Journal of Associated Medical Sciences 2026; 59(2): 72-83

JAMS
Journal of Associated Medical Sciences

Scopus Indexed & Thai-Journal Citation Index Centre (TCI)

Journal of Associated Medical Sciences

Journal homepage: https://www.tci-thaijo.org/index.php/bulletinAMS/index

Comprehensive benchmarking of machine learning models for blood glucose classification and prediction: new approach for improved hyperglycemia and hypoglycemia detection and prediction

Houda El Bouhissi[1*], Tatiana Ermakova[2], Rabie A. Ramadan[3], Djamila Ouaret[1]

[1]*University of Bejaia, Faculty of Exact Sciences, LIMED Laboratory, Bejaia, Algeria.*
[2]*School of Computing, Communication and Business, Hochschule für Technik und Wirtschaft, University of Applied Sciences for Engineering and Economics, Germany.*
[3]*Department of Information Systems, College of Economics, Management, and Information Systems, Nizwa, Sultanate of Oman.*

ARTICLE INFO

Article history:
Received 22 August 2025
Accepted as revised 9 January 2026
Available online 27 January 2026

Keywords:
Diabetes, CRNN, CNN, LSTM, bioinspired algorithm.

ABSTRACT

Background: Diabetes mellitus affects 463 million people worldwide and necessitates continuous blood glucose monitoring. Current glucose prediction systems often lack efficiency, and real-time prediction is essential for timely clinical intervention.

Objectives: This study aims to develop and validate a novel Convolutional Recurrent Neural Network (CRNN) enhanced with bio-inspired algorithms to improve blood glucose prediction and enable real-time detection of hypoglycemia and hyperglycemia.

Materials and methods: The proposed framework employs a CRNN architecture that combines Convolutional Neural Networks (CNNs) for feature extraction with Long Short-Term Memory (LSTM) layers for temporal sequence learning. The model was trained and evaluated using the HUPA-UCM diabetes dataset. Additionally, the study benchmarks the proposed model against 19 traditional Machine Learning (ML) algorithms and compares it with state-of-the-art methods from the literature.

Results: The proposed approach demonstrates superior predictive capability, consistently delivering promising results across multiple evaluation frameworks. The model achieves clinically acceptable prediction intervals, confirming its effectiveness in enhancing the accuracy and reliability of blood glucose prediction for diabetes management.

Conclusion: The findings demonstrate that the proposed CRNN model, enhanced with bio-inspired algorithms, provides an effective and reliable solution for real-time blood glucose prediction. By outperforming conventional ML methods and achieving clinically acceptable accuracy levels, the model shows strong potential for integration into intelligent diabetes management systems to support timely clinical decisions and improve patient outcomes.

* **Corresponding contributor.**
Author's Address: *University of Bejaia, Faculty of Exact Sciences, LIMED Laboratory, Bejaia, Algeria.*
E-mail address: houda.elbouhissi@gmail.com
doi: 10.12982/JAMS.2026.038
E-ISSN: 2539-6056

Introduction

Medical background

Diabetes is a complex metabolic disorder that has become a major global health challenge. In 2014, it affected over 8.5% of adults aged 18 and older.[1] The disease is characterized by impaired blood glucose regulation and is associated with a high risk of complications, including cardiovascular disease, nephropathy, retinopathy, and neuropathy.[2,3] In 2019,

diabetes accounted for 1.5 million deaths worldwide, nearly half of which occurred before the age of 70; an additional 460,000 kidney disease-related deaths and approximately 20% of cardiovascular deaths were attributable to hyperglycemia.[4] Diabetes develops due to insufficient insulin secretion or reduced tissue sensitivity to insulin. Insulin is produced by pancreatic β-cells and facilitates glucose transport into cells, where it is used for energy or stored as glycogen.[5] Insulin deficiency or resistance leads to hyperglycemia, whereas excessive secretion can cause hypoglycemia.[6] In non-diabetic individuals, fasting glucose is <1.26 gm/L (7 mmol/L); values of 1.26-2 gm/L indicate prediabetes, and 2 gm/L confirm diabetes.[7-9] An HbA1c level <7% is widely recommended to minimize the risk of long-term complications.[10,11] Two main types of diabetes are recognized: type 1 (T1D) and type 2 (T2D). T1D results from autoimmune destruction of β-cells, leading to absolute insulin deficiency; T2D is characterized by insulin resistance and reduced secretion. Less common forms include gestational diabetes and other specific types.[12,13] This study focuses on T1D, the most severe form, often developing in childhood or adolescence and requiring lifelong insulin therapy.[14]

Motivation and research objectives

Continuous blood glucose monitoring is required for effective management.[15] Previous studies have combined CNNs and LSTM networks, for example, in CNN-LSTM hybrid architectures and two-headed LSTM models.[16] Dilated RNNs (DRNNs) have also been explored, but there has been no explicit application of Convolutional Recurrent Neural Networks (CRNNs) targeting both hypo- and hyperglycemia in real time.[17] Indeed, single models are primarily focused on either temporal dependencies or localized glucose patterns but rarely integrate both aspects in an effective manner. RNNs can capture long-term trends for modeling temporal dependencies.

This study integrates both and presents a CRNN-based predictive model enhanced with a bio-inspired optimizer, applied to real-time glucose monitoring. CRNNs are particularly well suited for sequential data such as time series and can learn both short- and long-term dependencies, making them ideal for detecting hyper-and hypoglycemia using time-series data. While CRNNs are not entirely new, having been initially used in speech recognition and video processing, their application to healthcare and glucose prediction remains relatively recent. They have shown promise in biomedical signal analysis, including ECG, EEG, and continuous glucose monitoring (CGM) data. In addition, researchers have widely adopted bioinspired algorithms for optimizing tasks such as CNNs. These algorithms efficiently navigate high dimensional hyperparameter spaces without resorting to exhaustive and computationally expensive exploration. Unlike many previous studies that focus solely on blood glucose prediction, this work also addresses classification of hypo and hyperglycemia in real time, which is an essential feature for practical deployment. This study develops and validates its model using the recent HUPA-UCM dataset[18] which provides rich, multivariate time series data including CGM readings and insulin doses, thus offering a more comprehensive input space for modeling. The research also includes extensive benchmarking, following prior works such as Tena *et al*.,[19] who compared two ensemble networks with 10 others; Ruan *et al*.,[20] who evaluated 18 Machine Learning (ML) algorithms including logistic regression, SGD, k-nearest neighbors, decision tree, Gaussian naive Bayes, Bernoulli NB, multinomial NB, support vector machine, quadratic discriminant analysis, random forest, extra trees, linear discriminant analysis, AdaBoost, bagging, gradient boosting, XGBoost, and multi-layer perceptron; and Shi *et al*., who compared six algorithms including logistic regression, random forest, gradient boosting machine, deep neural networks, XGBoost, and Rulefit.[21] Furthermore, the approach emphasized in this study is predicated on clinically relevant prediction horizons and actionable alerts, with a view to aligning model outputs with the practical management needs of individuals suffering from diabetes. This study makes a significant contribution by addressing methodological shortcomings and enhancing the applicability of predictive models in real-world settings, thereby advancing the goal of personalized and proactive diabetes care. Hypoglycemia classification was evaluated for two thresholds: mild (BG<4.0 mmol/L) and severe (BG<3.0 mmol/L), consistent with clinical definitions. The classification metrics (Precision, Recall, F1, Accuracy, AUROC) were computed for each threshold using the predicted labels vs. true labels. Regression performance for continuous glucose values was evaluated with MAE, RMSE, and R^2 on the predicted 30-min-ahead values.

The paper is organized as follows: Section 2 reviews related work. Section 3 outlines the proposed methodology, Section 4 presents and discusses the results, and Section 5 concludes the paper.

Related work

Tsichlaki *et al*. conducted a systematic literature for hypoglycemia detection in T1D patients.[22] In the following, we present the most important related works according to hyperglycemia and hypoglycemia into three main categories: DL, ML and reinforcement-based approaches.

Deep learning

Jaloli and Cescon propose a CNN-LSTM model to predict blood glucose (BG) level in people with T1D for 30-, 60, and 90-minute prediction horizons (PH), based on historical glucose readings, meal data, and insulin doses.[16] They evaluated the approach on two datasets: Replace-BG, reflecting free-living conditions, and DIAdvisor, representing an in- hospital setting. Aiello *et al*. introduced a two-headed LSTM-based Deep Glucose Forecasting approach for predicting interstitial

glucose levels in people with T1D, using CGM data, carbohydrate intake, and recommended insulin therapy.[23] The architecture was trained on data from 100 virtual adult patients generated by the UVA/Padova simulator and was evaluated on both virtual and real patient datasets, demonstrating strong generalization performance on previously unseen data. Zhu *et al*. presented a dilated recurrent neural network (DRNN) model to generate 30-minute forecasts of glucose levels in individuals with T1DM.[17]

Additionally, they apply transfer learning to leverage data from multiple subjects. On the OhioT1DM dataset, their model achieved a root mean square error (RMSE) of 18.9 mg/dL, outperforming autoregressive models (ARX, 20.1 mg/dL), support vector regression (SVR, 21.7 mg/dL), and conventional neural networks for predicting glucose (NNPG, 22.9 mg/dL). Tena *et al*.[19] introduced two ensemble neural network models for blood glucose prediction in people with T1DM at three prediction horizons (PHs)-30, 60, and 120 minutes and compares their performance against ten recently proposed neural networks. All these models are evaluated on the same OhioT1DM dataset. Karagoz *et al*. presented a comparative analysis of transformer models for multi-horizon BG prediction, evaluating forecasts up to 4 hours ahead using input histories of up to one week.[24] They used the publicly available DCLP3 dataset (N=112), split 80%-10%-10% for training, validation, and testing, and the OhioT1DM dataset (N=12) as an external test set. For short-term prediction, the Crossformer model, a patch-wise transformer architecture, achieved the best 30-minute forecast with an RMSE of 15.6 mg/dL on the OhioT1DM dataset. Munoz-Organero reviewed the current state of the art in blood glucose prediction for T1DM patients and proposed, implemented, and validated a novel hybrid model.[25] Specifically, the differential equations describing carbohydrate and insulin absorption are modeled using a RNN implemented with LSTM cells. The results show promising performance. Alvarado *et al*. presented a method for predicting and detecting hypoglycemic events over a 24-hour time PH.[26] The approach combines wavelet transform analysis of glucose time series with CNNs and was validated using real-world data from 20 individuals with T1D. The results demonstrate strong performance, with accuracy, sensitivity, specificity, and precision all exceeding 88. Mirshekarian *et al*. presented an RNN approach using LSTM units to learn a physiological model of blood glucose.[27] Sun *et al*. introduced a sequential model consisting of one LSTM layer, one bidirectional LSTM layer, and several fully connected layers to predict blood glucose levels across multiple PHs.[28] The model was trained and tested on 26 retrospectively analyzed datasets from 20 real patients. Idrissi *et al*. presented sequential neural network architecture with an LSTM layer followed by two fully connected layers for blood glucose prediction.[29] Another approach proposed by Iacono *et al*. highlights the importance of warning systems while examining the significant problem of preventing hypoglycemia and hyperglycemia in the management of T1D.[30] The proposed work has investigated the prediction of future blood glucose levels using DL models, such as personalized LSTM models. Using LSTM models, the individualized alarm system demonstrates promising predictive performance within 40-minutes prediction horizon, achieving an RMSE of 6.45. It attains an F-Score of 78/79% for hypoglycemia detection and 83.87% for hyperglycemia detection. In addition, the work proposed by Li et al. investigates the use of CRNN model for blood glucose level forecasting in the management of type 1 diabetes.[31] The proposal combined a CNN to extract robust features from tile-aligned inputs sauch as CGM readings, insulin boluses, and meals, with an RNN to capture temporal dynamics and generate predictions in 30 and 60 minutes ahead.

Machine learning and statistical models

Kavakiotis *et al*. explored the use of data mining and ML in diabetes research, highlighting the critical role these techniques play in turning massive volumes of data, including genetic and clinical data from Electronic Health Records (EHRs), into useful insights.[32] The authors discovered that ML algorithms, in particular supervised learning techniques (85% of the cases), were widely used in diabetes research through a systematic literature review (focusing on articles from 2011 to 2016). In addition, Rghioui *et al*. presented an innovative system for remotely monitoring diabetes patients, leveraging contemporary technology like artificial intelligence (AI) and smart devices to optimize monitoring procedures and minimize related costs.[33] The results show that the J48 algorithm has a remarkable 99.17% accuracy, 99.47% sensitivity, and 99.32% precision in classification. In addition, Kodama *et al*. looked at ML systems trained for hypoglycemia diagnosis or prediction.[34] The findings showed that while current ML algorithms (SVM, XG-Boost, RF, LR, and ANN) showed moderate potential in predicting impending hypoglycemia, with sensitivity and specificity averaging 0.80, they were limited in their capacity to detect continuing hypoglycemia. In addition, the research conducted by Shi *et al*. aimed to create an ML model that could forecast a probability that older persons with diabetes would experience severe hypoglycemia (SH) and require hospitalization.[21] To train the ML model, they selected 258 predictors related to medical history, drugs, laboratory tests, and demographics using EHR data from a sizable sample of older persons in Hong Kong. Six distinct ML algorithms were evaluated for performance, and the findings showed that the XG-Boost model performed the best (AUROC=0.978). This research confirms that older people who are at a high risk of experiencing severe hypoglycemia could benefit from preventive intervention by having this ML model integrated into electronic health record systems.

Reinforcement learning

The purpose of the study of Wang *et al*. is to enhance individualized insulin titration in the treatment of T2D

using a model-based reinforcement learning framework called RL-DITR.[35] This framework analyzes the benefits of glycemic status using model-patient interaction to determine the ideal insulin prescription. In the development stage, RL-DITR achieved a mean absolute error of 1.10±0.03 U, which was superior to other DL models and conventional clinical techniques for optimizing insulin titration. With a mean absolute error of 1.18±0.09 U, a step-by-step clinical validation showed that inpatient glycemic management was improved in comparison to junior and intermediate physicians.

To summarize, recent studies on predicting hyperglycemia and hypoglycemia have explored various ML and DL approaches. CGM forecasting is dominated by temporal deep models, such LSTM and CRNN. CRNNs, for instance, have demonstrated impressive performance by integrating convolutional layers for local pattern extraction with recurrent layers for modeling temporal dependencies, achieving a low root mean square error (RMSE) at 30-60-minute horizons. Overall, deep temporal architecture remains the gold standard for CGM- based short-term prediction, hybrid models offer incremental improvements, and interpretable models remain valuable in safety-critical clinical contexts.

Materials and methods

CRNNs have shown strong potential in blood glucose prediction by combining convolutional layers (for spatial feature extraction) with recurrent layers (for modeling temporal dependencies). This integration enables better modeling of both localized glucose fluctuations and long-term trends, outperforming models that address only one aspect. Hyperparameter tuning using metaheuristics such as Particle Swarm Optimization (PSO), Grey Wolf Optimizer (GWO), Cuckoo Search (CS) and Ant Colony Optimization (ACO) can further enhance performance.[36, 37]

The overall workflow of the proposed approach is presented in Figure 1 and involves four main steps: “Data description” where we introduce the used dataset by outlining its source, size, features, and target variables relevant to the study. “Data pre-processing” which is a crucial step for data cleaning, transforming, and normalizing to ensure consistency and suitability for model training. “Optimization” where GWO optimization algorithm is applied to fine-tune parameters or select features that enhance model performance. And finally, “Proposed CRNN architecture” The designed Convolutional Recurrent Neural Network integrates convolutional layers for feature extraction and recurrent layers for temporal pattern learning.

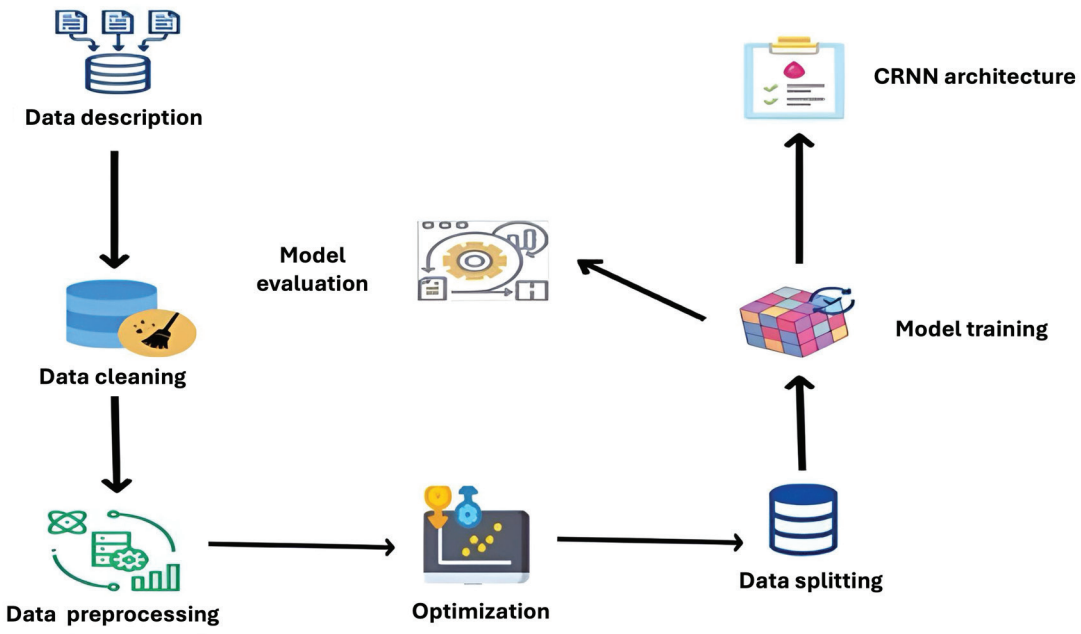

Figure 1. *Proposed approach architecture.*

Data description

Although our study relies on a single dataset HUPA-UCM diabetes dataset first made available in April 2024, comprises 309,392 rows and 8 columns (see Table 1).[17] This choice is justified by the fact that the dataset is recent, comprehensive, and high-quality, containing many samples and rich contextual information. Its size and diversity provide sufficient variability for robust model training and evaluation, reducing the risk of overfitting and enabling reliable generalization within the target population. Moreover, no other publicly available dataset offers the same level of completeness or alignment as the specific hyperglycemia-prediction task addressed in this work.

The dataset aggregates data from 25 type 1 diabetes (T1D) patients aged 18 to 65 years, collected between June 13, 2018, and May 18, 2022. Each record is time-stamped at 5-minute intervals, providing continuous time-series monitoring over several weeks per patient. The dataset contains glucose measurements obtained from a Continuous Glucose Monitor (CGM), physical activity indicators (steps, calories burned, and heart rate), insulin injections (basal rates and bolus volumes), and carbohydrate intake during the interval. Notably, the dataset shows initial glucose values above 300 mg/dL and later glucose values in the range of 110-128 mg/dL. The data enables forecasting tasks such as predicting blood glucose levels 30-60 minutes in advance and analyzing responses to diet and treatment. The temporal variable was converted to minutes to ensure consistency. No missing values are present, making the dataset well-suited for analytical and predictive tasks. The average glucose concentration is 141.43 mg/dL, with a standard deviation of 57.09, and values ranging from 40 to 444 mg/dL-indicating substantial glycemic variability. The average heart rate is 76.99 bpm (SD=85), and the mean number of steps per interval is 30.83. Mean energy expenditure is 8.81 calories per interval. In contrast, insulin delivery parameters (basal and bolus) and carbohydrate intake exhibit lower average values, consistent with typical clinical dosing and meal patterns.

Table 1. HUPA-UCM dataset description.

Attribute	Description
Time	Timestamp of each measurement in the format yyyy-MM-dd'T'HH:mm:ss
Glucose	Blood glucose concentration (mg/dL)
Calories	Calories burned during the time interval
Heart_rate	Heart rate (beats per minute)
Steps	Number of steps taken during the time interval
Basal_rate	Amount of basal insulin infused during the interval
Bolus_volume_delivered	Volume of bolus insulin injected during the interval
Carb_input	Carbohydrate intake during the interval (1 serving = 10 gm)

Data pre-processing

At this step, the dataset is prepared to ensure its quality and suitability for the modeling process. First, the missing values are identified and addressed using a statistical imputation method, the mean. Statistical measures such as the inter- quartile range (IQR) or z-score are used to detect outliers. These are then either removed or adjusted to reduce their influence. All variables are made comparable using either standardization or normalization techniques, which are used to perform feature scaling. This improves both model convergence and performance. Finally, the processed data is split into training and testing subsets, enabling the model to be evaluated impartially.

Optimization

Metaheuristic algorithms have been extensively adopted for optimizing CNNs due to their capacity to navigate high-dimensional hyperparameter spaces efficiently without resorting to exhaustive and computationally expensive searches. Recent studies have shown how well various metaheuristic algorithms can be used to optimize CNNs for healthcare tasks which improved accuracy and reduced error rates through optimized parameter selection. We have used GWO for our proposal because it is widely used for medical and prediction tasks. GWO optimization algorithm is a metaheuristic proposed by which mimics the leadership hierarchy and hunting mechanism of grey wolves in nature.[37] It is based on the hierarchy-based hunting characteristics of the Canidae family. The GWO principles focus on hunting behaviors highlight the dynamic nature of addressing convergence concerns. GWO algorithm uses the dynamic features of the Canidae population to solve this problem. This enhances the algorithm's capacity to attain superior global optimal convergence compared to other optimization techniques that often converge with local optimal solutions. Additionally, it fosters enhanced stability between the exploitation and exploration phases. When grey wolves are present, the hierarchy-based hunting phenomenon involves a leading grey wolf, a representative grey wolf, an obeying grey wolf and a scapegoat grey wolf. The leadership

hierarchy is simulated using four types of grey wolf: alpha, beta, delta and omega. In addition, three main steps of hunting are implemented.

These are searching for prey, encircling prey and attacking prey. These steps are used to perform optimization. The purpose of this step is to optimize the hyperparameters to achieve good accuracy in reduced time. In GWO, candidate solutions are modeled as wolves, and the best three solutions so far are considered α, β, and δ. Wolves encircle prey using position update equations that simulate hunting strategies. Two key coefficient vectors (A and C) control the movement toward prey and the balance between exploration and exploitation. The encircling is modeled by computing distances from each wolf to α, β, and δ, then estimating three possible positions. The new position is the average of these three positions, encouraging convergence toward optimal solutions. The parameter **a** decrease linearly from 2 to 0 over iterations, gradually shifting the search from exploration to exploitation. At each iteration, fitness is evaluated, and the top three wolves are updated.

Finally, the algorithm stops after a fixed number of iterations, returning the best-found solution (α). The hyperparameter search space for GWO was defined based on commonly reported ranges in the literature and guided by preliminary experiments to ensure efficient exploration and stable convergence. Specifically:

- Population size: 20-40 wolves, balancing search diversity and computational cost.
- Number of iterations: 30-50, allowing convergence without excessive runtime.
- Search bounds for model hyperparameters:
- Learning rate: 0.0001-0.01
- Number of neurons per layer: 16-256
- Batch size: 16-128

These ranges were chosen to cover realistic and meaningful values while avoiding unstable regions. We will clarify these details in the revised manuscript to ensure full transparency and reproducibility.

In this study, GWO is used exclusively for hyper parameter optimization rather than for training the model's weights. The CNN-LSTM hyperparameters—such as the number of filters, kernel size, LSTM units, dropout rate, and window length—are non-differentiable and form a highly non-convex search space, making gradient-based optimizers unsuitable. GWO provides a gradient-free global search strategy that can effectively explore this complex space and identify better-performing configurations, while the internal network weights are still trained using a standard gradient-based optimizer

Proposed CRNN architecture

CNN and RNN are combined into a powerful hybrid DL model. A main application of CNNs aims to learn to extract features from sequential data. CRNNs follow exactly this approach, as the connections of a CNN within the model learn relevant features in a timeseries input in a self-supervised manner. RNNs, mainly LSTMs: RNNs are used for modelling temporal dependencies that are part of the series. LSTMs solve the problem of vanishing gradients in RNNs so that long-term dependencies can be modeled. As shown in Figure 2, the CRNN model is built with a feedback mechanism for glucose prediction. The model receives the data via the input layer. This input passes through the CNN components. These components are designed to ex back mechanism refines the prediction process over time and allows the model to be used for real-time predictions.

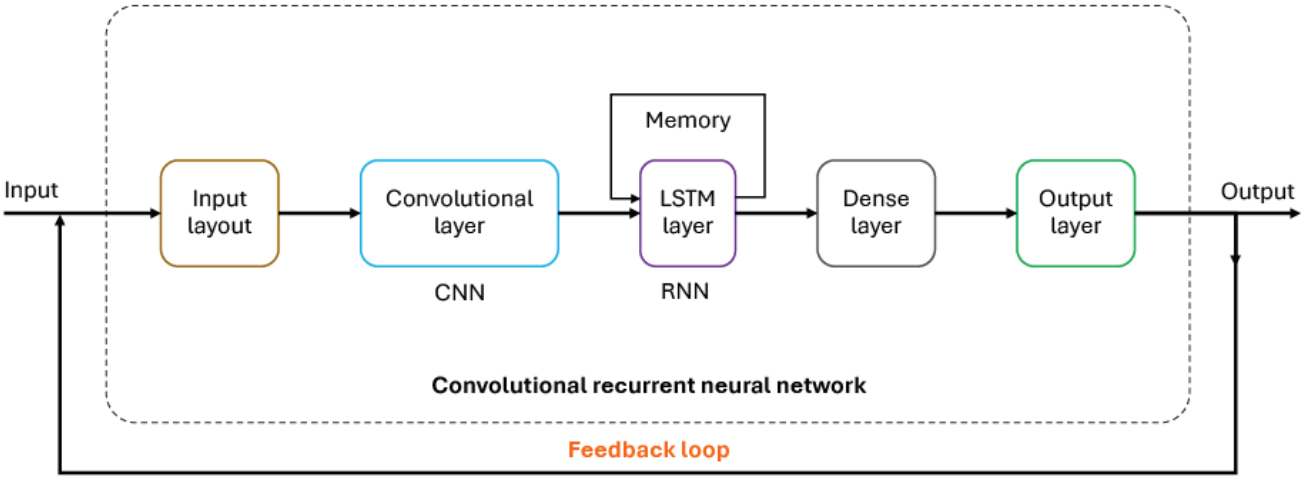

Figure 2. *The proposed CRNN architecture.*

Metrics

The following metrics were applied as the most used for evaluating classification models:

- Mean Absolute Error (MAE): This metric measures the average magnitude of errors without considering their direction.
- Mean Squared Error (MSE): This metric emphasizes larger errors by squaring the differences.
- Root Mean Squared Error (RMSE): These metric computes error in the same units as the target variable.
- Precision: The proportion of correctly predicted positive instances among all predicted positives.
- Accuracy measures the proportion of correctly predicted instances (both positive and negative) out of the total predictions made. It reflects the overall correctness of a classification model.
- Recall: measures the proportion of actual positive cases that are correctly identified by the model. It indicates how well the model detects true positives.
- F1-score: The F1-score is the harmonic mean of Precision and Recall, providing a balanced measure of a model's accuracy when both false positives and false negatives are important. It is especially useful for evaluating models on imbalanced datasets.
- Clarke Error Grid is a metric used to evaluate blood glucose predictions by classifying each predicted value based on its clinical impact on the patient. It distinguishes acceptable errors from potentially dangerous ones, assessing the safety of a model or device.

Results

In this section, the performance of the proposed model is evaluated in comparison to different baseline models in terms of hypoglycemia detection (classification) and prediction of continuous glucose values (regression). The training performance and model stability are analyzed. In addition, the current literature model is compared with clinical validations. The proposed model is trained and tested on the HUPA-UCM diabetes dataset. The model integrates layer-wise CNN layers for extracting short-term patterns and LSTM layers for capturing long-term dependencies in the time series data. The proposed architecture, as shown in Figure 3, is based on a 1D convolutional layer with 64 filters and a kernel size of 3. These layers process the input data, including the glucose measurements and the associated health metrics. This is followed by a max-pooling layer with a size of 2 to reduce dimensionality and explore salient features. In addition, the sequential structure contains two LSTM layers, with each layer consisting of 50 slices. The first 50 layers provide the sequences, and the second 50 layers provide the final output; they are designed to capture the temporal dependencies at different levels.

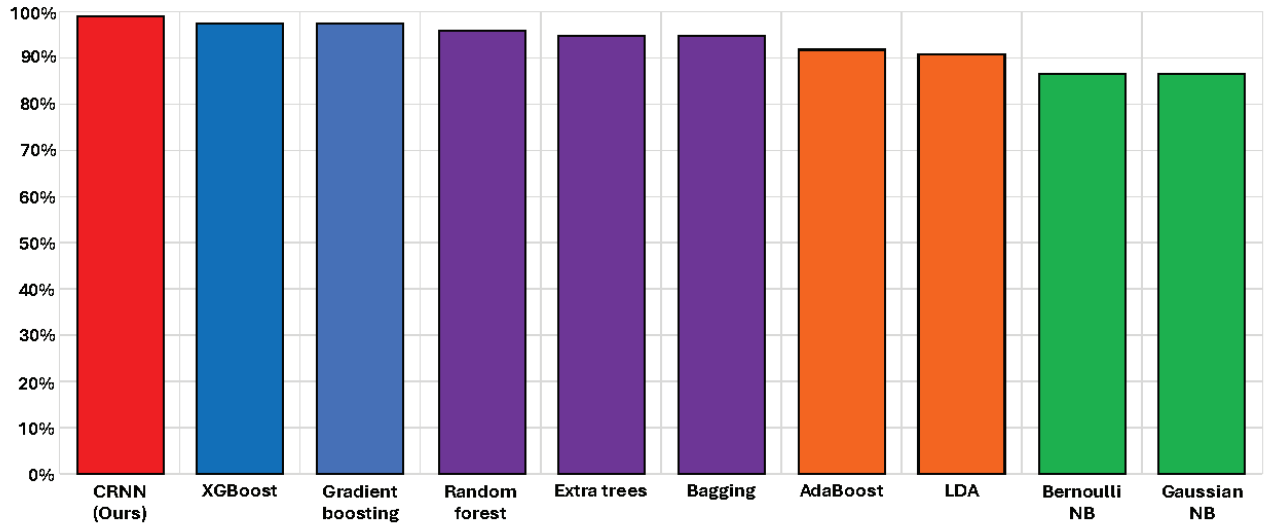

Figure 3. *Top 10 classification models.*

The proposed model (CRNN+GWO) is evaluated against 19 baseline ML algorithms and recent state-of-the-art methods demonstrating strong performance in both classification and regression tasks. As shown in Table 2, the proposal consistently outperforms all traditional ML approaches. As can be seen on Table 2, our model achieves improved precision (0.97), recall (0.98), and F1-score (0.98), which is a significant improvement over traditional ML approach. These improvements lead to an accuracy of 99%, outperforming all compared algorithms. This sets a new bench- mark for the classification of glucose events. In addition, the AUROC analysis showed that the hybrid model has a performance of 0.98, while the classification accuracy of CRNN improved by 1% compared to the best-performing ensemble methods (XGBoost and Gradient Boosting with 97%). Figure 3 shows the 10 most accurate models. The performance of ten models in the classification of hypoglycemia was compared with the proposed model demonstrating superiority in both accuracy and robustness. The area under the curve (AUROC) value for CRNN is 0.97, and the precision, recall, and F1 scores for both mild (BG<4) and severe (BG<3) hypoglycemia are all above 0.96. These results indicate that the proposed CRNN hybridized with GWO exhibits exceptional consistency and clinical reliability. It is noteworthy that an increase in the training epoch number from 50 to 100 results in a complete absence of change in the classification metrics. This serves to con- firm the stability and early convergence of the model.

Table 2. Classification performance comparison – hypoglycemia.

Model	AUROC	Precision (BG<4)	Recall (BG<4)	F1 score (BG<4)	Precision (BG<3)	Recall (BG<3)	F1 score (BG<3)	Classification accuracy
CRNN (50 Epochs) + GWO	0.98	0.97	0.98	0.98	0.98	0.98	0.98	0.99
CRNN (100 Epochs) + GWO	0.98	0.97	0.98	0.98	0.98	0.98	0.98	0.99
CRNN (50 Epochs)	0.97	0.96	0.97	0.96	0.96	0.97	0.96	0.98
CRNN (100 Epochs)	0.97	0.96	0.97	0.96	0.96	0.97	0.96	0.98
XGBoost	0.96	0.88	0.70	0.78	0.97	0.67	0.79	0.97
Gradient boosting	0.96	0.87	0.70	0.78	0.96	0.67	0.79	0.97
RF	0.94	0.86	0.67	0.75	0.96	0.66	0.78	0.95
Extra Trees	0.93	0.85	0.68	0.76	0.94	0.66	0.78	0.94
Bagging	0.93	0.84	0.70	0.76	0.93	0.67	0.78	0.94
AdaBoost	0.89	0.68	0.60	0.64	0.63	0.46	0.53	0.91
LDA	0.88	0.69	0.75	0.72	0.72	0.72	0.72	0.90
Bernoulli NB	0.82	0.60	0.60	0.60	0.47	0.67	0.55	0.86
Gaussian NB	0.81	0.47	0.68	0.56	0.33	0.81	0.47	0.85
Decision Tree	0.81	0.70	0.71	0.70	0.68	0.73	0.70	0.85
SVM	0.79	0.73	0.10	0.18	0.41	0.10	0.16	0.83
QDA	0.77	0.23	0.96	0.37	0.15	0.97	0.26	0.82
Passive aggressive	0.76	0.46	0.25	0.32	0.33	0.10	0.15	0.81
Multinomial NB	0.75	0.10	0.10	0.10	0.10	0.10	0.10	0.80
SGD	0.74	0.12	0.10	0.11	0.10	0.10	0.10	0.80
MLP	0.74	0.57	0.17	0.26	0.47	0.14	0.22	0.80
LR	0.73	0.48	0.10	0.17	0.39	0.10	0.16	0.79
KNN	0.62	0.40	0.18	0.25	0.30	0.15	0.20	0.70

In contrast, ensemble models such as XGBoost and Gradient Boosting demonstrate robust overall performance yet exhibit diminished recall values (0.67-0.70), particularly under BG3 conditions, signifying a diminished capacity to capture severe hypoglycemia events. Conventional methods, including RF, Extra Trees, and Bagging, demonstrate moderate performance, with Area Under the Receiver Operating Characteristic Curve (AUROC) values ranging from 0.93 to 0.94 and F1 scores approximating 0.75 to 0.78. This suggests a limited degree of sensitivity.

It has been demonstrated that simpler models, including SVM, LR, and NB variants, exhibit poor recall (<0.2) and unstable F1 scores. This renders them unreliable for real- world hypoglycemia detection. It is interesting to note that QDA achieves a high recall rate of approximately 0.97 but extremely low precision, which renders it susceptible to false positives. As a result, the

present study demonstrates that CRNN hybridized with GWO exhibits an optimal balance between detection capability and false alarm mitigation, rendering it highly conducive to the development of early hypoglycemia warning systems.

On the other hand, the regression analysis has shown that our model outperforms existing state-of-the-art approaches. Table 3 shows the comparison of regression performance with the current literature. As can be seen, our model achieved an improved RMSE of 10.0 mg/dL and an MAE) of 7.0 mg/dL with an R-squared value of 0.92, indicating good accuracy in predicting continuous glucose levels. In addition, clinical acceptability according to Clarke Error Grid Analysis is 92%, which is an excellent performance for continuous monitoring applications. The approach closest to our model is which gives an RMSE of 35.19 mg/dL for LSTM and 36.08 mg/dL for self-attention networks. The maximum clinical acceptability achieved by both algorithms was 91% using Clarke Error Grid analysis.

Table 3. Regression performance comparison – Blood glucose value prediction.

Model/Study	RMSE (mg/dL)	MAE (mg/dL)	R^2/Correlation	Dataset	Prediction horizon (minute)	Clinical acceptance	Real-time accuracy
Our CRNN (50 Epochs) + GWO	10	7	0.92	HUPA-UCM	Real-time (5-min)	92%	90%
Our CRNN (100 Epochs) + GWO	10	7	0.92	HUPA-UCM	Real-time (5-min)	92%	90%
[35]Bi-LSTM	19.49±5.42	14.93±4.20	0.43±0.2	Life-log data	Current time (0-min)	-	-
[33]LSTM	35.19	-	-	OhioT1DM	30-60	>91% (CEG)	-
[33]SAN	36.08	-	-	Multiple	30-60	>91% (CEG)	-
[33]CNN	~37-40	-	-	Multiple	30-60	>91% (CEG)	-
[33]TCN	~37-40	-	-	Multiple	30-60	>91% (CEG)	-

Table 4 shows the training performance analysis with 50 and 100 epochs. The analysis revealed optimal generalization properties where the improvement over RMSE during training was from 9.2 mg/dL (50 Epochs) to 8.8 mg/dL (100 Epochs). Also, the test set RMSE remained stable at 10.0 mg/dL for prediction tasks, which is an indicator of generalization without over-fitting. Comparing CRNN+GWO models trained for 50 and 100 epochs shows a high degree of model stability with only slight improvements in training performance. Training RMSE improved slightly from 9.2 mg/dL to 8.8 mg/dL (a 4.3% reduction), suggesting modest learning progress with additional epochs. However, the test RMSE and real-time RMSE remained unchanged at 10.0 and 15.0 mg/dL, respectively, strongly indicating that the model did not overfit de- spite the longer training period. Similarly, R-squared (0.92) stayed constant, confirming that the model's ability to explain variance was already optimized and consistent across epochs. The minor decrease in MAE (from 9.3 to 9.2 mg/dL) shows slightly better average prediction accuracy, but the effect is minimal. Classification metrics—F1 score, precision, recall, and accuracy—remained unchanged, all at strong values, con- firming that the model is highly stable in terms of event classification across both training configurations. The clinical acceptance score (0.92) and real-time accuracy (0.90) also showed no change, further reinforcing confidence in the model's performance when deployed in a clinical setting. Overall, the results suggest that the hybridized GWO based CRNN architecture achieves rapid convergence, with performance metrics stabilizing after 50 epochs. Extending training to 100 epochs does not degrade performance but offers diminishing returns, demonstrating that the model is efficient, stable, and robust under extended training.

Table 4. Training performance analysis.

Metric	CRNN (50 Epochs)	CRNN (100 Epochs)	Change
RMSE (Training set)	9.2 mg/dL	8.8 mg/dL	-0.4 (-4.3%)
RMSE (Test set – Prediction)	10.0 mg/dL	10.0 mg/dL	0.0 (0.0%)
RMSE (Real-time)	15.0 mg/dL	15.0 mg/dL	0.0 (0.0%)
R-squared (R^2)	0.92	0.92	0.00 (0.0%)
MAE (Prediction)	7.1 mg/dL	7.0 mg/dL	-0.1 (-1.4%)
MAE (Real-time)	10.0 mg/dL	10.0 mg/dL	0.0 (0.0%)
F1 Score	>0.95	>0.95	No significant change
Precision	0.97	0.97	No significant change
Recall	0.98	0.98	No significant change
Classification accuracy	0.99	0.99	No change
Clinical acceptance	0.92	0.92	No change
Real-time accuracy	0.90	0.90	No change

Table 5 compares our results with four recent studies using clinical validation metrics. The proposed CRNN model achieved 92% clinical acceptability and 99% accuracy for hypoglycemia detection, surpassing the best values reported in those studies. However, since our model was evaluated on the HUPA-UCM dataset whereas other works used different datasets, these comparisons are not one-to-one. We report them for context, but direct performance ranking should be interpreted with caution. Future evaluations on common datasets will be necessary for a fair comparison.

Table 5. Comparison of blood glucose prediction models.

Study & Model	Classification accuracy	Clinical acceptance	Real-time performance	Dataset	Key findings
Our CRNN (100 Epochs)+GWO	99% classification	92% clinical	90% real-time	HUPA-UCM	Superior across all evaluation modes
[33]LSTM	>91% (CEG Safe zone)	>91% (CEG)	Not reported	OhioT1DM	Best generalization across datasets
[33]SAN	>91% (CEG safe zone)	>91% (CEG)	Not reported	Multiple	Second-best performance
[35]LSTM	Physiologically sound	–	Not reported	OhioT1DM	SHAP-validated interpretability
[35]Bi-LSTM	–	Correlation: 0.43±0.2	Not reported	Life-log	Virtual CGM framework

Note: *CEG: Clarke error grid; CGM: ???, SAN: self-attention network, CRNN: convolutional recurrent neural network.*

An avenue for further improvement is the use of Transformer-based models, which have achieved state-of-the-art results in blood glucose prediction. These models leverage self-attention to capture long-term dependencies and have outperformed traditional CNN/LSTM in some scenarios. We did not evaluate Transformers in this study, as our focus was on CNN/LSTM hybrids; however, future work will compare the CRNN approach against Transformers to ensure the proposed model remains competitive with the latest deep learning techniques.

A limitation of this study is that the model was trained and tested on a single dataset (HUPA-UCM). No external or cross-population evaluation was performed, which may affect the generalizability of the results. In future work, we plan to validate the model on additional datasets (e.g., OhioT1DM) and across different patient populations, as well as exploring additional AI architectures and bio-inspired optimization algorithms to further enhance predictive performance.

Conclusion

In this paper, we present a CRNN architecture, enhanced with a GWO for hyperparameter tuning, applied to the task of predicting hyperglycemia and hypoglycemia. Among the tested configurations, the CRNN+GWO hybrid achieved the lowest RMSE and MAE, outperforming the baseline CRNN. While our model outperforms others in point estimates of accuracy, we acknowledge that we did not perform statistical significance testing. Thus,

differences, especially when within a few percentage points, should be interpreted with caution. Using a recent large-scale dataset, we benchmarked our results against recent state-of-the-art approaches. Our primary objective was to maximize accuracy in the early prediction of hypo and hyperglycemic events and to enable real-time detection. The proposed model achieved 99% accuracy on offline data with low mean squared and absolute errors, and nearly 90% accuracy in simulated real-time scenarios. It can trigger timely alerts for abnormal glucose levels, offering strong potential for improving patient health outcomes. The combination of CNN, RNN—particularly LSTM and GWO proved robust across DL tasks in this domain.

Funding

No funding was provided for this research.

Conflict of interest

There is no potential conflict of interest.

CRediT authorship contribution statement

Djamila Ouaret: conducted the experiments and analyzed the data; **Houda El Bouhissi**: supervised the project, drafted the manuscript, improved the contribution and served as the corresponding author; **Tatiana Emarkova**: drafted the manuscript, improved the contribution. **Rabie A. Ramadan**: drafted the manuscript, improved the contribution.

Acknowledgements

None

References

[1] ElSayed NA, McCoy RG, Aleppo G, Balapattabi K, Beverly EA, Briggs Early K, et al. Diagnosis and classification of diabetes: Standards of care in diabetes—2025. Diabetes Care. 2025; 48(Suppl 1): S27-S49. doi:10.2337/dc25-S002.

[2] Vaiyapuri T, Alharbi G, Muttipoll Dharmarajlu S, Bouteraa Y, Misra S, Venkata Naga Ramesh J, et al. IoT-Enabled early detection of diabetes diseases using deep learning and dimensionality reduction techniques. IEEE Access. 2024; 12: 143016-28.

[3] Albahli S. Predictive analytics for diabetic patient care: Leveraging AI to forecast readmission and hospital stays. Comput Model Eng Sci. 2025; 143(1): 1095-128. doi:10.32604/cmes.2025. 05882.

[4] Lin J, Song Z, Li Y, Chiang C, Hirakawa Y, Nakano Y, Hong YJ, Matsunaga M, Ota A, Tamakoshi K, Yatsuya H. Nonrestorative sleep and type 2 diabetes incidence: The Aichi workers' cohort study. J Epidemiol. 2024; 34(9): 428-33. doi:10.2188/jea.JE20230184.

[5] Bliss M. Who discovered insulin? In: The discovery of insulin. London: Palgrave Macmillan; 1982: pp 189-211.

[6] Umpierrez GE, Davis GM, ElSayed NA, Fadini GP, Galindo RJ, Hirsch IB, et al. Hyperglycemic crises in adults with diabetes: A consensus report. Diabetes Care. 2024; 47(8): 1257-75. doi:10.2337/dci24-0032.

[7] Yang T, Yu X, Tao R, Li H, Zhou J. Blood glucose prediction for type 2 diabetes using clustering-based domain adaptation. Biomed Signal Process Control. 2025; 105: 107629. doi:10.1016/j.bspc. 2025.107629.

[8] Mazgouti L, Laamiri N, Ben Ali J, El Amrani El Idrissi N, Di Costanzo V, Naeck R, et al. Optimization of blood glucose prediction with LSTM-XGBoost fusion and integration of statistical features for enhanced accuracy. Biomed Signal Process Control. 2025; 107: 107814. doi:10.1016/j.bspc.2025.107814.

[9] Gao J, Guo C, Liu Y, Li P, Zhang J, Liu M. Dynamic-static feature fusion with multi-scale attention for continuous blood glucose prediction. In: Proc IEEE Int Conf Acoust Speech Signal Process (ICASSP); 2025 Apr; Hyderabad; India: pp 1-5.

[10] Sharma SK, Zamani AT, Abdelsalam A, Muduli D, Alabrah AA, Parveen N, Alanazi SM. A diabetes monitoring system and healthmedical service composition model in cloud environment. IEEE Access. 2023; 11: 32804-19. doi:10.1109/ACCESS. 2023.3258549.

[11] Ghosh M, Bora V. Evolution in blood glucose monitoring: from invasive to noninvasive devices and sensors. Discover Med. 2025; 2: 74. doi:10.1007/ s44337025002731.

[12] El Bouhissi H, Al-Qutaish RE, Ziane A, Amroun K, Yaya N, Lachi M. Towards diabetes mellitus prediction based on machine learning. In: Proc Int Conf Smart Computing and Application (ICSCA); 2023: pp 1-6.

[13] El Bouhissi H, Ziane A, Rahmani L, Medbal M, Kostiuk M. RF-PSO: An optimized approach for diabetes prediction. In: Proc Int Conf Smart Technologies (ICST); 2023: pp 227-38.

[14] Tomic D, Harding JL, Jenkins AJ, Shaw JE, Magliano DJ. The epidemiology of type 1 diabetes mellitus in older adults. Nat Rev Endocrinol. 2025; 21(2): 92-104. doi:10.1038/s4157402401046z.

[15] Alfian G, Syafrudin M, Ijaz MF, Syaekhoni MA, Fitriyani NL, Rhee J. A personalized healthcare monitoring system for diabetic patients using BLEbased sensors. Sensors (Basel). 2018; 18(7): 2183. doi:10.3390/s18072183.

[16] Jaloli M, Cescon M. Longterm prediction of blood glucose levels in type 1 diabetes using a CNNLSTM based deep neural network. J Diabetes Sci Technol. 2023; 17(6): 1590-601. doi:10.1177/1932296822109 2785.

[17] 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; 4(3): 308-24. doi:10.1007/s41666020000682.

[18] Hidalgo JI, Alvarado J, Botella M, Aramendi A, Velasco JM, Garnica O. HUPAUCM diabetes dataset. Data Brief. 2024; 55: 110559. doi:10.1016/j.dib. 2024.110559.

[19] Tena F, Garnica O, Lanchares J, Fernándezde Vega

F, Hidalgo JI. Ensemble deep neural networks for blood glucose prediction. Sensors (Basel). 2021; 21(21): 7090. doi:10.3390/s21217090.

[20] Ruan Y, Bellot A, Moysova Z, Tan GD, Lumb A, Davies J, van der Schaar M, Rea R. Predicting the risk of inpatient hypoglycemia with machine learning using electronic health records. Diabetes Care. 2020; 43(7): 1504-11. doi:10.2337/dc191743.

[21] Shi M, Yang A, Lau ESH, Luk AOY, Ma RCW, Kong APS, Wong RSM, Chan JCM, Chan JCN, Chow E. A novel electronic health recordbased, machine-learning model to predict severe hypoglycemia leading to hospitalizations in older adults with diabetes: a territorywide cohort and modeling study. PLoS Med. 2024; 21(4): e1004369. doi:10. 1371/journal.pmed.1004369.

[22] Tsichlaki S, Koumakis L, Tsiknakis M. Type 1 Diabetes hypoglycemia prediction algorithms: Systematic review. JMIR Diabetes. 2022; 7(3): e34699. doi:10. 2196/34699.

[23] Aiello EM, Lisanti G, Magni L, Musci M, Toffanin C. Therapydriven deep glucose forecasting. Eng Appl Artif Intell. 2020; 87: 103255. doi:10.1016/j. engappai.2019.103255.

[24] Karagoz MA, Breton MD, Fathi AE. A comparative study of transformerbased models for multihorizon blood glucose prediction. arXiv. 2025: arXiv: 2505. 08821. doi:10.48550/arXiv.2505.08821.

[25] Munoz-Organero M. Deep physiological model for blood glucose prediction in T1DM patients. Sensors (Basel). 2020; 20(14): 3896. doi:10.3390/s2014 3896.

[26] Alvarado J, Velasco JM, Chavez F, Fernándezde Vega F, Hidalgo JI. Combining wavelet transform with convolutional neural networks for hypoglycemia prediction from CGM data. Chemometr Intell Lab Syst. 2023; 243: 105017. doi:10.1016/j.chemolab. 2023.105017.

[27] Mirshekarian S, Bunescu R, Marling C, Schwartz F. Using LSTMs to learn physiological models of blood glucose behavior. In: Proc IEEE EMBS Int Conf (EMBC); 2017; Jeju; Korea: pp 2887-91.

[28] Sun Q, Jankovic MV, Bally L, Mougiakakou SG. Predicting blood glucose with LSTM and Bi-LSTM deep neural networks. In: Proc Int Symp Neural Networks Appl (NEUREL); 2018; Belgrade: pp 1-5.

[29] El Idrissi T, Idri A, Abnane I, Bakkoury Z. Predicting blood glucose using an LSTM neural network. 2019 Federated Conference on Computer Science and Information Systems (FedCSIS). 2019: 35-41. doi:10.15439/2019F159.

[30] Iacono F, Magni L, Toffanin C. Personalized LSTM-based alarm systems for hypoglycemia and hyperglycemia prevention. Biomed Signal Process Control. 2023; 86: 105167. doi:10.1016/j.bspc.2023. 105167.

[31] Li K, Daniels J, Liu C, HerreroVinas P, Georgiou P. Convolutional recurrent neural networks for glucose prediction. IEEE J Biomed Health Inform. 2019; 24(2): 603-13. doi:10.1109/JBHI.2019.29084 88.

[32] Kavakiotis I, Tsave O, Salifoglou A, Maglaveras N, Vlahavas I, Chouvarda I. Machine learning and data mining methods in diabetes research. Comput Struct Biotechnol J. 2017; 15: 104-16. doi:10.1016/j.csbj.2016.12.005.

[33] Rghioui A, Lloret J, Sendra S, Oumnad A. A smart architecture for diabetic patient monitoring using machine learning algorithms. Healthcare (Basel). 2020; 8(3): 348. doi:10.3390/healthcare8030348.

[34] 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: a metaanalysis. JMIR Diabetes. 2021; 6(1): e22458. doi:10.2196/22458.

[35] Wang G, Liu X, Ying Z, Yang G, Chen Z, Liu Z, et al. Optimized glycemic control of type 2 diabetes with reinforcement learning: a proofofconcept trial. Nat Med. 2023; 29: 2633-42. doi:10.1038/ s41591023025529.

[36] LópezGómez JA, Romero FP, Angulo E. Bioinspired algorithms for the characterization of excellent performance in handball players: A datadriven methodology. Expert Syst Appl. 2025; 274: 126821. doi:10. 1016/ j.eswa.2025.126821.

[37] Mirjalili S, Saremi S, Mirjalili SM, dos S. Coelho L. Multiobjective grey wolf optimizer: a novel algorithm for multicriterion optimization. Expert Syst Appl. 2016; 47: 106-19. doi:10.1016/j.eswa. 2015.10.039.