Deep Learning Approach for Diagnosing Diabetes Progression from Continuous Glucose Monitoring Data

Anisha Raghu*[1], Ananya Raghu†[1], Giacomo Nebbia[2], Praveer Singh[3], and Jayashree Kalpathy-Cramer[4]

[1]Quarry Lane School, Dublin, CA, USA
[2]add
[3]Department of Ophthalmology, University of Colorado Anschutz Medical Campus, Aurora, CO, USA
[4]add

1 Abstract

Objective: To investigate whether continuous glucose monitoring (CGM)-derived glycemic patterns can distinguish prediabetes from oral-medication-treated diabetes, and to evaluate whether CGM provides discriminative value beyond HbA1c alone.

Materials and Methods: We use the AI-READI dataset to train models on CGM time-series data collected from Dexcom sensors. We compare HbA1c-only baselines, classical time-series models, models with handcrafted CGM features, and deep learning approaches. Our proposed architecture combines a 1D-CNN that learns robust embeddings of intra-day glucose patterns with a transformer that models variability and temporal structure across days. To improve robustness, we apply data augmentation and supervised pretraining of the 1D-CNN backbone. External validation is performed on the CGMacros dataset, including CGM data collected using Libre and Dexcom sensors on the same cohort of patients.

Results: CGM-based models outperform HbA1c-only baselines by 9% in balanced accuracy (BA). Deep learning approaches further improve performance. The proposed 1D-CNN + transformer model achieves a BA of 74% on AI-READI, outperforming handcrafted-feature models by 8%. On external validation with CGMacros, the model demonstrates strong generalization across sensors, achieving BA of 78% and 74% on Libre and Dexcom sensor data, showing consistent performance while outperforming handcrafted-feature models that show a lot of variation across sensors for the same cohort.

Discussion: Data augmentation significantly improves performance, providing robustness in performance across CGM sensors. Pretraining the 1D-CNN backbone further improves performance.

Conclusion: These findings indicate that CGM time-series modeling enables more accurate stratification of diabetes states than HbA1c and handcrafted CGM metrics alone, highlighting the value of deep temporal models for CGM-based risk assessment.

2 Introduction

Diabetes mellitus is one of the most prevalent chronic diseases worldwide, affecting more than 530 million adults, with projections rising to nearly 783 million by 2045 [1]. Even more concerning, over 98 million people currently live with pre-diabetes, a condition that often progresses silently to type 2 diabetes without early intervention [2]. Because symptoms may be mild or absent in the early stages, diabetes is frequently diagnosed late, by which point patients may already face complications such as neuropathy, retinopathy, and cardiovascular disease, conditions that are more difficult to manage

*†These authors contributed equally to this work. Work performed during an internship at QTIM Lab.

once established [3]. Traditionally, diagnosis relies on fasting plasma glucose, HbA1c, or oral glucose tolerance testing, but these capture only static snapshots of glucose control. In contrast, continuous glucose monitoring (CGM) provides dynamic, real-time data on glucose fluctuations and is increasingly used in diabetes management to adjust therapy and detect glycemic variability [4]. Because CGM collects near-continuous glucose readings, it can reveal hidden events such as nocturnal hypoglycemia and post-meal spikes that routine tests often miss [5]. Importantly, CGM-derived metrics such as time in range (TIR) have been shown to correlate with the risk of microvascular complications, including retinopathy and nephropathy, establishing CGM as a clinically meaningful measure of glycemic control rather than solely a monitoring tool [6]. Together, these insights help tailor treatments more precisely and support behavioral changes by giving patients clearer feedback on how daily choices affect their glucose levels. CGM-derived metrics are increasingly used to evaluate treatment response and stratify glycemic control in patients receiving glucose-lowering medications, providing clinically actionable information beyond HbA1c alone [7].

2.1 Contributions

- We directly predict diabetes onset from CGM time series data instead of using handcrafted features.
- We develop a novel pre-training method to train a 1D CNN backbone that is further fine tuned to build the final model.
- We show that one week of CGM data outperforms HbA1c measurements as a marker for detection of onset of diabetes.

2.2 Prior Work

Although HbA1c remains the most widely used clinical marker for diagnosing diabetes, its reliability is limited. Most prior studies using CGM data have focused on forecasting short-term glucose trajectories rather than diagnosing chronic conditions. In particular, CGM has been widely leveraged to predict near-future glucose levels, hypoglycemic events, or overall glycemic variability, using methods ranging from statistical models and machine learning [8] to deep learning frameworks such as bi-LSTMs and transformers [9], [10], [11]. Relatively limited work, however, has explored the use of CGM data for the direct prediction of diabetes status itself.

A recent study in *Nature Medicine* [12] proposed a diagnostic approach based on six handcrafted features derived from CGM traces: mean glucose level, expected maximum spike relative value, percentage of time above 150 mg/dl, spike resolution (time to absorb 50% of a glucose spike), expected daily number of spikes, and nocturnal hypoglycemia. These metrics summarize key aspects of glucose dynamics and were shown to distinguish normoglycemic, prediabetic, and diabetic participants. For further methodological details, we refer the reader to [12]. These features were then used in a predictive framework to assess diabetes status. While this approach demonstrated the potential of CGM-derived features, it relied on manually engineered summaries of the data rather than the full temporal signal.

To our knowledge, there are currently no published methods that attempt to use the raw CGM time series to discriminate between pre-diabetic and medication controlled diabetes patients. Addressing this gap is important because handcrafted features may discard informative temporal patterns that could be critical for accurate classification. Leveraging the full time series with deep learning has the potential to capture richer representations of glucose dynamics, enabling more robust and generalizable prediction.

3 Methods

3.1 Datasets

The AI-READI dataset [13] [14] is a large-scale, multimodal resource designed to advance diabetes research by linking systemic health data with high-resolution ocular imaging. The dataset provides extensive participant metadata that complements its multimodal physiological and imaging data. The AI-READI dataset has data for 1067 patients, out of which 1011 patients had CGM data for atleast

Study Group	Train	Validation	Test
Prediabetes or lifestyle-controlled diabetes (Group 1)	37	152	39
Oral / non-insulin injectable–controlled diabetes (Group 2)	41	220	44

Table 1: Distribution of participants across training, validation, and test splits in the AI-READI dataset.

one week. The distribution of patients from study group 1 and 2 (prediabetes/lifestyle controlled diabetes and oral/non-insulin controlled diabetes) across the train, validation, and test splits is shown in Table 1.

The CGMacros dataset [15] is used as an external test set for model evaluation. It consists of CGM time-series data collected from two commercially available sensors, Dexcom and Libre, which were worn simultaneously by each participant, enabling direct cross-sensor comparison. The dataset includes recordings from 45 individuals, spanning a clinically relevant spectrum of glycemic status: 16 prediabetic participants and 14 diabetic participants whose condition is managed through lifestyle interventions or metformin therapy. This composition makes CGMacros well suited for assessing model generalization across different sensors.

As shown in Figure 1, there is substantial overlap in HbA1c values between prediabetic/lifestyle controlled diabetes and patients needing medication for diabetes in the AI-READI dataset. This observation suggests that relying solely on HbA1c may lead to misclassification. Motivated by this limitation, we incorporate continuous glucose monitoring (CGM) time-series data to capture dynamic patterns of glucose variation for better detection of diabetes onset.

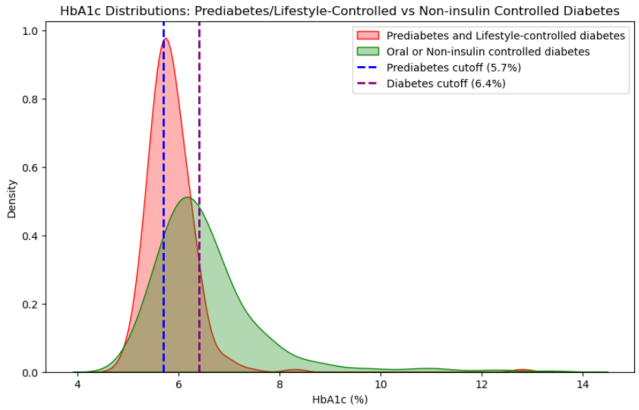

Figure 1: Distribution of HbA1c values for Group 1 (Prediabetic/Life Style Controlled Diabetes) and Group 2 (Oral or Non-insulin controlled diabetes). While HbA1c is commonly used as a diagnostic marker, many patients clinically labeled as diabetic have values below the 6.5% threshold, and some non-diabetic patients exceed it. This overlap suggests that HbA1c alone may not be a sufficient diagnostic indicator.

3.2 Model Architecture

To assess the added value of continuous glucose monitoring (CGM), we compare HbA1c-based baselines, classical CGM approaches using handcrafted features, and deep learning methods operating directly on raw CGM time-series data (Figure 2).

Machine Learning Models

- **HbA1c-only model:** A random forest classifier trained using HbA1c as the sole input feature.

- **Handcrafted CGM features:** A random forest classifier trained on summary CGM metrics extracted following [12].
- **PCA + XGBoost:** Principal component analysis applied to CGM time-series data, followed by an XGBoost classifier [16].
- **PCA + TabPFN:** Principal component analysis applied to CGM time-series data, followed by a TabPFN classifier [17].

Deep Learning Approaches

- **1D CNN (weekly):** A one-dimensional convolutional neural network trained end-to-end on a full week of CGM time-series data.
- **2D CNN (weekly):** A two-dimensional convolutional neural network trained on a full week of CGM data, treating intra-day time and inter-day structure as separate dimensions.
- **Transformer + 1D CNN (proposed):** A hybrid architecture in which a 1D CNN backbone learns intra-day glucose representations that are subsequently modeled across days using a transformer encoder.

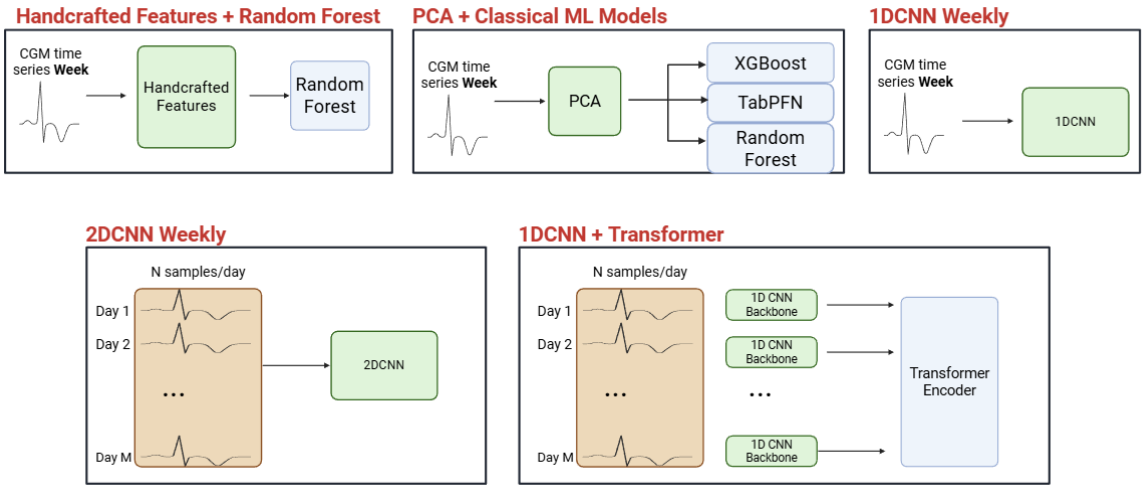

Figure 2: Overview of modeling approaches evaluated for CGM-based diabetes stratification. Methods include classical models using handcrafted CGM features with random forests, PCA-based dimensionality reduction followed by classical machine learning models, deep learning approaches using weekly 1D and 2D CNNs, and a hybrid architecture combining a 1D CNN backbone with a transformer encoder to model both intra-day glucose patterns and cross-day temporal structure.

3.3 Data Augmentation

Continuous glucose monitoring (CGM) signals exhibit substantial variability across sensors, even when collected from the same individual. Figure 3 visualizes representative differences between Dexcom and Libre sensors from the CGMacros dataset, including shifts in baseline glucose levels, differences in amplitude, and sensor-specific noise characteristics. These differences motivated the use of data augmentation to make models trained on AI-READI more robust to variability across CGM sensors.

To improve robustness to such sensor-dependent variability, we apply data augmentation during training on the AI-READI dataset. Augmentation is designed to simulate realistic variations observed across sensors and individuals, encouraging the model to learn sensor-invariant glycemic patterns. We apply augmentation to both CGM time-series data and selected clinical metadata, including high blood pressure status.

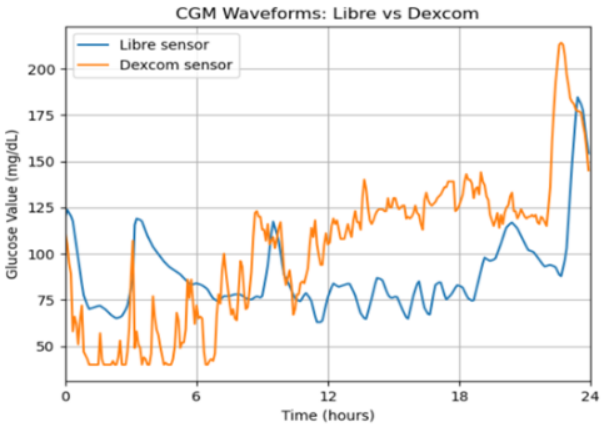

Figure 3: Representative variation in CGM time-series data between Dexcom and FreeStyle Libre sensors worn simultaneously by the same participant, illustrating differences in baseline, amplitude, and noise characteristics.

For the CGM signals, we apply several simple transformations, including shifting the overall glucose level, scaling the signal, adding random noise, and applying a temporal shift. These augmentations keep the overall shape of the glucose patterns the same while introducing realistic variation.

Figure 4 shows an example of CGM data before and after augmentation, where the augmented signal differs slightly from the original 24-hour glucose trace. These augmentations help the model handle variation in CGM data and perform better when tested on external datasets and different sensors.

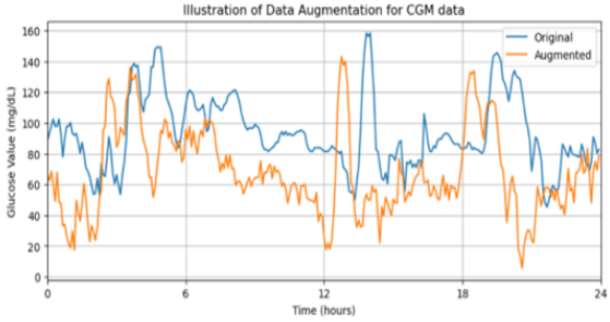

Figure 4: Example of CGM data before and after augmentation. The original 24-hour glucose trace (blue) is compared with an augmented version (orange) created by applying scaling, time shifts, and added noise.

3.4 Training

All models are trained using the training split of the AI-READI dataset, with hyper-parameters selected based on performance on the validation set. To ensure fair comparison across methods and reduce sensitivity to random initialization, we repeat each experiment 20 times with different random seeds and report the average performance across runs.

Hyper-parameter tuning is performed exclusively on the validation set and includes selection of number of PCA components, tree depth, number of trees for classical ML approaches and selection of learning rate and number of epochs for the deep learning model.

For the transformer+1D CNN model, We first pretrain the 1D CNN backbone to predict diabetes status from a single day of CGM data, allowing the model to learn robust glycemic representations. Pretraining convolutional neural networks on time-series data has been shown to improve downstream

classification performance [18]. For weekly modeling, each day's CGM signal is independently encoded using the pretrained 1D CNN, producing 1024-dimensional embeddings that are fused using a transformer encoder to capture cross-day temporal structure.

More detailed descriptions of the training procedure and hyper-parameter settings are provided in the released reference code [19].

4 Results

We first report results on the AI-READI dataset. Table 2 summarizes performance across all evaluated models, including HbA1c-based baselines, classical CGM approaches, and deep learning architectures. Figure 5 presents the balanced accuracy achieved by each model on AI-READI. The proposed 1DCNN + Transformer model achieves the highest balanced accuracy among all evaluated methods, outperforming both classical machine learning and other deep learning approaches.

We next evaluate model generalization on the CGMacros dataset. Figure 6 shows balanced accuracy across Dexcom and Libre sensors for the handcrafted-feature random forest and the 1DCNN + Transformer model. The proposed model demonstrates consistent performance across both sensors, while the handcrafted-feature model exhibits greater variability between sensor types.

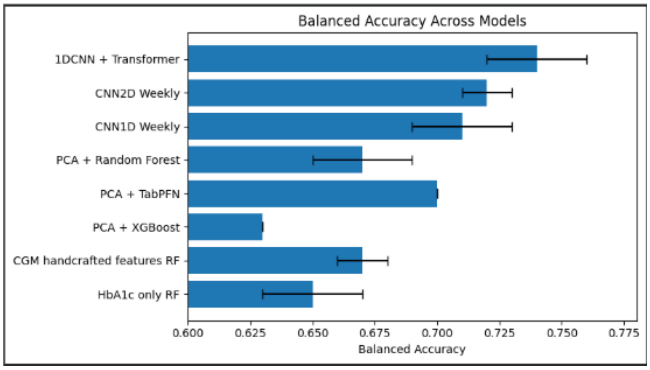

Figure 5: 1DCNN + Transformer model outperforms all other approaches on the AI-READI dataset, showing superior performance in diabetes stage prediction.

Table 2: Results for all models evaluated on the AI-READI dataset. Metrics are reported as mean ± standard deviation.

Model	AUC	AUPRC	Accuracy	BA	F1
HbA1c only RF	0.73 ± 0.005	0.76 ± 0.01	0.66 ± 0.01	0.65 ± 0.02	0.63 ± 0.02
CGM handcrafted features RF	0.76 ± 0.01	0.79 ± 0.01	0.67 ± 0.01	0.67 ± 0.01	0.72 ± 0.01
PCA + XGBoost	0.67	0.70	0.64	0.63	0.66
PCA + TabPFN	0.75	0.82	0.70	0.70	0.72
PCA + Random Forest	0.72 ± 0.02	0.75 ± 0.02	0.67 ± 0.02	0.67 ± 0.02	0.71 ± 0.03
CNN1D Weekly	0.80 ± 0.01	0.84 ± 0.00	0.70 ± 0.01	0.71 ± 0.02	0.68 ± 0.03
CNN2D Weekly	0.82 ± 0.01	0.85 ± 0.01	0.71 ± 0.01	0.72 ± 0.01	0.69 ± 0.03
1DCNN + Transformer	$\mathbf{0.82 \pm 0.01}$	$\mathbf{0.86 \pm 0.01}$	$\mathbf{0.74 \pm 0.02}$	$\mathbf{0.74 \pm 0.02}$	$\mathbf{0.73 \pm 0.02}$

5 Discussion

Overall, results show that models leveraging CGM time-series data outperform HbA1c-only and handcrafted-feature baselines, with the proposed 1DCNN + Transformer achieving the highest balanced accuracy. We further analyze the impact of key design choices through ablation studies.

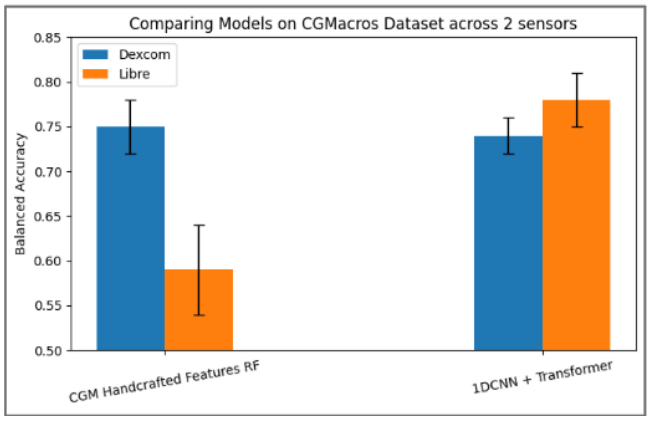

Figure 6: Balanced accuracy of CGM handcrafted-feature random forest and 1DCNN + Transformer models across Dexcom and Libre sensors. Our model is highly robust across sensors in contrast to hand-crafted feature based models.

5.1 Ablation Studies

5.1.1 Effect of Data Augmentation

To assess the role of data augmentation in cross-dataset and cross-sensor generalization, we evaluate multiple augmentation strategies when training on the AI-READI dataset and testing on the CGMacros dataset across Libre and Dexcom sensors. Table 3 reports balanced accuracy on the AI-READI dataset as well as on CGMacros for each augmentation setting.

Table 3: Effect of training data augmentation on balanced accuracy evaluated on the AI-READI dataset and the CGMacros dataset across Libre and Dexcom sensors.

Training Augmentation	AI-READI	CGMacros Libre	CGMacros Dexcom
Temporal shift, scale, noise & sensor bias	0.74	0.78	0.74
Scale, noise & sensor bias	0.73	0.64	0.68
None	0.70	0.52	0.66

5.1.2 Effect of 1D CNN Pretraining

We also evaluate the effect of pretraining the 1D CNN backbone on daily CGM data prior to weekly modeling. Table 4 compares performance of the 1DCNN + Transformer model with and without pretraining on the CGMacros dataset.

Table 4: Balanced accuracy of the 1DCNN + Transformer model with and without pretraining of the 1D CNN backbone on daily CGM data.

Model	Balanced Accuracy
1DCNN + Transformer with pretrained 1D CNN backbone	0.74
1DCNN + Transformer (no pretraining)	0.69

6 Conclusion

Overall, our findings demonstrate that modeling continuous glucose monitoring (CGM) data as time series provides a substantial advantage over traditional HbA1c-based approaches, yielding a 9% improvement in balanced accuracy. Deep learning methods consistently outperform both hand-crafted CGM features and classical machine learning models, with transformer-based architectures achieving the strongest performance at 74% balanced accuracy. Importantly, the transformer + 1D CNN model

generalizes robustly across external datasets, maintaining high performance on CGMacros for both Dexcom and Libre sensors, whereas hand-crafted features show limited transferability. We further show that targeted data augmentation is critical for robustness, improving balanced accuracy by up to 14%, and that pretraining the 1D CNN backbone provides an additional 5% gain. Together, these results highlight the value of end-to-end, pretrained deep temporal models for capturing clinically meaningful glucose dynamics and support CGM-based modeling as a more sensitive and generalizable framework for diabetes risk stratification.

7 Acknowledgements

We would like to thank the Quantitative Translational Imaging in Medicine (QTIM) Lab and the University of Colorado Anschutz Medical Campus for providing us with guidance and resources.

References

[1] Md Jamal Hossain, Md Al-Mamun, and Md Rabiul Islam. Diabetes mellitus, the fastest growing global public health concern: Early detection should be focused. *Health Sci. Rep.*, 7(3):e2004, March 2024.

[2] CDC. Prediabetes: Could it be you? infographic. https://www.cdc.gov/diabetes/communication-resources/prediabetes-statistics.html, February 2025. Accessed: 2025-9-1.

[3] Diabetes complications. *Diabetes Mellitus*, January 2008.

[4] Heather Hall, Dalia Perelman, Alessandra Breschi, Patricia Limcaoco, Ryan Kellogg, Tracey McLaughlin, and Michael Snyder. Glucotypes reveal new patterns of glucose dysregulation. *PLoS Biol.*, 16(7):e2005143, July 2018.

[5] Eden M Miller. Using continuous glucose monitoring in clinical practice. *Clin. Diabetes*, 38(5):429–438, December 2020.

[6] https://diabetesjournals.org/care/article-abstract/41/11/2370/36582/Association-of-Time-in-Range-as-Assessed-by?redirectedFrom=fulltext. Accessed: 2025-12-22.

[7] https://www.sciencedirect.com/science/article/abs/pii/S221385872200319. Accessed: 2025-12-22.

[8] William P T M van Doorn, Yuri D Foreman, Nicolaas C Schaper, Hans H C M Savelberg, Annemarie Koster, Carla J H van der Kallen, Anke Wesselius, Miranda T Schram, Ronald M A Henry, Pieter C Dagnelie, Bastiaan E de Galan, Otto Bekers, Coen D A Stehouwer, Steven J R Meex, and Martijn C G J Brouwers. Machine learning-based glucose prediction with use of continuous glucose and physical activity monitoring data: The maastricht study. *PLoS One*, 16(6):e0253125, June 2021.

[9] Min Hyuk Lim, Seunghyeok Jeong, Jaeheon Choi, Young-Hak Kim, Kyung-Ah Park, Jaehyun Lee, Min Gu Kang, Hyeong-Kyu Kim, Seong-Eun Kim, Hyeong Jin Kim, Nam Hoon Lee, Jin-Woo Yoon, Hyejung Yoo, and Hyeong Soo Kim. A deep learning framework for virtual continuous glucose monitoring and glucose prediction based on life-log data. *Scientific Reports*, 15(367):1–13, 2025.

[10] Junjie Luo, Abhimanyu Kumbara, Mansur Shomali, Rui Han, Anand Iyer, Ritu Agarwal, and Gordon Gao. A large sensor foundation model pretrained on continuous glucose monitor data for diabetes management. August 2025.

[11] Ebrahim Farahmand, Reza Rahimi Azghan, Nooshin Taheri Chatrudi, Eric Kim, Gautham Krishna Gudur, Edison Thomaz, Giulia Pedrielli, Pavan Turaga, and Hassan Ghasemzadeh. AttenGluco: Multimodal Transformer-Based blood glucose forecasting on AI-READI dataset. February 2025.

[12] Mattia Carletti, Jay Pandit, Matteo Gadaleta, Danielle Chiang, Felipe Delgado, Katie Quartuccio, Brianna Fernandez, Juan Antonio Raygoza Garay, Ali Torkamani, Riccardo Miotto, Hagai Rossman, Benjamin Berk, Katie Baca-Motes, Vik Kheterpal, Eran Segal, Eric J. Topol, Edward Ramos, and Giorgio Quer. Multimodal ai correlates of glucose spikes in people with normal glucose regulation, pre-diabetes and type 2 diabetes. *Nature Medicine*, July 2025.

[13] AI-READI Consortium. Flagship dataset of type 2 diabetes from the AI-READI project, 2024.

[14] AI-READI Consortium. AI-READI: rethinking AI data collection, preparation and sharing in diabetes research and beyond. *Nat. Metab.*, 6(12):2210–2212, December 2024.

[15] Ricardo Gutierrez-Osuna, David Kerr, Bobak Mortazavi, and Anurag Das. CGMacros: a scientific dataset for personalized nutrition and diet monitoring, January 2025.

[16] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, page 785–794. ACM, August 2016.

[17] Noah Hollmann, Samuel Müller, Katharina Eggensperger, and Frank Hutter. Tabpfn: A transformer that solves small tabular classification problems in a second, 2023.

[18] Kathan Kashiparekh. *ConvTimeNet: A Pre-Trained Deep Convolutional Neural Network for Time Series Classification*. 2019.

[19] Ananya Raghu and Anisha Raghu. Ai-readi cgm diabetes progression. https://github.com/ananyar1007/AIREADI_CGM.