low-amplitude ocular pulse[11]related to the heart rate with good accuracy in a majority of eyes. Cardiovascular system properties have influence on ocular blood flow[12]. Heart rate and arterial blood pressure monitoring data associated with Triggerfish measurements can be utilized to more accurately detect glaucoma. At the same time, predictive models involving wide range of sensor data based attributes supplemented with measurements of corneal biomechanical properties[13] have better performance metrics. Our approach involving Triggerfish sensor record that is aligned to data acquired using cardiovascular system monitoring device considers functional properties of the eye that can possibly lead to accurate assessment of conditions in early glaucoma stages. Identification of specific interactions of cardiovascular system and eye function during the diagnostic session divided according to the physiological circadian cycle properties will potentially allow more precise control of the disease.

[1]Poznan Supercomputing and Networking Center, Poznań, Poland. [2]Faculty of Computing and Telecommunications, Poznan University of Technology, Poznań, Poland. [3][Wasilewicz] Eye Clinic, Poznań, Poland. ✉email: hubert@man.poznan.pl

Triggerfish is a relatively new device and there are no extensive software tools for supporting complete analysis of its data output. Current cost of a single examination (disposable contact lens) is relatively high, can be many times higher than OCT scan (sample OCT imaging output is shown in Figure 1) and there are no specific multi-sensor data based clinical protocols ready to apply in glaucoma detection and treatment. The majority of ML-based systems for glaucoma diagnosis is intended for structural analysis of image data[14] and the range of implemented functions is usually limited. System designed for management and analysis of the sensor data can address many of the aforementioned issues in the context of glaucoma diagnosis. Deployment of the system will enable collection of accessible data that is typically stored in isolated/closed repositories. It will encourage research on the personalized approaches focused on the multi-sensor data and the exchange of new ideas.

ML models that are considered in this study involve sensor data acquired during 24-hour monitoring session. Prediction for a particular case can be explained in terms closely related to the basic eye or cardiac data properties in the selected time intervals (such as approximation of diastolic arterial pressure amplitude in the night sleep interval, etc.). Medical professionals can assess the estimated contribution of each model's attribute to the prediction (using the explanation generated in the system).

This paper proposes a comprehensive system for glaucoma diagnosis and collaborative research support. The system has the following basic aims:

1. Support of glaucoma detection and control using multiple ML techniques applied for the multi-sensor sensor data instead of common methods based on OCT imaging.
2. Provision of collaborative research platform for medical specialists in ophthalmology and data scientists that supports exploratory data analysis using multiple data visualization techniques such as box plots and heat maps.
3. Supply of data management services (focused mainly on time series acquired using Triggerfish contact lens sensor and devices for continuous monitoring of cardiovascular system properties) that facilitate clinical data collection.

Users of the system can be assigned with the following basic roles:

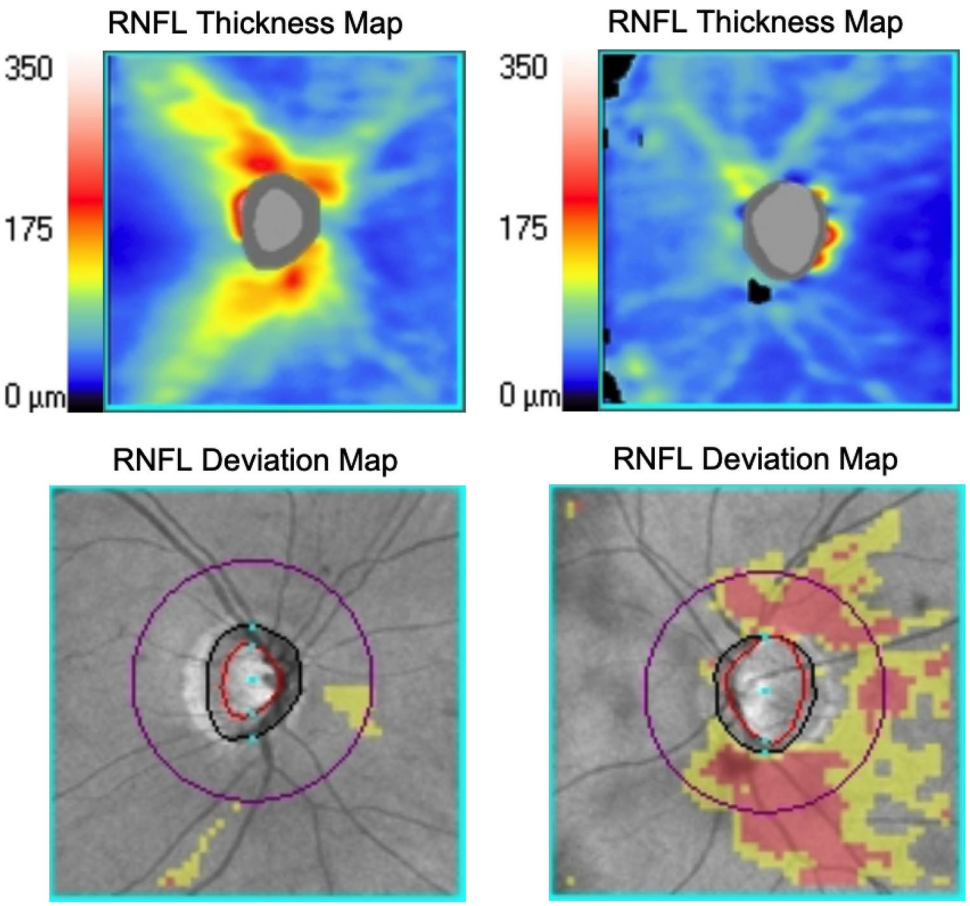

Fig. 1. Sample OCT imaging output providing topographic thickness information of the retinal nerve fiber layer (RNFL). Healthy eye images on the left. Images of the eye with glaucomatous optic neuropathy on the right (retinal nerve fiber layer asymmetry and decreased average thickness). Image courtesy of Robert Wasilewicz.

1. Ophthalmologists experienced in glaucoma diagnosis and clinical research. They can set new directions for the research and assess the practical value of the implemented solutions.
2. Medical doctors pursuing advanced training (e.g. clinical residency training program) in ophthalmology without prior experience in glaucoma detection and treatment. They can participate in collection and validation of clinical data as well as in research scenarios.
3. ML engineers or data scientists experienced in application of ML methods in biomedical data analysis. These specialists can collaborate with the doctors to design and develop new analytic components for the users.

General outline of the proposed approach is shown in Figure 2. The components of the diagram refer to data acquisition (Triggerfish and SOMNOtouch devices, clinical data), data management (storage, data collection, sharing), application and evaluation of ML models, data visualization, medical doctors and data science experts collaboration in clinical research related to glaucoma diagnosis and treatment.

The key contributions of the proposed approach for glaucoma diagnosis can be summarized as follows:

1. Machine learning approach involving Triggerfish CLS measurements and cardiac data considers functional properties of the eye that can lead to accurate assessment of conditions in early glaucoma stages. As a comparison, common imaging methods (including OCT) detect structural (morphological) changes which are typically related to progression of glaucoma.
2. Predictive model involving basic measurements of corneal biomechanical properties (such as corneal hysteresis) and sensor data based attributes provides AUC of 0.88. It can support glaucoma detection without direct IOP measurement.
3. The relationship between Triggerfish CLS measurements and cardiac sensor data in time intervals is quantified as a correlation coefficient. This representation is simple, easy to understand and use in analytical scenarios.

The paper is organized as follows. Section 2 presents predictive models based on the sensor data that have been provided for users of the system. Section 3 contains the overview of architecture and services of the system. Section 4 describes main application scenarios. In section 5 we discuss issues related to deployment of the

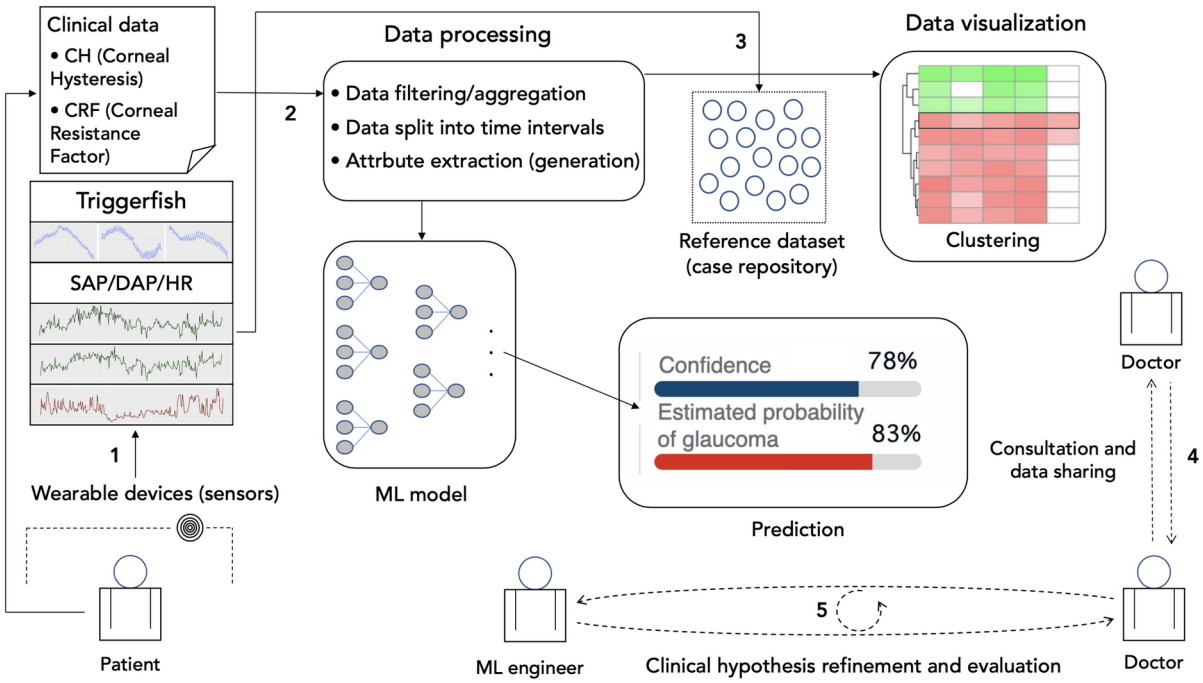

Fig. 2. Overview of data flow and interactions of users of the system. 1. Data acquired using Triggerfish CLS and devices for continuous monitoring of cardiovascular system parameters as well as clinical measurements (CH, CRF etc.) are entered into the system. 2. Raw data is checked and processed in order to generate attributes/features suitable as input for the predictive ML models and other analytical components of the system. 3. Selected cases can be included in the reference dataset after evaluation performed by an expert ophthalmologist (regarding diagnosis, annotation and completeness). Reference repository is used in development of ML models or statistical research. 4. Doctors can share their own cases to assess patient's examination results using data visualization functions, consult a diagnosis or treatment plan. 5. Collaborative development and evaluation of clinical hypotheses can be performed by doctors and data scientists. Clinical hypotheses may concern the properties of specific case groups, considering data characteristics such as the reciprocal relation of TF and cardiac signal. This approach can be seen as exploratory data analysis performed in iterative steps.

system and transdisciplinary research. Section 6 contains a summary of the paper and possible directions for development of the system.

Application of machine learning to diagnosis of glaucoma

Here we provide a summary of the ML techniques we use in glaucoma detection process. In the earlier paper[15] we described the details of methods we applied in development and evaluation of the predictive models. Such ML models based on sensor data are deployed in the system and are used to support the diagnosis of new cases.

Input data

Input data for our research was collected at [Wasilewicz] Eye Clinic in Poznań. Input dataset contains 138 cases (87 females and 51 males). This dataset was used to build predictive ML models described in this paper. Reference diagnosis labels of the cases were assigned by the experienced ophthalmologist. Additional verification of the diagnosis assignment was performed when data collection was completed. Figure 3 shows distribution of the basic clinical data for NORM (healthy) and POAG (glaucomatous neuropathy) cases. The following labels were assigned to the cases according to the diagnosis:

- 50 high tension glaucoma (POAG/HTG)
- 30 normal tension glaucoma (POAG/NTG)
- 58 control/healthy (NORM)

Primary open-angle glaucoma (POAG) is the most prevalent type of glaucoma that is commonly classified into NTG and HTG. IOP value is within the normal range for the population (IOP $\leq$21 mm Hg) in NTG, whereas elevated IOP is the main feature of HTG. Prevalence of glaucoma depends on many factors (e.g. age, gender). It was investigated in many studies for different populations[16,17]. Minority class is composed of common glaucoma types that we labelled as glaucomatous neuropathy. Standard random undersampling was applied during data collection process to obtain close case count of the majority and minority class. Randomly selected cases form the majority class (NORM) were skipped. The resulting distribution is more balanced and the received data seems suitable for application of the selected binary classification algorithms. We performed additional assessment of the results using SMOTE (Synthetic Minority Over-sampling Technique[18]) for the input dataset to oversample NORM class. AUC and accuracy of the models for synthetic data are similar to the cross-validation results for the original input data with a slightly greater variance.

Goldmann applanation tonometer[19]was used to measure initial IOP before application of Triggerfish contact lens sensor. 24-hour Triggerfish record is available for the each case. SOMNOtouch NIBP (Somnomedics noninvasive continuous blood pressure monitor) device[20] was used to record systolic/diastolic arterial pressure (SAP/DAP) and heart rate (HR) during 24-hour period (see Figure 5). Internal clock of Triggerfish and SOMNOtouch NIBP device were synchronized. In addition, corneal hysteresis (CH) and corneal resistance

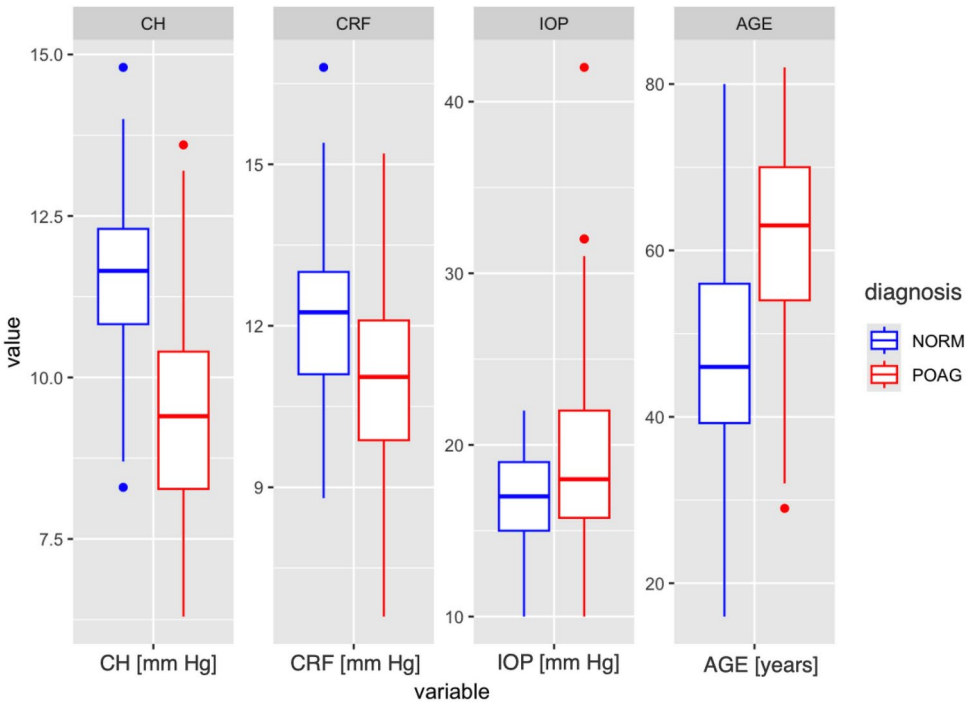

Fig. 3. Summary of the basic clinical data included in the input dataset. Box plots for NORM (healthy) and POAG (glaucomatous neuropathy) cases.

factor (CRF) were measured with Ocular Response Analyzer (Reichert) to quantify basic biomechanical properties of the cornea[21].

Data processing and attribute generation

At the beginning, we calculated separate median value for the each raw signal in time intervals overlapping Triggerfish series (bursts). Such low-dimensional representation captures the underlying characteristics of the high-dimensional input data[22]. Raw Triggerfish data can contain high, sharp peaks that are related to the eye blinks. As we consider only median approximation of Triggerfish burst values, we don't use any peak detection (or noise reduction) methods. Figure 4 shows median approximation of Triggerfish measurements which is performed before attribute generation.

We divided the 24-hour session into the consecutive time intervals on the basis of the physiological circadian cycle properties and the protocol proposed by Robert Wasilewicz, MD, PhD[23]. Main division of the 24-hour session contains the following base time points:

- start of the recording session (START)
- begin of the main/night sleep period (SLEEP)
- end of the main/night sleep period (WAKE)
- end of the recording session (END)

Main time intervals used in the generation of the attributes/features for ML algorithms are enumerated in Table 1.

Missing values in time intervals were imputed using constant (stepwise) interpolation for the available adjacent values.

We defined the following attributes that give the basic properties of Triggerfish (TF), cardiac sensor signal or their relation in specified time interval (prefix of the relevant attribute name is given in brackets):

1. (sum) Sum of TF values in the interval (numerical approximation of definite integral of TF over the time interval). We assumed constant TF value between consecutive bursts.
2. (slope) Slope (in radians) of linear regression line for TF fitted using ordinary least squares method in the interval.
3. (sec_deriv_integral) Sum of the numerical approximation of TF second derivative in the interval.
4. (ampl) Amplitude of the signal, i.e. difference of the maximal value and the minimal value in the interval. Additionally, we computed the modified amplitude of the signal as 95 % quantile minus the minimal value (flat_ampl). This is an estimate of total change of TF variability rate in the interval.
5. (cor) Correlation coefficient of TF and cardiac signal (SAP, DAP or HR) in the interval. We also divided correlation value range at points {−0.65,−0.25,0.25,0.65} and mapped such intervals onto five integer levels

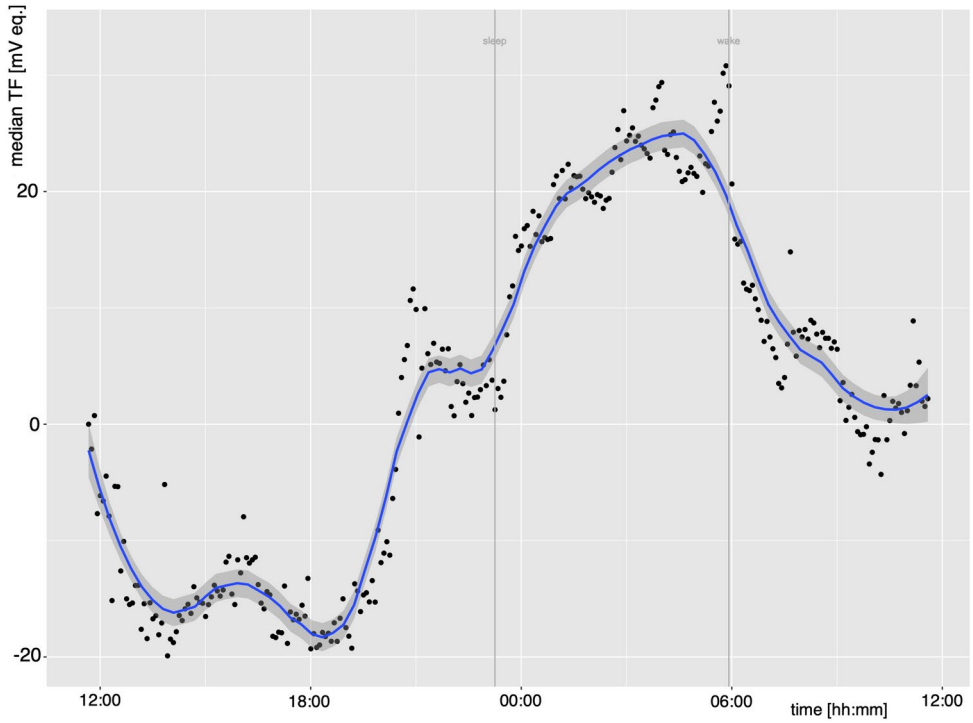

Fig. 4. 24-hour session Triggerfish CLS output for a normal (healthy) patient. Each black point represents median of one TF series (burst). Two grey vertical lines mark the beginning and the end of sleep period. Smoothed blue line is the loess (locally weighted polynomial regression) approximation of TF.

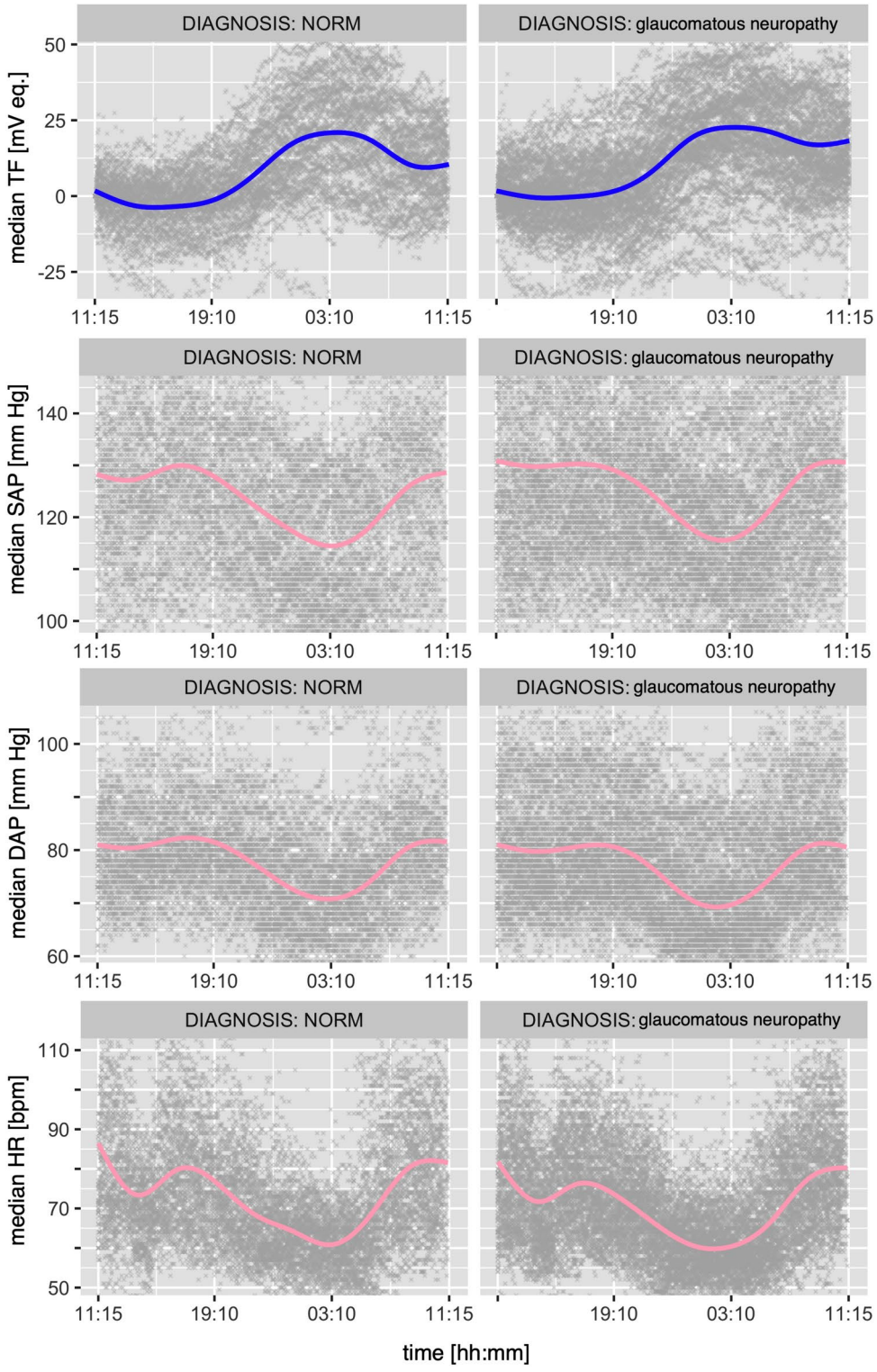

Fig. 5. 24-hour overview of attributes in the input dataset for healthy (NORM) and POAG cases (glaucomatous neuropathy). Each point represents median of the values within one series (one burst). Scatterplots contain smoothed color lines which are attribute approximation generated by loess function (locally weighted polynomial regression) for TF [mV equivalents], SAP/DAP [mm Hg], HR [beats per minute].

{−2,−1,0,1,2} (cor_level). We calculated Spearman's rank correlation coefficient (usually denoted as ρ) to describe monotonic relation of the signals. The above mapping can be seen as the assignment of a relative magnitude level for a given ρ value (i.e. strong negative, medium negative, none, medium positive, strong

Interval name	Time range
START_[SLEEP-5h]	START until SLEEP−5h
[SLEEP-5h]_SLEEP	SLEEP−5h until SLEEP
SLEEP_[SLEEP+2h]	SLEEP until SLEEP+2h
[SLEEP+2h]_WAKE	SLEEP+2h until WAKE
WAKE_[WAKE+2h]	WAKE until WAKE+2h
[WAKE+2h]_END	WAKE+2h until END

Table 1. Range of the main time intervals defined for 24-hour session.

positive). We also tested if the correlation is statistically significant (i.e. ρ is significantly different from zero) using p-value returned by the corr.test function from psych library in R.

We considered raw TF values and TF scaled linearly into [0,20] range separately for each case (TF^s) in 1. (sum) and 2. (slope). We used this range to facilitate comparison of TF records when relative values are more important than exact/raw measurements.

Algorithms

We investigate binary classification problem for the healthy cases labeled as NORM and the other cases in our dataset labeled as glaucomatous neuropathy (diagnosed as POAG/HTG or POAG/NTG). Many ML algorithms can be used to solve this problem[24,25]. We report results for logistic regression and XGBoost that are commonly applied in biomedical data analysis. Logistic regression is a fast algorithm and its results are quite easy to interpret. This classification algorithm determines linear decision boundaries. Logistic regression is based on properties of sigmoidal curve used for modelling the probability of belonging to a particular class for a given case. XGBoost (Extreme Gradient Boosting) is an ensemble learning method based on decision trees. In this algorithm, many shallow trees are built sequentially to improve predictive performance of the final model. We make use of H2O framework for development and deployment of the ML models. It is open source, distributed and scalable predictive analytics environment[26]. H2O Flow web interface was used for model evaluation and visualization of the results.

Choosing a subset of attributes to achieve high prediction performance of the models is associated with the significant computational cost as a consequence of the large number of possible subsets to check. We used LASSO regularization (Least Absolute Shrinkage and Selection Operator)[24]and RSM (Random Subspace Method)[27] for initial attribute selection. We applied LASSO regularization for generalized linear models (GLM) which is implemented in H2O library. RSM is based on fitting linear models on small randomly chosen subsets of attributes in order to create ranking list of attributes. Final subset of attributes is selected using information criteria or validation set. We applied parallel version of RSM implemented in regRSM library.

Evaluation of the models

Cross-validation (CV) resampling procedure was applied to estimate prediction performance of the models. Input dataset is randomly split into k equally sized subsets in standard k-fold CV. In each of the k steps we subsequently use one subset as validation/test data and the remaining subsets as training data. Prediction performance/quality (according to the selected metrics) is calculated for test data in the each step and average of the results is the final estimation. We have set CV fold size equal to 10 considering size of the input dataset. In addition, we repeated CV procedure 100 times to assess variance of the estimation for the each model.

There are many different metrics for evaluation of quality of binary classification results. We calculated the following metrics:

1. Area Under the ROC Curve (AUC): estimate based on properties of the plot for value pairs of TPR (true positive rate i.e. probability of positive prediction for truly positive case) and FPR (false positive rate i.e. probability of positive prediction for truly negative case) calculated for a range of different classification threshold values[28].
2. Brier score (mean squared loss): given by the formula $\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$, where N is the total cases count, y_i is the actual value (assigned to the reference class), $\hat{y}_i$ is the predicted value (probability) for the case
3. Accuracy: ratio of correctly classified cases to the total cases count.

Predictive models

In this section we present models deployed in our system that are used in glaucoma detection scenario. These models do not include IOP value that can be registered (e.g. by Goldmann applanation tonometer) no more than few times a day. Instead we consider Triggerfish and cardiac sensor data related attributes for 24-hour monitoring session.

In Table 2 we enumerated attributes for the models with highest AUC value estimated in repeated CV routine for the input dataset. Model $G_0$is based only on TF and cardiac data derived attributes. Additional cardiac attributes joined with TF data can improve predictive performance of model. Quantification of the reciprocal relation of TF and cardiac signal (e.g. correlation coefficient) seems particularly valuable[15]. Model $G_1$is based on sensor data supplemented with measurements of corneal biomechanical properties (CH, CRF)[29]. Such models can be a tool suitable for glaucoma detection regardless of direct IOP measurements. Sensor data

Model id	Attributes
B	IOP
G_0	slope_TF^s_[sleep-5h]_sleep, sec_deriv_integral_TF^s_sleep_[sleep+3h], sum_TF^s_[wake-4h]_wake, sum_TF^s_wake_[wake+5h], ampl_HR_start_sleep, flat_ampl_DAP_sleep_wake, cor_level_HR_[sleep+2h]_wake
G_1	CH, CRF, slope_TF^s_[sleep-5h]_sleep, sec_deriv_integral_TF_sleep_[sleep+3h], sum_TF^s_wake_[wake+5h], ampl_HR_start_sleep, flat_ampl_DAP_sleep_wake, cor_level_HR_[sleep+2h]_wake

Table 2. Summary of the attributes selected for the model G_0 involving only sensor data derived attributes and G_1 with corneal biomechanical measurements (CH and CRF). Baseline model B is based on the initial IOP measurement only.

Model id	Type	AUC	Brier score	Accuracy
B	logistic regression	0.60±0.01	0.24±0.01	0.59±0.01
G_0	logistic regression	0.77±0.01	0.20±0.01	0.73±0.01
	XGBoost	0.70±0.02	0.22±0.01	0.70±0.01
	naive Bayes	0.68±0.02	0.23±0.01	0.68±0.01
G_1	logistic regression	0.88±0.01	0.14±0.01	0.83±0.01
	XGBoost	0.85±0.01	0.15±0.01	0.81±0.01
	naive Bayes	0.84±0.01	0.16±0.01	0.82±0.01

Table 3. Estimation of performance metrics for the models from Table 2.

Metric name	Value for G_1
F1	0.86
F2	0.87
F0.5	0.84
precision	0.83
sensitivity (recall)	0.89
specificity	0.74
NPV (negative predictive value)	0.83

Table 4. Summary of model G_1 (logistic regression) performance metrics.

derived attributes are complementary to CH and CRF with the highest mean AUC of 0.88±0.01 for G_1. Best performance metrics estimated for the logistic regression model (see Table 3) are similar to the results reported in our previous paper[15]. Summary of G_1 metrics depending on classification threshold is shown in Table 4 (results for the max F1 threshold). We performed basic hyperparameter tuning using grid search. The following optimal values were determined for G_1 XGBoost model:

- number of trees (n_trees): 50
- learning rate which specifies shrinkage of the feature weights after each boosting step (eta): 0.21
- maximum depth to which each tree will be built (max_depth): 1

Important aspect of this approach is introduction of cardiac data derived attributes that can be seen as prospective modifiable risk factors of glaucoma.

System overview

ML-based system architecture

Increasing adoption of ML techniques has given rise to many challenges related to system development, deployment and management[30]. These challenges should be addressed along with the issues typical for standard software system engineering. ML-based system view is divided into ML subsystem and software subsystem in some high-level approaches[31]. Such view on architecture is in line with different characteristics, functions and key stakeholders of the each subsystem. Distinct nature of the each subsystem has an impact on requirements analysis and design assumptions. Development team can use different methodology and organizational principles regarding ML field with specific roles like data scientist, data engineer and domain expert. Typical concerns of the ML field include quality of data, hyperparameter tuning for algorithms, model performance assessment, visualization and explanation of the results. On the other hand, standard software engineering deals with the concerns like security, availability, testing, system maintenance and update. New important role in development team may be assigned to an expert that has experience in both software engineering and application of ML. Main

coordinator can more efficiently spot and manage design issues or trade-offs that arise due to the complexity of ML-based system.

Microservices architecture is getting common over the last years. This architectural paradigm assumes that a system is composed of many loosely coupled components (or services) which are independently deployable[32]. Different technology stack, programming languages and data sources can be used for different components. Such design capabilities seem to be especially valuable in development of complex ML-based systems. Scalability and update flexibility of this approach outweigh the potential performance gains of monolithic architecture[33].

Data integration

Early diagnosis of glaucoma is a challenging task. Diagnostic routine includes diverse examinations and the resulting data require appropriate interpretation. The use of wearable medical devices is constantly growing in many fields of health care. Continuous monitoring of the physiological signals can provide data essential for the development of reliable diagnostic methods and management standards for the disease.

One of the important aspects of data integration is combining data from different sources and delivering unified view of them for the users of the system[34]. We introduced notion of the case or, equivalently, set of examinations in the system to integrate series of patient examinations in time slot assigned for diagnosis. We focus mainly on the analysis of sensor data therefore typical case involves Triggerfish and SOMNOtouch data from one 24-hour session. Measurements like IOP, CH, CRF, CCT (central corneal thickness) and the other anatomical readings of the eye features are also included. Any relevant examination data could optionally be included in the case, e.g. OCT, fundus images or selected optometric test results. In particular, data that may be useful include numerical features determined for OCT tomograms such as disc area (DA) or cup to disc ratio (CDR). As the system was designed primarily for the sensor data processing, we consider any imaging data to be supplementary data only. Integration based on the notion of the case enables sharing of the sensor data and creating research datasets available for the selected users. Statistical analysis of such large datasets can lead to identification of patient subgroups that have specific characteristic related to diagnosis or management of a disease.

Personal data (such as name or personal ID number) of a case can be accessed only by the case owner. In any scenarios of data analysis and sharing, only anonymized content is available in the system.

Application services

The system is comprised of backend components, data storage and web application interface (see Figure 6). Core backend components are implemented in Java using Play Framework for Java. Frontend web interface is built using JavaScript libraries, including Plotly Graphing Library for making interactive plots of multiple data formats. Currently we use PostgreSQL as a relational database management system.

The scripting component is responsible for running Python and R code for basic processing of sensor data and application of the selected ML techniques. R language offers many libraries for data processing and the recent ML algorithms. Reference implementation of R (i.e. GNU R) works well in practice, but it is quite complex. Several attempts have been made to create an implementation of the R language that provides better performance while keeping compatibility with the original version of R[35]. Java environment is the basis for implementation of the main system, therefore we examined JVM-based interpreters for R (e.g. Renjin). Such approach could contribute to the closer integration of data processing components with the core system. Nevertheless, it is challenging to map semantics of scripting language like R that is significantly differ from the JVM bytecode. At this moment JVM-based interpreter for R requires additional modification/adjustment of the scripts and imposes constraints

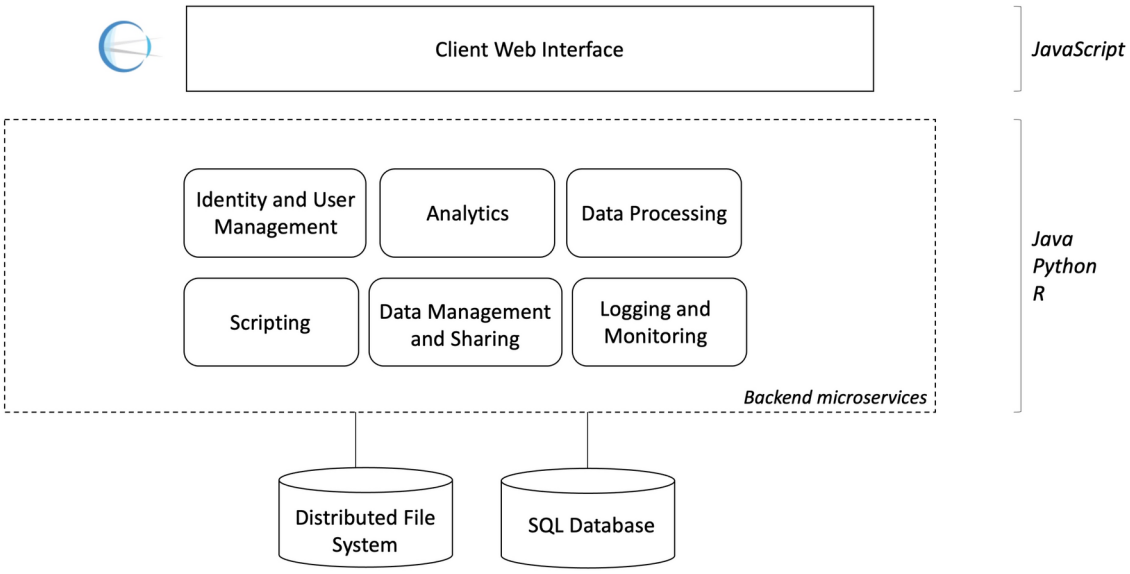

Fig. 6. Overview of the basic system components.

for specific R library versions. We aimed at running generic code in the scripting component with possibility of enhancement and fast deployment. Finally we implemented a solution based on Java ProcessBuilder class to execute the scripts as operating system processes. This approach enables easy script update and use of multiple libraries.

Application scenarios

In this section we describe application scenarios created on the basis of requirements elicited using knowledge of the domain experts and system engineers. Figure 2 shows doctors and ML engineers making use of the system in the scenarios involving diagnostic decision support and collaborative research.

Glaucoma diagnosis

Diagnosis is inherently individual assessment of a patient based on available data and clinical experience of medical doctor[36]. In the scenario focused on clinical decision support we consider supplementary information generated by ML model that can be used in evaluation of the most relevant diagnostic hypotheses.

Diagnostic support scenario is intended for the medical professionals. They can check basic output of ML model which is predicted probability of glaucomatous neuropathy (positive) diagnosis for a case. The result is shown as numeric probability attached to the single bar chart. Classification label is also shown. It is based on a comparison with optimal classification threshold determined for the model. Diagram with explanation of the model prediction is shown to allow interpretation of the result by a user (see Figure 7). Understanding and comparing how a model uses the attributes to make a given prediction can provide opportunity to get insight of its properties[37]. Currently we use LIME (Local Interpretable Model-agnostic Explanations) and DALEX visualization techniques for generation of explanation of particular prediction. LIME method assumes that every complex model is linear on a local scale and can be approximated with an interpretable model[38]. DALEX break-down plots are fast approximations of Shapley values[39]. Such local explanations enable evaluation of the practical usability of the model by ophthalmologists experienced in glaucoma diagnosis. This approach can lead to refinement of the model attributes by data scientists or ML engineers working together with the doctors.

Data visualization

Embedding model results in extended context of the available patient data can facilitate identification of specific features related to the course of disease and prognosis. Following recommendations of the domain experts we provided box plot view of distribution of data included in a set of selected cases. This type of plot is common in scientific papers and usually is correctly interpreted by the medical professionals. Box plots support visual comparison of data distribution properties across different sets of cases (see Figure 8). User can define set of cases using filtering by diagnosis or the range of value of the selected measurements (IOP, CH, CRF etc.). It is

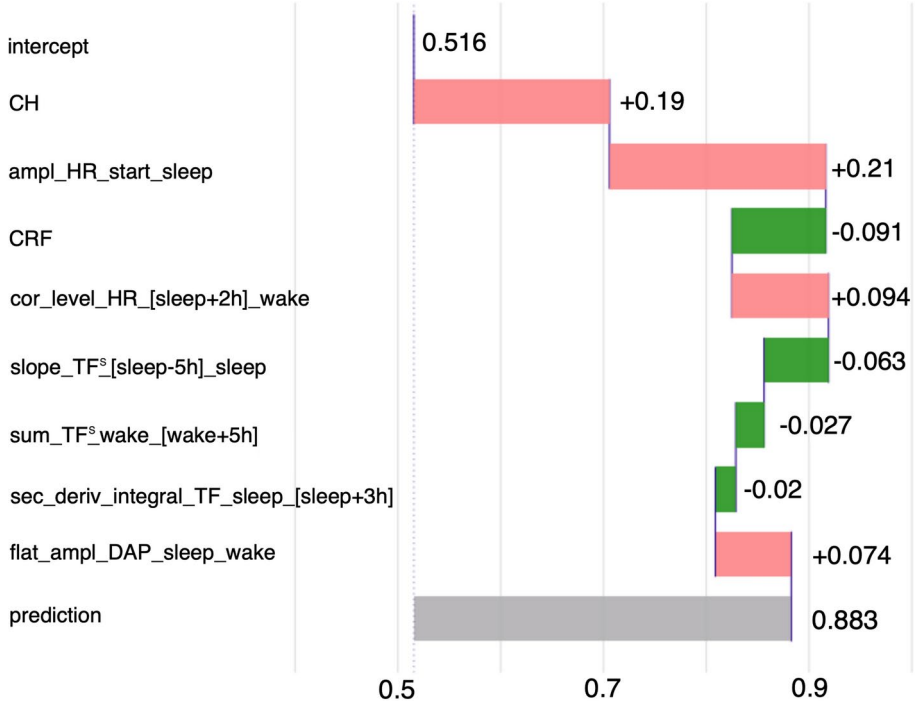

Fig. 7. Prediction for particular case can be decomposed onto model attributes using DALEX break-down profile generated on the basis of conditional responses of the model. User can assess the contribution of each attribute to the prediction (0.883) for the instance. Positive attribute contributions are shown as pink bars (e.g. CH), negative as green bars (e.g. CRF). Intercept can be interpreted as mean value (an estimate of the expected value of the model's predictions for all cases).

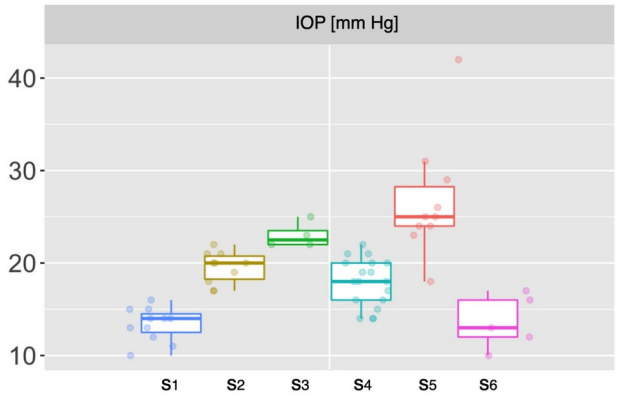

Fig. 8. Comparison of IOP [mm Hg] measurements distribution in the selected sets of cases. Box plots for the 6 sets of cases defined on the basis of diagnosis: s_1:NORM, s_2:suspected OH (ocular hypertension), s_3:OH, s_4:suspected POAG/HTG, s_5:POAG/HTG, s_6:POAG/NTG.

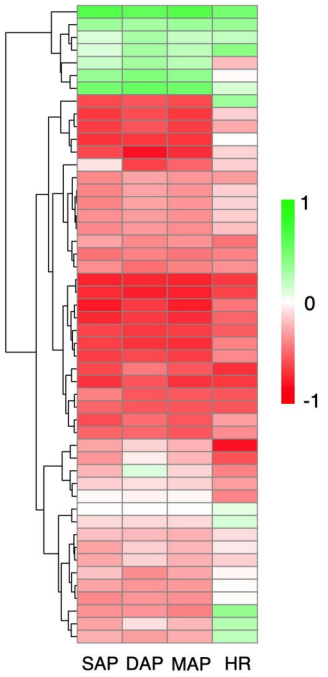

Fig. 9. Heat map view of the Spearman's rank correlation coefficient values for TF and cardiac sensor signal: SAP, DAP, MAP (mean arterial pressure), HR in the first 2 hours of the main/night sleep period (SLEEP_[SLEEP+2h]). View generated for the set of 50 cases.

also possible to generate heat map view of Triggerfish and cardiac signal correlations for a set of selected cases (in time intervals listed in Table 1). Rows (cases) in the heat map matrix can be ordered according to the result of hierarchical clustering for Euclidean distance (see Figure 9). Identification of subgroups of cases with significant positive or negative correlations and specific properties quantified by the other measurements can lead to more efficient diagnostic or treatment recommendations (related to e.g. cardiovascular system properties).

Collaborative research

Triggerfish and devices for continuous monitoring of cardiovascular system parameters have been introduced relatively recently into the clinical toolkit. Relation of Triggerfish CLS signal and cardiac activity with clinical data has not been extensively studied yet, so there are few detailed medical knowledge sources that could facilitate interpretation of such data for specific cases. Existing publications use different data analysis methods and usually report the properties of small patient groups, therefore can't be immediately used as the basis for guidelines of clinical routine. Exploratory data analysis is an approach that seems appropriate for examination of the available data in the multidisciplinary teams involving doctors and data scientists. It uses standard data

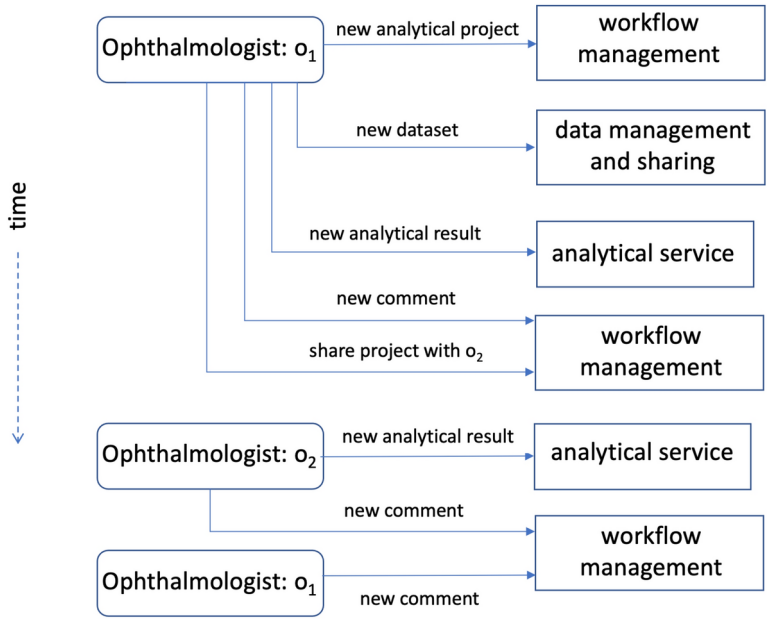

Fig. 10. Subsequent steps of collaborative research in the system. A researcher can create a new analytical project, select a specific dataset (e.g., NTG cases with low IOP< 15 mm Hg), generate box plots for some basic eye measurements (CH, CRF, IOP, etc.), write a summary of the results, then share the project with another researcher. The second researcher can generate heat maps for the available sensor data (e.g., correlation between Triggerfish and cardiac sensor signal: SAP, DAP, HR) for the considered dataset in selected time intervals, write a summary of these results, etc. In the following steps of exploratory data analysis performed for the given dataset, users can identify important features related to the development of the disease.

science techniques for comprehensive characterization of properties of the data. We can use statistical methods and various data visualization techniques to answer basic questions about patient data.

Collaborative research is a way of tackling complex problems. Clear communication is essential for effective collaboration. Our system enables creation of research projects that support activities related to development and application of personalized medicine techniques. Users working on the selected issue can add notes/comments containing multiple content in workspace of a project. Entries added by the users can contain text, images, links and embedded results of analytic functions provided by the system. Default order of entries is aligned with the main timeline of the project. Each analytic result includes its timestamp and the details of the set of cases for which it was created. Precise identification of the analytic results (such as box plots, heat maps and the other graphs) helps in tracking of discussion and development of the research conclusions.

Users can share access to selected cases from their repository. It encourages data collection and application of the analytic functions to assess specific properties of the data. Data sharing is one way to prompt more frequent use of sensor data in development of new glaucoma diagnosis and control standards based on continuous monitoring of eye and cardiovascular system properties. We would like to allow flexible exploration of data and exchange of new ideas. Constraints related to research workflow in the system are rather soft and primarily intended to improve coordination and reproducibility of the results. Figure 10 presents a detailed view of the steps in collaborative research in the system.

Discussion

Software system presented in this paper can be seen as a supporting tool for glaucoma diagnosis and collaborative research involving implementation of personalized medicine premises.

Personalized medicine assumes that individual patient data can be used to more precisely detect or treat a disease. Identification and evaluation of the relationship between time series recorded by multiple devices can be a way to better understanding the nature of the condition. Embedding sensor data in the context of clinical data maintained in the system can facilitate identification of specific physiological properties related to the course of the disease.

Transdisciplinary research

Transdisciplinarity concept has roots in discussions about the need for new forms of interdisciplinary collaboration[40]. It encompasses the following key characteristics:

- focus on specific, complex, real-world problems that are important for society
- transformative approach (i.e. support of action or change of the status quo)
- contemplation of broad context of the research and compatibility of its parts
- development of integrated knowledge that crosses disciplinary boundaries

Transdisciplinarity provides framework for the research projects and highlights importance of validation of the outcomes from different perspectives[41].

Glaucoma affects many millions of patients in the world and remains a major problem for the health care system. As the age is one of the significant risk factors, the disease is an important issue in the aging population. One of the aims of our research is development of new diagnostic methods based on multi-sensor data. These methods can supplement current options available for patients and change diagnosis and treatment standards.

Services of the system can be extended and accommodated to handle different data formats. New system scenarios can be proposed regarding possible data sources. For example, using genotyping arrays to find single nucleotide polymorphisms (SNP) across genome allow identification of genetic variants associated with risk of progression in particular glaucoma types[42]. Adding new system capabilities to handle such SNP data with consideration of its relationship with the other data will be consistent with transdisciplinarity assumptions related to compatibility and development of integrated knowledge. Involvement of users from many clinical and research fields is in line with the transdisciplinary attitude to the crossing of disciplinary boundaries.

Community adoption

Lack of adequate tools for management, sharing and analysis of collected data reduces productivity of the research. Design of our system is focused on the services for handling Triggerfish and cardiac sensor data. Data collection workflow in the system is similar to the approach common in clinical practice. It facilitates adoption by the eye doctors and customization of the services. We assume evolution of the system and incremental implementation of the functions that address new requirements arising in the diagnostic and analytic scenarios. As the users of the system have different experience and knowledge it is important to introduce guidelines for application of the ML techniques to diagnosis of glaucoma. Relevant issues in this context include:

- quality of sensor data in relation to patient activity during the day
- understanding limitations of the ML models and appropriate assessment of the predicted output

Establishment of the center of excellence for glaucoma research is planned in the near future at Poznan Supercomputing and Networking Center. It will provide high performance computational resources for application of the latest ML techniques in the field of ophthalmology, maintain the system and support collaborative development of personalized medicine standards.

Conclusions

In recent years, many new devices for continuous monitoring of patient health parameters have become available[43]. ML algorithms increasingly are able to efficiently process complex data. It makes possible to develop personalized approach in many fields of medicine. 24-hour Triggerfish record joined with cardiac sensor data can be used to more accurately diagnose or track progression of glaucoma. Nevertheless, large amount of data is required to build reliable ML models[44]. At this moment Triggerfish device is not commonly used in clinical practice and software tools for sensor data processing usually offer only simple analytic functions.

While relevant data availability is currently limited, we expect that inclusion of new users will lead to increase the size of the collection of cases in the system. Support for scenarios involving application of ML techniques for assessment of specific cases can encourage medical professionals to collect and share more sensor data for patients. Consequently, greater availability of the data can increase interest in collaborative research scenarios for novel approaches in glaucoma diagnosis and control.

Key contributions

The key contributions of our study include:

- provision of ML models involving sensor data to support diagnosis of glaucoma
- introduction of collaborative research platform that enables collection and use of sensor data acquired by Triggerfish and devices for continuous monitoring of cardiovascular system properties
- implementation of data exploration services that facilitate development and evaluation of clinical hypotheses

Future research

Possible extensions of the system include development of treatment monitoring services making use of sensor data. Another direction of research is introduction of deep learning models[45] (e.g. 1D convolutional neural networks) for glaucoma diagnosis or segmentation of Triggerfish time series (labeling features such as eye blinking).

Data availability

The datasets used and/or analysed during the current study are available from the corresponding author upon reasonable request.

Received: 20 May 2024; Accepted: 10 February 2025
Published online: 07 March 2025

References

1. Campbell, C. G. et al. The potential application of artificial intelligence for diagnosis and management of glaucoma in adults. *British Medical Bulletin* **134**(1), 21–33. https://doi.org/10.1093/bmb/ldaa012 (2020).

2. Guergueb, T. & Akhloufi, M. A. A Review of Deep Learning Techniques for Glaucoma Detection. *SN Computer Science* **4**, 274. https://doi.org/10.1007/s42979-023-01734-z (2023).
3. Maetschke, S., Antony, B. & Ishikawa, H. A feature agnostic approach for glaucoma detection in OCT volumes. *PLoS ONE* **14**(7). https://doi.org/10.1371/journal.pone.0219126 (2019).
4. Kim, J., Tran, L. & Chew, E. Y. et al. *Optic Disc and Cup Segmentation for Glaucoma Characterization Using Deep Learning*, 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS), pp. 489-494 https://doi.org/10.1109/CBMS.2019.00100 (2019).
5. Martin, K. R. et al. Use of Machine Learning on Contact Lens Sensor-Derived Parameters for the Diagnosis of Primary Open-angle Glaucoma. *American Journal of Ophthalmology* **194**, 46–53 (2018).
6. Yousefi, S. Clinical Applications of Artificial Intelligence in Glaucoma. *Journal of Ophthalmic & Vision Research* **18**(1), 97–112. https://doi.org/10.18502/jovr.v18i1.12730 (2023).
7. Mansouri, K., Weinreb, R. N. & Liu, J. H. K. Efficacy of a Contact Lens Sensor for Monitoring 24-H Intraocular Pressure Related Patterns. *PLoS ONE* **10**(5), 1–14 (2015).
8. Tojo, N., Abe, S., Ishida, M., Yagou, T. & Hayashi, A. The Fluctuation of Intraocular Pressure Measured by a Contact Lens Sensor in Normal-Tension Glaucoma Patients and Nonglaucoma Subjects. *Journal of Glaucoma* **26**(3), 195–200 (2017).
9. Liu, J. H. K., Mansouri, K. & Weinreb, R. N. Estimation of 24-Hour Intraocular Pressure Peak Timing and Variation Using a Contact Lens Sensor. *PLoS ONE* **10**(6) https://doi.org/10.1371/journal.pone.0129529 (2015).
10. Mansberger, S. L., Medeiros, F. & Gordon, M. Diagnostic Tools for Calculation of Glaucoma Risk. *Survey of Ophthalmology* **53**(6), 11–16 (2008).
11. Willekens, K. et al. Review on Dynamic Contour Tonometry and Ocular Pulse Amplitude. *Ophthalmic Research* **55**, 91–98 (2016).
12. Pillunat, K. R. et al. Nocturnal blood pressure in primary open-angle glaucoma. *Acta Ophthalmologica* **93**(8), 621–626 (2015).
13. Pallikaris, I., Tsilimbaris, M. K. & Dastiridou, A. I. *Ocular Rigidity* (Springer, Biomechanics and Hydrodynamics of the Eye, 2021).
14. Thompson, A. C., Jammal, A. A. & Medeiros, F. A. A review of deep learning for screening, diagnosis, and detection of glaucoma progression. *Translational Vision Science & Technology* **9**(2), 42 (2020).
15. Świerczyński, H., Pukacki, J., Szczęsny, S., Mazurek, C. & Wasilewicz, R. Sensor data analysis and development of machine learning models for detection of glaucoma. *Biomedical Signal Processing and Control* **86**(C) https://doi.org/10.1016/j.bspc.2023.105350 (2023).
16. Zhang, N., Wang, J. & Li, Y. et al. Prevalence of primary open angle glaucoma in the last 20 years: a meta-analysis and systematic review. *Nature Scientific Reports* **11**, 13762. https://doi.org/10.1038/s41598-021-92971-w (2021).
17. Chan, M.P., Broadway, D. C. & Khawaja A. P. et al. Glaucoma and intraocular pressure in EPIC-Norfolk Eye Study: cross sectional study, *BMJ*, **358**(8121) https://doi.org/10.1136/bmj.j3889 (2017).
18. Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P. SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research* **16**, 321–357 (2002).
19. Śródka, W. Goldmann applanation tonometry - not as good as gold. *Acta of Bioengineering and Biomechanics* **12**(2), 39–47 (2010).
20. Bilo, G. et al. Validation of the Somnotouch-NIBP noninvasive continuous blood pressure monitor according to the European Society of Hypertension International Protocol revision 2010. *Blood Pressure Monitoring* **20**(5), 291–294 (2015).
21. Susanna, C. N., Diniz-Filho, A. & Daga, F. B. A Prospective Longitudinal Study to Investigate Corneal Hysteresis as a Risk Factor for Predicting Development of Glaucoma. *American Journal of Ophthalmology* **187**, 148–152 (2018).
22. Gisler, Ch., Ridi, A. & Fauquex, M. *Towards Glaucoma Detection Using Intraocular Pressure Monitoring, 6th International Conference of Soft Computing and Pattern Recognition (SoCPaR)* (Tunis, Tunisia, 2014).
23. Wasilewicz, R. et al. 24 hour continuous ocular tonography Triggerfish and biorhythms of the cardiovascular system functional parameters in healthy and glaucoma populations. *Acta Ophthalmologica* **91**(s252), https://doi.org/10.1111/j.1755-3768.2013.2721.x (2013).
24. Hastie, T. & Tibshirani, R. *J* 2nd edn. (H. Friedman, The Elements of Statistical Learning, Springer), (2009).
25. Park, D.J., Park, M.W. & Lee, H. et al. Development of machine learning model for diagnostic disease prediction based on laboratory tests. *Nature Scientific Reports* **11**, 7567. https://doi.org/10.1038/s41598-021-87171-5 (2021).
26. Landry, M. & Bartz, A. *Machine Learning with R and H2O*, https://docs.h2o.ai/h2o/latest-stable/h2o-docs/booklets/RBooklet.pdf, H2O.ai, Inc., 7th edition, (2021).
27. Teisseyre, P., Kłopotek, R. A. & Mielniczuk, J. Random Subspace Method (RSM) for Linear Regression. *Computational Statistics* **31**, 943–972 (2016).
28. Fawcett, T. An introduction to ROC analysis. *Pattern Recognition Letters* **27**(8), 861–874 (2006).
29. Kaushik, S., Pandav, S. S. & Banger, A. Relationship between corneal biomechanical properties, central corneal thickness, and intraocular pressure across the spectrum of glaucoma. *American Journal of Ophthalmology* **153**(5), 840–849 (2012).
30. Sutton, R.T. & Pincock, D. et al. An overview of clinical decision support systems: benefits, risks, and strategies for success. *npj Digital Medicine* **3**(1) https://doi.org/10.1038/s41746-020-0221-y (2020).
31. Muccini, H. & Vaidhyanathan, K. *Software Architecture for ML-based Systems: What Exists and What Lies Ahead*, 2021 IEEE/ACM 1st Workshop on AI Engineering - Software Engineering for AI (WAIN), Madrid, Spain, pp. 121-128 https://doi.org/10.1109/WAIN52551.2021.00026 (2021).
32. Liu, G., Huang, B. & Liang, Z. et al. *Microservices: architecture, container, and challenges*, 2020 IEEE 20th International Conference on Software Quality, Reliability and Security Companion (QRS-C), Macau, China, pp. 629-635 https://doi.org/10.1109/QRS-C51114.2020.00107 (2020).
33. Al-Debagy, O. & Martinek, P. *A Comparative Review of Microservices and Monolithic Architectures*, 2018 IEEE 18th International Symposium on Computational Intelligence and Informatics (CINTI), Budapest, Hungary, pp. 000149-000154, https://doi.org/10.1109/CINTI.2018.8928192 (2018).
34. Myneni, S. & Patel, V. L. Organization of biomedical data for collaborative scientific research: A research information management system. *International Journal of Information Management* **30**(3), 256–264 (2010).
35. Stadler, L. et al. Optimizing R language execution via aggressive speculation. *SIGPLAN Notices* **52**(2), 84–95. https://doi.org/10.1145/3093334.2989236 (2017).
36. Hausmann, D. & Zulian, C. et al. Tracing the decision-making process of physicians with a Decision Process Matrix, *BMC Medical Informatics and Decision Making*, **16**(133) https://doi.org/10.1186/s12911-016-0369-1 (2016).
37. Rudin, C. et al. Interpretable Machine Learning: Fundamental Principles and 10 Grand Challenges. *Statistics Surveys* **16**, 1–85. https://doi.org/10.1214/21-SS133 (2022).
38. Ribeiro, M. et al. *"Why Should I Trust You?": Explaining the Predictions of Any Classifier*, 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16), New York, USA, https://doi.org/10.1145/2939672.2939778 (2016).
39. Biecek, P. DALEX: Explainers for Complex Predictive Models in R. *Journal of Machine Learning Research* **19**, 84:1-84:5 (2018).
40. Lawrence, M. G. et al. Characteristics, potentials, and challenges of transdisciplinary research. *One Earth* **5**(1), 44–61. https://doi.org/10.1016/j.oneear.2021.12.010 (2022).
41. Morton, W. L. et al. Architectures of adaptive integration in large collaborative projects. *Ecology and Society* **20**(4), 5. https://doi.org/10.5751/ES-07788-200405 (2015).
42. Mabuchi, F. et al. Genetic Variants Associated With the Onset and Progression of Primary Open-Angle Glaucoma. *American Journal of Ophthalmology* **215**, 135–140 (2020).

43. Anikwe, C. V. et al. Mobile and wearable sensors for data-driven health monitoring system: State-of-the-art and future prospect. *Expert Systems with Applications* **202**(117362), https://doi.org/10.1016/j.eswa.2022.117362 (2022).
44. Arora, A. et al. The value of standards for health datasets in artificial intelligence-based applications. *Nature Medicine* **29**, 2929–2938. https://doi.org/10.1038/s41591-023-02608-w (2023).
45. Zhang, A., Lipton, Z.C., Li, M. & Smola, A.J. *Dive into Deep Learning*, Cambridge University Press, 1st edition, https://D2L.ai (2023).

Acknowledgements
This work was supported by the Polish Ministry of Science and Higher Education, DWD/4/24/2020 funding for Applied Ph.D. Programme for 2020-2024.

Author contributions
Conceptualization: C.M., R.W., J.P.,H.Ś.; Data acquisition and management: R.W., J.P.; Manuscript writing and editing: H.Ś., S.S.; Software design and implementation: H.Ś.; Investigation: R.W.; Statistical analysis: H.Ś.; Supervision: S.S., C.M. All authors reviewed the manuscript.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to H.Ś.

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.