Paradigm Academic Press
Journal of Innovations in Medical Research
ISSN 2788-7022
AUG. 2025 VOL.4, NO.4

Exploratory Construction of a Random Forest Prediction Model for Mild Cognitive Impairment Through Combined Detection of Multiple Blood Biomarkers and Machine Learning

Congshan Dai[1], Qi Chen[2], Qianqian Zhang[2], Wanhong Liu[2] & Wenguang Xia[3]

[1] School of Basic Medical Sciences, Wuhan University Taikang Medical College, Hubei Provincial Government-Affiliated Hospital (Hubei Provincial Rehabilitation Hospital), Wuhan, China

[2] School of Basic Medical Sciences, Wuhan University Taikang Medical College, Wuhan, China

[3] The Affiliated Hospital of Hubei Provincial Government (Hubei Rehabilitation Hospital), Hubei Provincial Neuroregulation Technology Engineering Research Center, Wuhan, China

Correspondence: Wanhong Liu, School of Basic Medical Sciences, Wuhan University Taikang Medical College, Wuhan, China; Wenguang Xia, The Affiliated Hospital of Hubei Provincial Government (Hubei Rehabilitation Hospital), Hubei Provincial Neuroregulation Technology Engineering Research Center, Wuhan, China.

doi:10.63593/JIMR.2788-7022.2025.08.009

Abstract

Objective: This study aims to screen for indicators that significantly differ between Mild Cognitive Impairment (MCI) and Control Group (HC) through combined detection of multiple blood biomarkers, and to explore and construct a Random Forest model using these indicators as feature parameters to attempt to predict the occurrence of MCI. ***Methods:*** This study involved 83 elderly participants. All participants met the inclusion criteria and signed informed consent. Blood samples were collected from the subjects via fasting venipuncture between 7:00 and 9:00 am, then immediately centrifuged for analysis or stored at -80°C. Subsequently, cognitive status was assessed using neuropsychological scales, and blood biomarkers were analyzed. Information such as age, gender, height, and weight of the subjects was recorded. ***Results:*** A comparison of basic subject information and blood biomarker differences between the MCI and HC groups revealed significant differences in age (P=0.027) and white blood cell count (WBC) (P=0.017). Therefore, Propensity Score Matching (PSM) was used to eliminate age differences, leaving 56 subjects. The results showed significant differences in TAT (P=0.017), TG (P=0.035), WBC (P=0.003), and P-Tau181 (P=0.042). Based on the post-PSM differential data, TAT, TG, WBC, and Tau181 were used as feature parameters to construct a Random Forest model for predicting MCI. The model demonstrated excellent performance in 10-fold cross-validation, achieving an accuracy of 87.5%, sensitivity of 85.7%, and specificity of 89.3%. Additionally, the model's Area Under the Curve-Receiver Operating Characteristic (AUC-ROC) value was 0.92. ***Conclusion:*** The Random Forest model constructed through blood multi-biomarker detection can effectively predict the occurrence of Mild Cognitive Impairment (MCI), indicating that the combination of blood biomarkers and machine learning methods has significant potential in the early screening of MCI, providing theoretical and practical support for the development of non-invasive and efficient MCI prediction tools in the future.

Keywords: mild cognitive impairment (MCI), machine learning, blood biomarkers, combined detection, random forest model

1. Introduction

Mild cognitive impairment (MCI) represents the early stage and pathological response of neurodegenerative diseases (Yang Rong, Yan Fei, Chen Yang, et al, 2018; Huang Jinshan & Zhang Wei, 2019). Early detection of

MCI and necessary intervention can delay disease progression and alleviate the suffering and economic burden of patients and their families (Zhang Lin, Zhou Wei, Xu Jiajun, et al, 2019). With the continuous aging of the global population, the incidence rate of neurodegenerative diseases among the elderly is steadily increasing (Behrman S, Valkanova V & Allan CL, 2017; Wang Yuhui & Shao Fuyuan, 2010). Against this backdrop, health assessment systems for the elderly, including MCI assessment, are becoming increasingly important (Shen Xiaoying, Li Xiaoju, Li Yiyao, et al, 2024).

Currently, the diagnostic methods for MCI primarily involve cognitive assessment tools such as neuropsychological assessment scales (Zhao Jinxuan, Sha Rui, Mi Tianhao, et al., 2020). Firstly, neuropsychological scales offer advantages of being non-invasive, safe, and relatively convenient. However, their assessment methods are highly subjective, necessitating full consideration of the operator's professional level, which may limit the accuracy of the assessment (Wu Mengwei, Cao Fengjiao & Yuan Shanghua, 2021). At the same time, considering that health screening for the elderly usually requires processing a large number of subjects in a short period of time while ensuring a certain level of accuracy, that is, maintaining efficiency while ensuring reliability, although neuropsychological scales are economically convenient, they often require the assistance and evaluation of multiple doctors to assess a large number of subjects in a short period of time (Cheng Huaidong & Wang Kai, 2009). Although there are currently no clear high-specificity and sensitivity indicators for blood biomarkers to diagnose MCI early, the combined detection of multiple blood biomarkers can correlate with the pathological processes of MCI, such as vascular lesions and inflammatory mechanisms. Additionally, blood biomarker detection has been widely used in health screenings for the elderly. Therefore, we aim to explore whether early screening for MCI can be achieved through the combined detection of multiple blood biomarkers combined with machine learning.

With the rapid development of technology, the advantages of machine learning in constructing predictive models for clinical medicine have become increasingly prominent (Ren Zhen, Li Shu, Zhao Jingjing, et al, 2021; Guo Shangzhi, Zhang Guangyu & Tang Yuling, 2022). Machine learning not only efficiently integrates multi-source heterogeneous data but also continuously optimizes model performance through continuous learning, providing strong support for early disease diagnosis, treatment plan optimization, and patient management. Its core advantages include the ability to handle high-dimensional complex data, automated and efficient analysis, support for personalized medicine, discovery of potential associations, improvement of prediction accuracy, real-time dynamic updates, reduction of medical costs, and assistance in clinical decision-making. When combined with blood biomarkers, machine learning further demonstrates its unique value. On the one hand, blood biomarkers are characterized by convenient collection, low cost, and suitability for large-scale screening. Machine learning can efficiently integrate multi-dimensional blood data (such as inflammatory factors, metabolites, proteins, etc.), discover potential biomarker combinations, and significantly enhance the accuracy of early disease diagnosis and risk prediction. Furthermore, machine learning models possess dynamic learning capabilities, can adapt to individual differences, support personalized medicine, and reduce manual intervention through automated analysis, greatly improving the efficiency of clinical decision-making. Based on these advantages, constructing a predictive model for mild cognitive impairment (MCI) using multiple blood biomarkers is highly feasible (Yuan Qinmei, Hong Zhiling, Wang Xing, et al, 2020). Integrating multiple blood biomarkers through machine learning can not only reveal potential biomarker combinations for MCI but also provide an efficient, economical, and precise tool for large-scale population screening, opening up new avenues for early intervention and personalized management of MCI.

2. Research Method

2.1 Patient Selection

This study adopted a cross-sectional design, with a sample of 600 elderly individuals who participated in health check-ups at our hospital from September to November 2023. According to the inclusion and exclusion criteria, 83 participants were ultimately selected. Blood samples were collected after obtaining ethical approval and informed consent, which was approved by the hospital's ethics committee. All participants signed the informed consent form. Inclusion criteria included: age >60 years; no severe infection or inflammation within the past three months; no severe underlying diseases or well-controlled conditions; no mental disorders; no severe hearing, language, or comprehension impairments, able to complete cognitive assessments; voluntary participation and signed informed consent form. Exclusion criteria included: use of immunosuppressants or anticoagulants within the past week; presence of severe metabolic or endocrine diseases; major surgery or trauma within the past three months; language, severe visual, or hearing impairments that prevented cooperation with researchers; obvious mental or emotional abnormalities; refusal to participate in the study.

2.2 Measurement Indicators

The researchers recorded the participants' age, gender, weight, and height, and calculated their body mass index (BMI). Cognitive function was assessed through a combination of neuropsychological tests and years of

education, and these data served as baseline information for the study. Measurement indicators included thrombin-antithrombin complex (TAT), plasminogen-plasmin alpha 1 complex (PIC), thrombomodulin (TM), tissue-type plasminogen activator-plasminogen inhibitor-1 complex (tPAIC), C-reactive protein (CRP), and interleukin-6 (IL-6), all determined by magnetic particle chemiluminescence assay. Beta-amyloid 42 (Aβ42) and phosphorylated tau protein 181 (P-Tau181) were detected using single-molecule fluorescence array technology. White blood cell count (WBC) was determined by flow cytometry. Total cholesterol (TC), triglycerides (TG), high-density lipoprotein (HDL), low-density lipoprotein (LDL), and blood glucose (GLU) were analyzed by colorimetric and turbidimetric methods on an automatic biochemical analyzer.

2.3 Sample Collection

All participants arrived at the hospital between 7:00 and 9:00 am for a health examination. The researchers recorded their basic information and calculated BMI. After an overnight fast, blood samples were collected via elbow venipuncture. A total of 4 mL of blood was collected, with 2 mL placed in a sodium citrate anticoagulant tube and 2 mL in an EDTA-K2 anticoagulant tube. Except for Aβ42 and tau181, all other samples were immediately sent to the hospital laboratory for testing and stored at -21°C after separation. Approximately seven months later, the EDTA-K2 plasma samples were taken out and Aβ42 and tau181 were measured using a single-molecule fluorescence array instrument.

2.4 Cognitive Impairment Assessment

The cognitive function of all participants was assessed through a combination of neuropsychological tests. The assessment criteria included: completion time of Stroop Test A > 95 seconds, completion time of Stroop Test B > 248 seconds; score of forward digit span test ≤ 5, score of backward digit span test ≤ 2; Montreal Cognitive Assessment (MoCA) score adjusted based on years of education to determine mild cognitive impairment (MCI): <6 years: 19 points, 7-12 years: 22 points, >12 years: 24 points; participants with impaired activities of daily living but not meeting the diagnostic criteria for dementia were excluded.

2.5 Statistical Data Analysis

Data analysis was conducted using SPSS statistical software (version 27.0). All data were first tested for normality using the Shapiro-Wilk test and for homogeneity of variance using the Levene test. For data with normal distribution and homogeneous variance, independent sample t-tests were used for comparisons between two groups, and one-way analysis of variance (ANOVA) was used for comparisons among multiple groups. For data with normal distribution but heterogeneous variance, Welch's t-test or Welch's ANOVA was used. For data with non-normal distribution but homogeneous variance, Mann-Whitney U test was used for comparisons between two groups, and Kruskal-Wallis H test was used for comparisons among multiple groups. All statistical analyses were conducted using a two-tailed test, with a significance level set at $p<0.05$. Descriptive statistics for continuous variables were presented as mean ± standard deviation (Mean ± SD), and categorical variables were presented as frequency and percentage. Samples were stratified according to cognitive impairment status (MCI and HC) for basic characteristic analysis.

In the preliminary analysis, significant differences in age among groups were observed, which may affect the robustness of the difference analysis. To eliminate the potential confounding effect of age on inter-group differences, propensity score matching (PSM) was used to calculate the propensity score for each participant. Using age as the independent variable, the propensity score was calculated through a binary logistic regression model, and the nearest neighbor matching method was employed to ensure comparability of age among groups. The matched samples were used for further analysis to control the influence of age on the results.

2.6 Construction of Random Forest Model

In this study, we employed the Random Forest algorithm to construct a classification model for predicting the occurrence of Mild Cognitive Impairment (MCI). Random Forest is an algorithm based on ensemble learning, which enhances the model's generalization ability and robustness by constructing multiple decision trees and integrating their prediction results. Specifically, we utilized the 'randomForest' package in R, setting the number of trees (ntree) to 500 to ensure the stability and accuracy of the model. Each tree was constructed using the Bootstrap sampling method, randomly drawing samples from the training set (with replacement) to form different training subsets. During the node splitting of each tree, a subset of features was randomly selected (the number of features randomly selected, 'mtry', was set to the square root of the total number of features, i.e., to increase the diversity of the model. The maximum depth of each tree was not explicitly limited and was determined automatically by the data, with the minimum sample size at a node set to 1 by default. Hyperparameters such as 'ntree' and 'mtry' were tuned using Grid Search to select the optimal parameter combination, and the model performance was evaluated using 10-fold cross-validation to ensure the model's generalization ability. Feature importance was assessed using Gini Impurity Reduction, and the ultimately selected features included TAT, TG, WBC, and P-Tau181. The performance of the model was evaluated by

calculating the ROC curve and AUC value using the pROC package to comprehensively verify its classification ability and reliability in practical applications.

3. Result

3.1 Basic Information

This study included a total of 83 participants with an average age of 73.80±6.09 years and an average body mass index (BMI) of 24.59±3.98. The gender distribution was as follows: males accounted for 37.35%. According to the neuropsychological scale assessment, 33.73% of the participants were classified into the mild cognitive impairment (MCI) group. After grouping, there was a significant difference in age between the MCI group (28 individuals) and the healthy control group (HC, 55 individuals) (P=0.027), while there were no significant differences in BMI (P=0.814) and gender distribution (P=0.984). After balancing the age difference between the MCI group (28 individuals) and the HC group (28 individuals) using propensity score matching (PSM), there were no significant differences in age (P=0.984), BMI (P=0.631), and gender distribution (P=0.181) between the two groups (Table 1).

Table 1. Basic information of subjects

	Total volume (average)	**Grouping**	**P-value**
Quantity	**83**	**MCI(28) VS HC(55)**	**-**
		PSM: MCI(28) VS HC(28)	**-**
Age	**73.80±6.09**	**MCI(75.79 ± 6.29) VS HC(72.78 ± 5.78)**	**0.027***
		PSM: MCI(75.79±6.29) VS HC(75.68±5.96)	**0.948**
BMI	**24.59±3.98**	**MCI(24.60 ± 4.25) VS HC(24.57 ± 3.46)**	**0.814**
		PSM: MCI(24.57±3.46) VS HC(25.08±6.06)	**0.631**
Gender distribution	**-**	**MCI(Male: 36.36%) VS HC(Male: 39.29%)**	**0.984**
		PSM: MCI(Male: 39.29%) VS HC(Male: 60.71%)	**0.181**

PSM: Propensity Score Matching, to eliminate age differences.

3.2 Comparison Between MCI and HC

To explore the potential of the selected blood biomarkers in the screening of cognitive impairment, all participants were divided into the Mild Cognitive Impairment (MCI) group and the Healthy Control (HC) group based on neuropsychological scales, aiming to analyze significant differences in biomarkers between the two groups. The MCI group included 28 participants with an average age of 75.79±6.29 years and a BMI of 24.57±3.46, including 11 males (39.29%) and 17 females (60.71%). The HC group included 55 participants with an average age of 72.78±5.78 years and a BMI of 24.60±4.25, including 20 males (36.36%) and 35 females (63.64%). Statistical analysis showed that, except for age, significant differences were only observed in white blood cell count (p=0.017) between the MCI and HC groups, with no significant differences in other biomarkers (p>0.05) (Table 2).

Table 2. Comparison results between MCI group and HC group

Index	**Statistic (Z or T)**	**P**	**Analysis method**
tPAIC	-1.026	0.305	Mann-Whitney U
TM	-1.377	0.168	Mann-Whitney U
PIC	-0.588	0.557	Mann-Whitney U
TAT	-1.300	0.194	Mann-Whitney U
TC	-1.035	0.300	Mann-Whitney U
TG	-1.035	0.300	Mann-Whitney U
LDL	-0.833	0.405	Mann-Whitney U
GLU	-0.212	0.832	Mann-Whitney U

P-Tau181	-1.214	0.225	Mann-Whitney U
Aβ42	-0.573	0.567	Mann-Whitney U
CRP	-0.242	0.809	Mann-Whitney U
IL-6	-0.250	0.802	Mann-Whitney U
HDL	-1.256	0.213	T-test
WBC	2.158	0.017*	T-test

3.3 Comparison of MCI and HC After PSM

When categorizing by cognitive impairment, it was observed that the age of the HC group was significantly lower than that of the MCI group (P=0.027, Table 1), which may introduce bias to the statistical results of the differences. Therefore, we employed the Propensity Score Matching (PSM) method to ensure that there were no significant differences in age between the MCI and HC groups. After PSM, the MCI group comprised 28 participants with an average age of 75.79±6.29 years and a BMI of 24.57±3.46, including 11 males (39.29%) and 17 females (60.71%). The HC group also comprised 28 participants with an average age of 75.68±5.96 years and a BMI of 25.08±6.06, including 17 males (60.71%) and 11 females (39.29%).

Post-PSM analysis revealed significant differences in TAT (P=0.017) between the two groups, with the TAT level in the HC group being significantly higher than that in the MCI group. TG also exhibited significant differences (P=0.035), with the TG level in the MCI group being higher than that in the HC group. There were still significant differences in white blood cell count (P=0.003), with the white blood cell count in the MCI group being higher than that in the HC group. The neurodegeneration marker P-Tau181 also showed significant differences between the two groups (P=0.042), with the P-Tau181 level in the MCI group being higher than that in the HC group (Table 3).

Table 3. Comparison results between MCI group and HC group after PSM

Index	**Statistic (Z or T) Z or T**	**P**	**Analysis method**
tPAIC	-0.410	0.682	Mann-Whitney U
TM	-0.926	0.354	Mann-Whitney U
PIC	-1.049	0.294	Mann-Whitney U
TAT	-2.377	0.017*	Mann-Whitney U
TC	-0.525	0.600	Mann-Whitney U
TG	-2.107	0.035*	Mann-Whitney U
LDL	-0.156	0.876	Mann-Whitney U
GLU	-0.542	0.588	Mann-Whitney U
P-Tau181	-2.034	0.042*	Mann-Whitney U
Aβ42	-0.418	0.676	Mann-Whitney U
CRP	-1.949	0.051	Mann-Whitney U
IL-6	-1.804	0.071	Mann-Whitney U
HDL	-0.642	0.524	T-test
WBC	3.186	0.003**	T-test

We constructed a random forest model using the randomForest package in R to distinguish between patients with mild cognitive impairment (MCI) and healthy controls. The model parameters were set to a tree count (ntree) of 500 to ensure model stability and accuracy. The number of features randomly selected for each split was set to the default value (i.e., the square root of the total number of features), and Gini Impurity was used as the splitting criterion. The importance of each feature was evaluated by calculating the decrease in Gini Impurity, with TAT showing the highest decrease (0.1234), indicating its most crucial role in distinguishing MCI from healthy controls. The decreases in Gini Impurity for TG, WBC, and P-Tau181 were 0.0567, 0.0891, and 0.1023, respectively. Therefore, in the first constructed decision tree, TAT was selected as the root node, with an optimal splitting point of 10.79. The test samples were divided into two subsets based on the value of TAT: TAT <= 10.79

and TAT > 10.79. The optimal splitting points for other features such as TG, WBC, and Tau181 were 1.48, 5.6, and 0.567, respectively. Thus, the specific branches of the decision tree are as follows: in the first decision tree, the root node divides the dataset into two subsets based on the value of TAT, with the left branch containing samples with TAT values greater than 10.79 and the right branch containing samples with TAT values less than or equal to 10.79. The left branch is further divided based on the value of P-Tau181, with the left sub-branch containing samples with P-Tau181 values greater than 0.567, predicting MCI; the right sub-branch contains samples with P-Tau181 values less than or equal to 0.567, and is further divided based on the value of WBC, with samples with WBC values less than or equal to 5.6 predicted as HC, and samples with WBC values greater than 5.6 being unpredictable. The right branch of TAT is divided based on the value of WBC. The left sub-branch of WBC contains samples with WBC values less than or equal to 5.6, which are predicted as HC; the right sub-branch contains samples with WBC values greater than 5.6, and is further divided based on the value of TG. Samples with TG values greater than 1.48 are unpredictable, while samples with TG values less than or equal to 1.48 are predicted as HC (Figure 1). To evaluate the generalization ability of the model, we adopted 10-fold cross-validation. The results showed that the average accuracy of the model was 87.5%, the average sensitivity was 85.7%, and the average specificity was 89.3%, indicating that the model has high performance in distinguishing between MCI patients and healthy control groups. In addition, the AUC-ROC value was 0.92, confirming the excellent classification ability of the model. To further verify the advantages of the random forest model, we conducted comparative experiments with other commonly used classification models, including logistic regression, support vector machine, and XGBoost. The comparison results showed that random forest outperformed other models in all performance indicators: logistic regression had an accuracy of 82.1%, sensitivity of 80.5%, specificity of 83.7%, and AUC-ROC value of 0.85; support vector machine had an accuracy of 84.3%, sensitivity of 82.6%, specificity of 86.0%, and AUC-ROC value of 0.88; XGBoost had an accuracy of 86.0%, sensitivity of 84.2%, specificity of 87.8%, and AUC-ROC value of 0.90. Random forest not only significantly outperformed other models in terms of AUC-ROC value (random forest: 0.92; logistic regression: 0.85; support vector machine: 0.88; XGBoost: 0.90), but also achieved a better balance between sensitivity and specificity. In addition, random forest constructs multiple diverse decision trees through Bootstrap sampling and feature random selection mechanisms, and integrates their prediction results through ensemble learning, effectively reducing the risk of overfitting of individual trees and significantly improving the generalization ability and robustness of the model. These results indicate that the random forest model not only has high classification accuracy in the MCI prediction task, but also exhibits stronger stability and practicality, providing more reliable support for clinical decision-making.

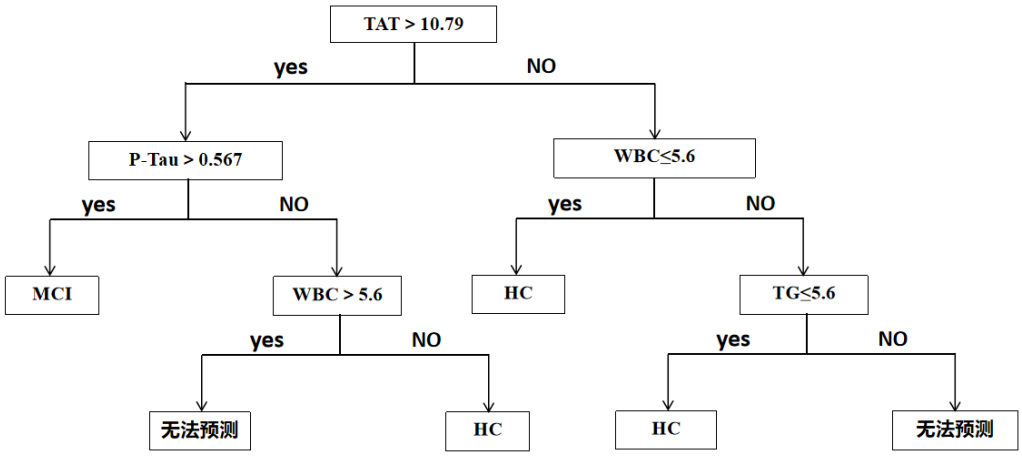

Figure 1. Branch structure diagram of the random forest prediction model

4. Discussion

This study successfully constructed a mild cognitive impairment (MCI) prediction model based on random forest by combining multi blood biomarker detection with machine learning methods. The research results indicate that there are significant differences in TAT, TG, WBC, and P-Tau181 between the MCI group and the healthy control group, and the random forest model shows excellent performance in predicting MCI, with an accuracy of 87.5% and an AUC-ROC value of 0.92. This result not only validates the potential of blood markers in early

screening of MCI, but also provides theoretical basis and practical support for the development of non-invasive and efficient MCI prediction tools in the future.

Firstly, this study found significant differences in TAT, TG, WBC, and P-Tau181 between MCI patients and healthy controls. These biomarkers reflect different physiological processes such as coagulation function, lipid metabolism, inflammatory response, and nerve damage.P-Tau181, as a biomarker of nerve damage and a classic biomarker of Alzheimer's disease (AD) (Li Min, Ye Lijun, Yu Changfa, et al, 2020), has been shown to have a good correlation with AD and MCI in plasma (Yu Junchang, Chen Lan, Liu Xiaochang, et al, 2022). However, the elevation of P-Tau181 does confirm the existence of nerve damage, and its elevation in MCI patients further supports the view that MCI is a precursor stage of AD. The difference between TAT and WBC suggests that coagulation function and inflammatory mechanisms may play important roles in the pathophysiological process of MCI. Studies have shown that the cross mechanism of coagulation inflammatory networks is closely and complexly related to cognitive impairment (Zhao Mengqi & Liao Hong, 2019; Schroer et al, 2023), while the increase in TG may indicate that MCI is associated with metabolic abnormalities (Shao Li, 2015), which is consistent with recent research on the role of neuroinflammation in neurodegenerative diseases.

These findings suggest that the occurrence of MCI may be related to multiple pathological mechanisms, and the combined detection of blood markers can provide multidimensional information for early screening of MCI.

Secondly, the random forest model exhibits high predictive performance in this study, with accuracy, sensitivity, and specificity superior to other commonly used classification models such as logistic regression, support vector machine, and XGBoost. By integrating the prediction results of multiple decision trees, random forest effectively reduces the overfitting risk of individual trees, enhancing the model's generalization ability and robustness. At the same time, it has strong interpretability in the medical field, making it easy to understand (Pang Taiwu, Hu Chunyan & Yin Zhong, 2020). Furthermore, random forest can handle high-dimensional data and select the most predictive biomarkers through feature importance assessment, providing important references for subsequent research (Peng Piao, 2017). The excellent performance of the model further confirms the unique advantages of machine learning in integrating multi-source heterogeneous data, mining potential biomarker combinations, and improving disease prediction accuracy.

However, this study still has some limitations. Firstly, the sample size is relatively small, especially after PSM, which further reduces the sample size and may affect the generalization ability of the model. Secondly, there is currently no gold standard blood biomarker for detecting MCI. Currently, the biomarker diagnosis of MCI mainly relies on neurodegenerative disease-related biomarkers such as P-Tau181, P-Tau217, and Aβ42 (Gao Song, Zhang Min & Lu Yuan, 2024; Zheng Yuan & Wen Zhongmin, 2014). Even the combined detection of multiple blood biomarkers used in this study cannot ensure high specificity and sensitivity for the diagnosis of MCI. Meanwhile, the pathological changes caused by MCI as an early stage of neurodegenerative disease may not be obvious (Wu Yue & Cheng Zhaohuo, 2013). Blood biomarkers, as the overall response of the systemic peripheral circulation of the body, are inherently susceptible to more interference in the context of aging, such as chronic low-grade inflammation and immune decline commonly found in the elderly (Li Xiaxia & Ma Lina, 2023; Liu Changhu, Hu Song, Mao Yongjun, et al, 2017), which further leads to greater interference from pathophysiological factors on the biomarkers in this study (Yang Ying, 2017; Gao Danni, Ni Xiaolin, Fang Sihang, et al, 2022), thus making it impossible to accurately locate MCI. Although the subjects were diagnosed and grouped based on neuropsychological scale scores for MCI, the inherent subjectivity may cast doubt on the diagnosis of MCI in the subjects, further increasing the uncertainty of this study. In addition, because this study is a cross-sectional design, it further leads to the inability to determine the causal relationship between blood biomarkers and MCI. Therefore, future research should consider using more biomarkers, expanding the sample size, and adopting a longitudinal design to further verify the high sensitivity, specificity, and long-term value of these biomarkers in predicting MCI. Despite these limitations, the results of this study provide a noninvasive and efficient method for early screening of MCI.

In summary, this study successfully constructed an MCI prediction model based on random forest by combining multi-biomarker joint detection with machine learning methods. The research results indicate that biomarkers have significant potential in the early screening of MCI, and the random forest model demonstrates excellent performance in predicting MCI. This achievement provides new ideas and methods for the early screening and intervention of MCI, and holds important clinical application value. Future research should further expand the sample size, adopt a longitudinal design, and explore more potential biomarkers to further enhance the predictive ability and clinical application value of the model.

5. Conclusion

This study explored the construction of a mild cognitive impairment (MCI) prediction model based on random forest through the combined detection of multiple blood biomarkers and machine learning methods. The research results showed that there were significant differences in TAT, TG, WBC, and P-Tau181 between MCI patients

and healthy control group. The random forest model demonstrated excellent performance in predicting MCI, with an accuracy rate of 87.5% and an AUC-ROC value of 0.92. This achievement not only validates the potential of blood biomarkers in early screening for MCI but also provides theoretical and practical support for the development of non-invasive and efficient MCI prediction tools in the future.

The innovation of this study lies in combining multiple blood biomarkers with machine learning. By efficiently integrating multidimensional data through a random forest model, it significantly enhances the accuracy and reliability of MCI prediction. The research results suggest that the combination of blood biomarkers and machine learning methods holds significant application value in the early screening of MCI, providing an efficient, economical, and accurate tool for large-scale population screening. Furthermore, this study provides an important reference for future exploration of more potential biomarker combinations and optimization of prediction models.

Despite the limitations of this study, such as a small sample size and a cross-sectional design, its findings provide new insights and methods for early screening and intervention of MCI. Future research should further expand the sample size, adopt a longitudinal design, and explore more potential biomarkers to further enhance the predictive power and clinical application value of the model. Overall, this study opens up new avenues for early diagnosis and personalized management of MCI, and holds significant scientific and clinical practical value.

Declaration

This manuscript is original, has not been published elsewhere, and is not under consideration by any other journal. All authors have approved the final version and declare no conflicts of interest.

Our experiment has been approved by the ethics committee, and this is the ethics approval number (2022) Lunshen No.1 (Research).

References

Behrman S, Valkanova V, Allan CL, (2017 May). Diagnosing and managing mild cognitive impairment. *Practitioner, 261*(1804), 17-20.

Cheng Huaidong, Wang Kai, (2009). Current status and issues in the application of neuropsychological scales in China. *Anhui Medical Journal, 30*(02), 143-144.

Gao Danni, Ni Xiaolin, Fang Sihang, et al, (2022). Analysis of chronic inflammatory diseases in middle-aged and elderly Chinese people. *Chinese Journal of Geriatric Health Care and Medicine, 20*(06), 19-22.

Gao Song, Zhang Min, Lu Yuan, (2024). Advances in research on biological markers of mild cognitive impairment. *Chinese Journal of General Practice, 22*(06), 1043-1046+1058.

Guo Shangzhi, Zhang Guangyu, Tang Yuling, (2022). Medical Big Data Analysis and Clinical Application Based on Machine Learning. *Computer Knowledge and Technology, 18*(12), 17-18. DOI: 10.14004/j.cnki.ckt.2022.0730

Huang Jinshan, Zhang Wei, (2019). Mild cognitive impairment and its risk factors. *Internal Medicine, 14*(04), 441-445+458.

Li Min, Ye Lijun, Yu Changfa, et al, (2020). Expression and significance of P-tau-181, IL-1β, and AD7c-NTP in patients with Alzheimer's disease. *China Medical Herald, 17*(11), 116-119. DOI: 10.20047/j.issn1673-7210.2020.11.029.

Li Xiaxia, Ma Lina, (2023). Current Research on Intrinsic Capacity and Chronic Low-Grade Inflammation in the Elderly. *Chinese Journal of Geriatric Multi-organ Diseases, 22*(12), 951-953.

Liu Changhu, Hu Song, Mao Yongjun, et al, (2017). Research progress on frailty in the elderly. *Chinese Journal of General Practice, 20*(16), 2025-2033.

Pang Taiwu, Hu Chunyan, Yin Zhong, (2020). An Improved Random Forest and Its Application in Medical Diagnosis. *Software, 41*(07), 159-163.

Peng Piao, (2017). Research on Variable Importance Measurement and Kernel Density Estimation Algorithm Based on Random Forests. Xiamen University.

Ren Zhen, Li Shu, Zhao Jingjing, et al, (2021). Research progress and prospects of machine learning applications in emergency medicine. *Chinese Journal of Emergency Medicine, 41*(03), 261-265.

Schroer AB, Ventura PB, Sucharov J, Misra R, Chui MKK, Bieri G, Horowitz AM, Smith LK, Encabo K, Tenggara I, Couthouis J, Gross JD, Chan JM, Luke A, Villeda SA, (2023, August). Platelet factors attenuate inflammation and rescue cognition in ageing. *Nature, 620*(7976), 1071-1079.

Shao Li, (2015). Research progress on mild cognitive impairment and its influencing factors. *Foreign Medical Sciences (Medical Geography), 36*(02), 148-154.

Shen Xiaoying, Li Xiaoju, Li Yiyao, et al, (2024). Epidemiological status and influencing factors of mild cognitive impairment in Chinese elderly aged ≥65 years. *Modern Preventive Medicine, 51*(11), 2013-2019+2042.

Wang Yuhui, Shao Fuyuan, (2010). Deepening the Understanding of Cognitive Dysfunction. *World Clinical Drugs, 31*(07), 385-389.

Wu Mengwei, Cao Fengjiao, Yuan Shanghua, (2021). Application of neuropsychological scales in health examination. *China Medical Science, 11*(18), 6-9+39.

Wu Yue, Cheng Zhaohuo, (2013). Research progress on mild cognitive impairment. *Chinese Journal of Geriatrics, 33*(09), 2215-2217.

Yang Rong, Yan Fei, Chen Yang, et al, (2018). Risk factors of mild cognitive impairment. *Chinese General Practice, 21*(12), 1397-1401.

Yang Ying, (2017). Study on the Correlation between Inflammatory Factor Levels and Frailty in Hospitalized Elderly Patients. Chengdu Medical University.

Yu Junchang, Chen Lan, Liu Xiaochang, et al, (2022). A controlled study on serum P-tau181 levels in patients with Alzheimer's disease, vascular dementia, and mild cognitive impairment. *Journal of Clinical Psychological Medicine, 32*(05), 370-373.

Yuan Qinmei, Hong Zhiling, Wang Xing, et al, (2020). Application of artificial intelligence in mental illness. *International Journal of Psychiatry, 47*(01), 4-7.

Zhang Lin, Zhou Wei, Xu Jiajun, et al, (2019). Research progress on intervention methods for mild cognitive impairment. *Chinese Journal of Rehabilitation Medicine, 34*(07), 869-874.

Zhao Jinxuan, Sha Rui, Mi Tianhao, et al., (2020). Advances in diagnosis and rehabilitation therapy of mild cognitive impairment. *Journal of Alzheimer's Disease and Related Disorders, 3*(02), 147-153.

Zhao Mengqi, Liao Hong, (2019). Advances in the study of the relationship between neuroinflammation and cognitive dysfunction in diseases. *Journal of China Pharmaceutical University, 50*(04), 497-504.

Zheng Yuan, Wen Zhongmin, (2014). Meta-analysis of total Tau protein and β42 amyloid protein in cerebrospinal fluid of patients with mild cognitive impairment. *Chinese Journal of Neurological and Psychiatric Diseases, 40*(06), 341-347.