

Original Paper

Machine Learning Models for Blood Glucose Level Prediction in Patients With Diabetes Mellitus: Systematic Review and Network Meta-Analysis

Kui Liu^{1*}, MMS; Linyi Li^{1*}, MMS; Yifei Ma¹, MMS; Jun Jiang¹, MD; Zhenhua Liu¹, MD; Zichen Ye¹, MMS; Shuang Liu¹, MBA; Chen Pu¹, MMS; Changsheng Chen², MD; Yi Wan¹, MD

¹Department of Health Service, Air Force Medical University, Xi'an, Shaanxi, China

²Department of Health Statistics, Air Force Medical University, Xi'an, Shaanxi, China

*these authors contributed equally

Corresponding Author:

Yi Wan, MD

Department of Health Service

Air Force Medical University

No 169, Changle West Road, Xincheng District

Xi'an, Shaanxi, 710032

China

Phone: 86 17391928966

Fax: 86 29 8471267

Email: wanyi@fmmu.edu.cn

Abstract

Background: Machine learning (ML) models provide more choices to patients with diabetes mellitus (DM) to more properly manage blood glucose (BG) levels. However, because of numerous types of ML algorithms, choosing an appropriate model is vitally important.

Objective: In a systematic review and network meta-analysis, this study aimed to comprehensively assess the performance of ML models in predicting BG levels. In addition, we assessed ML models used to detect and predict adverse BG (hypoglycemia) events by calculating pooled estimates of sensitivity and specificity.

Methods: PubMed, Embase, Web of Science, and Institute of Electrical and Electronics Engineers Explore databases were systematically searched for studies on predicting BG levels and predicting or detecting adverse BG events using ML models, from inception to November 2022. Studies that assessed the performance of different ML models in predicting or detecting BG levels or adverse BG events of patients with DM were included. Studies with no derivation or performance metrics of ML models were excluded. The Quality Assessment of Diagnostic Accuracy Studies tool was applied to assess the quality of included studies. Primary outcomes were the relative ranking of ML models for predicting BG levels in different prediction horizons (PHs) and pooled estimates of the sensitivity and specificity of ML models in detecting or predicting adverse BG events.

Results: In total, 46 eligible studies were included for meta-analysis. Regarding ML models for predicting BG levels, the means of the absolute root mean square error (RMSE) in a PH of 15, 30, 45, and 60 minutes were 18.88 (SD 19.71), 21.40 (SD 12.56), 21.27 (SD 5.17), and 30.01 (SD 7.23) mg/dL, respectively. The neural network model (NNM) showed the highest relative performance in different PHs. Furthermore, the pooled estimates of the positive likelihood ratio and the negative likelihood ratio of ML models were 8.3 (95% CI 5.7-12.0) and 0.31 (95% CI 0.22-0.44), respectively, for predicting hypoglycemia and 2.4 (95% CI 1.6-3.7) and 0.37 (95% CI 0.29-0.46), respectively, for detecting hypoglycemia.

Conclusions: Statistically significant high heterogeneity was detected in all subgroups, with different sources of heterogeneity. For predicting precise BG levels, the RMSE increases with a rise in the PH, and the NNM shows the highest relative performance among all the ML models. Meanwhile, current ML models have sufficient ability to predict adverse BG events, while their ability to detect adverse BG events needs to be enhanced.

Trial Registration: PROSPERO CRD42022375250; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=375250

(*JMIR Med Inform* 2023;11:e47833) doi: [10.2196/47833](https://doi.org/10.2196/47833)

KEYWORDS

machine learning; diabetes; hypoglycemia; blood glucose; blood glucose management

Introduction

Diabetes mellitus (DM) has become one of the most serious health problems worldwide [1], with more than 463 million (9.3%) patients in 2019; this number is predicted to reach 700 million (10.9%) in 2045 [2], which has resulted in growing concerns about the negative impacts on patients' lives and the increasing burden on the health care system [3]. Furthermore, previous studies have shown that without appropriate medical care, DM can lead to multiple long-term complications in blood vessels, eyes, kidneys, feet (ulcers), and nerves [4-7]. Adverse blood glucose (BG) events are one of the most common short-term complications, including hypoglycemia with BG < 70 mg/dL and hyperglycemia with BG > 180 mg/dL. Hyperglycemia in patients with DM may lead to lower limb occlusions and extremity nerve damage, further leading to decay, necrosis, and local or whole-foot gangrene, even requiring amputation [8,9]. Hypoglycemia can cause serious symptoms, including anxiety, palpitation, and confusion in a mild scenario and seizures, coma, and even death in a severe scenario [10,11]. Thus, there is an imminent need for preventing adverse BG events.

Machine learning (ML) models use statistical techniques to provide computers with the ability to complete assignments by training themselves without being explicitly programmed [12]. However, ML models for managing BG requires huge amounts of BG data, which cannot be satisfied by the multiple data points generated by the traditional finger-stick glucose meter [13]. With the introduction of the continuous glucose monitoring (CGM) device, which typically produces a BG reading every 5 minutes all day long, the size of the data set of BG readings is sufficient to be used in ML models [14].

Recently, there has been an immense surge in using ML technologies for predicting DM complications. Regarding BG management, previous studies have developed different types of ML models, including random forest (RF) models, support vector machines (SVMs), neural network models (NNMs), and autoregression models (ARMs), using CGM data, electronic health records (EHRs), electrocardiograph (ECG), electroencephalograph (EEG), and other information (ie, biochemical indicators, insulin intake, exercise, and meals) [10,15-20]. However, the performance of different models in these studies was not quite consistent. For instance, in terms of BG level prediction, Prendin et al [21] showed that the SVM achieved a lower root mean square error (RMSE) than the ARM, while Zhu et al [22] showed a different result.

Therefore, this meta-analysis aimed to comprehensively assess the performance of ML models in BG management in patients with DM.

Methods

Search Strategy and Study Selection

The study protocol has been registered in the international prospective register of systematic reviews (PROSPERO;

registration ID: CRD42022375250). Studies on BG levels or adverse BG event prediction or detection using ML models were eligible, with no restrictions on language, investigation design, or publication status. PubMed, Embase, Web of Science, and Institute of Electrical and Electronics Engineers (IEEE) Explore databases were systematically searched from inception to November 2022. Keywords used for study repository searches were (“machine learning” OR “artificial intelligence” OR “logistic model” OR “support vector machine” OR “decision tree” OR “cluster analysis” OR “deep learning” OR “random forest”) AND (“hypoglycemia” OR “hyperglycemia” OR “adverse glycaemic events”) AND (“prediction” OR “detection”). Details regarding the search strategies are summarized in [Multimedia Appendix 1](#). Manual searches were added to review reference lists in relevant studies.

Selection Criteria

Inclusion criteria were as follows: (1) participants in the studies were diagnosed with DM; (2) study endpoints were hypoglycemia, hyperglycemia, or BG levels; (3) the studies established at least 2 or more types of ML models for prediction of BG levels and 1 or more types of ML models for prediction or detection of adverse BG events; (4) the studies reported the performance of ML models with statistical or clinical metrics; (5) the studies contained the development and validation of ML models; and (6) study outcomes were means (SDs) of performance metrics of test data for prediction of BG levels and sensitivity and specificity of test data for prediction or detection of adverse BG events.

Exclusion criteria were as follows: (1) studies did not report on the derivation of ML models, (2) studies were based only on physiological or control-oriented ML models, (3) studies could not reproduce true positives, true positives, false negatives, and false positives for prediction or detection of adverse BG events, (4) studies were reviews, systematic reviews, animal studies, or irretrievable and repetitive papers, and (5) studies had unavailable full text or outcome metrics.

Authors KL and LYL screened and selected studies independently based on the criteria mentioned before. Authors KL and YM extracted and recorded the data from the selected studies. Conflicts were resolved by reaching a consensus. The study strictly followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) statement ([Multimedia Appendix 2](#)) [23-25].

Data Extraction and Management

Two reviewers independently carried out data extraction and quality assessment. If a single study included more than 1 extractable test results for the same ML model, the best result was extracted. If a single study included 2 or more models, the performance metrics of each model were extracted. For studies predicting BG levels, RMSEs based on different prediction horizons (PHs) were extracted. For studies predicting or detecting adverse BG events, the sensitivity, specificity, and

precision of reproducing the 2×2 contingency table were extracted.

Specifically, the following information was extracted:

- General characteristics: first author, publication year, country, data source, and study purpose (ie, predicting or detecting hypoglycemia)
- Experimental information: participants (type of DM, type 1 or 2), sample size (patients, data points, and hypoglycemia), demographic information, models, study place and time, model parameters (ie, input and PHs), model performance metrics, threshold of BG levels for hypoglycemia, and reference (ie, finger-stick)

Methodological Quality Assessment of Included Reviews

The Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool was applied to assess the quality of included studies based on patient selection (5 items), index test (3 items), reference standard (4 items), and flow and timing (4 items). All 4 domains were used for assessing the risk of bias, and the first 3 domains were used to assess the consensus of applicability. Each domain has 1 query in relation to the risk of bias or applicability consisting of 7 questions [26].

Data Synthesis and Statistical Analysis

The performance metrics of ML models used to predict BG levels, predict adverse BG events, and detect adverse BG events were assessed independently. The performance metrics were the RMSE of ML models in predicting BG levels and the sensitivity and specificity of ML models in predicting or detecting adverse BG events. A network meta-analysis was conducted for BG level-based studies to assess the global and local inconsistency between studies and plotted the surface under the cumulative ranking (SUCRA) curve of every model to calculate relative ranks. For event-based studies, pooled sensitivity, specificity, the positive likelihood ratio (PLR), and

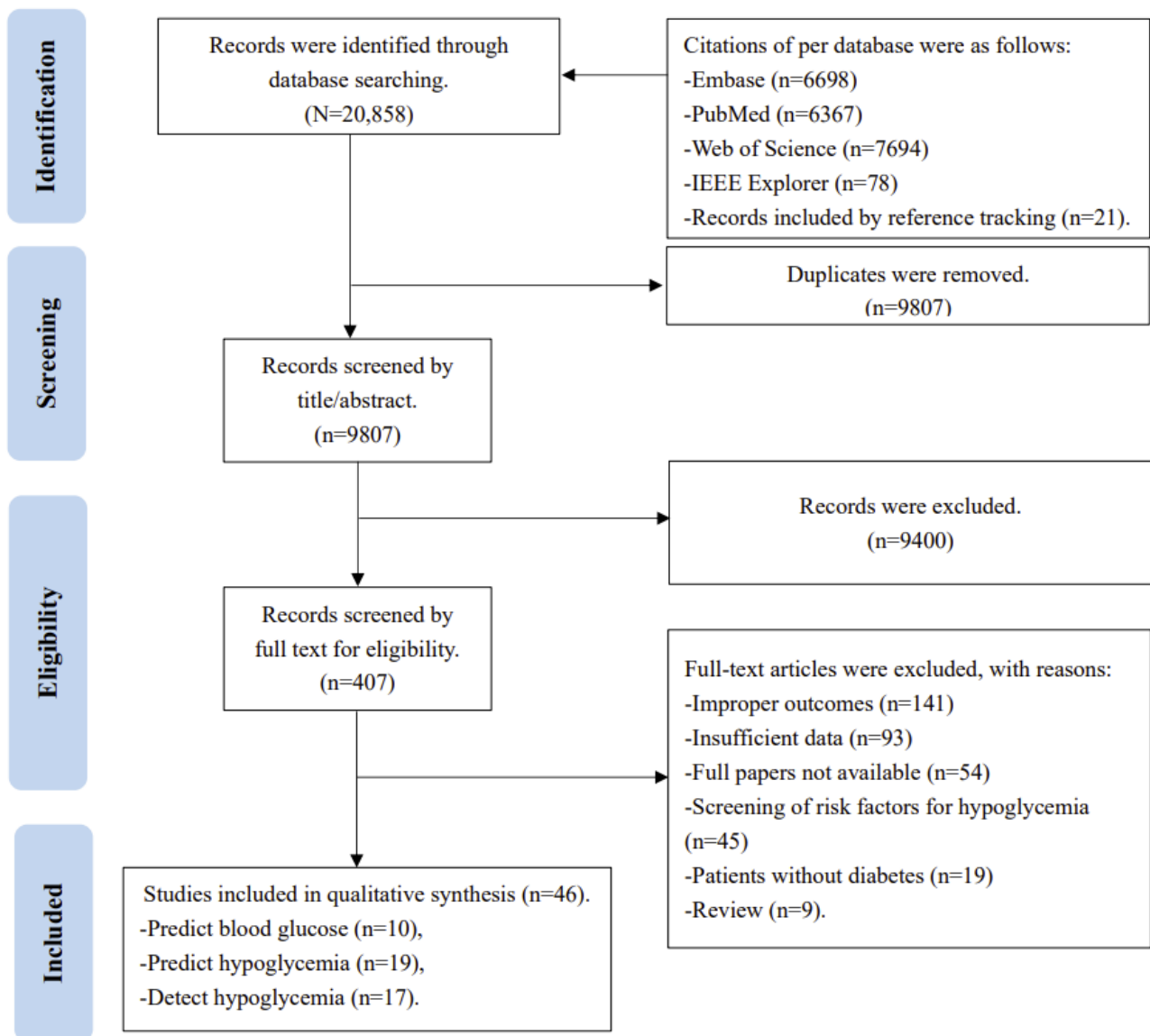
the negative likelihood ratio (NLR) with 95% CIs were calculated. Study heterogeneity was assessed by calculating I^2 values based on multivariate random-effects meta-regression that considered within- and between-study correlation and classifying them into quartiles (0% to <25% for low, 25% to <50% for low-to-moderate, 50% to <75% for moderate-to-high, and >75% for high heterogeneity) [27,28]. Furthermore, meta-regression was used to evaluate the source of heterogeneity for both BG level-based and adverse event-based studies. The summary receiver operating characteristic (SROC) curve of every model was also used to evaluate the overall sensitivity and specificity. Publication bias was assessed using the Deek funnel plot asymmetry test.

Furthermore, BG level-based studies were divided into 4 subgroups based on different PHs (15, 30, 45, 60 minutes), and adverse event-based studies were analyzed using different types of models (ie, NNM, RF, and SVM). A 2-sided P value of <.05 was considered statistically significant. All statistical analyses were performed using Stata 17 (Stata Corp) and Review Manager (RevMan; Cochrane) version 5.3.

Results

Search Results

A total of 20,837 studies were identified through systematically searching the predefined electronic databases; these also included 21 studies found using reference tracking [10,29-48]. Of the 20,837 studies, 9807 (47.06%) were retained after removing duplicates. After screening titles and abstracts, 9400 (95.85%) studies were excluded owing to reporting irrelevant topics or no predefined outcomes. The remaining 407 (4.15%) studies were retrieved for full-text evaluation. Of these, 361 (88.7%) studies were excluded for various reasons, and therefore 46 (11.3%) studies were included in the final meta-analysis (Figure 1).

Figure 1. Flow diagram of identifying and including studies. IEEE: Institute of Electrical and Electronics Engineers.

Description of Included Studies

As studies on hyperglycemia were insufficient for analysis, we selected studies on hypoglycemia to assess the ability of ML models to predict adverse BG events. In total, the 46 studies included 28,775 participants: n=428 (1.49%) for predicting

BG levels, n=28,138 (97.79%) for predicting adverse BG events, and n=209 (0.72%) for detecting adverse BG events. Of the 46 studies, 10 (21.7%) [20-22,49-55] predicted BG levels (Table 1), 19 (41.3%) [15,29-39,47,48,56-60] predicted adverse BG events (Table 2), and the remaining 17 (37%) [10,16,40-46,61-68] detected adverse BG events (Table 3).

Table 1. Baseline characteristics of BG^a level-based studies (N=10).

First author (year), country	Data source	Sample size		Demographic information	Object; setting	Model; PH ^b (minutes); input	Performance metrics
		Patients, n	Data points, n				
Pérez-Gandía (2010), Spain [20]	CGM ^c device	15	728	— ^d	T1DM ^e ; out	Models: NNM ^f , ARM ^g PH: 15, 30 Input: CGM data	RMSE ^h , delay
Prendin (2021) United States [21]	CGM device	Real (n=141)	350,000	Age	T1DM; out	ARM, autoregressive moving average (ARMA), autoregressive integrated moving average (ARIMA), SVM ⁱ , RF ^j feed-forward neural network (fNN), long short-term memory (LSTM) PH: 30 Input: CGM data	RMSE, coefficient of determination (COD) sensibility, delay, precision F_1 score, time gain
Zhu (2020) England [22]	Ohio T1DM, UVA/Pado-va T1D	Real (n=6), simulated (n=10)	1,036,800	—	T1DM; out	DRNN ^k , NNM, SVM, ARM PH:30 Input: BG level, meals, exercise, meal times	RMSE, mean absolute relative difference (MARD) time gain
D'Antoni (2020), Italy [49]	Ohio T1DM	6	—	Age, sex ratio	T1DM; out	ARJNN ^l , RF, SVM, autoregression (AR), one symbolic model (SAX), recurrent neural network (RNN), one neural network model (NARX), jump neural network (JNN), delayed feed-forward neural network model (DFFNN) PH: 15, 30 Input: CGM data	RMSE
Amar (2020), Israel [50]	CGM device, insulin pump	141	1,592,506	Age, sex ratio, weight, BMI, duration of DM	T1DM; in	ARM, gradually connected neural network (GCN), fully connected (FC [neural network]), light gradient boosting machine (LCBM), RF PH: 30, 60 Input: CGM data	RMSE, Clarke error grid (CEG)
Li (2020), England [51]	UVA/Pado-va T1D	Simulated (n=10)	51,840	—	T1DM; out	GluNet, NNM, SVM, latent variable with exogenous input (LVX), ARM PH: 30, 60 Input: BG level, meals, exercise	RMSE, MARD, time lag
Zecchin (2012), Italy [52]	UVA/Pado-va T1D, CGM device	Simulated (n=20), real (n=15)	—	—	T1DM; out	Neural network–linear prediction algorithm (NN-LPA), NN, ARM PH: 30 Input: meals, insulin	RMSE, energy of second-order differences (ESOD), time gain, J index
Mohebbi (2020), Denmark [53]	Cornerstones4Care platform	Real (n=50)	—	—	T1DM; in	LSTM, ARIMA PH: 15, 30, 45, 60, 90	RMSE, MAE
Daniels (2022), England [54]	CGM device	Real (n=12)	—	Sex ratio	T1DM; out	Convolutional recurrent neural network (CRNN), SVM PH: 30, 45, 60, 90, 120 Input: BG level, insulin, meals, exercise	RMSE, MAE, CEG, time gain
Alfian (2020), Korea [55]	CGM device	Real (n=12)	26,723	—	—	SVM, k-nearest neighbor k-nearest neighbor (kNN), DT ^m , RF, AdaBoost, XGBoost ⁿ , NNM PH: 15, 30 Input: CGM data	RMSE, glucose-specific root mean square error (gRMSE), R2 score, mean absolute percentage error (MAPE)

^aBG: blood glucose.^bPH: prediction horizon.^cCGM: continuous glucose monitoring.^dNot applicable.^eT1DM: type 1 diabetes mellitus.^fNNM: neural network model.^gARM: autoregression model.

^hRMSE: root mean square error.

ⁱSVM: support vector machine.

^jRF: random forest.

^kDRNN: dilated recurrent neural network.

^lARJNN: ARTiDe jump neural network.

^mDT: decision tree.

ⁿXGBoost: Extreme Gradient Boosting.

Table 2. Baseline characteristics of studies predicting adverse BG^a events (N=19).

First author (year), country	Data source	Sample size			Object; setting	Model	Time	Age (years), mean (SD)/range	Threshold
		Patients, n	Data points, n	Hypoglycemia, n					
Pils (2014), United States [39]	CGM ^b device	2	2518	152	T1DM ^c ; out	SVM ^d	All	— ^e	3.9
Seo (2019), Korea [15]	CGM device	104	7052	412	DM ^f ; out	RF ^g , SVM, k-nearest neighbor (kNN), logistic regression (LR)	Postprandial	52	3.9
Parcerisas (2022), Spain [29]	CGM device	10	67	22	T1DM; out	SVM	Nocturnal	31.8 (SD 16.8)	3.9
Stuart (2017), Greece [30]	EHRs ^h	9584	—	1327	DM; in	Multivariable logistic regression (MLR)	All	—	4
Bertachi (2020), Spain [31]	CGM device	10	124	39	T1DM; out	SVM	Nocturnal	31.8 (SD 16.8)	3.9
Elhadd (2020), Qatar [32]	—	13	3918	172	T2DM; out	XGBoost ⁱ	All	35-63	—
Mosquera-Lopez (2020), United States [33]	CGM device	10	117	17	T1DM; out	SVM	Nocturnal	33.7 (SD 5.8)	3.9
Mosquera-Lopez (2020), United States [33]	CGM device	20	2706	258	T1DM; out	SVM	Nocturnal	—	3.9
Ruan (2020), England [34]	EHRs	17,658	3276	703	T1DM; in	XGBoost, LR, stochastic gradient descent (SGD), kNN, DT ^j , SVM, quadratic discriminant analysis (QDA), RF, extra tree (ET), linear discriminant analysis (LDA), Adaboost, bagging	All	66 (SD 18)	4
Güemes (2020), United States [35]	CGM device	6	55	6	T1DM; out	SVM	Nocturnal	40-60	3.9
Jensen (2020), Denmark [36]	CGM device	463	921	79	T1DM; out	LDA	Nocturnal	43 (SD 15)	3
Oviedo (2019), Spain [37]	CGM device	10	1447	420	T1DM; out	SVM	Postprandial	41 (SD 10)	3.9
Toffanin (2019), Italy [38]	CGM device	20	7096	36	T1DM; out	Individual model-based	All	46	3.9

First author (year), country	Data source	Sample size			Object; setting	Model	Time	Age (years), mean (SD)/range	Thresh- old
		Patients, n	Data points, n	Hypoglycemia, n					
Bertachi (2018), United States [47]	CGM device	6	51	6	T1DM; out	NNM ^k	Nocturnal	40-60	3.9
Eljil (2014), United Arab Emirates [48]	CGM device	10	667	100	T1DM; out	Bagging	All	25	3.3
Dave (2021), United States [56]	CGM device	112	546,640	12,572	T1DM; out	RF	All	12.67 (SD 4.84)	3.9
Marcus (2020), Israel [57]	CGM device	11	43,533	5264	T1DM; out	Kernel ridge regression (KRR)	All	18-39	3.9
Reddy (2019), United States [58]	—	55	90	29	T1DM; out	RF	—	33 (SD 6)	3.9
Sampath (2016), Aus- tralia [59]	—	34	150	40	T1DM; out	Ranking aggre- gation (RA)	Nocturnal	—	—
Sudharsan (2015), United States [60]	—	—	839	428	T2DM; out	RF	All	—	3.9

^aBG: blood glucose.

^bCGM: continuous glucose monitoring.

^cT1DM: type 1 diabetes mellitus.

^dSVM: support vector machine.

^eNot applicable.

^fDM: diabetes mellitus.

^gRF: random forest.

^hEHR: electronic health record.

ⁱXGBoost: Extreme Gradient Boosting.

^jDT: decision tree.

^kNNM: neural network model.

Table 3. Baseline characteristics of studies detecting adverse BG^a events (N=17).

First author (year), country	Data source	Sample size			Object; setting	Model	Time	Age (years), mean (SD)/range	Threshold
		Patients, n	Data points, n	Hypo-glycemia, n					
Jin (2019), United States [10]	EHRs ^b	— ^c	4104	132	T1DM ^d ; in	Linear discriminant analysis (LDA)	All	—	—
Nguyen (2013), Australia [16]	EEG ^e	5	144	76	T1DM; in	Levenberg-Marquardt (LM), genetic algorithm (GA)	All	12-18	3.3
Chan (2011), Australia [40]	CGM ^f device	16	100	52	T1DM; experimental	Feed-forward neural network (fNN)	Nocturnal	14.6 (SD 1.5)	3.3
Nguyen (2010), Australia [41]	EEG	6	79	27	T1DM; experimental	Block-based neural network (BRNN)	Nocturnal	12-18	3.3
Rubega (2020), Italy [42]	EEG	34	2516	1258	T1DM; experimental	NNM ^g	All	55 (SD 3)	3.9
Chen (2019), United States [43]	EEG	—	300	11	DM ^h ; in	Logistic regression (LR)	All	—	—
Jensen (2013), Denmark [44]	CGM device	10	1267	160	T1DM; experimental	SVM ⁱ	All	44 (SD 15)	3.9
Skladnev (2010), Australia [45]	CGM device	52	52	11	T1DM; in	fNN	Nocturnal	16.1 (SD 2.1)	3.9
Iaione (2005), Brazil [46]	EEG	8	1990	995	T1DM; experimental	NNM	Morning	35 (SD 13.5)	3.3
Nuryani (2012), Australia [61]	ECG	5	575	133	DM; in	SVM, linear multiple regression (LMR)	All	16 (SD 0.7)	3.0
San (2013), Australia [62]	ECG	15	440	39	T1DM; in	Block-based neural network (BBNN), wavelet neural network (WNN), fNN, SVM	All	14.6 (SD 1.5)	3.3
Ling (2012), Australia [63]	ECG	16	269	54	T1DM; in	Fuzzy reasoning model (FRM), fNN, multiple regression-fuzzy inference system (MR-FIS)	Nocturnal	14.6 (SD 1.5)	3.3

First author (year), country	Data source	Sample size			Object; setting	Model	Time	Age (years), mean (SD)/range	Threshold
		Patients, n	Data points, n	Hypo- glycemia, n					
Ling (2016), Australia [64]	ECG	16	269	54	T1DM; in	Extreme learning machine-based neural network (ELM-NN), particle swarm optimization-based neural network (PSO-NN), MR-FIS, LMR, fuzzy inference system (FIS)	Nocturnal	14.6 (SD 1.5)	3.3
Nguyen (2012), Australia [65]	EEG	5	44	20	T1DM; in	NNM	—	12-18	3.3
Ngo (2020), Aus- tralia [66]	EEG	8	135	53	T1DM; in	BRNN	Nocturnal	12-18	3.9
Ngo (2018), Aus- tralia [67]	EEG	8	54	26	T1DM; in	BRNN	Nocturnal	12-18	3.9
Nuryani (2010), Australia [68]	ECG	5	27	8	T1DM; experi- mental	Fuzzy support vector machine (FSVM), SVM	Nocturnal	16 (SD 0.7)	3.3

^aBG: blood glucose.

^bEHR: electronic health record.

^cNot applicable.

^dT1DM: type 1 diabetes mellitus.

^eEEG: electroencephalograph.

^fCGM: continuous glucose monitoring.

^gNNM: neural network model.

^hDM: diabetes mellitus.

ⁱSVM: support vector machine.

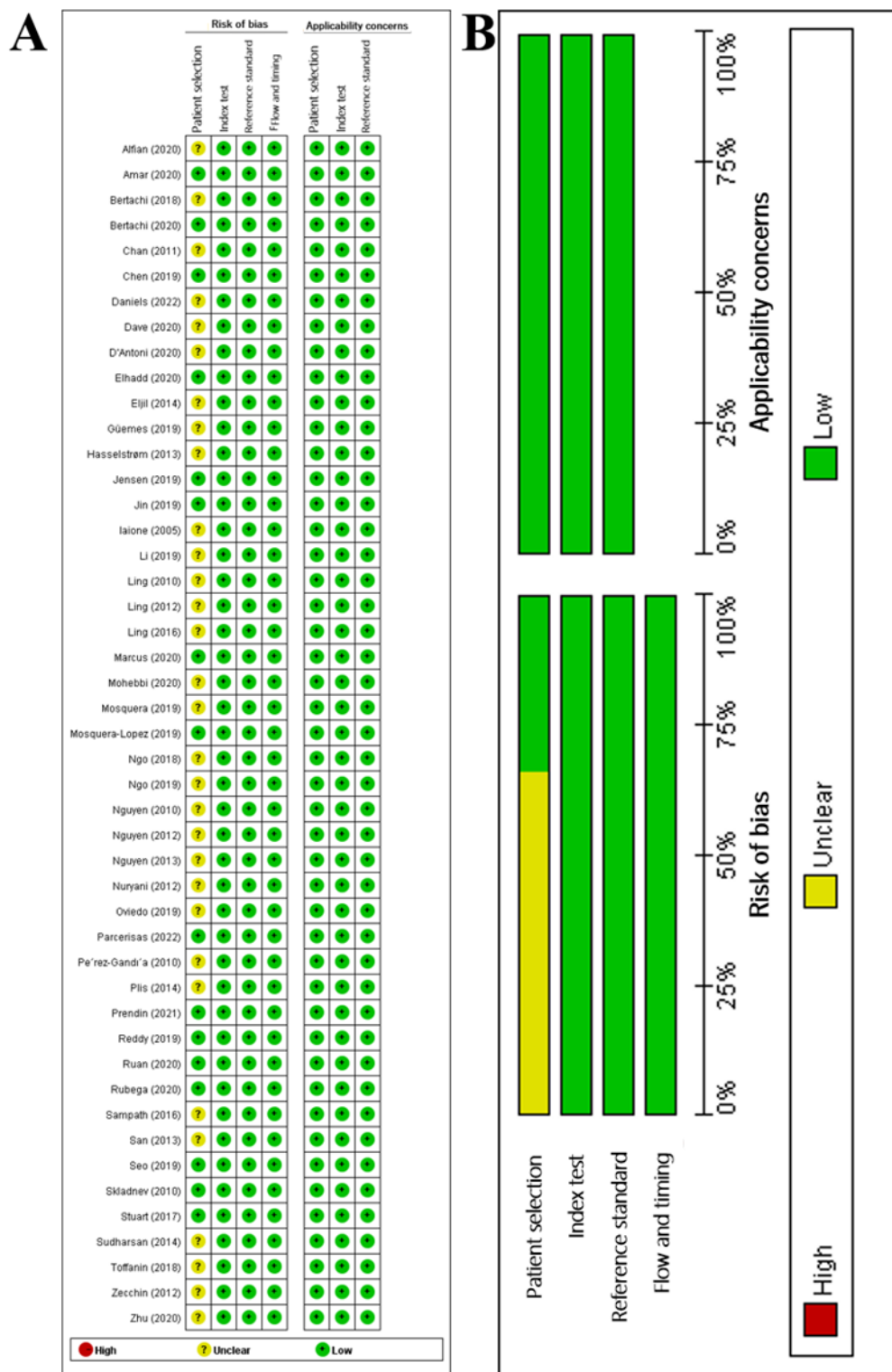
As shown in Tables 1-3, 40 (87%) studies [10,16,20-22,29,31,33-42,44-59,62-68] included participants with type 1 diabetes mellitus (T1DM), 2 (4.3%) studies [32,60] included participants with type 2 diabetes mellitus (T2DM), and the remaining 4 (8.7%) studies [15,30,43,61] did not specify the type of DM. Regarding the data source of ML models, CGM devices were involved in 22 (47.8%) studies [15,20,21,29,31,33,35-40,44,45,47,48,50,52,54-57], EEG signals were used in 8 (17.4%) studies [16,41-43,46,65-67], ECG signals were involved in 5 (10.9%) studies [61-64,68], EHRs were used in 3 (6.5%) studies [10,30,34], data generated by the UVA/Padova T1D simulator were used in 3 (6.5%) studies [22,51,52], the Ohio T1DM data set was used in 2 (4.3%) studies [22,49], and 4 (8.7%) studies [32,58-60] did not report the source of data. Regarding the setting of data collection, 24 (52.2%) studies [15,20-22,29,31-33,35-39,47-49,51,52,54,56-60] were conducted in an out-of-hospital setting, 13 (28.3%) studies [10,16,34,43,50,53,61-67] were conducted in an in-hospital setting, 6 (13%) studies [40-42,44,46,68] were conducted in an

experimental setting, and the remaining 1 (2.2%) study [55] did not specify the environment. Regarding when adverse BG events occurred in the 36 (78.3%) adverse event-based studies, 15 (41.7%) [29,31,33,35,36,40,41,45,47,59,63,64,66-68] reported nocturnal hypoglycemia, 16 (44.4%) [10,16,30,32,34,38,39,42-44,48,56,57,60-62] were not specific about the time of day, 2 (5.6%) [15,37] reported postprandial hypoglycemia, 1 (2.8%) [46] reported morning hypoglycemia, and the remaining 2 (5.6%) [58,65] did not report the time setting. To carry out the network meta-analysis of BG level-based studies, we chose the RMSE as the outcome to be compared.

Quality Assessment of Included Studies

The quality assessment results using the QUADAS-2 tool showed that more than half of all included studies did not report the patient selection criteria in detail, which led to low-quality patient selection (Figure 2). Furthermore, the diagnosis of hypoglycemia using blood or the CGM device was considered high quality in the reference test in our study.

Figure 2. Quality assessment of included studies. Risk of bias and applicability concerns graph (A) and risk of bias and applicability concerns summary (B).



Statistical Analysis

Machine Learning Models for Predicting Blood Glucose Levels

Network meta-analysis was conducted to evaluate the performance of different ML models. For PH=30 minutes, 10 (21.7%) studies [20-22,49-55] with 32 different ML models were included, and the network map is shown in Figure 3A.

The mean RMSE was 21.40 (SD 12.56) mg/dL. Statistically significant inconsistency was detected using the inconsistency test($I^2=87.11, P<.001$), as shown in the forest plot in Multimedia Appendix 1. Meta-regression indicated that I^2 for the RMSE was 60.75%, and the source of heterogeneity analysis showed that place and validation type were statistically significant ($P<.001$). The maximum SUCRA value was 99.1 for the dilated recurrent neural network (DRNN) model with a mean RMSE

of 7.80 (SD 0.60) mg/dL [22], whereas the minimum SUCRA value was 0.4 for 1 symbolic model with a mean RMSE of 71.4 (SD 21.9) mg/dL [49]. The relative ranks of the ML models are shown in Table 4, and the SUCRA curves are shown in Figure 4A. Publication bias was tested using the Egger test ($P=.503$), indicating no significant publication bias.

For PH=60 minutes, 4 (8.7%) studies [50,51,55] with 17 different ML models were included, and the network map is shown in Figure 3B. The mean RMSE was 30.01 (SD 7.23) mg/dL. Statistically significant inconsistency was detected using the inconsistency test ($I^2=8.82$, $P=.012$), as shown in the forest plot in Multimedia Appendix 3. Meta-regression indicated that none of the sample size, reference, place, validation type, and model type was a source of heterogeneity. The maximum SUCRA value was 97.8 for the GluNet model with a mean RMSE of 19.90 (SD 3.17) mg/dL [51], while the minimum SUCRA value was 4.5 for the decision tree (DT) model with a mean RMSE of 32.86 (SD 8.81) mg/dL [55]. The relative ranks of the ML models are shown in Table 5, and the SUCRA curves are shown in Figure 4B. No significant publication bias was detected using the Egger test ($P=.626$).

For PH=15 minutes, 3 (6.5%) studies [20,49,55] with 14 different ML models were included, and the network map is shown in Figure 3C. The mean RMSE was 18.88 (SD 19.71) mg/dL. Statistically significant inconsistency was detected using

the inconsistency test ($I^2=28.29$, $P<.001$), as shown in the forest plot in Multimedia Appendix 4. Meta-regression showed that I^2 was 41.28%, and the model type and sample size both were the source of heterogeneity, with $P=.002$ and $.037$, respectively. The maximum SUCRA value was 99.1 for the ARTiDe jump neural network (ARJNN) model with a mean RMSE of 9.50 (SD 1.90) mg/dL [49], while the minimum SUCRA value was 0.3 for the SVM with a mean RMSE of 13.13 (SD 17.30) mg/dL [55]. The relative ranks of the ML models are shown in Table 6, and SUCRA curves are shown in Figure 4C. Statistically significant publication bias was detected using the Egger test ($P=.003$).

For PH=45 minutes, only 2 (4.3%) studies [54,55] with 11 different ML models were included, and the network map is shown in Figure 3D. The mean RMSE was 21.27 (SD 5.17) mg/dL. Statistically significant inconsistency was detected using the inconsistency test ($I^2=6.92$, $P=.009$), as shown in the forest plot in Multimedia Appendix 5. Meta-regression indicated significant heterogeneity from the model type ($P=.006$). The maximum SUCRA value was 99.4 for the NNM with a mean RMSE of 10.65 (SD 3.87) mg/dL [55], while the minimum SUCRA value was 26.3 for the DT model with a mean RMSE of 23.35 (6.36) mg/dL [55]. The relative ranks of the ML models are shown in Table 7, and SUCRA curves are shown in Figure 4D. Statistically significant publication bias was detected using the Egger test ($P<.001$).

Figure 3. Network map of ML models for predicting BG levels in different PHs. PH=30 (A), 60 (B), 15 (C), and 45 minutes (D). ARIMA: autoregressive integrated moving average; ARM: autoregression model; ARMA: autoregressive moving average; ARJNN: ARTiDe jump neural network; BG: blood glucose; CRNN-MTL: convolutional recurrent neural network multitask learning; CRNN-MTL-GV: convolutional recurrent neural network multitask learning glycemic variability; CRNN-STL: convolutional recurrent neural network single-task learning; CRNN-TL: convolutional recurrent neural network transfer learning; DFFNN: delayed feed-forward neural network; DRNN: dilated recurrent neural network; DT: decision tree; FC: fully connected (neural network); fNN: feed-forward neural network; GCN: gradually connected neural network; JNN: jump neural network; kNN: k-nearest neighbor; LGBM: light gradient boosting machine; LSTM: long short-term memory; LVX: latent variable with exogenous input; ML: machine learning; NARX: one neural network model; NN-LPA: neural network-linear prediction algorithm; NNM: neural network model; PH: prediction horizon; RF: random forest; RNN: recurrent neural network; SAX: one symbolic model; SVR: support vector regression.

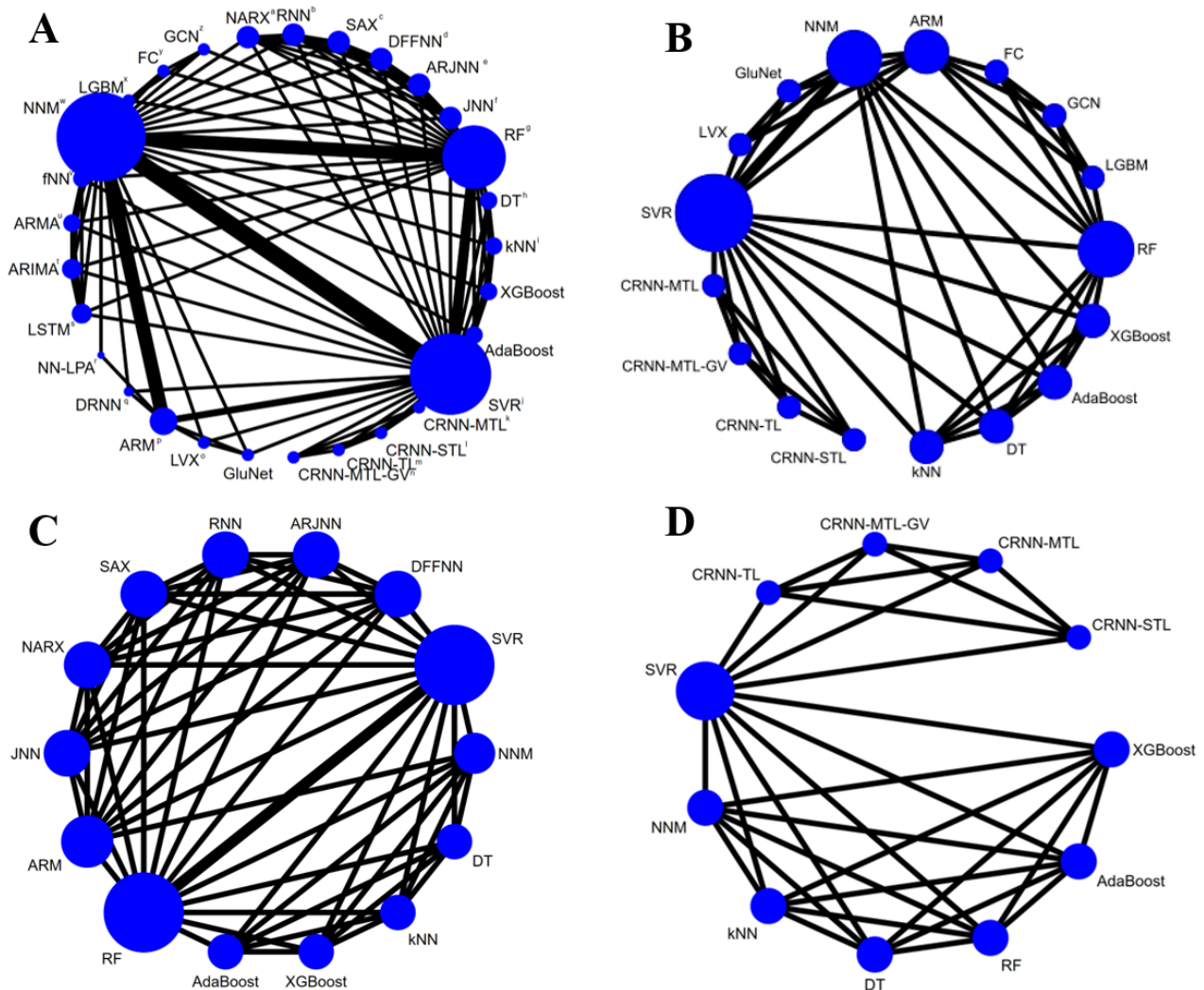


Table 4. Relative ranks of ML^a models for predicting BG^b levels in PH^c=30 minutes.

ML model	SUCRA ^d	Relative rank
NNM ^e	52.0	14.4
ARM ^f	39.6	17.9
ARJNN ^g	79.5	6.8
RF ^h	6.9	27.1
SVM ⁱ	73.3	8.5
One symbolic model (SAX)	0.4	28.9
Recurrent neural network (RNN)	19.0	23.7
One neural network model (NARX)	3.9	27.9
Jump neural network (JNN)	36.0	18.9
Delayed feed-forward neural network model (DFFNN)	15.8	24.6
Gradually connected neural network (GCN)	41.1	17.5
Fully connected (FC [neural network])	58.1	12.7
Light gradient boosting machine (LGBM)	69.3	9.6
DRNN ^j	99.1	1.2
Autoregressive moving average (ARMA)	54.3	13.8
Autoregressive integrated moving average (ARIMA)	46.6	16.0
Feed-forward neural network (fNN)	86.3	4.8
Long short-term memory (LSTM)	69.1	9.7
GluNet	96.4	2.0
Latent variable with exogenous input (LVX)	75.2	7.9
Neural network–linear prediction algorithm (NN-LPA)	60.0	12.2
Convolutional recurrent neural network multitask learning (CRNN-MTL)	77.5	7.3
Convolutional recurrent neural network multitask learning glycemic variability (CRNN-MTL-GV)	77.2	7.4
Convolutional recurrent neural network transfer learning (CRNN-TL)	71.8	8.9
Convolutional recurrent neural network single-task learning (CRNN-STL)	52.0	14.4
k-Nearest neighbor (kNN)	26.0	21.7
DT ^k	16.2	24.5
AdaBoost	18.0	24.0
XGBoost ^l	29.2	20.8

^aML: machine learning.

^bBG: blood glucose.

^cPH: prediction horizon.

^dSUCRA: surface under the cumulative ranking.

^eNNM: neural network model.

^fARM: autoregression model.

^gARJNN: ARTiDe jump neural network.

^hRF: random forest.

ⁱSVM: support vector machine.

^jDRNN: dilated recurrent neural network.

^kDT: decision tree.

^lXGBoost: Extreme Gradient Boosting.

Figure 4. SUCRA curves of ML models for predicting BG levels in different PHs. PH=30 (A), 60 (B), 15 (C), and 45 minutes (D). ARIMA: autoregressive integrated moving-average; ARM: autoregression model; ARMA: autoregressive moving average; ARJNN: ARTiDe jump neural network; BG: blood glucose; CRNN-MTL: convolutional recurrent neural networks multitask learning; CRNN-MTL-GV: convolutional recurrent neural networks multitask learning glycemic variability; CRNN-STL: convolutional recurrent neural networks single-task learning; CRNN-TL: convolutional recurrent neural networks transfer learning; DFFNN: delayed feed-forward neural network; DRNN: dilated recurrent neural network; DT: decision tree; FC: fully connected (neural network); fNN: feed-forward neural network; GCN: gradually connected neural network; JNN: jump neural network; kNN: k-nearest neighbor; LGBM: light gradient boosting machine; LSTM: long short-term memory; LVX: latent variable with exogenous input; ML: machine learning; NARX: one neural network model; NN-LPA: neural network-linear prediction algorithm; NNM: neural network model; PH: prediction horizon; RF: random forest; RNN: recurrent neural network; SAX: one symbolic model; SVR: support vector regression.

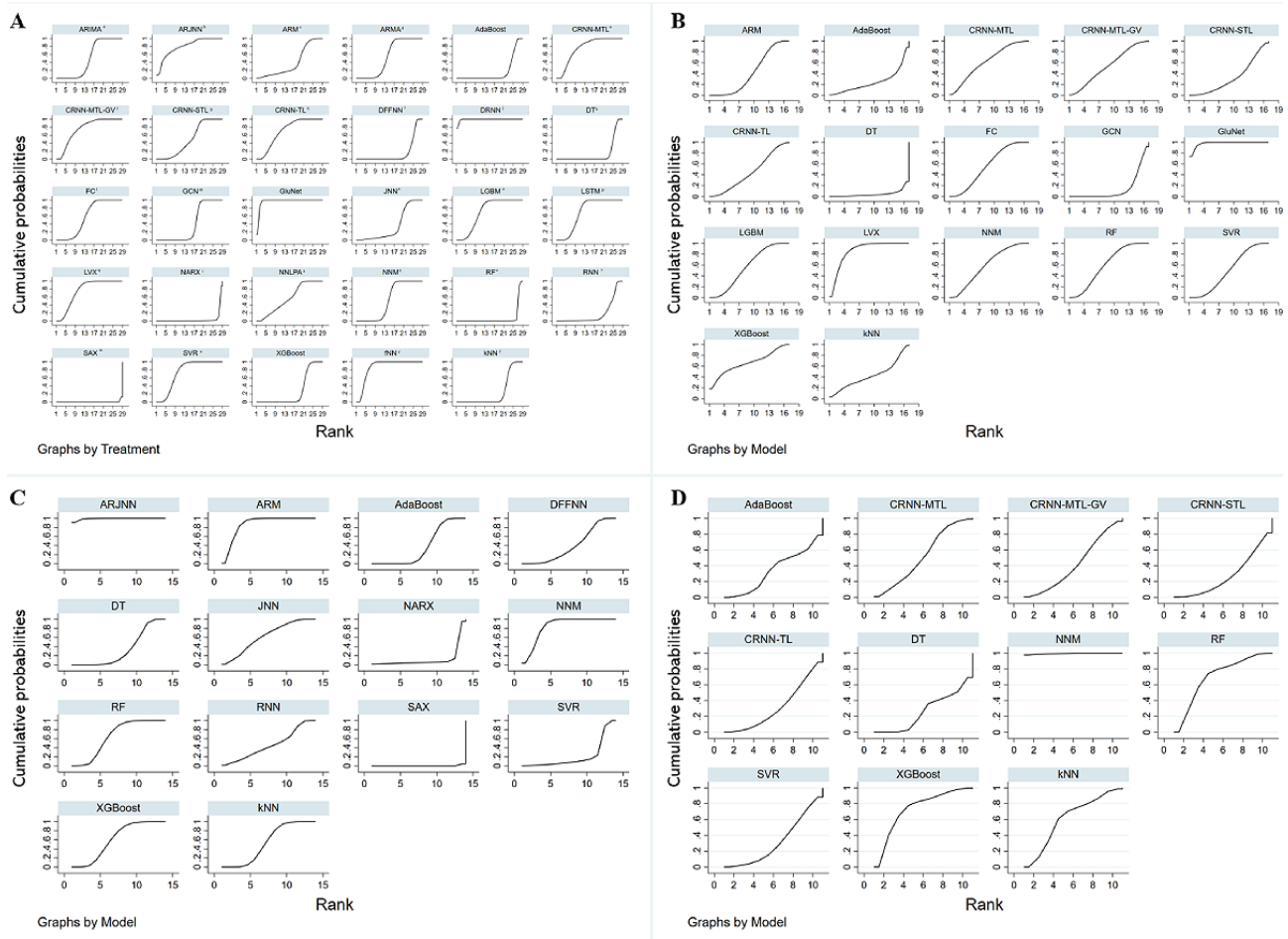


Table 5. Relative ranks of ML^a models for predicting BG^b levels in PH^c=60 minutes.

ML model	SUCRA ^d	Relative rank
ARM ^e	41.0	10.4
Gradually connected neural network (GCN)	14.2	14.7
Fully connected (FC [neural network])	55.7	8.1
Light gradient boosting machine (LGBM)	56.0	8.0
RF ^f	59.7	7.5
GluNet	97.8	1.4
NNM ^g	59.9	7.4
SVM ^h	49.5	9.1
Latent variable with exogenous input (LVX)	85.9	3.3
Convolutional recurrent neural network multitask learning (CRNN-MTL)	61.4	7.2
Convolutional recurrent neural network multitask learning glycemic variability (CRNN-MTL-GV)	54.2	8.3
Convolutional recurrent neural network transfer learning (CRNN-TL)	44.5	9.9
Convolutional recurrent neural network single-task learning (CRNN-STL)	32.5	11.8
k-Nearest neighbor (kNN)	42.5	10.2
DT ⁱ	4.5	16.3
AdaBoost	24.1	13.1
XGBoost ^j	66.5	6.4

^aML: machine learning.

^bBG: blood glucose.

^cPH: prediction horizon.

^dSUCRA: surface under the cumulative ranking.

^eARM: autoregression model.

^fRF: random forest.

^gNNM: neural network model.

^hSVM: support vector machine.

ⁱDT: decision tree.

^jXGBoost: Extreme Gradient Boosting.

Table 6. Relative ranks of ML^a models for predicting BG^b levels in PH^c=15 minutes.

ML model	SUCRA ^d	Relative rank
NNM ^e	84.4	3.0
ARM ^f	86.8	2.7
ARJNN ^g	99.1	1.1
RF ^h	64.6	5.6
SVM ⁱ	20.9	11.3
One symbolic model (SAX)	0.3	14.0
Recurrent neural network (RNN)	45.9	8.0
One neural network model (NARX)	11.8	12.5
Jump neural network (JNN)	62.2	5.9
Delayed feed-forward neural network model (DFNN)	39.6	8.9
k-Nearest neighbor (kNN)	53.7	7.0
DT ^j	33.3	9.7
AdaBoost	36.8	9.2
XGBoost ^k	60.8	6.1

^aML: machine learning.

^bBG: blood glucose.

^cPH: prediction horizon.

^dSUCRA: surface under the cumulative ranking.

^eNNM: neural network model.

^fARM: autoregression model.

^gARJNN: ARTiDe jump neural network.

^hRF: random forest.

ⁱSVM: support vector machine.

^jDT: decision tree.

^kXGBoost: Extreme Gradient Boosting.

Table 7. Relative ranks of ML^a models for predicting BG^b levels in PH^c=45 minutes.

ML model	SUCRA ^d	Relative rank
Convolutional recurrent neural network multitask learning (CRNN-MTL)	52.1	5.8
Convolutional recurrent neural network multitask learning glycemic variability (CRNN-MTL-GV)	41.8	6.8
Convolutional recurrent neural network transfer learning (CRNN-TL)	31.6	7.8
Convolutional recurrent neural network single-task learning (CRNN-STL)	27.5	8.2
SVM ^e	32.0	7.8
k-Nearest neighbor (kNN)	61.4	4.9
DT ^f	26.3	8.4
RF ^g	70.3	4.0
AdaBoost	34.1	7.6
XGBoost ^h	73.5	3.7
NNM ⁱ	99.4	1.1

^aML: machine learning.

^bBG: blood glucose.

^cPH: prediction horizon.

^dSUCRA: surface under the cumulative ranking.

^eSVM: support vector machine.

^fDT: decision tree.

^gRF: random forest.

^hXGBoost: Extreme Gradient Boosting.

ⁱNNM: neural network model.

Machine Learning Models for Predicting Hypoglycemia

ML models for predicting hypoglycemia (adverse BG events) involved 19 (41.3%) studies [15,29-39,47,48,56-60], with pooled estimates of 0.71 (95% CI 0.61-0.80) for sensitivity, 0.91 (95% CI 0.87-0.94) for specificity, 8.3 (95% CI 5.7-12.0) for the PLR, and 0.31 (95% CI 0.22-0.44) for the NLR. The heterogeneity between different ML models in these studies is shown in the forest plot in Figure 5, which was high for both sensitivity ($I^2=100%$, 95% CI 100%-100%) and specificity ($I^2=100%$, 95% CI 100%-100%). The SROC curve is shown in Figure 6A, with an area under the curve (AUC) of 0.91 (95% CI 0.88-0.93). According to the meta-regression results, the type of DM and time were statistically significant sources of heterogeneity for sensitivity while the type of DM, reference, data source, setting, and threshold were statistically significant sources of heterogeneity for specificity (Multimedia Appendix 6). No statistically significant publication bias was detected ($P=.09$). In addition to integral analysis for the hypoglycemia prediction model, we also carried out analysis of 4 subgroups based on the characteristics of the included studies, including the NNM, the RF, the SVM, and ensemble learning (RF, Extreme Gradient Boosting [XGBoost], bagging).

For the NNM, 3 (6.5%) studies [15,34,47] were included, with pooled estimates of 0.50 (95% CI 0.16-0.84) for sensitivity,

0.91 (95% CI 0.84-0.96) for specificity, 5.9 (95% CI 3.2-10.8) for the PLR, and 0.54 (95% CI 0.24-1.21) for the NLR. As shown in the forest plot in Figure 7A, I^2 values were 99.59% (95% CI 99.46%-99.71%) and 97.82% (95% CI 96.68%-98.86%) for sensitivity and specificity, respectively. The SROC curve is shown in Figure 6B, with an AUC of 0.90 (95% CI 0.87-0.92). Meta-regression results revealed that statistically significant heterogeneity was detected in all the factors between these studies (type of DM, reference, time, data source, setting, threshold) for sensitivity and 4 factors (reference, data source, setting, threshold) for specificity (Multimedia Appendix 7). No statistically significant publication bias was detected ($P=.86$).

For the RF, 5 (10.9%) studies [15,34,56,58,60] were included, with pooled estimates of 0.87 (95% CI 0.79-0.93) for sensitivity, 0.94 (95% CI 0.91-0.96) for specificity, 13.9 (95% CI 10.1-18.9) for the PLR, and 0.14 (95% CI 0.08-0.22) for the NLR. The forest plot in Figure 7B shows that statistically significant heterogeneity was detected in both sensitivity ($I^2=98.32%$, 95% CI 97.61%-99.02%) and specificity ($I^2=99.41%$, 95% CI 99.24%-99.58%). The SROC curve is shown in Figure 6C, with an AUC of 0.97 (95% CI 0.95-0.98). Meta-regression failed to run due to data instability or asymmetry. No statistically significant publication bias was detected ($P=.21$).

Figure 5. Sensitivity and specificity forest plots of ML models for predicting adverse BG events. The horizontal lines indicate 95% CIs. The square markers represent the effect value of a single study, and the diamond marker represents the combined results of all studies. The vertical line shows the line of no effects. BG: blood glucose; ML: machine learning.

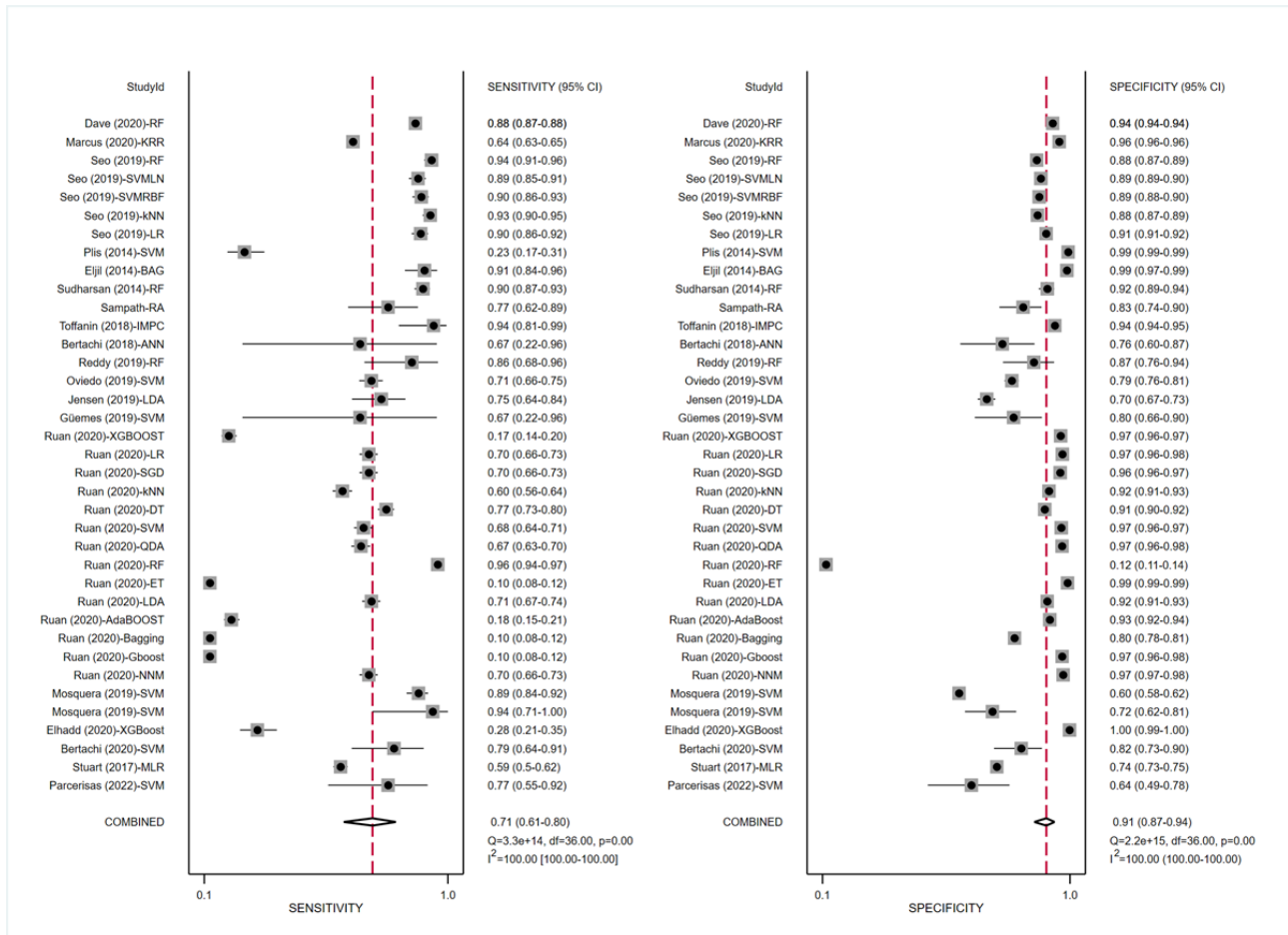


Figure 6. SROC curves of all ML algorithms (A), NNM algorithms (B), RF algorithms (C), SVM algorithms (D), and ensemble learning algorithms (E) for predicting adverse BG events. The hollow circles represent results of all studies, and the red diamonds represent the summary result of all studies. AUC: area under the curve; BG: blood glucose; ML: machine learning; NNM: neural network model; RF: random forest; SROC: summary receiver operating characteristic; SVM: support vector machine.

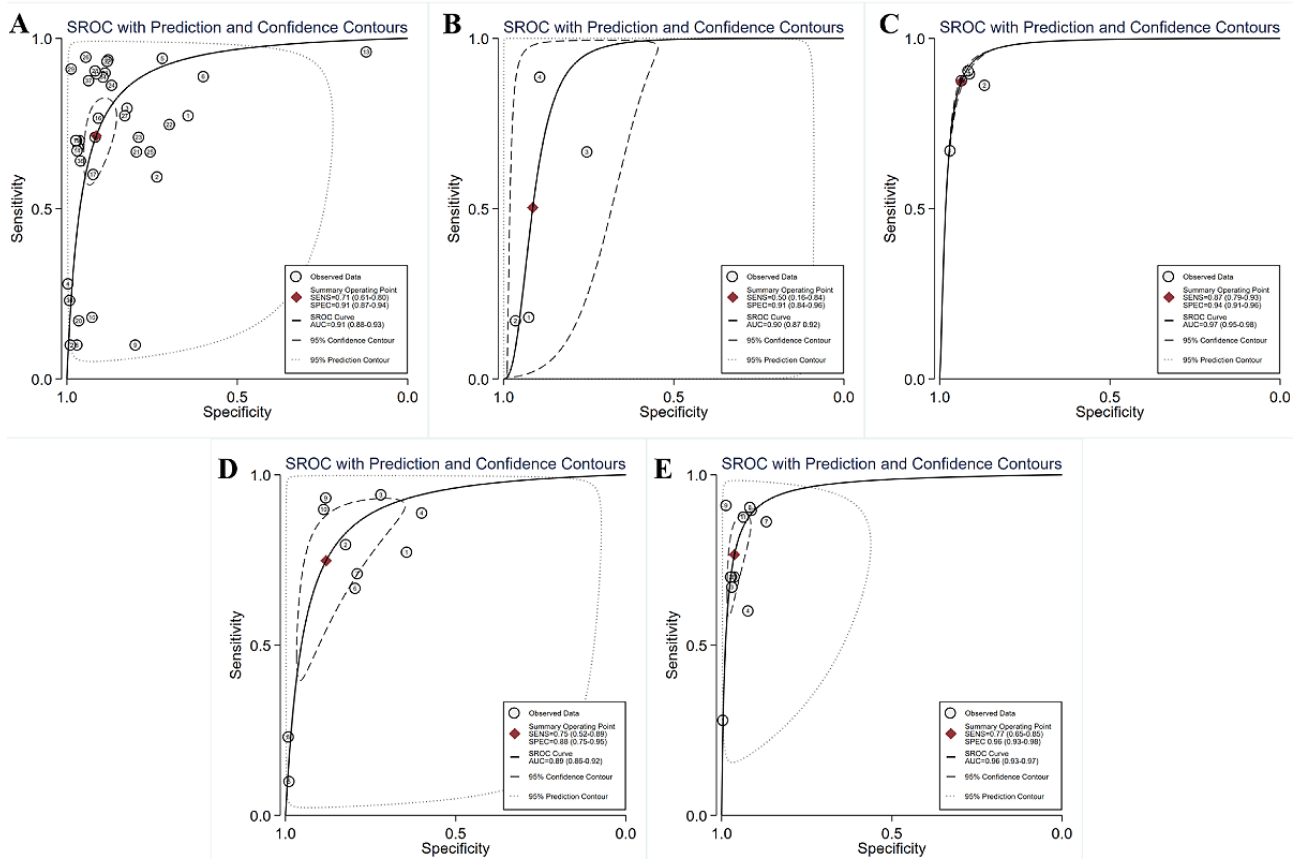
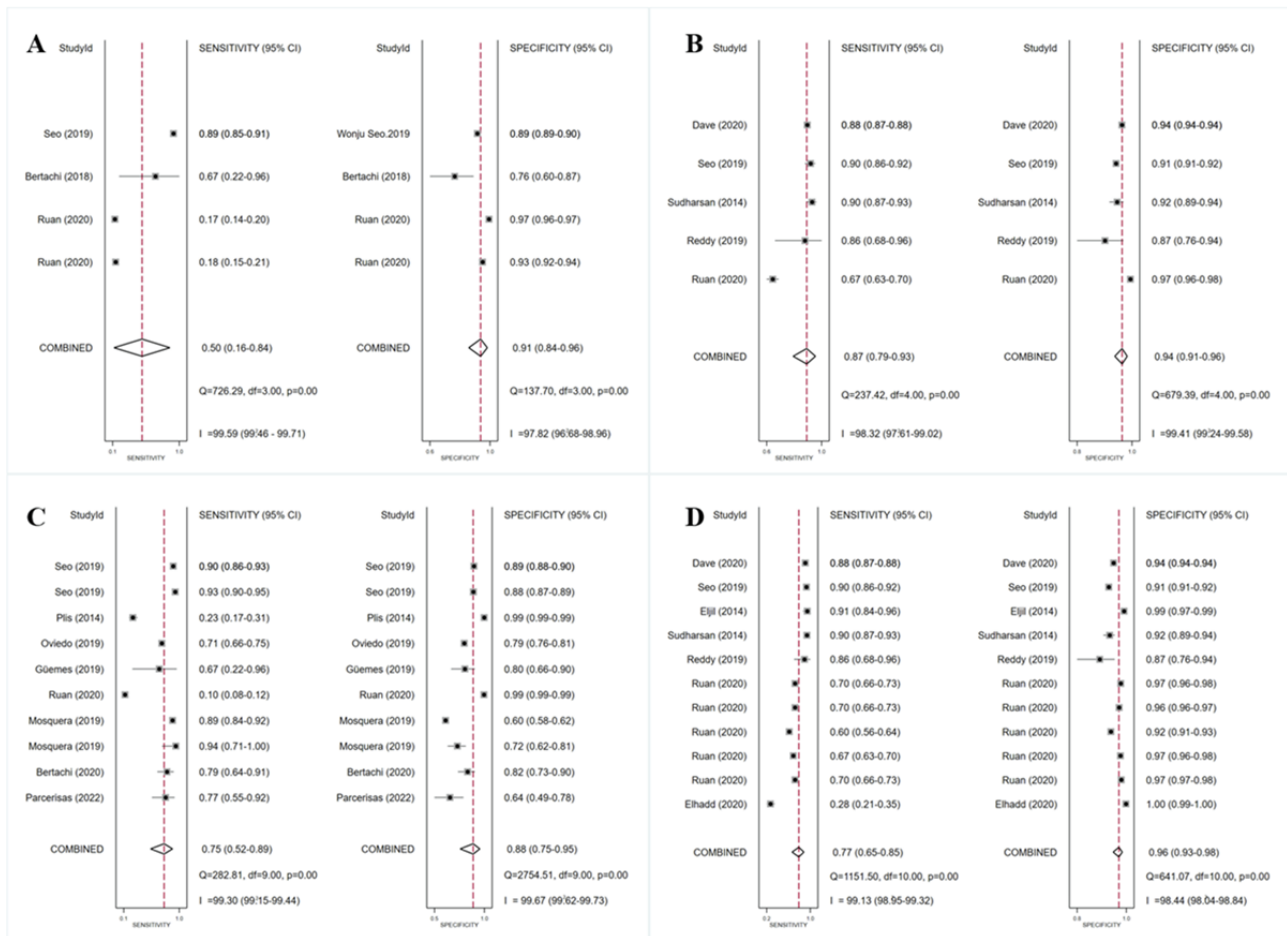


Figure 7. Sensitivity and specificity forest plots of NNM algorithms (A), RF models (B), SVM algorithms (C), and ensemble learning algorithms (D) for predicting adverse BG events. The horizontal lines indicate 95% CIs. The square markers represent the effect value of a single study, and the diamond marker represents the combined results of all studies. The vertical line shows the line of no effects. BG: blood glucose; NNM: neural network model; RF: random forest; SROC: summary receiver operating characteristic; SVM: support vector machine.



For the SVM, 8 (17.4%) studies [15,29,33-35,37,39,47] were involved, with pooled estimates of 0.75 (95% CI 0.52-0.89) for sensitivity, 0.88 (95% CI 0.75-0.95) for specificity, 6.3 (95% CI 3.4-11.7) for the PLR, and 0.29 (95% CI 0.15-0.55) for the NLR. Statistically significant heterogeneity was detected for both sensitivity ($I^2=99.30\%$, 95% CI 99.15%-99.44%) and specificity ($I^2=99.67\%$, 95% CI 99.62%-99.73%), as shown in Figure 7C. The SROC curve is shown in Figure 6D, with an AUC of 0.89 (95% CI 0.86-0.92). Meta-regression results showed that reference, time, data source, setting, and threshold were sources of heterogeneity for sensitivity, while reference, data source, setting, and threshold were sources of heterogeneity for specificity (Multimedia Appendix 8). Publication bias was not statistically significant ($P=.83$).

For ensemble learning models (RF, XGBoost, bagging), 7 (15.2%) studies [15,32,34,48,56,58,60] were involved, with pooled estimates of 0.77 (95% CI 0.65-0.85) for sensitivity, 0.96 (95% CI 0.93-0.98) for specificity, 20.4 (95% CI 12.5-33.3) for the PLR, and 0.24 (95% CI 0.16-0.37) for the NLR. Statistically significant heterogeneity was detected for both sensitivity ($I^2=99.13\%$, 95% CI 98.95%-99.32%) and specificity ($I^2=98.44\%$, 95% CI 98.04%-98.84%), as shown in Figure 7D. The SROC curve is shown in Figure 6E, with an AUC of 0.96 (95% CI 0.93-0.97). Meta-regression results showed that there was no source of heterogeneity for sensitivity, while the type

of DM, setting, and threshold were sources of heterogeneity for specificity (Multimedia Appendix 9). No statistically significant publication bias was detected ($P=.50$).

Machine Learning Models for Detecting Hypoglycemia

ML models for detecting hypoglycemia (adverse BG events) involved 17 (37%) studies [10,16,40-46,61-68], with pooled estimates of 0.74 (95% CI 0.70-0.78) for sensitivity, 0.70 (95% CI 0.56-0.81) for specificity, 2.4 (95% CI 1.6-3.7) for the PLR, and 0.37 (95% CI 0.29-0.46) for the NLR. The heterogeneity between different models in these studies is shown in the forest plots in Figure 8 and was high for both sensitivity ($I^2=92.80\%$, 95% CI 91.10%-94.49%) and specificity ($I^2=99.04\%$, 95% CI 98.82%-99.16%). The SROC curve is shown in Figure 9A, with an AUC of 0.77 (95% CI 0.73-0.81). Based on the meta-regression results, reference, time, data source, setting, and threshold were statistically significant sources of heterogeneity for sensitivity, while reference, data source, and threshold were statistically significant sources of heterogeneity for specificity (Multimedia Appendix 9). Statistically significant publication bias was detected ($P<.001$). In addition to integral analysis for the hypoglycemia detection model, we also carried out analysis of 2 subgroups based on the characteristics of the included studies, including the NNM and the SVM.

For the NNM, 11 (23.9%) studies [40-42,45,46,62-67] were involved, with pooled estimates of 0.76 (95% CI 0.70-0.80) for sensitivity, 0.67 (95% CI 0.49-0.82) for specificity, 2.3 (95% CI 1.4-3.9) for the PLR, and 0.36 (95% CI 0.27-0.48) for the NLR. The heterogeneity between different studies is shown in the forest plot in Figure 10A and was high for both sensitivity ($I^2=97.30\%$, 95% CI 96.62%-97.99%) and specificity ($I^2=98.23\%$, 95% CI 97.83%-98.62%). The SROC curve is shown in Figure 9B, with an AUC of 0.78 (95% CI 0.74-0.81). Based on the of meta-regression results, reference, time, data source, setting, and threshold were statistically significant sources of heterogeneity for sensitivity, while reference and setting were statistically significant sources of heterogeneity for specificity (Multimedia Appendix 10). Statistically significant publication bias was detected ($P<.001$).

For the SVM, 4 (8.7%) studies [10,44,61,62] were included, with pooled estimates of 0.80 (95% CI 0.73-0.86) for sensitivity, 0.65 (95% CI 0.41-0.83) for specificity, 2.3 (95% CI 1.2-4.4) for the PLR, and 0.31 (95% CI 0.18-0.51) for the NLR. The heterogeneity between different studies is shown in the forest plot in Figure 10B and was high for both sensitivity ($I^2=55.86\%$, 95% CI 11.96%-99.76%) and specificity ($I^2=99.02\%$, 95% CI 98.68%-99.36%). The SROC curve is shown in Figure 9C, with an AUC of 0.81 (95% CI 0.78-0.85). Meta-regression results indicated that reference, time, data source, setting, and threshold were statistically significant sources of heterogeneity for sensitivity, while reference, data source, setting, and threshold statistically significant sources of heterogeneity for specificity (Multimedia Appendix 11). No statistically significant publication bias was detected ($P=.31$).

Figure 8. Sensitivity and specificity forest plots of ML models for detecting adverse BG events. The horizontal lines indicate 95% CIs. The square markers represent the effect value of a single study, and the diamond marker represents the combined results of all studies. The vertical line shows the line of no effects. BG: blood glucose; ML: machine learning.

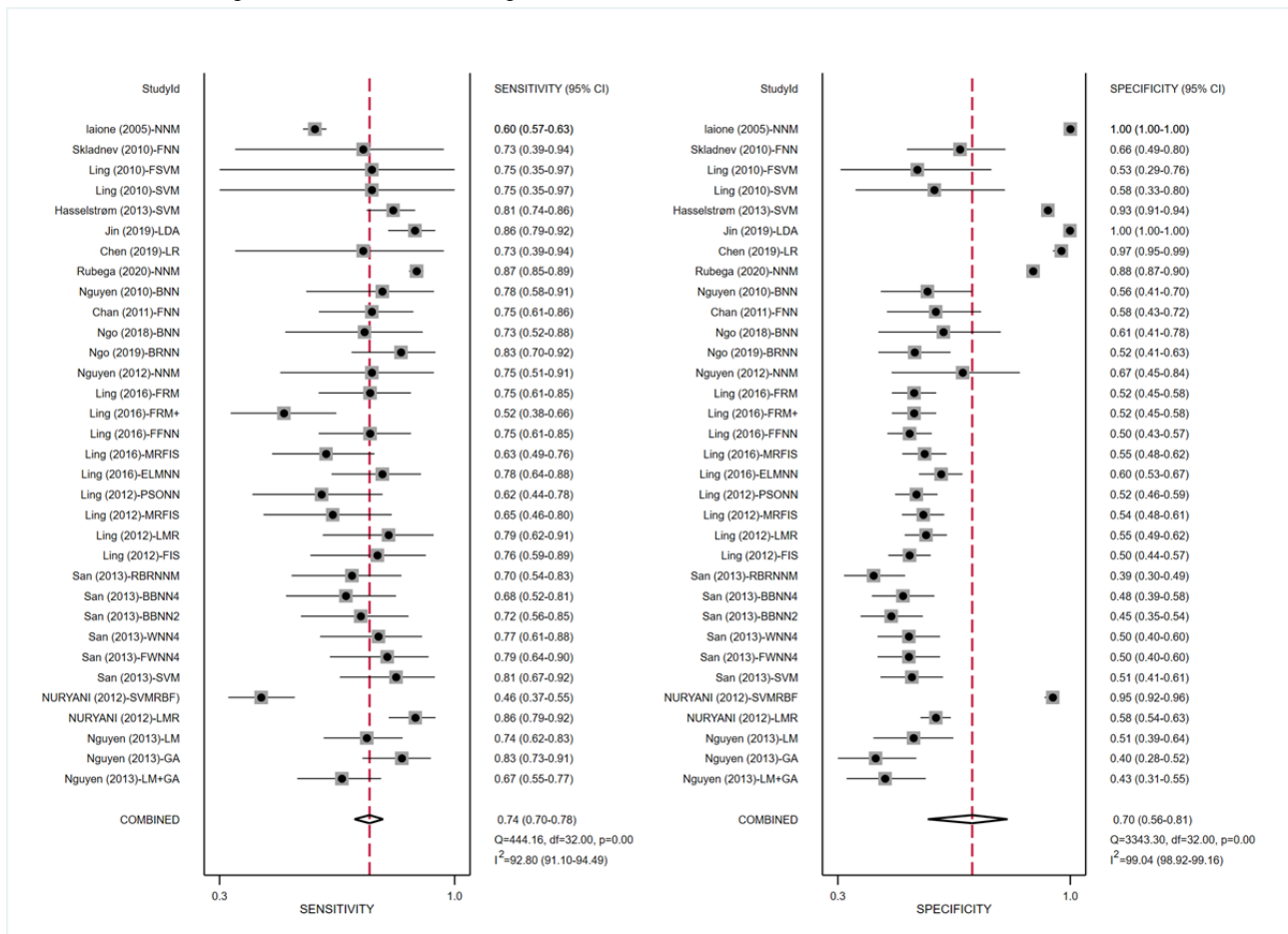


Figure 9. SROC curves of all ML algorithms (A), NNM algorithms (B), and SVM algorithms (C) for detecting adverse BG events. The hollow circles represent results of all studies, and the red diamonds represent the summary result of all studies. AUC: area under the curve; BG: blood glucose; ML: machine learning; NNM: neural network model; SROC: summary receiver operating characteristic; SVM: support vector machine.

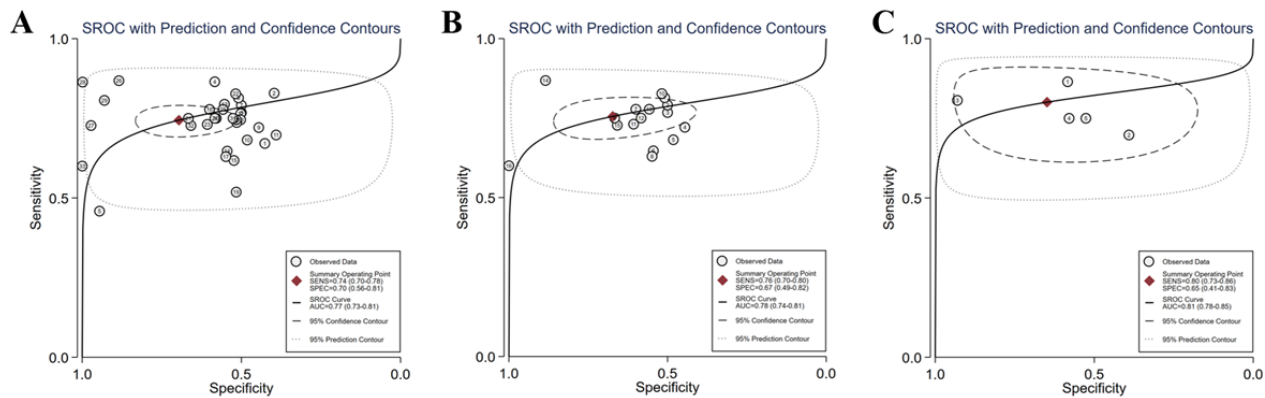
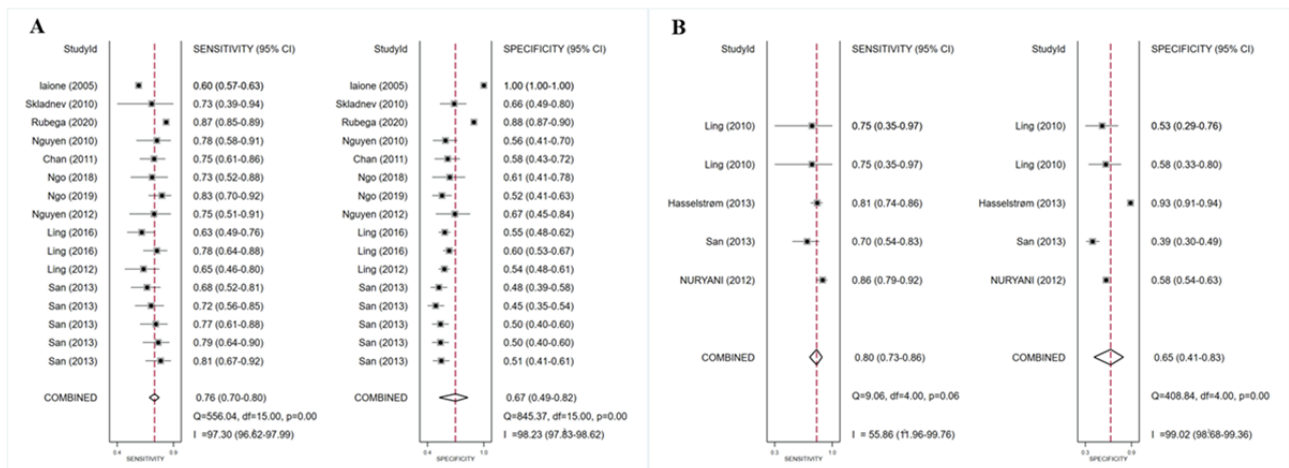


Figure 10. Sensitivity and specificity forest plots of NNM algorithms (A) and SVM algorithms (B) for detecting adverse BG events. The horizontal lines indicate 95% CIs. The square markers represent the effect value of a single study, and the diamond marker represents the combined results of all studies. The vertical line shows the line of no effects. BG: blood glucose; NNM: neural network model; SVM: support vector machine.



Discussion

Principal Findings

This meta-analysis systematically assessed the performance of different ML models in enhancing BG management in patients with DM based on 46 eligible studies. Comprehensive evidence obtained via exhaustive searching allowed us to assess the overall ability of the ML models in different scenarios, including predicting BG levels, predicting adverse BG events, and detecting adverse BG events.

Comparison to Prior Work

Obviously, the RMSE of ML models for predicting BG levels increased as the PH increased from 15 to 60 minutes, which indicates that the longer the PH, the larger the prediction error. Based on the results of relative ranking, among all the ML models for predicting BG levels, neural network-based models, including the DRNN, GluNet, ARJNN, and NNM, achieved the minimum RMSE and the maximum SUCRA in different PHs, indicating the highest relative performance. In contrast, the DT achieved the maximum RMSE and the minimum SUCRA in a PH of 60 and 45 minutes, indicating that lowest relative performance. Thus, for predicting BG levels, neural network-based algorithms might be an appropriate choice. We

found that time domain features combined with historical BG levels as input can further improve the performance of NNM algorithms [49,55]. However, the quality of training data for NNMs needs to be high; therefore, the requirements during data collection and preprocessing of raw data are high [22,51].

Regarding ML models for predicting adverse BG events, the pooled sensitivity, specificity, PLR, and NLR were 0.71 (95% CI 0.61-0.80), 0.91 (95% CI 0.87-0.94), 8.3 (95% CI 5.7-12.0), and 0.31 (95% CI 0.22-0.44), respectively. According to the *Users' Guide to Medical Literature*, with regard to diagnostic tests [69], a PLR of 5-10 should be able to moderately increase the probability of persons having or developing a disease and an NLR of 0.1-0.2 should be able to moderately decrease the probability of having or developing a disease after taking the index test. Hence, current ML models have relatively sufficient ability to predict the occurrence of hypoglycemia, especially RF algorithms with a PLR of 13.9 (95% CI 10.1-18.9) and an NLR of 0.14 (95% CI 0.08-0.22). On the contrary, although the PLR of NNM algorithms was 5.9 (95% CI 3.2-10.8), their sensitivity and NLR were 0.50 (95% CI 0.16-0.84) and 0.54 (95% CI 0.24-1.21), respectively, which is far from satisfactory. Although RF algorithms seem to be able to capture the complex, nonlinear patterns affecting hypoglycemia [56], it was still not enough to determine which algorithm shows the best

performance, as the test scenarios were quite different and there was high heterogeneity between studies.

Regarding ML models for detecting hypoglycemia, the pooled sensitivity, specificity, PLR, and NLR were 0.74 (95% CI 0.70-0.78), 0.70 (0.56-0.81), 2.4 (1.6-3.7), and 0.37 (0.29-0.46), respectively, which indicates that the algorithms generate small changes in probability [69]. Nevertheless, it does not mean that ML models combined with ECG or EEG monitoring, which we found in 13 of 17 studies, should not be further investigated. Considering patients with both DM and cardiovascular risk, or patients under intensive care and in a coma, combined ML models and ECG or EEG signals might be able to avoid deficits in physical and cognitive function and death caused by hypoglycemia [70].

Strengths and Limitations

The study has several limitations. First, although we developed a comprehensive search strategy, there was still a possibility of potential missing studies. To further increase the rate of literature retrieval, we included the main medical databases with a feasible search strategy, including PubMed, Embase, Web of Science, and IEEE Explore, and references from relevant studies were also screened for eligibility to avoid omissions. Second, statistically significant high heterogeneity was detected in all subgroups, with different sources of heterogeneity, including different types of DM, ML models, data sources, reference index, time and setting of data collection, and threshold of hypoglycemia, among studies. To address this issue, hierarchical analysis and meta-regression analysis were carried out in different subgroups to explore the possible sources of heterogeneity. Furthermore, for several studies that provided no required outcome measures or had inconsistent outcome measures, relevant estimation methods were used to calculate the indicators, which might have led to a certain amount of estimation error. However, the estimation error was small enough to be accepted owing to an appropriate estimation method, and the results of this study were further enriched. However, future studies are required to report all relevant outcome measures for further evaluation.

Future Directions

In future, more accurate ML models will be used for BG management, which will certainly improve the quality of life of patients with DM and reduce the burden of adverse BG events. First, as mentioned before, current ML models have relatively sufficient ability to predict BG levels and hypoglycemia, and the fact that an extended PH is more beneficial for increasing the time available for patients and

clinicians to respond still needs to be emphasized [15]. Hence, future studies should focus on enhancing the performance of ML models in longer PHs (ie, 60 minutes). Second, most of the raw data from CGM devices are highly imbalanced due to the low incidence of adverse BG events, which may lead to several performance distortions. Previous studies have reported several approaches to reduce the data imbalance, including oversampling [71] and cost-based learning [15]. However, to the best of our knowledge, few studies have investigated the effectiveness of those approaches in BG management models, which needs to be further studied in the future. Furthermore, the high variability of BG levels in the human body due to several factors, such as meal intake, high-intensity exercise, and insulin dosage, creates challenges for ML models; thus, future works need to integrate these factors with existing models to further enhance their accuracy [22,51]. It is also necessary to consider the computational complexity and convenience of use for patients and physicians. Moreover, several studies have implied that a combination of ML models and features extracted from CGM profiles can achieve better predictability compared to an ML model alone [15,56]. Recently, studies have focused on more novel deep learning models, such as transformers, which have also been proved clinically useful [72]. Therefore, further studies that focus on optimizing the structure of an ensemble method are needed to explore more models with a new structure. Lastly, it should be mentioned that although several studies have achieved high performance using relatively small data set [29,31,32,35,39,47,57], which can reduce the difficulty in model development, it also creates a concern about whether this will decrease the generalization ability of the models. Most of the models were developed and tested with a certain data set, and few of them have been prospectively validated in a clinical setting. Therefore, they need to be applied in clinical practice and be updated, as needed, to provide real-time feedback for the automatic collection of BG levels and generate a basis for prompt medical intervention [73].

Conclusion

In summary, in predicting precise BG levels, the RMSE increases with an increase in the PH, and the NNM shows the relatively highest performance among all the ML models. Meanwhile, according to the PLR and NLR, current ML models have sufficient ability to predict adverse BG (hypoglycemia) events, while their ability to detect adverse BG events needs to be enhanced. Future studies are required to focus on improving the performance and using ML models in clinical practice [70,73].

Acknowledgments

The study was funded by the National Natural Science Foundation of China (grant no. 82073663) and the Shaanxi Provincial Research and Development Program Foundation (grant nos. 2017JM7008 and 2022SF-245).

Data Availability

The data sets used and analyzed during the study are available from the corresponding author upon reasonable request.

Authors' Contributions

YW and CC conceived and designed the study. KL and LL undertook the literature review and extracted data. KL, LL, and JJ interpreted the data. KL, YM, and SL wrote the first draft of the manuscript, with revision by YW, ZL, CP, and ZY. All authors have read and approved the final version of the manuscript and had final responsibility for submitting it for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplemental plot1-forest (RMSE PH=30). PH: prediction horizon; RMSE: root mean square error.
[\[PNG File , 808 KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) checklist.
[\[PDF File \(Adobe PDF File\), 66 KB-Multimedia Appendix 2\]](#)

Multimedia Appendix 3

Supplemental plot2-forest (RMSE PH=60). PH: prediction horizon; RMSE: root mean square error.
[\[PNG File , 565 KB-Multimedia Appendix 3\]](#)

Multimedia Appendix 4

Supplemental plot3-forest (RMSE PH=15). PH: prediction horizon; RMSE: root mean square error.
[\[PNG File , 1014 KB-Multimedia Appendix 4\]](#)

Multimedia Appendix 5

Supplemental plot4-forest (RMSE PH=45). PH: prediction horizon; RMSE: root mean square error.
[\[PNG File , 838 KB-Multimedia Appendix 5\]](#)

Multimedia Appendix 6

Supplemental plot5 - metaregression (pre-all).
[\[PNG File , 130 KB-Multimedia Appendix 6\]](#)

Multimedia Appendix 7

Supplemental plot5-metaregression(pre-NN).
[\[PNG File , 136 KB-Multimedia Appendix 7\]](#)

Multimedia Appendix 8

Supplemental plot5-metaregression(pre-SVM).
[\[PNG File , 132 KB-Multimedia Appendix 8\]](#)

Multimedia Appendix 9

Supplemental plot5-metaregression(det-all).
[\[PNG File , 129 KB-Multimedia Appendix 9\]](#)

Multimedia Appendix 10

supplemental plot5-metaregression(det-NN).
[\[PNG File , 123 KB-Multimedia Appendix 10\]](#)

Multimedia Appendix 11

Supplemental plot5-metaregression(det-SVM).
[\[PNG File , 132 KB-Multimedia Appendix 11\]](#)

References

1. Oviedo S, Vehí J, Calm R, Armengol J. A review of personalized blood glucose prediction strategies for T1DM patients. *Int J Numer Method Biomed Eng* 2017 Jun;33(6):e2833 [doi: [10.1002/cnm.2833](https://doi.org/10.1002/cnm.2833)] [Medline: [27644067](https://pubmed.ncbi.nlm.nih.gov/27644067/)]
2. Saeedi P, Petersohn I, Salpea P, Malanda B, Karuranga S, Unwin N, et al. IDF Diabetes Atlas Committee. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: results from the International Diabetes Federation Diabetes Atlas, 9 edition. *Diabetes Res Clin Pract* 2019 Nov;157:107843 [doi: [10.1016/j.diabres.2019.107843](https://doi.org/10.1016/j.diabres.2019.107843)] [Medline: [31518657](https://pubmed.ncbi.nlm.nih.gov/31518657/)]
3. BMC Medicine. Diabetes education for better personalized management in pediatric patients. *BMC Med* 2023 Jan 24;21(1):30 [FREE Full text] [doi: [10.1186/s12916-022-02709-2](https://doi.org/10.1186/s12916-022-02709-2)] [Medline: [36690983](https://pubmed.ncbi.nlm.nih.gov/36690983/)]
4. Chen D, Wang M, Shang X, Liu X, Liu X, Ge T, et al. Development and validation of an incidence risk prediction model for early foot ulcer in diabetes based on a high evidence systematic review and meta-analysis. *Diabetes Res Clin Pract* 2021 Oct;180:109040 [doi: [10.1016/j.diabres.2021.109040](https://doi.org/10.1016/j.diabres.2021.109040)] [Medline: [34500005](https://pubmed.ncbi.nlm.nih.gov/34500005/)]
5. Li Y, Su X, Ye Q, Guo X, Xu B, Guan T, et al. The predictive value of diabetic retinopathy on subsequent diabetic nephropathy in patients with type 2 diabetes: a systematic review and meta-analysis of prospective studies. *Ren Fail* 2021 Dec;43(1):231-240 [FREE Full text] [doi: [10.1080/0886022X.2020.1866010](https://doi.org/10.1080/0886022X.2020.1866010)] [Medline: [33478336](https://pubmed.ncbi.nlm.nih.gov/33478336/)]
6. Wu B, Niu Z, Hu F. Study on risk factors of peripheral neuropathy in type 2 diabetes mellitus and establishment of prediction model. *Diabetes Metab J* 2021 Jul;45(4):526-538 [FREE Full text] [doi: [10.4093/dmj.2020.0100](https://doi.org/10.4093/dmj.2020.0100)] [Medline: [34352988](https://pubmed.ncbi.nlm.nih.gov/34352988/)]
7. Bellemo V, Lim G, Rim TH, Tan GSW, Cheung CY, Sadda S, et al. Artificial intelligence screening for diabetic retinopathy: the real-world emerging application. *Curr Diab Rep* 2019 Jul 31;19(9):72 [doi: [10.1007/s11892-019-1189-3](https://doi.org/10.1007/s11892-019-1189-3)] [Medline: [31367962](https://pubmed.ncbi.nlm.nih.gov/31367962/)]
8. Jain AMC, Ahmeti I, Bogoev M, Petrovski G, Milenkovicj T, Krstevska B, et al. A new classification of diabetic foot complications: a simple and effective teaching tool. *J Diab Foot Comp* 2012;4(1):1-5
9. Okonofua FE, Odimegwu C, Ajabor H, Daru PH, Johnson A. Assessing the prevalence and determinants of unwanted pregnancy and induced abortion in Nigeria. *Stud Fam Plann* 1999 Mar;30(1):67-77 [doi: [10.1111/j.1728-4465.1999.00067.x](https://doi.org/10.1111/j.1728-4465.1999.00067.x)] [Medline: [10216897](https://pubmed.ncbi.nlm.nih.gov/10216897/)]
10. Jin Y, Li F, Vimalananda VG, Yu H. Automatic detection of hypoglycemic events from the electronic health record notes of diabetes patients: empirical study. *JMIR Med Inform* 2019 Nov 08;7(4):e14340 [FREE Full text] [doi: [10.2196/14340](https://doi.org/10.2196/14340)] [Medline: [31702562](https://pubmed.ncbi.nlm.nih.gov/31702562/)]
11. Lipska KJ, Ross JS, Wang Y, Inzucchi SE, Minges K, Karter AJ, et al. National trends in US hospital admissions for hyperglycemia and hypoglycemia among Medicare beneficiaries, 1999 to 2011. *JAMA Intern Med* 2014 Jul;174(7):1116-1124 [FREE Full text] [doi: [10.1001/jamainternmed.2014.1824](https://doi.org/10.1001/jamainternmed.2014.1824)] [Medline: [24838229](https://pubmed.ncbi.nlm.nih.gov/24838229/)]
12. Zou Y, Zhao L, Zhang J, Wang Y, Wu Y, Ren H, et al. Development and internal validation of machine learning algorithms for end-stage renal disease risk prediction model of people with type 2 diabetes mellitus and diabetic kidney disease. *Ren Fail* 2022 Dec;44(1):562-570 [FREE Full text] [doi: [10.1080/0886022X.2022.2056053](https://doi.org/10.1080/0886022X.2022.2056053)] [Medline: [35373711](https://pubmed.ncbi.nlm.nih.gov/35373711/)]
13. Felizardo V, Garcia NM, Pombo N, Megdiche I. Data-based algorithms and models using diabetics real data for blood glucose and hypoglycaemia prediction - a systematic literature review. *Artif Intell Med* 2021 Aug;118:102120 [doi: [10.1016/j.artmed.2021.102120](https://doi.org/10.1016/j.artmed.2021.102120)] [Medline: [34412843](https://pubmed.ncbi.nlm.nih.gov/34412843/)]
14. Rodbard D. Continuous glucose monitoring: a review of recent studies demonstrating improved glycemic outcomes. *Diabetes Technol Ther* 2017 Jun;19(S3):S25-S37 [FREE Full text] [doi: [10.1089/dia.2017.0035](https://doi.org/10.1089/dia.2017.0035)] [Medline: [28585879](https://pubmed.ncbi.nlm.nih.gov/28585879/)]
15. Seo W, Lee Y, Lee S, Jin S, Park S. A machine-learning approach to predict postprandial hypoglycemia. *BMC Med Inform Decis Mak* 2019 Nov 06;19(1):210 [FREE Full text] [doi: [10.1186/s12911-019-0943-4](https://doi.org/10.1186/s12911-019-0943-4)] [Medline: [31694629](https://pubmed.ncbi.nlm.nih.gov/31694629/)]
16. Nguyen LB, Nguyen AV, Ling SH, Nguyen HT. Combining genetic algorithm and Levenberg-Marquardt algorithm in training neural network for hypoglycemia detection using EEG signals. *Annu Int Conf IEEE Eng Med Biol Soc* 2013;2013:5386-5389 [doi: [10.1109/EMBC.2013.6610766](https://doi.org/10.1109/EMBC.2013.6610766)] [Medline: [24110953](https://pubmed.ncbi.nlm.nih.gov/24110953/)]
17. Rodríguez-Rodríguez I, Rodríguez J, Woo WL, Wei B, Pardo-Quiles D. A comparison of feature selection and forecasting machine learning algorithms for predicting glycaemia in type 1 diabetes mellitus. *Appl Sci* 2021 Feb 16;11(4):1742 [doi: [10.3390/app11041742](https://doi.org/10.3390/app11041742)]
18. Wang Y, Wu X, Mo X. A novel adaptive-weighted-average framework for blood glucose prediction. *Diabetes Technol Ther* 2013 Oct;15(10):792-801 [FREE Full text] [doi: [10.1089/dia.2013.0104](https://doi.org/10.1089/dia.2013.0104)] [Medline: [23883406](https://pubmed.ncbi.nlm.nih.gov/23883406/)]
19. San PP, Ling SH, Soe NN, Nguyen HT. A novel extreme learning machine for hypoglycemia detection. *Annu Int Conf IEEE Eng Med Biol Soc* 2014;2014:302-305 [doi: [10.1109/EMBC.2014.6943589](https://doi.org/10.1109/EMBC.2014.6943589)] [Medline: [25569957](https://pubmed.ncbi.nlm.nih.gov/25569957/)]
20. Pérez-Gandía C, Facchinetti A, Sparacino G, Cobelli C, Gómez EJ, Rigla M, et al. Artificial neural network algorithm for online glucose prediction from continuous glucose monitoring. *Diabetes Technol Ther* 2010 Jan;12(1):81-88 [doi: [10.1089/dia.2009.0076](https://doi.org/10.1089/dia.2009.0076)] [Medline: [20082589](https://pubmed.ncbi.nlm.nih.gov/20082589/)]
21. Prendin F, Del Favero S, Vettoretti M, Sparacino G, Facchinetti A. Forecasting of glucose levels and hypoglycemic events: head-to-head comparison of linear and nonlinear data-driven algorithms based on continuous glucose monitoring data only. *Sensors (Basel)* 2021 Feb 27;21(5):1647 [FREE Full text] [doi: [10.3390/s21051647](https://doi.org/10.3390/s21051647)] [Medline: [33673415](https://pubmed.ncbi.nlm.nih.gov/33673415/)]

22. Zhu T, Li K, Chen J, Herrero P, Georgiou P. Dilated recurrent neural networks for glucose forecasting in type 1 diabetes. *J Healthc Inform Res* 2020 Sep 12;4(3):308-324 [[FREE Full text](#)] [doi: [10.1007/s41666-020-00068-2](https://doi.org/10.1007/s41666-020-00068-2)] [Medline: [35415447](#)]
23. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: the PRISMA statement. *PLoS Med* 2009 Jul 21;6(7):e1000097 [[FREE Full text](#)] [doi: [10.1371/journal.pmed.1000097](https://doi.org/10.1371/journal.pmed.1000097)] [Medline: [19621072](#)]
24. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* 2009 Jul 21;339:b2700 [[FREE Full text](#)] [doi: [10.1136/bmj.b2700](https://doi.org/10.1136/bmj.b2700)] [Medline: [19622552](#)]
25. Akl E, Altman D, Aluko P, Askie L, Beaton D, Berlin J. *Cochrane Handbook for Systematic Reviews of Interventions*. New York, NY: John Wiley & Sons; 2019.
26. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011 Oct 18;155(8):529-536 [[FREE Full text](#)] [doi: [10.7326/0003-4819-155-8-201110180-00009](https://doi.org/10.7326/0003-4819-155-8-201110180-00009)] [Medline: [22007046](#)]
27. White I. Multivariate random-effects meta-regression: updates to Mvmeta. *Stata J* 2011 Jul 01;11(2):255-270 [doi: [10.1177/1536867x1101100206](https://doi.org/10.1177/1536867x1101100206)]
28. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003 Sep 06;327(7414):557-560 [[FREE Full text](#)] [doi: [10.1136/bmj.327.7414.557](https://doi.org/10.1136/bmj.327.7414.557)] [Medline: [12958120](#)]
29. Parcerisas A, Contreras I, Delecourt A, Bertachi A, Beneyto A, Conget I, et al. A machine learning approach to minimize nocturnal hypoglycemic events in type 1 diabetic patients under multiple doses of insulin. *Sensors (Basel)* 2022 Feb 21;22(4):1665 [[FREE Full text](#)] [doi: [10.3390/s22041665](https://doi.org/10.3390/s22041665)] [Medline: [35214566](#)]
30. Stuart K, Adderley NJ, Marshall T, Rayman G, Sitch A, Manley S, et al. Predicting inpatient hypoglycaemia in hospitalized patients with diabetes: a retrospective analysis of 9584 admissions with diabetes. *Diabet Med* 2017 Oct 12;34(10):1385-1391 [doi: [10.1111/dme.13409](https://doi.org/10.1111/dme.13409)] [Medline: [28632918](#)]
31. Bertachi A, Viñals C, Biagi L, Contreras I, Vehí J, Conget I, et al. Prediction of nocturnal hypoglycemia in adults with type 1 diabetes under multiple daily injections using continuous glucose monitoring and physical activity monitor. *Sensors (Basel)* 2020 Mar 19;20(6):1705 [[FREE Full text](#)] [doi: [10.3390/s20061705](https://doi.org/10.3390/s20061705)] [Medline: [32204318](#)]
32. Elhadd T, Mall R, Bashir M, Palotti J, Fernandez-Luque L, Farooq F, et al. for PROFAST-Ramadan Study Group. Artificial intelligence (AI) based machine learning models predict glucose variability and hypoglycaemia risk in patients with type 2 diabetes on a multiple drug regimen who fast during ramadan (the PROFAST - IT Ramadan study). *Diabetes Res Clin Pract* 2020 Nov;169:108388 [[FREE Full text](#)] [doi: [10.1016/j.diabres.2020.108388](https://doi.org/10.1016/j.diabres.2020.108388)] [Medline: [32858096](#)]
33. Mosquera-Lopez C, Dodier R, Tyler NS, Wilson LM, El Youssef J, Castle JR, et al. Predicting and preventing nocturnal hypoglycemia in type 1 diabetes using big data analytics and decision theoretic analysis. *Diabetes Technol Ther* 2020 Nov;22(11):801-811 [[FREE Full text](#)] [doi: [10.1089/dia.2019.0458](https://doi.org/10.1089/dia.2019.0458)] [Medline: [32297795](#)]
34. Ruan Y, Bellot A, Moysova Z, Tan GD, Lumb A, Davies J, et al. Predicting the risk of inpatient hypoglycemia with machine learning using electronic health records. *Diabetes Care* 2020 Jul;43(7):1504-1511 [doi: [10.2337/dc19-1743](https://doi.org/10.2337/dc19-1743)] [Medline: [32350021](#)]
35. Guemes A, Cappon G, Hernandez B, Reddy M, Oliver N, Georgiou P, et al. Predicting quality of overnight glycaemic control in type 1 diabetes using binary classifiers. *IEEE J Biomed Health Inform* 2020 May;24(5):1439-1446 [[FREE Full text](#)] [doi: [10.1109/JBHI.2019.2938305](https://doi.org/10.1109/JBHI.2019.2938305)] [Medline: [31536025](#)]
36. Jensen MH, Dethlefsen C, Vestergaard P, Hejlesen O. Prediction of nocturnal hypoglycemia from continuous glucose monitoring data in people with type 1 diabetes: a proof-of-concept study. *J Diabetes Sci Technol* 2020 Mar;14(2):250-256 [[FREE Full text](#)] [doi: [10.1177/1932296819868727](https://doi.org/10.1177/1932296819868727)] [Medline: [31390891](#)]
37. Oviedo S, Contreras I, Quirós C, Giménez M, Conget I, Vehí J. Risk-based postprandial hypoglycemia forecasting using supervised learning. *Int J Med Inform* 2019 Jun;126:1-8 [doi: [10.1016/j.ijmedinf.2019.03.008](https://doi.org/10.1016/j.ijmedinf.2019.03.008)] [Medline: [31029250](#)]
38. Toffanin C, Aiello EM, Cobelli C, Magni L. Hypoglycemia prevention via personalized glucose-insulin models identified in free-living conditions. *J Diabetes Sci Technol* 2019 Nov;13(6):1008-1016 [[FREE Full text](#)] [doi: [10.1177/1932296819880864](https://doi.org/10.1177/1932296819880864)] [Medline: [31645119](#)]
39. Plis K, Bunesco R, Marling C, Shubrook J, Schwartz F. A machine learning approach to predicting blood glucose levels for diabetes management. 2014 Presented at: AAI-14: 2014 Association for the Advancement of Artificial Intelligence Workshop; 2014; Ohio
40. Chan K, Ling S, Dillon T, Nguyen H. Diagnosis of hypoglycemic episodes using a neural network based rule discovery system. *Expert Syst Appl* 2011 Aug 19;38(8):9799-9808 [[FREE Full text](#)] [doi: [10.1016/j.eswa.2011.02.020](https://doi.org/10.1016/j.eswa.2011.02.020)] [Medline: [37860015](#)]
41. Nguyen HT, Jones TW. Detection of nocturnal hypoglycemic episodes using EEG signals. *Annu Int Conf IEEE Eng Med Biol Soc* 2010;2010:4930-4933 [doi: [10.1109/IEMBS.2010.5627233](https://doi.org/10.1109/IEMBS.2010.5627233)] [Medline: [21096665](#)]
42. Rubega M, Scarpa F, Teodori D, Sejling A, Frandsen CS, Sparacino G. Detection of hypoglycemia using measures of EEG complexity in type 1 diabetes patients. *Entropy (Basel)* 2020 Jan 09;22(1):81 [[FREE Full text](#)] [doi: [10.3390/e22010081](https://doi.org/10.3390/e22010081)] [Medline: [33285854](#)]

43. Chen J, Lalor J, Liu W, Druhl E, Granillo E, Vimalananda VG, et al. Detecting hypoglycemia incidents reported in patients' secure messages: using cost-sensitive learning and oversampling to reduce data imbalance. *J Med Internet Res* 2019 Mar 11;21(3):e11990 [FREE Full text] [doi: [10.2196/11990](https://doi.org/10.2196/11990)] [Medline: [30855231](https://pubmed.ncbi.nlm.nih.gov/30855231/)]
44. Jensen MH, Christensen TF, Tarnow L, Seto E, Dencker Johansen M, Hejlesen OK. Real-time hypoglycemia detection from continuous glucose monitoring data of subjects with type 1 diabetes. *Diabetes Technol Ther* 2013 Jul;15(7):538-543 [doi: [10.1089/dia.2013.0069](https://doi.org/10.1089/dia.2013.0069)] [Medline: [23631608](https://pubmed.ncbi.nlm.nih.gov/23631608/)]
45. Skladnev VN, Ghevondian N, Tarnavskii S, Paramalingam N, Jones TW. Clinical evaluation of a noninvasive alarm system for nocturnal hypoglycemia. *J Diabetes Sci Technol* 2010 Jan 01;4(1):67-74 [FREE Full text] [doi: [10.1177/193229681000400109](https://doi.org/10.1177/193229681000400109)] [Medline: [20167169](https://pubmed.ncbi.nlm.nih.gov/20167169/)]
46. Iaione F, Marques JLB. Methodology for hypoglycaemia detection based on the processing, analysis and classification of the electroencephalogram. *Med Biol Eng Comput* 2005 Jul;43(4):501-507 [doi: [10.1007/BF02344732](https://doi.org/10.1007/BF02344732)] [Medline: [16255433](https://pubmed.ncbi.nlm.nih.gov/16255433/)]
47. Bertachi A, Biagi L, Contreras I, Luo N, Vehí J. Prediction of blood glucose levels and nocturnal hypoglycemia using physiological models and artificial neural networks. 2013 Presented at: 3rd International Workshop on Knowledge Discovery in Healthcare Data; July 13, 2018; Stockholm, Sweden
48. Eljil KAAS. Predicting Hypoglycemia in Diabetic Patients Using Machine Learning Techniques. United Arab Emirates: American University of Sharjah; 2014.
49. D'Antoni F, Merone M, Piemonte V, Iannello G, Soda P. Auto-regressive time delayed jump neural network for blood glucose levels forecasting. *Knowl Based Syst* 2020 Sep;203:106134 [doi: [10.1016/j.knosys.2020.106134](https://doi.org/10.1016/j.knosys.2020.106134)]
50. Amar Y, Shilo S, Oron T, Amar E, Phillip M, Segal E. Clinically accurate prediction of glucose levels in patients with type 1 diabetes. *Diabetes Technol Ther* 2020 Aug 01;22(8):562-569 [doi: [10.1089/dia.2019.0435](https://doi.org/10.1089/dia.2019.0435)] [Medline: [31928415](https://pubmed.ncbi.nlm.nih.gov/31928415/)]
51. Li K, Liu C, Zhu T, Herrero P, Georgiou P. GluNet: a deep learning framework for accurate glucose forecasting. *IEEE J Biomed Health Inform* 2020 Feb;24(2):414-423 [doi: [10.1109/jbhi.2019.2931842](https://doi.org/10.1109/jbhi.2019.2931842)]
52. Zecchin C, Facchinetti A, Sparacino G, De Nicolao G, Cobelli C. Neural network incorporating meal information improves accuracy of short-time prediction of glucose concentration. *IEEE Trans Biomed Eng* 2012 Jun;59(6):1550-1560 [doi: [10.1109/TBME.2012.2188893](https://doi.org/10.1109/TBME.2012.2188893)] [Medline: [22374344](https://pubmed.ncbi.nlm.nih.gov/22374344/)]
53. Mohebbi A, Johansen AR, Hansen N, Christensen PE, Tarp JM, Jensen ML, et al. Short term blood glucose prediction based on continuous glucose monitoring data. *Annu Int Conf IEEE Eng Med Biol Soc* 2020 Jul;2020:5140-5145 [doi: [10.1109/EMBC44109.2020.9176695](https://doi.org/10.1109/EMBC44109.2020.9176695)] [Medline: [33019143](https://pubmed.ncbi.nlm.nih.gov/33019143/)]
54. Daniels J, Herrero P, Georgiou P. A multitask learning approach to personalized blood glucose prediction. *IEEE J Biomed Health Inform* 2022 Jan;26(1):436-445 [doi: [10.1109/JBHI.2021.3100558](https://doi.org/10.1109/JBHI.2021.3100558)] [Medline: [34314367](https://pubmed.ncbi.nlm.nih.gov/34314367/)]
55. Alfian G, Syafrudin M, Anshari M, Benes F, Atmaji F, Fahrurrozi I, et al. Blood glucose prediction model for type 1 diabetes based on artificial neural network with time-domain features. *Biocybern Biomed Eng* 2020 Oct;40(4):1586-1599 [FREE Full text] [doi: [10.1016/j.bbe.2020.10.004](https://doi.org/10.1016/j.bbe.2020.10.004)]
56. Dave D, DeSalvo DJ, Haridas B, McKay S, Shenoy A, Koh CJ, et al. Feature-based machine learning model for real-time hypoglycemia prediction. *J Diabetes Sci Technol* 2021 Jul 01;15(4):842-855 [FREE Full text] [doi: [10.1177/1932296820922622](https://doi.org/10.1177/1932296820922622)] [Medline: [32476492](https://pubmed.ncbi.nlm.nih.gov/32476492/)]
57. Marcus Y, Eldor R, Yaron M, Shaklai S, Ish-Shalom M, Shefer G, et al. Improving blood glucose level predictability using machine learning. *Diabetes Metab Res Rev* 2020 Nov 14;36(8):e3348 [doi: [10.1002/dmrr.3348](https://doi.org/10.1002/dmrr.3348)] [Medline: [32445286](https://pubmed.ncbi.nlm.nih.gov/32445286/)]
58. Reddy R, Resalat N, Wilson LM, Castle JR, El Youssef J, Jacobs PG. Prediction of hypoglycemia during aerobic exercise in adults with type 1 diabetes. *J Diabetes Sci Technol* 2019 Sep;13(5):919-927 [FREE Full text] [doi: [10.1177/1932296818823792](https://doi.org/10.1177/1932296818823792)] [Medline: [30650997](https://pubmed.ncbi.nlm.nih.gov/30650997/)]
59. Sampath S, Tkachenko P, Renard E, Pereverzev SV. Glycemic control indices and their aggregation in the prediction of nocturnal hypoglycemia from intermittent blood glucose measurements. *J Diabetes Sci Technol* 2016 Nov;10(6):1245-1250 [FREE Full text] [doi: [10.1177/1932296816670400](https://doi.org/10.1177/1932296816670400)] [Medline: [27660190](https://pubmed.ncbi.nlm.nih.gov/27660190/)]
60. Sudharsan B, Peeples M, Shomali M. Hypoglycemia prediction using machine learning models for patients with type 2 diabetes. *J Diabetes Sci Technol* 2015 Jan;9(1):86-90 [FREE Full text] [doi: [10.1177/1932296814554260](https://doi.org/10.1177/1932296814554260)] [Medline: [25316712](https://pubmed.ncbi.nlm.nih.gov/25316712/)]
61. Nuryani N, Ling SSH, Nguyen HT. Electrocardiographic signals and swarm-based support vector machine for hypoglycemia detection. *Ann Biomed Eng* 2012 Apr;40(4):934-945 [doi: [10.1007/s10439-011-0446-7](https://doi.org/10.1007/s10439-011-0446-7)] [Medline: [22012087](https://pubmed.ncbi.nlm.nih.gov/22012087/)]
62. San PP, Ling SH, Nuryani N, Nguyen H. Evolvable rough-block-based neural network and its biomedical application to hypoglycemia detection system. *IEEE Trans Cybern* 2014 Aug;44(8):1338-1349 [doi: [10.1109/TCYB.2013.2283296](https://doi.org/10.1109/TCYB.2013.2283296)] [Medline: [24122616](https://pubmed.ncbi.nlm.nih.gov/24122616/)]
63. Ling SH, Nguyen HT. Natural occurrence of nocturnal hypoglycemia detection using hybrid particle swarm optimized fuzzy reasoning model. *Artif Intell Med* 2012 Jul;55(3):177-184 [doi: [10.1016/j.artmed.2012.04.003](https://doi.org/10.1016/j.artmed.2012.04.003)] [Medline: [22698854](https://pubmed.ncbi.nlm.nih.gov/22698854/)]
64. Ling SH, San PP, Nguyen HT. Non-invasive hypoglycemia monitoring system using extreme learning machine for type 1 diabetes. *ISA Trans* 2016 Sep;64:440-446 [doi: [10.1016/j.isatra.2016.05.008](https://doi.org/10.1016/j.isatra.2016.05.008)] [Medline: [27311357](https://pubmed.ncbi.nlm.nih.gov/27311357/)]
65. Nguyen LB, Nguyen AV, Ling SH, Nguyen HT. An adaptive strategy of classification for detecting hypoglycemia using only two EEG channels. *Annu Int Conf IEEE Eng Med Biol Soc* 2012;2012:3515-3518 [doi: [10.1109/EMBC.2012.6346724](https://doi.org/10.1109/EMBC.2012.6346724)] [Medline: [23366685](https://pubmed.ncbi.nlm.nih.gov/23366685/)]

66. Ngo CQ, Chai R, Nguyen TV, Jones TW, Nguyen HT. Electroencephalogram spectral moments for the detection of nocturnal hypoglycemia. *IEEE J Biomed Health Inform* 2020 May;24(5):1237-1245 [doi: [10.1109/JBHI.2019.2931782](https://doi.org/10.1109/JBHI.2019.2931782)] [Medline: [31369389](https://pubmed.ncbi.nlm.nih.gov/31369389/)]
67. Ngo CQ, Truong BCQ, Jones TW, Nguyen HT. Occipital EEG activity for the detection of nocturnal hypoglycemia. *Annu Int Conf IEEE Eng Med Biol Soc* 2018 Jul;2018:3862-3865 [doi: [10.1109/EMBC.2018.8513069](https://doi.org/10.1109/EMBC.2018.8513069)] [Medline: [30441206](https://pubmed.ncbi.nlm.nih.gov/30441206/)]
68. Nuryani N, Ling SH, Nguyen HT. Hypoglycaemia detection for type 1 diabetic patients based on ECG parameters using fuzzy support vector machine. 2010 Presented at: IJCNN 2010: 2010 International Joint Conference on Neural Networks; July 18-23, 2010; Barcelona, Spain [doi: [10.1109/ijcnn.2010.5596916](https://doi.org/10.1109/ijcnn.2010.5596916)]
69. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *JAMA* 1994 Mar 02;271(9):703-707 [doi: [10.1001/jama.271.9.703](https://doi.org/10.1001/jama.271.9.703)] [Medline: [8309035](https://pubmed.ncbi.nlm.nih.gov/8309035/)]
70. Kodama S, Fujihara K, Shiozaki H, Horikawa C, Yamada MH, Sato T, et al. Ability of current machine learning algorithms to predict and detect hypoglycemia in patients with diabetes mellitus: meta-analysis. *JMIR Diabetes* 2021 Jan 29;6(1):e22458 [FREE Full text] [doi: [10.2196/22458](https://doi.org/10.2196/22458)] [Medline: [33512324](https://pubmed.ncbi.nlm.nih.gov/33512324/)]
71. McShinsky R, Marshall B. Comparison of forecasting algorithms for type 1 diabetic glucose prediction on 30 and 60-minute prediction horizons. 2020 Presented at: KDH@ECAI 2020: 5th International Workshop on Knowledge Discovery in Healthcare Data co-located with 24th European Conference on Artificial Intelligence; August 29-30, 2020; Santiago de Compostela, Spain, and virtually
72. Deng Y, Lu L, Aponte L, Angelidi AM, Novak V, Karniadakis GE, et al. Deep transfer learning and data augmentation improve glucose levels prediction in type 2 diabetes patients. *NPJ Digit Med* 2021 Jul 14;4(1):109 [FREE Full text] [doi: [10.1038/s41746-021-00480-x](https://doi.org/10.1038/s41746-021-00480-x)] [Medline: [34262114](https://pubmed.ncbi.nlm.nih.gov/34262114/)]
73. Van Calster B, Steyerberg EW, Wynants L, van Smeden M. There is no such thing as a validated prediction model. *BMC Med* 2023 Feb 24;21(1):70 [FREE Full text] [doi: [10.1186/s12916-023-02779-w](https://doi.org/10.1186/s12916-023-02779-w)] [Medline: [36829188](https://pubmed.ncbi.nlm.nih.gov/36829188/)]

Abbreviations

ARM: autoregression model
ARJNN: ARTiDe jump neural network
AUC: area under the curve
BG: blood glucose
CGM: continuous glucose monitoring
DM: diabetes mellitus
DRNN: dilated recurrent neural network
DT: decision tree
ECG: electrocardiograph
EEG: electroencephalograph
EHR: electronic health record
ML: machine learning
NLR: negative likelihood ratio
NNM: neural network model
PH: prediction horizon
PLR: positive likelihood ratio
QUADAS-2: Quality Assessment of Diagnostic Accuracy Studies
RF: random forest
RMSE: root mean square error
SROC: summary receiver operating characteristic
SUCRA: surface under the cumulative ranking
SVM: support vector machine
T1DM: type 1 diabetes mellitus
T2DM: type 2 diabetes mellitus
XGBoost: Extreme Gradient Boosting

Edited by C Lovis; submitted 03.04.23; peer-reviewed by C Toffanin, S Lee; comments to author 30.07.23; revised version received 21.08.23; accepted 12.10.23; published 20.11.23

Please cite as:

Liu K, Li L, Ma Y, Jiang J, Liu Z, Ye Z, Liu S, Pu C, Chen C, Wan Y

Machine Learning Models for Blood Glucose Level Prediction in Patients With Diabetes Mellitus: Systematic Review and Network Meta-Analysis

JMIR Med Inform 2023;11:e47833

URL: <https://medinform.jmir.org/2023/1/e47833>

doi: [10.2196/47833](https://doi.org/10.2196/47833)

PMID:

©Kui Liu, Linyi Li, Yifei Ma, Jun Jiang, Zhenhua Liu, Zichen Ye, Shuang Liu, Chen Pu, Changsheng Chen, Yi Wan. Originally published in JMIR Medical Informatics (<https://medinform.jmir.org>), 20.11.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on <https://medinform.jmir.org/>, as well as this copyright and license information must be included.