Proceedings of Machine Learning Research vol 245:1–13, 2024 2024 International Conference on Machine Learning and Intelligent Computing

Construction and Evaluation of a Metabolic Syndrome Prediction Model Based on Classification Algorithms

Zhenghao Zhao
Department of Computer Technology and Application, Qinghai University, Xining, China

Qiongqiong Hu* 2010990036@QHU.EDU.CN
Department of Computer Technology and Application, Qinghai University, Xining, China

Editors: Nianyin Zeng and Ram Bilas Pachori

Abstract

Exploring the risk factors influencing Metabolic Syndrome (MS), constructing a risk prediction model based on multiple classification algorithms, comparing the predictive performance of different models for MS, and interpreting the models to derive specific MS prediction rules, providing scientific basis for MS prediction. A retrospective analysis was conducted on clinical data from 2,193 MS patients. Based on whether the patients developed MS, they were divided into a healthy group and an MS group. Statistical correlation tests were used to identify the risk factors associated with MS. Six classification algorithms, including decision trees, logistic regression, random forests, naive Bayes, K-nearest neighbors, and support vector machines, were employed to build an MS prediction model. The prediction model's performance was evaluated using R language by generating receiver operating characteristic (ROC) curves. Among the 2,193 MS patients, the incidence rate of MS was 34.66%. Significant differences (P¡0.05) were observed between the healthy group and the MS group in terms of age, marital status, income, ethnicity, waist circumference, body mass index (BMI), uric acid levels, blood glucose levels, high-density lipoprotein levels, and triglyceride levels. Blood glucose, waist circumference, BMI, and triglycerides showed a significant linear correlation with MS. The ROC curve results indicated that the random forest algorithm achieved an area under the curve (AUC) of 0.94 (95% CI: 0.914-0.957), logistic regression achieved an AUC of 0.90 (95% CI: 0.867-0.925), support vector machines achieved an AUC of 0.89 (95% CI: 0.859-0.920), decision trees achieved an AUC of 0.87 (95% CI: 0.831-0.905), K-nearest neighbors achieved an AUC of 0.81 (95% CI: 0.770-0.850), and naive Bayes achieved an AUC of 0.74 (95% CI: 0.694-0.785). The study results confirmed that factors such as age, marital status, waist circumference, BMI, blood glucose levels, and triglyceride levels are all risk factors for developing MS. Furthermore, the random forest and logistic regression models demonstrated excellent performance in predicting MS.

Keywords: Metabolic Syndrome, prediction model, classification algorithm, statistical analysis, Nomogram

1. Introduction

Metabolic Syndrome (MS) is a metabolic disorder characterized by the coexistence of multiple metabolic conditions. Its primary features include abdominal obesity, elevated fasting blood glucose, hypertension, increased triglycerides, and decreased high-density lipoprotein cholesterol (Mohseni-Takalloo et al., 2024; Tkachenko et al., 2023; HU, 2024). Approximately one-fourth of the global population is affected by MS, and its prevalence is rapidly increasing due to economic development and lifestyle changes. Asians are more prone to MS compared to individuals of European and American descent. MS patients have a significantly higher incidence and mortality rate of

cardiovascular diseases and type 2 diabetes compared to non-MS individuals, making it a significant public health concern worldwide (Chen and Lin, 2023; Zhang et al., 2024).

Numerous studies, both domestic and international, have investigated the factors and prediction models related to MS. Zheng et al. (2023) found a significant correlation between uric acid levels and the incidence of MS in elderly individuals. Li et al. (2018) suggested that the prevalence of MS is associated with age and the intake of fungi and algae. Huang et al. (2024) highlighted the impact of factors such as BMI, type 2 diabetes, fasting blood glucose, and triglycerides on MS. Tang et al. (2023) emphasized the close relationship between diet and the occurrence and development of MS. Furthermore, Zhang et al. developed an MS risk prediction model using K-nearest neighbors and logistic regression algorithms but did not provide detailed accuracy information. Chen and Lin (2023) visualized the model as a column chart using the R language, but the usage details were not elaborated.

Currently, research on MS remains relatively limited and lacks comprehensive analysis. Moreover, many prediction studies on classification models do not mention accuracy. In this study, we employed various classification algorithms and statistical methods to establish predictive models for MS and analyze the generated nomogram for interpretation. The aim was to comprehensively analyze and understand the multiple factors contributing to MS and construct accurate prediction models, providing scientific evidence for early diagnosis, treatment planning, and disease management. We conducted a comprehensive analysis of 13 indicators influencing MS and constructed six classification models for predicting MS occurrence, comparing and analyzing the performance of each model. Finally, using the R language, we visualized the nomogram of the predictive model and provided explanations for the prediction rules.

2. OBJECTIVES AND METHODS

2.1. Objectives

The research focused on a publicly available dataset of MS patient information obtained from a data repository. The dataset included demographic, clinical, and laboratory measurements, as well as information on the presence or absence of metabolic syndrome. Based on the presence or absence of MS, the dataset was divided into a healthy group and an MS group. After data preprocessing, the healthy group consisted of 1433 individuals, while the MS group comprised 760 individuals. Among the features to be analyzed, the first column containing the sequential identification number and the last column containing the MS label were excluded, leaving 13 remaining features. These features included age, gender, marital status, income status, race, waist circumference, Body Mass Index (BMI), albuminuria, albumin-to-creatinine ratio in urine, uric acid, blood glucose, high-density lipoprotein, and triglycerides. Among these features, gender, marital status, and race were categorical data, while the others were continuous data.

2.2. Methods

The correlation analysis between the factors and MS, as well as the comparison of classification prediction model performance, were conducted using statistical methods. For continuous data, normal distribution and homogeneity of variances were examined. If both assumptions were met, a t-test was applied. If the assumptions were not met, the Mann-Whitney U test was used. For categorical data, the chi-square test was employed. The comparison of classification prediction models was

performed using ROC curves, comparing the Area Under the Curve (AUC) for different models. A higher AUC value indicates better model performance. AUC values ranging from 0.7 to < 0.8 suggest acceptable discriminative ability, AUC values from 0.8 to < 0.9 indicate good discriminative ability, and AUC ≥ 0.9 suggests excellent discriminative ability.

Six classification algorithms were employed in this study. Decision tree is a classification and regression algorithm based on a tree-like structure. It recursively partitions the dataset to construct a tree, with each internal node representing a feature and each leaf node representing a class or value. It is characterized by interpretability and ease of understanding (Mi et al., 2024). Random forest is an ensemble learning algorithm composed of multiple decision trees. It exhibits good generalization and overfitting resistance, making it suitable for high-dimensional data and complex classification problems (Zhan et al., 2023). K-nearest neighbors is an instance-based classification algorithm. It is simple and intuitive, and performs well for problems with non-linear and complex decision boundaries ([13). Logistic regression is a statistical regression algorithm used for binary or multi-class problems. It is commonly used for prediction and classification tasks (Wichitaksorn et al., 2023). Support vector machine is a machine learning algorithm for binary classification and regression. It finds an optimal hyperplane in the feature space that maximizes the projection of samples from different classes, achieving classification or regression tasks (Song et al., 2024). Naive Bayes is a classification algorithm based on the Bayesian theorem and the assumption of feature independence. It is simple and efficient, performing well when dealing with large-scale datasets (Sui et al., 2023).

3. DATA PREPROCESSING

3.1. Data Exploration

The data in this study can be categorized into two types: continuous and categorical data. For the continuous data, descriptive statistics such as maximum, minimum, mean, and standard deviation were calculated. For the categorical data, the different types of categories were examined.

As shown in Figure 1, the distribution of all continuous data appears to be relatively dispersed. Particularly notable are the bar charts in (b) for income status and (e) for the albumin-to-creatinine ratio in urine. The bar chart of the albumin-to-creatinine ratio in urine not only exhibits a highly dispersed data distribution, but also a large range of values. Additionally, the majority of data points are smaller than 100, indicating the presence of outliers. The cleaning of these outliers is described in Section 3.2.

There are three types of categorical data: gender, marital status, and race. Marital status has several possible values, including single, married, widowed, divorced, separated, and other. Race has values such as white, Asian, black, Mexican-American, Hispanic, and other races. Gender has only two possible values, and no corresponding pie chart was generated during the exploratory phase.

3.2. Data Processing

During the data exploration process, it was observed that there were missing values in the variables of income status, waist circumference, BMI, and marital status. For income status, waist circumference, and BMI, the missing values were imputed using the mean value. However, for marital status, as it is a categorical variable, it was not possible to impute missing values with the mean, so the

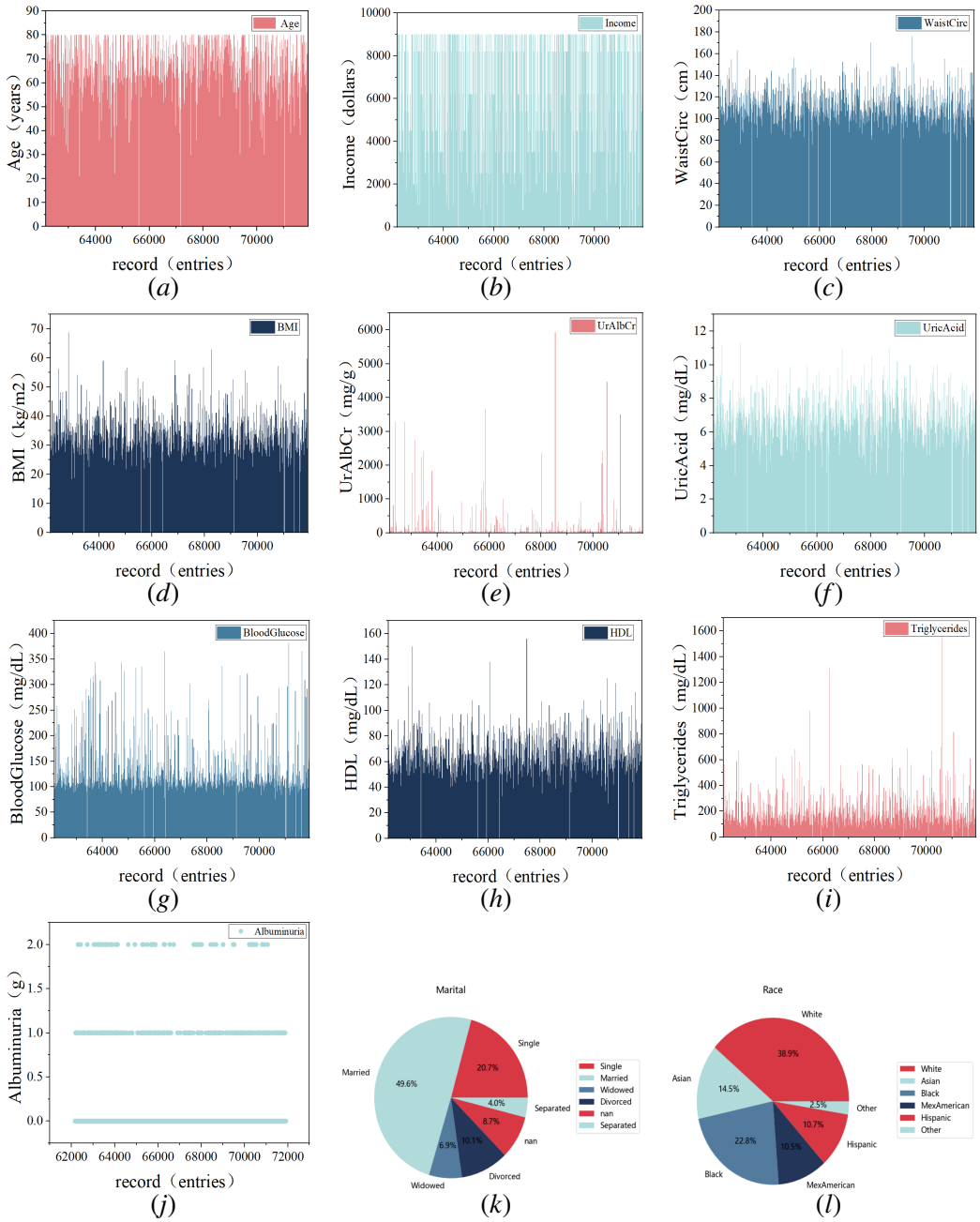

Figure 1: Bar Chart of Metric Data (a-i), Scatter Plot of Albuminuria (j), Pie Charts of Categorical Data (k-l).

entire records with missing marital status were directly deleted. After handling the missing values, the dataset was reduced from 2402 records to 2193 records.

During the data exploration, the presence of outliers was observed. Boxplots and quartiles were used to identify and handle outliers in this study. Considering that removing outliers for certain features may affect the classification performance, this study only focused on cleaning outliers for features with a small number of outlier records that did not exhibit clear distribution patterns and were unlikely to significantly impact the classification performance. For example, for the feature " albuminuria," as shown in Figure 1 (j), it can be clearly seen that it only has three distinct values: 0, 1, and 2. Removing outliers for this feature may affect the classification performance.

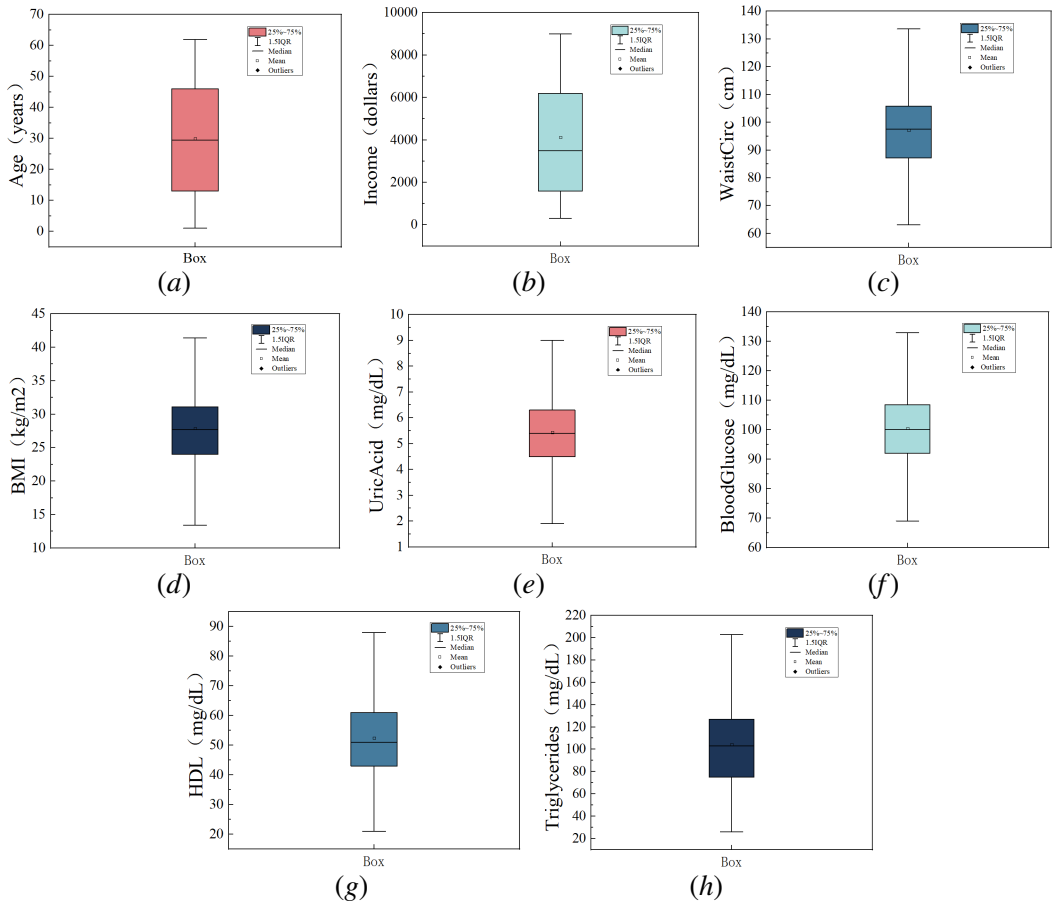

Figure 2: Boxplot after Outlier Removal.

As shown in Figure 2, after removing outliers, there are no outliers beyond the upper and lower boundaries. The horizontal line at the bottom represents the minimum value, while the horizontal line at the top represents the maximum value. The bottom boundary of the box represents the first quartile (Q1), the middle line inside the box represents the second quartile (Q2) or the median, and the top boundary of the box represents the third quartile (Q3).

4. RESULTS

4.1. Feature Analysis

The dataset was divided into two groups, the healthy group ($n = 1433$) and the MS group ($n = 760$), based on whether the patients had MS. For continuous variables, a normality test and a test for homogeneity of variances were conducted. If both assumptions were met, a t-test was used; otherwise, non-parametric statistical methods, specifically the Mann-Whitney U test, were employed. For categorical variables, a chi-square test was directly applied. The Shapiro-Wilk test was used for normality testing, and if the p-value was less than 0.05, the data were considered not normally distributed. The Levene test was used for testing homogeneity of variances, and if the p-value was less than 0.05, the data were considered to have unequal variances. The results of the normality test and homogeneity of variances test for the continuous variables are shown in Tables 1, 2 and 3. The statistical values are reported with three decimal places. The results indicate that for age (normality test: healthy group statistic = 0.950, $p < 0.05$; MS group statistic = 0.968, $p < 0.05$. Homogeneity of variances test: statistic = 23.535, p¡0.05), income status (normality test: healthy group statistic = 0.872, $p < 0.05$; MS group statistic = 0.872, $p < 0.05$. Homogeneity of variances test: statistic = 10.835, $p < 0.05$), waist circumference (normality test: healthy group statistic = 0.983, $p < 0.05$; MS group statistic = 0.986, $p < 0.05$. Homogeneity of variances test: statistic = 14.329, $p < 0.05$), BMI (normality test: healthy group statistic = 0.978, $p < 0.05$; MS group statistic=0.985, p¡0.05. Homogeneity of variances test: statistic = 2.371, $p > 0.05$), urine albumin (normality test: healthy group statistic = 0.339, $p < 0.05$; MS group statistic = 0.519, $p < 0.05$. Homogeneity of variances test: statistic = 53.475, $p < 0.05$), urine albumin-to-creatinine ratio (normality test: healthy group statistic = 0.089, $p < 0.05$; MS group statistic = 0.191, $p < 0.05$. Homogeneity of variances test: statistic = 15.041, $p < 0.05$), uric acid (normality test: healthy group statistic = 0.992, $p < 0.05$; MS group statistic = 0.993, $p < 0.05$. Homogeneity of variances test: statistic = 0.052, $p > 0.05$), blood glucose (normality test: healthy group statistic = 0.980, $p < 0.05$; MS group statistic = 0.971, $p < 0.05$. Homogeneity of variances test: statistic = 7.060, $p < 0.05$), high-density lipoprotein (normality test: healthy group statistic = 0.984, $p < 0.05$; MS group statistic = 0.952, $p < 0.05$. Homogeneity of variances test: statistic = 8.624, $p < 0.05$), and triglycerides (normality test: healthy group statistic = 0.973, $p < 0.05$; MS group statistic = 0.982, $p < 0.05$. Homogeneity of variances test: statistic = 2.369, $p > 0.05$), both groups showed p-values less than 0.05, indicating that the data did not follow a normal distribution. Therefore, the Mann-Whitney U non-parametric test was used. The detailed results for each variable are presented in Tables 1, 2 and 3.

The Mann-Whitney U test was performed for continuous variables, and the chi-square test was conducted for categorical variables. The results showed that age (statistic = 375894.0, $p < 0.05$), marital status (statistic = 31.608, $p < 0.05$), income status (statistic = 603344.5, $p < 0.05$), race (statistic = 30.394, $p < 0.05$), waist circumference (statistic = 229373.0, $p < 0.05$), BMI (statistic = 259422.5, $p < 0.05$), urine albumin (statistic = 484164.0, $p < 0.05$), urine albumin-to-creatinine ratio (statistic = 419901.5, $p < 0.05$), uric acid (statistic = 395531.5, $p < 0.05$), blood glucose (statistic = 205184.5, $p < 0.05$), high-density lipoprotein (statistic = 811214.0, $p < 0.05$), and triglycerides (statistic = 270777.5, $p < 0.05$) were all found to have p-values less than 0.05, indicating statistical significance. These factors were identified as having potential predictive value and were included in the MS risk prediction model. Gender, except for gender (statistic = 0.791, $p > 0.05$), did not emerge as a significant factor in the occurrence of MS, which is consistent with

Table 1: Comparison of Normality Tests for Indicators in the Healthy Group.

Indicator	Statistic	P-value
Age	0.950	1.3736440748629741e-21
Income	0.872	1.0375014526745268e-32
WaistCirc	0.983	5.8159028808180135e-12
BMI	0.978	4.372551786250255e-14
Albuminuria	0.339	0.0
UrAlbCr	0.089	0.0
UricAcid	0.992	2.462253121393587e-07
BloodGlucose	0.980	2.70676872842629e-13
HDL	0.984	1.708581019721489e-11
Triglycerides	0.973	1.2259649946697732e-15

Table 2: Comparison of Normality Tests for Indicators in the MS Group.

Indicator	Statistic	P-value
Age	0.968	9.802320413698773e-12
Income	0.872	1.6845256883998964e-24
WaistCirc	0.986	7.83736993525963e-07
BMI	0.985	6.165275863168063e-07
Albuminuria	0.519	4.006312309504652e-41
UrAlbCr	0.191	0.0
UricAcid	0.993	0.0009104780037887394
BloodGlucose	0.971	3.153376518238993e-11
HDL	0.952	5.825161621781597e-15
Triglycerides	0.982	3.3458071868608386e-08

Table 3: Comparison of Homogeneity of Variance Tests for Indicators.

Indicator	Statistic	P-value
Age	23.535	1.3127663478251223e-06
Income	10.853	0.0010115626164418881
WaistCirc	14.329	0.000157624764718148
BMI	2.371	0.12374040373244082
Albuminuria	53.475	3.6534535428452546e-13
UrAlbCr	15.041	0.00010829221810200355
UricAcid	0.052	0.8199224220393394
BloodGlucose	7.060	0.007940355953755026
HDL	8.624	0.003351997897306312
Triglycerides	2.369	0.12390777306333184

the findings reported by Li et al. (2018) The results of the tests for each feature are presented in Table 4.

Table 4: Results of Feature Indicator Tests.

Indicator	Statistic	P-value
Age	375894.0	6.18355869650907e-33
Sex	0.791	0.37366148292281753
Marital	31.608	2.3000862239794745e-06
Income	603344.5	2.8555282848974843e-05
Race	30.394	1.2334653165225025e-05
WaistCirc	229373.0	1.5950247817990886e-110
BMI	259422.5	8.619237426963492e-91
Albuminuria	484164.0	5.292041286920807e-13
UrAlbCr	419901.5	1.0218016607586587e-18
UricAcid	395531.5	4.458412855971216e-26
BloodGlucose	205184.5	5.622327716488222e-128
HDL	811214.0	1.0658175178794017e-79
Triglycerides	270777.5	7.112565837628927e-84

4.2. Correlation Analysis

Based on the Mann-Whitney U test conducted for the continuous variables, the Pearson correlation coefficient (r) was calculated to determine the more significant factors associated with MS. The Pearson correlation coefficient (r) indicates the strength and direction of linear association between variables. A correlation coefficient (r) with $|r| < 0.4$ indicates a low degree of linear correlation, $0.4 =< |r| < 0.7$ indicates a moderate degree of linear correlation, and $0.7 =< |r| < 1$ indicates a high degree of linear correlation. The correlation coefficients (r) are presented in Figure 3, and the corresponding values in the last column of the heatmap represent the correlation coefficient between each factor and MS. The results showed that age ($r = 0.25$), income (r$= -0.093$), urine albumin-to-creatinine ratio ($r = 0.085$), uric acid ($r = 0.23$), and high-density lipoprotein ($r = -0.38$) exhibited a low degree of linear correlation with MS. Blood glucose ($r = 0.49$), waist circumference ($r = 0.46$), BMI ($r = 0.42$), and triglycerides ($r = 0.42$) demonstrated a significant linear correlation with MS. In other words, blood glucose, waist circumference, BMI, and triglycerides are important factors associated with the occurrence of MS.

4.3. Classification Prediction Models

A total of 10 continuous variables selected through the Mann-Whitney U test and 2 categorical variables (marital status and race) identified through the chi-square test were employed as independent variables, while the occurrence of MS served as the dependent variable. By integrating six classification algorithms, an MS risk prediction model was constructed.

4.4. Classification Performance and Evaluation

In this study, the performance of the classification prediction models was evaluated based on three aspects: accuracy, time consumption, and AUC value. The random forest algorithm achieved an accuracy of 84%, with a time consumption of 352 milliseconds and an AUC value of 0.92 (95% CI: 0.899-0.945). The logistic regression algorithm attained an accuracy of 82%, with a time consumption of 15 milliseconds and an AUC value of 0.90 (95% CI: 0.865-0.923). The support vector machine algorithm yielded an accuracy of 80%, with a time consumption of 70 milliseconds and an AUC value of 0.89 (95% CI: 0.859-0.920). The decision tree algorithm achieved an accuracy of 79%, with a time consumption of 9 milliseconds and an AUC value of 0.86 (95% CI: 0.818-0.892). The K-nearest neighbors algorithm obtained an accuracy of 75%, with a time consumption of 11 milliseconds and an AUC value of 0.81 (95% CI: 0.770-0.854). The naive Bayes algorithm demonstrated an accuracy of 65%, with a time consumption of 3 milliseconds and an AUC value of 0.74 (95% CI: 0.689-0.786). According to the results, the random forest algorithm exhibited the highest accuracy and the largest AUC value, demonstrating superior classification ability. Although the random forest algorithm had a relatively higher time consumption of 352 milliseconds, it remained acceptable for this dataset. In contrast, the naive Bayes algorithm showed lower accuracy and a smaller AUC value, indicating relatively poorer performance. The numerical values of the model performance indicators are presented in Table 5.

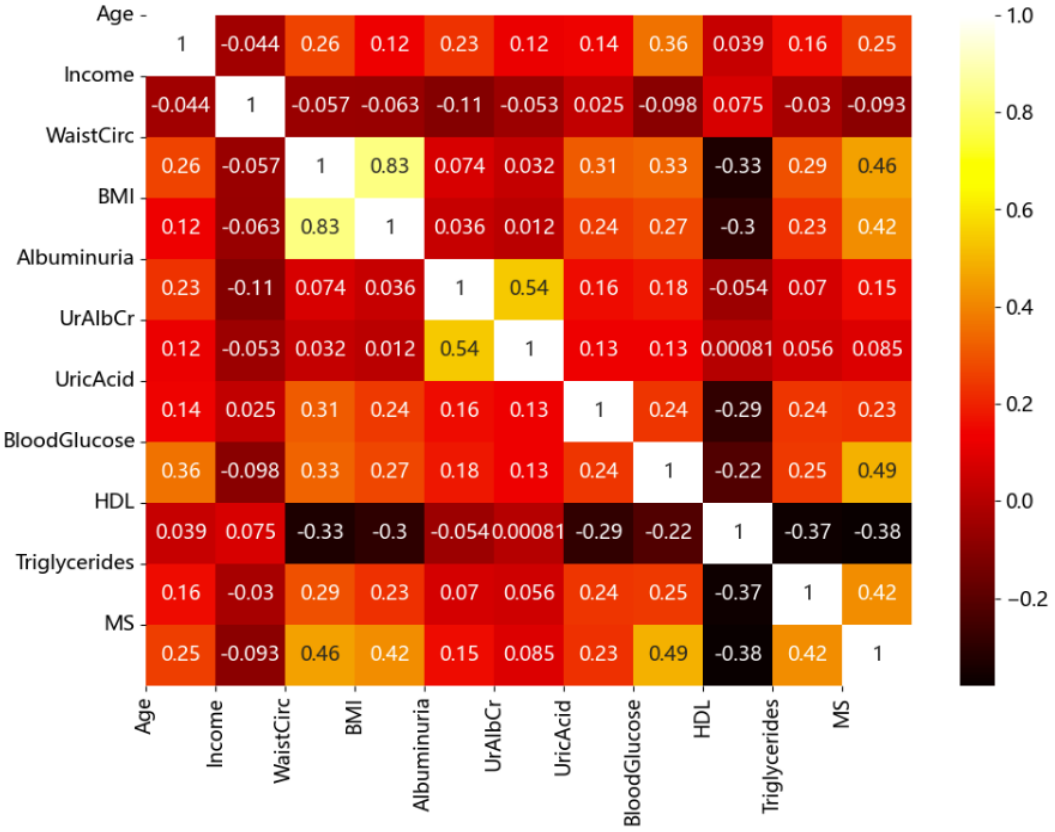

Figure 3: Correlation Heatmap.

Table 5: Performance indicators of the classification prediction models.

Algorithm	Accuracy (%)	Time Consumption (ms)	AUC
Random Forest	0.84	0.352	0.92(95%CI: 0.899õ.945)
Logistic Regression	0.82	0.015	0.90(95%CI: 0.865õ.923)
Support Vector Machine	0.80	0.070	0.89(95%CI: 0.859õ.920)
Decision Tree	0.79	0.009	0.86(95%CI: 0.818õ.892)
K-Nearest Neighbors	0.75	0.011	0.81(95%CI: 0.770õ.854)
Naive Bayes	0.65	0.003	0.74(95%CI: 0.689õ.786)

The closer the ROC curve is to the upper-left corner and the larger the area under the curve (AUC), the better the performance of the classification algorithm. As shown in Figure 4, the ROC curve of the random forest algorithm is closest to the upper-left corner and has the largest area under the curve, with an AUC value of 0.92, indicating the best performance. The logistic regression algorithm follows with an AUC value of 0.90.

4.5. Nomogram

The predictive nomogram for determining whether the research subjects have MS can be plotted using the R programming language and a logistic regression model. Due to the negligible contribution of income to the model, it can be essentially ignored. The nomogram after removing income is shown in Figure 5.

Each predictive variable can be assigned a specific score on the corresponding scoring axis. The scores for the 11 variables are then summed to obtain a total score. Finally, the total score is mapped onto the MS prediction axis through the Total Points scoring axis, yielding the probability of a patient having MS. For example, in the dataset, the first record has an age of 22, marital status as single, race as white, waist circumference of 81, BMI of 23.3, urine albumin of 0, urine albumin-to-creatinine ratio of 3.88, uric acid of 4.9, blood glucose of 92, HDL of 41, and triglycerides of 84. The total score is 146, which corresponds to a probability significantly less than 10% on the MS prediction axis. Therefore, this research subject is determined to be part of the healthy population. For detailed information, please refer to Figure 5.

5. CONCLUSION

MS is a common metabolic disorder that is closely associated with the risk of cardiovascular disease and type 2 diabetes, making it a significant public health concern worldwide.

In this study, it was found that blood glucose, waist circumference, BMI, and triglycerides exhibited significant linear correlations with MS. This suggests that these factors play important roles in the development of MS. These results are consistent with previous research and further validate the significance of blood glucose, waist circumference, BMI, and triglycerides in the occurrence of MS.

To predict the risk of MS, a predictive model using six classification algorithms was established in this study. Through comparative analysis, it was found that the random forest algorithm performed the best. This indicates that the random forest algorithm can achieve excellent performance in predicting MS.

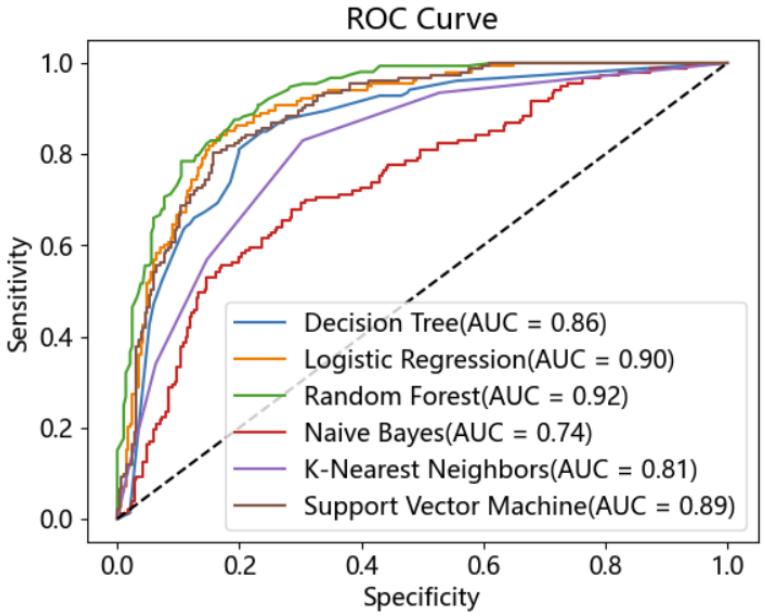

Figure 4: ROC curves of the predictive models.

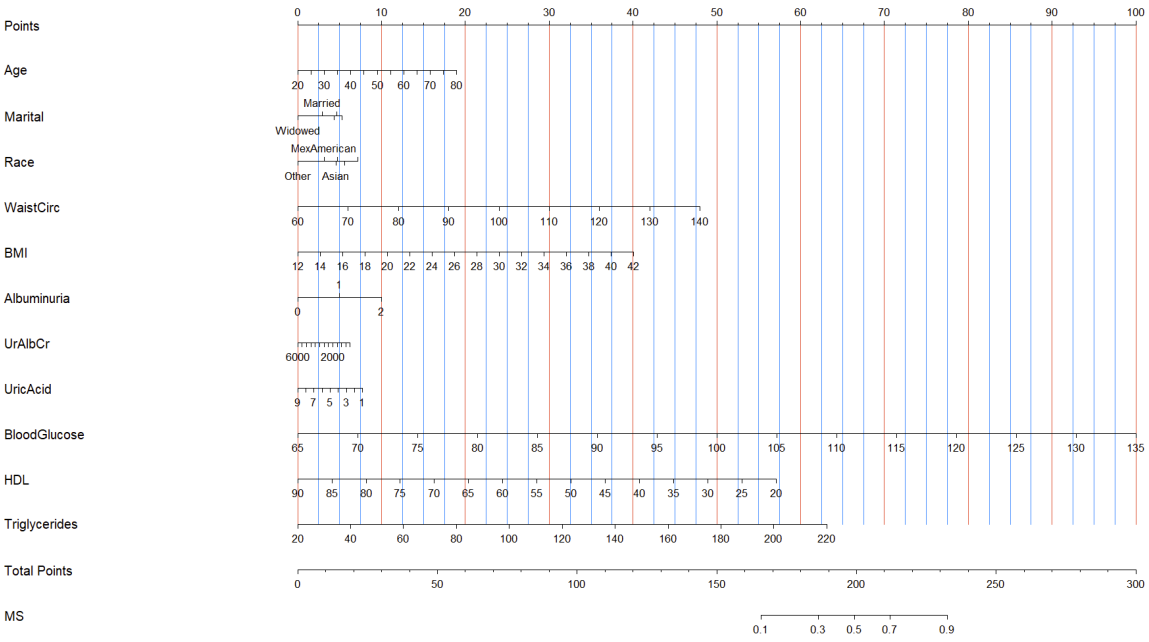

Figure 5: Nomogram.

Additionally, nomogram of the predictive model were created using the R programming language. These charts can aid in understanding the prediction rules and decision-making process of the model, providing a deeper interpretation and understanding of the predictive results.

These findings have important clinical implications for understanding and predicting the occurrence of MS, and they serve as a reference for further research and prevention. However, this study also has some limitations, such as its retrospective analysis using a single dataset and the exclusion of other potential risk factors. Therefore, future research could expand the sample size, consider more influential risk factors for MS, and conduct prospective studies to validate and generalize these findings.

Acknowledgments

This paper here was supported by Experimental (Practical) Course Development Project at Qinghai University in 2024, SYKC-2024-10.

References

Yabo Chen and Yufu Lin. Construction and assessment of a prediction model for the risk of helicobacter pylori infection in a population with metabolic syndrome. *China Medicine and Pharmacy*, 13:17–21, 2023. doi: 10.20116/j.issn2095-0616.2023.13.04.

Sheng-Shou HU. Epidemiology and current management of cardiovascular disease in china. *Journal of Geriatric Cardiology*, 21(4):387–406, 2024. doi: 10.26599/1671-5411.2024.04.001.

Qian Huang, Shaoyan Zheng, Zhiying Zhang, Danping Zhu, Xueting Fan, Biao Du, and Songjian Liu. A prediction model involving biomarkers for the risk of metabolic syndrome using the lasso regression. *Chinese Journal of Convalescent Medicine*, 33:1–5, 2024. doi: 10.13517/j.cnki.ccm.2024.01.001.

Yaru Li, Liyun Zhao, Dongmei Yu, Zhihong Wang, and Gangqiang Ding. Metabolic syndrome prevalence and its risk factors among adults in china: A nationally representative cross-sectional study. *PLOS ONE*, 13(6):1–16, 06 2018. doi: 10.1371/journal.pone.0199293.

Xiaoxi Mi, Lili Dai, Xuerui Jing, Jia She, Bjørn Holmedal, Aitao Tang, and Fusheng Pan. Accelerated design of high-performance mg-mn-based magnesium alloys based on novel bayesian optimization. *Journal of Magnesium and Alloys*, 12(2):750–766, 2024. doi: 10.1016/j.jma.2024.01.005.

Sahar Mohseni-Takalloo, Hadis Mohseni, Hassan Mozaffari-Khosravi, Masoud Mirzaei, and Mahdieh Hosseinzadeh. The effect of data balancing approaches on the prediction of metabolic syndrome using non-invasive parameters based on random forest. *BMC Bioinformatics*, 25(1):18, 2024. doi: 10.1186/s12859-024-05633-9.

Jie Song, Dongsheng Yu, Siwei Wang, Yanhe Zhao, Xin Wang, Lixia Ma, and Jiangang Li. Mapping soil organic matter in cultivated land based on multi-year composite images on monthly time scales. *Journal of Integrative Agriculture*, 23(4):1393–1408, 2024. doi: 10.1016/j.jia.2023.09.017.

Qi-ru Sui, Qin-huang Chen, Dan-dan Wang, and Zhi-gang Tao. Application of machine learning to the vs-based soil liquefaction potential assessment. *Journal of Mountain Science*, 20(8):2197–2213, 2023. doi: 10.1007/s11629-022-7809-4.

Rongsong Tang, Qun Wang, and Kun Yang. Research progress on time-restricted eating and metabolic syndrome. *Chinese Journal of Diabetes*, 31(11):852–857, 11 2023.

Kateryna Tkachenko, Isabel Esteban-Díez, José M. González-Sáiz, Patricia Pérez-Matute, and Consuelo Pizarro. Dual classification approach for the rapid discrimination of metabolic syndrome by ftir. *Biosensors*, 13(1), 2023. doi: 10.3390/bios13010015.

Nuttanan Wichitaksorn, Yingyue Kang, and Faqiang Zhang. Random feature selection using random subspace logistic regression. *Expert Systems with Applications*, 217:119535, 2023. doi: 10.1016/j.eswa.2023.119535.

Xianghao Zhan, Yiheng Li, Yuzhe Liu, Nicholas J. Cecchi, Samuel J. Raymond, Zhou Zhou, Hossein Vahid Alizadeh, Jesse Ruan, Saeed Barbat, Stephen Tiernan, Olivier Gevaert, Michael M. Zeineh, Gerald A. Grant, and David B. Camarillo. Machine-learning-based head impact subtyping based on the spectral densities of the measurable head kinematics. *Journal of Sport and Health Science*, 12(5):619–629, 2023. doi: 10.1016/j.jshs.2023.03.003.

Chao Zhang, Xinfeng Guo, and Shuwei Zhao. Correlation between metabolic syndrome and hearing loss in noise -exposed workers. *Occupational Health and Emergency Rescue*, 42:53–57, 2024. doi: 10.16369/j.oher.issn.1007-1326.2024.01.011.

Hui Zhang, Dandan Chen, Jing Shao, Leiwen TANG, Jingjie Wu, Erxu Xue, and Zhihong Ye.

Hui Zheng, Hui WANG, Hongli Yin, Donghua Yin, Yang Cheng, and Ying Wang. Correlation of serum uric acid level with metabolic syndrome in elderly aged. *Practical Geriatrics*, 37(11):1099–1102, 11 2023.