DOI: 10.19884/j.1672-5220.202302004

Application of Stacking Ensemble Learning in Clinical Fitting of Orthokeratology Lens for Myopia Correction

GONG Jiaming[1,2], LI Kangmei[1], HU Jun[2,3], CHEN Hao[4], CAO Qianqian[2,3*], WU Ge[4*]
1. College of Mechanical Engineering, Donghua University, Shanghai 201620, China
2. Institute of Artificial Intelligence, Donghua University, Shanghai 201620, China
3. Shanghai Industrial Big Data Engineering Technology Research Center, Shanghai 201600, China
4. Eye Hospital, Wenzhou Medical University, Wenzhou 325000, China

Abstract: Aiming at the problems of a large difficulty coefficient and tedious process in the clinical fitting of the orthokeratology (OK) lens, a stacking ensemble learning model is proposed to predict the parameters of the OK lens and realize its intelligent fitting. The feature set that is most suitable for the target variables is constructed by feature derivation based on *F*-test and feature selection under the variance-improved Boruta algorithm. A stacking ensemble learning prediction model is studied. The model uses random forest (RF), gradient boosting decision tree (GBDT) and support vector regression (SVR) as the first layer basic learners and linear regression (LR) as the second layer meta-learner. The experimental results show that the prediction indexes of the model are highly consistent with the clinical diagnosis results, which verifies that the model can be used as an effective auxiliary diagnosis method.

Key words: orthokeratology (OK) lens; feature engineering; stacking ensemble model; parameter prediction; intelligent fitting
CLC number: TP181 **Document code**: A
Article ID: 1672-5220(2024)02-0184-11

Open Science Identity (OSID)

0 Introduction

In recent years, myopia has become more and more serious, and it seems to be a very common phenomenon around the world. According to the current myopia incidence forecast, by 2050, the global myopia population will reach 4.949 billion (accounting for about 52% of the world's population), and the high myopia population will reach 925 million (accounting for about 10% of the world's population)[1-3]. Refractive surgery is one of the most effective methods to treat myopia by operating on the cornea or lens to correct ametropia. This includes conventional laser-assisted in situ keratomileusis (LASIK), all-femtosecond laser, all-laser or implantable collamer lens (ICL) implantation. It can help people with nearsightedness, farsightedness or astigmatism take off the lens[4-6]. However, refractive surgery has many demands on patients, and patients under the age of 18 are usually not eligible for surgery. Since the maintenance of visual acuity after refractive surgery is related to the stability of the refractive power before surgery, laser correction surgery requires the stability of myopia for nearly two years, and the annual change cannot exceed 50 degrees. Atropine is a non-selective antagonist of muscarinic acetylcholine receptors (mAChR). Myopia is associated with chronic inflammation of the eye, which can be controlled by atropine or by biochemical changes at the scleral level that interfere with eye development[7-8]. Topical atropine is used at different concentrations of drops for myopia control[9]. However, atropine has a high potential to induce acute glaucoma. With the use of atropine, pupil dilation leads to an increase in the total amount of light entering the eye, some of the light is partially refracted into the interior of the eye through the limbal part of the cornea and lens, leading to symptoms such as photophobia and glare in the eye[10]. Therefore, neither surgery nor medication is the ideal treatment for myopia.

Wearing orthokeratology (OK) lens is a very popular method for correcting low and moderate myopia, and it is especially effective in relieving children's myopia[11-13]. In its original form, wearing OK lens involves wearing a flat and hard lens made of polymethylmethacrylate (PMMA) while awake, intending to gradually flatten the cornea to reduce myopia and provide reasonable vision after lens removal. The OK lens can be worn during nighttime sleep. Regular use by the patient can maintain the remodeling effect of the cornea, thereby allowing the patient to see things more clearly during the day[14-15]. As a highly permeable rigid material with an anti-geometric design and direct contact with the eye, the risk of fitting is extremely high.

Received date: 2023-02-04
Foundation item: Shanghai Science and Technology Project, China (No. 20DZ2251400)
* Correspondence should be addressed to CAO Qianqian, email: qianqian_cao@dhu.edu.cn; WU Ge, email: wg_wmu@163.com
Citation: GONG J M, LI K M, HU J, et al. Application of stacking ensemble learning in clinical fitting of orthokeratology lens for myopia correction [J]. *Journal of Donghua University (English Edition)*, 2024, 41(2): 184-194.

Clinical fitting requires sufficient experience and has obvious disadvantages such as time-consuming and labor-intensive fitting process, poor fitting with multiple trial changes and double-blindness in fitting. This lengthens the fitting period for patients. Frequent changes of public trial lenses and dynamic fluorescein assessment can cause ocular discomfort and even ocular infection, and incorrect fitting can aggravate myopia symptoms[16-17]. Little is known worldwide about the actual fitting standards for the OK lens[18].

Artificial intelligence, machine learning and deep learning technology have been proven to be effective diagnostic tools for identifying various diseases in medical care services in various scientific studies. Machine learning is a branch of artificial intelligence. It provides a series of tools and methods to identify complex patterns in medical data, which can facilitate very complex and time-consuming tasks. Therefore, it has been applied in various medical fields[19-21]. At present, there are some research results about the application of artificial intelligence to the parameter prediction of the OK lens. In Ref. [22], the prediction of the OK lens parameters was achieved for the first time by traditional machine learning methods, realizing the application of machine learning to the field of the OK lens. Zhang et al.[23] put forward a fitting model of the OK lens based on ophthalmic data, which mainly adopts a linear fitting scheme based on important characteristic factors. A general formula for the OK lens fitting is developed. Zhang et al.[24] put forward a geometric fitting method, by torus fitting the corneal data points in the contact area between the cornea and the OK lens, to find the lens parameters of the OK lens that are most suitable for this area. The above machine learning and computer fitting techniques have been successfully applied to the prediction of the OK lens parameters. However, these studies have failed to fully mine the information in the ophthalmology data set, failed to study the methods to make the prediction model more accurate, and still have the limitations of low accuracy, time-consuming and low efficiency.

The purpose of this research is to study the high-precision intelligent fitting model of the OK lens, predict the alignment curve (AC) and target power (TP) parameters during the fitting of the OK lens for vision shaping treatment (VST), and realize the clinical intelligent diagnosis fitting, thus reducing the number of times to replace the public fitting pieces and shortening the fitting time.

1 Data Information Mining

1.1 Data sources and characteristics

The clinical data of 1 235 myopia patients from the Eye Hospital, Wenzhou Medical University were collected. To ensure variance stability of the data, all inspection measurements were performed by inspectors in the same department. In the collected data, there were 563 males and 672 females with a mean age of 12.00 ± 2.38 years (range from 7 to 21 years). Age (A) and Sex (S) are the input dimensions of the patients' basic information. Medmont corneal topography system (E300U, Australia) and other measuring instruments were used to obtain corneal parameters, including sag difference at 8 mm (SD8), flat keratometry (FK), steep keratometry (SK), difference keratometry (dK), eccentricity value (include flat eccentricity value (Fe) and steep eccentricity value (Se)), axial length (AL), anterior chamber depth (ACD), white to white (WTW), spherical equivalent refraction (SER), cylindrical refraction (CYL) and visual acuity (V). All the patients were fitted with an OK lens (Euclid, USA). All the patients were first-time OK lens wearers and were not affected by other ocular diseases. During the sampling process, there was no occlusion of the upper and lower eyelids and the corneal coverage exceeded 90%. Fluorescent staining patterns of the patients were all well-performed, and unacceptable conditions such as excessive lens center gap and insufficient bearing on the alignment curve were rejected, and the patients with poor treatment outcomes during subsequent follow-up periods were also excluded. This study was conducted in compliance with the tenets of the Declaration of Helsinki and was approved by the Medical Ethics Committee of Eye Hospital, Wenzhou Medical University. The basic clinical information for the subjects is shown in Table 1. Mean and SD in the second column represent the average value and standard deviation of the characters, and range in the third column represents the interval of the characters.

Table 1 Patient clinical information collected

Character	Mean ± SD	Range
A	12.00±2.38	7-21
SD8/μm	24.14±18.75	-66.00-83.00
FK/D	41.37±1.4	37.03-46.48
SK/D	42.71±1.53	38.13-47.42
dK/D	1.35±0.85	0.04-6.77
Fe	0.41±0.18	0-0.90
Se	0.42±0.16	0-0.93
AL/mm	25.09±0.82	22.16-27.19
ACD/mm	3.63±0.27	2.00-6.13
WTW/mm	12.12±0.39	10.90-13.10
SER/D	-2.84±1.15	-7.00--0.25
CYL/D	-0.42±0.06	-12.00-0
V	1.02±0.07	0.80-1.20

1.2 Feature engineering

Only important ophthalmic features were considered in the clinical fitting process, failing to fully consider the influence of other ophthalmic features on the final

parameters of the OK lens, and the dataset suffered from a serious problem of correlated feature scarcity. In this study, feature derivation and feature selection were performed on the original feature dataset to fully exploit the hidden information associated with the predictive variables.

1.2.1 Feature derivation

F-test analysis is carried out between the original 14-dimensional features and the predicted target variables to determine the degree of correlation between each dimensional feature and the predicted variables. The formula of *F*-test is as[25-26]

$$f = \frac{r_i^2(n-2)}{1-r_i^2}, \tag{1}$$

where r_i is calculated as

$$r_i = \frac{(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_i)^{\mathrm{T}}(\boldsymbol{y} - \bar{\boldsymbol{y}})}{d_{\mathrm{st}}(\boldsymbol{X}_i)\, d_{\mathrm{st}}(\boldsymbol{y})}, (i = 1, 2, \cdots, 14), \tag{2}$$

where $\boldsymbol{X}_i$ is the true value column vector of the ith dimensional feature sample; $\overline{\boldsymbol{X}}_i$ is the mean column vector of the ith dimensional feature sample; $\boldsymbol{y}$ represents the true value column vector of the predicted labels; $\bar{\boldsymbol{y}}$ represents the mean column vector of all predicted labels, and the numerator is the inner product of the two column vectors; $d_{\mathrm{st}}(\boldsymbol{X}_i)$ represents the variance of all samples of the ith dimensional feature; $d_{\mathrm{st}}(\boldsymbol{y})$ represents the variance of the corresponding predicted label samples.

According to the results of the *F*-test, these 14 dimensional features are given weight proportional to their *f* values, and different forms of feature derivation such as feature interaction, feature ratio, polynomial and trigonometric functions are carried out. The larger the weight, the more new features related to it will be derived. In this study, 100-dimensional features are generated by feature derivation. Figure 1 is a schematic diagram of the derived new features, where *A* represents A; *S* represents S, D_8 represents SD8; *H* represents FK; *K* represents SK; D_{d} represents dK; *B* represents Fe; *E* represents Se; *L* represents AL; *C* represents ACD; *W* represents WTW; R_{R} represents SER; *Y* represents CYL; *V* represents V.

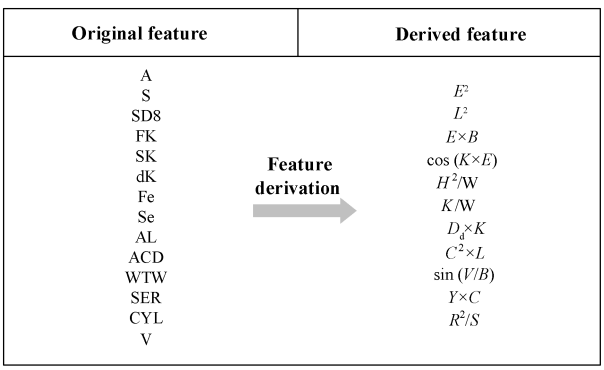


Fig. 1 Feature derivation diagram

1.2.2 Variance-improved Boruta feature selection

The feature selection used in this study is divided into a variance filtering stage and an improved Boruta selection wrapper stage. In the original Boruta algorithm, if all samples of the real features are used to construct shadow features, it will make the operation complexity of the whole process high[27-28]. This study improves the original Boruta algorithm by controlling the proportion of disturbed samples in real features, solves the problem of the high complexity of the original algorithm in time and space operation, and avoids the problem of overfitting the model caused by the excessive number of features obtained by screening. The specific steps to perform feature selection are as follows.

Input: new feature dataset D, feature set $F = \{f_i \mid i = 1, 2, \cdots, t\}$.

Output: optimal set of features F_{b}.

Step 1 Calculate the variance S_i^2 of each dimension in the new feature dataset D, arrange them in descending order, and delete the feature sets that are not in the range by setting the threshold p_1. In this study, p_1 is set to 0.3, and the filtered feature sets will be used as the primary dataset D_{new}.

Step 2 Assume that there are n_1 groups of samples and m_1 features in D_{new}, the dataset D_{new} is replicated and the shadow feature dataset D_{s} is constructed by randomly disrupting the number of samples according to $(n_1 p_2) m_1$. In this study, p_2 is set to 0.6.

Step 3 Merge dataset D_{new} with dataset D_{s} to obtain an extended feature dataset D_{w} with n_1 rows and $2m_1$ columns, run a random forest classifier on D_{w}, and calculate Z_{score} (the measure for judging feature importance, i.e., feature importance score) for real feature and shadow feature.

Z_{score} is calculated as[29]

$$Z_{\mathrm{score}} = \frac{Z_i - \overline{Z}}{d_{\mathrm{st}}(Z)}, \tag{3}$$

where Z_i is the feature Z_{score}; $\overline{Z}$ is the mean of the feature set Z_{score}; $d_{\mathrm{st}}(Z)$ is the standard deviation of the feature set Z_{score}.

Step 4 Mark the maximum Z_{score} of the shadow feature set as $Z_{\max}$, the features in the real feature set with Z_{score} greater than $Z_{\max}$ are marked as important, and less than $Z_{\max}$ are marked as unimportant and deleted.

Step 5 Repeat steps 3 – 5 above until the new feature set D is filtered and encapsulated.

Step 6 Return the optimal feature set F_{b}.

1.3 Feature engineering complexity analysis

In the feature derivation stage of this study, limited new features are generated on a specific number of feature sets, and time-consuming operations like sorting are not involved, so the time complexity of this stage is $O(1)$. Suppose that the feature dimension of the derived new feature set is m_2 and the number of samples is n_2. In the

variance filtering stage, there is a descending sorting operation for the feature variance S_i^2 of each dimension, and the average time complexity of quick sorting is $O(m_2\log(m_2))$. In the encapsulation stage, the running time complexity of the original traditional Boruta algorithm is $O((n_2!)^{m_2})$ [30]. In this study, the improved Boruta only disrupts the sample number of (m_2p_2) n_2 to build the shadow feature, so the time complexity of this stage is $O(((n_2p_2)!)^{m_2})$. The maximum time complexity of the feature engineering stage of this study is

$$O(1) + O(m_2(\log(m_2))) + O(((n_2p_2)!)m_2) \approx O(m_2^2). \quad (4)$$

In the feature engineering stage, the time complexity of the algorithm proposed in this study is linearly related to the square of the feature dimension. This algorithm can process the feature information in the original dataset faster, perform new feature derivation and optimal feature set selection, and mine the hidden information in the dataset with reduced time overhead.

2 Prediction Model Construction

2.1 Stacking ensemble model building solution

The stacking ensemble learning model can take advantage of each different base learner to avoid the problem of poor generalization of the unitary learner and thus improve the prediction accuracy[31-32]. In this study, we propose a stacking solution ensemble prediction model structure: random forest (RF), gradient boosting decision tree (GBDT) and support vector regression (SVR) as basic learners and linear regression (LR) as meta-learner. Both RF and GBDT consist of multiple regression decision trees, and their final prediction results are jointly determined by multiple decision trees in different learning strategy states, which have higher accuracy compared with other classical regression models. SVR only uses part of the support vector to make hyper-plane decisions, which does not need to rely on all the data and is more inclusive in terms of sample size. Since different learners can perform feature learning prediction from different spatial perspectives to a certain extent, the differences between different models can be complemented, so the overall prediction accuracy and generalization can be improved by stacking the ensemble.

RF consists of CART-type regression decision trees as weak learners, and the final output of the model is determined jointly by the individual weak learners in the forest that are not related to each other[33]. In this study, multiple CART-type regression decision trees are constructed, and the prediction results of all decision trees are averaged as the prediction results of the random forest by applying the Bagging integration strategy.

GBDT regression model, like RF, also consists of CART-type regression decision trees as weak learners, except that GBDT uses the Boosting integration strategy, and the final output of the model is determined by the continuous accumulation of the prediction results of each weak learner in the forest[34].

SVR, when applied to regression problems, maps the original feature vector to a high-dimensional feature space using a kernel function making it linearly regressible[35]. In this study, the support vector machine based learner is determined by finding the optimal hyper-plane corresponding to the set of features.

LR is trained to fit corresponding weight values to each dimensional feature, and the sample values are multiplied by the weight values plus a bias value, and all dimensional features are computed and then summed to the linear regression model. The multiple linear regression meta-learner studied in this paper has the prediction outputs of RF, GBDT and SVR as inputs and the prediction results of this stacking ensemble model as outputs.

2.2 Stacking ensemble model building process

The schematic diagram of the construction of the stacking ensemble model is shown in Fig. 2. The stacking ensemble model construction proposed in this study has four main steps.

Step 1 The feature engineered dataset is extracted by stratified sampling, with 80% of the dataset being used for model training and 20% for model testing.

Step 2 To prevent the occurrence of overfitting, each basic learner of RF, GBDT and SVR is trained in a five-fold cross-validation way on the training set, and the base learner hyper-parameters are optimized by using GridSearchCV, which limits a set of values with a range for each hyper-parameter to be tuned by different learners according to the characteristics of the dataset. The values are determined in steps for each round of change. The grid search will iterate through each set of possible hyper-parameters to try each combination and find the best hyper-parameter combination for each combination in training and cross-validation, so that each learner can achieve the best performance.

Step 3 The predicted output of the base learner from the previous step is used as the input features of the LR meta-learner for training. The training prediction of a single base learner may suffer from overfitting or unstable prediction bias, so the model prediction capability is further enhanced by balancing the disadvantages and advantages of each base learner with a simple structured LR meta-learner model.

Step 4 The stacking ensemble model is output and tested.

Algorithm 1 shows the specific steps of the whole stacking algorithm model implementation.

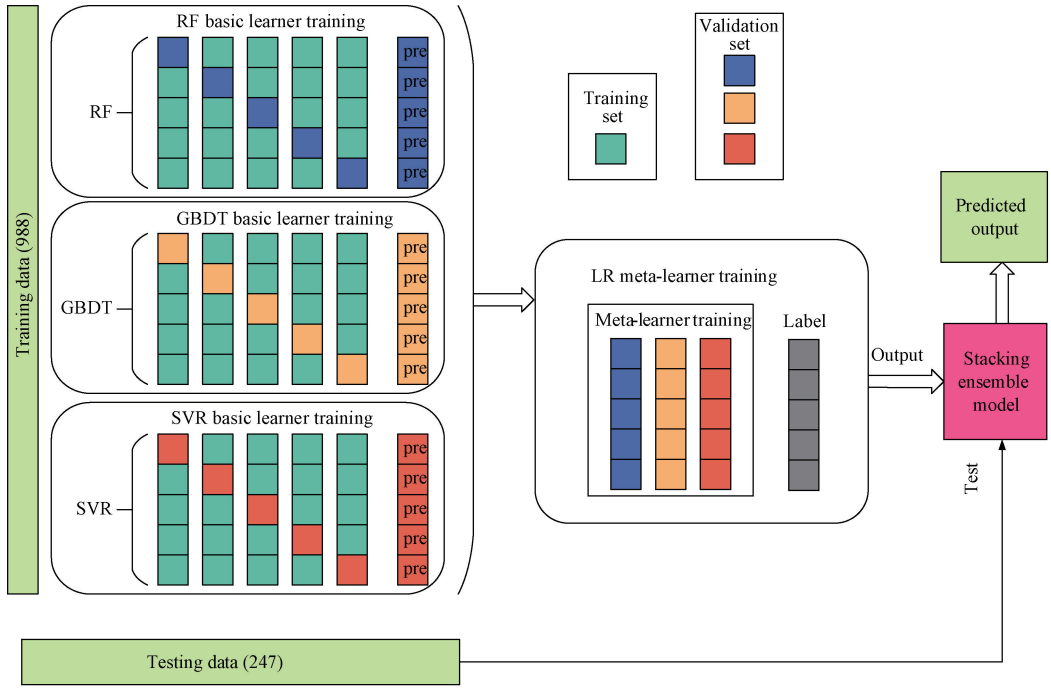


Fig. 2 Schematic diagram of stacking ensemble model construction

Algorithm 1 Stacking ensemble model

Input: training data $D_b = \{x_i, y_i\}$, $i \in \{1, 2, \cdots, n\}$; B_l contains RF, GBDT and SVR; M contains LR

Output: prediction model

1. //Step 1: data set division
2. 80% for training;
3. 20% for testing;
4. //Step 2: basic learners training
5. **for** $l = 1$ to 3 **do**
6. Learn B_l based on D_b
7. $B_l(x_i) \rightarrow x_i'$;
8. Hyper-parameters optimization;
9. **end for**
10. //Step 3: meta-leaner training
11. **for** $i = 1$ to n **do**
12. $D_m = \{x_i', y_i\}$, where $x_i' = \{B_1(x_i), B_2(x_i), B_3(x_i)\}$;
13. **end for**
14. Learn M based on D_m;
15. Hyper-parameters optimization;
16. Retun prediction model.

2.3 Predictive model complexity analysis

Suppose that the number of features is m_3 and the number of samples is n_3, and the number of CART trees in the base learner RF is k_1 and the number of CART trees in the base learner GBDT is k_2, the training time complexity of RF is $O(k_1 m_3 n_3 \log(n_3))$, that of GBDT is $O(k_2 m_3 n \log(n_3))$ and that of SVR is $O({n_3}^2)$. For the meta-learner LR, the number of features is 3 and the number of samples is n, so the training time complexity is $O(9n_3 + 3)$. Because the meta-learner is trained after the basic learner is trained separately in this study, the training mode of the whole ensemble model is in a string state, and the worst training time complexity of the whole model is

$$O(k_1 m_3 n_3 \log(n_3)) + O(k_2 m_3 n_3 \log(n_3)) + O({n_3}^2) + O(9n_3 + 3) \approx O({n_3}^2). \quad (5)$$

As can be seen from formula (5), the time complexity of model training is greatly influenced by SVR, and the final ensemble training of the model will not occupy too much consumption.

3 Results and Discussion

The full process of implementing the intelligent fitting of the OK lens in this study is shown in Fig. 3. The experimental environment is built based on Python 3.8, and the entire model building process is performed in the PyCaret open source package and Sklearn machine learning toolkit environment. The information of the device used for the experiment is Intel(R) Core(TM) i7-10870H CPU@2.20 GHz, Windows 10, 64-bit OS, and 16 G RAM. The experimental and analytical process is carried out in three parts: feature engineering results, system running time and predicted model performance.

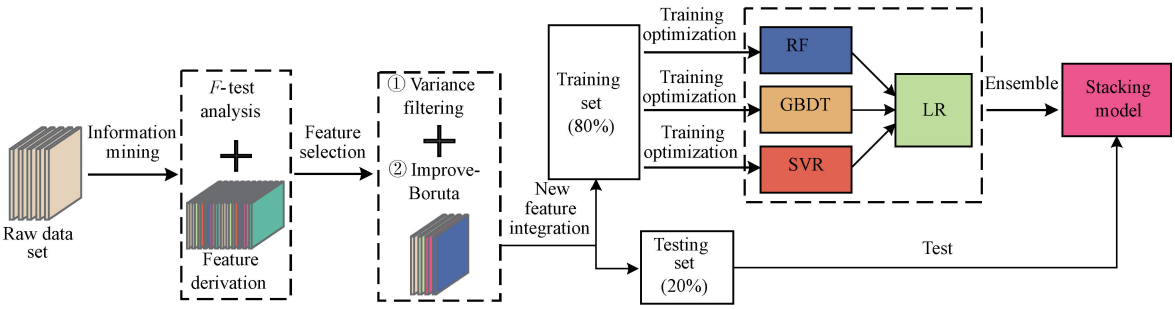


Fig. 3 Implementation process of intelligent fitting of OK lens

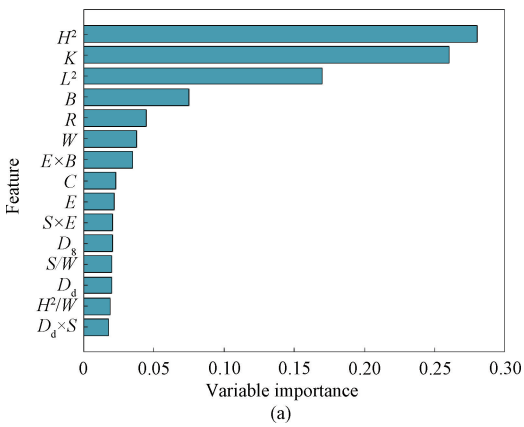
3.1 Feature engineering analysis

Before feature derivation, the results of the *F*-test for two different prediction tasks of the OK lens, AC and TP, are shown in Table 2. It can be learned that several feature dimensions such as FK, SK, AL and WTW set higher derived weights in the prediction study for AC. In the prediction study for TP, several feature dimensions such as SER, AL, Se and dK set higher derived weights.

Table 2 *F*-test analysis results

Prediction task	S	A	SD8/μm	FK/D	SK/D	dK/D	Fe
AC	56.0	0.9	0.1	819.0	777.0	6.54	9.6
TP	5.9	9.8	0.7	0.6	0.6	26.00	5.9
Prediction task	Se	AL/mm	ACD/mm	WTW/mm	SER/D	CYL/D	V
AC	0.2	623	0.86	167.0	7.7	0.9	1.8
TP	25.0	259	5.50	9.9	3 730.0	10.0	7.0

Hundred-dimensional new features were randomly derived according to the set weights, which were features selected by variance-improved Boruta feature selection, and Fig. 4 shows the set of features constructed for two different prediction tasks, AC and TP, and the importance factors corresponding to each dimensional feature. Among the constructed predicted AC feature dimensions, H^2, L^2, $E\times B$, $K\times E$, K/W, F/W and $D_d\times K$ are used as the derived new feature dimensions, and the original feature dimensions such as SK, Fe, SER, WTW, ACD, Se and dK are retained. Among the constructed predicted TP feature dimensions, R^2, $R^2\times R$, $R\times R^3$ and R^2/A are used as the derived new feature dimensions, and the original feature dimensions of SER, AL, ACD, A, FK and CYL are retained.



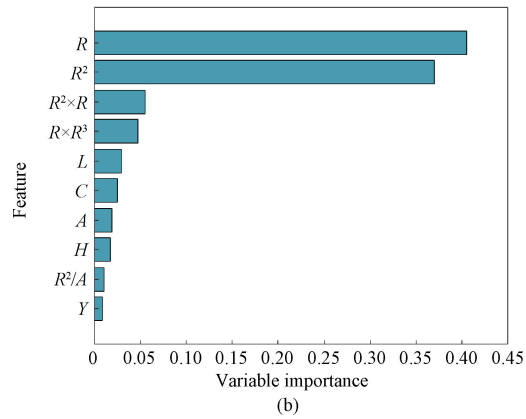


Fig. 4 Best feature set and corresponding importance factors of two tasks: (a) AC prediction; (b) TP prediction

3.2 System running time analysis

The main time-consuming information under feature engineering in this study is shown in Table 3. It can be seen that the variance-improved Boruta feature selection algorithm proposed in this study has a significant improvement in running speed compared to the traditional Boruta algorithm, which indicates that the algorithm proposed in this study reduces the computational time overhead of information mining.

The training optimization and final prediction running time consumption metrics of each learner in the stacking ensemble model are shown in Table 4. It can be seen that in the training optimization of the ensemble

model for prediction AC, the training optimization of the three base learners spends about 98% of the total elapsed time, and the ensemble time of the model only spends about 2%; in the training optimization of the ensemble model for prediction TP, the ensemble time of the model only spends less than 3%. Therefore, it can be shown that the stacking ensemble prediction model proposed in this study does not increase the time complexity of the whole system based on the ensemble. The running time of the stacking ensemble model for prediction AC and prediction TP is only 1.69 s and 1.52 s, respectively, verifying that it meets the time requirements for clinical applications.

Table 3 Feature engineering time consuming results

Prediction task	Algorithm	Running time/s
AC	Boruta	9.6
	Variance-improved Boruta	5.4
TP	Boruta	9.0
	Variance-improved Boruta	4.7

Table 4 Model training optimization time consumption results

Prediction task	Model	Training optimization time/min	Prediction running time/s
AC	RF	31.3	0.63
	GBDT	29.7	0.56
	SVR	45.5	0.31
	LR	2.2	0.19
	Stacking	108.7	1.69
TP	RF	17.8	0.54
	GBDT	15.5	0.55
	SVR	35.4	0.28
	LR	1.8	0.15
	Stacking	70.5	1.52

3.3 Model performance analysis

For the evaluation metrics of machine learning regression models, R-square (R^2) indicates the difference between the regression predicted value and the true value in the range of $(-\infty, 1]$. Mean squared error E_{MS} is used to evaluate the degree of variation of the data. Root mean squared error E_{RMS} is the arithmetic square root of E_{MS}, i.e., the standard deviation. Mean absolute error E_{MA} is the mean of the absolute error between the predicted value and the true value. The formulas are shown below[36-37]:

$$E_{MA} = \frac{1}{n}\sum_{i=1}^{n} |f_i - y_i|, \tag{6}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (f_i - y_i)^2}{\sum_{i=1}^{n} (\bar{y}_i - y_i)^2}, \tag{7}$$

$$E_{MS} = \frac{1}{n}\sum_{i=1}^{n} (f_i - y_i)^2, \tag{8}$$

$$E_{RMS} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (f_i - y_i)^2}, \tag{9}$$

where f_i is the predicted value of the model; y_i is the true value; $\bar{y}_i$ is the average of true values of test samples; n is the number of test samples.

The stacking ensemble prediction model after feature engineering information mining in this study is compared in accuracy with the models trained under the original dataset and the original machine learning model. The results of the training tests for AC and TP under the original dataset and the original machine learning model are indicated in Fig. 5. For the tests on AC, RF performed the best with R^2 of 0.70, E_{RMS} of 0.74, E_{MS} of 0.56 and E_{MA} of 0.63; for the tests on TP, SVR performed the best with R^2 of 0.72, E_{RMS} of 0.65, E_{MS} of 0.42 and E_{MA} of 0.44. The test results of the basic learner and stacking ensemble learning model in this study are shown in Fig. 6. In the tests for both AC and TP, the stacking ensemble model performs best, with its R^2 improving to 0.84 and 0.88, E_{RMS} decreasing to 0.50 and 0.36, E_{MS} decreasing to 0.25 and 0.13, and E_{MA} decreasing to 0.34 and 0.18 relative to the original machine learning model.

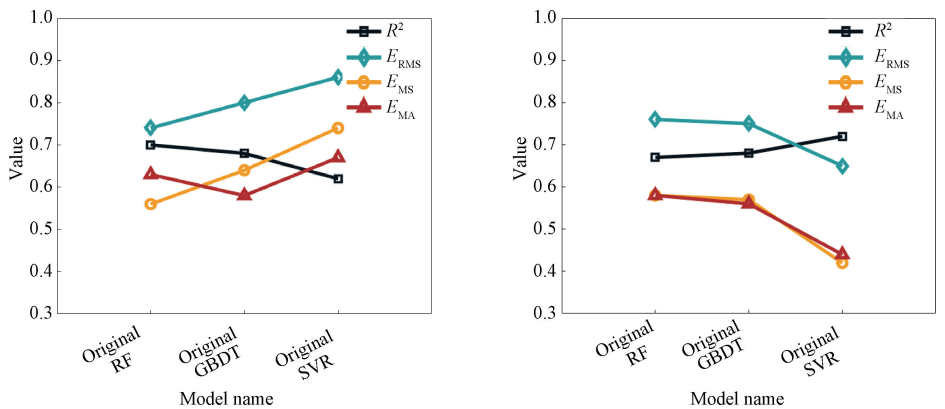


Fig. 5 Test results under original dataset and original machine learning model: (a) AC test results; (b) TP test results

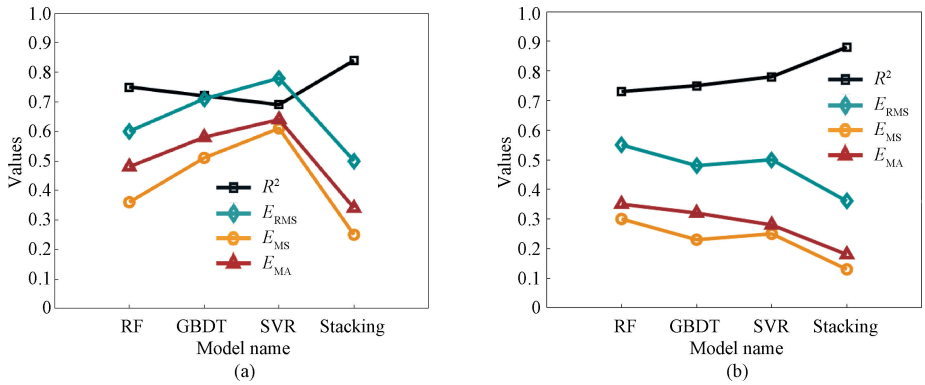


Fig. 6 Test results of this research model: (a) AC test results; (b) TP test results

To further evaluate the performance of the prediction model, the proposed model was validated for clinical application in this study. The AC and TP values were predicted by the stacking ensemble prediction model for 200 patients with clinical diagnostic optometric data, and then compared with the final sizing parameters. The Bland-Altman test was performed to test the consistency of the stacking ensemble prediction method and the clinical diagnosis method. The results of the Bland-Altman test are shown in Fig. 7. The results show that the fixation parameters obtained from the stacking ensemble prediction method and the clinical diagnosis method have a high similarity. The horizontal coordinate of Fig. 7 shows the mean of the clinical diagnostic values mean AC and mean TP and the model prediction values AC-P and TP-P, and the vertical coordinate shows the difference between the clinical diagnostic values and the model prediction values. The upper and lower dotted lines in Fig. 7. indicate the upper and lower limits of the 95% consistency boundary between them; the middle solid line indicates the mean value of the difference, which clearly shows that the majority of the predicted data values are within the 95% agreement confidence interval.

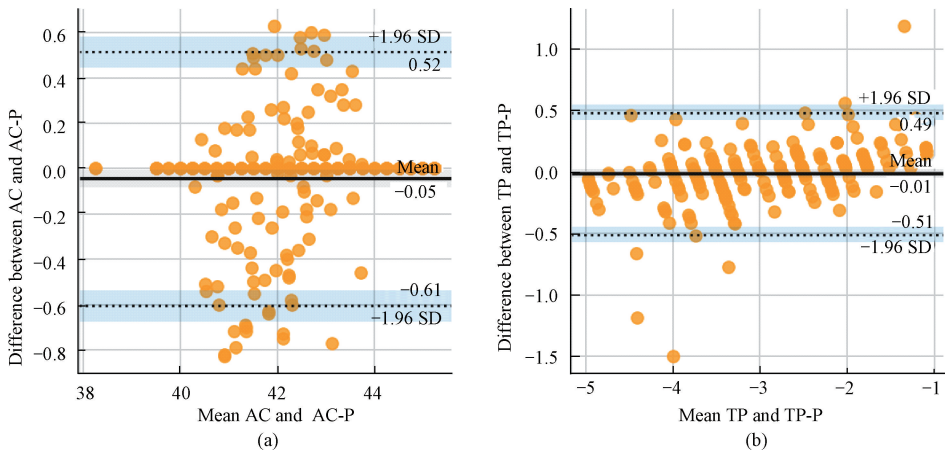


Fig. 7 Bland-Altman test results: (a) AC test results; (b) TP test results

Pearson's correlation coefficient is a measure of the linear relationship between two datasets[38-39]. Strictly speaking, Pearson correlation requires each dataset to be normally distributed. A correlation of −1 or +1 implies a perfect negative and positive linear relationship, and 0 indicates no correlation. It is calculated as

$$r_{xy} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2}\sqrt{\sum_i (y_i - \bar{y})^2}} = \frac{\mathrm{cov}(x, y)}{\sigma_x \sigma_y}, \quad (10)$$

where r_{xy} represents Pearson's correlation coefficient; x is the predicted value and $\bar{x}$ is the average value of all predicted values; y is the real value and $\bar{y}$ is the average of all real values; $\mathrm{cov}(x, y)$ is the covariance between the two variables; σ_x and σ_y are the internal variances of the two data dimensions of the stacking ensemble predictive value and the clinical diagnostic true value, respectively. The correlation evaluation indexes of the lens parameters obtained by the stacking ensemble predictive method and the clinical diagnostic method are shown in Table 5. The Bland-Altman diagram is used to calculate the 95% consistency confidence interval (CI95%) of the difference between the two methods. A two-tailed $P < 0.05$ was considered statistically significant. The results show that the lens parameters obtained by the two methods are highly consistent and similar.

Table 5 Evaluation index of correlation between predicted value and real value

Prediction task	Number	Pearson's correlation coefficient	CI95%	*P*
AC	200	0.971	[0.96, 0.98]	3.52×10^{-126}
TP	200	0.977	[0.97, 0.98]	5.33×10^{-136}

4 Conclusions

In this study, a stacking ensemble prediction model is proposed to achieve clinically intelligent diagnosis and fitting of the OK lens. Through feature derivation based on *F*-test and variance-improved Boruta feature selection, the hidden information in the original dataset is mined, which lays a foundation for improving the generalization of the model in the future. An ensemble learning model with RF, GBDT and SVR as basic learners and LR as meta-learner is constructed by stacking strategy, and each learner is optimized by hyper-parameter, which gives full play to the advantages of different models in parameter prediction. The experimental results show that the intelligent prediction model in this study has better accuracy than the original machine learning model. In clinical application, the Pearson's correlation coefficient of AC and TP with the final calibration parameters reaches 0.971 and 0.977, respectively, and the intelligent prediction running time is only 1.69 s and 1.52 s. The prediction model proposed in this study still has some errors in predicting the parameters of individual patients, and its generalization needs to be improved. Therefore, building a prediction model for unstable and small sample data will be the next research direction.

References

[1] LOUGHEED T. Myopia: the evidence for environmental factors [J]. *Environmental Health Perspectives*, 2014, 122(1): 47-63.

[2] MRUGACZ M, GAJECKA M, MRUKWA-KOMINEK E, et al. Myopia: risk factors, disease mechanisms, diagnostic modalities, and therapeutic options 2019 [J]. *Journal of Ophthalmology*, 2020, 2020: 1-2.

[3] LEE S S Y, LINGHAM G, SANFILIPPO P G, et al. Incidence and progression of myopia in early adulthood [J]. *JAMA Ophthalmology*, 2022, 140(2): 162-169.

[4] GAO H, MILES T P, TROCHE R, et al. Quality of vision following LASIK and PRK-MMC for treatment of myopia [J]. *Military Medicine*, 2022, 187(9/10): e1051-e1058.

[5] NUCCI P, DRACK A V. Refractive surgery for unilateral high myopia in children[J]. *Journal of American Association for Pediatric Ophthalmology and Strabismus*, 2001, 5(6): 348-351.

[6] KERSHNER R M. Clear corneal cataract surgery and the correction of myopia, hyperopia, and astigmatism [J]. *Ophthalmology*, 1997, 104(3): 381-389.

[7] PREPAS S B. Light, literacy and the absence of ultraviolet radiation in the development of myopia [J]. *Medical Hypotheses*, 2008, 70(3): 635-637.

[8] LIN H J, WEI C C, CHANG C Y, et al. Role of chronic inflammation in myopia progression: clinical evidence and experimental validation[J]. *EBioMedicine*, 2016, 10: 269-281.

[9] NICKLA D L, ZHU X Y, WALLMAN J. Effects of muscarinic agents on chick choroids in intact eyes and eyecups: evidence for a muscarinic mechanism in choroidal thinning[J]. *Ophthalmic and Physiological Optics*, 2013, 33(3): 245-256.

[10] GONG Q W, JANOWSKI M, LUO M, et al. Efficacy and adverse effects of atropine in childhood myopia: a meta-analysis [J]. *JAMA Ophthalmology*, 2017, 135(6): 624-630.

[11] CHO P, CHEUNG S W, EDWARDS M. The longitudinal orthokeratology research in children (LORIC) in Hong Kong: a pilot study on refractive changes and myopic control [J]. *Current Eye Research*, 2005, 30(1): 71-80.

[12] CHO P, CHEUNG S W. Retardation of myopia in orthokeratology (ROMIO) study: a 2-year randomized clinical trial [J]. *Investigative Ophthalmology & Visual Science*, 2012, 53(11): 7077-7085.

[13] WALLINE J J, JONES L A, SINNOTT L T. Corneal reshaping and myopia progression [J]. *The British Journal of Ophthalmology*, 2009, 93(9): 1181-1185.

[14] MORGAN P B, EFRON N, WOODS C A, et al. International survey of orthokeratology contact lens fitting [J]. *Contact Lens and Anterior Eye*, 2019, 42(4): 450-454.

[15] FU A C, QIN J, RONG J B, et al. Effects of orthokeratology lens on axial length elongation in unilateral myopia and bilateral myopia with anisometropia children [J]. *Contact Lens and Anterior Eye*, 2020, 43(1): 73-77.

[16] WONG T Y, FERREIRA A, HUGHES R, et al. Epidemiology and disease burden of pathologic myopia and myopic choroidal neovascularization: an evidence-based systematic review [J]. *American Journal of Ophthalmology*, 2014, 157(1): 9-25.

[17] ROSUERO MLG, BOMBIN AA, ROMERO R,

et al. Clinical tool to measure fluorescein patterns in orthokeratology[J]. *PeerJ*. 2022, 10: 1-15.
[18] RAJABI M T, HOSSEINI S S, GHORBANI Z, et al. Utility of orthokeratology contact lenses; efficacy of myopia correction and level of patient satisfaction in Iranian myopic/myope-astigmatic patients[J]. *Journal of Current Ophthalmology*, 2015, 27(3/4): 99-102.
[19] MANICKAM P, MARIAPPAN S A, MURUGESAN S M, et al. Artificial intelligence (AI) and internet of medical things (IoMT) assisted biomedical systems for intelligent healthcare[J]. *Biosensors*, 2022, 12(8): 562.
[20] HOGARTY D T, MACKEY D A, HEWITT A W. Current state and future prospects of artificial intelligence in ophthalmology: a review [J]. *Clinical & Experimental Ophthalmology*, 2019, 47(1): 128-139.
[21] ABDULLAH T A A, ZAHID M S M, ALI W. A review of interpretable ML in healthcare: taxonomy, applications, challenges, and future directions[J]. *Symmetry*, 2021, 13(12): 2439.
[22] FAN Y Z, YU Z K, TANG T, et al. Machine learning algorithm improves accuracy of ortho-K lens fitting in vision shaping treatment [J]. *Contact Lens and Anterior Eye*, 2022, 45(3): 101474.
[23] ZHANG Q T, XIE P Y, YANG L N, et al. A machine learning model on orthokeratology lens fitting based on the data of optometry examination [J]. *Chinese Journal of Ophthalmology*, 2019, 55(2): 105-110. (in Chinese)
[24] ZHANG L, ZHANG Y, LIU Y L, et al. A novel fitting algorithm for alignment curve radius estimation using corneal elevation data in orthokeratology lens trial[J]. *Contact Lens and Anterior Eye*, 2017, 40(6): 401-407.
[25] SANDERSON E, WINDMEIJER F. A weak instrument *F*-test in linear IV models with multiple endogenous variables [J]. *Journal of Econometrics*, 2016, 190(2): 212-221.
[26] ŠAŠIĆ S, VERIOTTI T, KOTECKI T, et al. Comparing the predictions by NIR spectroscopy based multivariate models for distillation fractions of crude oils by *F*-test[J]. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2023, 286: 122023.
[27] HE H Y, HUANG G Y, ZHANG B, et al. Research on boruta-ET-based anomalous traffic detection model[J]. *Security and Communication Networks*, 2022, 2022: 1-8.
[28] HEIDARI A A, AKHOONDZADEH M, CHEN H. A wavelet PM2.5 prediction system using optimized kernel rxtreme learning with Boruta-XGBoost feature selection [J]. *Mathematics*, 2022, 10(19): 3566.
[29] DEVORE G R. Computing the Z score and centiles for cross-sectional analysis: a practical approach [J]. *Journal of Ultrasound in Medicine*, 2017, 36(3): 459-473.
[30] DEMIR S, SAHIN E K. An investigation of feature selection methods for soil liquefaction prediction based on tree-based ensemble algorithms using AdaBoost, gradient boosting, and XGBoost [J]. *Neural Computing and Applications*, 2023, 35(4): 3173-3190.
[31] YIN F M, DU J, XU X Z, et al. Depression detection in speech using transformer and parallel convolutional neural networks [J]. *Electronics*, 2023, 12(2): 328.
[32] CHUNG D, YUN J, LEE J, et al. Predictive model of employee attrition based on stacking ensemble learning [J]. *Expert Systems with Applications*, 2023, 215: 119364.
[33] GOH V, CHOU Y J, LEE C C, et al. Predicting bacteremia among septic patients based on ED information by machine learning methods: a comparative study [J]. *Diagnostics*, 2022, 12(10): 2498.
[34] XIA M, WU Y Z, WANG Z J. Time delay identification in dynamical systems based on interpretable machine learning [J]. *Journal of Donghua University (English Edition)*, 2022, 39(4): 332-339.
[35] WU J H, WANG X K, WANG J L, et al. Method of soft-sensor modeling for fermentation process based on geometric support vector regression [J]. *Journal of Donghua University (English Edition)*, 2013, 30(1): 1-6.
[36] ELHADIDY A A, EL-BADAWY S M, ELBELTAGI E E. A simplified pavement condition index regression model for pavement evaluation[J]. *International Journal of Pavement Engineering*, 2021, 22(5): 643-652.
[37] THELWALL M, WILSON P. Regression for citation data: an evaluation of different methods [J]. *Journal of Informetrics*, 2014, 8(4): 963-971.
[38] PLEBANKIEWICZ E, GRACKI J. Analysis and prediction of universities' buildings' renovation costs using a regression model [J]. *Applied Sciences*, 2022, 13(1): 401.
[39] LI G Z, ZHANG A N, ZHANG Q Z, et al. Pearson correlation coefficient-based performance enhancement of broad learning system for stock price prediction [J]. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2022, 69(5): 2413-2417.

Stacking 集成学习应用于近视矫正中的角膜塑形镜临床验配

巩家铭[1,2]，李康妹[1]，胡　俊[2,3]，陈　浩[4]，曹倩倩[2,3*]，吴　戈[4*]

1. 东华大学 机械工程学院，上海 201620
2. 东华大学 人工智能研究院，上海 201620
3. 上海工业大数据工程技术研究中心，上海 201600
4. 温州医科大学附属眼视光医院，浙江 温州 325000

摘　要：针对角膜塑形（orthokeratology，OK）镜临床验配难度系数大和过程繁琐费力的问题，提出一种stacking 集成学习方法预测 OK 镜参数值，实现 OK 镜智能验配。通过基于 *F* 检验的特征衍生和基于方差-改进 Boruta 算法的特征选择，构建出最适合目标变量的特征集合。研究了以随机森林（random forest，RF）、梯度提升决策树（gradient boosting decision tree，GBDT）和支持向量回归（support vector regression，SVR）作为第一层基学习器，线性回归（linear regression，LR）作为第二层元学习器的 stacking 集成学习预测模型。实验结果表明模型预测结果和临床诊断结果高度一致，验证该模型可作为一种有效的辅助临床验配方法。

关键词：角膜塑形（OK）镜；特征工程；stacking 集成模型；参数预测；智能验配