Regenerative Therapy 15 (2020) 180–186

Contents lists available at ScienceDirect

Regenerative Therapy

journal homepage: http://www.elsevier.com/locate/reth

Original Article

Comparison of machine learning tools for the prediction of AMD based on genetic, age, and diabetes-related variables in the Chinese population

Shaofeng Hao [a, 1], Junye Bai [b, 1], Huimin Liu [c, d, 1], Lijun Wang [c, d], Tao Liu [e], Chaobin Lin [f], Xiangguang Luo [c, d], Junhui Gao [c, d], Jiangman Zhao [c, d], Huilin Li [a, *], Hui Tang [c, d, **]

[a] Department of Ophthalmology, Heji Hospital Affiliated to Changzhi Medical College, Changzhi 046011, China
[b] Department of Ophthalmology, General Hospital of Tisco, Sixth Hospital of Shanxi Medical University, Taiyuan 030008, China
[c] Shanghai Biotecan Pharmaceuticals Co., Ltd., Shanghai 201204, China
[d] Shanghai Zhangjiang Institute of Medical Innovation, Shanghai 201204, China
[e] Department of Ophthalmology, The Aviation Hanzhong 3201 Hospital, Xi'an Jiao Tong University, Hanzhong 723000, China
[f] Department of Ophthalmology, The First Hospital of Quanzhou Affiliated to Fujian Medical University, Quanzhou 362000, China

ARTICLE INFO

Article history:
Received 16 June 2020
Received in revised form
1 September 2020
Accepted 9 September 2020

Keywords:
AMD
SNPs
Age
Diabetes
Machine learning tools

ABSTRACT

Introduction: Age-related macular degeneration (AMD) is the main cause of visual impairment and the most important cause of blindness in older people. However, there is currently no effective treatment for this disease, so it is necessary to establish a risk model to predict AMD development.
Methods: This study included a total of 202 subjects, comprising 82 AMD patients and 120 control subjects. Sixty-six single-nucleotide polymorphisms (SNPs) were identified using the MassArray assay. Considering 14 independent clinical variables as well as SNPs, four predictive models were established in the training set and evaluated by the confusion matrix, area under the receiver operating characteristic (ROC) curve (AUROC). The difference distributions of the 14 independent clinical features between the AMD and control groups were tested using the chi-squared test. Age and diabetes were adjusted using logistic regression analysis and the "genomic-control" method was used for multiple testing correction.
Results: Three SNPs (rs10490924, OR = 1.686, genomic-control corrected p-value (GC) = 0.030; rs2338104, OR = 1.794, GC = 0.025 and rs1864163, OR = 2.125, GC = 0.038) were significant risk factors for AMD development. In the training set, four models obtained AUROC values above 0.72.
Conclusions: We believe machine learning tools will be useful for the early prediction of AMD and for the development of relevant intervention strategies.

1. Introduction

Age-related macular degeneration (AMD) is a progressive and degenerative eye disease that is a major cause of blindness in elderly individuals [1]. Early AMD is characterized by the presence of medium-size drusen or retinal pigment abnormalities without overt functional loss in the clinic. Late-stage AMD can be divided into two forms: geographic atrophy (dry AMD) or neovascular AMD (wet AMD or nAMD) [2,3]. The prevalence of dry AMD is higher than that of wet AMD, accounting for 80–85% of all cases; however, 90% of all clinical cases with loss of sight are wet AMD, and rapid visual acuity loss may occur within several months [4]. According to the diagnosis by indocyanine green angiography, wet AMD is further divided into the typical choroidal neovascular (CNV) subtype and the polypoidal choroidal vasculopathy (PCV) subtype [5].

AMD accounts for 8.7% of all blindness worldwide, particularly in people older than 60 years [6]. Its prevalence is likely to increase

* Corresponding author. Department of Ophthalmology, Heji Hospital Affiliated to Changzhi Medical College, No. 271, Taihang East Street, Luzhou District, Changzhi 046011, China.
** Corresponding author. Shanghai Biotecan Pharmaceuticals Co., Ltd., No. 180 Zhangheng Road, Shanghai 201204, China.
E-mail addresses: 88935560@qq.com (H. Li), tang11_23@126.com (H. Tang).
Peer review under responsibility of the Japanese Society for Regenerative Medicine.
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.reth.2020.09.001

as a consequence of exponential population aging. In China, the incidences of early AMD and advanced AMD in the elderly population are 4.7%–9.2% and 0.2%–1.9%, respectively [7,8]. The results of a meta-analysis show that the global incidences of early, late and any AMD are 8.01%, 0.37%, and 8.69%, respectively. There was a higher prevalence of early and any AMD in Europeans than in Asians (early: 11.2% *vs* 6.8%; any: 12.3% *vs* 7.4%); early, late, and any AMD were more prevalent in Europeans than in Africans (early: 11.2% *vs* 7.1%; late: 0.5% *vs* 0.3%; any: 12.3% *vs* 7.5%), while there was no difference in prevalence between Asians and Africans [6]. It is predicted that by 2020 and 2040, the number of global AMD patients will reach 196 million and 288 million, respectively [6]. Due to its socioeconomic impact together with the growing incidence and severity of the disease, AMD has become a major challenge in ophthalmology in recent years. Therefore, it is necessary to develop a reliable early warning method for AMD that could lead to the early intervention and treatment of AMD.

AMD is a clinically heterogeneous and genetically complex disease, with multiple environmental and genetic risk factors involved [9,10]. Epidemiological research suggests that AMD is a complex disease caused by the actions and interactions of multiple genes and environmental factors, such as age, light exposure, body mass index (BMI), and cigarette smoking [11,12]. Older age is a major risk factor for AMD, with more than 10% of people older than 80 years being affected by late AMD, while female sex has been inconsistently reported as a risk factor as well [13]. Cigarette smoking has been reported as a significant environmental risk factor and shown to modify the effect of ARMS2 A69S [14,15] but not all [16,17]. Other epidemiological studies have linked not only cigarette smoking and light exposure but also alcohol consumption, diet, drugs, lipids and hypertension to the risk of AMD [18–23]. Previous cataract surgery was reported as a risk factor in this disease [24]. Lim et al. also reported that major risk factors include cigarette smoking, nutritional factors, cardiovascular diseases, and genetic markers, including genes regulating complement, lipid, angiogenesis, and extracellular matrix pathways in AMD development [25]. Familial aggregation [26,27] and twin studies [28] have suggested that genetic variation may also play an important role in the disease, in which genetic factors account for a high proportion of the risk of the disease, up to 45–70% [29]. Several genetic loci have been associated with AMD, including two major loci in the complement factor H (*CFH*) gene on 1q32 and the *ARMS2/HTRA1* locus on the 10q26 gene cluster [30,31]. On the basis of large genome-wide association studies, high-density lipoprotein cholesterol (HDL-C) pathway genes have been implicated, including *LIPC* and *CETP*, and possibly *ABCA1* and *LPL* [32,33].

In this study, the Chinese population was used as the research object. Based on the published SNPs of AMD-related genes, 66 SNPs of 44 genes and 10 intergenic regions related to AMD were selected [6,34–49], and 14 clinical features were collected in case–control studies. This study aimed to analyze the relationship between SNPs and AMD, to identify SNPs related to AMD in the Chinese population and to construct a genetic structure network of SNPs susceptible to AMD disease in the Chinese population. The results regarding SNPs in AMD-related genes will help elucidate the molecular genetic mechanisms of AMD in the Chinese population. More importantly, machine learning tools combining SNPs and clinical features were used to predict the early diagnosis of clinical AMD, which is of great significance in reducing the incidence of AMD and improving the treatment rate of AMD.

2. Methods and materials

2.1. Patients and data collection

Our study recruited 82 patients with AMD and 120 healthy controls from the following 4 subcenters from January to December 2018: Heji Hospital Affiliated to Changzhi Medical College (68 control samples and 36 AMD patients), General Hospital of Tisco (50 control samples and 37 AMD patients), The Aviation Hanzhong 3201 Hospital (6 AMD patients) and The First Hospital of Quanzhou (2 control samples and 3 AMD patients). AMD was diagnosed according to the standard of the Clinical Age-Related Maculopathy Staging (CARMS) system [50]. Controls had no family relationship with the AMD patients, and they all underwent ocular examinations. Ethics approval and written informed consent were obtained from all participants in the study.

2.2. Obtaining relevant clinical data

According to the questionnaire, a total of 14 clinical risk factors were assessed and are listed in Table 1, including age, sex, BMI, hypertension, hyperlipidemia, diabetes, renal dysfunction, atherosclerosis, history of ophthalmic surgery, family history of AMD, long-term outdoor work, vegetarian status, smoking status and drinking status. Smoking or drinking status was defined as follows: no means never smoking/drinking; yes means including ex-smokers/drinkers and current smokers/drinkers. BMI was measured in kg/m^2 (underweight < 18.5, normal $= 18.5–23.9$, overweight $= 24–27.9$, obese ≥ 28).

2.3. DNA extraction and genotyping

Sixty-six SNPs were selected from 44genes and 10 intergenic regions that were previously reported to be associated with AMD [6,34–49] and some related studies on the susceptible loci of AMD in recent years were listed in Supplementary Table S1. A 4-ml peripheral blood sample was obtained from each participant for DNA analysis. Genomic DNA was extracted from whole blood using the Gold Mag-Mini Whole Blood Genomic DNA Purification Kit (GoldMag Co., Ltd., Xi'an City, China). The DNA concentration was measured using a NanoDrop 2000 spectrophotometer (Thermo Scientific, Fitchburg, WI, USA). We used the website https://agenacx.com to design multiplex primers for each SNP: forward PCR primer, reverse PCR primer, and UEP primer. The primers of the 66 SNPs are shown in Supplementary Table S2. The SNPs were genotyped with an Agena BioscienceTM MassArray ®Analyzer with a 384-well configuration (Agena, California, USA) using the standard protocol recommended by the manufacturer. Data management and analysis were performed using Typer Analyzer 4.0 software (Agena).

2.4. Model construction

A total of 120 AMD patients and 82 control subjects were included in the training set, and the odd ratio (OR) values of all SNPs were calculated using the software package PLINK (PLINK version 1.07) [51]. Three SNPs were identified to be significantly associated with AMD risk (Supplementary Table S3). Then, 4 models were established to predict AMD development, including logistic regression (LR), AdaBoost, random forest (RF) and XGBoost

Table 1
Demographics of AMD patients and control subjects.

Variables	AMD (82)	Control (120)	p value
Age (years)			<0.0001***
40–60	14	61	
61–80	55	53	
>80	13	6	
Sex			0.886
Male	40	61	
Female	42	59	
BMI (kg/m^2)			0.525
<18.5	5	3	
18.5–23.9	38	56	
24.0–27.9	33	48	
≥28	6	13	
AMD family history			0.880
Yes	2	2	
No	80	118	
Hypertension			0.978
Yes	42	60	
No	40	60	
Diabetes			<0.001**
Yes	25	66	
No	57	54	
Hyperlipidemia			0.927
Yes	27	40	
No	55	80	
Renal dysfunction			0.517
Yes	13	14	
No	69	106	
Long-term outdoor work			0.825
Yes	19	25	
No	63	95	
Vegetarian			0.953
Yes	19	27	
No	63	93	
Smoking			0.984
Yes	32	47	
No	50	73	
Drinking			0.304
Yes	25	45	
No	57	75	
Atherosclerosis			0.674
Yes	34	45	
No	48	75	
History of ophthalmic surgery			0.545
Yes	18	21	
No	64	99	

BMI, body mass index; **p < 0.001, ***p < 0.0001, chi square test.

models, considering clinical features and SNPs. Cross validation is a common solution when the available datasets are limited and k-fold is a common cross validation approach [52]. Therefore, k-fold (k = 4) cross validation was used in the construction of the four predictive models, which were compared with the use of the confusion matrix, area under the receiver operating characteristic (ROC) curve (AUROC). The AUROC is one of the most used metrics for evaluating binary classifiers and plots sensitivity against 1-specificity.

In the present study, four models were selected based on several currently and frequently adopted predictive model types. LR is most frequently used to predict the occurrence of an event in clinical research, such as the prediction of heart failure in patients [53]. The AdaBoost model is considered a generalized additive model and had good predictive performance in diabetes classification [54]. XGBoost has demonstrated excellent performance in clinical research due to its high efficiency and impressive accuracy [55,56]. The RF algorithm gradually converges with an increase in the number of trees and is more accurate because of the injected randomness. This algorithm has been used for predicting overall survival in breast cancer [57].

2.5. Statistical analysis

For all SNPs, the ORs and 95% confidence intervals (CIs) of the minor alleles of AMD patients and healthy controls were assessed adjusting for age and diabetes by logistic regression analysis assuming an additive genetic model using the PLINK package. Four models (LR, AdaBoost, RF and XGBoost) for prediction of AMD risk were constructed combining 3 SNPs with age and diabetes and were evaluated by the Python (version 3.7.0). The diagnostic values of 4 models were assessed by ROC analysis. The chi-squared test was used to assess the differences between groups and statistical calculations were performed using SPSS version 19.0. The "genomic-control" method was used for multiple testing correction at a level of $p < 0.05$.

3. Results

3.1. Characteristics of the subjects

Fourteen clinical characteristics of the patients with AMD and control patients are shown in Table 1. There were no statistically significant differences between the AMD and control subjects in terms of sex, BMI, hypertension, hyperlipidemia, renal dysfunction, atherosclerosis, history of ophthalmic surgery, AMD family history, long-term outdoor work, vegetarian status, smoking status or drinking status. Only age ($p < 0.001$) and diabetes ($p < 0.001$) were significantly different between the AMD and control groups.

3.2. Associations between the SNPs and AMD

We used the minor allele (A1) of each SNP in the additive genetic model in logistic regression analysis adjusting for age and diabetes and used the multiple-testing method "genomic-control" for multiple testing correction by PLINK software, which was estimated as previously described [58]. Multiple-testing was performed in PLINK to make the statistics more rigorous and accurate. Genomic-control is now regarded as the gold standard to eliminate the potential effect of population stratification [59]. In this study, we used the "genomic-control" method to correct for possible AMD-control differences in the genetic structure of our study population. The OR values and p values (unadjusted p-value (*UNADJ*) and genomic-control corrected p-value (GC)) were obtained by PLINK software (Supplementary Table S3). The results showed that 3 SNPs tended to be significantly associated with AMD: rs10490924, rs2338104 and rs1864163. These SNPs were risk factors in AMD development and are shown in the forest plots in Fig. 1.

3.3. Three risk SNPs in AMD

Among the 3 risk SNPs, the nonsynonymous mutation (A69S) SNP rs10490924 was located on chromosome 10q26 in exon 1 of the age-related maculopathy susceptibility 2 (LOC387715/ARMS2) gene, which was reported as a strong genetic risk factor for AMD [60]. Tong et al. performed a Human Genome Epidemiology (HuGE) systematic review and meta-analysis and reported that the *LOC387715/ARMS2* rs10490924 G→T polymorphism plays an important role in AMD diseases [61]. In the present study, we found that the T allele of rs10490924 (OR = 1.686, 95%CI = 1.079–2.635, GC = 0.030) was also a risk factor for AMD development. Rs2338104 is near the KCTD10/MVK gene, which is located on chromosome 12 and encodes enzymes involved in cholesterol synthesis and degradation, and has been reported to be associated with HDL-C and the risk of developing coronary artery disease [62,63]. Restrepo et al. used the Population Architecture using

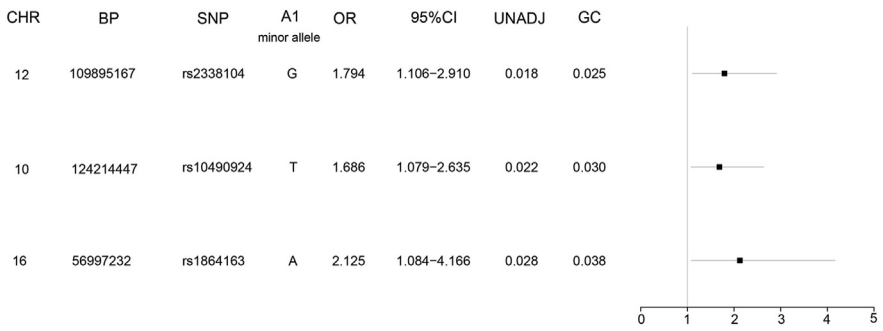

Fig. 1. Forest plots. Three SNPs associated with AMD (odds ratios (ORs) and 95% confidence intervals (CIs)). ORs are denoted by black boxes, and 95% CIs are denoted by the corresponding gray lines.

Genomics and Epidemiology (PAGE) study and reported that the G allele of rs2338104 was associated with AMD in African Americans and Mexican Americans [49]. In our current study, we found that the G allele of rs2338104 (OR = 1.794, 95% CI = 1.106–2.910, GC = 0.025) was also a risk factor for AMD development in the Chinese population. Rs1864163 near cholesteryl ester transfer protein (CETP) is a hydrophobic glycoprotein that has an established role in transporting cholesterol from peripheral tissues to the liver for elimination by exchanging the triglycerides of very low-density lipoprotein (LDL) and LDL against the cholesteryl esters of HDL [64]. Naseri et al. reported that the CETP of rs1864163 was associated with HDL-C [65]. A previous study reported lipid levels as a major risk factor for AMD and found that higher levels of HDL increased the risk of incident AMD [66]. In the present study, we found that the A allele of rs1864163 (OR = 2.125, 95% CI = 1.084–4.166, GC = 0.038) was a risk factor for AMD development.

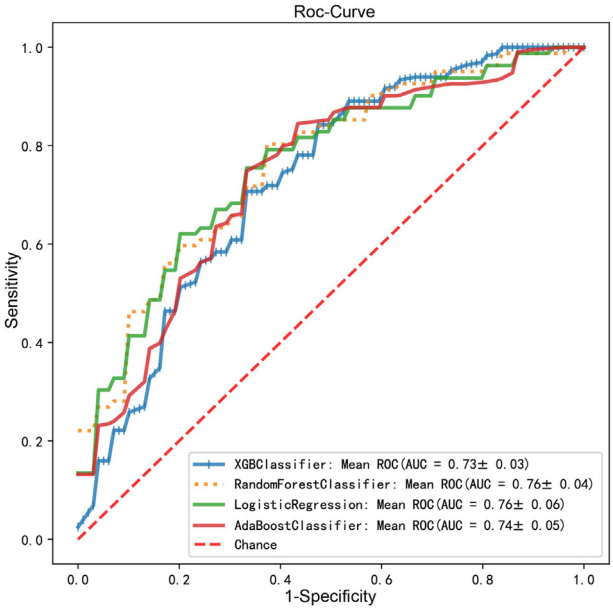

Fig. 2. Evaluation of the predictive models. The figure shows the average ROC curves of the 4 models in the training set. The mean AUC values with standard deviations of the different prediction models are shown in the box.

3.4. Comparison of the four prediction models in the training set

For each model, the optimal parameter values were chosen with a grid search approach. The model with the optimal parameters was chosen as the final model to be compared with the others. The predictive abilities of the prediction models were assessed using the AUROC. The average AUROC is shown in Fig. 2. All models had AUROC values above 0.72. As a useful tool, the ROC curve has been widely used to validate the performance of landslide susceptibility models. It is generally considered that if the AUC of the model is greater than 0.70, the model has high accuracy [67]. K-fold cross validation (k = 4) was used in the four models in Fig. 3.

4. Discussion

AMD is the most common cause of irreversible blindness worldwide [25], and there is currently no effective treatment for this disease [41]. A systematic review and meta-analysis indicated that AMD causes a substantial global burden [6]. Therefore, it is urgent to diagnose AMD in the early stage.

This study included a total of 202 subjects, which comprised 82 AMD patients and 120 controls. To identify the risk of AMD in order to achieve earlier diagnosis in suspected AMD patients, 14 independent clinical variables related to AMD were included in the study. Only two clinical features, namely, age and diabetes, were significantly different between the AMD and control groups, while other risk factors, such as smoking, AMD family history or ophthalmic surgery, were not significant between the two groups, probably due to the unbalanced small sample size. Smokers aged >40 years are two to four times more likely to develop AMD than nonsmokers of the same age [68]. A systematic review including 18 prospective and cross-sectional studies and six case–control studies involving 113,780 individuals identified age, smoking, previous cataract surgery, and a family history of AMD as strong risk factors for AMD [69]. AMD is most frequently found in Caucasians, followed by Hispanics and Asians with the lowest rate reported in African Americans [70]. Siblings of an affected individual have a threefold to six fold higher risk than those of the general population [71]. A total of 66 SNPs were identified, and 3 SNPs were significantly associated with AMD development, as determined using PLINK software. Our results showed that 3 SNPs (rs10490924, rs2338104 and rs1864163) were risk factors in AMD development, consistent with the findings of previous studies.

The present study aims to compare the performance of four machine learning techniques (MLTs) in the prediction of patients with AMD using data from the four subcenters. MLT applications have been reported in cardiology, especially for developing prediction models using both supervised and unsupervised methods

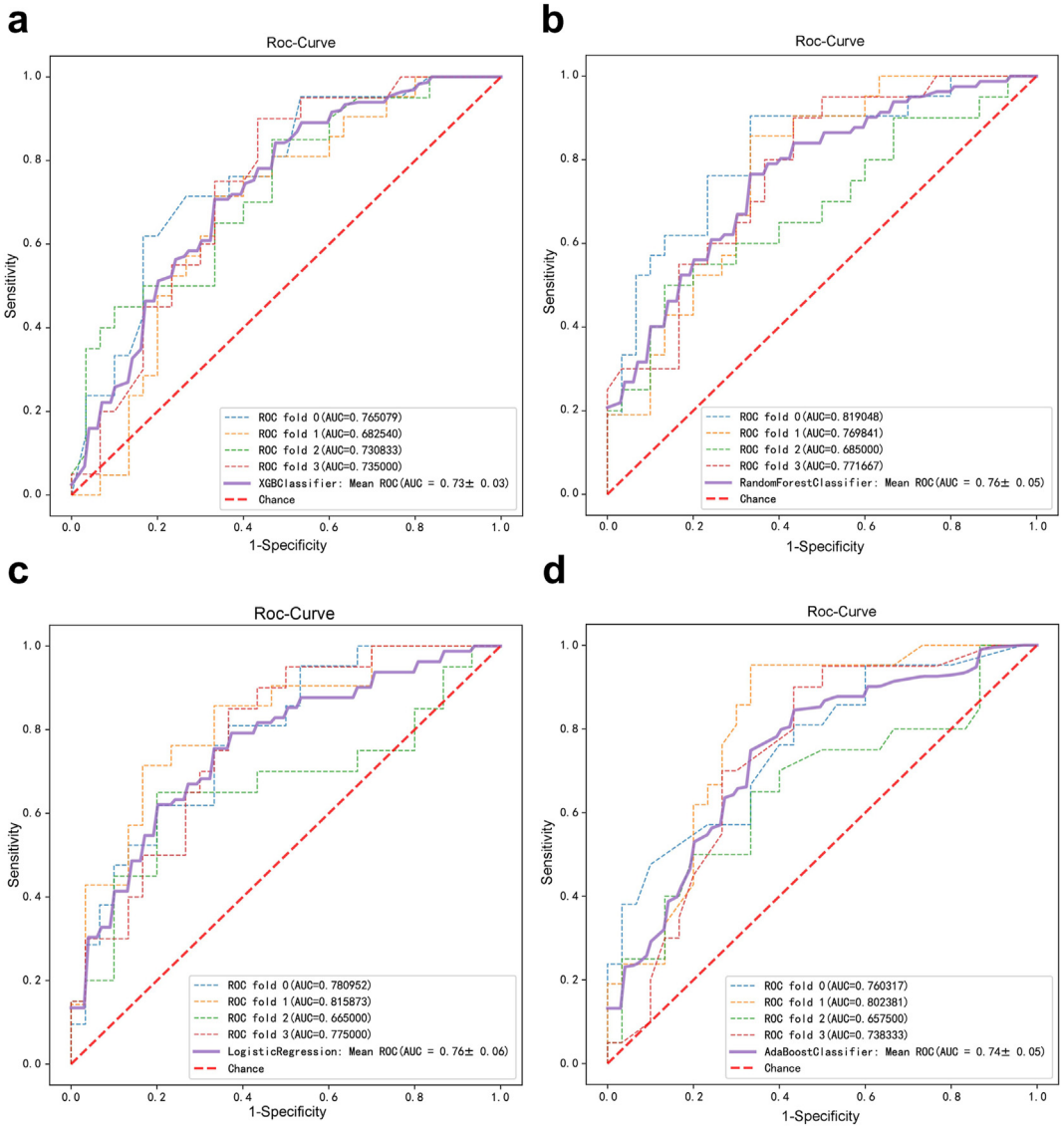

Fig. 3. K-fold (k = 4) cross validation was used in the XGBoost, RF, LR and AdaBoost models. a-d, k = 4 was used in the XGBoost, RF, LR and AdaBoost models.

[72]. In recent years, MLTs have also been increasingly used in the field of heart failure research [53] and chronic kidney disease progression [56]. The prediction abilities of the four machine learning models of LR, AdaBoost, XGBoost and RF were assessed. Two clinical features, namely, age and diabetes, and 3 SNPs were used as inputs for prediction modeling in the training set. To make full use of the data, the four models were constructed using the aforementioned dataset with 4-fold cross validation. All models obtained AUROCs above 0.72.

There were some limitations in this study. First, the sample size used was relatively small, and the total sample of patients with AMD and healthy controls was unbalanced. Second, the samples from four centers were seriously unbalanced; only six AMD samples were collected from The Aviation Hanzhong 3201 Hospital, and two control samples and three AMD samples were collected from the Quanzhou First Hospital. Third, we only selected 2 clinical features with significant differences between the two groups for inclusion in the prediction models, while other risk factors, such as smoking status, AMD family history, BMI, drinking status, hypertension, and hyperlipidemia, were not considered. Fourth, we only used four models in the training set but not in the validation set. It is urgent for us to collect samples for validation in the near future.

5. Conclusions

In summary, we constructed and evaluated AMD prediction models integrating 3 SNPs and 2 clinical factors. The four models showed AUROCs above 0.72 in the training set. Machine learning tools have the potential to aid in the early diagnosis and treatment of patients with AMD. There is still a way to go before the models can be applied in the clinic for AMD prediction, and they should be validated in a larger cohort.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Major Projects of Special Development Funds in Zhangjiang National Independent Innovation Demonstration Zone, Shanghai (ZJ2017-ZD-012).

Abbreviations

AMD	age-related macular degeneration
SNPs	single-nucleotide polymorphisms
OR	odd ratio
CI	confidence interval
CNV	choroidal neovascular
PCV	polypoidal choroidal vasculopathy
BMI	body mass index
CFH	complement factor H
LR	logistic regression
RF	random forest
AUROC	area under the receiver operating characteristic curve
MLTs	machine learning techniques
HDL-C	high-density lipoprotein cholesterol
LDL	low-density lipoprotein
CARMS	clinical age-related maculopathy staging
GC	genomic-control corrected p-value
UNADJ	unadjusted p-value

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.reth.2020.09.001.

References

[1] Hernandez-Zimbron LF, Zamora-Alvarado R, Ochoa-De la Paz L, Velez-Montoya R, Zenteno E, Gulias-Canizo R, et al. Age-related macular degeneration: new paradigms for treatment and management of amd. Oxid Med Cell Longev 2018;2018:8374647.
[2] Sayen A, Hubert I, Berrod JP. [age related macular degeneration]. Rev Prat 2011;61:159–64.
[3] Jager RD, Mieler WF, Miller JW. Age-related macular degeneration. N Engl J Med 2008;358:2606–17.
[4] Ayoub T, Patel N. Age-related macular degeneration. J R Soc Med 2009;102:56–61.
[5] Lim TH, Laude A, Tan CS. Polypoidal choroidal vasculopathy: an angiographic discussion. Eye 2010;24:483–90.
[6] Wong WL, Su X, Li X, Cheung CM, Klein R, Cheng CY, et al. Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and meta-analysis. Lancet Glob Health 2014;2:106–16.
[7] Chen SJ, Cheng CY, Peng KL, Li AF, Hsu WM, Liu JH, et al. Prevalence and associated risk factors of age-related macular degeneration in an elderly Chinese population in taiwan: the shihpai eye study. Invest Ophthalmol Vis Sci 2008;49:3126–33.
[8] Yang K, Liang YB, Gao LQ, Peng Y, Shen R, Duan XR, et al. Prevalence of age-related macular degeneration in a rural Chinese population: the handan eye study. Ophthalmology 2011;118:1395–401.
[9] Tuo J, Bojanowski CM, Chan CC. Genetic factors of age-related macular degeneration. Prog Retin Eye Res 2004;23:229–49.
[10] Scholl HP, Fleckenstein M, Charbel Issa P, Keilhauer C, Holz FG, Weber BH. An update on the genetics of age-related macular degeneration. Mol Vis 2007;13:196–205.
[11] Age-Related Eye Disease Study Research G. Risk factors associated with age-related macular degeneration. A case-control study in the age-related eye disease study: age-related eye disease study report number 3. Ophthalmology 2000;107:2224–32.
[12] Hyman L, Neborsky R. Risk factors for age-related macular degeneration: an update. Curr Opin Ophthalmol 2002;13:171–5.
[13] Smith W, Assink J, Klein R, Mitchell P, Klaver CC, Klein BE, et al. Risk factors for age-related macular degeneration: pooled findings from three continents. Ophthalmology 2001;108:697–704.
[14] Schmidt S, Hauser MA, Scott WK, Postel EA, Agarwal A, Gallins P, et al. Cigarette smoking strongly modifies the association of loc387715 and age-related macular degeneration. Am J Hum Genet 2006;78:852–64.
[15] Neuner B, Wellmann J, Dasch B, Dietzel M, Farwick A, Stoll M, et al. Loc387715, smoking and their prognostic impact on visual functional status in age-related macular degeneration-the muenster aging and retina study (mars) cohort. Ophthalmic Epidemiol 2008;15:148–54.
[16] Conley YP, Jakobsdottir J, Mah T, Weeks DE, Klein R, Kuller L, et al. Cfh, elovl4, plekha1 and loc387715 genes and susceptibility to age-related maculopathy: Areds and chs cohorts and meta-analyses. Hum Mol Genet 2006;15:3206–18.
[17] Seitsonen SP, Onkamo P, Peng G, Xiong M, Tommila PV, Ranta PH, et al. Multifactor effects and evidence of potential interaction between complement factor h y402h and loc387715 a69s in age-related macular degeneration. PloS One 2008;3:e3833.
[18] Klein R, Peto T, Bird A, Vannewkirk MR. The epidemiology of age-related macular degeneration. Am J Ophthalmol 2004;137:486–95.
[19] Rakic JM. [multifactorial influences on age-related macular degeneration]. Bull Soc Belge Ophtalmol 2006:9–11.
[20] Francis PJ, George S, Schultz DW, Rosner B, Hamon S, Ott J, et al. The loc387715 gene, smoking, body mass index, environmental associations with advanced age-related macular degeneration. Hum Hered 2007;63:212–8.
[21] Seddon JM, George S, Rosner B. Cigarette smoking, fish consumption, omega-3 fatty acid intake, and associations with age-related macular degeneration: the us twin study of age-related macular degeneration. Arch Ophthalmol 2006;124:995–1001.
[22] Chong EW, Kreis AJ, Wong TY, Simpson JA, Guymer RH. Alcohol consumption and the risk of age-related macular degeneration: a systematic review and meta-analysis. Am J Ophthalmol 2008;145:707–15.
[23] Fraser-Bell S, Wu J, Klein R, Azen SP, Varma R. Smoking, alcohol intake, estrogen use, and age-related macular degeneration in latinos: the los angeles latino eye study. Am J Ophthalmol 2006;141:79–87.
[24] Cugati S, Mitchell P, Rochtchina E, Tan AG, Smith W, Wang JJ. Cataract surgery and the 10-year incidence of age-related maculopathy: the blue mountains eye study. Ophthalmology 2006;113:2020–5.
[25] Lim LS, Mitchell P, Seddon JM, Holz FG, Wong TY. Age-related macular degeneration. Lancet 2012;379:1728–38.
[26] Klaver CC, Wolfs RC, Assink JJ, van Duijn CM, Hofman A, de Jong PT. Genetic risk of age-related maculopathy. Population-based familial aggregation study. Arch Ophthalmol 1998;116:1646–51.
[27] Luo L, Harmon J, Yang X, Chen H, Patel S, Mineau G, et al. Familial aggregation of age-related macular degeneration in the Utah population. Vis Res 2008;48:494–500.
[28] Grizzard SW, Arnett D, Haag SL. Twin study of age-related macular degeneration. Ophthalmic Epidemiol 2003;10:315–22.
[29] Seddon JM, Cote J, Page WF, Aggen SH, Neale MC. The us twin study of age-related macular degeneration: relative roles of genetic and environmental influences. Arch Ophthalmol 2005;123:321–7.
[30] Haines JL, Hauser MA, Schmidt S, Scott WK, Olson LM, Gallins P, et al. Complement factor h variant increases the risk of age-related macular degeneration. Science 2005;308:419–21.
[31] Jakobsdottir J, Conley YP, Weeks DE, Mah TS, Ferrell RE, Gorin MB. Susceptibility genes for age-related maculopathy on chromosome 10q26. Am J Hum Genet 2005;77:389–407.
[32] Reynolds R, Rosner B, Seddon JM. Serum lipid biomarkers and hepatic lipase gene associations with age-related macular degeneration. Ophthalmology 2010;117:1989–95.
[33] Neale BM, Fagerness J, Reynolds R, Sobrin L, Parker M, Raychaudhuri S, et al. Genome-wide association study of advanced age-related macular degeneration identifies a role of the hepatic lipase gene (lipc). Proc Natl Acad Sci U S A 2010;107:7395–400.
[34] Edwards AO, Ritter 3rd R, Abel KJ, Manning A, Panhuysen C, Farrer LA. Complement factor h polymorphism and age-related macular degeneration. Science 2005;308:421–4.
[35] Contreras AV, Zenteno JC, Fernandez-Lopez JC, Rodriguez-Corona U, Falfan-Valencia R, Sebastian L, et al. Cfh haplotypes and arms2, c2, c3, and cfb alleles show association with susceptibility to age-related macular degeneration in mexicans. Mol Vis 2014;20:105–16.
[36] Morrison MA, Magalhaes TR, Ramke J, Smith SE, Ennis S, Simpson CL, et al. Ancestry of the Timorese: age-related macular degeneration associated genotype and allele sharing among human populations from throughout the world. Front Genet 2015;6:238.
[37] Cheng CY, Yamashiro K, Chen LJ, Ahn J, Huang L, Huang L, et al. New loci and coding variants confer risk for age-related macular degeneration in east asians. Nat Commun 2015;6:6063.
[38] Holliday EG, Smith AV, Cornes BK, Buitendijk GH, Jensen RA, Sim X, et al. Insights into the genetic architecture of early stage age-related macular degeneration: a genome-wide association study meta-analysis. PloS One 2013;8:e53830.
[39] Fritsche LG, Chen W, Schu M, Yaspan BL, Yu Y, Thorleifsson G, et al. Seven new loci associated with age-related macular degeneration. Nat Genet 2013;45:433–9. 439e431-432.
[40] McKay GJ, Dasari S, Patterson CC, Chakravarthy U, Silvestri G. Complement component 3: an assessment of association with amd and analysis of gene-gene and gene-environment interactions in a northern Irish cohort. Mol Vis 2010;16:194–9.
[41] Tian J, Yu W, Qin X, Fang K, Chen Q, Hou J, et al. Association of genetic polymorphisms and age-related macular degeneration in Chinese population. Invest Ophthalmol Vis Sci 2012;53:4262–9.

[42] Lu F, Zhao P, Fan Y, Tang S, Hu J, Liu X, et al. An association study of serping1 gene and age-related macular degeneration in a han Chinese population. Mol Vis 2010;16:1–6.
[43] Sun Y, Li S, Li H, Yang F, Bai Y, Zhao M, et al. Tnfrsf10a-loc389641 rs13278062 but not rest-c4orf14-polr2b-igfbp7 rs1713985 was found associated with age-related macular degeneration in a Chinese population. Invest Ophthalmol Vis Sci 2013;54:8199–203.
[44] Cipriani V, Leung HT, Plagnol V, Bunce C, Khan JC, Shahid H, et al. Genome-wide association study of age-related macular degeneration identifies associated variants in the tnxb-fkbpl-notch4 region of chromosome 6p21.3. Hum Mol Genet 2012;21:4138–50.
[45] Wang Y, Wang M, Han Y, Zhang R, Ma L. Abca1 rs1883025 polymorphism and risk of age-related macular degeneration. Graefes Arch Clin Exp Ophthalmol 2016;254:323–32.
[46] Huang L, Li M, Ma X, Li Y, Zhang C, Sun Y, et al. Rs4711751 and rs1999930 are not associated with neovascular age-related macular degeneration or polypoidal choroidal vasculopathy in the Chinese population. Ophthalmic Res 2014;52:102–6.
[47] Fourgeux C, Dugas B, Richard F, Bjorkhem I, Acar N, Bron AM, et al. Single nucleotide polymorphism in the cholesterol-24s-hydroxylase (cyp46a1) gene and its association with cfh and loc387715 gene polymorphisms in age-related macular degeneration. Invest Ophthalmol Vis Sci 2012;53:7026–33.
[48] Ma L, Tang FY, Chu WK, Young AL, Brelen ME, Pang CP, et al. Association of toll-like receptor 3 polymorphism rs3775291 with age-related macular degeneration: a systematic review and meta-analysis. Sci Rep 2016;6:19718.
[49] Restrepo NA, Spencer KL, Goodloe R, Garrett TA, Heiss G, Buzkova P, et al. Genetic determinants of age-related macular degeneration in diverse populations from the page study. Invest Ophthalmol Vis Sci 2014;55:6839–50.
[50] Fagerness JA, Maller JB, Neale BM, Reynolds RC, Daly MJ, Seddon JM. Variation near complement factor i is associated with risk of advanced amd. Eur J Hum Genet 2009;17:100–4.
[51] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. Plink: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559–75.
[52] Vabalas A, Gowen E, Poliakoff E, Casson AJ. Machine learning algorithm validation with a limited sample size. PloS One 2019;14:e0224365.
[53] Lorenzoni G, Sabato SS, Lanera C, Bottigliengo D, Minto C, Ocagli H, et al. Comparison of machine learning techniques for prediction of hospitalization in heart failure patients. J Clin Med 2019;8.
[54] Chen P, Pan C. Diabetes classification model based on boosting algorithms. BMC Bioinf 2018;19:109.
[55] Yu B, Qiu W, Chen C, Ma A, Jiang J, Zhou H, et al. Submito-xgboost: predicting protein submitochondrial localization by fusing multiple feature information and extreme gradient boosting. Bioinformatics 2020;36:1074–81.
[56] Xiao J, Ding R, Xu X, Guan H, Feng X, Sun T, et al. Comparison and development of machine learning tools in the prediction of chronic kidney disease progression. J Transl Med 2019;17:119.
[57] Hu C, Steingrimsson JA. Personalized risk prediction in clinical oncology research: applications and practical issues using survival trees and random forests. J Biopharm Stat 2018;28:333–49.
[58] Li M, Wang W, Zhang L, Xin W, Zhao Y, Huan L, et al. Genetic polymorphisms in sox17 associated with intracranial aneurysm in Chinese han people: a genotype-phenotype study. Neuropsychiatric Dis Treat 2019;15:779–83.
[59] Castano R, Bosse Y, Endam LM, Filali-Mouhim A, Desrosiers M. C-met pathway involvement in chronic rhinosinusitis: a genetic association analysis. Otolaryngol Head Neck Surg 2010;142:665–71. e661-662.
[60] Kanda A, Chen W, Othman M, Branham KE, Brooks M, Khanna R, et al. A variant of mitochondrial protein loc387715/arms2, not htra1, is strongly associated with age-related macular degeneration. Proc Natl Acad Sci U S A 2007;104:16227–32.
[61] Tong Y, Liao J, Zhang Y, Zhou J, Zhang H, Mao M. Loc387715/htra1 gene polymorphisms and susceptibility to age-related macular degeneration: a huge review and meta-analysis. Mol Vis 2010;16:1958–81.
[62] Kathiresan S, Willer CJ, Peloso GM, Demissie S, Musunuru K, Schadt EE, et al. Common variants at 30 loci contribute to polygenic dyslipidemia. Nat Genet 2009;41:56–65.
[63] Willer CJ, Sanna S, Jackson AU, Scuteri A, Bonnycastle LL, Clarke R, et al. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. Nat Genet 2008;40:161–9.
[64] Brown ML, Inazu A, Hesler CB, Agellon LB, Mann C, Whitlock ME, et al. Molecular basis of lipid transfer protein deficiency in a family with increased high-density lipoproteins. Nature 1989;342:448–51.
[65] Naseri P, Khodakarim S, Guity K, Daneshpour MS. Familial aggregation and linkage analysis with covariates for metabolic syndrome risk factors. Gene 2018;659:118–22.
[66] van Leeuwen R, Klaver CC, Vingerling JR, Hofman A, van Duijn CM, Stricker BH, et al. Cholesterol and age-related macular degeneration: is there a link? Am J Ophthalmol 2004;137:750–2.
[67] Luo X, Lin F, Zhu S, Yu M, Zhang Z, Meng L, et al. Mine landslide susceptibility assessment using ivm, ann and svm models considering the contribution of affecting factors. PloS One 2019;14:e0215134.
[68] McCarty CA, Mukesh BN, Fu CL, Mitchell P, Wang JJ, Taylor HR. Risk factors for age-related maculopathy: the visual impairment project. Arch Ophthalmol 2001;119:1455–62.
[69] Chakravarthy U, Wong TY, Fletcher A, Piault E, Evans C, Zlateva G, et al. Clinical risk factors for age-related macular degeneration: a systematic review and meta-analysis. BMC Ophthalmol 2010;10:31.
[70] Chou R, Dana T, Bougatsos C, Grusing S, Blazina I. Screening for impaired visual acuity in older adults: updated evidence report and systematic review for the us preventive services task force. J Am Med Assoc 2016;315:915–33.
[71] Biesemeier A, Taubitz T, Julien S, Yoeruek E, Schraermeyer U. Choriocapillaris breakdown precedes retinal degeneration in age-related macular degeneration. Neurobiol Aging 2014;35:2562–73.
[72] Johnson KW, Torres Soto J, Glicksberg BS, Shameer K, Miotto R, Ali M, et al. Artificial intelligence in cardiology. J Am Coll Cardiol 2018;71:2668–79.