Nature Communications

https://doi.org/10.1038/s41467-026-70891-5

Article in Press

Routine blood tests and machine learning identify complications in high myopia

Received: 4 June 2025

Accepted: 8 March 2026

Cite this article as: Li, S., Ren, J., Wang, F. *et al.* Routine blood tests and machine learning identify complications in high myopia. *Nat Commun* (2026). https://doi.org/10.1038/s41467-026-70891-5

Shengjie Li, Jun Ren, Fenglin Wang, Jianing Wu, Yingzhu Li, Xuanxuan Wang, Mengyu Zhang, Henggui Hu, Yunxiao Song, Wenjun Cao, Xingtao Zhou & Meiyan Li

Routine Blood Tests and Machine Learning Identify Complications in High Myopia

Shengjie Li[1,2,3,4#*], Jun Ren[1#], Fenglin Wang[5], Jianing Wu[1], Yingzhu Li[1], Xuanxuan Wang[6], Mengyu Zhang[7], Henggui Hu[7], Yunxiao Song[8], Wenjun Cao[1,2,3,4*], Xingtao Zhou[2,3,4*], Meiyan Li[2,3,4*]

1 Department of Clinical Laboratory, Eye & ENT Hospital, Fudan University, Shanghai, 200031, China

2 Eye Institute and Department of Ophthalmology, Eye & ENT Hospital, Fudan University, Shanghai 200031, China

3 NHC Key laboratory of Myopia and Related Eye Diseases; Key Laboratory of Myopia and Related Eye Diseases, Chinese Academy of Medical Sciences, Shanghai, 200031, China

4 Shanghai Key Laboratory of Visual Impairment and Restoration, Shanghai, 200031, China

5 College of Life Sciences, Nankai University, Tianjin, 300071, China.

6 Department of Clinical Laboratory, the First Affiliated Hospital of Anhui Medical University, Anhui, 230022, China

7 Department of Clinical Laboratory, Anhui Wanbei Electricity Group General Hospital, Suzhou 234000, China

8 Department of Clinical Laboratory, Shanghai Xuhui Central Hospital, Fudan University, Shanghai, 200031, China

[#]These authors (Shengjie Li, and Jun Ren) contributed equally.

*Corresponding author:

Shengjie Li, Eye & ENT Hospital, Shanghai Medical College, Fudan University, Fenyang Road 83th, Shanghai 200031, China. E-mail: lishengjie6363020@163.com

Meiyan Li, Eye & ENT Hospital, Shanghai Medical College, Fudan University, Fenyang Road 83th, Shanghai 200031, China. E-mail: limeiyan0406073@126.com

Wenjun Cao, Eye & ENT Hospital, Shanghai Medical College, Fudan University, Fenyang Road 83th, Shanghai 200031, China. E-mail: wgkjyk@aliyun.com

Xingtao Zhou, Eye & ENT Hospital, Shanghai Medical College, Fudan University, Fenyang Road 83th, Shanghai 200031, China. E-mail: doctzhouxingtao@163.com

Abstract

High myopia can lead to cataract, glaucoma, retinal detachment, choroidal neovascularization, and macular degeneration, causing irreversible vision loss. Imaging detects these complications, but population screening is limited by equipment, and specialist availability. Here we show that a machine learning model using routine blood test results identifies people at increased risk of complications related to high myopia during standard health examinations. We develop the model in a multicentre study of 10,661 participants and validate it in two independent cohorts. The model shows high accuracy across centres (area under the receiver operating characteristic curve=0.9010-0.9649) and flags individuals who receive a clinical diagnosis in a hospital-based prospective follow-up study of 5,067 participants. In a community screening study of 311,254 adults, the model increases the yield of detected complications among those referred for ophthalmic assessment (positive predictive value=74%). This scalable blood-based approach supports opportunistic screening and earlier referral in primary care and community settings.

Introduction

Myopia is a global epidemic, with high myopia identified as a growing concern because of its associated risk of complications[1,2]. Myopia, particularly high myopia, is not merely a refractive error but is associated with markedly increased risks of vision-threatening complications[3,4], including cataracts (3–8-fold[5,6]), glaucoma (~7-fold[7]), retinal detachment (12–39[8]), myopic macular degeneration (~20[9,10]), and choroidal neovascularization (~10[8]), compared with non-myopic eyes. While the risk of complications increases with increasing myopia, the population impact varies: lower degrees are more prevalent and thus contribute greatly to the overall burden, whereas high myopia, although less common, is associated with a much higher individual risk and accounts for more than half of pathological myopia. Some complications, such as retinal detachment and glaucoma, may benefit from early detection and timely intervention (e.g., prophylactic laser for high-risk retinal lesions), whereas effective treatments for myopic macular degeneration are lacking, and this condition is the leading cause of irreversible blindness in working-age populations, particularly in urbanized regions[11]. The substantial economic burden associated with these complications and their impacts on quality of life underscore the critical need for effective early screening strategies[12].

Current diagnostic paradigms rely heavily on specialized imaging modalities such as optical coherence tomography (OCT), fundus autofluorescence, and perimetry. While these techniques provide detailed structural information of the eyes, they present substantial limitations in real-world clinical practice. First, considerable barriers exist to accessing these technologies, with approximately 78% of low- and middle-income countries lacking basic OCT infrastructure[13]. Second, considerable variability has been documented in clinician interpretations of the imaging findings, with studies reporting up to 22% variability in glaucoma detection using OCT parameters[14]. Third, the resource-intensive nature of these examinations limits their applicability for population-level screening, resulting in delayed diagnoses, a critical concern given that nearly 40% of myopic macular lesions are asymptomatic in the early stages[15]. Recent advances in portable fundus cameras and smartphone-based imaging offer cost-effective, accessible, and point-of-care alternatives for retinal assessment[16,17]. These technologies facilitate the early detection and monitoring of

retinal pathologies, including diabetic retinopathy, age-related macular degeneration, and retinal vascular disorders, thereby improving accessibility to essential diagnostic services. Nevertheless, challenges remain regarding image quality, standardization, regulatory approval, and medicolegal considerations, highlighting the continued need for complementary screening approaches such as blood-based risk stratification models.

Emerging evidence has demonstrated the underutilized potential of routine clinical blood tests in ocular risk assessments. Haematological markers such as the lymphocyte count, basophil count, mean platelet volume, and the neutrophil-to-lymphocyte ratio can predict the risk of retinal detachment[18,19], whereas metabolic parameters such as the level of haemoglobin A1c (HbA1c) are correlated with maculopathy progression[20], and higher levels of inflammatory markers such as IL-6 are clearly associated with an increased glaucoma risk[21]. These findings suggest that routinely available clinical laboratory data may provide valuable insights into the risks of myopia-related complications; nevertheless, they remain systematically underused in current screening protocols. From a cost perspective, blood-based screening is also highly efficient. In health examination centres and primary care settings, the required haematological parameters are already determined through routine blood tests, enabling opportunistic screening for high myopia-related complications at no additional expense. Even if blood tests were performed specifically for ocular screening, the median cost is approximately 300 RMB per person, which is lower than that of fundus-based imaging (400–600 RMB). Additionally, blood tests provide concurrent systemic health information, offering dual-purpose clinical utility. These features underscore blood-based models as a complementary, scalable, and low-cost strategy for identifying high myopia-related complications.

Routine clinical laboratory tests are among the most common diagnostic methods used in medical practice and provide valuable and timely insights into various diseases[22]. As blood cells interact with different tissues and organs, these tests can serve as powerful diagnostic tools. Their key advantages—low cost, accessibility, consistency, and widespread use in primary healthcare settings—make them indispensable in routine medical evaluations[23]. The integration of machine learning (ML) with routinely available clinical laboratory data represents a transformative approach to address the challenges

mentioned above. Recent advances have shown that ML algorithms can effectively predict high myopia-related complications by leveraging routinely available blood data[18,19,24]. These technological advancements could help address the urgent need for scalable screening tools, particularly in resource-limited settings, by enabling opportunistic screening during routine health examinations, where the burden of high myopia–related complications is increasing most rapidly.

This large-scale study was conducted in three phases—a retrospective multicentre case–control study, a hospital-based prospective cohort study, and a community-based cross-sectional study—and leveraged routine blood test data combined with advanced machine learning architectures to establish a paradigm for opportunistic screening during routine health examinations for high myopia–related complications. In the retrospective multicentre case–control phase, we analysed data from a diverse cohort of 15,000 patients across five tertiary care centres in Asia to develop a robust predictive model integrating haematological, metabolic, and inflammatory markers. The performance of the model was validated through two large sample cohort studies: a hospital-based prospective cohort study and a community-based cross-sectional study. We hypothesize that blood-based screening during routine health examinations will be clinically advantageous over a direct ophthalmic examination in settings where specialized imaging is unavailable or impractical, particularly in health examination centres and community primary care facilities. This approach prioritizes scalability and generalizability, aligns with the principles of precision medicine, and holds promise for reducing the global burden of preventable blindness associated with high myopia.

Results

Overview of the case–control study

The structure and methodology of the case–control study are outlined in Figure 1. Between Jan. 1, 2016, and Dec. 30, 2024, 10,661 individuals who met the inclusion criteria were enrolled at one of five hospitals in China. The Eye & ENT cohort (discovery cohort) included 5,102 participants. The first external validation cohort included 3,125 participants from Xuhui Central Hospital of Fudan University and Putuo People's

Hospital, and the second external validation cohort included 2,434 participants from Wanbei Coal–Electricity Group General Hospital and the First Affiliated Hospital of Anhui Medical University. The numbers of patients with high myopia overall and associated with cataracts, macular degeneration, choroidal neovascularization, glaucoma, and retinal detachment in the discovery cohort, validation cohort 1, and validation cohort 2 are illustrated in Figure 1. Across all cohorts, patients with uncomplicated high myopia were generally younger than those with complicated high myopia, with the differences reaching statistical significance in the discovery cohort, validation cohort 1, and validation cohort 2 ($P < 0.001$). No significant differences in the sex distribution were observed between the uncomplicated and complicated high myopia groups in any cohort ($P > 0.05$). Detailed information about the variables of interest in the discovery cohort, validation cohort 1, and validation cohort 2 is presented in Tables S1–S3. Approximately 70% of the routine laboratory parameters differed significantly among the six groups (the overall high myopia group and the five complication groups). Effect sizes were calculated as follows to better assess their potential clinical relevance: η^2 for continuous variables from one-way ANOVA (0.01 = small effect, 0.06 = medium effect, and ≥0.14 = large effect) and Cramér's V for categorical variables (0.1 = small effect, 0.3 = medium effect, and ≥0.5 = large effect). Notably, 38.71% (discovery cohort), 31.15% (validation cohort 1), and 37.7% (validation cohort 2) of the parameters showed a medium or large effect size, suggesting potential clinical relevance beyond statistical significance.

Development of screening models based on all blood test variables

In the discovery cohort, all 61 variables from the routine clinical blood tests were employed to train screening models based on the 10 ML algorithms. The performance of the models developed from the different ML algorithms is shown in Figure 2A, and the detailed results are provided in Figure S1 and Table S4. In terms of the area under the receiver operating characteristic curve (AUC) (Figure 2B), the performance of the random forest (RF) (macroaverage AUC = 0.9707 [95% CI=0.9480–0.9891]) and extreme gradient boosting (XGBoost) (macroaverage AUC = 0.9601 [95% CI=0.9452–0.9806]) models was significantly better (Table S5; $P < 0.05$) than that of the other models according to the DeLong test, whose macroaverage AUC values ranged from 0.6007 to 0.9561. The results of the DeLong test (Table S5) also showed no significant

difference (P = 0.184) between the RF and XGBoost algorithms.

As shown in Figure 2C, the RF model achieved AUCs of 0.9725 (95% CI = 0.9523–0.9901), 0.9782 (95% CI = 0.9581–0.9940), 0.9752 (95% CI = 0.9539–0.9918), 0.9721 (95% CI = 0.9504–0.9888), 0.9617 (95% CI = 0.9391–0.9815), and 0.9643 (95% CI = 0.9422–0.9834); negative predictive value (NPV)s of 0.9249 (95% CI = 0.8972–0.9489), 0.9333 (95% CI = 0.9041–0.9568), 0.9631 (95% CI = 0.9397–0.9843), 0.9698 (95% CI = 0.9451–0.9895), 0.9867 (95% CI = 0.9701–0.9971), and 0.9813 (95% CI = 0.9595–0.9962); and positive predictive value (PPV)s of 0.7626 (95% CI = 0.7271–0.7942), 0.7922 (95% CI = 0.7581–0.8239), 0.9072 (95% CI = 0.8773–0.9348), 0.9516 (95% CI = 0.9278–0.9735), 0.9231 (95% CI = 0.8929–0.9499), and 0.8868 (95% CI = 0.8535–0.9173) for identifying high myopia overall and high myopia associated with cataracts, macular degeneration, choroidal neovascularization, glaucoma, and retinal detachment, respectively. A confusion matrix (Figure 2D) was created to visualize the performance of the RF model, revealing sensitivities of 0.8133 (95% CI = 0.7784–0.8469), 0.8966 (95% CI = 0.8681–0.9248), 0.7586 (95% CI = 0.7212–0.7929), 0.7108 (95% CI = 0.6749–0.7450), 0.6857 (95% CI = 0.6491–0.7192), and 0.7581 (95% CI = 0.7228–0.7906) and specificities of 0.9008 (95% CI = 0.8722–0.9271), 0.8603 (95% CI = 0.8251–0.8919), 0.9878 (95% CI = 0.9715–0.9970), 0.9961 (95% CI = 0.9827–0.9996), 0.9976 (95% CI = 0.9849–0.9997), and 0.9924 (95% CI = 0.9739–0.9991) for identifying high myopia overall and high myopia associated with cataracts, macular degeneration, choroidal neovascularization, glaucoma, and retinal detachment, respectively.

The XGBoost model (Figure 2E) achieved AUCs of 0.9620 (95% CI = 0.9526–0.9713), 0.9728 (95% CI = 0.9657–0.9799), 0.9694 (95% CI = 0.9592–0.9787), 0.9588 (95% CI = 0.9393–0.9752), 0.9595 (95% CI = 0.9331–0.9829), and 0.9720 (95% CI = 0.9568–0.9856); NPVs of 0.9174 (95% CI = 0.8983–0.9356), 0.9122 (95% CI = 0.8917–0.9328), 0.9589 (95% CI = 0.9467–0.9711), 0.9646 (95% CI = 0.9536–0.9747), 0.9868 (95% CI = 0.9796–0.9928), and 0.9802 (95% CI = 0.9718–0.9880); and PPVs of 0.7300 (95% CI = 0.6868–0.7755), 0.7746 (95% CI = 0.7405–0.8091), 0.8416 (95% CI = 0.7757–0.8990), 0.8594 (95% CI = 0.7879–0.9265), 1.000 (95% CI = 1.0000–1.0000), and 0.9388 (95% CI = 0.8824–0.9864) for identifying high myopia overall and high myopia associated with cataracts, macular degeneration, choroidal

neovascularization, glaucoma, and retinal detachment, respectively. The confusion matrix (Figure 2F) revealed sensitivities of 0.7967 (95% CI = 0.7520–0.8373), 0.8621 (95% CI = 0.8312–0.8931), 0.7328 (95% CI = 0.6667–0.8000), 0.6627 (95% CI = 0.5750–0.7471), 0.6857 (95% CI = 0.5428–0.8065), and 0.7419 (95% CI = 0.6469–0.8298) and specificities of 0.8846 (95% CI = 0.8618–0.9057), 0.8510 (95% CI = 0.8254–0.8748), 0.9784 (95% CI = 0.9689–0.9866), 0.9884 (95% CI = 0.9818–0.9936), 1.000 (95% CI = 1.0000–1.0000), and 0.9962 (95% CI = 0.9925–0.9988) for identifying high myopia overall and high myopia associated with cataracts, macular degeneration, choroidal neovascularization, glaucoma, and retinal detachment, respectively.

Additionally, the area under the precision–recall curve (PR-AUC) values of the RF (0.8917 [95% CI = 0.8786–0.9108]) and XGBoost (0.8788 [95% CI = 0.8514–0.8906]) models (Figure S2) were significantly higher ($P < 0.05$) than those of the other models (Figure S2) according to the DeLong test, confirming the superior performance of the models developed based on these two algorithms. Given these results, the RF and XGBoost algorithms were selected for the subsequent steps in the development of the screening model.

Identification of the final model developed based on nine features

Feature reduction was performed on the data from the discovery cohort to optimize the performance of the models and increase their clinical applicability. The number of features was sequentially reduced from 61 to 1 based on their importance rankings to determine the minimum number of features that yielded optimal predictive capability.

The all-feature RF model was significantly better than the 8-feature model ($P = 0.0227$) but not significantly better than the 9-feature model ($P = 0.0970$) (Figure 3A, Table S6). In other words, the macroaverage changes in the AUC value indicated that the RF model incorporating nine features (EOS, PCT, BASO%, M, UA, LPB, BASO, EOS%, and TG levels; macroaverage AUC = 0.9649 [95% CI=0.9528–0.9770]) retained nearly the highest predictive ability. For the XGBoost algorithm, the all-feature model was significantly better than the 19-feature model ($P < 0.0010$, macroaverage AUC = 0.9493) but not significantly better than the 20-feature model ($P = 0.1860$, macroaverage AUC = 0.9621) (Figure 3A, Table S6). Based on the AUC (0.9649 vs. 0.9621) and the number of included features (9 vs. 20), the nine-feature RF model was selected

as the final screening model.

The global SHapley Additive exPlanations (SHAP) analysis provided insights into the overall behaviour of the RF model, with SHAP summary plots (Figure 3B) illustrating the ranked contributions of each feature based on their average SHAP values in descending order. The top nine contributing features were the EOS, PCT, BASO%, M, UA, LPB, BASO, EOS%, and TG levels. The distributions and correlations of these nine features are presented in Figure S3A. Strong correlations were observed between the EOS and EOS% and between the BASO and BASO%, whereas no significant correlations were identified among the other variables.

The 9-feature RF model achieved a macroaverage AUC of 0.9649 (95% CI=0.9528–0.9770) (Figure 3C) and an PR-AUC of 0.8812 (95% CI=0.8045–0.9579) (Figure 3D). At the subgroup level, the RF model achieved AUCs of 0.9736 (95% CI = 0.9480–0.9880), 0.9790 (95% CI = 0.9342–0.9885), 0.9705 (95% CI = 0.9272–0.9772), 0.9589 (95% CI = 0.9348–0.9706), 0.9475 (95% CI = 0.8976–0.9788), and 0.9599 (95% CI = 0.9214–0.9746); NPVs of 0.9387 (95% CI = 0.8915–0.9926), 0.9333 (95% CI = 0.8821–0.9692), 0.9650 (95% CI = 0.9150–0.9797), 0.9669 (95% CI = 0.9465–0.9759), 0.9867 (95% CI = 0.9478–0.9974), and 0.9824 (95% CI = 0.9523–0.9894); and PPVs of 0.8602 (95% CI = 0.8075–0.9178), 0.8237 (95% CI = 0.7660–0.8618), 0.7860 (95% CI = 0.7436–0.8160), 0.7451 (95% CI = 0.6978–0.7964), 0.7273 (95% CI = 0.6941–0.7578), and 0.7869 (95% CI = 0.7352–0.8387) for identifying high myopia overall and high myopia associated with cataracts, macular degeneration, choroidal neovascularization, glaucoma, and retinal detachment, respectively (Figure 3C). The confusion matrix (Figure 3E) revealed sensitivities of 0.8423 (95% CI = 0.7987–0.8823), 0.8934 (95% CI = 0.8398–0.9297), 0.7759 (95% CI = 0.7182–0.8305), 0.6867 (95% CI = 0.6486–0.7280), 0.6857 (95% CI = 0.6513–0.7428), and 0.7742 (95% CI = 0.7265–0.8065) and specificities of 0.9463 (95% CI = 0.8967–0.9943), 0.8864 (95% CI = 0.8383–0.9175), 0.9689 (95% CI = 0.9257–0.9755), 0.9832 (95% CI = 0.9416–0.9940), 0.9915 (95% CI = 0.9546–1.0000), and 0.9849 (95% CI = 0.9517–0.9929) for identifying high myopia overall and high myopia associated with cataracts, macular degeneration, choroidal neovascularization, glaucoma, and retinal detachment, respectively. The optimal cut-off values of the final RF model for the detection of high myopia (Figure S4A) and high myopia

associated with cataracts (Figure S4B), macular degeneration (Figure S4C), choroidal neovascularization (Figure S4D), glaucoma (Figure S4E), and retinal detachment (Figure S4F) were 0.47, 0.62, 0.16, 0.18, 0.11, and 0.12, respectively. A detailed description of the performance, including sensitivity, specificity, PPV, NPV, accuracy, and F1 score, of the final RF model is provided in Table 1.

We subsequently performed a subgroup analysis after excluding patients with active systemic inflammatory conditions (n= 126), recent infections (n=240), immunosuppressive medication use (n= 64), diabetes (n= 277), or hypertension (n=455), conditions that could affect haematological biomarkers to assess the performance of the final RF model.

The final RF model achieved a macroaverage AUC of 0.9560 (95% CI: 0.9236–0.9884), which was not significantly different from that of the model including all patients (DeLong test, P = 0.056). As shown in Figure S5A, the RF model achieved AUCs of 0.9701 (95% CI = 0.9377–1.0000), 0.9751 (95% CI = 0.9427–1.0000), 0.9685 (95% CI = 0.9361–1.0000), 0.9536 (95% CI = 0.9212–0.9860), 0.9236 (95% CI = 0.8912–0.9560), and 0.9423 (95% CI = 0.9099–0.9747) for identifying high myopia overall and high myopia associated with cataracts, macular degeneration, choroidal neovascularization, glaucoma, and retinal detachment, respectively. Other performance metrics, including sensitivity, specificity, PPV, NPV, accuracy, and F1 score, are summarized in Table S7, with macroaverage values of 0.7639, 0.9557, 0.9229, 0.9625, 0.9250, and 0.8250, respectively.

External validation of the final RF model

Models developed in the discovery cohort require validation with independent validation cohorts to ensure robustness. Therefore, two external validation cohorts were used to evaluate the performance of the 9-feature RF model. In validation cohort 1, the RF model achieved a macroaverage AUC of 0.9020 (95% CI = 0.8815–0.9225; Figure 4A) and an PR-AUC of 0.7797 (95% CI = 0.7091–0.8504; Figure 4B). As shown in Figure 4A, the RF model achieved AUCs of 0.9021 (95% CI = 0.8543–0.9505), 0.9133 (95% CI = 0.8626–0.9640), 0.8951 (95% CI = 0.8505–0.9437), 0.8948 (95% CI = 0.8629–0.9390), 0.9324 (95% CI = 0.8739–0.9649), and 0.8745 (95% CI = 0.8319–0.9193) for identifying high myopia overall and high myopia associated with cataracts, macular degeneration, choroidal neovascularization, glaucoma, and retinal

detachment, respectively. The confusion matrix (Figure 4C) further illustrates the model's performance, revealing sensitivities of 0.7229 (95% CI = 0.6743–0.7656), 0.8482 (95% CI = 0.7973–0.9016), 0.5888 (95% CI = 0.5396–0.6430), 0.5840 (95% CI = 0.5208–0.6374), 0.7481 (95% CI = 0.6910–0.7890), and 0.5592 (95% CI = 0.5028–0.6036) and specificities of 0.8702 (95% CI = 0.8292–0.9084), 0.7610 (95% CI = 0.7012–0.8011), 0.9864 (95% CI = 0.9422–0.9943), 0.9886 (95% CI = 0.9546–0.9958), 0.9996 (95% CI = 0.9757–1.0000), and 0.9940 (95% CI = 0.9655–1.0000) for identifying high myopia overall and high myopia associated with cataracts, macular degeneration, choroidal neovascularization, glaucoma, and retinal detachment, respectively.

The distributions and correlations of the nine blood test variables in validation cohort 1 are presented in Figure S3B. Strong correlations were observed between the EOS and EOS% and between BASO and BASO%, whereas no significant correlations were found among the other variables, suggesting limited multicollinearity and supporting the independent contributions of these features to model performance.

Similar results were observed in validation cohort 2, where the RF model achieved a macroaverage AUC of 0.9010 (95% CI = 0.8761–0.9259; Figure 4D) and an PR-AUC of 0.7841 (95% CI = 0.7057–0.8625; Figure 4E). As shown in Figure 4D, the RF model achieved AUCs of 0.9064 (95% CI = 0.8604–0.9370), 0.9164 (95% CI = 0.8678–0.9569), 0.8988 (95% CI = 0.8520–0.9389), 0.8752 (95% CI = 0.8167–0.9216), 0.9351 (95% CI = 0.8876–0.9651), and 0.8743 (95% CI = 0.8440–0.9277) for identifying high myopia overall and high myopia associated with cataracts, macular degeneration, choroidal neovascularization, glaucoma, and retinal detachment, respectively. The confusion matrix (Figure 4F) confirmed these findings, with sensitivities of 0.7456 (95% CI = 0.7152–0.7969), 0.8551 (95% CI = 0.8099–0.9115), 0.6007 (95% CI = 0.5342–0.6616), 0.5696 (95% CI = 0.5041–0.6110), 0.7890 (95% CI = 0.7427–0.8284), and 0.6145 (95% CI = 0.5490–0.6812) and specificities of 0.8774 (95% CI = 0.8188–0.9281), 0.7828 (95% CI = 0.7458–0.8198), 0.9812 (95% CI = 0.9437–0.9924), 0.9894 (95% CI = 0.9620–0.9975), 0.9995 (95% CI = 0.9667–1.0000), and 0.9930 (95% CI = 0.9463–1.0000) for identifying high myopia overall and high myopia associated with cataracts, macular degeneration, choroidal neovascularization, glaucoma, and retinal detachment, respectively.

The distributions and correlations of these nine blood test variables in validation cohort 2 are presented in Figure S3C. Strong correlations were observed between the EOS and EOS% and between the BASO and BASO%, whereas no significant correlations were found among the other variables. A detailed description of the performance, including sensitivity, specificity, PPV, NPV, accuracy, and F1 score, of the final RF model in the validation cohort is provided in Table 1.

A subgroup analysis of validation cohort 1 was performed after excluding patients with active systemic inflammatory conditions (n= 81), recent infections (n=144), immunosuppressive medication use (n= 25), diabetes (n= 222), or hypertension (n=367) to assess the performance of the final RF model. The final RF model achieved a macroaverage AUC of 0.9010, which was not significantly different from that of the model including all patients (DeLong test, P = 0.868). As shown in Figure S5B, the RF model achieved AUCs of 0.9064, 0.9164, 0.8988, 0.8752, 0.9351, and 0.8743 for identifying high myopia overall and high myopia associated with cataracts, macular degeneration, choroidal neovascularization, glaucoma, and retinal detachment, respectively. Other performance metrics, including sensitivity, specificity, PPV, NPV, accuracy, and F1 score, are summarized in Table S8, with macroaverage values of 0.6758, 0.9327, 0.8407, 0.9398, 0.7279, and 0.7359, respectively.

Furthermore, a subgroup analysis of validation cohort 2 was performed after excluding patients with active systemic inflammatory conditions (n= 33), recent infections (n=76), immunosuppressive medication use (n= 18), diabetes (n= 109), or hypertension (n=138) to assess the performance of the final RF model. The final RF model achieved a macroaverage AUC of 0.9008, which was not significantly different from that of the model including all patients (DeLong test, P = 0.976). As shown in Figure S5C, the RF model achieved AUCs of 0.9042, 0.9135, 0.8890, 0.8833, and 0.9456 for identifying high myopia overall and high myopia associated with cataracts, macular degeneration, choroidal neovascularization, glaucoma, and retinal detachment, respectively. Other performance metrics, including sensitivity, specificity, PPV, NPV, accuracy, and F1 score, are summarized in Table S9, with macroaverage values of 0.6834, 0.9332, 0.8486, 0.9398, 0.7286, and 0.7429, respectively.

In the sensitivity analysis excluding UA, the only kidney function indicator included in the model, the final

RF model maintained robust performance (Table S10), achieving macroaverage AUCs of 0.8589 and 0.8613 in validation cohorts 1 (Figure S6A) and 2 (Figure S6B), respectively.

Furthermore, a comparison between the final nine-feature RF model and other machine learning algorithms—constructed using the complete set of blood test variables as input features—was conducted in validation cohorts 1 (Table S11, Figure S7) and 2 (Table S12, Figure S8) using key performance metrics. As shown in Figure 4G, in validation cohort 1, the final RF model outperformed all the other models, including those utilizing the full set of blood test variables. Notably, its performance was comparable to that of the conventional RF model trained on the complete feature set. Similar results were observed in validation cohort 2 (Figure 4H), further confirming the robustness of the final RF model.

Taken together, these findings indicate that the nine-feature RF model demonstrated strong screening performance across all the validation cohorts, consistent with the findings in the discovery cohort.

Evaluation of the performance of the final RF model with DCA and calibration curves

DCA of the final RF model indicated that it consistently offered net benefits for the discovery cohort and both validation cohorts over a range of threshold probabilities. In the discovery cohort (Figure S9A), validation cohort 1 (Figure S9B), and validation cohort 2 (Figure S9C), the RF model outperformed the “treat none” strategy, indicating its practical utility in clinical decision-making.

The calibration curves further confirmed the model’s robust screening performance. In the discovery cohort (Figure 5A), the RF model showed good overall calibration, particularly for high myopia and high myopia associated with macular degeneration. High myopia associated with cataracts was slightly overestimated in the mid-to-high probability range, whereas the model tended to underestimate the likelihood of high myopia associated with choroidal neovascularization, glaucoma, and retinal detachment, as indicated by calibration curves consistently falling below the ideal diagonal line, suggesting a conservative prediction tendency for these categories.

Similar calibration performance was observed in validation cohort 1 (Figure 5B) and validation cohort 2 (Figure 5C), further supporting the reliability and generalizability of the model across different populations.

Model interpretability and clinical application

Given the challenges clinicians face in adopting diagnostic models with limited interpretability, we employed SHAP to elucidate the decision-making process of our final model by quantifying the contribution of each feature to the model predictions. The SHAP summary plot (Figure 6A) presents the relative importance values of the nine selected features based on their mean absolute SHAP values, providing a global interpretation of model behaviour. SHAP dependence plots were generated to illustrate how individual features influence the model output and to further improve the interpretability. Each plot depicts the relationship between the actual values of the features and their corresponding SHAP values.

We developed a freely accessible online web application to facilitate seamless integration into clinical practice (Figure 6B), which is available at https://high-myopia.streamlit.app/. This interactive platform allows clinicians to input the nine required feature values, after which the application automatically computes and displays the predicted risk based on the RF model.

Hospital-based prospective validation of the final RF model

A total of 5,067 participants with high myopia were enrolled in the hospital-based prospective cohort study to screen for high myopia-related complications (Figure 7A). Of these participants, 4,572 met the eligibility criteria and were assessed in real time with the online web application developed for assessing the risk of high myopia-related complications. According to the screening results of the model, 1,523 individuals were classified as high risk. Among them, 127 declined to cooperate, and 84 withdrew voluntarily, resulting in 211 individuals not being referred for further ophthalmic evaluation.

Among them, 1,312 high-risk participants were referred to the Department of Ophthalmology for further evaluation, resulting in the confirmation of 938 cases of high myopia-related complications. An additional 2,270 individuals identified as being at low risk of complications were also referred for further ophthalmic evaluation, and 347 cases of high myopia-related complications were identified at EENT Hospital.

Figure 7B presents the confusion matrix of the final RF model for detecting the five high myopia-related complications. The model achieved a mean sensitivity of 70.0% (95% CI=66.43%–74.22%), a specificity of

94% (95% CI=90.18%–98.22%), a PPV of 86% (95% CI=82.33%–89.87%), and an NPV of 94% (95% CI=91.60%–97.04%), as detailed in Figure 7C.

The final RF model yielded false positive rates of 14.22% for cataracts, 0.23% for macular degeneration, 0.52% for choroidal neovascularization, 0% for glaucoma, and 0.18% for retinal detachment, indicating a low proportion of false alarms. In terms of downstream costs, the higher false positive rate for cataracts may lead to some additional routine ophthalmologic visits, while the very low false positive rates for macular degeneration, choroidal neovascularization, and retinal detachment minimize unnecessary procedures. Overall, the model helps allocate ophthalmic resources efficiently, focusing on confirmatory evaluations of high-risk individuals and limiting unnecessary healthcare expenditure.

Community-based validation of the final RF model

A total of 311,254 participants were included in the community-based cross-sectional study, among whom 1,905 were identified as having high myopia (Figure 7D). Of these, 1,894 met the inclusion criteria and were assessed in real time with the online web application. According to the predictions of the RF model, 267 individuals were classified as high risk.

A total of 210 high-risk individuals were subsequently referred to the Department of Ophthalmology for further clinical evaluation, which confirmed 148 cases of high myopia-related complications. Owing to the cross-sectional nature of the study, follow-up validation was not performed for participants in the low-risk group.

The final RF model had favourable PPVs (Figure 7D) in detecting the five high myopia-related complications, with the highest PPV observed for the prediction of macular degeneration (0.795), followed by glaucoma (0.75), retinal detachment (0.75), choroidal neovascularization (0.733), and cataracts (0.65), indicating strong performance in identifying true positive cases within the high myopia population.

Discussion

In this study, we developed and validated a machine learning framework leveraging routine blood test data to enable opportunistic screening for vision-threatening complications of high myopia during routine health examinations, thereby addressing critical limitations of current imaging-based diagnostic paradigms. Our key innovation lies in showing that a readily obtainable set of nine routine blood biomarkers—EOS, PCT, BASO%, M, UA, LPB, BASO, EOS%, and TG levels—can predict five major complications with robust accuracy and the validation of the model through a hospital-based prospective cohort study and a community-based cross-sectional study. These findings advance ocular risk stratification by establishing haematological features as accessible proxies for microstructural pathology, particularly in resource-limited settings.

Routine blood indices, biochemical indices, and coagulation indices can reflect a wide range of physiological and pathological states in the body[23,25], providing information on aspects such as inflammation, blood circulation, metabolic status, and tissue injury. In this study, the selected biomarkers were associated with known mechanisms of myopic complications. The inclusion of eosinophil and basophil counts and percentages aligns with recent evidence showing that localized inflammation drives retinal detachment and choroidal neovascularization through IL-4/IL-13-mediated vascular leakage[26–28]. IL-4 and IL-13, key Th2-type cytokines, play critical roles in mediating localized inflammation in the eye. They promote the recruitment and activation of eosinophils and basophils, increasing vascular permeability, which may contribute to the development of retinal detachment and choroidal neovascularization. The evidence suggests that these cytokines induce the endothelial expression of adhesion molecules, facilitating leukocyte adhesion and migration and thereby amplifying inflammatory responses in the retina and choroid. Importantly, targeting IL-4/IL-13 signalling with anti-inflammatory therapies, such as monoclonal antibodies against IL-4 or IL-13, could reduce local inflammation and vascular leakage, decreasing the risk of disease progression[29,30]. While preclinical studies support this approach, the safety and efficacy of such interventions in clinical settings remain to be established, and future prospective studies are warranted to evaluate their therapeutic potential.

UA, a potent antioxidant, paradoxically showed a protective association in our model, although this finding

is consistent with its dual role in neutralizing oxidative stress in early macular degeneration while exacerbating inflammation at higher concentrations[31]. In this study, we also revealed that abnormalities in triglyceride and lipoprotein levels were associated with complications of high myopia. Systemic dyslipidaemia-induced blood lipid deposition plays a critical role in the pathogenesis of macular degeneration, where lipid dysregulation promotes lipid accumulation, oxidative stress, and chronic inflammation in Bruch's membrane, ultimately leading to retinal pigment epithelial dysfunction and disease progression[32,33]. The prominence of the thrombocytocrit and the level of lipoprotein B corroborates the roles of thrombotic and dysregulated lipid pathways in the progression of myopic glaucoma[34]. This biological agreement between the identified biomarkers and known mechanisms strengthens the clinical plausibility of the model, suggesting that systemic shifts in haematometabolic parameters may precede detectable fundus changes by months to years.

The utility of routine blood biomarkers for screening high myopia-related complications is further supported by their inherent stability and established systemic–ocular interplay. First, routine blood parameters exhibit high intraindividual consistency over time, as evidenced by longitudinal cohort studies showing a <15% coefficient of variation for haematological indices such as the eosinophil count and thrombocytocrit across intervals of years[23]. This stability allows reliable risk stratification, in contrast to the transient imaging artefacts detectable in OCT scans. This systemic–ocular axis operates bidirectionally—not only do blood markers reflect ocular pathology but retinal imaging itself can also predict certain systemic conditions[35–40]. Boris Babenko *et al*.[36] reported strong correlations between ocular imaging features and peripheral blood parameters; specifically, conjunctival vascular density, scleral hue, and subtle haemorrhagic manifestations, including fluctuations in haemoglobin levels, platelet counts, and inflammatory markers, were shown to reflect systemic physiological changes. Cheung *et al*.[41] developed fully automated artificial intelligence and deep learning-based software for analysing retinal vessels revealing that a narrower retinal arteriolar calibre was significantly associated with elevated blood pressure, whereas a wider venular calibre was associated with a higher body mass index, elevated haemoglobin A1c levels, and current smoking status. Moreover, algorithms based on digital colour fundus photography have shown good performance in predicting hypertension, hyperglycaemia, and dyslipidaemia, with AUCs of 0.766, 0.880 and 0.703, respectively[42].

These ocular features may thus serve as noninvasive surrogates for underlying haematological or inflammatory conditions, underscoring the potential of integrated systemic–ocular diagnostic approaches. Taken together, this bidirectional crosstalk suggests that blood-based screening could serve not as a replacement for retinal exams but as a scalable first-line triage tool, particularly in regions lacking imaging infrastructure.

Our blood-based opportunistic screening during routine health examinations framework offers notable advantages over AI-driven fundus imaging for detecting high myopia-related complications, particularly in resource-limited settings. Although deep learning models based on retinal photographs have achieved strong diagnostic performance (AUCs of 0.89–0.95 for myopic maculopathy), their widespread implementation is constrained by the need for high-resolution imaging equipment, which is unavailable in more than 90% of rural clinics in low-income countries, and the requirement for trained personnel. These challenges are further compounded by the frequent need for pharmacologic pupil dilation in highly myopic eyes with peripheral retinal pathology. In contrast, routine blood testing is almost universally accessible, with 98% of primary care facilities worldwide equipped to perform basic haematological assessments. Moreover, blood collection requires no ophthalmic expertise and remains feasible in conditions that compromise fundus imaging, such as media opacities (e.g., cataracts) or severe fundus distortions associated with high myopia.

Nevertheless, our model had a nonnegligible false-positive rate, suggesting that its optimal role is as a prescreening tool in community clinics, primary care facilities, or health examination centres to identify individuals at high risk who should undergo a confirmatory ophthalmic evaluation. In particular, within health examination centres, individuals with high myopia could be screened at no additional cost or resource burden, since the required haematological parameters are already included in standard blood test panels. By leveraging routinely collected data, our model enables efficient risk stratification for multiple high myopia-related complications, thereby reducing missed diagnoses and supporting the more rational allocation of limited ophthalmic resources.

Cost considerations further support the practical utility of our model. In health examination centres, the

haematological parameters required are already incorporated into standard blood panels, which means that our approach can identify individuals at risk of multiple high myopia-related complications without introducing any additional expense. Similarly, in community and primary care settings, patients often undergo routine blood testing for nonophthalmic conditions, and incorporating these results into our model allows opportunistic case finding at no extra cost. When equipment amortization and personnel expenditures are accounted for, the median cost per blood-based screening is approximately ¥300 (≈ $43 or €37), which is considerably lower than that of most ophthalmic imaging modalities. Beyond ophthalmic risk prediction, blood tests also provide concurrent systemic health information (e.g., uric acid levels for gout and triglyceride levels for dyslipidaemia), thus offering dual-purpose clinical benefits that cannot be achieved through fundus imaging alone.

Moreover, we acknowledge the rapid development of point-of-care ophthalmic imaging technologies. Smartphone-based fundus photography and portable fundus cameras, for example, have emerged as cost-effective, portable, and accessible methods for the early detection of retinal diseases such as diabetic retinopathy, age-related macular degeneration, and retinal vascular disorders, significantly broadening access to diagnostic services[16,17,43]. Rather than competing with these imaging-based modalities, our model is best positioned as a complementary tool: by leveraging routinely collected blood test data, it enables opportunistic, low-cost risk stratification for high myopia-related complications, thereby facilitating targeted referral for a confirmatory ophthalmic evaluation and promoting the more efficient allocation of limited healthcare resources.

This blood-based screening model is specifically designed for opportunistic case-finding during routine health examinations in community and primary care settings where specialized ophthalmic imaging is unavailable. The target population includes asymptomatic adults with high myopia (≤ –6.0 D) who do not routinely access ophthalmologic care, estimated to represent approximately 60-70% of individuals with high myopia in Chinese populations[44,45]. By integrating this approach into standard health check-up workflows, early identification of high-risk individuals can be achieved without the need for specialized ophthalmic equipment, thereby facilitating timely referral for confirmatory imaging and intervention.

Although the RF models displayed high overall discriminative performance, the sensitivities for detecting macular degeneration, choroidal neovascularization, and retinal detachment were relatively low (<0.6). Clinically, missed cases in these categories could delay diagnosis and timely intervention for vision-threatening complications. This discrepancy can be explained by several factors. First, decision thresholds were determined using the Youden index, which aims to balance sensitivity and specificity rather than maximizing sensitivity alone. Second, class imbalance existed in the dataset, with some conditions being less frequent than others, which likely reduced the sensitivity for these less common categories. Collectively, these factors explain why the sensitivity did not fully reflect the exceptionally high AUC values. Furthermore, in the community-based cross-sectional study, the detection of 347 complication cases among low-risk individuals indicated a substantial false-negative rate. Recognizing this limitation is important for interpreting model performance and for guiding clinical application, particularly when considering potential threshold adjustments or future model refinements to optimize sensitivity without compromising overall accuracy.

The final model yielded a false positive rate of 14.22% for cataracts, which could translate into a substantial number of unnecessary ophthalmic referrals in large-scale applications. For instance, in a screening population of 10,000 individuals with high myopia, approximately 1,422 false positive referrals would require confirmatory ophthalmic evaluation, potentially increasing healthcare burden and costs. To mitigate this, several optimization strategies may be considered. First, adjusting decision thresholds for specific complications could improve specificity without markedly compromising sensitivity. Second, implementing a two-stage screening workflow—a blood-based ML pre-screen (Stage 1) followed by a clinical risk score (Stage 2) incorporating established cataract risk factors such as diabetes, prolonged ultraviolet exposure, smoking, and obesity—may further reduce false positives. Individuals exceeding stage-specific thresholds would be referred for confirmatory ophthalmic examination, while those with intermediate risk could undergo short-interval follow-up. Third, restricting model application to older high-myopia populations (⩾50 years), in whom cataracts are most prevalent and established clinical referral pathways already exist, could enhance the practical value and cost-effectiveness of opportunistic screening. These refinements should be evaluated in future prospective implementation studies to balance diagnostic sensitivity with healthcare

resource utilization.

Several limitations of this study should be acknowledged. First, although the multicentre case–control study provided a large and diverse dataset for model development, its retrospective and noninterventional nature may introduce selection bias. Second, the hospital-based prospective cohort and community-based cross-sectional studies were conducted exclusively in Asian populations, which may limit the generalizability of our findings to other ethnic groups and geographic regions. External validation is warranted in African, European, Latin American, and other ancestry groups, as hematological parameters and the prevalence of myopia-related complications may vary across populations. Third, the cross-sectional design of the community-based cohort precludes the assessment of temporal or causal relationships between routine blood test markers and high myopia-related complications. Furthermore, the fact that only participants who screened positive underwent confirmatory ophthalmic examinations, while model-negative individuals were not systematically assessed, introduces a potential verification bias that may lead to an overestimation of the sensitivity, specificity, and predictive values. In particular, the lack of systematic evaluation among model-negative participants prevents an accurate estimation of the false negative and negative predictive values, which is essential for assessing the clinical utility of the model. Fourth, the proportion of patients with high myopia-related complications was higher than that of patients with uncomplicated high myopia in both the discovery and validation cohorts, reflecting the tertiary referral nature of the participating hospitals; although a class–SMOTE strategy was applied during model training and the model performed well in both hospital- and community-based settings, the overrepresentation of complication cases may still limit the generalizability to populations with lower prevalence. Fifth, while our model displayed promising predictive performance, we did not conduct a formal cost-effectiveness or health economic analysis to quantify potential clinical or economic benefits. Sixth, although we incorporated multiple haematological, metabolic, and inflammatory markers, other potentially relevant biomarkers were not included, and future studies integrating multiomics data may further improve the predictive accuracy. Finally, a key limitation is the inability to determine whether the model preferentially detects early-stage or advanced disease, which has implications for clinical utility and screening timing. Theoretically, preferential detection of early-stage disease would provide greater clinical benefit by enabling timely intervention to prevent irreversible visual

loss, whereas uniform performance across all stages would enhance population-level case identification. Future prospective studies with detailed staging and longitudinal follow-up are needed to define the stage-specific performance and optimal application window of this blood-based model.

Our framework provides a promising strategy for opportunistic screening of high myopia-related complications during routine health examinations, serving as a practical first-line approach in health examination centres and community primary care settings. With performance approaching that of specialized imaging models, this strategy may be particularly valuable in low-resource settings, where the burden of preventable blindness is disproportionately high. Future integration with telemedicine platforms and portable diagnostic tools could further support the development of more accessible and scalable presymptomatic care models, contributing greatly to global efforts to mitigate myopia-related visual impairment.

Future studies should explore the integration of blood-based biomarkers with smartphone-assisted fundus imaging, potentially establishing a cost-effective, two-tiered screening strategy suitable for wide-scale deployment. Moreover, mechanistic studies are needed to determine whether altering the levels of the identified biomarkers—such as the use of fibrates to reduce triglyceride levels—can mitigate the risk of myopic complications, thereby opening new therapeutic avenues.

Methods

Ethical considerations

This study received ethical approval from the institutional review boards of the participating institutions: Xuhui Central Hospital (Approval No. 2025-017), the Eye & ENT Hospital of Fudan University (Approval No. 2025124), Wanbei Coal–Electricity Group General Hospital (Approval No. WBZY-LLWYH-2025-020), and the First Affiliated Hospital of Anhui Medical University (Approval No. PJ2025-06-34). This study was conducted in accordance with the Declaration of Helsinki and the Ethical Guidelines for Medical and Health Research Involving Human Subjects. This study was conducted in three phases. First, a multicentre

retrospective case–control study was performed to develop and internally and externally validate the machine learning-based model using routine blood test data from patients with high myopia, both with and without complications. Second, the clinical performance of the model was prospectively validated in a hospital-based cohort. Finally, the model was further evaluated in a community-based cross-sectional study to examine its utility for population-level screening. Given that the multicentre case–control study and community-based cross-sectional study was noninterventional and retrospective in nature, the institutional review boards waived the requirement for informed consent. In the prospective, hospital-based cohort, all participants provided written informed consent prior to sample collection and study enrolment.

Study design and participants

Multicentre, case–control study

A multicentre, case–control study was conducted across five hospitals in China between January 1, 2016, and December 30, 2024. Patients from the Eye & ENT Hospital of Fudan University composed the discovery cohort. The first validation cohort consisted of patients from Xuhui Central Hospital of Fudan University and Putuo District Central Hospital, and the second validation cohort included patients from Wanbei Coal–Electricity Group General Hospital and The First Affiliated Hospital of Anhui Medical University. The overall study design is illustrated in Figure 1. Sex of participants was recorded as male or female according to hospital registration data, which were based on self-reported information at the time of admission. As sex were not directly relevant to the study objectives, no sex-based analyses were performed.

Individuals were excluded if they met any of the following criteria to minimize the effects of potential confounding factors: a history of haematologic diseases or cancers; systemic illnesses such as acute infections, metabolic syndrome, or autoimmune diseases; pregnancy or intraocular surgery within the past six months; aged under 18 years; or a lack of available laboratory test results. In the subgroup analyses, patients with active systemic inflammatory conditions, recent infections, immunosuppressive medication use, diabetes, or hypertension were excluded to further validate the robustness of the model. This information was collected through a review of medical records and laboratory data, and individuals meeting one or more criteria were removed from the analysis prior to model development.

During the multicentre, case–control study phase, all patients with high myopia who visited one of the five participating hospitals between January 1, 2016, and December 30, 2024, were included, without sampling. The discovery cohort comprised 1,464 patients with uncomplicated high myopia and patients with high who were divided into the following complication-specific subgroups: 1,914 patients with cataracts, 698 with macular degeneration, 478 with choroidal neovascularization, 185 with glaucoma, and 363 with retinal detachment. External validation cohort 1 included 937 patients with uncomplicated high myopia, 986 patients with high myopia complicated with cataracts, 463 patients with high myopia complicated with macular degeneration, 335 patients with high myopia complicated with choroidal neovascularization, 114 patients with high myopia complicated with glaucoma, and 290 patients with high myopia complicated with retinal detachment. External validation cohort 2 included 787 patients with uncomplicated high myopia and with high myopia presenting with the following complications: 771 patients with cataracts, 255 with macular degeneration, 192 with choroidal neovascularization, 130 with glaucoma, and 299 with retinal detachment. No participants had two or more of these five high myopia complications. Venous blood samples were collected from patients newly diagnosed with high myopia or high myopia-related complications within 24 hours of presentation.

Large-sample cohort study

The performance of the screening model was validated through two large-sample cohort studies: a hospital-based prospective cohort study and a community-based cross-sectional study. A prospective screening program was initiated at the Eye & ENT Hospital of Fudan University and Xuhui Central Hospital of Fudan University. Briefly, all patients with high myopia who presented to the hospital from January 2025 to May 2025 were sequentially enrolled, with approximately 5,000 patients recruited during the study period. Routine blood tests were performed for all eligible participants. Individuals who were identified as positive for high myopia-related complications according to the established screening model were further evaluated with ophthalmic examinations to confirm the presence of these complications. Furthermore, individuals identified as negative for high myopia-related complications were further assessed through ophthalmic examinations.

A community-based cross-sectional screening program was launched in Xuhui District, Shanghai, in July 2024[46]. Residents of Xuhui District with high myopia were invited to participate through two local community health service centres, the Changqiao Community health service centre and the Caohejing Community health service centre, with whom we have maintained strong collaborative relationships over the past decade. Interested individuals were screened for eligibility according to predefined inclusion and exclusion criteria. Eligible participants were provided with detailed information about the study objectives, procedures, and potential risks. Approximately 300,000 residents of Xuhui District were included. Routine blood tests were performed for all eligible participants as part of their routine physical examinations; those who were identified as positive for high myopia-related complications according to the model were further evaluated and confirmed to be positive through ophthalmic examinations conducted at the Eye & ENT Hospital of Fudan University or Xuhui Central Hospital of Fudan University. Furthermore, individuals who were identified as negative for high myopia-related complications were not further evaluated with ophthalmic examinations.

The inclusion criteria were as follows: no prior medical history of high myopia-related complications; ability to understand, sign, and date a written informed consent form; good physical condition; and clarity of consciousness.

The exclusion criteria were as follows: a history of haematologic diseases or any form of cancer; pregnancy or intraocular surgery within the past six months; and aged under 18 years.

The study exit criteria were as follows: development of serious comorbid conditions such as significant cardiovascular, hepatic, or renal diseases or malignancies; noncooperation during follow-up diagnostic confirmation for high myopia-related complications; voluntary withdrawal for any reason; and death from any cause.

Eye examinations and diagnostic criteria

The procedures for the detailed ophthalmic examinations have been described previously[21,47,48]. All the enrolled patients underwent comprehensive eye evaluations, which included intraocular pressure

measurements, slit-lamp biomicroscopy, optical coherence tomography, B-scan ultrasonography, ultrawide fundus photography, ultra-widefield autofluorescence, and fluorescein angiography. In our study, refractive error was measured using cycloplegic refraction, and high myopia was defined as a refractive error of ≤ −6.0 dioptres. Uncomplicated high myopia was defined as high myopia in the absence of related complications, including cataracts, myopic macular degeneration, choroidal neovascularization, glaucoma, and retinal detachment.

High myopia-related complications were identified based on the following standard ophthalmologic diagnostic criteria:

Cataract—diagnosed by slit-lamp biomicroscopy, defined as lens opacity graded according to the Lens Opacities Classification System III and confirmed to cause a best-corrected visual acuity reduction not attributable to other ocular diseases;

Macular degeneration—diagnosed by fundus photography and/or OCT according to the Meta-Analysis for Pathologic Myopia classification, including diffuse chorioretinal atrophy, patchy atrophy, and macular atrophy;

Choroidal neovascularization—confirmed by OCT and fluorescein angiography or OCT angiography, showing subretinal hyperreflective material, exudation, and leakage consistent with neovascular membranes secondary to pathological myopia;

Glaucoma—diagnosed by a combination of characteristic optic nerve head changes (increased cup-to-disc ratio and rim thinning), corresponding visual field defects on standard automated perimetry, and intraocular pressure measurements, according to the European Glaucoma Society guidelines;

Retinal detachment—confirmed by a dilated fundus examination and/or ultrasonography, showing separation of the neurosensory retina from the retinal pigment epithelium, with retinal breaks identified in most cases.

All diagnoses were made by board-certified ophthalmologists at the participating centres, and patients with secondary causes unrelated to high myopia were excluded.

Data sources

For this multicentre cohort study, routine clinical laboratory data were extracted from the patients' electronic medical records. Each institution's principal investigator collected information on 61 laboratory parameters per patient, which were classified into three major categories: blood cell analysis parameters, biochemical analysis parameters, and coagulation analysis parameters.

At presentation, ophthalmologists first evaluated whether patients had any prior diagnosis or treatment for high myopia-related complications. Those with previously diagnosed complications or a history of treatment for such conditions were excluded to ensure that the laboratory measurements reflected the onset of newly identified disease. For patients newly diagnosed with high myopia or high myopia-related complications, venous blood samples were collected within 24 hours of presentation. Laboratory tests obtained during stable or follow-up periods were not included in this analysis. Blood samples were obtained via standard venipuncture and analysed at multiple clinical laboratories, including the Eye & ENT Hospital of Fudan University, Putuo District Central Hospital, the First Affiliated Hospital of Anhui Medical University, Shanghai Xuhui Central Hospital, and Wanbei Coal–Electricity Group General Hospital. The procedures for the detailed laboratory analyses have been described previously[18]. The blood cell analyses included 24 variables, the biochemical analyses included 31 variables, and the coagulation analyses included 6 variables, as detailed in Table S13.

The selection of these haematological, biochemical, and coagulation parameters was guided by two main considerations: availability and biological plausibility supported by prior research. First, in terms of availability, all 61 parameters are routinely measured in standard clinical practice, which greatly enhances the feasibility and potential clinical utility of our model. Second, regarding biological plausibility, prior evidence supports the relevance of these parameters in ocular risk assessments. Recent studies have demonstrated that routinely available blood test data can be effective at predicting high myopia-related complications, particularly retinal detachment[18,19]. Blood markers of inflammation[49], glucose levels[50], and lipid levels[51] have been linked to an increased risk of ocular diseases, including high myopia and retinal detachment. In addition, several haematological, biochemical, and coagulation parameters have been

reported to be associated with the onset or severity of other eye diseases, such as glaucoma[52–54], cataract[55–57], and macular degeneration[58,59]. Therefore, based on the prior evidence, we developed a machine learning–based screening model using routine blood test data to identify individuals at risk for high myopia-related complications (cataracts, macular degeneration, choroidal neovascularization, glaucoma, and retinal detachment).

Blood Sample Quality Control

Interlaboratory and intralaboratory quality control measures were implemented to ensure the reliability and consistency of laboratory results across multiple centres. All the centres participated in external proficiency testing organized by the Clinical Laboratory Center using standardized survey samples that were analysed locally and evaluated centrally, with all the centres meeting the required criteria. In addition, standardized internal quality control (IQC) materials were measured daily before patient samples were processed to ensure assay precision, and the coefficient of variation (CV) was maintained within 5%. These procedures ensured the accuracy, precision, and comparability of the haematological data across all the study sites.

Machine learning (ML) model development

A total of 61 features (detailed in Table S13) were utilized to develop diagnostic models. For each participant, if more than 20% of the 61 laboratory variables were missing (i.e., 13 variables), the participant was excluded from the study. Conversely, participants with ≤20% missing values (i.e., 12 variables) were retained, and any remaining missing values for these variables were imputed using the medians[60]. Ten ML algorithms were employed to ensure a diverse and robust modelling approach, namely, the RF, AdaBoost, decision tree (DT), gradient boosting (GB), k-nearest neighbours (KNN), light gradient boosting machine (LightGBM, LGBM), logistic regression (LR), multilayer perceptron (MLP), naïve Bayes (NB), and XGBoost algorithms. Multiple classification methods were used to screen for different subtypes of high myopia complications. We applied a Synthetic Minority Oversampling Technique (SMOTE) approach to adjust for an unequal class distribution and mitigate the effect of class imbalance during model development. Each algorithm was selected based on its ability to balance interpretability and predictive power and capture both linear and nonlinear interactions among features. A combination of grid search and manual fine-tuning was

implemented to identify the optimal hyperparameters and optimize model performance. The final hyperparameters for each of the ten machine learning models are detailed in Table S14. All modelling procedures were conducted in Python (v3.11), and the corresponding code is available at https://github.com/fudanRenjun/high-myopia/tree/master.

Model performance evaluation

The models were assessed with multiple performance metrics, including precision, recall, PPV, NPV, F1 score, and accuracy. Furthermore, their classification capabilities were evaluated by calculating AUC and PR-AUC. The DeLong test[61], a nonparametric statistical method, was used to compare AUC values across the different machine learning models. The optimal cut-off value for the models was determined as the value that maximized the Youden index and the sum of sensitivity and specificity. Based on this threshold, the sensitivity, specificity, accuracy, PPV, and NPV were calculated to evaluate model performance[62]. Decision curve analysis[63] and calibration curves[64] were used to further assess the performance of the model.

Feature selection and model interpretation

In the discovery cohort, feature selection was performed with the SHAP method[65]. The process refined the predictive model by narrowing the 61 input features to the most critical predictors while maintaining optimal predictive performance. The feature selection process was embedded within cross-validation to avoid information leakage and ensure a fair comparison among the models. The DeLong test[61] was further applied to compare the AUC values of the models before and after feature selection, validating the impact of this dimensionality reduction process. The SHAP analysis also provides both global interpretability and local interpretability, offering insights into overall model behaviour and individualized predictions.

To facilitate clinical adoption, the final predictive model was deployed as a web-based application (https://high-myopia.streamlit.app/) using the Streamlit framework in Python to facilitate clinical adoption. By inputting the nine selected features (Table S15)—EOS, PCT, BASO%, M, UA, LPB, BASO, EOS%, and TG levels—which are routinely available in typical clinical settings, the clinician receives probability estimates for different high myopia-related complications along with a personalized force plot that visually

explains the prediction for individual patients. This interactive tool enhances the interpretability of the model and supports clinical decision-making.

Statistical analysis

All the statistical analyses were performed using GraphPad Prism (v10) and Python (v3.11). The Kolmogorov–Smirnov test was applied to assess the normality of the data distribution. For continuous variables with a normal distribution, one-way analysis of variance (ANOVA) was used, whereas the Kruskal–Wallis test was used for non-normally distributed data. Categorical variables were analysed using the chi-square test, where appropriate. A false discovery rate correction was applied using the Benjamini–Hochberg method to account for multiple comparisons and control for type I error. Effect sizes were also calculated: η^2 (eta squared) for continuous variables from one-way ANOVA and Cramér's V for categorical variables. Continuous data are presented as the means ± standard deviations (SDs), whereas categorical data are summarized as counts and percentages. A two-tailed P value < 0.05 was considered to indicate statistical significance.

Statistics & Reproducibility

No statistical method was used to predetermine sample size. For each participant, if more than 20% of the 61 laboratory variables were missing (i.e., 13 variables), the participant was excluded from the study. The experiments were not randomized. The Investigators were not blinded to allocation during experiments and outcome assessment.

Data Availability: The de-identified data of discovery cohort and validation cohort in this study have been deposited in the Source data file. The corresponding acronyms for these cohorts are listed in Tables S1 and S2, respectively. Access to comprehensive, individual-level clinical data is restricted to protect participant privacy and comply with the conditions of informed consent. The individual-level clinical data may be made available upon request to the corresponding author, accompanied by a study protocol and evidence of ethical approval. All requests will be reviewed by the corresponding author (Meiyan Li) in

consultation with the cohort leadership and relevant study committees, and a response is typically provided within 12 weeks. Approved users will be required to sign a Data Use Agreement specifying that the data are to be used solely for the approved project and that any resulting publications must acknowledge the data source. The data will be made available only for scientific research purposes. Once access has been granted, the data will remain available for six months. The data generated in this study are provided in the Supplementary Information/Source Data file.

Code Availability: The analysis code was written in Python (version 3.11) and relies on standard open-source libraries (NumPy, Pandas, scikit-learn, XGBoost), used in compliance with their MIT and BSD licenses. All reused components retain their original license and attribution. The analysis codes are available on GitHub at (https://github.com/fudanRenjun/high-myopia/tree/master) archived with Zenodo[66] under https://doi.org/10.5281/zenodo.18382509.

References

1. Flitcroft, D. I. *et al.* IMI - Defining and Classifying Myopia: A Proposed Set of Standards for Clinical and Epidemiologic Studies. *Invest Ophthalmol Vis Sci* **60**, M20–M30 (2019).
2. Morgan, I. G., Ohno-Matsui, K. & Saw, S.-M. Myopia. *Lancet* **379**, 1739–1748 (2012).
3. Zhang, X. *et al.* Optic neuropathy in high myopia: Glaucoma or high myopia or both? *Prog Retin Eye Res* **99**, 101246 (2024).
4. Flitcroft, D. I. The complex interactions of retinal, optical and environmental factors in myopia aetiology. *Prog Retin Eye Res* **31**, 622–660 (2012).
5. Kanthan, G. L., Mitchell, P., Rochtchina, E., Cumming, R. G. & Wang, J. J. Myopia and the long-

term incidence of cataract and cataract surgery: the Blue Mountains Eye Study. *Clin Exp Ophthalmol* **42**, 347–353 (2014).

6. Haarman, A. E. G. *et al.* The Complications of Myopia: A Review and Meta-Analysis. *Invest Ophthalmol Vis Sci* **61**, 49 (2020).
7. Wang, Y. X. *et al.* High myopia as risk factor for the 10-year incidence of open-angle glaucoma in the Beijing Eye Study. *Br J Ophthalmol* **107**, 935–940 (2023).
8. Du, Y. *et al.* Complications of high myopia: An update from clinical manifestations to underlying mechanisms. *Adv Ophthalmol Pract Res* **4**, 156–163 (2024).
9. Foo, L. L. *et al.* Predictors of myopic macular degeneration in a 12-year longitudinal study of Singapore adults with myopia. *Br J Ophthalmol* **107**, 1363–1368 (2023).
10. Zheng, F. *et al.* Quantitative OCT angiography of the retinal microvasculature and choriocapillaris in highly myopic eyes with myopic macular degeneration. *Br J Ophthalmol* **106**, 681–688 (2022).
11. Liu, H. *et al.* Economic evaluation of combined population-based screening for multiple blindness-causing eye diseases in China: a cost-effectiveness analysis. *Lancet Glob Health* **11**, e456–e465 (2023).
12. Fricke, T. R. *et al.* Global prevalence of visual impairment associated with myopic macular degeneration and temporal trends from 2000 through 2050: systematic review, meta-analysis and modelling. *Br J Ophthalmol* **102**, 855–862 (2018).
13. World report on vision. https://www.who.int/publications/i/item/9789241516570.

14. Cheung, N. *et al.* Prevalence and Associations of Retinal Emboli With Ethnicity, Stroke, and Renal Disease in a Multiethnic Asian Population: The Singapore Epidemiology of Eye Disease Study. *JAMA Ophthalmol* **135**, 1023–1028 (2017).

15. Igarashi-Yokoi, T. *et al.* Prognostic Factors for Axial Length Elongation and Posterior Staphyloma in Adults With High Myopia: A Japanese Observational Study. *Am J Ophthalmol* **225**, 76–85 (2021).

16. Chen, Y. *et al.* Acceptability, applicability, and cost-utility of artificial-intelligence-powered low-cost portable fundus camera for diabetic retinopathy screening in primary health care settings. *Diabetes Res Clin Pract* **223**, 112161 (2025).

17. Ahn, S. J. & Kim, Y. H. Clinical Applications and Future Directions of Smartphone Fundus Imaging. *Diagnostics (Basel)* **14**, 1395 (2024).

18. Li, S. *et al.* Developing and validating a clinlabomics-based machine-learning model for early detection of retinal detachment in patients with high myopia. *J Transl Med* **22**, 405 (2024).

19. Li, S. *et al.* Development and validation of a routine blood parameters-based model for screening the occurrence of retinal detachment in high myopia in the context of PPPM. *EPMA J* **14**, 219–233 (2023).

20. Perais, J. *et al.* Prognostic factors for the development and progression of proliferative diabetic retinopathy in people with diabetic retinopathy. *Cochrane Database Syst Rev* **2**, CD013775 (2023).

21. Li, S. *et al.* Association Between 17-β-Estradiol and Interleukin-8 and Visual Field Progression

in Postmenopausal Women with Primary Angle Closure Glaucoma. *Am J Ophthalmol* **217**, 55–67 (2020).

22. Horton, S. *et al.* The Top 25 Laboratory Tests by Volume and Revenue in Five Different Countries. *Am J Clin Pathol* **151**, 446–451 (2019).
23. Foy, B. H. *et al.* Haematological setpoints are a stable and patient-specific deep phenotype. *Nature* **637**, 430–438 (2025).
24. Lundberg, S. M. *et al.* From Local Explanations to Global Understanding with Explainable AI for Trees. *Nat Mach Intell* **2**, 56–67 (2020).
25. Bossi, E. *et al.* Revolutionizing Blood Collection: Innovations, Applications, and the Potential of Microsampling Technologies for Monitoring Metabolites and Lipids. *Metabolites* **14**, 46 (2024).
26. Gause, W. C., Wynn, T. A. & Allen, J. E. Type 2 immunity and wound healing: evolutionary refinement of adaptive immunity by helminths. *Nat Rev Immunol* **13**, 607–614 (2013).
27. Rosenberg, H. F., Dyer, K. D. & Foster, P. S. Eosinophils: changing perspectives in health and disease. *Nat Rev Immunol* **13**, 9–22 (2013).
28. Nishiguchi, K. M. *et al.* C9-R95X polymorphism in patients with neovascular age-related macular degeneration. *Invest Ophthalmol Vis Sci* **53**, 508–512 (2012).
29. Leung, D. Y. M. *et al.* Dupilumab Inhibits Vascular Leakage of Blood Proteins Into Atopic Dermatitis Skin. *J Allergy Clin Immunol Pract* **11**, 1421–1428 (2023).
30. Gao, L., Jiang, W., Liu, H., Chen, Z. & Lin, Y. Receptor-selective interleukin-4 mutein attenuates laser-induced choroidal neovascularization through the regulation of macrophage polarization

in mice. *Exp Ther Med* **22**, 1367 (2021).

31. Ames, B. N., Cathcart, R., Schwiers, E. & Hochstein, P. Uric acid provides an antioxidant defense in humans against oxidant- and radical-caused aging and cancer: a hypothesis. *Proc Natl Acad Sci U S A* **78**, 6858–6862 (1981).
32. van Leeuwen, E. M. *et al.* A new perspective on lipid research in age-related macular degeneration. *Prog Retin Eye Res* **67**, 56–86 (2018).
33. Li, B., Goss, D., Miller, J. W., Lin, J. B. & Vavvas, D. G. Systemic Dyslipidemia in Age-related Macular Degeneration: An Updated Systematic Review and Meta-analysis. *Ophthalmol Sci* **4**, 100341 (2024).
34. Ma, Y., Li, S., Shao, M., Cao, W. & Sun, X. Platelet Parameters and Their Relationships With the Thickness of the Retinal Nerve Fiber Layer and Ganglion Cell Complex in Primary Open-Angle Glaucoma. *Front Neurol* **13**, 867465 (2022).
35. Qiu, J. *et al.* Development and Validation of a Multimodal Multitask Vision Foundation Model for Generalist Ophthalmic Artificial Intelligence. *NEJM AI* https://doi.org/10.1056/AIoa2300221 (2024) doi:10.1056/AIoa2300221.
36. Babenko, B. *et al.* A deep learning model for novel systemic biomarkers in photographs of the external eye: a retrospective study. *Lancet Digit Health* **5**, e257–e264 (2023).
37. Wang, J. *et al.* Artificial intelligence-enhanced retinal imaging as a biomarker for systemic diseases. *Theranostics* **15**, 3223–3233 (2025).
38. Xiao, W. *et al.* Screening and identifying hepatobiliary diseases through deep learning using

ocular images: a prospective, multicentre study. *Lancet Digit Health* **3**, e88–e97 (2021).

39. Mitani, A. *et al.* Detection of anaemia from retinal fundus images via deep learning. *Nat Biomed Eng* **4**, 18–27 (2020).
40. Joo, Y. S. *et al.* Non-invasive chronic kidney disease risk stratification tool derived from retina-based deep learning and clinical factors. *NPJ Digit Med* **6**, 114 (2023).
41. Cheung, C. Y. *et al.* A deep-learning system for the assessment of cardiovascular disease risk via the measurement of retinal-vessel calibre. *Nat Biomed Eng* **5**, 498–508 (2021).
42. Zhang, L. *et al.* Prediction of hypertension, hyperglycemia and dyslipidemia from retinal fundus photographs via deep learning: A cross-sectional study of chronic diseases in central China. *PLoS One* **15**, e0233166 (2020).
43. Rajalakshmi, R., Prathiba, V., Arulmalar, S. & Usha, M. Review of retinal cameras for global coverage of diabetic retinopathy screening. *Eye (Lond)* **35**, 162–172 (2021).
44. Peng, Y. *et al.* Eye care use among rural adults in China: the Handan Eye Study. *Ophthalmic Epidemiol* **20**, 274–280 (2013).
45. Liang, Y. B. *et al.* Refractive errors in a rural Chinese adult population the Handan eye study. *Ophthalmology* **116**, 2119–2127 (2009).
46. Li, S. *et al.* A noninvasive machine learning model using a complete blood count for screening of primary vitreoretinal lymphoma. *Nat Commun* **16**, 10667 (2025).
47. Li, S. *et al.* Metabolomics identifies and validates serum androstenedione as novel biomarker for diagnosing primary angle closure glaucoma and predicting the visual field progression. *Elife*

12, RP91407 (2024).

48. Li, S. *et al.* Serum metabolite biomarkers for the early diagnosis and monitoring of age-related macular degeneration. *J Adv Res* S2090-1232(24)00434-X (2024) doi:10.1016/j.jare.2024.10.001.
49. Long, Q., Ye, J., Li, Y., Wang, S. & Jiang, Y. C-reactive protein and complement components in patients with pathological myopia. *Optom Vis Sci* **90**, 501–506 (2013).
50. Arndt, C. *et al.* Increased intravitreal glucose in rhegmatogenous retinal detachment. *Eye (Lond)* **37**, 638–643 (2023).
51. Luo, S. *et al.* Quantitative proteomics analysis of human vitreous in rhegmatogenous retinal detachment associated with choroidal detachment by data-independent acquisition mass spectrometry. *Mol Cell Biochem* **477**, 1849–1863 (2022).
52. Li, S. *et al.* Association of systemic inflammation indices with visual field loss progression in patients with primary angle-closure glaucoma: potential biomarkers for 3P medical approaches. *EPMA J* **12**, 659–675 (2021).
53. Li, S. *et al.* Association between coagulation function and patients with primary angle closure glaucoma: a 5-year retrospective case-control study. *BMJ Open* **7**, e016719 (2017).
54. Shao, M., Li, Y., Teng, J., Li, S. & Cao, W. Association Between Serum Lipid Levels and Patients With Primary Angle-Closure Glaucoma in China: A Cross Sectional, Case-Control Study. *Front Med (Lausanne)* **8**, 618970 (2021).
55. Shao, M. *et al.* Association Between Serum Complement C3 Levels and Age-Related Cataract.

Invest Ophthalmol Vis Sci **58**, 4934–4939 (2017).

56. Li, S. *et al.* Association between serum lipids concentration and patients with age-related cataract in China: a cross-sectional, case-control study. *BMJ Open* **8**, e021496 (2018).
57. Li, B. *et al.* Predictive role of the peripheral blood inflammation indices neutrophil-to-lymphocyte ratio (NLR), platelet-to-lymphocyte ratio (PLR), and systemic immunoinflammatory index (SII) for age-related cataract risk. *PLoS One* **19**, e0313503 (2024).
58. Wang, Y. *et al.* The Association between the Lipids Levels in Blood and Risk of Age-Related Macular Degeneration. *Nutrients* **8**, 663 (2016).
59. Pinna, A. *et al.* Complete Blood Cell Count-Derived Inflammation Biomarkers in Men with Age-Related Macular Degeneration. *Ocul Immunol Inflamm* **27**, 932–936 (2019).
60. Berkelmans, G. F. N. *et al.* Population median imputation was noninferior to complex approaches for imputing missing values in cardiovascular prediction models in clinical practice. *J Clin Epidemiol* **145**, 70–80 (2022).
61. Grandini, M., Bagli, E. & Visani, G. Metrics for Multi-Class Classification: an Overview. Preprint at https://doi.org/10.48550/arXiv.2008.05756 (2020).
62. Hanley, J. A. & McNeil, B. J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **143**, 29–36 (1982).
63. Kerr, K. F., Brown, M. D., Zhu, K. & Janes, H. Assessing the Clinical Impact of Risk Prediction Models With Decision Curves: Guidance for Correct Interpretation and Appropriate Use. *J Clin Oncol* **34**, 2534–2540 (2016).

64. Steyerberg, E. W. *et al.* Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* **21**, 128–138 (2010).

65. Hu, J. *et al.* Identification and validation of an explainable prediction model of acute kidney injury with prognostic implications in critically ill children: a prospective multicenter cohort study. *EClinicalMedicine* **68**, 102409 (2024).

66. fudanRenjun. fudanRenjun/high-myopia: high-myopia. Zenodo https://doi.org/10.5281/zenodo.18382509 (2026).

Acknowledgements: This work was supported by the National Key Research and Development Program of China (2024YFC2510805, XT Z), Shanghai Municipal Health Commission (2023ZZ02019, XT Z), and the National Natural Science Foundation of China (82302582, SJ L). The sponsors or funding organizations played no role in the design or execution of this research. We gratefully acknowledge the Putuo District Central Hospital of Shanghai University for providing the sample data used in this study.

Author Contributions: W.J.C., X.T.Z, M.Y.L., and S.J.L. conceptualized and designed this study. S.J.L., Y.Z.L., J.N.W., F.L.W., X.X.W., M.Y.Z., H.G.H., Y.X.S., and J.R. performed most experiments. W.J.C., X.T.Z, and M.Y.L. performed partial experiments. S.J.L., J.N.W., F.L.W., and J.R. finished the acquisition and analysis of data. S.J.L., J.R., F.L.W., and J.N.W. prepared figures, performed the statistical analysis. S.J.L. wrote the original draft. W.J.C., X.T.Z, M.Y.L., and S.J.L. reviewed and supervised the manuscript. All the authors have read and approved the final manuscript.

Competing Interests: The authors declare no competing interests.

Table 1. Nine features-based RF Model Performance in the discovery cohort and validation cohorts

	Disease	Sensitivity (95%CI)	Specificity (95%CI)	Accuracy (95%CI)	PPV (95%CI)	NPV (95%CI)	F1 (95%CI)	AUC (95%CI)	PRAUC (95%CI)
Discovery Cohort 1	HM	0.8423 (0.7987, 0.8823)	0.9463 (0.8967, 0.9943)	0.9171 (0.8799, 0.9478)	0.8602 (0.8075, 0.9178)	0.9387 (0.8915, 0.9926)	0.8512 (0.8117, 0.8843)	0.9736 (0.9480, 0.9880)	0.9495 (0.8912, 0.9933)
	HM associated with cataract	0.8934 (0.8398, 0.9297)	0.8864 (0.8383, 0.9175)	0.8890 (0.8530, 0.9410)	0.8237 (0.7660, 0.8618)	0.9333 (0.8821, 0.9692)	0.8571 (0.8070, 0.9109)	0.9790 (0.9342, 0.9885)	0.9709 (0.9427, 0.9788)
	HM associated with MD	0.7759 (0.7182, 0.8305)	0.9689 (0.9257, 0.9755)	0.9428 (0.8960, 0.9737)	0.7965 (0.7429, 0.8375)	0.9650 (0.9150, 0.9797)	0.7860 (0.7436, 0.8160)	0.9705 (0.9272, 0.9772)	0.9034 (0.8528, 0.9520)
	HM associated with CN	0.6867 (0.6486, 0.7280)	0.9832 (0.9416, 0.9940)	0.9544 (0.9109, 0.9617)	0.8143 (0.7694, 0.8728)	0.9669 (0.9465, 0.9759)	0.7451 (0.6978, 0.7964)	0.9589 (0.9348, 0.9706)	0.8350 (0.7926, 0.8748)
	HM associated with glaucoma	0.6857 (0.6513, 0.7428)	0.9915 (0.9546, 1.0000)	0.9790 (0.9439, 0.9847)	0.7742 (0.7180, 0.8307)	0.9867 (0.9478, 0.9974)	0.7273 (0.6941, 0.7578)	0.9475 (0.8976, 0.9788)	0.7795 (0.7436, 0.8299)
	HM associated with RD	0.7742 (0.7265, 0.8065)	0.9849 (0.9517, 0.9929)	0.9696 (0.9276, 0.9790)	0.8000 (0.7568, 0.8489)	0.9824 (0.9523, 0.9894)	0.7869 (0.7352, 0.8387)	0.9599 (0.9214, 0.9746)	0.8488 (0.7957, 0.8823)
	Mean	0.7764 (0.6894, 0.8633)	0.9602 (0.9187, 1.0000)	0.9420 (0.9065, 0.9775)	0.8115 (0.7808, 0.8422)	0.9622 (0.9391, 0.9853)	0.7923 (0.7364, 0.8482)	0.9649 (0.9528, 0.9770)	0.8812 (0.8045, 0.9579)
Validation cohort 1	HM	0.7229 (0.6743, 0.7656)	0.8702 (0.8292, 0.9084)	0.8298 (0.7907, 0.8767)	0.6779 (0.6468, 0.7272)	0.8926 (0.8541, 0.9308)	0.6997 (0.6500, 0.7546)	0.9021 (0.8543, 0.9505)	0.8288 (0.7767, 0.8682)
	HM associated with cataract	0.8482 (0.7973, 0.9016)	0.7610 (0.7012, 0.8011)	0.7930 (0.7330, 0.8493)	0.6735 (0.6156, 0.7194)	0.8961 (0.8361, 0.9302)	0.7508 (0.6975, 0.7876)	0.9133 (0.8626, 0.9640)	0.8746 (0.8337, 0.9333)
	HM associated with MD	0.5888 (0.5396, 0.6430)	0.9864 (0.9422, 0.9943)	0.9359 (0.8910, 0.9891)	0.8630 (0.8088, 0.9175)	0.9428 (0.8843, 0.9950)	0.7000 (0.6619, 0.7469)	0.8951 (0.8505, 0.9437)	0.7497 (0.7118, 0.7872)
	HM associated with CN	0.5840 (0.5208, 0.6374)	0.9886 (0.9546, 0.9958)	0.9505 (0.9256, 0.9616)	0.8424 (0.7994, 0.8731)	0.9581 (0.9261, 0.9640)	0.6898 (0.6320, 0.7355)	0.8948 (0.8629, 0.9390)	0.7268 (0.6859, 0.7643)
	HM associated with glaucoma	0.7481 (0.6910, 0.7890)	0.9996 (0.9757, 1.0000)	0.9861 (0.9379, 0.9954)	0.9902 (0.9665, 0.9974)	0.9860 (0.9415, 0.9964)	0.8523 (0.8130, 0.8934)	0.9324 (0.8739, 0.9649)	0.8028 (0.7624, 0.8561)
	HM associated with RD	0.5592 (0.5028, 0.6036)	0.9940 (0.9655, 1.0000)	0.9577 (0.9256, 0.9655)	0.8939 (0.8424, 0.9381)	0.9612 (0.9286, 0.9747)	0.6880 (0.6354, 0.7293)	0.8745 (0.8319, 0.9193)	0.6958 (0.6650, 0.7501)
	Mean	0.6752 (0.5539, 0.7965)	0.9333 (0.8309, 1.0000)	0.9088 (0.8269, 0.9908)	0.8235 (0.6921, 0.9549)	0.9395 (0.9000, 0.9789)	0.7301 (0.6628, 0.7974)	0.9020 (0.8815, 0.9225)	0.7797 (0.7091, 0.8504)
Validation cohort 2	HM	0.7456 (0.7152, 0.7969)	0.8774 (0.8188, 0.9281)	0.8389 (0.7816, 0.8880)	0.7149 (0.6720, 0.7737)	0.8932 (0.8408, 0.9397)	0.7299 (0.6953, 0.7731)	0.9064 (0.8604, 0.9370)	0.8438 (0.7902, 0.8797)
	HM associate	0.8551	0.7828	0.8086	0.6867	0.9065	0.7617	0.9164	0.8766

d with cataract	(0.8099, 0.9115)	(0.7458, 0.8198)	(0.7669, 0.8517)	(0.6497, 0.7418)	(0.8732, 0.9464)	(0.7264, 0.8107)	(0.8678, 0.9569)	(0.8318, 0.9290)
HM associated with MD	0.6007 (0.5342, 0.6616)	0.9812 (0.9437, 0.9924)	0.9330 (0.8846, 0.9756)	0.8224 (0.7896, 0.8766)	0.9443 (0.8867, 0.9784)	0.6943 (0.6536, 0.7296)	0.8988 (0.8520, 0.9389)	0.7511 (0.7140, 0.8046)
HM associated with CN	0.5696 (0.5041, 0.6110)	0.9894 (0.9620, 0.9975)	0.9477 (0.9146, 1.0000)	0.8562 (0.8169, 0.9096)	0.9542 (0.9064, 0.9600)	0.6841 (0.6251, 0.7152)	0.8752 (0.8167, 0.9216)	0.7010 (0.6427, 0.7384)
HM associated with glaucoma	0.7890 (0.7427, 0.8284)	0.9995 (0.9667, 1.0000)	0.9896 (0.9679, 1.0000)	0.9885 (0.9388, 0.9981)	0.9897 (0.9671, 0.9951)	0.8776 (0.8404, 0.9104)	0.9351 (0.8876, 0.9651)	0.8257 (0.7719, 0.8682)
HM associated with RD	0.6145 (0.5490, 0.6812)	0.9930 (0.9463, 1.0000)	0.9637 (0.9393, 0.9781)	0.8800 (0.8447, 0.9207)	0.9685 (0.9309, 0.9813)	0.7237 (0.6732, 0.7553)	0.8743 (0.8440, 0.9277)	0.7063 (0.6683, 0.7636)
Mean	0.6957 (0.5733, 0.8182)	0.9372 (0.8444, 1.0000)	0.9136 (0.8373, 0.9899)	0.8248 (0.7079, 0.9416)	0.9427 (0.9041, 0.9813)	0.7452 (0.6713, 0.8192)	0.9010 (0.8761, 0.9259)	0.7841 (0.7057, 0.8625)

HM, high myopia; RD, retinal detachment; MD, macular degeneration; CN, choroidal neovascularization; CI, Confidence Interval; PPV, positive predictive value; NPV, negative predictive value; AUC, area under the receiver operating characteristic curve; PRAUC, area under the precision-recall curve; RF, random forest; F1, the harmonic mean of precision and recall.

Figure Legends

Figure 1. Overview of the blood-based machine learning pipeline for screening high myopia-related complications

This flowchart outlines the development and validation of a machine learning model using routine blood test parameters to screen for HM and its complications. After the exclusion criteria were applied, the data from 5,102 patients were used for model development, and the data from two independent cohorts (n = 3,125 and 2,434) were used for validation. Five HM-related complications were included. Models constructed from multiple algorithms were compared through key metrics (AUC, PR-AUC, sensitivity, specificity, PPV, NPV, and F1 score). A random forest model based on nine selected features showed optimal performance and was evaluated with a ROC curve analysis and confusion matrices. A web-based application was developed to facilitate clinical use. HM, high myopia; RD, retinal detachment; MD, macular degeneration; CN, choroidal neovascularization; PPV, positive predictive value; NPV, negative predictive value; AUC, area under the receiver operating characteristic curve; PR-AUC, area under the precision–recall curve; RF, random forest; ROC, receiver operating characteristic curve; F1 score, the harmonic mean of precision and recall; UA, uric acid; EOS, eosinophil; PCT, thrombocytocrit; BASO%, basophil percentage, TG, triglyceride.

Figure 2. Performance of machine learning models for screening high myopia-related complications using the results of routine laboratory tests in the discovery cohort

(A) Comparison of multiple machine learning algorithms using key performance metrics, including AUC, sensitivity, specificity, PPV, NPV, accuracy, and F1 score. The results obtained through fivefold cross-validation are reported as the means ± standard deviations. Source data are provided as a Source Data file.

(B) Heatmap comparing the AUC values of different models. The DeLong test, a nonparametric statistical method, was used to assess differences between the areas under the ROC curves.

(C) ROC curves of the RF model for screening high myopia-related complications.

(D) Confusion matrix heatmap illustrating the accuracy, sensitivity, and specificity of the predictions from the RF model.

(E) ROC curves of the XGBoost model for screening high myopia-related complications.

(F) Confusion matrix heatmap illustrating the accuracy, sensitivity, and specificity of the predictions from the XGBoost model.

HM, high myopia; RD, retinal detachment; MD, macular degeneration; CN, choroidal neovascularization; PPV, positive predictive value; NPV, negative predictive value; AUC, area under the receiver operating characteristic curve; ROC, receiver operating characteristic curve; F1 score, the harmonic mean of precision and recall; RF, random forest; DT, decision tree; GB, gradient boosting; KNN, k-nearest neighbours; LGBM, light gradient boosting machine; LR, logistic regression; MLP, multilayer perceptron; NB, naïve Bayes; XGBoost, extreme gradient boosting.

Figure 3. Feature characteristics and identification of the final model for screening high myopia-related complications in the discovery cohort

(A) AUC values for models with different numbers of features, showing that the all-feature RF model performed better than the 8-feature model did (P = 0.0227) but did not perform significantly better than the 9-feature model (P = 0.0970). With respect to the XGBoost model, the all-feature model outperformed the 19-feature model (P < 0.0010) but did not significantly outperform the 20-feature model (P = 0.1860). The DeLong test (two-sided) was used to assess differences between the areas under the ROC curves. Source data are provided as a Source Data file.

(B) Feature importance ranking scores for all variables associated with complications of high myopia.

(C) ROC curves of the RF model based on the top nine features for screening high myopia-related complications.

(D) PR-AUC curves of the RF model based on the top nine features.

(E) Confusion matrix heatmap illustrating the accuracy, sensitivity, and specificity of the predictions from the RF model based on the nine selected features.

HM, high myopia; RD, retinal detachment; MD, macular degeneration; CN, choroidal neovascularization; PPV, positive predictive value; NPV, negative predictive value; AUC, area under the receiver operating characteristic curve; ROC, receiver operating characteristic curve; PR-AUC, area under the precision–recall curve; RF: random forest; XGB, extreme gradient boosting; EOS, eosinophils; PCT, thrombocytocrit; BASO%, basophil percentage; M, monocytes; UA, uric acid; LPB, apolipoprotein b; BASO, basophils; EOS%, eosinophil percentage; TG, triglyceride; P, phosphorus; LDL, low-density lipoprotein; HB, hemoglobin; GLB, globulin; PA, prealbumin; MPV, mean platelet volume; PT, prothrombin time; DD, d-dimer; FIB, fibrinogen; GLU, glucose; LDH, lactic dehydrogenase, SHAP, SHapley Additive exPlanations.

Figure 4. External validation of the final RF model in validation cohorts 1 and 2

(A) ROC curves of the RF model based on the top nine features for screening high myopia-related complications in validation cohort 1.

(B) PR-AUC curves of the RF model based on the top nine features in validation cohort 1.

(C) Confusion matrix heatmap illustrating the accuracy, sensitivity, and specificity of the predictions from the RF model based on the nine selected features in validation cohort 1.

(D) ROC curves of the RF model based on the top nine features for screening high myopia-related complications in validation cohort 2.

(E) PR-AUC curves of the RF model based on the top nine features in validation cohort 2.

(F) Confusion matrix heatmap illustrating the accuracy, sensitivity, and specificity of the predictions from the RF model based on the nine selected features in validation cohort 2.

(G) A comparison of multiple machine learning algorithms was conducted in validation cohort 1 using key performance metrics, including the AUC, sensitivity, specificity, PPV, NPV, accuracy, and F1 score. Except

for the final RF model, the performance of all the other models was analysed using the complete set of blood test variables as input features. Source data are provided as a Source Data file.

(H) A comparison of multiple machine learning algorithms was conducted in validation cohort 2 using key performance metrics, including the AUC, sensitivity, specificity, PPV, NPV, accuracy, and F1 score. Except for the final RF model, the performance of all the other models was analysed using the complete set of blood test variables as input features. Source data are provided as a Source Data file.

HM, high myopia; RD, retinal detachment; MD, macular degeneration; CN, choroidal neovascularization; PPV, positive predictive value; NPV, negative predictive value; AUC, area under the receiver operating characteristic curve; ROC, receiver operating characteristic curve; PR-AUC, area under the precision–recall curve; RF, random forest; DT, decision tree; GB, gradient boosting; KNN, k-nearest neighbours; LGBM, light gradient boosting machine; LR, logistic regression; MLP, multilayer perceptron; NB, naïve Bayes; XGBoost, extreme gradient boosting.

Figure 5. Calibration curves for the RF model in the discovery cohort and validation cohorts 1 and 2

(A) Calibration curve of the RF model for the discovery cohort, showing good calibration for HM and HM associated with MD at higher predicted probabilities (>0.6). The model slightly overestimated HM with cataracts and underestimated HM with choroidal neovascularization, glaucoma, and RD, as indicated by the curves below the ideal line.

(B) Calibration curve of the RF model for validation cohort 1, with performance similar to that in the discovery cohort.

(C) Calibration curve of the RF model for validation cohort 2, confirming the reliability and generalizability of the model.

HM, high myopia; RD, retinal detachment; MD, macular degeneration; CN, choroidal neovascularization;

RF: random forest.

Figure 6. Analysis of the clinical utility of the model

(A) SHAP summary plot depicting the relative importance of the nine selected features based on their mean absolute SHAP values, providing a global understanding of the model's behaviour.

(B) SHAP dependence plots showing the influence of individual features on the model's predictions. We developed a freely accessible online web application to support clinical integration, which is available at https://high-myopia.streamlit.app/. This interactive platform allows clinicians to enter the nine required feature values and then automatically calculates and displays the predicted risk based on the random forest model.

HM, high myopia; RD, retinal detachment; MD, macular degeneration; CN, choroidal neovascularization. SHAP, SHapley Additive exPlanations, UA, uric acid; EOS%, eosinophil percentage; TG, triglycerides; EOS, eosinophils; LPB, apolipoprotein B; M, monocytes; PCT, plateletcrit; BASO%, basophil percentage; BASO, basophils.

Figure 7. External validation of the RF screening model in hospital-based and community-based cohorts.

(A) Screening workflow of the hospital-based prospective cohort study.

(B) Confusion matrix heatmap depicting the prediction performance (accuracy, sensitivity, and specificity) of the RF model in the hospital-based cohort.

(C) Summary of the prediction metrics (accuracy, sensitivity, specificity, PPV, and NPV) of the RF model in the hospital-based cohort. Source data are provided as a Source Data file.

(D) Screening workflow of the community-based cross-sectional study.

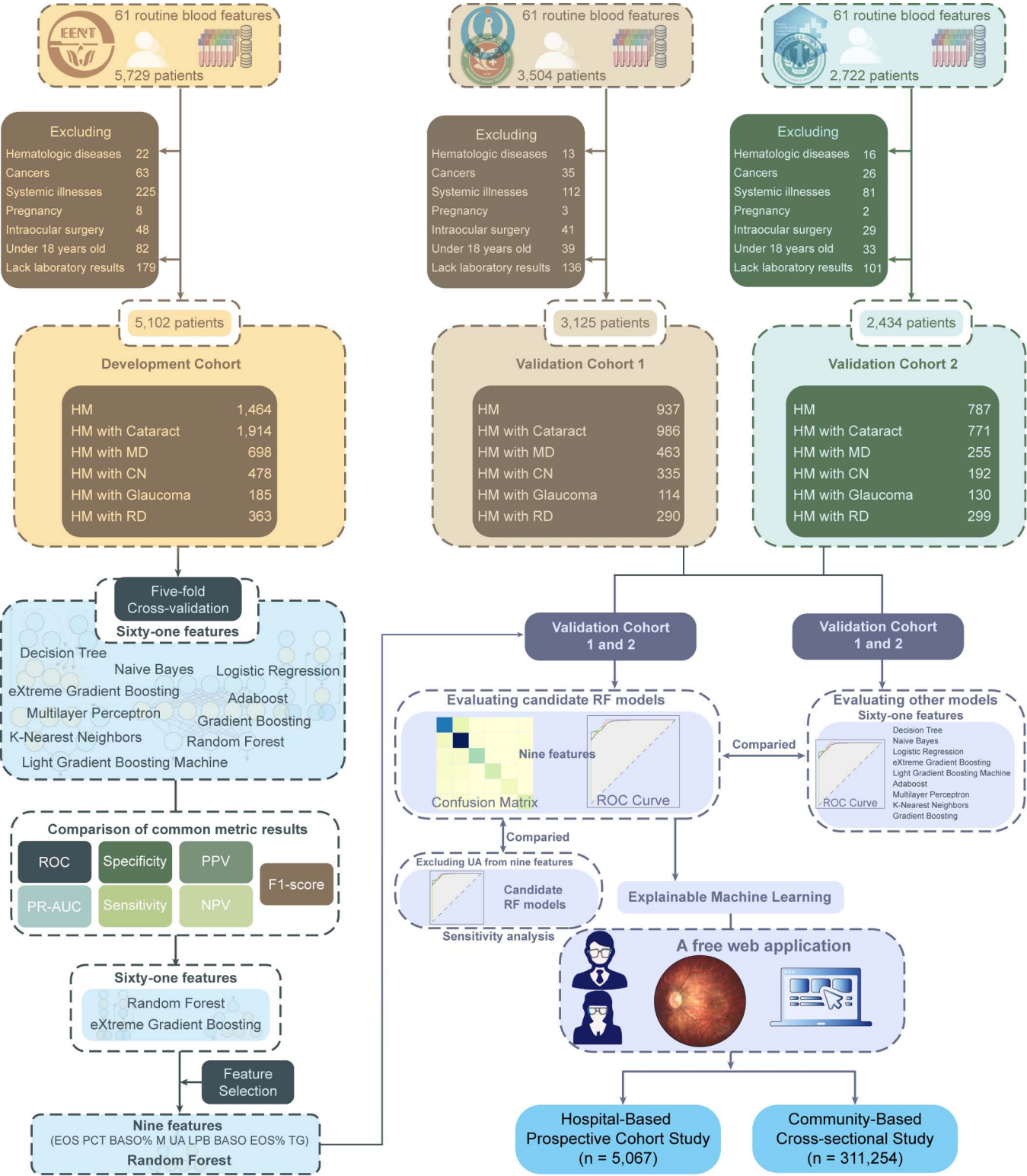
HM, high myopia; RD, retinal detachment; MD, macular degeneration; CN, choroidal neovascularization;

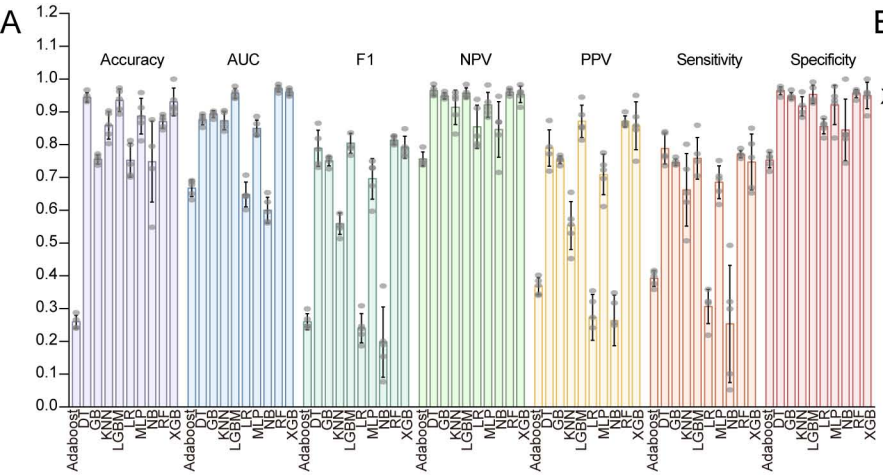
PPV, positive predictive value; NPV, negative predictive value; RF: random forest.

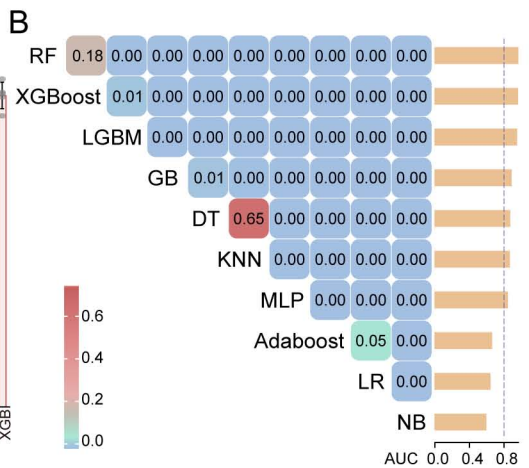
Editor's Summary

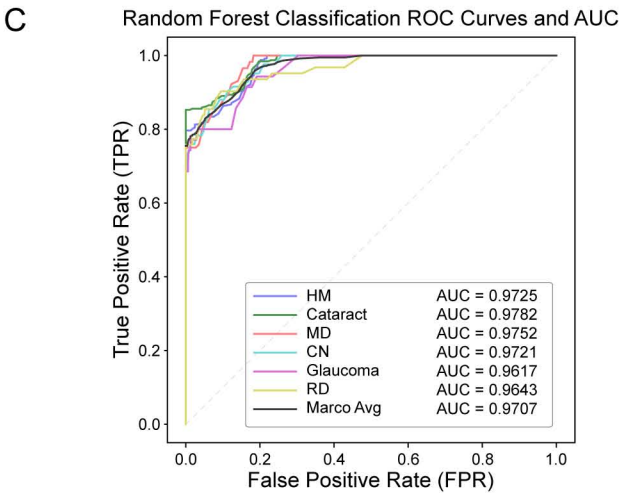
Authors develop a machine learning model using routine blood tests that identify people at higher risk of complications from high myopia. Their results support opportunistic screening during health checks to enable earlier referral.

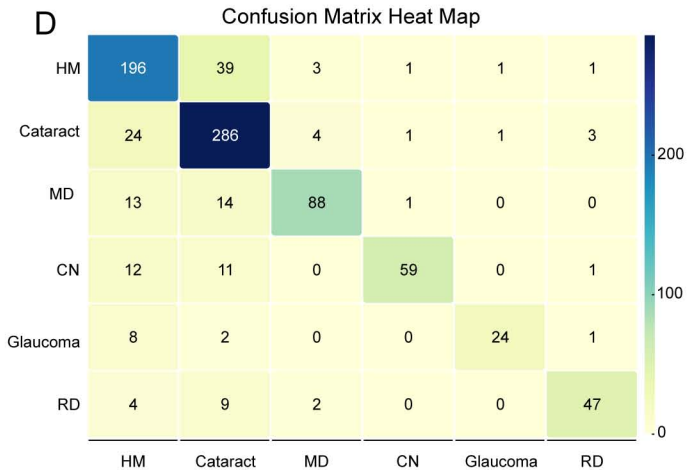
Peer review information: *Nature Communications* thanks Mengtian Kang, Carla Lança and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

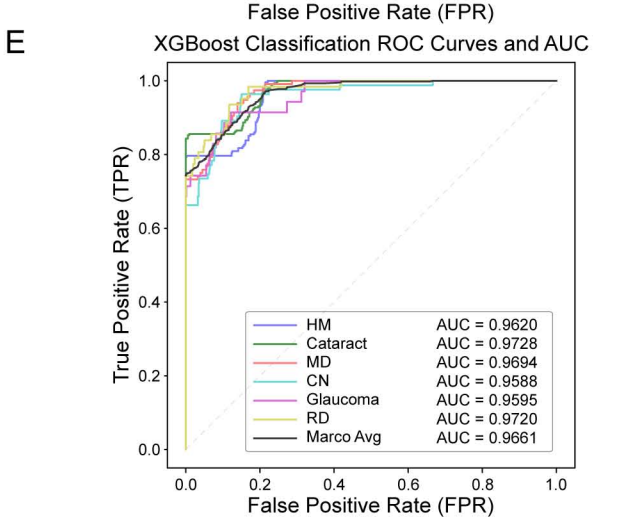

61 routine blood features
5,729 patients
61 routine blood features
3,504 patients
61 routine blood features
2,722 patients
Excluding
Hematologic diseases 22
Cancers 63
Systemic illnesses 225
Pregnancy 8
Intraocular surgery 48
Under 18 years old 82
Lack laboratory results 179
Excluding
Hematologic diseases 13
Cancers 35
Systemic illnesses 112
Pregnancy 3
Intraocular surgery 41
Under 18 years old 39
Lack laboratory results 136
Excluding
Hematologic diseases 16
Cancers 26
Systemic illnesses 81
Pregnancy 2
Intraocular surgery 29
Under 18 years old 33
Lack laboratory results 101
5,102 patients
Development Cohort
HM 1,464
HM with Cataract 1,914
HM with MD 698
HM with CN 478
HM with Glaucoma 185
HM with RD 363
3,125 patients
Validation Cohort 1
HM 937
HM with Cataract 986
HM with MD 463
HM with CN 335
HM with Glaucoma 114
HM with RD 290
2,434 patients
Validation Cohort 2
HM 787
HM with Cataract 771
HM with MD 255
HM with CN 192
HM with Glaucoma 130
HM with RD 299
Five-fold Cross-validation
Sixty-one features
Decision Tree
Naive Bayes
Logistic Regression
eXtreme Gradient Boosting
Adaboost
Multilayer Perceptron
Gradient Boosting
K-Nearest Neighbors
Random Forest
Light Gradient Boosting Machine
Validation Cohort 1 and 2
Validation Cohort 1 and 2
Evaluating candidate RF models
Nine features
Confusion Matrix
ROC Curve
Compared
Evaluating other models
Sixty-one features
ROC Curve
Decision Tree
Naive Bayes
Logistic Regression
eXtreme Gradient Boosting
Light Gradient Boosting Machine
Adaboost
Multilayer Perceptron
K-Nearest Neighbors
Gradient Boosting
Comparison of common metric results
ROC
Specificity
PPV
F1-score
PR-AUC
Sensitivity
NPV
Compared
Excluding UA from nine features
Candidate RF models
Sensitivity analysis
Explainable Machine Learning
Sixty-one features
Random Forest
eXtreme Gradient Boosting
A free web application
Feature Selection
Nine features
(EOS PCT BASO% M UA LPB BASO EOS% TG)
Random Forest
Hospital-Based Prospective Cohort Study (n = 5,067)
Community-Based Cross-sectional Study (n = 311,254)

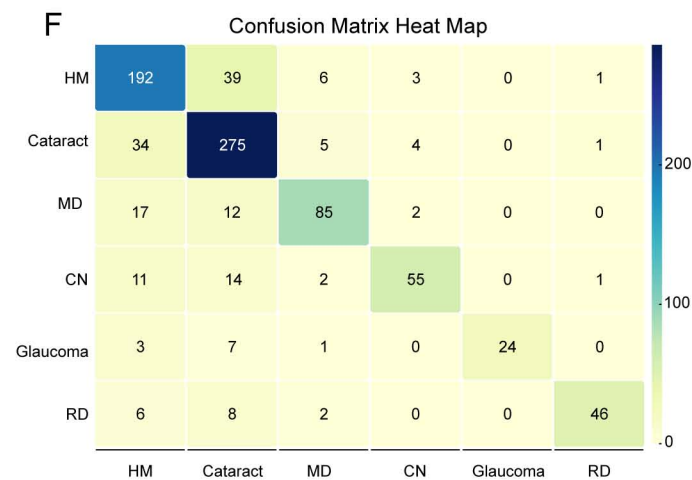

A
Accuracy
AUC
F1
NPV
PPV
Sensitivity
Specificity
Adaboost
DT
GB
KNN
LGBM
LR
MLP
NB
RF
XGB

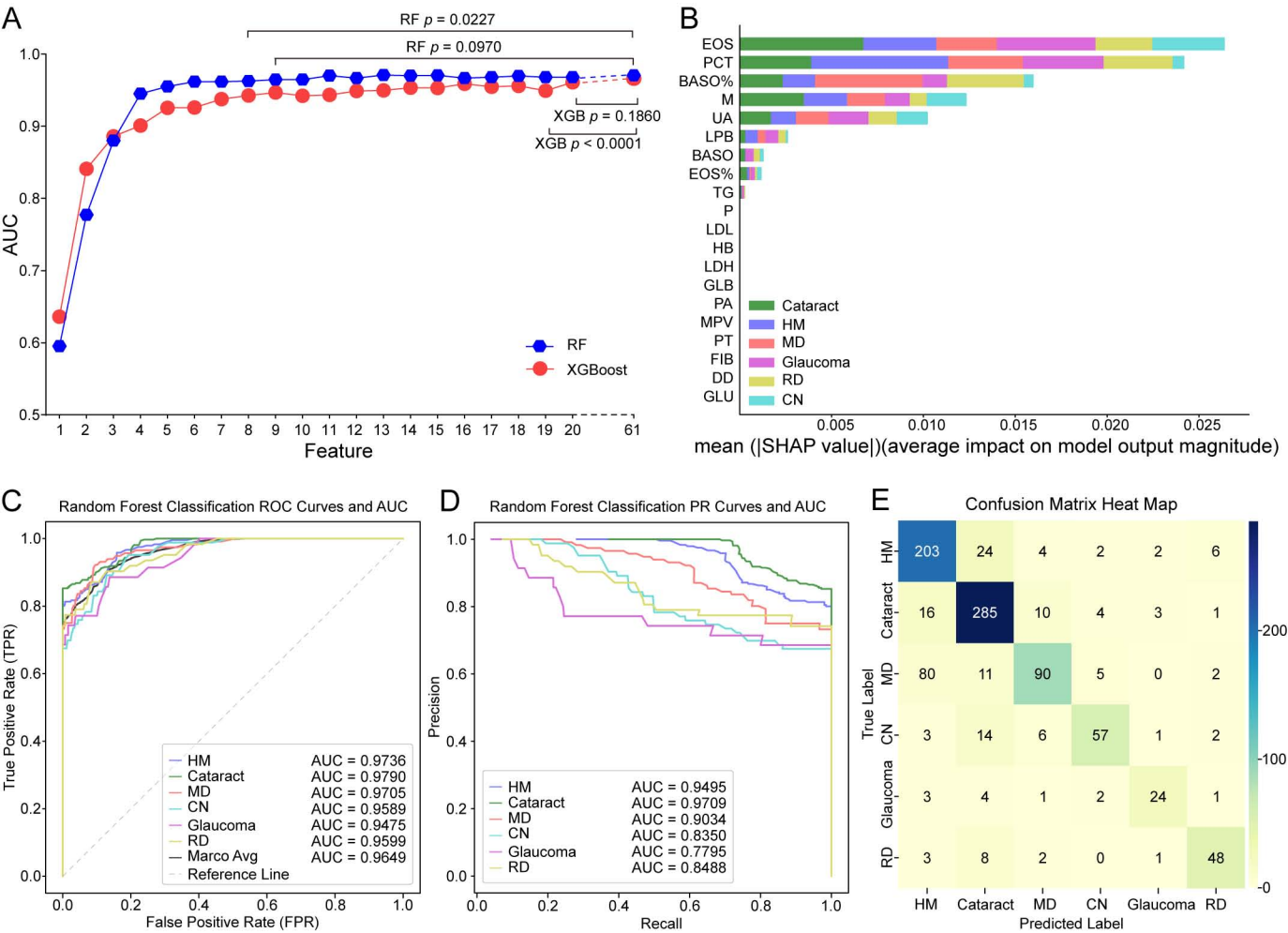

B
RF 0.18 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
XGBoost 0.01 0.00 0.00 0.00 0.00 0.00 0.00 0.00
LGBM 0.00 0.00 0.00 0.00 0.00 0.00 0.00
GB 0.01 0.00 0.00 0.00 0.00 0.00
DT 0.65 0.00 0.00 0.00 0.00
KNN 0.00 0.00 0.00 0.00
MLP 0.00 0.00 0.00
Adaboost 0.05 0.00
LR 0.00
NB
AUC 0.0 0.4 0.8

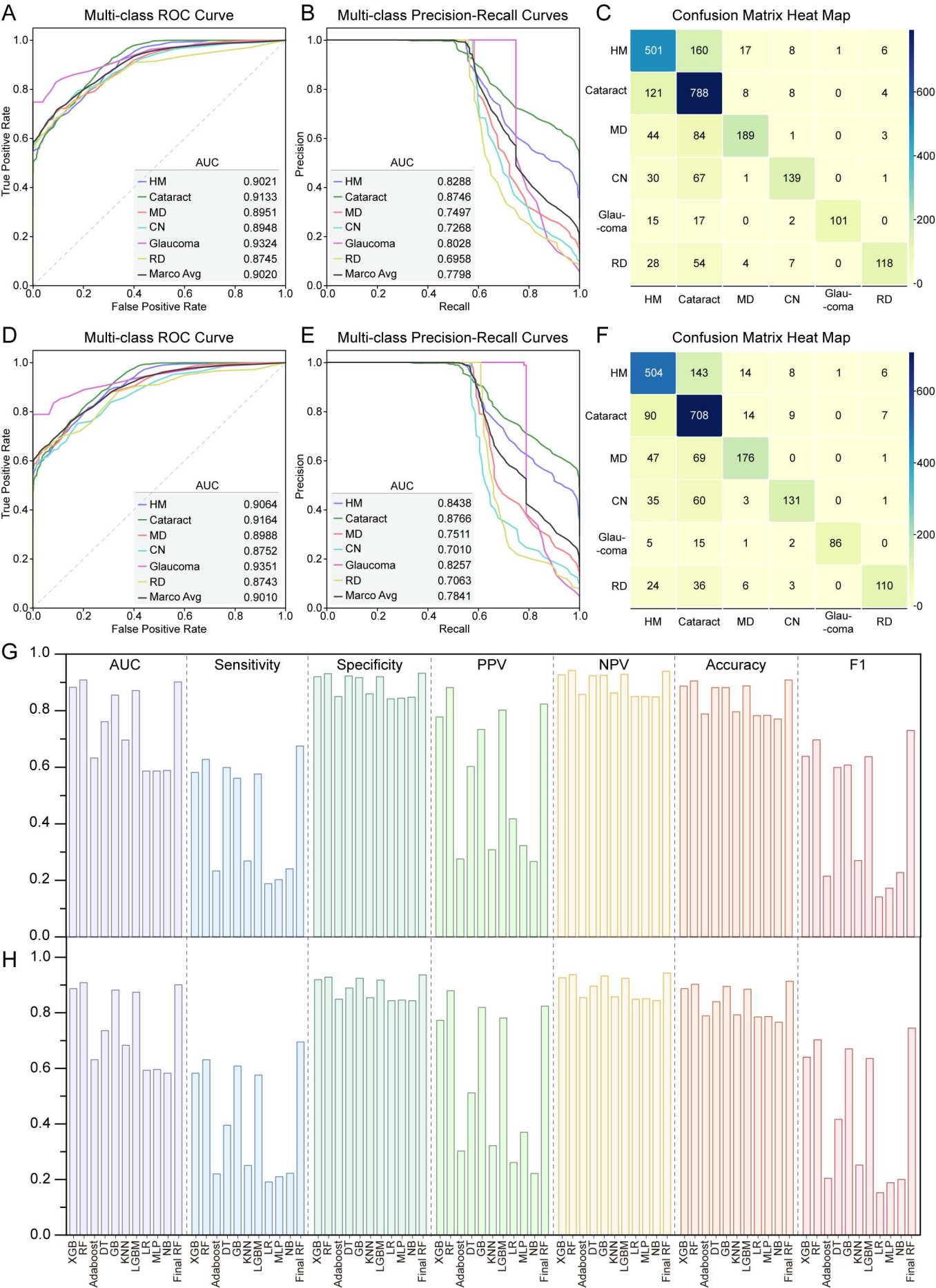

C
Random Forest Classification ROC Curves and AUC
True Positive Rate (TPR)
False Positive Rate (FPR)
HM AUC = 0.9725
Cataract AUC = 0.9782
MD AUC = 0.9752
CN AUC = 0.9721
Glaucoma AUC = 0.9617
RD AUC = 0.9643
Marco Avg AUC = 0.9707

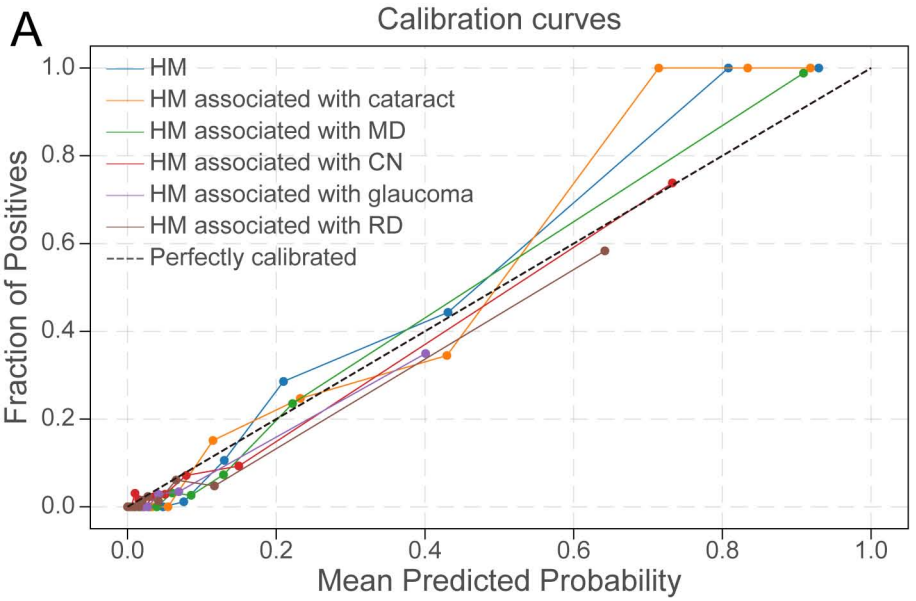

D
Confusion Matrix Heat Map
HM 196 39 3 1 1 1
Cataract 24 286 4 1 1 3
MD 13 14 88 1 0 0
CN 12 11 0 59 0 1
Glaucoma 8 2 0 0 24 1
RD 4 9 2 0 0 47
HM Cataract MD CN Glaucoma RD

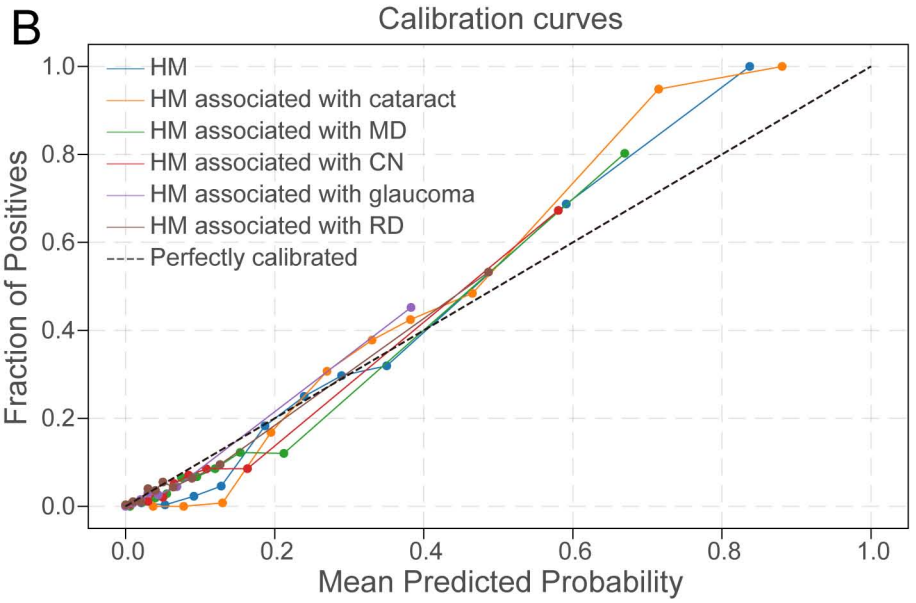

E
XGBoost Classification ROC Curves and AUC
True Positive Rate (TPR)
False Positive Rate (FPR)
HM AUC = 0.9620
Cataract AUC = 0.9728
MD AUC = 0.9694
CN AUC = 0.9588
Glaucoma AUC = 0.9595
RD AUC = 0.9720
Marco Avg AUC = 0.9661

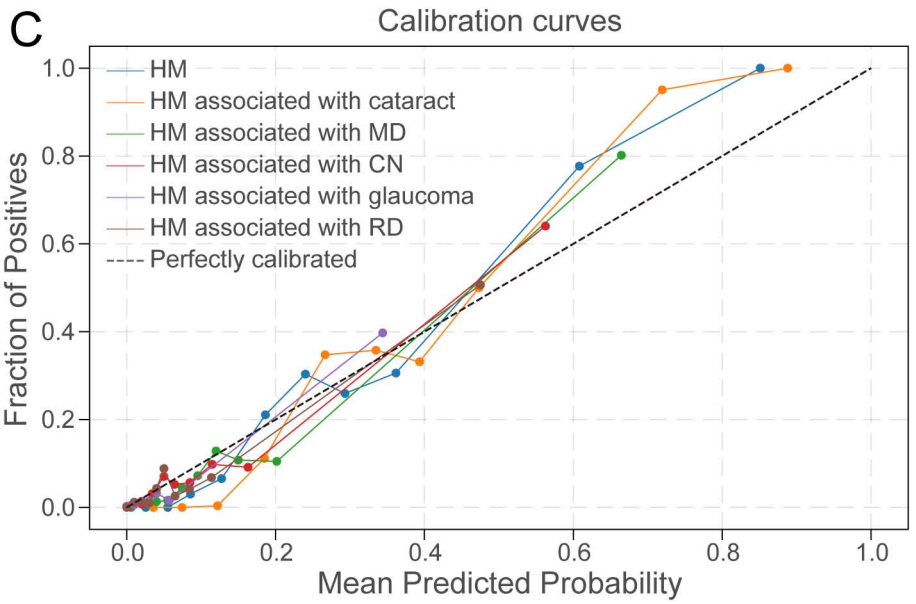

F
Confusion Matrix Heat Map
HM 192 39 6 3 0 1
Cataract 34 275 5 4 0 1
MD 17 12 85 2 0 0
CN 11 14 2 55 0 1
Glaucoma 3 7 1 0 24 0
RD 6 8 2 0 0 46
HM Cataract MD CN Glaucoma RD


A
RF p = 0.0227
RF p = 0.0970
XGB p = 0.1860
XGB p < 0.0001
AUC
Feature
RF
XGBoost
B
EOS
PCT
BASO%
M
UA
LPB
BASO
EOS%
TG
P
LDL
HB
LDH
GLB
PA
MPV
PT
FIB
DD
GLU
Cataract
HM
MD
Glaucoma
RD
CN
mean (|SHAP value|)(average impact on model output magnitude)
C
Random Forest Classification ROC Curves and AUC
True Positive Rate (TPR)
False Positive Rate (FPR)
HM AUC = 0.9736
Cataract AUC = 0.9790
MD AUC = 0.9705
CN AUC = 0.9589
Glaucoma AUC = 0.9475
RD AUC = 0.9599
Marco Avg AUC = 0.9649
Reference Line
D
Random Forest Classification PR Curves and AUC
Precision
Recall
HM AUC = 0.9495
Cataract AUC = 0.9709
MD AUC = 0.9034
CN AUC = 0.8350
Glaucoma AUC = 0.7795
RD AUC = 0.8488
E
Confusion Matrix Heat Map
True Label
Predicted Label
HM 203 24 4 2 2 6
Cataract 16 285 10 4 3 1
MD 80 11 90 5 0 2
CN 3 14 6 57 1 2
Glaucoma 3 4 1 2 24 1
RD 3 8 2 0 1 48

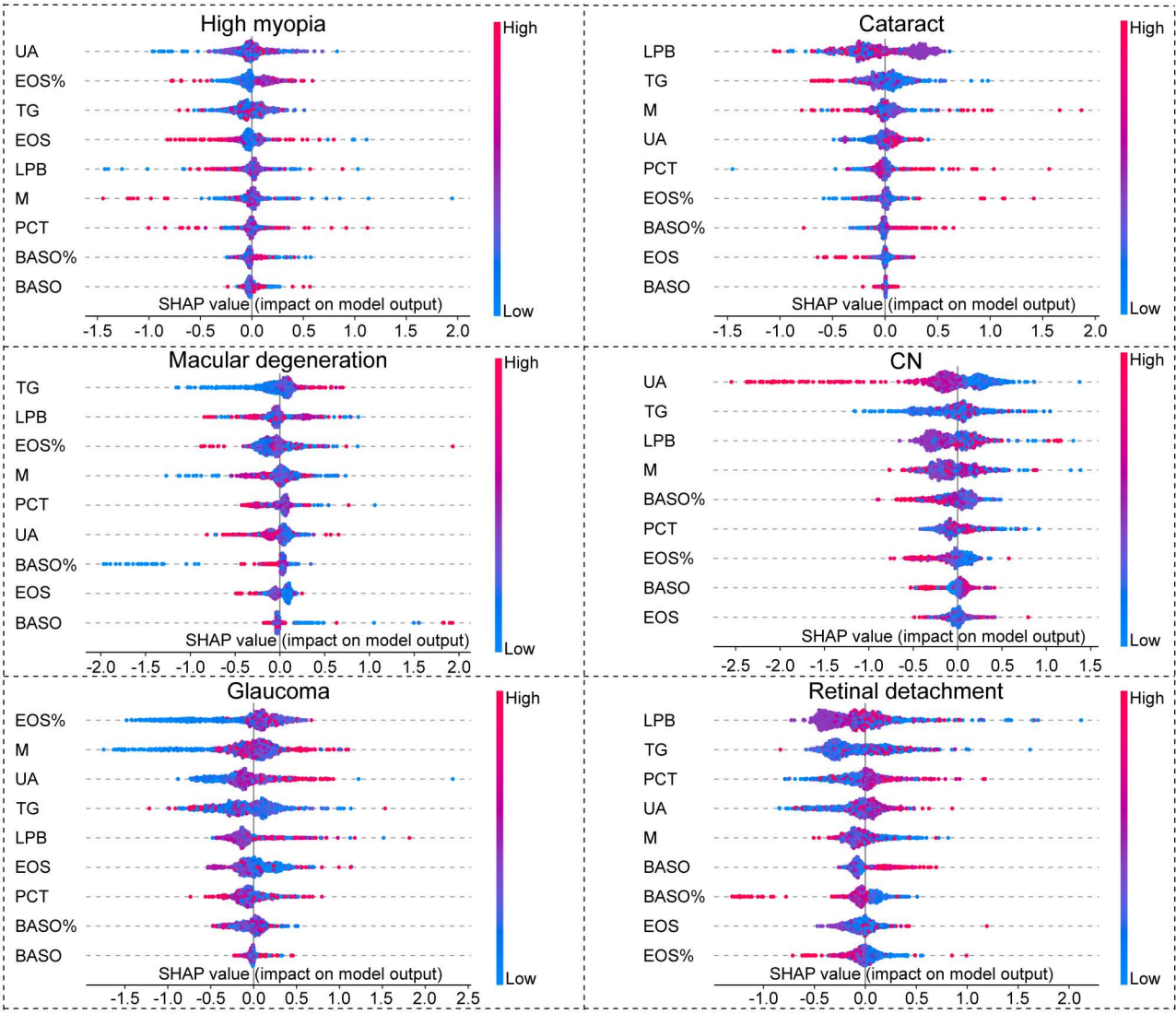

A
Multi-class ROC Curve
True Positive Rate
False Positive Rate
AUC
HM 0.9021
Cataract 0.9133
MD 0.8951
CN 0.8948
Glaucoma 0.9324
RD 0.8745
Marco Avg 0.9020
B
Multi-class Precision-Recall Curves
Precision
Recall
AUC
HM 0.8288
Cataract 0.8746
MD 0.7497
CN 0.7268
Glaucoma 0.8028
RD 0.6958
Marco Avg 0.7798
C
Confusion Matrix Heat Map
HM 501 160 17 8 1 6
Cataract 121 788 8 8 0 4
MD 44 84 189 1 0 3
CN 30 67 1 139 0 1
Glau--coma 15 17 0 2 101 0
RD 28 54 4 7 0 118
HM Cataract MD CN Glau--coma RD
D
Multi-class ROC Curve
True Positive Rate
False Positive Rate
AUC
HM 0.9064
Cataract 0.9164
MD 0.8988
CN 0.8752
Glaucoma 0.9351
RD 0.8743
Marco Avg 0.9010
E
Multi-class Precision-Recall Curves
Precision
Recall
AUC
HM 0.8438
Cataract 0.8766
MD 0.7511
CN 0.7010
Glaucoma 0.8257
RD 0.7063
Marco Avg 0.7841
F
Confusion Matrix Heat Map
HM 504 143 14 8 1 6
Cataract 90 708 14 9 0 7
MD 47 69 176 0 0 1
CN 35 60 3 131 0 1
Glau--coma 5 15 1 2 86 0
RD 24 36 6 3 0 110
HM Cataract MD CN Glau--coma RD
G
AUC
Sensitivity
Specificity
PPV
NPV
Accuracy
F1
H
XGB RF Adaboost DT GB KNN LGBM LR MLP NB Final RF


A
Calibration curves
HM
HM associated with cataract
HM associated with MD
HM associated with CN
HM associated with glaucoma
HM associated with RD
Perfectly calibrated
Fraction of Positives
Mean Predicted Probability
0.0
0.2
0.4
0.6
0.8
1.0

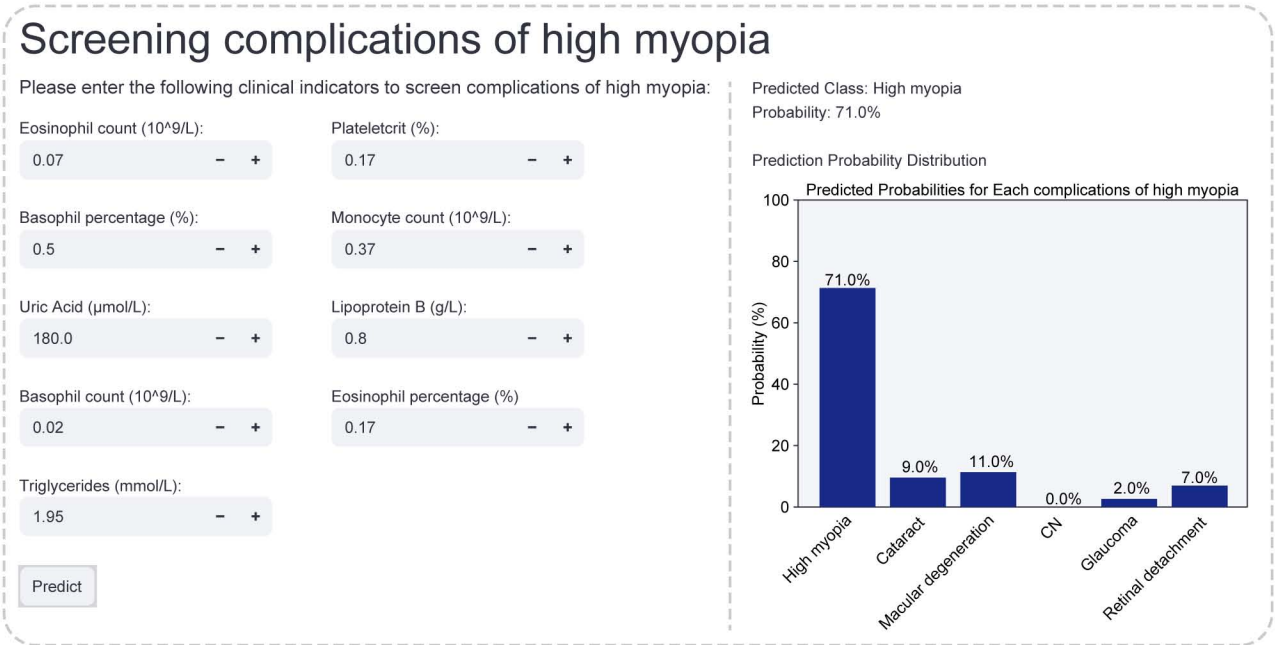

B
Calibration curves
HM
HM associated with cataract
HM associated with MD
HM associated with CN
HM associated with glaucoma
HM associated with RD
Perfectly calibrated
Fraction of Positives
Mean Predicted Probability
0.0
0.2
0.4
0.6
0.8
1.0

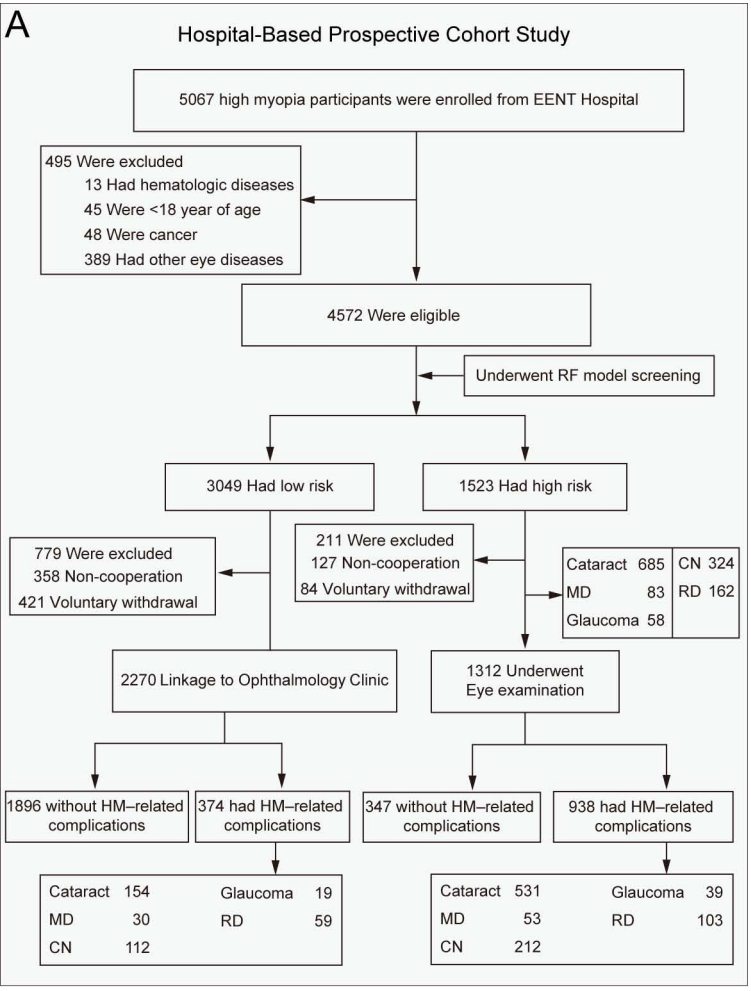

C
Calibration curves
HM
HM associated with cataract
HM associated with MD
HM associated with CN
HM associated with glaucoma
HM associated with RD
Perfectly calibrated
Fraction of Positives
Mean Predicted Probability
0.0
0.2
0.4
0.6
0.8
1.0

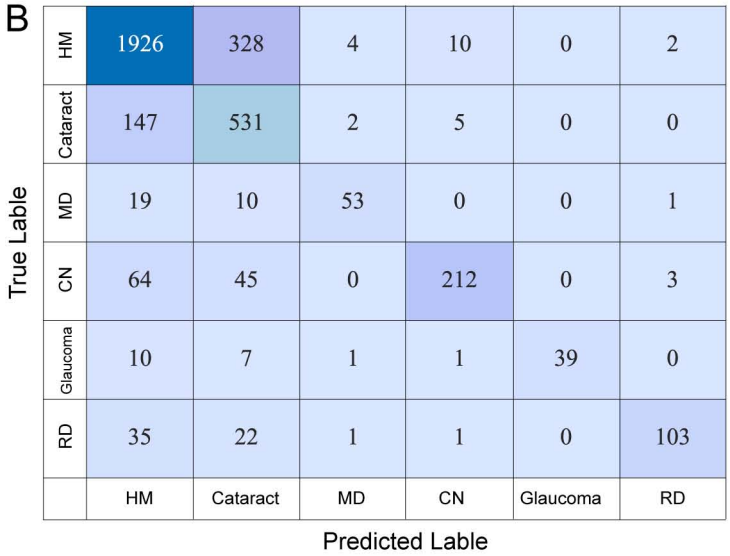
A

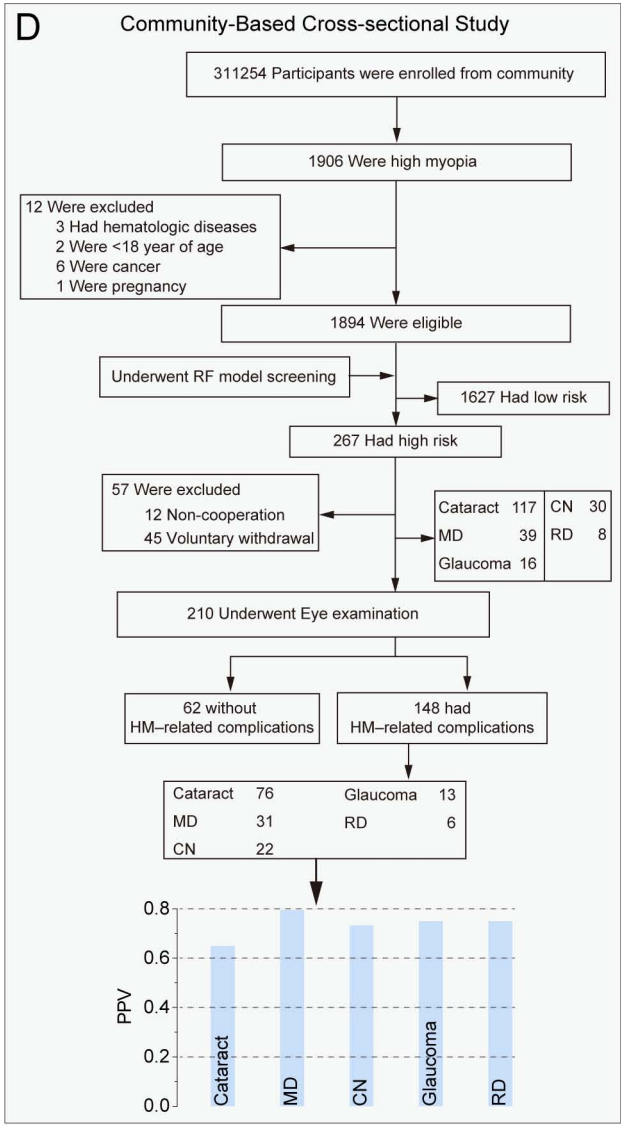

High myopia
UA
EOS%
TG
EOS
LPB
M
PCT
BASO%
BASO
SHAP value (impact on model output)
-1.5 -1.0 -0.5 0.0 0.5 1.0 1.5 2.0
High
Low
Cataract
LPB
TG
M
UA
PCT
EOS%
BASO%
EOS
BASO
SHAP value (impact on model output)
-1.5 -1.0 -0.5 0.0 0.5 1.0 1.5 2.0
High
Low
Macular degeneration
TG
LPB
EOS%
M
PCT
UA
BASO%
EOS
BASO
SHAP value (impact on model output)
-2.0 -1.5 -1.0 -0.5 0.0 0.5 1.0 1.5 2.0
High
Low
CN
UA
TG
LPB
M
BASO%
PCT
EOS%
BASO
EOS
SHAP value (impact on model output)
-2.5 -2.0 -1.5 -1.0 -0.5 0.0 0.5 1.0 1.5
High
Low
Glaucoma
EOS%
M
UA
TG
LPB
EOS
PCT
BASO%
BASO
SHAP value (impact on model output)
-1.5 -1.0 -0.5 0.0 0.5 1.0 1.5 2.0 2.5
High
Low
Retinal detachment
LPB
TG
PCT
UA
M
BASO
BASO%
EOS
EOS%
SHAP value (impact on model output)
-1.0 -0.5 0.0 0.5 1.0 1.5 2.0
High
Low

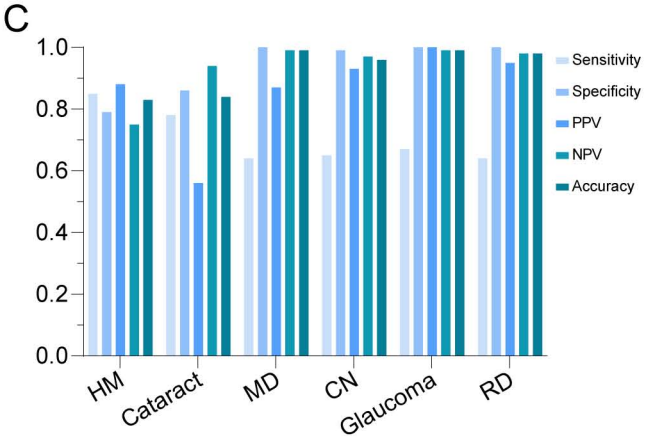
B


Screening complications of high myopia
Please enter the following clinical indicators to screen complications of high myopia:
Eosinophil count (10^9/L):
0.07
Plateletcrit (%):
0.17
Basophil percentage (%):
0.5
Monocyte count (10^9/L):
0.37
Uric Acid (μmol/L):
180.0
Lipoprotein B (g/L):
0.8
Basophil count (10^9/L):
0.02
Eosinophil percentage (%)
0.17
Triglycerides (mmol/L):
1.95
Predict
Predicted Class: High myopia
Probability: 71.0%
Prediction Probability Distribution
Predicted Probabilities for Each complications of high myopia
Probability (%)
100
80
60
40
20
0
71.0%
9.0%
11.0%
0.0%
2.0%
7.0%
High myopia
Cataract
Macular degeneration
CN
Glaucoma
Retinal detachment


A
Hospital-Based Prospective Cohort Study
5067 high myopia participants were enrolled from EENT Hospital
495 Were excluded
13 Had hematologic diseases
45 Were <18 year of age
48 Were cancer
389 Had other eye diseases
4572 Were eligible
Underwent RF model screening
3049 Had low risk
1523 Had high risk
779 Were excluded
358 Non-cooperation
421 Voluntary withdrawal
211 Were excluded
127 Non-cooperation
84 Voluntary withdrawal
Cataract 685 CN 324
MD 83 RD 162
Glaucoma 58
2270 Linkage to Ophthalmology Clinic
1312 Underwent Eye examination
1896 without HM–related complications
374 had HM–related complications
347 without HM–related complications
938 had HM–related complications
Cataract 154 Glaucoma 19
MD 30 RD 59
CN 112
Cataract 531 Glaucoma 39
MD 53 RD 103
CN 212


B
True Lable
HM 1926 328 4 10 0 2
Cataract 147 531 2 5 0 0
MD 19 10 53 0 0 1
CN 64 45 0 212 0 3
Glaucoma 10 7 1 1 39 0
RD 35 22 1 1 0 103
HM Cataract MD CN Glaucoma RD
Predicted Lable


D
Community-Based Cross-sectional Study
311254 Participants were enrolled from community
1906 Were high myopia
12 Were excluded
3 Had hematologic diseases
2 Were <18 year of age
6 Were cancer
1 Were pregnancy
1894 Were eligible
Underwent RF model screening
1627 Had low risk
267 Had high risk
57 Were excluded
12 Non-cooperation
45 Voluntary withdrawal
Cataract 117 CN 30
MD 39 RD 8
Glaucoma 16
210 Underwent Eye examination
62 without HM–related complications
148 had HM–related complications
Cataract 76 Glaucoma 13
MD 31 RD 6
CN 22
PPV
0.0
0.2
0.4
0.6
0.8
Cataract
MD
CN
Glaucoma
RD


C
1.0
0.8
0.6
0.4
0.2
0.0
Sensitivity
Specificity
PPV
NPV
Accuracy
HM
Cataract
MD
CN
Glaucoma
RD