Liu et al. BMC Endocrine Disorders (2025) 25:291
https://doi.org/10.1186/s12902-025-02104-x

BMC Endocrine Disorders

RESEARCH

Open Access

Construction and validation of a hypoglycemia risk prediction model for hospitalized type 2 diabetes patients based on machine learning

Caixia Liu[1†], Zhaoxu Huang[2†], Tuonan Liu[3], Yangyuan Ge[1], Jie Yuan[1], Yue Lin[2], Chunli Wang[1], Jinjuan Zhang[1], Xiaoli Wang[1], Yan Hua[3*†] and Rui Qi[1*†]

Abstract

Background To compare three machine learning algorithms for constructing a hypoglycemia risk prediction model in hospitalized type 2 diabetes patients, identify the optimal model, and validate it to provide decision-making support for early clinical identification of high-risk patients.

Methods A case-control study design was adopted, retrospectively collecting clinical data from 1,167 hospitalized type 2 diabetes patients in the endocrinology department of a tertiary hospital from January to December 2024. Patients were divided into a hypoglycemia group (220 cases) and a non-hypoglycemia group (947 cases). After screening predictive variables using LASSO regression, the data were randomly split into a training set (934 cases) and a validation set (233 cases) at an 8:2. The training set was used to construct prediction models using Logistic Regression, Random Forest (RF), and Extreme Gradient Boosting (XGBoost) algorithms, with internal validation performed on the validation set to assess predictive performance. The optimal model was determined by comprehensively evaluating the Area Under the ROC Curve (AUC) and F1 score. The SHAP (Shapley Additive Explanations) method was applied for interpretability analysis.

Results The incidence of hypoglycemia was 18.85% (220/1,167). LASSO regression identified nine key predictive variables: random C-peptide, insulin-containing fluid infusion, BMI, length of hospital stay, age, renal dysfunction, albumin level, lipohypertrophy, and insulin antibodies, all of which were statistically significant ($P < 0.05$). Validation results showed that the XGBoost model exhibited the best predictive performance in both the training set

†Liu Cai-xia and Huang Zhao-xu contributed equally to this work as co-first authors.

†Qi Rui and Hua Yan contributed equally to this paper as co-last authors.

*Correspondence:
Yan Hua
huayan1112@126.com
Rui Qi
rachelqqi@126.com

Full list of author information is available at the end of the article

(AUC = 0.853) and the validation set (AUC = 0.910), outperforming the other models significantly. SHAP analysis revealed the contribution of each feature to the prediction.

Conclusion The prediction model developed with the XGBoost algorithm demonstrated superior discriminative performance, providing a reliable tool for clinical identification of high-risk hypoglycemia in hospitalized type 2 diabetes patients.

Clinical trial number Not applicable.

Keywords Hospitalized type 2 diabetes, Hypoglycemia, Prediction model, Machine learning, XGBoost algorithm

Introduction

In recent years, factors such as population aging, sedentary behavior, and obesity have significantly contributed to the increase in the prevalence of type 2 diabetes mellitus (T2DM) [1]. The global number of cases increased from 151 million in 2000 to 425 million in 2017, with projections indicating a rise to 629 million by 2045 [2]. T2DM is characterized by insulin resistance and β-cell dysfunction, with chronic hyperglycemia damages organs such as the heart, kidneys, and eyes, resulting in macrovascular and microvascular complications [3]. Hypoglycemia, an acute complication of diabetes, is one of the major obstacles in blood glucose control. The American Diabetes Association (ADA) classifies hypoglycemia into five categories: severe hypoglycemia (requiring assistance to recover), symptomatic hypoglycemia (blood glucose ≤ 3.9 mmol/L), asymptomatic hypoglycemia (blood glucose ≤ 3.9 mmol/L without symptoms), probable hypoglycemia (symptoms present but blood glucose not measured), and relative hypoglycemia (symptoms present but blood glucose >3.9 mmol/L) [4]. Studies indicate that around 46.5% of patients with T2DM experience monthly episodes of hypoglycemia [5], with 31.6% of insulin-treated patients at risk of severe hypoglycemia, particularly those on intensive therapy or short-acting insulin [6]. The harms of hypoglycemia include cognitive dysfunction (e.g., confusion, coma) and cardiovascular events (e.g., arrhythmias) [7, 8]. Psychologically, the fear of hypoglycemia may overshadow concerns about long-term complications, impacting not only patients but also their families and friends [9]. Recurrent hypoglycemia impairs defense mechanisms and increases mortality risk. Additionally, each hypoglycemic episode incurs additional medical costs of approximately £1,200, posing a significant economic burden [10]. Therefore, accurate identification and management of hypoglycemia are critical for preventing adverse events and have become key indicators of diabetes care quality [11]. Scientific assessment and intervention can enhance care safety and patient quality of life.

Machine learning, which utilizes algorithmic models to discern patterns from data and generate predictions, has been widely applied in medicine, particularly in diabetes management. It can extract crucial insights from clinical data to aid physicians in diagnosis and risk assessment, particularly in repetitive and uncertain medical scenarios [12]. Common machine learning algorithms include Logistic Regression, Support Vector Machine (SVM), Random Forest (RF), and Extreme Gradient Boosting (XGBoost) [13, 14]. In diabetes care, machine learning has been used to predict disease progression, blood glucose fluctuations, and complication risks. For instance, the Eye Art system employs machine learning to screen for diabetic retinopathy, achieving sensitivity and specificity levels exceeding 91% [15]. Similarly, algorithms like Random Forest have been utilized to forecast the risk of diabetes onset and track blood glucose trends, facilitating early intervention [16]. For example, Random Forest or gradient boosting algorithms can amalgamate data from various sources (such as electronic health records and real-time monitoring) to develop high-precision prediction models, enabling healthcare providers to take proactive measures to mitigate hypoglycemic events [17, 18].

Given the critical need for proactive hypoglycemia management in hospitalized T2DM patients, this study aims to develop and validate a robust prediction tool leveraging machine learning. We specifically sought to compare the predictive performance of three distinct algorithms—Logistic Regression, Random Forest, and XGBoost—in forecasting hypoglycemia risk. These algorithms were selected to represent a spectrum of modeling techniques, from a traditional linear classifier to more complex, ensemble-based tree methods [19, 20]. The models were constructed using clinically relevant variables filtered through LASSO regression from a comprehensive dataset [21]. The primary objective was to identify the optimal model through rigorous internal validation, evaluating metrics such as the Area Under the ROC Curve (AUC) and F1 score [22]. Furthermore, we employed the SHAP (SHapley Additive exPlanations) framework to enhance the interpretability of the best-performing model [22]. The ultimate goal is to provide a validated, high-performance, and clinically interpretable tool to assist healthcare providers in the early identification of high-risk inpatients, thereby facilitating timely interventions and improving patient safety.

Subjects and methods

Study subjects

A case-control design was employed to retrospectively gather data from T2DM patients hospitalized in the endocrinology department of the First Affiliated Hospital of Air Force Medical University from January to December 2024. Inclusion criteria: (1) Diagnosis of T2DM and hypoglycemia met; (2) Hospitalized status. Exclusion criteria: (1) Admission for hypoglycemia, diabetic ketoacidosis, or hyperosmolar hyperglycemic state; (2) Presence of infectious diseases, acute coronary syndrome, malignancies, anemia, or renal failure; (3) Missing or duplicate medical records. In addition, for participants under the age of 16, we have provided comprehensive information regarding the study to their parents or legal guardians and obtained written informed consent from them.

Considering the need for 5–10 samples per variable for predictive modeling and an estimated hypoglycemia incidence of 20% [23], the minimum sample size was calculated as 875 cases (accounting for 20% potential data loss). Ultimately, the study comprised 1,167 cases, with 220 (18.85%) in the hypoglycemia group and 947 (81.15%) in the non-hypoglycemia group.

Study variables and data extraction methods

Based on the literature review, discussions within the research team, and two rounds of expert inquiries, a screening form for risk factors of hypoglycemia in hospitalized type 2 diabetic patients was designed, modified, and finalized. The form includes a total of 28 risk factors. General information: age, gender, BMI, educational level, length of hospital stay. Diabetes-related information: duration of diabetes, diabetes treatment plan (basal insulin, intensive insulin, oral medication). Laboratory-related tests: C-peptide, insulin antibody, glycosylated hemoglobin, total cholesterol, triglycerides, high-density lipoprotein cholesterol, low-density lipoprotein cholesterol, hemoglobin, free triiodothyronine, free thyroxine, total triiodothyronine, total thyroxine, thyroid stimulating hormone, vitamin D. Other comorbidities: abnormal liver function, abnormal kidney function, fat hyperplasia, hypothyroidism, creatinine abnormality.

The relevant data of type 2 diabetic patients were extracted from the hospital information system (HIS), including general patient information, diabetes-related information, treatment plans, and laboratory tests. If the patient had a blood glucose level of less than 3.9 mmol/L, they were classified as the hypoglycemia occurrence group.

Statistical methods

SPSS 25.0 and R 4.3 were used for data analysis. Categorical data were described using frequencies and percentages, with group comparisons conducted using chi-square tests. Continuous variables (e.g., age, BMI, random C-peptide) were categorized based on clinical reference values. Missing data exceeding 20% were excluded, while data with less than 20% missing values were imputed using the mode.

Prediction model construction

LASSO regression analysis was performed in R software (version 4.3) to identify factors influencing hypoglycemia. To rigorously evaluate model performance and prevent overfitting, the entire dataset was randomly split into a training set (80% of the data) and an internal hold-out test set (20% of the data). Three machine learning models—Logistic Regression, Random Forest, and XGBoost (Extreme Gradient Boosting)—were developed with these algorithms. The training set was exclusively used for model training and parameter tuning. To further optimize model parameters and assess stability during training, 5-fold cross-validation was employed on the training set. The hold-out test set, which was not involved in any aspect of model building or training, was strictly reserved for the final performance evaluation. After removing result labels from the previously divided test set data, it was fed into the trained machine learning model to obtain metrics such as prediction sensitivity, specificity, PPV, NPV, F1 score, and accuracy. The model was evaluated using the ROC curve and the area under the curve (AUC), and the model with the best predictive performance was selected. The SHAP (Shapley Additive exPlanations) method was employed to interpret the model results, evaluate the significance of each feature in the decision-making process, and offer visual insights into the role of each feature in decision-making.

Results

Baseline characteristics

A total of 1,167 hospitalized patients with type 2 diabetes were included in this study, of whom 220 (18.85%) experienced hypoglycemia, while 947 (81.15%) did not. Univariate analysis revealed several factors significantly associated with hypoglycemia ($P < 0.05$), including age, body mass index (BMI), treatment regimen, administration of insulin-containing fluids, C-peptide levels, insulin antibody levels, liver and renal function status, presence of lipodystrophy, serum albumin, triglycerides, high-density lipoprotein (HDL) cholesterol, hemoglobin, free triiodothyronine (FT3), and length of hospital stay.

Patients at higher risk for hypoglycemia were younger (aged 13–39), had a BMI < 18.5 kg/m², serum albumin < 30 g/L, hemoglobin levels between 60 and 90 g/L, and FT3 < 3.85 pmol/L. Metabolic indicators associated with increased hypoglycemia risk included C-peptide < 0.33 nmol/L, insulin antibody levels > 20%, triglycerides ≤ 1.7 mmol/L, and HDL cholesterol > 2.06 mmol/L.

Regarding treatment, intensive insulin therapy and the absence of insulin-containing fluid administration were significantly linked to a higher incidence of hypoglycemia. Additionally, patients with abnormal liver function, impaired renal function, lipodystrophy, or a hospital stay of 7–14 days showed an increased risk. No significant association was found between hypoglycemia and other factors (Table S1, $P > 0.05$).

Analysis of the characteristics of hypoglycemic events

To further elucidate the patterns of hypoglycemic episodes, a detailed analysis was performed on the hypoglycemia group ($n = 220$) regarding severity, frequency, and timing. As shown in Fig. 1, the majority of hypoglycemic events (192 patients, 87.3%) were mild (blood glucose 3.0-3.9 mmol/L), while 28 patients (12.7%) experienced severe hypoglycemia (blood glucose < 3.0 mmol/L). In terms of frequency, 178 patients (80.9%) had a single episode, 32 (14.5%) had two episodes, and 10 (4.5%) had three or more episodes. Crucially, the temporal distribution of hypoglycemia revealed that the postprandial period was the most vulnerable window, with 102 cases (46.4%) occurring within 2 h after a meal. This was followed by fasting (46 cases, 20.9%), preprandial (41 cases, 18.6%), bedtime (18 cases, 8.2%), and nocturnal periods (13 cases, 5.9%). This comprehensive profile indicates that our model captures a full spectrum of hypoglycemic events, with a particular concentration in the postprandial phase, providing critical insights for timing-specific clinical interventions.

Variable selection via LASSO regression

Hypoglycemia occurrence was used as the dependent variable (0 = no hypoglycemia, 1 = hypoglycemia), and the

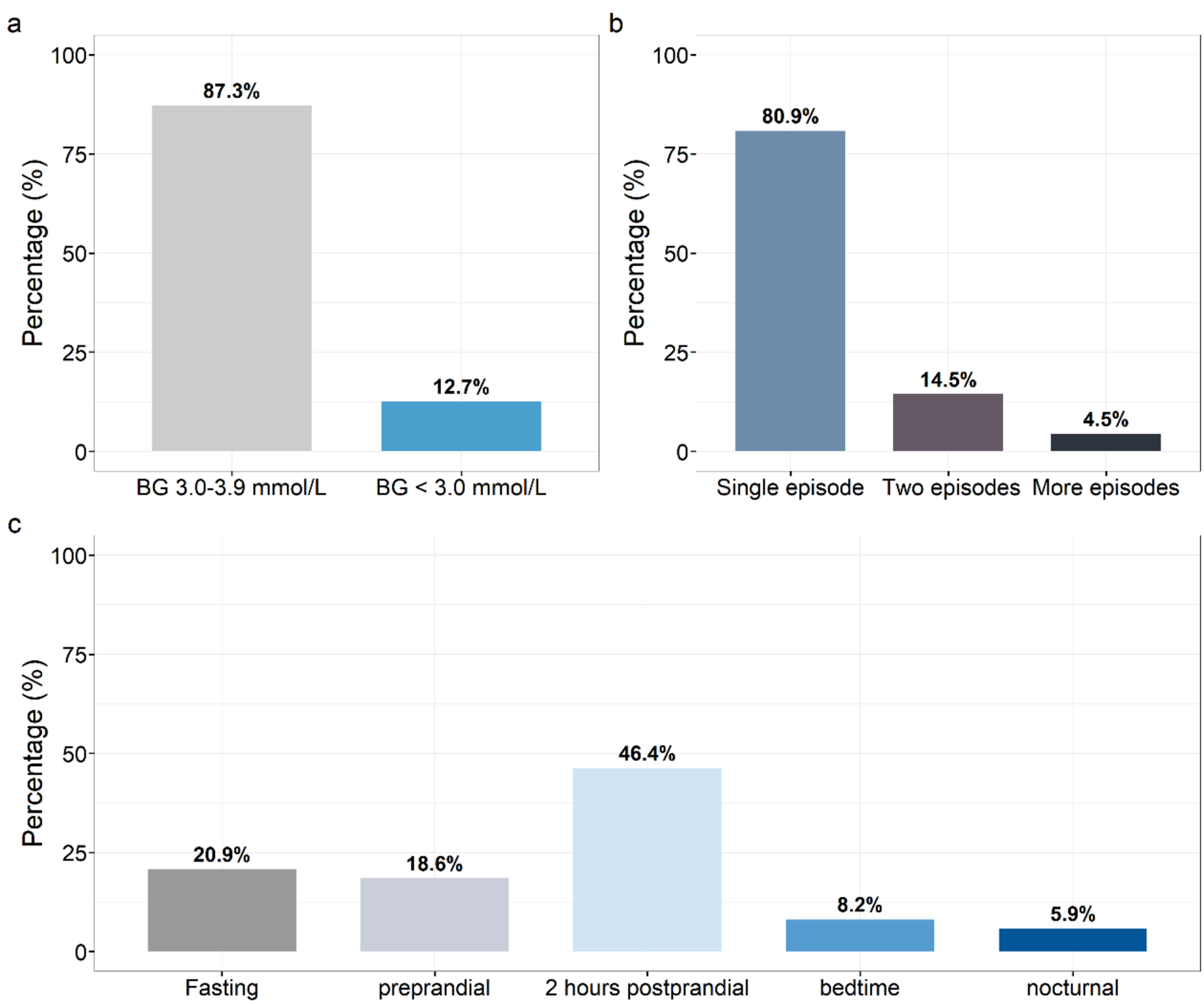

Fig. 1 Distribution Analysis of Hypoglycemic Events. (**a**) Severity of Hypoglycemia; (**b**) Frequency of Hypoglycemia; (**c**) Timing of Hypoglycemia

previously identified high-risk indicators were included as independent variables. To reduce dimensionality and eliminate multicollinearity, Least Absolute Shrinkage and Selection Operator (Lasso) regression was applied. This technique penalizes the regression coefficients, shrinking some to zero, thereby performing variable selection. The Lasso path plot identified two key regularization parameters: Lambda.min ($\lambda = 0.0083$), which provides the minimum cross-validated error, and Lambda.1se ($\lambda = 0.0278$), representing the most regularized model within one standard error of the minimum. To achieve a more parsimonious model and enhance generalizability, Lambda.1se was chosen as the optimal cutoff (Figure S1-S2). Based on this criterion, nine predictors were selected for model construction: random C-peptide level, administration of insulin-containing fluids, body mass index (BMI), length of hospital stay, age, impaired renal function, albumin classification, presence of lipodystrophy, and insulin antibody classification (Fig. 2).

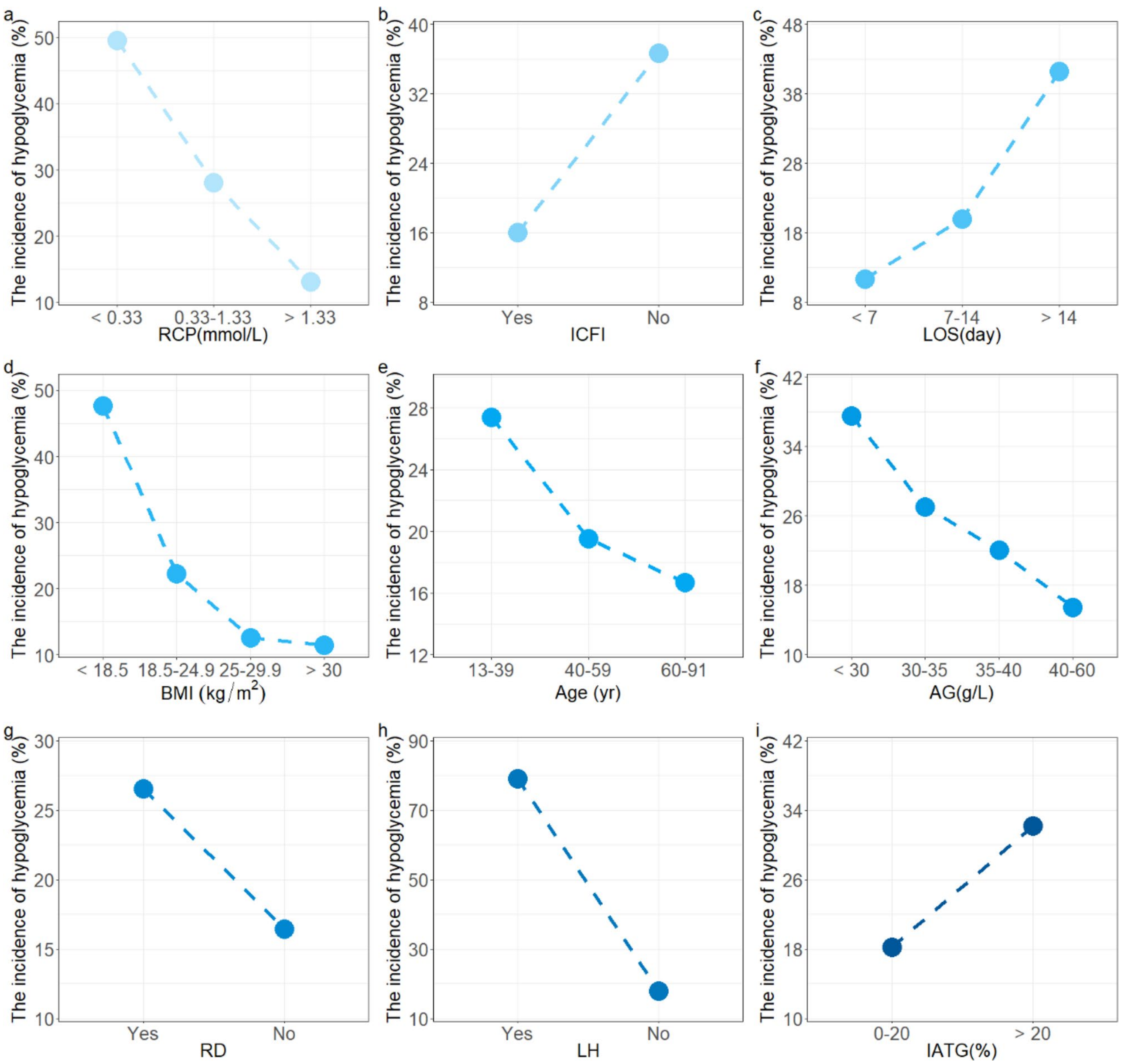

Fig. 2 Lasso Regression Model Identifying Nine Key Predictors of Hypoglycemia. Using hypoglycemia occurrence as the dependent variable (0 = no hypoglycemia, 1 = hypoglycemia), Lasso regression was applied for dimensionality reduction and variable selection. This method penalizes less informative variables by shrinking their coefficients to zero. Based on the optimal regularization parameter (Lambda.1se), nine predictive factors were selected: Random C-Peptide(RCP), Insulin-containing Fluid Infusion(ICFI), Body mass index (BMI), Length of Hospital Stay(LOS), age(Age), Renal Dysfunction(RD), Albumin Grade(AG), Lipohypertrophy(LH), Insulin AntibodyTiter Grade(IATG)

Using hypoglycemia occurrence as the dependent variable (0 = no hypoglycemia, 1 = hypoglycemia), Lasso regression was applied for dimensionality reduction and variable selection. This method penalizes less informative variables by shrinking their coefficients to zero.

Model performance

Using the nine predictive variables identified through Lasso regression, three machine learning models (logistic regression, random forest, and XGBoost) were developed to predict in-hospital hypoglycemia. For model training, 80% of the patient dataset was randomly selected, and 5-fold cross-validation was employed to assess model performance and generalizability.

Model performance was evaluated using confusion matrices and the area under the receiver operating characteristic curve (AUC), along with sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), F1 score, and overall accuracy. Among the models, XGBoost demonstrated the best predictive performance, achieving AUC values of 0.853 on the training set and 0.910 on the testing set. These values were notably higher than those of the logistic regression model (AUC = 0.833 and 0.824, respectively) and the random forest model (AUC = 0.844 and 0.862, respectively) (Table 1).

In both the training and testing sets, the XGBoost model consistently outperformed the other two models across all evaluation metrics (Fig. 3; Table 1), indicating superior accuracy, robustness, and stability. Therefore, XGBoost was selected as the optimal model for predicting in-hospital hypoglycemia.

Key drivers of hypoglycemia

To enhance the interpretability of the XGBoost-based predictive model for in-hospital hypoglycemia risk among patients with type 2 diabetes, SHAP (SHapley Additive exPlanations) analysis was employed. This method provides a unified approach to explain the contribution of each feature to individual predictions. The SHAP summary plot indicated that random C-peptide (RCP), insulin-containing fluid infusion (ICFI), length of hospital stay (LOS), and body mass index (BMI) were the most influential variables in predicting hypoglycemia risk (Fig. 4). Among them, lower RCP levels and the absence of insulin-containing fluid infusion were associated with a higher risk, while longer hospital stays and lower BMI also contributed significantly to increased hypoglycemia susceptibility. These findings highlight the clinical relevance of metabolic and treatment-related factors in identifying high-risk patients and support the robustness and transparency of the XGBoost model.

Discussion

Model performance and clinical relevance of hypoglycemia

In this study, we developed and validated a hypoglycemia risk prediction model for hospitalized T2DM patients using three machine learning algorithms. The observed hypoglycemia incidence of 18.85% aligns with previous reports [19], reinforcing hypoglycemia as a persistent clinical challenge in inpatient settings.

Our comparative analysis demonstrated the superior performance of the XGBoost model, which achieved an AUC of 0.910 in the validation set, significantly outperforming both Logistic Regression (AUC = 0.824) and Random Forest (AUC = 0.862). This robust discriminative ability, coupled with superior F1 scores across both training and validation cohorts, confirms XGBoost's exceptional capacity to capture complex, non-linear interactions among clinical variables for this prediction task [22]. The algorithm's ensemble structure, which sequentially builds multiple decision trees to correct previous errors, proves particularly effective in handling the heterogeneous nature of clinical data, including mixed variable types and potential missing values.

The model's clinical utility is further enhanced through the implementation of the SHAP framework, which effectively addresses the "black-box" concern often associated with sophisticated machine learning models. By providing quantitative, individualized explanations for each prediction, SHAP transforms the model from an opaque predictor into an interpretable clinical decision support tool. This transparency is crucial for fostering clinician trust and facilitating the integration of such tools into routine care pathways, ultimately bridging the gap between algorithmic performance and practical clinical application.

Table 1 Predictive performance of models for hypoglycemia in hospitalized T2DM patients

	Accuracy	Sensitivity	Specificity	PPV	NPV	F1 score
XGBoost-Train	0.853	0.975	0.352	0.861	0.771	0.914
XGBoost-Test	0.910	0.985	0.526	0.915	0.870	0.948
RF-Train	0.844	0.957	0.427	0.863	0.862	0.903
RF-Test	0.862	0.924	0.489	0.885	0.858	0.917
Logistic-Train	0.833	0.967	0.197	0.842	0.667	0.905
Logistic-Test	0.824	0.962	0.278	0.840	0.650	0.897

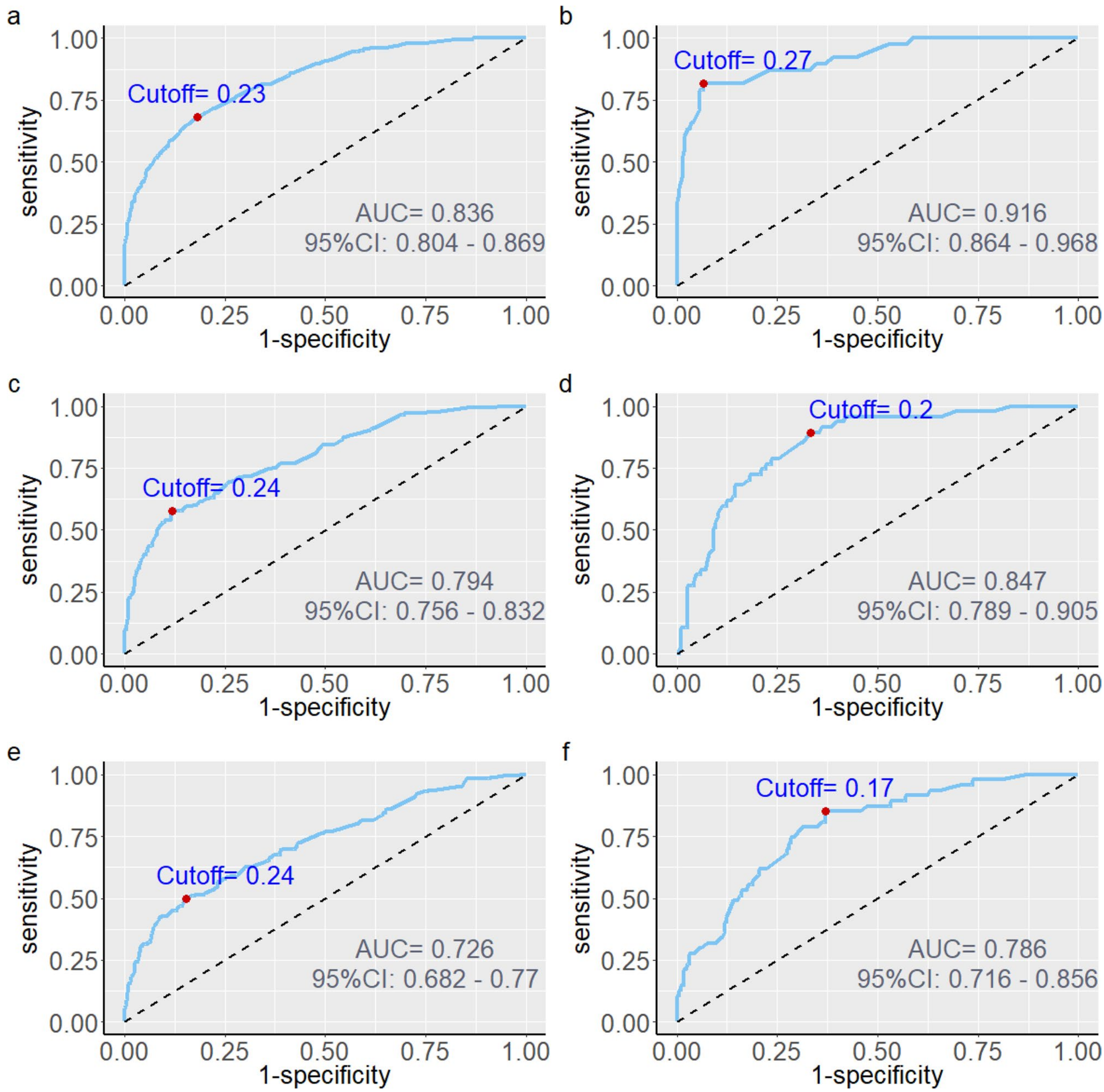

Fig. 3 Hypoglycemia Risk Prediction Model for Hospitalized Patients with Type 2 Diabetes

High hypoglycemia incidence in hospitalized T2DM patients

In this study, we found that the incidence of hypoglycemia among hospitalized patients with T2DM was as high as 18.85%, which was similar to the previous research results (the incidence of hypoglycemia among hospitalized T2DM patients was 18.79%) [24]. This phenomenon may be related to multiple factors. Firstly, older patients and those with a longer disease duration are more likely to experience hypoglycemia (Table 2), which may be because as patients age and the disease progresses, their pancreatic function gradually declines, their insulin sensitivity increases, and their autonomic nerve function is impaired, resulting in less obvious hypoglycemic symptoms and being more easily overlooked [25]. Secondly, patients with lower body weight are more likely to experience hypoglycemia, which may be related to malnutrition and weakened insulin resistance, leading to enhanced insulin action [26]. Additionally, patients with diabetic nephropathy (especially stage 4–5) and diabetic autonomic neuropathy have a significantly increased risk of hypoglycemia, as renal dysfunction affects insulin metabolism and clearance, prolonging the action time of insulin in the body [25]. The treatment plan is also an important

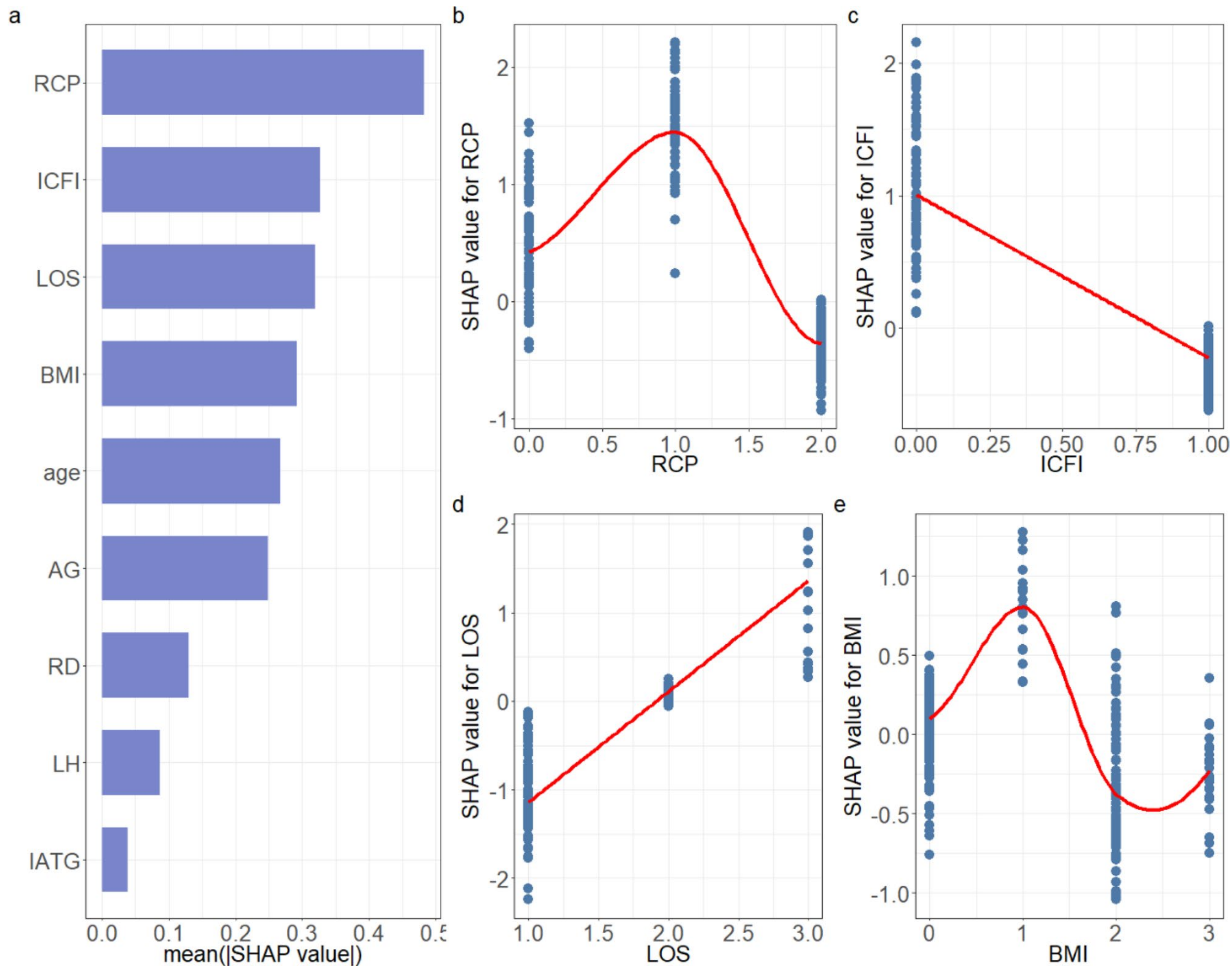

Fig. 4 Contribution of Features to Prediction Results in the In-Hospital Hypoglycemia Risk Model for Patients with Type 2 Diabetes

factor influencing the incidence of hypoglycemia, especially for patients using insulin or sulfonylurea drugs (such as glibenclamide), who have a higher incidence of hypoglycemia [27]. Moreover, patients with lower self-management ability and knowledge level of diabetes are also important factors for hypoglycemia occurrence, lacking awareness and ability to respond to hypoglycemic symptoms, which can prevent patients from taking timely measures when their blood sugar drops. During hospitalization, the treatment plan, dietary control, and examination arrangements of patients may change, all of which can increase the risk of hypoglycemia. For example, treatment adjustments in the early stage of hospitalization, dietary restrictions, and frequent examinations may lead to irregular eating patterns in patients, thereby triggering hypoglycemia [27].

Risk factors for hypoglycemia in hospitalized patients with T2DM

This study identified nine key predictors of hypoglycemia, which can be understood through several clinical themes: endogenous insulin capacity, nutritional and metabolic reserve, treatment-related factors, and clinical complexity. The most influential predictor was random C-peptide. A low level (< 0.33 ng/ml), indicating β-cell failure and consequent reliance on exogenous insulin, was a significant independent risk factor (OR = 2.661, 95% CI 1.550–4.568), a finding consistent with established pathophysiology [28]. Nutritional status, reflected by BMI and albumin, was also critical. Lower values were independently associated with heightened hypoglycemia risk. A BMI in the 25–29.9 kg/m² range was protective (OR = 0.578), suggesting that greater metabolic reserve contributes to glycemic stability, whereas low BMI and hypoalbuminemia may signal malnutrition, increasing vulnerability [27–30]. Treatment-related factors revealed important nuances. The use of insulin-containing fluid

Table 2 Table of hypoglycemia in hospitalized patients with type 2 diabetes

Project	Number of cases(1167)	Group of without hypoglycemia(947)	Group with hypoglycemia(220)	χ^2	*P*
Gender				3.34	0.07
Male	767	634	133		
Female	400	313	87		
Age(years)				6.29	0.04
13–39	95	69	26		
40–59	533	429	104		
60–91	539	449	90		
BMI				28.85	< 0.001
< 18.5	21	11	10		
18.5–24.9	688	535	153		
25-29.9	397	347	50		
> 30	61	54	7		
Treatment plan				79.28	< 0.001
Basal insulin	476	413	63		
Rapid-acting insulin	402	274	128		
Oral medication(excluding secretagogues)	271	249	22		
Oral medication(containing secretagogues)	18	11	7		
Intravenous insulin solution				38.66	< 0.001
Yes	1006	845	161		
No	161	102	59		
C-peptide(nmol/L)				96.62	< 0.001
< 0.33	109	55	54		
0.33–1.33	185	133	52		
> 1.33	873	759	114		
Insulin antibody(%)				6.79	0.009
0–20	1111	909	202		
> 20	56	38	18		
Liver function abnormality				5.97	0.02
Yes	42	28	14		
No	1125	919	206		
Kinder function abnormality				14.11	< 0.001
Yes	279	205	74		
No	888	742	146		
Fat hyperplasia				45.60	< 0.001
Yes	19	4	15		
No	1148	943	205		
Albumin(g/L)				16.99	< 0.001
40–60	668	565	103		
35–40	430	335	95		
30–35	37	27	10		
< 30	32	20	12		
Triglycerides(umol/L)				8.15	0.004
0-1.7	797	628	168		
> 1.7	370	318	52		
High-density liportein cholesterol(mmol/L)				14.14	< 0.001
< 0.91	330	281	49		
0.91–2.06	821	658	163		
> 2.06	16	8	8		
Hemoglobin(g/L)				16.35	< 0.001
110–150、120–160	1050	864	186		
90–110、90–120	105	78	27		
60–90	12	5	7		

Table 2 (continued)

Project	Number of cases(1167)	Group of without hypoglycemia(947)	Group with hypoglycemia(220)	χ^2	*P*
Free triiodothyronine(pmol/L)				9.03	0.01
< 3.85	235	175	60		
3.85–6.3	907	750	157		
> 6.3	25	22	3		
Vitamin D				6.14	0.05
< 20	814	674	140		
20–30	284	223	61		
> 30	69	50	19		
Hospital stay days				13.67	0.001
< 7	195	173	22		
7–14	955	764	191		
> 7	17	10	7		

was a protective factor (OR = 0.306), which may reflect its application in more controlled clinical scenarios, despite known challenges with insulin adsorption in infusion systems [25, 27]. Conversely, lipohypertrophy was a profound risk factor (OR = 13.625), as it leads to erratic insulin absorption, and the presence of insulin antibodies likely alters insulin pharmacokinetics, both increasing hypoglycemia risk [31]. Clinical complexity was captured by age, renal function, and length of stay. Advanced age and renal impairment (OR = 1.583) are well-documented risk factors due to impaired counter-regulatory responses and reduced clearance of glucose-lowering agents [32–35]. The association between a longer LOS and increased hypoglycemia risk should be interpreted with caution. While it may reflect the cumulative impact of clinical complexity, treatment changes, and nutritional disruptions, a longer LOS also increases the opportunity for hypoglycemia detection. Therefore, its association likely represents a composite of true risk and surveillance bias [9, 22, 36]. The selection of predictors such as lipohypertrophy and insulin antibodies may reflect unique characteristics of our single-center cohort. While statistically significant here, their generalizability should be confirmed through external validation. This underscores that variable selection in data-driven models can be cohort-sensitive, highlighting the need to verify not only the model's performance but also the stability of this specific predictor set across diverse clinical settings.

Clinical implications, model strengths and limitations

The primary clinical value of our model lies in its ability to proactively identify high-risk patients. This is particularly relevant given our finding that the majority of hypoglycemic events (87.3%) were mild; while often overlooked, these episodes are critical indicators of underlying glycemic instability. Furthermore, the pronounced peak in hypoglycemia during the postprandial period (46.4%) suggests that factors such as inappropriate prandial insulin dosing or delayed meal intake are key drivers in our cohort. This temporal pattern underscores the necessity for clinicians to extend their focus beyond fasting glucose control to include meticulous review of mealtime insulin regimens and patient education.

The strong predictive performance of the XGBoost model can be attributed to its inherent algorithmic strengths, which are particularly suited to capturing the complex, non-linear relationships among clinical variables such as age, BMI, and C-peptide levels-interactions that traditional Logistic Regression may overlook. As an ensemble method, XGBoost enhances accuracy and resists overfitting, making it effective for heterogeneous clinical datasets. Moreover, the application of the SHAP framework effectively addresses the "black-box" concern often associated with machine learning models by providing clear, quantitative insights into feature contributions, thereby enhancing interpretability and fostering clinical trust.

However, this study has several limitations that should be considered. First, its retrospective, single-center design may introduce selection bias and limit the generalizability of the findings. Second, the model relies solely on structured electronic health data and does not incorporate dynamic or behavioral variables-such as continuous glucose monitoring (CGM) trends, dietary intake, or physical activity-which are known to influence hypoglycemia risk. Additionally, the inclusion of the total LOS as a predictor constrains the model's utility for early risk assessment at admission, as this information is not available prospectively. Future research should focus on external validation across diverse clinical settings and aim to integrate real-time, patient-level data-such as "days since admission" for dynamic risk updates-to enhance predictive accuracy and support personalized intervention strategies.

Conclusion

This study establishes an effective prediction framework for hypoglycemia in hospitalized T2DM patients using eight key clinical variables. The XGBoost model demonstrated excellent performance in integrating these diverse features, providing a reliable tool for early identification of high-risk individuals. The model's primary clinical value lies in enabling proactive risk stratification, particularly through its ability to detect mild hypoglycemic events (87.3% of cases) that often go unrecognized yet indicate underlying glycemic instability. The pronounced postprandial pattern (46.4% of events) further highlights the need to focus on modifiable factors like inappropriate prandial insulin dosing. These findings support a shift from reactive correction to proactive prevention in inpatient glucose management. For clinical implementation, we recommend a stratified approach: intensifying monitoring for patients with low C-peptide or renal impairment, conducting regular injection site assessments for those with lipohypertrophy, and relaxing glycemic targets for elderly or malnourished patients. Increasing monitoring frequency after the third hospitalization day and implementing CGM for high-risk subgroups represent practical applications of our findings. Study limitations include the retrospective single-center design and reliance on structured data, which may affect generalizability and omit behavioral factors. Future research should focus on external validation across diverse populations and incorporate real-time monitoring data to enhance predictive capability and support personalized interventions.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12902-025-02104-x.

Supplementary Material 1

Acknowledgements
This study is supported by Research Project of the "Clinical Medicine + X" Research Center of the Air Force Military Medical University (No.: LHJJ24HL08).

Author contributions
Concept and design: Qi Rui, Huang Zhao-xu and Hua Yan, data collection and analysis: Liu Cai-xia, Ge Yang-yuan, Yuan Jie, Wang Chun-li, Zhang Jin-juan, Wang Xiao-li, dragting of the article: Liu Cai-xia, and Liu Tuo-nan, critacle revision of the article for important intellectual content: Lin yue, study supervision: Yuan Jie. All the authors approved the final article.

Funding
The "Clinical Medicine + X" Research Center of the Air Force Military Medical University (No.: LHJJ24HL08).

Data availability
All data support the findings of this study are included in this manuscript and its supplementary information files.

Declarations

Ethical approval and consent to participate
The Medical Ethics Committee of the First Affiliated Hospital of Air Force Medical University approved the study protocol (KY20242029-C-1). All eligible participants signed an informed consent form, which stated that the study was conducted entirely voluntarily, anonymously, and in confidence, and that they had the freedom to decline or withdraw from the study at any time. All methods were carried out in accordance with relevant guidelines and regulations. In addition, for participants under the age of 16, we have provided comprehensive information regarding the study to their parents or legal guardians and obtained written informed consent from them. Our research also adhered to the relevant provisions of the Declaration of Helsinki.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Endocrinology, Xijing Hospital, The Fourth Military Medical University, Shaanxi Xi'an 710032, China
[2]Department of Anesthesiology, Xijing Hospital, The Fourth Military Medical University, Shaanxi Xi'an 710032, China
[3]Nursing Department, The Fourth Military Medical University, Shaanxi Xi'an 710032, China

Received: 21 July 2025 / Accepted: 10 November 2025
Published online: 22 November 2025

References
1. Tinajero MG, Malik VS. An update on the epidemiology of type 2 diabetes: A global perspective. Endocrinol Metab Clin North Am. 2021;50(3):337–55.
2. Yun JS, Ko SH. Current trends in epidemiology of cardiovascular disease and cardiovascular risk management in type 2 diabetes. Metabolism. 2021;123:154838.
3. Ramírez-Alarcón K, Victoriano M, Mardones L, et al. Phytochemicals as potential epidrugs in type 2 diabetes mellitus. Front Endocrinol (Lausanne). 2021;12:656978.
4. Lowe RN, Williams B, Claus LW. Diabetes: how to manage patients experiencing hypoglycaemia. Drugs Context. 2022;11:2021–9.
5. Sun Q, Tielin Z, Shuanglian N, Dongmei W. Practical discussion on the construction of the medical quality management information platform in our hospital. China Hosp Manage. 2022;42(08):67–9.
6. Zhang Hongmei Z, Liting Z, Suqin Z. Wenjing. Progress in the Application of Real-World Research in Nursing Management. Nursing Research. 2022. 36(14): 2543–2545. [7] Amiel SA. The consequences of hypoglycaemia. Diabetologia. 2021. 64(5): 963–970.
7. Amiel SA. The consequences of hypoglycaemia. Diabetologia. 2021;64(5):963–970.
8. Freeland B. Hypoglycemia in diabetes mellitus. Home Healthc now. 2017;35(8):414–9.
9. Silbert R, Salcido-Montenegro A, Rodriguez-Gutierrez R, Katabi A, McCoy RG. Hypoglycemia among patients with type 2 diabetes: Epidemiology, risk Factors, and prevention strategies. Curr Diab Rep. 2018;18(8):53.
10. Ruan Y, Tan GD, Lumb A, Rea RD. Importance of inpatient hypoglycaemia: impact, prediction and prevention. Diabet Med. 2019;36(4):434–43.
11. Rodriguez-Gutierrez R, Salcido-Montenegro A, Singh-Ospina NM, et al. Documentation of hypoglycemia assessment among adults with diabetes during clinical encounters in primary care and endocrinology practices. Endocrine. 2020;67(3):552–60.
12. Garg A, Mago V. Role of machine learning in medical research: a survey - ScienceDirect. Comput Sci Rev. 2021;40.
13. Mora T, Roche D, Rodríguez-Sánchez B. Predicting the onset of diabetes-related complications after a diabetes diagnosis with machine learning algorithms. Diabetes Res Clin Pract. 2023;204:110910.
14. Zou LX, Wang X, Hou ZL, Sun L, Lu JT. Machine learning algorithms for diabetic kidney disease risk predictive model of Chinese patients with type 2 diabetes mellitus. Ren Fail. 2025;47(1):2486558.

15. Bhaskaranand M, Ramachandra C, Bhat S, et al. The value of automated diabetic retinopathy screening with the eyeart system: A study of more than 100,000 consecutive encounters from people with diabetes. Diabetes Technol Ther. 2019;21(11):635–43.
16. Casanova R, Saldana S, Simpson SL, et al. Prediction of incident diabetes in the Jackson heart study using High-Dimensional machine learning. PLoS ONE. 2016;11(10):e0163942.
17. Hou Jinjie X, Rong X, Xiaran Y, Jiaxin, Xia Tian. Construction and validation of a prediction model for hypoglycemia risk in elderly patients with type 2 diabetes mellitus. J Nurs. 2024;39(24):31–6.
18. Zuo Dan Z, Xili DX. Construction and validation of a prediction model for hypoglycemia risk in patients with type 2 diabetes. J Nurs. 2021;36(01):30–3.
19. Seo W, Lee YB, Lee S, Jin SM, Park SM. A machine-learning approach to predict postprandial hypoglycemia. BMC Med Inf Decis Mak. 2019;19(1):210.
20. Bergford S, Riddell MC, Jacobs PG, et al. The type 1 diabetes and exercise initiative: predicting hypoglycemia risk during exercise for participants with type 1 diabetes using repeated measures random forest. Diabetes Technol Ther. 2023;25(9):602–11.
21. Xu XY, Wu HY, Wei Q. Obesity-related indices as predictors of lower extremity arterial disease in type 2 diabetes mellitus. Endocrine. 2025;87(2):554–61.
22. Jia H, Zhang J. Machine learning-based prediction of hypoglycemia severity in hospitalized diabetic patients. Front Endocrinol (Lausanne). 2025;16:1634358.
23. Shah BR, Walji S, Kiss A, James JE, Lowe JM. Derivation and validation of a Risk-Prediction tool for hypoglycemia in hospitalized adults with diabetes: the hypoglycemia during hospitalization (HyDHo) score. Can J Diabetes. 2019;43(4):278–e2821.
24. Zheng Caixia. Construction of a low blood sugar prediction model for hospitalized type 2 diabetic patients based on machine learning algorithms. J Nurs. 2024.
25. Li M, Zhao M, Yan H, Guo H, Shi B. Clinical characteristics and influencing factors of hypoglycemia in hospitalized patients with type 2 diabetes mellitus: A cross-sectional study. Nurs Open. 2023;10(10):6827–35.
26. Guzmán G, Martínez V, Yara JD et al. Glycemic control and hypoglycemia in patients treated with insulin pump therapy: an observational study. J Diabetes Res. 2020. 2020: 1581726.
27. Yunir E, Nugraha A, Rosana M, et al. Risk factors of severe hypoglycemia among patients with type 2 diabetes mellitus in outpatient clinic of tertiary hospital in Indonesia. Sci Rep. 2023;13(1):16259.
28. González-Vidal T, Rivas-Otero D, Gutiérrez-Hurtado A, et al. Hypoglycemia in patients with type 2 diabetes mellitus during hospitalization: associated factors and prognostic value. Diabetol Metab Syndr. 2023;15(1):249.
29. Yun JS, Park YM, Han K, Cha SA, Ahn YB, Ko SH. Association between BMI and risk of severe hypoglycaemia in type 2 diabetes. Diabetes Metab. 2019;45(1):19–25.
30. Agrawal S, Makuch S, Dróżdż M, et al. The impact of hypoglycemia on patients with diabetes mellitus: A Cross-Sectional analysis. J Clin Med. 2022;11(3):626.
31. Gentile S, Satta E, Strollo F, et al. Insulin-induced skin lipohypertrophies: A neglected cause of hypoglycemia in dialysed individuals with diabetes. Diabetes Metab Syndr. 2021;15(4):102145.
32. Pratiwi C, Mokoagow MI, Made Kshanti IA, Soewondo P. The risk factors of inpatient hypoglycemia: A systematic review. Heliyon. 2020;6(5):e03913.
33. Akirov A, Amitai O, Masri-Iraqi H, et al. Predictors of hypoglycemia in hospitalized patients with diabetes mellitus. Intern Emerg Med. 2018;13(3):343–50.
34. Sircar M, Bhatia A, Munshi M. Review of hypoglycemia in the older adult: clinical implications and management. Can J Diabetes. 2016;40(1):66–72.
35. Bellary S, Kyrou I, Brown JE, Bailey CJ. Type 2 diabetes mellitus in older adults: clinical considerations and management. Nat Rev Endocrinol. 2021;17(9):534–48.
36. Criner KE, Kim HN, Ali H, et al. Hypoglycemia symptoms are reduced in hospitalized patients with diabetes. J Diabetes Complications. 2021;35(10):107976.

Publisher's note