Digital Health Trends Journal 2026;2(2):91-103
doi:10.34172/dhtj.28
https://dhtjournal.ae

Digital Health Trends Journal

Original Article

Metabolic Symphony: A Comparison on Machine Learning and Ensemble Algorithms of Lifestyle Promotion Project Data for Predicting Metabolic Syndrome Onset

Senobar Naderian[1], Zeinab Nikniaz[2], Mahdieh Abbasalizad Farhangi[3], Elahe Movagharnia[4], Leila Nikniaz[5*], Taha Samad-Soltani[6*]

[1]Student Research Committee, Tabriz University of Medical Sciences, Tabriz, Iran
[2]Liver and Gastrointestinal Diseases Research Center, Tabriz University of Medical Sciences, Tabriz, Iran
[3]Tabriz Health Services Management Research Center, Tabriz University of Medical Sciences, Tabriz, Iran
[4]Student Research Committee, Tabriz University of Medical Sciences, Tabriz, Iran
[5]Tabriz Health Services Management Research Center, Tabriz University of Medical Sciences, Tabriz, Iran
[6]Department of Health Information Technology, School of Management and Medical Informatics, Tabriz University of Medical Sciences, Tabriz, Iran

***Corresponding Authors:** Leila Nikniaz, Email: nikniazleila@gmail.com, Taha Samad-Soltani, Email: Samadsoltani@tbzmed.ac.ir

Abstract

Introduction: Metabolic syndrome (MetS) is a complex cluster of interconnected metabolic irregularities that seriously increase the risk of cardiovascular disorders and type 2 diabetes mellitus. Forecasting MetS and its associated factors with machine-learning (ML) models offers a promising approach to analyze datasets and uncover patterns associated with MetS risk. Accordingly, this study examined the performance of ML models in predicting MetS using distinctive combinations of feature selection and normalization techniques.

Methods: To this end, decision tree, K-nearest neighbors (KNN), Naïve Bayes, neural networks, random forest (RF), and support vector machine (SVM) were employed, evaluating their accuracy, precision, sensitivity, and specificity. Moreover, feature selection methods included the Chi2 score, F score, and mutual information, while normalization techniques comprised MinMax, StandardScaler (STD), and Robust.

Results: Our analysis revealed that ensemble models, particularly those combining SVM, RF, and KNN with mutual information and STD normalization, outperformed individual models in terms of accuracy and F1 scores, reaching an accuracy of 0.85% and an impressive F1 score of 0.93. Furthermore, the results revealed common features, such as age, body mass index, waist circumference, high blood pressure, diabetes, and specific nutritional intakes per day that consistently influenced MetS risk across different models.

Conclusion: Overall, our findings highlight the efficacy of ML models in predicting MetS and the importance of considering both clinical and lifestyle factors in predictive modeling. Using large datasets and advanced algorithms, healthcare practitioners can identify individuals at high risk of MetS and implement interventions to moderate their risk and improve health outcomes.

Keywords: Metabolic syndrome, Machine learning, Ensemble learning, Lifestyle

Received: January 18, 2026, **Revised:** March 21, 2026, **Accepted:** April 2, 2026, **ePublished:** June 14, 2026

Introduction

Metabolic syndrome (MetS), similarly called syndrome X or obesity syndrome, is a medical disorder that is indicated by a combination of disorders, including insulin resistance (IR), abdominal obesity, abnormal lipid levels, and hypertension. The World Health Organization characterizes MetS as a pathological state where an individual experiences multiple health issues simultaneously.[1] It should be noted that the incidence of this condition has been steadily increasing reaching approximately 25% worldwide, with the occurrence of nearly 3% in children and 5% in adolescents, displaying slight variations across different countries and regions.[2,3]

The increase rate of MetS results from a complicated interaction between genetic and acquired factors, culminating in inflammation and cardiovascular diseases. The global increasing prevalence of MetS underscores the impact of environmental and lifestyle factors, including excessive calorie consumption and insufficient physical activity. In addition, high caloric intake has been revealed as a primary factor triggering most of the pathways associated with MetS and diabetes [4] as a metabolic risk factor, with visceral adiposity playing a central role. Among various mechanisms, IR, neurohormonal activation, and chronic inflammation are emphasized as key contributors in the onset, advancement, and transformation of MetS

© 2026 The Author(s); This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

to cardiovascular diseases.[5,6,7] Clinicians and researchers have prioritized the early identification of risks, which subsequently leads to prompt treatment and enhanced cardiovascular outcomes.[2]

Artificial intelligence (AI) is a broad term indicating the utilization of a computer to simulate intelligent actions with slight human involvement.[8,9] In recent years, AI, particularly machine learning (ML), has underwent significant growth in data analysis and computing, enabling applications to operate intelligently.[10] ML aims to instruct machines on efficient data handling by enabling them to learn from the data.[11] The efficacy and efficiency of an ML solution typically relies on the nature and attributes of the data, as well as the implementation of the learning algorithms.[12] Algorithms refer to a set of methods enabling computers to automate data-driven model programming and construct models through systematic pattern detection in statistically significant data.[13]

The integration of AI into clinical practice demonstrates great promise as an area of advancement.[14] In the medical domain, ML methods have the potential to improve the accuracy of disease diagnosis and classification by analyzing clinical data, thereby introducing a novel paradigm in medical practices.[15]

The goal of ensemble learning, a ML method, is to increase prediction accuracy by combining predictions from various models. The goal of this method is to reduce the possibility of overgeneralization-related prediction mistakes.[16] In general, ensemble approaches effectively limit prediction mistakes by using a diverse range of independent models. Essentially, the ensemble approach creates a combined prediction by combining individual contributions. The ensemble behaves and produces results as though it were a single model even though it consists of several underlying models.[17,18] More precisely, ensemble models mainly aim to combine several weak learners into robust learners in order to increase overall model accuracy.[19] Bias, variability, and noises are common causes of discrepancies between real and anticipated values in ML models.[20] Notable methods in this field include bagging, boosting, stacking, and voting; these methods combine the outputs of several base learners to provide better predictive performance.[21,22] Voting is one of the ensembled algorithms that is utilized the most.[19] Voting classifiers increase accuracy and robustness by combining predictions from separate models.[18,22] Overall, further research should determine the precise function that ensemble approaches play in predicting MetS, even if they have demonstrated promise in a number of other areas, such as disease detection and prediction.

In a recent study by Tavares et al, some ML techniques were used to predict MetS with the best performance of the light gradient boosting machine (LGBM) with 87.8% sensitivity and an area under the receiver operating characteristic curve (AUC-ROC) of 0.86 using the gathered dataset of 17,182 individuals over 17 years of a checkup program.[23] Based on the analysis of the ML models obtained using the data gathered from a cohort study of middle-aged Koreans, tree-based ML algorithms, including random forest (RF) and XGBoost, showed the most efficacy in identifying MetS with the highest AUC of 0.851 of XGBoost after applying the synthetic minority oversampling technique (SMOTE).[24] Additionally, an ML model using proteomics data and age almost accurately predicted the probability of developing MetS in 10 years, with apolipoproteins, immune-related proteins, and coagulation-related proteins being strongly correlated with its development.[25] Similarly, Eyvazlou et al. developed an artificial neural network model to predict MetS among employees of an Iranian oil refinery using biomedical, lifestyle, sleep-related, and work-related variables. Their ANN model achieved an accuracy of 89%, outperforming logistic regression, which achieved 72% accuracy.[26] Moreover, in a study used different algorithms to create and validate a predictive model for MetS over 2 years, employing an individual's initial health condition and body weight after the mentioned duration. They could achieve the best predictive performance with an AUC of 0.879 of XGBoost.[27] Similarly, a study was designed to develop a MetS prediction model using genetic and clinical factors through ML models. Two models were created: one with clinical information only and another with additional generic information. In the test analysis, the Naïve Bayes (NB) classification model incorporating genetic information demonstrated better performance, with privileged sensitivity and specificity compared to the clinical-only model.[28] In addition, another research explored the connection between diabetes and MetS risk factors and compared ML methods using the data gathered over years to challenge the common belief associating diabetes with low high-density lipoprotein (HDL). Instead, increased HDL levels confirmed a positive correlation with diabetes onset.[29]

Using only noninvasive data, a MetS prediction model was created in 2023. The researchers created three unique synthetic features from four fundamental pieces of data, namely, waist circumference (WC), systolic and diastolic blood pressure (BP), gender, in order to extract meaningful characteristics. They verified that the decision tree (DT) model was the most suitable for the realistic prediction of MetS by testing many classification techniques. With an AUC of 0.889, a recall of 0.855%, and a specificity of 0.773%, the suggested model performed well. Considering that it simply employed four fundamental properties, the model was straightforward and simple to understand.[30] Further, in another study, different supervised ML models were tested to estimate the likelihood of acquiring MetS. Furthermore, the accuracy and predictive power of the models that employ the SMOTE were demonstrated. Using the SMOTE with 10-fold cross-validation, the assessment of the ML models revealed the advantage of the stacking ensemble algorithm over the other algorithms, with the accuracy, precision, recall, and F1 score values of 0.898, as well as an AUC value of 0.965.[31]

To the best of our knowledge, no recent study has so far predicted MetS incidence in Iran using ML methods and

the LPP dataset. Hence, the current study aims to develop an effective predictive model for MetS by employing ML techniques on relevant data.

Methods

The Lifestyle Promotion Project (LPP) attends as a longitudinal community-based scheme aimed at preventing and supervising non-communicable diseases (NCDs) within East Azerbaijan Province, Iran. The plan displays in two main phases: a cross sectional prevalence study (phase I) conducted from February 2014 to April 2014 about NCDs and their accompanying risk factors, and a prospective follow-up study (phase II) originated in February 2016. The phase I survey included 3,000 participants aged 15–65 years from 1,500 family units, while phase II followed up on lifestyle interventions originated in phase I. This thorough methodology guarantees a community-based and evidence-based strategy for NCD prevention, which is potentially appropriate in other developing countries. More information about every step of this project has been declared in the study protocol.[32]

Dataset

The study employed the LPP dataset, obtaining a broad range of risk factors for NCDs based on the World Health Organization format. This dataset included socio-demographic information, Angina status, smoking habits, physical activity, anxiety levels, dietary patterns, food security, food safety, biochemical measurements, daily intakes, biomedical parameters, and details of lifestyle promotion interventions.

Data Preprocessing and Integration

Data were gathered from four distinct sources with unique columns, demanding a thorough merging process. Biomedical assessment data, information from questionnaires, and physical examinations were synchronized. The inclusion criteria focused on patients with at least one of the biomedical test results: Fast blood sugar, ferritin, anemia, alanine transaminase, cholesterol, HDL, hemoglobin, aspartate aminotransferase, serum vitamin D status, leading to a consolidated dataset with 548 rows and 8,814 columns.

Handling Missing Values

Null and missing data were tackled applying python packages NumPy and Pandas. Columns were converted to numeric values, and empty columns were detected and removed. Moreover, numerical values were assigned to empty cells based on expert opinions. The final dataset comprised 502 rows and 132 columns for features and targets.

Target Selection

One specific column named "metabolic syndrome" was targeted, which was categorized in presence or absence classes to capture the aspects of MetS.

Normalization

The dataset underwent scaling using three normalization methods: StandardScaler (STD), MinMax, and Robust normalization. STD transformed data to have a mean and standard deviation of 1, MinMaxScaler rescaled data to a specific range, and RobustScaler adjusted for outliers. Three scaler data frames were instantiated and fitted to the data, and target columns were incorporated.

Feature Importance

It was analyzed using chi-square, analysis of variance F-value, and mutual information algorithms, which were all motivated by earlier research on feature selection strategies in medical data analysis. Sikri et al showed how feature ranking is affected by pre-processing data, which is crucial to meeting the chi-square method's assumptions.[33] A mutual information criterion-based feature selection technique was presented by Sulaiman and Labadin, who demonstrated how well it worked to enhance ML model performances.[34] Furthermore, Hoque et al presented a greedy feature selection technique based on mutual information theory that demonstrated excellent classification accuracy over several datasets.[35] Our choice of feature selection methods was guided by this research, which also offered insightful information about how to use them for medical data analysis. For this purpose, feature selection objects were created and fitted to imputed data, and the top 10 features were selected for each target variable using the SelectKBest class from the sklearn.feature_selection module.

A graphic depiction of the entire process is shown in Figure 1, representing how various processes and pathways within each interact to accomplish our objective.

Machine-Learning Models

Different algorithms, including DT, KNN, NB, RF, SVM, and NN, were used to classify the dataset. Each technique

Figure 1. Feature-Outcome Analysis and Effective Classification Using a Unified Methodology Framework

involved different steps, such as setting a random seed, loading the dataset, extracting target columns, converting features, separating data, determining class weights, training the classifier model, assessing performance with various metrics, and visualizing the outcomes. Three feature selection techniques were applied for every target variable, which were followed by three normalization techniques. Nine different combinations of normalization and feature selection methods were utilized to assess the function of each algorithm for the target variables. Accuracy, precision, recall, F1 score, and specificity were among the metrics.

Accuracy indicated the strong general correctness of model prediction. It is the proportion of cases that were accurately predicted to all instances.

In addition, precision measures the accuracy of positive predictions. It is defined as the proportion of accurately anticipated positive observations to all positive predictions.

Moreover, sensitivity or recall estimates the ability of the model to capture all the positive instances. It is the ratio of correctly predicted positive observations to the total actual positives.

Additionally, specificity gauges how well the model can account for every bad example. It is the proportion of all actual negative observations to all correctly predicted negative observations.

A thorough examination of model performance measures was made possible by investigating a 3x3 combination of feature selection and normalization.

Using the same feature selection and normalization techniques, an ensemble learning algorithm was trained in addition to previously mentioned ML algorithms. Individual base classifiers, such as DT and NB, and KNN, RF and SVM, were trained on the preprocessed dataset using the ensemble learning method, ensuring uniformity in feature selection and normalization across the ensemble. Predictions of these base classifiers were then aggregated using a "soft" voting scheme, in which the probabilities predicted by each base classifier are averaged, or a "hard" voting strategy, in which the final prediction is decided by a majority vote. Then, the performance of the ensemble classifiers was estimated on unseen data through the use of cross-validation techniques. The accuracy and F1 score were computed and then compared with the results from each classifier separately. Confusion matrices were also produced to offer a thorough examination of the performance of the ensemble model.

A publicly available dataset from Kaggle entitled "Metabolic Syndrome: A Comprehensive Dataset on Risk Factors and Health Indicators" was utilized for external validation. This dataset encompassed the target variable of interest, namely, MetS. The accuracy and other metrics of interest of the trained models on this external dataset were documented as well.

The originality of our approach is illustrated in Figure 2 as a result of the interplay between the feature selection and normalization techniques. This innovative approach not only contributes to the existing body of research on MetS prediction but also allows for a focused extraction of the most frequently occurring features, shedding light on their significance in the context of MetS in the Iranian population.

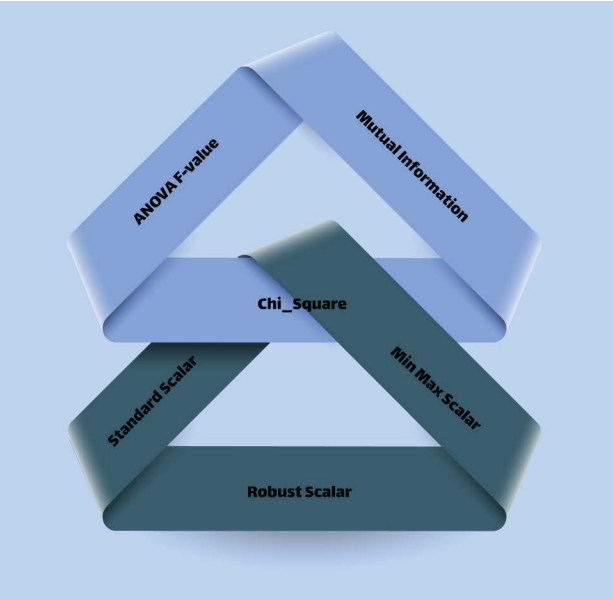

Figure 2. Uniqueness of the Technique

The unique approach, which combines feature selection and normalization techniques, is displayed in Figure 2.

Availability of Data and Materials

The codes and datasets supporting the study's conclusions are accessible via the GitHub repository linked here. Three versions of the dataset resulting from different normalization methods (csv1: MinMax, csv2: STD, and csv3: RobustScaler) are available.

Ethical Approval

Notably, this study utilized existing data from the LPP study that was conducted in East Azerbaijan Province, Iran, ensuring ethical considerations, participant privacy, and confidentiality. The study aligns with ethical standards and regulation, prioritizing the responsible use of data for scientific inquiry.

Results

To maximize each model's performance for a single target, the MetS category (2 classes), it was attempted to find out the most effective and optimized combination of feature selection and normalization methods. The best feature sets and algorithm metrics for each method are provided in Table 1.

Table 1 provides an understanding of the performance metrics related to different combinations of feature selection and normalization techniques, each correlated to a specific set. Particularly, the Chi2 score with MinMax normalization resulted in a similar feature set as the Chi2 score with STD, with only two differences in features "Angina grade" and "FBS." Additionally,

the F score combinations with MinMax, Robust, and STD normalization methods returned the same feature set as the Chi2 score with STD. Moreover, the mutual information combination with MinMax, STD, and Robust normalization methods all resulted in unique sets, highlighting the consistency of these combinations in terms of selected features.

Decision Tree

Table 2 evaluates the performance of distinct feature selection and normalization method combinations for the DT algorithm. F score with all 3 normalization methods obtained the highest accuracy of 0.78%, precision of 0.84%, sensitivity of 0.77%, and specificity of 0.91%. In addition, the Chi2 score with STD normalization developed a competitive accuracy of 0.74%. Robust normalization with the Chi2 score sorted a sensitivity of 0.99% but exchanged an accuracy of 0.68% and a specificity of 0.43%. Furthermore, mutual information with the combination of each normalization methods gained an accuracy of 0.78%, pointing out a specificity of 0.97%. The outcomes demonstrated that the F score and mutual information, with the combination of each normalization methods, constantly provided balanced and accurate outcomes across different feature sets, while Robust normalization with the Chi2 score outshone in sensitivity at the expenditure of general accuracy and specificity.

K Nearest Neighbor

Table 3 investigates the performance of distinct feature selection and normalization combinations for the KNN algorithm. Particularly, mutual information with STD normalization showed powerful overall performance, achieving an accuracy of 0.76% and perfect precision, sensitivity, and specificity. The combination of mutual information and score with other normalization methods also produced powerful results, just differing in accuracy. Additionally, the Chi2 score with MinMax normalization

Table 1. Feature Selection and Normalization Combinations With Respective Sets

Feature Selection	Normalization	Label	Obtained Features From "Feature Selection" and "Normalization" Combination
Chi2 score	MinMax	Set-1	['age', 'BMI (body mass index) category', 'Female waist circumference' (cm), 'Male waist circumference' (cm), 'Hypertension categorical' (1 = normal, 2 = pre-hypertension, 3 = normal with medicine, 4 = pre-hypertension with medication, 5 = grade 1, 6 = grade 2), 'Angina grade' (severe/not severe), 'Diabetic' (1 = taking medicine 2 = not taking medicine), 'Blood glucose level' (normal/prediabetic/diabetic), 'Waist-to-height ratio', 'Total blood pressure' (yes/no)]
	STD	Set-2	['age', 'BMI (body mass index) category', 'Female waist circumference' (cm), 'Male waist circumference' (cm), 'Hypertension categorical' (1 = normal, 2 = pre-hypertension, 3 = normal with medicine, 4 = pre-hypertension with medication, 5 = grade 1, 6 = grade 2), 'FBS' (mg/dL), 'Diabetic' (1 = taking medicine 2 = not taking medicine), 'Blood glucose level' (normal/prediabetic/diabetic), 'Waist-to-height ratio', 'Total blood pressure' (yes/no)]
	Robust	Set-3	['FBS' (mg/dL), 'Semi-solid oil per capita intake (intake/g/day)', 'EPA' (intake, mg/day), 'Iron intake' (mg/day), 'Fluoride' (intake/mg/day), 'Oleic fatty acid intake' (g/day), 'Copper' (intake/mcg/day), 'Chromium' (intake/mcg/day), 'Atocopherol' (intake/mg/day), 'Caffeine' (intake/mg/day)]
F score	MinMax STD Robust	Set-4	['**age**', '**BMI** (body mass index) category', '**Female waist circumference**' (cm), 'Male waist circumference' (cm), '**Hypertension categorical**' (1 = normal, 2 = pre-hypertension, 3 = normal with medicine, 4 = pre-hypertension with medication, 5 = grade 1, 6 = grade 2), 'FBS' (mg/dL), '**Blood glucose level**' (normal/prediabetic/ diabetic), '**Dyslipidemia**' (yes/no), '**Waist-to-height ratio**', '**Total blood pressure**' (yes/no)]
Mutual information	MinMax STD Robust	Set-5	['**Occupation**' (employed or self-employed, student, and unemployed), 'Female waist circumference' (cm), 'Hypertension categorical' (1 = normal, 2 = pre-hypertension, 3 = normal with medicine, 4 = pre-hypertension with medication, 5 = grade 1, 6 = grade 2), 'Having angina' (yes/no), '**FBS**' (mg/dL), 'Dyslipidemia (yes/no)', 'Waist-to-height ratio', 'Hypertension grade 2' (SBP > 140/DBP > 90, yes/no), '**Folate' (intake/mcg/day), 'Calcium' (intake/mg/day)**]

Note. STD: StandardScaler; FBS: Fast blood sugar; EPA: Eicosapentaenoic acid; SBP: Systolic blood pressure; DBP: Diastolic blood pressure.

Table 2. Comparison of Feature Selection and Normalization Techniques in the Decision Tree Model for Predicting Metabolic Syndrome

Feature Selection	Normalization	Feature Sets	Accuracy	Precision	Sensitivity	Specificity
Chi2 score	MinMax	Set-1	0.72	0.89	0.87	0.93
	STD	Set-2	0.74	0.88	0.86	0.91
	Robust	Set-3	0.68	0.77	0.99	0.43
F score	MinMax STD Robust	Set-4	0.78	0.89	0.88	0.91
Mutual information	MinMax STD Robust	Set-5	0.78	0.84	0.77	0.97
External validation/F score	MinMax STD Robust	ExSet-4	0.86	0.87	0.82	0.89

Note. STD: StandardScaler.

displayed powerful overall performance, attaining an accuracy of 0.74% and surpassing in precision (0.98%), sensitivity (0.97%), and specificity (0.98%). Remarkably, the use of STD and Robust normalizations generally led to similar or somewhat lower performance across all feature selection methods.

Naïve Bayes

Table 4 provides data related to the functioning of different feature selection and normalization combinations for the NB algorithm. Noticeably, the combination of F score with each of 3 normalization methods achieved the highest performance of an accuracy of 0.76% and well-proportioned precision, sensitivity, and specificity among various features. The Chi2 score with STD normalization showed powerful general performance, attaining an accuracy of 0.76% and representing balanced precision (0.76%), sensitivity (0.80%), and specificity (0.68%). Contrarily, chi 2 combination with Robust normalization normally caused reduced performance. Similarly, mutual information with each of the normalization methods demonstrated competitive outcomes, obtaining an accuracy of 0.73% with well-balanced precision, sensitivity, and specificity.

Neural Network

Table 5 presents the performance of distinct feature selection and normalization combinations applied to the NN algorithm. Particularly, the F score with MinMax normalization distinguished itself, reaching an accuracy of 0.83%, with well-balanced precision (0.83%), sensitivity (0.87%), and specificity (0.74%) among various features. Robust normalization with the F score also shaped competitive outcomes, with an accuracy of 0.99% and near-perfect precision, sensitivity, and specificity. However, it is worth mentioning that this combination caused the high standard deviations, implying potential flexibility in the results. The Chi2 score with MinMax normalization revealed powerful performance, specifically in sensitivity (0.89%), but with lower accuracy (0.82%) competed to the F score. Further, mutual information with MinMax normalization showed robust performance, gaining an accuracy of 0.84% with balanced precision, sensitivity, and specificity. The choice of feature selection and normalization combination approach considerably affects model function and performance, and in this context, the F score with MinMax normalization arises as a capable combination for accomplishing a stability between accuracy and generalization across diverse features.

Random Forest

Table 3. Comparison of Feature Selection and Normalization Techniques in the K-Nearest Neighbors Model for Predicting Metabolic Syndrome

Feature Selection	Normalization	Feature Sets	Accuracy	Precision	Sensitivity	Specificity
Chi2 score	MinMax	Set-1	0.74	0.98	0.97	0.98
	STD	Set-2	0.70	1.0	1.0	1.0
	Robust	Set-3	0.63	0.96	0.97	0.95
F score	MinMax	Set-4	0.75	1.0	1.0	1.0
	STD		0.73	1.0	1.0	1.0
	Robust		0.73	1.0	1.0	1.0
Mutual information	MinMax	Set-5	0.75	1.0	1.0	1.0
	STD		0.76	1.0	1.0	1.0
	Robust		0.72	1.0	1.0	1.0
External validation/mutual information	STD	ExSet-5	0.79	0.89	0.83	0.93

Note. STD: StandardScaler.

Table 4. Comparison of Feature Selection and Normalization Techniques in the Naïve Bayes Model for Predicting Metabolic Syndrome

Feature Selection	Normalization	Feature Sets	Accuracy	Precision	Sensitivity	Specificity
Chi2 score	MinMax	Set-1	0.75	0.76	0.79	0.69
	STD	Set-2	0.76	0.76	0.80	0.68
	Robust	Set-3	0.53	0.64	0.90	0.26
F score	MinMax	Set-4	0.76	0.77	0.72	0.84
	STD					
	Robust					
Mutual information	MinMax	Set-5	0.73	0.73	0.63	0.91
	STD					
	Robust					
External validation/F score	MinMax	ExSet-4	0.79	0.78	0.54	0.92
	STD					
	Robust					

Note. STD: StandardScaler.

Table 6 represents the functioning of discrete feature selection and normalization combinations across the RF algorithm. The Chi2 score with MinMax normalization distinguished itself, achieving an accuracy of 0.75% with imposing precision (0.98%), sensitivity (0.97%), and specificity (0.98%). Notably, the F score with MinMax normalization also brought powerful outcomes, highlighting a flawless accuracy of 0.76% and gaining the maximum precision, sensitivity, and specificity among varied features. Robust normalization with Chi2 score and mutual information led to a slightly lower accuracy (0.69% and 0.78%, respectively), implying some trade-off in functioning. It is worth pointing out that the Chi2 score with Robust normalization caused the most stable results, with lower standard deviations, indicating stability in the predictions. In this context, the Chi2 score with MinMax normalization and the F score with MinMax normalization appears as capable combinations for accomplishing high accuracy and well-proportioned performance across various features.

Support Vector Machine

Table 7 demonstrates the results of different feature selection and normalization combinations in the SVM

Table 5. Comparison of Feature Selection and Normalization Techniques in the Neural Network Model for Predicting Metabolic Syndrome

Feature Selection	Normalization	Feature Sets	Accuracy	Precision	Sensitivity	Specificity
Chi2 score	MinMax	Set-1	0.82	0.82	0.89	0.68
	STD	Set-2	0.98	0.98	0.98	0.97
	Robust	Set-3	0.81	0.80	0.93	0.58
F score	MinMax	Set-4	0.83	0.83	0.87	0.74
	STD		0.98	0.98	0.99	0.97
	Robust		0.99	0.99	1.0	0.99
Mutual information	MinMax	Set-5	0.84	0.83	0.94	0.64
	STD		0.98	0.98	0.97	0.98
	Robust		0.99	0.99	1.0	1.0
External validation/mutual information	Robust	ExSet-5	0.86	0.86	0.78	0.90

Note. STD: StandardScaler.

Table 6. Comparison of Feature Selection and Normalization Techniques in the Random Forest Model for Predicting Metabolic Syndrome

Feature Selection	Normalization	Feature Sets	Accuracy	Precision	Sensitivity	Specificity
Chi2 score	MinMax	Set-1	0.75	0.98	0.97	0.98
	STD	Set-2	0.74	1.0	1.0	1.0
	Robust	Set-3	0.69	0.97	0.97	0.95
F score	MinMax	Set-4	0.76	1.0	1.0	1.0
	STD					
	Robust					
Mutual information	MinMax	Set-5	0.78	1.0	1.0	1.0
	STD		0.79	1.0	1.0	1.0
	Robust		0.78	1.0	1.0	1.0
External validation/mutual information	STD	ExSet-5	0.88	1.0	1.0	1.0

Note. STD: StandardScaler.

Table 7. Comparison of Feature Selection and Normalization Techniques in the Support Vector Machine Model for Predicting Metabolic Syndrome

Feature Selection	Normalization	Feature Sets	Accuracy	Precision	Sensitivity	Specificity
Chi2 score	MinMax	Set-1	0.67	0.73	0.63	0.89
	STD	Set-2	0.71	0.76	0.73	0.81
	Robust	Set-3	0.62	0.60	0.96	0.14
F score	MinMax	Set-4	0.71	0.74	0.62	0.95
	STD		0.72	0.77	0.68	0.91
	Robust		0.72	0.79	0.74	0.87
Mutual information	MinMax	Set-5	0.75	0.73	0.61	0.95
	STD		0.76	0.80	0.72	0.93
	Robust		0.74	0.77	0.69	0.90
External validation/mutual information	STD	ExSet-5	0.85	0.88	0.78	0.93

Note. STD: StandardScaler.

model. The Chi2 score with MinMax normalization revealed an acceptable accuracy of 0.67%, with distinguished precision (0.73%) and specificity (0.89%). Nonetheless, its sensitivity (0.63%) indicated some encounters in properly recognizing positive instances. The F score with MinMax normalization provided enhanced accuracy (0.71%) and well-balanced performance across precision (0.74%), sensitivity (0.62%), and specificity (0.95%). Enigmatically, Robust normalization with the Chi2 score could create a particularly high sensitivity of 0.96% but at the expense of precision (0.60%), representing a higher rate of false positives. Meanwhile, mutual information with MinMax normalization distinguished itself with a solid accuracy of 0.75% and well-balanced precision (0.73%), sensitivity (0.61%), and specificity (0.95%). Generally, the F score with MinMax normalization surfaces as a promising combination, suggesting a good cooperation between accuracy and well-balanced performance among different metrics.

Generally, the Chi2 score with MinMax normalization and the F score with MinMax normalization stand out as strong options for DT and KNN algorithms, reliably producing accurate and balanced results across a range of characteristics. These combinations show how well they work together to provide competitive specificity, sensitivity, accuracy, and precision.

When compared to NB, NN, RF, and SVM models, the F score with MinMax normalization shows the best accuracy and balanced performance, making it a promising combination. It is imperative to acknowledge that the Robust normalization of the NN model with the F score displays competitive results, but with larger standard deviations, implying possible uncertainty in the outcomes.

With appropriate feature selection and normalization combinations, the NN, RF, and SVM algorithms outperform the other models in general, exhibiting good accuracy and well-proportioned precision, sensitivity, and specificity. On the other hand, there are certain performance trade-offs between the Chi2 score with Robust normalization for RF and the robust normalization of the DT with the Chi2 score.

Briefly, the F score with MinMax normalization is a powerful tool that works well with a variety of models and strikes a good balance between accuracy and balanced performance measures.

Ensemble Algorithms

Table 8 provides the ensemble analysis. Models sharing feature selection methods and metrics were combined. Using the F score with MinMax, STD, and Robust combinations, DT and NB models demonstrated an ensemble accuracy of 0.84% with MinMax, while results for STD and Robust were not provided. Conversely, the ensemble of SSVM, RF, and KNN, utilizing mutual information with STD normalization, outperformed with an accuracy of 0.85% and an impressive F1 score of 0.93.

When ML models were quantitatively compared to predict MetS, ensemble models outperformed individual models in terms of accuracy and F1 scores. In particular, the mutual information with STD normalization ensemble of SVM, RF, and KNN could perform better than individual models in the prediction challenge, with an F1 score of 0.93 and an accuracy of 0.85%. However, a few standout models also performed well, such as the NN with the F score and MinMax normalization.

Discussion

This study investigated the application of different ML models in predicting the MetS and its related factors through implementing various combinations of normalization and feature selection methods. The Chi2 score with MinMax normalization and the F score with MinMax normalization were effective for DT and KNN algorithms, providing accurate and balanced results. In NB, NN, RF, and SVM models, the F score with MinMax normalization showed the best accuracy and balanced performance. However, there were performance trade-offs between the Chi2 score and DT's robust normalization. While the exploration of a 3*3 combination of normalization and feature selections was valuable for comparing metrics, the primary emphasis was on extracting the most frequently occurring features influencing MetS for in-depth discussions.

Model Functioning

Among all models, multi-layer perception modeling exhibited the best overall performance with the accuracy of 0.99% and sensitivity and specificity of 1%, followed by RF with the overall optimized accuracy of 0.79%, and 1.0 for other metrics. Different studies have employed different models of NNs to predict MetS. For example, the use of ANN in studies resulted in accuracies of 85–95% and even lower.[24,25,26] While the application of multi-layer perception modeling and CNN led to an accuracy of 79% and 77%, respectively.[22,27] Likewise, some studies reported the RF as one of the best functioning algorithms in predicting MetS with an average reported accuracy of 79–82%.[22,25,27,28,29] Following these two models, the remaining ML models employed in this investigation demonstrated

Table 8. Ensemble Analysis of Shared Feature Selection Methods and Metrics in Machine Learning Models

Model	Feature Selection Method	Metric	Ensemble Accuracy	F1 Score
Decision tree and Naïve Bayes	F score	MinMax STD Robust	0.84	0.83
Support vector machine, random forest, and KNN	Mutual information	STD	0.85	0.93

Note. STD: StandardScaler; KNN: K-nearest neighbor.

nearly identical performance in terms of accuracies; however, regarding precision, sensitivity, and specificity metrics, the KNN's performance had nearly reached its peak, and the DT and SVM models were positioned thereafter. Nonetheless, the sensitivity was observed to be significantly lower than the specificity in most of the research comparing the three DT models of KNN and SVM.[16,25,26,27,30] Additionally, the specificity level ranged from 70% to 90%.[16,25,26,27]

These dissimilarities in performance among studies are considered to be associated with differences in the study population, whether ML models were constructed using basic clinical data entirely, and disparities in the use of genetic or blood test data.

With the exception of the NN, ensemble approaches outperformed simple individual ML models in our investigation, confirming results from earlier research in the field. The boosting ensemble method, LGBM, also confirmed outstanding results, including an impressive AUC-ROC of 0.86, a sensitivity of 87.8%, and a specificity of 70.2%. Based on these performance parameters, the LGBM model is useful in identifying people who are at risk of developing MetS because it has acceptable discrimination and calibration.[23] Furthermore, the models' accuracy and predictive power were enhanced by the application of the SMOTE. According to the evaluation of the ML models, the stacking ensemble algorithm outperformed other algorithms when paired with SMOTE, showing great precision, recall, and F1 score values of 0.898 in addition to its astounding accuracy of 89.35%. Further, the model's overall performance was shown by the AUC score, which was noticeably high at 0.965.[31] This implies that variables like dataset properties and modelling distinctions influence how effective ensemble approaches are in healthcare prediction tasks.

Feature Selection and Clinical Relevance

Figure 3 depicts frequently repeated elements across many models.

The incidence of MetS is multiplex, requiring a variety of factors like lifestyle factors, including obesity, high BP, high blood sugar, abnormal cholesterol levels, and confirmed clinical metrics.[31,32]

According to reports, these factors are the top 5 critical features in the MetS incidence by the models of ANN,[33-37] NB,[20] KNN, MPN,[38,39] SVM,[38- 39] DT, SVM.[40-42] Additionally, existing research using ML techniques to predict the likelihood of MetS has investigated the influence of known factors on the development of this disease. By utilizing the LPP dataset to analyze risk factors in the Iranian population, our current research adds to this expanding understanding. Our results support the importance of well-established variables, such as age, body mass index (BMI) and WC, in MetS risk. These benchmark indicators highlight the essential roles of age, obesity, and dyslipidemia in the pathophysiology of MetS.[43,44] Aging is accompanied by changes in body composition, metabolism, and hormonal balance and

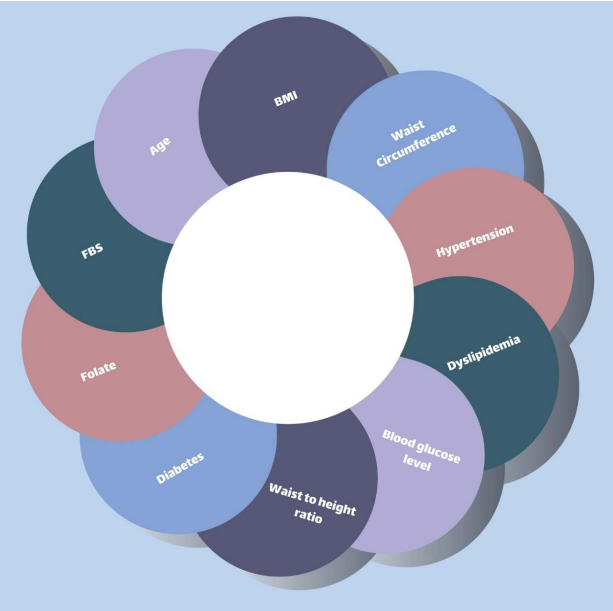

Figure 3. Key Metabolic Syndrome Components: Commonalities Among Models

a tendency of developing visceral obesity. Moreover, excessive body weight and WC, associated with higher BMI, contributes to low-grade inflammation and IR.[45]

In addition to these traditional markers, our analysis looks at modifiable lifestyle factors to further explore the complex nature of MetS risk. Leveraging the dataset provided by LPP, it was possible to evaluate the impact of high BP, diabetes, and blood glucose level, angina, and some nutritional intakes per day, such as 'eicosapentaenoic acid (EPA) intake (mg/day)', 'iron intake', 'fluoride intake (mg/day)', 'oleicfat intake', 'copper intake (mcg/day)', 'chromium intake (mcg/day)', 'Atocopherol intake (mg/day)', 'caffeine intake (mg/day)', 'folate intake (mcg/day)', and 'calcium intake (mg/day)'. This all-encompassing strategy is in line with the increasing focus on controlling and evaluating MetS risk by utilizing a combination of clinical, lifestyle, and genetic variables.[45-47] The dysregulation of the renin-angiotensin-aldosterone system and IR are common causes of elevated BP, which is a characteristic of MetS. According to Das,[48] this dysregulation leads to greater sodium retention and vasoconstriction, exacerbating hypertension. In addition, IR and poor glucose metabolism, which are caused by dysfunctional adipose tissue and persistent low-grade inflammation, are key components of the pathogenesis of diabetes in the MetS. Moreover, angina, which is frequently found in MetS, results from the buildup of atherosclerosis plaques in coronary arteries. The accumulation of this plaque lowers the oxygen flow in the myocardial ischemia that is made worse by underlying metabolic disorders.[49] Model SVM became a well predictor in ML research on the associated features of the MetS by emphasizing the likelihood of some factors, such as BP, blood glucose level, and diabetes. To the best of our knowledge, no ML research has so far identified angina as a MetS predictor, most probably because of the different origin and factors of processed datasets. Nonetheless, numerous clinical

studies have substantiated the association. As reported by Izadpanah et al, patients with unstable angina pectoris and MetS had higher levels of serum renalase, suggesting a possible link between renalase and the pathogenesis of MetS and angina.[50] Though their combined effects on inflammatory pathways, lipid metabolism, oxidative stress, and insulin sensitivity, other previously mentioned intake factors seem to have an impact on the MetS. The anti-inflammatory qualities of EPA can help mitigate systematic inflammation; existing research in animal studies indicates that EPA decreases palmitate-induced or high-fat-diet-induced Cer lipotoxicity and IR, while evidence in humans is significantly deficient.[51,52]

The function of iron in transferring oxygen, with excess iron potentially promoting oxidative damage and impairing insulin sensitivity, and the copper intake may modulate MetS risk through antioxidant defense mechanisms and glucose metabolism. As reported, copper and iron concentrations increased with an increase in the number of metabolic factors (P for trend < 0.001).[53] Elevated serum ferritin has been associated with a higher prevalence of metabolic syndrome and with its clinical components. Srivastav et al. also reported that serum ferritin was positively correlated with insulin resistance, measured by HOMA-IR, but inversely correlated with serum insulin levels.[54]

There is a hypothesis that fluoride may disrupt signaling pathways; as for instance, a correlation was reported between the levels of fluoride level and the higher hypertriglyceridemia and hypertension.[55] In addition, oleic fat intake may affect lipid metabolism and adipose tissue function, potentially affecting lipid profiles and adipokine secretion.[56] This fatty acid, which is a monounsaturated omega-9 fatty acid, is often included in studies examining fatty acid compositions and their association with MetS. Comparing fatty acid profiles in plasma between patients with MetS highlights differences in fatty acid synthesis pathways; with the activation of auxiliary metabolic pathways for oleic acid observed in MetS, potentially inhibiting the synthesis of beneficial fatty acids (e.g., EPA, intensified synthesis of EPA is observed, suggesting protective mechanisms against metabolic dysregulation.[57] While α -tocopherol may help mitigate oxidative stress, the association between serum α-tocopherol levels and MetS in the Korean population was dose-dependent, with higher levels of these vitamins associated with the increased risk of MetS and its components (e.g., hypertriglyceridemia, high fasting glucose, and high BP).[58] Chromium intake has been associated with improved insulin sensitivity and glucose metabolism.[59-61] However, the effect of MetS on humans remains controversial; a Chinese case report indicated that plasma chromium levels may be attributed to high WC, triglyceride, high glucose levels, and potentially MetS risk.[62] Although extensive research has investigated the correlation between caffeine intake and MetS, the conclusions are incompatible. According to some research, caffeine intake may impact energy metabolism and insulin sensitivity.[63] Nonetheless, the significant beneficial effects of caffeine consumption on several components of MetS (e.g., WC, triglyceride level, BP and fasting blood glucose level) reduce body fat and IR.[64] There are also some reports of the absence of a significant association between caffeine consumption and MetS risk.[65-67] Folate may influence one-carbon metabolism and DNA methylation, where it plays a crucial role as a methyl donor, potentially mitigating IR and reducing the risk of associated conditions, such as stroke.[68] Although this relationship remains unclear, there is a report that a higher folate intake may contribute to a lower MetS score, lower plasma fasting glucose levels, and higher HDL cholesterol,[69] and even high fasting glucose levels in different populations.[70]

Conclusion

Overall, our results confirmed the importance of applying ML methods for the prediction of MetS and the identification of its underlying risk factors. The findings of this study further demonstrated how particular feature selection and normalisation combinations (e.g., the Chi2 score and MinMax normalisation or F score and MinMax normalization) help produce balanced and accurate predictions across a variety of ML models. Furthermore, by including both conventional clinical measures and modifiable lifestyle factors, these models become more predictive and offer a more thorough method of assessing MetS risk. Accordingly, subsequent investigations should concentrate on integrating more extensive and varied datasets to augment the precision and applicability of prognostic models, enabling improved approaches for MetS prevention and management.

Authors' Contribution

Conceptualization: Senobar Naderian, Taha Samad-Soltani.
Data Curation: Leila Nikniaz, Zeinab Nikniaz.
Formal Analysis: Senobar Naderian, Mahdieh Abbasalizad.
Investigation: Elaheh Movagharnia, Taha Samad-Soltani.
Methodology: Senobar Naderian.
Project Administration: Leila Nikniaz.
Resources: Leila Nikniaz, Zeinab Nikniaz.
Software: Senobar Naderian.
Supervision: Taha Samad-Soltani.
Validation: Elaheh Movagharnia.
Writing–Original Draft: Senobar Naderian.
Writing–Review & Editing: Taha Samad-Soltani, Mahdieh Abbasalizad.

Availability of Data and Materials

Final datasets from the LPP study (csv1, csv2, and csv3) are accessible via the GitHub repository (https://github.com/senonaderian/metabolicsyndrome_prediction.git).

Competing Interests

The authors declare they have no competing interests related to this publication.

Consent for Publication

Not applicable.

Ethical Approval

The original LPP study (approval No. 1394.383) followed ethical guidelines, including obtaining informed consent from participants. For illiterate participants, consent was obtained from their legally authorized representatives. In addition, all experimental protocols

were approved by the Ethics Committee of Tabriz University of Medical Sciences.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

1. Madan K, Paliwal S, Sharma S, Kesar S, Chauhan N, Madan M. Metabolic syndrome: the constellation of co-morbidities, a global threat. Endocr Metab Immune Disord Drug Targets 2023;23(12):1491-504. doi:10.2174/1871530323666230309144825
2. Pavithra H, Naik PR. Prevalence of metabolic syndrome and its risk factors among adults in a rural area of Dakshina Kannada district. Indian J Community Med 2023;48(6):861-6. doi:10.4103/ijcm.ijcm_743_22
3. Noubiap JJ, Nansseu JR, Lontchi-Yimagou E, Nkeck JR, Nyaga UF, Ngouo AT, et al. Global, regional, and country estimates of metabolic syndrome burden in children and adolescents in 2020: a systematic review and modelling analysis. Lancet Child Adolesc Health 2022;6(3):158-70. doi:10.1016/s2352-4642(21)00374-6
4. Sowers JR. Recommendations for special populations: diabetes mellitus and the metabolic syndrome. Am J Hypertens 2003;16(11 Pt 2):41S-5S. doi:10.1016/j.amjhyper.2003.07.009
5. Rochlani Y, Pothineni NV, Kovelamudi S, Mehta JL. Metabolic syndrome: pathophysiology, management, and modulation by natural compounds. Ther Adv Cardiovasc Dis 2017;11(8):215-25. doi:10.1177/1753944717711379
6. Fahed G, Aoun L, Bou Zerdan M, Allam S, Bou Zerdan M, Bouferraa Y, et al. Metabolic syndrome: updates on pathophysiology and management in 2021. Int J Mol Sci 2022;23(2):786. doi:10.3390/ijms23020786
7. Bovolini A, Garcia J, Andrade MA, Duarte JA. Metabolic syndrome pathophysiology and predisposing factors. Int J Sports Med 2021;42(3):199-214. doi:10.1055/a-1263-0898
8. Hamet P, Tremblay J. Artificial intelligence in medicine. Metabolism 2017;69S:S36-40. doi:10.1016/j.metabol.2017.01.011
9. Sarker IH, Furhad MH, Nowrozy R. AI-driven cybersecurity: an overview, security intelligence modeling and research directions. SN Comput Sci 2021;2(3):173. doi:10.1007/s42979-021-00557-0
10. Mahesh B. Machine learning algorithms-a review. Int J Sci Res 2020;9(1):381-6. doi:10.21275/art20203995
11. Pugliese R, Regondi S, Marini R. Machine learning-based approach: global trends, research directions, and regulatory standpoints. Data Sci Manag 2021;4:19-29. doi:10.1016/j.dsm.2021.12.002
12. Pineda-Jaramillo JD. A review of machine learning (ML) algorithms used for modeling travel mode choice. Dyna (Medellin, Colombia) 2019;86(211):32-41. doi:10.15446/dyna.v86n211.79743
13. Briganti G, Le Moine O. Artificial intelligence in medicine: today and tomorrow. Front Med (Lausanne) 2020;7:27. doi:10.3389/fmed.2020.00027
14. Ishaq A, Sadiq S, Umer M, Ullah S, Mirjalili S, Rupapara V, et al. Improving the prediction of heart failure patients' survival using SMOTE and effective data mining techniques. IEEE Access 2021;9:39707-16. doi:10.1109/ACCESS.2021.3064084
15. Alhumaidi NH, Dermawan D, Kamaruzaman HF, Alotaiq N. The Use of Machine Learning for Analyzing Real-World Data in Disease Prediction and Management: Systematic Review. JMIR Med Inform 2025;13:e68898. doi:10.2196/68898
16. Ali R, Hardie RC, Narayanan BN, De Silva S. Deep learning ensemble methods for skin lesion analysis towards melanoma detection. In: 2019 IEEE National Aerospace and Electronics Conference (NAECON). Dayton, OH: IEEE; 2019. p. 311-6. doi:10.1109/NAECON46414.2019.9058245
17. Zubair Hasan KM, Zahid Hasan M. Performance evaluation of ensemble-based machine learning techniques for prediction of chronic kidney disease. In: Emerging Research in Computing, Information, Communication and Applications. Singapore: Springer; 2019. p. 415-26. doi:10.1007/978-981-13-5953-8_34
18. Nahar N, Ara F, Neloy MA, Barua V, Hossain MS, Andersson K. A comparative analysis of the ensemble method for liver disease prediction. In: 2019 2nd International Conference on Innovation in Engineering and Technology (ICIET). Dhaka: IEEE; 2019. p. 1-6. doi:10.1109/ICIET48527.2019.9290507
19. Lakshmanarao A, Srisaila A, Kiran TS. Heart disease prediction using feature selection and ensemble learning techniques. In: 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV). Tirunelveli: IEEE; 2021. p. 994-8. doi:10.1109/ICICV50876.2021.9388482
20. Shorewala V. Early detection of coronary heart disease using ensemble techniques. Inform Med Unlocked 2021;26:100655. doi:10.1016/j.imu.2021.100655
21. Filiz E. Evaluation of match results of five successful football clubs with ensemble learning algorithms. Res Q Exerc Sport 2023;94(3):773-82. doi:10.1080/02701367.2022.2053647
22. Gong H, Wang M, Zhang H, Elahe MF, Jin M. An explainable AI approach for the rapid diagnosis of COVID-19 using ensemble learning algorithms. Front Public Health 2022;10:874455. doi:10.3389/fpubh.2022.874455
23. Daniel Tavares L, Manoel A, Henrique Rizzi Donato T, Cesena F, André Minanni C, Miwa Kashiwagi N, et al. Prediction of metabolic syndrome: a machine learning approach to help primary prevention. Diabetes Res Clin Pract 2022;191:110047. doi:10.1016/j.diabres.2022.110047
24. Cai X, Xue Z, Zeng FF, Tang J, Yue L, Wang B, et al. Population serum proteomics uncovers a prognostic protein classifier for metabolic syndrome. Cell Rep Med 2023;4(9):101172. doi:10.1016/j.xcrm.2023.101172
25. Eyvazlou M, Hosseinpouri M, Mokarami H, Gharibi V, Jahangiri M, Cousins R, et al. Prediction of metabolic syndrome based on sleep and work-related risk factors using an artificial neural network. BMC Endocr Disord 2020;20(1):169. doi:10.1186/s12902-020-00645-x
26. Lee S, Lee H, Choi JR, Koh SB. Development and validation of prediction model for risk reduction of metabolic syndrome by body weight control: a prospective population-based study. Sci Rep 2020;10(1):10006. doi:10.1038/s41598-020-67238-5
27. Choe EK, Rhee H, Lee S, Shin E, Oh SW, Lee JE, et al. Metabolic syndrome prediction using machine learning models with genetic and clinical information from a nonobese healthy population. Genomics Inform 2018;16(4):e31. doi:10.5808/GI.2018.16.4.e31
28. Perveen S, Shahbaz M, Keshavjee K, Guergachi A. Metabolic syndrome and development of diabetes mellitus: predictive modeling based on machine learning techniques. IEEE Access 2019;7:1365-75. doi:10.1109/ACCESS.2018.2884249
29. Kim J, Mun S, Lee S, Jeong K, Baek Y. 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 2022;22(1):664. doi:10.1186/s12889-022-13131-x
30. Datta S, Schraplau A, Da Cruz HF, Sachs JP, Mayer F, Böttinger E. A machine learning approach for non-invasive diagnosis of metabolic syndrome. In: 2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE). Athens: IEEE; 2019. p. 933-40. doi:10.1109/BIBE.2019.00175.
31. Trigka M, Dritsas E. Predicting the occurrence of metabolic syndrome using machine learning models. Computation 2023;11(9):170. doi:10.3390/computation11090170

32. Tabrizi JS, Farahbakhsh M, Sadeghi-Bazargani H, Nikniaz L. Prevention and control of non-communicable diseases in Iranian population: life style promotion project phase II: study protocol. Iran J Public Health 2018;47(9):1397-405.
33. Sikri A, Singh NP, Dalal S. Chi-square method of feature selection: Impact of pre-processing of data. Int J Intell Syst Appl Eng 2023;11(3s):241-8.
34. Sulaiman MA, Labadin J. Feature selection based on mutual information. In: 2015 9th International Conference on IT in Asia (CITA). Sarawak: IEEE; 2015. p. 1-6. doi:10.1109/CITA.2015.7349827
35. Hoque N, Bhattacharyya DK, Kalita JK. MIFS-ND: a mutual information-based feature selection method. Expert Syst Appl 2014;41(14):6371-85. doi:10.1016/j.eswa.2014.04.019
36. Shin D. Prediction of metabolic syndrome using machine learning approaches based on genetic and nutritional factors: a 14-year prospective-based cohort study. BMC Med Genomics. 2024;17:224. doi:10.1186/s12920-024-01998-1
37. Xu W, Zhang Z, Hu K, Fang P, Li R, Kong D, et al. Identifying metabolic syndrome easily and cost effectively using non-invasive methods with machine learning models. Diabetes Metab Syndr Obes 2023;16:2141-51. doi:10.2147/dmso.S413829
38. Zhang H, Chen D, Shao J, Zou P, Cui N, Tang L, et al. Machine learning-based prediction for 4-year risk of metabolic syndrome in adults: a retrospective cohort study. Risk Manag Healthc Policy 2021;14:4361-8. doi:10.2147/rmhp.S328180
39. Park JE, Mun S, Lee S. Metabolic syndrome prediction models using machine learning and Sasang constitution type. Evid Based Complement Alternat Med 2021;2021:8315047. doi:10.1155/2021/8315047
40. Shin H, Shim S, Oh S. Machine learning-based predictive model for prevention of metabolic syndrome. PLoS One 2023;18(6):e0286635. doi:10.1371/journal.pone.0286635
41. Yang H, Yu B, P OU, Li X, Lai X, Zhang G, et al. Machine learning-aided risk prediction for metabolic syndrome based on 3 years study. Sci Rep 2022;12(1):2248. doi:10.1038/s41598-022-06235-2
42. Karimi-Alavijeh F, Jalili S, Sadeghi M. Predicting metabolic syndrome using decision tree and support vector machine methods. ARYA Atheroscler 2016;12(3):146-52.
43. Grundy SM, Cleeman JI, Daniels SR, Donato KA, Eckel RH, Franklin BA, et al. Diagnosis and management of the metabolic syndrome: an American Heart Association/National Heart, Lung, and Blood Institute Scientific Statement. Circulation 2005;112(17):2735-52. doi:10.1161/circulationaha.105.169404
44. Hruby A, Hu FB. The epidemiology of obesity: a big picture. Pharmacoeconomics 2015;33(7):673-89. doi:10.1007/s40273-014-0243-x
45. Yu CS, Lin YJ, Lin CH, Wang ST, Lin SY, Lin SH, et al. Predicting metabolic syndrome with machine learning models using a decision tree algorithm: retrospective cohort study. JMIR Med Inform 2020;8(3):e17110. doi:10.2196/17110
46. Gharipour M, Nezafati P, Sadeghian L, Eftekhari A, Rothenberg I, Jahanfar S. Precision medicine and metabolic syndrome. ARYA Atheroscler 2022;18(4):1-10. doi:10.22122/arya.2022.26215
47. Ahmed M, Kumari N, Mirgani Z, Saeed A, Ramadan A, Ahmed MH, et al. Metabolic syndrome; definition, pathogenesis, elements, and the effects of medicinal plants on it's elements. J Diabetes Metab Disord 2022;21(1):1011-22. doi:10.1007/s40200-021-00965-2
48. Das UN. Renin-angiotensin-aldosterone system in insulin resistance and metabolic syndrome. J Transl Int Med 2016;4(2):66-72. doi:10.1515/jtim-2016-0022
49. da Silva AA, do Carmo JM, Li X, Wang Z, Mouton AJ, Hall JE. Role of hyperinsulinemia and insulin resistance in hypertension: metabolic syndrome revisited. Can J Cardiol 2020;36(5):671-82. doi:10.1016/j.cjca.2020.02.066
50. Izadpanah P, Asadian F, Jangjou A. Association of serum renalase levels and renalase rs10887800 polymorphism with unstable angina pectoris patients having metabolic syndrome. Diabetes Metab Syndr Obes 2020;13:3249-59. doi:10.2147/dmso.S265773
51. Walchuk C, Wang Y, Suh M. The impact of EPA and DHA on ceramide lipotoxicity in the metabolic syndrome. Br J Nutr 2021;125(8):863-75. doi:10.1017/s0007114520003177
52. Lopez-Huertas E. The effect of EPA and DHA on metabolic syndrome patients: a systematic review of randomised controlled trials. Br J Nutr 2012;107 Suppl 2:S185-94. doi:10.1017/s0007114512001572
53. Lu CW, Lee YC, Kuo CS, Chiang CH, Chang HH, Huang KC. Association of serum levels of zinc, copper, and iron with risk of metabolic syndrome. Nutrients 2021;13(2):548. doi:10.3390/nu13020548
54. Srivastav SK, Mir IA, Bansal N, Singh PK, Kumari R, Deshmukh A. Serum ferritin in metabolic syndrome-mechanisms and clinical applications. Pathophysiology 2022;29(2):319-25. doi:10.3390/pathophysiology29020023
55. Schneider-Matyka D, Gutowska I, Panczyk M, Grochans E, Szkup M. Elevated serum fluoride levels in perimenopausal women are related to the components of metabolic syndrome. Eur Rev Med Pharmacol Sci 2021;25(17):5474-82. doi:10.26355/eurrev_202109_26656
56. Satogami K, Tseng PT, Su KP, Takahashi S, Ukai S, Li DJ, et al. Relationship between polyunsaturated fatty acid and eating disorders: systematic review and meta-analysis. Prostaglandins Leukot Essent Fatty Acids 2019;142:11-9. doi:10.1016/j.plefa.2019.01.001
57. Szczuko M, Kaczkan M, Drozd A, Maciejewska D, Palma J, Owczarzak A, et al. Comparison of fatty acid profiles in a group of female patients with chronic kidney diseases (CKD) and metabolic syndrome (MetS)-similar trends of changes, different pathophysiology. Int J Mol Sci 2019;20(7):1719. doi:10.3390/ijms20071719
58. Kim T, Kang J. Association between serum retinol and α-tocopherol levels and metabolic syndrome in Korean general population: analysis of population-based nationally representative data. Nutrients 2020;12(6):1689. doi:10.3390/nu12061689
59. Sreejayan N, Dong F, Kandadi MR, Yang X, Ren J. Chromium alleviates glucose intolerance, insulin resistance, and hepatic ER stress in obese mice. Obesity (Silver Spring) 2008;16(6):1331-7. doi:10.1038/oby.2008.217
60. Wang ZQ, Zhang XH, Russell JC, Hulver M, Cefalu WT. Chromium picolinate enhances skeletal muscle cellular insulin signaling in vivo in obese, insulin-resistant JCR:LA-cp rats. J Nutr 2006;136(2):415-20. doi:10.1093/jn/136.2.415
61. Zhang Q, Sun X, Xiao X, Zheng J, Li M, Yu M, et al. The effect of maternal chromium status on lipid metabolism in female elderly mice offspring and involved molecular mechanism. Biosci Rep 2017;37(2):BSR20160362. doi:10.1042/bsr20160362
62. Chen S, Zhou L, Guo Q, Fang C, Wang M, Peng X, et al. Association of plasma chromium with metabolic syndrome among Chinese adults: a case-control study. Nutr J 2020;19(1):107. doi:10.1186/s12937-020-00625-w
63. González-Domínguez R, Mateos RM, Lechuga-Sancho AM, González-Cortés JJ, Corrales-Cuevas M, Rojas-Cots JA, et al. Synergic effects of sugar and caffeine on insulin-mediated metabolomic alterations after an acute consumption of soft drinks. Electrophoresis 2017;38(18):2313-22. doi:10.1002/elps.201700044
64. Hino A, Adachi H, Enomoto M, Furuki K, Shigetoh Y, Ohtsuka M, et al. Habitual coffee but not green tea consumption is inversely associated with metabolic syndrome: an epidemiological study in a general Japanese population.

Diabetes Res Clin Pract 2007;76(3):383-9. doi:10.1016/j.diabres.2006.09.033
65. Wilsgaard T, Jacobsen BK. Lifestyle factors and incident metabolic syndrome. The Tromsø Study 1979-2001. Diabetes Res Clin Pract 2007;78(2):217-24. doi:10.1016/j.diabres.2007.03.006
66. Wan CJ, Lin LY, Yu TH, Sheu WH. Metabolic syndrome associated with habitual indulgence and dietary behavior in middle-aged health-care professionals. J Diabetes Investig 2010;1(6):259-65. doi:10.1111/j.2040-1124.2010.00055.x
67. Chang CS, Chang YF, Liu PY, Chen CY, Tsai YS, Wu CH. Smoking, habitual tea drinking and metabolic syndrome in elderly men living in rural community: the Tianliao old people (TOP) study 02. PLoS One 2012;7(6):e38874. doi:10.1371/journal.pone.0038874
68. Ashok T, Puttam H, Tarnate VCA, Jhaveri S, Avanthika C, Trejo Treviño AG, et al. Role of vitamin B12 and folate in metabolic syndrome. Cureus 2021;13(10):e18521. doi:10.7759/cureus.18521
69. Navarrete-Muñoz EM, Vioque J, Toledo E, Oncina-Canovas A, Martínez-González M, Salas-Salvadó J, et al. Dietary folate intake and metabolic syndrome in participants of PREDIMED-Plus study: a cross-sectional study. Eur J Nutr 2021;60(2):1125-36. doi:10.1007/s00394-020-02364-4
70. Koo YS, Lee YJ, Park JM. Inverse association of serum folate level with metabolic syndrome and its components in Korean premenopausal women: findings of the 2016-2018 Korean National Health Nutrition examination survey. Nutrients 2022;14(4):880. doi:10.3390/nu14040880