

# Predicting visual acuity with machine learning in treated ocular trauma patients

Zhi-Lu Zhou<sup>1,2</sup>, Yi-Fei Yan<sup>3,4</sup>, Jie-Min Chen<sup>2</sup>, Rui-Jue Liu<sup>2</sup>, Xiao-Ying Yu<sup>2</sup>, Meng Wang<sup>2</sup>, Hong-Xia Hao<sup>2,5</sup>, Dong-Mei Liu<sup>2</sup>, Qi Zhang<sup>3,4</sup>, Jie Wang<sup>1</sup>, Wen-Tao Xia<sup>2</sup>

<sup>1</sup>Department of Forensic Medicine, Guizhou Medical University, Guiyang 550009, Guizhou Province, China

<sup>2</sup>Shanghai Key Laboratory of Forensic Medicine, Shanghai Forensic Service Platform, Institute of Forensic Science, Ministry of Justice, Shanghai 200063, China

<sup>3</sup>The SMART (Smart Medicine and AI-based Radiology Technology) Lab, Shanghai Institute for Advanced Communication and Data Science, Shanghai University, Shanghai 200444, China

<sup>4</sup>School of Communication and Information Engineering, Shanghai University, Shanghai 200444, China

<sup>5</sup>Basic Medical College, Jiamusi University, Jiamusi 154007, Heilongjiang Province, China

**Co-first authors:** Zhi-Lu Zhou, Yi-Fei Yan, and Jie-Min Chen

**Correspondence to:** Wen-Tao Xia. No.1347, Guangfu West Road, Putuo, Shanghai 200063, China. xiawt@ssfjd.cn; Jie Wang. No.9, Beijing Road, Yunyan, Guiyang 50009, Guizhou Province, China. Wj6400@gmc.edu.cn; Qi Zhang. School of Communication and Information Engineering, Shanghai University, Shanghai 200444, China. zhangq@t.shu.edu.cn

Received: 2022-05-05 Accepted: 2023-05-29

## Abstract

• **AIM:** To predict best-corrected visual acuity (BCVA) by machine learning in patients with ocular trauma who were treated for at least 6mo.

• **METHODS:** The internal dataset consisted of 850 patients with 1589 eyes and an average age of 44.29y. The initial visual acuity was 0.99 logMAR. The test dataset consisted of 60 patients with 100 eyes collected while the model was optimized. Four different machine-learning algorithms (Extreme Gradient Boosting, support vector regression, Bayesian ridge, and random forest regressor) were used to predict BCVA, and four algorithms (Extreme Gradient Boosting, support vector machine, logistic regression, and random forest classifier) were used to classify BCVA in patients with ocular trauma after treatment for 6mo or longer. Clinical features were obtained from outpatient records, and ocular parameters were extracted from optical coherence tomography images and fundus

photographs. These features were put into different machine-learning models, and the obtained predicted values were compared with the actual BCVA values. The best-performing model and the best variable selected were further evaluated in the test dataset.

• **RESULTS:** There was a significant correlation between the predicted and actual values [all Pearson correlation coefficient (PCC)>0.6]. Considering only the data from the traumatic group (group A) into account, the lowest mean absolute error (MAE) and root mean square error (RMSE) were 0.30 and 0.40 logMAR, respectively. In the traumatic and healthy groups (group B), the lowest MAE and RMSE were 0.20 and 0.33 logMAR, respectively. The sensitivity was always higher than the specificity in group A, in contrast to the results in group B. The classification accuracy and precision were above 0.80 in both groups. The MAE, RMSE, and PCC of the test dataset were 0.20, 0.29, and 0.96, respectively. The sensitivity, precision, specificity, and accuracy of the test dataset were 0.83, 0.92, 0.95, and 0.90, respectively.

• **CONCLUSION:** Predicting BCVA using machine-learning models in patients with treated ocular trauma is accurate and helpful in the identification of visual dysfunction.

• **KEYWORDS:** ocular trauma; predicting visual acuity; best-corrected visual acuity; visual dysfunction; machine learning  
**DOI:10.18240/ijo.2023.07.02**

**Citation:** Zhou ZL, Yan YF, Chen JM, Liu RJ, Yu XY, Wang M, Hao HX, Liu DM, Zhang Q, Wang J, Xia WT. Predicting visual acuity with machine learning in treated ocular trauma patients. *Int J Ophthalmol* 2023;16(7):1005-1014

## INTRODUCTION

Ocular trauma is the leading cause of blindness in young adults and children<sup>[1]</sup>. Approximately 19 million people worldwide are visually impaired or even blind on account of ocular trauma<sup>[2]</sup>. Owing to the complexity and diversity of ocular trauma and the great difference in post-treatment recovery, clinicians and forensic doctors often pay great attention to the recovery of post-traumatic visual acuity

(VA). To better understand visual recovery after surgery or trauma, many researchers have investigated the relationship between macular<sup>[3-4]</sup>, retinal<sup>[5-6]</sup>, and vitreous<sup>[7]</sup> injuries caused by diseases or trauma. Hou *et al*<sup>[8]</sup> found that macular injury involving the fovea or a thin optic nerve fiber layer results in severe vision loss. Phillips *et al*<sup>[9]</sup> found that in the lens and anterior chamber, timely treatment and effective control can effectively help the recovery of VA. Liang *et al*<sup>[10]</sup> found that postoperative VA was positively correlated with cube volume and cube average thickness (CAT).

With the advent of electrophysiology, new progress has been made in the examination of visual functional conductivity. Studies have found that VA can be assessed by the amplitude waveform and latency period of the pattern visual evoked potential (PVEP) in the scope of vision<sup>[11]</sup>. However, in forensic medicine, the actual VA cannot be determined when performing visual function tests because of the patient's lack of cooperation. Various methods have been proposed to address this problem. At present, the most common are methods of the fogging test, distance transformation, and electrophysiology, but there are many subjective influencing factors of the first two. The waveform of PVEP can be affected by several factors, such as unstable resistance, misalignment of the eye when the patient annotates the stimulus screen, and poor cooperation. Therefore, the results of the image visual evoked potentials therefore need to be combined with the findings of the ocular examination, the site of the ocular injury and the magnitude of the acting forces. During literature review period, it is found that the ocular trauma score (OTS) takes full account of structural changes to the eye, and has been used to predict VA in patients with ocular injuries since the year of 2002<sup>[12]</sup>. Some researchers argue that OTS can provide objective, real, and effective information for the prognosis of ocular trauma, which is also found to be closely related to the severity of ocular trauma and prognosis<sup>[13-14]</sup>. Xiang *et al*<sup>[15]</sup> found that OTS helped identify situations that disguise or exaggerate VA loss in patients with ocular trauma.

Interdisciplinary cross-collaboration has become a new trend, and research on the application of artificial intelligence to ophthalmology is maturing. Chen *et al*<sup>[16]</sup> used an artificial neural network-based machine-learning algorithm to automatically predict VA after ranibizumab treatment for diabetic macular edema. Huang *et al*<sup>[17]</sup> predicted VA and best-corrected visual acuity (BCVA) using a feed-forward artificial neural networks (ANN) and an error back-propagation learning algorithm in patients with retinopathy of prematurity. Rohm *et al*<sup>[18]</sup> used five different machine-learning algorithms to predict VA in patients with neovascular age-related macular degeneration after ranibizumab injections. Wei *et al*<sup>[19]</sup> developed a deep learning algorithm based on optical

coherence tomography (OCT), to predict VA after cataract surgery in highly myopic eyes. Murphy *et al*<sup>[20]</sup> developed a fully automated three-dimensional image analysis method for measuring the minimum linear diameter of macular holes and derived an inferred formula to predict postoperative VA in idiopathic macular holes.

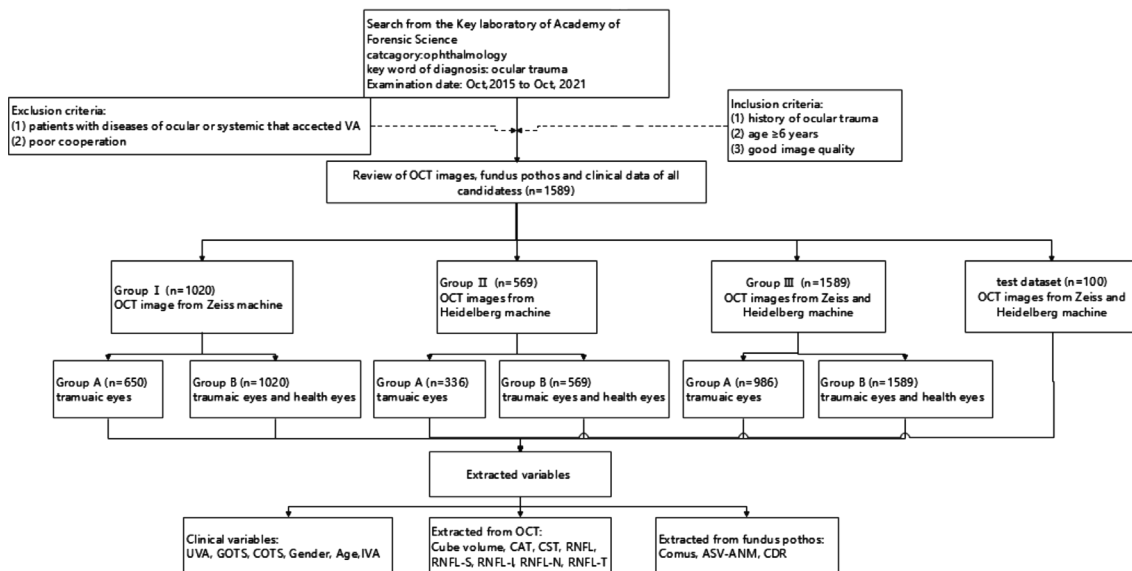
However, most of the above studies are restricted to single-disease studies and no information has been reported so far in relation to the assessment of VA after ocular trauma. This study utilized the relationship between eyeball structure and vision, extracted features from ophthalmology examination, and introduces OTS scores combined with machine-learning techniques to develop a model for predicting BCVA. In addition, the weight of each feature in the model was visualized using a (Shapley additive exPlanations) map to explore the importance of the above features in the task of BCVA prediction.

## MATERIALS AND METHODS

**Ethical Approval** As this study was a retrospective analysis experiment, our ethics committee ruled that approval was not required for this study, and the requirement for individual consent for this study was waived.

**Materials** All internal experimental data were obtained from the Key Laboratory of the Academy of Forensic Science, Shanghai, China. The cases were reviewed to evaluate eligibility based on clinical data, OCT images (Heidelberg, Germany; Carl Zeiss, Goeschwitzer Strasse, Germany), and fundus photographs (Carl Zeiss, Goeschwitzer Strasse, Germany). As of October 2021, the datasets comprised 1589 eyes, including 986 traumatic eyes and 603 healthy eyes. Our inclusion criteria were as follows: 1) differing degrees of ocular trauma; 2) the time between injury was at least 6mo; 3) the therapy records were complete; 4) the BCVA after recovery was proven to be real. Patients with ocular or other systemic diseases that likely affected the VA and poor cooperation were excluded. The test dataset was collected while these models training and validating, was also obtained from the Key Laboratory of the Academy of Forensic Science (Shanghai, China) using the same inclusion and exclusion criteria. From January 2022 to April 2023, the test dataset comprised 100 eyes, including 71 traumatic eyes and 29 healthy eyes, after removing the cases that do not meet the inclusion criteria.

**Optical Coherence Tomography Images** OCT images of the internal dataset were obtained *via* two different devices: Heidelberg and Zeiss. Because the OCT images were obtained from two machines, and to avoid the effect of fusion of data from two different machines, we divided the data into group I (data from Zeiss), group II (data from Heidelberg), and group III (data from Zeiss and Heidelberg, namely all data). Each group was further divided into a traumatic group (group A)



**Figure 1 Process of sample selection and source of the variables** OCT: Optical coherence tomography; UVA: Uncorrected visual acuity; GOTS: Grading of ocular trauma score; COTS: Classification of ocular trauma score; IVA: Initial vision acuity; CAT: Cube average thickness; CST: Cube subfield thickness; RNFL: The average thickness of retinal nerve fiber layer; RNFL-S: Superior RNFL; RNFL-I: Inferior RNFL; RNFL-N: Nasal RNFL; RNFL-T: Temporal RNFL; ASV-ANM: Areas the specific value of abnormal and normal macula; CDR: Cup to disc ratio.

and a healthy and traumatic group (group B). For example, IA represented the traumatic group of Zeiss, IIB represented the traumatic and healthy eye group of Heidelberg, and IIIA represented the traumatic eyes of all data.

OCT images of the test dataset were also obtained *via* the Zeiss and Heidelberg device, which consisted of 100 eyes.

**Feature Extraction** We obtained 17 variables (Figure 1), which consisted of six variables from clinical data, eight extracted from OCT images, and three extracted from fundus photos. Table 1 shows the OTS scoring and grading process<sup>[12]</sup>. BCVA was obtained using a projector-eye chart (NIDEK; Aichi magistrate, Japan). For the convenience of statistical analysis, we converted the decimal VA to logarithm of minimal angle of resolution (logMAR) VA<sup>[21]</sup>. To extract variables from fundus photographs, we used the fundus photo reader software RadiAnt DICOM Viewer (Medixant, Poznan, Poland) to measure the area or length that we needed. To decrease the error, the data were measured by the same person, and the average value of the two measurements was taken. The main variable extraction process is shown in Figure 2.

**Vision Acuity Conversion** The VA was converted to its logMAR equivalent, with counting fingers being assigned a value of 1.9, hand motion 2.3, light perception 2.7, and no light perception 3.0<sup>[19,21]</sup>.

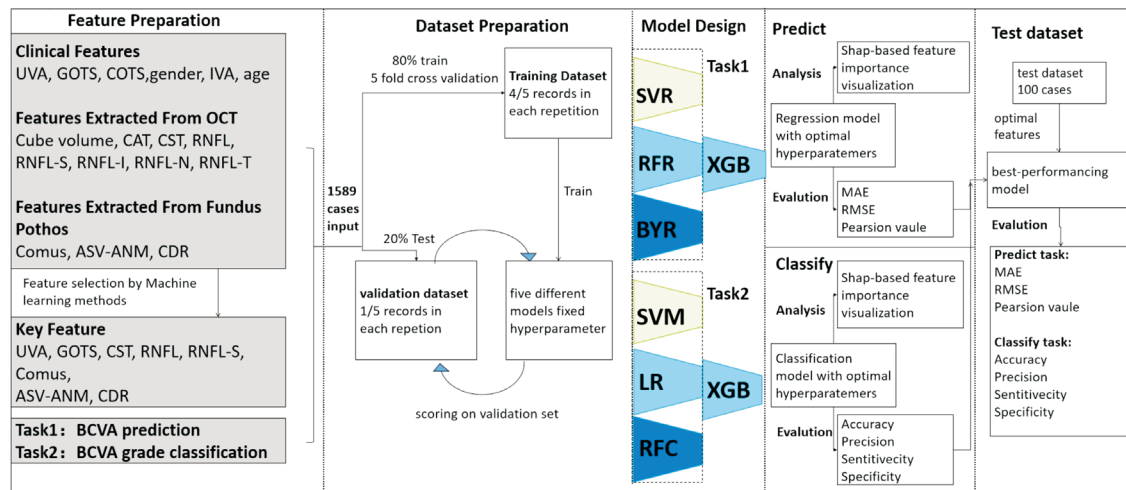
**Overview of Machine-Learning Analytic Systems** We proposed a BCVA analysis system based on machine-learning methods, including the prediction and grading of BCVA using the Extreme Gradient Boosting (XGB) model, and combined it with the method of model post-hoc interpretation, namely SHAP, to analyze the importance of input features of the model.



**Figure 2 Process of variable extraction from fundus photos** Area 1 is the abnormal area of the macula, 2 is the normal area of the macula, and 1/2 is areas the specific value of abnormal and normal macula (ASV-ANM). Area 3 denotes abnormal areas of the optic disk, 4 denotes normal areas of the optic disk, and 3/4 denotes Comus. Area 6 denotes the vertical diameter of the optic cup, 5 denotes the vertical diameter of the optic disc, and 6/5 denotes the cup to disc ratio (CDR).

A flowchart of the experiment is shown in Figure 3. First, to determine the importance of each feature, all available features were used to predict BCVA using the XGB model and SHAP method, and then features were filtered by combining the least absolute shrinkage selection operator (LASSO) with an independent sample *t*-test.

Second, to complete the regression and classification task the features obtained after screening were randomly divided into the training dataset and the validation dataset by a five-fold cross-validation method using the four models. Then, we performed ablation experiments on screened features and investigated the role of each feature in the corresponding task



**Figure 3 The flowchart of our experiment** UVA: Uncorrected visual acuity; GOTS: Grading of ocular trauma score; COTS: Classification of ocular trauma score; IVA: Initial vision acuity; CAT: Cube average thickness; CST: Cube subfield thickness; RNFL: The average thickness of retinal nerve fiber layer; RNFL-S: Superior RNFL; RNFL-I: Inferior RNFL; RNFL-N: Nasal RNFL; RNFL-T: Temporal RNFL; ASV-ANM: Areas the specific value of abnormal and normal macula; CDR: Cup to disc ratio; BCVA: Best-corrected visual acuity; SVR: Support vector regression; RFR: Random forest regressor; BYR: Bayesian ridge; XGB: Extreme gradient boosting; SVM: Support vector machine; LR: Logistic regression; RFC: Random forest classifier; MAE: Mean absolute error; RMSE: Root mean square error.

**Table 1 The input variables' statistics**

Variable	IA	IB	IIA	IIB	IIIA	IIIB	Test dataset
No. of left eyes	345 (53.1%)	508 (49.8%)	182 (54.2%)	279 (49.0%)	527 (54.4%)	787 (49.5%)	49 (49.0%)
No. of right eyes	305 (46.9%)	512 (50.2%)	154 (45.8%)	290 (51.0%)	459 (45.6%)	802 (50.5%)	51 (51.0%)
Age ( $y, x_1$ )	44.59±15.49	44.06±15.10	44.82±13.20	44.70±13.65	44.67±14.74	44.29±14.60	42.29±15.35
Male (gender, $x_2$ )	519 (79.8%)	821 (80.5%)	261 (77.7%)	456 (80.1%)	780 (79%)	1277 (80.4%)	79 (709.0%)
Female (gender, $x_2$ )	131 (20.2%)	199 (19.5%)	75 (22.3%)	113 (19.9%)	206 (21%)	312 (19.6%)	22 (22.0%)
IVA (logMAR, $x_3$ )	1.45±1.02	0.95±1.06	1.34±1.02	0.84±1.02	1.47±0.96	0.99±0.99	1.14±1.26
UVA (logMAR, $x_4$ )	1.28±1.01	0.87±0.99	1.04±0.98	0.67±0.91	1.22±0.99	0.82±0.96	0.97±1.08
GOTS ( $x_5$ )	3.34±1.06	3.89±1.15	3.50±1.06	4.05±1.09	78.49±17.02	86.11±16.82	4.24±1.17
COTS ( $x_6$ )	77.49±17.35	85.19±17.41	80.45±16.21	87.76±15.58	3.39±1.06	3.95±1.13	86.89±17.15
CST ( $x_7$ )	251.67±18.58	249.44±93.91	247.85±94.22	238.64±75.43	248.49±111.49	226.72±106.20	247.02±55.81
Cube volume ( $x_8$ )	9.83±1.10	10.04±3.22	-	-	9.83±1.10	10.04±3.22	9.59±1.40
CAT ( $x_9$ )	273.09±30.66	275.68±26.07	-	-	273.09±30.66	275.68±26.07	217.52±125.43
RNFL ( $x_{10}$ )	88.81±23.55	92.16±20.63	93.73±26.06	97.46±23.02	89.94±25.47	93.59±22.61	91.91±24.50
RNFL-S ( $x_{11}$ )	108.04±35.94	115.62±45.42	113.03±36.35	119.12±32.89	107.97±38.45	116.28±42.10	112.26±32.10
RNFL-I ( $x_{12}$ )	110.00±35.90	116.19±32.07	117.39±37.35	125.62±55.34	110.70±38.96	118.97±42.93	112.32±29.54
RNFL-N ( $x_{13}$ )	65.99±17.74	66.82±16.27	64.95±29.52	67.02±26.31	64.57±23.79	66.55±20.94	69.66±30.85
RNFL-T ( $x_{14}$ )	72.53±35.48	71.71±29.23	80.34±30.03	80.35±25.41	74.74±34.31	74.45±28.69	71.79±27.95
Comus ( $x_{15}$ )	0.39±0.25	0.42±0.34	0.42±0.24	0.39±0.23	0.39±0.23	0.28±0.24	0.33±0.21
ASV-ANM ( $x_{16}$ )	0.79±0.32	0.73±0.37	0.66±0.38	0.68±0.37	0.79±0.34	0.75±0.34	0.27±0.09
CDR ( $x_{17}$ )	0.44±0.22	0.42±0.19	0.42±0.23	0.41±0.19	0.44±0.23	0.41±0.19	0.43±0.17
BCVA (logMAR, $y$ )	0.86±1.05	0.55±0.94	0.74±0.98	0.44±0.83	0.90±0.98	0.57±0.88	0.96±1.51

$x_i$  is considered the input variable to the model and  $y$  is the output of the model. Due to the logMAR value being lower represents the vision acuity higher, and is sometimes shown as negative values, the standard deviation values may be higher than the mean values. IVA: Initial vision acuity; UVA: Uncorrected visual acuity; GOTS: Grading of ocular trauma score; COTS: Classification of ocular trauma score; CST: Cube subfield thickness; CAT: Cube average thickness; RNFL: The average thickness of retinal nerve fiber layer; RNFL-S: Superior RNFL; RNFL-I: Inferior RNFL; RNFL-N: Nasal RNFL; RNFL-T: Temporal RNFL; ASV-ANM: Areas the specific value of abnormal and normal macula; CDR: Cup to disc ratio; BCVA: Best-corrected visual acuity; logMAR: Logarithm of the minimum angle of resolution.

using SHAP. Finally, the eligible test dataset was used further validate the best-performing model and best variables.

**Feature Selection** The optimal combination of features was selected using the SHAP method combined with the LASSO and an independent sample *t*-test. SHAP is an additively explanatory model inspired by cooperative game theory, in which all features are considered “contributors” to the model. The model generates a contribution value, the Shapley value, for each predicted sample, which is the value assigned to each feature in the sample (*i.e.*, the importance of each feature in the model). The LASSO method is widely used in model improvement and selection, and it makes the non-importance coefficient of features parallel to zero to select features by compressing the coefficient of the features and selecting the punishment function. However, the independent sample *t*-test is a common method in statistical analysis that can be used to test whether the difference between the means of the two types of samples is significant.

The specific methods were 1) inputting all features of group III into the model to predict BCVA and the Shapley value of this model corresponding to all features was calculated by the SHAP method and ranked. 2) Input all features of group III into the model, and the feature that is screened by the LASSO method is viewed as group L. 3) Divide group III data into two groups with logMAR of 0.3 as the critical value, and perform an independent sample *t*-test on all features of these two groups, and the feature with a significant difference as group T. 4) The intersection of groups L and T was determined, and the feature with too small a Shapley value was derived as the final feature.

**Extreme Gradient Boosting** Extreme Gradient Boosting (XGB) is a tree-integrated model widely used in Kaggle competitions and many other machine-learning competitions with good results. This inference was computed based on the residuals of the upper model. XGB is an optimized gradient tree boosting system that improves computational speed through algorithmic innovations, such as parallel and distributed computation and approximates greedy search, which is controlled by adding regularization coefficients and residual learning to the loss function. In addition, XGB can learn sparse data and has good generalization ability.

**Regression and Classification Model** The features used in this experiment were extracted based on small samples of clinical variables, using OCT images. Considering the high applicability of machine learning to small sample sets, we selected XGB, support vector regression (SVR), Bayesian ridge (BYR), and random forest regressor (RFR) regression models using the filtered features as model inputs and a grid search approach to find the optimal hyperparameters to predict BCVA for each of the three data sets. For the BCVA

classification task, we classified the BCVA of all patients into two categories using logMAR equal to 0.3 as the threshold value. For patients with logMAR less than or equal to 0.3, we assigned label 1, and for patients with logMAR greater than 0.3, we assigned label 0. For this binary classification task, we also used four machine-learning models: XGB, Support Vector Machine (SVM), Logistic Regression Classifier (LRC), and Random Forest Classifier (RFC), to filter the features as input, and grid search hyperparameters to complete the classification of BCVA.

**Evaluation Standards** For the three experimental datasets, we used a five-fold cross-validation method to separate the training and test sets. For the prediction of corrected VA, we used Pearson correlation coefficient (PCC), Mean absolute error (MAE), and root mean square error (RMSE) to measure the accuracy of model prediction, where  $y_i$  is the true value of corrected VA,  $\hat{y}_i$  is the predicted corrected VA, and  $n$  is the number of samples.

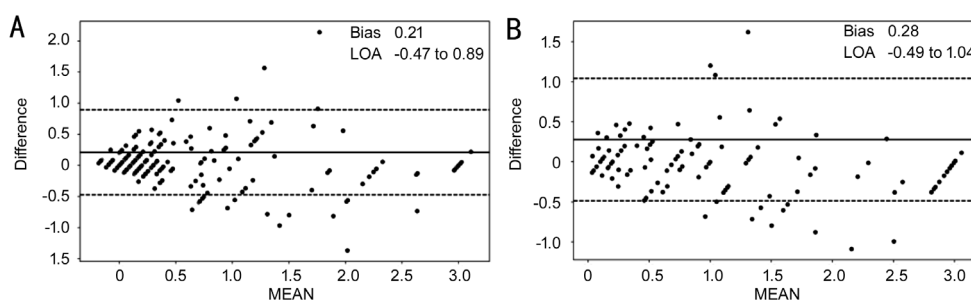
$$\left\{ \begin{array}{l} MAE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \\ RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \\ PCC = \frac{\sum_{i=1}^n \left( y_i - \frac{1}{n} \sum_{i=1}^n y_i \right) \left( \hat{y}_i - \frac{1}{n} \sum_{i=1}^n \hat{y}_i \right)}{\sqrt{\sum_{i=1}^n \left( y_i - \frac{1}{n} \sum_{i=1}^n y_i \right)^2} \sqrt{\sum_{i=1}^n \left( \hat{y}_i - \frac{1}{n} \sum_{i=1}^n \hat{y}_i \right)^2}} \end{array} \right. \quad (1)$$

For the classification of corrected VA classes, we used accuracy, sensitivity, specificity, and precision as evaluation metrics, TP, TN, FP, and FN denote the numbers of true positives, true negatives, false positives, and false negatives, respectively, and samples with logMAR >0.30 represent positive samples.

$$\left\{ \begin{array}{l} Precision = \frac{TP}{TP + FP} \\ Sensitivity = \frac{TP}{TP + FN} \\ Specificity = \frac{TN}{TN + FP} \\ Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \end{array} \right. \quad (2)$$

**Statistics** The experiments were performed on a Dell computer with an Intel (R) Core (TM) i7-10870H CPU @ 2.20 GHz and 32 GB RAM. The model development was performed using Python (version 3.10) with the sci-kit-learn library (version 1.0.2), and statistical analyses were performed using a commercially available statistical software package (SPSS Statistics; IBM Chicago, USA).

The distribution of variables was described by calculating the mean and standard deviation of each continuous variable for



**Figure 4 Consistency of BCVAs between prediction and ground truth** Agreement assessed using Bland-Altman for the predicted value of BCVA and the gold standard in groups A (A) and B (B). In the plots, the solid lines represent the actual mean difference (bias), and dotted lines show 95% limits of agreement. BCVA: Best-corrected visual acuity.

**Table 2 The performance predicted in four models**

Groups	XGB model			SVR model		
	PCC	MAE	RMSE	PCC	MAE	RMSE
Group IA	0.88±0.04	0.32±0.03	0.45±0.05	0.77±0.07	0.34±0.03	0.47±0.04
Group IB	0.91±0.02	0.22±0.02	0.34±0.02	0.90±0.02	0.23±0.02	0.36±0.04
Group IIA	0.90±0.02	0.30±0.04	0.40±0.05	0.84±0.05	0.32±0.05	0.46±0.08
Group IIB	0.92±0.02	0.20±0.02	0.32±0.03	0.88±0.05	0.23±0.03	0.38±0.1
Group IIIA	0.84±0.02	0.31±0.02	0.42±0.03	0.87±0.06	0.31±0.03	0.47±0.11
Group IIIB	0.92±0.01	0.21±0.01	0.32±0.02	0.88±0.07	0.22±0.01	0.40±0.11

XGB: Extreme gradient boosting; SVR: Support vector regression; PCC: Pearson correlation coefficient; MAE: Mean absolute error; RMSE: Root mean square error.

**Table 3 The performance predicted in four models**

Groups	BYR model			RFR model		
	PCC	MAE	RMSE	PCC	MAE	RMSE
Group IA	0.76±0.05	0.36±0.05	0.49±0.07	0.84±0.03	0.30±0.04	0.41±0.03
Group IB	0.88±0.02	0.28±0.01	0.39±0.01	0.91±0.03	0.20±0.01	0.34±0.03
Group IIA	0.76±0.12	0.40±0.04	0.58±0.17	0.86±0.04	0.30±0.03	0.42±0.06
Group IIB	0.84±0.07	0.29±0.03	0.45±0.10	0.90±0.05	0.20±0.01	0.33±0.02
Group IIIA	0.83±0.07	0.38±0.01	0.53±0.09	0.89±0.01	0.29±0.02	0.43±0.03
Group IIIB	0.87±0.02	0.28±0.01	0.41±0.05	0.91±0.03	0.20±0.02	0.33±0.03

BYR: Bayesian ridge; RFR: Random forest regressor; PCC: Pearson correlation coefficient; MAE: Mean absolute error; RMSE: Root mean square error.

all data. Continuous variables of group A were compared with those of group B using the independent samples *t*-test method. The consistency between the predicted and actual values was verified using the Bland-Altman diagram. PCC was used to analyze the correlation between the predicted and actual values.

**RESULTS**

The characteristics of the input variables are listed in Table 1. The average age of all the groups was close to 44y. The number of men was higher than that of women in each group. The optimal variables for predicting BCVA include uncorrected VA (UVA), grading of ocular trauma score (GOTS), cube subfield thickness (CST), the average thickness of retinal nerve fiber layer (RNFL), superior RNFL (RNFL-S), Comus, areas the specific value of abnormal and normal macula (ASV-ANM), and cup to disc ratio (CDR) were obtained after feature selection.

**Best-Corrected Visual Acuity Prediction Performance Analysis** In Tables 2 and 3, we compare the performance of the four machine-learning algorithms after using five-

fold cross-validation when classifying the dataset into three groups. XGB obtained better results in Groups IA, IIA, and IIIA, with MAEs of 0.32±0.03, 0.30±0.04, and 0.32±0.02, respectively, and RMSEs of 0.45±0.05, 0.40±0.05 and 0.42±0.03, respectively. The best results were obtained for the random forest model in three data groups, IB, IIB, and IIIB, with MAEs of 0.20±0.01, 0.20±0.01, 0.20±0.02, and RMSEs 0.24±0.03, 0.33±0.02, 0.33±0.03, respectively.

Figure 4 shows a Bland-Altman plot assessing the agreement between the model predictions and the ground truth. The 95% confidence intervals and mean deviations for the consistency of predictions for group III are presented separately. The 95% confidence intervals for groups IIIA and IIIB are -0.47 to 0.89 logMAR and -0.49 to 1.04, with mean deviations of 0.21 and 0.28 respectively.

**Best-Corrected Visual Acuity Grade Classification Performance Analysis** The experimental results were evaluated using five-fold cross-validation and a grid search approach to determine the best parameters for the model.

**Table 4 The performance of the classified model**

Groups	XGB model				SVM model			
	Sensitivity	Precision	Specificity	Accuracy	Sensitivity	Precision	Specificity	Accuracy
Group IA	0.91±0.05	0.86±0.05	0.69±0.11	0.84±0.03	0.84±0.07	0.85±0.08	0.70±0.15	0.79±0.07
Group IB	0.82±0.06	0.80±0.02	0.89±0.02	0.86±0.03	0.74±0.04	0.81±0.04	0.90±0.03	0.84±0.02
Group IIA	0.93±0.01	0.86±0.06	0.76±0.09	0.86±0.04	0.87±0.02	0.87±0.06	0.80±0.08	0.84±0.04
Group IIB	0.84±0.03	0.82±0.04	0.90±0.03	0.88±0.02	0.74±0.11	0.87±0.06	0.94±0.03	0.86±0.05
Group IIIA	0.92±0.02	0.86±0.03	0.71±0.07	0.85±0.03	0.88±0.04	0.86±0.03	0.72±0.06	0.82±0.02
Group IIIB	0.82±0.03	0.82±0.04	0.90±0.02	0.87±0.01	0.73±0.02	0.86±0.04	0.93±0.02	0.86±0.01

XGB: Extreme Gradient Boosting; SVM: Support vector machine. Sensitivity: Number of correctly predicted eyes with BCVA≤0.3 logMAR / total number of eyes with actual BCVA≤0.3 logMAR. Precision: The number of correctly predicted eyes with BCVA≤0.3 logMAR / total number of predicted eyes with BCVA≤0.3 logMAR. Specificity: The number of correctly predicted eyes with BCVA>0.3 logMAR / total number of eyes with actual BCVA>0.3 logMAR. Accuracy: The total number of correctly predicted eyes / total number of eyes.

**Table 5 The performance of the classified model**

Groups	LR model				RFC model			
	Sensitivity	Precision	Specificity	Accuracy	Sensitivity	Precision	Specificity	Accuracy
Group IA	0.87±0.06	0.85±0.04	0.67±0.12	0.81±0.03	0.86±0.02	0.86±0.05	0.71±0.07	0.82±0.03
Group IB	0.75±0.07	0.83±0.07	0.92±0.03	0.86±0.03	0.73±0.05	0.86±0.04	0.94±0.02	0.86±0.02
Group IIA	0.88±0.03	0.88±0.02	0.79±0.05	0.85±0.03	0.89±0.04	0.86±0.03	0.76±0.04	0.84±0.03
Group IIB	0.77±0.06	0.83±0.07	0.92±0.03	0.86±0.03	0.74±0.02	0.81±0.04	0.90±0.02	0.84±0.02
Group IIIA	0.87±0.02	0.86±0.01	0.74±0.04	0.82±0.01	0.89±0.04	0.88±0.04	0.77±0.06	0.85±0.03
Group IIIB	0.75±0.04	0.83±0.05	0.91±0.03	0.85±0.01	0.73±0.03	0.81±0.05	0.91±0.02	0.84±0.02

LR: Logistic regression; RFC: Random forest classifier.

The qualitative results shown in Tables 4 and 5 indicate that XGB obtained better results for all groups using the same combination of features than the other three methods. Sensitivity, precision, specificity, and accuracy for group IIIA classification were 0.92±0.02, 0.86±0.03, 0.71±0.07, and 0.85±0.03, respectively, and in group IIIB 0.82±0.03, 0.82±0.04, 0.90±0.02, and 0.87±0.01, respectively. The sensitivity of group A was higher than the specificity in all three datasets, indicating that the prediction accuracy of BCVA ≤0.3 (logMAR) was higher than the prediction accuracy of BCVA>0.3 (logMAR).

**Test Dataset** Since the XGB model has the most optimal performance for the performing the regression and classification tasks, the test dataset was used to determine the prediction performance and the accurate classification performance of the XGB model. As shown in Tables 6 and 7, the XGB model demonstrated stable promising results with an MAE of 0.20, RMSE of 0.29, and PCC of 0.96. The sensitivity, precision, specificity, and accuracy were 0.83, 0.92, 0.95, and 0.90, respectively. The Figure 5 shown the confusion matrixes of XGB model in group IIIB of internal dataset and test dataset.

To understand the role of each feature in the model, we visualized the importance of the XGB model features that obtained the best results using SHAP, where importance refers to the extent to which each feature contributes to the model's predicted results. As shown in Figures 6 and 7, the UVA

**Table 6 The result of regression between the internal dataset and test dataset**

XGB model	PCC	MAE	RMSE
Internal dataset (group IIIB)	0.92±0.01	0.21±0.01	0.32±0.02
Test dataset	0.96	0.20	0.29

XGB: Extreme gradient boosting; PCC: Person correlation coefficient; MAE: Mean absolute error; RMSE: Root mean square error. Group IIIB represents the all samples in internal dataset.

**Table 7 The result of classification between the internal dataset and test dataset**

XGB model	Sensitivity	Precision	Specificity	Accuracy
Internal dataset (group IIIB)	0.82±0.03	0.82±0.04	0.90±0.02	0.87±0.01
Test dataset	0.83	0.92	0.95	0.90

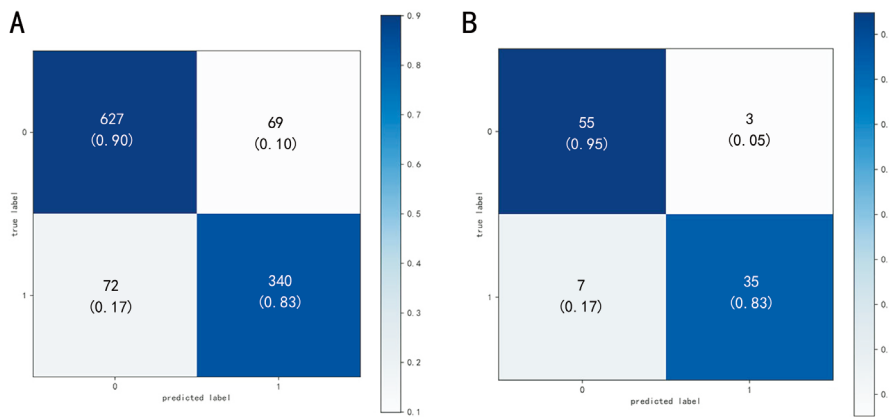
XGB: Extreme gradient boosting; Group IIIB represents the all samples in internal dataset.

played a key role in both the XGB prediction and classification models, and the importance of the remaining features varied in both tasks, with GOTS, RNFL-S, and CST showing a greater contribution to the models in both tasks.

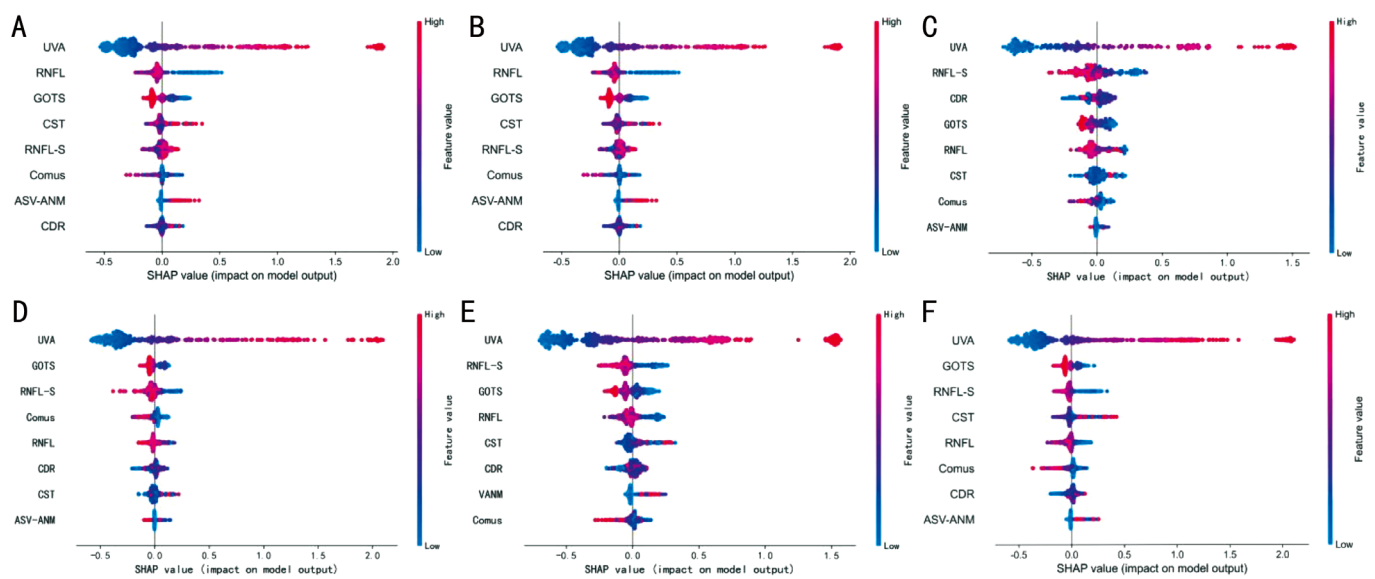
## DISCUSSION

To ensure judicial justice, we should clarify the cause of visual injury and confirm that the BCVA obtained by ophthalmology examination is reliable. In China, VA decline caused by accidental or intentional injury can obtain compensation

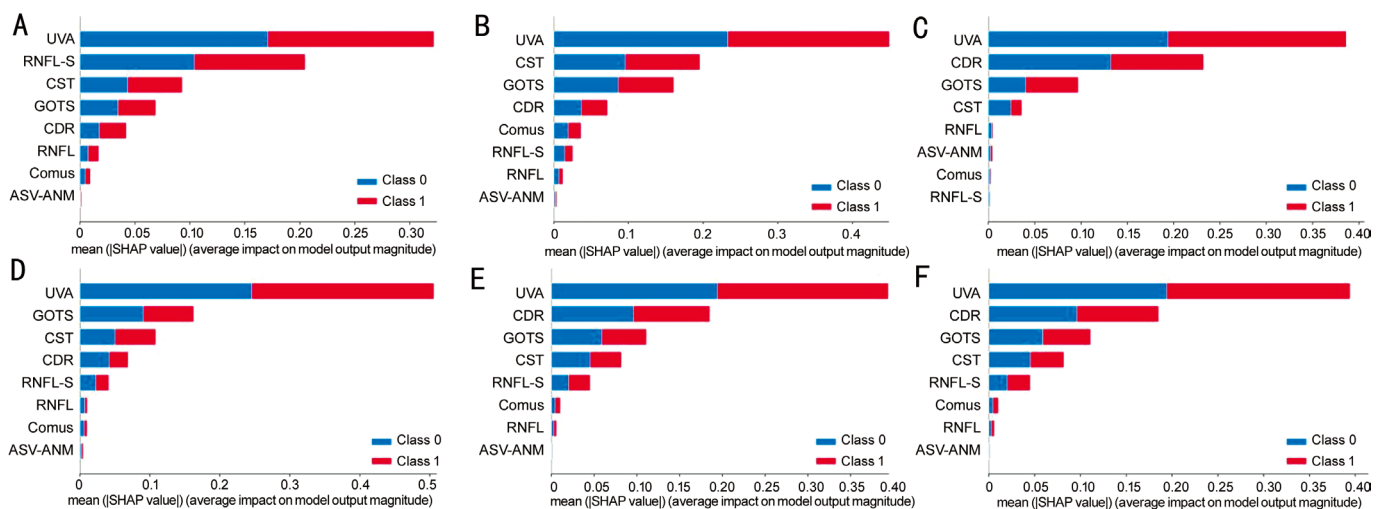




**Figure 5** The confusion matrixes of XGB model: A: All samples in the internal dataset of the XGB model; B: The test dataset of the XGB model, the predicted label = 0 with true label = 0 represents the correct prediction of visual acuity > 0.3 logMAR, the predicted label = 1 with true label = 1 represents the correct prediction of visual acuity ≤ 0.3 logMAR. XGB: Extreme gradient boosting.



**Figure 6** Plot of weights of the different features for the BCVA prediction task. The global feature importance plots for all groups with the horizontal coordinates indicating the Shapley value for each sample corresponding to each feature and the color of each point representing the magnitude of this sample feature value. A, B, C, D, E, and F represent group IA, IB, IIA, IIB, IIIA, and IIIB, respectively. BCVA: Best-corrected visual acuity.



**Figure 7** Plot of weights of different features of the BCVA classification task. BCVA classification task in all groups, where class 1 represents BCVA ≤ 0.3 and class 2 represents BCVA > 0.3. A, B, C, D, E, and F represent group IA, IB, IIA, IIB, IIIA, and IIIB, respectively. BCVA: Best-corrected visual acuity.



or impose punishment on the perpetrator. Therefore, some patients disguise severe vision loss or blindness. VA in ocular trauma is usually worse in patients with posterior segment involvement. Therefore, a forecast model that can assess BCVA might be helpful for the accurate judgment of forensic workers. Although there are patients with poor cooperation in practice, because high-resolution OCT and fundus photographs may reveal morphological changes, the factors affecting VA can be identified. In recent years, machine learning has been extensively applied, and it has been found that an OCT scan of the macula could provide millions of morphological parameters affecting VA<sup>[22-23]</sup>. Previous studies have mostly focused on the diagnosis and classification of eye diseases. Due to the complex and changeable nature of ocular trauma, such studies are relatively few.

Several studies have developed machine-learning algorithms to predict VA in patients with ocular or systemic diseases. Some of them used OCT images of the macular<sup>[19]</sup>, and some used clinical data and measurement features from OCT (such as central retinal thickness)<sup>[24]</sup>. Others use basic information, such as disease type or condition, age, and sex<sup>[25]</sup>.

As shown in Table 1, the initial VA, UVA, GOTS, classification of ocular trauma score, RNFL, ASV-ANM, CDR, and BCVA were worse in group A than in group B. Our results could identify several rules of thumb: men easily suffered injuries compared to women, and the RNFL was more easily influenced than the macula lutea. The data from OCT images, fundus photos, and clinical information were used to predict BCVA in patients with ocular trauma using the XGB, SVR, BYR, and RFR models, and another four models for the accomplishment of the classification task. The results reveal that these models can predict BCVA in most patients with ocular trauma and shows promising performance. As expected, the best predictor variables by the auto-selected model were including UVA, GOTS, CST, RNFL, RNFL-S, comus<sup>[26]</sup>, ASV-ANM, and CDR. This outcome coincides with the consensus that the VA is closely related to eyeball structure. We can observe that the predicted values are well correlated with the ground truth values ( $P>0.7$ ), and the Bland-Altman plot shows good consistency between the gold standard and predicted values. The XGBoost model had the best performance in Group A, and the RFR model had the optimal results in Group B. In forensic clinical assessment of visual function, the recovery vision after ocular trauma below 0.3 logMAR can be used as a basis for assessing the degree of impairment and disability. To improve the efficiency and accuracy of identifying pseudo visual loss, this experiment was combined with the corresponding conditions of the visual function assessment, and finally a dichotomous experiment with a 0.3 logMAR cut-off was performed. From the classification

results, the XGB model had the highest accuracy in all groups, and sensitivity was always greater than specificity in group A. We speculate that this is a problem of sample imbalance, and we found that in group A, there were more eyes with  $BCVA \leq 0.3$  logMAR than eyes with  $BCVA > 0.3$  logMAR, but in group B, the increase in the number of healthy eyes leads to an increase in the number of eyes with  $BCVA > 0.3$  logMAR. Finally, to prove the generalization of the model, we combined OCT images captured by the two machines to predict VA and compared the outcomes with those of groups I and II; no significant difference in outcomes predicted and classification between each group was observed. We tested the model with additional data to determine how well it eventually performed the regression and classification tasks, and the test set was not involved in the training or gradient descent process of the model. It therefore makes sense to use an independent test set to test the best model against the best variables, and our results show that our model also performs well on the unexposed test set. The results of the test dataset showed that the regression and classification model also showed a stable and promising performance on the test dataset with MAE of 0.20 and RMSE of 0.29. The classification performance was also good.

The advantages of this experiment are as follows. Our experiments were designed based on the relationship between changes in the eye structure and VA. This innovative experimental design may help evaluate VA after injury. We divided the data into three groups to avoid errors due to OCT images obtained from different systems. Our study has some limitations. First, the data extracted from the OCT images and fundus photos were artificial, and several effective features were lost. Second, the sample size needs to be increased to further improve the robustness and generalizability of the machine-learning models. Finally, an error between the predicted and actual values still exists.

In the future, we plan to directly input OCT images and fundus photos into the model for VA prediction and continually increase the sample size to optimize the model. It is also expected to develop an open platform by using real-world clinical data for optimized software, which means inputting ophthalmic images into the model to directly obtain the VA and assist in accurate diagnosis.

In conclusion, owing to the complex and changeable conditions of ocular trauma, the prognosis of vision is difficult to clarify. This is based on the relationship between changes in eye structure and vision and the increasing application of artificial intelligence in ophthalmology. We used four different machine-learning models to predict the BCVA and found useful variables to predict BCVA. It can be used to predict VA and may be helpful for the auxiliary analysis of postoperative VA in clinical ophthalmology.

## ACKNOWLEDGEMENTS

**Foundations:** Supported by National Key R&D Program of China (No.2022YFC3302001); the Human Injury and Disability Degree Classification (No.SF20181312); the National Natural Science Foundation of China (No.62071285).

**Conflicts of Interest:** Zhou ZL, None; Yan YF None; Chen JM, None; Liu RJ, None; Yu XY, None; Wang M, None; Hao HX, None; Liu DM, None; Zhang Q, None; Wang J, None; Xia WT, None.

## REFERENCES

- Liu Y, Q in B, Liu S W. Epidemiology characteristics and ocular trauma score in eye injuries. *Guoji Yanke Zazhi (Int Eye Sci)* 2013;13(10): 1993-1997.
- Flynn TH, Fennessy K, Horgan N, Walsh B, O'Connell E, Cleary P, Beatty S. Ocular injury in hurling. *Br J Sports Med* 2005;39(8):493-496; discussion 496.
- Zheng Y, Xiang ZY, Yang YY, Chen AJ, Chen Q. Predictive value of ectopic inner foveal layers for visual acuity after idiopathic macular epiretinal membrane surgery. *Chin J Ophthalmol and Otorhinolaryngol* 2021;21:9-15.
- Jiang H, Han M, Li QJ. The application of optical coherence tomography in traumatic macular disease. *Chin J Ocular Trauma* 2003;25(8):515-517 .
- Liu YC, Xia WT, Zhu GY, Liu RJ, Liu H, Chen JM. Image of the traumatic optic atrophy detected with optical coherence tomography and the vision function defection. *Fa Yi Xue Za Zhi* 2009;25(2):92-94.
- Aalok K. Effects of blunt trauma of eye on retinal nerve fibre layer. *Int Eye Res* 2021;2(4):228-232.
- Qin ZX, He Y, Xu YF, Yu T, Liu Y, Wu N. Pediatric ocular trauma with pars Plana vitrectomy in Southwest of China: clinical characteristics and outcomes. *Int J Ophthalmol* 2021;14(9):1321-1326.
- Hou CH, Lu YC, Pu C, Chang YH, Lin KK, Lee JS, Chen KJ. Outcomes and longitudinal trend of traumatic cataract wound dehiscence in patients with blunt ocular injury. *Sci Rep* 2021;11(1):18191.
- Phillips HH, Blegen Iv HJ, Anthony C, Davies BW, Wedel ML, Reed DS. Pars Plana vitrectomy following traumatic ocular injury and initial globe repair: a retrospective analysis of clinical outcomes. *Mil Med* 2021;186(Suppl 1):491-495.
- Liang GL, Liu J, Wu ZY, Xu JH. The significance and application value of preoperation SD-OCT for vision prediction after cataract surgery. *ZhengJiang medicine* 2018;40:171-173.
- Gundogan FC, Mutlu FM, Altinsoy HI, Tas A, Oz O, Sobaci G. Pattern visual evoked potentials in the assessment of objective visual acuity in amblyopic children. *Int Ophthalmol* 2010;30(4):377-383.
- Kuhn F, Maisiak R, Mann L, Mester V, Morris R, Witherspoon CD. The ocular trauma score (OTS). *Ophthalmol Clin North Am* 2002;15(2):163-165.vi.
- Unal MH, Aydin A, Sonmez M, Ayata A, Ersanli D. Validation of the ocular trauma score for intraocular foreign bodies in deadly weapon-related open-globe injuries. *Ophthalmic Surg Lasers Imaging Retina* 2008;39(2):121-124.
- Uysal Y, Mutlu FM, Sobaci G. Ocular Trauma Score in childhood open-globe injuries. *J Trauma* 2008;65(6):1284-1286.
- Xiang J, Guo ZM, Wang X, Yu LL, Liu H. Application of ocular trauma score in mechanical ocular injury in forensic medicine. *Fa Yi Xue Za Zhi* 2015;31(5):352-355.
- Chen SC, Chiu HW, Chen CC, Woung LC, Lo CM. A novel machine learning algorithm to automatically predict visual outcomes in intravitreal ranibizumab-treated patients with diabetic macular edema. *J Clin Med* 2018;7(12):475.
- Huang CY, Kuo RJ, Li CH, Ting DS, Kang EY, Lai CC, Tseng HJ, Yang LY, Wu WC. Prediction of visual outcomes by an artificial neural network following intravitreal injection and laser therapy for retinopathy of prematurity. *Br J Ophthalmol* 2020;104(9):1277-1282.
- Rohm M, Tresp V, Müller M, Kern C, Manakov I, Weiss M, Sim DA, Priglinger S, Keane PA, Kortuem K. Predicting visual acuity by using machine learning in patients treated for neovascular age-related macular degeneration. *Ophthalmology* 2018;125(7):1028-1036.
- Wei L, He WW, Wang JR, Zhang KK, Du Y, Qi J, Meng JQ, Qiu XD, Cai L, Fan Q, Zhao ZN, Tang YT, Ni S, Guo HK, Song YX, He XX, Ding DY, Lu Y, Zhu XJ. An optical coherence tomography-based deep learning algorithm for visual acuity prediction of highly myopic eyes after cataract surgery. *Front Cell Dev Biol* 2021;9:652848.
- Murphy DC, Nasrullo AV, Lendrem C, Graziado S, Alberti M, la Cour M, Obara B, Steel DHW. Predicting postoperative vision for macular hole with automated image analysis. *Ophthalmol Retina* 2020;4(12):1211-1213.
- Lange C, Feltgen N, Junker B, Schulze-Bonsel K, Bach M. Resolving the clinical acuity categories "hand motion" and "counting fingers" using the Freiburg Visual Acuity Test (FrACT). *Graefes Arch Clin Exp Ophthalmol* 2009;247(1):137-142.
- Chung YW, Choi MY, Kim JS, Kwon JW. The association between macular thickness and axial length in myopic eyes. *Biomed Res Int* 2019;2019:8913582.
- Abdolrahimzadeh S, Parisi F, Plateroti AM, Evangelista F, Fencia V, Scuderi G, Recupero SM. Visual acuity, and macular and peripapillary thickness in high myopia. *Curr Eye Res* 2017;42(11):1468-1473.
- Waldstein SM, Vogl WD, Bogunovic H, Sadeghipour A, Riedl S, Schmidt-Erfurth U. Characterization of drusen and hyperreflective foci as biomarkers for disease progression in age-related macular degeneration using artificial intelligence in optical coherence tomography. *JAMA Ophthalmol* 2020;138(7):740-747.
- Devalla SK, Renukanand PK, Sreedhar BK, Subramanian G, Zhang L, Perera S, Mari JM, Chin KS, Tun TA, Strouthidis NG, Aung T, Thiéry AH, Girard MJA. DRUNET: a dilated-residual U-Net deep learning network to segment optic nerve head tissues in optical coherence tomography images. *Biomed Opt Express* 2018;9(7):3244-3265.
- Liu YC, Xia WT, Zhu GY, Zhou XT, Fan LH, Liu RJ, Chen JM. The relationship between eyeball structure and visual acuity in high myopia. *Fa Yi Xue Za Zhi* 2010;26(3):169-172.