**Artificial Intelligence**

# A Framework for Automating Psychiatric Distress Screening in Ophthalmology Clinics Using an EHR-Derived AI Algorithm

**Samuel I. Berchuck[1], Alessandro A. Jammal[2], David Page[3], Tamara J. Somers[4], and Felipe A. Medeiros[2,3]**

[1] Department of Statistical Science, Duke University, Durham, NC, USA
[2] Duke Eye Center and Department of Ophthalmology, Duke University, Durham, NC, USA
[3] Department of Biostatistics and Bioinformatics, Duke University, Durham, NC, USA
[4] Department of Psychiatry and Behavioral Sciences, Duke University, Durham, NC, USA

**Purpose:** In patients with ophthalmic disorders, psychosocial risk factors play an important role in morbidity and mortality. Proper and early psychiatric screening can result in prompt intervention and mitigate its impact. Because screening is resource intensive, we developed a framework for automating screening using an electronic health record (EHR)-derived artificial intelligence (AI) algorithm.

**Methods:** Subjects came from the Duke Ophthalmic Registry, a retrospective EHR database for the Duke Eye Center. Inclusion criteria included at least two encounters and a minimum of 1 year of follow-up. Presence of distress was defined at the encounter level using a computable phenotype. Risk factors included available EHR history. At each encounter, risk factors were used to discriminate psychiatric status. Model performance was evaluated using area under the receiver operating characteristic (ROC) curve and area under the precision–recall curve (PR AUC). Variable importance was presented using odds ratios (ORs).

**Results:** Our cohort included 358,135 encounters from 40,326 patients with an average of nine encounters per patient over 4 years. The ROC and PR AUC were 0.91 and 0.55, respectively. Of the top 25 predictors, the majority were related to existing distress, but some indicated stressful conditions, including chemotherapy (OR = 1.36), esophageal disorders (OR = 1.31), central pain syndrome (OR = 1.25), and headaches (OR = 1.24).

**Conclusions:** Psychiatric distress in ophthalmology patients can be monitored passively using an AI algorithm trained on existing EHR data.

**Translational Relevance:** When paired with an effective referral and treatment program, such algorithms may improve health outcomes in ophthalmology.

## Introduction

In patients with ophthalmic disorders, psychosocial risk factors play an important role in morbidity and mortality.[1] The prevalence of psychiatric distress (i.e., anxiety and depression) in ophthalmic diseases is high; in studies of cataracts, glaucoma, diabetic retinopathy, and age-related macular degeneration (the leading causes of blindness and vision loss in the United States) the prevalence of psychiatric distress ranged from 5% to 57%.[2–5] Similar prevalence was noted for other common ophthalmic disorders such as dry eyes.[6] The presence of psychiatric distress in ophthalmic disorders is associated with worse medication and follow-up adherence,[7] disease comprehension,[8] and vision-related quality of life,[9] as well as increased morbidity[10] and health care costs.[11]

Proper and early screening of psychiatric distress can result in prompt intervention and can mitigate

1

negative outcomes.[12] However, traditional approaches to psychiatric screening present burdens related to cost and time requirements.[13] There are examples of practices and clinics that have implemented strategies that limit burdens, including routinely using brief self-report questionnaires. For example, oncology and cardiology clinics have tested a two-stage approach.[14] This approach begins with large-scale prescreening of patients using a brief self-report questionnaire[15] in oncology using the National Comprehensive Cancer Network (NCCN) distress thermometer[16] and in cardiology using the two-item patient health questionnaire.[17] Then, only patients who test positive on the prescreening instrument are screened further using a more formal assessment. This approach limits time and cost burdens, as only a subset of high-risk patients receive formal assessment.[18] Nonetheless, the two-stage approach is not widely used, due to patient reluctance, time consumption, and a lack of personnel to administer the questionnaires.

In recent years, there has been an increasing focus on automating screening for distress with the assistance of emerging technologies, including artificial intelligence (AI)-driven emotion recognition,[19] to alleviate challenges associated with routine screening.[20] These methods, however, require prospective data collection, which is not conducive to screening in clinics where resources are limited.[21] To overcome this limitation, we propose developing an automated prescreening measure of psychiatric distress that is based on available and existing electronic health records (EHR) data. EHR data have been used extensively to develop computable phenotypes of, and predictions for, incident medical disorders.[22,23] Identifying existing and incident cases of psychiatric distress is critical, as interventions can be tailored to improve patient distress attributed to vision-related diseases.[7]

In this study, we developed an automated AI algorithm to predict psychiatric distress among a large cohort of patients attending the Duke Eye Center, a tertiary referral center. We hypothesized that the AI algorithm would have high accuracy to identify distress, indicating that prescreening could be performed automatically at scale, with formal assessment reserved for a predetermined subset of high-risk patients.

## Methods

This was a retrospective cohort study using patients from the Duke Ophthalmic Registry, which consisted of adults at least 18 years of age who were evaluated at the Duke Eye Center or its satellite clinics from 2012 to 2021. The Duke University Institutional Review Board approved this study with a waiver of informed consent due to the retrospective nature of this work. All methods adhered to the tenets of the Declaration of Helsinki for research involving human subjects and were conducted in accordance with regulations of the Health Insurance Portability and Accountability Act.

Patients were included in the cohort if they had at least two encounters and at least 1 year of follow-up to the Duke Eye Center main site. Eligible encounters included any in which the patient was at least 18 years of age and that occurred between June 2013 and October 2021.

### Psychiatric Distress Outcome

Psychiatric distress (often shortened to distress throughout the manuscript) was defined as a binary indicator at the encounter level using an existing EHR phenotype from the Phenotype Knowledge-Base (PheKB) that defined depression and anxiety.[24] Psychiatric distress is one factor of the multifactorial psychosocial distress and was chosen as an outcome in this study because it can be reliably measured from EHR data. Versions of this algorithm have been shown to be associated with the nine-item patient health questionnaire, with area under the receiver operating characteristic (ROC) curve of 0.70 to 0.80.[25,26] Distress was defined at the encounter level, because anxiety and depression are not permanent conditions and can be recurrent.[27,28]

Distress was defined using International Classification of Diseases (ICD) diagnostic codes, medical history, and Current Procedural Terminology (CPT) procedure codes. The detailed definitions using these codes and medications are found in the pseudo-code for the phenotype available at PheKB. ICD codes for depression and anxiety can be found in Supplementary Tables S1 and S2, respectively. Medications used to identify depression and anxiety can be found in Supplementary Tables S3 and S4, which contain a list of generic and brand names of antidepressant and antianxiety medications. CPT codes are included in Supplementary Table S5 and included procedure codes for delivering psychotherapy.

For each encounter, distress was defined if an eligible diagnostic or procedure code or medication occurred within a window of 180 days around the encounter date. For a diagnostic code to indicate distress it had to occur on at least two distinct calendar days that are at least 30 days apart and not more
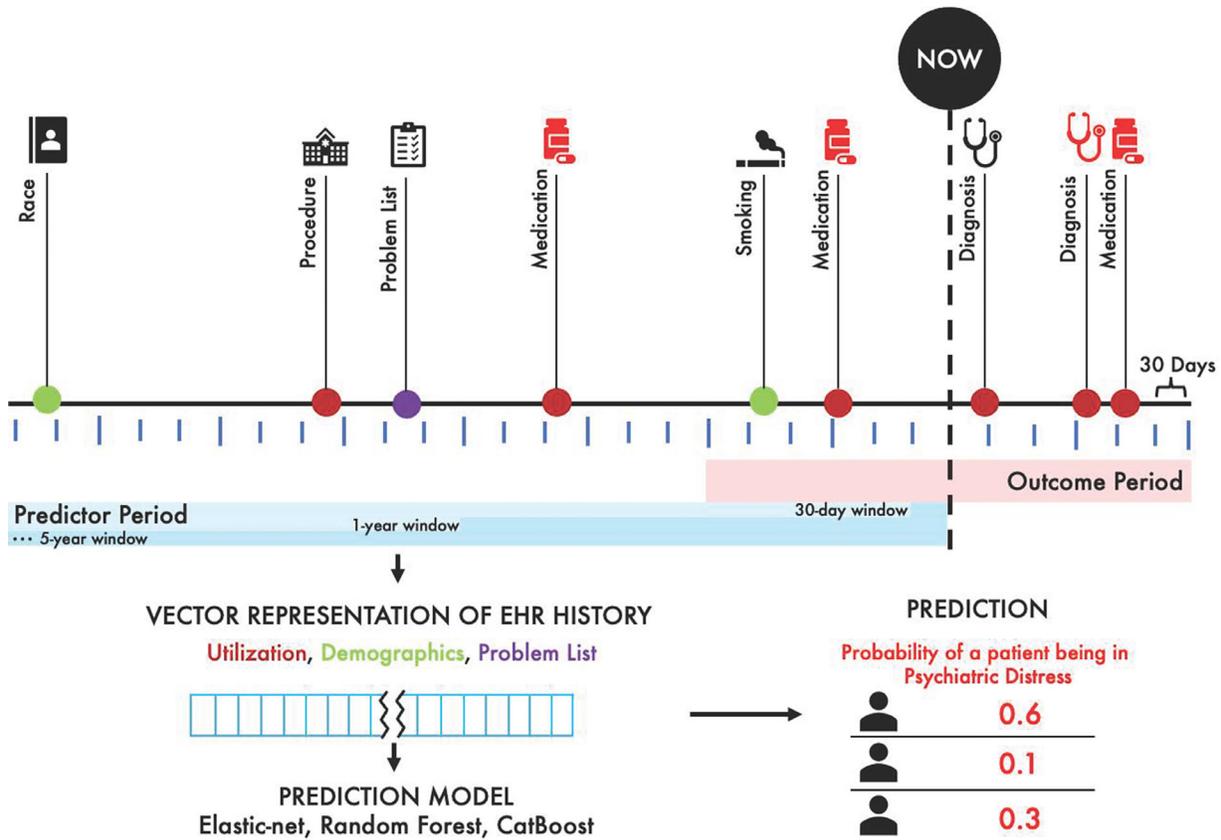
**Figure 1.** Visualizing the modeling framework. Both the predictors and outcome are defined based on the encounter date as an anchor. The outcome is defined using data collected in a 180-day window around the encounter (*pink area*). The predictor is defined using all EHR data collected prior to the encounter (*blue area*) and is broken into three phases of 3 months, 1 year, and 5 years. *Red EHR items* correspond to ones that qualify for the distress outcome phenotype (e.g., antidepressant medication). In this example, the patient has a diagnostic code and medication (both *red*) in the outcome period indicating that the patient had distress at the time of the encounter. Importantly, these occurred within 30 days of each other. The EHR history is converted into a vectorized form and fed into a machine learning algorithm (here, the elastic-net model). The algorithm then outputs a probability of distress for each encounter.

than 180 days apart. This is intended to avoid interpreting as an event "rule-out" codes that only appear in a patient's record once for a brief period (i.e., <30 days). The 180-day feature is intended to acknowledge that "rule-out" coding may appear more than once in a patient's medical record. This rule is stricter than the more common approach that only requires the presence of a single code. However, accounting for the nature of the EHR data is likely to lead to a higher positive predictive value. For a medication to indicate distress, it had to occur within 30 days of a corresponding ICD diagnostic code. For example, an anti-anxiety medication had to occur within 30 days of an ICD code indicating anxiety. There were no additional criteria for a CPT code to indicate distress, only that it occurred within 180 days of the encounter date. A visualization of the modeling framework can be found in Figure 1, and the flow chart in Figure 2 shows how patient distress was defined.

## Risk Factors

For each encounter, risk factors of distress were defined based on the available EHR history for that patient (i.e., any data available by the encounter). The risk factors were broken up into three groups: utilization, demographics, and problem list. The algorithm had 1840 variables as input.

### Utilization

Utilization contained predictors that quantify a patient's use of healthcare services and included diagnostic and procedure codes, medications, and clinical encounters. Diagnostic (ICD) and procedure (CPT) codes were grouped based on the Clinical Classifications Software (CCS) developed by the Agency for Healthcare Research and Quality.[29] All ICD codes were categorized using Version 9 ICD codes; thus, all Version 10 ICD codes were first mapped to Version
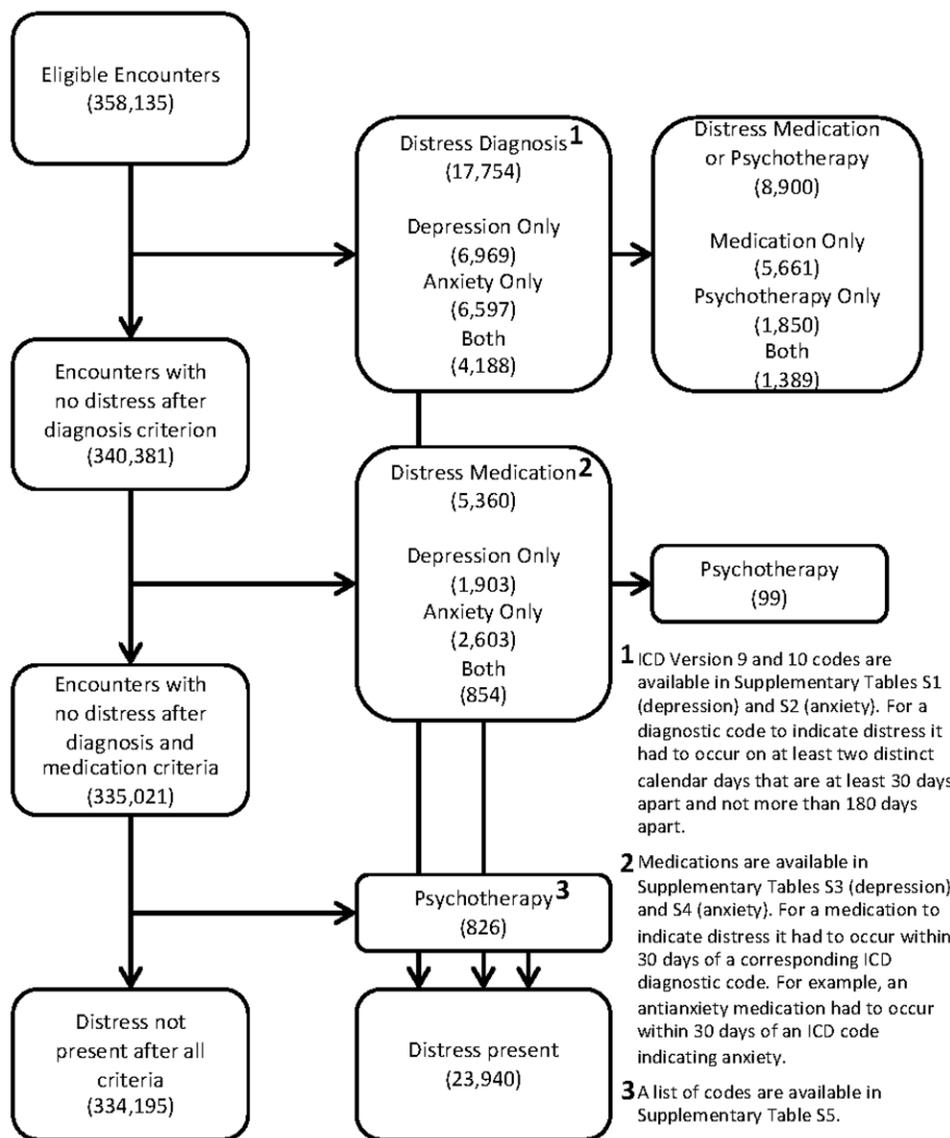
**Figure 2.** Flow chart demonstrating how patient distress was defined at the encounter level. Diagnosis and procedure codes come from the ICD.

9 using the general equivalence mappings (Centers for Medicare and Medicaid Services). There were 253 and 239 categories of diagnostic and procedure non-zero codes, respectively. Medications were grouped using the second level of the Anatomical Therapeutic Chemical (ATC) classification. This yielded 81 different drug subgroups. There were 35 distinct clinical encounter types.

For each encounter, utilization variables were coded as binary indicators of the variable occurring in a time window prior to the encounter (e.g., encounter to an oncology clinic). Three windows were used: within the past 3 months, 1 year, and 5 years. This yielded three binary variables for each utilization variable.

For example, consider a patient at an ophthalmology encounter who had an oncology encounter 2 years prior. This patient would have three variables representing their prior utilization of oncology clinics. The variables representing within the past 3 months and 1 year would be zero, indicating no utilization, and the variable representing the past 5 years would be one, representing utilization.

Because the definition of the outcome included a 180-day window around the encounter, variables that were identified in the 180-day window prior to the encounter and were used to define the outcome were not included when defining the predictors. This was done because the presence of these codes indicates a

deterministic relationship with the outcome. In this setting, our algorithm is not needed, and the patient can be assumed to be distressed. Codes that were removed include CCS diagnostic groups (adjustment disorders, alcohol-related disorders, anxiety disorders, disorders usually diagnosed in infancy, childhood, or adolescence, mood disorders, and substance-related disorders), CCS procedure group (psychological and psychiatric evaluation and therapy), and ATC subgroups (N06A, N05A, N05B). As an example, if an ICD code for anxiety showed up 90 days prior to the encounter, the CCS group anxiety disorders would be zeroed out for all three follow-up windows. This zeroing out applied to the presence of a single code, as opposed to the stricter rule defined in the outcome above. This was done to avoid allowing the AI algorithm to learn a near deterministic map. However, if there was an additional ICD code for anxiety at 181 days prior to the encounter, the anxiety variable would be one for both 1 and 5 years. We did not remove these variables that preceded the 180-day window, because previous distress is a predictor of future distress.

### Demographics

The demographic risk factors included age at the encounter (years), sex (male, female), race (Caucasian/white, African American/black, Asian, multiracial, other), ethnicity (non-Hispanic Latino, Hispanic Latino), marital status (married, single), income level, education, and binary behavior indicators of prior use of alcohol, smoking, and illicit drugs. The first level of the categorical variables was used as the reference category. Income level and education were obtained from U.S. Census Bureau's American Community Survey for 2006 to 2011. Income level was measured by per capita income in the past 12 months and was race specific. Education was measured by the percentage of residents who achieved a high-school education and was sex specific. Census data were assigned to patients based on the Zip Code in which they lived. In the models, age and education were scaled by 10, and income was scaled by 10,000. All three continuous predictors were mean centered. Of the demographic variables, race, ethnicity, marriage, income, and education had missing values at a rate less than 10%. These missing values were imputed using single mean imputation, and all patients were included in the final analysis.

### Problem List

Problem list items included any mention of depression (key words: depressed, depression, major depression) or anxiety (anxiety, obsessive compulsive disorder, panic attacks, panic disorders, post-traumatic stress disorder). Problem list items were again coded temporally, using 3 months, 1 year, and 5 years. For both anxiety and depression, an indicator was created to signal if at least one measure was present in each of the temporal ranges. Because problem list items were not used to define the PheKB outcome, we did not remove any problem list items within 180 days of the encounter.

### Training the Model

To predict psychiatric distress using the EHR history upon encounter at the Duke Eye Center, we used three machine learning classification models: elastic-net,[30] Random Forest,[31] and CatBoost.[32] Elastic-net is a regularized linear model that penalizes overfitting by shrinking the regression coefficients toward zero. The log-likelihood function to be minimized is

$$-\frac{1}{n} \sum_{i=1}^{n} y_i(\beta_0 + x_i^T \boldsymbol{\beta}) - \log(1 + e^{\beta_0 + x_i^T \boldsymbol{\beta}})$$
$$+ \lambda[(1-\alpha)\|\boldsymbol{\beta}\|_2^2/2 + \alpha\|\boldsymbol{\beta}\|_1],$$

where $x_i$ is a $p$-dimensional vector of the EHR history; $y_i$ is an indicator of distress for encounter $i = 1, \ldots, n$; $\boldsymbol{\beta}$ is a vector of regression coefficients; and $\beta_0$ is an intercept. The penalty term $\lambda$ represents the degree of penalization, and the elastic-net term $\alpha$ bridges the gap between a least absolute shrinkage and selection operator (LASSO; $\alpha = 1$) and Ridge regression ($\alpha = 0$). It is known that Ridge regression shrinks coefficients of correlated predictors toward each other, whereas LASSO tends to pick one and zero-out the others. The potential weakness of a linear model is that variables may interact with one another in ways we do not know beforehand. Decision trees are often a successful method in such situations, but they can have overfitting problems; the risk of overfitting is reduced by ensemble methods such as random forests or gradient boosting, both of which we also employ. The Random Forest algorithm uses an ensemble of decision trees to make robust predictions, where the output of the Random Forest algorithm is given by the class selected by the most trees. CatBoost is a gradient boosting algorithm for binary classification trees that uses ordered boosting to overcome overfitting and allows for categorical features to be handled natively.

Although ridge regression can be accomplished by traditional gradient descent, the other (absolute value) penalty term in the elastic-net algorithm means we instead must use a different method; the most efficient currently is cyclical coordinate descent, which

we performed in glmnet.[33] To both tune the pair ($\lambda$, $\alpha$) and estimate model accuracy in an unbiased manner, we used nested cross-validation, with the inner loop tuning ($\lambda$, $\alpha$) via a grid search to minimize internal cross-validation error minus one standard error,[34] and with the outer loop estimating model performance including area under the curve (AUC) for both ROC and precision–recall (PR) curves. The Random Forest algorithm was run with 500 trees and one random split for each candidate splitting variable. CatBoost was run with default parameters using the cross-entropy loss. All three models were implemented using 10-fold cross-validation. An additional test dataset was not used. To prevent data leakage, sampling was performed at the patient level, and the percent of distress was balanced across folds.

## Statistical Analysis

The overall performance of the models was evaluated using the ROC and PR AUC, estimated by cross-validation as in the preceding paragraph. All performance metrics are presented as a mean and standard deviation (SD) across the 10 cross-validation folds. PR curves plot precision (i.e., positive predictive value) against recall (i.e., sensitivity) and are useful when there is imbalance in the cases and controls. A PR curve depends only on predictions of the minority class, as precision and recall do not depend on true negatives.

These summaries are presented overall and for subgroups, including within each subspecialty at the Duke Eye Center, diseases including primary open-angle glaucoma (POAG), diabetic retinopathy, age-related macular degeneration (AMD), and cataracts, and demographics. Disease diagnoses were based on ICD diagnostic code definitions from previous studies[35–37] and had to occur within 30 days of a corresponding clinic encounter (POAG from the glaucoma clinic, AMD and diabetic retinopathy from the vitreous retinal clinic, cataracts from any clinic). Finally, to determine the importance of the time interval for prediction (i.e., 3 months, 1 year, 5 years), we examined overall performance for models that included only data from each time interval.

Additionally, sensitivity is presented across varying levels of specificity. Sensitivity is presented for all cases of distress and a subset of cases determined to be incident or new. A new case was required to be the first case present for a patient and to have no prior encounters to a psychiatry clinic. Furthermore, the patient could have no problem list, medication, or procedure items prior to the encounter that suggested any psychiatric diagnoses. We highlight sensitivity at a specificity

of 70%. This value comes from published data from a two-stage screening approach, where we used the proportion of patients referred from the prescreening questionnaire who ended up not having distress in a more formal evaluation (i.e., specificity).[14,15] Finally, for the elastic-net model, the largest 25 coefficients in absolute value were presented, along with their odds ratios (ORs). The ORs were averaged across the 10 cross-validation folds. The non-zero demographic predictors were also presented. OR *P* values were not presented, as they cannot be computed reliably for the elastic-net model.[38]

The summaries for the cohort are presented with continuous variables presented as mean and standard deviation and categorical variables as counts and percentages. Hypothesis tests are presented across distress group; categorical variables were tested using a $\chi^2$ test, and continuous variables were tested using a Wilcoxon rank-sum test. Patient data were anonymized, and all statistical analyses were conducted using R 4.0.5 (R Foundation for Statistical Computing, Vienna, Austria) within the Protected Analytics Computing Environment (PACE). PACE is a secure virtual network space developed by Duke University for the analysis of identifiable protected health information. The R packages glmnet, ranger, and catboost were used to carry out the models.[30,39,40]

## Results

The study cohort consisted of 358,135 encounters from 40,326 patients with an average $\pm$ SD of $9 \pm 10$ encounters per patient over $4 \pm 2$ years of follow-up. The average age of the patients was $60 \pm 17$ years, with a breakdown of 23,762 (59%) females, 27,323 (68%) Caucasian/white, 10,573 (26%) African American/black, and the rest Asian, multiracial, and other races. There were 6069 (15%) patients with at least one encounter with corresponding distress. Full summary details at the patient level can be found in Table 1. Encounter level summaries can be found in Table 2, with the top seven predictors by base rate presented for each utilization category, along with problem list items.

The optimal tuning parameters in the elastic-net model were found to be $\alpha = 0.08 \pm 0.03$ and $\lambda = 0.04 \pm 0.01$, indicating a preference for Ridge regression. The original number of predictors included was 1840, and after regularization only 292 remained that were non-zero in at least one cross-validation fold.

The ROC and PR curves for the three machine learning algorithms are presented in Figure 3. The

**Table 1.** Summary of Demographics Presented Across Patient Distress Indicators

| Variable | All | Distress | Other | *P* |
|---|---|---|---|---|
| Sample size, *n* (%) | 40,326 (100) | 6069 (15) | 34,257 (85) | — |
| Number of encounters | | | | <0.001 |
|    Mean ± SD | 8.88 ± 9.85 | 9.90 ± 11.15 | 8.70 ± 9.59 | |
|    Median (min, max) | 6 (2, 134) | 6 (2, 133) | 6 (2, 134) | |
| Follow-up (y), mean ± SD | 3.83 ± 2.09 | 4.15 ± 2.08 | 3.77 ± 2.08 | <0.001 |
| Age at first encounter (y), mean ± SD | 60.17 ± 16.69 | 60.3 ± 15.79 | 60.15 ± 16.84 | 0.489 |
| Gender (female), *n* (%) | 23,762 (59) | 4326 (71) | 19,436 (57) | <0.001 |
| Race, *n* (%) | | | | <0.001 |
|    Caucasian/white | 27,323 (68) | 4238 (70) | 23,085 (67) | |
|    African American/black | 10,573 (26) | 1575 (26) | 8998 (26) | |
|    Asian | 1346 (3) | 103 (2) | 1243 (4) | |
|    Multiracial | 410 (1) | 45 (1) | 365 (1) | |
|    Other | 674 (2) | 108 (2) | 566 (2) | |
| Ethnicity (Hispanic/Latino), *n* (%) | 1160 (3) | 169 (3) | 991 (3) | 0.672 |
| Marital status (single), *n* (%) | 16,775 (42) | 3117 (51) | 13,658 (40) | <0.001 |
| Alcohol use, *n* (%) | 19,813 (49) | 3376 (56) | 16,437 (48) | <0.001 |
| Smoking use, *n* (%) | 16,528 (41) | 3011 (50) | 13,517 (39) | <0.001 |
| Illicit drug use, *n* (%) | 1276 (3) | 420 (7) | 856 (2) | <0.001 |
| Annual income ($1000), mean ± SD | 32.98 ± 17.95 | 32.66 ± 17.17 | 33.04 ± 18.09 | 0.747 |
| Education (%), mean ± SD | 0.87 ± 0.13 | 0.87 ± 0.13 | 0.87 ± 0.13 | 0.287 |

A patient was defined as having psychosocial distress if they had at least one distress encounter. The only temporally varying variables are alcohol, smoking, and illicit drug use, which are taken to be any use across the entire EHR history and follow-up. *P* values represent hypothesis tests across distress group, with categorical variables tested using a $\chi^2$ test and continuous variables tested using a Wilcoxon rank-sum test.

intervals correspond to 95% cross-validation confidence intervals. The mean ± SD ROC AUCs for elastic-net, CatBoost, and Random Forest were 0.912 ± 0.007, 0.918 ± 0.007, and 0.913 ± 0.007, respectively, with PR AUCs of 0.547 ± 0.032, 0.575 ± 0.031, and 0.552 ± 0.033. For a PR curve, a non-informative classifier would yield an AUC equal to the prevalence of distress in the population, 7% at the encounter level. The improvements from CatBoost and Random Forest were minimal compared to elastic-net and within the range of cross-validation error. Because the elastic-net model was comparable in terms of performance with the more complex algorithms and yields interpretable feature importance values as OR, the remaining results are presented using the elastic-net model.

Table 3 includes the ROC and PR AUCs across subspecialty, diseases, and demographics, along with the base rate and prevalence of distress within each subgroup. AUC performances ranged from 0.87 to 0.94 for ROC curves and 0.52 to 0.63 for PR curves. The ROC and PR AUCs for neuro-ophthalmology, the subspecialty with the highest level of distress at 12.2%, were 0.89 and 0.60, respectively. For POAG and AMD (the diseases with the highest rates of distress at 7.4% and 7.3%, respectively), the ROC and PR AUCs were 0.91 and 0.90 and 0.57 and 0.56, respectively. Finally, Supplementary Figure S1 and Table S6 present results of the elastic-net model with only predictors from 3 months, 1 year, and 5 years prior to the encounter.

Figure 4 presents the sensitivity values for existing and new distress across a continuum of specificity values. At a specificity level of 0.70, the sensitivity values were 0.92 ± 0.01 and 0.71 ± 0.06, respectively, for existing and new distress. Table 4 presents the top 25 predictors of distress from the elastic-net model. The full list of non-zero predictors is given in Supplementary Table S7. Finally, in Table 5, we present the non-zero coefficients for the demographic and risky-behavior predictors. In Table 5 and Supplementary Table S7, ORs that rounded to 1.00 have an additional column indicating the direction of the association.

## Discussion

In this study, we introduced an AI algorithm that automates prescreening of psychiatric distress

**Table 2.** Summary of Utilization and Problem Lists Calculated Using the Entire EHR History Presented Across Encounter Distress Indicators

| Variable | All, n (%) | Distress, n (%) | Other, n (%) |
|---|---|---|---|
| Sample size | 358,135 (100) | 23,940 (7) | 334,195 (93) |
| CCS diagnostic groups | | | |
|   Other eye disorders | 246,823 (69) | 17,744 (74) | 229,079 (69) |
|   Other aftercare | 245,736 (69) | 21,974 (92) | 223,762 (67) |
|   Retinal detachments, defects, vascular occlusion, and retinopathy | 181,054 (51) | 11,925 (50) | 169,129 (51) |
|   Cataract | 177,833 (50) | 12,613 (53) | 165,220 (49) |
|   Residual codes, unclassified | 169,926 (47) | 19,795 (83) | 150,131 (45) |
|   Glaucoma | 146,924 (41) | 8,781 (37) | 138,143 (41) |
|   Other screening for suspected conditions (not mental disorders or infectious disease) | 130,440 (36) | 15,714 (66) | 114,726 (34) |
| CCS procedure groups | | | |
|   Ophthalmologic and otologic diagnosis and treatment | 334,998 (94) | 22,536 (94) | 312,462 (93) |
|   Other diagnostic procedures (interview, evaluation, consultation) | 325,570 (91) | 23,710 (99) | 301,860 (90) |
|   Other therapeutic procedures | 285,630 (80) | 23,221 (97) | 262,409 (79) |
|   Laboratory—chemistry and hematology | 249,917 (70) | 23,081 (96) | 226,836 (68) |
|   Other laboratory | 201,153 (56) | 21,777 (91) | 179,376 (54) |
|   Microscopic examination (bacterial smear, culture, toxicology) | 173,743 (49) | 20,418 (85) | 153,325 (46) |
|   Anesthesia | 157,886 (44) | 13,652 (57) | 144,234 (43) |
| ATC drug groups | | | |
|   Ophthalmologicals | 210,847 (59) | 17,299 (72) | 193,548 (58) |
|   Nasal preparations | 176,427 (49) | 15,274 (64) | 161,153 (48) |
|   Antibacterials for systemic use | 165,231 (46) | 15,970 (67) | 149,261 (45) |
|   Analgesics | 128,451 (36) | 14,443 (60) | 114,008 (34) |
|   Otologicals | 122,304 (34) | 12,784 (53) | 109,520 (33) |
|   Stomatological preparations | 117,248 (33) | 13,919 (58) | 103,329 (31) |
|   Corticosteroids, dermatological preparations | 109,898 (31) | 13,046 (54) | 96,852 (29) |
| Encounters | | | |
|   Ophthalmology | 282,412 (79) | 19,270 (80) | 263,142 (79) |
|   General surgery | 149,017 (42) | 13,361 (56) | 135,656 (41) |
|   General medicine | 130,842 (37) | 16,804 (70) | 114,038 (34) |
|   Radiology | 126,622 (35) | 15,244 (64) | 111,378 (33) |
|   Lab | 90,576 (25) | 11,741 (49) | 78,835 (24) |
|   Orthopedics | 83,819 (23) | 11,510 (48) | 72,309 (22) |
|   Emergency medicine | 81,827 (23) | 11,822 (49) | 70,005 (21) |
| Problem list | | | |
|   Anxiety | 14,412 (4) | 5145 (21) | 9267 (3) |
|   Depression | 14,230 (4) | 5062 (21) | 9168 (3) |

The top seven variables are presented for each utilization type and are ranked by their proportion in the entire sample size.

using existing EHR data. Our findings suggest that prescreening of distress can be accomplished at scale, eliminating previous hurdles to scalability in the two-stage approach including patient reluctance, time consumption, and a lack of personnel to administer the questionnaires. This finding is particularly important in an ophthalmology setting, where patients have high levels of distress, yet there is no existing infrastructure for distress screening.

In our study, 15% of patients had at least one encounter with distress. This value is consistent with previously reported prevalence of anxiety and depression in patients with ophthalmic disorders, which ranged from 5% to 57%.[2–5] Our prevalence is likely on the lower end, as our inclusion criteria did not include a disease diagnosis. At the encounter level, the prevalence of distress was 7%, with neuro-ophthalmology having the highest rate of distress at 12.2%. This finding
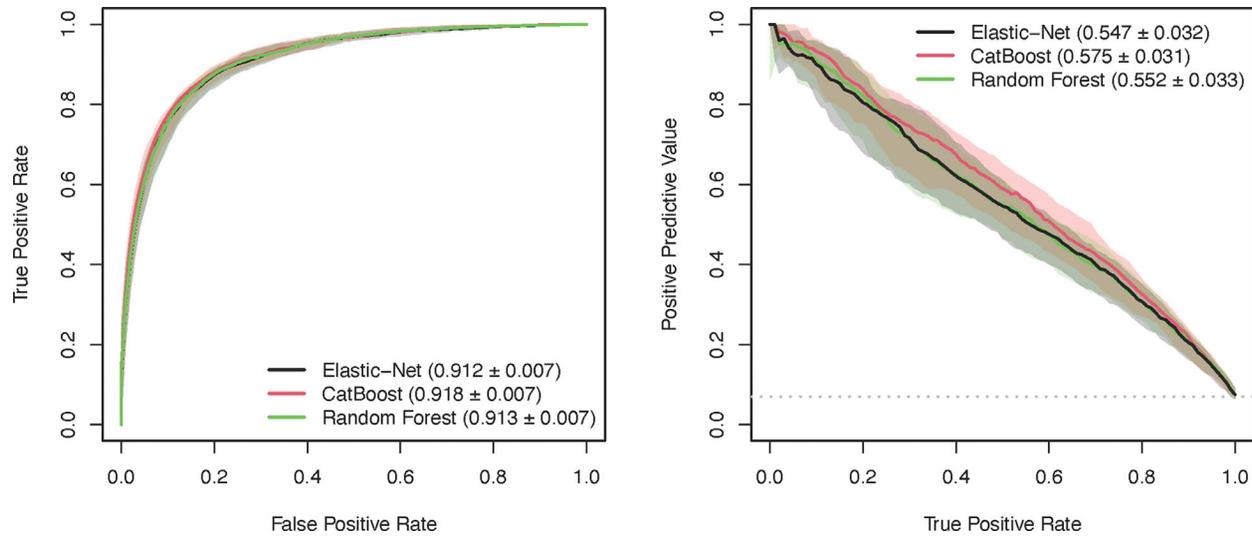
**Figure 3.** ROC and PR curves for the elastic-net, CatBoost, and Random Forest algorithms. In parentheses are mean ± SD for ROC and PR AUCs across cross-validation folds. Intervals represent 95% cross-validation confidence intervals. The *horizontal line* on the PR curve represents the prevalence of distress across encounters (7%).

is consistent with literature that indicates the overlap of neuro-ophthalmology and psychiatric conditions.[41] Of the diseases we included in our study, distress was highest among patients with POAG (7.4%) and AMD (7.3%). This is consistent with previous findings, as patients with POAG and AMD are at higher risk for anxiety and depression.[42,43]

Our algorithm had high classification accuracy, with ROC and PR AUCs of 0.912 ± 0.007 and 0.547 ± 0.032, respectively. The performance was consistent across varying subspecialties, diseases, and demographics. Of note, for the high-distress subgroups (i.e., neuro-ophthalmology, POAG, and AMD), the PR AUCs are on the upper end of performance with values of 0.60 ± 0.06, 0.57 ± 0.06, and 0.56 ± 0.15, respectively.

We also presented sensitivity values at a fixed specificity of 0.70, which has meaning based in literature comparing brief pre-screening surveys to a gold-standard psychiatric assessment of distress. For example, Cull et al.[14] reported a sensitivity and specificity of 0.85 and 0.71, respectively, when comparing the Hospital Anxiety and Depression Scale to a gold-standard psychiatric interview in oncology patients. Another study, using a meta-analysis, found a sensitivity of 0.81 (0.79–0.82) at a specificity of 0.72 when comparing the NCCN Distress Thermometer to the Hospital Anxiety and Depression Scale.[15] Neither study distinguished between existing and new distress. In our study, at a specificity level of 0.70 the sensitivity values were 0.92 ± 0.01 and 0.71 ± 0.06, respectively, for existing and new distress. These results are promising and indicate that our modeling framework

may be able to replicate the operating characteristics of existing pre-screening surveys for both existing and new distress.

In ophthalmology clinics, identifying both patients who have already been treated for distress and those with new distress is an important task. For patients with existing distress who have already been treated, the algorithm can be viewed as a computable phenotype that collects existing EHR data and returns a summary statistic that represents level of distress. If the PheKB phenotype is a valid measure of distress, this viewpoint should hold. This is important in the context of ophthalmic disorders, as there is evidence that interventions can be tailored to specific eye disorders to improve well-being. For example, a recent study in visually impaired glaucoma patients demonstrated that a social work intervention decreased distress.[7] This intervention provided support for these patients that was tailored to eye-related distress, including procuring closed-circuit televisions. Thus, the algorithm removes any barriers to identifying distressed patients during routine care and sets up a system where distressed patients can be referred to an intervention that is tailored to patients with eye disorders.

This is particularly impactful in an ophthalmology context, where resources and priorities do not permit screening for distress, even if it is already present in the EHR. Our algorithm performs well in this setting, with a sensitivity of 0.92 at a specificity of 0.70. For patients not currently being treated for distress, the algorithm can be viewed as a prediction model that identifies incident distress. At a specificity level of 0.70,

**Table 3.** Performance Metrics Presented Across Subgroups

| Metric | Encounters, *n* (%) | Distress, *n* (%) | ROC Curve, Mean ± SE | PR Curve, Mean ± SE |
|---|---|---|---|---|
| All | 358,135 (100) | 23,940 (6.7) | 0.91 ± 0.01 | 0.55 ± 0.03 |
| Subspecialty at Duke Eye Center | | | | |
|   Comprehensive/general | 50,935 (14.2) | 4408 (8.7) | 0.90 ± 0.01 | 0.56 ± 0.03 |
|   Cornea | 37,911 (10.6) | 2213 (5.8) | 0.92 ± 0.03 | 0.55 ± 0.10 |
|   Glaucoma | 80,782 (22.6) | 4532 (5.6) | 0.91 ± 0.02 | 0.52 ± 0.05 |
|   Low vision | 5315 (1.5) | 416 (7.8) | 0.93 ± 0.03 | 0.59 ± 0.11 |
|   Neurology | 9587 (2.7) | 1168 (12.2) | 0.89 ± 0.03 | 0.60 ± 0.06 |
|   Ocular immunology | 4811 (1.3) | 338 (7.0) | 0.94 ± 0.04 | 0.60 ± 0.20 |
|   Oculoplastics oncology | 5461 (1.5) | 386 (7.1) | 0.91 ± 0.03 | 0.53 ± 0.13 |
|   Ophthalmology equipment | 1619 (0.5) | 137 (8.5) | 0.87 ± 0.08 | 0.56 ± 0.19 |
|   Pediatrics | 6710 (1.9) | 371 (5.5) | 0.94 ± 0.02 | 0.56 ± 0.14 |
|   Surgical comprehensive | 13,404 (3.7) | 1241 (9.3) | 0.89 ± 0.02 | 0.55 ± 0.06 |
|   Vision correction | 535 (0.1) | 43 (8.0) | 0.89 ± 0.11 | 0.63 ± 0.31 |
|   Vision rehab performance | 5115 (1.4) | 374 (7.3) | 0.90 ± 0.04 | 0.54 ± 0.07 |
|   Vitreous retinal | 135,950 (38.0) | 8313 (6.1) | 0.91 ± 0.01 | 0.54 ± 0.05 |
| Disease | | | | |
|   POAG | 43,233 (12.1) | 3718 (7.4) | 0.91 ± 0.02 | 0.57 ± 0.06 |
|   Diabetic retinopathy | 8032 (2.2) | 552 (5.8) | 0.90 ± 0.05 | 0.46 ± 0.13 |
|   AMD | 18,270 (5.1) | 1426 (7.3) | 0.90 ± 0.03 | 0.56 ± 0.15 |
|   Cataracts | 30,831 (8.6) | 2378 (6.4) | 0.90 ± 0.03 | 0.48 ± 0.10 |
| Demographics | | | | |
|   Male | | | | |
|     Caucasian/white | 103,604 (28.9) | 6960 (6.8) | 0.91 ± 0.02 | 0.56 ± 0.05 |
|     African American/black | 36,840 (10.3) | 2498 (6.9) | 0.91 ± 0.02 | 0.51 ± 0.06 |
|     Other | 8020 (2.2) | 572 (7.2) | 0.92 ± 0.03 | 0.57 ± 0.14 |
|   Female | | | | |
|     Caucasian/white | 141,416 (39.5) | 9389 (6.7) | 0.91 ± 0.01 | 0.54 ± 0.05 |
|     African American/black | 59,058 (16.5) | 3952 (6.3) | 0.92 ± 0.01 | 0.56 ± 0.04 |
|     Other | 9197 (2.6) | 569 (6.3) | 0.91 ± 0.03 | 0.52 ± 0.17 |
| Age (y) | | | | |
|   <40 | 35,033 (9.8) | 2090 (6.3) | 0.91 ± 0.02 | 0.53 ± 0.09 |
|   40–50 | 29,693 (8.3) | 2029 (7.0) | 0.91 ± 0.02 | 0.53 ± 0.10 |
|   50–60 | 54,771 (15.3) | 3429 (6.8) | 0.91 ± 0.02 | 0.53 ± 0.10 |
|   60–70 | 92,222 (25.8) | 6400 (6.7) | 0.92 ± 0.01 | 0.55 ± 0.05 |
|   70–80 | 90,435 (25.3) | 5930 (6.6) | 0.90 ± 0.01 | 0.53 ± 0.07 |
|   >80 | 55,981 (15.6) | 4062 (6.7) | 0.92 ± 0.02 | 0.57 ± 0.07 |

Performance metrics include AUCs for ROC and PR. The mean across cross-validation folds is presented, along with standard errors. The prevalence of distress is also presented within each subgroup.

our model has a sensitivity of 0.83 for new cases, indicating that the performance is still adequate in patients without psychiatric indicators in their EHR history. This is a particularly impactful finding because our modeling framework uses EHR data for both the predictors and outcome. Therefore, based on the design of our model, it follows that we can identify existing distress at a high rate. The fact that we can identify incident distress with no EHR history indicates that the model can be used more generally to predict distress. This is further illuminated in the non-zero variables in the elastic-net model.

Of the top 25 predictors in Table 4, we see that the majority are related to existing distress, including the first variable, an encounter to a psychiatry clinic within the past 3 months. This reinforces that our model performs well for existing cases of distress. Also present, however, are variables that are
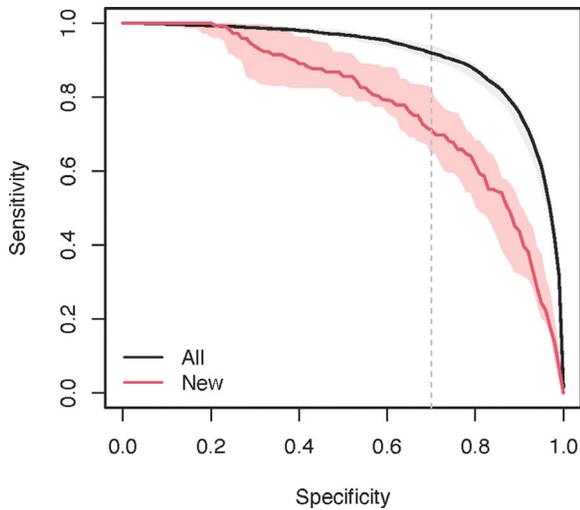
**Figure 4.** Sensitivity values for existing and new distress across a continuum of specificity values. New distress is defined as any distress encounter that was the first distress encounter for each patient and was not preceded by an encounter to a psychiatry clinic. The *vertical line* represents 70% specificity, which we used to compare our results to previous studies.

associated with distress but are not direct indicators of healthcare utilization related to distress, including suffocation and strangulation (OR = 1.88); self-inflicted injuries, including attempt of suicide (1.58); chemotherapy (1.36); esophageal disorders, including esophagitis (1.31); other nervous system disorders, including central pain syndrome (1.25); and headaches and migraines (1.24). The presence of these variables is more evidence that our model is not simply identifying existing characteristic healthcare utilization for distress.

Although our model performs adequately for incident distress, we could likely improve the performance by including diagnostic tests, including clinical measures of disease severity. We did not do this initially, because we restricted ourselves to data input by clinicians and billing codes from Duke University's EHR system Epic. In the future, we will expand our algorithm to include these diagnostic test data, including imaging data. These measures are not readily available and must be extracted from individual instru-

**Table 4.** ORs for the Top 25 Predictors of Distress Using All Predictor Types With Variables Ordered by the Absolute Value of Their Coefficient

| Variable | Distress, n (%) | Other, n (%) | OR |
|---|---|---|---|
| Intercept | — | — | 0.01 |
| Enc: Psychiatry (3 mo) | 2730 (11.40) | 473 (0.14) | 3.71 |
| Dx: Adjustment disorders (1 y) | 722 (3.02) | 484 (0.14) | 2.06 |
| Problem list: Anxiety (3 mo) | 453 (1.89) | 355 (0.11) | 2.00 |
| Dx: Anxiety disorders (1 y) | 7184 (30.01) | 9073 (2.71) | 1.94 |
| Problem list: Depression (1 y) | 1576 (6.58) | 1680 (0.50) | 1.91 |
| Dx: E Codes: Suffocation (1 y) | 13 (0.05) | 14 (0.00) | 1.88 |
| Problem list: Depression (3 mo) | 453 (1.89) | 330 (0.10) | 1.81 |
| Dx: Mood disorders (1 y) | 6926 (28.93) | 7477 (2.24) | 1.79 |
| Problem list: Anxiety (1 y) | 1586 (6.62) | 1674 (0.50) | 1.77 |
| Problem list: Depression (5 y) | 5145 (21.49) | 9267 (2.77) | 1.69 |
| Enc: Psychiatry (1 y) | 3874 (16.18) | 1664 (0.50) | 1.67 |
| Proc: Psychological and psychiatric evaluation and therapy (5 y) | 7465 (31.18) | 12,436 (3.72) | 1.63 |
| Dx: Attention-deficit, conduct, and disruptive behavior disorders (3 mo) | 314 (1.31) | 340 (0.10) | 1.59 |
| Dx: Suicide and intentional self-inflicted injury (3 mo) | 155 (0.65) | 39 (0.01) | 1.58 |
| Dx: Anxiety disorders (5 y) | 12,597 (52.62) | 33,728 (10.09) | 1.57 |
| Problem list: Anxiety (5 y) | 5062 (21.14) | 9168 (2.74) | 1.50 |
| Dx: Mood disorders (5 y) | 10,431 (43.57) | 21,587 (6.46) | 1.47 |
| Meds: Psychoanaleptics (5 y) | 12,278 (51.29) | 39,555 (11.84) | 1.45 |
| Dx: Miscellaneous mental health disorders (3 mo) | 823 (3.44) | 1060 (0.32) | 1.43 |
| Meds: Psychoanaleptics (1 y) | 4283 (17.89) | 7928 (2.37) | 1.40 |
| Dx: Maintenance chemotherapy, radiotherapy (3 mo) | 579 (2.42) | 1417 (0.42) | 1.36 |
| Dx: Esophageal disorders (3 mo) | 4390 (18.34) | 13,267 (3.97) | 1.31 |
| Dx: Other nervous system disorders (3 mo) | 6107 (25.51) | 18,381 (5.50) | 1.25 |
| Dx: Residual codes, unclassified (3 mo) | 8841 (36.93) | 33,193 (9.93) | 1.25 |
| Dx: Headache, including migraine (3 mo) | 1264 (5.28) | 2674 (0.80) | 1.24 |

Enc, encounter type; Dx, diagnosis group; Proc, procedure group; Meds, medication group.

**Table 5.** Demographic Predictors With Non-Zero Coefficients

| Variable | Importance | Distress | Other | OR | Direction of Association[a] |
|---|---|---|---|---|---|
| Race (African American/Black), *n* (%) | 49 | 5713 (23.86) | 90,185 (26.99) | 0.90 | |
| Illicit drug use (yes), *n* (%) | 90 | 1530 (6.39) | 8488 (2.54) | 1.04 | |
| Gender (female), *n* (%) | 91 | 16,975 (70.91) | 192,696 (57.66) | 1.04 | |
| Marriage (single), *n* (%) | 184 | 12,452 (52.01) | 137,734 (41.21) | 1.01 | |
| Alcohol use (yes), *n* (%) | 185 | 12,782 (53.39) | 154,320 (46.18) | 1.01 | |
| Smoking use (yes), *n* (%) | 225 | 12,166 (50.82) | 144,757 (43.32) | 1.00 | + |
| Age (per 10 y), mean ± SD | 229 | 63.99 ± 15.86 | 64.68 ± 16.22 | 1.00 | − |
| Race (Asian), *n* (%) | 251 | 303 (1.27) | 9238 (2.76) | 1.00 | − |

Presented are ORs, importance rankings, and summaries across encounter type (distress vs. other).
[a]The direction of the association for ORs that rounded to 1.00.

ments. The current model can be applied in a more general EHR system without additional data collection and curation. Furthermore, our patients may have been receiving psychiatric care outside the Duke University Health System that we did not have access to when defining our distress outcome. We tried to minimize this by limiting our patient population to those attending the Duke Eye Center main site, which is in Durham, NC, where the Duke University main hospital and majority of outpatient clinics are located. To fully overcome this limitation, a formal external validation of our AI algorithm should be performed using a gold-standard assessment of distress. Finally, prior to this algorithm being employed outside of Duke, it should be trained with data from multiple health centers to permit generalizability.

In our study, we looked at the performance of three machine learning models, including two nonlinear ones (CatBoost and Random Forest), with the linear model elastic-net winning due to interpretability. In the future, it would be beneficial to also include a state-of-the-art deep neural network model, although there is a tradeoff in interpretability and performance. This will become more important for natural language processing approaches that can be used to incorporate clinical progress notes as predictors in the EHR history. Furthermore, there are improvements that can be considered for the actual model structure; for example, the models would likely be improved if they accounted for dependencies introduced by encounters belonging to the same patient.

Finally, because our model has demonstrated that it can identify patients with distress, an important question then becomes what to do with these patients? There is substantial evidence that screening alone is not enough and that an efficient referral system and evidence-based treatment are necessary.[44] Thus, devel-

oping a system for identifying and treating patients with distress in ophthalmology clinics will require buy-in from the patient, provider, payer, and healthcare system. Future studies will have to focus on the development of referral systems that are acceptable to patients and providers, as well as interventions, including vision and behavioral interventions, that could improve patients' quality of life and are focused on improving distress related to specific vision-related disorders. A special focus will have to be on guaranteeing that patients will have access to appropriate care, regardless of demographics and distress severity.

In conclusion, our study demonstrated that prescreening for distress in ophthalmology patients can be automated using an AI algorithm trained on existing EHR data. The algorithm identified distress in patients already being treated, and in those with incident distress. These findings suggest that screening for distress in ophthalmology clinics is feasible and may reduce negative health outcomes in patients.

translational vision science & technology

## References

1. McCusker S, Koola MM. Association of Ophthalmologic Disorders and Depression in the Elderly. *Prim Care Companion CNS Disord*. 2015; 17(4):10.4088/PCC.14r01731, https://doi.org/10.4088/PCC.14r01731.

2. Congdon N, O'Colmain B, Klaver CCW, et al. Causes and prevalence of visual impairment among adults in the United States. *Arch Ophthalmol*. 2004;122(4):477–485, https://doi.org/10.1001/archopht.122.4.477.

3. Zheng Y, Wu X, Lin X, Lin H. The prevalence of depression and depressive symptoms among eye disease patients: a systematic review and meta-analysis. *Sci Rep*. 2017;7(1):46453, https://doi.org/10.1038/srep46453.

4. Diniz-Filho A, Abe RY, Cho HJ, Baig S, Gracitelli CPB, Medeiros FA. Fast visual field progression is associated with depressive symptoms in patients with glaucoma. *Ophthalmology*. 2016;123(4):754–759.

5. Zhang X, Olson DJ, Le P, Lin F-C, Fleischman D, Davis RM. The association between glaucoma, anxiety, and depression in a large population. *Am J Ophthalmol*. 2017;183:37–41.

6. Clarke A, Rumsey N, Collin JRO, Wyn-Williams M. Psychosocial distress associated with disfiguring eye conditions. *Eye (Lond)*. 2003;17(1):35–40, https://doi.org/10.1038/sj.eye.6700234.

7. Hark LA, Madhava M, Radakrishnan A, et al. Impact of a social worker in a glaucoma eye care service: a prospective study. *Health Soc Work*. 2019;44(1):48–56, https://doi.org/10.1093/hsw/hly038.

8. Kong XM, Zhu WQ, Hong JX, Sun XH. Is glaucoma comprehension associated with psychological disturbance and vision-related quality of life for patients with glaucoma? A cross-sectional study. *BMJ Open*. 2014;4(5):e004632, https://doi.org/10.1136/bmjopen-2013-004632.

9. Chen X, Lu L. Depression in diabetic retinopathy: a review and recommendation for psychiatric management. *Psychosomatics*. 2016;57(5):465–471, https://doi.org/10.1016/j.psym.2016.04.003.

10. Berchuck SI, Jammal A, Mukherjee S, Somers T, Medeiros FA. Impact of anxiety and depression on progression to glaucoma among glaucoma suspects. *Br J Ophthalmol*. 2021;105(9):1244–1249, https://doi.org/10.1136/bjophthalmol-2020-316617.

11. Grenard JL, Munjas BA, Adams JL, et al. Depression and medication adherence in the treatment of chronic diseases in the United States: a meta-analysis. *J Gen Intern Med*. 2011;26(10):1175–1182, https://doi.org/10.1007/s11606-011-1704-y.

12. Cunningham SC, Aizvera J, Wakim P, Felber L. Use of a self-reported psychosocial distress screening tool as a predictor of need for psychosocial intervention in a general medical setting. *Soc Work Health Care*. 2018;57(5):315–331, https://doi.org/10.1080/00981389.2018.1437499.

13. Mitchell AJ, Kaar S, Coggan C, Herdman J. Acceptability of common screening methods used to detect distress and related mood disorders—preferences of cancer specialists and non-specialists. *Psycho-Oncology*. 2008;17(3):226–236, https://doi.org/10.1002/pon.1228.

14. Cull A, Gould A, House A, et al. Validating automated screening for psychological distress by means of computer touchscreens for use in routine oncology practice. *Br J Cancer*. 2001;85(12):1842–1849, https://doi.org/10.1054/bjoc.2001.2182.

15. Ma X, Zhang J, Zhong W, et al. The diagnostic role of a short screening tool—the distress thermometer: a meta-analysis. *Support Care Cancer*. 2014;22(7):1741–1755, https://doi.org/10.1007/s00520-014-2143-1.

16. Donovan KA, Jacobsen PB. Progress in the implementation of NCCN guidelines for distress management by member institutions. *J Natl Comp Canc Netw*. 2013;11(2):223–226, https://doi.org/10.6004/jnccn.2013.0029.

17. Lichtman JH, Bigger JT, Blumenthal JA, et al. Depression and coronary heart disease. *Circulation*. 2008;118(17):1768–1775, https://doi.org/10.1161/CIRCULATIONAHA.108.190769.

18. Fann JR, Berry DL, Wolpin S, et al. Depression screening using the Patient Health Questionnaire-9 administered on a touch screen computer. *Psycho-Oncology*. 2009;18(1):14–22, https://doi.org/10.1002/pon.1368.

19. Kelly D, Curran K, Caulfield B. Automatic prediction of health status using smartphone-derived behavior profiles. *IEEE J Biomed Health Inform*. 2017;21(6):1750–1760, https://doi.org/10.1109/JBHI.2017.2649602.

20. Rana R, Latif S, Gururajan R, et al. Automated screening for distress: a perspective for the future. *Eur J Cancer Care*. 2019;28(4):e13033, https://doi.org/10.1111/ecc.13033.

translational vision science & technology

21. Latif S, Rana R, Khalifa S, Jurdak R, Epps J, Schuller BW. Multi-task semi-supervised adversarial autoencoding for speech emotion recognition. *arXiv*. 2020, https://doi.org/10.48550/arXiv.1907.06078.

22. Hu Z, Melton GB, Arsoniadis EG, Wang Y, Kwaan MR, Simon GJ. Strategies for handling missing clinical data for automated surgical site infection detection from the electronic health record. *J Biomed Inform*. 2017;68:112–120.

23. Simon GE, Johnson E, Lawrence JM, et al. Predicting suicide attempts and suicide deaths following outpatient visits using electronic health records. *Am J Psychiatry*. 2018;175(10):951–960.

24. PheKB. Depression. Available at: https://phekb.org/phenotype/depression. Accessed September 22, 2022.

25. Huang SH, LePendu P, Iyer SV, Tai-Seale M, Carrell D, Shah NH. Toward personalizing treatment for depression: predicting diagnosis and severity. *J Am Med Inform Assoc*. 2014;21(6):1069–1075, https://doi.org/10.1136/amiajnl-2014-002733.

26. Ingram WM, Baker AM, Bauer CR, et al. Defining major depressive disorder cohorts using the EHR: multiple phenotypes based on ICD-9 codes and medication orders. *Neurol Psychiatry Brain Res*. 2020;36:18–26, https://doi.org/10.1016/j.npbr.2020.02.002.

27. Monroe SM, Harkness KL. Major depression and its recurrences: life course matters. *Ann Rev Clin Psychol*. 2022;18(1):329–357, https://doi.org/10.1146/annurev-clinpsy-072220-021440.

28. Scholten WD, Batelaan NM, van Balkom AJ, Wjh Penninx B, Smit JH, van Oppen P. Recurrence of anxiety disorders and its predictors. *J Affect Disord*. 2013;147(1–3):180–185, https://doi.org/10.1016/j.jad.2012.10.031.

29. Elixhauser A, Steiner C, Palmer L. Clinical Classifications Software (CCS). Available at: https://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp. Accessed September 22, 2022.

30. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010;33(1):1–22.

31. Breiman L. Random forests. *Mach Learn*. 2001;45:5–32, https://doi.org/10.1023/A:1010933404324.

32. Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. CatBoost: unbiased boosting with categorical features. *arXiv*. 2019, https://doi.org/10.48550/arXiv.1706.09516.

33. Tibshirani R, Bien J, Friedman J, et al. Strong rules for discarding predictors in LASSO-type problems. *J R Stat Soc Series B Stat Methodol*. 2012;74(2):245–266, https://doi.org/10.1111/j.1467-9868.2011.01004.x.

34. Breiman L, Friedman JH, Olshen RA, Stone CJ. *Classification and Regression Trees*. New York: Routledge; 1984.

35. Newton KM, Peissig PL, Kho AN, et al. Validation of electronic medical record-based phenotyping algorithms: results and lessons learned from the eMERGE network. *J Am Med Inform Assoc*. 2013;20(e1):e147–e154, https://doi.org/10.1136/amiajnl-2012-000896.

36. Jammal AA, Thompson AC, Mariottoni EB, et al. Rates of glaucomatous structural and functional change from big data: the Duke Glaucoma Registry study. *Am J Ophthalmol*. 2020;222:238–247, https://doi.org/10.1016/j.ajo.2020.05.019.

37. Simonett JM, Sohrab MA, Pacheco J, et al. A validated phenotyping algorithm for genetic association studies in age-related macular degeneration. *Sci Rep*. 2015;5(1):12875, https://doi.org/10.1038/srep12875.

38. Bujak R, Daghir-Wojtkowiak E, Kaliszan R, Markuszewski MJ. PLS-based and regularization-based methods for the selection of relevant variables in non-targeted metabolomics data. *Front Mol Biosci*. 2016;3:35, https://doi.org/10.3389/fmolb.2016.00035.

39. Wright MN, Ziegler A. ranger: a fast implementation of random forests for high dimensional data in C++ and R. *J Stat Softw*. 2017;77(1):1–17, https://doi.org/10.18637/jss.v077.i01.

40. Dorogush AV, Ershov V, Gulin A. CatBoost: gradient boosting with categorical features support. *arXiv*. 2018, https://doi.org/10.48550/arXiv.1810.11363.

41. Newman NJ. Neuro-ophthalmology and psychiatry. *Gen Hosp Psychiatry*. 1993;15(2):102–114, https://doi.org/10.1016/0163-8343(93)90106-x.

42. Mabuchi F, Yoshimura K, Kashiwagi K, et al. High prevalence of anxiety and depression in patients with primary open-angle glaucoma. *J Glaucoma*. 2008;17(7):552–557.

43. Williams RA. The psychosocial impact of macular degeneration. *Arch Ophthalmol*. 1998;116(4):514–520, https://doi.org/10.1001/archopht.116.4.514.

44. Carlson LE. Screening alone is not enough: the importance of appropriate triage, referral, and evidence-based treatment of distress and common problems. *J Clin Oncol*. 2013;31(29):3616–3617, https://doi.org/10.1200/jco.2013.51.4315.