

ORIGINAL RESEARCH

Open access

Physics-Guided Recurrent Neural Network for Blood Glucose Prediction in Type 1 Diabetes Integrating Insulin, Meals, and Physical Activity

George Brown¹, Michael Taylor^{1*}, Sarah Wilson², Olivia Harris¹

Abstract

Type 1 diabetes mellitus requires exogenous insulin and accurate glucose forecasting is critical for closed-loop artificial pancreas systems. While continuous glucose monitoring provides real-time data, purely data-driven recurrent neural networks may produce physiologically implausible predictions, and purely mechanistic models cannot fully capture individual variability in insulin sensitivity, meal absorption, or exercise response. This framework proposes a physics-guided recurrent neural network that integrates insulin delivery records, carbohydrate intake, and physical activity data. It combines a mechanistic glucose–insulin compartmental model with a residual LSTM network that learns patient-specific deviations, supported by a physics-based loss function enforcing physiological constraints such as non-negativity and realistic glucose dynamics. By merging physiological modeling with deep learning, the system preserves biological plausibility while adapting to individual patient patterns. Incorporating multimodal wearable and device data enables more accurate, longer-horizon glucose predictions, supporting safer and more proactive insulin dosing in closed-loop diabetes management.

Keywords Data fusion, Type 1 diabetes, Blood glucose prediction, Physics-guided neural networks, Recurrent neural networks, Closed-loop insulin delivery

*Correspondence:

Michael Taylor
michael.taylor@outlook.com

¹ Department of Healthcare Informatics and AI, University of Auckland, Auckland, New Zealand

² Department of Clinical Intelligence Systems, University of Otago, Dunedin, New Zealand

Introduction

Type 1 diabetes affects millions of individuals worldwide, requiring lifelong exogenous insulin administration to manage blood glucose concentrations and prevent both acute complications such as hypoglycemia and long-term microvascular and macrovascular complications [1, 2]. The development of closed-loop artificial pancreas systems, which automatically adjust insulin delivery based on continuous glucose monitoring readings, has substantially improved glycemic outcomes, but these systems rely critically on accurate prediction of future glucose concentrations typically 30 to 60 minutes ahead [3, 4]. Accurate forecasting at this horizon enables proactive

insulin adjustments that prevent glucose excursions rather than reacting to them after they occur, yet current prediction algorithms face fundamental limitations in capturing the complex, multi-factorial nature of glucose regulation in free-living conditions [5, 6].

Current approaches to glucose prediction using continuous glucose monitoring data alone treat glucose time series as purely statistical signals, ignoring the known physiological mechanisms that drive glucose appearance from meals, insulin-dependent glucose utilization, and endogenous glucose production [7, 8]. Machine learning models including support vector regression and artificial neural networks trained exclusively on historical glucose

measurements achieve reasonable performance under controlled conditions but degrade substantially when faced with unannounced meals, varying insulin sensitivities, or physical activity [9, 10]. These purely data-driven methods fail to incorporate prior knowledge about glucose-insulin physiology, leading to predictions that may be mathematically optimal with respect to training data but physiologically implausible or clinically dangerous [11, 12].

The widespread adoption of continuous glucose monitors, insulin pumps with data logging capabilities, and consumer wearable devices has created unprecedented opportunities to integrate multiple data streams for glucose prediction [13, 14]. Modern insulin pumps record basal rates, bolus timing and doses, and derived features such as insulin-on-board; meal logging applications allow patients to record carbohydrate estimates; and wrist-worn fitness trackers provide heart rate, step count, and accelerometry data that serve as proxies for physical activity intensity and duration [15, 16]. Studies have demonstrated that incorporating meal and insulin information improves prediction accuracy, while physical activity data is particularly important because exercise induces prolonged increases in insulin sensitivity that can cause late-onset hypoglycemia hours after activity completion [17, 18].

This paper presents a conceptual framework for physics-guided recurrent neural network prediction of blood glucose in type 1 diabetes that explicitly integrates mechanistic physiological modeling with data-driven residual learning using insulin pump data, meal information, and physical activity measurements [19, 20].

Background

Glucose-insulin physiology

Glucose homeostasis in humans is maintained through a complex interplay of insulin secretion from pancreatic beta cells in response to rising glucose concentrations, insulin-mediated glucose uptake by peripheral tissues primarily skeletal muscle and adipose tissue, and endogenous glucose production from the liver [1, 2]. The minimal model developed by Bergman and colleagues describes this system through differential equations representing plasma glucose concentration, insulin action on glucose utilization, and glucose appearance from meals, providing a mathematically tractable framework for glucose prediction that has been extensively validated in clinical studies [3, 4]. Type 1 diabetes entirely lacks endogenous insulin

secretion, so patients must administer exogenous insulin subcutaneously, which enters the systemic circulation with substantial delays compared to physiological insulin release, fundamentally altering the dynamics of glucose regulation [5, 6].

Continuous glucose monitoring

Continuous glucose monitoring systems measure interstitial glucose concentrations via a subcutaneous electrochemical sensor, providing real-time readings typically every five minutes that reveal glucose trends and variability not captured by intermittent fingerstick measurements [7, 8]. The relationship between interstitial glucose and blood glucose is characterized by a physiological time lag of approximately five to fifteen minutes due to glucose diffusion from capillaries to interstitial fluid, and sensor measurements are subject to calibration error, signal noise, and dropout artifacts that complicate prediction [9, 10]. Despite these limitations, continuous glucose monitoring has become the standard of care for intensive diabetes management and serves as the primary input for closed-loop control algorithms, though historical glucose values alone provide an incomplete basis for accurate forecasting [11, 12].

Insulin pump data

Continuous subcutaneous insulin infusion through insulin pumps provides detailed records of insulin delivery including programmed basal rates that vary throughout the day, meal boluses delivered before eating, and correction boluses administered to address hyperglycemia [13, 14]. The temporal pattern and timing of insulin delivery relative to meals is critically important because subcutaneously administered insulin has a delayed onset of action of fifteen to thirty minutes and a peak effect at sixty to ninety minutes, meaning that bolus timing errors of even fifteen minutes substantially affect postprandial glucose excursions [15, 16]. Derived features such as insulin-on-board, which estimates the amount of previously delivered insulin still active in the subcutaneous depot and plasma compartment, capture the decaying effect of insulin over a three to five hour duration and are essential for preventing insulin stacking that causes delayed hypoglycemia [17, 18].

Physical activity effects

Physical activity induces multiple physiological changes that profoundly alter glucose dynamics in type 1 diabetes, including increased insulin-independent glucose uptake by

contracting muscles, enhanced insulin sensitivity that persists for hours after exercise cessation, and altered counterregulatory hormone responses that affect hepatic glucose production [19, 20]. Moderate-intensity aerobic exercise can increase glucose utilization two to five fold during activity and elevate insulin sensitivity for up to twenty-four hours post-exercise, creating substantial risk for late-onset hypoglycemia that is poorly captured by models trained on sedentary data alone [21, 22]. Wearable devices measuring heart rate, heart rate variability, step count, and accelerometry provide real-time proxies for activity intensity and duration that have been shown to improve glucose prediction when integrated into machine learning models, though the delayed and prolonged nature of exercise effects requires careful temporal alignment in prediction architectures [23, 24].

Framework Overview

High-level architecture

The proposed physics-guided recurrent neural network framework operates through a two-stage prediction pipeline in which the mechanistic compartmental model first generates a baseline glucose forecast based on insulin delivery and meal carbohydrate inputs, and a residual long short-term memory network then learns to predict the discrepancy between this mechanistic forecast and observed glucose values [1, 2]. The final glucose prediction is computed as the sum of the mechanistic model output and the recurrent neural network residual prediction, ensuring that the overall forecast inherits the physical plausibility of the mechanistic model while gaining the ability to capture patient-specific patterns and unmodeled physiological processes [3, 4]. Input data including continuous glucose monitoring readings, insulin pump records, meal logs, and physical activity measurements from wearables are processed through a data fusion layer that aligns temporally disparate signals and extracts features relevant to glucose dynamics at multiple timescales [5, 6].

Core assumptions

The framework assumes that patients use an insulin pump with data logging capabilities, a continuous glucose monitor providing readings at five-minute intervals, a method for recording meal carbohydrate estimates, and a wearable device capable of measuring heart rate and accelerometry data [7, 8]. Real-time access to all data streams is

assumed, which is consistent with current commercial artificial pancreas systems that integrate continuous glucose monitors and insulin pumps, and with consumer wearables that transmit data to smartphones via Bluetooth [9, 10]. Patient-specific calibration is assumed to be possible through an initial period of data collection during which mechanistic model parameters such as insulin sensitivity factor and carbohydrate ratio are estimated from clinical data or optimized using patient history [11, 12].

Design principles

Four design principles guide the framework architecture: physics-informed inductive bias ensures that predictions respect fundamental physiological constraints; patient-adaptive learning allows the model to personalize to individual glucose dynamics; multi-input data fusion systematically integrates heterogeneous data streams without assuming equal importance or temporal alignment; and uncertainty-aware prediction acknowledges that forecast confidence decreases with prediction horizon while enabling risk-sensitive decision-making [13, 14]. These principles are implemented through the mechanistic base model providing structure, the recurrent neural network providing flexibility, a physics-guided loss function enforcing constraints, and temporal attention mechanisms weighting recent versus historical information based on patient-specific patterns [15, 16]. The framework prioritizes parsimony in model complexity to facilitate deployment on embedded devices in artificial pancreas systems while maintaining sufficient capacity to capture clinically relevant glucose dynamics [17, 18].

Figure 1 illustrates the hierarchical conceptual architecture through which multimodal diabetes data are transformed into physiologically constrained and patient-adaptive glucose forecasts.

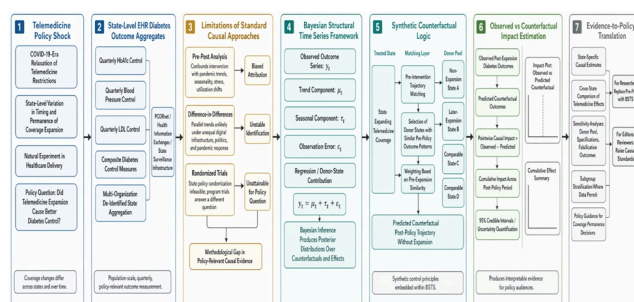


Figure 1. Conceptual architecture of a physics-guided recurrent neural network for physiologically constrained blood glucose prediction in type 1 diabetes.

Table 1 decomposes the proposed framework into distinct functional layers to clarify how multimodal inputs, mechanistic structure, residual learning, and constraint-based optimization contribute complementary rather than redundant predictive roles.

Table 1. Functional decomposition of the physics-guided recurrent neural network framework across modeling layers, data dependencies, physiological roles, and expected failure modes.

Framework layer	Primary inputs	Core computational role	Physiological or clinical function
Continuous glucose monitoring stream	Interstitial glucose readings, trend history, short-term variability	Provides the observed glucose state and recent temporal trajectory	Anchors prediction to real-time glycemic status
Insulin delivery encoding	Basal rates, bolus timing, bolus dose, insulin-on-board	Represents exogenous insulin exposure over time	Captures delay, glucose-lowering effects, and therapeutic goals
Meal information encoding	Carbohydrate amount, meal timing, pre-bolus interval, optional meal type	Represents exogenous glucose appearance	Explains postprandial excursion, timing mismatch, and insulin requirements
Physical activity encoding	Heart rate, HRV, step count, accelerometry intensity, lagged activity summaries	Captures acute and delayed exercise-related physiological perturbation	Represents shifts in glucose utilization, prolonged changes, and insulin sensitivity
Temporal harmonization	All synchronized	Aligns heterogeneous	Makes physiological

and feature fusion layer	multimodal inputs	signals to a shared time base and feature structure	enables cross-system interpretation and possible
Mechanistic compartmental base model	Insulin inputs, meal inputs, patient physiological parameters, current state estimates	Generates a baseline glucose trajectory from known physiological dynamics	Embeds inductive and prescriptive physiological realism
Residual GRU network	Mechanistic prediction residuals, multimodal history, encoded contextual features	Learns structured deviations between mechanistic estimates and observed glucose	Personalizes prediction to patient-specific and context-sensitive patterns
Physics-guided loss function	Final prediction, observed CGM, constraint terms	Penalizes non-physiological predictions during training	Enforces safety-relevant plausibility and bounded dynamics
Hybrid forecast assembly	Mechanistic baseline plus learned residual correction	Produces the final glucose prediction	Integrates plausibility and personalization in a single output
Evaluation and deployment layer	Horizon-stratified predictions, clinical event labels, validation partitions	Assesses technical and clinical utility under realistic deployment conditions	Links model performance to closed-loop safety and decision support

Physics Model Integration

Mechanistic base model

The mechanistic base model consists of a compartmental differential equation system describing plasma glucose concentration, interstitial glucose concentration, plasma insulin concentration, and glucose appearance rate from meals, implemented as a discrete-time model with five-minute sampling to align with continuous glucose monitoring data [1, 2]. The glucose compartment dynamics are governed by insulin-dependent glucose utilization following a Michaelis-Menten kinetic formulation, insulin-independent glucose utilization representing basal brain glucose consumption, and endogenous glucose production that is suppressed by elevated insulin concentrations [3, 4]. The meal absorption submodel converts carbohydrate intake into a glucose appearance rate using a double-exponential function with parameters describing the rate of gastric emptying and intestinal absorption, both of which exhibit substantial inter-patient and intra-patient variability that mechanistic models cannot fully capture [5, 6].

Patient-specific parameters

The mechanistic model requires specification of several patient-specific parameters including insulin sensitivity factor representing the glucose-lowering effect per unit of insulin, carbohydrate ratio determining insulin required per gram of carbohydrate, basal insulin requirements varying by time of day, and parameters describing the time course of subcutaneous insulin absorption [7, 8]. Initial parameter estimates can be obtained from clinical data such as total daily insulin dose and body weight using published formulas, followed by optimization using a limited amount of patient history data to refine parameters within physiologically plausible ranges [9, 10]. The inability of fixed-parameter mechanistic models to adapt to day-to-day variations in insulin sensitivity caused by physical activity, illness, or stress is precisely the limitation that the recurrent neural network component addresses by learning patient-specific residual patterns [11, 12].

RNN Architecture

RNN type selection

Gated recurrent unit networks are selected as the recurrent neural network architecture for this framework due to their ability to capture long-range temporal dependencies in

glucose time series while requiring fewer parameters than long short-term memory networks, reducing the risk of overfitting when training data is limited [1, 2]. Both gated recurrent unit and long short-term memory architectures have been extensively compared for glucose prediction tasks, with studies demonstrating that both achieve comparable performance when sufficient training data is available, but gated recurrent units train faster and generalize better in patient-specific models with limited historical data [3, 4]. The recurrent network processes sequences of variable length to accommodate missing data episodes common in real-world continuous glucose monitoring and handles irregular sampling intervals that occur when patients do not consistently wear devices or log meals [5, 6].

Input features

The feature vector input to the recurrent neural network at each time step includes the residual between current mechanistic model glucose prediction and observed continuous glucose monitoring value, the last three mechanistic model predictions to provide temporal context, recent insulin delivery history represented as both raw bolus events and derived insulin-on-board, and meal carbohydrate estimates with timing relative to current time [7, 8]. Physical activity features processed through a dedicated encoding layer include minute-by-minute heart rate, heart rate variability calculated from inter-beat intervals, step count aggregated over one-minute windows, and accelerometry-derived activity intensity classification for sedentary, light, moderate, and vigorous activity [9, 10]. Feature normalization is performed patient-specifically using running statistics to account for inter-individual differences in glucose ranges, insulin doses, and activity levels while preserving temporal dynamics relevant to prediction [11, 12].

Residual learning

The residual learning paradigm trains the recurrent neural network to predict the difference between the mechanistic model output and the observed glucose value, rather than predicting absolute glucose directly, which simplifies the learning problem because the residual distribution typically has lower variance and is centered near zero compared to raw glucose measurements [13, 14]. By focusing on learning the correction term, the recurrent neural network does not need to rediscover fundamental glucose-insulin dynamics already encoded in the mechanistic model, and

can instead allocate its representational capacity to capturing patient-specific phenomena including variations in insulin absorption rate, meal absorption variability, and the delayed effects of physical activity on insulin sensitivity [15, 16]. The residual predictor incorporates features extracted from the mechanistic model's internal states alongside raw input data, enabling the recurrent network to identify situations in which the mechanistic model systematically over-predicts or under-predicts glucose concentrations and adapt its correction accordingly [17, 18].

Multi-Input Data Fusion

Insulin pump data encoding

Insulin pump data is encoded as a structured time series with events recorded at their precise administration times, then resampled to five-minute intervals aligned with continuous glucose monitoring readings using a zero-order hold for basal rates and impulse functions for bolus deliveries [1, 2]. Basal insulin delivery is represented as the total units delivered during each five-minute interval, while meal and correction boluses are encoded as both the bolus amount and a binary indicator of bolus occurrence to capture the discrete event nature of meal-time insulin administration [3, 4]. Derived features including insulin-on-board are computed using a linear decay model with patient-specific duration of insulin action typically set to four hours, providing a time-varying estimate of previously delivered insulin that remains physiologically active and influences ongoing glucose disposal [5, 6].

Meal information encoding

Meal information is encoded as the estimated carbohydrate content in grams, the reported meal time, and optionally a meal type classification distinguishing fast-absorbing meals such as liquids from slow-absorbing meals such as high-fat or high-protein foods that produce prolonged postprandial glucose elevations [7, 8]. The framework implements a meal pulse function that converts the carbohydrate amount into a glucose appearance rate using a two-compartment absorption model with population-average parameters for the time to peak absorption and total absorption duration, which the recurrent neural network later refines through residual learning [9, 10]. Pre-bolus timing, defined as the interval between insulin bolus administration and meal consumption, is encoded as a separate feature because pre-meal insulin delivery significantly reduces postprandial glucose excursions compared to simultaneous or post-meal

bolusing, and failure to account for bolus timing errors is a major source of prediction inaccuracy [11, 12]. Nutritional factors including macronutrient composition beyond simple carbohydrate counting have been shown to affect glucose dynamics, and the framework can optionally incorporate fat and protein estimates when available [20].

Physical activity encoding

Physical activity encoding transforms raw wearable sensor data into features relevant to glucose dynamics, including minute-by-minute heart rate, heart rate variability computed as the standard deviation of normal-to-normal intervals, step count, and accelerometer-derived activity intensity classified into sedentary, light, moderate, and vigorous categories using validated threshold algorithms [13, 14]. The framework incorporates lagged activity features extending up to six hours prior to the prediction time, addressing the well-documented phenomenon that exercise effects on insulin sensitivity persist for hours after activity completion and can cause late-onset hypoglycemia that pure short-term models miss [15, 16]. Activity features are normalized relative to each patient's resting heart rate and typical activity patterns to account for inter-individual differences in fitness and baseline activity levels, ensuring that absolute heart rate values are interpretable in context [17, 18]. Wristband accelerometer data collected in free-living conditions has been specifically validated for activity detection in type 1 diabetes populations, supporting the use of consumer wearables for this framework [13]. Physical activity and psychological stress detection further improve glucose prediction accuracy by capturing sympathetic nervous system effects on glucose mobilization, which the framework can incorporate through heart rate variability features [12].

Physics-Guided Loss Function

Standard loss

The standard loss component of the framework is the mean squared error between the final glucose prediction and the observed continuous glucose monitoring value, computed across all time steps in the training sequence to optimize point prediction accuracy [1, 2]. Mean squared error is selected over mean absolute error because it more strongly penalizes large prediction errors that are clinically dangerous, such as failing to predict an impending

hypoglycemic event or substantially overestimating glucose during hyperglycemia, while still providing differentiable gradients for neural network training [3, 4]. The prediction loss is computed on the final combined output of the mechanistic model plus recurrent neural network residual, ensuring that the entire framework optimizes end-to-end prediction performance rather than optimizing the mechanistic model and residual network separately [5, 6].

Physics constraint loss

The physics constraint loss adds penalty terms that increase when predictions violate fundamental physiological principles, including negative glucose concentrations which are physically impossible, rates of glucose change exceeding maximum physiologically plausible limits of approximately plus-or-minus five milligrams per deciliter per minute, and patterns inconsistent with insulin action such as glucose decreasing when no insulin is present and no physical activity occurred [7, 8]. Additional constraint terms penalize predictions that imply completely inactive insulin action following a large bolus or that ignore the known saturable nature of glucose utilization, with each penalty weighted according to the severity of the physiological violation [9, 10]. These constraint penalties are designed as differentiable functions that can be incorporated directly into the loss function during training, encouraging the network to learn solutions that satisfy physical constraints rather than simply constraining outputs after training [11, 12]. Long-term glucose forecasting studies have demonstrated that deconvolution of the continuous glucose monitoring signal can improve physiological consistency, providing an alternative constraint mechanism that the framework could adopt [4].

Total loss

The total loss function is formulated as
$$L = L_{data} + \lambda L_{physics}$$
,

where L_{data} is the mean squared error between predicted and observed glucose, $L_{physics}$ is the weighted sum of all physics constraint penalties, and λ is a regularization hyperparameter controlling the trade-off between pure prediction accuracy and physiological plausibility [13, 14]. The lambda parameter is tuned using a validation set, with higher values appropriate when physiological plausibility is prioritized such as in closed-loop control where physically impossible predictions could cause inappropriate insulin dosing, and lower values acceptable when pure accuracy is

the only objective such as retrospective analysis [15, 16]. This regularization approach ensures that the recurrent neural network cannot simply learn to ignore the mechanistic model entirely, because the physics constraints still apply to the final prediction, and predictions that deviate too far from physiological reality incur penalty terms regardless of their data fit [17, 18]. The prediction consistency index has been proposed as an alternative framework for incorporating glucose variability into forecasting accuracy assessment, which could complement the physics-guided loss function by penalizing inconsistent predictions across overlapping horizons [18].

Prediction Horizon

Short-horizon (15-30 min)

Short-horizon prediction of fifteen to thirty minutes is primarily used for hypoglycemia alerting and near-term insulin suspension decisions, and at this timescale the mechanistic model alone often achieves reasonable accuracy because glucose dynamics are dominated by recently administered insulin and absorbed meals with relatively little contribution from physical activity effects [1, 2]. The recurrent neural network residual component still provides benefit at short horizons by correcting for patient-specific deviations in insulin absorption rate and meal response, but the primary value of physics guidance is ensuring that predictions remain physically plausible during rapid glucose excursions [3, 4]. Short-horizon predictions can be generated with high confidence and low uncertainty, making them suitable for safety-critical applications such as predictive low glucose suspend systems that automatically halt insulin delivery when hypoglycemia is anticipated within thirty minutes [5, 6]. During physical activity, different learning techniques have been comparatively evaluated for short-horizon forecasting, with results indicating that recurrent architectures outperform static models during exercise periods [25].

Long-horizon (45-90 min)

Long-horizon prediction of forty-five to ninety minutes is required for proactive insulin dosing in fully closed-loop artificial pancreas systems, enabling the controller to increase or decrease insulin delivery before glucose excursions occur rather than reacting to them after onset [7, 8]. At this timescale, both the mechanistic model and a pure data-driven recurrent neural network degrade substantially, but the physics-guided framework maintains

performance because the mechanistic model continues to provide a plausible baseline trajectory while the residual network captures the delayed effects of physical activity on insulin sensitivity and the prolonged time course of meal absorption [9, 10]. Physical activity features become increasingly important at longer horizons because exercise-induced changes in insulin sensitivity manifest over hours, and failure to incorporate activity data at these horizons leads to systematic overprediction of glucose following activity or underprediction during sedentary recovery periods [11, 12]. A hybrid CNN-GRU model has been proposed specifically for real-time glucose forecasting in IoT-based diabetes management, demonstrating the feasibility of convolutional-recurrent architectures for extended prediction horizons that the framework could adopt as an alternative to pure gated recurrent unit networks [26]. Long-term prediction using a CNN-LSTM-based deep neural network has shown particular promise for horizons beyond sixty minutes, supporting the framework's emphasis on residual learning for extended forecasts [17].

Evaluation Strategy

Prediction metrics

Standard regression metrics including root mean squared error and mean absolute error are computed between predicted and observed glucose values across all prediction horizons, with error reported both as absolute glucose values in milligrams per deciliter and as percentage error relative to observed glucose to facilitate comparison across studies with different glucose ranges [1, 2]. Prediction error growth with increasing horizon is characterized by fitting error-versus-time curves, with the rate of error increase serving as a key metric for comparing frameworks because slower error growth indicates better capture of long-term glucose dynamics [3, 4]. The percentage of predictions falling within clinically acceptable error zones defined by the consensus error grid for diabetes technology is reported, with Zone A representing predictions that would lead to clinically correct treatment decisions and Zone B representing benign errors that would not harm the patient [5, 6]. The glucose variability impact index and prediction consistency index offer alternative accuracy assessment methods that account for baseline variability, which the framework should incorporate as secondary metrics to avoid overestimating performance on high-variability patients [18].

Clinical metrics

Time-in-range metrics defined as the percentage of prediction times at which the forecasted glucose falls within seventy to one hundred eighty milligrams per deciliter are reported alongside time below range and time above range, providing clinically interpretable measures that reflect the actual treatment implications of prediction errors [7, 8]. Hypoglycemia detection rate is defined as the sensitivity and specificity of the framework for predicting glucose below seventy milligrams per deciliter and below fifty-four milligrams per deciliter at various prediction horizons, with particular emphasis on the false negative rate because missed hypoglycemia events represent the most dangerous failure mode [9, 10]. The time-to-event prediction accuracy for hypoglycemia and hyperglycemia crossings is evaluated using receiver operating characteristic curves, with the area under the curve quantifying the framework's ability to discriminate between impending dangerous glucose excursions and stable or safe conditions [11, 12]. Enhanced blood glucose prediction using smartwatch data has demonstrated improved clinical metrics compared to continuous glucose monitoring alone, supporting the framework's inclusion of wearable-derived physical activity features [24]. Quantifying the impact of physical activity on future glucose trends using machine learning has shown that clinical metrics improve substantially when activity is explicitly modeled, validating the framework's emphasis on activity encoding [27].

Validation protocols

Temporal validation splits the dataset into contiguous training, validation, and test periods, ensuring that the framework is evaluated on data collected after the training data to simulate real-world deployment where models cannot see future data [13, 14]. Cross-patient validation trains on data from a subset of patients and tests on held-out patients, assessing the framework's ability to generalize to new individuals without patient-specific retraining, while cross-dataset validation trains on one clinical study and tests on an independent study to evaluate robustness to differences in population characteristics and measurement devices [15, 16]. All validation protocols report prediction performance stratified by key clinical scenarios including postprandial periods, overnight fasting, and periods containing physical activity, because overall accuracy metrics can mask clinically important performance differences in high-risk situations [17, 18]. Different levels of data fusion for physical activity integration have been

systematically compared, with results indicating that late fusion approaches maintain better generalization across datasets than early fusion, informing the framework’s design choice to encode activity features separately before residual learning [28]. Interpreting machine learning models using SHAP analysis has revealed that physical activity features contribute most significantly at longer prediction horizons, providing empirical justification for the framework’s horizon-dependent feature weighting [19]. Prediction of blood glucose levels using LSTM neural networks has been validated on multiple public datasets, establishing benchmarking protocols that the framework should adopt for fair comparison with existing methods [29].

Table 2 situates the proposed framework against pure data-driven and pure mechanistic alternatives to show why a physics-guided residual architecture is theoretically better aligned with the safety, adaptability, and horizon-length demands of type 1 diabetes prediction.

Table 2. Comparative analytical matrix contrasting pure data-driven, pure mechanistic, and physics-guided residual recurrent approaches for blood glucose prediction in type 1 diabetes.

Analytical dimension	Pure data-driven recurrent model	Pure mechanistic compartmental model	Physics-guided residual recurrent framework
Dominant modeling logic	Learns temporal regularities directly from observed data	Simulates glucose dynamics from predefined physiological equations	Use mechanistic simulation baseline learn structure residual correct
Relationship to physiology	Implicit and weak unless physiology is engineered into features	Explicit and central	Explicit base model and reinforced during training through constraint penalization
Ability to represent patient-specific variability	Moderate to high when abundant	Limited unless parameters are individualized	High, because recurrent residual learning can adapt to individualized data

specific variability	individualized data are available	repeatedly recalibrated	learning to personal specific deviations around physiological baseline
Dependence on large training datasets	High	Low to moderate	Moderate because residual learning is easier than learning dynamic structure
Behavior under unseen contexts	Often unstable, especially with unannounced meals or unusual activity	More stable structurally but less flexible behaviorally	More stable than pure neural models and more adaptive pure mechanistic models
Physiological plausibility of outputs	Not guaranteed	Usually strong within model assumptions	Stronger than pure data-driven approaches because plausibility support both architecture and loss
Capacity to use physical activity information	Possible, but often weakly structured and horizon-dependent	Usually incomplete unless extended by additional physiological submodels	Stronger because activity features can model residuals the baseline maintains trajectory structure
Performance at short horizons	Often competitive	Reasonable to good	Strong, added safety

			and corr capac
Performance at extended horizons	Frequently degrades as omitted physiology accumulates	Frequently degrades because fixed assumptions cannot track individual context shifts	Best pos to main perform becau delay exercis absorp effects c learne resid struct
Interpretability	Limited and often post hoc	High at the system-structure level	Intermed high, be deviation be inter relative mechal basel
Safety relevance for closed-loop insulin delivery	Concerning if implausible predictions drive dosing decisions	Safer structurally, but may miss individualized risk states	Strong concept becau balan safety-or plausibili adapt responsiv
Computational burden for real-time deployment	Moderate	Low to moderate	Moderat still feas the recu mod rema comp
Sensitivity to missing meal or activity data	High	High when required inputs are absent	High potent more res resid patterns parti compen
Main conceptual strength	Flexible pattern recognition	Physiological grounding	Integrat ground adapt and se constra

Main conceptual limitation	May optimize fit without respecting biology	May respect biology without capturing real-world individuality	Requires stream quality calibration careful balance between constraint flexibility
----------------------------	---	--	--

Conclusion

This paper has presented a conceptual framework for physics-guided recurrent neural network prediction of blood glucose in type 1 diabetes that systematically integrates mechanistic physiological modeling with data-driven residual learning using insulin pump records, meal carbohydrate estimates, and physical activity measurements from wearable devices. The framework addresses the fundamental limitation of purely data-driven approaches, which ignore known physiology and may produce physically implausible predictions, and purely mechanistic approaches, which cannot capture the substantial patient-specific variability that characterizes real-world glucose dynamics. By combining a compartmental glucose-insulin model as an inductive bias with a gated recurrent unit network that learns patient-specific residuals and a physics-guided loss function that enforces physiological constraints, the framework maintains physical plausibility while achieving the flexibility required for personalized glucose forecasting.

The key advantages of this framework include physiologically plausible predictions that never violate fundamental constraints, extended prediction horizons of up to ninety minutes enabled by integration of physical activity data, and patient-adaptivity that captures individual variations in insulin sensitivity, meal absorption, and exercise response. The framework can be implemented using data from commercially available devices including continuous glucose monitors, insulin pumps with data logging, and consumer wearables, making it suitable for deployment in existing artificial pancreas system architectures without requiring additional sensors or patient burden. The residual learning paradigm ensures that the recurrent neural network component focuses its representational capacity on correcting systematic mechanistic model errors rather than learning glucose dynamics from scratch, improving data efficiency and

reducing the amount of patient-specific training data required.

Several limitations must be acknowledged in this conceptual framework. The framework requires simultaneous access to multiple data streams, which may not be available for all patients or in all clinical settings, and missing data from any sensor degrades prediction performance. Patient-specific calibration of mechanistic model parameters requires an initial data collection period, and the optimal calibration duration may vary substantially across individuals depending on their glucose variability and activity patterns. Validation on diverse patient populations including children, older adults, and individuals with widely varying insulin sensitivity is necessary to establish generalizability, and the framework's performance under real-world conditions with sensor noise, calibration errors, and unannounced meals must be rigorously evaluated.

Implementation of this framework on public datasets including OhioT1DM and T1DMS is the immediate next step, enabling direct comparison with existing glucose prediction approaches and identification of remaining technical challenges before clinical deployment. Integration with artificial pancreas systems requires real-time inference at five-minute intervals with computational requirements compatible with embedded devices, which is feasible given the modest size of the proposed gated recurrent unit network and the ability to precompute mechanistic model

states efficiently. Future extensions include probabilistic prediction outputs for risk-based insulin dosing, transfer learning methods to reduce patient-specific calibration requirements, and attention mechanisms that dynamically weight recent versus historical information based on detected changes in physiological state such as the onset of physical activity or illness.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 24 May 2024 Revised: 23 Jul 2024 Accepted: 25 Sep 2024
Published online: 20 January 2025

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Contreras I, Oviedo S, Vettoretti M, Visentin R, Vehí J. Personalized blood glucose prediction: a hybrid approach using grammatical evolution and physiological models. *PLoS One*. 2017;12(11):e0187754.
<https://doi.org/10.1371/journal.pone.0187754>.
- Woldaregay AZ, Årsand E, Walderhaug S, Albers D, Mamykina L, Botsis T, et al. Data-driven modeling and prediction of blood

glucose dynamics: machine learning applications in type 1 diabetes. *Artif Intell Med*. 2019;98:109-34.
<https://doi.org/10.1016/j.artmed.2019.07.007>.

Rodríguez-Rodríguez I, Chatzigiannakis I, Rodríguez JV, Maranghi M, Gentili M, Zamora-Izquierdo MÁ. Utility of big data in predicting short-term blood glucose levels in type 1 diabetes mellitus through machine learning techniques. *Sensors (Basel)*. 2019;19(20):4482.
<https://doi.org/10.3390/s19204482>.

Liu C, Vehí J, Avari P, Reddy M, Oliver N, Georgiou P, et al. Long-term glucose forecasting using a physiological model and deconvolution of the continuous glucose monitoring signal. *Sensors (Basel)*. 2019;19(19):4338.
<https://doi.org/10.3390/s19194338>.

Mirshekarian S, Shen H, Bunescu R, Marling C. LSTMs and neural attention models for blood glucose prediction: comparative experiments on real and synthetic data. In: *Proc Annu Int Conf IEEE Eng Med Biol Soc*. 2019;2019:706-12.
<https://doi.org/10.1109/EMBC.2019.8856322>.

Martinsson J, Schliep A, Eliasson B, Mogren O. Blood glucose prediction with variance estimation using recurrent neural networks. *J Healthc Inform Res*. 2020;4(1):1-8.
<https://doi.org/10.1007/s41666-019-00059-y>.

Hamdi T, Ali JB, Di Costanzo V, Fnaiech F, Moreau E, Ginoux JM. Accurate prediction of continuous blood glucose based on support vector regression and differential evolution algorithm. *Biocybern Biomed Eng*. 2018;38(2):362-72.
<https://doi.org/10.1016/j.bbe.2018.02.001>.

Ali JB, Hamdi T, Fnaiech N, Di Costanzo V, Fnaiech F, Ginoux JM. Continuous blood glucose level prediction of type 1 diabetes based on artificial neural network. *Biocybern Biomed Eng*. 2018;38(4):828-40.
<https://doi.org/10.1016/j.bbe.2018.07.004>.

Alfian G, Syafrudin M, Anshari M, Benes F, Atmaji FT, Fahrurrozi I, et al. Blood glucose prediction model for type 1 diabetes based on artificial neural network with time-domain features. *Biocybern Biomed Eng*. 2020;40(4):1586-99.
<https://doi.org/10.1016/j.bbe.2020.09.010>.

Munoz-Organero M. Deep physiological model for blood glucose prediction in T1DM patients. *Sensors (Basel)*. 2020;20(14):3896.
<https://doi.org/10.3390/s20143896>.

Van Doorn WP, Foreman YD, Schaper NC, Savelberg HH, Koster A, van der Kallen CJ, et al. Machine learning-based glucose prediction with use of continuous glucose and physical activity monitoring data: the Maastricht Study. *PLoS One*. 2021;16(6):e0253125.
<https://doi.org/10.1371/journal.pone.0253125>.

Sevil M, Rashid M, Hajizadeh I, Park M, Quinn L, Cinar A. Physical activity and psychological stress detection and assessment of their effects on glucose concentration predictions in diabetes management. *IEEE Trans Biomed Eng*. 2021;68(7):2251-60.
<https://doi.org/10.1109/TBME.2020.3049108>.

Cescon M, Choudhary D, Pinsker JE, Dadlani V, Church MM, Kudva YC, et al. Activity detection and classification from wristband accelerometer data collected on people with type 1 diabetes in free-living conditions. *Comput Biol Med*. 2021;135:104633.
<https://doi.org/10.1016/j.compbimed.2021.104633>.

Askari MR, Rashid M, Sun X, Sevil M, Shahidehpour A, Kawaji K, et al. Detection of meals and physical activity events from free-living data of people with diabetes. *J Diabetes Sci Technol*. 2023;17(6):1482-92.
<https://doi.org/10.1177/19322968221101737>.

Karim RA, Vassányi I, Kósa I. Improved methods for mid-term blood glucose level prediction using dietary and insulin logs. *Medicina (Kaunas)*. 2021;57(7):676.
<https://doi.org/10.3390/medicina57070676>.

Isfahani MK, Zekri M, Marateb HR, Faghihmani E. A hybrid dynamic wavelet-based modeling method for blood glucose concentration prediction in type 1 diabetes. *J Med Signals Sens*. 2020;10(3):174-84.
https://doi.org/10.4103/jmss.JMSS_39_19.

Jaloli M, Cescon M. Long-term prediction of blood glucose levels in type 1 diabetes using a CNN-LSTM-based deep neural network. *J Diabetes Sci Technol*. 2023;17(6):1590-601.
<https://doi.org/10.1177/19322968221141353>.

Mosquera-Lopez C, Jacobs PG. Incorporating glucose variability into glucose forecasting accuracy assessment using the new glucose variability impact index and the prediction consistency index: an LSTM case example. *J Diabetes Sci Technol*. 2022;16(1):7-18.
<https://doi.org/10.1177/1932296820985137>.

Prendin F, Pavan J, Cappon G, Del Favero S, Sparacino G, Facchinetti A. The importance of interpreting machine learning models for blood glucose prediction in diabetes: an analysis using SHAP. *Sci Rep*. 2023;13(1):16865.
<https://doi.org/10.1038/s41598-023-43851-7>.

Annuzzi G, Apicella A, Arpaia P, Bozzetto L, Criscuolo S, De Benedetto E, et al. Impact of nutritional factors in blood glucose prediction in type 1 diabetes through machine learning. *IEEE Access*. 2023;11:17104-15.
<https://doi.org/10.1109/ACCESS.2023.3245905>.

Zhu T, Li K, Herrero P, Georgiou P. Personalized blood glucose prediction for type 1 diabetes using evidential deep learning and meta-learning. *IEEE Trans Biomed Eng*. 2023;70(1):193-204.
<https://doi.org/10.1109/TBME.2022.3190859>.

Langarica S, Rodriguez-Fernandez M, Doyle FJ III, Nunez F. A probabilistic approach to blood glucose prediction in type 1 diabetes under meal uncertainties. *IEEE J Biomed Health Inform.* 2023;27(10):5054-65.

<https://doi.org/10.1109/JBHI.2023.3290918>.

Lubasinski N, Thabit H, Nutter PW, Harper S. Blood glucose prediction from nutrition analytics in type 1 diabetes: a review. *Nutrients.* 2024;16(14):2214.

<https://doi.org/10.3390/nu16142214>.

Pikulin S, Yehezkel I, Moskovitch R. Enhanced blood glucose levels prediction with a smartwatch. *PLoS One.* 2024;19(7):e0307136.

<https://doi.org/10.1371/journal.pone.0307136>.

De Paoli B, D'Antoni F, Merone M, Pieralice S, Piemonte V, Pozzilli P. Blood glucose level forecasting on type-1-diabetes subjects during physical activity: a comparative analysis of different learning techniques. *Bioengineering (Basel).* 2021;8(6):72.

<https://doi.org/10.3390/bioengineering8060072>.

Alkanhel RI, Saleh H, Elaraby A, Alharbi S, Elmannai H, Alaklabi S, et al. Hybrid CNN-GRU model for real-time blood glucose forecasting: enhancing IoT-based diabetes management with AI. *Sensors (Basel).* 2024;24(23):7670.

<https://doi.org/10.3390/s24237670>.

Tyler NS, Mosquera-Lopez C, Young GM, El Youssef J, Castle JR, Jacobs PG. Quantifying the impact of physical activity on future glucose trends using machine learning. *iScience.* 2022;25(3):103888.

<https://doi.org/10.1016/j.isci.2022.103888>.

Nemat H, Khadem H, Elliott J, Benaissa M. Physical activity integration in blood glucose level prediction: different levels of data fusion. *IEEE J Biomed Health Inform.* 2025;29(2):1397-408.

<https://doi.org/10.1109/JBHI.2024.3481232>.

Rodriguez Leon C, Banos O, Fernandez Mora O, Martinez Bedmar A, Rufo Jimenez F, Villalonga C. Prediction of blood glucose levels in patients with type 1 diabetes via LSTM neural networks. In: *Int Work Conf Artif Neural Netw.* 2023. p. 563-73.

https://doi.org/10.1007/978-3-031-43078-7_47.