

Construction and Multi-Model Performance Evaluation of an Intelligent Screening Model for Meige Syndrome Based on Machine Learning

Saiqun Zhang

Zhengzhou University

Mengle Peng

Zhengzhou University

Gang Liu

the Third People's Hospital of Henan Province

Yongjun Wu

wuyongjun135@126.com

Zhengzhou University

Research Article

Keywords: Meige syndrome, clinical prediction model, multinomial logistic regression, support vector machine, random forest, artificial neural network

Posted Date: October 22nd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11021999/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Objective

To develop and validate an intelligent screening model for Meige syndrome (MS) based on routine clinical inquiry data, enabling quantitative differentiation among MS patients, patients with other ocular diseases, and normal individuals.

Methods

A total of 3,996 clinical cases were enrolled. Least Absolute Shrinkage and Selection Operator (LASSO) regression was applied to rank the importance of 12 candidate variables. The predictive performance of four models—multinomial logistic regression, support vector machine (SVM), random forest (RF), and artificial neural network (ANN)—was systematically compared with emphasis on goodness-of-fit, predictive accuracy, and probability calibration.

Results

Statistically significant differences across all 12 variables were observed among the three groups (all $P < 0.05$). Importance ranking identified history of other craniofacial conditions, age, sleep quality, history of drinking, and history of thyroid disease as the five most important predictors. Among simplified models built on these five variables, ANN demonstrated optimal overall performance: testing set accuracy of 0.777, AUC of 0.864, and the best probability calibration stability. RF ranked second (accuracy 0.776, AUC 0.858) but exhibited systematic calibration deviation in high-probability intervals. Multinomial logistic regression and SVM showed comparatively limited efficacy. The simplified model closely approximated the full-variable model in diagnostic performance, and decision curve analysis suggested possible meaningful clinical utility.

Conclusion

The simplified ANN model, requiring no laboratory testing and completable within one minute, may serve as a practical auxiliary tool for early MS screening in primary healthcare settings.

1 Introduction

Meige syndrome (MS) was named after the French neurologist Henry Meige and represents a form of segmental craniocervical dystonia. Its cardinal manifestations include blepharospasm, oromandibular dystonia, and spasmodic torticollis^[1–3], which can substantially impair visual function and reduce health-related quality of life^[4–6]. Domestic epidemiological data indicate that the peak incidence occurs

between 55 and 60 years of age, with the prevalence in females approximately two to three times that in males^[7]. Because the clinical features of MS overlap substantially with those of common ocular conditions such as dry eye disease and blepharitis, as well as other neurological disorders^[8–10], accurate diagnosis remains a significant clinical challenge. The delay from symptom onset to definitive diagnosis can extend to several years, seriously impeding early intervention and prognostic improvement.

Currently, the diagnosis of MS still lacks specific objective biomarkers, and clinical diagnosis is primarily based on a comprehensive evaluation of characteristic clinical manifestations^[11, 12]. Recent investigations have attempted to construct predictive nomograms based on blood biomarkers, such as coagulation markers^[13], inflammatory markers^[14] and oxidative stress markers^[15]. These models achieved training set AUC values of 0.748, 0.767, and 0.930, respectively, and testing set AUC values of 0.746, 0.735, and 0.914, respectively, preliminarily demonstrating the diagnostic potential of objective indicators. However, these models largely depend on laboratory testing, limiting their applicability in primary healthcare settings. To overcome this, a nomogram model based on clinical inquiry data was developed, achieving training set and validation set AUC values of 0.789 and 0.800, respectively, establishing a methodological foundation for non-invasive risk assessment^[16]. However, a critical limitation persists: all existing studies (whether based on blood or inquiry data) have adopted a binary classification framework that can only distinguish MS patients from healthy individuals. This fails to address the clinical need to differentiate MS from visually similar ocular diseases. Misdiagnosis of MS as dry eye disease, blephariti, or similar conditions not only leads to ineffective treatment but also delays timely interventions, such as botulinum toxin injection. Accordingly, this study integrated 12 routine clinical inquiry items (including demographic characteristics, lifestyle factors, and medical history) to develop and validate a three-class predictive model for MS, thereby providing a convenient and feasible reference tool for its early screening and differential diagnosis.

2 Materials and methods

2.1 Study design and population

This study adopted a retrospective cross-sectional design, consecutively enrolling patients and the physical examination population who visited the Third People's Hospital of Henan Province between 2020 and 2025. A total of 5,284 clinical cases were initially collected and classified into three groups: the MS group, comprising primary MS patients confirmed by two specialists in Meige syndrome according to established clinical diagnostic criteria; the other eye disease (OED) group, comprising patients with non-MS ocular conditions (including dry eye disease and blepharitis) confirmed by ophthalmologists; and the normal control (NC) group, comprising individuals verified as free of neurological and ocular organic disease by routine health screening. After data integrity verification, the numbers of eligible cases with complete variable data were 1,984, 1,968, and 1,332 in the MS, OED, and NC groups, respectively. As the NC group had the smallest eligible sample ($n = 1,332$), and to ensure balanced three-group composition and minimize the potential confounding effect of class imbalance on model training, 1,332 consecutively

enrolled cases with complete data were also selected from the MS and OED groups in chronological order of visit. Accordingly, 1,332 participants per group were included, yielding a total analytical sample of 3,996 cases.

General inclusion criteria for all participants were age ≥ 18 years. General exclusion criteria for all groups included: comorbid, clearly diagnosed neurodegenerative diseases (e.g., Parkinson's disease, Huntington's disease); and missing key clinical variables.

Based on these general criteria, participants were classified into three groups with specific criteria as follows: MS group: Inclusion required the fulfillment of the clinical diagnostic criteria for primary MS. Exclusion criteria included secondary MS (e.g., caused by prolonged use of neuroleptic agents, cerebrovascular lesions, or encephalitis); OED group: Inclusion required ophthalmologist-confirmed non-MS ocular conditions (including dry eye disease and blepharitis). Exclusion criteria included any form of MS or suspected MS; NC group: Inclusion required verification as free of neurological and ocular organic diseases via routine health screening. Exclusion criteria included a diagnosis or history of any ocular or neurological disorders.

2.2 Data collection

The following 12 clinical variables were extracted from the hospital's electronic medical record (EMR) system. Demographic variables: gender (male/female) and age (years). Lifestyle variables: mental stress (poor/moderate), sleep (poor/moderate), history of smoking (yes/no), and history of drinking (yes/no). Medical history variables: hypertension (yes/no), diabetes (yes/no), cardiovascular disease (yes/no), cerebrovascular disease (yes/no), thyroid disease (yes/no), and other craniofacial conditions involving the oral cavity, eyes, nose, and/or brain (yes/no).

2.3 Predictive model development

The full sample was randomly partitioned into a training set ($n = 2,997$) and a testing set ($n = 999$) at a 3:1 ratio. Four classification models were developed: multinomial logistic regression, SVM, RF, and ANN. Hyperparameter specifications for the three-class task were as follows: multinomial logistic regression employed the multinomial framework with the lbfgs solver and a regularization strength of $C = 0.1$; SVM used a linear kernel function with $C = 1.0$ and balanced class weights; RF comprised 300 decision trees with a maximum depth of 5, a minimum of 15 samples required to split a node, a minimum of 6 samples per leaf node, and balanced class weights; The ANN used three hidden layers with ReLU activation and a maximum of 1,000 training epochs. The neuron counts were dynamically determined according to the number of input features (n), set to $1.5n$, n , and $\max(5, n/2)$; for the 12-feature configuration in this study, this corresponded to 18, 12, and 6 neurons, respectively. All models underwent five-fold cross-validation to ensure robustness.

Model performance was evaluated across multiple dimensions: accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), AUC (computed for the MS group using a one-versus-rest strategy), calibration curves, decision curve analysis(DCA)^[17, 18], and F2 score, which was

adopted as the primary evaluation metric ($\beta = 2$, emphasizing recall over precision) given that missed diagnoses in early MS screening may delay diagnosis and worsen prognosis. The overall study design, data analysis, and reporting followed the TRIPOD statement^[19].

2.4 Statistical analysis

Conventional statistical analyses were performed using SPSS 31.0, and machine learning models were constructed and evaluated using the Scikit-learn framework in Python 3.11. Normally distributed continuous data are presented as mean $\pm$ standard deviation, with intergroup comparisons performed using one-way analysis of variance. Non-normally distributed continuous data are expressed as median and interquartile range [M (P₂₅, P₇₅)], with intergroup comparisons performed using the Kruskal–Wallis test. Categorical data are presented as frequencies and percentages, with intergroup comparisons performed using the chi-square (χ^2) test. Variables with statistically significant differences were entered into multinomial logistic regression to calculate odds ratios (ORs) and 95% confidence intervals (CIs). This regression analysis was intended to evaluate the independent association strength of each variable after adjustment for confounders, serving as a descriptive associative analysis; it differs fundamentally from multinomial logistic regression model in the machine learning framework, which is oriented toward predictive performance. All statistical tests were two-sided, with a significance threshold of $\alpha = 0.05$.

3 Results

Figure 1 illustrates the complete analytical workflow from initial cohort enrollment to final model selection. Among all evaluated algorithms, ANN demonstrated the highest diagnostic efficacy.

3.1 Baseline characteristics of study participants

Baseline clinical characteristics of the three groups are detailed in Table 1. Compared with the other two groups, the MS group was predominantly composed of middle-aged to elderly women; had the highest proportion of poor sleep and the lowest proportions of smokers and alcohol users; and showed the highest prevalences of thyroid disease and cerebrovascular disease among the three groups, whereas the prevalences of hypertension, diabetes, and cardiovascular disease were lower than those of the OED group but higher than those of the NC group. Statistically significant differences in all variables were detected across the three groups.

All 12 variables were entered into multinomial logistic regression to further evaluate their independent discriminatory capacity after adjustment for confounders (Fig. 2). Age, sleep, history of smoking, history of cardiovascular disease, history of cerebrovascular disease, history of thyroid disease, and history of other craniofacial conditions were statistically significant in both the OED-vs-MS and NC-vs-MS comparisons. History of hypertension and diabetes were significant only in the OED-vs-MS comparison, mental stress and history of drinking were significant only in the NC-vs-MS comparison, and gender was not significant in either comparison.

Table 1

Comparison of baseline clinical characteristics among the three groups of study participants.

Variable		MS (n = 1,332)	OED (n = 1,332)	NC (n = 1,332)	χ^2	P-value
Age		58(52,65)	64(53,73)	38(30,49)	1311.462	<0.001
Gender	Female	913(68.5%)	720(54.1%)	672(50.5%)	100.095	<0.001
	Male	419(31.5%)	612(45.9%)	660(49.5%)		
Mental stress	Poor	167(12.5%)	109(8.2%)	236(17.7%)	54.333	<0.001
	Moderate	1165(87.5%)	1223(91.8%)	1096(82.3%)		
Sleep	Poor	509(38.2%)	80(6.0%)	126(9.5%)	566.965	<0.001
	Moderate	823(61.8%)	1252(94.0%)	1206(90.5%)		
Smoking	Yes	143(10.7%)	263(19.7%)	236(17.7%)	44.127	<0.001
	No	1189(89.3%)	1069(80.3%)	1096(82.3%)		
Drinking	Yes	177(13.3%)	261(19.6%)	627(47.1%)	439.746	<0.001
	No	1155(86.7%)	1071(80.4%)	705(52.9%)		
Hypertension	Yes	395(29.7%)	518(38.9%)	179(13.4%)	222.668	<0.001
	No	937(70.3%)	814(61.1%)	1153(86.6%)		
Diabetes	Yes	172(12.9%)	353(26.5%)	64(4.8%)	254.780	<0.001
	No	1160(87.1%)	979(73.5%)	1268(95.2%)		
Cardiovascular disease	Yes	133(10.0%)	244(18.3%)	63(4.7%)	127.651	<0.001
	No	1199(90.0%)	1088(81.7%)	1269(95.3%)		
Cerebrovascular disease	Yes	146(11.0%)	117(8.8%)	20(1.5%)	99.354	<0.001
	No	1186(89.0%)	1215(91.2%)	1312(98.5%)		
Thyroid disease	Yes	386(29.0%)	105(7.9%)	97(7.3%)	324.134	<0.001
	No	946(71.0%)	1227(92.1%)	1235(92.7%)		
Other craniofacial conditions	Yes	498(37.4%)	1007(75.6%)	16(1.2%)	1564.109	<0.001
	No	834(62.6%)	325(24.4%)	1316(98.8%)		

Note: Following data integrity verification, the number of participants with poor mental stress(n = 236, 17.7%) and the number with history of smoking (n = 236, 17.7%) in the normal control group were identical. Secondary verification against the original database confirmed this to be a statistical coincidence attributable to similar distributions of these two behavioral characteristics within this subgroup population, and was not the result of data entry error.

3.2 Comparison of model performance

To evaluate the collective predictive capacity of the 12 baseline variables, four machine learning models were constructed using the full cohort (n = 3,996) to perform three-class differentiation among the MS, OED, and NC groups. Table 2 presents the performance of the four full-variable (12-variable) models on the training and testing sets, serving as a reference baseline for subsequent simplified models and as a basis for variable importance analysis. Using testing-set accuracy and AUC as primary evaluation metrics, model performance ranked from highest to lowest as follows: ANN, RF, multinomial logistic regression, and SVM.

Table 2
Performance comparison of four predictive models on training and testing sets.

Model	Dataset	Accuracy	Sensitivity	Specificity	PPV	NPV	F2 Score	AUC
Logistic	train	0.773	0.628	0.846	0.671	0.820	0.636	0.841
	test	0.766	0.619	0.839	0.658	0.815	0.626	0.841
SVM	train	0.740	0.675	0.773	0.598	0.826	0.658	0.816
	test	0.752	0.712	0.772	0.609	0.843	0.689	0.819
RF	train	0.793	0.717	0.831	0.679	0.854	0.709	0.872
	test	0.772	0.685	0.815	0.650	0.838	0.677	0.858
ANN	train	0.826	0.740	0.869	0.738	0.870	0.739	0.904
	test	0.774	0.706	0.808	0.647	0.846	0.693	0.860

3.3 Variable importance assessment and ranking

Although the full-variable models demonstrated robust diagnostic performance, requiring 12 clinical inputs may increase the assessment burden and limit efficiency in primary healthcare settings. To construct a more streamlined and practical screening tool, feature reduction was necessary. To quantitatively assess the relative contribution of each candidate variable to the three-class decision, LASSO regression was applied within multinomial logistic regression framework to rank the 12 variables by importance. All variables were standardized using StandardScaler prior to modeling to eliminate scale disparities between the continuous variable (age) and binary variables, ensuring comparability of regression coefficients. The optimal regularization parameter was determined through five-fold cross-validation over a candidate parameter space to obtain the model with the best generalization performance^[20–22]. The absolute values of regression coefficients for each variable under each pairwise comparison were then extracted, and their mean value was used as an approximate quantification of each variable's relative contribution, on the basis of which variables were ranked and presented in a bar chart (Fig. 3).

The results showed that other craniofacial conditions exerted the greatest influence on model classification decisions and played a dominant role in the three-class differentiation, a result consistent with the finding that this variable showed the largest intergroup differences in the baseline analysis. Age, sleep, history of drinking, and history of thyroid disease constituted the next most important predictors in descending order, with the remaining variables contributing progressively less. Based on these findings, the five highest-ranked variables were selected for inclusion in the simplified model, balancing predictive performance with clinical feasibility.

3.4 Five-fold cross-validation results for the five-variable simplified models

3.4.1 Overall predictive performance

The five selected variables were incorporated into each of the four modeling approaches. To further assess the generalization stability of each model, five-fold cross-validation was performed, and comprehensive performance metrics are presented in Fig. 4. Using testing set accuracy and AUC as primary metrics, model performance ranked from highest to lowest as follows: ANN, RF, multinomial logistic regression, and SVM, a ranking consistent with the full-variable models. Five-fold cross-validation results were aligned with testing set trends for all models, indicating stable performance.

3.4.2 Discriminative ability

After five-fold cross-validation, all four models achieved AUC values ≥ 0.795 on the validation sets, indicating satisfactory overall discriminative ability (see supplementary Figs. 1). In terms of validation set AUC, RF (0.852) slightly outperformed ANN (0.850), while multinomial logistic regression (0.825) and SVM (0.795) ranked comparatively lower. It should be noted that on the independent testing set, ANN achieved a higher AUC (0.864) than RF (0.858), yielding a slightly different ranking order across the two evaluation frameworks. Overall, both ANN and RF demonstrated strong discriminative capacity.

3.4.3 Calibration performance

Calibration curves generated after five-fold cross-validation revealed notable differences in probability calibration among the four models (see supplementary Figs. 2). Overall, ANN demonstrated the best calibration performance, followed by RF, while SVM and multinomial logistic regression showed comparatively inferior calibration.

3.4.4 Clinical decision value

DCA derived from five-fold cross-validation results demonstrated that, across threshold probabilities from 0 to 0.8, the net benefit of all four models was positive (see supplementary Figs. 3), indicating that utilizing these models for early MS screening can effectively assist physicians in accurately identifying true MS cases for timely interventions while minimizing the misclassification of non-MS individuals.

4 Discussion

This study is the first to apply a three-class classification framework to MS predictive modeling, addressing the methodological gap left by existing studies that have been limited to binary classification, and enabling direct quantitative differentiation between MS and ocular diseases with overlapping symptomatology. Although previously published blood biomarker models reported high AUC values, direct numerical comparisons are methodologically constrained, as those studies focused on the less complex binary task of distinguishing MS patients from healthy controls. Tackling the inherently more challenging three-class differentiation inherently alters the baseline metrics. Nevertheless, the simplified model presented here accomplishes this complex task using only five routine clinical inquiry items and requires no laboratory testing, substantially improving accessibility and practicality in primary healthcare settings.

Among all variables, other craniofacial conditions demonstrated the strongest discriminative power across the three groups: its prevalence was as high as 75.6% in the OED group but only 1.2% in the NC group, with differences of substantial statistical and clinical significance. Rather than directly reflecting the underlying pathological mechanisms of MS, this finding likely highlights the complex clinical overlap within craniofacial disorders. As MS is a typical craniofacial dystonia characterized by progressive symptom spread from the eyelids to the face and neck, concurrent craniofacial conditions may arise as secondary complications of prolonged spasms or reflect shared regional neuromuscular vulnerabilities^[23, 24]. Age is also an important discriminatory variable. Prior research has shown that the peak incidence of MS clusters around 55–60 years of age^[25], whereas the NC group in this study had a median age of only 38 years, conferring strong population-level discriminatory power upon this variable. Furthermore, the differences in sleep status among the three groups present clear clinical implications: the proportion of poor sleep quality in the Meige syndrome group was significantly higher than that in the other two groups, which is consistent with clinical observations that central nervous system dysfunction affects the sleep regulatory network^[26, 27].

It is noteworthy that the proportion of participants reporting poor mental stress was higher in the NC group (17.7%) than in the MS group (12.5%) or the OED group (8.2%). This seemingly counterintuitive finding can be explained from two perspectives: on the one hand, the NC group predominantly consisted of working-age adults (median age 38 years) who are objectively exposed to concurrent occupational and family-related stressors; on the other hand, MS patients with prolonged illness may develop adaptive psychological coping mechanisms that stabilize subjective stress ratings. Nonetheless, this variable was assessed using a patient-rated binary scale with limited information granularity, restricting deeper interpretation. Despite this, mental stress remained statistically significant in multinomial logistic regression for discriminating between the MS group and the NC group ($OR = 1.921, P < 0.05$). The substantial difference in history of drinking between the MS group (13.3%) and the NC group (47.1%; $OR = 5.454, P < 0.05$) was primarily attributable to demographic compositional differences: the MS group was predominantly composed of older adult women (68.5% female, median age 58 years), whereas the NC group consisted mainly of working-age adults with higher rates of alcohol exposure. Additionally,

clinical observations indicate that certain dystonia subtypes exhibit transient symptomatic responsiveness to alcohol^[28, 29]; this symptom-relieving effect may be associated with the observed differences in history of drinking, although the causal direction remains unclear and warrants further investigation. Notably, the prevalence of thyroid disease in the MS group was 29.0%, significantly exceeding that of the OED (7.9%) and NC (7.3%) groups. While the exact causal links remain to be fully elucidated, previous reports suggest that thyroid dysfunction and associated autoimmune inflammation may be intimately linked to the clinical manifestation of craniocervical dystonia^[30, 31]. These findings additionally underscore the potential clinical utility of routine thyroid function screening for patients with MS.

Regarding the overall performance of the four predictive models, ANN achieved the best results (testing set AUC = 0.864, accuracy = 0.777), followed by RF (AUC = 0.858), with multinomial logistic regression (AUC = 0.841) and SVM (AUC = 0.819) performing comparatively less well. This ranking was highly consistent across the training set, testing set, and cross-validation set, reflecting result stability rather than random variation. The advantage of ANN stems primarily from its feature self-learning capability and multilayer nonlinear mapping mechanism, which effectively captures complex interaction effects among multidimensional clinical features^[32–34]. Recent clinical prediction studies have also confirmed that neural network architectures tend to provide superior fit and predictive accuracy compared with traditional linear models when processing medical inquiry data characterized by complex nonlinear interaction effects^[35, 36]. RF performed well in terms of discriminative ability but showed systematic deviation in the high-probability interval of the calibration curve, an inherent limitation arising from the aggregated averaging mechanism of ensemble tree models^[37, 38]. Multinomial logistic regression offers strong clinical interpretability but its linear assumptions may prove insufficient for capturing the combinatorial effects of multidimensional risk factors in MS^[39, 40]. When handling the large-sample data in this study, which consisted predominantly of discrete categorical variables, SVM encountered difficulty constructing an optimal three-class hyperplane through kernel function mapping, limiting its discriminative efficacy^[41–43].

Of particular note, the simplified ANN (five variables, AUC = 0.864) marginally outperformed the full-variable ANN (12 variables, AUC = 0.860), with the two models showing a high degree of consistency across accuracy, calibration curves, and DCA (see Figs. 4–7). This result corroborates the necessity of the feature selection step: by eliminating low-contribution variables, parameter optimization becomes more focused on core discriminative information, yielding more robust generalization. From a clinical translation perspective, the simplified model requires only five readily obtainable items, which enhances its potential feasibility for initial triage in primary healthcare settings. Therefore, it constitutes the preferred implementation approach.

The principal strengths of this study are as follows. First, this is the first three-class predictive model for MS constructed exclusively from clinical inquiry information, as previous studies have predominantly focused on binary classification frameworks. Second, the relatively large sample ($n = 3,996$) with

balanced three-group allocation effectively controlled for the potential impact of class imbalance on model performance.

Several limitations should be acknowledged. First, as the model was developed from single-center retrospective data, only internal validation was performed, which may have introduced selection bias; multicenter, prospective, and independent external cohort validation is needed to confirm predictive efficacy and clinical applicability. Second, the MS and OED groups were formed by chronologically selecting the first 1,332 eligible cases, precluding the inclusion of all eligible participants during the study period, which may have introduced some time-truncation bias. Third, some subjective variables (e.g., mental stress status, sleep) were self-rated binary items, resulting in notable information loss and susceptibility to recall bias and subjective error. Future research should conduct prospective external validation in independent cohorts and further improve model generalizability through systematic hyperparameter optimization.

5 Conclusion

This study was the first to introduce a three-class classification framework into MS predictive modeling. Based on 3,996 multi-group clinical cases, feature importance analysis selected five optimal inquiry variables: history of other craniofacial conditions, age, sleep, history of drinking, and history of thyroid disease. These variables were used to develop and internally validate a simplified predictive model capable of three-way differentiation among MS, other eye diseases, and normal controls without reliance on laboratory testing. Among the four modeling approaches, ANN demonstrated the best overall performance, with a testing set AUC of 0.864 and accuracy of 0.777. RF ranked second in discriminative ability (AUC 0.858, accuracy 0.776) but showed systematic calibration deviation in high-probability intervals. Multinomial logistic regression and SVM performed comparatively less well. Five-fold cross-validation results were highly consistent with testing set trends, and decision curve analysis confirmed genuine clinical decision benefits for all models. The simplified ANN model developed in this study requires no laboratory testing, can be implemented within one minute, and may serve as a practical and effective auxiliary tool for early screening of MS in primary healthcare settings.

Declarations

This study was conducted in accordance with the Declaration of Helsinki, and was approved by the Ethics Committee of the Ethics Committee of the Third People's Hospital of Henan Province (Approval No. 2025SZSYLCYJ0807). The requirement for informed consent was waived because of the retrospective nature of the study and anonymized data analysis.

Funding Declaration

The authors have no relevant financial or non-financial interests to disclose. This study was funded by the Third People's Hospital of Henan Province. Clinical trial number: not applicable.

Author Contribution

Saiqun Zhang wrote the manuscript. Mengle Peng and Gang Liu and Yongjun Wu reviewed the manuscript.

References

1. Zhou Y, Huo H, Yang S, et al. Acupuncture for Meige syndrome: A case report. *Explore*. 2024;20:103068. <https://doi.org/10.1016/j.explore.2024.103068>.
2. Zhou Y, Yang Z, Liu D, et al. Characteristics of the cortico-striato-thalamo-cerebellar structural covariance network in Meige syndrome. *Neuroscience*. 2025;580:351–8. <https://doi.org/10.1016/j.neuroscience.2025.06.067>.
3. Li X, Ma Y, Lin H, et al. Computed tomography navigation radiofrequency ablation for Meige's syndrome: A game-changer in treatment. *Asian J Surg*. 2024;48:281–6. <https://doi.org/10.1016/j.asjsur.2024.09.042>.
4. Liu J, Li L, Li Y, et al. Metabolic Imaging of Deep Brain Stimulation in Meige Syndrome. *Front Aging Neurosci*. 2022;14:848100. <https://doi.org/10.3389/fnagi.2022.848100>.
5. Zheng H, Wu L, Tian S, et al. Effect of botulinum toxin type A on non-motor symptoms and quality of life in Meige syndrome. *Front Neurol*. 2023;14:1115482. <https://doi.org/10.3389/fneur.2023.1115482>.
6. Liu J, Ding H, Xu K, et al. Micro lesion effect of pallidal deep–brain stimulation for meige syndrome. *Sci Rep*. 2022;12:19980. <https://doi.org/10.1038/s41598-022-23156-2>.
7. Wei S, Liu X. Chinese Expert Consensus on the Diagnosis and Treatment of Meige Syndrome (2025). *Chin J Ocular Fundus Dis*. 2025;41:418–27. <https://doi.org/10.3760/cma.j.cn511434-20250414-00169>.
8. Yatika C, Saket T, Ragini P. Diagnostic Dilemma: Unraveling Meige Disorder Mistaken for Functional Neurological Disorder. *Cureus*. 2024;16:e61465. <https://doi.org/10.7759/cureus.61465>.
9. Teng X, Qu Q, Shu Y, et al. Genetic screening in patients of Meige syndrome and blepharospasm. *Neurol Sci*. 2022;43:3683–94. <https://doi.org/10.1007/s10072-022-05900-8>.
10. Zhu G, Yin Z, Chen Y, et al. Subthalamic Nucleus Oscillatory Characteristics in Meige, Cervical Dystonia and Generalized Dystonia. *Ann Clin Transl Neurol*. 2025;12:1082–92. <https://doi.org/10.1002/acn3.70040>.
11. Wang Y, Yang A, Song Z, et al. Abnormal brain networks in Meiges syndrome based on centrality analysis and functional network connectivity: a cross-sectional analysis. *Brain Imaging Behav*. 2025;19:899–910. <https://doi.org/10.1007/s11682-025-01000-1>.
12. Li M, Li J, Li M, et al. The Associations Between Depression, Subjective Sleep Quality, and Health-Related Quality of Life in Patients with Meige Syndrome: A Cross-Sectional Study. *Nat Sci Sleep*. 2025;17:3053–63. <https://doi.org/10.2147/NSS.S558119>.

13. Chen Y, Sun Y, Zhang X, et al. A nomogram based on coagulation markers for predicting meige syndrome risk. *Front Aging Neurosci.* 2025;17:1665276. <https://doi.org/10.3389/fnagi.2025.1665276>.
14. Fu R, Lian W, Zhang B, et al. Development and Validation of a Nomogram Based on Inflammatory Markers for Risk Prediction in Meige Syndrome Patients. *J Inflamm Res.* 2024;17:7721–31. <https://doi.org/10.2147/JIR.S481649>.
15. Zhu Y, Fu R, Wang Z, et al. Development and validation of predictive models for meige syndrome patients based on oxidative stress markers. *Front Immunol.* 2025;16:1536109. <https://doi.org/10.3389/fimmu.2025.1536109>.
16. Liu G, Guo Q, Xiang J, et al. Analysis of risk factors for Meige syndrome and construction and validation of a clinical prediction nomogram model. *Front Neurol.* 2026;17:1755240. <https://doi.org/10.3389/fneur.2026.1755240>.
17. Andrew JV, Ford H. Decision curve analysis to evaluate the clinical benefit of prediction models. *Spine J.* 2021;21:1643–8. <https://doi.org/10.1016/j.spinee.2021.02.024>.
18. Ben VC, Gary SC, Andrew JV, et al. Evaluation of performance measures in predictive artificial intelligence models to support medical decisions: overview and guidance. *Lancet Digit Health.* 2025;7:100916. <https://doi.org/10.1016/j.landig.2025.100916>.
19. Gary SC, Karel GMM, Paula D, et al. TRIPOD + AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ.* 2024;385:e078378. <https://doi.org/10.1136/bmj-2023-078378>.
20. Huang T, He W, Xie Y, et al. A LASSO-derived clinical score to predict severe acute kidney injury in the cardiac surgery recovery unit: a large retrospective cohort study using the MIMIC database. *BMJ Open.* 2022;12:e060258. <https://doi.org/10.1136/bmjopen-2021-060258>.
21. Quan X, Deng Y, Liu Z, et al. LASSO-derived nomogram prediction model for lymph node metastasis in colorectal cancer: a retrospective analysis. *PeerJ.* 2025;13:e19148. <https://doi.org/10.7717/peerj.19148>.
22. Cui W, Bin X, Ning L, et al. Construction and validation of a regularized logistic regression model for predicting late recurrence after ablation in persistent atrial fibrillation: a clinical parameter approach. *Front Cardiovasc Med.* 2025;12:1713560. <https://doi.org/10.3389/fcvm.2025.1713560>.
23. Karen G, Christine K. Dystonia updates: definition, nomenclature, clinical classification, and etiology. *J Neural Transm (Vienna).* 2021;128:395–404. <https://doi.org/10.1007/s00702-021-02314-2>.
24. Jamir PR, Andrew M, Shi Y, et al. Neurobiology of Dystonia: Review of Genetics, Animal Models, and Neuroimaging. *Brain Sci.* 2025;15:767. <https://doi.org/10.3390/brainsci15070767>.
25. Ma H, Qu J, Ye L, et al. Blepharospasm, Oromandibular Dystonia, and Meige Syndrome: Clinical and Genetic Update. *Front Neurol.* 2021;12:630221. <https://doi.org/10.3389/fneur.2021.630221>.
26. Yi M, Li J, Liu G, et al. Mental health and quality of life in patients with craniofacial movement disorders: A cross-sectional study. *Front Neurol.* 2022;13:938632. <https://doi.org/10.3389/fneur.2022.938632>.

27. Xie H, Shan M, Wang S, et al. Deep brain stimulation of the subthalamic nucleus for primary Meige syndrome: clinical outcomes and predictive factors. *J Neurosurg.* 2024;140:1650–63. <https://doi.org/10.3171/2023.11.JNS232075>.
28. Anna K, Selma S, Karen G, et al. Functional and Molecular Properties of DYT-SGCE Myoclonus-Dystonia Patient-Derived Striatal Medium Spiny Neurons. *Int J Mol Sci.* 2021;22:3565–86. <https://doi.org/10.3390/ijms22073565>.
29. Diana NK, Steven JF, Kristina S. Alcohol responsiveness in laryngeal dystonia: a survey study. *J Neurol.* 2015;262:1548–56. <https://doi.org/10.1007/s00415-015-7751-2>.
30. Gamze K-B, Laura MS, Ami R, et al. Thyroid disease in cervical dystonia. *Parkinsonism Relat Disord.* 2023;107:105274. <https://doi.org/10.1016/j.parkreldis.2022.105274>.
31. Susanne AS, Lea T, Nicole R. Thyroid Disorders and Movement Disorders-A Systematic Review. *Mov Disord Clin Pract.* 2023;10:360–8. <https://doi.org/10.1002/mdc3.13656>.
32. Mao W, Jiayu Lyu, Deep KV, et al. Application of artificial neural networks in detection and diagnosis of gastrointestinal and liver tumors. *World J Clin Cases.* 2020;8:3971–7. <https://doi.org/10.12998/wjcc.v8.i18.3971>.
33. Celina SS, Irina-Georgeta S, Cristina MG, et al. Exploring the Intersection of Artificial Intelligence and Clinical Healthcare: A Multidisciplinary Review. *Diagnostics (Basel).* 2023;13:1995–2032. <https://doi.org/10.3390/diagnostics13121995>.
34. Abdullahi AI, Sahalu BJ, Hussaini M, et al. DeepClassifier: A Data Sampling-Based Hybrid BiLSTM-BiGRU Neural Network for Enhanced Type 2 Diabetes Prediction. *Comput Model Eng Sci.* 2026;146:37–70. <https://doi.org/10.32604/cmesci.2026.076187>.
35. Bowen C, Chi X, Chen N, et al. Prediction of STN-DBS for Parkinson's disease by uric acid-related brain function connectivity: A machine learning study based on resting state function MRI. *Front Aging Neurosci.* 2023;15:1105107. <https://doi.org/10.3389/fnagi.2023.1105107>.
36. Ding X, Frederick S, Richard K, et al. Comparison of neural network and logistic regression for dementia prediction: results from the PREADViSE trial. *J Gerontol Geriatr.* 2021;69:137–46. <https://doi.org/10.36150/2499-6564-n311>.
37. Oya Kg, Serhat EA, Menelaos P, et al. Evaluating the sample size requirements of tree-based ensemble machine learning techniques for clinical risk prediction. *Stat Methods Med Res.* 2025;34:1356–72. <https://doi.org/10.1177/09622802251338983>.
38. Yubin P, Joyce CH. (2020) CaliForest: Calibrated Random Forest for Health Data. *Proc ACM Conf Health Inference Learn (2020)* 2020:40–50. <https://doi.org/10.1145/3368555.3384461>
39. Paula D, Ma J, Constanza LAN, et al. Methodological conduct of prognostic prediction models developed using machine learning in oncology: a systematic review. *BMC Med Res Methodol.* 2022;22:101–17. <https://doi.org/10.1186/s12874-022-01577-x>.
40. Fumi I, Koutarou M, Ryu M, et al. Predictive Performance of Machine Learning-Based Models for Poststroke Clinical Outcomes in Comparison With Conventional Prognostic Scores: Multicenter, Hospital-Based Observational Study. *JMIR AI.* 2024;3:e46840. <https://doi.org/10.2196/46840>.

41. Mark AZ, Jonathan RB, Ronald JJ. Using information theory to optimize a diagnostic threshold to match physician-ordering practice. J Biomed Inf. 2021;117:103756.
<https://doi.org/10.1016/j.jbi.2021.103756>.

42. Rosita G, Stefania F, Danilo L, et al. An Overview on the Advancements of Support Vector Machine Models in Healthcare Applications: A Review. Information. 2024;15:235–71.
<https://doi.org/10.3390/info15040235>.

43. Jin X, Wang Y, Wang J, et al. A Machine Learning Approach to Differentiate Cold and Hot Syndrome in Viral Pneumonia Integrating Traditional Chinese Medicine and Modern Medicine: Machine Learning Model Development and Validation. JMIR Med Inf. 2025;13:e64725.
<https://doi.org/10.2196/64725>.

Figures

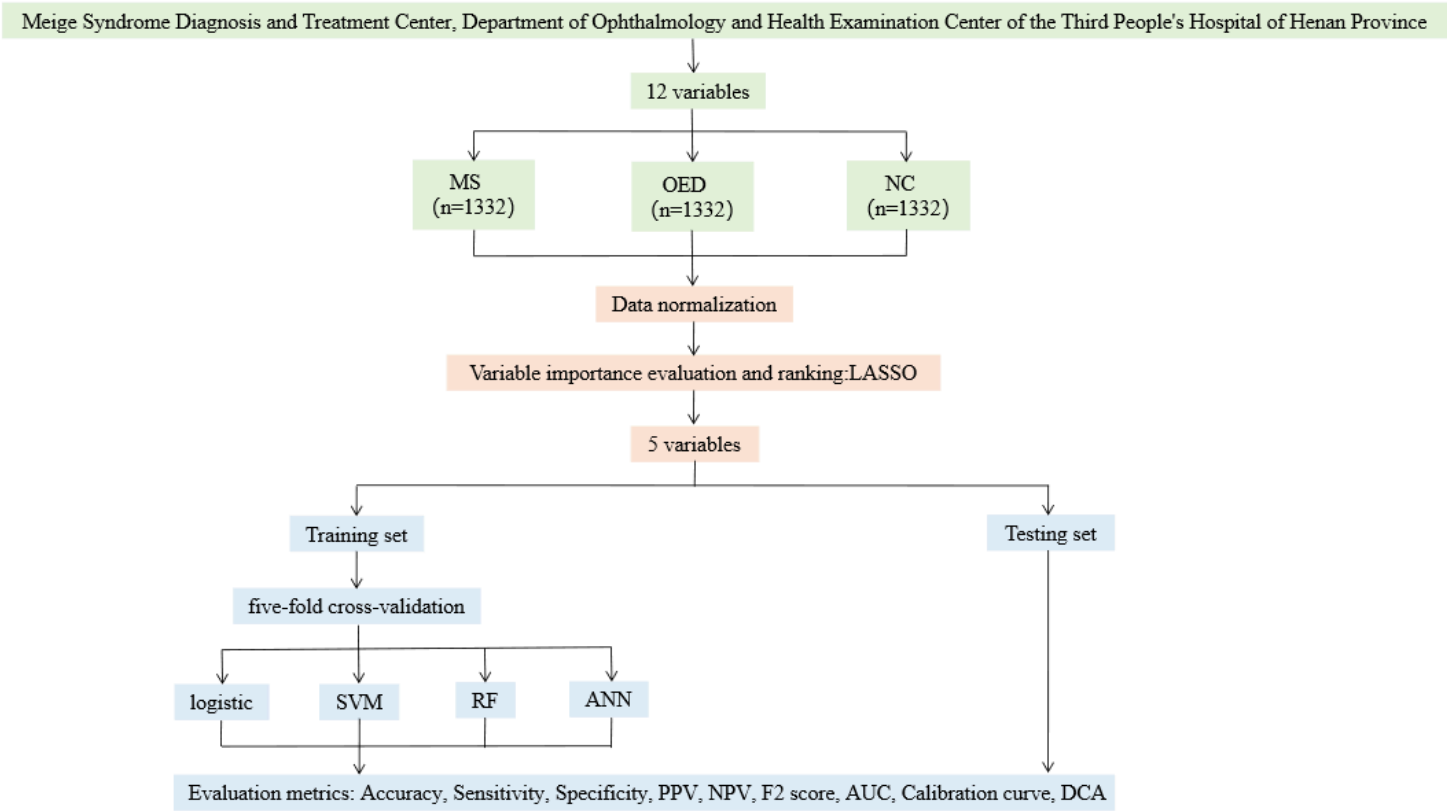


Figure 1

Flowchart of the study design and analytical process for the intelligent screening model of MS.

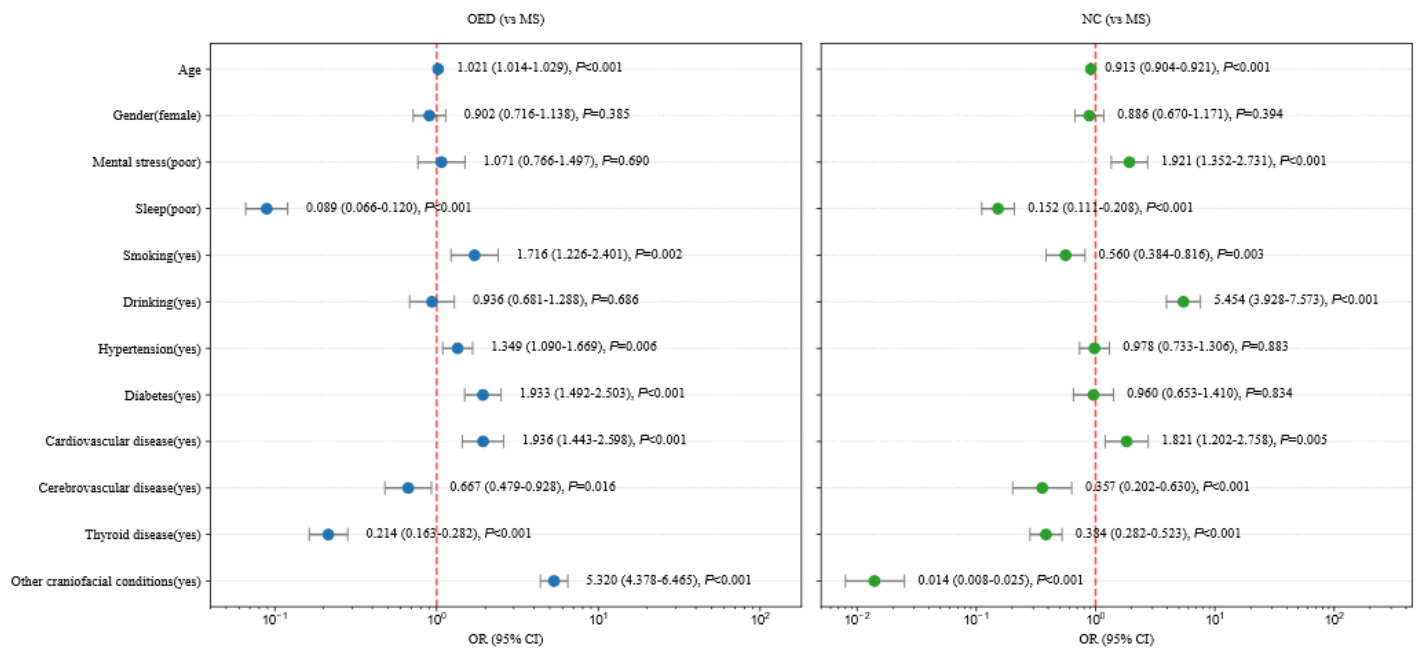


Figure 2

Forest plot of multivariable multinomial logistic regression analysis, with MS as the reference group.

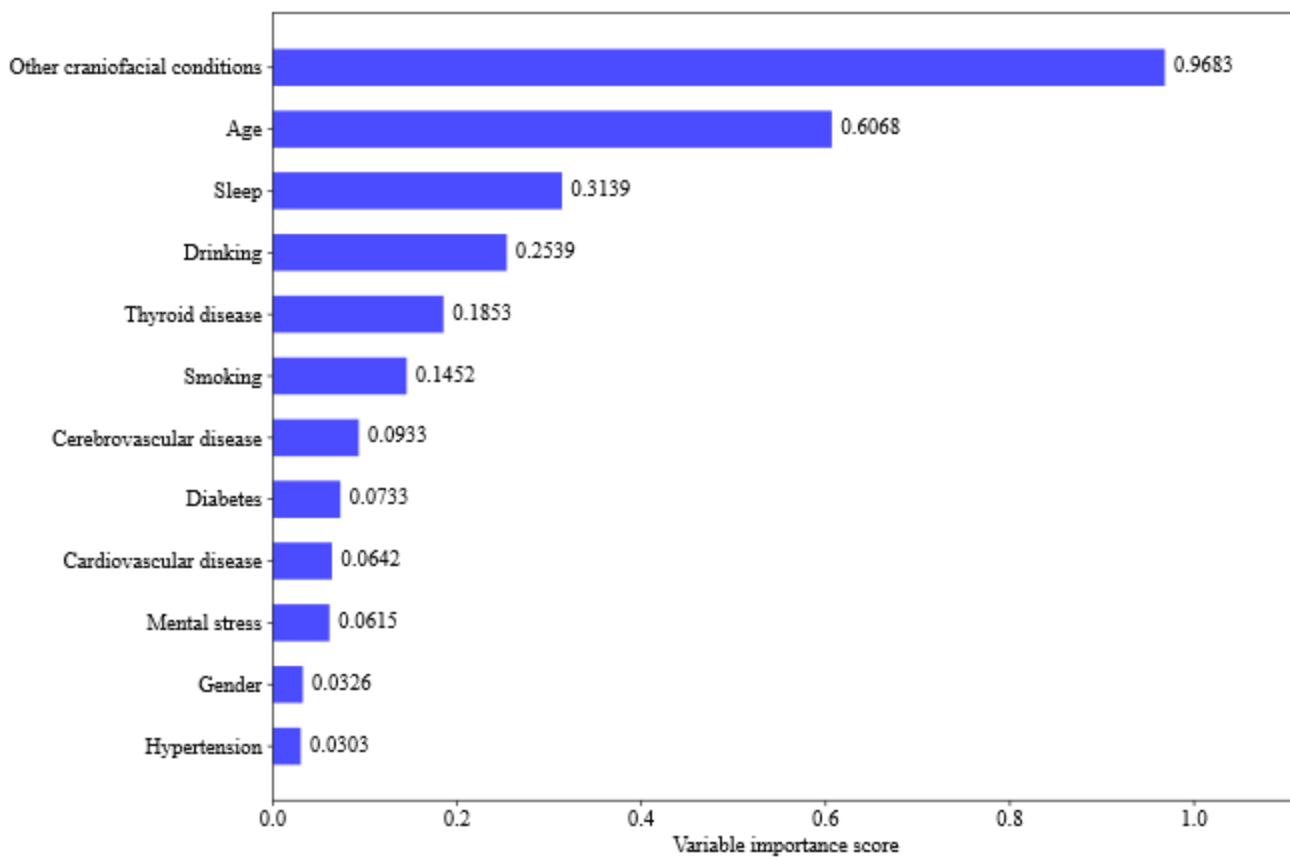


Figure 3

Bar chart of LASSO coefficient-based importance ranking for all predictor variables.

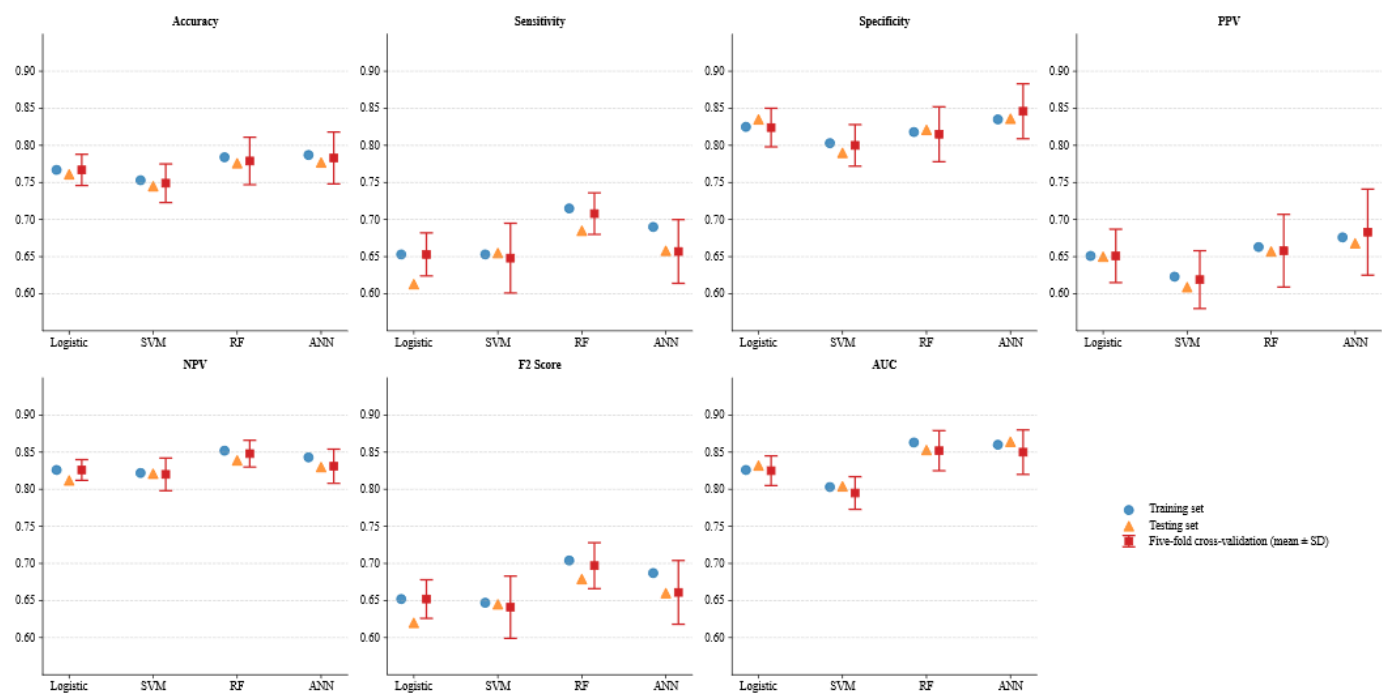


Figure 4

Performance comparison of the five-variable simplified models across different datasets.

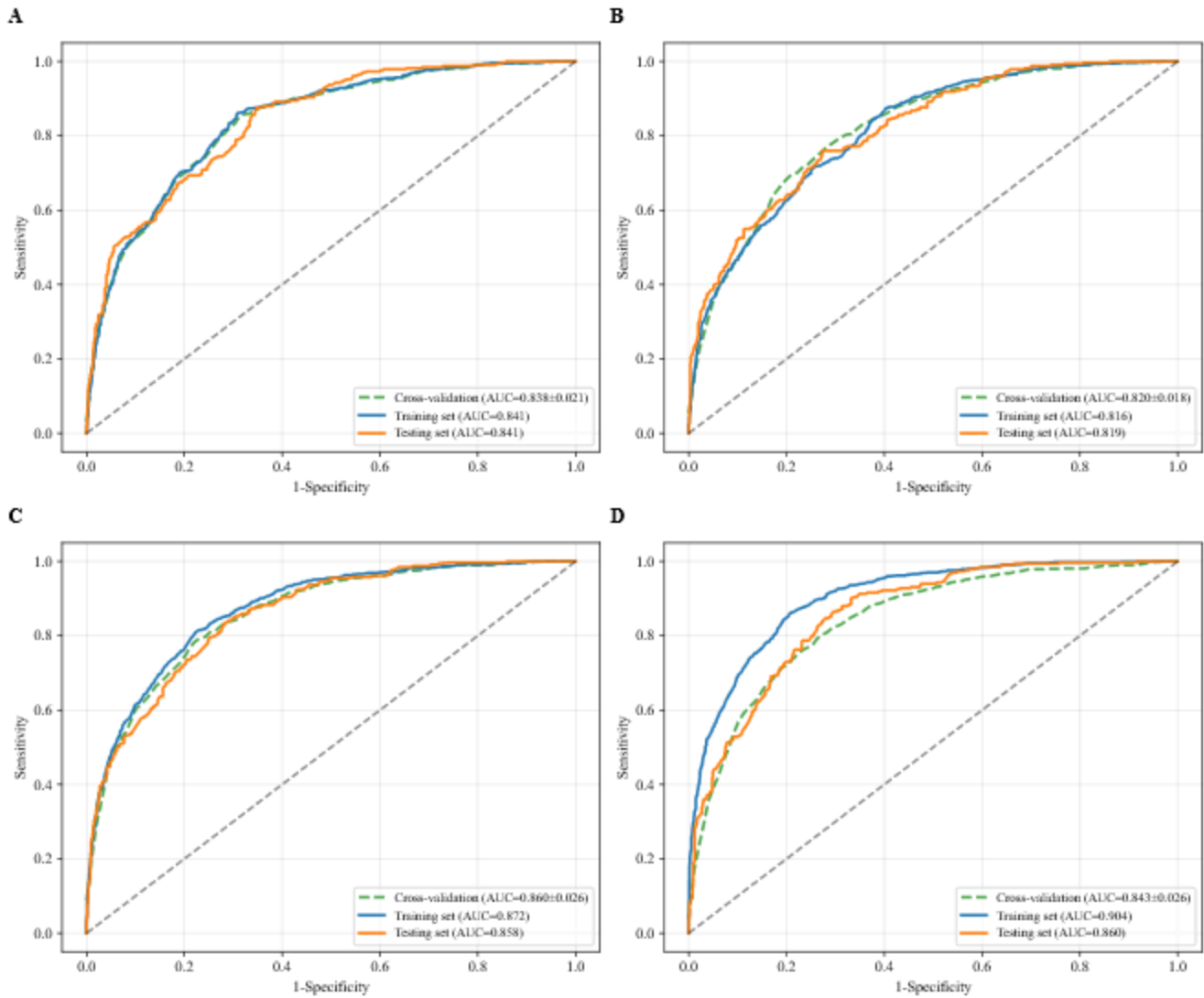


Figure 5

ROC curves of the full-variable models across different datasets. (A) Multinomial logistic regression; (B) SVM; (C) RF; (D) ANN.

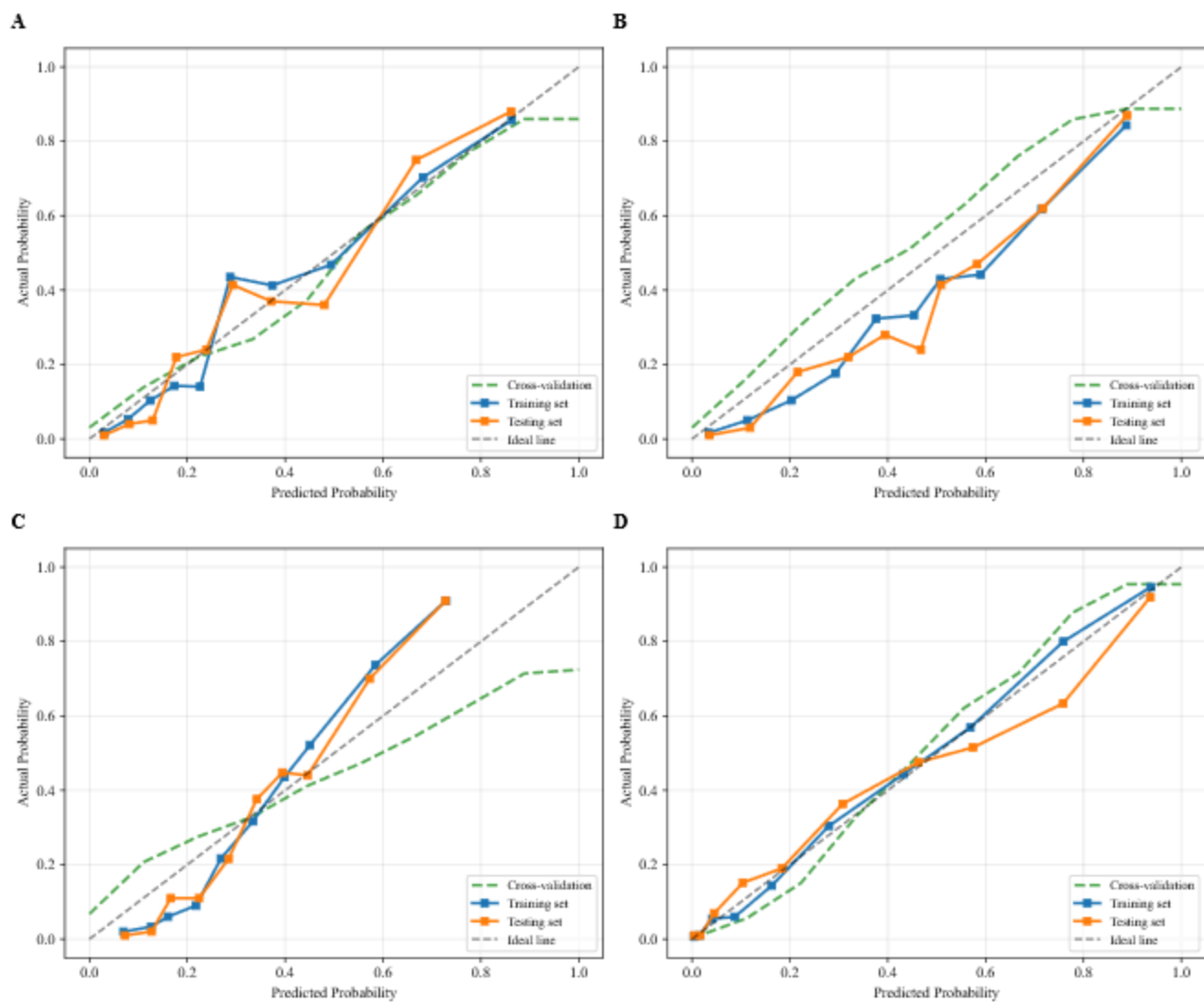


Figure 6

Calibration curve of the full-variable models across different datasets. (A) Multinomial logistic regression; (B) SVM; (C) RF; (D) ANN.

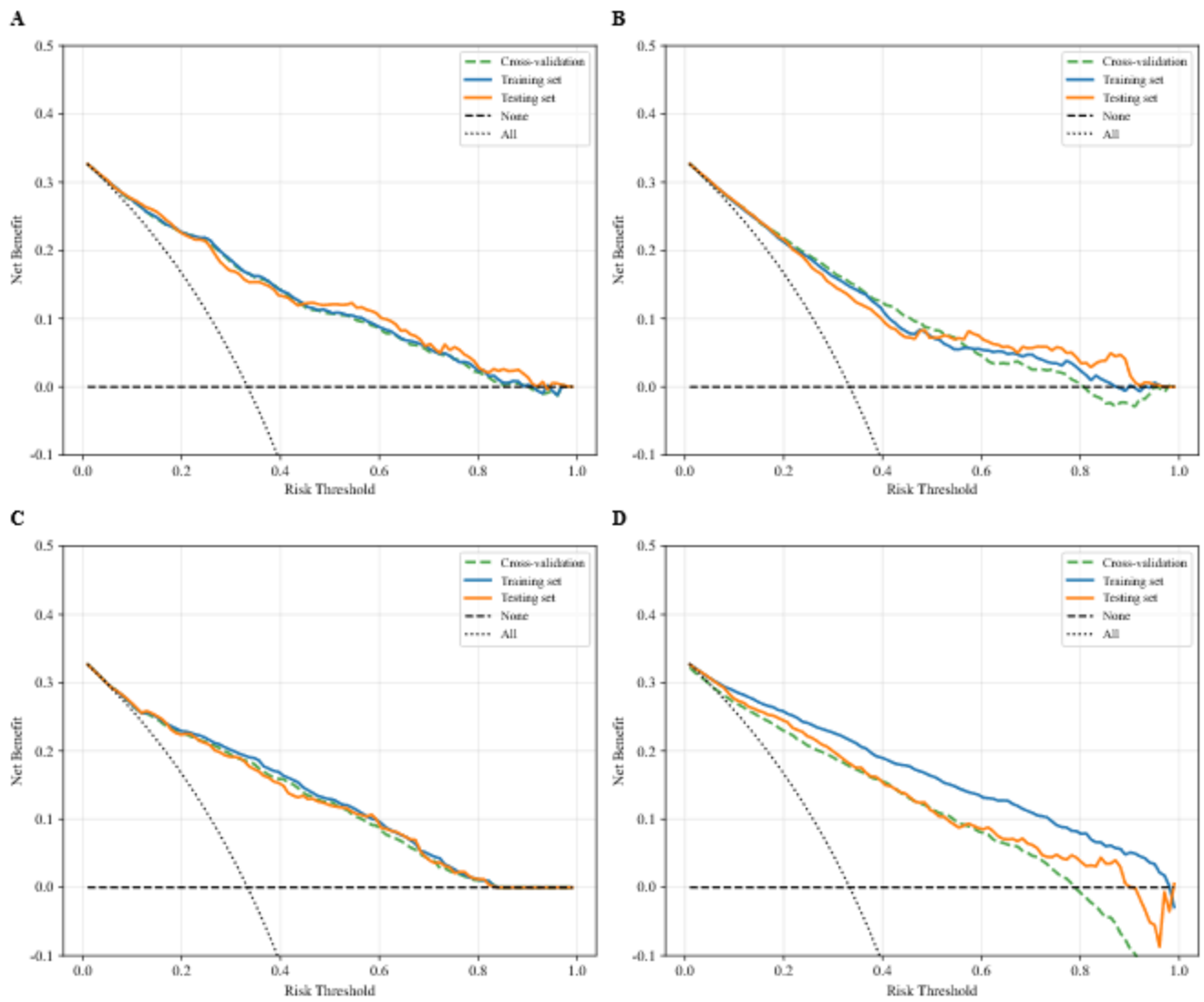


Figure 7

DCA curve of the full-variable models across different datasets. (A) Multinomial logistic regression; (B) SVM; (C) RF; (D) ANN.