

ORIGINAL RESEARCH ARTICLE

A clinically weighted loss function for risk-sensitive blood glucose forecasting

Joseph Passant^{1*} and Andrea Corradini^{2*}¹Department of Computer Science, School of Computer Science and Informatics, University of Liverpool, Liverpool, United Kingdom²Department of Digital Business & Software Engineering, Management Center Innsbruck, Innsbruck, Austria

Abstract

Blood glucose forecasting models have the potential to improve glycemic control and reduce the burden of diabetes management. However, most prior work has focused on optimizing statistical accuracy, which does not necessarily align with clinical safety and efficacy. This study introduces a clinically weighted root mean squared error (CW-RMSE) loss function that incorporates clinical risk into model optimization by weighting prediction errors according to continuous glucose error grid analysis (CG-EGA) classification and glycemic range to reduce clinically harmful errors. WinFormer, an Informer-based prediction model employing windowed attention, was trained using the REPLACE-BG dataset and either root mean squared error (RMSE) or CW-RMSE loss functions to predict blood glucose values over a 120-min prediction horizon. Compared to RMSE loss, CW-RMSE loss produced significant improvements in CG-EGA classification in hypoglycemia (accurate predictions: 83.2% to 88.0%; erroneous predictions: 13.2% to 8.1%) while maintaining performance for predictions of values in euglycemic (accurate predictions: 92.6% to 92.4%; erroneous predictions: 2.7% to 2.4%) and hyperglycemic ranges (accurate predictions: 91.2% to 90.6%; erroneous predictions: 4.1% to 4.1%). In comparison to RMSE loss, CW-RMSE loss also led to significant improvements in RMSE and mean absolute error (MAE) in hypoglycemia (RMSE: 39.8 mg/dL to 29.8 mg/dL; MAE: 26.6 mg/dL to 16.4 mg/dL), but worse RMSE and MAE in the euglycemia range (RMSE: 31.8 mg/dL to 33.8 mg/dL; MAE: 23.1 mg/dL to 25.1 mg/dL) and the hyperglycemia range (RMSE: 44.9 mg/dL to 52.3 mg/dL; MAE: 30.0 mg/dL to 34.8 mg/dL). These findings demonstrate that cost-sensitive learning can effectively guide blood glucose forecasting models toward more clinically meaningful and risk-aware optimization.

***Corresponding authors:**Joseph Passant
(jgpassant@proton.me);
Andrea Corradini
(andrea.corradini@mci.edu)

Citation: Passant J, Corradini A.
A clinically weighted loss function
for risk-sensitive blood glucose
forecasting. *Artif Intell Health*.
2026;3(4):026030003.
doi: 10.36922/AIH026030003

Received: January 12, 2026**Revised:** February 12, 2026**Accepted:** February 25, 2026**Published online:** May 26, 2026

Copyright: © 2026 Author(s).
This is an Open-Access article
distributed under the terms of the
Creative Commons Attribution
License, permitting distribution,
and reproduction in any medium,
provided the original work is
properly cited.

Publisher's Note: AccScience
Publishing remains neutral with
regard to jurisdictional claims in
published maps and institutional
affiliations.

Keywords: Blood glucose forecasting; Machine learning; Cost-sensitive learning; Type 1 diabetes

1. Introduction

Diabetes is a broad class of conditions that affect an individual's ability to regulate their blood glucose (BG). Type 1 diabetes (T1D) is a distinct condition that results from an autoimmune response that destroys insulin-producing pancreatic β -cells, resulting in absolute insulin deficiency and the need for lifelong exogenous insulin therapy to

maintain BG levels within a healthy euglycemic range (70–179 mg/dL), avoiding both hypoglycemia (≤ 69 mg/dL) and hyperglycemia (≥ 180 mg/dL).¹ People with T1D should spend at least 70% of their time in the target glucose range and less than 4% of their time in hypoglycemia.² However, optimal glycemic control is extremely challenging, with only 25% of people with diabetes meeting glycemic targets.³ Poor glycemic control can increase the risk of both acute and chronic complications. Hypoglycemia initially presents with symptoms such as sweating, rapid heart rate, light-headedness, confusion, tremors, and visual changes; if left untreated, it can progress to coma and death.¹ In contrast, hyperglycemia increases the risk of long-term complications such as nephropathy, neuropathy, retinopathy, cardiovascular disease, and diabetic foot infections. Sustained hyperglycemia may lead to diabetic ketoacidosis, an acute and potentially life-threatening condition characterized by hyperglycemia, ketonuria, electrolyte disturbances, and metabolic acidosis.¹ T1D also has a profound impact on mental health and quality of life. Hypo- and hyperglycemic events are disruptive and burdensome, and fear of hypoglycemia and concern about long-term complications, combined with the challenge of balancing management of T1D with everyday life, contribute to coping difficulties and an increased prevalence of anxiety and depression among people with T1D.^{4–6}

Continuous glucose monitors (CGMs) measure glucose concentration within the interstitial fluid between cells and use it to estimate BG, typically providing readings every five minutes. CGMs have become standard in diabetes management,¹ largely replacing direct BG measurements through “finger prick” blood sampling. CGMs have been shown to reduce hemoglobin A1c, increase time in range, and reduce time in hypoglycemia.⁷ CGM users also report improved diabetes-related quality of life, which is attributed in part to quicker responses to hypo- and hyperglycemia.⁷ Hybrid closed-loop insulin delivery systems are also becoming increasingly established. These devices integrate CGM data with an algorithm-controlled insulin pump to automatically adjust insulin delivery in response to changes in BG.

The near-real-time insights into BG trends and patterns available from CGM devices offer an opportunity for further improvement of T1D management. For example, the development of predictive algorithms could provide BG forecasts, enabling proactive, rather than reactive, management decisions. This advancement holds potential to reduce the frequency and severity of glycemic excursions, thereby reducing negative effects on physical health, and alleviating stress and anxiety associated with

T1D management.⁸

Several factors challenge the development and evaluation of BG prediction models, and addressing these is necessary to achieve clinically acceptable performance. BG data is inherently imbalanced. Euglycemic values represent the majority class, hyperglycemic values occur less frequently, and hypoglycemic values represent a significant minority class. If models do not address this imbalance, they may overfit to the majority class and perform poorly on minority classes. Importantly, clinical risk also varies across glycemic ranges. Hypoglycemia poses an acute and immediate threat to life, while hyperglycemia poses longer-term health risks. Prediction errors in these regions may cause incorrect treatment and management decisions that exacerbate risk and undermine a model’s clinical efficacy. In contrast, prediction errors for euglycemic values are likely to be less consequential.

A further challenge lies in achieving high model performance over longer prediction horizons (PHs). Early studies primarily focused on short PHs (e.g., 30 min), but more recent work increasingly targets PHs of up to 120 min. Extending the PH may offer additional clinical benefit; Ehrmann *et al.*⁹ reported that people with T1D believe predictions provided 120 min in advance would reduce fear of hypoglycemia and diabetes-related distress to a greater extent than predictions provided only 30 minutes in advance.

A significant advantage of CGMs is the frequency of BG readings, which not only affords individuals knowledge of their current BG but also the patterns and trends in BG over time. This contextual information is important for management decisions. Multi-horizon BG prediction models are also likely to provide greater clinical benefit because they offer richer contextual information (i.e., predicting each data point at five-minute intervals up to the PH) than models that predict only a single future time point. However, multi-horizon forecasting over prolonged PHs compounds difficulties created by class imbalance during model development. Hypoglycemic episodes are typically treated and corrected rapidly, implying that individuals spend relatively little time in this range during each episode. A multi-horizon model with a 120-min PH predicts 24 consecutive time points, yet a hypoglycemic excursion might account for only a few consecutive data points. As a result, not only are hypoglycemic sequences underrepresented overall, but the proportion of hypoglycemic data within those sequences is also underrepresented compared to euglycemic and hyperglycemic data points. This creates a dual imbalance that challenges sequence classification (e.g., determining the threshold for labeling a sequence as hypo- or

hyperglycemic) and the effectiveness of traditional balancing strategies.

This study aims to develop and evaluate a novel cost-sensitive learning (CSL) loss function that addresses this class and risk imbalance by penalizing prediction errors more heavily in hypoglycemic and hyperglycemic ranges as well as clinically harmful predictions across all glycemic regions, and to assess its impact on both statistical and clinical measures of performance compared to standard root mean squared error (RMSE) loss.

2. Related work

Researchers have studied both machine learning techniques and neural network-based approaches for BG prediction, with numerous comparative studies evaluating their relative performance. Liu *et al.*¹⁰ conducted a comprehensive systematic review and meta-analysis of 56 BG prediction models across 30 studies to assess the comparative performance of the included models. Their analysis revealed that neural networks, including recurrent neural networks (RNNs), GluNet, ARTiDe jump, and standard neural network models, achieved the lowest RMSE values across varying PHs and concluded that neural network-based algorithms consistently outperform simpler machine learning approaches. More recent work has highlighted the specific potential of self-attention networks.¹¹ A variety of self-attention networks adapted from Transformer¹² and Informer^{13,14}-based architectures have been developed for BG prediction. Sergazinov *et al.*¹⁵ reported that gluformer, which combined a Transformer architecture with an infinite mixture model, outperformed models such as autoregressive integrated moving average (ARIMA), RNN, and random forest. Zhu *et al.*¹⁶ applied an edge-based temporal fusion transformer for multi-horizon prediction of BG levels, showing notable improvements in accuracy at 30- and 60-min PHs, and outperforming support vector regression (SVR), long short-term memory (LSTM), and deep RNN models. TimeGlu employed a novel encoder-decoder architecture incorporating bi-LSTM and additive attention mechanisms to extract high-dimensional features, and it outperformed double exponential smoothing, ARIMA, BATS, and TBATS in terms of mean absolute error (MAE) and mean absolute percentage error.¹⁷ Zhu *et al.*¹⁸ developed GPFormer, an Informer-based model designed for multi-horizon glucose prediction, which achieved lower RMSE for all PHs and data sets compared to baseline models (N-BEATS, Bi-LSTM, SVR, and ARIMA). Lastly, Xue *et al.*¹⁹ introduced BGFormer, another specialized Informer-based architecture, which led to lower RMSE compared to both “Vanilla” Transformer and Informer models at 60- and 90-min PHs.

Most studies optimize for and report on statistical measures of loss, with RMSE and MSE appearing to be among the most commonly used loss functions and RMSE being a *de facto* choice for model evaluation and comparison.^{10,20} MSE and RMSE penalize larger errors more heavily but assume all errors carry equal consequence. This is inappropriate for BG forecasting, where the clinical significance of prediction errors varies across glycemic ranges. MSE and RMSE are further confounded by the inherent class imbalance in BG data. By seeking to minimize global error across all training points, MSE and RMSE will naturally favor performance in the euglycemic range, at the expense of hypoglycemia and hyperglycemia. As a result, models may achieve strong global statistical performance while potentially masking limitations in clinical efficacy.

To ensure clinical efficacy and safety, evaluation must include clinical metrics that reflect the potential impact of prediction error on T1D management decisions. Clarke error grid analysis (CEGA) categorizes glucose estimates into zones based on clinical risk²¹ and has been widely used to assess the clinical accuracy of BG meters as well as the evaluation of BG prediction models. This grid categorizes individual BG estimates into five zones according to the degree of clinical risk that would result from acting on an erroneous reading. However, CEGA was developed for static glucose readings and does not account for the direction and rate of change of BG, which are essential considerations for continuous BG data. Continuous glucose error grid analysis (CG-EGA) was developed to extend CEGA to specifically quantify the clinical accuracy of CGM systems.²² It achieves this through combining the assessment of both the error of the predicted value compared to the reference value and the error of the predicted rate of change compared to the reference rate of change. CG-EGA has since been proposed as an important metric in the evaluation of BG forecasting models.²⁰ Although CG-EGA incorporates glycemic range into its classification, studies often report only global CG-EGA performance summaries, which may mask important differences between hypoglycemic, euglycemic, and hyperglycemic ranges. Evaluation that includes both statistical and clinical metrics across glycemic ranges may offer more meaningful insight into a model’s performance and clinical efficacy.

Over-sampling minority classes or under-sampling majority classes are common strategies to address class imbalance.^{23,24} While these methods can improve data balance, the characteristics of BG data and the demands of multi-horizon forecasting may warrant alternative training approaches. CSL describes a group of algorithms and

strategies designed to account for the misclassification costs of prediction errors. Araf *et al.*²⁵ offers a detailed taxonomy and overview of CSL in medical domains, concluding that CSL techniques enable effective handling of severely class-imbalanced data, leading to more balanced predictions and more equitable performance across classes. One approach to CSL uses instance weighting, where higher weights are assigned to instances where misclassification is of greater consequence.²⁵ Despite this promise, few studies have applied CSL grounded in clinical risk to regression-based tasks such as BG forecasting models. As such, CSL offers a potential solution to the challenges posed by BG data and the limitations of RMSE and MSE loss functions.

3. Data and methods

3.1. Dataset

This study used the REPLACE-BG dataset,^{26,27} which consists of CGM data collected in a six-month randomized trial for 226 people with T1D from 14 endocrinology practices in the United States. The dataset comprises data from 112 females and 114 males, with ages ranging from 19 to 78. Among them, 207 were of White non-Hispanic ethnicity, whereas 19 were from Hispanic, Latino, black or African American, or Asian ethnicity. Data was collected using the Dexcom G4 Platinum Continuous Glucose Monitoring system, which provided measures of interstitial fluid glucose levels in the range of 40 mg/dL to 400 mg/dL at five-minute intervals. The number of CGM data samples for individuals ranged from 16,274 to 94,089.

3.2. Data processing

The REPLACE-BG dataset was split into training, validation, and test sets as follows:

- The train and test sets were first established by splitting datasets for respective individuals in chronological order, using an 80:20 split ratio.
- The validation set was then created from the resulting training set by further dividing it, again using an 80:20 split ratio and preserving the chronological order of the data.

This resulted in an initial training:validation:test set split ratio of 64:16:20. Overlapping slices of 96 consecutive data points, representing 8 h of CGM data, were created for each patient dataset, using a stride of 2 to maximize slices representing the minority hypoglycemic class. Slices containing two or more consecutive missing values were discarded to preserve data quality. Singular missing values were filled using linear interpolation.

Sequences were categorized into glycemic classes based on the following rules applied to the target sequence:

- If the target sequence contained ≥ 6 data points ≤ 69 mg/dL, it was classified as a hypoglycemic sequence
- If the target sequence contained ≥ 6 data points ≥ 180 mg/dL, it was classified as a hyperglycemic sequence
- Otherwise, the sequence was classified as euglycemic

Euglycemic and hyperglycemic sequences were randomly undersampled to achieve a 2:1:2 ratio between hypoglycemic, euglycemic, and hyperglycemic target sequences. BG values were normalized using z-score normalization.

Each respective sequence of 96 data points was divided into encoder inputs (72), decoder inputs (36), and target (24) outputs. The decoder input consists of a start token comprising 12 complete data samples and 24 data points of the target sequence with BG values zeroed. The target length of 24 data samples represents a PH of 120 minutes. After data processing, the training set consisted of 275,485 sequences, the validation set 127,568 sequences, and the test set 141,896 sequences.

3.3. Model development

3.3.1. Architecture

This study used the WinFormer model (Figure 1).²⁸ WinFormer is adapted from Informer^{13,14} and BGFormer.¹⁹ The primary differentiating characteristic of WinFormer is the application of windowed attention in self-attention mechanisms in both the encoder and decoder, in comparison to ProbSparse (Informer) and microscale/windowed and dual attention mechanisms (BGFormer). Full attention is employed within the encoder-decoder cross attention. The windowed attention mechanism restricts self-attention computation to fixed-size overlapping local windows. Attention scores for overlapping positions included in more than one window are aggregated using mean aggregation. The same training hyperparameters were used for all model variants to enable direct comparison of training conditions (see Table 1).

3.3.2. Clinically weighted loss function

We developed a clinically weighted RMSE loss function (CW-RMSE), which applied specific weightings based on glycemic region and CG-EGA classification (see Section 3.4.2), as described below:

- Predictions were classified using CG-EGA as accurate predictions (APs), benign errors (BEs), or erroneous predictions (EPs).
- RMSE served as the base loss term.
- A weight of 2.0 was applied to the RMSE value for BEs and 6.0 to EPs.
- A weight of 1.5 was applied to predictions in the

hypoglycemic range in addition to the weighting described in step three, with the application being multiplicative.

- Gradient clipping was applied using an L2-norm threshold of 2.0.

Table 1. Model training hyperparameters

Hyperparameter	Value
Embedding dimensions	128
Feed-forward dimension	1,024
Attention heads	8
Encoder layers	2
Decoder layers	1
Dropout	0.2
Encoder attention window size	12
Encoder attention window stride	6
Decoder attention window size	6
Decoder attention window stride	3
Activation function	GeLU
Epochs	15
Iterations per epoch	500
Batch size	256
Learning rate	0.0001

Weight values were selected heuristically through an exploratory grid search followed by manual refinement, with the primary objective of reducing clinically harmful EPs and enhancing performance in the hypoglycemic range. Weighting of the CG-EGA classes was therefore designed to prioritize the reduction of EPs over BEs while preserving APs. In addition, glycemic range weighting was applied exclusively to hypoglycemic samples to reflect the asymmetry in clinical risk. Gradient clipping was introduced to support training stability and mitigate the risk of hard decision boundaries resulting from CG-EGA classification and glycemic range thresholds.

3.4. Evaluation metrics

3.4.1. Statistical evaluation metrics

To evaluate the statistical performance of the respective

models, RMSE and MAE were used. RMSE measures the square root of the mean of squared differences between predicted and reference values. It penalizes larger errors more heavily than smaller errors. For a set of n predictions with reference values y_i and predictions $\hat{y}_i$, RMSE is defined as:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (1)$$

On the other hand, MAE measures the mean of the absolute differences between predicted and reference values, weighting all errors linearly. MAE is defined as:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2)$$

Lower RMSE and MAE indicate better statistical accuracy of predicted BG values.

Uncertainty of RMSE and MAE was quantified using the standard error of the estimate (SEE). For MAE, SEE was calculated as the standard deviation of per-sample absolute errors divided by the square root of the number of observations. For RMSE, SEE was calculated first by calculating the SEE of the mean squared error and then converting this to the RMSE scale.

3.4.2. Clinical evaluation metrics

Continuous glucose error grid analysis was used to evaluate the clinical performance of respective models by jointly evaluating point accuracy and rate-of-change accuracy following Kovatchev *et al.*²² and adapted for the purposes of this study from de Bois.²⁹

Rate error grid analysis (R-EGA) assesses the difference between predicted and reference rates of BG change between consecutive data points and classifies each prediction into one of eight rate-error categories: A, B, and upper and lower variants of C, D, and E (Figure 2).

Point error grid analysis (P-EGA) classifies each prediction into one of five point-error categories (A to E) based on the deviation between the reference and predicted glucose value relative to the glycemic range (Figure 3). Boundaries of the point-error categories are defined according to the reference rate of BG change. The upper boundaries of A_p , uB_p , and uD_p are expanded by 10 mg/dL if the reference rate of BG change is falling at a rate between -1 and -2 mg/dL/min and by 20 mg/dL when falling faster than -2 mg/dL/min. Similarly, when the reference rate of BG change is increasing at a rate between 1 – 2 mg/dL/min, the lower boundaries of A_p , lB_p , and lD_p are reduced by 10 mg/dL, and by 20 mg/dL when the reference rate of BG change is rising faster than 2 mg/dL/min.

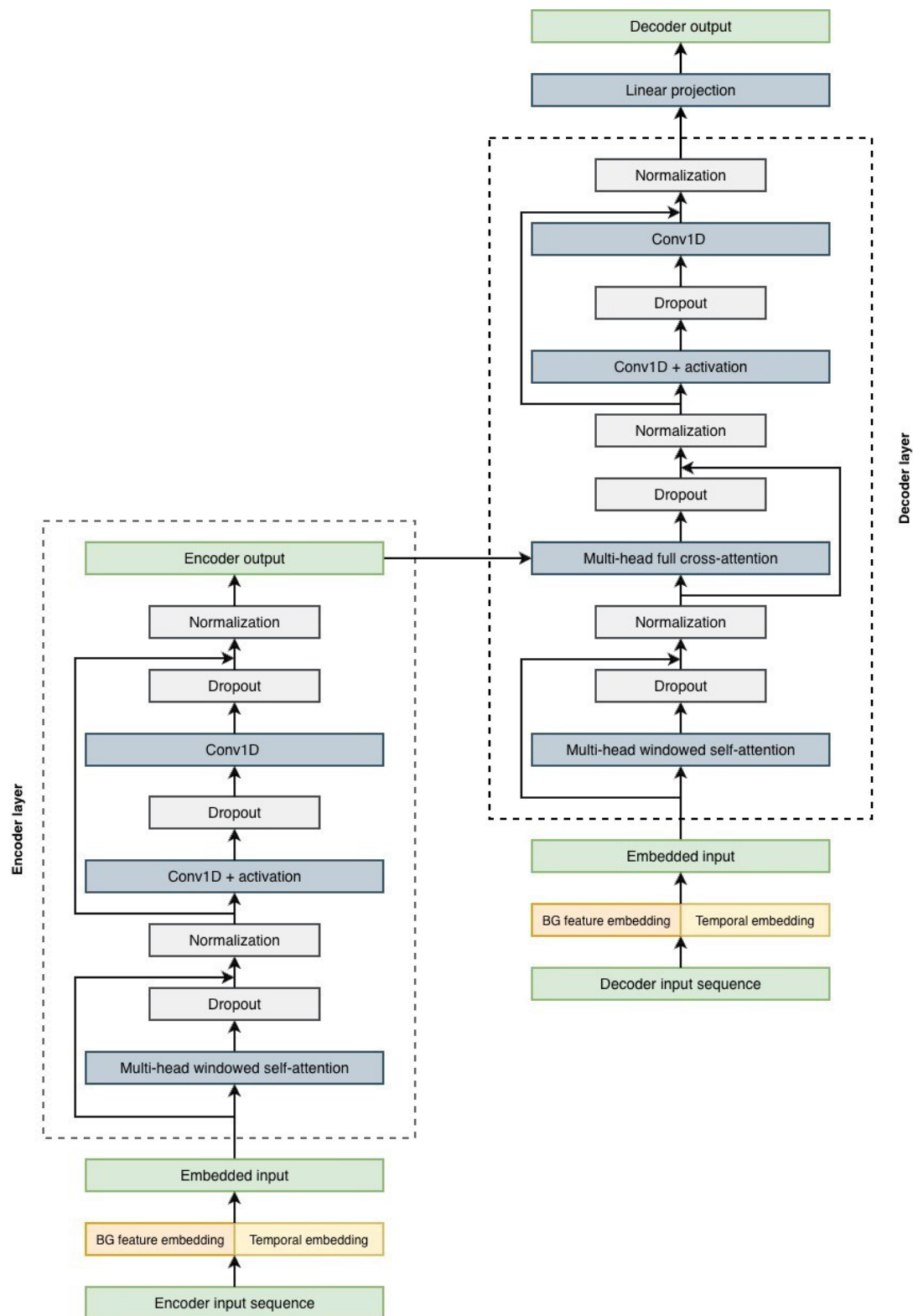


Figure 1. Schematic architecture of WinFormer
Abbreviation: BG: Blood glucose.

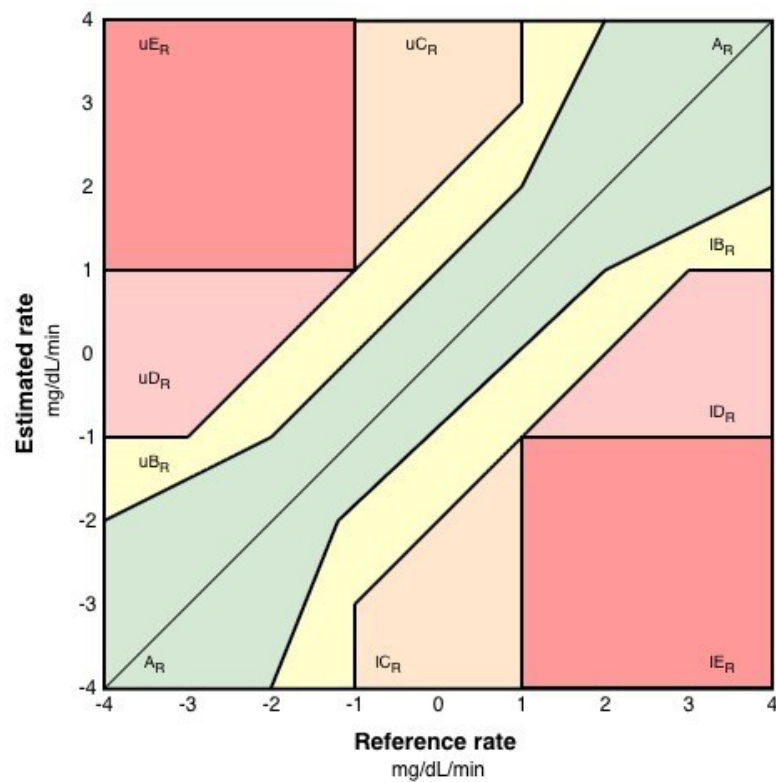


Figure 2. Schematic illustration of the rate error grid

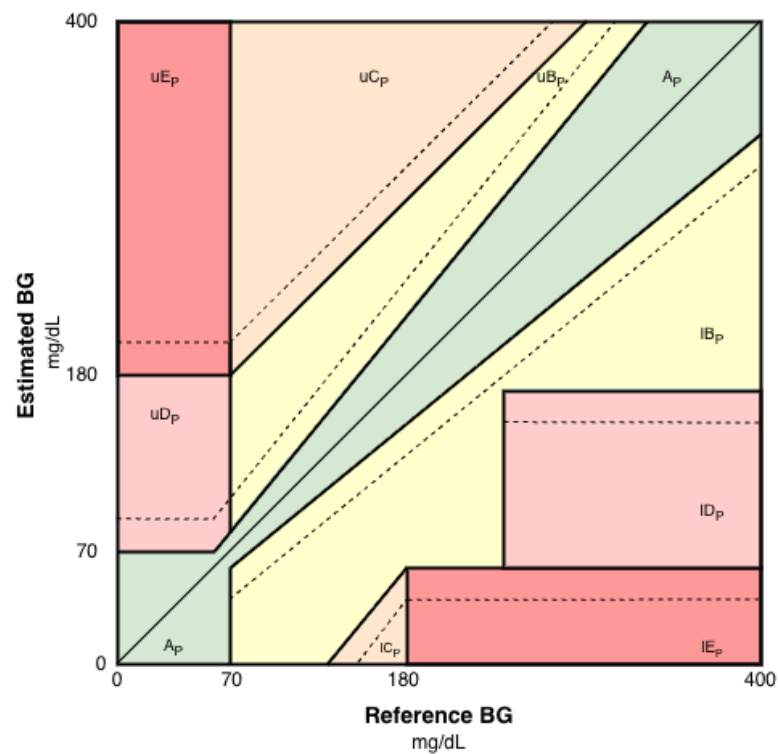


Figure 3. Schematic illustration of the point error grid
Abbreviation: BG: Blood glucose.

Measures of R-EGA and P-EGA are combined to produce a single classification (CG-EGA) for each BG prediction (Figure 4):

- APs would lead to correct treatment decisions.
- BEs would lead to benign or no treatment errors.
- EPs could lead to potentially dangerous treatment decisions.

3.5. Implementation and statistical analysis

This work investigates the influence of CW-RMSE on BG prediction. Two models were implemented for comparative evaluation:

- The baseline model trained using a standard RMSE loss function
- An experimental model trained using the proposed CW-RMSE loss function

Both models were trained and evaluated on identical data using identical hyperparameters and computational environment. Early stopping was not applied to ensure consistent training durations.

We retained model checkpoints achieving the lowest percentage of global EPs on the validation set, consistent with this study's objective of optimizing for clinical rather than purely statistical performance.

All evaluation metrics were assessed globally and across

hypoglycemic (≤ 69 mg/dL), euglycemic (70–179 mg/dL), and hyperglycemic (≥ 180 mg/dL) ranges, and at each five-minute interval over the course of the 120-minute PH to assess glycemic range and time-dependent performance.

Inferential statistical tests were conducted to determine the significance of performance differences between models. Independent *t*-tests were performed on per-sample squared errors and per-sample absolute errors to assess whether model errors differed significantly. Chi-square tests were used to assess differences in CG-EGA classification distributions between models globally and within respective glycemic regions. A *p*-value of 0.05 was used to determine the threshold for statistical significance in all statistical tests.

4. Results

4.1. Statistical evaluation metrics

On the hold-out test set, training with the CW-RMSE resulted in higher global statistical error compared to standard RMSE loss (Figure 5). Global RMSE increased from $36.7 \text{ mg/dL} \pm 0.02$ (SEE) to $40.4 \pm 0.03 \text{ mg/dL}$, and MAE increased from $25.4 \pm 0.01 \text{ mg/dL}$ to $27.8 \pm 0.02 \text{ mg/dL}$. Independent *t*-tests on per-sample squared and absolute errors confirmed these differences were statistically significant (RMSE: $t = -104.6$, $p < 0.0001$; MAE: $t = -113.3$, $p < 0.0001$).

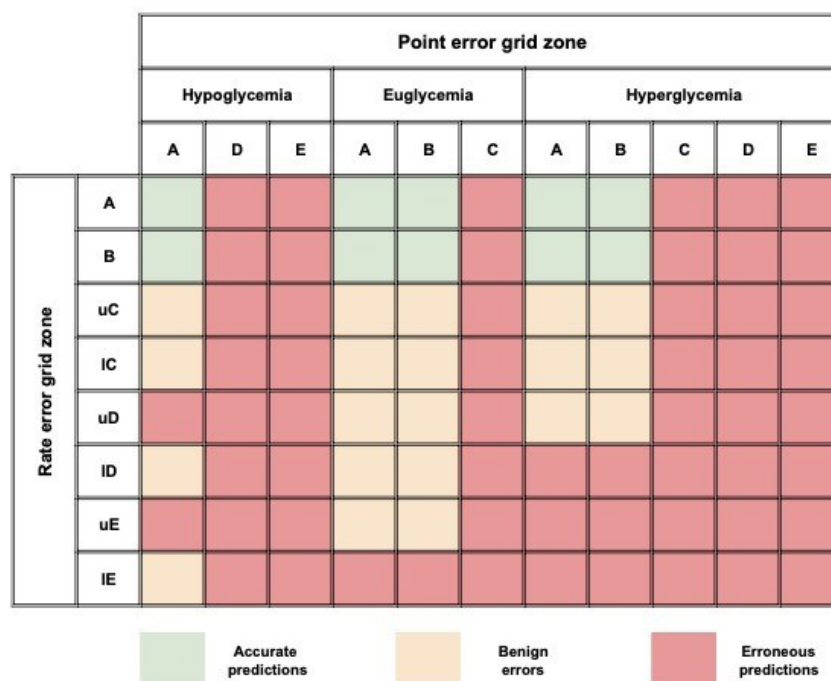


Figure 4. Continuous glucose error grid analysis matrix combining rate error grid analysis and point error grid analysis

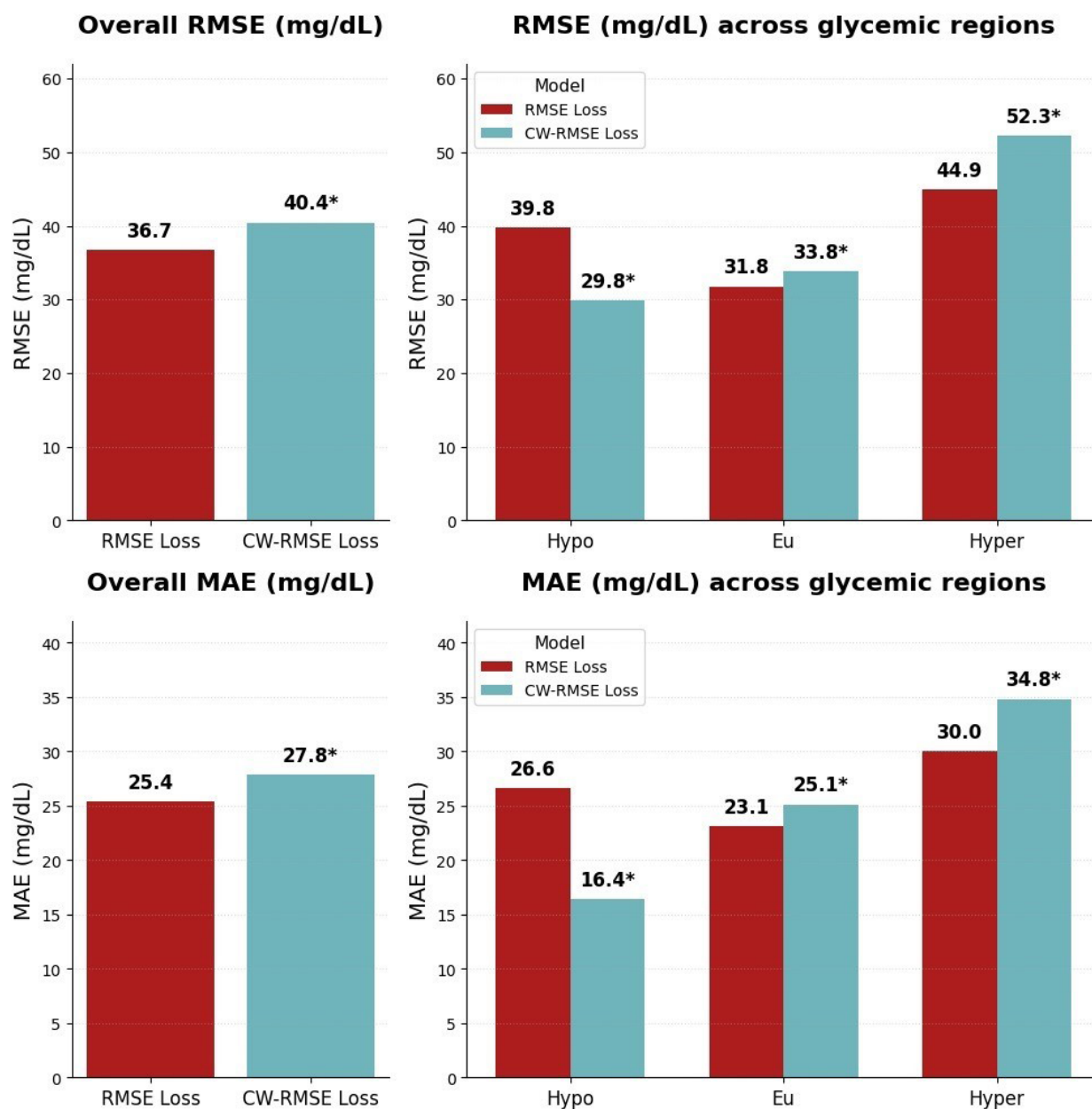


Figure 5. Root mean squared error (RMSE) and mean absolute error (MAE) for WinFormer models trained with RMSE and clinically weighted root mean squared error (CW-RMSE) loss functions. * denotes significant difference vs. RMSE loss determined via independent *t*-test ($p < 0.0001$). Abbreviations: Eu: Euglycemia; Hyper: Hyperglycemia; Hypo: Hypoglycemia.

When analyzed by glycemic range, the CW-RMSE model achieved substantially lower error in the hypoglycemic range, with RMSE decreasing from $39.8 \text{ mg/dL} \pm 0.13$ to $29.8 \pm 0.14 \text{ mg/dL}$ and MAE from $26.6 \pm 0.08 \text{ mg/dL}$ to $16.4 \pm 0.07 \text{ mg/dL}$ (RMSE: $t = 52.3$, $p < 0.0001$; MAE: $t = 97.0$, $p < 0.0001$). In contrast, the CW-RMSE model showed higher error in both the euglycemic (RMSE: $31.8 \pm 0.02 \text{ mg/dL}$ vs. $33.8 \pm 0.02 \text{ mg/dL}$, $t = -70.2$, $p <$

0.0001 ; MAE: $23.1 \pm 0.01 \text{ mg/dL}$ vs. $25.1 \pm 0.02 \text{ mg/dL}$, $t = -96.6$, $p < 0.0001$) and hyperglycemic ranges (RMSE: $44.9 \pm 0.05 \text{ mg/dL}$ vs. $52.3 \pm 0.06 \text{ mg/dL}$, $t = -97.2$, $p < 0.0001$; MAE: $30.0 \pm 0.03 \text{ mg/dL}$ vs. $34.8 \pm 0.04 \text{ mg/dL}$, $t = -96.7$, $p < 0.0001$).

These findings illustrate a trade-off in model optimization: the CW-RMSE loss successfully prioritized accuracy in the safety-critical hypoglycemic region but

at the expense of reduced statistical performance in euglycemic and hyperglycemic ranges.

Figure 6 shows the global RMSE and MAE for each prediction time point from 5 to 120 minutes, demonstrating a largely consistent difference between CW-RMSE and RMSE loss functions across time points.

4.2. Clinical evaluation metrics

Statistically significant differences were observed in the distribution of CG-EGA classifications, both globally and across each glycemic region ($p < 0.0001$ for all comparisons; Table 2).

Table 2. Chi-square analysis of the effect of RMSE and CW-RMSE loss functions on continuous glucose error grid analysis classifications

Category	Chi-square statistic	<i>p</i> -value
Global	1,577.188	<0.0001
Hypoglycemia	1,858.401	<0.0001
Euglycemia	740.062	<0.0001
Hyperglycemia	427.814	<0.0001

The most clinically meaningful improvements occurred in the hypoglycemic range, where the CW-RMSE model increased the percentage of APs (AP%) by 4.8 percentage points (from 83.2% to 88.0%) and reduced the percentage of EPs (EP%) by 5.1 percentage points (13.2% to 8.1%). In contrast, differences in the euglycemic and hyperglycemic ranges were negligible. In the euglycemic range, AP% decreased by 0.2 percentage points, while EP% improved marginally by 0.3 percentage points. In the hyperglycemic range, AP% reduced by 0.6 percentage points, and EP% was unchanged at 4.1% (Figure 7).

Performance for euglycemic and hyperglycemic values was consistent across PHs for RMSE and CW-RMSE models (Figure 8). However, within the hypoglycemic range, model performance diverged beyond approximately 30 min, with the CW-RMSE model maintaining higher AP% and lower EP% as the PH increased to 120 min (Figure 8).

4.3. Ablation analysis results

Ablation analysis isolated the individual and combined effects of hypoglycemic and CG-EGA-based weighting strategies on model performance (Table 3).

Hypoglycemic range weighting alone led to a small

improvement in hypoglycemic RMSE and MAE relative to the RMSE model (RMSE: 39.8 ± 0.13 mg/dL vs. 37.1 ± 0.13 mg/dL; MAE: 26.6 ± 0.08 mg/dL vs. 23.8 ± 0.08 mg/dL) but a negligible change in clinical outcomes (EP%: 13.2% vs. 13.1%; AP%: 83.2% vs. 83.1%). Global measures of RMSE, MAE, EP%, and AP% each remained relatively unchanged (RMSE: 36.7 ± 0.02 mg/dL unchanged; MAE: 25.4 ± 0.01 mg/dL vs. 24.8 ± 0.01 mg/dL; EP%: 3.5% vs. 3.6%; AP%: 91.8% unchanged).

Continuous glucose error grid analysis only weighting led to larger improvements compared to the RMSE model in the hypoglycemic range, reducing RMSE and MAE (RMSE: 39.8 ± 0.13 mg/dL vs. 30.6 ± 0.14 mg/dL; MAE: 26.6 ± 0.08 mg/dL vs. 17.4 ± 0.07 mg/dL) and EP and AP% (EP%: 13.2% vs. 9.4%; AP%: 83.2% vs. 86.5%). Global EP% and AP% differences were negligible (EP%: 3.5% vs. 3.4%; AP%: 91.8% vs. 91.4%). However, both global RMSE and MAE increased (RMSE: 36.7 ± 0.02 mg/dL vs. 40.3 ± 0.03 mg/dL; MAE: 25.4 ± 0.01 mg/dL vs. 28.1 ± 0.02 mg/dL).

The CW-RMSE model, integrating both hypoglycemic and CG-EGA strategies, produced the largest overall improvement in hypoglycemic RMSE, MAE, EP%, and AP% (RMSE: 29.8 ± 0.14 mg/dL; MAE: 16.4 ± 0.07 mg/dL; EP%: 8.1%; AP%: 88%) and the lowest global EP% (EP%: 3.2%). Global AP% was nearly identical to that of the RMSE and hypoglycemic-weighting-only models, differing by only 0.1 percentage points (AP%: 91.7% vs. 91.8%). However, the CW-RMSE model also produced the highest global RMSE (40.4 ± 0.03 mg/dL) and second-highest global MAE (27.8 ± 0.02 mg/dL).

5. Discussion

5.1. CW-RMSE vs. RMSE loss functions

Training with CW-RMSE produced statistically significant differences in statistical and clinical evaluation metrics. From a clinical perspective, CW-RMSE led to meaningful improvements in hypoglycemic prediction performance, with increased AP% and reduced EP% in this safety-critical range. In contrast, AP% showed marginal reductions in euglycemic and hyperglycemic ranges, while EP% improved slightly in euglycemia and was maintained in hyperglycemia. These findings indicated that CW-RMSE enhanced hypoglycemic safety while maintaining performance across euglycemic and hyperglycemic samples, consistent with the aims of the weighting strategy.

Statistically, CW-RMSE reduced RMSE and MAE in hypoglycemia but increased error in euglycemia and hyperglycemia. As CW-RMSE glycemic range weighting was applied only to hypoglycemic samples, optimization

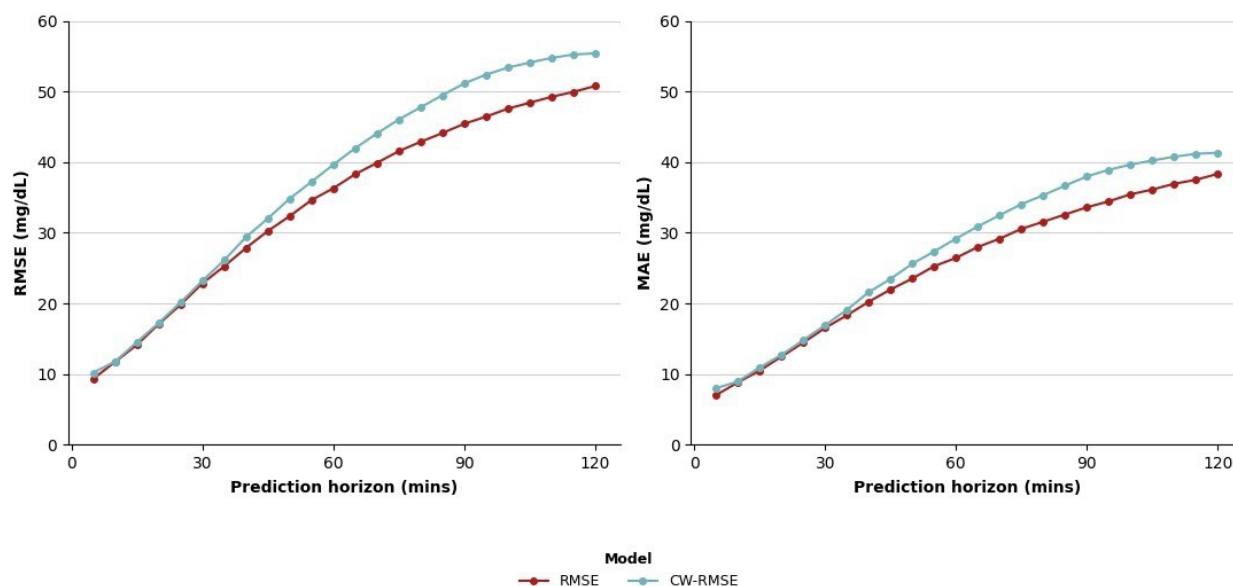


Figure 6. Root mean squared error (RMSE) and mean absolute error (MAE) for WinFormer models trained with RMSE and clinically weighted root mean squared error (CW-RMSE) loss functions by prediction horizon.

Table 3. Ablation analysis results

Model variant	Hypo RMSE (mg/dL)	Hypo MAE (mg/dL)	Global RMSE (mg/dL)	Global MAE (mg/dL)	Hypo EP%	Hypo AP%	Global EP%	Global AP%
RMSE	39.8	26.6	36.7	25.4	13.2%	83.2%	3.5%	91.8%
Hypo weighting only	37.1	23.8	36.7	24.8	13.1%	83.1%	3.6%	91.8%
CG-EGA weighting only	30.6	17.4	40.3	28.1	9.4%	86.5%	3.4%	91.4%
CW-RMSE	29.8	16.4	40.4	27.8	8.1%	88.0%	3.2%	91.7%

Abbreviations: AP%: Accurate prediction percentage; CG-EGA: Continuous glucose error grid analysis; CW-RMSE: Clinically weighted root mean squared error; EP%: Erroneous prediction percentage; Hypo: Hypoglycemia; RMSE: Root mean squared error.

was intentionally biased toward the reduction of error in this range. The resulting increase in euglycemia and hyperglycemia is therefore consistent with the redistribution of error away from hypoglycemia and toward euglycemia and hyperglycemia. Despite this increase, it is notable that CG-EGA classification performance was maintained in these regions. Further optimization of CW-RMSE weighting strategies may enable additional improvements in both clinical and statistical performance.

Clinical performance was broadly similar between models up to a PH of approximately 30 minutes. Beyond this point, differences emerged in hypoglycemic predictions. CW-RMSE attenuated the typical horizon-related increase in CG-EGA EP% and decline in CG-EGA

AP% relative to the RMSE loss-trained model.

In their reviews of BG prediction research, Ahmed *et al.*³⁰ and Liu *et al.*¹⁰ both highlighted that model performance declines as PH increases, likely because of the complex interactions among the various factors affecting glucose dynamics, and emphasized the importance of improving model performance at longer PHs. The comparative performance improvement observed in this study at longer PHs for hypoglycemic BG values is important for T1D management, supporting preventative actions such as insulin and carbohydrate intake modulation that increase the likelihood of averting hypoglycemic episodes entirely, or at least, attenuating their severity and duration.

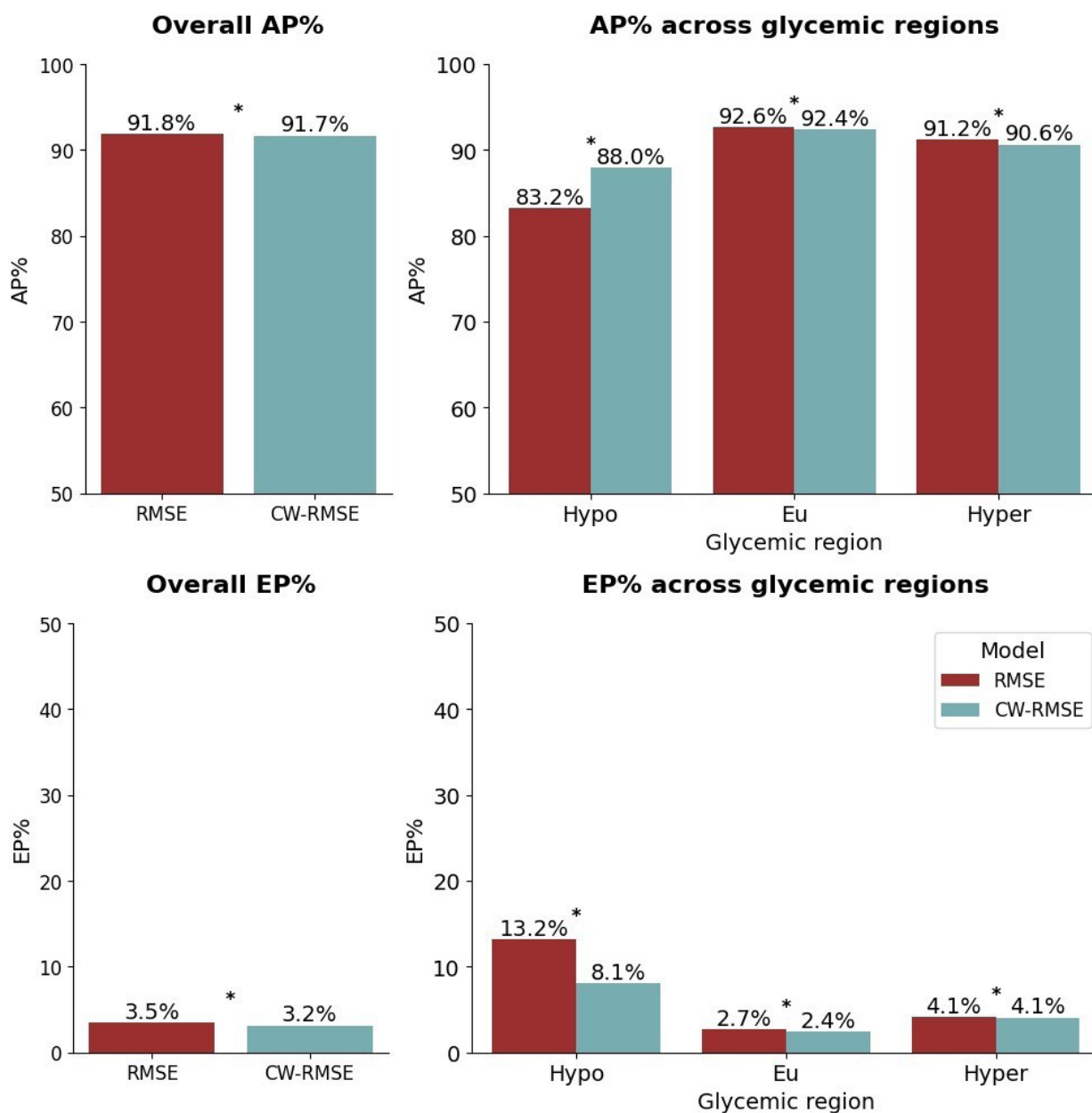


Figure 7. AP% and EP% for WinFormer models trained with RMSE and CW-RMSE loss functions. * denotes a statistically significant difference in classification distribution between models assessed via a χ^2 test ($p < 0.0001$).

Abbreviations: AP%: Accurate prediction percentage; CW-RMSE: Clinically weighted root mean squared error; EP%: Erroneous prediction percentage; Eu: Euglycemia; Hyper: Hyperglycemia; Hypo: Hypoglycemia; RMSE: Root mean squared error.

Ablation analyses demonstrated that while hypoglycemic range weighting alone can reduce clinical and statistical error, the magnitude of the effect is unlikely to be clinically meaningful. In contrast, CG-EGA-based weighting in isolation produced larger improvements in EP%, AP%, RMSE, and MAE in the hypoglycemic range as well as for EP% globally, albeit at the expense of increased global RMSE and MAE.

Clinically weighted root mean squared error integrated each of these approaches, yielding the strongest clinical and statistical hypoglycemic performance while largely maintaining global clinical performance. This highlights the complementary relationship between the respective weighting strategies, potentially reflecting the different modes through which they address error.

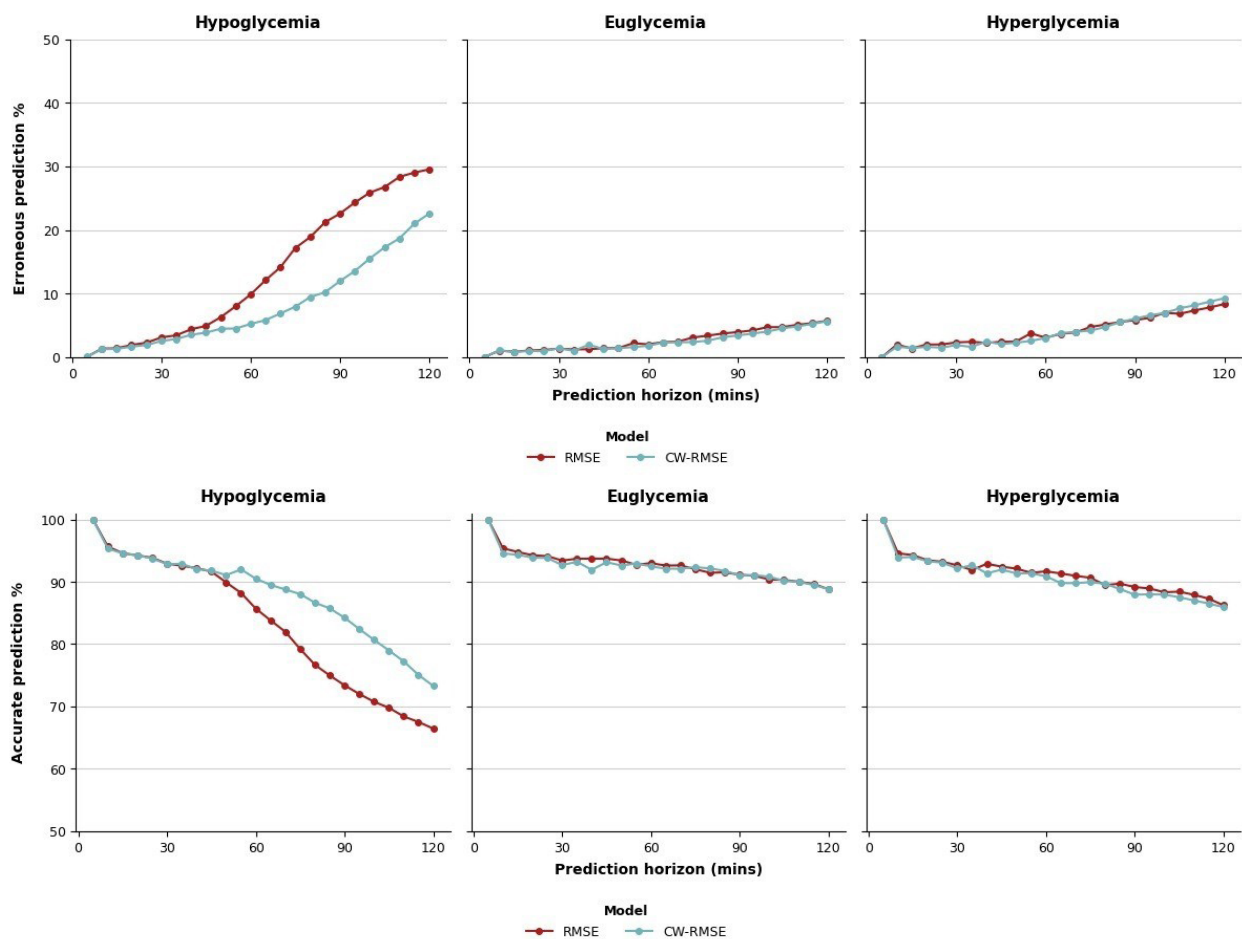


Figure 8. Erroneous prediction percentage and accurate prediction percentage of continuous glucose error grid analysis by glycemic region and prediction horizon.

Abbreviations: CW-RMSE: Clinically weighted root mean squared error; RMSE: Root mean squared error.

5.2. Cost-sensitive learning for blood glucose forecasting

This study demonstrates that CSL can steer the BG forecasting model optimization toward clinically safer behavior. By weighting hypoglycemic samples and clinically harmful CG-EGA classification, error was redistributed away from safety-critical hypoglycemia predictions while preserving global clinical accuracy. CSL thus helps mitigate class and risk imbalance within diabetes data and forecasting; however, imbalance in predictions was reduced rather than eliminated, even with the application of majority class undersampling and CSL in combination.

Araf *et al.*²⁵ noted that CSL is used less often in regression tasks in the medical domain than in classification tasks. This study demonstrated that it can be applied effectively in a medical time-series setting. Araf *et al.*²⁵ also noted two pitfalls of CSL in the medical domain; firstly, that clinical

costs are often uncertain or ill-specified, and secondly, that CSL may lead to overfitting to minority outcomes if costs are inadequately defined or heavily weighted towards the minority class. This study used the CG-EGA for continuous glucose predictions, anchoring model optimization to a recognized and well-defined measure of clinical risk.

Weight selection in CW-RMSE was heuristic with the objective of demonstrating proof-of-concept for risk-aware optimization rather than identifying optimal weight values. CW-RMSE weights and gradient clipping threshold were selected through exploratory grid search followed by manual refinement, guided by training stability and subjective assessments of clinically meaningful trade-offs, particularly the reduction in clinically harmful EPs balanced against changes in AP rates and improvements in hypoglycemic range with reduced performance in euglycemic and hyperglycemic ranges. While this approach

yielded stable optimization and meaningful changes in model performance, more systematic weighting strategies may enable further improvements in future work.

Clinically weighted root mean squared error loss introduces hard boundaries that can result in discontinuities and unstable gradients near thresholds. Future work should explore alternative CSL weighting strategies. Approaches that smooth loss boundaries, apply dynamic weighting assignment, or approaches such as focal loss that adapt cost based on the varying difficulty or significance of individual samples may support further improvements in learning and model performance.

Although this study demonstrated the efficacy of CSL, other approaches may also successfully enhance optimization for clinical accuracy. For example, Katsarou *et al.*³¹ used ensemble selection to directly minimize hypoglycemic EP%, lowering harmful errors while maintaining performance elsewhere. This underscores the broader principle that models should be designed, selected, and evaluated around clinically consequential endpoints.

5.3. Blood glucose forecasting evaluation

Two priorities for future model evaluation are identified through this study. Firstly, the divergence between statistical and clinical performance underscores the limitation of the sole evaluation of BG forecasting models through statistical metrics, such as RMSE and MAE. Clinical evaluation metrics should be integral to the evaluation of BG forecasting models designed for clinical applications. Many studies evaluate BG forecasting models based solely, or primarily, on statistical metrics,^{15,17,32-37} and to date, statistical metrics such as RMSE have provided the primary means of performance comparison between articles.^{10,30} Secondly, global model evaluation in isolation can mask the behavior of models in safety-critical minority classes such as hypoglycemia. Given the inherent class imbalance and asymmetrical risk across glycemic ranges, glycemic range-specific reporting of statistical and clinical metrics is essential to understand risk distribution and clinical efficacy.

5.4. The role of clinicians and patients

The development of a BG forecasting model involves numerous subjective decisions. To date, few studies have investigated the views of patients on broad approaches to data processing, model design, training, evaluation, and practical use of BG prediction models. This information is critical to the development of effective predictive models that can support T1D management in daily living. Stawarz *et al.*³⁸ sought to co-design ML-enhanced technology possibilities for supporting T1D management practices,

identifying that the perceived value and function of this technology are context dependent. Further, Barth *et al.*³⁹ directly engaged with T1D users using BG prediction applications, showing that predictions can support management decisions in everyday contexts while also identifying considerations such as trust, agency, and understanding that influence whether people act on predictions.

To the authors' knowledge, no study has systematically presented the viewpoints of clinicians involved in diabetes care (i.e., diabetes specialists) on the design and role of BG forecasting models in supporting improved clinical outcomes. Further research is warranted to gain input into specific design decisions to ensure that BG forecasting models support clinical requirements and user values. Clinician and patient inputs are essential for informing key design and development decisions in BG prediction models, including:

- The relative value of regression vs. classification-based models
- Which clinical evaluation metrics to prioritize
- Relative risk across glycemic ranges, and how this should shape loss function design, class/range weighting, and checkpoint criteria
- Required PH duration and the relationship with safe/acceptable error.
- The relative importance of model personalization vs. population models.
- The importance of uncertainty quantification (e.g., prediction intervals, risk scores) and interpretability.
- Tolerance for false positives versus missed events, including acceptable false-alarm rates.
- Factors that engender trust in model predictions, including interpretability, transparency, and alignment with clinical reasoning.
- Factors that support long-term utilization of BG forecasting models for decision support, including cognitive demand, interface design, customization, and the abstraction of complex metrics into intuitive and actionable outputs

5.5. Limitations

The present study evaluated CW-RMSE only in comparison to RMSE loss within a single model architecture. This approach enabled isolation of the effects of the respective loss functions while avoiding confounding influences. However, future work should extend this evaluation to include alternative risk-aware objectives and assess their behavior across different model architectures to better understand how different methods may affect model optimization and clinical efficacy in BG forecasting.

Validation was performed using a single dataset and at a population level. The large sample size supported stable estimation of population-level effects, but does not address individual variation, potentially masking important inter-individual differences in model performance. Future work should assess the generalizability of model performance through cross-dataset validation and explore patient-level analysis. Such an evaluation would support a clearer understanding of how population-level performance may translate to support and utility for individuals and real-world use cases.

Lastly, the fact that the weighting strategy was not systematically evaluated is also a limitation of the study. Although this approach was sufficient to demonstrate the potential of CW-RMSE, future work should systematically optimize the weighting scheme, potentially including patient-specific optimization strategies, which would support translation to real-world deployment.

6. Conclusion

This study demonstrated that orienting the training objective of BG forecasting models towards clinical accuracy through CW-RMSE can meaningfully improve hypoglycemic safety without degrading clinical performance in euglycemia or hyperglycemia in comparison to the sole optimization of statistical accuracy. This supports the application of CSL in regression-based BG forecasting.

Further, the study provides evidence to support the critical importance of including clinical performance metrics in the evaluation of BG forecasting models, as well as evaluating BG forecasting models across glycemic regions to understand the distribution of errors and clinical risk.

Finally, this study draws attention to the relative lack of input from clinicians and people with T1D into model development, which is essential to align model development with clinical objectives and user values to support real-world care.

Acknowledgments

None.

Funding

None.

Conflict of interest

The authors declare they have no competing interests.

Author contributions

Conceptualization: Joseph Passant

Data curation: All authors

Formal analysis: Joseph Passant

Investigation: Joseph Passant

Methodology: Joseph Passant

Project administration: All authors

Supervision: Andrea Corradini

Writing—original draft: Joseph Passant

Writing—review & editing: All authors

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

Data is available from the corresponding author upon reasonable request.

References

1. Lucier J, Mathias PM. Type 1 diabetes. In: *StatPearls*. Treasure Island, FL: StatPearls Publishing; 2024. Last accessed on February 11, 2026. <https://www.ncbi.nlm.nih.gov/books/NBK507713/>
2. Battelino T, Danne T, Bergenstal RM, *et al.* Clinical targets for continuous glucose monitoring data interpretation: recommendations from the international consensus on time in range. *Diabetes Care*. 2019;42(8):1593–1603.
doi: 10.2337/dci19-0028
3. Hankosky ER, Schapiro D, Gunn KB, Lubelczyk EB, Mitroi J, Nelson DR. Gaps remain for achieving HbA1c targets for people with type 1 or type 2 diabetes using insulin: results from NHANES 2009–2020. *Diabetes Ther*. 2023;14(6):967–975.
doi: 10.1007/s13300-023-01399-0
4. Duinkerken E, Snoek FJ, de Wit M. The cognitive and psychological effects of living with type 1 diabetes: a narrative review. *Diabet Med*. 2020;37(4):555–563.
doi: 10.1111/dme.14216
5. Bollepalli S, Smith LB, Vasquez A, Rodriguez H, Vehik K. Addressing the burdens of type 1 diabetes in youth. *Clin Pract*. 2012;9(4):409–424.
doi: 10.2217/cpr.12.33
6. Joensen LE, Almdal TP, Willaing I. Associations between patient characteristics, social relations, diabetes

- management, quality of life, glycaemic control and emotional burden in type 1 diabetes. *Prim Care Diabetes*. 2016;10(1):41-50.
doi: 10.1016/j.pcd.2015.06.007
7. Lin R, Brown F, James S, Jones J, Ekinici E. Continuous glucose monitoring: a review of the evidence in type 1 and type 2 diabetes mellitus. *Diabet Med*. 2021;38(5):e14528.
doi: 10.1111/dme.14528
8. DiMeglio LA, Evans-Molina C, Oram RA. Type 1 diabetes. *Lancet*. 2018;391(10138):2449-2462.
doi: 10.1016/S0140-6736(18)31320-5
9. Ehrmann DE, Laviola L, Priesterroth LS, Hermanns N, Babion N, Glatzer T. Fear of hypoglycaemia and diabetes distress: expected reduction by glucose prediction. *J Diabetes Sci Technol*. 2024;18(5):1027-1034.
doi: 10.1177/19322968241267886
10. Liu K, Li L, Ma Y, *et al*. Machine learning models for blood glucose level predictions in patients with diabetes mellitus: systematic review and network meta-analysis. *JMIR Med Inform*. 2023;11(1):e47833.
doi: 10.2196/47833
11. Langarica S, de la Vega D, Cariman N, *et al*. Deep learning-based glucose prediction models: a guide for practitioners and a curated dataset for improved diabetes management. *IEEE Open J Eng Med Biol*. 2024;5(5):467-475.
doi: 10.1109/OJEMB.2024.3365290
12. Vaswani A, Shazeer N, Parmar N, *et al*. Attention is all you need. *arXiv*. Preprint posted June 12, 2017.
doi: 10.48550/arXiv.1706.03762
13. Zhou H, Zhang S, Peng J, *et al*. Informer: beyond efficient transformer for long sequence time-series forecasting. *AAAI*. 2021;35(12):11106-11115.
doi: 10.1609/aaai.v35i12.17325
14. Zhou H. Informer2020 [software]. GitHub. Published 2020. Last accessed on February 11, 2026. <https://github.com/zhouhaoyi/Informer2020>
15. Sergazinov R, Armandpour M, Gaynanova I. Gluformer: transformer-based personalized glucose forecasting with uncertainty quantification. In: Proceedings of the ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); June 4-10, 2023; Rhodes Island, Greece. IEEE; 2023:1-5.
doi: 10.1109/icassp49357.2023.10096419
16. Zhu T, Chen T, Kuang L, Zeng J, Li K, Georgiou P. Edge-based temporal fusion transformer for multi-horizon blood glucose prediction. In: Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS); May 21-25, 2023; Monterey, CA. IEEE; 2023.
doi: 10.1109/ISCAS46773.2023.10181448
17. Cheng M, Diao X, Cui Y, Liu W, Cheng S. Toward short-term glucose prediction solely based on CGM time series. *arXiv*. Preprint posted April 18, 2024.
doi: 10.48550/arXiv.2404.11924
18. Zhu T, Afentakis I, Li K, *et al*. Multi-Horizon Glucose Prediction Across Populations With Deep Domain Generalization. *IEEE J Biomed Health Inform*. 2025;29(8):5424-5437.
doi: 10.1109/jbhi.2024.3428921
19. Xue Y, Guan S, Jia W. BGformer: an improved Informer model to enhance blood glucose prediction. *J Biomed Inform*. 2024;157(1):104715.
doi: 10.1016/j.jbi.2024.104715
20. de Bois M, El Yacoubi MA, Ammi M. GLYFE: review and benchmark of personalised glucose predictive models in type 1 diabetes. *Med Biol Eng Comput*. 2021;60(1):1-17.
doi: 10.1007/s11517-021-02437-4
21. Clarke W, Cox D, Gonder-Frederick LA, Carter W, Pohl SL. Evaluating clinical accuracy of systems for self-monitoring of blood glucose. *Diabetes Care*. 1987;10(5):622-628.
doi: 10.2337/diacare.10.5.622
22. Kovatchev BP, Cox DJ, Gonder-Frederick LA, Clarke WL. Evaluating the accuracy of continuous glucose-monitoring sensors: continuous glucose-error grid analysis illustrated by TheraSense Freestyle Navigator data. *Diabetes Care*. 2004;27(8):1922-1928.
doi: 10.2337/diacare.27.8.1922
23. Devi D, Biswa SK, Purkayastha B. A review on solution to class imbalance problem: undersampling approaches. In: Proceedings of the International Conference on Computational Performance Evaluation (ComPE); July 2-4, 2020; Shillong, India. IEEE; 2020.
doi: 10.1109/ComPE49325.2020.9200087
24. Saripuddin M, Suliman A, Sameon SS, Jorgenson BN. Random undersampling on imbalanced time series data for anomaly detection. In: Proceedings of the 4th International Conference on Machine Learning and Machine Intelligence (MLMI 2021); September 17-19, 2021; Hangzhou, China. ACM; 2021:151-156.
doi: 10.1145/3490725.3490748
25. Araf I, Idri A, Chair I. Cost-sensitive learning for imbalanced medical data: a review. *Artif Intell Rev*. 2024;57(4).
doi: 10.1007/s10462-023-10652-8
26. Aleppo G, Ruedy KJ, Riddlesworth TD, *et al*. REPLACE-BG: a randomized trial comparing continuous glucose monitoring with and without routine blood glucose monitoring in adults with well-controlled type 1 diabetes. *Diabetes Care*. 2017;40(4):538-545.

- doi: 10.2337/dc16-2482
27. Jaeb Center for Health Research. REPLACE-BG public dataset [data set]. 2017. Last accessed on February 11, 2026. <https://public.jaeb.org/dataset/546>
28. Passant J. *Informer-based blood glucose prediction: a comparative analysis of attention mechanisms and training methods*. Master's dissertation. Liverpool, UK: University of Liverpool; 2024. Last accessed on May 22, 2026. https://github.com/JosephPassant/Blood-Glucose-Prediction/blob/main/report/passant_ibgp.pdf
29. de Bois M. CG-EGA [software]. GitHub. Published 2020. Last accessed on February 11, 2026. <https://github.com/dotXem/CG-EGA>
30. Ahmed B, Ali ME, Masud MM, Nazni M. Recent trends and techniques of blood glucose level prediction for diabetes control. *Smart Health*. 2024;32:100457.
doi: 10.1016/j.smhl.2024.100457
31. Katsarou DN, Georga EI, Christou MA, *et al*. Optimizing hypoglycaemia prediction in type 1 diabetes with ensemble machine learning modeling. *BMC Med Inform Decis Mak*. 2025;25(1):46.
doi: 10.1186/s12911-025-02867-2
32. Alshehri OS, Alshehri OM, Samma H. Blood glucose prediction using RNN, LSTM and GRU: a comparative study. In: Proceedings of the IEEE International Conference on Advanced Systems and Emergent Technologies (IC_ASET); April 27-29, 2024; Hammamet, Tunisia. IEEE; 2024.
doi: 10.1109/IC_ASET61847.2024.10596176
33. Annuzzi G, Arpaia P, Bozzetto L, Criscuolo S, Giugliano S, Pesola M. Assessing the features on blood glucose level prediction in type 1 diabetes patients through explainable artificial intelligence. In: Proceedings of the IEEE International Conference on Metrology for Extended Reality, Artificial Intelligence and Neural Engineering (MetroXRINE); October 25-27, 2023; Milan, Italy. IEEE; 2023.
doi: 10.1109/MetroXRINE58569.2023.10405831
34. Butunoi BP, Stolojescu-Crisan C, Negru V. Short-term glucose prediction in type 1 diabetes. *Procedia Comput Sci*. 2024;238(10):41-48.
doi: 10.1016/j.procs.2024.05.194
35. Nordin MSM, Mahmud F. Univariate and multivariate time series blood glucose prediction with LSTM deep learning model. *Evol Electr Electron Eng*. 2024;5(1):276-285. Last accessed on February 11, 2026. <https://publisher.uthm.edu.my/periodicals/index.php/eeee/article/view/15439>.
doi: 10.30880/eeee.2024.05.01.035
36. Rehman NU, Contreras I, Benyeto A, Vehi J. The impact of missing continuous blood glucose samples on machine learning models for predicting postprandial hypoglycaemia: an experimental analysis. *Mathematics*. 2024;12(10):1567.
doi: 10.3390/math12101567
37. Mohamed Youssuff AR, Zainulabedin Hasan M, Anand R, Rajasekhara Babu M. Leveraging deep learning models for continuous glucose monitoring and prediction in diabetes management: toward enhanced blood sugar control. *J Syst Assur Eng Manag*. 2024;15(6):2077-2084.
doi: 10.1007/s13198-023-02200-y
38. Stawarz K, Katz D, Ayobi A, *et al*. Co-designing opportunities for human-centred machine learning in supporting type 1 diabetes decision making. *Int J Hum Comput Stud*. 2023;173:103003.
doi: 10.1016/j.ijhcs.2023.103003
39. Barth CM, Bernard J, Huang EM. "It's like a glimpse into the future": exploring the role of blood glucose prediction technologies for type 1 diabetes self-management. In: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24); May 11-16, 2024; Honolulu, HI. ACM; 2024:1-21.
doi: 10.1145/3613904.3642234