International Journal of Trend in Research and Development, Volume 13(2), ISSN: 2394-9333
www.ijtrd.com

Research on Cognitive Function among Older Adults Based on Machine Learning

[1]Chang Chunyi and [2]Wang Jingyi,
[1,2]School of Statistics, Beijing Wuzi University, Beijing, China

***Abstract*—**With the continuous acceleration of population aging, the size of the population aged 60 and above in China has continued to expand by the end of 2024, and the proportion of elderly people has been steadily increasing. In this context, cognitive decline has gradually become an important factor affecting the quality of life and social participation of older adults, while also imposing long-term pressure on family caregiving systems and the allocation of social medical resources. Therefore, systematically identifying the influencing factors of cognitive function among middle-aged and older adults, revealing their mechanisms of action, and improving risk identification capabilities are of great significance for actively responding to population aging and promoting healthy aging.

Secondly, in terms of cognitive impairment risk prediction, all machine learning models demonstrate satisfactory classification performance. Compared with traditional models such as logistic regression and support vector machines, ensemble learning methods including random forest, XGBoost, and LightGBM show overall superior performance in terms of recall and discriminative ability. By integrating the prediction results of multiple base learners, the Stacking ensemble model further improves recall while maintaining a stable AUC, achieving a more balanced overall performance. These results indicate that the multi-model ensemble strategy can enhance the screening effectiveness for high-risk populations while balancing risk identification capability and model stability.

Keywords*—Older Adults; Cognitive Function; Machine Learning*

I. INTRODUCTION

A. Research Background

As the process of population aging continues to accelerate and the size of the elderly population steadily expands, aging has become a significant challenge for China's social development. Changes in the demographic structure have brought increasingly prominent health issues related to aging, among which cognitive decline has gradually emerged as a crucial factor affecting the quality of life, social participation ability, and independent living capacity of the elderly. Cognitive function encompasses the abilities of memory, attention, thinking, and language comprehension exhibited by individuals in the process of information acquisition, processing, and application, serving as an important foundation for maintaining normal social life. However, with advancing age, the cognitive function of the elderly generally shows a declining trend, with some individuals further developing into cognitive impairment, placing considerable pressure on individuals, families, and the social healthcare system. Therefore, systematically identifying the influencing factors of cognitive function in the elderly, revealing its mechanisms of action, and enhancing risk identification capabilities are of great significance for promoting healthy aging.

B. Research Objectives and Significance

The decline in cognitive function among the elderly not only affects their quality of life and independence but also imposes heavy burdens on families and society. Research on the cognitive function of the elderly holds not only academic value but also practical significance. In recent years, the state and society have continuously introduced relevant policies aimed at advancing healthy aging strategies and improving the quality of life for the elderly. Against this backdrop, academia has gradually increased its focus on cognitive function. By constructing effective risk prediction models, rapid preliminary screening and early warning of cognitive impairment risk for the elderly can be achieved in community and family settings, gaining valuable time for timely intervention. Given the current reality of China's increasingly deepening population aging, it is urgent to adhere to a problem-oriented approach, closely follow the principal social contradictions, and conduct in-depth explorations of the cognitive function of the elderly from a multi-dimensional perspective. This not only helps provide theoretical support for public policies in an aging society but also offers a scientific basis for improving the quality of life of the elderly, carrying significant importance.

C. Research Status

In recent years, machine learning technology, with its ability to handle high-dimensional data and uncover complex nonlinear relationships, has been widely applied in the field of cognitive impairment risk prediction. Scholars both domestically and internationally have conducted extensive methodological explorations based on different data sources and algorithmic frameworks. Cho et al. (2025), based on data from the South Korean Chronic Cerebrovascular Disease and Alzheimer's Disease Biobank, identified single nucleotide polymorphism sites associated with dementia through genome-wide association analysis. They constructed risk prediction models for the transition from mild cognitive impairment to dementia using six machine learning algorithms, including Random Forest, K-Nearest Neighbors, Artificial Neural Networks, Support Vector Machines, XGBoost, and LightGBM, providing a methodological reference for genomics-based personalized cognitive risk assessment. Zhang Lei et al. (2025) focused on hospitalized elderly patients with multimorbidity, comprehensively applying nine machine learning algorithms to construct cognitive impairment prediction models. Through feature importance evaluation, they identified key decision factors affecting cognitive function, including age, number of comorbidities, education level, and cerebrovascular diseases. Yang Jin (2025), targeting patients with chronic heart failure, screened variables using LASSO regression and constructed cognitive impairment prediction models based on four algorithms: Logistic Regression, XGBoost, Support Vector Machines, and Artificial Neural Networks. Additionally, a web-based calculator was developed to enable visualized prediction of cognitive function risk. Qin Xinyi (2023),

utilizing data from the Survey of Health, Aging and Retirement in Europe, addressed data imbalance issues by comparing different sampling methods. Six machine learning models were constructed, including Random Forest, XGBoost, and a two-layer Stacking ensemble model. By comparing model performance under different feature selection and sampling methods, the optimal prediction model for cognitive impairment was identified. Zhang Hengchuan (2024), based on longitudinal data from the Chinese Longitudinal Healthy Longevity Survey, employed four feature selection algorithms to screen predictive factors, with a focus on constructing a two-layer Stacking ensemble learning model. The first layer of base models included Logistic Regression, Linear Discriminant Analysis, K-Nearest Neighbors, and Naive Bayes, while the second layer meta-model adopted Logistic Regression for integration. External validation was conducted to assess the model's generalization predictive ability for cognitive impairment in the elderly.

II. RESEARCH DESIGN AND QUESTIONNAIRE PROCESSING

A. *Data Source and Processing*

The data for this study originates from the survey data published in 2023 by the China Health and Retirement Longitudinal Study (CHARLS) project, implemented by the National School of Development at Peking University. This survey collects multi-dimensional micro-level data on Chinese middle-aged and elderly households and individuals, covering various aspects such as demographic and economic background, health status, medical insurance, and retirement and pension, possessing high national representativeness and academic authority. This study focuses on elderly respondents aged 60 and above from this project. Data cleaning was completed using Stata 17.0. To ensure the reliability of the research results, listwise deletion was applied to handle missing values, meaning cases with missing values on key variables were removed. After processing missing values and outliers, this study further excluded observations with obvious logical errors in key variables, ultimately determining 5,797 samples as valid analytical samples for subsequent empirical analysis.

B. Data Source and Processing

This paper primarily analyzes the influencing factors of cognitive ability among the elderly, using the cognitive ability score as the dependent variable. Referring to existing studies, it focuses on examining the effects of five categories of independent variables: lifestyle habits, health status, internet participation, socioeconomic status, and educational attainment. Among these, lifestyle habits, health status, and internet participation are measured through three observed variables each; socioeconomic status is measured through four observed variables. The selection and specification of specific variables are as follows:

Table 1: Indicator Construction

Latent Variable	Observed Variable
Lifestyle Habits	Alcohol Consumption
	Exercise
	Sleep Score
Health Status	Chronic Diseases
	Self-rated Memory
	Headache
Internet Participation	Internet Participation Purpose Score
	Electronic Device Ownership Score
	Specific Internet Use Score
Socioeconomic Status	Social Security Level
	Medical Insurance Leve
	Total Household Income
	Retirement
Socio demographic Characteristics	Gender
	Age
	General Residence
	Education Level

III. MACHINE LEARNING-BASED RISK PREDICTION OF COGNITIVE IMPAIRMENT

A. Model Specification Methods and Data Processing

1. Binary Classification Processing of Cognitive Impairment Risk Data

Considering the complex pathways influencing cognitive function and the significant nonlinear relationships among variables, this study transforms the cognitive function prediction problem from a continuous regression task into a binary classification task, namely, the prediction of cognitive impairment risk. Regarding variable definition, referring to common statistical processing methods used in previous studies, the discrimination threshold is set as the mean of the overall cognitive function score minus one standard deviation: individuals with cognitive function scores below this threshold are classified as having cognitive impairment, whereas those with scores above the threshold are classified as having no cognitive impairment.

In the machine learning predictive analysis, unlike the latent variable-based modeling approach used in structural equation modeling, this study employs observed variables for direct modeling. Based on the 5,797 valid samples obtained after data cleaning described earlier, 17 feature variables closely related to cognitive function are selected as model inputs. These include alcohol consumption, sleep, exercise, chronic diseases, self-rated memory, headache, total household income, retirement status, pension insurance, medical insurance, educational attainment, age, gender, and urban/rural residence category. On this basis, the dataset is divided into a training set and a test set, with the training set containing 4,637 samples and the test set containing 1,160 samples, providing a complete feature matrix for subsequent model training and prediction.

2. Handling of Data Imbalance

In cognitive impairment risk prediction, there is a significant imbalance in the distribution of sample categories. Based on the binary classification criteria mentioned above, there are 1,068 individuals with cognitive impairment in the sample, accounting for 18.42% of the total sample, while 4,729 individuals have no cognitive impairment, accounting for 81.58%, exhibiting typical characteristics of class imbalance. Under these circumstances, if model training is conducted directly, classification models tend to favor the majority class (no-risk samples) due to their larger number, thereby weakening the ability to identify high-risk individuals with cognitive impairment, manifested as prediction bias for the minority

class or decreased sensitivity. Given that the focus of this study is on effectively identifying populations at risk of cognitive impairment, it is necessary to address the sample imbalance issue before model training.

To this end, this study employs the Adaptive Synthetic Sampling (ADASYN) method to process the training data. ADASYN is an oversampling technique developed based on the SMOTE method. Its core idea is to adaptively generate different numbers of synthetic samples according to the distribution characteristics and classification difficulty of minority class samples in the feature space, thereby more accurately characterizing the true distribution of minority class samples.

3. Model Parameter Selection and Tuning

In machine learning research, when comparing the predictive performance of different classification models, the key lies in obtaining a reasonable and stable estimate of the model's generalization error. Since model training is based on a finite sample, and the sample partitioning involves a certain degree of randomness, if performance evaluation is conducted using only a single training set and test set split, the results are often susceptible to the influence of the data partitioning method, making it difficult to comprehensively reflect the model's generalization ability. Five-fold cross-validation, by repeatedly partitioning the sample into training and validation sets, allows the model's predictive performance to be evaluated on different subsets, thereby effectively reducing the impact of randomness in sample partitioning. It is a commonly used and relatively robust method for model performance evaluation and comparison. By synthesizing the results of five validations, a more stable and reliable estimate of the model's predictive performance can be obtained, providing a basis for performance comparison among different models.

In the process of model parameter selection and performance evaluation, this study adopts a strategy combining grid search with five-fold cross-validation to optimize the models. Specifically, the grid search method constructs parameter combinations within a pre-set range of hyperparameters and trains models under different parameter configurations. Simultaneously, the predictive performance of the model under each parameter configuration is evaluated through five-fold cross-validation, with the average of the five validation results serving as the performance evaluation metric for that parameter combination. By systematically traversing different hyperparameter combinations and integrating the cross-validation results, the parameter configuration with the optimal predictive performance is ultimately selected for model training and comparative analysis. This method reduces the influence of empirical parameter selection, ensures the reproducibility of the parameter selection process, and enhances the stability of model performance evaluation, providing methodological support for subsequent model predictive performance comparison and result analysis.

B. Model Evaluation Metrics

In the context of cognitive impairment risk prediction, the research objective is not merely to pursue overall classification accuracy, but rather to focus more on the model's ability to identify high-risk individuals. Due to the typically low proportion of cognitive impairment samples in the general population, the data exhibits significant class imbalance characteristics. If only overall metrics such as accuracy are used, even if the model performs well on the majority class samples, it may mask its insufficient ability to identify high-risk populations. Therefore, this study, starting from the two dimensions of minority class identification performance and the model's overall discriminative ability, selects Recall, F_2 -score, and the Area Under the Receiver Operating Characteristic Curve (AUC) as the model performance evaluation metrics.

1. Confusion Matrix and Recall

In the cognitive impairment risk prediction task, the model's output results can be categorized into two types: predicting the presence of cognitive impairment risk and predicting the absence of cognitive impairment risk. By comparing the model's predictions with the true labels, a confusion matrix for the binary classification problem can be constructed, which can be represented as:

Table 2: Confusion Matrix Diagram

	Predicted No Risk	**Predicted Risk**
Actual No Risk	True Negative (TN)	False Positive (FP)
Actual Risk	False Negative (FN)	True Positive (TP)

Among these, TP represents the number of individuals correctly predicted by the model as being at high risk for cognitive impairment; FP represents the number of individuals without risk who are misclassified by the model as high risk; FN represents the number of individuals who actually have cognitive impairment risk but are missed by the model; and TN represents the number of individuals correctly predicted by the model as having no risk. The confusion matrix intuitively reflects the model's classification performance on individuals with and without risk and provides the foundation for calculating subsequent evaluation metrics.

Based on the confusion matrix, Recall can be defined to measure the model's ability to identify truly high-risk individuals. Its calculation formula is:

$$R\square call \quad = \frac{TP}{TP+FN} (1)$$

This metric reflects the proportion of individuals who actually have cognitive impairment risk that are successfully identified by the model. The higher the recall rate, the stronger the model's coverage of high-risk populations, and the lower the probability of high-risk individuals being missed. In the application scenario of cognitive impairment risk prediction, the research objective is not simply to pursue the overall consistency of prediction results, but to identify potential high-risk individuals as much as possible. Especially in situations where the sample class distribution is imbalanced, if the model is overly biased towards predicting the majority class, it tends to cause a decrease in recall rate, thereby increasing the risk of missing high-risk individuals. Therefore, this study takes recall as an important reference metric in model performance evaluation to highlight the model's practical application value in cognitive impairment high-risk screening tasks.

In binary classification problems, in addition to recall, precision is also one of the common evaluation metrics defined based on the confusion matrix. It is used to measure the proportion of individuals predicted by the model as high risk who are actually at high risk. Its calculation formula is:

$$Precision \quad = \frac{TP}{TP+FP} (2)$$

2. F-score

As indicated in the previous subsection, in the risk prediction application scenario studied in this paper, the focus of model performance evaluation is on identifying individuals with genuine risk as much as possible, that is, improving the model's recall rate for high-risk samples. In this context, relying solely on a single evaluation metric makes it difficult to comprehensively characterize the model's overall performance across different prediction outcomes. Therefore, it is necessary to introduce a composite metric that can reflect the trade-offs between different evaluation objectives. The F-score metric is a type of comprehensive evaluation indicator that assigns different weights to various evaluation objectives through a parameter, thereby reflecting the research focus in model performance assessment. Its general form is defined as:

$$F_\beta = \frac{(1+\beta^2)\times Recall \times Precision}{\beta^2 \times Precision + Recall}, \quad \beta > 0 \quad (3)$$

In cognitive impairment risk screening research, the practical application goal of the model is not merely to pursue consistency in prediction results across the overall sample, but to emphasize adequate coverage of potential high-risk populations. Based on this research background, this study sets β=2 to strengthen the metric's emphasis on recall rate, resulting in the F2 metric, thereby making the model performance evaluation more aligned with the actual needs of cognitive impairment risk prediction. Its expression is:

$$F_2 = \frac{5\times Recall \times Precision}{4\times Precision + Recall} \quad (4)$$

3. ROC and AUC

In binary classification prediction problems, relying solely on evaluation metrics based on a single threshold may not fully reflect the model's overall performance under different discrimination criteria. Therefore, the ROC curve is widely used to describe the overall discriminative ability of classification models under various threshold conditions. The ROC curve depicts the changing relationship between the true positive rate and the false positive rate at different thresholds, reflecting the model's ability to distinguish between the two classes of samples from an overall perspective. The ROC curve takes the False Positive Rate (FPR) as the x-axis and the True Positive Rate (TPR) as the y-axis. The true positive rate is used to measure the proportion of at-risk individuals correctly identified by the model, and its calculation formula is:

$$TPR = \frac{TP}{TP+FN} \quad (5)$$

The false positive rate is used to measure the proportion of individuals without risk who are incorrectly classified by the model as having risk. Its calculation formula is:

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP+TN}} \quad (6)$$

To quantitatively characterize the ROC curve, the Area Under the Curve (AUC) is commonly used as a comprehensive evaluation metric for the model's overall discriminative ability. AUC is defined as the area enclosed by the ROC curve and the horizontal axis, with a value range of [0, 1]. The larger the value, the stronger the model's ability to distinguish between individuals with and without risk. Compared to evaluation metrics that rely on a single classification threshold, AUC can comprehensively reflect model performance under different discrimination thresholds, offering better stability and comparability. It is particularly suitable for prediction tasks with imbalanced class distributions. Therefore, this study adopts AUC as the core metric for model performance evaluation, comparing and analyzing different models from the perspective of overall discriminative ability, and conducting comprehensive assessments in conjunction with other metrics.

B. Single Model Construction

1. Logistic Regression Model

Logistic Regression (LR) is a classical statistical learning method primarily used for solving binary classification problems. It is widely applied in fields such as medical statistics, public health, and health economics. In cognitive function research, Logistic Regression is often used to analyze the relationship between various individual characteristics and the risk of cognitive impairment, providing an important quantitative analysis tool for risk identification and early screening of cognitive dysfunction.

This model uses the LogisticRegression function from Python's scikit-learn library to classify and predict the risk of cognitive impairment. The key parameters of the model mainly include the regularization coefficient C, the type of regularization penalty (penalty), the maximum number of iterations (max_iter), the optimization algorithm (solver), and the class weight setting (class_weight). Here, C is the inverse of the regularization strength; a smaller value indicates stronger regularization constraints, which helps suppress model overfitting. The penalty parameter is used to set the regularization method. This paper adopts L1 regularization to enhance model sparsity and improve feature interpretability. Considering the requirements of the L1 regularization form for the solver, 'liblinear' is selected as the parameter optimization algorithm during model training. The max_iter parameter limits the maximum number of iterations for model training to ensure sufficient convergence. Simultaneously, to mitigate the impact of imbalanced sample class distribution on model performance, class_weight = 'balanced' is introduced to automatically adjust the weight allocation of different classes during training. The optimal parameter settings for the Logistic Regression model are shown in the table:

Table 3: Logistic Regression Model Parameter Settings

Parameter Name	Parameter Value
C	0.02
penalty	L1
max_iter	1000
solver	liblinear
class_weight	balanced

The predictive performance of the Logistic Regression model was evaluated on the test set. The results showed that the model achieved a recall of 0.8458, an F_2 score of 0.6123, and an AUC of 0.7635. Overall, this model demonstrates a relatively high detection capability in high-risk disease screening tasks and exhibits a good ability to distinguish between individuals at risk.

2. Support Vector Machine Model

The Support Vector Machine (SVM) is a supervised learning model based on the principle of structural risk minimization in statistical learning theory. Its core idea is to construct an optimal separating hyperplane in the feature space to maximize the margin between two classes of samples, thereby enhancing the model's generalization ability while controlling empirical risk. Compared to models that only focus on minimizing training error, SVM balances model complexity and classification performance by introducing a margin constraint, making it particularly

suitable for medical prediction tasks with limited sample sizes and high-dimensional features. Due to the heterogeneity of individual characteristics and the fact that different risk states often lack a strictly linearly separable structure, SVM typically adopts a soft margin form. This allows a small number of samples to violate the margin constraint, enhancing the model's robustness to noise and outliers.

This model is constructed using the SVM from the sklearn machine learning library. During the model building process, the performance of SVM is mainly influenced by the form of the kernel function and its related parameters. Considering the potential nonlinear characteristics of the relationships among variables in the cognitive impairment risk prediction task, this paper focuses on three key types of parameters: the kernel function type, the kernel coefficient, and the penalty factor. The kernel parameter defines the similarity measure for samples in the feature space. This paper selects the linear kernel and the Radial Basis Function (RBF) kernel to balance model expressiveness and computational efficiency. The penalty coefficient C balances margin maximization and training error minimization: a smaller C value results in weaker penalty for misclassifications and stronger model generalization; conversely, a larger C value leads to a higher degree of model fit to the training samples. The kernel coefficient controls the influence scope of the RBF kernel, affecting the impact of individual samples on the decision boundary. This paper adopts the commonly used 'scale' and 'auto' settings in scikit-learn for configuration. The optimal parameter settings for the SVM model are shown in the table:

Table 4: SVM Model Parameter Settings

Parameter Name	Parameter Value
C	0.01
kernel	rbf
gamma	scale

The predictive performance of the Support Vector Machine model was evaluated on the test set. The results showed that the model achieved a recall of 0.8178, an F2 score of 0.6018, and an AUC of 0.7657. Overall, this model can effectively identify high-risk individuals in the cognitive impairment risk screening task, demonstrating a stable overall discriminative performance while maintaining a relatively high detection capability.

3. ***Random Forest model***

Random Forest (RF) is a supervised learning algorithm based on the concept of ensemble learning. Its core idea is to enhance the model's generalization ability by constructing multiple decision trees and aggregating their prediction results. This model employs the Bagging framework, which involves drawing bootstrap samples (with replacement) from the original training set to generate multiple training subsets, and independently training one decision tree on each subset. To further reduce the correlation among base learners, Random Forest, during the node splitting process of each decision tree, does not use all p features but randomly selects m features as candidate partitioning variables, effectively reducing the model's variance.

This model is constructed using the Random Forest algorithm from Python's scikit-learn machine learning library to classify and predict the risk of cognitive impairment. The performance of the Random Forest model is primarily influenced by the number of decision trees and parameters related to the tree structure. Considering the complex relationships among variables and the significant nonlinear characteristics in the cognitive impairment risk prediction task, this paper focuses on five key parameters: the number of decision trees (n_estimators), the maximum number of features considered for splitting (max_features), the maximum depth of the tree (max_depth), the minimum number of samples required to split an internal node (min_samples_split), and the minimum number of samples required at a leaf node (min_samples_leaf). The number of decision trees determines the total count of trees in the forest; appropriately increasing the number of trees can help reduce the variance of the model's predictions, thereby improving overall stability, but an excessive number of trees also increases computational cost. The maximum number of features controls how many features are considered as candidates for splitting at each node; by randomly selecting a subset of features in the feature subspace, it can effectively reduce the correlation between different decision trees and enhance the model's generalization ability. The maximum depth limits the growth depth of a single decision tree, structurally preventing the model from overfitting the training samples. The minimum samples for split specifies the minimum number of samples required for an internal node to be split further, while the minimum samples at leaf node restricts the minimum number of samples a leaf node can contain. These last two parameters constrain the complexity of the decision tree from the perspective of sample size. By jointly optimizing the parameters mentioned above, the model's stability and generalization performance are improved while ensuring its good discriminative ability for individuals at risk of cognitive impairment. The optimal parameter settings for the Random Forest model are shown in the table:

Table 5: RF Model Parameter Settings

Parameter Name	Parameter Value
n_estimators	302
max_features	sqrt
max_depth	5
min_samples_split	5
min_samples_leaf	6

The predictive performance of the Random Forest model was evaluated on the test set. The results indicated that this model performs well in the cognitive impairment risk prediction task, achieving a recall of 0.8692, an F2 score of 0.6090, and an AUC of 0.7693. It can be observed from the results that the Random Forest model possesses strong sensitivity in identifying high-risk individuals, effectively reducing the risk of missing potential cases of cognitive impairment. Simultaneously, its relatively high AUC value indicates that the model maintains stable discriminative ability across different classification thresholds, and its overall performance meets the application requirements for cognitive impairment risk screening.

4. ***XGBoost model***

XGBoost (Extreme Gradient Boosting) is an efficient gradient-boosted decision tree algorithm, representing an improvement and extension of the traditional Gradient Boosting Decision Tree (GBDT) method. This approach progressively builds multiple base learners using a forward stagewise additive model and combines them through weighted summation, thereby integrating several weak learners into a single strong learner. In each iteration, the newly added decision tree is used to fit the residuals between the previous model's predictions and the true labels,

continuously enhancing the overall model's predictive performance.

The XGBoost algorithm, based on the Python platform, was used to construct a gradient-boosted decision tree model for the classification and prediction of cognitive impairment risk. Specifically, this paper selects key hyperparameters for optimization from three aspects: model scale, learning process, and random sampling mechanism. These include the number of decision trees (n_estimators), learning rate (learning_rate), maximum tree depth (max_depth), minimum child node weight (min_child_weight), feature sampling ratio per tree (colsample_bytree), and subsample ratio (subsample). The number of trees and the learning rate collectively determine the model's iteration strength and convergence pace: the former controls the scale of base learners, while the latter adjusts the contribution of each newly generated tree to the overall model's prediction. The maximum tree depth and minimum child node weight are primarily used to constrain the structural complexity of individual trees, mitigating overfitting risk through both tree structure limitations and sample weight constraints. Meanwhile, the feature sampling ratio and subsample ratio introduce moderate randomness by randomly drawing from the feature space and the sample set, thereby reducing model variance and enhancing generalization ability. The optimal parameter settings for the XGBoost model are shown in the table:

Table 6: XGBoost Model Parameter Settings

Parameter Name	Parameter Value
n_estimators	252
learning_rate	0.01
max_depth	3
min_child_weight	3
colsample_bytree	0.83
subsample	0.74

The predictive performance of the XGBoost model was evaluated on the test set. The results showed that it achieved a recall of 0.8645, an F2 score of 0.6074, and an AUC of 0.7684. Overall, the XGBoost model is capable of maintaining good comprehensive performance while ensuring a relatively high detection capability, meeting the dual requirements of sensitivity and stability in the cognitive impairment risk screening scenario.

5. ***LightGBM model***

LightGBM is an efficient implementation form of the gradient boosting decision tree model. It maintains consistency with the previously discussed models in terms of the objective function optimization framework, iteratively constructing multiple decision trees to gradually reduce the loss function value and achieve the characterization of complex nonlinear relationships. Since its optimization objective form is the same as XGBoost, this paper will not repeat the formula derivation. Regarding the tree structure generation mechanism, LightGBM adopts a leaf-wise growth strategy. Unlike the traditional level-wise expansion method, this strategy, during each split iteration, selects the node among all current leaf nodes that yields the maximum loss reduction for expansion, allowing the loss function to be optimized to the greatest extent in local regions. This approach can achieve lower training error under the same tree depth constraint, thereby enhancing the model's expressive power and its ability to characterize complex variable relationships. LightGBM, while retaining the core idea of gradient boosting optimization, enhances local fitting capability through the leaf-wise growth mechanism, improving prediction accuracy while ensuring model generalization performance. It is well-suited for classification tasks characterized by complex variable relationships and significant nonlinear features, such as cognitive impairment risk prediction.

The LightGBM algorithm, based on the Python platform, was used to construct a gradient boosting decision tree model for the classification and prediction of cognitive impairment risk. Specifically, this paper selects key hyperparameters for optimization from three aspects: model scale, learning process, structural complexity control, and feature random sampling mechanism. These include the number of decision trees (n_estimators), learning rate (learning_rate), maximum number of leaves (num_leaves), minimum number of samples per leaf (min_child_samples), and the feature sampling ratio per tree (colsample_bytree). The number of trees and the learning rate collectively determine the model's iteration strength and convergence pace: the former controls the scale of base learners, while the latter adjusts the contribution of each newly generated tree to the overall model's prediction. They work in conjunction to balance the model's fitting ability and training stability. The maximum number of leaves and the minimum number of samples per leaf are primarily used to constrain the structural complexity of individual trees: the former limits the scale of leaves that can be generated per tree, controlling model complexity from an overall capacity perspective; the latter, by setting the minimum number of samples required for a leaf node, prevents local overfitting caused by too few samples, thereby enhancing the model's robustness. The feature sampling ratio, by randomly drawing from the feature space during the construction of each tree, introduces moderate randomness, reduces the impact of correlations among features on the model, and consequently decreases model variance and enhances generalization ability. The optimal parameter settings for the LightGBM model are shown in the table:

Table 7: LightGBM Model Parameter Settings

Parameter Name	Parameter Value
n_estimators	367
learning_rate	0.01
num_leaves	49
min_child_samples	29
colsample_bytree	0.7

The predictive performance of the LightGBM model was evaluated on the test set. The results showed that it achieved a recall of 0.8738, an F2 score of 0.6048, and an AUC of 0.7657. Overall, the LightGBM model maintains a high detection capability while balancing overall discriminative performance. Its recall rate of 0.8738 is the highest among all models, indicating that this model possesses the strongest sensitivity in identifying individuals at high risk of cognitive impairment.

C. Stacking Ensemble Model

1. Brief Introduction to Model Principles

To further enhance the comprehensive performance of the cognitive impairment risk prediction model, this study introduces the Stacking ensemble learning method based on the construction of single models. Stacking (Stacked Generalization) is a hierarchical model fusion strategy. Its

core idea is to build a two-level learning structure based on the parallel learning of multiple models. The prediction results of multiple base learners for the samples are used as new input features, and then a meta-leamer relearns and integrates these prediction results to form the final prediction output. Unlike simple averaging or manual weighting methods, Stacking automatically learns the optimal combination relationships among different models through a data-driven approach, enabling the model to capture the structural differences and complementary information between various prediction patterns at a higher level, thereby reducing the bias and variance risks of single models while improving overall generalization ability.

In the specific implementation process, this study constructs a fusion framework with a two-layer structure. The first layer consists of multiple base learners, with each model trained independently on the same training data to output the predicted probability for each sample. To avoid information leakage and overfitting issues, five-fold stratified cross-validation is used during the training phase to generate cross-validated prediction results: first, the training data is divided into five mutually exclusive subsets; in each iteration, four folds of data are used to train the base learners, and predictions are made on the remaining one fold. After five iterations, all training samples obtain prediction probabilities generated by models trained on data that did not include them. Subsequently, the cross-validated prediction results from each base learner are concatenated column-wise to form a new meta-feature matrix. The second-layer meta-learner is trained using this meta-feature matrix as input, learning the combination relationships among the prediction results of the different base models to achieve a weighted integration of the final output. In the testing phase, each base learner is first refitted on the complete training set, then generates prediction probabilities for the test samples, which are then integrated by the trained meta-learner to obtain the final prediction results.The schematic diagram is shown in the figure below:

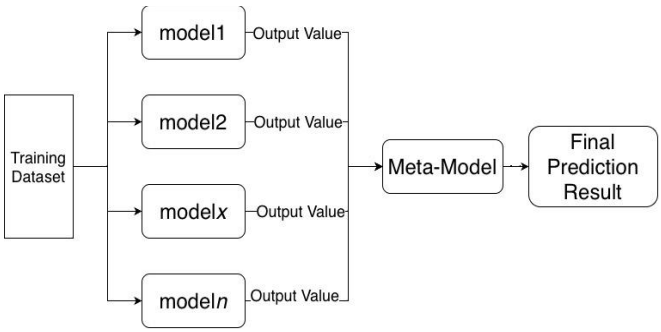

Figure1: Stacking Model Schematic Diagram.

2. Establishment of a Cognitive Risk Model Based on Stacking

In the construction process of the Stacking ensemble model, the selection of both base learners and the meta-learner significantly impacts the overall performance and stability of the model. To ensure a reasonable fusion framework structure, this study first systematically evaluated the recall capability, comprehensive discriminative performance, and stability of each model based on the results of single-model experiments. Given that cognitive impairment risk screening falls within a public health application scenario, where the ability to identify high-risk individuals is a priority, models with relatively high recall performance and stable overall performance were prioritized during the candidate model screening phase. After comprehensive comparison, Random Forest, XGBoost, and LightGBM demonstrated superior levels in terms of recall and overall discriminative ability, exhibiting good generalization capabilities; therefore, they were included in the candidate set for the first-layer base learners.

Analyzed from the perspective of model structure, Random Forest is based on the Bagging framework, reducing variance and enhancing model stability through the parallel construction of multiple trees. XGBoost and LightGBM are based on the Boosting framework, progressively improving model expressiveness by iteratively optimizing residuals, effectively capturing nonlinear relationships and complex interaction effects among features. The three types of models exhibit significant differences in their modeling mechanisms, possessing good structural complementarity that contributes to enhancing the overall robustness of the ensemble model. Regarding the selection of the meta-learner, considering that the first-layer base learners already possess strong nonlinear modeling capabilities, if a complex model were to be used in the second layer, it might increase the risk of overfitting and reduce the generalization stability of the model. Within the Stacking framework, the primary role of the meta-learner is to perform a reweighted integration of the predicted probabilities output by each base model; its task is essentially a probability combination problem in a low-dimensional feature space. Therefore, this study selects the Logistic Regression model, which has a relatively simple structure, strong generalization ability, and good interpretability, as the meta-learner. On one hand, Logistic Regression can linearly integrate the outputs of different base models, achieving optimized weighting at the probability level. On the other hand, its simple parameter structure helps control model complexity, enhancing the stability and generalizability of the ensemble model in practical application scenarios.

After determining the candidate set of first-layer base learners, this study further constructed Stacking ensemble structures with different combinations and systematically compared their predictive performance. During the model construction process, to ensure the rationality of parameter settings and the stability of results, five-fold stratified cross-validation combined with grid search was used to optimize the hyperparameters of each base learner. Specifically, different parameter combinations were traversed within a pre-set parameter range, the performance of each combination was evaluated through five-fold cross-validation, and the parameter configuration with the optimal comprehensive metrics was selected for model training. The performance of each ensemble structure, after parameter optimization, was evaluated on an independent test set. To objectively assess the application effectiveness of different ensemble strategies in cognitive impairment risk screening, this study adopted recall, F_2 score, and AUC as evaluation metrics. The prediction results for the different ensemble structures are shown in the following table:

Table 8: Prediction Results of Stacking Ensemble Models

model	Recall	F2	AUC
LGBM+RF	0.8785	0.6092	0.7693
XGBoost+LGBM	0.8645	0.5945	0.7689
XGBoost+RF	0.8832	0.6081	0.7682
XGBoost+LGBM+RF	0.8785	0.6100	0.7693

To compare the performance of different ensemble structures in cognitive impairment risk prediction, a comprehensive analysis was conducted on the recall, F2 score, and AUC metrics of each model. The results show that the AUC values of all ensemble models are generally close,

all at similar levels, indicating relatively small differences in overall discriminative ability among the different structures. Among them, LGBM+RF and XGBoost+LGBM+RF achieved slightly higher AUC, demonstrating relatively stable overall discriminative capability. In terms of recall, the XGBoost+RF model achieved the highest value, showing its strong ability to identify high-risk individuals; the recall performance of LGBM+RF and the three-model ensemble structure was also relatively similar. Combined with the results of the F_2 metric, it can be seen that the three-model ensemble structure performs relatively more balancedly in terms of both recall and precision.

Considering all metrics comprehensively, the three-model Stacking structure maintains stable performance in overall discriminative ability, risk identification capability, and performance balance, without any significant shortcomings. Compared to the two-model ensemble structures, its comprehensive performance is more balanced, with smaller performance fluctuations. Therefore, the three-model Stacking ensemble structure constructed with XGBoost, LightGBM, and Random Forest was ultimately selected as the optimal prediction model for this study.

D. Comparative Analysis of Model Results

After completing the parameter tuning and structure construction for each model, to systematically compare the performance of different algorithms in the cognitive impairment risk prediction task, this study conducted a unified evaluation of Logistic Regression, Support Vector Machine, Random Forest, XGBoost, LightGBM, and the Stacking ensemble model. Model performance was measured across three dimensions: recall, F2 score, and AUC, to comprehensively reflect each model's ability to identify high-risk individuals, performance balance, and overall discriminative capability. The prediction results for each model are shown in the table below:

Table 9: Comparison of Prediction Results Across Different Models

model	Recall	F2	AUC
LR	0.8458	0.6123	0.7635
SVM	0.8178	0.6018	0.7657
RF	0.8692	0.6090	0.7693
XGBoost	0.8645	0.6074	0.7684
LGBM	0.8738	0.6048	0.7657
Stacking	0.8785	0.6100	0.7693

From the perspective of overall discriminative ability, both the Random Forest and the Stacking ensemble model achieved an AUC of 0.7693, the highest level among this group of models; XGBoost achieved 0.7684; Support Vector Machine and LightGBM achieved 0.7657; and Logistic Regression achieved 0.7635. It can be observed that the ensemble model maintained discriminative ability comparable to the best single model without any performance degradation, demonstrating relatively stable overall discrimination capability. In terms of recall, LightGBM, which performed well among the single models, achieved 0.8738, while the Stacking ensemble model further improved this to 0.8785, the highest value among all models. Compared to single models, the ensemble strategy achieved a certain degree of optimization in identifying high-risk individuals, enhancing the model's coverage of the potential at-risk population. Regarding the F_2 score, Logistic Regression achieved 0.6123, the highest among the single models; the Stacking ensemble model achieved 0.6100, a small gap from the best single model, while being higher than the remaining single models. This indicates that while improving recall, the ensemble model did not significantly compromise overall model balance.

Synthesizing the above results, it can be found that each single model has its own advantages on different metrics, but it is difficult for any single one to achieve optimal performance simultaneously across multiple evaluation dimensions. The Stacking ensemble model, while maintaining the highest level of AUC, achieved a further improvement in recall and maintained a relatively stable F_2 performance, resulting in a more balanced overall performance. Although the degree of improvement in each metric is limited, the ensemble model shows no significant shortcomings across multiple dimensions, reflecting the optimization role and application potential of multi-model information integration in the task of cognitive impairment risk prediction.

CONCLUSION

This chapter focused on the issue of cognitive impairment risk prediction by constructing and comparing multiple machine learning models, systematically evaluating the performance and stability of different algorithms in screening tasks. At the single-model level, Logistic Regression, Support Vector Machine, Random Forest, XGBoost, and LightGBM models were established respectively, and a comprehensive analysis of model performance was conducted from multiple dimensions including recall, F2 score, and AUC. Building on this foundation, this chapter further constructed a Stacking ensemble model based on Random Forest, XGBoost, and LightGBM, enhancing overall performance through the integration of multi-model prediction results. The results showed that while maintaining a high AUC level, the ensemble model achieved a further improvement in recall, with overall performance being more balanced compared to single models. In summary, the research in this chapter validates the application value of multi-model ensemble methods in the task of cognitive impairment risk prediction, providing methodological support for subsequent model promotion and practical application.

References

[1] Cho M, Ban HJ, Nam HR, Chu CH, Jeon JP, Kim SC. A machine learning framework for classifying dementia risk in mild cognitive impairment: evidence from a Korean genome-wide association study cohort. Alzheimers Res Ther. 2025 Nov 10;17(1):241.

[2] Zhang L, Hu XT, Chen W, et al. Construction of a prediction model for mild cognitive impairment in multimorbid elderly based on machine learning algorithms. Journal of New Medicine, 2025, 35(04): 409-418.

[3] Yang J. Construction of a risk prediction model for mild cognitive impairment in patients with chronic heart failure based on machine learning [D]. Chengdu Medical College, 2025.

[4] Qin XY. Research on cognitive impairment in the elderly based on machine learning and interpretability [D]. Soochow University, 2023.

[5] Zhang HC. Establishment and validation of a risk prediction model for mild cognitive impairment in Chinese elderly based on ensemble learning methods [D]. Anhui Medical University, 2024.

[6] Yang WY, Gao XF, Xiao H, et al. Construction and validation of a nomogram model for predicting cognitive impairment risk in hypertensive patients based on CHARLS data. Journal of Chongqing Medical University, 2025, 50(10): 1329-1337.

[7] Chen C. Research on risk prediction models for mild cognitive impairment in middle-aged and elderly Chinese population [D]. Chongqing Medical University, 2024.
[8] Guo LN, Fan GS. Prediction of surface soil bulk density based on grid search and cross-validation support vector machine. Chinese Journal of Soil Science, 2018, 49(03): 512-518.
[9] Fan B. Research on diabetes risk prediction model based on convolutional neural network [D]. Nanjing University of Posts and Telecommunications, 2022.
[10] Lu NC. Research on credit default risk assessment based on K-means-SMOTE and improved Stacking method [D]. Yangzhou University, 2025.
[11] Ji YZ. Application of machine learning classification models in corporate bankruptcy risk assessment [D]. China University of Petroleum (Beijing), 2022.
[12] Zhu ST. Research on cerebrovascular disease prediction based on ensemble learning [D]. Soochow University, 2023.
[13] Yuan K. Research on liver cancer prediction model based on interpretable machine learning [D]. Jiangxi University of Finance and Economics, 2022.
[14] Chen YH. Personal loan default prediction and SHAP explanation based on Stacking ensemble model under imbalanced data [D]. Jiangxi University of Finance and Economics, 2025.
[15] Chen SY. Research on hyperuricemia prediction in Chongqing based on Stacking ensemble model [D]. Chongqing Medical University, 2024.
[16] Dagliati A, Marini S, Sacchi L, Cogni G, Teliti M, Tibollo V, De Cata P, Chiovato L, Bellazzi R. Machine Learning Methods to Predict Diabetes Complications. J Diabetes Sci Technol. 2018 Mar;12(2):295-302.
[17] Guo Y. Research on credit risk problems based on Stacking model [D]. Shanxi University, 2024.