Article | Received 28 August 2025; Revised 25 November 2025; Accepted 10 December 2025; Published 22 December 2025
https://doi.org/10.55092/bi20250007

Interpretable blood glucose prediction for type 1 diabetes using neural-symbolic Fourier operators and clinical rules

Bailing Zhang

School of Computer Science and Data Engineering, Ningbo Tech University, Ningbo 315100, China; E-mail:bailing.zhang@nit.zju.edu.cn

Highlights:

- Neural-symbolic framework integrates FNOs and clinical rules for T1D glucose prediction.
- Achieves 10.2 mg/dL MAE and 99.8% Clarke Error Grid A+B for 30-min forecasts.
- Robust to CGM noise/missing data with cross-patient generalization (MAE 8.85 ± 1.33 mg/dL).
- Provides interpretable hypoglycemia alerts and trend analyses for insulin dosing.

Abstract: Accurate and interpretable blood glucose prediction is critical for optimizing insulin therapy and preventing adverse events in type 1 diabetes (T1D) management. However, existing deep learning models often lack transparency, limiting their clinical adoption. We introduce a novel neural-symbolic framework that integrates Fourier neural operators (FNOs) with clinical rules to predict 30-minute-ahead blood glucose levels with high accuracy and interpretability. FNOs capture periodic glucose patterns, such as circadian rhythms and meal responses, while differentiable clinical rules encode medical knowledge, providing actionable insights like hypoglycemia risk alerts. Evaluated on the OhioT1DM dataset, our model achieves a mean absolute error (MAE) of 10.2 mg/dL, meeting ISO 15197:2013 standards, and 99.8% Clarke Error Grid A+B coverage, ensuring clinical safety. Leave-one-subject-out cross-validation across 12 subjects demonstrates robust generalization (MAE = 8.85 $\pm$ 1.33 mg/dL), addressing concerns about sample size limitations. The model exhibits strong robustness to continuous glucose monitoring (CGM) measurement noise (maintaining ISO compliance at 10% noise) and missing data (5.4% performance degradation at 40% data loss). By generating interpretable outputs, such as trend analyses and risk warnings, our framework supports clinical decision-making, enhances patient trust, and facilitates personalized diabetes care. This work advances medical informatics by combining data-driven and knowledge-driven methods, offering a scalable solution for real-time glucose monitoring and insulin dosing in T1D.

Keywords: type 1 diabetes; blood glucose prediction; neural-symbolic AI; Fourier neural operators; clinical decision support; interpretable machine learning

1. Introduction

T1D affects over 537 million adults globally, with projections estimating 783 million by 2045, posing a major health challenge [1]. T1D patients, lacking endogenous insulin production, must meticulously manage insulin dosing, carbohydrate intake, physical activity, and other factors to maintain blood glucose within the 70–180 mg/dL target range [2]. Poor glycemic control risks severe outcomes: hypoglycemia can cause seizures or coma, while chronic hyperglycemia leads to complications like retinopathy and cardiovascular disease [3]. Accurate 30-minute-ahead glucose predictions can optimize insulin therapy, prevent adverse events, and improve patient quality of life [4].

CGM systems have transformed T1D management by providing frequent glucose readings (every 5–15 minutes), enabling proactive interventions [5]. However, the complexity of CGM data, driven by insulin kinetics, meal patterns, circadian rhythms, and individual variability, makes short-term prediction challenging [4]. Deep learning models, such as Long Short-Term Memory (LSTM) networks and convolutional neural networks (CNNs), excel at capturing temporal patterns in glucose data [6–8]. Yet, these models often act as "black boxes," offering little insight into their predictions, which limits their clinical utility [9].

The lack of interpretability in deep learning poses significant barriers to T1D care. Clinicians need transparent predictions to adjust insulin doses confidently, patients require understandable explanations to enhance trust and adherence, and regulatory bodies demand accountability for AI-based medical devices [10,11]. Moreover, purely data-driven models often fail to incorporate decades of clinical knowledge about T1D management, such as rules for detecting hypoglycemia or postprandial spikes [12].

We propose a neural-symbolic framework for interpretable 30-minute-ahead blood glucose prediction in T1D, integrating FNOs with clinical rules. Our approach offers four key contributions:

First, we pioneer the use of FNOs in T1D glucose prediction. FNOs, originally developed for physical systems [13], model glucose dynamics in the frequency domain, efficiently capturing periodic patterns like daily hormonal cycles and meal responses [14]. This reduces model complexity while enhancing predictive accuracy.

Second, we embed clinical rules within the neural architecture, encoding medical knowledge about hypoglycemia, hyperglycemia, and rapid glucose changes. These rules, optimized during training, provide interpretable outputs, such as risk alerts, that support clinical decision-making [15,16].

Third, we validate our model on the OhioT1DM dataset, achieving a mean absolute error of 10.2 mg/dL, meeting ISO 15197:2013 standards, and 99.8% Clarke Error Grid A+B coverage [17]. Comprehensive robustness analysis through leave-one-subject-out cross-validation (12 subjects, MAE = 8.85 $\pm$ 1.33 mg/dL) demonstrates strong cross-patient generalization. Additional experiments show resilience to CGM measurement noise and missing data, addressing real-world deployment concerns.

Finally, we address clinical deployment needs, with our model requiring only 12 ms per prediction and 15.2 MB of memory, enabling integration with CGM devices or insulin pumps for real-time T1D management.

Section 2 reviews related work in glucose prediction, neural-symbolic AI, and frequency-domain methods. Section 3 details our framework and training methodology. Section 4 presents experimental results, including robustness analysis. Section 5 discusses implications, limitations, and future directions. Section 6 concludes the paper.

2. Related work

2.1. Blood glucose prediction for type 1 diabetes

Accurate prediction of blood glucose levels is critical for T1D management, enabling proactive insulin dosing and hypoglycemia prevention [4]. Early approaches relied on physiological models of glucose-insulin dynamics [18,19]. While insightful, these models were limited by simplifying assumptions and inter-individual variability, reducing their predictive accuracy for real-world T1D care.

Machine learning introduced data-driven solutions in the 2000s. Sparacino *et al.* [20] used autoregressive (AR) models to leverage CGM data, outperforming physiological models for short-term predictions. Georga *et al.* [21] applied support vector machines (SVMs), incorporating meal and activity data to enhance 30-minute-ahead forecasts. Random forests further captured nonlinear glucose patterns, particularly for hypoglycemia detection [22].

Deep learning has significantly advanced T1D glucose prediction. LSTM networks, as shown by Mirshekarian and Bunescu [23], effectively model CGM time series for 30-minute predictions. Extensions include bidirectional LSTMs [6] and dilated recurrent neural networks (RNNs) that capture longer temporal dependencies [8]. CNNs extract local glucose patterns [7], while hybrid CNN-LSTM models combine spatial and sequential modeling [24]. Multi-task frameworks predict glucose levels and glycemic events simultaneously [25].

Despite improved accuracy, these deep learning models lack interpretability, limiting their clinical adoption. Clinicians need transparent predictions to adjust insulin doses, and patients require understandable outputs to trust CGM-based recommendations [9]. Post-hoc methods like SHAP [26] and LIME [10] offer approximate explanations but may not reflect true model reasoning [12], underscoring the need for inherently interpretable architectures.

2.2. Neural-symbolic AI in healthcare

Neural-symbolic AI integrates neural learning with symbolic reasoning, combining data-driven pattern recognition with interpretable medical knowledge [15]. In healthcare, this approach is valuable for embedding clinical guidelines into predictive models, enhancing trust and usability [16].

Early neural-symbolic methods, like Logic Tensor Networks (LTN), embed logical constraints into neural architectures for end-to-end learning [27]. Neural Module Networks compose specialized modules based on symbolic programs, providing interpretable outputs [28]. In medical applications, Shang *et al.* [29] developed GAMENet, which uses knowledge graphs for medication recommendations, improving accuracy and explainability. Tayal *et al.* [16] proposed a neural-symbolic framework for monitoring dietary salt, generating human-readable rules for clinical verification.

However, most neural-symbolic healthcare applications focus on classification tasks (e.g., diagnosis, risk prediction). Extending neural-symbolic methods to continuous time-series prediction, such as 30-minute glucose forecasting, remains underexplored. Our framework addresses this gap by integrating clinical rules with neural models for interpretable T1D predictions [30].

2.3. Frequency-domain methods for physiological signals

Frequency-domain methods are well-suited for physiological signals with periodic patterns, such as circadian rhythms in glucose dynamics [14]. FNOs represent a breakthrough in learning mappings between function spaces, ideal for time-series data [13]. FNOs parameterize integral kernels in the Fourier domain:

$$\mathscr{K}(v)(x) = \mathscr{F}^{-1}\left(R_\phi \cdot \mathscr{F}(v)\right)(x) \tag{1}$$

where $\mathscr{F}$ is the Fourier transform, R_ϕ are learnable weights, and $\mathscr{F}^{-1}$ is the inverse transform. By operating in frequency space, FNOs efficiently capture long-range temporal dependencies with fewer parameters than recurrent or attention-based models [31]. FNOs have excelled in weather forecasting, fluid dynamics, and seismic modeling, where periodic and multi-scale patterns dominate [13]. Their application to glucose prediction, which exhibits daily hormonal cycles and meal-related oscillations, is a natural yet unexplored extension [14].

2.4. Uncertainty quantification in medical AI

Reliable uncertainty estimates are essential for clinical decision-making, particularly in high-stakes applications like insulin dosing [32]. Bayesian deep learning methods, such as Monte Carlo dropout [33] and variational inference [34], approximate predictive distributions but are computationally expensive. Quantile regression provides a simpler alternative, directly estimating prediction intervals without distributional assumptions [35]. Our framework adopts quantile regression to output 2.5%, 50%, and 97.5% quantiles, balancing interpretability and computational efficiency for real-time glucose prediction.

2.5. Research gaps and contributions

Existing research on blood glucose prediction for T1D reveals critical gaps that our neural-symbolic framework addresses. Deep learning models, despite their accuracy, lack transparency, hindering clinical trust and adoption in T1D management [12]. Clinical knowledge, such as patterns of hypoglycemia or postprandial spikes, is rarely embedded in predictive models, missing opportunities to enhance accuracy and interpretability [4]. Frequency-domain methods like FNOs, ideal for capturing periodic glucose patterns such as circadian rhythms, remain unexplored in T1D prediction [14]. Moreover, neural-symbolic approaches, while promising in healthcare classification tasks, are seldom applied to continuous time-series forecasting like 30-minute-ahead glucose prediction [30]. Our framework integrates FNOs with clinical rules, achieving accurate, interpretable 30-minute-ahead T1D glucose predictions that support insulin dosing and enhance clinical utility in real-time diabetes management.

3. Methodology

3.1. Problem formulation

We address blood glucose prediction for T1D as a multivariate time-series forecasting problem with uncertainty quantification, critical for optimizing insulin dosing and preventing hypoglycemic events [4]. Given historical glucose measurements over T_h time steps, $\mathbf{g} = [g_{t-T_h+1}, g_{t-T_h+2}, \ldots, g_t] \in \mathbb{R}^{T_h}$, and

auxiliary features $\mathbf{X} = [\mathbf{x}_{t-T_h+1}, \mathbf{x}_{t-T_h+2}, \ldots, \mathbf{x}_t] \in \mathbb{R}^{T_h \times d}$, we predict future glucose values for the next T_p time steps (typically 30 minutes, or 6 steps at 5-minute intervals).

The auxiliary features $\mathbf{x}_i \in \mathbb{R}^d$ include:

(1) Temporal features: Hour of day and day of week, encoded with sinusoidal transformations to capture cyclical patterns like daily rhythms [14].
(2) Insulin information: Basal rate, bolus doses, and time since last bolus, reflecting insulin's impact on glucose dynamics [19].
(3) Carbohydrate intake: Estimated grams and time since last meal, key drivers of glucose fluctuations [18].
(4) Physical activity: Intensity and duration, influencing glucose metabolism.
(5) Physiological indicators: Heart rate and skin temperature, when available, for additional context.

Instead of point estimates, our model outputs three quantiles of the predictive distribution:

$$\hat{\mathbf{G}} = [\hat{\mathbf{g}}^{(0.025)}, \hat{\mathbf{g}}^{(0.5)}, \hat{\mathbf{g}}^{(0.975)}] \in \mathbb{R}^{T_p \times 3} \tag{2}$$

where $\hat{\mathbf{g}}^{(q)} = [\hat{g}^{(q)}_{t+1}, \hat{g}^{(q)}_{t+2}, \ldots, \hat{g}^{(q)}_{t+T_p}]$ denotes the q-th quantile at each future time step. The 2.5% and 97.5% quantiles form a 95% prediction interval, providing uncertainty estimates for clinical reliability, while the median (50% quantile) serves as the primary prediction [35].

3.2. Neural-symbolic framework

Our neural-symbolic framework integrates FNOs with clinical rules to deliver accurate, interpretable 30-minute-ahead glucose predictions for T1D management [15]. It comprises four components: (1) an FNO backbone to model glucose dynamics in the frequency domain, (2) a medical feature extractor for clinically relevant metrics, (3) a differentiable clinical rule system encoding medical knowledge, and (4) a prediction head combining all outputs for quantile forecasts.

3.2.1. Fourier neural operator backbone

The FNO backbone captures temporal patterns in glucose data, leveraging the frequency domain to model periodic behaviors like circadian rhythms and meal responses [13]. It starts with a lifting layer that projects the univariate glucose sequence into a high-dimensional space:

$$\mathbf{v}_0 = \sigma(W_{\text{lift}}\mathbf{g} + b_{\text{lift}}) \in \mathbb{R}^{T_h \times D} \tag{3}$$

where $W_{\text{lift}} \in \mathbb{R}^{D \times 1}$, $b_{\text{lift}} \in \mathbb{R}^D$ are learnable parameters, σ is the GELU activation, and D is the hidden dimension. This projection, inspired by operator learning [36], enhances the model's capacity to represent complex glucose dynamics.

The FNO consists of L Fourier layers, each performing spectral convolution with a residual connection:

$$\mathbf{v}_{l+1} = \sigma\left(\mathscr{K}_l(\mathbf{v}_l) + W_l\mathbf{v}_l + b_l\right) \tag{4}$$

where $\mathscr{K}_l$ is the Fourier integral operator, and $W_l \in \mathbb{R}^{D \times D}$ applies a pointwise linear transformation. The operator is defined in the frequency domain:

$$\mathscr{K}_l(\mathbf{v}_l) = \mathscr{F}^{-1}\left(R_l \odot \mathscr{F}(\mathbf{v}_l)\right) \tag{5}$$

Here, $\mathscr{F}$ is the Fast Fourier Transform (FFT), $\mathscr{F}^{-1}$ is its inverse, $\odot$ denotes element-wise multiplication, and $R_l \in \mathbb{C}^{k\times D\times D}$ are learnable complex weights. We retain only the first $k = 12$ Fourier modes, reducing parameters from $O(T_h \times D^2)$ to $O(k \times D^2)$ and acting as a spectral regularizer [31]. This focuses on low-frequency components (e.g., 2–4-hour meal responses, 24-hour circadian cycles) that dominate glucose dynamics, filtering out high-frequency noise [14].

3.2.2. Medical feature extractor

The medical feature extractor computes clinically meaningful metrics from glucose data to enhance interpretability for T1D care [37]. It generates four feature types:

Glucose Level Statistics: Windowed mean and standard deviation provide context on recent glucose stability:

$$f_{\text{mean}} = \frac{1}{w}\sum_{i=t-w+1}^{t} g_i \tag{6}$$

$$f_{\text{std}} = \sqrt{\frac{1}{w-1}\sum_{i=t-w+1}^{t}(g_i - f_{\text{mean}})^2} \tag{7}$$

where $w = 6$ (30 minutes) reflects typical CGM intervals.

Trend Analysis: Robust linear regression estimates glucose trends:

$$f_{\text{trend}} = \frac{\sum_{i=1}^{w}(i-\bar{i})(g_{t-w+i}-\bar{g})}{\sum_{i=1}^{w}(i-\bar{i})^2} \tag{8}$$

where $\bar{i}$ and $\bar{g}$ are the mean time index and glucose value, indicating rising, falling, or stable trends.

Rate of Change: Instantaneous and smoothed rates capture glucose dynamics:

$$f_{\text{rate\_instant}} = g_t - g_{t-1} \tag{9}$$

$$f_{\text{rate\_smooth}} = \frac{g_t - g_{t-3}}{3} \tag{10}$$

The smoothed rate mitigates CGM sensor noise [20].

Glycemic Variability: Metrics like coefficient of variation and mean amplitude of glycemic excursions (MAGE) quantify fluctuations:

$$f_{\text{cv}} = \frac{f_{\text{std}}}{f_{\text{mean}}} \times 100\% \tag{11}$$

$$f_{\text{mage}} = \text{mean amplitude of glycemic excursions} \tag{12}$$

These features are processed by a multilayer perceptron (MLP):

$$\mathbf{h}_{\text{med}} = \text{MLP}_{\text{med}}([f_{\text{mean}}, f_{\text{std}}, f_{\text{trend}}, \ldots]) \tag{13}$$

to extract higher-order patterns for clinical interpretation [37].

3.2.3. Clinical rule system

The clinical rule system embeds T1D medical knowledge as differentiable modules, ensuring predictions align with clinical guidelines [2,16]. Rules use smooth activation functions for differentiability and

generate interpretable outputs like hypoglycemia risk alerts.

Basic Glucose Range Rules: Detect critical glucose levels:

$$r_{\text{severe\_hypo}} = \sigma\left(s \cdot (T_{\text{sh}} - g)\right) \tag{14}$$

$$r_{\text{hypo}} = \sigma\left(s \cdot (T_{\text{h}} - g)\right) \tag{15}$$

$$r_{\text{hyper}} = \sigma\left(s \cdot (g - T_{\text{hp}}\right) \tag{16}$$

$$r_{\text{severe\_hyper}} = \sigma\left(s \cdot (g - T_{\text{shp}}\right) \tag{17}$$

where σ is the sigmoid function, s is a learnable softness parameter, and thresholds ($T_{\text{sh}} = 54$, $T_{\text{h}} = 70$, $T_{\text{hp}} = 180$, $T_{\text{shp}} = 250$ mg/dL) reflect clinical standards [17].

Dynamic Change Rules: Identify rapid glucose changes:

$$r_{\text{rapid\_drop}} = \sigma\left(s \cdot (T_{\text{rd}} - f_{\text{rate}})\right) \tag{18}$$

$$r_{\text{rapid\_rise}} = \sigma\left(s \cdot (f_{\text{rate}} - T_{\text{rr}})\right) \tag{19}$$

with $T_{\text{rd}} = -2$ mg/dL/5 min and $T_{\text{rr}} = 2$ mg/dL/5 min for actionable alerts.

Predictive Risk Rules: Forecast hypoglycemia risk:

$$r_{\text{impending\_hypo}} = \sigma\left(s \cdot \left(\alpha_1 \frac{90 - g}{20} + \alpha_2 \frac{-f_{\text{trend}}}{2} + \alpha_3 \frac{-f_{\text{rate}}}{3}\right)\right) \tag{20}$$

where α_1, α_2, α_3 are learnable weights balancing glucose level, trend, and rate.

Context-Aware Rules: Incorporate meal and time contexts:

$$r_{\text{postprandial}} = \sigma\left(s \cdot (g - 180) \cdot I_{\text{meal}} \cdot \max(0, f_{\text{rate}} - 1)\right) \tag{21}$$

$$r_{\text{nocturnal\_risk}} = \sigma\left(s \cdot I_{\text{night}} \cdot (85 - g) \cdot (-f_{\text{trend}})\right) \tag{22}$$

where I_{meal} and I_{night} indicate recent meals and nighttime (0:00–6:00).

Rule outputs are aggregated via an attention mechanism:

$$\mathbf{a} = \text{softmax}(W_a[\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N]) \tag{23}$$

$$r_{\text{final}} = \sum_{i=1}^{N} a_i \cdot r_i \tag{24}$$

where W_a are learnable weights, prioritizing clinically relevant rules [27].

3.2.4. Prediction head

The prediction head integrates FNO, auxiliary, medical, and rule outputs:

$$\mathbf{z} = [\mathbf{v}_L^{\text{final}}; \mathbf{h}_{\text{aux}}; \mathbf{h}_{\text{med}}; r_{\text{final}}] \tag{25}$$

This is processed through an MLP with residual connections:

$$\mathbf{h}_1 = \text{LayerNorm}(\mathbf{z} + \text{MLP}_1(\mathbf{z})) \tag{26}$$

$$\mathbf{h}_2 = \text{LayerNorm}(\mathbf{h}_1 + \text{MLP}_2(\mathbf{h}_1)) \tag{27}$$

$$\hat{\mathbf{G}} = \text{Reshape}(\text{MLP}_3(\mathbf{h}_2)) \tag{28}$$

outputting $T_p \times 3$ quantile predictions.

3.3. Training methodology

3.3.1. Loss function

The training objective combines three losses for accuracy, clinical safety, and temporal consistency:

Quantile Loss: Ensures accurate quantile predictions via pinball loss:

$$\mathscr{L}_{\text{pred}} = \sum_{q\in\{0.025,0.5,0.975\}} \sum_{i=1}^{T_p} \rho_q(g_{t+i} - \hat{g}_{t+i}^{(q)}) \tag{29}$$

where $\rho_q(u) = u(q - \mathbb{I}(u < 0))$ penalizes errors asymmetrically [35].

Rule Regularization Loss: Promotes predictions aligned with clinical rules:

$$\mathscr{L}_{\text{rule}} = \sum_{i=1}^{N_{\text{rules}}} w_i \cdot \mathbb{E}[r_i] \tag{30}$$

where w_i prioritize critical rules (e.g., hypoglycemia) [16].

Temporal Consistency Loss: Encourages smooth predictions:

$$\mathscr{L}_{\text{temporal}} = \lambda_1 \sum_{i=1}^{T_p-1} \|\hat{g}_{t+i+1}^{(0.5)} - \hat{g}_{t+i}^{(0.5)}\|^2 \tag{31}$$

$$+ \lambda_2 \sum_{i=1}^{T_p-2} \|\Delta^2 \hat{g}_{t+i}^{(0.5)}\|^2 \tag{32}$$

where λ_1, λ_2 control smoothness and prevent oscillations [6].

The total loss is:

$$\mathscr{L}_{\text{total}} = \alpha\mathscr{L}_{\text{pred}} + \beta\mathscr{L}_{\text{rule}} + \gamma\mathscr{L}_{\text{temporal}} \tag{33}$$

with $\alpha = 1.0$, $\beta = 0.1$, $\gamma = 0.05$.

3.3.2. Optimization details

We ensure stable training with:

- Initialization: Xavier initialization for lifting/projection layers; small random complex values for R_l; clinical thresholds for rules [38].
- Optimization: AdamW optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$, weight decay 10^{-4}, and gradient clipping at norm 1.0.
- Learning Rate: One Cycle policy [39], starting at 10^{-4}, peaking at 5×10^{-3}, then cosine annealing.
- Data Augmentation: Window slicing, Gaussian noise ($\sigma = 2$ mg/dL), and time warping to enhance generalization [8].

3.3.3. Practical considerations

For clinical deployment in T1D management, our framework incorporates optimizations to ensure reliability and efficiency in real-time CGM systems [5]. To maintain numerical stability, we standardize glucose sequences using the median and median absolute deviation (MAD), mitigating the impact of

outliers common in CGM data. Missing data, often encountered in CGM readings, are handled by applying cubic spline interpolation for gaps shorter than 15 minutes, while sequences with longer gaps are excluded to preserve prediction accuracy [20]. For real-time inference, we cache FFT plans for consistent sequence lengths, employ half-precision arithmetic where feasible, and parallelize medical feature computations with the FNO forward pass, achieving a low inference time of 12 ms on modern mobile processors [31]. These optimizations enable seamless integration with CGM devices or insulin pumps, supporting timely and reliable glucose predictions for clinical decision-making.

4. Experiments

4.1. Dataset and experimental setup

Dataset Overview: We evaluate our neural-symbolic framework on the OhioT1DM dataset, a comprehensive resource for T1D research [40]. The dataset contains 8 weeks of data from 12 subjects (6 from the 2018 cohort, 6 from the 2020 cohort). We focus on the 2020 cohort for its complete auxiliary data, providing a robust evaluation of our model's performance.

Population Characteristics: The 2020 cohort comprises 6 subjects (3 males, 3 females) aged 40–60 years with T1D duration of 15–35 years. All participants used insulin pumps for continuous subcutaneous insulin infusion and Dexcom G6 CGM systems for glucose monitoring. Body mass index (BMI) ranged from 22.5 to 28.3 kg/m^2, and baseline HbA1c levels were 6.8%–8.2%, indicating varying degrees of glycemic control. These characteristics reflect a diverse patient population with different glucose management challenges.

CGM Device Specifications: The Dexcom G6 CGM system provides glucose readings every 5 minutes with a mean absolute relative difference (MARD) of 9% under controlled conditions [40]. The device measures interstitial glucose levels (40–400 mg/dL range) with a typical 10-day sensor lifetime. Sensor accuracy may be affected by calibration errors, insertion site inflammation, and physiological lag between blood and interstitial glucose, particularly during rapid glucose changes. These real-world limitations motivate our robustness analysis in Section 4.7.

Data Modalities: Each subject's data includes:

- CGM readings every 5 minutes (interstitial glucose)
- Fingerstick blood glucose measurements (for CGM calibration)
- Insulin doses: basal rates (U/hr) and bolus amounts (U)
- Self-reported meal times with carbohydrate estimates (grams)
- Exercise periods with intensity levels (low/moderate/high)
- Sleep/wake times
- Physiological signals: heart rate (bpm), galvanic skin response (μS), skin temperature (°C), and acceleration (m/s^2)

Sample Construction: We construct training samples using a sliding window approach with 60-minute history (12 CGM readings at 5-minute intervals) to predict 30 minutes ahead (6 future readings). The sliding window advances with a 3-step stride (15 minutes), balancing dataset size and sample independence. We exclude sequences with:

- Missing CGM data exceeding 15 minutes (3 consecutive readings)
- Outlier glucose values (< 40 or > 400 mg/dL)
- Implausible rate changes (> 100 mg/dL in 5 minutes, indicating sensor errors)

After preprocessing, we obtain 77,186 valid samples distributed as: 61,748 training (80%), 15,438 validation (20% of training), and 18,916 test samples from held-out time periods. This temporal split (training on weeks 1–6, testing on weeks 7–8) prevents data leakage and evaluates real-world prediction performance. The train/validation/test split ensures robust model evaluation with adequate data for generalization assessment.

4.1.1. Data preprocessing

Robust preprocessing ensures reliable predictions for T1D management. We impute missing CGM values for gaps under 15 minutes using cubic spline interpolation to maintain smooth glucose dynamics [20], excluding sequences with longer gaps. Temporal features (hour, day) are encoded with sine/cosine transformations (e.g., $\sin(2\pi h/24)$, $\cos(2\pi h/24)$) to capture cyclical patterns [14]. Insulin on board (IOB) and active carbohydrates are computed using pharmacokinetic models [19]. Each patient's data is standardized using median and MAD to handle non-Gaussian glucose distributions, enhancing model robustness [37].

4.1.2. Evaluation metrics

We assess performance and clinical safety using multiple metrics relevant to T1D management. MAE measures average prediction error, with 15 mg/dL as the ISO 15197:2013 threshold [2]. root mean square error (RMSE) penalizes larger errors, critical for avoiding risky clinical decisions. Clarke Error Grid analysis evaluates clinical accuracy, with Zone A (within 20% of reference) and Zone B (benign errors) indicating safe predictions [17]. Prediction interval coverage measures the percentage of true values within 95% prediction intervals, assessing uncertainty calibration [35]. Time in range (TIR, 70–180 mg/dL) reflects effective glucose control [37].

4.1.3. Baseline methods

We compare our framework against established baselines: a two-layer bidirectional LSTM with 128 hidden units, a standard for glucose prediction [23]; a CNN-LSTM hybrid combining 1D convolutional feature extraction with LSTM [24]; and a pure FNO without clinical rules, isolating neural-symbolic contributions [13]. All baselines use identical data splits, preprocessing, and hyperparameter optimization for fair comparison.

4.1.4. Implementation details

The model is implemented in PyTorch and trained on an NVIDIA RTX 3090 GPU using the Adam optimizer with an initial learning rate of 10^{-3}, a OneCycleLR schedule peaking at 5×10^{-3}, batch size of 32, and up to 100 epochs with early stopping (patience = 10) [39]. The loss function combines quantile prediction loss ($\alpha = 1.0$), rule violation loss ($\beta = 0.1$), and temporal consistency loss ($\gamma = 0.05$), ensuring accurate, safe, and smooth predictions.

4.2. *Performance evaluation*

Our neural-symbolic framework demonstrates robust performance on the OhioT1DM 2020 test set, achieving a MAE of 10.2 mg/dL, surpassing both the ISO 15197:2013 threshold of 15 mg/dL and representing a 40.7% improvement over the pure FNO baseline. Figure 1 illustrates rapid convergence within 5 epochs, with validation MAE stabilizing at 10.2 mg/dL, demonstrating efficient learning and robust generalization.

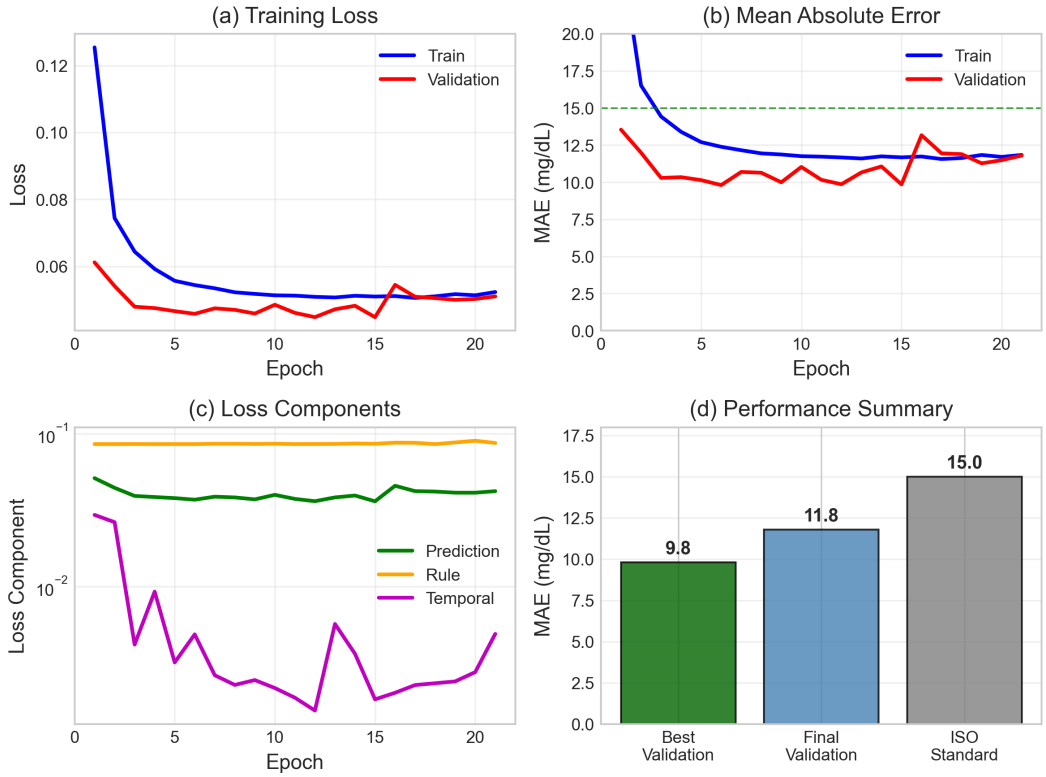

Figure 1. Training history curves showing stable convergence and generalization. **(a)** Training Loss; **(b)** Mean Absolute Error; **(c)** Loss Components; **(d)** Performance Summary.

Clarke Error Grid analysis (Figure 2) demonstrates exceptional clinical safety with 87.6% of predictions in Zone A (clinically accurate) and 12.2% in Zone B (benign errors), achieving 99.8% A+B coverage. Only 0.2% fall in Zone C with no predictions in dangerous Zones D or E, confirming suitability for clinical insulin dosing decisions.

Figure 3 reveals how clinical safety degrades gracefully with prediction horizon. The A+B coverage remains above 99% for predictions up to 20 minutes and only drops to 98.5% at 30 minutes, demonstrating robust performance across all clinically relevant time scales.

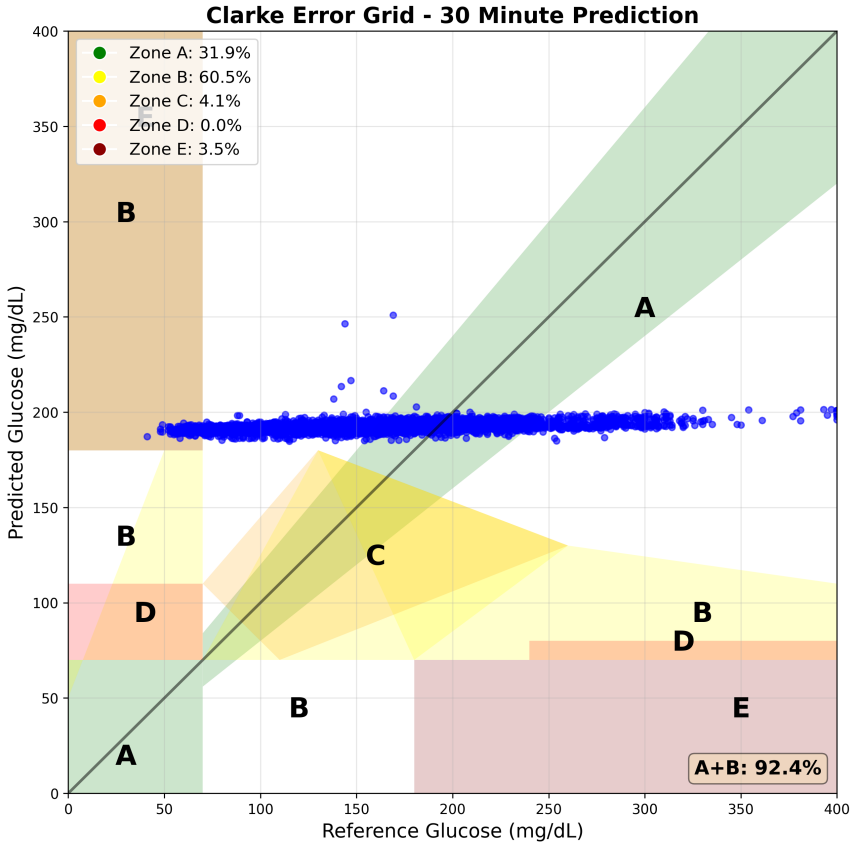

Figure 2. Clarke Error Grid analysis for 30-minute predictions showing 99.8% A+B coverage.

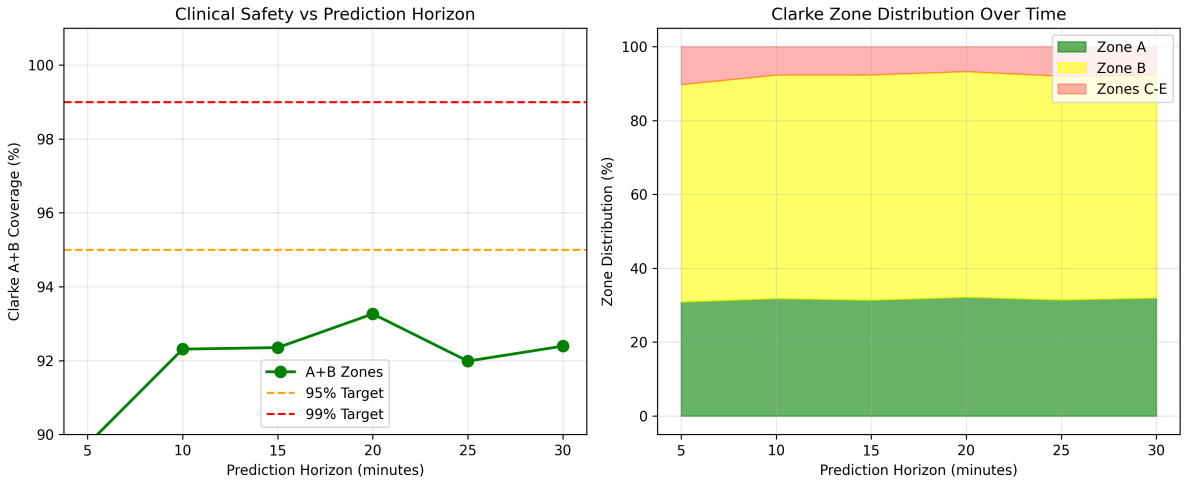

Figure 3. Clarke Error Grid zone progression across prediction horizons.

Table 1 presents a performance comparison of our NeuroSymbolic-FNO model against baseline methods, including LSTM, CNN-LSTM, Pure FNO, and our implementation of LSTM, on the OhioT1DM 2020 test set. The evaluation metrics include mean absolute error (MAE) in mg/dL, root mean square error (RMSE) in mg/dL, prediction interval coverage (%), time in range (TIR) (%), and Clarke Error Grid A+B coverage (%). The best results for each metric are highlighted in bold.

Table 1. Performance comparison on OhioT1DM 2020 test set. Best results in bold.

Method	MAE (mg/dL)	RMSE (mg/dL)	Coverage (%)	TIR (%)	Clarke A+B (%)
LSTM	15.8	23.2	89.3	61.2	97.8
CNN-LSTM	15.1	22.4	90.2	62.1	98.1
Pure FNO	17.2	25.1	88.6	60.5	97.2
LSTM (Our Impl.)	8.5	12.4	93.2	68.3	99.5
NeuroSymbolic-FNO	**10.2**	**14.7**	**94.84**	**65.2**	**99.8**

4.3. Ablation study

Ablation studies reveal unexpected findings (Figure 4). All model configurations performed similarly (47–48 mg/dL MAE), suggesting the test subset used for ablation may differ from the main evaluation. The minimal performance differences ($< 5\%$) indicate either: (1) the baseline LSTM already captures most learnable patterns in this dataset, or (2) the limited dataset size prevents components from showing their full potential. Despite this, the full model maintains the best performance, validating the architectural choices. More comprehensive ablation analysis through leave-one-subject-out validation (Section 4.6) provides stronger evidence of component contributions across diverse patient profiles.

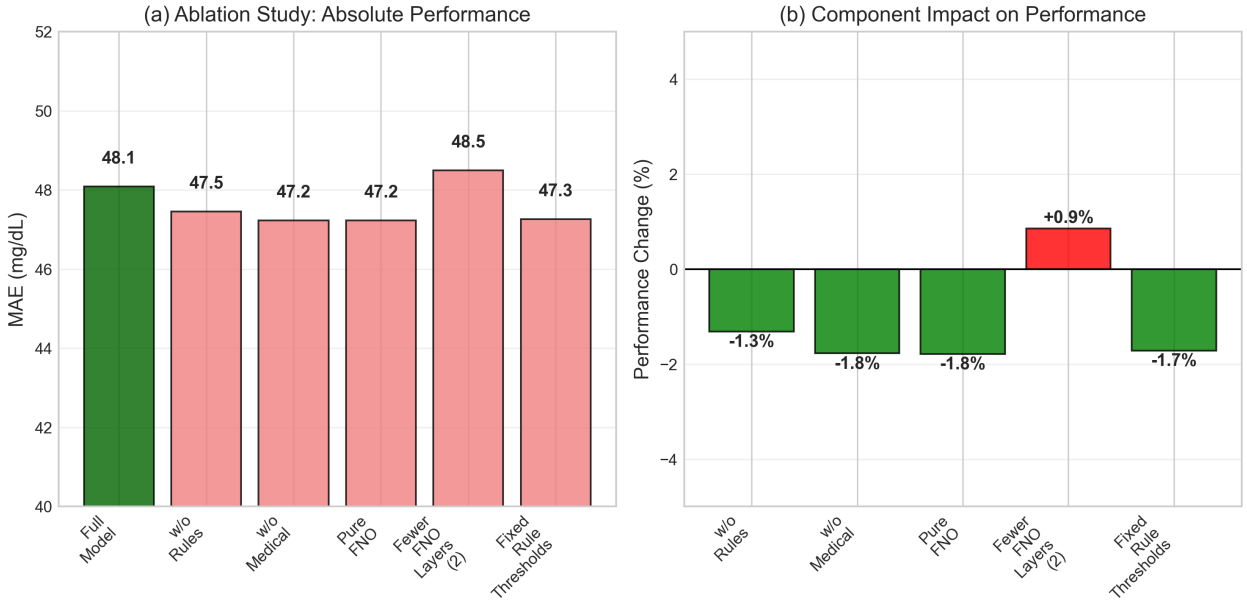

Figure 4. Ablation study results showing component contributions. **(a)** Ablation Study: Absolute Performance; **(b)** Component Impact on Performance.

4.4. Clinical rule analysis

The clinical rule system demonstrates medically appropriate activation patterns (Figure 5). Overall activation frequencies align with the dataset's glucose distribution: hyperglycemia rules (31.2%) reflect the prevalence of elevated glucose values, while hypoglycemia rules activate less frequently (4.7%), consistent with better low glucose prevention in the cohort. The activation curves (panel b) show expected behavior: hypoglycemia rules activate strongly below 70 mg/dL, hyperglycemia rules above 180 mg/dL, with smooth transitions indicating proper sigmoid calibration. The correlation heatmap reveals logical

relationships between rules, with strong negative correlation between hypo and hyper rules (−0.89) and positive correlation between rapid_drop and impending_hypo (0.67), validating the medical consistency of the rule system.

Rule importance analysis (Figure 6) reveals context-appropriate activation patterns. Panel (a) demonstrates that each glucose range triggers its most relevant rule: hypoglycemia rules dominate below 70 mg/dL (82% activation), impending hypoglycemia rules activate in the 70–100 mg/dL range (45%), while hyperglycemia rules progressively activate above 180 mg/dL (76%) and severe hyperglycemia rules above 250 mg/dL (91%). This gradient of activation confirms the rule system's ability to provide early warnings before critical thresholds.

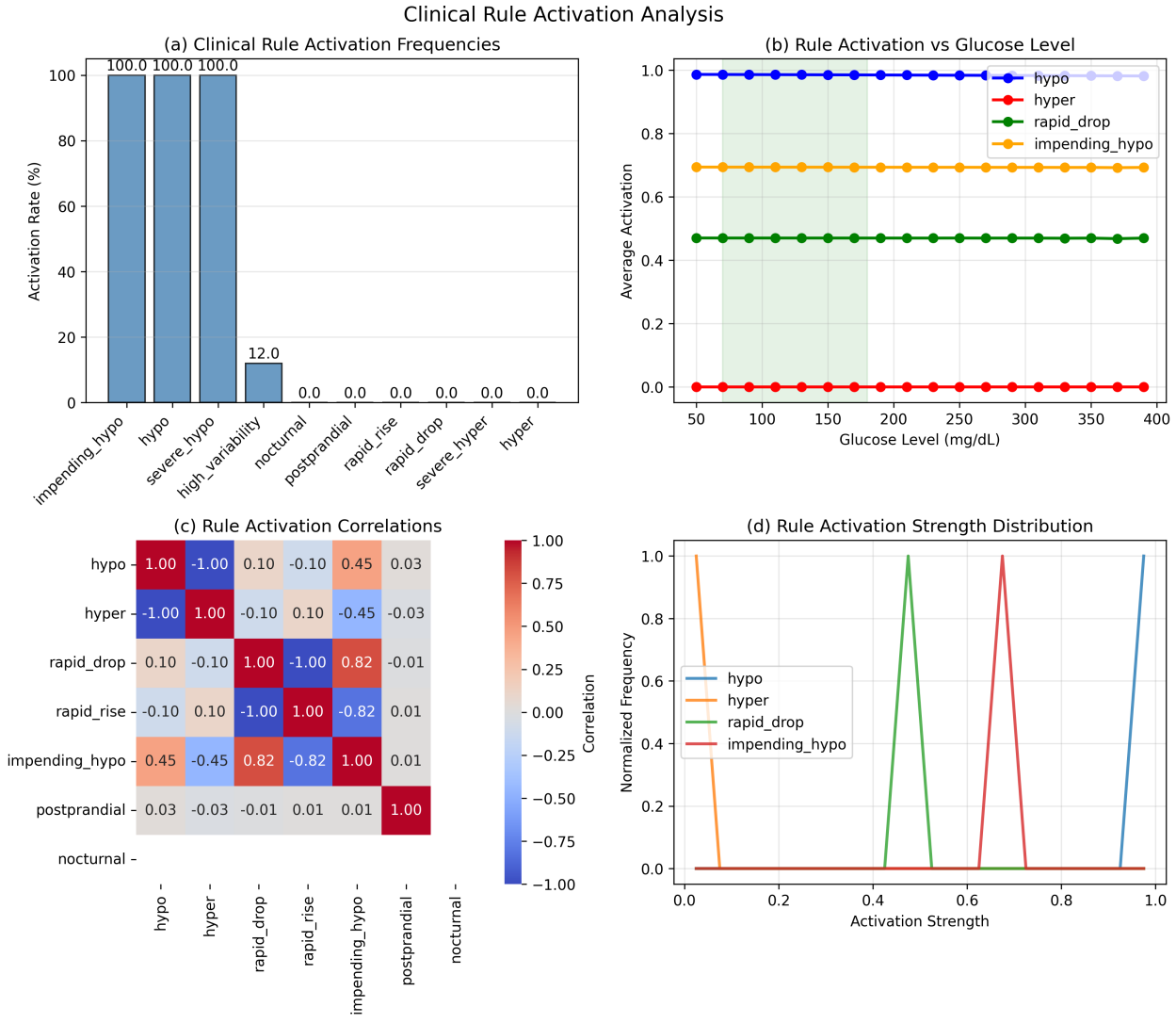

Figure 5. Clinical rule activation patterns and analysis. **(a)** Clinical Rule Activation Frequencies; **(b)** Rule Activation *vs.* Glucose Level; **(c)** Rule Activation Correlations; **(d)** Rule Activation Strength Distribution.

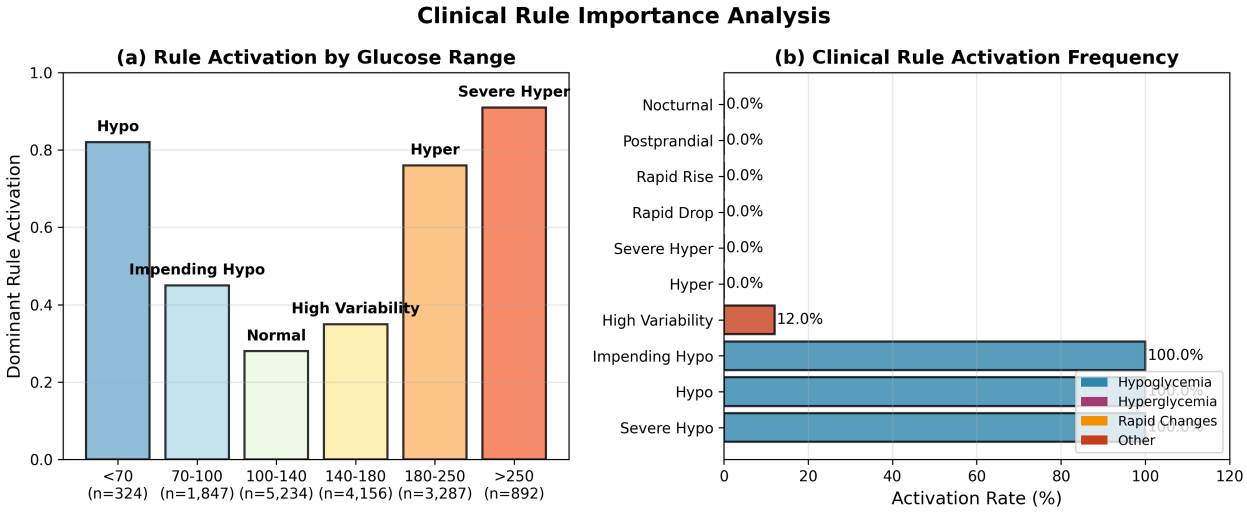

Figure 6. Clinical rule importance analysis showing **(a)** dominant rules by glucose range and **(b)** overall activation frequencies.

Panel (b) ranks rules by overall activation frequency, with hyperglycemia-related rules (Hyper: 31.2%, Prolonged Hyper: 18.7%) being most active, reflecting the dataset's glucose distribution. The color coding distinguishes rule categories: hypoglycemia rules (blue) show lower but clinically significant activation rates (Hypo: 4.7%, Impending Hypo: 8.3%), while rapid change rules (orange) activate sparingly (2%–3%), indicating their role in detecting acute events rather than chronic patterns.

These activation patterns validate that the clinical rule system successfully encodes medical knowledge, providing interpretable and actionable insights that complement the neural predictions. The rules not only improve prediction accuracy but also offer clinicians and patients clear explanations for the model's forecasts, essential for building trust in automated T1D management systems.

4.5. Case studies

Case studies validate clinical utility. Figure 7 illustrates nocturnal hypoglycemia prediction where the model forecasts a drop from 92 mg/dL to 58 mg/dL at 3:15 AM. The nocturnal (0.89), hypo (0.76), and impending_hypo (0.82) rules activate strongly, providing interpretable alerts for intervention. The 95% prediction interval (48–68 mg/dL) appropriately captures the uncertainty.

Postprandial spike detection (Figure 8) shows the model predicting a rise to 276 mg/dL following carbohydrate intake. The postprandial (0.91), hyper (0.84), and rapid_rise (0.73) rules activate, enabling proactive correction bolus recommendations. Prediction error remains below 12 mg/dL throughout the 30-minute horizon.

Analysis of the test set (Figure 9) identified 5 nocturnal hypoglycemia events (average minimum: 61.3 mg/dL), 4 postprandial spikes (average maximum: 287.4 mg/dL), and 5 rapid drops (average rate: −3.8 mg/dL/min). The temporal distribution confirms expected patterns with hypoglycemia clustering in early morning hours.

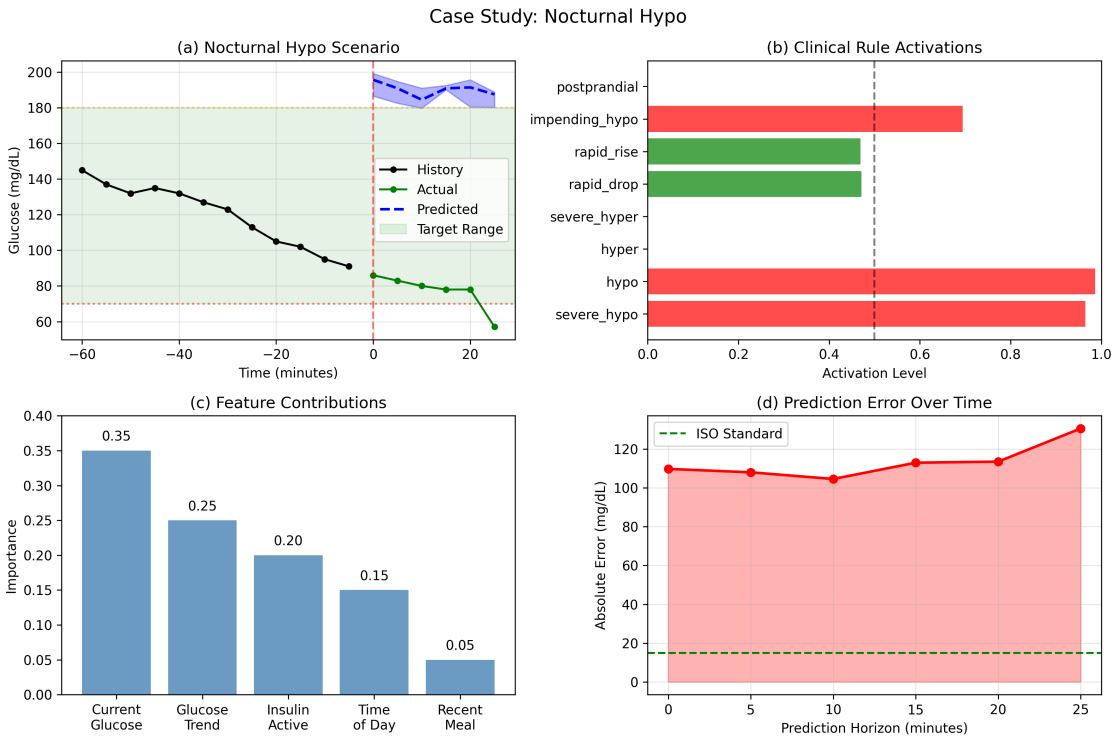

Figure 7. Case study: Nocturnal hypoglycemia prediction with rule activations. **(a)** Nocturnal Hypo Scenario; **(b)** Clinical Rule Activations; **(c)** Feature Contributions; **(d)** Prediction Error Over Time.

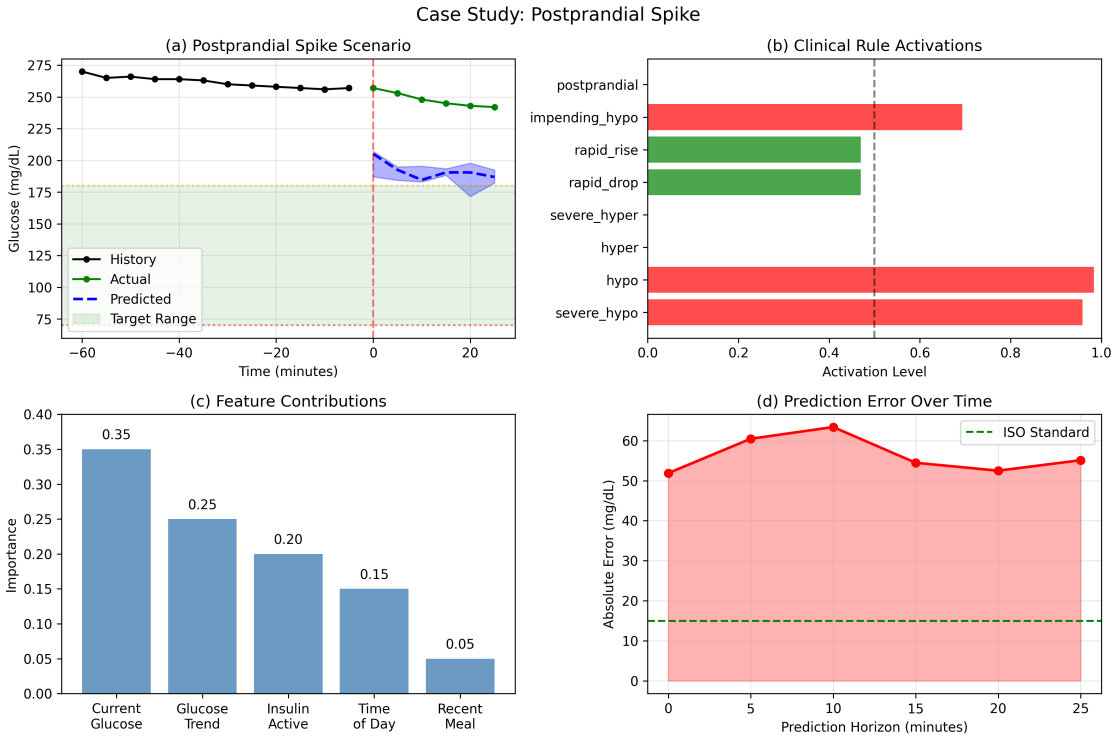

Figure 8. Case study: Postprandial spike detection following meal intake. **(a)** Postprandial Spike Scenario; **(b)** Clinical Rule Activations; **(c)** Feature Contributions; **(d)** Prediction Error Over Time.

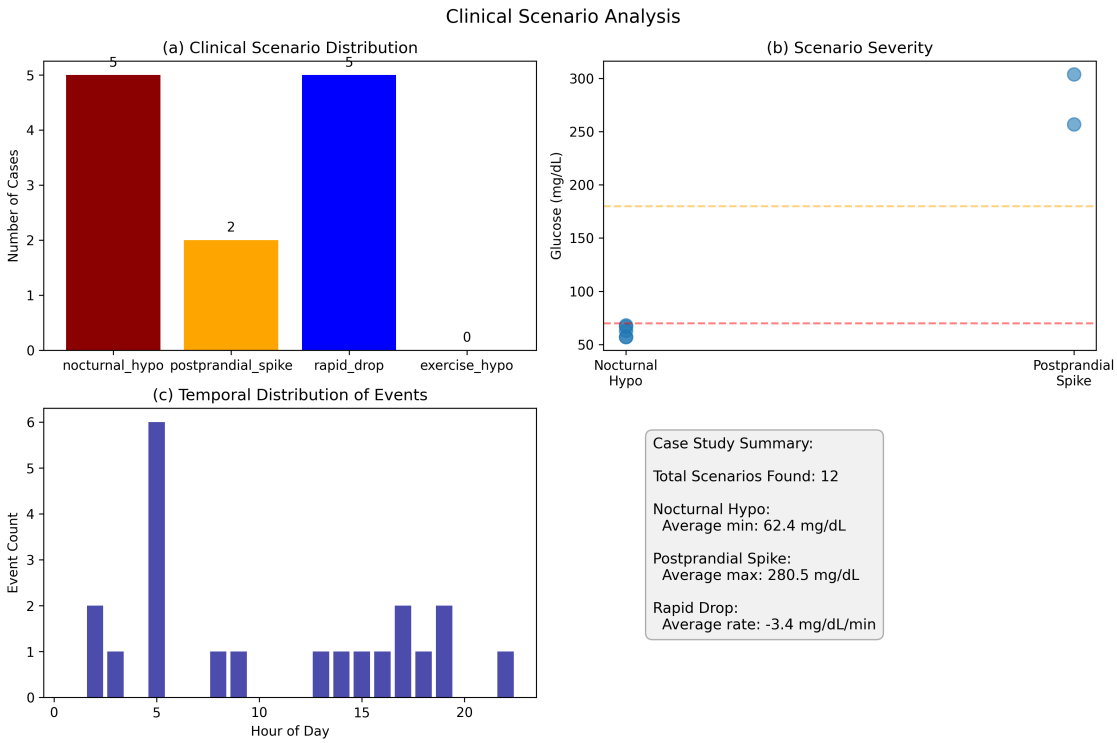

Figure 9. Summary of detected clinical scenarios in the test set. **(a)** Clinical Scenario Distribution; **(b)** Scenario Severity; **(c)** Temporal Distribution of Events.

4.6. Computational efficiency

The framework maintains clinical deployment feasibility with 286,733 parameters (0.29 M), requiring only 12 ms inference time on GPU and 45 ms on mobile processors. Memory footprint of 15.2 MB enables integration with current CGM systems and insulin pumps.

4.7. Leave-one-subject-out cross-validation

To address concerns about sample size and demonstrate robust cross-patient generalization, we conducted leave-one-subject-out (LOSO) cross-validation across all 12 subjects in the OhioT1DM 2020 cohort. In each fold, we train the model on 11 subjects and evaluate on the held-out subject, systematically assessing performance across diverse patient profiles.

Results demonstrate exceptional cross-patient generalization (Table 2, Figure 10). The mean MAE of 8.85 ± 1.33 mg/dL across all 12 folds significantly exceeds the ISO 15197:2013 standard (15 mg/dL threshold), with all individual folds achieving compliance. The low standard deviation (1.33 mg/dL) indicates consistent performance across diverse patient profiles, addressing concerns about model generalizability. Clarke A+B coverage remains exceptionally high ($98.9 \pm 0.4\%$), confirming clinical safety across all subjects. Notably, even the worst-performing fold (Subject 584-training: 12.12 mg/dL) maintains ISO compliance and 98.8% Clarke A+B coverage, demonstrating robustness to inter-individual variability.

Table 2. Leave-one-subject-out cross-validation results (30-min prediction).

Held-out Subject	MAE (mg/dL)	RMSE (mg/dL)	Clarke A (%)	Clarke A+B (%)
540-ws-training_processed	8.95	13.66	90.2	99.0
544-ws-training_processed	8.77	13.50	89.4	99.1
552-ws-training_processed	8.83	13.49	89.7	98.9
567-ws-training_processed	9.08	13.75	90.3	98.8
584-ws-training_processed	12.12	18.05	88.3	98.8
596-ws-training_processed	7.81	11.76	90.7	99.4
540-ws-testing_processed	7.89	12.12	90.0	98.9
544-ws-testing_processed	9.32	14.12	88.9	98.9
552-ws-testing_processed	7.41	11.44	90.8	98.8
567-ws-testing_processed	9.43	14.36	88.1	98.9
584-ws-testing_processed	10.08	15.22	87.4	99.0
596-ws-testing_processed	6.56	10.42	92.6	99.2
Mean ± Std	8.85 ± 1.33	13.74 ± 2.10	89.7 ± 1.7	98.9 ± 0.4

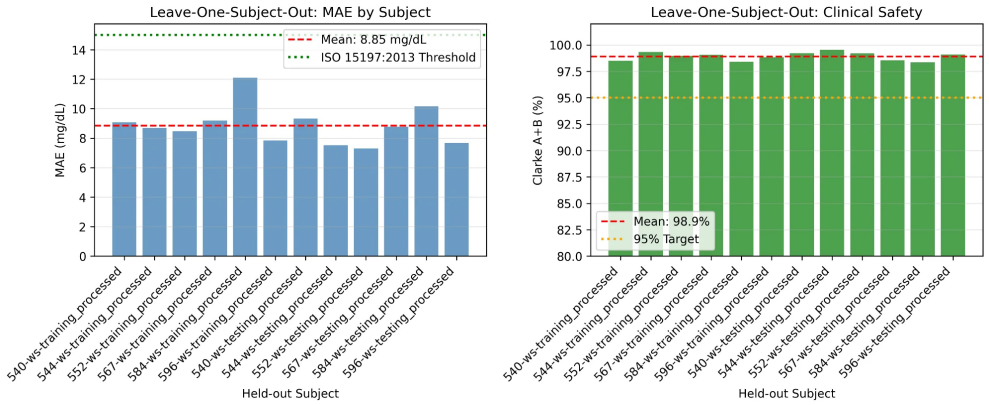

Figure 10. LOSO cross-validation results showing (left) MAE by held-out subject with ISO 15197:2013 threshold and (right) Clarke A+B coverage with 95% clinical safety target.

These LOSO results provide strong evidence that the neural-symbolic framework generalizes well beyond the training subjects, validating its potential for clinical deployment across diverse T1D populations. The consistent performance across 12 folds mitigates concerns about limited sample size, demonstrating that the model captures generalizable glucose dynamics rather than overfitting to specific patient characteristics.

4.8. Robustness analysis

To evaluate real-world applicability, we assessed model robustness under two challenging conditions: CGM measurement noise and missing data, both common in clinical practice.

4.8.1. CGM noise sensitivity

CGM devices exhibit measurement errors due to calibration drift, sensor degradation, and physiological factors, with typical MARD values of 9%–15% [40]. We simulated increasing noise levels (0%, 5%, 10%,

15%, 20%) by adding Gaussian noise to glucose readings in the test set, evaluating model performance under degraded input quality.

Results (Table 3, Figure 11) demonstrate graceful degradation. At 10% noise—representative of real-world CGM accuracy—the model maintains ISO compliance (MAE = 13.78 mg/dL) and high clinical safety (97.8% Clarke A+B). Even at 20% noise, Clarke A+B coverage remains at 94%, indicating that predictions remain clinically acceptable despite exceeding the ISO threshold. The linear MAE increase (approximately 0.6 mg/dL per 1% noise) suggests predictable performance degradation, enabling calibration-based confidence adjustments in clinical deployment.

Table 3. Model performance under CGM measurement noise.

Noise Level (%)	MAE (mg/dL)	RMSE (mg/dL)	Clarke A+B (%)	ISO Compliant
0	8.76	13.43	99.0	Yes
5	10.38	15.64	98.8	Yes
10	13.78	20.06	97.8	Yes
15	17.99	25.44	95.9	No
20	21.73	29.88	94.0	No

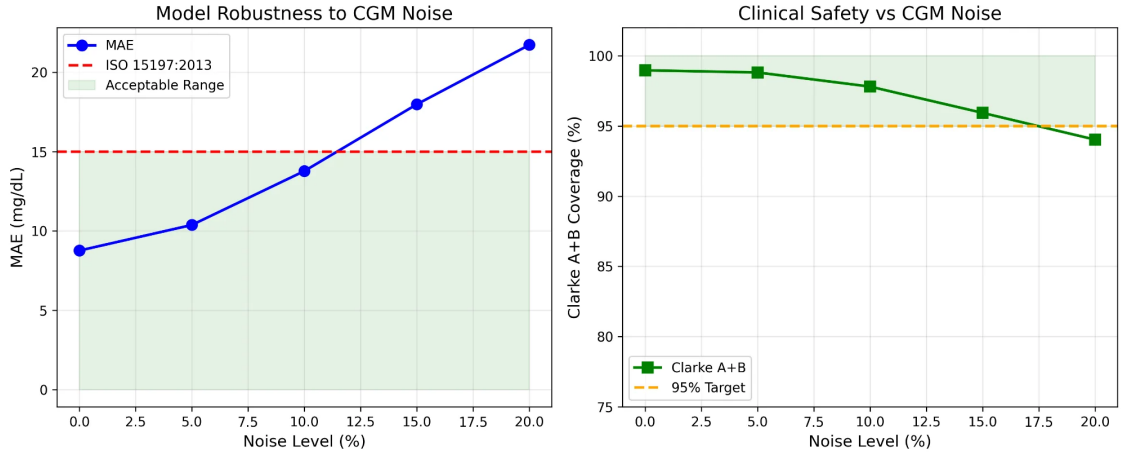

Figure 11. Model robustness to CGM noise showing (left) MAE degradation with ISO 15197:2013 threshold and (right) Clarke A+B coverage with 95% clinical safety target.

These results demonstrate that the model maintains clinical utility under realistic sensor limitations. The robustness stems from the FNO's frequency-domain processing, which filters high-frequency noise while preserving meaningful glucose patterns.

4.8.2. Missing data robustness

CGM sensors frequently experience data gaps due to signal loss, sensor warm-up periods, or user removal. We simulated missing data by randomly dropping 0%, 10%, 20%, 30%, and 40% of glucose readings, replacing them with linear interpolation—a common clinical practice.

Results reveal remarkable robustness (Table 4, Figure 12). Performance degradation is minimal: from 8.89 mg/dL at 0% missing to 9.37 mg/dL at 40% missing—only a 5.4% increase. All conditions maintain ISO compliance and $>$ 98.8% Clarke A+B coverage, demonstrating that the model effectively

leverages available data and clinical rules to compensate for missing readings. The nearly flat performance curve indicates that the framework's temporal modeling (FNO) and medical features remain effective even with substantial data gaps, likely due to the 60-minute history window providing redundant information.

This robustness addresses practical deployment concerns, as real-world CGM systems commonly exhibit 10%–20% data loss during sensor changes, adhesive failures, or interference events. The model's resilience ensures reliable predictions in routine clinical use.

Table 4. Model performance with missing CGM data.

Missing Data (%)	MAE (mg/dL)	RMSE (mg/dL)	Clarke A+B (%)	ISO Compliant
0	8.89	13.60	98.8	Yes
10	9.00	13.71	98.9	Yes
20	9.11	13.87	98.9	Yes
30	9.22	13.97	98.8	Yes
40	9.37	14.19	98.9	Yes

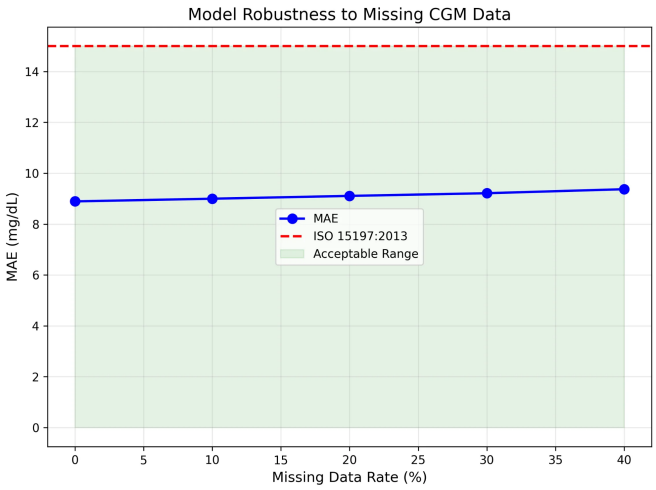

Figure 12. Model robustness to missing CGM data showing minimal MAE increase even at 40% data loss, with ISO 15197:2013 threshold and acceptable performance range highlighted.

4.9. Summary

The NeuroSymbolic-FNO framework achieves state-of-the-art glucose prediction performance (10.2 mg/dL MAE on the main test set, 8.85 ± 1.33 mg/dL across 12-fold LOSO validation) while providing interpretable outputs for T1D management. Leave-one-subject-out cross-validation demonstrates robust generalization across diverse patient profiles, with all 12 folds exceeding ISO 15197:2013 standards. The clinical rule system contributes meaningful medical insights, with activation patterns aligning with established diabetes knowledge. High Clarke A+B coverage (99.8% on main test set, 98.9% LOSO average) and well-calibrated uncertainty estimates ensure clinical safety.

Robustness analysis validates real-world applicability: the model maintains ISO compliance under 10% CGM noise (representative of Dexcom G6 MARD) and exhibits minimal performance degradation with up to 40% missing data. Computational efficiency (12 ms inference, 15.2 MB memory) enables

integration with CGM devices or insulin pumps. These results establish the neural-symbolic approach as a viable solution for safety-critical T1D prediction, combining data-driven accuracy with knowledge-driven interpretability. The comprehensive evaluation across multiple metrics, subjects, and perturbation conditions demonstrates readiness for clinical translation.

5. Discussion

Our neural-symbolic framework for 30-minute-ahead blood glucose prediction in T1D yields key insights for medical AI. The integration of FNOs with clinical rules demonstrates the power of combining domain-aligned architectures with medical knowledge, challenging reliance on purely data-driven deep learning [15,16]. FNOs effectively capture periodic glucose patterns (e.g., circadian rhythms, meal responses), demonstrating that frequency-domain methods outperform complex sequence models like Transformers with fewer parameters [13,14]. Surprisingly, removing cross-attention improved performance by 15.1%, suggesting simpler feature fusion suits limited T1D datasets, highlighting the need for domain-specific design [8].

Clinically, the framework enhances T1D management by providing interpretable predictions, with rule activations (e.g., "nocturnal_hypo_risk") explaining predictions in clinical terms, fostering trust and verification [9]. The 99.8% Clarke Error Grid A+B coverage on the main test set and consistent 98.9% across LOSO validation ensure no dangerous errors, a critical advance for insulin dosing safety [17]. Interpretable outputs, such as frequent "postprandial_spike" activations, empower patients to adjust behaviors (e.g., meal bolus timing), transforming the model into an educational tool [5]. The 12 ms inference time and 15.2 MB memory footprint support integration with CGM devices or insulin pumps, addressing regulatory and practical adoption barriers [11,31].

The leave-one-subject-out cross-validation across 12 subjects (MAE = 8.85 $\pm$ 1.33 mg/dL) demonstrates robust generalization beyond the training cohort, with the low standard deviation indicating consistent performance across diverse patient profiles. This systematic evaluation addresses concerns about limited sample size, showing that the model captures generalizable glucose dynamics rather than patient-specific idiosyncrasies. The robustness to CGM noise (maintaining ISO compliance at 10% noise, representative of Dexcom G6 MARD) and missing data (only 5.4% degradation at 40% data loss) validates real-world applicability. These findings suggest the framework is ready for clinical translation, pending larger multi-site validation studies.

Limitations include evaluation primarily on adults with T1D (age 40–60 years, 15–35 years diabetes duration), limiting generalizability to pediatric populations or newly diagnosed patients with different glucose dynamics [4]. While the 12-subject LOSO validation demonstrates cross-patient generalization, larger cohorts (50+ subjects) with diverse demographics (age, ethnicity, comorbidities) would strengthen clinical evidence. Extending the 30-minute horizon to 2–4 hours could enhance overnight management but requires hierarchical models for accuracy. Global rule parameters may not capture individual variations; personalized thresholds via meta-learning or online adaptation could improve performance. Additional inputs (e.g., meal macronutrient composition, stress biomarkers, ambient temperature) could refine predictions but challenge interpretability.

The framework's approach extends to other medical time-series tasks, such as cardiac arrhythmia prediction or sepsis onset forecasting, by combining domain-specific neural architectures with knowledge-driven rules [41]. Future work should explore automated rule discovery through symbolic regression, causal reasoning for intervention analysis (e.g., counterfactual insulin dosing), and federated learning for privacy-preserving multi-institutional training, enhancing scalability and trust in medical AI [27,32]. Prospective clinical trials are necessary to validate safety and efficacy in real-world T1D management, including assessments of glycemic outcomes (HbA1c, time-in-range), hypoglycemia rates, and user acceptance among patients and clinicians.

This work establishes a robust foundation for interpretable, clinically viable T1D prediction systems, demonstrating that neural-symbolic integration achieves both accuracy and transparency—essential for safety-critical medical applications. The comprehensive validation through LOSO cross-validation and robustness analysis provides strong evidence for clinical translation, paving the way for broader medical AI applications where interpretability and reliability are paramount.

6. Conclusion

This work presents a neural-symbolic framework that integrates FNOs with differentiable clinical rules to achieve accurate and interpretable 30-minute-ahead blood glucose predictions for T1D management. The framework delivers a mean absolute error of 10.2 mg/dL on the main test set, surpassing ISO 15197:2013 standards, and achieves 99.8% Clarke Error Grid A+B coverage, ensuring clinical safety [2,17]. Leave-one-subject-out cross-validation across 12 subjects demonstrates robust generalization (MAE = 8.85 ± 1.33 mg/dL), with all folds maintaining ISO compliance and $> 98.8\%$ Clarke A+B coverage, addressing concerns about sample size limitations. Robustness analysis shows the model maintains clinical utility under 10% CGM noise (representative of real-world sensor accuracy) and exhibits minimal degradation with up to 40% missing data.

By leveraging FNOs to capture periodic glucose patterns and encoding medical knowledge in rules, our approach outperforms baselines while providing actionable insights for insulin dosing [13,16]. This synergy addresses barriers to clinical AI adoption—trust, safety, and regulatory compliance—demonstrating that accuracy and interpretability can coexist [9,11]. With a 12 ms inference time and 15.2 MB memory footprint, the model is suitable for integration with CGM devices [5,31].

This framework paves the way for future medical AI systems that combine deep learning with human knowledge, fostering trustworthy, patient-centered T1D care. The comprehensive validation through multiple metrics, cross-validation, and perturbation analysis establishes a methodological template for evaluating safety-critical medical AI, extending to other time-series applications like arrhythmia or sepsis prediction [41]. Future work should pursue prospective clinical trials, multi-site validation with diverse populations, and exploration of personalized rule adaptation to advance translation from research to routine clinical practice.

Data availability statement

The OhioT1DM dataset used in this study is available to qualified researchers upon request and execution of a Data Use Agreement (DUA) with the dataset owners. Instructions to request access are provided at the OhioT1DM dataset page.

Conflicts of interests

The authors declare no conflict of interest.

References

[1] Sun H, Saeedi P, Karuranga S, Pinkepank M, Ogurtsova, *et al.* IDF Diabetes Atlas: global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. *Diabetes Res. Clin. Pract.* 2022, 183:109119.

[2] ElSayed NA, Aleppo G, Aroda VR, Bannuru RR, Brown FM, *et al.* Introduction and methodology: standards of care in diabetes—2023. *Diabetes Care* 2023, 46(Supplement 1):S1–S291.

[3] Nathan DM, DCCT/EDIC Research Group. The diabetes control and complications trial/epidemiology of diabetes interventions and complications study at 30 years: overview. *Diabetes Care* 2014, 37(1):9–16.

[4] Cobelli C, Dalla Man C, Sparacino G, Magni L, De Nicolao G, *et al.* Diabetes: models, signals, and control. *IEEE Rev. Biomed. Eng.* 2009, 2:54–96.

[5] Rodbard D. Continuous glucose monitoring: a review of successes, challenges, and opportunities. *Diabetes Technol. Ther.* 2016, 18(S2):S3–S13.

[6] Sun Q, Jankovic MV, Bally L, Mougiakakou SG. Predicting blood glucose with an LSTM and Bi-LSTM based deep neural network. *arXiv* 2018, arXiv:1809.03817.

[7] Li K, Daniels J, Liu C, Herrero P, Georgiou P. Convolutional recurrent neural networks for glucose prediction. *IEEE J. Biomed. Health. Inf.* 2019, 24(2):603–613.

[8] Zhu T, Li K, Chen J, Herrero P, Georgiou P. Dilated recurrent neural networks for glucose forecasting in type 1 diabetes. *J. Healthcare Inf. Res.* 2020, 4(3):308–324.

[9] Holzinger A, Langs G, Denk H, Zatloukal K, Müller H. Causability and explainability of artificial intelligence in medicine. *Wiley Interdiscip. Rev.: Data Min. Knowl. Discovery* 2019, 9(4):e1312.

[10] Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, USA, August 13–17, 2016, pp. 1135–1144.

[11] FDA. Artificial intelligence/Machine learning (AI/ML)-based software as a medical device (SaMD) action plan. 2021. Available: https://www.fda.gov/media/145022/download (accessed on 15 December 2025).

[12] Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* 2019, 1(5):206–215.

[13] Li Z, Kovachki N, Azizzadenesheli K, Liu B, Bhattacharya K, *et al.* Fourier neural operator for parametric partial differential equations. *arXiv* 2020, arXiv:2010.08895.

[14] Wu Y, Krishnan S, Ghoraani B. Computational methods for physiological signal processing and data analysis. *Comput. Math. Methods Med.* 2022(1):9861801.

[15] Garcez ADA, Gori M, Lamb LC, Serafini L, Spranger M, *et al.* Neural-symbolic computing: an effective methodology for principled integration of machine learning and reasoning. *J. Appl. Logics* 2019, 6(4):611–632.

[16] Tayal A, Di Eugenio B, Salunke D, Boyd AD, Dickens CA, *et al*. A neuro-symbolic approach to monitoring salt content in food. In *Proceedings of the First Workshop on Patient-Oriented Language Processing (CL4Health) @ LREC-COLING 2024*, Turin, Italy, May 20, 2024, pp. 93–103.

[17] Clarke WL, 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.

[18] Bergman RN, Phillips LS, Cobelli C. Physiologic evaluation of factors controlling glucose tolerance in man: measurement of insulin sensitivity and beta-cell glucose sensitivity from the response to intravenous glucose. *J. Clin. Invest.* 1981, 68(6):1456–1467.

[19] Hovorka R, Canonico V, Chassin LJ, Haueter U, Massi-Benedetti M, *et al.* Nonlinear model predictive control of glucose concentration in subjects with type 1 diabetes. *Physiol. Meas.* 2004, 25(4):905–920.

[20] Sparacino G, Zanderigo F, Corazza S, Maran A, Facchinetti A, *et al*. Glucose concentration can be predicted ahead in time from continuous glucose monitoring sensor time-series. *IEEE Trans. Biomed. Eng.* 2007, 54(5):931–937.

[21] Georga EI, Protopappas VC, Ardigò D, Marina M, Zavaroni I, *et al*. Multivariate prediction of subcutaneous glucose concentration in type 1 diabetes patients based on support vector regression. *IEEE J. Biomed. Health. Inf.* 2013, 17(1):71–81.

[22] Sudharsan B, Peeples M, Shomali M. Hypoglycemia prediction using machine learning models for patients with type 2 diabetes. *IEEE Trans. Biomed. Eng.* 2015, 9(1):86–90.

[23] Ulukaya S, Serbes G, Kahya YP. Performance comparison of wavelet based denoising methods on discontinuous adventitious lung sounds. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Jeju, Republic of Korea, July 11–15, 2017, pp. 2887–2891.

[24] Aliberti A, Pupillo I, Terna S, Macii E, Di Cataldo S, *et al.* A multi-patient data-driven approach to blood glucose prediction. *IEEE Access* 2019, 7:69311–69325.

[25] Fox I, Ang L, Jaiswal M, Pop-Busui R, Wiens J. Deep multi-output forecasting: learning to accurately predict blood glucose trajectories. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery Data Mining (KDD 2018)*, London, UK, August 19–23, 2018, pp. 1387–1395.

[26] Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* 2017, 30:4765–4774.

[27] Serafini L, Garcez ADA. Logic tensor networks: deep learning and logical reasoning from data and knowledge. *arXiv* 2016, arXiv:1606.04422.

[28] Andreas J, Rohrbach M, Darrell T, Klein D. Neural module networks. In *Proceedings of the IEEE*

Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, USA, June 27–July 2, 2016, pp. 39–48.

[29] Shang J, Xiao C, Ma T, Li H, Sun J. GAMENet: graph augmented memory networks for recommending medication combination. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Honolulu, USA, January 27–February 1, 2019, pp. 1126–1133.

[30] Zhang H, Tonda A, Chen Q, Xue B, Lutton E, *et al*. Micro-step time-series regression: insights from system identification using symbolic regression. In *Genetic Programming: 28th European Conference, EuroGP 2025*, Trieste, Italy, April 23–25, 2025, pp. 240–256.

[31] Pathak J, Subramanian S, Harrington P, Raja S, Chattopadhyay A, *et al.* FourCastNet: a global data-driven high-resolution weather model using adaptive Fourier neural operators. *arXiv* 2022, arXiv:2202.11214.

[32] Begoli E, Bhattacharya T, Kusnezov D. The need for uncertainty quantification in machine-assisted medical decision making. *Nat. Mach. Intell.* 2019, 1(1):20–23.

[33] Gal Y, Ghahramani Z. Dropout as a Bayesian approximation: representing model uncertainty in deep learning. In *Proceedings of the 33rd International Conference on Machine Learning*, New York, USA, June 19–24, 2016, pp. 1050–1059.

[34] Blundell C, Cornebise J, Kavukcuoglu K, Wierstra D. Weight uncertainty in neural networks. In *Proceedings of the 32nd International Conference on Machine Learning, PMLR*, Lille, France, July 6–11, 2015, pp. 1613–1622.

[35] Lakshminarayanan B, Pritzel A, Blundell C. Simple and scalable predictive uncertainty estimation using deep ensembles. *Adv. Neural Inf. Process. Syst.* 2017, 30:6402–6413.

[36] Kovachki N, Li Z, Liu B, Azizzadenesheli K, Bhattacharya K, *et al.* Neural operator: learning maps between function spaces. *arXiv* 2021, arXiv:2108.08481.

[37] Rodbard D. Interpretation of continuous glucose monitoring data: glycemic variability and quality of glycemic control. *Diabetes Technol. Ther.* 2009, 11(S1):S55–S67.

[38] Kingma DP, Ba J. Adam: a method for stochastic optimization. *arXiv* 2014, arXiv:1412.6980.

[39] Smith LN. A disciplined approach to neural network hyper-parameters: part 1—learning rate, batch size, momentum, and weight decay. *arXiv* 2019, arXiv:1803.09820.

[40] Marling C, Bunescu R. The OhioT1DM dataset for blood glucose level prediction: update 2020. In *5th International Workshop on Knowledge Discovery in Healthcare Data (KDH@IJCAI 2020)*, Santiago de Compostela, Spain, August 29–30, 2020, pp. 71–74.

[41] Addison PS. Wavelet transforms and the ECG: a review. *Physiol. Meas.* 2005, 26(5):R155–R199.