GLUCONV: CONVINCING CONVOLUTIONS MEET TRUSTWORTHY GLUCOSE MONITOR FOR CONTINUOUS DIABETES MANAGEMENT

Sijie Xiong[1] *Tao Sun*[1] *Haiqiao Liu*[1] *Cheng Tang*[1] *Yinlong Hu*[2] *Atsushi Shimada*[1]

[1]Faculty of Information Science and Electrical Engineering, Kyushu University, Fukuoka, Japan
[2]College of Artificial Intelligence and Automation, Hohai University, Changzhou, China
Email: sijiexiongkyushu@gmail.com, xiong.sijie.630@s.kyushu-u.ac.jp, tang@ait.kyushu-u.ac.jp

ABSTRACT

Diabetes is considered as one of the most prevalent and serious chronic diseases worldwide, posing a significant threat to global public health. While current wearable continuous glucose monitors (CGMs) adopt machine learning (ML) kernels because of their responsiveness and high-quality of accuracy, ML approaches cannot warrant trustworthiness in daily supervision, which is a mandatory requirement in medical monitoring. However, deep learning (DL) algorithms with reliability credits fall short in responsiveness and accuracy. To overcome these challenges, we propose GluConv, an efficient glucose forecasting and monitoring framework that fully builds upon convolutions. Both inter-variate correlation and temporal dependency are captured through multi-scale convolutional layers. Our extensive experiments demonstrate that GluConv remarkably outperforms the excellent ML/DL baseline models on trustworthiness and responsiveness, while maintaining acceptable accuracy. This makes GluConv **Conv**incing (accurate and trustworthy), **Conv**olutional (responsive), and **Conv**enient (robust). Codes will be available on **GitHub**.

***Index Terms*—** Diabetes, CGMs, Glucose Monitoring, Responsiveness, Trustworthiness

1. INTRODUCTION

Diabetes is recognized as one of the most widespread chronic diseases globally, with billions of individuals affected daily. Accurate and reliable predictions of glucose dynamics can provide substantial benefits in preventing complications [1, 2]. In addition, continuous glucose monitors (CGMs) require strong responsiveness to offer patients with real-time monitoring and notify them to adjust their lifestyle in time [3]. However, glucose variation behaviors are inherently complex and challenging to forecast [4, 5]. As a result, maintaining accuracy, reliability, and responsiveness concurrently poses an inherent trade-off challenge.

Driven by the accuracy and rapid inference capabilities, machine learning (ML) models have been extensively integrated into CGMs, offering practical relief for patients [6]. However, MLs frequently ignore trustworthiness, which generates beautiful lies and misleads users [4]. Moreover, most ML models operate with fixed parameters and demand substantial time for evolution, overlooking responsiveness.

To overcome these two drawbacks of MLs, deep learning methods (DLs) have been introduced as efficient alternatives [7]. The combined extraction of temporal dependencies and inter-variable correlations in DLs yields features that are more balanced and robustly cross-validated. However, current CNN (Convolutional Neural Network)-based and RNN (Recurrent Neural Network)-based architectures exhibit limitations in concurrently modeling global and local information, leading to decrease in accuracy [3, 8]. Moreover, Transformer-based models with global attention are disturbed by redundant tokens, which results in weak adaption capability in responsiveness [4]. On the other hand, UNet structures have demonstrated great performance in accuracy and trustworthiness in medical image segmentation [9]. This observation inspires the application of such a principle to glucose monitoring.

Building on the computational efficiency of convolutions and the reliability enhancements of UNet, we propose GluConv, a novel architecture designed to balance these three priorities. Our key contributions are summarized as follows:

- **Efficient UNets:** We propose two efficient UNet algorithms based on convolutions for temporal dependencies and variate correlations: Time Channel Block and Variate Channel Block. The architecture optimizes the fusion of both global and local features at multi-scales.
- **GluConv & Performance Superiority** Based on previous blocks, we propose GluConv and extensive experiments demonstrate that GluConv has superiority over benchmarks. GluConv is deemed as a **Conv**incing, **Conv**olutional, and **Conv**enient model in terms of accuracy, trustworthiness, responsiveness, robustness.

2. METHODOLOGY

This work was supported by JST CREST Grant Number JPMJCR22D1 and JSPS KAKENHI Grant Number JP22H00551 and JP25K21340, Japan.

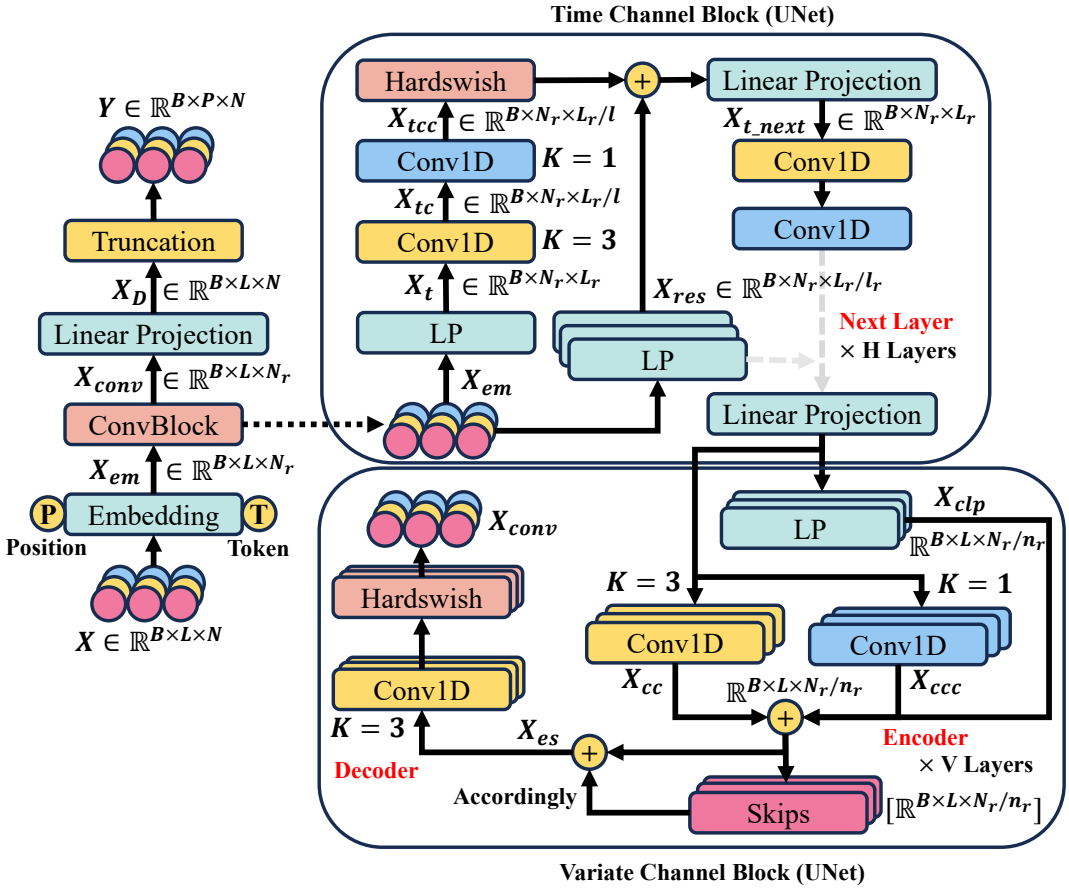

Fig. 1. The architecture of GluConv. ConvBlock contains two UNets: Time Channel Block and Variate Channel Block. Time Channel Block has H layers, and Variate Channel Block has V layers. l and n_r stand for layer coefficient 2^i. K stands for convolutional kernel size. LP represents the combination of a linear projection and a permutation. Conv1D denotes the 1-D convolution. Skips stand for a list of processed tensors in reverse direction. B is the batch size. L is the historical window length while P is the forecasting window length. N_r is the variate-embedded dimension. N is the number of variates. L_r is the temporal dimension. All activation functions are set as Hardswish. Details of the two UNets can refer to **Algorithm 1 and 2**.

Table 1. The accuracy and uncertainty performance comparison across all models. The performance enhancements are marked in **bold**. Underline marks deterioration. − represents there is no such measurement on models.

Method	Broll		Colas		Dubosson		Hall		Weinstock	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
ARIMA	11.75 (-13.54%)	9.71 (-19.46%)	5.91 (**+2.37%**)	4.87 (**+4.93%**)	18.75 (-6.30%)	14.58 (-16.05%)	8.22 (**+9.02%**)	6.97 (**+7.89%**)	15.87 (**+2.03%**)	13.34 (**+2.40%**)
Linear	11.98 (-11.40%)	9.83 (-18.01%)	5.33 (-8.31%)	4.41 (-4.99%)	15.69 (-27.02%)	11.90 (-42.18%)	7.86 (**+4.87%**)	6.62 (**+3.02%**)	15.58 (**+0.16%**)	13.09 (**+0.53%**)
XGBoost	9.76 (-36.65%)	8.72 (-33.03%)	6.18 (**+6.58%**)	5.32 (**+12.97%**)	17.57 (-13.41%)	15.42 (-9.73%)	7.49 (**+0.21%**)	6.52 (**+1.53%**)	15.36 (-1.21%)	13.04 (**+0.15%**)
Latent ODE	14.96 (**+10.84%**)	13.05 (**+11.11%**)	5.64 (-2.28%)	4.84 (**+4.34%**)	17.38 (-14.69%)	15.12 (-11.90%)	7.71 (**+2.97%**)	6.61 (**+2.87%**)	15.06 (-3.28%)	12.72 (-2.36%)
Transformer	14.04 (**+4.95%**)	12.28 (**+5.54%**)	5.97 (**+3.31%**)	5.24 (**+11.64%**)	15.71 (-26.90%)	12.98 (-30.35%)	8.18 (**+8.58%**)	7.07 (**+9.19%**)	14.15 (-9.90%)	11.91 (-9.32%)
TFT	12.43 (-7.35%)	10.23 (-13.39%)	5.51 (-4.71%)	4.47 (-3.58%)	17.50 (-13.90%)	14.53 (-16.45%)	8.12 (**+7.93%**)	6.76 (**+5.03%**)	15.25 (-1.95%)	12.50 (-4.16%)
Gluformer	16.70 (**+20.09%**)	14.82 (**+21.73%**)	6.94 (**+16.67%**)	6.03 (**+23.22%**)	23.48 (**+15.10%**)	20.70 (**+18.26%**)	8.17 (**+8.47%**)	7.04 (**+8.81%**)	15.94 (**+2.44%**)	13.65 (**+4.62%**)
GluTANN	15.88 (**+15.98%**)	14.13 (**+17.91%**)	7.13 (**+19.03%**)	6.23 (**+25.68%**)	21.85 (**+8.78%**)	19.40 (**+12.78%**)	8.27 (**+9.58%**)	7.11 (**+9.70%**)	15.69 (**+0.89%**)	13.30 (**+2.11%**)
GluConv	13.34 (-)	11.60 (-)	5.77 (-)	4.63 (-)	19.93 (-)	16.92 (-)	7.48 (-)	6.42 (-)	15.55 (-)	13.02 (-)
Deterioration	**32**	**Enhancement**	**48**							
Uncertainty	Likelihood / Calibration		Likelihood / Calibration		Likelihood / Calibration		Likelihood / Calibration		Likelihood / Calibration	
ARIMA	- / -		- / -		- / -		- / -		- / -	
Linear	-9.95 / -0.15 (**+78.49%** / **+184.25%**)		-9.16 / 0.15 (**+81.55%** / **16.41%**)		-10.11 / 0.17 (**+103.46%** / **+27.27%**)		-9.53 / 0.10 (**+83.84%** / **+41.62%**)		-10.22 / 0.11 (**+75.93%** / **+12.18%**)	
XGBoost	-10.03 / 0.07 (**+78.66%** / -80.53%)		-9.36 / 0.09 (**+81.94%** / -39.32%)		-10.22 / 0.07 (**+103.42%** / -76.63%)		-9.56 / 0.08 (**+83.89%** / **+27.03%**)		-10.28 / 0.11 (**+76.07%** / **+12.18%**)	
Latent ODE	-28.75 / 0.38 (**+92.56%** / **+66.74%**)		-8.80 / 0.24 (**+80.80%** / **+47.75%**)		-30.19 / 0.44 (**+101.16%** / **+71.90%**)		-18.19 / 0.44 (**+91.53%** / **+86.73%**)		-30.08 / 0.40 (**+91.82%** / **+75.85%**)	
Transformer	-9.98 / 0.19 (**+78.56%** / **+33.49%**)		-9.30 / 0.09 (**+81.83%** / -39.32%)		-10.22 / 0.07 (**+103.42%** / -76.63%)		-9.56 / 0.12 (**+83.89%** / **+51.35%**)		-10.12 / 0.11 (**+75.69%** / **+12.18%**)	
TFT	- / 0.15 (- / **+15.75%**)		- / 0.09 (- / -39.32%)		- / 0.26 (- / **+52.45%**)		- / 0.08 (- / **+27.03%**)		- / 0.08 (- / -20.76%)	
Gluformer	-1.96 / 0.11 (-9.18% / **-14.88%**)		-1.61 / 0.10 (-4.97% / -25.39%)		-1.17 / 0.12 (**+129.91%** / -3.03%)		-1.44 / 0.06 (-6.94% / **+2.71%**)		-2.41 / 0.09 (-2.07% / -7.34%)	
GluTANN	-1.97 / 0.14 (-8.63% / **+9.74%**)		-1.77 / 0.08 (**+4.52%** / -56.74%)		-2.02 / 0.12 (**+117.33%** / -3.03%)		-1.63 / 0.05 (**+5.52%** / -16.75%)		-2.43 / 0.09 (-1.23% / -7.34%)	
GluConv	-2.14 / 0.13 (- / -)		-1.69 / 0.13 (- / -)		+0.35 / 0.12 (- / -)		-1.54 / 0.06 (- / -)		-2.46 / 0.10 (- / -)	
Deterioration	**21**	**Enhancement**	**44**	**Plateau**	**0**	**–**	**15**			

Algorithm 1 TIME CHANNEL BLOCK

Input: $X_{em} \in \mathbb{R}^{B \times N_r \times L_r}$
Output: $\in \mathbb{R}^{B \times L \times N_r}$
$X_t \in \mathbb{R}^{B \times N_r \times L_r} = \text{Linear}(\text{Permute}(X_{em}))$
for i = 0 to H **do**
  $X_{res} = \text{Permute}(\text{Linear}(X_{em}))$ with $l = 2^i$
  $X_{tcc} = \text{Conv1D}(\text{Conv1D}(X_t))$ with $l = 2^i$
  $temp = \text{Hardswish}(X_{tcc}) + X_{res}$ with $l = 2^i$
  $X_t = X_{t\_next} = \text{Linear}(temp) \in \mathbb{R}^{B \times N_r \times L_r}$ with **Up-sampling**
Return $Output = \text{Linear}(X_t) \in \mathbb{R}^{B \times L \times N_r}$

Algorithm 2 VAIRATE CHANNEL BLOCK

Input: Output from Time Channel Block $\in \mathbb{R}^{B \times L \times N_r}$
Output: $X_{conv} \in \mathbb{R}^{B \times L \times N_r}$
$Res = \text{Clone}(Input)$; $Skips = [Res]$; $x = Input$
for i = 0 to V **do**
  $X_{clp} = \text{Permute}(\text{Linear}(x))$ with $n_r = 2^i$
  $X_{cc} = \text{Conv1D}(\text{Permute}(x))$ with $n_r = 2^i$
  $X_{ccc} = \text{Conv1D}(\text{Permute}(x))$ with $n_r = 2^i$
  $x = X_{cc} + X_{ccc} + X_{clp}$; $Skips = \text{append}(x)$
  $x = \text{Permute}(x)$
$Skips = \text{Reverse}(Skips)$
for i = 0 to V **do** with **Up-sampling**
  $x = \text{Permute}(\text{Conv1D}(X_{es} = x + Skips[i]))$
Return $X_{conv} = x \in \mathbb{R}^{B \times L \times N_r}$

Overall Architecture of GluConv The architecture of GluConv is presented in Figure 1. Blood glucose data X are firstly processed with positional and token embeddings. Then, we design and employ ConvBlock to extract fine-grained features X_{conv} from X_{em}. The details of ConvBlock are decomposed into two algorithms in subsequent sections: Time Channel Block and Variate Channel Block. X_{conv} recovers to the original variate number N from N_r by a linear projection and turns to be X_D. Finally, we truncate and adopt the last P time steps to generalize the prediction Y.

Time Channel Block We mimic the UNet structure [9] and DWConv [10] to construct this block. We employ the Conv1D with $K = 3$ to enhance the receptive fields of X_t and the one with $K = 1$ to capture the local features X_{tcc}. Moreover, we leverage a stack of LP blocks to provide the residual information X_{res} according to different layers. Finally, a linear projection acts as the decoder to recover dimensions X_{t_next} for the next layer until the completion of H layers. The pseudo-algorithm is presented in **Algorithm 1**.

Variate Channel Block This block further processes the features transmitted from Time Channel Block. Similarly, we take Conv1Ds as the multi-scale Encoder with LP providing residual information. We store the sum of X_{cc} and X_{ccc} in a list Skips in a reverse chronological order. Then, we employ another multi-scale Conv1Ds as the Decoder to process the sum of the element of Skips and $X_{cc} + X_{ccc}$ accordingly for different layers. After Hardswish activations, we output X_{conv} to the main flow of GluConv. The pseudo-algorithm is included in **Algorithm 2**.

3. EXPERIMENT SETTING

Dataset Pre-processing Scheme We apply GluConv to 5 well-recognized diabetes datasets collected by CGMs: Broll [11], Colas [12], Dubosson [13], Hall [14], and Weinstock [15]. To protect clinic relevance, simple linear interpolation fulfills missing values and all values are restricted within 20mg/dL to 400mg/dL. In addition, there is no extreme fluctuation beyond 40mg/dL per 15-minute interval. Dubosson and Weinstock contain recordings from patients with Type-I diabetes only. Broll collects data from Type-II patients. Colas and Hall are hybrid ones with both Type-I and Type-II data. At subject levels, Broll involves 5 patients, while Colas and Weinstock samples over 190 patients. The different characteristics of datasets are deemed to replicate most real-world situations, which guarantees rationales and effectiveness of our experiments.

Baseline Models To comprehensively evaluate the position of GluConv in MLs and DLs, 8 excellent baseline models are employed. The **ML models**: ARIMA [16], Linear Regression (Linear) [17], and XGBoost [17]. The **DL models**: Transformer [18], Temporal Fusion Transformer (TFT) [17], Latent ODE [8], Gluformer [19], GluTANN [4]. Notably, Gluformer (proposed in ICASSP and employed in ICLR) is the most advanced DL model. It is also the main competitor to GluConv. The diversity of benchmark models can ensure a fair and balanced evaluation of GluConv on accuracy, trustworthiness, and responsiveness.

Main Task Traditional experiments follow the assumption that glucose data and their predictions obey the Gaussian distribution. However, CGMs are easily influenced by sampling environments, including physical temperature, biological variations, electromagnetic interference, etc. To better mimic the real world, we add irregular random noises during the learning process. If the dataset distribution is asymmetric, we choose Laplace distributions as the noise base distribution [20], while we select Gaussian distributions [21] in symmetric cases. The Bernoulli dropout function is used to introduce irregularity. All experiments are fine-tuned by Optuna [22] and all hyper-parameters can be found on Github.

Evaluation Metrics To measure **ACCURACY**, two metrics are employed herein: the root mean square error (**RMSE**) and the mean absolute error (**MAE**). Since all samples display a right-skew distribution [23], we abandon the mean but the median as the center when computing the RMSE and MAE. To measure **TRUSTWORTHINESS**, we take the log-likelihood [3] and the calibration [7] as the metrics. For models that estimate the distribution over future values, $\hat{P}^{(i)}_{j,k+L+1:k+L+T}$:

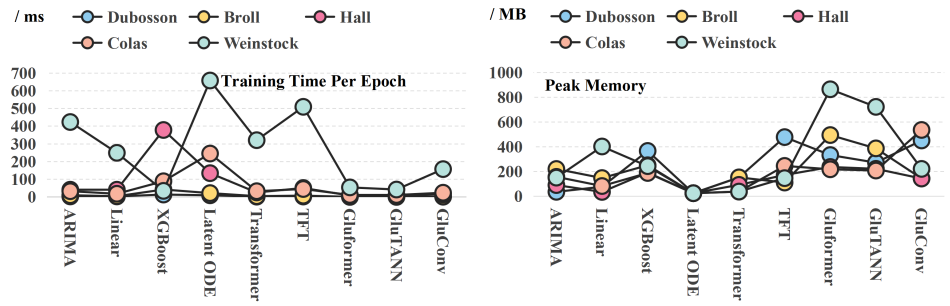

Fig. 2. The TTE and PMA indicators of each model. The existence of Skips increases memory occupation.

$\mathbb{R}^T \to [0, 1]$, the log-likelihood is defined as Eq. (1) [3].

$$\log L_{i,j,k} = \log \hat{P}^{(i)}_{j,k+L+1:k+L+T} \left(y^{(i)}_{j,k+L+1:k+L+T} \right), \quad (1)$$

where $\hat{P}$ is the parametric predictive distribution and $\log L$ is the logarithmic likelihood. Higher values indicate better trustworthiness of the predictive values. We cut up $M = 10$ quintiles to differentiate confidences over different distributional regions. For M confidence intervals: $0 < p_1 < \cdots < p_M < 1$, we estimate the true confidence level $\hat{p}_m$ using R test input-output pairs by Eq. (2).

$$\hat{p}_m = \frac{\left| \left\{ y^{(i)}_{j,k+L+t} \,\middle|\, \hat{F}^{(i)}_{j,k+L+t} \left(y^{(i)}_{j,k+L+t} \right) \leq p_m \right\} \right|}{R}, \quad (2)$$

where $\hat{F}$ denotes the cumulative distribution function (CDF) of predictions. Then, the calibration Cal_t at time step t across all M levels is computed as Eq. (3).

$$Cal_t = \sum_{m=1}^{M} (p_m - \hat{p}_m)^2. \quad (3)$$

Finally, we measure the **RESPONSIVENESS** by training time per epoch (**TTE**) and the peak memory allocation (**PMA**) delivered by GPU.

4. PERFORMANCE EVALUATION

Accuracy The upper region of Table 1 reveals the accuracy performance. GluConv achieves 48 enhancements out of 80 cases. Notably, GluConv beats all benchmarks in Hall, which demonstrates that GluConv is excellent at handling hybrid diabetes scenarios. In addition, GluConv surpasses two advanced models in all datasets: Gluformer (ICLR 2024) and GluTANN (CAI 2025). The increases in most cases are within $10\% - 20\%$. The encoder-decoder structure and the skip connection contribute to such enhancements, while no attention mechanism is employed, helping GluConv avoid redundant tokens. Moreover, the multi-scale structure empowers GluConv with widened receptive fields and the ability to integrate original information with well-extracted features. Admittedly, GluConv deteriorates in the other cases, especially in Dubosson. The simplicity and few subjects reduce the inter-correlations, leading to decreases in accuracy. Given Increase (48) v.s. Decrease (32), we consider that benefits outweigh costs, and GluConv has an acceptable accuracy credits.

Trustworthiness The lower part of Table 1 demonstrates a spotlight performance of GluConv in trustworthiness. It achieves 44 enhancements out of 65 situations. Numerous increases go beyond 75%. ML models fall short in log-likelihood and calibrations. This means that the predictions generated from MLs are less likely to occur and deviate severely from the reality. However, the strong performance of GluConv indicates that the potential of DLs to produce more probable and reliable forecasting and early signals for CGMs and patients. Given the introduction of irregular noise, GluConv also confirms that it has greater robustness in real-world situations and is a better candidate for CGMs.

Responsiveness As presented in Figure 2, GluConv remains an excellent speed and competitive to Gluformer and GluTANN. The simplicity of convolutions contribute greatly to the responsiveness. On the other hand, indicated by the PMA, the moderate peak memory indicates that GluConv is a practical model for CGMs, and the encoder-decoder structure expands the memory occupations. Although it underperforms MLs, compared to other DLs, GluConv displays more balanced. In summary, we consider that GluConv has an excellent responsiveness performance and memory occupation.

5. CONCLUSION

In this work, we propose a convolution-based encoder-decoder model, GluConv, for glucose monitoring. The design is motivated by the need for enhancements in accuracy and responsiveness of DLs, and trustworthiness of MLs. Experimental results demonstrate that the multi-scale convolutions provide higher receptive fields and reduce redundant tokens, leading to increases in accuracy. The encoder-decoder structure and skip connections strengthen trustworthiness, with responsiveness ensured by simple convolutions. Overall, GluConv is a convolutional, convincing, and convenient model for glucose monitors and medical assistance diagnoses. For future, the UNets can be optimized to reduce peak memory.

6. REFERENCES

[1] Natalie Nanayakkara, Andrea J. Curtis, Stephane Heritier, Adelle M. Gadowski, Meda E. Pavkov, Timothy Kenealy, David R. Owens, Rebecca L. Thomas, Soon Song, Jencia Wong, Juliana C.N. Chan, Andrea O.Y. Luk, Giuseppe Penno, Linong Ji, Viswanathan Mohan, Anandakumar Amutha, Pedro Romero-Aroca, Danijela Gasevic, Dianna J. Magliano, Helena J. Teede, John Chalmers, and Sophia Zoungas, "Impact of age at type 2 diabetes mellitus diagnosis on mortality and vascular complications: Systematic review and meta-analyses," *Diabetologia*, vol. 64, pp. 275–287, Feb. 2021.

[2] Silvia Oviedo, Josep Vehí, Remei Calm, and Joaquim Armengol, "A review of personalized blood glucose prediction strategies for t1dm patients," *International Journal for Numerical Methods in Biomedical Engineering*, vol. 33, no. 6, pp. e2833, Jul. 2017.

[3] Renat Sergazinov, Elizabeth Chun, Valeriya Rogovchenko, Nathaniel Fernandes, Nicholas Kasman, and Irina Gaynanova, "GlucoBench: Curated list of continuous glucose monitoring datasets with prediction benchmarks," in *International Conference on Learning Representations*, Jan. 2024.

[4] Sijie Xiong, Youhao Xu, Cheng Tang, Jianing Wang, Shuqing Liu, and Atsushi Shimada, "GluTANN: Transformer-based continuous glucose monitoring model with ANN attention," in *International Conference on Artificial Intelligence*, May. 2025.

[5] Arnold C. T. Ng, Victoria Delgado, Barry A. Borlaug, Jeroen J. Bax, and Amit Sharma, "Diabesity: The combined burden of obesity and diabetes on heart disease and the role of imaging," *Nature Reviews Cardiology*, vol. 18, pp. 291–304, Apr. 2021.

[6] Hun-Sung Kim and Kun-Ho Yoon, "Lessons from use of continuous glucose monitoring systems in digital healthcare," *Endocrinology and Metabolism*, vol. 35, no. 3, pp. 541–548, Jul. 2020.

[7] Volodymyr Kuleshov, Nathan Fenner, and Stefano Ermon, "Accurate uncertainties for deep learning using calibrated regression," in *Proceedings of the International Conference on Machine Learning*, Jul. 2018, vol. 80, pp. 2796–2804.

[8] Khaled Alkilane, Yihang He, and Der-Horng Lee, "MixMamba: Time series modeling with adaptive expertise," *Information Fusion*, vol. 112, pp. 102589, Jul. 2024.

[9] Fabian Isensee, Paul F. Jaeger, Simon A. A. Kohl, Jens Petersen, and Klaus H. Maier-Hein, "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, pp. 203–211, Feb. 2021.

[10] Pengfei Zhang, Eric Lo, and Baotong Lu, "High performance depthwise and pointwise convolutions on mobile devices," in *Proceedings of the AAAI Conference on Artificial Intelligence*, Apr. 2020, vol. 32, pp. 6795–6802.

[11] Steven Broll, Jacek Urbanek, David Buchanan, Elizabeth Chun, John Muschelli, Naresh M Punjabi, and Irina Gaynanova, "Interpreting blood GLUcose data with R package iglu," *PloS One*, vol. 16, no. 4, pp. e0248560, Apr. 2021.

[12] Ana Colás, Luis Vigil, Borja Vargas, David Cuesta-Frau, and Manuel Varela, "Detrended fluctuation analysis in the prediction of type 2 diabetes mellitus in patients at risk: Model optimization and comparison with other metrics," *PloS One*, vol. 14, no. 12, pp. e0225817, 2019.

[13] Fabien Dubosson, Jean-Eudes Ranvier, Stefano Bromuri, Jean-Paul Calbimonte, Juan Ruiz, and Michael Schumacher, "The open D1NAMO dataset: A multi-modal dataset for research on non-invasive type 1 diabetes management," *Informatics in Medicine Unlocked*, vol. 13, pp. 92–100, Oct. 2018.

[14] Heather Hall, Dalia Perelman, Alessandra Breschi, Patricia Limcaoco, Ryan Kellogg, Tracey McLaughlin, and Michael Snyder, "Glucotypes reveal new patterns of glucose dysregulation," *PLoS Biology*, vol. 16, no. 7, pp. e2005143, 2018.

[15] Ruth S Weinstock, Stephanie N DuBose, Richard M Bergenstal, Naomi S Chaytor, Christina Peterson, Beth A Olson, Medha N Munshi, Alysa JS Perrin, Kellee M Miller, Roy W Beck, et al., "Risk factors associated with severe hypoglycemia in older adults with type 1 diabetes," *Diabetes Care*, vol. 39, no. 4, pp. 603–610, Oct. 2016.

[16] Federico Garza, Max Mergenthaler Canseco, Cristian Challú, and Kin G Olivares, "Statsforecast: Lightning fast forecasting with statistical and econometric models," *PyCon Salt Lake City, Utah, US*, vol. 2022, pp. 66, 2022.

[17] Julien Herzen, Francesco Lässig, Samuele Giuliano Piazzetta, Thomas Neuer, Léo Tafti, Guillaume Raille, Tomas Van Pottelbergh, Marek Pasieka, Andrzej Skrodzki, Nicolas Huguenin, et al., "Darts: User-friendly modern machine learning for time series," *Journal of Machine Learning Research*, vol. 23, no. 124, pp. 1–6, 2022.

[18] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, Dec. 2017, vol. 30.

[19] Renat Sergazinov, Mohammadreza Armandpour, and Irina Gaynanova, "Gluformer: Transformer-based personalized glucose forecasting with uncertainty quantification," in *International Conference on Acoustics, Speech and Signal Processing*, Jun. 2023, pp. 1–5.

[20] Axel Brando, Jose A Rodriguez, Jordi Vitria, and Alberto Rubio Muñoz, "Modelling heterogeneous distributions with an uncountable mixture of asymmetric laplacians," in *Advances in Neural Information Processing Systems*, Dec. 2019, vol. 32.

[21] Carl Rasmussen, "The infinite gaussian mixture model," in *Advances in Neural Information Processing Systems*, Dec. 1999, vol. 12.

[22] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Jul. 2019, pp. 1314–1324.

[23] Marios Anthimopoulos, Joachim Dehais, Sergey Shevchik, Botwey H. Ransford, David Duke, Peter Diem, and Stavroula Mougiakakou, "Computer vision-based carbohydrate estimation for Type 1 patients with diabetes using smartphones," *Journal of Diabetes Science and Technology*, vol. 9, no. 3, pp. 507–515, Apr. 2015.