Statistical Analysis and Prediction Model Study of Metabolic Syndrome Risk Factors

Sian Zhu, Danxuan Zhang, Chengzhi Liu, Xiaoling Zhou, and Jun Wang

***Abstract*—This paper aims to systematically analyze the risk factors associated with metabolic syndrome (MetS) and to develop a logistic regression model for risk assessment. A total of 536 participants were enrolled and their demographic and clinical data were collected, including age, gender, high uric acid levels, uric acid overflow, and obesity status. Chi-squared tests were used to identify significant associations between MetS and various risk factors, such as age, gender, hyperuricemia, and obesity status. A logistic regression model was constructed to predict the probability of MetS occurrence, resulting in three different equations for the total population, men and women. The results showed that the model effectively captured the relationships between the identified risk factors and the likelihood of developing MetS. By providing a scientific basis for targeted prevention strategies, this paper contributes to the understanding of MetS and highlights the need for public health interventions aimed at reducing its prevalence and associated health burden.**

***Index Terms*—metabolic syndrome, logistic regression, risk factors, chi-squared test.**

I. Introduction

THE metabolic syndrome (MetS) is a clinical syndrome that encompasses a variety of metabolic abnormalities. It was first proposed by Professor Reaven of Stanford University in 1988 and has since emerged as a major public health concern due to its substantial impact on the global population ([1], [2], [3]). Characterized by abdominal obesity, atherogenic dyslipidaemia, elevated blood pressure, a prothrombotic and proinflammatory state, insulin resistance, and elevated glucose levels, MetS significantly increases the risk of developing chronic diseases such as Type 2 diabetes and cardiovascular disease ([4], [5]). The complex and multifaceted origins of MetS, which include genetic and acquired factors, are not fully understood; however, sedentarism and unbalanced dietary patterns have been suggested to play a fundamental role in its development. Epidemiological data indicate that the prevalence of MetS is not only persistently high, but also shows an alarming upward trend worldwide. In China, MetS has become a serious challenge for public health policy, with the prevalence among Chinese adults increasing from 11.0% between 2010 and 2012 to 31.1% in 2023 ([6]). This trend is not isolated to China, as global estimates suggest that about 3% of children and 5% of adolescents will have MetS in 2020, highlighting the urgent need for prevention strategies ([7]).

Although the pathogenic mechanisms of metabolic syndrome are not fully understood, clinical observations have shown that MetS is closely associated with several non-communicable diseases, such as Type 2 diabetes, cardiovascular disease, and periodontal disease ([8], [9]). These diseases have become major global health issues because of their significant impact on public health and socioeconomic burden. Therefore, early diagnosis and accurate assessment of MetS are crucial to prevent the onset of related diseases. Furthermore, the pathogenesis of MetS involves multiple genetic and acquired entities ([10], [11]). This syndrome is associated with a pro-inflammatory state, oxidative stress, haemodynamic dysfunction and ischaemia - conditions that often overlap in "dysmetabolic" patients ([2]). An increase in reactive oxygen species overloads the antioxidant systems, leading to oxidative stress and subsequent cellular dysfunction. Furthermore, oxidative stress and inflammation are associated with cellular senescence and cardiovascular disease, suggesting that cardiovascular disease should not be considered solely as the result of classical MetS risk factors ([2]). Lifestyle modification, especially dietary habits, is the primary therapeutic strategy for the management of MetS ([12]). Both the Mediterranean diet and the dietary approaches to stop hypertension have been supported by scientific evidence as beneficial dietary patterns for the prevention and treatment of MetS ([12]). In addition, intermittent fasting has been shown to improve cardiometabolic risk factors and alter gut microbiota in patients with MetS, providing mechanistic insights into the prevention of adverse outcomes ([13]).

Given the complexity of the MetS, accurate assessment of individual risk and disease progression remains a major challenge. Traditional diagnostic criteria may not capture the intricate interplay between metabolic factors, so predictive models are needed that can provide a more comprehensive risk assessment. Various computational and statistical approaches, including traditional statistical models and machine learning techniques, have been used to analyze the impact of these factors. For example, Markov models simulate disease progression over time [14]; clustering methods identify metabolic subgroups for personalized risk assessment [15], [16]; Bayesian networks uncover probabilistic dependencies between risk factors [17], [18]; Bayesian logistic regression has shown practical value in classifying disease-related risk, as in recent studies predicting COVID-19 mortality using comorbidity data [19]; and machine learning models such as

Manuscript received January 17, 2025; revised May 28, 2025.

This work was supported in part by Guangxi University of Chinese Medicine Joint Fund Project under Grant 2023L2066.

S. A. Zhu is an undergraduate student of School of Mathematics and Finance, Hunan University of Humanities, Science and Technology, Loudi 417000, P. R. China (e-mail: 2363621167@qq.com).

D. X. Zhang is an associate chief physician of Liuzhou Traditional Chinese Medical Hospital, Liuzhou 545001, P. R. China (e-mail: 279711128@qq.com).

C. Z. Liu is an associate professor of School of Mathematics and Finance, Hunan University of Humanities, Science and Technology, Loudi 417000, P. R. China (corresponding author to provide e-mail: it-rocket@163.com).

X. L. Zhou is a chief physician of Liuzhou Traditional Chinese Medical Hospital, Liuzhou 545001, P. R. China (corresponding author to provide e-mail: zxl-lz@163.com).

J. Wang is an undergraduate student of School of Mathematics and Finance, Hunan University of Humanities, Science and Technology, Loudi 417000, P. R. China (e-mail: 1918838015@qq.com).

random forest have demonstrated strong predictive capabilities [20], [21], [22], [23], [24].

Despite these advances, challenges remain, particularly the need for interpretable models that can be effectively integrated into clinical decision making. As deep learning techniques often lack transparency, there is growing interest in hybrid models that balance predictive power with interpretability.

In this study, we propose a mathematical modeling framework to systematically analyze the risk factors of MetS, providing a scientific basis for its prevention and management. By improving early detection and risk assessment, this approach aims to enhance clinical decision making and optimize intervention strategy development, ultimately contributing to more effective MetS management.

II. Data acquisition

A total of 536 participants were included in this study. The data set includes demographic and clinical variables relevant to the assessment of MetS, such as age, gender, presence of hyperuricemia, uric acid levels, and overweight status. The baseline distribution of these variables is summarized in Table I. All data were collected during routine health examinations and were approved by the relevant ethical review board.

As shown in Table I, the majority of participants were under 60 years of age (96.08%), with a slightly higher proportion of males than females. The prevalence of hyperuricemia was 48.88%, with a significantly higher incidence in men. Uric acid excess (defined as a level above 180 $\mu mol/L$) was observed in 8.02% of the population. Overweight individuals accounted for 36.00% of the sample.

III. Analysis of influencing factors

A. Chi-squared test procedure

To identify variables significantly associated with MetS, we performed a chi-squared test to examine the relationships between MetS and several categorical variables, including gender, age, hyperuricemia, and lifestyle factors such as smoking and alcohol consumption. For clarity and reproducibility, the detailed procedure of the chi-squared test is presented below.

1. Hypothesis formulation: Null hypothesis (H_0): The investigated variables (e.g. gender, age group and lifestyle habits) are independent of the presence of MetS. Alternative hypothesis (H_1): The investigated variables are not independent of the presence of MetS.

2. Data preparation and contingency table construction: Collect relevant data and construct a contingency table, listing the observed frequencies of all possible combinations of MetS presence and each influencing factor.

3. Calculate expected frequencies: For each cell in the contingency table, calculate the expected frequencies assuming the null hypothesis is true. The formula for calculating expected frequencies is:

$$E_{ij} = \frac{(h_i \times l_j)}{n},$$

where E_{ij} is the expected frequency for cell (i, j), h_i is the total frequency of the i-th row, l_j is the total frequency of the j-th column, and n is the total sample size.

4. Calculate the chi-squared statistic: Using the observed and expected frequencies, calculate the chi-squared statistic to quantify the differences between them. The formula for the chi-squared statistic is

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

where r is the number of rows in the contingency table, c is the number of columns, O_{ij} is the observed frequency, and E_{ij} is the expected frequency.

5. Find the P-value and make a decision: Use the calculated chi-squared statistic and its corresponding degrees of freedom to determine the P-value. Then, compare the P-value with a predetermined significance level (e.g., 0.05). If the P-value is less than the significance level, we reject the null hypothesis in favor of the alternative hypothesis, indicating that there is a significant difference between the observed data and the null hypothesis. This means that there may be a strong association between MetS and the factors under study. The smaller the P-value, the greater the deviation between the observed and expected frequencies, indicating a stronger association.

The steps described above were applied to each candidate variable in the data set. The results of the chi-squared tests are presented and interpreted in the following section.

B. Chi-squared test results and interpretation

The results of the chi-squared test for each variable are shown in Table II. These results indicate statistically significant associations between age, gender, hyperuricemia, uric acid overflow, and overweight status with the occurrence of MetS, suggesting that these variables are important risk indicators.

In particular, gender ($\chi^2 = 52.73, p < 0.001$), age group ($\chi^2 = 44.82, p < 0.001$) and hyperuricemia ($\chi^2 = 41.60, p < 0.001$) have both high chi-squared statistics and highly significant P-values, indicating that they are the most influential in discriminating between individuals with and without MetS. Overweight status also shows a strong association ($\chi^2 = 35.19, p < 0.001$), further emphasizing the importance of weight management in the prevention of MetS. Uric acid excess, although statistically significant ($p = 0.0108$), has a more modest chi-squared value ($\chi^2 = 6.46$), suggesting a relatively weaker effect.

In comparison, hypertension ($\chi^2 = 3.42, p \approx 0.06$) and kidney health ($\chi^2 = 2.78, p \approx 0.06$) are not statistically significant at the conventional 0.05 level. Although their P-values are close to the threshold, the small chi-squared values indicate a weaker association, which may warrant further investigation to clarify their potential role in the development of the MetS.

Overall, these results provide both statistical and quantitative evidence to support the prioritisation of certain variables – particularly gender, age and hyperuricemia – in subsequent modelling and targeted prevention strategies.

IV. MetS risk assessment model

Having identified the significant influencing factors, we proceed to define the dependent and independent variables for model construction. In this section, we perform a forward

TABLE I
BASIC INFORMATION ABOUT THE SUBJECTS

Variable		Total	Male	Female
Age	Under 60	515 (96.08%)	310 (57.84%)	205 (38.25%)
	Over 60	21 (3.92%)	9 (1.68%)	12 (2.24%)
Has hyperuricemia	Yes	262 (48.88%)	202 (37.69%)	60 (11.19%)
	No	274 (51.12%)	117 (21.83%)	157 (29.29%)
Uric acid excess	Below 180	241 (91.98%)	183 (69.85%)	58 (22.14%)
	Above 180	21 (8.02%)	19 (7.25%)	2 (0.76%)
Overweight	Yes	193 (36.00%)	150 (27.99%)	43 (8.02%)
	No	343 (64.00%)	169 (31.53%)	174 (32.46%)

TABLE II
RESULTS OF CHI-SQUARED TEST.

Variable	χ^2 Statistic	P-value
Age	53.8459	<0.001
Gender	60.1366	<0.001
Hyperuricemia	49.1454	<0.001
Uric acid excess	13.0989	0.0108
Overweight	55.5715	<0.001
Hypertension	3.5471	0.0596
Kidney health	3.5483	0.0596

stepwise logistic regression analysis on a dataset of 536 subjects, using age, elevated uric acid levels, uric acid excess, gender, and obesity as predictor variables. Given the strong association between gender and MetS, we not only develop a general risk assessment model, but also construct gender-specific models for male and female participants to improve predictive performance.

A. Construction of the logistic regression model

The logistic regression model is a statistical approach to binary classification problems that estimates the probability of an event occurring. It has been widely used in fields such as finance, health and social sciences. In this study, we use logistic regression ([25], [26]) to predict the probability of developing MetS.

Let Y be the dependent variable, which takes the value 0 (indicating no disease) or 1 (indicating the presence of disease), and let $X = (X_1, X_2, X_3, X_4, X_5)$ be the vector of independent variables. The logistic regression model can then be expressed as:

$$P(Y = 1|X) = \frac{1}{1 + e^{-q}}.$$

Where $P(Y = 1|X)$ is the probability that the dependent variable Y is 1 given the independent variables X, and q is the value of the linear predictor, which can be written as:

$$q = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5.$$

The parameters $\beta_0, \beta_1, \cdots, \beta_5$ are estimated from the data. In logistic regression, the likelihood function quantifies the probability of the observed data given specific parameter values and is used to estimate the parameters.

B. Construction of the likelihood function

Because logistic regression deals with binary outcomes, the likelihood function is typically based on the Bernoulli distribution. To simplify the calculation, the natural logarithm of the likelihood function is taken, resulting in the log-likelihood function.

For a single observation, the log-likelihood function is given by

$$\ln(P(y_i|X_i; \beta_0, \beta_1, \cdots, \beta_5)) = y_i \ln(p_i) + (1-y_i) \ln(1-p_i).$$

where y_i is the observed outcome for the i-th individual ($y_i = 1$ for presence, $y_i = 0$ for absence of disease), and p_i is the predicted probability.

For the whole sample, the overall log-likelihood function is

$$\ln(L(P)) = \sum_{i=1}^{m} \left[y_i \ln(p_i) + (1 - y_i) \ln(1 - p_i) \right],$$

where m is the number of observations.

C. Parameter estimation

The goal of parameter estimation is to find the values that maximize the log-likelihood function and thereby achieve the best model fit. In this study, we utilized MATLAB's built-in function 'fitglm' to perform the estimation efficiently. Based on the estimation results, we derived three different logistic regression models for the total population, males and females, as presented below:

For the general population:

$$L(P_1) = -3.0112 + 2.3607X_1 + 0.3886X_2 + 2.1060X_3 + 1.0594X_4 + 0.9437X_5.$$

For men:

$$L(P_2) = -1.4043 + 1.2353X_1 + 0.2512X_2 + 2.2134X_3 + 0.8272X_5.$$

For women:

$$L(P_3) = -4.0647 + 4.3730X_1 + 0.6835X_2 + 1.7881X_3 + 1.3378X_5.$$

In these equations, X_1 indicates age, X_2 indicates high uric acid levels, X_3 indicates uric acid excess, X_4 indicates gender and X_5 indicates obesity status.

These models provide a probabilistic framework for estimating the likelihood of MetS based on key risk factors, thereby facilitating individualized risk assessment and informing targeted intervention strategies.

TABLE III
PERFORMANCE OF LOGISTIC REGRESSION AT DIFFERENT THRESHOLDS.

Metric	Threshold	Overall	Male	Female
Accuracy	0.4	0.7332	0.6489	0.8341
	0.5	0.7258	0.6426	0.8525
	0.6	0.7090	0.6270	0.8571
Precision	0.4	0.8190	0.7311	0.8814
	0.5	0.7629	0.6464	0.8726
	0.6	0.7221	0.6025	0.8660
Recall rate	0.4	0.7607	0.5210	0.9293
	0.5	0.8433	0.7006	0.9674
	0.6	0.9031	0.8443	0.9837

V. PERFORMANCE EVALUATION

A. Results of the logistic regression models

To evaluate the classification performance of the logistic regression models, we introduced different thresholds (0.4, 0.5 and 0.6) to convert predicted probabilities into binary outcomes. Table III shows the accuracy, precision and recall rate at different thresholds for the total population and for male and female subgroups.

As shown in Table III, overall accuracy shows a slight decrease as the threshold increases from 0.4 to 0.6. A similar trend is observed for precision, suggesting that a higher threshold may lead to more false positives and a reduced ability to correctly identify true cases. Despite this, recall shows a significant increase when the threshold is raised to 0.6, suggesting that the model becomes more effective at capturing true positives and reducing false negatives.

These results highlight the trade-off between precision and recall rate, and emphasize the importance of choosing an appropriate classification threshold. In clinical practice, minimizing false negatives is particularly important as missed diagnoses can have serious health consequences.

Therefore, increasing the threshold to 0.6 improves the recall rate and increases the sensitivity of the model – an important consideration when the goal is to detect as many true cases as possible for timely medical intervention.

As noted in [4], MetS is strongly associated with cardiovascular disease and Type 2 diabetes. In this context, high recall rate is essential to ensure that at-risk individuals are accurately identified and provided with timely care. To further improve recall performance, we chose 0.6 as the operating threshold in our subsequent analysis, which balances diagnostic sensitivity with acceptable accuracy and precision. The results in Table III clearly demonstrate this trade-off and support our decision.

B. Comparison with other machine learning models

To further validate the effectiveness of the logistic regression model, we conducted a comparative evaluation with several commonly used machine learning models, including random forest (RF), support vector machine (SVM), k-nearest neighbors (KNN), and logistic regression (LR). For the LR models, we report performance across three threshold settings (0.4, 0.5, and 0.6). Evaluation metrics include precision, recall rate, and F1 score, as shown in Fig. 1. In addition, the receiver operating characteristic (ROC) curves for the LR, RF, and SVM models are shown in Fig.

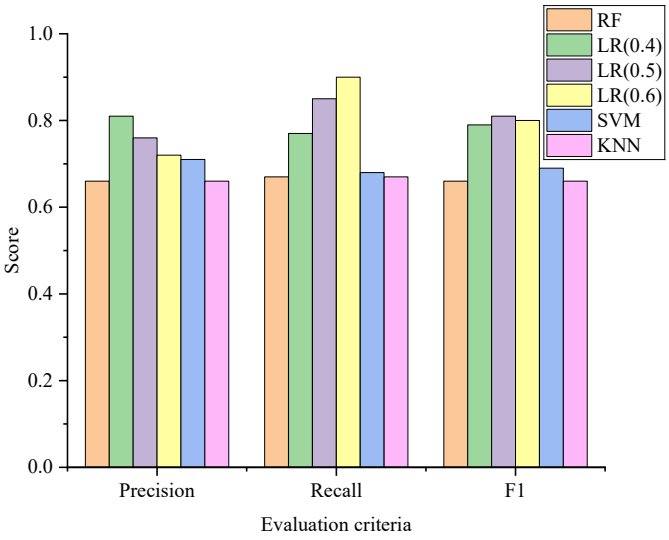

Fig. 1. Precision, recall, and F1 score of LR compared to RF, SVM, and KNN.

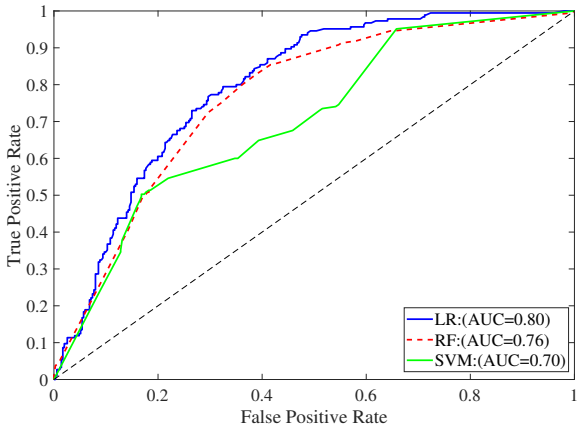

Fig. 2. ROC curves of LR, RF, SVM, and KNN.

2, and their corresponding AUC values are 0.80 for LR, 0.78 for RF, and 0.70 for SVM. The 95% confidence intervals (CI) for these AUCs are [0.7571, 0.8333] for LR, [0.7245, 0.8053] for RF, and [0.6575, 0.7477] for SVM.

Our results demonstrate that while RF, SVM, and KNN exhibit satisfactory performance on certain evaluation metrics, they do not consistently outperform LR. Notably, LR achieves the highest F1 score, reflecting a superior balance between precision and recall. In addition, LR achieves the highest area under the ROC curve (AUC = 0.80), with a relatively narrow 95% CI of [0.7571, 0.8333], suggesting robust and reliable classification performance. Although RF also performs well (AUC = 0.78, 95% CI: [0.7245, 0.8053]), its confidence interval partially overlaps with that of LR, indicating that the difference may not be statistically significant at the 95% level. In contrast, SVM has a lower AUC (0.70) with a non-overlapping CI of [0.6575, 0.7477], providing statistical evidence that LR significantly outperforms SVM. Taken together, these results highlight LR as the most stable and interpretable model among those evaluated, offering a favorable trade-off between predictive power and clinical applicability for MetS risk assessment.

VI. Conclusion

In this paper, we have successfully established and validated a logistic regression model to assess the risk factors associated with MetS. The results provide a comprehensive overview of the relationships between several variables and the prevalence of the syndrome. The results of our chi-squared test and logistic regression analyses have revealed significant statistical associations between MetS and several factors, with particularly strong associations identified for gender, age, and hyperuricemia. These findings underscore the importance of these variables in the development of MetS and suggest that targeted interventions focusing on these factors could be effective in reducing the incidence of the syndrome.

Our model, which incorporates these influential factors, provides a valuable tool for early diagnosis and risk assessment of MetS. By identifying individuals at higher risk, timely medical intervention can be facilitated, potentially preventing the progression of MetS and its associated complications, such as Type 2 diabetes and cardiovascular disease.

Further work could focus on refining the model and exploring additional risk factors that may influence MetS. Integrating genetic and lifestyle data into future models could enhance predictive accuracy and enable a more personalized approach to risk assessment. Furthermore, applying the model in clinical settings could help healthcare professionals make informed decisions about prevention strategies and interventions.

References

[1] G. M. Reaven, "Banting Lecture 1988: Role of insulin resistance in human disease," *Diabetes*, vol. 37, no. 12, pp. 1595–1607, 1988.

[2] J. L. Silveira Rossi, S. M. Barbalho, R. Reverete de Araujo, et al., "Metabolic syndrome and cardiovascular diseases: Going beyond traditional risk factors," *Diabetes/Metabolism Research and Reviews*, vol. 38, no. 3, pp. e3502, 2021.

[3] A. Bovolini, J. Garcia, M. A. Andrade, et al., "Metabolic syndrome pathophysiology and predisposing factors," *International Journal of Sports Medicine*, vol. 42, no. 3, pp. 199–214, 2021.

[4] K. G. M. M. Alberti, R. H. Eckel, S. M. Grundy, et al., "Harmonizing the metabolic syndrome: a joint interim statement of the international diabetes federation task force on epidemiology and prevention; national heart, lung, and blood institute; American heart association; world heart federation; international atherosclerosis society; and international association for the study of obesity," *Circulation*, vol. 120, no. 16, pp. 1640–1645, 2009.

[5] R. H. Eckel, J. M. Jakicic, J. D. Ard, et al., "2013 AHA/ACC guideline on lifestyle management to reduce cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines," *Journal of the American College of Cardiology*, vol. 63, no. 25, pp. 2960–2984, 2014.

[6] Chinese Medical Doctor Association Integrated Traditional Chinese and Western Medicine Physician Branch Endocrinology and Metabolism Disease Professional Committee, N. Qing, "Guidelines for the integration of syndrome and disease in the diagnosis and treatment of metabolic syndrome," *World Traditional Chinese Medicine*, vol. 18, no. 22, pp. 3157–3166, 2023.

[7] J. J. Noubiap, J. R. Nansseu, E. Lontchi-Yimagou, et al., "Global, regional, and country estimates of metabolic syndrome burden in children and adolescents in 2020: a systematic review and modelling analysis," *The Lancet Child & Adolescent Health*, vol. 6, no. 3, pp. 158–170, 2022.

[8] Chinese Medical Association Diabetes Branch, "Chinese guidelines for the prevention and treatment of Type 2 diabetes (17) — metabolic syndrome," *Chinese Community Doctor*, vol. 28, no. 11, pp. 1, 2012.

[9] S. Jepsen, J. Suvan, J. Deschner, "The association of periodontal diseases with metabolic syndrome and obesity," *Periodontology 2000*, vol. 83, no. 1, pp. 125–153, 2020.

[10] F. F. Costa, W. R. Rosário, A. C. R. Farias, et al., "Metabolic syndrome and COVID-19: An update on the associated comorbidities and proposed therapies," *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, vol. 14, no. 5, pp. 809–814, 2020.

[11] L. Rinaldi, P. C. Pafundi, R. Galiero, et al., "Mechanisms of non-alcoholic fatty liver disease in the metabolic syndrome. A narrative review," *Antioxidants*, vol. 10, no. 270, 2021.

[12] S. Castro-Barquero, A. M. Ruiz-León, M. Sierra-Pérez, et al., "Dietary strategies for metabolic syndrome: a comprehensive review," *Nutrients*, vol. 12, no. 10, pp. 2983, 2020.

[13] F. Yao, Y. Bo, L. Zhao, et al., "Prevalence and influencing factors of metabolic syndrome among adults in China from 2015 to 2017," *Nutrients*, vol. 13, no. 12, pp. 4475, 2021.

[14] L. C. Hwang, C. H. Bai, S. L. You, et al., "Description and prediction of the development of metabolic syndrome: a longitudinal analysis using a Markov model approach," *PloS One*, vol. 8, no. 6, pp. e67436, 2013.

[15] J. L. Kuk, C. I. Ardern, "Age and sex differences in the clustering of metabolic syndrome factors: association with mortality risk," *Diabetes Care*, vol. 33, no. 11, pp. 2457–2461, 2010.

[16] Y. Kim, H. Hwang, "Clustering effects of metabolic factors and the risk of metabolic syndrome," *Journal of Obesity & Metabolic Syndrome*, vol. 27, no. 3, pp. 166, 2018.

[17] H. S. Park, S. B. Cho, "Evolutionary attribute ordering in Bayesian networks for predicting the metabolic syndrome," *Expert Systems with Applications*, vol. 39, no. 4, pp. 4240–4249, 2012.

[18] H. Qi, R. Liu, C. C. Dong, et al., "Identifying influencing factors of metabolic syndrome in patients with major depressive disorder: A real-world study with Bayesian network modeling," *Journal of Affective Disorders*, vol. 362, pp. 308–316, 2024.

[19] F. Yanuar, M. Q. Shobri, R. F. Mabrur, et al., "Classification of death risk for COVID-19 patients using Bayesian logistic regression and naive Bayes classifier," *IAENG International Journal of Computer Science*, vol. 50, no. 3, pp. 915–920, 2023.

[20] A. Worachartcheewan, W. Shoombuatong, P. Pidetcha, et al., "Predicting metabolic syndrome using the random forest method," *The Scientific World Journal*, vol. 2015, no. 1, pp. 581501, 2015.

[21] F. Szabo de Edelenyi, L. Goumidi, S. Bertrais, et al., "Prediction of the metabolic syndrome status based on dietary and genetic parameters, using random forest," *Genes & Nutrition*, vol. 3, pp. 173–176, 2008.

[22] C. S. Yu, Y. J. Lin, C. H. Lin, et al., "Predicting metabolic syndrome with machine learning models using a decision tree algorithm: retrospective cohort study," *JMIR Medical Informatics*, vol. 8, no. 3, pp. e17110, 2020.

[23] J. Kim, S. Mun, S. Lee, et al., "Prediction of metabolic and pre-metabolic syndromes using machine learning models with anthropometric, lifestyle, and biochemical factors from a middle-aged population in Korea," *BMC Public Health*, vol. 22, no. 1, pp. 664, 2022.

[24] S. Mohseni-Takalloo, H. Mohseni, H. Mozaffari-Khosravi, et al., "The effect of data balancing approaches on the prediction of metabolic syndrome using non-invasive parameters based on random forest," *BMC Bioinformatics*, vol. 25, no. 1, pp. 18, 2024.

[25] A. Boateng, "A review of the logistic regression model with emphasis on medical research," *Journal of Data Analysis and Information Processing*, vol. 7, pp. 190–207, 2019.

[26] T. G. Nick and K. M. Campbell, "Logistic regression," *Topics in Biostatistics*, vol. 404, pp. 273–301, 2007.