AMERICAN ACADEMY OF OPHTHALMOLOGY®

Artificial Intelligence Models to Identify Patients at High Risk for Glaucoma Using Self-reported Health Data in a United States National Cohort

Rohith Ravindranath, MS, Joel Naor, MD, MS, Sophia Y. Wang, MD, MS

Purpose: Early glaucoma detection is key to preventing vision loss, but screening often requires specialized eye examination or photography, limiting large-scale implementation. This study sought to develop artificial intelligence models that use self-reported health data from surveys to prescreen patients at high risk for glaucoma who are most in need of glaucoma screening with ophthalmic examination and imaging.

Design: Cohort study.

Participants: Participants enrolled from May 1, 2018, to July 1, 2022, in the nationwide All of Us Research Program who were ≥18 years of age, had ≥2 eye-related diagnoses in their electronic health record (EHR), and submitted surveys with self-reported health history.

Methods: We developed models to predict the risk of glaucoma, as determined by EHR diagnosis codes, using 3 machine learning approaches: (1) penalized logistic regression, (2) XGBoost, and (3) a fully connected neural network. Glaucoma diagnosis was identified based on International Classification of Diseases codes extracted from EHR data. An 80/20 train–test split was implemented, with cross-validation employed for hyperparameter tuning. Input features included self-reported demographics, general health, lifestyle factors, and family and personal medical history.

Main Outcome Measures: Models were evaluated using standard classification metrics, including area under the receiver operating characteristic curve (AUROC).

Results: Among the 8205 patients, 873 (10.64%) were diagnosed with glaucoma. Across models, AUROC scores for identifying which patients had glaucoma from survey health data ranged from 0.710 to 0.890. XGBoost achieved the highest AUROC of 0.890 (95% confidence interval [CI]: 0.860–0.910). Logistic regression followed with an AUROC of 0.772 (95% CI: 0.753–0.795). Explainability studies revealed that key features included traditionally recognized risk factors for glaucoma, such as age, type 2 diabetes, and a family history of glaucoma.

Conclusions: Machine and deep learning models successfully utilized health data from self-reported surveys to predict glaucoma diagnosis without additional data from ophthalmic imaging or eye examination. These models may eventually enable prescreening for glaucoma in a wide variety of low-resource settings, after which high-risk patients can be referred for targeted screening using more specialized ophthalmic examination or imaging.

Financial Disclosure(s): The author(s) have no proprietary or commercial interest in any materials discussed in this article. *Ophthalmology Science 2025;5:100685*

Supplemental material available at www.ophthalmologyscience.org.

Glaucoma is a serious eye condition that progressively damages the optic nerve, eventually leading to irreversible vision loss and blindness if not treated. Glaucoma is the leading cause of blindness, both globally[1–3] and in the United States,[4] impacting 80 million individuals worldwide and 3 million in the United States.[4] Glaucomatous vision loss occurs gradually over time, progressing from asymptomatic in early stages to blindness in advanced glaucoma.[5,6] Alarmingly, 50% to 80% of glaucoma cases remain undiagnosed.[5,6] Moreover, Black and Hispanic/Latino communities are disproportionately affected by glaucoma.[7,8] These high-risk groups not only experience a higher prevalence of undetected and untreated glaucoma but also tend to be diagnosed at more advanced stages of the disease.[6,10] They are more likely to suffer vision loss after diagnosis and are generally underrepresented in glaucoma research.[9,11]

Screening for glaucoma plays a crucial role in enabling early detection and treatment. Research conducted on local population-based cohorts has found that the rate of positive glaucoma screenings is around 1% to 2% in Whites,[12] 2% in Hispanics,[13] and as high as 5% in Blacks.[12] Despite the importance of early glaucoma treatment for preventing blindness, the United States Preventive Services Task Force has highlighted several obstacles to implementing

https://doi.org/10.1016/j.xops.2024.100685
ISSN 2666-9145/25

broad-based screening programs.[14] Diagnosing glaucoma necessitates evaluating both the structure and function of the optic nerve. Most earlier studies have focused on stratifying glaucoma risk through imaging or functional tests,[15–19] and "most tests require specialized equipment and are performed in an eye specialty setting,"[14] making broad-scale screening or diagnosis less practical. Additionally, there is a shortage of dependable methods or validated instruments to identify individuals at higher risk of developing glaucoma, who would be the best candidates for more intensive screening.[14] Further research is needed to develop and validate risk assessment tools for early detection of those at heightened risk and to guide more effective screening strategies.[14]

Previous studies aimed at identifying high-risk patients during the prescreening stage have generally been limited in scope, focusing on race/ethnicity and family history.[20–22] Without the ability to target equipment- and resource-intensive screening efforts to the highest risk populations, it has been challenging to conduct randomized controlled trials with the scale and follow-up necessary to definitively demonstrate the benefits of glaucoma screening on long-term vision-related quality of life, even though glaucoma treatment is known to prevent vision loss over the patient's lifetime.[14]

The National Institute of Health's All Of Us[23] project provides a unique chance to create glaucoma prescreening tools designed to identify patients with a high likelihood of glaucoma across a broad and diverse United States population by applying artificial intelligence to self-reported survey data. This approach addresses the above key critical challenges in glaucoma screening. The All Of Us research project collects extensive health information from over a quarter million individuals across the country, including a myriad of surveys containing self-reported health data. All Of Us also prioritizes including a particularly diverse population, especially those traditionally underrepresented in clinical studies.[24] The aim of this study was to develop artificial intelligence-driven prediction algorithms that can analyze the self-reported survey data to identify patients with a high risk of glaucoma. These algorithms could eventually be used as prescreening tools to automatically detect patients with a higher probability of glaucoma, thereby directing resource-intensive, imaging-based screening programs toward these individuals.

Methods

We conducted a retrospective cross-sectional study using the All of Us Research Program to develop predictive models that use only self-reported health data to predict the risk of any kind of glaucoma. This study, using deidentified data, was exempt from Stanford University's Institutional Review Board approval and adheres to the tenets of the Declaration of Helsinki.

Data Source

The All of Us Research Program has developed a comprehensive longitudinal database that encompasses clinical, environmental, lifestyle, and genetic information from >1 270 000 individuals across the United States. United States residents aged 18 and older can participate by enrolling through the program's website (JoinAllOfUs.org) or via >60 health care provider organizations. Data are collected from various sources such as electronic health records (EHRs), surveys, physical measurements, and biospecimen collection to ensure a comprehensive and accurate representation of participants' health information. Upon completion of the core surveys (The Basics, Lifestyle, and Overall Health), all participants are invited to complete additional optional surveys such as Personal and Family History. All collected data are standardized using the Observational Medical Outcomes Partnership Common Data Model for observational health research.[25]

Inclusion/Exclusion Criteria

This study used data from the All Of Us, version 7 cohort, the most recent release at the time of analysis, covering participants enrolled from May 1, 2018, to July 1, 2022. Participants had ≥2 separate encounters with 2 separate eye-related diagnoses in their EHRs, indicating they likely visited an eye care provider where glaucoma status could have been assessed. Eye-related findings and diagnoses were identified using a hierarchical ontology of Observational Medical Outcomes Partnership codes, which are based on the Systemized Nomenclature of Medicine (SNOMED) codes (4038502, eye/vision finding), allowing the selection of parent codes that automatically include all related child codes. To train a model to identify patients with confirmed glaucoma, the study included those with definitive glaucoma diagnosis codes and those without any glaucoma-related diagnosis codes; patients diagnosed with suspect or borderline glaucoma were excluded (see Table S1 for related concept codes, available at www.ophthalmologyscience.org). Additionally, patients with no prior eye-related diagnoses were excluded, as it was impossible to determine if they had undiagnosed glaucoma. The study's cohort design and inclusion/exclusion criteria are summarized in Figure 1.

Feature Engineering

Model input features included information from Overall Health, The Basics, Lifestyle, and Personal and Family History surveys. The Personal and Family History survey collects information about the medical history of participants and participants' immediate biological family members. The Lifestyle survey collects information about participants' daily habits and behaviors, aiming to understand how these factors impact health. The Overall Health survey collects information regarding participants' reported levels of individual health, including overall general health, physical and mental health, and quality of life. The full text of the questions investigated in this study are listed in Table S2 (available at www.ophthalmologyscience.org), as well as their answer choices and short descriptions by which we will refer to these questions. Questions selected mainly focused on eye conditions, cardiovascular conditions, and hormone/endocrine conditions, as they have been known to be associated with glaucoma through previous literature.[26,27] A personal history of glaucoma was not included as a question in the survey, because in practice, those patients who already know they have glaucoma would not be the population taking this survey to be screened for glaucoma.

Information about demographics and additional covariates were identified from The Basics survey. Age was considered as a continuous covariate. Race\ethnicity was categorized as non-Hispanic White, Hispanic, non-Hispanic Black, non-Hispanic Asian, other race\ethnicity, and unknown. Sex at birth was categorized as male, female, and unknown. Education was categorized as did not graduate high school, high school graduate or graduate equivalency degree, some college, college graduate, advanced degree, and unknown. Gender was categorized as cisgender man, cisgender woman, nonbinary or transgender, and

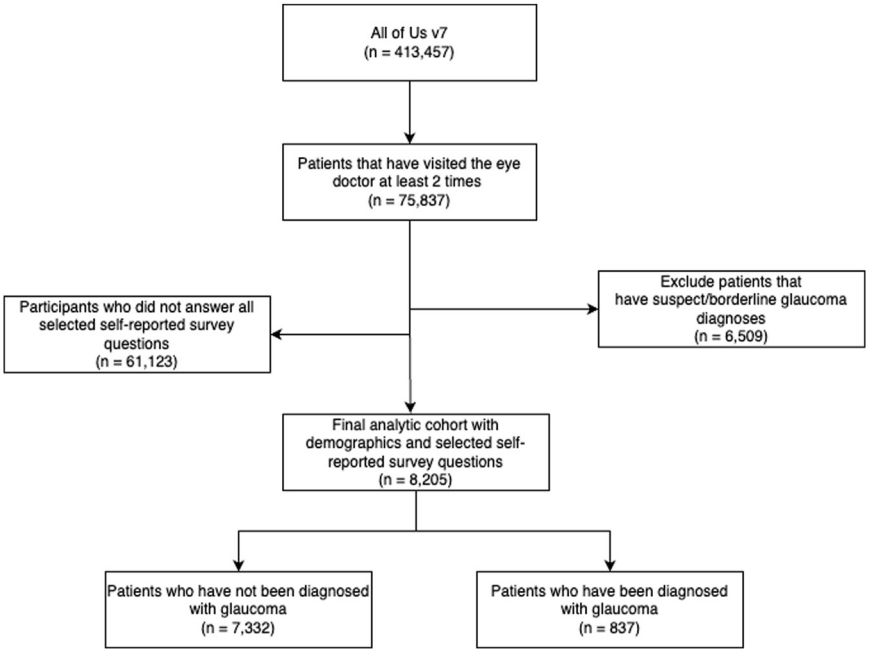

Figure 1. Feature engineering and cohort construction. Flowchart depicting the cohort design, with inclusion and exclusion criteria for the study population.

unknown via the 2-step process from Resiner Poteat et al, 2016.[28] Sexual identity was categorized as straight/heterosexual, everything else, and unknown. Income was categorized as <$25 000, $25 000 to $49 999, $50 000 to $74 999, $75 000 to $99 999, $100 000 to $149 999, $150 000+, and unknown. Relationship status was categorized as married or living together, never married, widowed, divorced or separated, or unknown. Insurance coverage was categorized as yes, no, and unknown. Employment status was categorized as employed for wages, self-employed, out of work/unable to work, retired, other (student/homemaker), or unknown. All measures were converted to one-hot encoded vectors and used as input into the model.

Modeling

Machine learning models were fitted on the training data using the Python *sklearn v1.2.1* package.[29] These models included penalized logistic regression (L2 penalization) and XGBoost. Hyperparameters were tuned using fivefold cross-validation on the training set to optimize the area under the receiver operating characteristic curve (AUROC). Grid search was used to tune penalization for logistic regression, and random search was used for XGBoost. A summary of hyperparameters is in Table S3 (available at www.ophthalmologyscience.org).

A fully connected deep learning network (FCN) was also built using Python TensorFlow v2.11.0 package.[30] The FCN used the same structured features as the machine learning models, constructed with 3 dense layers, 2 dropout layers, and an output layer with a sigmoid activation function (Input(61) -> Dense(128) -> Dropout(0.4) -> Dense(64) -> Dense(32) -> Dropout(0.4) -> Output(1, sigmoid)).

Evaluation

We assessed our classification models using standard metrics, including accuracy, balanced accuracy, recall, precision, F1 score (which balances recall and precision), AUROC, and the area under the precision—recall curve, all evaluated on the test set. Accuracy measures the proportion of correct predictions, while balanced accuracy corrects for data bias by averaging sensitivity and specificity, effectively weighting each sample by the inverse prevalence of its true class. During cross-validation, we fine-tuned the classification threshold for each model to optimize the F1 score on the test set. We used bootstrap simulation to generate 95% confidence intervals (CIs) for all metrics, performing 1000 training and test set iterations to produce 1000 sets of model statistics. This method enables us to empirically evaluate variability in model performance, enhancing result transparency. Additionally, we compared the model's performance (AUROC score) stratified by race/ethnicity to gain insights regarding the model's inherent biases. Finally, we performed explainability analyses using Shapley values, also known as SHapley Additive exPlanations values, to determine feature importance.[31,32] This method determines feature importance based on the magnitude of feature attributions through a game theory—based approach. Shapley values quantify the marginal contribution of each feature to the model's predictions, considering all possible subsets of features. We computed SHapley Additive exPlanations values for the XGBoost model on the test set.

Results

Study Population

Population characteristics for the study cohort of 8205 participants are summarized in Table 4. Patients who were diagnosed with glaucoma represented 10.64% (N = 873) of the cohort. The overall mean age was 62.72 years (standard deviation: 14.88). The majority of the cohort was female (67.54%, N = 5542). The majority of the cohort was non-Hispanic White (79.57%, N = 6529). Non-Hispanic Black participants formed 7.04% (N = 578)

Table 4. Population Characteristics

	Glaucoma Patients		Nonglaucoma Patients		Total	
	N = 873		*N = 7332*		*N = 8205*	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
Age	70.65	10.28	61.77	15.06	62.72	14.88
	N	**%**	**N**	**%**	**N**	**%**
Sex at birth						
Male	348	39.86%	2308	31.48%	2656	32.37%
Female	523	59.91%	5019	68.45%	5542	67.54%
Unknown	2	0.23%	5	0.07%	7	0.09%
Race/ethnicity						
Non-Hispanic White	644	73.77%	5885	80.26%	6529	79.57%
Hispanic	54	6.19%	525	7.16%	579	7.06%
Non-Hispanic Black	99	11.34%	479	6.53%	578	7.04%
Non-Hispanic Asian	22	2.52%	126	1.72%	148	1.80%
Non-Hispanic Middle Eastern/North African	9	1.03%	48	0.65%	57	0.69%
Non-Hispanic Native Hawaiian/Pacific Islander	1	0.11%	6	0.08%	7	0.09%
Other race/ethnicity	9	1.03%	68	0.93%	77	0.94%
Unknown	35	4.01%	195	2.66%	230	2.80%
Education						
Did not graduate high school	13	1.49%	124	1.69%	137	1.67%
High school graduate/graduate equivalency degree	85	9.74%	685	9.34%	770	9.38%
Some college	227	26.00%	1911	26.06%	2138	26.06%
College graduate or above	250	28.64%	2173	29.64%	2423	29.53%
Advanced degree	288	32.99%	2356	32.13%	2644	32.22%
Unknown	10	1.15%	83	1.13%	93	1.13%
Gender						
Cisgender man	343	39.29%	2274	31.01%	2617	31.90%
Cisgender woman	516	59.11%	4952	67.54%	5468	66.64%
Transgender/nonbinary	6	0.69%	42	0.57%	48	0.59%
Unknown	8	0.92%	64	0.87%	72	0.88%
Sexual identity						
Straight/heterosexual	812	93.01%	6636	90.51%	7448	90.77%
Everything else	46	5.27%	575	7.84%	621	7.57%
Unknown	15	1.72%	121	1.65%	136	1.66%
Employment status						
Employed for wages	186	21.31%	2630	35.87%	2816	34.32%
Self-employed	45	5.15%	506	6.90%	551	6.72%
Out of work/unable to work	109	12.49%	1045	14.25%	1154	14.06%
Retired	497	56.93%	2688	36.66%	3185	38.82%
Other (student/homemaker)	21	2.41%	370	5.05%	391	4.77%
Unknown	15	1.72%	93	1.27%	108	1.32%
Income						
<25k	93	10.65%	918	12.52%	1011	12.32%
25K–49.9k	140	16.04%	1142	15.58%	1282	15.62%
50K–74.9k	143	16.38%	1127	15.37%	1270	15.48%
75K–99.9k	126	14.43%	944	12.88%	1070	13.04%
100–149.9k	109	12.49%	1110	15.14%	1219	14.86%
150k+	143	16.38%	1260	17.18%	1403	17.10%
Unknown	119	13.63%	831	11.33%	950	11.58%
Marital status						
Married or living with partner	541	61.97%	4475	61.03%	5016	61.13%
Never married	93	10.65%	1070	14.59%	1163	14.17%
Widowed	74	8.48%	493	6.72%	567	6.91%
Divorced/separated	152	17.41%	1218	16.61%	1370	16.70%
Unknown	13	1.49%	76	1.04%	89	1.08%
Health insurance						
No	6	0.69%	80	1.09%	86	1.05%
Yes	857	98.17%	7196	98.15%	8053	98.15%
Unknown	10	1.15%	56	0.76%	66	0.80%

SD = standard deviation.

of the cohort, and Hispanics formed 7.06% (N = 579) of the cohort.

Model Performance

Receiver operating characteristic curves and precision–recall curves for all models on the entire cohort are shown in Figure 2, evaluated on the test set. The XGBoost model achieved the highest performance with an AUROC score of 0.890 (95% CI: 0.860–0.910). Logistic regression achieved the second highest performance with an AUROC score of 0.772 (95% CI: 0.753–0.795), while the FCN model performed the worst with an AUROC score of 0.710 (95% CI: 0.683–0.722). Classification metrics on the overall cohort are summarized in Table 5, with individual classification thresholds tuned on a validation set to maximize the F1 score. XGBoost performed the best with regard to accuracy, balanced accuracy, precision, and recall with scores of 0.904 (95% CI: 0.878–0.923), 0.771 (95% CI: 0.753–0.781), 0.572 (95% CI: 0.555–0.596), and 0.586 (95% CI: 0.554–0.592), respectively. Logistic regression performed the best with respect to accuracy, precision, and F1 with scores of 0.790 (95% CI: 0.765–0.811), 0.281 (95% CI: 0.267–0.291), and 0.372 (95% CI: 0.358–0.391), respectively. The FCN model had the best recall with a score of 0.730 (95% CI: 0.704–0.755).

A secondary analysis was performed to analyze our model's performance stratified by race/ethnicity. Table 6 reports the AUROC scores stratified by race/ethnicity for all models. XGBoost had the best performance for all subgroups. Non-Hispanic White had an AUROC score of 0.894 (95% CI: 0.883–0.902), Hispanic had a score of 0.846 (95% CI: 0.832–0.858), and non-Hispanic Black had a score of 0.916 (95% CI: 0.901–0.923).

Explainability

We performed explainability analyses to determine which input features contributed most to the model predictions, calculating Shapley values for the XGBoost model (Fig 3). The most important features included various demographic and eye conditions, such as age, race (non-Hispanic Black), cataracts, and refractive error (myopia and hyperopia). Notably, family history of glaucoma (mother, father, and sibling) was also among the most important features, as was type 2 diabetes.

Discussion

In this large nationwide multicenter study, we developed artificial intelligence prediction algorithms to identify patients with a high probability of glaucoma diagnosis using only self-reported survey data. Despite not using ophthalmic data, EHR, or imaging to predict glaucoma risk, the best-performing model achieved an AUROC of 0.890 (95% CI: 0.860–0.910) for identifying patients with glaucoma as well as being the most calibrated between the 3 models. When stratified by race/ethnicity, we observed that XGBoost performed the best across all subpopulations. Explainability studies demonstrated that important features included those traditionally known to be risk factors for glaucoma, as well as other systemic features.

There has been some previous work to develop models and tools that use nonimaging data to identify individuals at high risk for certain diagnoses such as glaucoma,[20–22] diabetic retinopathy,[33,34] and other medical diseases.[35–38] Most of these studies primarily rely upon ophthalmic clinical examination findings in order to risk-stratify patients, thus generally requiring ophthalmic equipment and

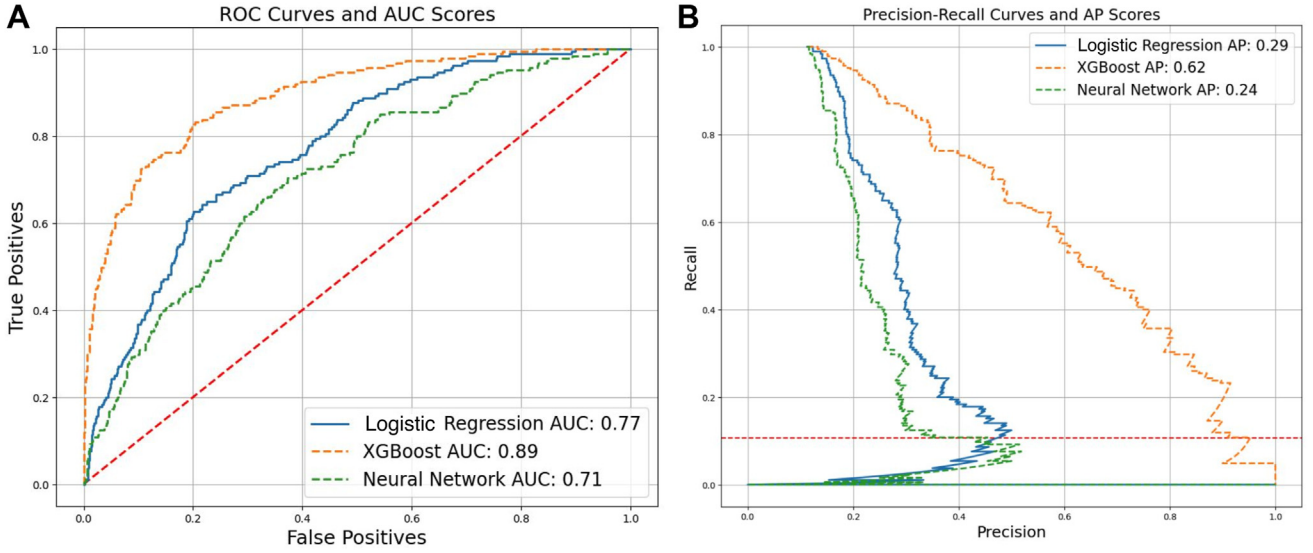

Figure 2. Receiver operating characteristic and precision–recall curves. This figure shows receiver operating characteristic curves (ROCs) (left) and precision–recall curves (right) for all models, evaluated on the entire test set. AP = area under the precision-recall curve; AUC = area under the receiver operating curve.

Table 5. Model Performance Metrics

	AUROC	ACC	Balanced ACC	Recall	Precision	F1
Logistic regression	0.772 (0.753–0.795)*	0.790 (0.765–0.811)	0.686 (0.670–0.698)	0.551 (0.535–0. 562)	0.281 (0.267–0.291)	0.372 (0.358–0.391)
XGBoost	**0.890 (0.86–0.910)**	**0.904 (0.878–0.923)**	**0.771 (0.753–0.781)**	0.600 (0.593–0.611)	**0.572 (0.555–0.596)**	**0.586 (0.554–0. 592)**
Fully connected network (FCN)	0.710 (0.683–0.722)	0.664 (0.634–0.689)	0.692 (0.673–0.713)	**0.730 (0.704–0.755)**	0.211 (0.199–0.225)	0.328 (0.308–0.342)

ACC = accuracy; AUROC = area under the receiver operating characteristic curve.
Bolded terms indicate best performance across all models.
*95% confidence intervals shown in parentheses.

ophthalmic examination by qualified personnel. For instance, Laroche et al[22] created a cost-effective, minimally invasive tool to estimate glaucoma risk based on age, intraocular pressure, and central corneal thickness. This tool was tested on 104 eyes—both normal and glaucomatous—from a single clinic in New York, achieving an AUROC of 0.81 in distinguishing between glaucoma patients and controls within that group.[22] In contrast, our models do not rely on any ophthalmic clinical data, making them suitable for use in nonophthalmic environments or in patients who have no prior eye care history. To assess one's risk of glaucoma and recommend further screening, a patient could simply fill out a survey. Moreover, our model was trained and assessed using a large, diverse, multicenter cohort across the United States, with separate training and evaluation populations to ensure performance metrics are robust and less prone to overfitting.

A similar approach using patients' self-reported health data to determine ophthalmic risk was used for the Diabetic Retinopathy Risk Assessment Tool (DRRisk). The DRRisk tool is available online and thus patients may use an online survey to provide inputs for the algorithm, which utilizes 14 features, including systolic blood pressure, hemoglobin A1c, hemoglobin, sex, ethnicity, diastolic blood pressure, age, and others. The DRRisk model achieved an AUROC of 0.8 for predicting diabetic retinopathy. Notably, the features for this calculator do require the patients' laboratory results and physical examination measurements, whereas our glaucoma risk model does not require this information. Because diabetes is a systemic disease, there are many known systemic markers of poor control that might be reasonably predictive of diabetic retinopathy. Despite these differences, it is striking that we were nevertheless able to achieve better results for predicting glaucoma risk, achieving an AUROC of 0.890 using only self-reported survey data.

Our glaucoma prescreening method effectively focuses screening efforts on high-risk patients by using routinely collected survey data. This targeted approach allows us to refer those most likely to have glaucoma for specialized ophthalmic screening, optimizing time and financial resources by concentrating imaging-based screening on those who need it most. General population screening for glaucoma typically yields a screen-positive rate—or glaucoma prevalence—of around 5% at most, underscoring the significant challenge of widespread screening for this condition.[13,39–41] For instance, the Proyecto VER study found a glaucoma prevalence of 1.97% (95% CI: 1.58%–2.36%) in a population-based sample of Hispanic adults >40 years old.[13] Similarly, the Rotterdam Study reported an overall prevalence of 1.10% (95% CI: 1.09, 1.11).[39]

Conducting a randomized controlled trial for glaucoma screening is challenging due to the low screen-positive rate in the general population, despite the known benefits of treatment in preventing vision loss. Our models achieved a precision of 0.572 when the predicted probability threshold was set to maximize the balance between recall (sensitivity) and positive predictive value (precision). At this threshold, deploying such a model to screen only model-positive patients could increase the screen-positive rate 10-fold compared with screening the general population. This

Table 6. Model Performance with Area Under the Receiver Operating Characteristic Curve Stratified by Race/Ethnicity

	XGBoost	Logistic Regression	Fully Connected Network
Non-Hispanic White	**0.894 (0.883–0.902)***	0.779 (0.761–792)	0.726 (0.714–0.733)
Hispanic	**0.846 (0.832–0.858)**	0.692 (0.683–0.709)	0.552 (0.548–0.562)
Non-Hispanic Black	**0.916 (0.901–0.923)**	0.851 (0.844–0.868)	0.775 (0.764–0.788)
Non-Hispanic Asian	**0.826 (0.814–0.846)**	0.674 (0.663–0.683)	0.590 (0.583–0.622)
Other/unknown	**0.843 (0.821–0.855)**	0.753 (0.736–0.761)	0.626 (0.601–0.639)

Bolded terms indicate best performance across all models.
*95% confidence intervals shown in parentheses.

targeted approach could optimize screening resource allocation and make future clinical trials of screening more feasible. While there is a tradeoff between recall and precision, the model allows adjustment of the "positive" prediction threshold to align with available screening capacity. For example, more patients could be screened, increasing sensitivity but also approaching the "screen everybody" approach, or fewer but higher probability patients could be screened, increasing positive predictive value. Although model performance on sensitivity and positive predictive value could still be improved, this prescreening approach still provides a substantial improvement over the current state, where an absence of systematic methods for identifying which patients to screen for glaucoma is limiting the ability to establish any screening programs dedicated to glaucoma.

Another strength of this study is the investigation of model explainability. Some features that our model has identified as important are known risk factors for glaucoma while others are novel. Age and non-Hispanic Black

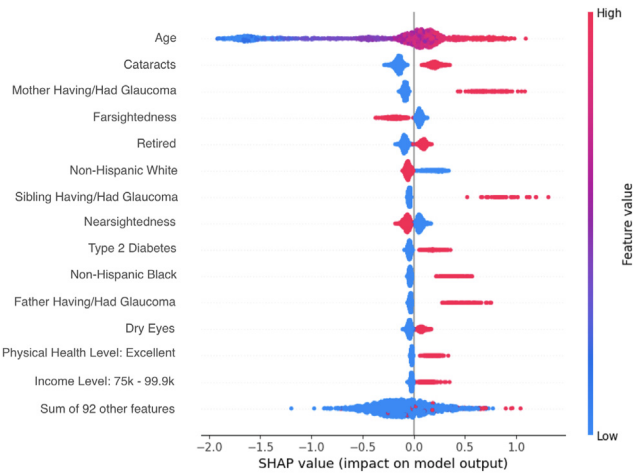

Figure 3. Model explainability with Shapley feature importance. The figure illustrates the Shapley value for the most important features in predicting whether a patient is at high risk for glaucoma, using the XGBoost model and calculated across the test set. Positive Shapley values indicate an influence toward predicting surgery, while negative values suggest an influence toward predicting no surgery. Points represent individual observations in the test set, with the color of each point reflecting the relative value of that feature for that patient. SHAP = SHapley Additive exPlanations.

populations are known to be accepted risk factors for glaucoma since older patients are more likely to have glaucoma progression than younger patients at similar intraocular pressure and individuals with African descent are known to be a predictor as there is higher prevalence, earlier presentation, and faster progression of primary open-angle glaucoma.[11,20] We also observed that having immediate family members who have been diagnosed with glaucoma is an important factor for our model and is a well-established risk factor for developing glaucoma. Having a first-degree relative (parent, sibling, or child) with glaucoma significantly elevates the risk of developing glaucoma, increasing it by approximately ninefold.[42] Moreover, the lifetime risk of glaucoma for first-degree relatives of glaucoma patients is around 22%, compared with just 2.3% for those without a family history of the condition.[42] Siblings of individuals with glaucoma have a 10.4% prevalence of the disease, in contrast to only 0.7% among siblings of those without glaucoma.[42] Farsightedness (hyperopia) and nearsightedness (myopia) are also important features in our model and are known risk factors for glaucoma, with myopia being more strongly associated with an increased risk of open-angle glaucoma.[43,44] People with myopia are approximately twice as likely to develop glaucoma compared with the general population and those with high myopia (optical prescription > -6.0 diopter) have a 14 times higher risk of developing glaucoma compared with the general population.[44] Type 2 diabetes was also found to be an important feature for the model, and this is a long-debated risk factor for glaucoma.[26,45] Interestingly, having cataract in your immediate family or yourself was also an important feature for the model, but it is not a known risk factor for most types of glaucoma. Patients with established eye care, such as those in the training population, may be simultaneously more likely to be diagnosed with cataract and glaucoma, yielding such an association. Both glaucoma and cataracts are also age-related, yielding some correlation between the 2 diseases.

There are several additional limitations to this study. The participants included were those with ≥ 2 eye-related diagnoses, meaning they would have seen an eye doctor for an examination and either been diagnosed with glaucoma or not. There is the potential for ascertainment bias, as individuals with conditions such as diabetes, cataract, or refractive error (e.g., myopia or hyperopia) are more likely to seek eye care. While these conditions may not all be direct risk factors for glaucoma, their presence increases

the likelihood of a person visiting an eye doctor, leading to the overrepresentation of such individuals in the dataset. They may also be more likely in the course of their eye care to be diagnosed with glaucoma, even though all participants in our cohort had eye care. Overall, the generalizability of the model to a true screening population where eye care is less frequent and conditions like these may be underrepresented remains to be studied. Furthermore, there was no ability to confirm the specialty of the providers submitting these diagnoses, which was another reason why we only included patients with $\geq$2 eye-related diagnoses, a stricter criterion than relying on a single diagnosis. We also could not include all participants in the All of Us study, as some individuals without a history of eye care could have undiagnosed glaucoma. This may affect the model's performance in a general population, especially as the prevalence of glaucoma in an eye care population is higher than would be expected in a true screening population. Future studies should test the model in a true screening population where all patients, regardless of prior eye care, are evaluated for glaucoma. In such a population, the prevalence of glaucoma would be low, and patients would not be aware of their glaucoma status; thus, the performance of the model under these conditions should be tested. Another important limitation is that the All of Us dataset lacks ophthalmic clinical data, imaging, and testing necessary for establishing a definitive glaucoma diagnosis clinically. Therefore, we relied on billing codes for diagnosing glaucoma, which may carry miscoding errors. It is possible that some participants with a glaucoma diagnosis may not have had the condition (i.e., they were glaucoma suspects). To help the model learn to accurately identify glaucoma risk, we excluded participants diagnosed only with borderline or suspect glaucoma. However, some glaucoma suspects may still have been included due to miscoding. Future validation studies should use data registries that include dedicated ophthalmic clinical data as additional steps in validating this model.

In summary, our research shows that machine and deep learning models can successfully analyze self-reported survey data—covering aspects such as lifestyle, general health, and family medical history—to identify individuals at high risk of glaucoma, independent of ophthalmic imaging or EHR data. Notably, our XGBoost model outperformed other methods, including deep learning approaches. Explainability studies confirmed that many key input features were clinically sensible. Models like ours, which pinpoint high-risk glaucoma patients using only self-reported survey data, could ultimately enable more targeted and efficient glaucoma screenings for those most at risk.

Footnotes and Disclosures

Originally received: September 26, 2024.
Final revision: December 10, 2024.
Accepted: December 13, 2024.
Available online: December 17, 2024. Manuscript no. XOPS-D-24-00397.

Department of Ophthalmology, Byers Eye Institute, Stanford University, Palo Alto, California.

Disclosure(s):

All authors have completed and submitted the ICMJE disclosures form.

The author(s) have no proprietary or commercial interest in any materials discussed in this article.

Financial Support was provided by the National Eye Institute K23EY03263501 (S.Y.W.), Career Development Award from Research to Prevent Blindness (S.Y.W.), unrestricted departmental grant from Research to Prevent Blindness (all authors), and departmental grant National Eye Institute P30-EY026877 (all authors). The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

HUMAN SUBJECTS: No human subjects were included in this study. This study, using deidentified data, was exempt from Stanford University Institutional Review Board approval and adheres to the tenets of the Declaration of Helsinki.

No animal subjects were used in this study.

Author Contributions:

Conception and design: Wang, Ravindranath, Naor

Data collection: Ravindranath

Analysis and interpretation: Wang, Ravindranath, Naor

Obtained funding: Wang

Overall responsibility: Wang, Ravindranath, Naor

Support for Open Access publication was provided by the Department of Ophthalmology, Stanford University.

Abbreviations and Acronyms:
AUROC = area under the receiver operating characteristic curve; **CI** = confidence interval; **EHR** = electronic health record; **FCN** = fully connected deep learning network.

Keywords:
Glaucoma screening, Machine learning, Deep learning, Electronic health records.

Correspondence:
Sophia Wang, MD, MS, Byers Eye Institute, Stanford University, 2370 Watson Ct, Palo Alto, CA 94303. E-mail: sywang@stanford.edu.

References

1. Resnikoff S, Pascolini D, Etya'ale D, et al. Global data on visual impairment in the year 2002. *Bull World Health Organ*. 2004;82:844–851.
2. Kingman S. Glaucoma is second leading cause of blindness globally. *Bull World Health Organ*. 2004;82:887–888.
3. Zhang N, Wang J, Li Y, Jiang B. Prevalence of primary open angle glaucoma in the last 20 years: a meta- analysis and systematic review. *Sci Rep*. 2021;11:13762.
4. Davuluru SS, Jess AT, Kim JSB, et al. Identifying, understanding, and addressing disparities in glaucoma care in the United States. *Transl Vis Sci Technol*. 2023;12:18.

5. Gupta P, Zhao D, Guallar E, et al. Prevalence of glaucoma in the United States: the 2005-2008 national health and nutrition examination survey. *Invest Ophthalmol Vis Sci*. 2016;57:2905−2913.
6. Shaikh Y, Yu F, Coleman AL. Burden of undetected and untreated glaucoma in the United States. *Am J Ophthalmol*. 2014;158:1121−1129.e1. Elsevier.
7. Rodriguez J, Sanchez R, Munoz B, et al. Causes of blindness and visual impairment in a population-based sample of U.S. Hispanics. *Ophthalmology*. 2002;109:737−743.
8. Giangiacomo A, Coleman AL. The epidemiology of glaucoma. In: Grehn F, Stamper R, eds. *Glaucoma*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2009:13−21.
9. Halawa OA, Jin Q, Pasquale LR, et al. Race and ethnicity differences in disease severity and visual field progression among glaucoma patients. *Am J Ophthalmol*. 2022;242:69−76.
10. Kang JH, Wang M, Frueh L, et al. Cohort study of race/ethnicity and incident primary open-angle glaucoma characterized by autonomously determined visual field loss patterns. *Transl Vis Sci Technol*. 2022;11:21.
11. Allison K, Patel DG, Greene L. Racial and ethnic disparities in primary open-angle glaucoma clinical trials: a systematic review and meta-analysis. *JAMA Netw Open*. 2021;4:e218348.
12. Tielsch JM, Sommer A, Katz J, et al. Racial variations in the prevalence of primary open-angle glaucoma. The Baltimore Eye Survey. *JAMA*. 1991;266:369−374.
13. Quigley HA, West SK, Rodriguez J, et al. The prevalence of glaucoma in a population-based study of Hispanic subjects: Proyecto VER. *Arch Ophthalmol*. 2001;119:1819−1826.
14. US Preventive Services Task Force, Mangione CM, Barry MJ, Nicholson WK. Screening for primary open-angle glaucoma: US preventive Services Task Force recommendation statement. *JAMA*. 2022;327:1992−1997.
15. Karvonen E, Stoor K, Luodonpää M, et al. Combined structure-function analysis in glaucoma screening. *Br J Ophthalmol*. 2022;106:1689−1695.
16. Tatemichi M, Nakano T, Tanaka K, et al. Performance of glaucoma mass screening with only a visual field test using frequency-doubling technology perimetry. *Am J Ophthalmol*. 2002;134:529−537.
17. Wilson MR, Khanna S. The value of different screening techniques for glaucoma. *Curr Opin Ophthalmol*. 1994;5:69−75.
18. Li G, Fansi AK, Harasymowycz P. Screening for glaucoma using GDx-VCC in a population with ≥1 risk factors. *Can J Ophthalmol*. 2013;48:279−285.
19. de Vries MM, Stoutenbeek R, Müskens RPHM, Jansonius NM. Glaucoma screening during regular optician visits: the feasibility and specificity of screening in real life. *Acta Ophthalmol*. 2012;90:115−121.
20. Guedes RAP, Guedes VMP, Chaoubah A. Focusing on patients at high-risk for glaucoma in Brazil: a pilot study. *J Fr Ophtalmol*. 2009;32:640−645.
21. Paudyal I, Yadav R, Parajuli A, et al. Screening of accompanying first degree relatives of patients with primary open angle glaucoma. *Nepal J Ophthalmol*. 2022;14:4−9.
22. Laroche D, Rickford K, Mike EV, et al. A novel, low-cost glaucoma calculator to identify glaucoma patients and stratify management. *J Ophthalmol*. 2022;2022:5288726.
23. All of Us Research Program Investigators, Denny JC, Rutter JL, Goldstein DB. The "all of us" research program. *N Engl J Med*. 2019;381:668−676.
24. Mapes BM, Foster CS, Kusnoor SV, et al. All of us research program. Diversity and inclusion for the all of us research program: a scoping review. *PLoS One*. 2020 Jul 1;15:e0234962.
25. OMOP CDM v5.3. https://ohdsi.github.io/CommonDataModel/cdm53.html. Accessed January 18, 2024.
26. Song BJ, Aiello LP, Pasquale LR. Presence and risk factors for glaucoma in patients with diabetes. *Curr Diab Rep*. 2016;16:124.
27. Kamboj A, Lause M, Kumar P. Ophthalmic manifestations of endocrine disorders-endocrinology and the eye. *Transl Pediatr*. 2017;6:286−299.
28. Reisner SL, Poteat T, Keatley J, et al. Global health burden and needs of transgender populations: a review. *Lancet*. 2016;388:412−436.
29. "Learn," scikit. https://scikit-learn.org/1.2/. Accessed February 2, 2024.
30. "Tensorflow," TensorFlow. https://www.tensorflow.org/. Accessed February 2, 2024.
31. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. In: Guyon I, Luxburg UV, Bengio S, et al., eds. *Advances in Neural Information Processing Systems*. 30. Curran Associates, Inc.; 2017.
32. Lundberg S. shap. Github. https://github.com/slundberg/shap. Accessed December 23, 2023.
33. Ogunyemi OI, Gandhi M, Lee M, et al. Detecting diabetic retinopathy through machine learning on electronic health record data from an urban, safety net healthcare system. *JAMIA Open*. 2021;4:ooab066.
34. "Related retinopathy risk test," Diabetes. https://diabetes.org/retinopathy-risk-test. Accessed December 18, 2023.
35. Sheth S, Lee P, Bajaj A, et al. Implementation of a machine-learning algorithm in the electronic health record for targeted screening for familial hypercholesterolemia: a quality improvement study. *Circ Cardiovasc Qual Outcomes*. 2021;14:e007641.
36. Anderson AE, Kerr WT, Thames A, et al. Electronic health record phenotyping improves detection and screening of type 2 diabetes in the general United States population: a cross-sectional, unselected, retrospective study. *J Biomed Inform*. 2016;60:162−168.
37. Shao Y, Zeng QT, Chen KK, et al. Detection of probable dementia cases in undiagnosed patients using structured and unstructured electronic health records. *BMC Med Inform Decis Mak*. 2019;19:128.
38. Wang L, Laurentiev J, Yang J, et al. Development and validation of a deep learning model for earlier detection of cognitive decline from clinical notes in electronic health records. *JAMA Netw Open*. 2021;4:e2135174.
39. Klein BE, Klein R, Sponsel WE, et al. Prevalence of glaucoma. The beaver dam eye study. *Ophthalmology*. 1992;99:1499−1504.
40. Dielemans I, Vingerling JR, Wolfs RC, et al. The prevalence of primary open-angle glaucoma in a population-based study in The Netherlands. The Rotterdam Study. *Ophthalmology*. 1994;101:1851−1855.
41. Coffey M, Reidy A, Wormald R, et al. Prevalence of glaucoma in the west of Ireland. *Br J Ophthalmol*. 1993;77:17−21.
42. McMonnies CW. Glaucoma history and risk factors. *J Optom*. 2017;10:71−78.
43. Chen SJ, Lu P, Zhang WF, Lu JH. High myopia as a risk factor in primary open angle glaucoma. *Int J Ophthalmol*. 2012;5:750−753.
44. Shen L, Melles RB, Metlapally R, et al. The association of refractive error with glaucoma in a multiethnic population. *Ophthalmology*. 2016;123:92−101.
45. Zhao YX, Chen XW. Diabetes and risk of glaucoma: systematic review and a Meta-analysis of prospective cohort studies. *Int J Ophthalmol*. 2017;10:1430−1435.