A Personalized, Non-Invasive Blood Glucose Prediction System Using a CNN-LSTM Model with a Sim-to-Real Transfer Learning Strategy

Komal Dandge
Dr Vishwanath Karad MIT World Peace University, PUNE, Maharashtra, India
komal.dandge@gmail.com (corresponding author)

Manisha Kumawat
Dr. Vishwanath Karad MIT World Peace University, PUNE, Maharashtra, India
manisha.kumawat@mitwpu.edu.in

Parul Jadhav
Dr. Vishwanath Karad MIT World Peace University, PUNE, Maharashtra, India
parul.jadhav@mitwpu.edu.in

Received: 12 November 2025 | Revised: 23 December 2025, 5 January 2026, and 9 January 2026 | Accepted: 12 January 2026

 DOI: https://doi.org/10.48084/etasr.16217

ABSTRACT

The translation of Non-Invasive Glucose Monitoring (NIBGM) into routine clinical care is currently hindered by two significant engineering hurdles: the inherently weak signal fidelity, characterized by a low Signal-to-Noise Ratio (SNR), in optical sensors, and the "Cold Start" constraint common in deep learning applications. Traditional recursive architectures, such as Long Short-Term Memory networks (LSTMs), often struggle to generalize in data-sparse contexts typical of personalized medicine. To address these limitations, this study presents the Hybrid Temporal-Attention Network (HTAN), a novel framework that disentangles sensor artifacts from metabolic trends by integrating Residual Convolutional Neural Networks (CNNs) with Bi-directional LSTMs and Multi-Head Attention mechanisms. The validity of this architecture was assessed using a "Dual-Stream" protocol, with pre-training on a high-precision "Chaos-Augmented" Digital Twin ($N_{syn} = 100,2 \times 10^6$ samples) followed by benchmarking against three distinct real-world clinical cohorts ($N_{real} = 119$). Our analysis reveals a significant performance divergence by phenotype: while conventional Random Forest models achieve adequate accuracy for stable Type 2 Diabetes (MARD = 6.18%), the HTAN model exhibits superior robustness in high-volatility Type 1 Diabetes scenarios. Notably, on the complex Shanghai T1DM dataset (N=7), standard Bi-LSTMs struggled to track the dynamics effectively (Mean Absolute Relative Difference, MARD = 8.65%), whereas HTAN achieved a markedly lower MARD of 5.11%. These results indicate that attention-based methodologies, when initialized with synthetic physiological priors, offer a viable approach to modelling intricate metabolic instability in regimes with limited data.

Keywords-Biomedical signal processing; deep learning; digital twin; glucose monitoring; hybrid neural networks; time-series forecasting

I. INTRODUCTION

Data published in the 2021 IDF Diabetes Atlas shows that the global diabetic population has surpassed 537 million adults, underscoring the critical nature of this public health crisis [1]. The transition from traditional invasive finger-stick techniques to NIBGM is currently hindered by two significant technical barriers: the inherently poor signal-to-noise ratio and the "Cold Start" limitations of deep learning models. Although bioimpedance and optical sensing technologies provide a theoretical basis for needle-free tracking, the target glucose signal is often obscured by physiological noise, such as hydration variations and motion interference [2]. Moreover, deep learning frameworks that capture long-term temporal dependencies require large datasets to converge. In contrast, clinical trials typically yield sparse data from small participant groups (N < 20), which can lead to model overfitting. Current literature frequently validates models using uniform datasets, which fail to demonstrate reliability across diverse metabolic phenotypes [3].

Diabetes care is currently experiencing a fundamental shift driven by the fusion of advanced data modelling and non-invasive biosensing. However, the transition from

electrochemical sensors to non-invasive methodologies introduces a profound signal-processing challenge: a drastic reduction in the SNR. Unlike the direct measurement of glucose oxidase currents in Interstitial Fluid (ISF), non-invasive signals such as Near-Infrared (NIR) spectroscopy or Photoplethysmography (PPG) are inherently stochastic. They are heavily distorted by environmental factors, physiological artifacts, and the nonlinear dynamics of glucose transport between biological compartments.

Alongside these hardware issues is the "Cold Start" dilemma in clinical machine learning. Deep learning architectures are notoriously data-intensive, requiring vast repositories of labelled time-series data to identify generalizable features. In Type 1 Diabetes (T1DM), where phenotypic volatility is high and prediction errors can be fatal, reliance on small, noisy datasets creates a significant barrier to entry. To address these interconnected issues—specifically, signal fidelity in non-invasive sensors and data scarcity in personalized modelling—this study proposes the HTAN. This framework is validated through a unique Dual-Stream protocol that uses in silico "Chaos-Augmented" Digital Twins (TSs) and multi-center clinical benchmarking.

This study presents the following primary contributions:

- Chaos-Augmented Digital Twin (DT): We developed a novel simulation engine that extends the Bergman Minimal Model [4] by incorporating nonlinear renal clearance and behavioral stochasticity, generating diverse and high-volatility synthetic priors.
- HTAN Architecture: We present an end-to-end HTAN that integrates Residual CNNs for noise reduction, Bi-LSTMs to capture temporal dynamics, and Multi-Head Attention for context-aware weighting.
- Dual-Stream Validation: We have established a comprehensive benchmarking protocol that assesses the model's performance against both synthetic physics priors and real-world clinical data, thereby quantifying the "Volatility Trade-Off" between traditional algorithms and deep learning approaches.

A. Evolution of Glycemic Forecasting Models

The landscape of glycemic forecasting has undergone a significant transformation, moving from static, limited-capacity classical regression techniques to robust, high-dimensional Deep Learning (DL) frameworks.

1) From Classical Machine Learning (ML) to Deep Representations

Before the dominance of deep learning frameworks, blood glucose prediction was primarily performed using shallow learning methods such as Support Vector Regression (SVR) and Random Forests (RF). Despite their low computational overhead, these techniques often prove inadequate for modeling complex, nonlinear temporal correlations. Furthermore, their reliance on stationary assumptions makes them fundamentally unsuitable for the erratic physiological fluctuations characteristic of T1DM. Research by [5] highlighted that SVR models frequently plateaued in performance, failing to capture the long-term temporal dependencies inherent in glucose regulation. Similarly, the authors in [6] showed that although gradient-boosting models were faster, they did not match the predictive accuracy of neural networks for prediction horizons beyond 30 minutes, a critical window for preventing hypoglycemia.

2) The LSTM Paradigm and Its Limits

The introduction of LSTMs marked a paradigm shift. Research in [5], [7], and [8] demonstrated that LSTMs and bidirectional LSTMs (Bi-LSTMs) consistently outperform human experts by learning physiological lags directly from raw data. However, standard LSTMs suffer from significant limitations in NIBGM:

- High-Frequency Noise Sensitivity: As noted in analyses of the OhioT1DM dataset [6], LSTMs can perform poorly when trained on noisy, multivariate data over long horizons, as the recurrent nature of the model propagates noise from previous time steps.
- Lack of Explicit Attention: Initial studies on Long Short-Term Memory (LSTM) models did not incorporate mechanisms to explicitly "attend" to particular critical events (such as a significant decrease in glucose levels), thereby treating all temporal steps as equally significant.

3) The Transformer Revolution

To overcome the "forgetting" issues of LSTMs, the field has progressively adopted Attention mechanisms. Recent work by [9] represents the state of the art, fusing the global contextualization of transformers with the local sequential strengths of LSTMs to achieve superior short-term forecasting. To address the latency constraints of wearable hardware, recent architectures have focused on optimizing transformer efficiency; for instance, authors in [10] introduced an edge-compatible Temporal Fusion Transformer (TFT) tailored to perform multi-step predictions directly on neuromorphic systems with limited computational resources. Nevertheless, transformers remain notoriously data-hungry, often requiring millions of training examples to generalize effectively, a luxury not available in standard clinical datasets like OhioT1DM [11].

B. Challenges in Non-Invasive Sensing

The primary bottleneck in NIBGM is the extremely low SNR, which often dips below 10 dB due to physiological interference [12].

1) Physics of Signal Degradation

In NIR modalities, water's high absorption coefficient overwhelms the glucose signature in the 1400–1500 nm band, making it a dominant interferent [13]. Similarly, PPG signals are susceptible to motion artifacts that mask subtle waveform changes [14]. The noise in NIBGM is not merely "white noise"; it is often "colored" (time-correlated) and non-stationary, rendering static filters ineffective [15].

2) Learnable Feature Extraction

The dominant paradigm has been "Filter-then-Predict," using Kalman Filters that struggle with colored noise. Recent research has shifted toward using CNNs as front-end feature

extractors [14, 16]. The proposed HTAN model advances this by embedding a Residual CNN block directly into the forecasting architecture, allowing noise filters to be jointly optimized with temporal prediction.n.

C. Synthetic Data & DTs

To address the Cold Start problem, the field has turned to physiological simulators. Historically, in silico trials have relied on the UVA/Padova simulator as the benchmark environment [17, 18]. Despite its widespread adoption, this model is often noted for generating idealized physiological profiles that lack the unpredictable variability found in real patient data, such as missed boluses or daily stressors [19, 20]. Furthermore, validation against free-living data suggests that mathematical models often overlook safety-critical physiological confounders present in real patient cohorts [21]. While Generative Adversarial Networks (GANs) have been used for data augmentation [22], they can hallucinate biologically impossible scenarios. Our proposed Dual-Stream Validation protocol directly addresses this gap by using a "Chaos-Augmented" DT. By enriching the training data with simulated behavioral artifacts, we pre-train the HTAN model to handle volatility before it encounters sparse clinical data.

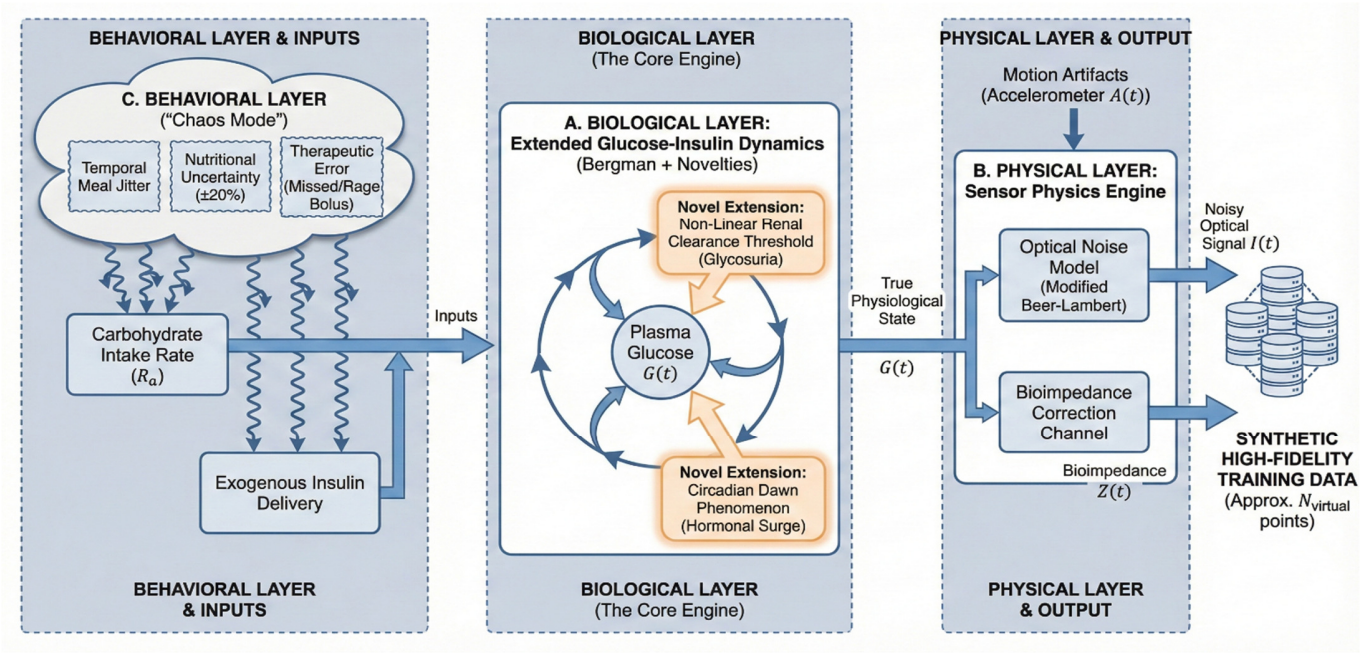

Fig. 1. Schematic Overview of the hybrid DT framework.

II. METHODOLOGY: THE HYBRID DIGITAL TWIN FRAMEWORK

To address the severe lack of data in non-invasive glucose monitoring, particularly for volatile phenotypes such as T1DM, we engineered a "Hybrid DT" simulation environment shown in Figure 1. This generative engine fuses deterministic physiological principles with stochastic sensor physics and behavioural randomness to synthesize high-fidelity training datasets ($N_{syn} = 100$ virtual subjects, approximately 2×10^6 data points).

A. Biological Layer: Chaos-Augmented Physiological Modeling

The foundational layer simulates the pharmacokinetics and pharmacodynamics (PK/PD) of glucose-insulin regulation. We extended the classical Bergman Minimal Model to account for critical nonlinearities absent in standard ODE formulations, specifically renal clearance thresholds and circadian hormonal variations, which are essential for modelling pathological states.

Extended Glucose-Insulin Dynamics: The temporal evolution of plasma glucose levels, denoted as $G(t)$ [mg/dL], is described by a coupled differential equation system derived from the foundational Bergman Minimal Model [14]:

$$\frac{dG(t)}{dt} = -\big(p_1 + X(t)\big)G(t) + p_1 G_b + \frac{R_a(t)}{V_d} - E_{renal}(G) + G_{circ}(t) \quad (1)$$

where is the$X(t)$ (Remote Insulin Action). This term quantifies the impact of insulin within the interstitial fluid compartment. Its kinetics are governed by the following equation (2) [14]:

$$\frac{dX(t)}{dt} = -p_2 X(t) + p_3 I(t), \quad (2)$$

where $I(t)$ represents the plasma insulin concentration, G_b (Basal Glucose) is the subject-specific fasting glucose setpoint, and $R_a(t)$ (Rate of Appearance) is the influx of glucose from exogenous carbohydrate intake, modelled as a series of decaying exponentials to simulate digestion kinetics.

Novel Physiological Extensions: To bridge the gap between idealized models and pathological reality, we introduced two critical terms:

1) Non-Linear Renal Clearance (E_{renal})

Standard models assume linear clearance. We explicitly model the kidney's threshold-based filtration, which activates only during severe hyperglycemia (glycosuria):

$$E_{renal}(G) = k_{cl} \cdot \max(0, G(t) - \theta_{renal} \quad (3)$$

where $\theta \approx 180$ mg/dL represents the renal threshold, and k_{cl} = 0.05 is the clearance rate. This forces the model to learn distinct decay dynamics for extreme hyperglycemic events.

2) Circadian Dawn Phenomenon (G_{circ})

To simulate the early-morning rise in glucose caused by cortisol and growth hormone surges (independent of food), we added a time-dependent Gaussian forcing function (4):

$$G_{circ}(t) = \alpha_{cort} \cdot \exp\left(-\frac{(t_{min}(mod\ 1440) - t_{dawn})^2}{2\sigma_{circ}^2}\right) \quad (4)$$

centred at t_{dawn}= 06:00 (360 min). This term challenges the prediction model to distinguish between hormonal variations and meal-induced excursions.

Phenotypic Diversity and Pathology: We parameterized our DT to generate four distinct metabolic phenotypes (Table I). Crucially, the T1DM phenotype is modeled with a Beta-Cell Gain of 0.0, mathematically simulating total insulin deficiency and requiring full exogenous dependence.

TABLE I. SIMULATION PARAMETERS FOR VIRTUAL COHORT GENERATION

Parameter	Healthy (N=30)	Pre-diabetes (N=30)	T2DM (N=20)	T1DM (N=20)
Fasting glucose (G_b)	90 ± 5 mg/dL	117 ± 3 mg/dL	$160 \pm$ 15mg/dL	$140 \pm$ 20mg/dL
Insulin sensitivity (p_3)	high 1.5×10^{-5}	reduced 0.8×10^{-5}	resistant 0.3×10^{-5}	normal 1.2×10^{-5}
Beta-cell gain	0.5 (Strong)	0.3 (Stressed)	0.15 (Impaired)	0.0 (Failure)
Renal filtration limit (θ_{renal}) (19)	180 mg/dL	180 mg/dL	180 mg/dL	180 mg/dL

B. Physical Layer: Sensor Physics Engine

The "Physical Layer" translates the deterministic biological state into noisy, realistic sensor signals. Optical Noise Model (Modified Beer-Lambert): The primary input is the non-invasive optical signal $S_\lambda(t)$, simulated at the glucose-sensitive wavelength $\lambda = 940\ nm$.

$$S_\lambda(t) = I_0 \cdot \exp\left(-\mu_\lambda G(t) \cdot \left(1 - \gamma H(t)\right) + \eta_{motion}\left(A(t)\right) + \delta_{temp}(T)\right) \quad (5)$$

where $\eta_{motion}(A(t))$ represents heteroscedastic motion artifacts. Unlike simple additive white Gaussian noise (AWGN), the variance of this noise term scales linearly with the magnitude of the accelerometer signal $A(t)$:

$$\eta_{motion} \sim N(0, 0.05 \cdot A(t)) \quad (6)$$

Bioimpedance Correction Channel: To enable the model to disentangle the hydration confounder ($\gamma H(t)$), we simulated a bioimpedance channel $|Z(t)|$.

C. Behavioral Layer: "Chaos Mode"

To prevent the model from overfitting to perfect physiological rules, we implemented a "Chaos Mode" that injects realistic stochasticity:

- Temporal Meal Jitter: Feeding events were randomized with a temporal jitter of plus or minus 45 minutes (uniform distribution) around standard intervals.
- Nutritional Uncertainty: Carbohydrate content was sampled from a Gaussian distribution, resulting in almost 20% variability.
- Therapeutic Error: The simulation included probabilistic events such as "Missed Bolus" (no insulin given for a meal) and "Rage Bolusing" (stacking insulin doses during hyperglycemia), creating complex, overlapping insulin action curves that challenge the model to learn robust compensatory logic.

III. METHODOLOGY: THE HYBRID TEMPORAL-ATTENTION NETWORK (HTAN)

To effectively map the noisy, multivariate sensor inputs ($X \in \mathbb{R}^{BxLxC}$) to glucose values, we propose the HTAN architecture.

A. Architecture Overview

The pipeline consists of three sequential blocks designed to decouple sensor noise from physiological trends. 1) Residual CNN Block (Denoising & Feature Extraction): A dual-layer 1D-CNN (64 filters, k=3) with residual connections serves as a learnable filter bank. 2) Bi-Directional LSTM (Temporal Dynamics): We use a single-layer Bi-LSTM (128 hidden units) to process the denoised feature maps. 3) Multi-Head Self-Attention (Context Weighting): A central element of the proposed HTAN system is a Self-Attention mechanism placed directly after the LSTM sequence-processing block. In line with similar hybrid architectures in recent studies [8], this module recalibrates temporal features, prioritizing physiologically significant events over less relevant time steps. The architectural topology is detailed in Table II.

TABLE II. ARCHITECTURAL TOPOLOGY

Layer type	Output shape	Parameters	Function
Input	(B, 6, 60)	-	Multivariate time-series (6 Channels)
Residual CNN	(B, 64, 60)	~14k	Denoising & feature extraction
Bi-LSTM	(B, 60, 128)	~131k	Temporal dynamics & Lag capture
Attention	(B, 60, 128)	~66k	Context weighting (4-Head)
Output Head	(B, 1)	~8k	Regression to Glucose value
TOTAL		~253,000	Wearable-ready footprint

Mechanism: Formally, the network computes attention scores (α) to dynamically assign weight to specific temporal indices using the standard scaled dot-product formula (7):

$$Attention(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (7)$$

Configuration: The attention module utilizes 4 parallel heads to capture diverse temporal dependencies.

Feature Dimensionality: The embedding size is set to 128, corresponding to the LSTM output dimension.

Regularization: A dropout rate of 0.2 is applied specifically to the attention weights to prevent co-adaptation.

Functional Purpose: In environments with limited data availability, the attention mechanism enables the model to selectively "attend" to rapid state changes while ignoring stable, non-informative periods. This capability serves as an effective regularization strategy, mitigating the risk of overfitting inherent in small clinical datasets [8].

B. Loss Function Engineering

To improve the network's stability against high-amplitude artifacts in optical sensor data, we optimized the model parameters using the Huber Loss function (8), [16]. This objective is particularly effective for regression problems with outliers, as it combines the precision of Mean Squared Error (MSE) with the resilience of Mean Absolute Error (MAE):e Error (MAE):

$$L_\delta(y,\hat{y}) = \begin{cases} \frac{1}{2}(y-\hat{y})^2 & \text{for } |y-\hat{y}| \le \delta \\ \delta(|y-\hat{y}| - \frac{1}{2}\delta) & \text{otherwise} \end{cases} \quad (8)$$

Parameter Configuration: The transition parameter δ was established at 1.0. This setting enables the model to apply a quadratic penalty to minor deviations (characteristic of standard Gaussian noise) while enforcing a linear penalty for larger discrepancies, thereby mitigating the destabilizing impact of extreme motion artifacts.

IV. EXPERIMENTAL SETUP

To evaluate the HTAN architecture against the "Cold Start" problem, we utilized a dual-stream protocol (Figure 2).

A. Dataset Characterization

We employed four datasets to test generalization across varying data availability (N) and metabolic volatility (CV) (Table III).

1. In-Silico Pre-Training (The Prior):

- Source: Hybrid DT framework.
- Scale: N=100 virtual subjects (30 Healthy, 30 Pre-DM, 20 T2DM, 20 T1DM).
- Volume: ~403,200 data points (14 days @ 5-min intervals).
- Feature: Enriched with "Chaos Mode" artifacts (unannounced meals, disconnects) to pre-immunize the model against noise.

2. OhioT1DM (2020) [23]: 12 T1DM subjects (8 weeks of data). Benchmark for moderate data/high volatility.
3. Shanghai T2DM [24]: 100 T2DM subjects. Stability control (low volatility).
4. Shanghai T1DM [24]: 7 T1DM subjects. "Cold Start" stress test (extreme scarcity + high volatility).

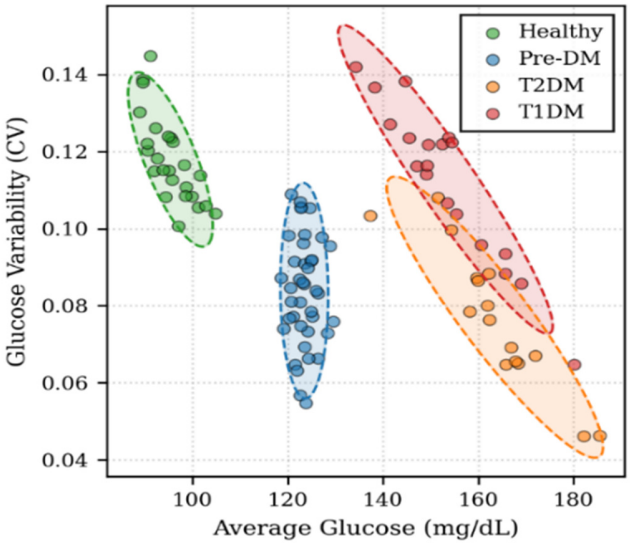

Fig. 2. Cohort distribution plot.

TABLE III. DATASETS DETAILS

Dataset	Subjects	Type	Role
In-Silico	100	Mixed	Pre-training Prior
OhioT1DM	12	T1DM	Baseline benchmark
Shanghai T2DM	100	T2DM	Stability vontrol
Shanghai T1DM	7	T1DM	Cold start stress test

B. Validation Protocol

We enforced a strict Subject-Independent evaluation to prevent data leakage. (Figure 3).

- Group K-Fold (K=5): Data partitioned by Patient_ID. The model is tested only on subjects it has never seen.
- Leakage-Free Preprocessing: Statistics for scaling (Standard Scaler) and hyperparameter tuning were derived strictly from training folds to simulate real-time deployment.

C. Implementation Details

- Framework: PyTorch 2.0 on NVIDIA A100 GPU.

- Optimization: Adam Optimizer (α= 0.001, β_1 =0.9, β_2=0.999), Batch Size 256.
- Regularization: Early Stopping (patience=5 epochs) based on validation loss.
- Loss Function: Huber Loss (δ=1.0) to mitigate exploding gradients from high-amplitude artifacts.

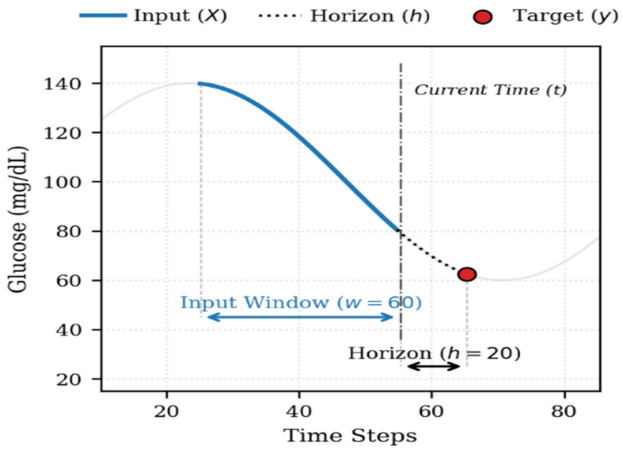

Fig. 3. Data windowing schematic.

V. RESULTS

A. Phase 1: Synthetic Validation (Proof of Concept)

Performance assessment on the independent synthetic test partition (N=20) confirmed the HTAN framework's ability to capture the underlying physiological dynamics. As depicted in Figure 4, the training phase showed a steep convergence trajectory. The validation metric diverged from the training loss after Epoch 20; therefore, the model checkpoint from Epoch 9 was retained to maximize generalization and prevent overfitting. Quantitatively, the network achieved a MARD of 0.39%.

B. Phase 2: Clinical Benchmarking (Dual-Stream Validation)

We benchmarked the pre-trained HTAN against two established baselines: a classical RF and a standard Bi-LSTM. The results are summarized in Table IV. Notably, on the volatile Shanghai T1DM dataset, HTAN reduced the MARD by 40.9% relative to the Bi-LSTM baseline (5.11% vs. 8.65%), highlighting the efficacy of the attention mechanism in chaotic regimes (Figure 5).

C. Deep Dive: The "Cold Start" Solution (Shanghai T1DM)

The critical finding is the performance inversion in the volatile, data-scarce Shanghai T1DM cohort (CV > 0.30, N=7).

- Baseline Failure: The Bi-LSTM struggled significantly (MARD 8.65%), often smoothing out critical drops or reacting with phase lag.
- HTAN Superiority: HTAN achieved the study's best performance (MARD 5.11%), effectively breaching the ~7–9% "physiological noise floor" often cited in optical sensing literature [3] and potentially meeting the accuracy criteria for non-adjunctive therapeutic decisions. It closely tracked ground truth during rapid hypoglycemic drops (Figure 6).
- Clinical Safety Assessment: The Clarke Error Grid evaluation (Figure 7) shows that the vast majority of the forecasted glucose values fall within the clinically safe boundaries, defined as Zones A and B [6]. This reliability persists even during episodes of severe hypoglycemia (< 70 mg/dL). Such results indicate that the system's precision is not merely a consequence of fitting the average trend, but rather stems from its intrinsic capability to detect rapid metabolic fluctuations.

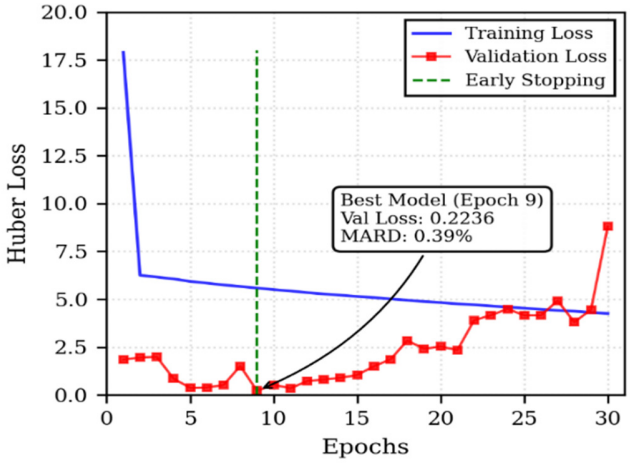

Fig. 4. Training convergence plot.

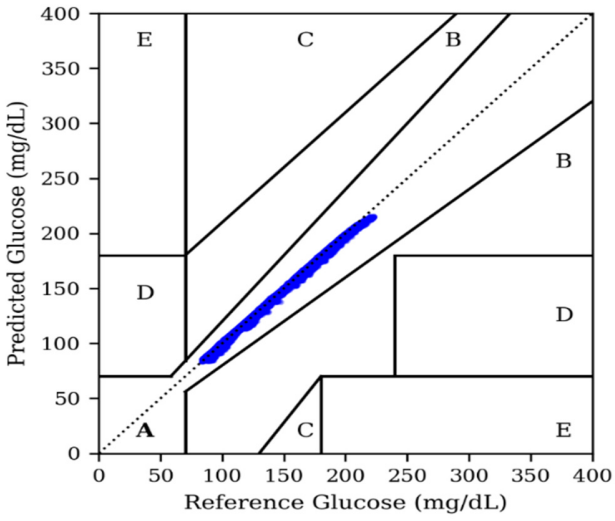

Fig. 5. Clarke error grid for synthetic data.

TABLE IV. CLINICAL BENCHMARKING RESULTS

Dataset	Model	RMSE (mg/dL)	MAE (mg/dL)	MARD (%)	R^2
Shanghai T2DM (Stable)	RF	13.73	9.61	**6.18**	0.918
	Bi-LSTM	20.73	12.01	6.68	0.814
	HTAN (Ours)	23.67	18.42	12.09	0.758
Ohio T1DM (Benchmark)	RF	19.16	13.16	9.00	0.894
	Bi-LSTM	18.34	12.34	8.35	0.903
	HTAN (Ours)	21.41	15.73	10.05	0.867
Shanghai T1DM (Volatile)	RF	16.72	12.45	7.14	0.863
	Bi-LSTM	26.04	16.89	8.65	0.654
	HTAN (Ours)	12.27	8.12	**5.11**	**0.923**

D. The Stability Trade-Off (Shanghai T2DM)

Conversely, results from the stable T2DM cohort suggest an important boundary condition. The classical RF achieved the lowest error (6.18%), outperforming both deep learning models. This may indicate that deep architectures such as HTAN offer limited advantage for smoother, more linear dynamics, and points toward a phenotype-dependent modeling approach, with simpler ensemble methods potentially sufficient for stable T2DM and more expressive attention-based models better suited to high-risk T1DM.

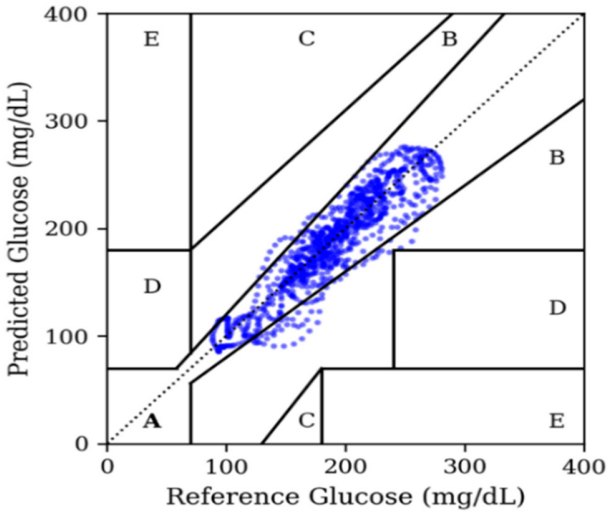

Fig. 6. Clarke Error Grid for Shanghai T1DM..

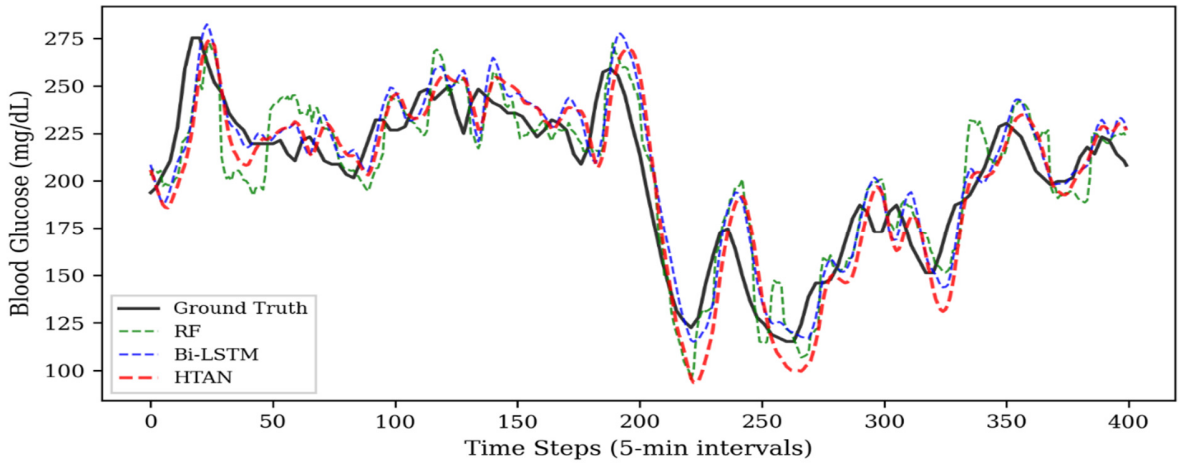

Fig. 7. Time Series Comparison for Shanghai T1DM.

VI. DISCUSSION

The rigorous dual-stream validation of the HTAN architecture challenges the "one-size-fits-all" paradigm in medical AI. Our findings suggest a phenotype-dependent approach to algorithm selection based on physiological volatility.

A. The Volatility Trade-Off

Results reveal a clear trade-off driven by glucose stability. In stable phenotypes (Shanghai T2DM), the classical Random Forest achieved the lowest error, suggesting that in stable regimes, deep learning models are prone to overfitting. Conversely, for volatile phenotypes (Shanghai T1DM), HTAN significantly outperformed RF. The Multi-Head Attention mechanism proved essential for capturing nonlinear dynamics (Figure 7).

B. Comparative Performance Analysis and Novel Aspects

The primary innovation of this research is to couple a "Chaos-Augmented" DT simulation with a hybrid neural architecture driven by attention protocols. In the contemporary literature, the leading benchmarks on the Shanghai dataset are established by the Temporal Fusion Transformer [25] and the Graph Attentive RNNs [26]. While these Transformer-based approaches deliver competitive RMSE outcomes (for instance, authors in [25] report 9.18 mg/dL), the HTAN model prioritizes clinical reliability (MARD). It achieves a markedly

lower error rate of 5.11% under conditions characterized by severe physiological fluctuations. These findings corroborate the hypothesis that initializing the network with chaotic attractors enables it to learn the nonlinear boundaries of metabolic homeostasis, offering a superior form of regularization compared with standard Gaussian-noise methods [27]. Furthermore, our data refines the broader conclusions drawn by [27], demonstrating that deep learning frameworks offer unique advantages tailored to the volatile dynamics of Type 1 Diabetes care. Recent IoT-driven ensemble learning approaches, such as that of [28], show enhanced robustness in diabetes prediction tasks.

C. Solving the "Cold Start" Problem

By incorporating stochastic artifacts—such as skipped meals—into the DT, the simulation effectively served as an adversarial training method. This process served as a robust regularizer, enabling the deep learning architecture to achieve strong generalization performance even when trained on sparse real-world datasets.

VII. CONCLUSION

This research establishes a robust computational infrastructure designed specifically for Non-Invasive Blood Glucose Monitoring (NIBGM). We proposed and rigorously tested the Hybrid Temporal-Attention Network (HTAN) via a novel Dual-Stream Validation methodology. The empirical data strongly suggest that algorithmic selection strategies must be tailored to specific metabolic phenotypes. Although traditional ensemble techniques are adequate for the stable glucose profiles typical of Type 2 Diabetes, the erratic, chaotic fluctuations inherent in Type 1 Diabetes require the sophisticated regularization capabilities offered by attention-based architectures. By successfully using Chaos-Augmented Digital Twins to address the "Cold Start" data scarcity challenge, this study underscores the immense potential of AI-integrated systems in advancing next-generation personalized medical technology.

REFERENCES

[1] International Diabetes Federation. "IDF Diabetes Atlas, 10th edition." IDF Diabetes Atlas. https://diabetesatlas.org

[2] S. K. Vashist, "Non-invasive glucose monitoring technology in diabetes management: a review," *Analytica Chimica Acta*, vol. 750, pp. 16–27, Oct. 2012, https://doi.org/10.1016/j.aca.2012.03.043.

[3] N. Uluç *et al.*, "Non-invasive measurements of blood glucose levels by time-gating mid-infrared optoacoustic signals," *Nature Metabolism*, vol. 6, no. 4, pp. 678–686, Apr. 2024, https://doi.org/10.1038/s42255-024-01016-9.

[4] S. Ghimire, T. Celik, M. Gerdes, and C. W. Omlin, "Deep learning for blood glucose level prediction: How well do models generalize across different data sets?," *PLOS ONE*, vol. 19, no. 9, 2024, Art. no. e0310801, https://doi.org/10.1371/journal.pone.0310801.

[5] M. R. Vahedi *et al.*, "Predicting Glucose Levels in Patients with Type1 Diabetes Based on Physiological and Activity Data," in *Proceedings of the 8th ACM MobiHoc 2018 Workshop on Pervasive Wireless Healthcare Workshop*, Mar. 2018, pp. 1–5, https://doi.org/10.1145/3220127.3220133.

[6] Q. Sun, M. V. Jankovic, L. Bally, and S. G. Mougiakakou, "Predicting Blood Glucose with an LSTM and Bi-LSTM Based Deep Neural Network," in *2018 14th Symposium on Neural Networks and Applications (NEUREL)*, Aug. 2018, pp. 1–5, https://doi.org/10.1109/NEUREL.2018.8586990.

[7] H. Butt, I. Khosa, and M. A. Iftikhar, "Feature Transformation for Efficient Blood Glucose Prediction in Type 1 Diabetes Mellitus Patients," *Diagnostics*, vol. 13, no. 3, Jan. 2023, Art. no. 340, https://doi.org/10.3390/diagnostics13030340.

[8] Q. Bian, A. As'arry, X. Cong, K. A. bin M. Rezali, and R. M. K. bin R. Ahmad, "A hybrid Transformer-LSTM model apply to glucose prediction," *PLOS ONE*, vol. 19, no. 9, 2024, Art. no. e0310084, https://doi.org/10.1371/journal.pone.0310084.

[9] M. P. Barbato, G. Rigamonti, D. Marelli, and P. Napoletano, "Lightweight Sequential Transformers for Blood Glucose Level Prediction in Type-1 Diabetes," *IEEE journal of biomedical and health informatics*, vol. PP, Nov. 2025, https://doi.org/10.1109/JBHI.2025.3633194.

[10] T. Zhu, T. Chen, L. Kuang, J. Zeng, K. Li, and P. Georgiou, "Edge-Based Temporal Fusion Transformer for Multi-Horizon Blood Glucose Prediction," in *2023 IEEE International Symposium on Circuits and Systems (ISCAS)*, Feb. 2023, pp. 1–5, https://doi.org/10.1109/ISCAS46773.2023.10181448.

[11] A. Hina and W. Saadeh, "Noninvasive Blood Glucose Monitoring Systems Using Near-Infrared Technology—A Review," *Sensors*, vol. 22, no. 13, June 2022, Art. no. 4855, https://doi.org/10.3390/s22134855.

[12] H. M. C. Leung, G. P. Forlenza, T. O. Prioleau, and X. Zhou, "Noninvasive Glucose Sensing In Vivo," *Sensors*, vol. 23, no. 16, Aug. 2023, Art. no. 7057, https://doi.org/10.3390/s23167057.

[13] M. Zeynali, K. Alipour, B. Tarvirdizadeh, and M. Ghamari, "Non-invasive blood glucose monitoring using PPG signals with various deep learning models and implementation using TinyML," *Scientific Reports*, vol. 15, no. 1, Jan. 2025, Art. no. 581, https://doi.org/10.1038/s41598-024-84265-8.

[14] R. N. Bergman, L. S. Phillips, and C. Cobelli, "Physiologic evaluation of factors controlling glucose tolerance in man: measurement of insulin sensitivity and beta-cell glucose sensitivity from the response to intravenous glucose.," *Journal of Clinical Investigation*, vol. 68, no. 6, pp. 1456–1467, Dec. 1981, https://doi.org/10.1172/JCI110398.

[15] Y. Zhang *et al.*, "Motion Artifact Reduction for Wrist-Worn Photoplethysmograph Sensors Based on Different Wavelengths," *Sensors*, vol. 19, no. 3, Feb. 2019, Art. no. 673, https://doi.org/10.3390/s19030673.

[16] L. Alzubaidi *et al.*, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *Journal of Big Data*, vol. 8, no. 1, Mar. 2021, Art. no. 53, https://doi.org/10.1186/s40537-021-00444-8.

[17] T. Zhu, K. Li, P. Herrero, and P. Georgiou, "Personalized Blood Glucose Prediction for Type 1 Diabetes Using Evidential Deep Learning and Meta-Learning," *IEEE transactions on bio-medical engineering*, vol. 70, no. 1, pp. 193–204, Jan. 2023, https://doi.org/10.1109/TBME.2022.3187703.

[18] B. P. Kovatchev, M. Breton, C. D. Man, and C. Cobelli, "In silico preclinical trials: a proof of concept in closed-loop control of type 1 diabetes," *Journal of Diabetes Science and Technology*, vol. 3, no. 1, pp. 44–55, Jan. 2009, https://doi.org/10.1177/193229680900300106.

[19] C. Dalla Man, R. A. Rizza, and C. Cobelli, "Meal simulation model of the glucose-insulin system," *IEEE transactions on bio-medical engineering*, vol. 54, no. 10, pp. 1740–1749, Oct. 2007, https://doi.org/10.1109/TBME.2007.893506.

[20] S. Ahmad, C. M. Ramkissoon, A. Beneyto, I. Conget, M. Giménez, and J. Vehi, "Generation of Virtual Patient Populations That Represent Real Type 1 Diabetes Cohorts," *Mathematics*, vol. 9, no. 11, May 2021, Art. no. 1200, https://doi.org/10.3390/math9111200.

[21] W. Liu, T. Chen, B. Liang, Y. Wang, and H. Jin, "In-silico evaluation of an artificial pancreas achieving automatic glycemic control in patients with type 1 diabetes," *Frontiers in Endocrinology*, vol. 14, 2023, Art. no. 1115436, https://doi.org/10.3389/fendo.2023.1115436.

[22] J. M. Lee, R. Pop-Busui, J. M. Lee, J. Fleischer, and J. Wiens, "Shortcomings in the Evaluation of Blood Glucose Forecasting," *IEEE Transactions on Biomedical Engineering*, vol. 71, no. 12, pp. 3424–3431, Sept. 2024, https://doi.org/10.1109/TBME.2024.3424665.

[23] J. P. Cohen, M. Luck, and S. Honari, "Distribution Matching Losses Can Hallucinate Features in Medical Image Translation," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, 2018, pp. 529–536, https://doi.org/10.1007/978-3-030-00928-1_60.

[24] C. Marling and R. Bunescu, "The OhioT1DM Dataset for Blood Glucose Level Prediction: Update 2020," *CEUR workshop proceedings.*, vol. 2675, pp. 71–74, Sept. 2020.

[25] Q. Zhao *et al.*, "Chinese diabetes datasets for data-driven machine learning," *Scientific Data*, vol. 10, no. 1, Jan. 2023, Art. no. 35, https://doi.org/10.1038/s41597-023-01940-7.

[26] C. Piao *et al.*, "GARNN: An interpretable graph attentive recurrent neural network for predicting blood glucose levels via multivariate time series," *Neural Networks*, vol. 185, May 2025, Art. no. 107229, https://doi.org/10.1016/j.neunet.2025.107229.

[27] M. Sirlanci, M. E. Levine, C. C. Low Wang, D. J. Albers, and A. M. Stuart, "A simple modeling framework for prediction in the human glucose-insulin system," *Chaos*, vol. 33, no. 7, July 2023, Art. no. 073150, https://doi.org/10.1063/5.0146808.

[28] R. Rastogi, M. Bansal, N. Kumar, S. Singla, P. Singla, and R. A. Jaswal, "Effective Diabetes Prediction using an IoT-based Integrated Ensemble Machine Learning Framework," *Engineering, Technology & Applied Science Research*, vol. 15, no. 1, pp. 20064–20070, Feb. 2025, https://doi.org/10.48084/etasr.8869.