

Personalized Multi-Horizon Glucose Forecasting in Type 1 Diabetes via Deep Learning

Alessandro Santopaolo^{1*}[0009–0003–0992–8534],
Ilaria Basile^{2*}[0009–0002–8158–1363], Alessandro Giuseppi¹[0000–0001–5503–8506],
and Giovanna Sannino²[0000–0001–7856–8761]

¹ Department of Computer, Control, and Management Engineering “Antonio
Ruberti” (DIAG), Sapienza University of Rome, Italy
`{santopaolo,giuseppi}@diag.uniroma1.it`

² Institute for High Performance and Networking (ICAR),
National Research Council of Italy (CNR), Naples, Italy
`{ilaria.basile,giovanna.sannino}@icar.cnr.it`

Abstract. This study presents a Deep Learning framework for forecasting blood glucose levels at a time horizon of 45 and 60 minutes using ElectroCardioGraphy (ECG)–derived features alongside baseline glucose readings. Using data from the DINAMO dataset, we extracted and normalized Heart Rate (HR) and Heart Rate Variability (HRV) metrics and categorized future glucose levels into clinically meaningful bins. A feedforward neural network was trained to classify these categories. The resulting model demonstrated promising classification performance on both prediction horizons, with balanced predictive capability across most glucose categories. To enhance transparency and clinical relevance, we employed SHapley Additive exPlanations (SHAP) to assess feature importance and interpret the model’s predictions. SHAP analysis revealed that HR and HRV metrics—particularly `hr_max`, `hr_min`, `hr_std`, and `nmi_mean`—provide valuable complementary information to current glucose levels for predicting short-term glycemic trends. This highlights the potential of ECG–derived features as predictive biomarkers for short-term glycemic changes. Overall, the proposed framework offers a promising and explainable approach to glucose forecasting, supporting transparency, physiological insight, and trust—key factors for the adoption of predictive models in real-time clinical monitoring.

Keywords: Blood Glucose Levels Forecasting · Heart Rate Variability
· Type 1 Diabetes · Deep Learning · Explainable AI

1 Introduction

Type 1 Diabetes (T1D) is a chronic autoimmune disease characterized by insulin deficiency, caused by the immune-mediated destruction of the pancreatic β -cells. The resulting imbalance of glycidic metabolism leads to a condition of chronic

* These authors contributed equally to this work.

hyperglycemia, which, if not properly managed, can lead to serious microvascular and macrovascular complications [7, 1].

Given the lack of effective treatment for this disease, the only possible therapy is the exogenous administration of insulin in order to maintain glycaemic balance. The patient must follow this replacement therapy throughout his life, combining it with careful self-management that includes continuous monitoring of Blood Glucose Levels (BGLs) and prevention of severe complications [9].

The ability to predict glucose levels in the short term (30-60 min) has great clinical relevance, allowing the prevention of critical events and optimizing insulin therapy, by allowing timely and personalized adjustments to insulin dosing. In this context, predictive models based on Machine Learning (ML) and Deep Learning (DL) techniques have shown great potential in analyzing glucose levels. These models serve as decision support tools in closed-loop insulin delivery systems such as the artificial pancreas [15]. However, in a critical area such as health, non-transparency of DL models is a significant barrier that can hinder their adoption. In this context, the application of Explainable AI (XAI) techniques makes it possible to illustrate the rationale behind the model’s predictions, furthering its comprehensibility [18, 21].

In this work, we propose a DL framework for predicting glucose levels at a 45-minute and 60-minute time horizons. The model leverages BGLs provided by Continuous Glucose Monitors (CGMs) along with Heart Rate Variability (HRV) metrics and Heart Rate (HR) parameters extracted from the Electro-CardioGraphic (ECG) signal acquired by wearable devices. The introduction of ECG-derived features as an additional input aims to enable a noninvasive approach to glucose monitoring. The ease of use of wearable ECG devices, such as smartwatches or fitness bands, makes them perfectly suited to the patient’s daily life, limiting the discomfort and invasiveness of glucose measurements.

Several studies have shown that fluctuations in BGLs can affect the electrical activity of the heart, particularly during hypo- or hyperglycemic events [22, 16]. This is particularly relevant in patients with T1D in whom one of the serious complications that can result from vascular alterations to the Autonomic Nervous System (ANS) is the Cardiovascular Autonomic Neuropathy (CAN). Since HRV reflects the dynamic interplay between sympathetic and parasympathetic activity, its modulation by glycemic variations makes it a promising biomarker for glucose trend prediction [22, 16]. Furthermore, it has been shown that in patients with T1D, hyperglycaemia and hyperinsulinaemia are associated with a high resting HR, while intensive diabetes management is associated with a lower HR [17, 23]. Hence, leveraging HR and HRV data offers a valuable, non-invasive means to enhance glucose level forecasting and support diabetes management.

2 Related Work

Blood glucose level prediction has emerged as a critical research area in diabetes management, driven by the urgent need to prevent severe hypoglycemic and hyperglycemic episodes that can lead to life-threatening complications. The

advancement of ML techniques has substantially enhanced glucose prediction accuracy, with Support Vector Machines (SVMs) and Random Forest algorithms successfully applied to glucose forecasting tasks [24, 11]. However, these traditional approaches typically require extensive manual feature engineering and may not fully capture complex interactions between multiple physiological variables. In contrast, DL techniques have shown greater potential by automatically learning complex patterns and feature interactions without explicit programming. Artificial Neural Networks (ANNs) have demonstrated their ability to model non-linear relationships between input features and glucose levels, effectively processing multi-dimensional feature vectors containing various physiological parameters [19, 3]. Moreover, several studies have shown that simple DL architectures can achieve competitive performance while maintaining computational efficiency suitable for real-time applications [25, 20].

However, the “black box” nature of DL networks poses significant challenges for clinical adoption, where understanding prediction rationale is crucial for building trust and ensuring patient safety. This has led to the increasingly important application of XAI techniques in healthcare [2].

The need for explainability becomes even more crucial when integrating other physiological features in predicting BGLs to understand how these features influence the model’s decision. In fact, recent studies have demonstrated that incorporating additional physiological signals beyond glucose measurements can substantially improve prediction accuracy. The integration of ECG-derived features represents a particularly promising approach [4, 6]. Cordeiro et al. [6] developed a DL architecture for hyperglycemia detection using ECG-derived features, including HRV parameters in time and frequency domains, validated on a private dataset containing ECG recordings and glucose measurements from over a thousand subjects. Arbi et al. [4] created mathematical equations combining corrected QT duration, T-wave amplitude, and HR for glucose level estimation using the D1NAMO dataset, employing recurrent convolutional neural networks for ECG segmentation and achieving very low estimation errors.

Based on these promising results, this work addresses critical gaps by demonstrating the predictive value of ECG-derived features for glucose forecasting using feedforward neural networks, providing comprehensive explainability analysis using SHapley Additive exPlanations (SHAP) to understand feature contributions, and evaluating multi-horizon prediction capabilities to support different clinical decision-making timeframes, thereby bridging the gap between advanced DL techniques and interpretability requirements in diabetes management.

3 Methodology

3.1 Dataset

In this work, we used the D1NAMO dataset [10], a publicly accessible dataset containing data acquired for three weeks in real-life conditions by 29 subjects (20 healthy people and 9 people with T1D). The healthy population is composed

of 16 men and 4 women with an average age of 33 years, whereas the diabetic population is composed of 6 men and 3 women with an average age of 38 years. In this study, we focused only on data from the diabetic population because our intention was to develop an approach capable of identifying critical episodes, such as hyperglycaemic or hypoglycaemic events. This dataset includes glucose measurements, food picture annotations, and ECG, breathing, and accelerometer signals. Physiological signals were acquired using the *Zephyr BioHarness 3 band*, while BGLs were measured every five minutes in diabetic subjects using the *iPro2 Professional CGM sensor*.

The dataset used for the prediction task in this paper was derived from the D1NAMO dataset, using glucose measurements and HRV parameters in both time and frequency domains calculated from the ECG signal provided by the dataset. The HRV features used in this work will be described in detail in the section 3.3. Our dataset contains information from only one individual with T1D from the D1NAMO collection: patient 008. Data from the remaining diabetic patients (001, 002, 003, 004, 005, 006, 007, and 009) were excluded due to technical limitations: glucose measurements were not properly time-aligned with ECG recordings, ECG signals were too noisy due to improper band placement, or patients had insufficient observations for effective deep learning model training. Patient 008 was the only subject meeting all requirements for data quality, temporal synchronization, and adequate sample size.

3.2 Pre-Processing

The ECG signal provided by the D1NAMO dataset required a pre-processing phase before the HRV features could be extracted. Firstly, due to time misalignments between the ECG signal and glucose measurements, only data points with synchronized ECG and BGLs were considered for analysis.

Additionally, due to the noisy nature of wearable ECG signals, filtering and quality assessment were also conducted prior to feature extraction. The ECG signal filtering and R-peaks identification were performed using dedicated functions from the Neurokit2 library [14]. Specifically, the filter applied to the ECG signal was a fifth-order high-pass filter with a cutoff frequency of 0.5 Hz, followed by powerline filtering at 50 Hz.

The subsequent procedure for extracting the QRS complexes and evaluating the ECG Signal Quality Index (SQI) was carried out according to the method described in [5]. At the end of this processing, the output consists of 2-minute ECG signal windows with a 1-minute overlap.

3.3 Feature Extraction

Following the pre-processing phase of the ECG signal, the features derived from HR and HRV were extracted. Given the length of the ECG segments extracted (2 minutes), the analysis focused only on parameters that can be evaluated over ultra-short recording periods (<5 min), calculated using the *pyHRV* toolbox [13]. In the frequency domain, this limited the analysis to parameters calculated

in the High-Frequency (HF) band (0.15–0.40 Hz) and Low-Frequency (LF) band (0.04–0.15 Hz). In the time domain, the triangular HRV index, which requires 24-hour recordings, was excluded.

Therefore, the dataset structure includes the following 21 input variables:

- **glucose_t0**: Glucose value at the baseline [mmol/L];
- **hr_mean**: Average HR computed over the window [bpm];
- **hr_max**: Maximum HR recorded in the window [bpm];
- **hr_min**: Minimum HR recorded in the window [bpm];
- **hr_std**: Standard deviation of HR over the window [bpm];
- **nni_mean**: Mean of NN intervals [ms];
- **sdnn**: Standard Deviation of NN intervals [ms];
- **rmssd**: Root Mean Square of Successive Differences [ms];
- **sdsd**: Standard Deviation of Successive Differences [ms];
- **nn50/nn20**: Number of successive intervals differing by more than 50 ms or 20 ms) [n.u.];
- **pNN50/pNN20**: Percentage of successive intervals differing by more than 50 ms or 20 ms [%];
- **peak_lf/peak_hf**: Peak frequencies in the LF band and HF band;
- **abspower_lf/abspower_hf**: Absolute powers in the LF band and HF band [ms²];
- **logpower_lf/logpower_hf**: Logarithmic powers in the LF band and HF band [log];
- **normpower_hf/normpower_lf**: Normalized powers in the LF and HF band [n.u.].

3.4 Glycemic Class Labeling

Glucose prediction was formulated as a multiclass classification task at 45- and 60-minute horizons, with each future value assigned to one of seven glycemic categories defined according to the classification scheme proposed in [8]. This classification scheme includes two hypoglycemic, three euglycemic, and two hyperglycemic classes, allowing early detection of deviations from normoglycemia and facilitating clinical interpretability. Class labels were assigned to each sample based on the forecasted glucose value at the respective time horizon. Table 1 presents the complete classification scheme with corresponding glucose ranges and sample distribution across all glycemic classes.

3.5 Model Architecture

To model the relationship between ECG-derived features and future glycemic states, we employed a feedforward neural network (FNN) trained as a multiclass classifier. This network is particularly suitable for forecasting tasks thanks to its ability to capture dependencies in sequential data [12].

The network takes as input a 21-dimensional vector corresponding to the features described in Section 3.3. The architecture consists of an input layer followed by two fully connected hidden layers. The first hidden layer comprises

Table 1: Glycemic class definitions and number of instances used for multi-class labeling.

Class Name	Class ID	Range (mmol/L)	Range (mg/dL)	Samples
Very Low	0	< 3.0	< 54	44
Low	1	3.0 – 3.9	54 – 70	40
Normal (lower range)	2	3.9 – 5.0	70 – 90	132
Normal	3	5.0 – 7.8	90 – 140	262
Normal (upper range)	4	7.8 – 10.0	140 – 180	204
High	5	10.0 – 13.9	180 – 250	255
Very High	6	> 13.9	> 250	35

64 neurons with ReLU activation, followed by a dropout layer (rate = 0.3) to reduce overfitting. The second hidden layer contains 32 neurons, also activated via ReLU. The output layer includes 7 neurons with softmax activation, corresponding to the seven glycemic classes defined in Section 3.4.

This architecture was identified through a grid search procedure conducted over a predefined set of hyperparameters. The search space included the number of units in the first hidden layer (32, 64), dropout rates (0.2, 0.3, 0.5) and learning rates (0.001, 0.0005). For each combination, model performance was evaluated using 5-fold cross-validation, and the average classification accuracy on the validation sets was used as the selection criterion. The configuration yielding the highest mean accuracy—64 units in the first hidden layer, a dropout rate of 0.3, and a learning rate of 0.0005—was retained for final training.

The model was trained using the Adam optimizer and a sparse categorical cross-entropy as loss function. To mitigate overfitting, an early stopping criterion was applied, terminating training when validation performance ceased to improve over a defined number of epochs.

4 Experimental Setup

We partitioned the dataset using an 80/20 train-test split. Stratified 5-fold cross-validation was applied within the training set to maintain class balance during hyperparameter optimization.

Model performance was evaluated using the classification metrics: overall accuracy, precision, recall, and F1-score for each glucose class. Confusion matrices were analyzed to identify misclassification patterns between glucose categories.

For interpretability analysis, we implemented SHAP with the DeepExplainer method. We used 100 background samples from the training data to establish baseline feature contributions, then computed SHAP values for 100 test samples. These 100 instances are drawn randomly but with the aim of mirroring the overall class distribution. Global feature importance was determined by averaging absolute SHAP values across instances.

All computations were performed on Google Colab using an NVIDIA Tesla T4 GPU (16 GB VRAM).

5 Results and Discussion

In this section, we report the performance of the proposed model evaluated at the 45-minute and 60-minute prediction horizons. Specifically, we present the classification results in terms of precision, recall, F1 score and support for each glycemic class (Tables 2 and 3). Furthermore, we include confusion matrices (Figure 1a and Figure 1b) to visualize misclassification patterns, along with SHAP summary graphs (Figure 2a and Figure 2b) to assess feature importance and model interpretability.

Class	Prec.	Rec.	F1	Supp.
0	0.90	0.82	0.86	11
1	0.72	0.60	0.65	10
2	0.78	0.77	0.77	26
3	0.70	0.85	0.77	48
4	0.75	0.60	0.67	40
5	0.86	0.96	0.91	51
6	1.00	1.00	1.00	8
Acc.	0.81			
Macro	0.82	0.80	0.80	—
Weighted	0.79	0.81	0.79	—

Table 2: Classification performance metrics for 45-minute glucose forecasting

Class	Prec.	Rec.	F1	Supp.
0	0.82	0.82	0.82	11
1	0.75	0.60	0.67	10
2	0.69	0.69	0.69	26
3	0.63	0.77	0.69	48
4	0.68	0.43	0.52	40
5	0.81	0.90	0.85	51
6	1.00	1.00	1.00	8
Acc.	0.73			
Macro	0.77	0.74	0.75	—
Weighted	0.73	0.73	0.72	—

Table 3: Classification performance metrics for 60-minute glucose forecasting

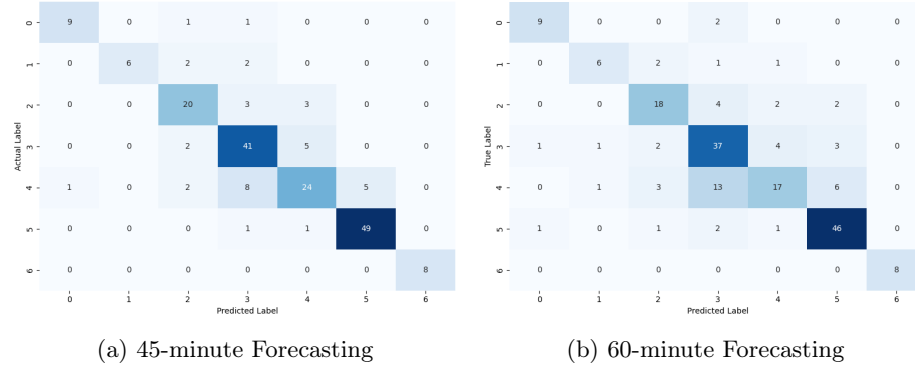


Fig. 1: Confusion matrices showing classification performance across 7 glycemic classes for 45-minute (a) and 60-minute (b) forecasting.

The proposed deep learning framework, tested on subject 008, demonstrated promising predictive performance across both temporal horizons, achieving 81% classification accuracy at 45 minutes and 73% at 60 minutes, with corresponding macro-averaged F1 scores of 0.80 and 0.75 respectively. It is notable that the system demonstrated improved performance for extreme glycemic states (Classes 0, 1, 5 and 6). The temporal degradation from 45 to 60 minutes followed expected

patterns, with Class 4 experiencing the most pronounced decline in F1 score (0.67 to 0.52), reflecting the inherent uncertainty in extended physiological forecasting.

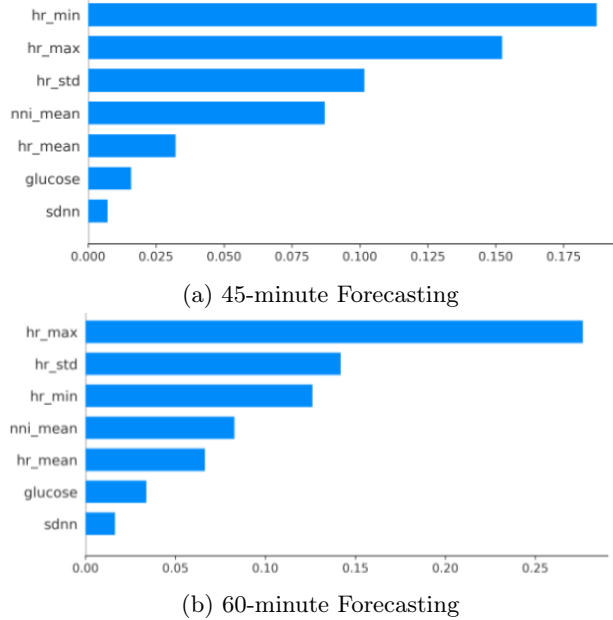


Fig. 2: Comparison of SHAP summary plots for 45-minute (a) and 60-minute (b) glycemic forecasting. The x-axis represents the mean absolute SHAP value.

The not-balanced distribution of glycemic classes in the test set, combined with the absence of imbalance-mitigation techniques, likely inflates the performance metrics of less represent class— e.g. the F1 score achieved for Class 6.

Confusion matrices showed that misclassifications mainly occurred between adjacent glucose categories, indicating physiologically plausible errors that maintain clinical safety. This suggests the model learned meaningful decision boundaries aligned with the continuous nature of glucose dynamics.

SHAP analysis identified **hr_max**, **hr_min**, **hr_std**, and **nni_mean** as the primary ECG-derived contributors to prediction accuracy, alongside **glucose_t0**, the baseline glucose measurement. These ECG-derived features capture complementary information about metabolic state transitions, highlighting their utility as predictive biomarkers for short-term glycemic changes. The findings support integration of wearable ECG monitoring in diabetes management protocols.

6 Conclusions and Future Work

This study presents a DL framework for personalized glucose forecasting in T1D combining ECG-derived features with glucose levels. The model demonstrated

promising classification performance for both 45-minute and 60-minute prediction horizons. SHAP analysis revealed that HR parameters and HRV metrics significantly contribute to prediction accuracy, often outweighing current glucose levels as predictive biomarkers. Despite its potential and explainability, the framework’s main limitation is the use of data from only one subject with adequate ECG-glucose synchronization, restricting generalizability. Future work should focus on expanding datasets via multi-site studies and federated learning, and on integrating lifestyle factors (e.g., nutrition, insulin, physical activity, sleep) to improve accuracy and clinical relevance. Emerging tools like mobile health and Large Language Models (LLMs) could further enable real-time interpretation and personalized recommendations, aiming for a non-invasive, actionable glucose monitoring system for daily diabetes management.

Acknowledgments. The authors thank the project “Digital Driven Diagnostics, prognostics and therapeutics for sustainable Health care”-D34Health (B53C22006120001), funded by the National Complementary Plan within the National Recovery and Resilience Plan, that is part of the NextGenerationEU programme of the Italian MUR.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Addissouky, T.A., Ali, M.M., El Sayed, I.E.T., Wang, Y.: Type 1 diabetes mellitus: retrospect and prospect. *Bulletin of the National Research Centre* **48**(1), 42 (2024)
2. Ahmad, M.A., Eckert, C., Teredesai, A.: Interpretable machine learning in health-care. In: *Proceedings of the 2018 ACM international conference on bioinformatics, computational biology, and health informatics*. pp. 559–560 (2018)
3. Ali, J.B., Hamdi, T., Fnaiech, N., Di Costanzo, V., Fnaiech, F., Ginoux, J.M.: Continuous blood glucose level prediction of type 1 diabetes based on artificial neural network. *Biocybernetics and Biomedical Engineering* **38**(4), 828–840 (2018)
4. Arbi, K.F., Soulimane, S., Saffih, F.: Non-invasive method for blood glucose monitoring using ecg signal. *Polish Journal of Medical Physics and Engineering* **29**(1), 1–9 (2023)
5. Basile, I., Sannino, G.: Blood glucose level prediction in type 1 diabetes: A comparative analysis of interpretable artificial intelligence approaches. *Results in Engineering* **25**, 103681 (2025)
6. Cordeiro, R., Karimian, N., Park, Y.: Hyperglycemia identification using ecg in deep learning era. *Sensors* **21**(18), 6263 (2021)
7. Daneman, D.: Type 1 diabetes. *The Lancet* **367**(9513), 847–858 (2006)
8. De Falco, I., Della Cioppa, A., Koutny, T., Ubl, M., Krcma, M., Scafuri, U., Tarantino, E.: A federated learning-inspired evolutionary algorithm: Application to glucose prediction. *Sensors* **23**(6), 2957 (2023)
9. DiMeglio, L.A., Evans-Molina, C., Oram, R.A.: Type 1 diabetes. *The Lancet* **391**(10138), 2449–2462 (2018)
10. Dubosson, F., Ranvier, J.E., Bromuri, S., Calbimonte, J.P., Ruiz, J., Schumacher, M.: The open d1namo dataset: A multi-modal dataset for research on non-invasive type 1 diabetes management. *Informatics in Medicine Unlocked* **13**, 92–100 (2018)

11. Georga, E.I., Protopappas, V.C., Polyzos, D., Fotiadis, D.I.: Evaluation of short-term predictors of glucose concentration in type 1 diabetes combining feature ranking with regression models. *Medical & biological engineering & computing* **53**, 1305–1318 (2015)
12. Gil-Gamboa, A., Paneque, P., Trull, O., Troncoso, A.: Medium-term water consumption forecasting based on deep neural networks. *Expert Systems with Applications* **247**, 123234 (2024)
13. Gomes, P.: pyHRV: Python toolbox for heart rate variability analysis (2021), <https://github.com/rhenanbartels/pyhrv>, accessed: 2025-05-20
14. Makowski, D., Pham, T., Lau, Z.J., Brammer, J.C., Lespinasse, F., Pham, H., Schölzel, C., Chen, S.A.: Neurokit2: A python toolbox for neurophysiological signal processing. *Behavior research methods* pp. 1–8 (2021)
15. Martínez-Delgado, L., Muñoz-Organero, M., Queipo-Alvarez, P.: Using absorption models for insulin and carbohydrates and deep learning to improve glucose level predictions. *Sensors* **21**(16), 5273 (2021)
16. Olde Bekkink, M., Koeneman, M., de Galan, B.E., Bredie, S.J.: Early detection of hypoglycemia in type 1 diabetes using heart rate variability measured by a wearable device. *Diabetes care* **42**(4), 689–692 (2019)
17. Paterson, A.D., Rutledge, B.N., Cleary, P.A., Lachin, J.M., Crow, R.S., Control, D., of Diabetes Interventions, C.T., Group, C.R.: The effect of intensive diabetes treatment on resting heart rate in type 1 diabetes: the diabetes control and complications trial/epidemiology of diabetes interventions and complications study. *Diabetes care* **30**(8), 2107–2112 (2007)
18. Pawar, U., O’shea, D., Rea, S., O’reilly, R.: Explainable ai in healthcare. In: 2020 international conference on cyber situational awareness, data analytics and Assessment (CyberSA). pp. 1–2. IEEE (2020)
19. Rajagopal, S., Koshy, S., Sagar, R.: A comparative study of deep learning techniques for the prediction of blood glucose level in type-1 diabetic patients
20. Rodríguez-Rodríguez, I., Chatzigiannakis, I., Rodríguez, J.V., Maranghi, M., Gentili, M., Zamora-Izquierdo, M.Á.: Utility of big data in predicting short-term blood glucose levels in type 1 diabetes mellitus through machine learning techniques. *Sensors* **19**(20), 4482 (2019)
21. Sindiramutty, S.R., Tee, W.J., Balakrishnan, S., Kaur, S., Thangaveloo, R., Jazri, H., Khan, N.A., Gharib, A., Manchuri, A.R.: Explainable ai in healthcare application. In: *Advances in Explainable AI Applications for Smart Cities*, pp. 123–176. IGI Global Scientific Publishing (2024)
22. Singh, J.P., Larson, M.G., O’Donnell, C.J., Wilson, P.F., Tsuji, H., Lloyd-Jones, D.M., Levy, D.: Association of hyperglycemia with reduced heart rate variability (the framingham heart study). *Am J Cardiol.* **86**(3), 309–312 (2000)
23. Torchinsky, M.Y., Gomez, R., Rao, J., Vargas, A., Mercante, D.E., Chalew, S.A.: Poor glycemic control is associated with increased diastolic blood pressure and heart rate in children with type 1 diabetes. *Journal of diabetes and its complications* **18**(4), 220–223 (2004)
24. Zecchin, C., Facchinetti, A., Sparacino, G., De Nicolao, G., Cobelli, C.: Neural network incorporating meal information improves accuracy of short-time prediction of glucose concentration. *IEEE Transactions on Biomedical Engineering* **59**(6), 1550–1560 (2012)
25. Zhu, T., Li, K., Herrero, P., Chen, J., Georgiou, P.: A deep learning algorithm for personalized blood glucose prediction. In: *KDH@ IJCAI*. pp. 64–78 (2018)