

1 **Title:** Design and Validation of an AI-Assisted Sequential Screening Framework for
2 Psychological Distress in Glaucoma

3
4 **Authors:** Natalie A. Chou, BA;^{*1} Youngsoo Baek, PhD;^{*2} Feilin Feng;³ Kaiwen Lu;² Eun Young
5 Choi, MD;⁴ Hannah M. Fisher, PhD;¹ Davina Malek, MD;⁵ Alessandro Jammal, MD, PhD;⁶
6 Tamara J. Somers, PhD;¹ Kelly W. Muir, MD;⁷ Felipe Medeiros, MD, PhD;⁵ Samuel I. Berchuck,
7 PhD^{2,3}

8
9 **Corresponding Author:** Samuel I. Berchuck, PhD

10
11 **Financial Support:** Research reported in this publication was supported by the National Eye
12 Institute of the National Institutes of Health (Bethesda, Maryland) under Awards Number
13 R00EY033027 (SIB), R01EY029885 (FAM), and R01EY036593 (FAM). The sponsor or funding
14 organization had no role in the design or conduct of this research. The content is solely the
15 responsibility of the authors and does not necessarily represent the official views of the National
16 Institutes of Health.

17
18 **Conflicts of Interest:** NC: None; YB: None; FF: None; KL: None; EYC: None; HMF: None; DM:
19 None; AAJ: None; TS: None; KM: None; FAM: AbbVie (C), Annexon (C); Carl Zeiss Meditec
20 (C), Enavate Sciences (C), Galimedix (C); Google Inc. (F); Heidelberg Engineering (F), nGoggle
21 Inc. (P), Novartis (F); ONL Therapeutics (C), Perfuse Therapeutics (C), Perceive Bio (C), Stealth
22 Biotherapeutics (C); Stuart Therapeutics (C), Thea Pharmaceuticals (C), Reichert (C, F).; SIB:
23 None.

24
25 **Running Head:** AI-Assisted Distress Screening in Glaucoma

26
27 **Address for reprints:** Department of Biostatistics & Bioinformatics, Duke University, 2424
28 Erwin Rd, Durham, NC, 27710, E-mail: sib2@duke.edu.

29
30 **Key Words:** Primary open-angle glaucoma; Psychological distress screening; Machine learning;
31 Electronic health records; Clinical decision support

* Equal contribution.

¹ Department of Psychiatry and Behavioral Sciences, Duke University, Durham, North Carolina, USA.

² Department of Biostatistics & Bioinformatics, Duke University, Durham, North Carolina, USA.

³ Department of Statistical Science, Duke University, Durham, North Carolina, USA.

⁴ Kaiser Permanente, Santa Clara Homestead Medical Center, Santa Clara, California, USA.

⁵ Bascom Palmer Eye Institute, University of Miami, Miami, Florida, USA.

⁶ Massachusetts Eye and Ear, Harvard University, Boston, Massachusetts, USA.

⁷ Duke Eye Center and Department of Ophthalmology, Duke University, Durham, North Carolina, USA.

32 **ABSTRACT**

33

34 **Purpose:** Psychological distress is highly prevalent in glaucoma and is associated with worse
35 adherence, reduced quality of life, and faster disease progression. However, distress is rarely
36 assessed in ophthalmology settings due to time, workflow, and staffing constraints. We evaluated
37 two artificial intelligence (AI)-based screening strategies, designed to efficiently identify
38 distressed primary open angle glaucoma (POAG) patients during routine care, aiming to achieve
39 effective, resource conscious, low burden clinical screening.

40

41 **Design:** Hybrid retrospective cohort and prospective cross-sectional study.

42

43 **Participants:** The retrospective cohort included >3,000 POAG patients from the Duke
44 Ophthalmic Registry. Prospective validation was conducted in a separate 300 POAG patient cohort
45 who completed patient-reported distress screening.

46

47 **Methods:** Using retrospective data, a neural network model was trained to predict an electronic
48 health record (EHR)-derived computable phenotype of distress (“silver standard”). Prospective
49 validation used the 8-item Patient Health Questionnaire (PHQ-8) as the “gold standard.” Three
50 screening strategies were compared against PHQ-8: (1) universal PHQ-2 screening (two-item
51 screener administered to all patients), (2) AI-only screening (fully automated EHR-based
52 screener), and (3) sequential screening, (only patients flagged as high risk by AI screener
53 completed the PHQ-2). Performance metrics included sensitivity, specificity, positive predictive
54 value (PPV), negative predictive value (NPV), accuracy, and screening burden.

55

56 **Main Outcome Measures:** Sensitivity; specificity; PPV; NPV; accuracy; proportion of patients
57 requiring secondary screening (screening burden).

58

59 **Results:** Distress prevalence was 17% (PHQ-8 > 6). Universal PHQ-2 screening (> 0) achieved
60 high sensitivity (0.96) but lower specificity (0.73) and PPV (0.41), while requiring screening of all
61 patients. The AI-assisted sequential approach substantially reduced screening burden while
62 maintaining strong diagnostic performance. By administering PHQ-2 to ~25% of patients,
63 sequential screening achieved sensitivity 0.64, specificity 0.93, PPV 0.64, NPV 0.93, and accuracy
64 0.88, representing a ~50% increase in PPV compared to PHQ-2 alone. AI-only screening reduced
65 burden further but did not achieve comparable sensitivity or predictive performance.

66

67 **Conclusions:** AI-assisted sequential screening enables scalable, resource efficient identification
68 of psychological distress in glaucoma care, substantially reducing screening burden while
69 preserving clinically meaningful performance. This framework offers a practical pathway for
70 integrating distress screening into routine ophthalmology workflows and improving the
71 identification and referral of at-risk patients.

72 Primary open angle glaucoma (POAG) is a chronic, progressive optic neuropathy that affects over
73 65 million people aged 40-80 worldwide,¹ and remains the leading cause of irreversible vision
74 loss.² As such, the uncertainty about future vision loss and the disease's gradual progression
75 contribute to considerable psychological burden in patients.³ While management primarily targets
76 intraocular pressure (IOP) reduction to slow functional decline, factors such as emotional
77 wellbeing, social support, and mental health, are increasingly recognized as critical in disease
78 evaluation and care.^{4,5}

79 The chronic course of POAG often leads to psychological distress including anxiety,
80 depression, fear of blindness, and social isolation, all of which substantially reduce quality of life.⁶⁻
81 ⁸ Psychological distress prevalence in POAG patients is notably high, with up to 25% experiencing
82 anxiety and 30% experiencing depression, rates more than 10% higher than the general
83 population.⁶ Disease severity influences these outcomes, as worse functional disease severity⁸ and
84 faster rates of progression⁹ correlate with greater depression. Importantly, psychological distress
85 impacts disease management and prognosis. Patients reporting distress frequently show poor
86 glaucoma comprehension,¹⁰ missed follow up visits,¹¹ and suboptimal medication adherence,¹²
87 contributing to long term visual field loss. Psychological distress may additionally exert a direct
88 physiologic effect, as stress can elevate IOP, the only modifiable risk factor in glaucoma.⁴
89 Experimental studies in primates¹³ and humans¹⁴ confirm that acute stress can increase IOP.
90 Addressing psychological distress early in the treatment pathway could improve adherence,
91 outcomes, and quality of life, but routine screening in glaucoma care is not common.

92 Early identification of psychological distress enables timely intervention and reduces
93 adverse outcomes.¹⁵ However, conventional screening approaches are often limited by substantial
94 time and cost demands.¹⁶ To address these challenges, several medical specialties have adopted

95 more efficient strategies, including brief self-report questionnaires and two stage screening
96 frameworks. In oncology and cardiology, for example, initial large scale pre-screening with short
97 instruments, such as the National Comprehensive Cancer Network (NCCN) Distress
98 Thermometer¹⁷ or the two item Patient Health Questionnaire¹⁸, is followed by more comprehensive
99 assessment only in patients with positive results.¹⁹ This approach reduces resource burden while
100 targeting high risk patients for further evaluation.²⁰ In glaucoma care, however, routine
101 psychological distress screening remains uncommon despite its potential to identify vulnerable
102 patients and enable early intervention, representing a critical gap in supportive disease
103 management. The most direct solution would be universal administration of validated patient
104 reported outcome (PRO) measures. Our group has taken steps toward this goal, including recent
105 work validating the Distress Thermometer in patients with POAG, supporting the feasibility and
106 clinical relevance of routine distress screening in this population.²¹ Nevertheless, universal
107 screening has not been widely implemented in ophthalmology practice due to persistent workflow,
108 time, and resource constraints. As a result, there is a need for scalable approaches that approximate
109 the benefits of universal screening while remaining feasible in routine care. In this context, we
110 propose an AI-assisted screening framework designed to reduce screening burden while preserving
111 performance comparable to universal screening.

112 Electronic health record (EHR) data have been widely used for predictive modeling across
113 clinical domains,²²⁻²⁵ including work demonstrating accurate prediction of incident suicide risk.²⁶
114 These advances highlight the potential of artificial intelligence (AI) to enable scalable, population
115 level identification of patients at risk for adverse outcomes. In prior work, we developed machine
116 learning models using EHR data from a large retrospective glaucoma cohort and demonstrated
117 strong discrimination for a computable distress phenotype (area under the receiver operating

118 characteristic [ROC] curve > 0.90).²⁷ However, because patient reported distress measures were
119 not available in that setting, outcomes were defined using an EHR-based proxy (hereafter referred
120 to as a “silver standard”). While such approaches are useful for large scale model development,
121 understanding clinical impact requires evaluation against validated PROs that directly measure
122 psychological distress.

123 To address this gap, we conducted a prospective study with PRO measures of distress,
124 enabling evaluation against a gold standard in a clinical setting. We evaluated whether AI-based
125 screening strategies can approximate the performance of universal screening while reducing
126 implementation burden. Specifically, we compared three approaches for identifying psychological
127 distress: (1) universal screening using the two item Patient Health Questionnaire (PHQ-2), (2) fully
128 automated AI-only screening based on EHR data, and (3) an AI-assisted sequential screening
129 strategy in which the algorithm is used to pre-screen patients for targeted questionnaire
130 administration. This sequential approach leverages AI to reduce screening burden by limiting
131 manual questionnaire screening to a subgroup of high-risk patients, rather than relying on
132 algorithmic predictions alone. This design enables direct evaluation of the tradeoff between
133 screening burden and diagnostic performance across approaches. Together, these analyses provide
134 a practical framework for implementing a scalable, clinically actionable distress screening in
135 glaucoma care.

136

137 **METHODS**

138 **Study Design**

139 This study used data from both a large retrospective EHR cohort and a prospective cohort with
140 PROs. The retrospective cohort was used to develop machine learning models using an EHR-

141 derived phenotype of distress (silver standard), while the prospective cohort was used to evaluate
142 model performance against PRO measures of distress (gold standard). Retrospective data were
143 obtained from the Duke Ophthalmic Registry, which includes patients evaluated at the Duke Eye
144 Center and affiliated satellite clinics between 2012 and 2021. Eligible patients were adults aged
145 ≥ 18 years with POAG. Prospective data were obtained from a study examining patient attitudes
146 toward psychological distress screening in glaucoma clinics.^{21,28} This study was approved by the
147 Duke University Institutional Review Board (Pro00108155). The retrospective component was
148 conducted under a waiver of informed consent due to the use of existing clinical data, and all
149 participants in the prospective study provided informed consent. All procedures adhered to the
150 tenets of the Declaration of Helsinki and complied with the Health Insurance Portability and
151 Accountability Act.

152 **Figure 1** provides an overview of the study design, including development of the AI
153 algorithm in the retrospective cohort and external validation in the prospective cohort. **Figure 2**
154 summarizes the three screening strategies evaluated: (1) universal screening, (2) AI-only
155 screening, and (3) AI-assisted sequential screening.

156

157 **Retrospective Cohort and Model Development**

158 The retrospective cohort was derived from the Duke Ophthalmic Registry and included adults
159 (aged ≥ 18 years) with POAG who were evaluated at the Duke Eye Center and affiliated satellite
160 clinics between 2012 and 2021. Patients were required to have at least two encounters and a
161 minimum of one year of follow up. To align with the prospective validation cohort, analyses were
162 restricted to encounters in the glaucoma clinic. POAG was defined using International

163 Classification of Diseases diagnostic codes (H40.1), and the unit of analysis was the clinical
164 encounter.

165 Psychological distress was defined using an encounter-level EHR-derived computable
166 phenotype based on a validated Phenotype KnowledgeBase (PheKB) depression algorithm, as
167 described in prior work.²⁷ Because this definition relies on proxy clinical data rather than direct
168 patient report, it is referred to as a silver standard measure of distress. Predictors were constructed
169 from structured EHR data available prior to each encounter and included demographics, healthcare
170 utilization (diagnoses, procedures, medications, and encounters), and problem-list information,
171 resulting in a high-dimensional feature set for model development (1,844 variables).
172 Socioeconomic variables, including income and education, were derived from U.S. Census data
173 and linked at the Zip Code level. Income level was measured by per capita income in the past 12
174 months and was race specific. Education was measured by the percentage of residents who
175 achieved a high school education and was sex specific. Insurance status (commercial, Medicaid,
176 Medicare, uninsured, other) was also included.

177 Two predictive models were trained to estimate encounter-level probabilities of
178 psychological distress: an elastic net classifier and a feedforward neural network. The elastic net
179 model was fit using patient-level 10-fold nested cross-validation to select optimal regularization
180 parameters and prevent overfitting. The neural network consisted of a four-layer architecture with
181 ReLU activation functions and was trained using the Adam optimizer,²⁹ with early stopping to
182 improve generalization.³⁰ To enhance stability, predictions were averaged across multiple random
183 initializations. Model performance was evaluated using the area under the receiver operating
184 characteristic (ROC) curve and precision-recall (PR) curve in held-out data constructed via patient-

185 level splitting. These models generated the predicted probabilities used for AI-only screening and
186 as the first stage of the sequential screening strategy.

187

188 **Prospective Cohort and Outcome Assessment**

189 Patients with POAG were recruited from the Duke Eye Center glaucoma clinics between
190 September 2022 and August 2023. Study procedures have been described previously and are
191 briefly summarized here.^{21,28} Eligibility required age ≥ 18 years, a confirmed POAG diagnosis, at
192 least one prior glaucoma clinic visit, and a minimum of one year of follow up. Exclusion criteria
193 included recent glaucoma or cataract surgery within 6 months, reported or suspected serious
194 mental illness (e.g., psychosis), or legal blindness, as determined by provider or chart review.
195 Eligible patients were contacted by email prior to scheduled clinic visits, and consenting
196 participants completed a set of questionnaires.

197 Demographic and clinical characteristics were collected to describe the cohort. Patient
198 reported measures included age, sex, race, ethnicity, marital status, education, income, and
199 employment status. Clinical data were obtained from the Duke Ophthalmic Registry and
200 supplemented by manual chart review. Eye specific measures included intraocular pressure (IOP)
201 and standard automated perimetry mean deviation (SAP MD), reported for the better eye.

202 Psychological distress was assessed using the Patient Health Questionnaire-8 (PHQ-8),
203 which served as the gold standard measure of distress.¹⁸ The PHQ-8 includes all items from the
204 PHQ-9 except for the question assessing suicidal or self-injurious thoughts; this item was excluded
205 to avoid triggering risk assessments that could not be reliably and sufficiently managed in a remote
206 research setting. Prior validation studies have demonstrated that the removal of this item has
207 minimal impact on total scores and does not alter established cutoffs for depression severity.^{31,32}

208 The PHQ-8 evaluates the frequency of depressive symptoms over the prior two weeks, with total
209 scores ranging from 0 to 24. Consistent with prior work, clinically relevant psychological distress
210 was defined as a PHQ-8 score > 6 .^{33,34} The PHQ-8 was used for evaluation purposes and was not
211 part of the screening strategies assessed in this study.

212

213 **Screening Strategies**

214 We evaluated three approaches for identifying psychological distress: (1) universal questionnaire-
215 based screening, (2) AI-only screening, and (3) an AI-assisted sequential screening strategy.
216 Universal screening was conducted using the two item Patient Health Questionnaire (PHQ-2), a
217 brief validated screener for depressive symptoms. In accordance with American Heart Association
218 (AHA) recommendations for depression screening,¹⁸ any positive response (score > 0) was
219 considered indicative of a positive screen. In the AI-only approach, our EHR-based algorithm was
220 used to assign each patient a predicted probability of psychological distress based on routinely
221 collected clinical data. Patients with predicted probabilities above a specified threshold were
222 classified as distressed (more details in the following Statistical Evaluation section). This approach
223 is fully automated and does not require administration of patient questionnaires, relying solely on
224 EHR-derived predictions.

225 Finally, the AI-assisted sequential strategy combines algorithmic risk stratification with
226 targeted questionnaire-based assessment. In this approach, the AI algorithm is first used to identify
227 patients at elevated risk for psychological distress. Only these high-risk patients then undergo
228 questionnaire-based screening using the PHQ-2, and patients with a positive PHQ-2 result are
229 classified as distressed. This design follows established two stage screening frameworks used in

230 other clinical domains, such as oncology and cardiology,^{35,36} while reducing the burden of
231 universal screening by limiting questionnaire administration to a subset of high-risk patients.

232

233 **Statistical Evaluation**

234 Summary statistics for cohort characteristics are presented using mean and standard deviation (SD)
235 for continuous variables and counts and percentages for categorical variables. Hypothesis tests are
236 presented, with categorical variables tested using a Chi-square test, and continuous variables tested
237 using a Wilcoxon rank sum test. Model discrimination was evaluated using ROC and precision-
238 recall (PR) curves in both the training (retrospective) and test (prospective) datasets. In the training
239 data, performance was assessed against the EHR-derived silver standard. In the test data,
240 performance was evaluated against both the silver standard and the patient reported gold standard.
241 Areas under the curve were computed for both ROC and PR curves.

242 Screening strategies were evaluated using the prospective cohort with PROs as the gold
243 standard (PHQ-8 > 6). Diagnostic performance was assessed using sensitivity, specificity, positive
244 predictive value (PPV), negative predictive value (NPV), and overall accuracy. We considered
245 two complementary evaluation approaches reflecting distinct clinical objectives. First, to assess
246 whether AI-only screening could serve as a substitute for universal questionnaire-based screening,
247 thresholds for the AI model were selected to match the specificity of PHQ-2 screening (score > 0).
248 This enabled a direct comparison of sensitivity, predictive values, and accuracy under comparable
249 operating conditions. Results are presented for both elastic net and neural network models. Second,
250 for the AI-assisted sequential strategy, we aimed not to replicate the operating characteristics of
251 universal screening, but to evaluate the tradeoff between screening burden and diagnostic
252 performance. In this setting, thresholds for the AI model were selected using Youden's index³⁷ to

253 balance sensitivity and specificity. The PHQ-2 threshold indicating distress was fixed at any score
254 > 0. Screening burden was defined as the proportion of patients requiring questionnaire-based
255 screening, which varies with the chosen AI threshold; operating characteristics were evaluated as
256 a function of this proportion to characterize the tradeoff between burden and performance.
257 Operating characteristics for the sequential strategy were further examined across demographic
258 subgroups, including age, sex, and race.

259 Variable importance was computed for both models. For the elastic net model, the largest
260 25 coefficients in absolute value were presented along with their odds ratios (OR). OR p-values
261 were not presented, since they cannot be computed reliably for the elastic net model.³⁸ The non-
262 zero demographic predictors were also presented. For a measure of variable importance in neural
263 network, Shapley values³⁹ were computed on a random subset of 1,000 encounters from the EHR
264 data. Variables with the top 25 absolute Shapley values for positive prediction were presented,
265 along with a list of demographic variables with the top 10 largest values.

266 All analyses were conducted using R (Version 4.0.5, R Core Team, Vienna, Austria) and
267 Python (Version 3.13, Python Software Foundation) within the Protected Analytics Computing
268 Environment (PACE) at Duke University. The glmnet and PyTorch packages were used for model
269 development.⁴⁰⁻⁴²

270

271 **RESULTS**

272 The retrospective cohort consisted of 31,043 encounters from 3,149 patients with a mean (SD) of
273 9.9 (7.3) encounters per patient over 2.6 (2.2) years of follow up. The mean age of the patients at
274 first encounter was 68.6 (13.2) years with a breakdown of 1,790 (57%) females, 1,651 (52%)
275 White, 1,337 (42%) Black/African American, and the rest Asian, multiracial, and other races.

276 There were 1,240 (39%) patients with at least one encounter with a positive silver standard distress
277 phenotype. The prospective cohort consisted of 300 patients with an average age of 68.4 (12.5)
278 years, with a breakdown of 155 (52%) females, 242 (81%) White, and 46 (15%) Black/African
279 American. Full summary details for the retrospective and prospective cohorts can be found in
280 **Table 1** and **Table 2**, respectively.

281 **Figure 3** presents ROC and PR curves for both machine learning models evaluated in the
282 training (retrospective) and test (prospective) data, against both silver and gold standard outcomes.
283 Both models demonstrated strong discrimination when predicting the silver standard, with ROC
284 AUCs of 0.93 (neural network) and 0.86 (elastic net) in the training data, and 0.83 and 0.86,
285 respectively, in the test data. PR AUCs were 0.89 and 0.79 in training, and 0.79 for both models
286 in the test data. When evaluated against the gold standard in the test data, discrimination was lower,
287 reflecting the increased difficulty of predicting PROs. However, the neural network retained
288 meaningful predictive signal, with ROC AUC of 0.76 and PR AUC of 0.39, and outperformed the
289 elastic net (ROC AUC 0.64; PR AUC 0.30). For comparison, universal PHQ-2 screening achieved
290 higher discrimination (ROC AUC 0.94; PR AUC 0.80). Given the superior performance of the
291 neural network for the gold standard outcome, subsequent analyses focus on the neural network
292 model, with elastic net results reported for completeness.

293 **Table 3** compares the diagnostic performance of the three screening approaches. Universal
294 screening using PHQ-2 achieved high sensitivity (0.96) and NPV (0.99), with specificity of 0.73,
295 PPV of 0.41, and accuracy of 0.77, but required questionnaire administration for all patients. In
296 contrast, AI-only screening eliminated the need for questionnaire-based screening but did not
297 achieve comparable diagnostic performance. The neural network-based screening, with the
298 threshold fixed to achieve the same specificity of 0.73 (anyone with a predicted probability of

299 distress $\geq 21\%$), yielded sensitivity of 0.68, PPV of 0.33, NPV of 0.92, and accuracy of 0.72. The
300 sequential screening strategy, using the Youden's index-based threshold (anyone with a predicted
301 probability of distress $\geq 26\%$) provided a balance between these approaches. When using the
302 neural network, sequential screening achieved the highest specificity (0.93), PPV (0.64), and
303 accuracy (0.88), while maintaining high NPV (0.93), with only a modest reduction in sensitivity
304 (0.64). This improvement was achieved while limiting questionnaire-based screening to 28% of
305 patients, compared to 100% under universal screening. **Table 4** presents the diagnostic
306 characteristics across subgroups of age, sex and race for the sequential screening method relying
307 on the neural network. The same breakdown is given for the elastic net model in **Supplementary**
308 **Table S1**. These findings highlight a tradeoff between sensitivity and operational feasibility, with
309 sequential screening achieving improved precision while substantially reducing screening burden.

310 **Figure 4** illustrates the tradeoff between screening burden and diagnostic performance for
311 the neural network sequential strategy. The vertical dashed line indicates the operating point
312 selected using Youden's index, which corresponds to the results in **Table 3**. As the operating
313 threshold varies, the proportion of patients requiring questionnaire-based screening changes, with
314 corresponding changes in sensitivity, specificity, and predictive values. These curves demonstrate
315 that the operating point can be adjusted to achieve different balances between screening burden
316 and diagnostic performance. **Supplementary Figure S1** shows the same curves for elastic net.

317 Variable importance measures for each model are reported in the **Supplementary Tables**
318 **S2-S4**. **Table S2** presents a list of variables with the 25 highest absolute Shapley values for the
319 neural network predicting positive psychological distress; **Supplementary Table S3** contains the
320 top 10 demographic predictors for the same model. Finally, **Supplementary Table S4** shows the
321 top 25 predictors from the elastic net model ordered by the largest coefficients in absolute value.

322

323 **DISCUSSION**

324 This study developed and validated AI-based screening strategies for identifying psychological
325 distress in glaucoma care, including both a fully automated AI-only approach and an AI-assisted
326 sequential screening framework. Our findings demonstrate that while AI-only screening does not
327 achieve performance comparable to universal PHQ-2 administration, it captures meaningful
328 predictive signal that can be leveraged to improve efficiency when combined with targeted PRO
329 screening. The sequential approach improved specificity, PPV, and overall accuracy, while
330 reducing the need for questionnaire-based screening by approximately fourfold, supporting a
331 practical balance between diagnostic performance and real world feasibility.

332 A central challenge of implementing psychological distress screening in ophthalmology is
333 balancing its diagnostic performance with the practical constraints of clinical workflows.
334 Universal PHQ-2 screening maximizes sensitivity but requires screening all patients, which may
335 be difficult to sustain in busy clinics with limited staff and competing priorities. Conversely, fully
336 automated approaches reduce burden, but rely solely on EHR-derived information and may not
337 capture patient reported distress with sufficient accuracy. These constraints motivate approaches
338 that can selectively target screening efforts while preserving meaningful clinical performance.

339 Consistent with this framing, the AI-only approach demonstrated uniformly lower
340 performance than universal PHQ-2 screening across sensitivity, PPV, NPV, and accuracy when
341 evaluated at matched specificity. In contrast, the sequential framework improved PPV and overall
342 accuracy while substantially reducing screening burden. Although universal screening achieved
343 the highest sensitivity, it required manual questionnaire administration for all patients. AI-only
344 screening eliminated this burden, but at the cost of reduced sensitivity and predictive performance.

345 The neural network-based sequential approach, borrowing strengths of both strategies, achieved
346 the highest specificity and accuracy, and markedly increased PPV to 0.64, compared to 0.41 for
347 PHQ-2 alone (approximately a 50% relative increase). Subgroup analyses (**Table 4**) further show
348 that PPV remained generally high across demographic groups, indicating that the improved
349 precision of the sequential approach was consistent across populations.

350 A key advantage of the sequential screening framework is its higher PPV and specificity,
351 ensuring that patients flagged by the method are more likely to benefit from downstream
352 psychological resources. While both universal and sequential approaches aim at identifying
353 subsets of patients for referral, the sequential framework reduces the number of patients who
354 require initial questionnaire-based screening, focusing clinical attention earlier in the workflow on
355 a smaller, higher-risk group. The sequential approach therefore yields a more targeted cohort for
356 referral (e.g., to a social worker), optimizing the limited availability of mental health service
357 providers and increasing the likelihood of meaningful intervention. Efficient allocation is
358 particularly critical in under resourced settings, as it promotes equity by prioritizing those with
359 greater need, reduces stigma associated with false positives, and strengthens provider confidence
360 in referrals. Moreover, this framework can be readily adapted to other chronic diseases
361 characterized by similar workflow and psychological care challenges⁴³ and aligns with value based
362 care principles, emphasizing efficient, personalized screening that balances diagnostic precision
363 with real world feasibility. Finally, any AI model that outputs predicted probabilities of
364 psychological distress can be seamlessly integrated into this framework, enabling further
365 improvements in performance as models evolve.

366 Reduced sensitivity in the sequential approach reflects an intentional and pragmatic
367 tradeoff to improve feasibility. As illustrated in **Figure 4**, thresholds that maximize sensitivity

368 come at the cost of substantially lower specificity and PPV, whereas modestly higher thresholds
369 yield meaningful gains in predictive value and screening efficiency. In this context, higher PPV is
370 particularly valuable, as patients identified by the sequential method are more likely to have true
371 psychological distress, increasing clinical confidence and the likelihood of effective follow up.
372 Furthermore, higher specificity resulting from algorithmic preselection implies that most patients
373 not at risk avoid unnecessary and potentially distressing assessment, preserving scarce mental
374 health resources.

375 These tradeoffs are especially relevant in ophthalmology, where universal PRO-based
376 screening is often impractical due to workflow constraints, limited staff capacity, and the absence
377 of standardized referral pathways.⁴⁴ In such settings, high sensitivity approaches, while ideal in
378 theory, may become “assessment without action,” where positive screens do not reliably lead to
379 follow up care.⁴⁵ The sequential framework instead aligns with real world clinical practice by
380 focusing screening and downstream care on a smaller subset of high-risk patients, increasing the
381 likelihood that identified patients receive meaningful intervention. Evidence from glaucoma
382 patients indicates strong interest in both distress screening and referral for psychological support,
383 underscoring the importance of designing workflows that translate screening into actionable care.²⁸

384 Identifying psychological distress is only the first step; effective implementation requires
385 clear pathways for follow up. In academic and large health systems, referral to embedded social
386 workers or integrated behavioral health providers represents the ideal pathway. However, in many
387 ophthalmology practices, particularly private or resource limited settings, alternative approaches
388 are needed. Physicians may recommend peer support resources or condition specific support
389 groups for vision loss, such as those offered through the Glaucoma Research Foundation or local
390 vision rehabilitation programs. Coordination with the patient’s primary care provider provides

391 another accessible pathway, enabling further evaluation and referral within an established care
392 relationship. Brief educational materials addressing distress, coping strategies, and available
393 community resources can also be provided at the point of care. As noted in prior work, integrating
394 mental health screening and referral into ophthalmology, even in resource limited settings, can
395 meaningfully improve patient wellbeing and engagement in care.⁴⁶ Future implementation studies
396 should evaluate the effectiveness of these pathways across diverse practice settings.

397 Several limitations warrant consideration. First, the sequential approach reduces
398 sensitivity, and some patients with psychological distress may be missed under resource
399 constrained screening, highlighting the need for careful calibration of thresholds as clinical
400 workflows evolve. Second, the prospective validation cohort was predominantly White, which
401 may limit generalizability to more diverse populations, so subgroup differences in performance
402 warrant further study in larger and more representative cohorts. Third, although external validation
403 was conducted using prospective data, model development and validation were performed within
404 a single health system, and evaluation in multisite settings is needed to establish broader
405 generalizability. Finally, reliance on an EHR-derived phenotype in the training phase may not fully
406 capture patient reported distress and may introduce misclassification, underscoring the importance
407 of incorporating additional gold standard PRO data in future work. Continued evaluation in
408 prospective implementation studies will be important to assess workflow integration, patient and
409 provider acceptability, and the impact of screening on downstream clinical care.

410 In conclusion, this study presents a practical, resource aware framework for screening
411 psychological distress in glaucoma care using EHR-derived AI. While AI-only screening does not
412 replace universal PRO-based screening, an AI-assisted sequential screening strategy provides a
413 meaningful balance between diagnostic performance and feasibility. By substantially reducing

414 burden while improving PPV and specificity, this approach offers a realistic and scalable pathway
415 for integrating psychological risk assessment into routine ophthalmology care and may be
416 adaptable to other chronic diseases with similar workflow and behavioral health challenges.

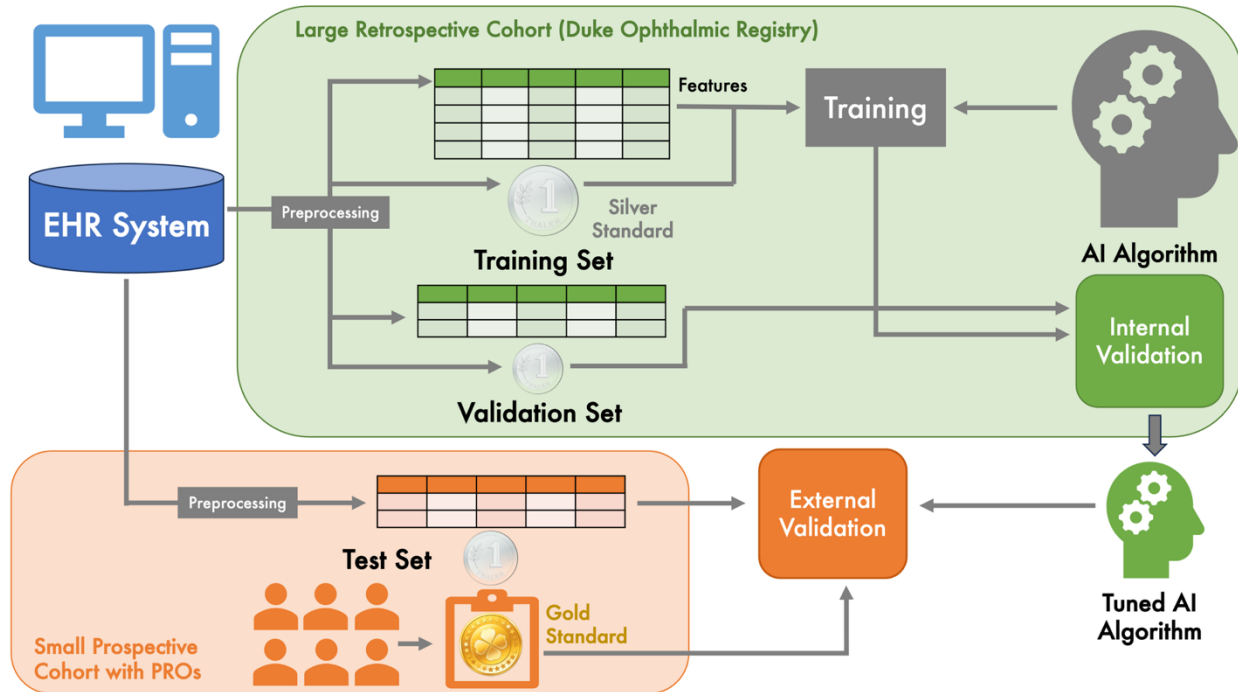
417 **REFERENCES**

- 418 1. Tham Y-C, Li X, Wong TY, Quigley HA, Aung T, Cheng C-Y. Global Prevalence of
419 Glaucoma and Projections of Glaucoma Burden through 2040. *Ophthalmology*.
420 2014;121(11):2081-2090. doi:10.1016/j.ophtha.2014.05.013
- 421 2. Weinreb RN, Aung T, Medeiros FA. The pathophysiology and treatment of glaucoma: a
422 review. *JAMA*. 2014;311(18):1901-1911.
- 423 3. Nelson P, Aspinall P, Pappasouliotis O, Worton B, O'Brien C. Quality of life in glaucoma
424 and its relationship with visual function. *Journal of Glaucoma*. 2003;12(2):139-150.
- 425 4. Kaluza G, Maurer H. Stress and intraocular pressure in open angle glaucoma. *Psychology*
426 *& Health*. 1997/10// 1997;12(5):667-675. doi:10.1080/08870449708407413
- 427 5. Berchuck S, Jammal A, Mukherjee S, Somers T, Medeiros FA. Impact of anxiety and
428 depression on progression to glaucoma among glaucoma suspects. *British Journal of*
429 *Ophthalmology*. 2021;105(9):1244-1249. doi:10.1136/bjophthalmol-2020-316617
- 430 6. Zhang X, Olson DJ, Le P, Lin F-C, Fleischman D, Davis RM. The association between
431 glaucoma, anxiety, and depression in a large population. *American Journal of Ophthalmology*.
432 2017;183:37-41.
- 433 7. Abe RY, Diniz-Filho A, Costa VP, Gracitelli CPB, Baig S, Medeiros FA. The Impact of
434 Location of Progressive Visual Field Loss on Longitudinal Changes in Quality of Life of Patients
435 with Glaucoma. *Ophthalmology*. 2016/3// 2016;123(3):552-557.
436 doi:10.1016/j.ophtha.2015.10.046
- 437 8. Skalicky S, Goldberg I. Depression and quality of life in patients with glaucoma: a cross-
438 sectional analysis using the Geriatric Depression Scale-15, assessment of function related to
439 vision, and the Glaucoma Quality of Life-15. *Journal of Glaucoma*. 2008;17(7):546-551.
- 440 9. Diniz-Filho A, Abe RY, Cho HJ, Baig S, Gracitelli CPB, Medeiros FA. Fast visual field
441 progression is associated with depressive symptoms in patients with glaucoma. *Ophthalmology*.
442 2016;123(4):754-759.
- 443 10. Kong XM, Zhu WQ, Hong JX, Sun XH. Is glaucoma comprehension associated with
444 psychological disturbance and vision-related quality of life for patients with glaucoma? A cross-
445 sectional study. *BMJ Open*. 2014/5// 2014;4(5):e004632-e004632. doi:10.1136/bmjopen-2013-
446 004632
- 447 11. Lee B, Fudemberg S, Hark L, et al. Factors contributing to nonadherence to follow-up
448 appointments in a resident glaucoma clinic versus primary eye care clinic. *Patient Preference*
449 *and Adherence*. 2016/1// 2016;19-19. doi:10.2147/PPA.S89336
- 450 12. Salman M, Andrews C, Heister M, Darnley-Fisch D, Newman-Casey PA. Psychosocial
451 Predictors of Glaucoma Medication Adherence among the Support, Educate, Empower (SEE)
452 Personalized Glaucoma Coaching Pilot Study Participants. *American Journal of Ophthalmology*.
453 2020/2// 2020;doi:10.1016/j.ajo.2020.02.009
- 454 13. Turner DC, Miranda M, Morris JS, Girkin CA, Downs JC. Acute Stress Increases
455 Intraocular Pressure in Nonhuman Primates. *Ophthalmology Glaucoma*. 2019/7// 2019;2(4):210-
456 214. doi:10.1016/j.ogla.2019.03.010
- 457 14. Abe RY, Silva TC, Dantas I, et al. Can Psychologic Stress Elevate Intraocular Pressure in
458 Healthy Individuals? *Ophthalmology Glaucoma*. 2020/7// 2020;doi:10.1016/j.ogla.2020.06.011
- 459 15. Cunningham SC, Aizvera J, Wakim P, Felber L. Use of a self-reported psychosocial
460 distress screening tool as a predictor of need for psychosocial intervention in a general medical
461 setting. *Social Work in Health Care*. 2018/5// 2018;57(5):315-331.
462 doi:10.1080/00981389.2018.1437499

- 463 16. Mitchell AJ, Kaar S, Coggan C, Herdman J. Acceptability of common screening methods
464 used to detect distress and related mood disorders—preferences of cancer specialists and non-
465 specialists. *Psycho-Oncology*. 2008/3// 2008;17(3):226-236. doi:10.1002/pon.1228
- 466 17. Donovan KA, Jacobsen PB. Progress in the Implementation of NCCN Guidelines for
467 Distress Management by Member Institutions. *Journal of the National Comprehensive Cancer*
468 *Network*. 2013/2// 2013;11(2):223-226. doi:10.6004/jnccn.2013.0029
- 469 18. Lichtman JH, Bigger JT, Blumenthal JA, et al. Depression and Coronary Heart Disease.
470 *Circulation*. 2008/10// 2008;118(17):1768-1775. doi:10.1161/CIRCULATIONAHA.108.190769
- 471 19. Cull A, Gould A, House A, et al. Validating automated screening for psychological
472 distress by means of computer touchscreens for use in routine oncology practice. *British Journal*
473 *of Cancer*. 2001/12// 2001;85(12):1842-1849. doi:10.1054/bjoc.2001.2182
- 474 20. Fann JR, Berry DL, Wolpin S, et al. Depression screening using the Patient Health
475 Questionnaire-9 administered on a touch screen computer. *Psycho-Oncology*. 2009/1//
476 2009;18(1):14-22. doi:10.1002/pon.1368
- 477 21. Chou NA, Choi EY, Fisher HM, et al. The Validity of the Distress Thermometer in
478 Patients with Primary Open-Angle Glaucoma. *Ophthalmology Glaucoma*.
479 2026;doi:10.1016/j.ogla.2026.02.006
- 480 22. Churpek MM, Yuen TC, Park SY, Gibbons R, Edelson DP. Using electronic health
481 record data to develop and validate a prediction model for adverse outcomes on the wards.
482 *Critical care medicine*. 2014;42(4):841-841.
- 483 23. Roumia M, Steinhubl S. Improving cardiovascular outcomes using electronic health
484 records. *Current cardiology reports*. 2014;16(2):451-451.
- 485 24. Neill DB. Using artificial intelligence to improve hospital inpatient care. *IEEE Intelligent*
486 *Systems*. 2013;28(2):92-95.
- 487 25. Hu Z, Melton GB, Arsoniadis EG, Wang Y, Kwaan MR, Simon GJ. Strategies for
488 handling missing clinical data for automated surgical site infection detection from the electronic
489 health record. *Journal of biomedical informatics*. 2017;68:112-120.
- 490 26. Simon GE, Johnson E, Lawrence JM, et al. Predicting suicide attempts and suicide deaths
491 following outpatient visits using electronic health records. *American Journal of Psychiatry*.
492 2018;175(10):951-960.
- 493 27. Berchuck SI, Jammal AA, Page D, Somers TJ, Medeiros FA. A Framework for
494 Automating Psychiatric Distress Screening in Ophthalmology Clinics Using an EHR-Derived AI
495 Algorithm. *Translational Vision Science & Technology*. 2022;11(10)doi:10.1167/tvst.11.10.6
- 496 28. Choi EY, Chou N, Fisher HM, et al. Patient Attitudes toward Distress Screening and
497 Referral in Glaucoma Care. *Ophthalmology Glaucoma*. 2025;doi:10.1016/j.ogla.2025.09.004
- 498 29. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. 2014/12// 2014;
- 499 30. Raskutti G, Wainwright MJ, Yu B. Early stopping and non-parametric regression: an
500 optimal data-dependent stopping rule. *J Mach Learn Res*. 2014;15(1):335–366.
- 501 31. Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure.
502 Slack Incorporated Thorofare, NJ; 2002. p. 509-515.
- 503 32. Wu Y, Levis B, Riehm KE, et al. Equivalency of the diagnostic accuracy of the PHQ-8
504 and PHQ-9: a systematic review and individual participant data meta-analysis. *Psychological*
505 *Medicine*. 2019;50(8):1368-1380. doi:10.1017/s0033291719001314
- 506 33. Gómez-Gómez I, Benítez I, Bellón J, et al. Utility of PHQ-2, PHQ-8 and PHQ-9 for
507 detecting major depression in primary health care: a validation study in Spain. *Psychological*
508 *Medicine*. 2022;53(12):5625-5635. doi:10.1017/s0033291722002835

- 509 34. Villarreal-Zegarra D, Barrera-Begazo J, Otazú-Alfaro S, Mayo-Puchoc N, Bazo-Alvarez
510 JC, Huarcaya-Victoria J. Sensitivity and specificity of the Patient Health Questionnaire (PHQ-9,
511 PHQ-8, PHQ-2) and General Anxiety Disorder scale (GAD-7, GAD-2) for depression and
512 anxiety diagnosis: a cross-sectional study in a Peruvian hospital population. *BMJ Open*.
513 2023;13(9)doi:10.1136/bmjopen-2023-076193
- 514 35. Brauer ER, Lazaro S, Williams CL, et al. Implementing a Tailored Psychosocial Distress
515 Screening Protocol in a Head and Neck Cancer Program. *The Laryngoscope*. 2021;132(8):1600-
516 1608. doi:10.1002/lary.30000
- 517 36. Jha MK, Qamar A, Vaduganathan M, Charney DS, Murrough JW. Screening and
518 Management of Depression in Patients With Cardiovascular Disease. *Journal of the American*
519 *College of Cardiology*. 2019;73(14):1827-1845. doi:10.1016/j.jacc.2019.01.041
- 520 37. Fluss R, Faraggi D, Reiser B. Estimation of the Youden Index and its Associated Cutoff
521 Point. *Biometrical Journal*. 2005/8// 2005;47(4):458-472. doi:10.1002/bimj.200410135
- 522 38. Bujak R, Dagher-Wojtkowiak E, Kaliszczan R, Markuszewski MJ. PLS-Based and
523 Regularization-Based Methods for the Selection of Relevant Variables in Non-targeted
524 Metabolomics Data. *Frontiers in Molecular Biosciences*. 2016;3doi:10.3389/fmolb.2016.00035
- 525 39. Shapley LS. 17. A Value for n-Person Games. In: Harold William K, Albert William T,
526 eds. *Contributions to the Theory of Games, Volume II*. Princeton University Press; 1953:307-
527 318.
- 528 40. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models
529 via Coordinate Descent. *Journal of statistical software*. 2010;33(1):1-22.
- 530 41. Wright MN, Ziegler A. ranger : A Fast Implementation of Random Forests for High
531 Dimensional Data in C++ and R. *Journal of Statistical Software*.
532 2017;77(1)doi:10.18637/jss.v077.i01
- 533 42. Dorogush AV, Ershov V, Gulin A. CatBoost: gradient boosting with categorical features
534 support. 2018/10// 2018;
- 535 43. Deshields TL, Wells-Di Gregorio S, Flowers SR, et al. Addressing distress management
536 challenges: Recommendations from the consensus panel of the American Psychosocial Oncology
537 Society and the Association of Oncology Social Work. *CA: A Cancer Journal for Clinicians*.
538 2021;71(5):407-436. doi:10.3322/caac.21672
- 539 44. Fudemberg SJ, Amarasekera DC, Silverstein MH, et al. Overcoming Barriers to Eye
540 Care: Patient Response to a Medical Social Worker in a Glaucoma Service. *Journal of*
541 *Community Health*. 2016/8// 2016;41(4):845-849. doi:10.1007/s10900-016-0162-1
- 542 45. Donovan KA, Grassi L, Deshields TL, Corbett C, Riba MB. Advancing the science of
543 distress screening and management in cancer care. *Epidemiology and Psychiatric Sciences*.
544 2020;29doi:10.1017/s2045796019000799
- 545 46. Demmin DL, Silverstein SM. <p>Visual Impairment and Mental Health: Unmet Needs
546 and Treatment Options</p>. *Clinical Ophthalmology*. 2020;Volume 14:4229-4251.
547 doi:10.2147/ophth.S258783
- 548

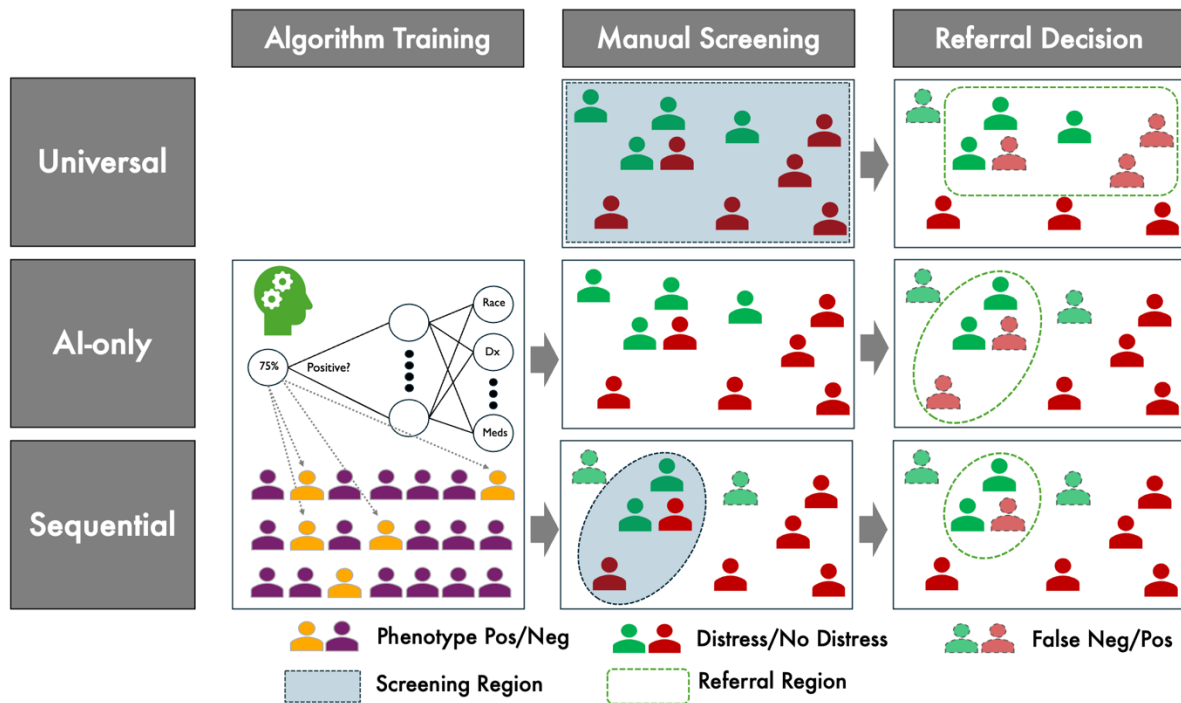
549 **Figure 1.** Schematic of Data Collection and Machine Learning Workflow. In the first phase, an
550 AI model was trained on electronic health record (EHR) data from a retrospective cohort (Duke
551 Ophthalmic Registry) to predict a silver standard outcome derived from EHR data. Model tuning
552 and internal validation were performed within the retrospective dataset. In the second phase, the
553 trained model was externally validated in a prospective cohort, based on its predictions for both
554 silver standard outcomes and gold standard patient-reported outcomes (PROs).



555

556

557 **Figure 2.** Comparison of Screening Strategies. Three approaches to identifying psychological
 558 distress are shown: (1) *Universal* screening, in which a patient-reported outcome (PRO) is
 559 administered to all patients; (2) *AI-only* screening, which relies exclusively on the trained AI model
 560 (**Figure 1**) to classify patients; and (3) *Sequential* screening, in which the AI model first identifies
 561 a subset of high-risk patients who then undergo targeted screening with a brief PRO. In Stage 1
 562 (algorithm training), yellow and purple figures represent patients with positive and negative silver
 563 standard distress phenotypes. In Stage 2 (manual screening), green and red figures represent
 564 patients with and without distress based on the gold standard. Blue shading covers patients who
 565 receive PROs under each strategy. Green dashed line covers patients ultimately classified as
 566 having distress for referral.



568 **Table 1.** Patient-level summary statistics for the training set data derived from a retrospective
 569 cohort. Summaries are presented across the silver standard distress status (at least one encounter
 570 with a positive EHR-derived distress phenotype). P-values represent hypothesis tests across
 571 distress. Continuous variables were tested using Wilcoxon rank sum test and categorical variables
 572 with Chi-square exact test.

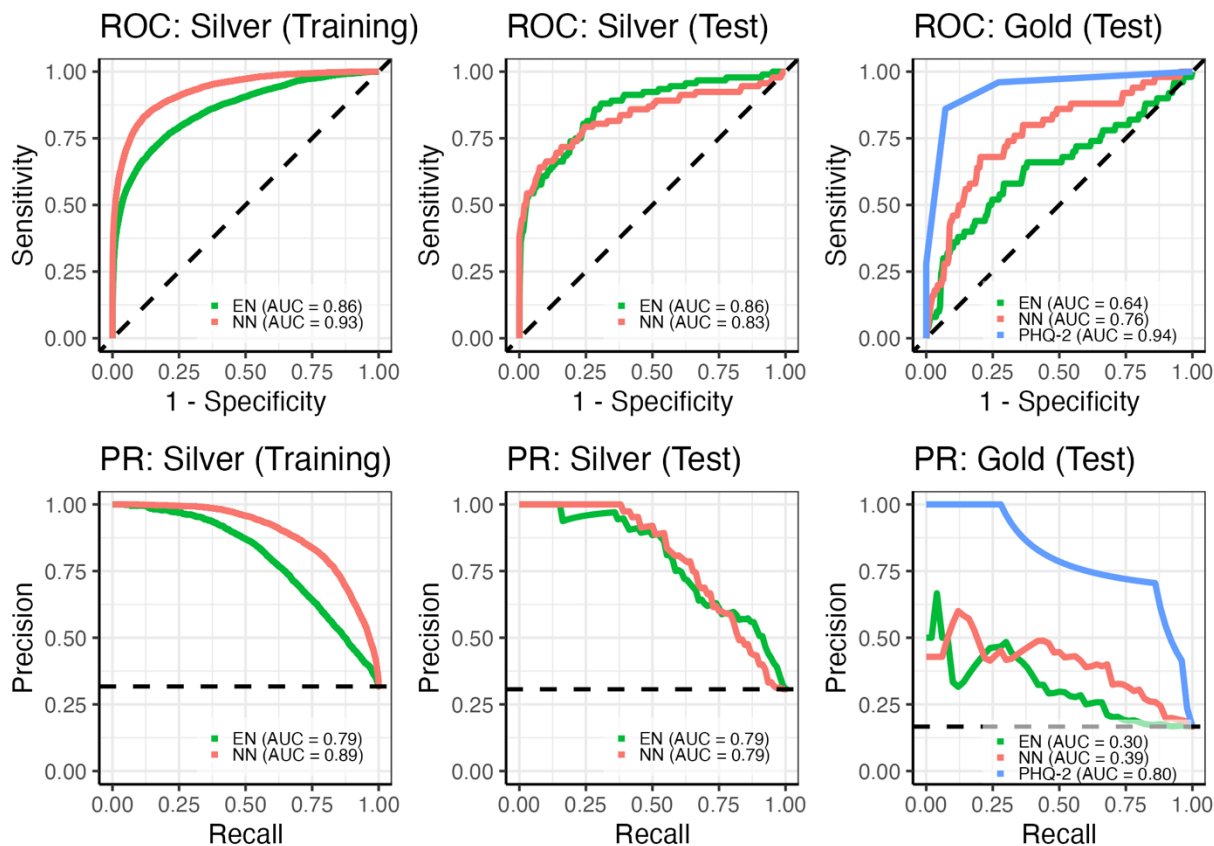
Variable	All	Distress	No Distress	P-Values
n	3,149	1,240 (39%)	1,909 (61%)	
Number of encounters				
Mean (SD)	9.86 (7.32)	9.92 (7.37)	9.82 (7.28)	
Median [min, max]	8 [1, 46]	8 [1, 41]	8 [1, 46]	
Follow-up in years, Mean (SD)	4.1 (2.50)	4.06 (2.49)	4.13 (2.51)	
Age (years)	68.57 (13.17)	69.08 (12.48)	68.23 (13.59)	0.075
Race (%)				0.002
White	1,651 (52%)	698 (56%)	953 (50%)	
Black/African American	1,337 (42%)	487 (39%)	850 (45%)	
Other	161 (5%)	55 (4%)	106 (6%)	
Sex (Female, %)	1,790 (57%)	788 (64%)	1,002 (52%)	<0.001
Ethnicity (Hispanic/Latino, %)	55 (2%)	20 (2%)	35 (2%)	0.679
Marital Status (Single, %)	1,479 (47%)	639 (52%)	840 (44%)	<0.001
Alcohol Use (%)	1,568(50%)	617(50%)	951(50%)	1.000
Smoking Use (%)	1,448(46%)	639(52%)	809(42%)	<0.001
Illicit Drug Use (%)	117(4%)	73(6%)	44(2%)	<0.001
High School Education in %, Mean (SD)	87 (0.13)	87 (0.12)	87 (0.13)	0.826
Annual Income in \$1k, Mean (SD)	32.28 (17.32)	32.47 (17.43)	32.16 (17.25)	0.628
Insurance				<0.001
Commercial (%)	611 (19%)	186 (15%)	425 (22%)	
Medicaid (%)	97 (3%)	47 (4%)	50 (3%)	
Medicare (%)	2,399 (76%)	995 (80%)	1,404 (74%)	
Uninsured (%)	42 (1%)	12 (1%)	30 (2%)	

574 **Table 2.** Summary statistics for the test set data derived from a prospective cohort. Summaries are
 575 presented across gold standard distress status (PHQ-8 > 6). P-values represent hypothesis tests
 576 across distress. Continuous variables were tested using Wilcoxon rank sum test and categorical
 577 variables with Chi-square test.

Variable	All	Distress	No Distress	P-Values
n	300	50 (17%)	250 (83%)	
PHQ-8, Mean (SD)	3.51 (3.99)	10.74 (3.89)	2.07 (1.89)	
PHQ-2				
Mean (SD)	0.71 (1.15)	2.52 (1.49)	0.34 (0.61)	
>0 (%)	39	96	27	
Age (years)	68.36 (12.45)	62.7 (16)	69.49 (11.32)	<0.001
Race (%)				0.064
Caucasian/white	242 (81%)	35 (70%)	207 (83%)	
Black/African American	46 (15%)	13 (26%)	33 (13%)	
Other	12 (4%)	2 (4%)	10 (4%)	
Sex (Female, %)	155 (52%)	29 (58%)	126 (50%)	0.355
Ethnicity (Hispanic/Latino, %)	5 (2%)	0 (0%)	5 (2%)	0.594
Marital Status (Single, %)	90 (30%)	24 (48%)	66 (26%)	0.004
Education (High school or lower, %)	73 (24%)	17 (34%)	56 (22%)	0.103
Income (Less than 100k/year)				0.010
%	139 (55%)	31 (74%)	108 (51%)	
Missing	48 (16%)	8 (16%)	40 (16%)	
Employment (%)				0.001
Retired	173 (58%)	25 (50%)	148 (59%)	
Employed	102 (34%)	13 (26%)	89 (36%)	
Disability	14 (5%)	7 (14%)	7 (3%)	
Other	11 (4%)	5 (10%)	6 (2%)	
IOP (mm Hg)				0.026
Mean (SD)	11.02 (3.59)	10.31 (3.79)	11.16 (3.54)	
Missing	2 (1%)	0 (0%)	2 (1%)	
Visual Acuity (logMAR)				0.962
Mean (SD)	0.18 (0.32)	0.26 (0.45)	0.16 (0.29)	
Missing	2 (1%)	0 (0%)	2 (1%)	
SAP MD (dB)				0.713
Mean (SD)	-4.07 (5.86)	-4.63 (6.19)	-3.95 (5.79)	
Missing	29 (10%)	6 (11%)	23 (9%)	
Canceled Appointment (%)				0.166
Not cancelled	249 (83%)	37 (74%)	212 (85%)	
Patient	34 (11%)	9 (18%)	25 (10%)	
Provider	17 (6%)	4 (8%)	13 (5%)	

578 dB = Decibel; IOP = intraocular pressure; mm Hg = Millimeters mercury; logMAR = logarithm of minimum angle of resolution;
 579 PHQ-2 = patient health questionnaire 2-item scale; PHQ-8 = PHQ 8-item scale; SAP MD = Standard automated perimetry mean
 580 deviation; SD = Standard deviation

581 **Figure 3.** Receiver operating characteristic (ROC; top row) and precision-recall (PR; bottom row)
582 curves for distress prediction models in training and test data. The left and middle columns show
583 performance for predicting the silver standard distress phenotype in the training and test sets,
584 respectively, comparing neural network (NN) and elastic net (EN) models. The right column shows
585 performance for predicting gold-standard distress (PHQ-8 > 6) in the test set, comparing NN, EN,
586 and PHQ-2 screening. Area under the curve (AUC) values are reported for each model and
587 outcome.



588
589

590 **Table 3.** Operating characteristics of manual, AI-only, and sequential screening strategies for
 591 identifying elevated distress (PHQ-8 > 6). Manual screening was defined as PHQ-2 > 0. AI-only
 592 strategies (neural network / elastic net) were evaluated under two different thresholding methods,
 593 applied to each predicted probability of distress: (1) thresholds selected to match the specificity
 594 of manual screening (specificity = 0.73) to enable direct comparison, and (2) thresholds selected
 595 using Youden’s index to optimize discrimination. Sequential screening used the AI model
 596 (threshold selected by Youden’s index) as a first-stage triage tool, with PHQ-2 administered only
 597 to patients identified as high risk. Proportion requiring screening denotes the fraction of
 598 participants receiving PHQ-2.

Screening Approach	Specificity	Sensitivity	PPV	NPV	Accuracy	Proportion Requiring Screening	Threshold
Manual							
PHQ-2	0.73	0.96	0.41	0.99	0.77	1.00	Score > 0
AI-only (specificity matched to PHQ-2)							
Neural Network	0.73	0.68	0.33	0.92	0.72	0.00	Probability ≥ 21%
Elastic Net	0.73	0.52	0.28	0.88	0.69	0.00	Probability ≥ 29%
AI-only (threshold based on Youden’s index)							
Neural Network	0.80	0.68	0.40	0.93	0.78	0.00	Probability ≥ 26%
Elastic Net	0.71	0.58	0.28	0.89	0.69	0.00	Probability ≥ 28%
Sequential							
Neural Network	0.93	0.64	0.64	0.93	0.88	0.28	Probability ≥ 26%
Elastic Net	0.90	0.54	0.52	0.91	0.84	0.34	Probability ≥ 28%

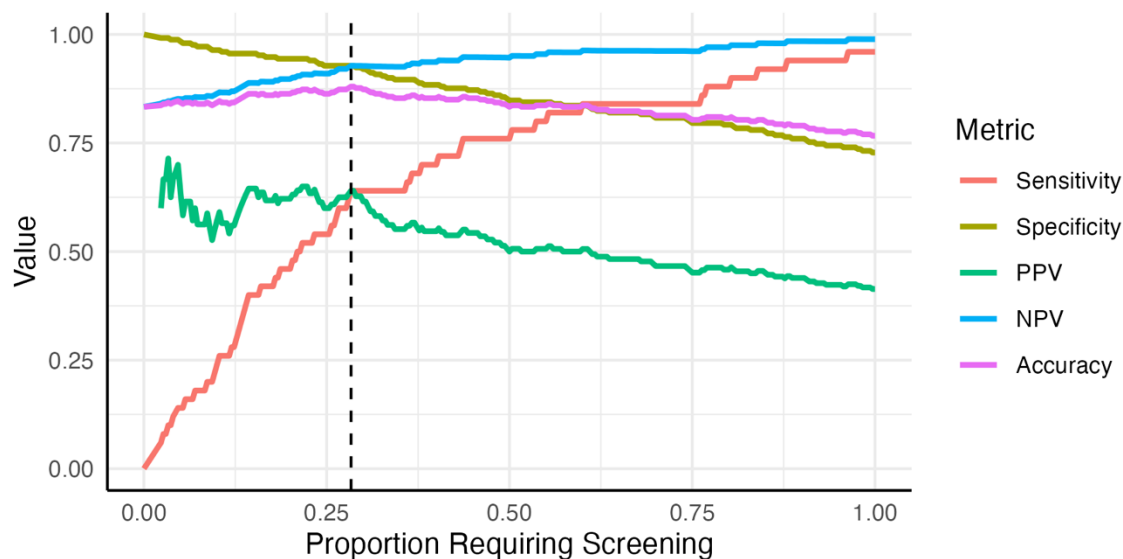
599 NPV = Negative Predicted Value; PHQ-2 = Patient Health Questionnaire 2-item scale; PPV = Positive Predicted Value

600 **Table 4.** Operating characteristics of the neural network sequential screening strategy by
601 demographic subgroup. Counts, number of patients with distress, and distress prevalence are also
602 reported for each subgroup.

Variable	n	Distress	Specificity	Sensitivity	PPV	NPV	Accuracy
Gender							
Male	145	21 (14%)	93	48	53	91	86
Female	155	29 (19%)	93	76	71	94	90
Race							
White	242	35 (14%)	91	71	58	95	88
Black/African American	46	13 (28%)	100	46	100	83	85
Other	12	2 (17%)	100	50	100	91	92
Age (years)							
<50	23	9 (39%)	93	67	86	81	83
50-60	33	6 (18%)	96	40	67	88	86
60-70	97	19 (20%)	94	75	75	94	90
70-80	130	15 (12%)	92	57	47	94	88
>80	37	5 (14%)	91	60	50	94	86

603 NPV = Negative Predicted Value; PPV = Positive Predicted Value.

604 **Figure 4.** Operating characteristics of the neural network sequential screening strategy as a
605 function of the proportion requiring manual screening. The vertical dashed line indicates the
606 threshold corresponding to Youden's index, which was used to compute the metrics in **Tables 3**
607 and **4**.



608

609 PPV = Positive Predicted Value; NPV = Negative Predicted Value.

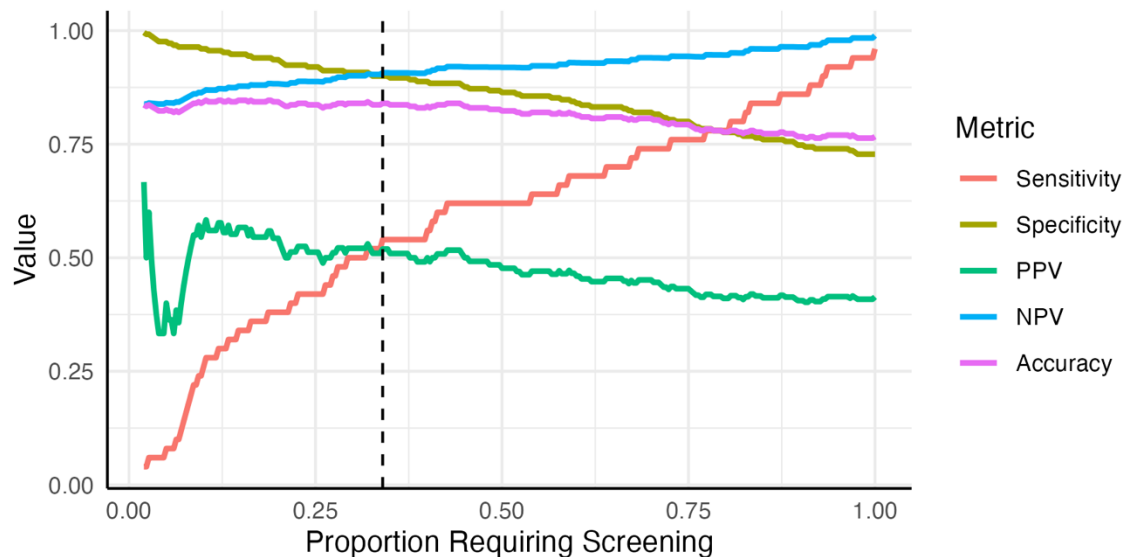
610

611 **Supplementary Table S1.** Operating characteristics of the elastic net sequential screening strategy
 612 by demographic subgroup. Counts, number of patients with distress, and distress prevalence are
 613 also reported for each subgroup.

Variable	n	Distress	Specificity	Sensitivity	PPV	NPV	Accuracy
Gender							
Male	145	21 (14%)	93	29	40	88	83
Female	155	29 (19%)	87	72	57	93	85
Race							
White	242	35 (14%)	89	57	47	92	84
Black/African American	46	13 (28%)	94	46	75	82	80
Other	12	2 (17%)	100	50	100	91	92
Age (years)							
<50	23	9 (39%)	93	44	80	72	74
50-60	33	6 (18%)	91	40	50	88	82
60-70	97	19 (20%)	91	56	60	90	84
70-80	130	15 (12%)	90	43	35	92	85
>80	37	5 (14%)	88	100	56	100	89

614 NPV = Negative Predicted Value; PPV = Positive Predicted Value.

615 **Supplementary Figure S1.** Operating characteristics of the elastic net sequential screening
616 strategy as a function of the proportion requiring manual screening. The vertical dashed line
617 indicates the threshold corresponding to Youden's index, which was used to compute the metrics
618 in **Tables 3** and **S1**.



619

620 NPV = Negative Predicted Value; PPV = Positive Predicted Value.

621 **Supplementary Table S2.** Predictors with the largest absolute Shapley values for positive
 622 prediction. All utilization variables presented used a 5-year window prior to the encounter. Also
 623 presented are the number of encounters with each predictor present, for both encounters with and
 624 without distress, along with corresponding proportions.

Variable	Distress	No Distress	Shapley Value
Diagnosis Group: Anxiety Disorders	3,818 (0.39)	870 (0.04)	0.21
Diagnosis Group: Other Nervous System Disorders	5,081 (0.52)	5,899 (0.28)	0.15
Diagnosis Group: Glaucoma	9,227 (0.94)	19,626 (0.93)	0.14
Procedure Group: Ophthalmologic And Otologic Diagnosis And Treatment	9,654 (0.98)	20,722 (0.98)	0.14
Diagnosis Group: Other Eye Diagnosis Group	6,820 (0.69)	13,045 (0.62)	0.13
Diagnosis Group: Unclassified	7,884 (0.80)	11,455 (0.54)	0.11
Encounter Type: General Medicine	8,523 (0.87)	15,339 (0.72)	0.10
Medication: Antithrombotic Agents	3,481 (0.35)	5,465 (0.26)	0.08
Procedure Group: Medications (Injections, Infusions And Other Forms)	6,318 (0.64)	9,358 (0.44)	0.07
Medication: Antibacterials For Systemic Use	7,676 (0.78)	12,976 (0.61)	0.07
Medication: Nasal Preparations	8,289 (0.84)	15,818 (0.75)	0.06
Diagnosis Group: Other Non-Traumatic Joint Disorders	6,181 (0.63)	8,614 (0.41)	0.06
Diagnosis Group: Genitourinary Symptoms And Ill-Defined Conditions	4,168 (0.42)	4,927 (0.23)	0.05
Medication: Antiemetics And Antinauseants	2,295 (0.23)	1,957 (0.09)	0.05
Medication: Beta Blocking Agents	6,185 (0.63)	11,939 (0.56)	0.05
Race: Black/African-American	3,846 (0.39)	9,284 (0.44)	0.05
Medication: Ophthalmologicals	9,208 (0.94)	18,546 (0.87)	0.05
Medication: Throat Preparations	2,734 (0.28)	2,804 (0.13)	0.05
Encounter Type: Wellness	2,548 (0.26)	2,175 (0.10)	0.05
Problem List: Anxiety	1,502 (0.15)	225 (0.01)	0.04
Diagnosis Group: Mood Disorders	2,915 (0.30)	439 (0.02)	0.04
Diagnosis Group: Retinal Detachments; Defects; Vascular Occlusion; And Retinopathy	3,482 (0.35)	7,290 (0.34)	0.03
Diagnosis Group: Esophageal Disorders	4,848 (0.49)	5,625 (0.27)	0.03
Diagnosis Group: Other Lower Respiratory Disease	5,435 (0.55)	6,324 (0.30)	0.03
Diagnosis Group: Immunizations And Screening For Infectious Disease	6,158 (0.63)	9,526 (0.45)	0.03

625

626

627 **Supplementary Table S3.** Demographic predictors with the largest absolute Shapley values for
628 positive prediction.

Variable	Distress	No Distress	Shapley Value
Race: Black/African-American	3,846 (0.39)	9,284 (0.44)	0.05
Problem List: Anxiety (5y)	1,611 (0.16)	231 (0.01)	0.04
Problem List: Depression (5y)	1,560 (0.16)	225 (0.01)	0.03
Marital Status: Single	5,079 (0.52)	9,193 (0.43)	0.01
Problem List: Anxiety (1y)	295 (0.03)	46 (0.00)	0.01
Gender: Male	3,459 (0.35)	10,187 (0.48)	-0.01
Insurance Status: Uninsured	453 (0.05)	1341 (0.06)	0.01
Alcohol Use: Yes	4,934 (0.50)	10,442 (0.49)	0.00
Age (Years)	71.51 (12.48)	70.38 (12.86)	0.00
Problem List: Depression (1y)	321 (0.03)	33 (0.00)	0.00

629

630 **Supplementary Table S4.** Odds ratios (ORs) for the top 25 predictors of distress for elastic net,
 631 ordered by coefficient absolute values. In parentheses is the time window prior the encounter that
 632 the variable corresponds to, one year (1y) or five years (5y).

Variable	Distress	No Distress	OR
Problem List: Depression (5y)	1,497 (0.15)	225 (0.01)	4.15
Problem List: Anxiety (5y)	1,502 (0.15)	225 (0.01)	2.93
Diagnosis Group: Anxiety disorders (5y)	3,818 (0.39)	870 (0.04)	1.23
Gender: Male	3,459 (0.35)	10,187 (0.48)	0.85
Drug Use: Yes	588 (0.06)	526 (0.02)	1.16
Smoking: Yes	5,034 (0.51)	8,817 (0.42)	1.13
Diagnosis Group: Mood Disorders (5y)	2,915 (0.30)	439 (0.02)	1.11
Diagnosis Group: Chest Pain (5y)	3,896 (0.40)	4,030 (0.19)	1.03
Encounter Type: Emergency Medicine (1y)	3,200 (0.33)	2,948 (0.14)	1.03
Medication: Psychoanaleptics (5y)	4,797 (0.49)	1,584 (0.07)	1.03
Procedure Group: Other non-OR therapeutic procedures, male genital (5y)	76 (0.01)	19 (0.00)	1.03
Diagnosis Group: Chronic obstructive pulmonary disease and bronchiectasis (1y)	620 (0.06)	363 (0.02)	1.02
Encounter Type: General Medicine (3m)	5,062 (0.51)	7,027 (0.33)	1.02
Diagnosis Group: Unclassified (1y)	5,084 (0.52)	6,062 (0.29)	1.02
Diagnosis Group: Substance-related disorders (5y)	647 (0.07)	192 (0.01)	1.02
Encounter Type: Wellness (5y)	2,548 (0.26)	2,175 (0.10)	1.02
Encounter Type: Radiology (5y)	6,966 (0.71)	10,841 (0.51)	1.02
Encounter Type: Infectious Diseases (5y)	425 (0.04)	358 (0.02)	1.01
Procedure Group: Spinal fusion (5y)	570 (0.06)	365 (0.02)	1.01
Procedure Group: Esophageal dilatation (5y)	198 (0.02)	57 (0.00)	1.01
Procedure Group: Cancer chemotherapy (3m)	6,450 (0.66)	9,878 (0.47)	1.01
Procedure Group: Other OR therapeutic procedures on skin and breast (5y)	2,691 (0.27)	2,304 (0.11)	1.01
Encounter Type: Infectious Diseases (1y)	208 (0.02)	101 (0.00)	1.01
Diagnosis Group: Mood disorders (1y)	1,282 (0.13)	130 (0.01)	1.01
Diagnosis Group: Fluid and electrolyte disorders (5y)	1,301 (0.13)	828 (0.04)	1.01

633