Heliyon 10 (2024) e29030

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Glu-Ensemble: An ensemble deep learning framework for blood glucose forecasting in type 2 diabetes patients

Yechan Han [a], Dae-Yeon Kim [b], Jiyoung Woo [c], Jaeyun Kim [c,*]

[a] Department of Medical Science, Soonchunhyang University, Asan, 31538, Republic of Korea
[b] Department of Internal Medicine, Soonchunhyang University Cheonan Hospital, Cheonan, 31151, Republic of Korea
[c] Department of AI and Big Data, Soonchunhyang University, Asan, 31538, Republic of Korea

ARTICLE INFO

Keywords:
Type 2 diabetes
Blood glucose forecasting
Deep learning
Ensemble method
Error grid analysis

ABSTRACT

Diabetes is a chronic metabolic disorder characterized by elevated blood glucose levels, posing significant health risks such as cardiovascular disease, and nerve, kidney, and eye damage. Effective management of blood glucose is essential for individuals with diabetes to mitigate these risks. This study introduces the Glu-Ensemble, a deep learning framework designed for precise blood glucose forecasting in patients with type 2 diabetes. Unlike other predictive models, Glu-Ensemble addresses challenges related to small sample sizes, data quality issues, reliance on strict statistical assumptions, and the complexity of models. It enhances prediction accuracy and model generalizability by utilizing larger datasets and reduces bias inherent in many predictive models. The framework's unified approach, as opposed to patient-specific models, eliminates the need for initial calibration time, facilitating immediate blood glucose predictions for new patients. The obtained results indicate that Glu-Ensemble surpasses traditional methods in accuracy, as measured by root mean square error, mean absolute error, and error grid analysis. The Glu-Ensemble framework emerges as a promising tool for blood glucose level prediction in type 2 diabetes patients, warranting further investigation in clinical settings for its practical application.

1. Introduction

Diabetes is a chronic disease marked by persistently high levels of glucose in the blood, stemming from insufficient insulin production or significant insulin resistance, which can lead to various health complications [1,2]. The condition is broadly classified into two primary types: type 1 diabetes (T1DM) and type 2 diabetes (T2DM). Type 1 diabetes is an autoimmune condition where the body's immune system mistakenly attacks and destroys the insulin-producing cells in the pancreas, typically manifesting in childhood and necessitating lifelong insulin therapy [3,4]. Type 2 diabetes, which is more prevalent, is characterized by insulin resistance or inadequate insulin production, often associated with aging, obesity, and a lack of physical activity [5–7]. Inadequately controlled or untreated diabetes can result in a wide array of health complications [8–15]. Moreover, uncontrolled blood glucose levels can both increase the risk of infection and prolong the healing of wounds [16–18].

Blood glucose forecasting serves as a pivotal tool for individuals with diabetes to enhance their management strategies [19–21]. Precise predictions of blood glucose levels empower individuals to make well-informed decisions regarding insulin administration, dietary choices, and physical activities. This, in turn, facilitates better control of blood glucose and mitigates the risk of complications

* Corresponding author.
E-mail addresses: ychan.sch@gmail.com (Y. Han), c99851@schmc.ac.kr (D.-Y. Kim), jywoo@sch.ac.kr (J. Woo), kimym38@sch.ac.kr (J. Kim).

https://doi.org/10.1016/j.heliyon.2024.e29030
Received 15 October 2023; Received in revised form 28 March 2024; Accepted 28 March 2024
Available online 4 April 2024

associated with diabetes [22–24]. Various approaches, including statistical models [25–27], machine learning algorithms [28–31], and artificial neural networks [32–39], have been explored for predicting blood glucose levels. Statistical models, which utilize mathematical equations to project future blood glucose levels based on past data, insulin doses, and other pertinent factors, are among the simplest and most prevalent methods. Despite their simplicity, accessibility, and speed, the precision of these models is contingent upon the quality of input data and the robustness of the mathematical framework. However, their applicability may be limited for individuals with complex or rapidly varying blood glucose profiles. Machine learning algorithms, which employ sophisticated computational techniques to analyze extensive datasets and forecast future blood glucose levels, offer greater accuracy than statistical models. These algorithms are adept at handling more intricate and variable glucose dynamics. Currently, many smartphone apps and wearable devices incorporate machine learning algorithms for blood glucose prediction. However, their reliability is constrained by the need for substantial data volumes to enhance accuracy. In contrast, artificial neural networks excel in managing erratic glucose fluctuations by adeptly modeling the nonlinear relationships within the data [40,41]. Case studies in blood glucose prediction have shown that artificial neural networks surpass conventional methodologies in performance [42], highlighting their potential in providing more reliable and nuanced blood glucose forecasts.

Among the 63 blood glucose prediction studies published from January 2014 to June 2020, a mere eight incorporated over 100 patients in their analyses [42], and notably, only a single study accounted for patients with T2DM. The prevalence of small sample sizes in such research poses significant limitations, primarily hindering the ability to capture the diverse variability inherent in blood glucose levels, which in turn compromises the generalizability of findings. The fluctuations in blood glucose are influenced by a multitude of factors including age, gender, ethnicity, and existing health conditions, making the development of a comprehensive model that accurately reflects the entire population a complex endeavor. In sum, the reliability and applicability of blood glucose prediction models in research are constrained by several challenges, such as the quality of the data, underlying statistical assumptions, and the intricacies of the models employed.

In this study, a blood glucose prediction model that addresses the previously discussed limitations was devised. An integrated model was constructed by training it with extensive blood glucose data from a diverse group of patients. For the prediction phase, an ensemble approach that synthesizes the outputs of multiple models to enhance predictive accuracy was employed. The key contributions of this study are outlined as follows: (1) this study primarily targets T2DM, a topic less extensively explored compared to T1DM. (2) The patient dataset is expanded beyond what has been typical in prior studies on blood glucose prediction, thereby reducing the potential for model bias through increased sample size. (3) A unified model that supersedes the need for individual patient models is introduced, thereby eliminating the initial synchronization period required for each patient's model setup and facilitating immediate blood glucose predictions for new patients. (4) An ensemble method that amalgamates various prediction models is implemented, aiming to diminish the bias inherent in predictions made by single models, thereby enhancing the overall predictive accuracy of the proposed approach.

The remainder of this paper is organized as follows. Section 2 describes the research and techniques used in this study. Section 3 describes the research methodology in detail. Section 4 discusses the experimental results. Section 5 presents a discussion of the results, while Section 6 provides the conclusions and a scope for future research.

2. Literature review

Historically, diabetes management necessitated regular fingerstick blood tests for blood glucose monitoring. However, the advent of continuous glucose monitoring (CGM) devices has transformed this practice by facilitating the continuous measurement and recording of estimated glucose values (EGVs) at frequent intervals. This advancement has provided researchers with a wealth of data, which has been instrumental in the development of predictive models for blood glucose levels and early warning systems for hypoglycemia. Numerous studies focusing on blood glucose prediction have been undertaken to date. This section provides an overview of the various methodologies employed in blood glucose prediction and assesses their efficacy.

2.1. Statistical model

Dasanayake et al. [25] introduced a technique for predicting blood glucose levels in T1DM patients engaged in physical activities, utilizing subspace identification methods. This approach integrates activity monitoring with measurements of subcutaneous blood glucose, and personalization is achieved through semi-definite programming. The effectiveness of the model was confirmed using data from 15 individuals with T1DM, showcasing its potential to enhance glucose management and facilitate the early identification of hypoglycemic events. The study underscored the significance of incorporating activity sensors into artificial pancreas systems and recommended the concurrent use of an insulin-glucose model with the exercise model for optimal results. While acknowledging possible challenges such as delays in sensor readings and the variability in individuals' heart rate responses to exercise, the research shed light on the linear relationship between physical activity and blood glucose levels, offering valuable perspectives for advancing glucose regulation in T1DM patients.

Novara et al. [26] introduced a blind identification method to model T1DM and reconstruct unobserved input signals. This technique was tested in a study involving five patients, utilizing data from CGM systems and insulin pumps. The findings revealed that this method enhances prediction accuracy by accurately recovering disturbances, with consistent results across all participants. This suggests the approach's viability in managing the challenge of unmeasured signals in diabetic patients. Additionally, the study underscored the critical role of mathematical modeling in planning diabetes treatment and demonstrated the blind identification method's capacity to accommodate variations in patient metabolism and lifestyle.

2.2. Machine learning model

Hamdi et al. [28] developed a machine learning-based model to forecast blood glucose levels in T1DM patients using solely CGM data. This model employed support vector regression (SVR) combined with differential evolution (DE) algorithms for optimizing SVR parameters and determining effective values. The model's performance was validated on real CGM data from 12 patients, achieving high prediction accuracy with root mean square errors (RMSE) of 9.44, 10.78, 11.82, and 12.95 mg/dL for prediction horizons of 15, 30, 45, and 60 min, respectively. These results indicate that the integration of SVR and DE algorithms can effectively handle nonlinear and complex data series, offering a promising solution for predicting blood glucose levels in T1DM patients.

Rodríguez-Rodríguez et al. [29] conducted a study to determine the minimal dataset necessary for precise blood glucose prediction in T1DM patients, employing machine learning techniques and wearable technology. The study monitored blood glucose levels in 25 T1DM patients and utilized predictive models such as ARIMA, support vector machines, and random forests to forecast blood glucose values. Findings indicated that accurate predictions could be achieved with a dataset comprising 24 to 72 historical blood glucose measurements collected over a 6-h period. Notably, the random forest model was identified as particularly effective in predicting blood glucose levels, highlighting its potential for application in diabetes management.

2.3. Deep learning model

Deng et al. [32] developed a predictive methodology utilizing deep learning to forecast blood glucose fluctuations in patients with T2DM. To address the challenges posed by limited and unbalanced datasets, the authors implemented transfer learning and data augmentation strategies. The efficacy of the proposed model was evaluated on a cohort of 40 T2DM patients, where it demonstrated a prediction accuracy exceeding 95% and a sensitivity over 90% for a 1-h prediction horizon, underscoring its clinical utility. Additionally, the model's applicability was validated on the OhioT1DM dataset for T1DM patients, indicating its versatility across diabetes types. While the study's approach was predominantly data-centric, the authors suggested future enhancements could involve integrating domain-specific knowledge, such as dietary habits, physical activity levels, and stress factors, to further refine the model's predictive accuracy.

Li et al. [34,35] proposed the GluNet framework as a deep-learning approach for accurate blood glucose prediction in patients with type 1 diabetes. Essentially, GluNet utilizes a personalized deep neural network to predict the probabilistic distribution of future CGM measurements based on historical data that could include blood glucose measurements, meal information, insulin doses, and other similar factors. The proposed model was empirically analyzed using blood glucose data from 16 T1DM patients, and the ABC4D and OhioT1DM datasets were utilized. Experimental results of the model displayed significant improvements over that of existing methods such as SVR and autoregression with exogenous input (ARX) in terms of RMSE and time lag for the 30- and 60-min prediction periods. However, the algorithm is highly dependent on the quality of the training data, and its performance can be influenced significantly by missing or highly variable data.

Khadem et al. [38] proposed a nested deep ensemble learning framework for personalized blood glucose level forecasting in type 1 diabetes patients. They used Ohio type 1 diabetes datasets and employed meta-learning analysis to optimize the lag length in predictions. Models were developed using multilayer perceptron and long short-term memory networks and evaluated for accuracy and clinical effectiveness. The findings revealed enhanced prediction accuracy, effectively overcoming challenges related to analyzing time-series data for forecasting blood glucose levels. The study's methodology based on enhanced predictive modeling has considerable potential for diabetes management.

Fitzgerald et al. [39] focused on the development and evaluation of continuous-time recurrent neural networks (CTRNNs) for forecasting blood glucose levels in intensive care unit (ICU) settings. The effectiveness of these models, which were designed to handle irregularly timed data inputs, was compared with that of traditional autoregressive models and gradient boosted trees (GBTs). Additionally, a combination of simulated and real electronic medical record (EMR) data was utilized. The findings indicated that while CTRNNs are promising—particularly in specific scenarios such as insulin treatment—GBT models often demonstrate comparable or superior performance with easier and faster training processes.

Butt et al. [53] developed a novel approach in their research, involving a multi-layered Long Short-Term Memory (LSTM)-based recurrent neural network for predicting blood glucose levels in patients with type 1 diabetes. The study introduces a new method for transforming event-based data into discriminative continuous features for use in blood glucose prediction. The methodology is evaluated using the Ohio T1DM dataset on three patients, achieving promising results. It suggests potential applications in closed-loop systems for precise insulin delivery to improve glycemic control in type 1 diabetes patients.

Prendin et al. [54] emphasize the crucial role of model interpretability in Type 1 Diabetes (T1D) management, particularly for decision support systems (DSS) that forecast glucose levels and suggest insulin boluses. Their case study contrasts two long-short term memory neural networks (LSTM), p-LSTM and np-LSTM, which, despite similar predictive accuracies, lead to distinct therapeutic decisions. Using SHAP for analysis, they demonstrate that only p-LSTM captures the physiological relationship between inputs and glucose prediction, advocating for its use in DSS. This work underlines the importance of choosing models that are not only accurate but also physiologically sound and explainable to enhance patient care.

The research reviewed in this section highlights the promising capabilities of diverse methodologies for blood glucose level prediction in diabetic patients. Nonetheless, the small sample sizes employed in some studies may constrain the generalizability of their results. To enhance the reliability and applicability of prediction models, it is imperative to use larger datasets for validation purposes. Addressing these challenges, the present study seeks to develop a more accurate and generalizable blood glucose prediction model for diabetic patients by leveraging an expanded dataset and incorporating data from various sources. By analyzing blood glucose readings

from a broader patient cohort, the model's precision may be refined and its potential for widespread clinical use may be extended.

3. Glu-ensemble architecture

The Glu-Ensemble algorithm proposed in this study is structured into three distinct phases. Initially, the collected blood glucose data are partitioned to create a dataset for training purposes. Following this, the training dataset is utilized to educate a deep learning model, which is then capable of forecasting blood glucose levels by applying the insights gained during the training process. In the concluding phase, an ensemble technique is applied, leveraging the predictions from the deep learning model to enhance the overall accuracy of the forecasts. Fig. 1 illustrates the tripartite architecture underpinning the Glu-Ensemble approach.

3.1. Phase 1: data splitting

In this initial phase, the process of data splitting was undertaken to convert the collected blood glucose readings into a format suitable for training the model. Fig. 2 illustrates the methodology for segmenting the blood glucose data, which were recorded at 5-min intervals, into distinct sequences to serve as the training dataset. Within the Glu-Ensemble framework, these sequences of blood glucose values form the sole input for the model. As depicted in the figure, the blood glucose time-series data were divided into sequences of varying durations, specifically 60-, 120-, and 180-min intervals. Opting for longer sequences furnishes the model with more extensive information, albeit at the expense of increased learning time. It is important to note that when dealing with sequences of different lengths, the label assigned to each sequence should correspond to the label of the sequence with the maximum length. This label represents the blood glucose level that the model aims to predict, based on the information contained within the sequence; in the illustrated example, this would be the blood glucose value forecasted to occur 30 min later. The choice of sequence length and interval spacing is flexible and can be tailored to meet the specific requirements of the study being conducted.

3.2. Phase 2: deep learning prediction

Five deep learning models were used in Glu-Ensemble: Recurrent neural network (RNN), long short-term memory (LSTM), stack LSTM, bidirectional (Bi)LSTM, and gated recurrent unit (GRU). The aforementioned models are used for time-series forecasting in various fields and are known to have good forecasting performance [43–45]. An RNN is a neural network that has loops to enable information to persist over time and has been widely used in time-series forecasting tasks. LSTM is a type of RNN that better handles the vanishing gradient problem in traditional RNNs when dealing with long sequences. Stack LSTM is an extension of LSTM that enables multiple layers to be stacked on top of each other, which can help improve the performance of a model by learning more complex patterns in the data. BiLSTM is a variant of LSTM that processes the input sequence in both the forward and backward directions, thereby enabling the model to capture information from both past and future time steps. The GRU is another variant of the RNN with a gating mechanism that enables it to either update or forget information, selectively, from the previous time step.

To apply these models to the blood glucose prediction task, the input sequence obtained from the data splitting phase was fed into the model that was trained to predict the blood glucose value at a specified time. The training process involved minimizing the difference between the predicted and actual blood glucose values using a suitable loss function. Once the model has been trained, it may be used to predict new blood glucose data. Overall, the use of multiple deep learning models in Glu-Ensemble enabled a more robust and accurate prediction of blood glucose values. By comparing the performances of these models, researchers can choose the best model or a combination of models for their specific needs. The reason for using similar models is that each model approaches the same

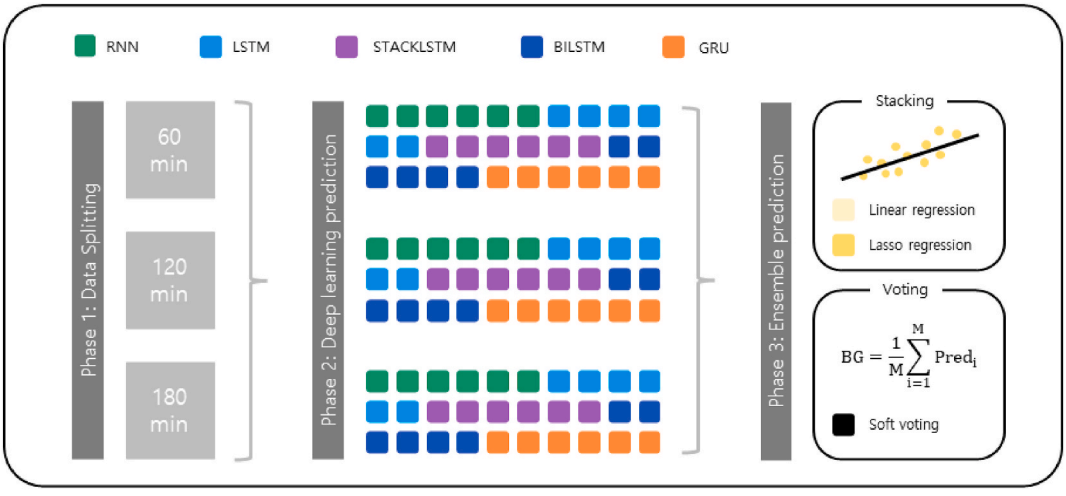

Fig. 1. Framework of glu-ensemble.

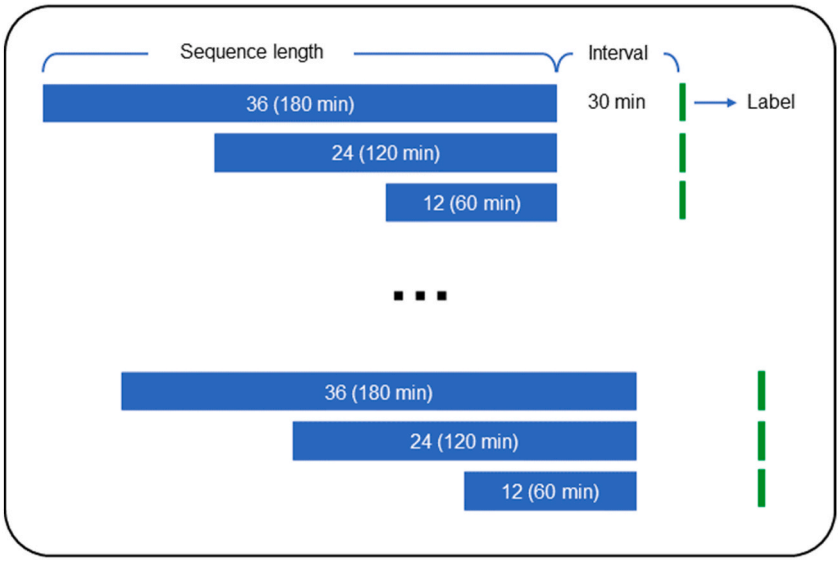

Fig. 2. Data splitting examples.

problem in a different way and brings its own unique strengths. By using a variety of similar models, system diversity may be achieved, and the strengths of each model may be combined to obtain optimal results. Additionally, employing similar models together allows for mutual reinforcement and enables better understanding and improved performance through comparison and evaluation between models.

3.3. Phase 3: ensemble prediction

Owing to the high variability and complexity of blood glucose data, achieving high accuracy and robustness using a single deep learning model is difficult. Ensembles provide a solution to this problem by combining multiple models to achieve improved performance and robustness. The idea behind ensemble prediction is that by combining predictions from different models, one can reduce the risk of overfitting and capture various aspects of the data that a single model might overlook. Utilizing ensembles to combine predictions from multiple deep learning models can lead to more accurate and robust predictions of blood glucose levels, which is bound to have important clinical implications for patients with diabetes.

Stacking and soft voting were employed during the ensemble prediction phase, where stacking combines the predictions of individual deep learning models using a regression model as a meta-model. This meta-model learns to combine predictions to minimize overall error and enhance the accuracy of the final prediction. Stacking model designers commonly employ linear, Lasso, and Ridge regression as meta-models for capturing various types of relationships within the data [46,47]. This study focused solely on Linear and Lasso regression models. Linear regression captures linear relationships between input and output variables, whereas Lasso simplifies the model by minimizing weight values through regularization. Ridge regression, known for its high computational complexity, was excluded from experimentation in this study. Through regularization these regressions help process complex data and combine diverse predictions, leading to improved generalization and predictive performance in ensemble models. Soft voting is a technique used in ensemble learning in which the predicted probabilities from multiple models are averaged to make a final prediction. In contrast to hard voting, which considers the majority vote of the predicted classes, soft voting considers the confidence level or probability of each model's prediction. In the case of regression models, soft voting involves averaging the predicted regression values from multiple models. The resulting average can provide a more accurate prediction than the prediction of any individual model because it considers the collective knowledge and expertise of the ensemble of models. Consequently, using an ensemble approach and customizing individual models, more accurate and robust predictions of blood glucose values can be made.

3.4. Performance evaluation methods

To assess the effectiveness of the Glu-Ensemble model, three evaluation metrics were employed: RMSE, mean absolute error (MAE), and error grid analysis (EGA) [48]. RMSE and MAE are standard metrics for gauging the precision of predictions in time-series models, defined by equations (1) and (2):

$$\mathrm{RMSE} = \sqrt{\sum_{k=1}^{n} \frac{(\hat{y}_k - y_k)^2}{n}} \tag{1}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{k=1}^{n} |\widehat{y_k} - y_k| \tag{2}$$

where $(\widehat{y_k} - y_k)$ represents the difference between the actual and predicted values, and n represents the number of predicted values. The RMSE and MAE are overall measures of the difference between predicted and actual values. Smaller RMSE and MAE values indicate that the model's predictions are closer to the actual values. In this study, the RMSE and MAE were calculated for the predicted blood glucose values at each time point for T2DM patients in the dataset.

The gold standard measure of accuracy, EGA, is a graphical method for evaluating the clinical significance of blood glucose predictions [45]. It categorizes the blood glucose spectrum into zones reflecting their clinical impact and employs color coding to indicate the associated risk levels of the predictions [45]. For this study, EGA was executed to evaluate the blood glucose predictions at different time points for T2DM patients (see Fig. 3). The EGA grid zones are structured according to the Parkes Error Grid (PEG), a commonly utilized framework for evaluating the clinical implications of blood glucose readings in T2DM patients [48,49]. The PEG segments the blood glucose range into five zones (A–E), where each zone corresponds to varying degrees of clinical risk associated with the glucose readings. Zones A and B are deemed clinically acceptable, while zones C through E signify escalating risk levels. EGA offers a nuanced examination of prediction accuracy, especially in scenarios where RMSE may not completely reflect the clinical importance of the predictions.

4. Experimental results

All the codes are freely accessible at https://github.com/SCH-YcHan/Glu-Ensemble. Please note that information related to patient privacy and security has been excluded from the presented data.

4.1. Data description

This research received ethical approval from the Institutional Review Board of Soon Chun Hyang University Hospital Cheonan (IRB Protocol Number: SCHCA 2019-11-048). Data collection spanned from July 2019 to March 2023 at the Cheonan Hospital of Soonchunhyang University and was systematically stored within an Apache-PHP-MySQL (APM) database framework. The study encompassed 140 patients diagnosed with type 2 diabetes, with blood glucose levels being monitored at 5-min intervals through the Dexcom G5 and Dexcom G6 Mobile systems. The demographic and clinical characteristics of the participating patients are detailed in Table 1.

4.2. Preprocessing

Some of the collected blood glucose data were recorded as either high or low, and these values were replaced by 400 and 60, respectively, to maintain consistency across the data. Moreover, among the patients who had majority of their blood glucose values recorded as high, those with more than 10% of their blood glucose values recorded as high (eight total patients) were excluded from the study to minimize the impact of repeated values. In some cases, blood glucose readings could not be taken at 5-min intervals owing to circumstances beyond the control of this study during the measurement period. To address this, only blood glucose values from the longest continuous period were extracted as this would provide the most accurate representation of the patient's blood glucose levels. This approach ensured that irregularities or fluctuations in blood glucose levels were excluded from the analysis. During the study,

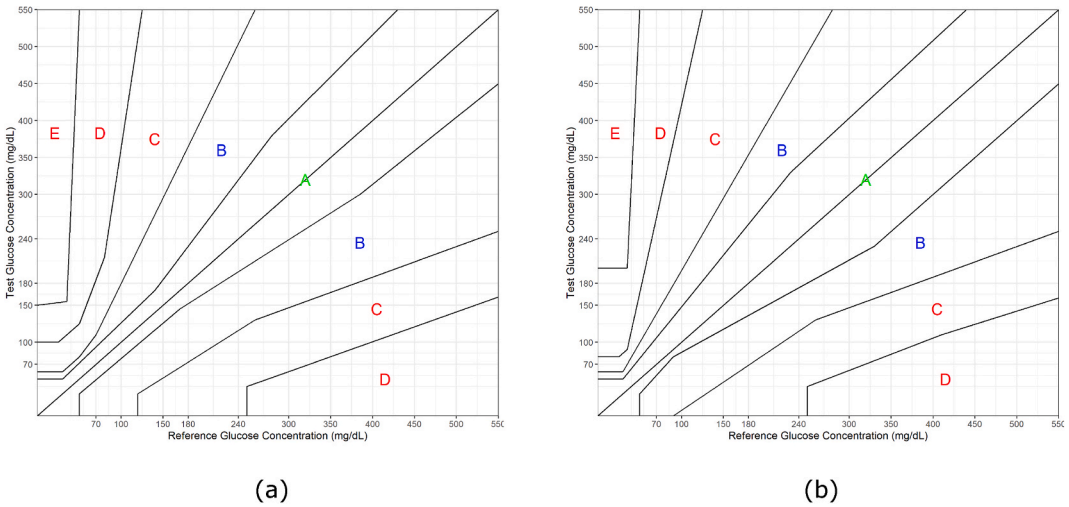

Fig. 3. Parkes error grid by diabetes type [50]: (a) Type 1 diabetes, (b) Type 2 diabetes.

Table 1
Patient demographics.

Sex	Age	Count	Total count
Male	20–29	6	83
	30–39	7	
	40–49	13	
	50–59	20	
	60–69	25	
	70–79	11	
	80–89	1	
Female	20–29	2	57
	30–39	11	
	40–49	10	
	50–59	17	
	60–69	13	
	70–79	4	
	80–89	0	

some patients ended up with less than 100 (500-min interval) blood glucose readings. To ensure that the dataset was sufficiently robust for the model, such patients with fewer than 100 readings (two patients) were also excluded from the study. The final dataset included 130 patients with type 2 diabetes. Fig. 4 shows the preprocessing flowchart.

4.3. Parameter setting

The sequence length, batch size, and hidden size of the training data were adjusted to construct a base model for the ensemble. This process was performed for each model, which resulted in 90 base models. As the parameters are tuned, the number of base models can be adjusted, and the performance of the final prediction can vary. Here, the models exhibited convergence of loss values over 100 epochs. Table 2 lists the parameters used.

4.4. Model validation

The model's performance was assessed using the k-fold cross-validation technique, widely recognized in machine learning for its robustness. To guarantee the exclusive inclusion of each patient in either the training or testing set, patient-level stratification was employed when creating the folds. Prior to initiating the 10-fold cross-validation, patient data were randomized to mitigate any potential bias. The dataset was segmented into 10 equal parts, with nine serving for training purposes and one reserved for testing

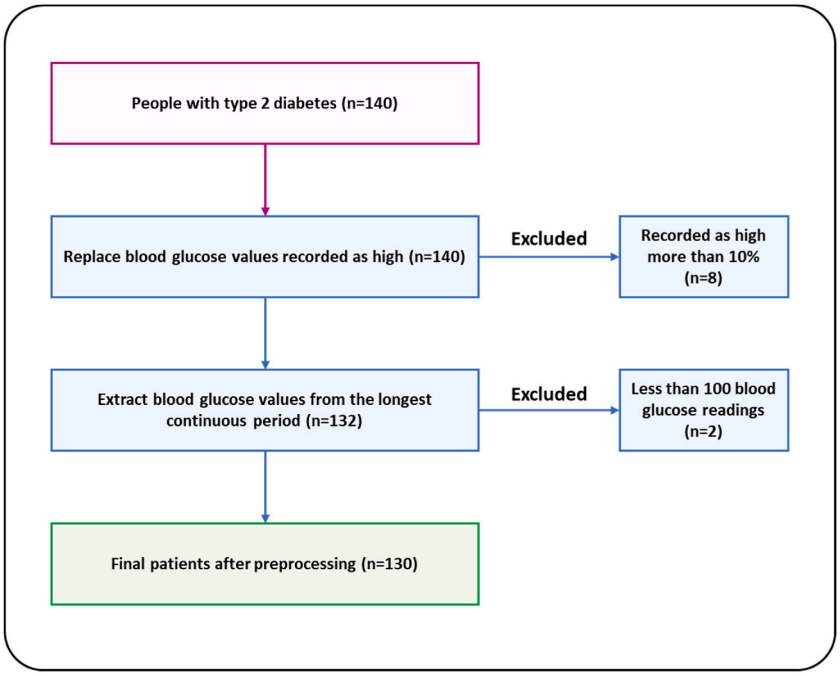

Fig. 4. Flowchart of preprocessing.

Table 2
List of parameters.

Parameter	Description	Values
Sequence length	Length of training data	12, 24, 36
Batch size	Number of samples to be propagated over the network	128, 256
Hidden size	Number of internal representations of past information	1, 2, 3 times the sequence length
Epoch	Number of training iterations	100

during each iteration. This procedure was iterated ten times, ensuring each segment functioned as the testing set once. The aggregated outcomes from each testing phase were then analyzed to derive a comprehensive performance metric for the model, encompassing the entire patient cohort in the study. Fig. 5 illustrates the methodology of the 10-fold cross-validation process applied in this evaluation.

4.5. Benchmark model

The objective of this study was to create a model capable of predicting blood glucose levels 30 min into the future. To gauge the predictive performance of the proposed model, the naïve method was used as a baseline for comparison. The naïve method is a straightforward forecasting technique that predicts future values will remain the same as the most recent observed value. In the context of this study, this implies that the blood glucose level forecasted for 30 min ahead is assumed to be identical to the current blood glucose reading. Fig. 6 illustrates the application of the naïve method to the considered dataset, where predictions are essentially the blood glucose readings from 30 min earlier, rendering them ineffective for practical forecasting. Consequently, any developed blood glucose prediction model is expected to outperform this rudimentary naïve method.

4.6. Evaluation of prediction outcomes using RMSE and MAE

In the study of the Glu-Ensemble framework for predicting blood glucose levels in patients with type 2 diabetes, a detailed analysis of experimental results highlights the framework's superior performance over the benchmark predictive model. Tables 3 and 4 present the RMSE and MAE for each prediction method by fold. The experimental results clearly indicate that the linear, lasso, and soft methods had better prediction performances than the naïve method. Among the methods, the naïve method has the largest average RMSE and MAE for all folds. The soft method, which averages the predicted values of 90 base models, has the best prediction performance with regard to the RMSE and MAE.

The RMSE analysis in Table 3, an essential metric for gauging prediction accuracy, illustrates that Glu-Ensemble consistently surpasses the other models. Specifically, within the stacking model, the linear method registers an average RMSE of 21.46 with a standard deviation of 2.00, the lasso method shows slight improvement with an average RMSE of 21.33 and a standard deviation of 1.98, and the soft method, part of Glu-Ensemble, sets a new benchmark by recording the lowest average RMSE of 21.15 and a standard deviation of 1.96, emphasizing its enhanced accuracy in blood glucose prediction. In contrast, the performance of the naïve method, used as a benchmark, exhibited the highest average RMSE of 24.16 and a standard deviation of 2.29, highlighting its lesser reliability and precision.

Transitioning to the analysis of MAE in Table 4, which sheds light on the average magnitude of errors in predictions, the superior performance of Glu-Ensemble is further verified. The linear method yields an average MAE of 14.54 and a standard deviation of 1.20, while the lasso method is slightly better with an average MAE of 14.42 and a standard deviation of 1.17. Yet again, the soft method, integral to Glu-Ensemble, excels by achieving the lowest average MAE of 14.38 and the smallest standard deviation of 1.11, further validating its precision and consistency in predicting blood glucose levels. The naïve method trails with the highest average MAE of

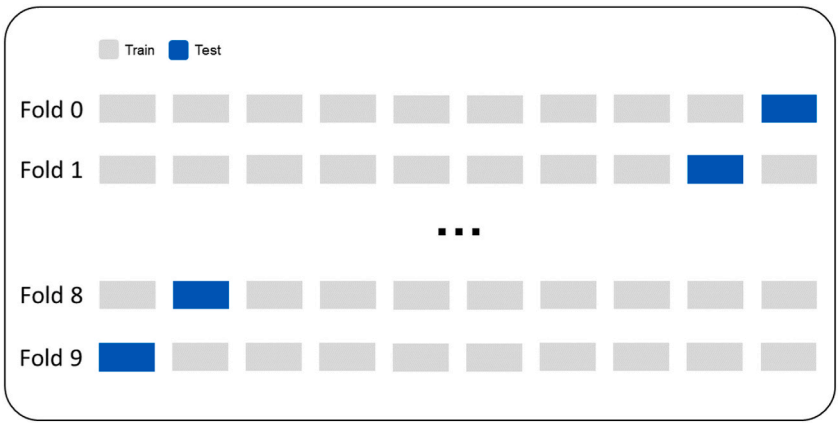

Fig. 5. 10-Fold cross validation.

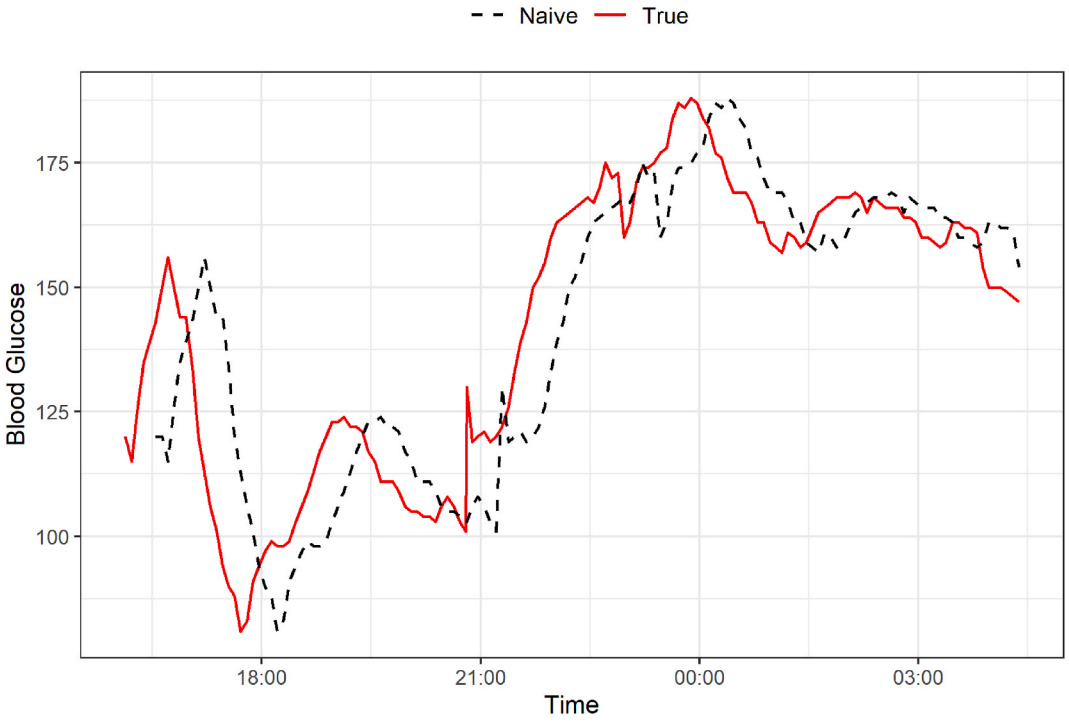

Fig. 6. Naïve method in sample data.

Table 3
RMSE of prediction models by fold.

Method	Fold0	Fold1	Fold2	Fold3	Fold4	Fold5	Fold6	Fold7	Fold8	Fold9	Mean (Std)
Linear	20.02	20.6	22.55	20.66	18.44	20.34	25.41	21.43	21.4	23.79	21.46 (2)
Lasso	19.93	20.5	22.42	20.5	18.37	20.2	25.27	21.35	21.26	23.59	21.33 (1.98)
Soft	19.81	20.17	22.25	20.25	18.14	20.36	25.16	21.26	21.01	23.1	21.15 (1.96)
Naive	22.82	23.12	25.81	23.16	20.61	23.56	29.11	23.83	23.75	25.82	24.16 (2.29)

Table 4
MAE of prediction models by fold.

Method	Fold0	Fold1	Fold2	Fold3	Fold4	Fold5	Fold6	Fold7	Fold8	Fold9	Mean (Std)
Linear	13.3	13.84	14.87	14.4	13.03	13.97	16.25	14.49	14.36	16.84	14.54 (1.2)
Lasso	13.25	13.76	14.77	14.21	12.97	13.86	16.11	14.4	14.2	16.67	14.42 (1.17)
Soft	13.29	13.65	14.66	14.31	12.91	14.08	16.13	14.51	13.91	16.31	14.38 (1.11)
Naive	14.94	15.37	16.95	16.22	14.28	16.09	18.76	16.08	15.45	18.22	16.24 (1.4)

16.24 and a standard deviation of 1.40, reflecting its comparative inaccuracy.

Figs. 7 and 8 show boxplots of the RMSE and MAE for the different prediction methods, respectively. Evidently, the prediction performance of the ensemble-based blood glucose prediction model is much better than that of the naïve method. Through this analysis, the Glu-Ensemble framework is firmly established as a highly effective and reliable tool for forecasting blood glucose levels in type-2 diabetes patients, significantly outperforming the benchmark method.

4.7. Evaluation of prediction outcomes using EGA

Table 5 presents the distribution of predictions across the different zones of the PEG when utilizing the prediction methodology, with the data aggregated from each test fold to validate the results. These findings underscore the clinical superiority of the ensemble method over the naïve approach. Notably, the soft method within the ensemble framework achieved the highest proportion of predictions in zone A, which signifies clinically accurate decisions, at 97.03%. Conversely, it had the lowest incidence of predictions in zone C, indicative of overcorrection, at only 0.047%. In contrast, the naïve method recorded the lowest percentage of predictions in zone A at 95.16% and the highest in zone C at 0.09%, showing a marked discrepancy in clinical accuracy. Additionally, the frequency

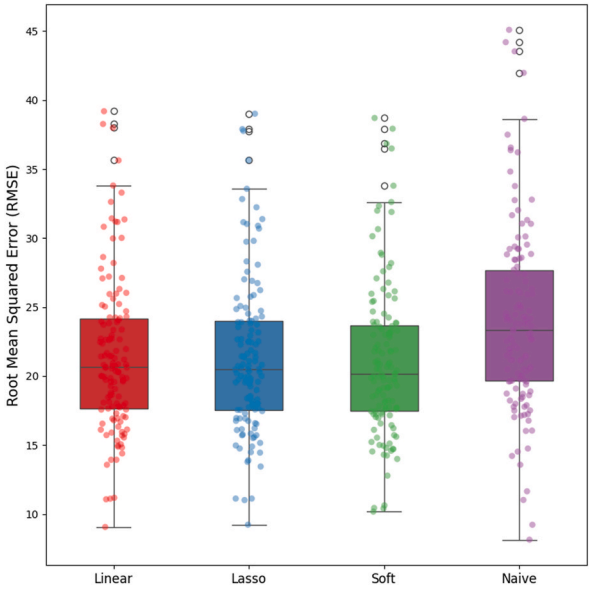

Fig. 7. RMSE boxplot for different forecasting methods.

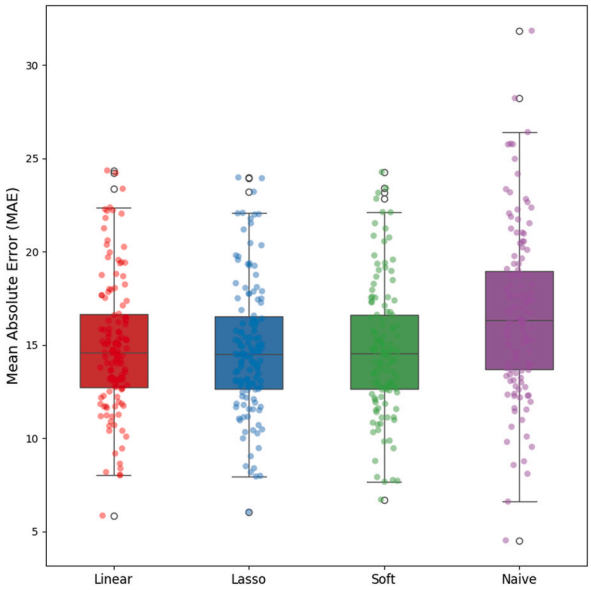

Fig. 8. MAE boxplot for different forecasting methods.

Table 5
Parkes error grid table by forecasting methods.

Method	A	B	C	D	E
Linear	96.878	3.072	0.05	0	0
Lasso	96.907	3.042	0.051	0	0
Soft	97.03	2.923	0.047	0	0
Naive	95.161	4.749	0.09	0	0

of zone C predictions with the naïve method was nearly double that of the other methods, further highlighting its limitations in comparison.

Fig. 9 shows the PEG for each prediction method. In the PEG of the naïve method, there are more points in the C zone in comparison to the other ensemble methods. This indicates that the ensemble method is more clinically relevant than the naïve method. In addition, the ensemble prediction method has a narrower range of predicted values (y-axis) than the naïve method. This is because the variance of the final prediction model is reduced by utilizing multiple models instead of a single model. Therefore, using the ensemble method to predict blood glucose results in more accurate predictions and effectively reduces the variance in the prediction model.

5. Discussion and validation

With access to large-scale public datasets, machine learning models can achieve remarkable performances in blood glucose prediction research. However, the privacy regulations in clinical settings prevent access to sufficient restriction-free data. This hampers machine-learning-based hyperparameter optimization and sensitivity analysis regarding input variables, thereby constraining advancements in these areas. Here, the effects of statistical metrics and hyperparameter optimization in stacking models are considered. Finally, the sensitivity analysis of these models is considered.

5.1. Statistical metrics

The choice of evaluation metrics for assessing blood glucose prediction indicators requires careful deliberation. The selection of RMSE, MAE, and EGA as predictive metrics in this study is underpinned by their respective merits in quantifying different aspects of predictive performance. RMSE and MAE are chosen for their effectiveness in capturing the magnitude of prediction errors and providing valuable insights into the overall accuracy of regression models (see Tables 3 and 4). RMSE emphasizes larger errors through its squared term, while MAE measures the average absolute error between predicted and actual values. These metrics offer straightforward interpretations and are widely employed in evaluating regression model performances. Additionally, EGA is incorporated as a complementary metric to assess the clinical relevance of blood glucose predictions (see Table 5 and Fig. 9). EGA categorizes prediction errors into clinically meaningful zones, allowing for a qualitative assessment of the impact of predictions on clinical decision-making. By considering the clinical significance of prediction errors, EGA provides valuable insights into the practical utility of the predictive models. In summary, this study shows that linear, lasso, and soft methods surpass the naïve method in terms of accuracy, with the soft method and ensemble-based blood glucose prediction model showing superior performance.

While metrics like mean absolute percentage error (MAPE) and R-squared value offer alternative perspectives on prediction accuracy and explanatory power, their inclusion may not substantially enhance the insights afforded by RMSE, MAE, and EGA in blood glucose predictions. The interpretability and clinical relevance of EGA supplement the quantitative assessments provided by RMSE and MAE (see Figs. 7 and 8). Therefore, as primary evaluation metrics in this study, the combination of RMSE, MAE, and EGA comprehensively assesses the predictive performance of blood glucose prediction models, capturing both quantitative accuracy and clinical relevance without introducing unnecessary complexity.

5.2. Hyperparameter optimization in stacking models

In this study, it is asserted that extensive hyperparameter tuning becomes unnecessary when employing a stacking model integrated with a multitude of base models. This assertion is rooted in the fundamental characteristics of stacking models, which amalgamate predictions from a diverse array of base models to bolster overall performance. Unlike standalone models, where hyperparameter optimization is pivotal for achieving peak performance, stacking models capitalize on the collective intelligence garnered from multiple diverse models, thereby diminishing the necessity for meticulous hyperparameter tuning [51].

This is explained by stacking models' ability to amalgamate predictions derived from diverse models, thereby accommodating a broader spectrum of data patterns and intricacies. Through the collective power of numerous base models, stacking exhibits resilience, which mitigates sensitivity to hyperparameter discrepancies [46]. Consequently, the present study posits that while hyperparameter tuning remains crucial in many machine learning scenarios, ensembling stacking models with several base models alleviates the computational overhead associated with hyperparameter searching. This streamlines the model's development and enhances its applicability in real-world settings.

5.3. Sensitivity analysis in stacking models

This study proposed the argument that sensitivity analysis regarding deep learning input values is unnecessary when utilizing a stacking model. This research proposition stems from stacking models' ability to combine predictions from diverse base models, thus reducing their sensitivity to input variations. By design, stacking models amalgamate predictions from multiple base models, thereby mitigating the individual sensitivities of each base model and providing more stable and generalizable predictions. As a result, the need for sensitivity analysis regarding input values is alleviated, as stacking models exhibit robustness against variations in input data.

This study is supported by the literature on stacking models and sensitivity analysis in machine learning. Sill et al. [52] discussed the robustness of stacking models in handling diverse data conditions, highlighting their adaptive ability to combine predictions from different models. Breiman [46] investigated the effectiveness of ensemble methods, including stacking, in improving model stability and generalization. Therefore, the present study asserts that stacking models reduce the necessity for sensitivity analysis regarding

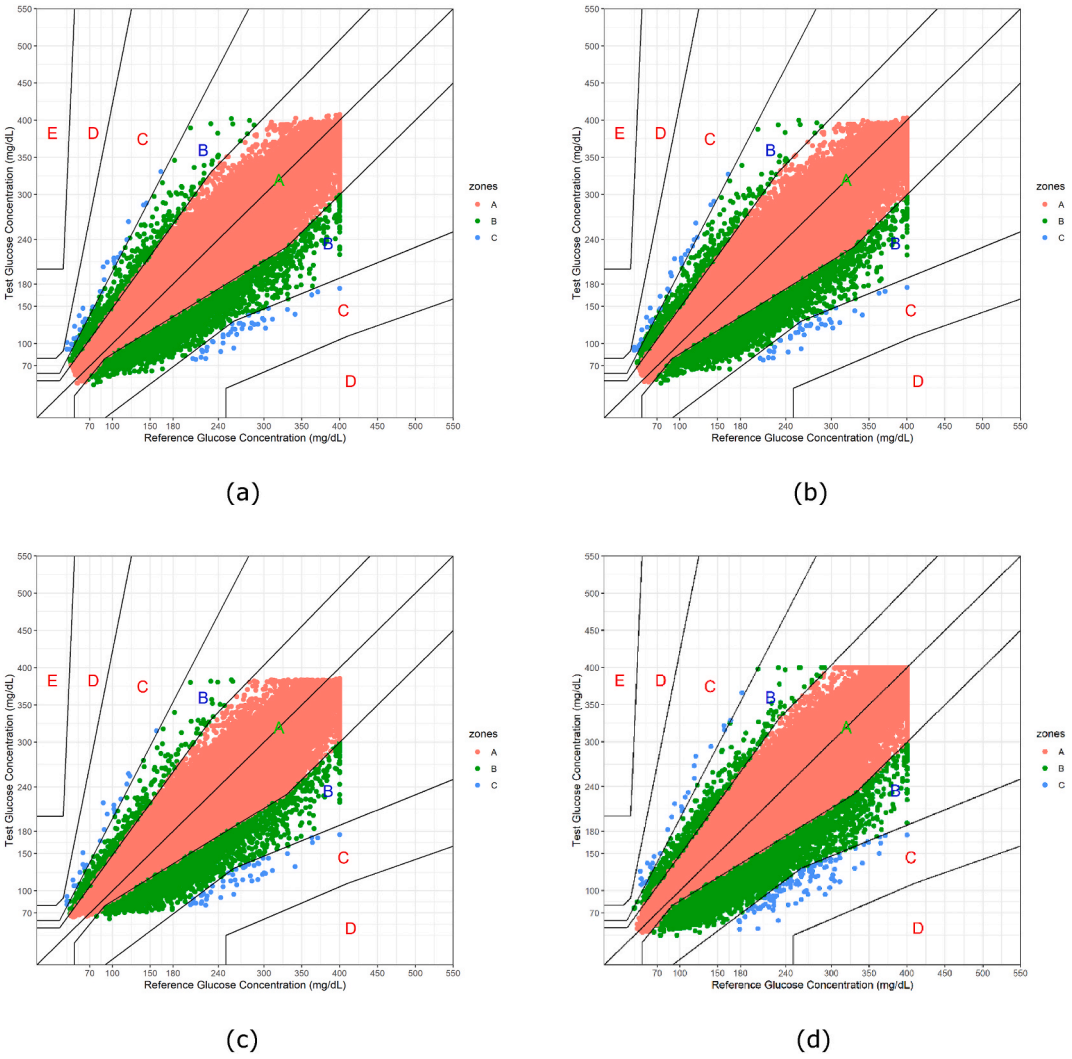

Fig. 9. Parkes error grid results for different forecasting methods: (a) Linear, (b) Lasso, (c) Soft, (d) Naïve method.

input values and are robust against variations in input data, thereby enhancing their applicability in real-world scenarios.

However, the present study, which focuses on the experimental application of stacking models, suggests that the realm of purely theoretical validation requires further research. While such theoretical investigations are foundational in understanding the mathematical and sensitivity analysis aspects of machine learning [55–58], empirical validation has not been conducted in a manner that the findings of the present study suggest. This discrepancy underscores a crucial limitation of the current work—the need for a more nuanced approach that bridges the gap between theoretical robustness and empirical applicability. To address this limitation and reinforce the credibility of the findings obtained in this study, future research will focus on conducting comprehensive empirical analyses. This will involve collecting and analyzing real-world data sets to validate the practical effectiveness of stacking models in mitigating input sensitivity and enhancing prediction accuracy.

6. Conclusions

In this study, the Glu-Ensemble framework was introduced, designed to predict blood glucose levels in patients with T2DM by leveraging ensemble techniques within deep learning predictive models. The empirical evidence demonstrated that the proposed framework surpasses the naïve approach in predictive accuracy, as measured by RMSE and MAE, and also excels in clinical relevance when assessed through EGA. Notably, the soft ensemble method, which averages predictions from 90 base models, was identified as the most effective, yielding the highest accuracy across RMSE, MAE, and EGA metrics. These findings advocate for the Glu-Ensemble model's utility as a superior alternative to the naïve method for forecasting blood glucose levels in T2DM patients, offering a more dependable tool for managing this condition.

The Glu-Ensemble framework, while promising, requires enhancements to address certain limitations. The training dataset

comprised blood glucose readings ranging from 60 to 400 mg/dL, yet real-life scenarios may present values outside this range, either <60 or >400 mg/dL. Moreover, the framework's reliance on a singular, integrated model, as opposed to patient-specific models, may lead to suboptimal performance in cases with a limited number of patients. Enhanced predictive accuracy is anticipated with the inclusion of more extensive datasets for training. Another challenge pertains to the model's adaptability to individual patient variations. The ensemble methods currently employed may not yield accurate predictions for patients whose characteristics significantly diverge from the mean of the training dataset. To improve precision and reliability across a diverse patient spectrum, it may be necessary to develop personalized models or integrate patient-specific factors into the Glu-Ensemble framework, thereby ensuring more tailored and effective blood glucose predictions.

Funding

The following are results of a study on the Leaders in INdustry-university Cooperation 3.0 Project, supported by the Ministry of Education and National Research Foundation of Korea (No. 1345356224). This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2022R1A2C1092808), and this work was supported by the Soonchunhyang University Research Fund.

Data availability statement

Has data associated with your study been deposited into a publicly available repository? Yes. All data from this study are in the article, code, supplementary material, or referenced. Datasets are available on GitHub at https://github.com/SCH-YcHan/Glu-Ensemble or upon request from the author.

CRediT authorship contribution statement

Yechan Han: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Dae-Yeon Kim:** Formal analysis, Data curation, Conceptualization. **Jiyoung Woo:** Writing – review & editing, Resources, Formal analysis, Data curation, Conceptualization. **Jaeyun Kim:** Writing – review & editing, Validation, Supervision, Resources, Methodology, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] American Diabetes Association, Diagnosis and classification of diabetes mellitus, Diabetes Care 32 (Suppl 1) (2009) S62–S67, https://doi.org/10.2337/dc09-S062.
[2] R. Balaji, R. Duraisamy, M.P. Kumar, Complications of diabetes mellitus: a review, Drug Invent, Today Off. 12 (2019).
[3] L. Castaño, G.S. Eisenbarth, Type-I diabetes: a chronic autoimmune disease of human, mouse, and rat, Annu. Rev. Immunol. 8 (1990) 647–679, https://doi.org/10.1146/annurev.iy.08.040190.003243.
[4] M.P. Morran, G.S. Omenn, M. Pietropaolo, Immunology and genetics of type 1 diabetes, Mt. Sinai J. Med. 75 (2008) 314–327, https://doi.org/10.1002/msj.20052.
[5] B.B. Kahn, J.S. Flier, Obesity and insulin resistance, J. Clin. Invest. 106 (2000) 473–481, https://doi.org/10.1172/JCI10842.
[6] S.E. Kahn, R.L. Hull, K.M. Utzschneider, Mechanisms linking obesity to insulin resistance and type 2 diabetes, Nature 444 (2006) 840–846, https://doi.org/10.1038/nature05482.
[7] N. Lascar, J. Brown, H. Pattison, A.H. Barnett, C.J. Bailey, S. Bellary, Type 2 diabetes in adolescents and young adults, Lancet Diabetes Endocrinol. 6 (2018) 69–80, https://doi.org/10.1016/S2213-8587(17)30186-9.
[8] L.J. Bennion, S.M. Grundy, Effects of diabetes mellitus on cholesterol metabolism in man, N. Engl. J. Med. 296 (1977) 1365–1371, https://doi.org/10.1056/NEJM197706162962401.
[9] O. Langer, Y. Yogev, O. Most, E.M. Xenakis, Gestational diabetes: the consequences of not treating, Am. J. Obstet. Gynecol. 192 (2005) 989–997, https://doi.org/10.1016/j.ajog.2004.11.039.
[10] D. Mennickent, A. Rodríguez, M. Farías-Jofré, J. Araya, E. Guzmán-Gutiérrez, Machine learning-based models for gestational diabetes mellitus prediction before 24–28 weeks of pregnancy: a review, Artif. Intell. Med. 132 (2022) 102378, https://doi.org/10.1016/j.artmed.2022.102378.
[11] J.L. Harding, M.E. Pavkov, D.J. Magliano, J.E. Shaw, E.W. Gregg, Global trends in diabetes complications: a review of current evidence, Diabetologia 62 (2019) 3–16, https://doi.org/10.1007/s00125-018-4711-2.
[12] J.W. Russell, L.A. Zilliox, Diabetic neuropathies, continuum lifelong learn, Neurol. 20 (5 Peripheral Nervous System Disorders) (2014) 1226–1240, https://doi.org/10.1212/01.CON.0000455884.29545.d2.
[13] A. Vinik, J. Ullal, H.K. Parson, C.M. Casellini, Diabetic neuropathies: clinical manifestations and current treatment options, Nat. Clin. Pract. Endocrinol. Metabol. 2 (2006) 269–281, https://doi.org/10.1038/ncpendmet0142.
[14] R.Z. Alicic, M.T. Rooney, K.R. Tuttle, Diabetic kidney disease: challenges, progress, and possibilities, Clin. J. Am. Soc. Nephrol. 12 (2017) 2032–2045, https://doi.org/10.2215/CJN.11491116.
[15] H. Ahsan, Diabetic retinopathy–biomolecules and multiple pathophysiology, Diabetes Metabol. Syndr. 9 (2015) 51–54, https://doi.org/10.1016/j.dsx.2014.09.011.
[16] S.O. Butler, I.F. Btaiche, C. Alaniz, Relationship between hyperglycemia and infection in critically ill patients, Pharmacotherapy 25 (2005) 963–976, https://doi.org/10.1592/phco.2005.25.7.963.
[17] A.P. Furnary, K.J. Zerr, G.L. Grunkemeier, A. Starr, Continuous intravenous insulin infusion reduces the incidence of deep sternal wound infection in diabetic patients after cardiac surgical procedures, Ann. Thorac. Surg. 67 (1999) 352–360, https://doi.org/10.1016/s0003-4975(99)00014-4.

[18] G. Gazal, Management of an emergency tooth extraction in diabetic patients on the dental chair, Saudi Dent. J. 32 (2020) 1–6, https://doi.org/10.1016/j.sdentj.2019.07.004.
[19] A.A. Aljumah, M.G. Ahamad, M.K. Siddiqui, Application of data mining: diabetes health care in young and old patients, J. King Saud Univ. Comput. Inf. Sci. 25 (2013) 127–136, https://doi.org/10.1016/j.jksuci.2012.10.003.
[20] P.M. Desai, M.E. Levine, D.J. Albers, L. Mamykina, Pictures worth a thousand words: reflections on visualizing personal blood glucose forecasts for individuals with type 2 diabetes, in: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, 2018, pp. 1–13, https://doi.org/10.1145/3173574.3174112.
[21] A.Z. Woldaregay, E. Årsand, S. Walderhaug, D. Albers, L. Mamykina, T. Botsis, G. Hartvigsen, Data-driven modeling and prediction of blood glucose dynamics: machine learning applications in type 1 diabetes, Artif. Intell. Med. 98 (2019) 109–134, https://doi.org/10.1016/j.artmed.2019.07.007.
[22] M. Bernardini, M. Morettini, L. Romeo, E. Frontoni, L. Burattini, Early temporal prediction of type 2 diabetes risk condition from a general practitioner electronic health record: a multiple instance boosting approach, Artif. Intell. Med. 105 (2020) 101847, https://doi.org/10.1016/j.artmed.2020.101847.
[23] S. Ellahham, Artificial intelligence: the future for diabetes care, Am. J. Med. 133 (2020) 895–900, https://doi.org/10.1016/j.amjmed.2020.03.033.
[24] J. Vehí, I. Contreras, S. Oviedo, L. Biagi, A. Bertachi, Prediction and prevention of hypoglycaemic events in type-1 diabetic patients using machine learning, Health Inf. J. 26 (2020) 703–718, https://doi.org/10.1177/1460458219850682.
[25] I.S. Dasanayake, D.E. Seborg, J.E. Pinsker, F.J. Doyle, E. Dassau, Empirical dynamic model identification for blood-glucose dynamics in response to physical activity, in: Proc. IEEE Conf. Decis. Control, Proc. IEEE Conf. Decis. Control 54th IEEE Conference on Decision and control, (CDC), IEEE Publications, 2015, pp. 3834–3839, https://doi.org/10.1109/CDC.2015.7402815.
[26] C. Novara, N.M. Pour, T. Vincent, G. Grassi, A nonlinear blind identification approach to modeling of diabetic patients, IEEE Trans. Control Syst. Technol. 24 (2015) 1092–1100, https://doi.org/10.1109/TCST.2015.2462734.
[27] P. Tkachenko, G. Kriukova, M. Aleksandrova, O. Chertov, E. Renard, S.V. Pereverzyev, Prediction of nocturnal hypoglycemia by an aggregation of previously known prediction approaches: proof of concept for clinical application, Comput. Methods Progr. Biomed. 134 (2016) 179–186, https://doi.org/10.1016/j.cmpb.2016.07.003.
[28] T. Hamdi, J.B. Ben Ali, V. Di Costanzo, F. Fnaiech, E. Moreau, J.M. Ginoux, Accurate prediction of continuous blood glucose based on support vector regression and differential evolution algorithm, Biocybern. Biomed. Eng. 38 (2018) 362–372, https://doi.org/10.1016/j.bbe.2018.02.005.
[29] I. Rodríguez-Rodríguez, I. Chatzigiannakis, J.V. Rodríguez, M. Maranghi, M. Gentili, M.Á. Zamora-Izquierdo, Utility of big data in predicting short-term blood glucose levels in type 1 diabetes mellitus through machine learning techniques, Sensors (Basel) 19 (2019) 4482, https://doi.org/10.3390/s19204482.
[30] J. Xie, Q. Wang, Benchmarking machine learning algorithms on blood glucose prediction for type I diabetes in comparison with classical time-series models, IEEE Trans. Bio Med. Eng. 67 (2020) 3101–3124, https://doi.org/10.1109/TBME.2020.2975959.
[31] A. Zale, N. Mathioudakis, Machine learning models for inpatient glucose prediction, Curr. Diabetes Rep. 22 (2022) 353–364, https://doi.org/10.1007/s11892-022-01477-w.
[32] Y. Deng, L. Lu, L. Aponte, A.M. Angelidi, V. Novak, G.E. Karniadakis, C.S. Mantzoros, Deep transfer learning and data augmentation improve glucose levels prediction in type 2 diabetes patients, npj Digit. Med. 4 (2021) 109, https://doi.org/10.1038/s41746-021-00480-x.
[33] M. Jaloli, M. Cescon, Long-term prediction of blood glucose levels in type 1 diabetes using a cnn-lstm-based deep neural network, J. Diabetes Sci. Technol. 17 (2022) 1590–1601, https://doi.org/10.1177/19322968221092785.
[34] K. Li, J. Daniels, C. Liu, P. Herrero, P. Georgiou, Convolutional recurrent neural networks for glucose prediction, IEEE J. Biomed. Health Inform 24 (2020) 603–613, https://doi.org/10.1109/JBHI.2019.2908488.
[35] K. Li, C. Liu, T. Zhu, P. Herrero, P. Georgiou, GluNet: a deep learning framework for accurate glucose forecasting, IEEE J. Biomed. Health Inform 24 (2020) 414–423, https://doi.org/10.1109/JBHI.2019.2931842.
[36] J. Martinsson, A. Schliep, B. Eliasson, O. Mogren, Blood glucose prediction with variance estimation using recurrent neural networks, J. Healthc. Inform. Res. 4 (2020) 1–18, https://doi.org/10.1007/s41666-019-00059-y.
[37] J. Chaki, S.T. Ganesh, S.K. Cidham, S.A. Theertan, Machine learning and artificial intelligence based Diabetes Mellitus detection and self-management: a systematic review, J. King Saud Univ.-Comput. Inform. Sci. 34 (2022) 3204–3225.
[38] H. Khadem, H. Nemat, J. Elliott, M. Benaissa, Blood glucose level time series forecasting: nested deep ensemble learning lag fusion, Bioengineering 10 (4) (2023) 487, https://doi.org/10.3390/bioengineering10040487.
[39] O. Fitzgerald, O. Perez-Concha, B. Gallego-Luxan, A. Metke-Jimenez, L. Rudd, L. Jorm, Continuous time recurrent neural networks: overview and benchmarking at forecasting blood glucose in the intensive care unit, J. Biomed. Inf. 146 (2023) 104498, https://doi.org/10.1016/j.jbi.2023.104498.
[40] A. Prieto, B. Prieto, E.M. Ortigosa, E. Ros, F. Pelayo, J. Ortega, I. Rojas, Neural networks: an overview of early research, current frameworks and new challenges, Neurocomputing 214 (2016) 242–268, https://doi.org/10.1016/j.neucom.2016.06.014.
[41] H. Wang, Z. Lei, X. Zhang, B. Zhou, J. Peng, A review of deep learning for renewable energy forecasting, Energy Convers. Manag. 198 (2019) 111799, https://doi.org/10.1016/j.enconman.2019.111799.
[42] V. Felizardo, N.M. Garcia, N. Pombo, I. Megdiche, Data-based algorithms and models using diabetics real data for blood glucose and hypoglycaemia prediction–a systematic literature review, Artif. Intell. Med. 118 (2021) 102120, https://doi.org/10.1016/j.artmed.2021.102120.
[43] J.F. Torres, D. Hadjout, A. Sebaa, F. Martínez-Álvarez, A. Troncoso, Deep learning for time series forecasting: a survey, Big Data 9 (2021) 3–21, https://doi.org/10.1089/big.2020.0159.
[44] M.Z. Wadghiri, A. Idri, T. El Idrissi, H. Hakkoum, Ensemble blood glucose prediction in diabetes mellitus: a review, Comput. Biol. Med. 147 (2022) 105674, https://doi.org/10.1016/j.compbiomed.2022.105674.
[45] W.L. Clarke, S. Anderson, L. Farhy, M. Breton, L. Gonder-Frederick, D. Cox, B. Kovatchev, Evaluating the clinical accuracy of two continuous glucose sensors using continuous glucose–error grid analysis, Diabetes Care 28 (2005) 2412–2417, https://doi.org/10.2337/diacare.28.10.2412.
[46] L. Breiman, Stacked regressions, Mach. Learn. 24 (1996) 49–64, https://doi.org/10.1007/BF00117832.
[47] X. Li, J. Luo, X. Jin, Q. He, Y. Niu, Improving soil thickness estimations based on multiple environmental variables with stacking ensemble methods, Rem. Sens. 12 (21) (2020) 3609, https://doi.org/10.3390/rs12213609.
[48] J.L. Parkes, S.L. Slatin, S. Pardo, B.H. Ginsberg, A new consensus error grid to evaluate the clinical significance of inaccuracies in the measurement of blood glucose, Diabetes Care 23 (2000) 1143–1148, https://doi.org/10.2337/diacare.23.8.1143.
[49] L. Zhang, R. Wang, Z. Li, J. Li, Y. Ge, S. Wa, S. Huang, C. Lv, Time-series neural network: a high-accuracy time-series forecasting method based on Kernel filter and time attention, Information 14 (9) (2023), https://doi.org/10.3390/info14090500, 500.
[50] A. Pfützner, D.C. Klonoff, S. Pardo, J.L. Parkes, Technical aspects of the Parkes error grid, J. Diabetes Sci. Technol. 7 (2013) 1275–1281, https://doi.org/10.1177/193229681300700517.
[51] B. Pavlyshenko, Using stacking approaches for machine learning models, in: 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP), 2018, pp. 255–258, https://doi.org/10.1109/DSMP.2018.8478522.
[52] J. Sill, G. Takács, L. Mackey, D. Lin, Feature-weighted linear stacking, arXiv preprint arXiv:0911.0460, https://doi.org/10.48550/arXiv.0911.0460, 2009.
[53] H. Butt, I. Khosa, M.A. Iftikhar, Feature transformation for efficient blood glucose prediction in type 1 diabetes mellitus patients, Diagnostics 13 (3) (2023) 340, https://doi.org/10.3390/diagnostics13030340.
[54] F. Prendin, J. Pavan, G. Cappon, S. Del Favero, G. Sparacino, A. Facchinetti, The importance of interpreting machine learning models for blood glucose prediction in diabetes: an analysis using SHAP, Sci. Rep. 13 (2023) 16865, https://doi.org/10.1038/s41598-023-44155-x.
[55] P. Zhang, A novel feature selection method based on global sensitivity analysis with application in machine learning-based prediction model, Appl. Soft Comput. 85 (2019) 105859, https://doi.org/10.1016/j.asoc.2019.105859.

[56] E. Dupuis, D. Novo, I. O'Connor, A. Bosio, Sensitivity analysis and compression opportunities in dnns using weight sharing, April, in: 2020 23rd International Symposium on Design and Diagnostics of Electronic Circuits & Systems, DDECS) IEEE, 2020, pp. 1–6, https://doi.org/10.1109/DDECS50862.2020.9095658, 2020.
[57] R. Asheghi, S.A. Hosseini, M. Saneie, A.A. Shahri, Updating the neural network sediment load models using different sensitivity analysis methods: a regional application, J. Hydroinf. 22 (2020) 562–577, https://doi.org/10.2166/hydro.2020.098.
[58] A.A. Shahri, C. Shan, S. Larsson, A novel approach to uncertainty quantification in groundwater table modeling by automated predictive deep learning, Nat. Resour. Res. 31 (2022) 1351–1373, https://doi.org/10.1007/s11053-022-10051-w.