Diabetes, Metabolic Syndrome and Obesity

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

Application and Analysis of Random Forest and Support Vector Classification in Risk Prediction of Childhood Obesity and Hyperuricemia

Yuhang Wang[1,*], Shuang Shi[1,*], Xinghua Wei[1], Yanjing Wu[1], Yunlong Shi[1], Jin Cai[2]

[1]Graduate School, Nantong University, Nantong, Jiangsu, People's Republic of China; [2]Department of Pediatrics, Affiliated Hospital of Nantong University, Nantong, Jiangsu, People's Republic of China

*These authors contributed equally to this work

Correspondence: Jin Cai, Department of Pediatrics, Affiliated Hospital of Nantong University, Nantong, Jiangsu, People's Republic of China, Email wyhhh124@163.com

Background: The concurrent rise of childhood obesity and hyperuricemia presents a serious public health concern. These conditions interact through complex metabolic mechanisms and significantly increase long-term risks of cardiometabolic diseases. Machine learning (ML) offers an effective framework for constructing efficient risk prediction models in pediatric populations.

Objective: This study aimed to develop and evaluate two ML models—Random Forest (RF) and Support Vector Classification (SVC)—to predict the risk of childhood obesity and hyperuricemia by integrating clinical and biochemical variables.

Methods: A total of 101 children were enrolled, including 60 with obesity and 41 with obesity plus hyperuricemia. Data preprocessing involved recursive feature elimination (RFE), ROSE-based oversampling, and feature standardization. Both RF and SVC models were trained and evaluated using area under the ROC curve (AUC), precision-recall curves, and calibration curves. SHAP (Shapley Additive Explanations) analysis was conducted to interpret feature contributions.

Results: Both models demonstrated strong predictive performance, with AUCs reaching 0.96. The SVC model achieved slightly higher average precision and recall, making it more suitable for community- or school-based screening of high-risk children. In contrast, the RF model exhibited superior calibration, suggesting its greater utility in clinical decision-making where probabilistic risk estimation guides personalized follow-up or intervention planning. SHAP analysis identified glomerular filtration rate (GFR), high-density lipoprotein cholesterol (HDL-C), and apolipoprotein B (ApoB) as key predictors, some exhibiting nonlinear associations with disease risk.

Conclusion: RF and SVC models offer reliable tools for early risk prediction of obesity and hyperuricemia in children, each tailored to distinct clinical scenarios. These findings support early identification and targeted intervention. Future studies will explore the integration of metabolomic data and ensemble approaches to further enhance model performance and clinical applicability.

Keywords: childhood obesity, hyperuricemia, machine learning, random forest, support vector classification, SHAP

Introduction

With the rapid expansion of medical data, the application of clinical prediction models has grown increasingly common. Although traditional approaches such as regression analysis and logistic regression remain widely used, they often fall short in addressing the nonlinear relationships and temporal complexities inherent in large-scale clinical datasets.[1] In recent years, driven by advances in big data and computational technologies, machine learning (ML) has become an increasingly important tool in medical research, owing to its powerful analytical capacity and outstanding predictive performance.[2] The significance of this research lies in demonstrating the advantages of machine learning (ML) in clinical prediction, as ML not only offers greater predictive accuracy but also provides more efficient tools for early disease detection and clinical management, highlighting its potential applications in modern medicine.

Received: 1 February 2025
Accepted: 22 June 2025
Published: 7 July 2025

In recent years, the dual epidemic of childhood obesity and hyperuricemia has drawn increasing public health attention. These two conditions not only pose serious threats to children's physical health, but may also interact through complex pathophysiological mechanisms, significantly increasing the risk of developing metabolic and cardiovascular diseases later in adulthood.[3] Obesity induces endoplasmic reticulum stress in the liver and promotes hepatic steatosis, significantly suppressing the synthesis of functional apolipoprotein AI (ApoAI). Under chronic metabolic stress and pro-inflammatory stimulation, both the transcription and translation of ApoAI are downregulated.[4] This process compromises the cholesterol-carrying capacity and reverse transport function of high-density lipoprotein (HDL), thereby markedly increasing the risk of atherosclerotic plaque formation. In addition, hepatic lipid metabolism is further disrupted by the excessive influx of free fatty acids (FFAs), which exacerbates lipid overload in hepatocytes. The accumulation of FFAs can induce excessive production of reactive oxygen species (ROS), activate the NF-κB signaling pathway, and promote the release of pro-inflammatory cytokines.[5] At the same time, it triggers the release of cytochrome c from mitochondria, thereby promoting hepatocyte apoptosis.

Uric acid–induced fat accumulation and disruptions in energy metabolism not only impair the metabolic functions of the liver and adipose tissue, but also promote visceral fat deposition and reduce insulin sensitivity through neuroendocrine feedback mechanisms. This mode of action positions uric acid as a central regulator of metabolic homeostasis, highlighting its role in the coordinated modulation of multiple metabolic pathways and reinforcing its significance as a key target in obesity and metabolism-related disease research.[6] Similarly, this pathway is further supported in the context of obstructive sleep apnea syndrome (OSAS), where hypoxic conditions and sympathetic activation not only enhance uric acid synthesis but also exacerbate metabolic dysregulation by altering neuroendocrine system function. Sleep disturbances impair normal energy balance by reducing the secretion of leptin and adiponectin, thereby increasing hunger and promoting fat accumulation.[7,8] In a hyperuricemic internal environment, leptin signaling is impaired, weakening its ability to regulate appetite and energy expenditure. This leads to increased food intake and enhanced fat storage, thereby indirectly accelerating disease progression through the aforementioned mechanisms. Collectively, these findings suggest that uric acid is not merely a consequence of obesity-related metabolic disorders, but also serves as a key regulatory factor actively involved in their pathogenesis.

To better understand the origins of these disorders and to develop effective strategies for prevention and treatment, it is essential to construct accurate prediction models for childhood obesity and its associated comorbidities.

Previous studies by Vergeer, L and Kenney, EL have explored the relationship between childhood obesity and factors such as economic conditions, dietary costs, living environment, and electronic device usage. These investigations incorporated several lifestyle variables and employed basic statistical analyses to draw sociological conclusions.[9] In the prediction of childhood obesity risk and probability, Liu et al developed obesity prediction models using various machine learning (ML) approaches, including Random Forest and XGBoost. Their models integrated multidimensional variables such as socioeconomic status and psychological-behavioral factors. The optimal model achieved an area under the curve (AUC) of 0.96, demonstrating the significant advantage of ML in handling nonlinear relationships and complex variables.[10] These models also employed causal inference to reveal relationships between variables, thereby providing scientific support for obesity intervention strategies.[11]

Although most existing predictive models rely on selected sociological and demographic factors to estimate obesity risk, the etiology and long-term outcomes of obesity are multifactorial and complex. As a major public health crisis of the 21st century, obesity substantially increases the risk of developing other primary lifestyle-related diseases, including metabolic syndrome, early-onset cardiovascular diseases, musculoskeletal disorders, neurodegenerative diseases, and autoimmune conditions.[12,13]

Childhood obesity is closely associated with metabolic syndrome (MetS), a cluster of metabolic, vascular, and inflammatory abnormalities that significantly increases the risk of developing type 2 diabetes (DM2) and cardiovascular disease (CVD) in adulthood.[14] Therefore, beyond predicting obesity alone, other studies have explored the integration of diverse disease-related data and modeling strategies. For example, Miller and Hsin-Yao Wang developed genetically oriented childhood obesity risk models using multiple algorithms, including support vector classification (SVC), applying machine learning techniques to characterize the genetic features associated with obesity.[15] Amit Das employed neural

networks to predict nonalcoholic fatty liver disease (NAFLD) in obese children, demonstrating the potential of machine learning in capturing complex nonlinear features.[16]

Although numerous studies have applied machine learning methods to explore disease risk, relatively little attention has been given to the predictive role of biochemical factors in the progression of obesity and the early stages of metabolic syndrome. As highlighted in the review by Mahmood Safaei, future research should focus on multivariate data analysis that integrates lifestyle, environmental factors, and multiple metabolic characteristics.[17]

Therefore, this study aims to develop an efficient and interpretable risk prediction model for childhood obesity and hyperuricemia based on multivariate clinical and biochemical features, with the goal of providing a scientific basis for early disease prediction and intervention, and promoting timely prevention and management.

Methods

This study aimed to construct and evaluate two machine learning models—Support Vector Classification (SVC) and Random Forest (RF)—to predict clinical occurrences of obesity and hyperuricemia. Modeling and analysis were performed using R software (version 4.2.2), with key steps including data preprocessing, feature selection, model training, and evaluation.

From January 2023 to June 2024, eligible pediatric participants were recruited from the outpatient and inpatient departments of the Affiliated Hospital of Nantong University. The study included two groups: the obesity group (n = 60) and the obesity combined with hyperuricemia group (n = 41). The inclusion criteria were: (1) age between 6 and 16 years; (2) diagnosis of obesity based on the World Health Organization (WHO) BMI percentile criteria for children, defined as a body mass index (BMI) at or above the 95th percentile for age and sex, where BMI = weight [kg] / height² [m²]; and (3) hyperuricemia defined according to the pediatric serum uric acid reference standards published by the Mayo Clinic, with levels exceeding the age- and sex-specific thresholds;[18] (4) All participants had not received prior treatment related to the study conditions and had no history of other major diseases.Exclusion criteria included: (1) presence of other metabolic disorders or major chronic illnesses; (2) incomplete data or samples with significant abnormalities. A total of 101 participants were ultimately enrolled and classified into two groups—obesity and obesity with hyperuricemia—based on anthropometric and biochemical indicators.

All participants had fasting blood samples (3–4 mL) collected in the morning at the hospital's physical examination center. Height and weight were recorded to calculate body mass index (BMI). Blood samples were drawn into serum separation tubes with clot activator, and serum was isolated for biochemical analysis. The measured indicators included free triiodothyronine (FT3), free thyroxine (FT4), aspartate aminotransferase (AST), alanine aminotransferase (ALT), total bilirubin (TBIL), glomerular filtration rate (GFR), apolipoprotein A1 (ApoA1), apolipoprotein B (ApoB), lipoprotein(a) [Lp(a)], and high-density lipoprotein cholesterol (HDL-C).

Data Preprocessing

Missing values were imputed using the median. Recursive feature elimination (RFE) combined with random forest was employed to select key features. The dataset was split into training and testing sets in an 8:2 ratio, and all features were standardized (mean = 0, standard deviation = 1). To address class imbalance, the ROSE package was used for oversampling, and zero-variance features were removed.

Model Training

The RF and SVC models were applied for prediction, and their parameters were optimized via grid search. For the Random Forest model, the mtry parameter was adjusted, while the SVC model utilized a radial basis function (RBF) kernel to optimize the C and sigma parameters. Model stability was evaluated using 10-fold cross-validation.

Model Performance Evaluation

Evaluation metrics included the ROC curve and corresponding AUC value on the test set. Precision-recall (PR) curves were also plotted, and the average precision (AP) was computed. Additionally, calibration curves and radar charts were used to present classification metrics such as accuracy, precision, recall, F1 score, and AUC.

Model Interpretability Analysis

SHAP values were used to assess the contribution of each feature in the model, and SHAP dependence plots were generated to reveal their specific impact on prediction. A heatmap of interaction strength was also constructed to visualize dependencies among features and further elucidate inter-variable relationships.

Results

Baseline Characteristic Comparison

Table 1 presents the baseline comparisons of multiple clinical and biochemical indicators between the obesity group and the obesity with hyperuricemia group. While most variables did not show statistically significant differences between the groups, serum uric acid levels and glomerular filtration rate (GFR) were significantly altered in the combined group, suggesting that hyperuricemia may be associated with early impairment of renal function. These differences are not only of potential pathophysiological relevance but were also identified as important features in subsequent machine learning models.

Additionally, variables such as body mass index (BMI), lipid metabolism, insulin function, and thyroid-related indicators exhibited trends of difference between groups, though not statistically significant. The distribution of sex was comparable between the two groups, ruling out sex as a confounding factor in the modeling process. Overall, the intergroup comparisons provided a foundation for feature selection and reflected potential metabolic abnormalities associated with hyperuricemia.

Feature Selection and Model Construction

We initially trained the Random Forest model using the full set of features. Based on the feature importance scores derived from the model, features with minimal predictive contribution were iteratively removed. The performance of various feature combinations was evaluated through five-fold cross-validation to identify the optimal feature set. Ultimately, 15 features with the highest predictive value were selected, including serum creatinine (Cr), BMI, and GFR.

Model Parameter Optimization

The Random Forest (RF) model constructs multiple decision trees by randomly selecting subsets of features and aggregates their predictions through majority voting, which enhances model generalizability and reduces the risk of overfitting. We used the caret package for training and applied grid search to optimize the mtry parameter, ultimately determining the optimal value as 2. The Support Vector Classification (SVC) model was implemented with a radial basis function (RBF) kernel. Grid search was employed to tune the penalty parameter C and the kernel parameter sigma, with the optimal combination identified as C = 1 and sigma = 0.1.

Table 1 Baseline Characteristics of Children in the Obesity Group and the Obesity + Hyperuricemia Group

Variables	Obesity (n = 60)	Obesity + Hyperuricemia (n = 41)	Statistic	*P*
Age(years), Mean ± SD	10.24 ± 1.84	10.43 ± 1.52	t=−0.40	0.689
BMI (kg/m²), Mean ± SD	92.80 ± 219.99	27.49 ± 4.73	t=1.48	0.144
Uric acid (μmol/L), Mean ± SD	336.79 ± 51.83	477.92 ± 91.85	t=−6.69	<0.001
GFR (mL/min/1 73m²), Mean ± SD	243.93 ± 21.57	205.07 ± 24.30	t=5.98	<0.001
HDL-C (mmol/L), Mean ± SD	71.87 ± 105.38	92.03 ± 130.80	t=−0.60	0.551
ApoB (g/L), Mean ± SD	2.03 ± 1.03	1.95 ± 1.27	t=0.26	0.797
C-peptide (ng/mL), Mean ± SD	2.60 ± 1.41	3.35 ± 1.33	t=−1.92	0.060
Lp(a) (mg/dL), Mean ± SD	52.93 ± 103.48	60.89 ± 100.49	t=−0.28	0.784
FT3 (pmol/L), Mean ± SD	6.66 ± 1.20	6.52 ± 0.71	t=0.50	0.617
Sex (%)			χ^2=1.28	0.258
Female	10 (40.00)	14 (56.00)		
Male	15 (60.00)	11 (44.00)		

Abbreviations: t, *t*-test; χ^2, Chi-square test; SD, standard deviation.

Model Evaluation and Results

Figure 1 displays the ROC curves of the RF and SVC models on the test set. Both models achieved an AUC of 0.96, indicating excellent discrimination between the obesity group and the obesity combined with hyperuricemia group. The balanced performance between true positive and false positive rates suggests strong classification capability.

Figure 2 presents the PR curves. The average precision (AP) for RF was 0.964, while for SVC it was slightly higher at 0.965, demonstrating that SVC performed marginally better, especially under high recall conditions. Nevertheless, overall performance between the two models was very similar.

Figure 3 shows the calibration curves. The RF model demonstrated calibration very close to the ideal line in the high-probability region, indicating superior reliability. The SVC model tended to overpredict in the lower probability range but improved as prediction probability increased. Overall, RF outperformed SVC in calibration stability.

Figure 4 compares five performance metrics—Accuracy, Precision, Recall, F1 Score, and AUC—between RF and SVC. SVC showed slight advantages in Recall and F1 Score, while RF was slightly lower in Precision. However, both models demonstrated robust and stable performance across all metrics.

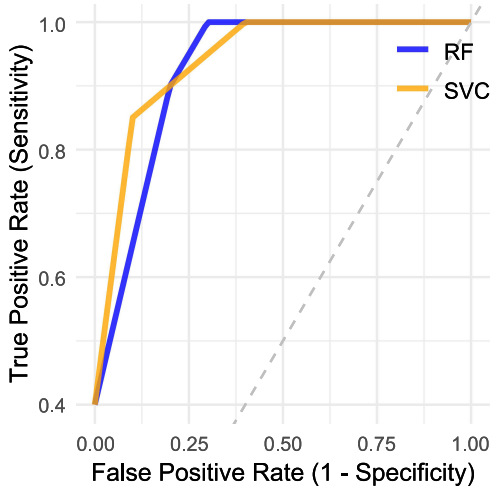

Figure 1 ROC curves of the RF and SVC models on the test set.

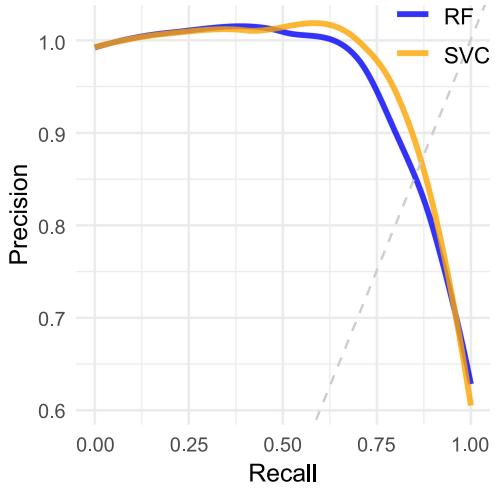

Figure 2 PR curves of the RF and SVC models.

https://doi.org/10.2147/DMSO.S519284

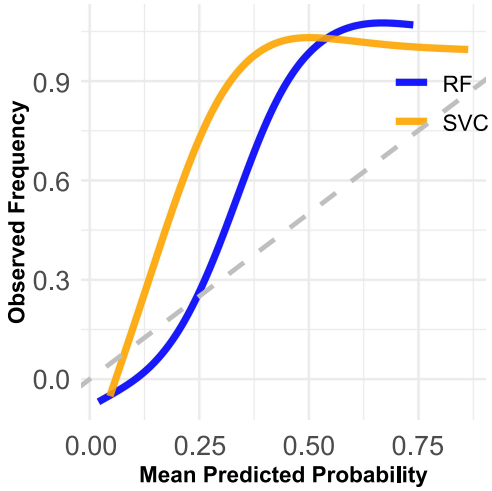

Figure 3 Calibration curves of the RF and SVC models.

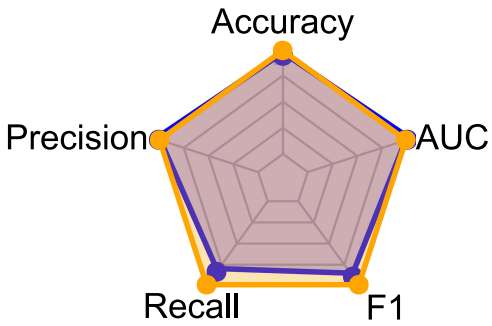

Figure 4 Performance comparison of RF and SVC in classification metrics.

Model Interpretation and Feature Analysis

Figure 5 illustrates the SHAP analysis results for the RF model. GFR had the strongest negative contribution (−0.067; +0.026), suggesting that higher GFR values were associated with reduced risk of combined obesity and hyperuricemia. ApoB and CysC showed nearly equal positive and negative contributions, implying variable impacts across samples. HDL-C primarily had negative contributions (−0.029; +0.021), aligning with its known protective effect and indicating its potential to reduce disease risk.

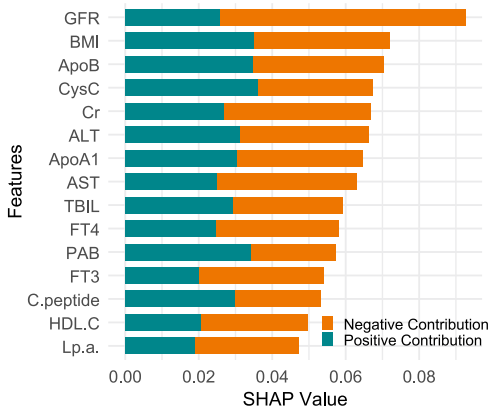

Figure 5 SHAP value ranking and feature contributions in the RF model.

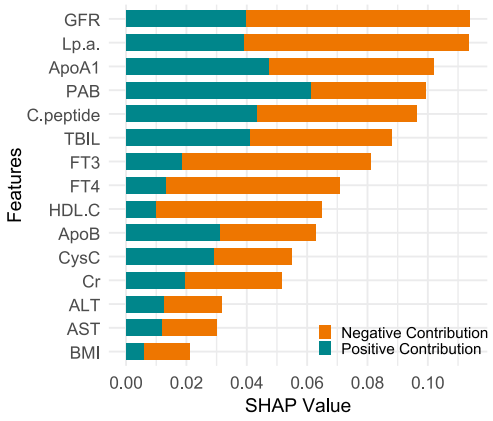

Figure 6 SHAP value ranking and feature contributions in the SVC model.

Figure 6 presents the SHAP values for the SVC model. GFR (−0.074; +0.040) and Lp(a) (−0.075; +0.039) were the most significant negative contributors, suggesting that higher levels were associated with lower risk. PAB had the most pronounced negative contribution (−0.038; +0.061), while ApoA1 (−0.055; +0.047) and FT3 (−0.062; +0.019) also demonstrated notable influence on model predictions.

Figure 7 shows SHAP dependence plots. BMI emerged as one of the strongest predictors, with SHAP values increasing sharply alongside BMI, indicating a strong positive association with disease risk. Other consistent contributors included Cr, CysC, and ApoB. GFR had largely negative SHAP values, particularly at low levels. C-peptide exhibited nonlinear effects: mildly negative in low to mid ranges and strongly positive at higher levels, possibly reflecting compensatory metabolic activity.

Figure 8 highlights SHAP dependence stratified by biochemical levels. Low HDL-C indicated metabolic disturbances, but the effect diminished or reversed with higher levels. Low Lp(a) levels suggested metabolic protection, while higher levels were associated with increased SHAP values and pro-inflammatory risk. For FT3 and FT4, lower values corresponded to elevated risk, while mid-to-high levels showed a slight reversal. AST and ALT showed negative contributions at low levels; however, due to sparse high-range data, SHAP values approached zero, limiting interpretability in that range.

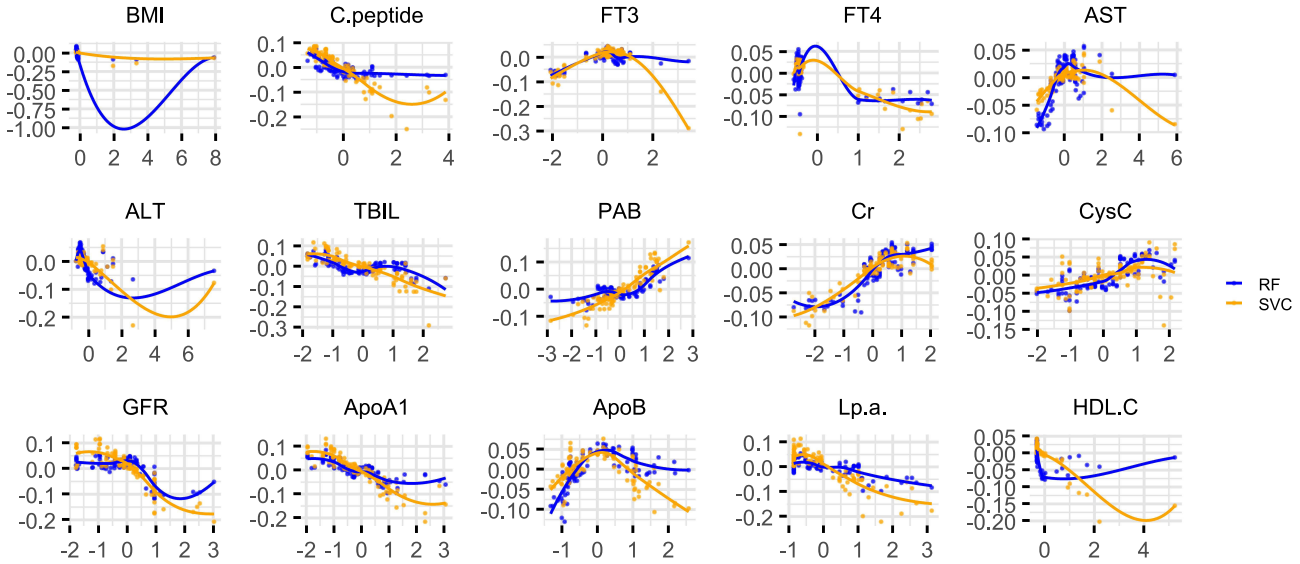

Figure 7 SHAP dependency plots of feature values and SHAP outputs.

https://doi.org/10.2147/DMSO.S519284

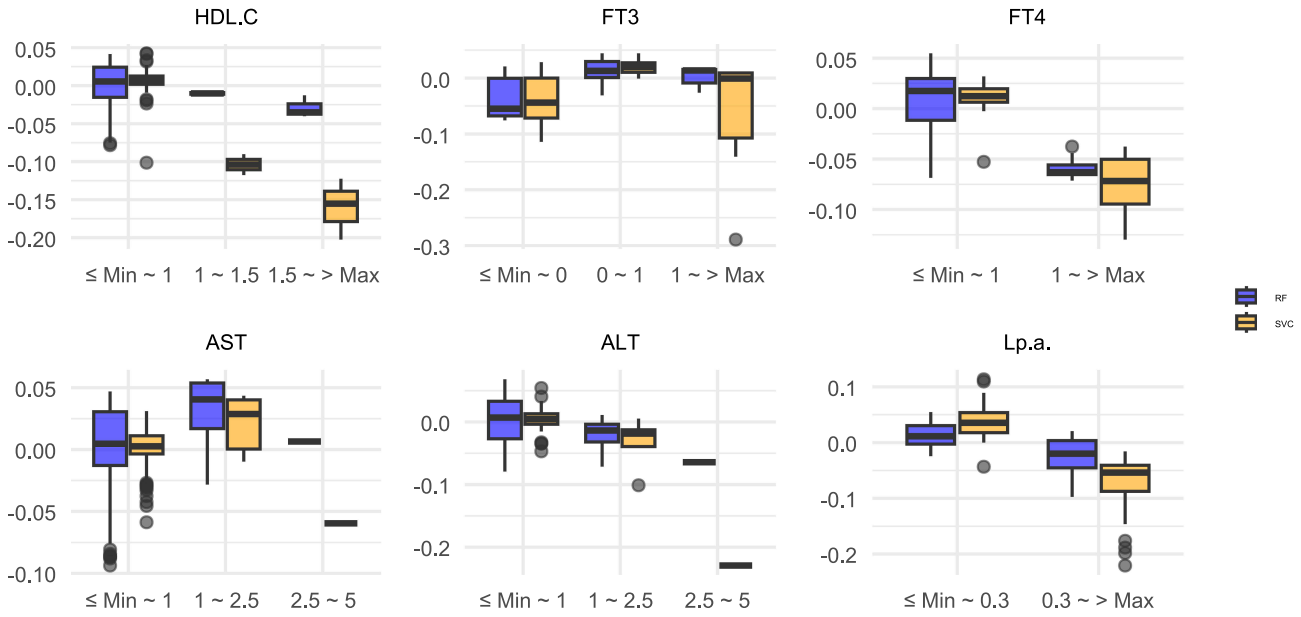

Figure 8 Binned feature analysis combined with SHAP interpretation.

Figure 9 compares the interaction intensity matrices for RF and SVC. The RF model showed a relatively even distribution, capturing complex inter-variable relationships, especially between BMI and variables such as C-peptide, GFR, and ApoB. In contrast, SVC demonstrated more localized interaction intensity, notably between ApoA1 and TBIL or Lp(a), reflecting its strength in capturing local structures in high-dimensional spaces. Lipid-related features (HDL-C, Lp(a), ApoB, ApoA1) and renal markers (ALT, CysC, Cr) exhibited strong internal interactions, especially in the SVC model. These patterns suggest that high-level metabolic indicators may jointly influence disease prediction. However, due to the complexity of these interactions, there is potential for redundancy and noise, which we aim to address in future modeling through feature fusion or dimensionality reduction strategies.

Discussion

In this study, we employed two classical machine learning models—Random Forest (RF) and Support Vector Classification (SVC)—to classify and predict childhood obesity and obesity combined with hyperuricemia. Both models demonstrated strong classification capabilities, with area under the curve (AUC) values of 0.96, indicating high accuracy and generalizability in identifying children with obesity and those with comorbid hyperuricemia. However, based on the precision-recall (PR) curves and comparative radar plots, SVC slightly outperformed RF in average precision (0.965 vs 0.964) and also showed superior performance in recall and F1-score.

Despite this, calibration curve analysis revealed that the RF model achieved better calibration in predicting risk across the entire probability range. Its predicted probabilities closely aligned with actual outcomes, suggesting that RF offers more reliable risk estimation. In contrast, the SVC model exhibited marked overestimation in low-probability regions, reflecting suboptimal calibration and the potential for misclassification of high-risk individuals in clinical settings.

Overall, the superior calibration of the RF model makes it more suitable for tasks requiring accurate probability estimation, such as risk stratification and clinical decision-making, thereby providing more dependable support for healthcare professionals. On the other hand, the relatively stronger classification performance of the SVC model suggests its utility in screening-based applications. However, its limited calibration restricts its role in precise risk prediction, highlighting the need for further optimization to improve its applicability in probability-based clinical scenarios.

In this study, we also utilized SHAP (Shapley Additive Explanations) values to analyze the influence of various clinical and metabolic features on uric acid (UA) levels. The findings were consistent with prior studies conducted in general populations. Uric acid is primarily produced in the liver and excreted by the kidneys, playing a key role in the

Figure 9 Heatmaps of feature interaction intensity in the RF and SVC models.

human antioxidant defense system.[19] In both models, GFR (glomerular filtration rate) showed a strong negative SHAP value, indicating the critical role of renal function in uric acid metabolism. This was particularly evident in children with obesity and hyperuricemia, where reduced GFR was closely associated with impaired uric acid clearance. Moreover, a decline in GFR also limits the kidney's ability to process lipid metabolic by-products. ApoA1 exerts an anti-atherosclerotic effect via reverse cholesterol transport (RCT), whereas ApoB promotes lipid deposition and inflammatory responses through the low-density lipoprotein pathway, thereby influencing uric acid levels and further contributing to atherogenesis and inflammation.[20,21] According to the study by Markova, Irena, dysregulation of lipoprotein metabolism can lead to lipotoxicity-induced lipid accumulation in the kidneys, triggering the secretion of renal inflammatory mediators such as MCP-1 and IL-6. Simultaneously, this process suppresses the secretion of epidermal growth factor (EGF), thereby impairing GFR. These interactions ultimately form a vicious cycle among lipid deposition, inflammation, and deteriorating renal filtration function.[22]

The significance of PAB in the model for obesity combined with hyperuricemia suggests a disturbance in protein metabolism. According to the study by Wunderlich, F. Thomas, this may be attributed to its role in activating inflammatory mediators such as IL-6, which could, in turn, impair the body's capacity to regulate uric acid metabolism.[23] C-peptide, as a marker of pancreatic beta-cell function, is closely associated with insulin resistance. Insulin resistance promotes enhanced lipolysis in adipose tissue, leading to elevated levels of free fatty acids (FFAs). These FFAs can activate the NF-κB signaling pathway, thereby exacerbating systemic inflammatory responses.[24,25]

https://doi.org/10.2147/DMSO.S519284

Furthermore, under conditions of insulin resistance, uric acid can induce excessive production of nitric oxide (NO) through oxidative stress. This excess NO may, in turn, influence the regulation of uric acid metabolism.[26]

Thyroid function plays a crucial regulatory role in obesity and uric acid metabolism by modulating basal metabolic rate and energy balance, thereby influencing both the production and excretion of uric acid.[27] It also indirectly regulates uric acid metabolism by enhancing the sensitivity of adipose tissue to insulin, thereby modulating insulin's metabolic effects.[28]

The levels of ALT and AST exhibited a complex nonlinear relationship, which is consistent with existing studies. Better liver function has been associated with a lower risk of metabolic diseases, particularly in relation to long-term systolic blood pressure variability.[29] However, in the high-level range, our model failed to effectively learn significant predictive patterns, likely due to the limited sample size within this interval. This contrasts with the findings of Ambrosi, Bruno et al, and we speculate that this discrepancy may stem from limitations in our sample inclusion criteria, which may not have adequately captured the role of liver function markers at higher levels.[30] As a byproduct of muscle metabolism, creatinine levels are positively correlated with muscle mass. Conversely, muscle mass is inversely associated with abnormal metabolic states such as hyperuricemia and insulin resistance.[31]

As a conventional indicator of body fat accumulation, BMI demonstrated predictive capacity in both models. However, its relative importance in the context of obesity combined with hyperuricemia was comparatively lower. This finding aligns with current research trends in metabolic syndrome, where factors such as fat distribution—particularly visceral adiposity—and inflammatory markers like CRP and TNF-α are considered more significant contributors to disease progression.[32] An increase in BMI concurrently leads to the expansion of adipose tissue, which directly impairs hepatic insulin sensitivity and disrupts uric acid clearance pathways.[33]

SHAP analysis highlighted the pivotal roles of GFR, HDL-C, and ApoB in the predictive models. In future work, we plan to incorporate binning plots to further illustrate their nonlinear risk patterns. This analytical approach not only enhances the interpretability of the results but also improves clinical comprehension of model-based decisions. Compared to traditional statistical models, both RF and SVC are better suited for handling high-dimensional, nonlinear features without relying on distributional assumptions, thus maintaining strong modeling performance. The application of RFE and ROSE enabled efficient training with fewer variables, facilitating rapid model development and risk identification in small-sample clinical studies.

The binning analysis further identified HDL-C and Lp(a) as the most influential features within the model. Elevated levels of HDL-C demonstrated protective effects, likely attributable to its antioxidative, anti-inflammatory, and reverse cholesterol transport functions. However, this relationship was not strictly linear—distinct functional patterns were observed across its different subtypes (notably HDL-C2 and HDL-C3), which exhibited opposite biological effects.[34] In contrast, elevated levels of Lp(a) were associated with a significantly increased risk, primarily through its pro-atherogenic effects, inhibition of fibrinolytic activity, and promotion of inflammatory responses.[35]

In summary, the prediction models for obesity and obesity combined with hyperuricemia underscore the critical roles of renal function, lipid metabolism, insulin resistance, endocrine regulation, and inflammatory response in pediatric metabolic syndrome. These findings elucidate the complex biological mechanisms underlying obesity and hyperuricemia and provide a scientific basis for early clinical intervention and management strategies.

In this study, we constructed two prediction models: Random Forest (RF) and Support Vector Classification (SVC). RF demonstrates advantages in capturing overall feature trends; however, due to its inherent randomness, the interpretation of feature importance may be unstable, with some variables showing inconsistent contributions across samples. In contrast, SVC, being more sensitive to decision boundaries, captures subtle inter-sample variations more effectively—especially in features such as GFR and Lp(a)—but it is also more prone to overfitting when dealing with noisy or irregular data. Thus, future work may explore ensemble modeling approaches to integrate the strengths of both algorithms, such as weighted averaging or stacking, to enhance predictive robustness and accuracy in identifying children at risk of obesity with hyperuricemia.

Importantly, model selection should align with specific clinical scenarios. SVC performs slightly better in average precision and recall, making it more suitable for sensitivity-oriented preliminary screening, such as identifying high-risk children in community or school-based health assessments. RF, on the other hand, offers superior calibration and is more

applicable to probability-driven clinical decision-making tasks, such as planning individualized interventions or follow-up strategies.

Despite leveraging SHAP analysis and binning-based interpretation to explore complex clinical and metabolic predictors of uric acid levels, this study has several limitations. The relatively small sample size constrains the model's ability to fully learn intricate variable interactions. Additionally, the current models and conclusions have not yet been validated using an independent external dataset, limiting the assessment of model generalizability. In future work, we aim to address these limitations by expanding the sample size and incorporating more sophisticated algorithms such as XGBoost and LightGBM, which offer enhanced expressiveness and calibration performance for complex data. Furthermore, we recognize the importance of model interpretability and transparency. While RF and SVC offer notable performance advantages, their "black-box" nature may hinder clinical trust and adoption. To mitigate this, we propose incorporating local interpretability tools such as LIME (Local Interpretable Model-Agnostic Explanations) to generate case-specific feature contributions that clinicians can readily understand, thereby increasing model transparency and acceptance.

Currently, we have collected metabolomics data for a subset of our study population. In the next phase, we plan to integrate this data into our prediction framework to further investigate metabolic interactions among key variables, enhancing mechanistic insight and the interpretability of model outputs.

Conclusion

In this study, we employed Random Forest (RF) and Support Vector Classification (SVC) models to predict the risk of childhood obesity and obesity combined with hyperuricemia. Both models demonstrated high classification performance. SVC showed a slight advantage in classification metrics, particularly in terms of average precision and recall on the Precision-Recall curve. In contrast, RF exhibited more stable calibration performance and more accurately predicted the probability of hyperuricemia in obese children.

SHAP analysis highlighted the key roles of renal function (eg, GFR), lipid metabolism (eg, ApoB, HDL-C), insulin resistance, endocrine regulation (eg, FT3, FT4), and inflammatory response. Notably, variables such as GFR and HDL-C exhibited protective effects, while lipoproteins (eg, ApoB, Lp(a)) and C-peptide were associated with increased disease risk through multiple mechanisms. The binned feature analysis further revealed nonlinear risk patterns, emphasizing the complex influence of these variables across different clinical ranges.

Given its superior precision and recall, the SVC model may be more suitable for early screening scenarios. In contrast, the RF model—with its superior calibration—may better support probability-based clinical decision-making, such as guiding individualized interventions or monitoring strategies. Our findings offer a valuable reference for the early identification of children at risk for obesity and hyperuricemia and could facilitate earlier intervention to reduce the incidence and long-term consequences of metabolic disorders.

Future work will aim to expand the sample size, incorporate ensemble learning methods, and improve model interpretability using techniques such as LIME. Additionally, integrating metabolomics data may further enhance the predictive power and clinical applicability of these models.

Ethical Statement

This study was conducted in accordance with the ethical standards of the Declaration of Helsinki and was approved by the Ethics Committee of the Affiliated Hospital of Nantong University. Written informed consent was obtained from all participants through their parents or legal guardians.

Disclosure

The authors declare no conflicts of interest in this work.

References

1. Ziauddeen N, Roderick PJ, Macklon NS, Alwan NA. Predicting childhood overweight and obesity using maternal and early life risk factors: a systematic review. *Obesity Rev*. 2018;19(3):302–312. doi:10.1111/obr.12640

2. Bi Q, Goodman KE, Kaminsky J, Lessler J. What is machine learning? A primer for the epidemiologist. *Am J Epidemiol*. 2019;188(12):2222–2239. doi:10.1093/aje/kwz189
3. Goel A, Reddy S, Goel P. Causes, Consequences, and Preventive Strategies for Childhood Obesity: a Narrative Review. *Cureus*. 2024;16(7): e64985–e64985. doi:10.7759/cureus.64985
4. Guo Q, Zhang C, Wang Y. Overexpression of apolipoprotein A-I alleviates endoplasmic reticulum stress in hepatocytes. *Lipids Health Dis*. 2017;16 (1):105. doi:10.1186/s12944-017-0497-3
5. Teng Y, Xu L, Li W, Liu P, Tian L, Liu M. Targeting reactive oxygen species and fat acid oxidation for the modulation of tumor-associated macrophages: a narrative review. *Front Immunol*. 2023;14:1224443. doi:10.3389/fimmu.2023.1224443
6. Seyed-Sadjadi N, Berg J, Bilgin AA, Grant R. Visceral fat mass: is it the link between uric acid and diabetes risk? *Lipids Health Dis*. 2017;16 (1):142.
7. McHill AW, Hull JT, Klerman EB. Chronic circadian disruption and sleep restriction influence subjective hunger, appetite, and food preference. *Nutrients*. 2022;14(9):1800. doi:10.3390/nu14091800
8. Obeidat AA, Ahmad MN, Haddad FH, Azzeh FS. Leptin and uric acid as predictors of metabolic syndrome in Jordanian adults. *Nutr Res Pract*. 2016;10(4):411–417. doi:10.4162/nrp.2016.10.4.411
9. Kenney EL, Gortmaker SL. United States adolescents' television, computer, videogame, smartphone, and tablet use: associations with sugary drinks, sleep, physical activity, and obesity. *J Pediatr*. 2017;182:144–149. doi:10.1016/j.jpeds.2016.11.015
10. Liu HY, Leng Y, Wu YC, Chau PH, Chung TWH, Fong DYT. Robust identification key predictors of short- and long-term weight status in children and adolescents by machine learning. *Front Public Health*. 2024;12:1.
11. Kim C, Costello FJ, Lee KC, Li Y, Li CY. Predicting factors affecting adolescent obesity using general Bayesian network and what-if analysis. *Int J Environ Res Public Health*. 2019;16(23):4684. doi:10.3390/ijerph16234684
12. Wang Y, Wang L, Qu W. New national data show alarming increase in obesity and noncommunicable chronic diseases in China. *Eur J Clin Nutr*. 2017;71(1):149–150. doi:10.1038/ejcn.2016.171
13. Suchindran C, North KE, Popkin BM, Gordon-Larsen P. Association of adolescent obesity with risk of severe obesity in adulthood. *JAMA*. 2010;304(18):2042–2047. doi:10.1001/jama.2010.1635
14. Zimmet P, Alberti KG, Kaufman F, et al. The metabolic syndrome in children and adolescents - an IDF consensus report. *Pediatr Diabetes*. 2007;8 (5):299–306. doi:10.1111/j.1399-5448.2007.00271.x
15. Wang H-Y, Chang S-C, Lin W-Y, et al. Machine learning-based method for obesity risk evaluation using single-nucleotide polymorphisms derived from next-generation sequencing. *J Comput Biol*. 2018;25(12):1347–1360. doi:10.1089/cmb.2018.0002
16. Das A, Connell M, Khetarpal S. Digital image analysis of ultrasound images using machine learning to diagnose pediatric nonalcoholic fatty liver disease. *Clin Imaging*. 2021;77:62–68. doi:10.1016/j.clinimag.2021.02.038
17. Safaei M, Sundararajan EA, Driss M, Boulila W, Shapi'i A. A systematic literature review on obesity: understanding the causes & consequences of obesity and reviewing various machine learning approaches used to predict obesity. *Comput Biol Med*. 2021;136:104754. doi:10.1016/j.compbiomed.2021.104754
18. Laboratories MC. Uric acid, serum - pediatric catalog. 2021. Available from: https://pediatric.testcatalog.org/show/URIC. Accessed November 21, 2024.
19. Foster C, Smith L, Alemzadeh R. Excess serum uric acid is associated with metabolic syndrome in obese adolescent patients. *J Diabetes Metab Disord*. 2020;19(1):535–543. doi:10.1007/s40200-020-00507-2
20. Smoak KA, Aloor JJ, Madenspacher J, et al. Myeloid differentiation primary response protein 88 couples reverse cholesterol transport to inflammation. *Cell Metab*. 2010;11(6):493–502. doi:10.1016/j.cmet.2010.04.006
21. Reyes-Soffer G, Ginsberg HN, Berglund L, et al. Lipoprotein (a): a genetically determined, causal, and prevalent risk factor for atherosclerotic cardiovascular disease: a scientific statement from the American Heart Association. *Arterioscler Thromb Vasc Biol*. 2022;42(1):e48–e60. doi:10.1161/ATV.0000000000000147
22. Markova I, Miklankova D, Hüttl M, et al. The effect of lipotoxicity on renal dysfunction in a nonobese rat model of metabolic syndrome: a urinary proteomic approach. *J Diabetes Res*. 2019;2019(1):8712979. doi:10.1155/2019/8712979
23. Wunderlich FT, Ströhle P, Könner AC, et al. Interleukin-6 signaling in liver-parenchymal cells suppresses hepatic inflammation and improves systemic insulin action. *Cell Metab*. 2010;12(3):237–249. doi:10.1016/j.cmet.2010.06.011
24. Shi H, Kokoeva MV, Inouye K, Tzameli I, Yin H, Flier JS. TLR4 links innate immunity and fatty acid–induced insulin resistance. *J Clin Invest*. 2006;116(11):3015–3025. doi:10.1172/JCI28898
25. Shin Y-E, Choi JW, Park YI, Kim H-K. 7,8-Dihydroxyflavone attenuates inflammatory response and insulin resistance induced by the paracrine interaction between adipocytes and macrophages. *Int J Mol Sci*. 2023;24(4):3520. doi:10.3390/ijms24043520
26. King C, Lanaspa MA, Jensen T, Tolan DR, Sánchez-Lozada LG, Johnson RJ. Uric Acid as a Cause of the Metabolic Syndrome. *Contrib Nephrol*. 2018;192:88–102.
27. Desideri G, Bocale R, D'Amore AM, et al. Thyroid hormones modulate uric acid metabolism in patients with recent onset subclinical hypothyroidism by improving insulin sensitivity. *Int Emerg Med*. 2020;15:67–71. doi:10.1007/s11739-019-02065-9
28. Ambrosi B, Masserini B, Iorio L, et al. Relationship of thyroid function with body mass index and insulin-resistance in euthyroid obese subjects. *J Endocrinol Invest*. 2010;33:640–643. doi:10.1007/BF03346663
29. Liu PY, Lin YK, Chen KW, et al. Association of liver transaminase levels and long-term blood pressure variability in military young males: the CHIEF Study. *Int J Environ Res Public Health*. 2020;17(17). doi:10.3390/ijerph17176094
30. Nakatsu Y, Seno Y, Kushiyama A, et al. The xanthine oxidase inhibitor febuxostat suppresses development of nonalcoholic steatohepatitis in a rodent model. *Am J Physiol Gastrointest Liver Physiol*. 2015;309(1):G42–G51. doi:10.1152/ajpgi.00443.2014
31. Chaudhari UK, Newcomb JD, Hansen BC. Serum creatinine progressively decreases with obesity and type 2 diabetes in nonhuman primates. *Diabetes*. 2018;67(Supplement_1). doi:10.2337/db18-1931-P
32. Carbone F, Lattanzio MS, Minetti S, et al. Circulating CRP levels are associated with epicardial and visceral fat depots in women with metabolic syndrome criteria. *Int J Mol Sci*. 2019;20(23):5981. doi:10.3390/ijms20235981
33. Kern PA, Ranganathan S, Li C, Wood L, Ranganathan G. Adipose tissue tumor necrosis factor and interleukin-6 expression in human obesity and insulin resistance. *Am J Physiol Endocrinol Metab*. 2001;280(5):E745–751. doi:10.1152/ajpendo.2001.280.5.E745

34. Okuda LS, Castilho G, Rocco DD, Nakandakare ER, Catanozi S, Passarelli M. Advanced glycated albumin impairs HDL anti-inflammatory activity and primes macrophages for inflammatory response that reduces reverse cholesterol transport. *BBA*. 2012;1821(12):1485–1492. doi:10.1016/j.bbalip.2012.08.011
35. Stulnig TM, Morozzi C, Reindl-Schwaighofer R, Stefanutti C. Looking at Lp (a) and related cardiovascular risk: from scientific evidence and clinical practice. *Curr Atheroscler Rep*. 2019;21:1–9. doi:10.1007/s11883-019-0803-9