Exploring Nutritional Influence on Blood Glucose Forecasting for Type 1 Diabetes Using Explainable AI

Giovanni Annuzzi, Andrea Apicella, Pasquale Arpaia, *Senior, IEEE*, Lutgarda Bozzetto, Sabatina Criscuolo, Egidio De Benedetto, *Senior, IEEE*, Marisa Pesola, Roberto Prevete

***Abstract*— Type 1 diabetes mellitus (T1DM) is characterized by insulin deficiency and blood sugar control issues. The state-of-the-art solution is the artificial pancreas (AP), which integrates basal insulin delivery and glucose monitoring. However, APs are unable to manage postprandial glucose response (PGR) due to limited knowledge of its determinants, requiring additional information for accurate bolus delivery, such as estimated carbohydrate intake. This study aims to quantify the influence of various meal-related factors on predicting postprandial blood glucose levels (BGLs) at different time intervals (15 min, 60 min, and 120 min) after meals by using deep neural network (DNN) models. The prediction models incorporate preprandial blood glucose values, insulin dosage, and various meal-related nutritional factors such as intake of energy, carbohydrates, proteins, lipids, fatty acids, fibers, glycemic index, and glycemic load as input variables. The impact of input features was assessed by exploiting eXplainable Artificial Intelligence (XAI) methodologies, specifically SHapley Additive exPlanations (SHAP), which provide insights into each feature's contribution to the model predictions. By leveraging XAI methodologies, this study aims to enhance the interpretability and transparency of BGL prediction models and validate clinical literature hypotheses. The findings can aid in the development of decision-support tools for individuals with T1DM, facilitating PGR management and reducing the risks of adverse events. The improved understanding of PGR determinants may lead to advancements in AP technology and improve the overall quality of life for T1DM patients.**

***Index Terms*— predictive models, postprandial blood glucose response, machine learning, explainable artificial intelligence, interpretability, meal-related features**

This work was carried out as part of the 'AI4PG' project, under the *FRA-Finanziamento per la Ricerca di Ateneo* initiative, financially supported by the University of Naples Federico II. This work is also supported by the European Union - FSE-REACT-EU, PON Research and Innovation 2014-2020 DM1062/2021 contract number 18-I-15350-2, and was partially supported by the Ministry of University and Research, PRIN research project "BRIO – BIAS, RISK, OPACITY in AI: design, verification and development of Trustworthy AI.", Project no. 2020SSKZ7R . Furthermore, we acknowledge financial support from the PNRR MUR project PE0000013-FAIR (CUP: E63C22002150007).

G. Annuzzi and L. Bozzetto are with the Department of Clinical Medicine and Surgery, University of Naples Federico II, Naples, Italy (e-mail: annuzzi@unina.it)

A. Apicella, P. Arpaia, S. Criscuolo, E. De Benedetto, M. Pesola, and R. Prevete are with the Department of Electrical Engineering and Information Technology (DIETI), University of Naples Federico II, Naples, Italy (e-mail:egidio.debenedetto@unina.it)

I. Introduction

TYPE 1 diabetes mellitus (T1DM) is an autoimmune disorder characterized by the destruction of pancreatic β-cells, inadequate insulin production, and compromised glycemia regulation [1]. As a result, the primary objective in T1DM management is achieving glycemia control via exogenous insulin administration. The accurate determination of insulin dosage at specific time points presents a significant challenge, as patients face fluctuations in metabolic needs [2]. Recent years have witnessed remarkable technological advancements with the emergence of closed-loop systems, commonly known as the artificial pancreas (AP). These innovative systems aim to replicate physiological insulin release through automated, glucose-responsive insulin delivery. Comprising an insulin pump, a continuous glucose monitor (CGM), and a control algorithm built on heuristics and theoretical knowledge, the AP systems are designed to optimize glycemia control by minimizing both hyperglycemic and hypoglycemic episodes [3]. In spite of their successful automation of basal insulin delivery, AP systems encounter challenges in efficiently addressing postprandial glucose regulation (PGR), a critical concern for individuals with T1DM who must manually determine the preprandial insulin dose based on meal information [4].

To overcome this challenge, the integration of innovative decision support algorithms into T1DM therapy can potentially facilitate PGR management. Advanced methods of artificial intelligence (AI), specifically machine learning (ML), focusing on predicting future blood glucose levels (BGLs) offer a potential approach for improving diabetes treatment and mitigating adverse events [5], [6]. Existing literature highlights the potential of ML-based approaches, including artificial neural networks, to predict future BGLs, enabling early detection of hypo- and hyperglycemic events, and optimizing insulin administration [6]–[8]. This knowledge can be integrated into closed-loop systems, like AP, to improve insulin delivery adequacy. However, predicting BGLs remains challenging due to various factors beyond insulin infusions, including meal intake, physical activity, sleep patterns, and emotional states, which impact real glucose signals [9]. Despite the importance of these factors, there is a scarcity of models that incorporate them into their framework, also considering the lack of real data availability. Prior studies [10], [11] investigated the

role of nutritional factors as input components in forecasting postprandial BGLs over various prediction horizons (PHs). These reported a significant impact of nutritional factors on the predicted BGLs, with increasing PHs showing more influence. Nonetheless, the aforementioned works, as the majority of BGLs prediction models in the literature, lack interpretability, rendering them as *black boxes* for ML scientists, healthcare practitioners, and patients. Indeed, in the field of decision-support models for medicine, there is a growing demand for transparent models capable of generating dependable and interpretable predictions [12]. In this regard, eXplainable Artificial Intelligence (XAI) has emerged as a crucial research area. XAI encompasses a range of techniques and methodologies with the primary objective of providing human-understandable explanations for the decisions and predictions generated by AI models [13], [14]. As a result, various methodologies and algorithms have been developed to unveil the inner workings of AI models and their decision-making processes [15].

Within this context, the present study aims to investigate the impact of specific input features on BGLs prediction by employing XAI methodologies. To achieve this objective, a postprandial BGL prediction model is proposed, having as features BGL values, the quantity of insulin administered during mealtime, microboluses of insulin provided by the AP system before the meal, as well as meal-related attributes (e.g., intake of energy, carbohydrates, proteins, lipids, fatty acids, fiber, cholesterol, glycemic index, and glycemic load). This investigation specifically involves the development of three distinct prediction models, with PHs of $15\,\mathrm{min}$, $60\,\mathrm{min}$, and $120\,\mathrm{min}$ after the meal. To assess the influence of each input feature and to confirm clinical evidence on the importance of nutritional factors, the SHapley Additive exPlanations (SHAP) approach is exploited [16]. In this regard, it is worth mentioning that the aim of this research is the assessment of the relative influence of various nutritional factors, and not towards the development of an advanced predictive model for glucose levels. Thus, in the proposed approach, the predictive model serves as a means to an end, rather than the primary object. According to our current understanding, the comprehensive impact of various nutritional factors on BGLs in T1DM appears as not thoroughly addressed with respect to XAI, but only from a clinical point of view, in the existing literature [17]–[19].

The work is structured as follows. Section II presents a thorough literature review on ML-powered BGL prediction systems and XAI applications, highlighting the current state of knowledge and gaps in the field. Sections III and Section IV outline the methodology employed, including data collection and study design. Section V reports the obtained results, providing detailed insights into the measured parameters and their significance. Building upon this, Section VI presents a comprehensive discussion, interpreting the findings within the context of existing diabetes management strategies. Finally, Section VII summarizes the key findings and paves the way to future works.

TABLE I
PREDICTION PERFORMANCE COMPARISON IN CURRENT LITERATURE. PH [PREDICTION HORIZON], RMSE [ROOT MEAN SQUARED ERROR]

Study	Type of inputs	PH (min)	RMSE (mg/dL)
Annuzzi et al. [10]	CGM data, statistical attributes, insulin, meal related information	30 60	8.0 ± 0.6 21.3 ± 1.6
Daniels et al. [21]	CGM data, insulin, physiological signals, meal intake	30 60 120	18.8 ± 2.3 31.8 ± 3.9 47.2 ± 4.6
Alfian et al. [23]	CGM data, statistical attributes	30 60	6.6 ± 2.4 15.3 ± 5.9
Jaloli et al. [27]	CGM data, insulin, carbohydrates	30 60	9.8 ± 1.2 18.3 ± 2.8
Li et al. [31]	CGM data, insulin, carbohydrates	30 60	19.3 ± 2.8 31.8 ± 3.5

II. RELATED WORKS

In recent years, ML has disclosed new perspectives in AP systems, providing the opportunity to successfully extract knowledge from data. Particularly, ML methodologies focused on predicting future BGLs have emerged as a promising approach for empowering individuals with T1DM [6], [9], [20], [21], effectively addressing the limitations associated with adverse events [5]. In the literature, several ML-based strategies, such as deep neural networks (DNNs), have demonstrated potential in BGL prediction and early detection of hypo- and hyperglycemic events, leading to improved preprandial insulin administration. In particular, numerous studies have investigated BGL prediction using different neural network models, including Feed Forward Neural Network (FFNN) [10], [22]–[24], Long Short Term Memory (LSTM) [25]–[30], and Convolutional Neural Network (CNN) [31], [32]. Although these ML models achieve satisfactory performance in predicting BGLs (see Table I), their lack of interpretability remains a significant issue [33].

Indeed, in the field of medical ML research, an ongoing debate revolves around the importance of transparent models capable of producing reliable and interpretable predictions [34]. The process of defining a medical problem that can be effectively addressed through ML, acquiring relevant data, cleansing the data, and refining the ML model to achieve optimal performance still requires substantial effort. As a result, researchers often overlook or neglect the incorporation of explainability methods. However, there has been a noticeable surge in the number of scientific papers focusing on several applications of XAI, such as image classification [35]–[37] and natural language processing [35], [38], [39]. Surely, in recent years, the healthcare domain has also been affected, especially involving biomedical [40] and tabular data [34]. This growing interest is motivated not only by the necessity of explaining existing *black-box* models and developing interpretable *white-box* models, but also by the critical importance of proactively addressing concerns related to discrimination and biases within datasets and model training [33], in order to prevent disparities and promote equitable representation of the patients' popula-

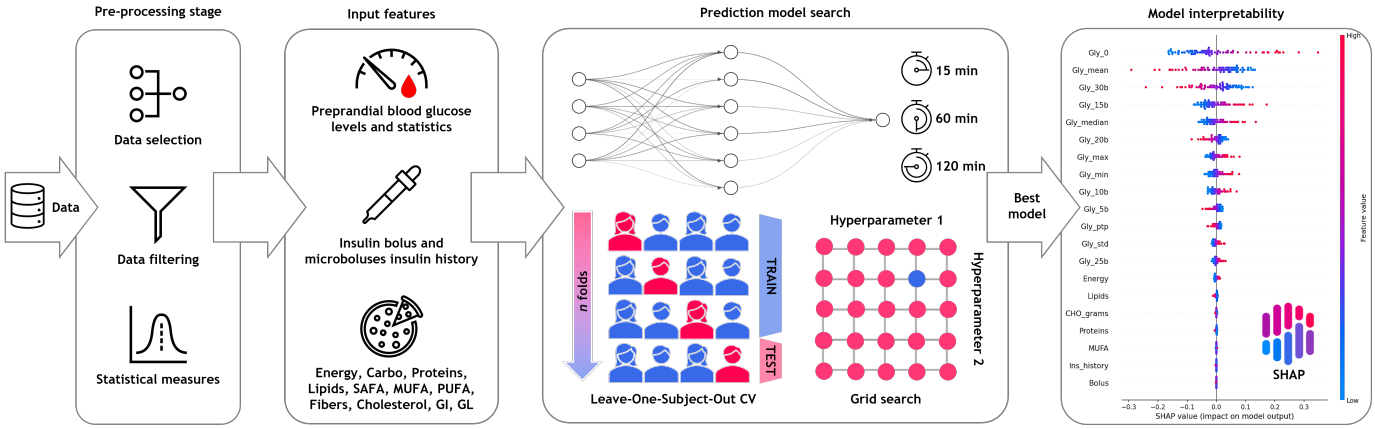

Fig. 1. Pipeline of the proposed approach. Data from 15 T1DM patients are selected from the dataset and filtered to reduce noise. Eight statistical attributes are computed from the preprandial glycemic history and considered input features along with preprandial glycemic values, administered insulin, and meal-related factors. For each of the three PHs, a fully-connected DNN model is optimized via a grid search of hyperparameters and validated using the LOSO-CV strategy. Then, the top-performing model is subjected to the XAI method for output interpretation. Shapley values are computed for all input features to evaluate their influence on the prediction.

tion in AI-powered systems.

In this context, the SHAP method [16] has emerged as rigorous approach and wide-ranging applicability. As a model-agnostic technique, SHAP allows the quantification of each feature's contribution to the prediction process. Notably, SHAP efficiently provides both global and local explanations, unifying disparate approaches such as Local Interpretable Model-agnostic Explanations (LIME) [35], DeepLIFT [41], and Layer-wise Relevance Propagation (LRP) [42]. Furthermore, a significant advantage of SHAP over other XAI techniques is its versatility, as it can be employed in both classification [43] and regression problems [44].

In diabetes management, XAI applications in the current literature mostly relates to predicting early asymptomatic stages of diabetes for early diagnosis [45], [46] or assessing the risks of adverse events [47], [48] in an interpretable manner. As for BGLs prediction, a personalized bidirectional LSTM model equipped with interpretability tools has been proposed in [49] using data from six T1DM patients. Specifically, the Shapley values related to CGM measurements, administered insulin, physical activity, and meal carbohydrates were calculated on the network output at PHs of $30\,\mathrm{min}$ and $60\,\mathrm{min}$. Physical activity had limited effects on the regression performance and was therefore disregarded, while CGM values and carbohydrates showed a positive impact on the predicted glucose value, in contrast to the negative impact of insulin. In [50] a decision tree algorithm has been proposed to predict various characteristics of PGR in women with gestational diabetes. The Shapley values were computed to identify the most significant contributors among CGM measurements, responses from habit surveys, demographic information, and meal-related features. Findings revealed the glycemic load, amount of carbohydrates, type of meal, amount of starch, and food consumed 6 hours before the current meal as the prominent factors. Despite these compelling studies, the application of XAI to AP research remains largely unexplored, primarily attributed to the complexity of the BGL prediction problem and the novelty of the explainability trend in this field.

III. Materials and Methods

This section provides an overview of the dataset used and presents the proposed method. As mentioned in Section I, the objective of this study was to assess the influence of input features on the predictive capability of BGLs. In order to achieve this goal, an XAI methodology was employed. The schematic representation of the proposed method is illustrated in Figure 1.

A. Dataset Description

This study was conducted on the *AI4PG* dataset, provided by the Diabetes Outpatient Clinic of Federico II University Hospital in Naples, Italy [25]. The utilization of this dataset in the present study received the necessary ethical approval from the Ethical Committee of University of Naples Federico II (Registration number 338/20).

The dataset comprises tabular data collected from 25 individuals diagnosed with T1DM equipped with a closed-loop system, the Medtronic MiniMed 670G [51]. Within the patients' group, there were 12 males and 13 females, with an average age of 40 ± 12 years and a duration of diabetes of 15 ± 12 years. In addition to the measured CGM values and the administered insulin doses, the dataset incorporates information regarding the patients dietary habits. More in detail, the CGM values and insulin data were automatically extracted from the closed-loop system, whereas nutritional data were obtained from a 7-day food record that is the golden standard for measuring dietary habits. All foods and drinks consumed (including dressings) were reported providing as much detail as possible (i.e. cooking methods, brands names). Energy intake, nutrient composition, glycaemic index, and glycaemic load were estimated using a validated software (Metadieta) [52]. Each entry within the dataset corresponds to a specific meal, encompassing a range of CGM measurements $(\mathrm{mg/dL})$ capturing pre- and postprandial BGLs. Overall, the dataset comprises 1264 meals, including breakfasts, lunches, and dinners, with CGM glycemia values available at 5-min intervals from $30\,\mathrm{min}$ before the meal to $120\,\mathrm{min}$ after. Multiple

features associated with each meal are estimated and reported, including the intake of energy (kcal), the quantities of carbohydrates (g), proteins (g), lipids (g), fatty acids (g), fibers (g), and cholesterol (mg), as well as the corresponding glycemic index (GI) and glycemic load (GL) values. Furthermore, the dataset provides insights into insulin delivery. Specifically, it includes information on the microboluses administered by the AP, spanning three hours before the meal until mealtime, along with manual boluses (MBs) of insulin delivered at mealtime (in I.U.). The MB insulin dosage was calculated by the patient, considering the quantity of carbohydrates and meal intake.
Participants were not required to provide data regarding their physical activity levels. Nevertheless, over the observation period, no cases of temporary blood glucose targets indicative of physical exercise sessions were observed.

B. Proposed Method

To investigate the influence of specific features on BGLs prediction, three ML-based systems employing Feed Forward Neural Networks (FFNN) [53] to forecast postprandial BGLs in patients individuals diagnosed with T1DM are proposed. Different PHs (15 min, 60 min, and 120 min) after a meal are considered.
The FFNNs took as input features:

- a 30 min window of BGLs (Gly_30b - Gly_0) along with associated statistical attributes such as minimum (Gly_min), maximum (Gly_max), mean (Gly_mean), standard deviation (Gly_std), peak-to-peak difference (Gly_ptp), median (Gly_median), kurtosis (Gly_kurt), and skewness (Gly_skew);
- information regarding the insulin dosages:
  - manually-administered bolus (MB) of insulin ($Bolus$) at mealtime;
  - cumulative sum of microboluses delivered by the closed-loop system worn by patients in the three-hour interval preceding the meal ($Ins\_history$), as an absolute measure of the system's insulin delivery;
- meal-related information:
  - energy intake ($Energy$);
  - carbohydrates ($Carbo$), glycemic index (GI), glycemic load (GL);
  - $Proteins$;
  - $Fibers$;
  - $Lipids$, monounsaturated fatty acids ($MUFA$), polyunsaturated fatty acids ($PUFA$), saturated fatty acid ($SAFA$), $Cholesterol$.

The data were preprocessed by selecting for each patient a number of observations between 30 and 100. Then, the preprandial BGLs were filtered with the Savinsky-Golay technique [54] to smooth out noise and leave the dynamics unchanged. Fully-connected DNNs were exploited to forecast BGLs. In order to determine optimal hyperparameter values of the FFNN at different PHs, a grid search strategy [53] was adopted. Ultimately, three separate models were derived, specifically tailored to predict at distinct PHs: 15 min, 60 min, and 120 min. Each model was validated by using Leave-One-Subject-Out Cross-Validation (LOSO-CV) strategy. Among various validation methods, such as holdout and k-fold, the LOSO-CV approach stands out as one of the most robust validation methods for inter-subject analysis. Indeed, being the test set composed of data belonging to a subject not seen during the training, LOSO-CV provides a reliable evaluation of the model performance on new and unencountered data. Additional details concerning the preprocessing step and experimental setup are reported in Section IV. The assessment of prediction performance for each postprandial PH was carried out through Root Mean Square Error (RMSE), defined by the following equation:

$$RMSE = \sqrt{\frac{1}{N}\sum_{N}(\hat{y}_t - y_t)^2} \tag{1}$$

where $\hat{y}_t$ and y_t represent the predicted and measured BGLs at time instant t, respectively; and N denotes the total number of observations in the dataset.

As the central focus of this study lies in achieving an interpretable decision-support system for T1DM management, special attention was given to addressing the crucial aspect of interpretability. To this purpose, the SHAP technique [16] was employed to analyze the impact of features on the output model.

C. Model Interpretability: SHAP

SHAP [16] is a method to explain individual predictions, providing relevance scores to each input feature. This methodology utilizes Shapley values, derived from coalitional game theory, to attribute the contribution of each feature to the final prediction. The basic principle of SHAP technique is to decompose the model output into the cumulative impacts of individual features. For complex models such as DNNs, the adopted SHAP method (i.e., Kernel SHAP) relies on weighted linear regression to compute the importance of each feature. In particular, it employs a simpler explanation model g, which serves as an interpretable approximation of the original prediction model. Given M input features, the Shapley value explanation $g(z')$ is conceptually derived using a linear additive model [16]:

$$g(z') = \phi_0 + \sum_{j=1}^{M} \phi_j z'_j, \tag{2}$$

where z'_j denotes the presence (1) or absence (0) of the feature j; ϕ_j is the Shapley value representing the relative feature contribution; and ϕ_0 is the base value when all input features are absent (0).
By assigning a value to each input, SHAP allows for a comprehensive understanding of how and to what extent each feature influenced the final prediction. This is achieved through systematic analysis of various feature combinations, calculating their individual effects on predictions when combined with others. In practice, for each combination, certain inputs are held constant at their actual values, while the features under evaluation are randomly perturbed.

To estimate the global relevance of each input on the model's outcome, it is possible to compute the feature importance. Specifically, the absolute Shapley values of feature j

were averaged across the data as follows:

$$I_j = \frac{1}{n} \sum_{i=1}^{n} |\phi_j^{(i)}| \tag{3}$$

where n is the number of instances in the dataset.

One notable characteristic of SHAP is its model-agnostic nature, making it applicable to a wide range of ML models. Additionally, SHAP offers consistent explanations, i.e. provides the same explanation for a given model and dataset, and effectively accommodates complex model behaviors, including interactions among features. This adaptability highlights its effectiveness in capturing intricate relationships within the data.

IV. Experiments

In this section, the conducted experiments are illustrated, together with the preprocessing and the experimental setup adopted. The main steps of the proposed pipeline are reported in Figure 1. After a first data preprocessing step (data selection, filtering, and statistical measures calculation), a Leave-One-Subject-Out Cross-Validation (LOSO-CV) model validation procedure was carried out with a hyperparameter optimization. Finally, found the best model, SHAP technique was exploited for interpretability.

A. Preprocessing Step

A subgroup of 15 subjects was selected from the initial cohort of 25 patients in the *AI4PG* dataset. The selection criterion was a minimum of 30 recorded meals per person. In cases where the number of recorded meals exceeded 100, only the first 100 meals were considered for the analysis. This rigorous selection aimed to ensure an adequate number of data points for reliable and robust analysis, resulting in a final dataset of 1036 meal records from 15 subjects. For each dataset entry, the Savitzky-Golay filtering technique [54] with a first-order polynomial and a 15-step sliding window was employed to smooth the BGLs trends. Specifically, BGL values measured each 5 min from 30 min prior to the meal until the mealtime were utilized as inputs for the FFNN model. Additionally, 8 statistical measures, computed from the preprandial BGLs values for each meal entry, were incorporated as additional input: minimum, maximum, mean, standard deviation, median, peak-to-peak difference, kurtosis, and skewness. In addition, microboluses administered by the AP system in the 3 h before the meal were summed to obtain a single quantity that takes into account the amount of basal insulin. Finally, all features were scaled using min-max scaling strategy (see Section IV-B for more details).

B. Experimental Setup

After the preprocessing, a ML model validation procedure was conducted. As mentioned in Section III-B, FFNNs were employed to predict postprandial BGLs at different PHs (15 min, 60 min, 120 min). In order to identify the optimal hyperparameters for the FFNN models, a grid search strategy was exploited. More specifically, Table II provides the tuned hyperparameters and the search spaces. Relatively shallow neural networks were considered, with a number of layers ranging from 1 to 3. Incorporation of a regularization term, employing weight decay with a penalty on L2 norm, was employed while systematically varying the weight decay values, as illustrated in Table II. The maximum number of epochs was established at 1000, with a patience value of 10 for the stopping criteria, as a usual trade-off between computational complexity and the model's generalization capability [53]. Ultimately, three distinct models were derived, each specifically tailored for predicting BGLs at different PHs.

TABLE II
THE SEARCH SPACE ADOPTED DURING THE GRID SEARCH FOR TUNING HYPERPARAMETERS

Tuned Hyperparameters	Search Space
number of hidden layers	{1, 2, 3}
number of neurons in each layer	{8, 16, 32, 64, 128}
optimization algorithm	{Stochastic Gradient Descent (SGD) [55], Adam [56]}
activation function	{Rectified Linear Unit (ReLU) [57], hyperbolic tangent function (tanh), Variable Activation Function (VAF) [58]}
learning rate	{0.0001, 0.0005, 0.001, 0.005, 0.01}
weight decay parameter (penalty L2)	{0.0001, 0.001, 0.01}

The proposed method was validated using the LOSO-CV strategy to effectively address inter-subject variability. This iterative approach includes training the model on *n-1* subjects, where *n* indicates the total number of subjects in the dataset, and considering the subject excluded during the training process as a test to assess the model. Moreover, in each iteration of the LOSO-CV strategy, all data underwent min-max scaling, taking into account the minimum and maximum values from the training data. The data were scaled to adjust the different scales of the involved features.

Each model was evaluated on the test set by using RMSE as described in Section III-B.

Subsequent to the identification of the best-performing models at 15 min, 60 min, and 120 min after the meal, the FFNNs were retrained, wherein data from the patient with the best RMSE was excluded to form the test set. Finally, the interpretation of the models' output was accomplished through the utilization of the SHAP, described in Section III-C.

V. Results

A. Prediction Results

The model's performance was evaluated by computing the RMSE between the real and predicted BGLs generated by the FFNNs at different PH after the meal (15 min, 60 min, 120 min). The mean and standard deviation of the RMSEs are summarized in Table III, alongside the selection of the best model. The results were obtained by averaging the RMSEs across n folds of the LOSO-CV strategy, where n represents the number of patients considered, in this case, 15.

Final performance in terms of RMSE was found to be comparable to previous literature (e.g. [21], [23], [27], [31] in Table I). However, it is essential to acknowledge that

TABLE III
PERFORMANCE AT DIFFERENT PREDICTION HORIZON (PH) OF THE PROPOSED FFNN MODELS BY MEANS OF ROOT MEAN SQUARED ERROR (RMSE) AND SELECTED HYPERPARAMETERS VALUES.

PH (min)	**RMSE** (mg/dL) **(mean ± std)**	**Selected Hyperparameters**
15	2.53 ± 0.43	number of hidden layers = 2 number of neurons = {32,16} optimization algorithm = Adam activation function = tanh learning rate = 0.0001 L2 penalty = 0.001 batch size = 32
60	24.74 ± 4.27	number of hidden layers = 3 number of neurons = {32,16,8} optimization algorithm = Adam activation function = tanh learning rate = 0.0001 L2 penalty = 0.01 batch size = 32
120	50.15 ± 7.70	number of hidden layers = 3 number of neurons = {32,16,8} optimization algorithm = Adam activation function = tanh learning rate = 0.0001 L2 penalty = 0.01 batch size = 16

differences in data utilization, experimental conditions, preprocessing techniques, and hyperparameter configurations among various studies may influence the outcomes, making an unbiased comparison of the proposed methods challenging.

Moreover, unlike other studies in the literature that often employed hold-out or k-fold cross-validation strategies [10], [20], [23], [59], [60], this investigation opted for the LOSO-CV approach. LOSO-CV is considered one of the most reliable validation methods for inter-subject analysis, as it accounts for the variability among subjects.

B. Interpretability Results

As discussed in Section I, the aim of this study is to quantify the impact of input features, especially meal-related features, on postprandial BGLs prediction by using XAI methodologies. Specifically, the SHAP technique (see Section III-C) was employed to interpret each trained FFNNs.

First, feature importance analysis was conducted for assessing the global influence of each specific input feature on the models' output. Specifically, following Eq. 3, the absolute Shapley values of feature j were averaged across the data of a single patient. Subsequently, the average importance of each feature was computed over 15 patients. The calculated feature importance for each PH are presented in Figure 2, providing a concise visualization of the average contribution of individual features to the model output. The length of each bar reflects the mean of the absolute Shapley values per feature across all data, helping identify the most influential factors. Higher feature importance values for a feature correspond to heightened impacts on increasing or decreasing the final prediction. In Figure 2.a, it can be observed that the most influential features are those related to glycemia, with the exception of *Gly_kurt* and *Gly_skew*, which appear negligible, along with the marginal impact of administered insulin and nutritional factors. The situation begins to change at $PH = 60\,\mathrm{min}$ in Figure 2.b, with an increasing importance of certain factors such as *Carbo*, *Lipids*, *Energy*, *Bolus*, and *GL*. Finally, in Figure 2.c, there is a further decrease in the effect of glycemic features and an increase in meal-related factors. The main exception is represented by *Gly_25b*, which remains highly impactful at $PH = 120\,\mathrm{min}$.

However, feature importance does not allow a thorough understanding of the features' effects. In order to evaluate the relationship between the value of the input variable and the prediction output, summary plots have been exploited and reported in Figure 3. This representation takes advantage of the importance of features to order them in a descending manner, and at the same time shows the relationship between the value of the input variable and its impact on prediction. Specifically, summary plots of Shapley values, as shown in Figure 3 for a single subject, provide a global perspective on feature importance and its underlying drivers by showing the distribution of individual feature contributions. As observed, certain glycemic features (i.e., *Gly_0*) exhibit a positive correlation for all the PHs, wherein higher input values correspond to higher predictions. Conversely, an inverse pattern is noticed for *Gly_30b*, wherein higher input values lead to lower predictions, and lower input values result in higher predictions. In Figure 3.a, *Carbo* does not appear as a row because its influence at $PH = 15\,\mathrm{min}$ is negligible, just like the other meal-related factors at the bottom. On the other hand, as already observed from the feature importance plots, it becomes essential to take into account the meal-related factors as the PH extends. In Figures 3.b and 3.c, it is evident that at lower values of *Carbo*, the impact is concentrated on very low negative SHAP values. However, as the feature value grows, the intensity of the impact towards positive SHAP values also grows, resulting in higher predicted outputs. Interestingly, *Bolus* and *GL* exhibit the opposite behavior.

Ultimately, Shapley correlation matrices were employed for gaining deeper insights into feature interplay. For the sake of brevity, Shapley correlation matrices for a single subject are presented in Figure 4. These grids reveal intricate feature relationships by calculating correlations between Shapley values for pairwise feature combinations. Unlike standard correlation matrices computed on input features, Shapley value correlations consider individual feature effects on the prediction, thus exposing potential counter-intuitive relationships between features. The heatmaps in Figure 4 illustrate patterns of alignment (red) or opposition (blue) in feature contributions on the output for all three PHs on the best fold. Notably, the intensity of values in the top-left area of the heatmaps indicates strong correlations among features related to glycemic history. Specifically, Figure 4.b and 4.c demonstrate consistency between the effects of *Gly_30b* and *Gly_25b*, while the behavior changes from *Gly_20b* to *Gly_0*. Interestingly, the effects of nutritional factors related to meals show limited correlations with glycemia. Stronger SHAP correlations are observed between

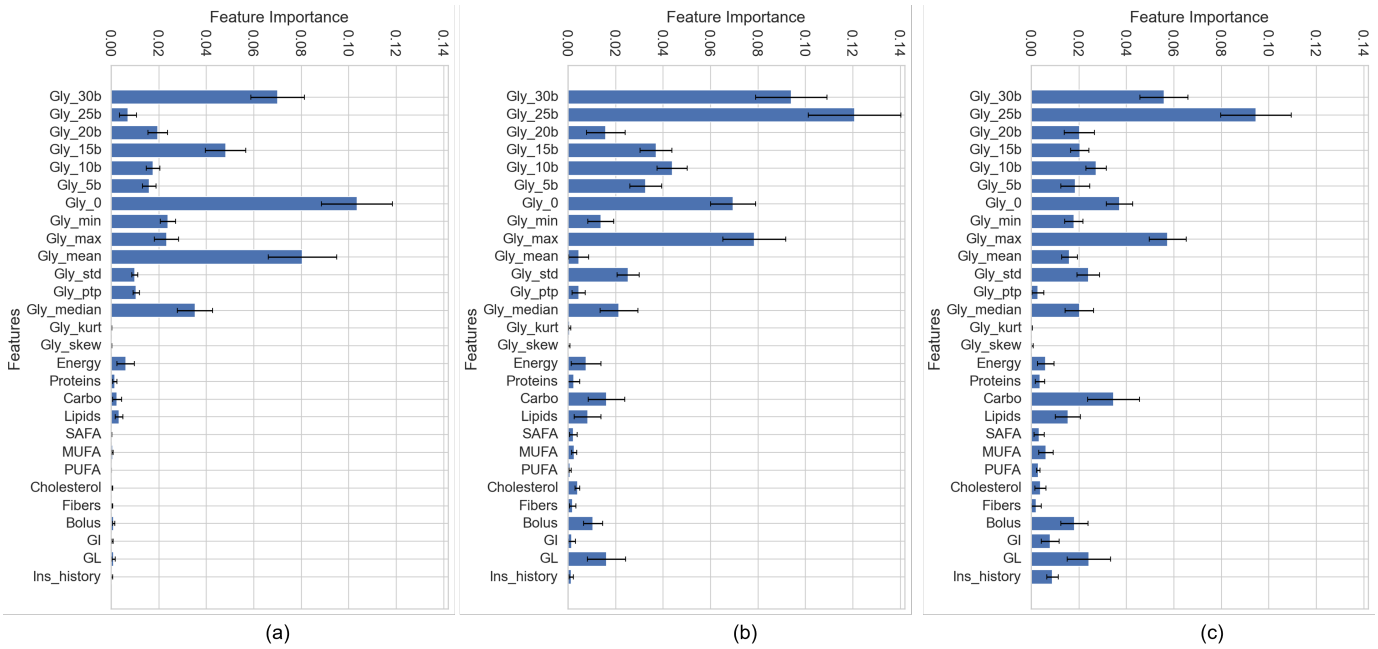

Fig. 2. Boxplots for feature importance analysis at (a) $15\,\mathrm{min}$, (b) $60\,\mathrm{min}$, and (c) $120\,\mathrm{min}$ after mealtime. Each feature bar corresponds to the average importance computed over 15 subjects, and the error bars indicate the associated standard deviation.
Before the meal: Gly 30b to Gly 5b - glycemia history from 30 min to 5 min before the meal; Gly median - median of glycemia history from 30 min before meal to mealtime; Gly mean - mean of glycemia history; Gly std - standard deviation of glycemia history; Gly ptp - peak-to-peak difference of glycemia history; Gly min - minimum of glycemia history; Gly max - maximum of glycemia history; Gly kurt - kurtosis of glycemia history; Gly skew - skewness of glycemia history; Ins history - sum of microboluses delivered by HCLS within the three-hour period preceding the meal. At mealtime: Gly 0 - glycemia value; Bolus - manual bolus of insulin. Meal-related: GL - glycemic load; Carbo - grams of carbohydrates; GI - glycemic index; Proteins - grams of proteins; Lipids - grams of lipids; SAFA - grams of saturated fatty acid; MUFA - grams of mono-unsaturated fatty acids; PUFA - grams of poly-unsaturated fatty acid; Cholesterol - milligrams of cholesterol; Fibers - grams of fibers; Energy - kcal of energy.

Energy and meal nutritional components, such as *Carbo* and *Lipids*. At $PH = 120\,\mathrm{min}$, these correlation values demonstrate increased strength, as evidenced by the intensification of values in the bottom-right area of the heatmap. Additionally, the effect of insulin administered before the meal (*Ins_history*) is partially correlated with glycemic history (from *Gly_30b* to *Gly_median*), probably given its computation based on CGM measurements.

VI. Discussion

The present study evaluated the performance of the proposed FFNN models for predicting BGLs. The models' RMSE performance, as depicted in Table III, demonstrated a level of accuracy that was comparable to the findings reported in Table I of previous studies. However, it is crucial to acknowledge that direct comparisons between different studies should be interpreted with caution. Several factors can influence the outcomes, including variations in data sources, experimental conditions, and preprocessing methods. Nevertheless, the results demonstrated that the FFNNs effectively predicted BGLs and exhibited performance comparable to state-of-the-art approaches.

This investigation used the LOSO-CV approach to account for inter-subject variability, by using each individual subject's data as an independent validation set while training on the remaining subjects' data. By doing so, the model can better generalize to new, unseen data and minimize the risk of overfitting. As a result, the proposed model offers valuable insights, showing promising performance and generalization capability in blood glucose prediction.

As stated in Section I, the primary objective of this study was to assess the impact of different features, particularly those associated with meals, on BGL predictions using a post-hoc XAI. Indeed, the significance of nutritional factors on PGR as reported in the literature [61] can lead to the hypothesis that these nutritional factors may have an impact on BGL predictions, although several aspects remain unclear, such as the extent to which these factors are subject dependent. Consequently, in the proposed approach, feature selection (FS) before training was intentionally excluded. As a matter of fact, FS would have restricted the exploration, as our focus was on using XAI to unveil the impact of features. To retrospectively evaluate each feature's contribution and expand understanding without preconceived notions, SHAP method was employed. The study reports interpretability results in the form of feature importance (Figure 2) and summary plots (Figure 3), allowing for a comprehensive understanding of the model's predictive factors.

Figures 2.a and 3.a show feature importance for $PH = 15\,\mathrm{min}$ after the meal. As observed, the influence of $Bolus$ and the majority of meal-related features on the postprandial predicted BGLs was found to be negligible. However, the preprandial glycemia history within the $30\,\mathrm{min}$ window had a significant impact on the postprandial glucose response (PGR). Specifically, higher glycemia values at mealtime (Gly_0) were associated with higher predicted BGLs, showing a high cor-

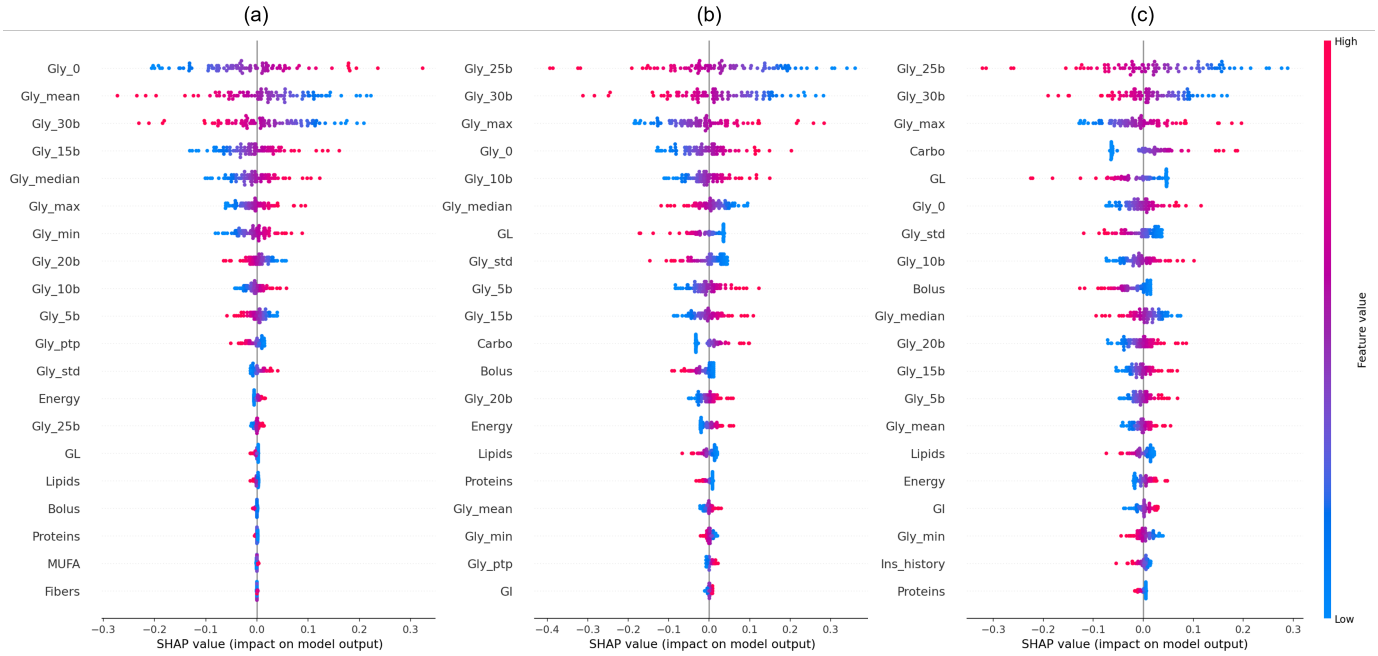

Fig. 3. Best-fold summary plots of SHAP values at (a) $15\,\mathrm{min}$, (b) $60\,\mathrm{min}$, and (c) $120\,\mathrm{min}$ after mealtime for a single subject. Features are listed on the vertical axis in descending order of importance, with the most significant factors at the top, while the horizontal axis illustrates the variation of the Shapley values. For a single observation, the influence of a feature is visualized through a point along the relative row, whereby the color of the point corresponds to its feature value, and its position is determined by its positive or negative Shapley value. As the influence of the feature on the output strengthens, the point progressively moves away from the gray vertical line, which represents zero impact.
Before the meal: Gly 30b to Gly 5b - glycemia history from 30 min to 5 min before the meal; Gly median - median of glycemia history from 30 min before meal to mealtime; Gly mean - mean of glycemia history; Gly std - standard deviation of glycemia history; Gly ptp - peak-to-peak difference of glycemia history; Gly min - minimum of glycemia history; Gly max - maximum of glycemia history; Gly kurt - kurtosis of glycemia history; Gly skew - skewness of glycemia history; Ins history - sum of microboluses delivered by HCLS within the three-hour period preceding the meal. At mealtime: Gly 0 - glycemia value; Bolus - manual bolus of insulin. Meal-related: GL - glycemic load; Carbo - grams of carbohydrates; GI - glycemic index; Proteins - grams of proteins; Lipids - grams of lipids; SAFA - grams of saturated fatty acid; MUFA - grams of mono-unsaturated fatty acids; PUFA - grams of poly-unsaturated fatty acid; Cholesterol - milligrams of cholesterol; Fibers - grams of fibers; Energy - kcal of energy.

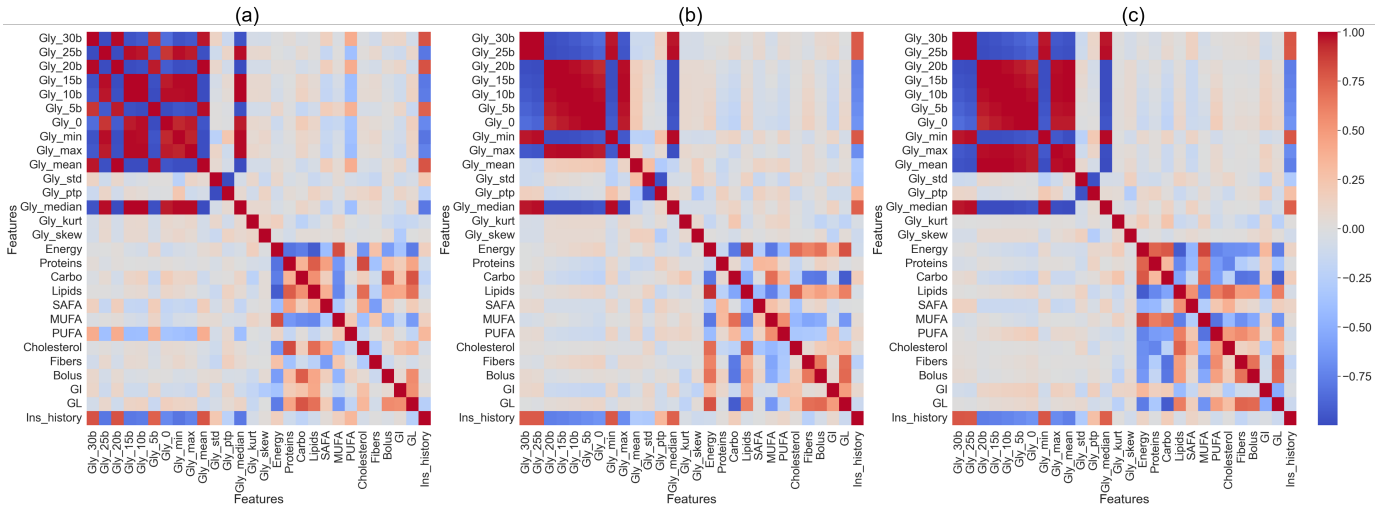

Fig. 4. Best-fold correlation matrices between SHAP values corresponding to each feature pair at (a) $15\,\mathrm{min}$, (b) $60\,\mathrm{min}$, and (c) $120\,\mathrm{min}$ after mealtime for a single subject. The intensity of red denotes a strong positive correlation, while the intensity of blue signifies a strong negative correlation between effects.

relation between the preceding and immediately following BGLs. This observation may align with physiological dynamics in real-life scenarios, wherein, within $15\,\mathrm{min}$ from the meal, the effects of *Bolus* and meal intake may not have fully manifested. This can be attributed to the time required for insulin and nutritional factors to circulate in the body and the inherent delays in glucose readings by the CGM device [62], [63]. As a matter of fact, clinical studies have shown that although insulin levels peak within $40\,\mathrm{min}$ to $60\,\mathrm{min}$ after injection, the maximum insulin action is observed

approximately $100\,\mathrm{min}$ after injection [63], [64].

Figures 2.b and 3.b illustrate the results obtained by using SHAP technique for a PH of $60\,\mathrm{min}$ after the meal. In this case, the $Carbo$ intake of the meal, as well as GL, appear to hold greater significance. Specifically, postprandial blood glucose is notably influenced by GL. Serving as a product of GI and $Carbo$, GL enables the simultaneous description of both the quality and quantity of carbohydrates in a meal. Considering that the impact on blood glucose is affected not only by the quantity but also the quality of carbohydrates consumed, the higher impact of GL seems appropriate, as it accounts for both aspects [65]. Furthermore, higher values of $Carbo$ lead to an increase in the predicted BGLs, accurately capturing the impact of the meal on glycemic dynamics. For $PH = 60\,\mathrm{min}$, $Bolus$ exhibits a greater impact on BGL prediction, aligning with the actual influence of insulin on glycemic levels, as it typically begins to take effect approximately $60\,\mathrm{min}$ after injection [63]. More in detail, an increase in $Bolus$ values at mealtime exerts an adverse effect on the predicted BGLs (see Figure 3.b), aligning with the well-known influence of insulin, which lowers glycemia values. Furthermore, the findings suggest that nutritional factors, in general, have a more significant impact on BGL prediction after $1\,\mathrm{h}$ from the meal. This observation can be interpreted as further confirmation that BGL values are influenced by nutritional factors over the medium term, consistent with previous research [17].

Finally, Figures 2.c and 3.c, show SHAP results for $PH = 120\,\mathrm{min}$. As can be observed, in this instance, the importance of the carbohydrate intake $Carbo$ appears to be more significant. However, the nutritional factor driving better blood glucose prediction at 2 h becomes GL, for the observed relative increased predictive value of GI. The predictive power of lipids intake $Lipids$ also becomes relevant, which is expected to influence the late postprandial blood glucose response in people with T1DM [18], [66].

The obtained results contribute to advancing the transition from physiological knowledge to clinical practice. This represents a critical step toward a better understanding of PGR determinants in individuals with T1DM and could offer valuable implications for advancing AP technology and developing decision-support tools for T1DM patients. Some aspects, however, still need to be addressed. First, the ML models used strongly depend on training data, and the availability of high-quality data is essential to get accurate predictions. One of the biggest challenges faced in this study was the limited public availability of real-world data, especially data that included detailed meal-related information beyond just carbohydrates. Most studies in the literature [9], [67] use synthetic data (i.e. UVA/Padova simulator [68]), or real public datasets (i.e. Ohio dataset [69], DirectNet [70]) that provide only blood sugar or only the amount of carbohydrates and insulin bolus as meal-related information. As our primary objective was to explore the impact of various nutritional factors on blood glucose, it was necessary to collect data with an experimental campaign. This is not an easy task in practical AP applications, since it would pose a heavy burden on the patients. However, it is worth noting that ongoing research to develop applications for automatic identification of food composition [71], [72] could help gather information on nutrients. Another aspect that should be addressed relates the input features used for the prediction of postprandial glycemic levels: in fact, these have been selected based on theoretical and physiological considerations [23], [61], and available data, but may not capture all relevant variables. Indeed, beyond nutritional factors, it would be of considerable interest to explore the influence of variables associated with physical activity and the psychological well-being of patients [73], [74]. In this regard, XAI-based features impact findings could be used to carry out feature selection to improve model performance by identifying the most influential input features [75]–[77]. Thirdly, in this study, PH until 120 min after the meal was explored, and it seems sufficient to start disentangling the different times at which the various nutrients exerted their influence on blood glucose. However, extending the postprandial period up to six hours would be an interesting avenue for investigation; this would make it more likely to capture later postprandial events such as hypoglycemia and hyperglycemia [61], [78].

VII. Conclusion

This study focused on addressing the challenges of managing PGR for individuals with T1DM through the use of DNN models. Despite the effectiveness of current AP technology in integrating basal insulin delivery and glucose monitoring, it falls short in managing PGR due to an incomplete understanding of its determinants. Consequently, this research aimed to quantify the influence of various input features on predicting postprandial BGLs at different time intervals after meals. By incorporating preprandial glycemic history, insulin dosage, and a range of nutritional factors as input variables, the models showed satisfactory performance in predicting glucose levels. Although these findings may not be immediately applicable due to current limitations, this study holds the potential to advance our understanding of the factors that influence postprandial glucose response in individuals with Type 1 Diabetes (T1DM) and the development of Artificial Pancreas (AP) technology.

To enhance interpretability, the SHAP explainability method was utilized, revealing the significant influence of meal-related factors like carbohydrates, lipids, and glycemic load on postprandial glucose levels up to two hours after a meal. By bridging the gap in our understanding of PGR determinants, this research contributes to the advancement of T1DM care, offering a deeper comprehension of PGR determinants. It holds the potential to offer valuable insights for the advancement of AP technology and the development of more effective and personalized approaches to diabetes management.

References

[1] N. A. ElSayed, G. Aleppo, V. R. Aroda, R. R. Bannuru, F. M. Brown, D. Bruemmer, B. S. Collins, M. E. Hilliard, D. Isaacs, E. L. Johnson, *et al.*, "2. classification and diagnosis of diabetes: standards of care in diabetes—2023," *Diabetes care*, vol. 46, no. Supplement_1, pp. S19–S40, 2023.

[2] J. Kaur and E. R. Seaquist, "Hypoglycaemia in type 1 diabetes mellitus: risks and practical prevention strategies," *Nature Reviews Endocrinology*, vol. 19, no. 3, pp. 177–186, 2023.

[3] M. Nwokolo and R. Hovorka, "The artificial pancreas and type 1 diabetes," *The Journal of Clinical Endocrinology & Metabolism*, vol. 108, no. 7, pp. 1614–1623, 2023.

[4] R. Lupoli, F. Pisano, and B. Capaldo, "Postprandial glucose control in type 1 diabetes: Importance of the gastric emptying rate," *Nutrients*, vol. 11, no. 7, p. 1559, 2019.

[5] T. Zhu, C. Uduku, K. Li, P. Herrero, N. Oliver, and P. Georgiou, "Enhancing self-management in type 1 diabetes with wearables and deep learning," *npj Digital Medicine*, vol. 5, no. 1, p. 78, 2022.

[6] T. Zhu, K. Li, P. Herrero, and P. Georgiou, "Deep learning for diabetes: a systematic review," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 7, pp. 2744–2757, 2020.

[7] I. Contreras and J. Vehi, "Artificial intelligence for diabetes management and decision support: literature review," *Journal of medical Internet research*, vol. 20, no. 5, p. e10775, 2018.

[8] I. Dankwa-Mullan, M. Rivo, M. Sepulveda, Y. Park, J. Snowdon, and K. Rhee, "Transforming diabetes care through artificial intelligence: the future is here," *Population health management*, vol. 22, no. 3, pp. 229–242, 2019.

[9] M. De Bois, M. A. E. Yacoubi, and M. Ammi, "Glyfe: review and benchmark of personalized glucose predictive models in type 1 diabetes," *Medical & Biological Engineering & Computing*, vol. 60, no. 1, pp. 1–17, 2022.

[10] G. Annuzzi, A. Apicella, P. Arpaia, L. Bozzetto, S. Criscuolo, E. De Benedetto, M. Pesola, R. Prevete, and E. Vallefuoco, "Impact of nutritional factors in blood glucose prediction in type 1 diabetes through machine learning," *IEEE Access*, vol. 11, pp. 17104–17115, 2023.

[11] R. A. Karim, I. Vassányi, and I. Kósa, "After-meal blood glucose level prediction using an absorption model for neural network training," *Computers in Biology and Medicine*, vol. 125, p. 103956, 2020.

[12] A. M. Antoniadi, Y. Du, Y. Guendouz, L. Wei, C. Mazo, B. A. Becker, and C. Mooney, "Current challenges and future opportunities for xai in machine learning-based clinical decision support systems: a systematic review," *Applied Sciences*, vol. 11, no. 11, p. 5088, 2021.

[13] T. Miller, "Explanation in artificial intelligence: Insights from the social sciences," *Artificial intelligence*, vol. 267, pp. 1–38, 2019.

[14] D. Minh, H. X. Wang, Y. F. Li, and T. N. Nguyen, "Explainable artificial intelligence: a comprehensive review," *Artificial Intelligence Review*, pp. 1–66, 2022.

[15] P. P. Angelov, E. A. Soares, R. Jiang, N. I. Arnold, and P. M. Atkinson, "Explainable artificial intelligence: an analytical review," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 11, no. 5, p. e1424, 2021.

[16] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *Advances in neural information processing systems*, vol. 30, 2017.

[17] C. Vetrani, I. Calabrese, L. Cavagnuolo, D. Pacella, E. Napolano, S. Di Rienzo, G. Riccardi, A. A. Rivellese, G. Annuzzi, and L. Bozzetto, "Dietary determinants of postprandial blood glucose control in adults with type 1 diabetes on a hybrid closed-loop system," *Diabetologia*, vol. 65, no. 1, pp. 79–87, 2022.

[18] K. J. Bell, C. E. Smart, G. M. Steil, J. C. Brand-Miller, B. King, and H. A. Wolpert, "Impact of fat, protein, and glycemic index on postprandial glucose control in type 1 diabetes: implications for intensive diabetes management in the continuous glucose monitoring era," *Diabetes care*, vol. 38, no. 6, pp. 1008–1015, 2015.

[19] E. Papakonstantinou, C. Oikonomou, G. Nychas, and G. D. Dimitriadis, "Effects of diet, lifestyle, chrononutrition and alternative dietary interventions on postprandial glycemia and insulin resistance," *Nutrients*, vol. 14, no. 4, p. 823, 2022.

[20] C. Pérez-Gandía, A. Facchinetti, G. Sparacino, C. Cobelli, E. Gómez, M. Rigla, A. de Leiva, and M. Hernando, "Artificial neural network algorithm for online glucose prediction from continuous glucose monitoring," *Diabetes technology & therapeutics*, vol. 12, no. 1, pp. 81–88, 2010.

[21] J. Daniels, P. Herrero, and P. Georgiou, "A multitask learning approach to personalized blood glucose prediction," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 1, pp. 436–445, 2021.

[22] G. Annuzzi, L. Bozzetto, A. Cataldo, S. Criscuolo, and M. Pesola, "Predicting and monitoring blood glucose through nutritional factors in type 1 diabetes by artificial neural networks," *Acta IMEKO*, vol. 12, no. 2, pp. 1–7, 2023.

[23] G. Alfian, M. Syafrudin, M. Anshari, F. Benes, F. T. D. Atmaji, I. Fahrurrozi, A. F. Hidayatullah, and J. Rhee, "Blood glucose prediction model for type 1 diabetes based on artificial neural network with time-domain features," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 4, pp. 1586–1599, 2020.

[24] S. L. Cichosz, T. Kronborg, M. H. Jensen, and O. Hejlesen, "Penalty weighted glucose prediction models could lead to better clinically usage," *Computers in Biology and Medicine*, vol. 138, p. 104865, 2021.

[25] L. Angrisani, G. Annuzzi, P. Arpaia, L. Bozzetto, A. Cataldo, A. Corrado, E. De Benedetto, V. Di Capua, R. Prevete, and E. Vallefuoco, "Neural network-based prediction and monitoring of blood glucose response to nutritional factors in type-1 diabetes," in *2022 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, pp. 1–6, IEEE, 2022.

[26] J. Carrillo-Moreno, C. Pérez-Gandía, R. Sendra-Arranz, G. García-Sáez, M. E. Hernando, and Á. Gutiérrez, "Long short-term memory neural network for glucose prediction," *Neural Computing and Applications*, vol. 33, pp. 4191–4203, 2021.

[27] M. Jaloli and M. Cescon, "Long-term prediction of blood glucose levels in type 1 diabetes using a cnn-lstm-based deep neural network," *Journal of Diabetes Science and Technology*, p. 19322968221092785, 2022.

[28] S. Mirshekarian, R. Bunescu, C. Marling, and F. Schwartz, "Using lstms to learn physiological models of blood glucose behavior," in *2017 39th Annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, pp. 2887–2891, IEEE, 2017.

[29] 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)*, pp. 1–5, IEEE, 2018.

[30] A. Aliberti, I. Pupillo, S. Terna, E. Macii, S. Di Cataldo, E. Patti, and A. Acquaviva, "A multi-patient data-driven approach to blood glucose prediction," *IEEE Access*, vol. 7, pp. 69311–69325, 2019.

[31] K. Li, C. Liu, T. Zhu, P. Herrero, and P. Georgiou, "Glunet: A deep learning framework for accurate glucose forecasting," *IEEE journal of biomedical and health informatics*, vol. 24, no. 2, pp. 414–423, 2019.

[32] K. Li, J. Daniels, C. Liu, P. Herrero, and P. Georgiou, "Convolutional recurrent neural networks for glucose prediction," *IEEE journal of biomedical and health informatics*, vol. 24, no. 2, pp. 603–613, 2019.

[33] P. Linardatos, V. Papastefanopoulos, and S. Kotsiantis, "Explainable ai: A review of machine learning interpretability methods," *Entropy*, vol. 23, no. 1, p. 18, 2020.

[34] J. Allgaier, L. Mulansky, R. L. Draelos, and R. Pryss, "How does the model make predictions? a systematic literature review on the explainability power of machine learning in healthcare," *Artificial Intelligence in Medicine*, vol. 143, p. 102616, 2023.

[35] M. T. Ribeiro, S. Singh, and C. Guestrin, """ why should i trust you?" explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 1135–1144, 2016.

[36] A. Apicella, F. Isgrò, R. Prevete, A. Sorrentino, and G. Tamburrini, "Explaining classification systems using sparse dictionaries," *ESANN 2019 - Proceedings, 27th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, p. 495 – 500, 2019.

[37] A. Apicella, S. Giugliano, F. Isgrò, and R. Prevete, "Exploiting auto-encoders and segmentation methods for middle-level explanations of image classification systems," *Knowledge-Based Systems*, vol. 255, p. 109725, 2022.

[38] D. Croce, D. Rossini, and R. Basili, "Auditing deep learning processes through kernel-based explanatory models," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 4037–4046, 2019.

[39] K. Qian, M. Danilevsky, Y. Katsis, B. Kawas, E. Oduor, L. Popa, and Y. Li, "Xnlp: A living survey for xai research in natural language processing," in *26th International Conference on Intelligent User Interfaces-Companion*, pp. 78–80, 2021.

[40] A. Apicella, F. Isgrò, A. Pollastro, and R. Prevete, "Toward the application of xai methods in eeg-based systems," in *XAI.it 2022 - Italian Workshop on Explainable Artificial Intelligence (AIxIA)*, 2022.

[41] A. Shrikumar, P. Greenside, and A. Kundaje, "Learning important features through propagating activation differences," in *International conference on machine learning*, pp. 3145–3153, PMLR, 2017.

[42] G. Montavon, A. Binder, S. Lapuschkin, W. Samek, and K.-R. Müller, "Layer-wise relevance propagation: an overview," *Explainable AI: interpreting, explaining and visualizing deep learning*, pp. 193–209, 2019.

[43] S. Knapič, A. Malhi, R. Saluja, and K. Främling, "Explainable artificial intelligence for human decision support system in the medical domain,"

Machine Learning and Knowledge Extraction, vol. 3, no. 3, pp. 740–770, 2021.

[44] S. Letzgus, P. Wagner, J. Lederer, W. Samek, K.-R. Müller, and G. Montavon, "Toward explainable artificial intelligence for regression models: A methodological perspective," *IEEE Signal Processing Magazine*, vol. 39, no. 4, pp. 40–58, 2022.

[45] C. J. Ejiyi, Z. Qin, J. Amos, M. B. Ejiyi, A. Nnani, T. U. Ejiyi, V. K. Agbesi, C. Diokpo, and C. Okpara, "A robust predictive diagnosis model for diabetes mellitus using shapley-incorporated machine learning algorithms," *Healthcare Analytics*, vol. 3, p. 100166, 2023.

[46] I. Tasin, T. U. Nabil, S. Islam, and R. Khan, "Diabetes prediction using machine learning and explainable ai techniques," *Healthcare Technology Letters*, vol. 10, no. 1-2, pp. 1–10, 2023.

[47] C. Mosquera-Lopez, K. L. Ramsey, V. Roquemen-Echeverri, and P. G. Jacobs, "Modeling risk of hypoglycemia during and following physical activity in people with type 1 diabetes using explainable mixed-effects machine learning," *Computers in Biology and Medicine*, vol. 155, p. 106670, 2023.

[48] A. Della Cioppa, I. De Falco, T. Koutny, U. Scafuri, M. Ubl, and E. Tarantino, "Reducing high-risk glucose forecasting errors by evolving interpretable models for type 1 diabetes," *Applied Soft Computing*, vol. 134, p. 110012, 2023.

[49] G. Cappon, L. Meneghetti, F. Prendin, J. Pavan, G. Sparacino, S. Del Favero, A. Facchinetti, *et al.*, "A personalized and interpretable deep learning based approach to predict blood glucose concentration in type 1 diabetes.," in *KDH@ ECAI*, pp. 75–79, 2020.

[50] E. A. Pustozerov, A. S. Tkachuk, E. A. Vasukova, A. D. Anopova, M. A. Kokina, I. V. Gorelova, T. M. Pervunina, E. N. Grineva, and P. V. Popova, "Machine learning approach for postprandial blood glucose prediction in gestational diabetes mellitus," *IEEE Access*, vol. 8, pp. 219308–219321, 2020.

[51] Medtronic, "MiniMed 670G; Medtronic MiniMed, USA." `https://www.medtronicdiabetes.com/customer-support/minimed-670g-system-support`. Available online July 20, 2023.

[52] "Metadieta software." `https://www.metadieta.it/`. [Online; accessed 17-October-2023].

[53] C. M. Bishop and N. M. Nasrabadi, *Pattern recognition and machine learning*, vol. 4. Springer, 2006.

[54] A. Savitzky and M. J. Golay, "Smoothing and differentiation of data by simplified least squares procedures.," *Analytical chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964.

[55] Y. Goldberg, "Neural network methods for natural language processing," *Synthesis lectures on human language technologies*, vol. 10, no. 1, pp. 1–309, 2017.

[56] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[57] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th international conference on machine learning (ICML-10)*, pp. 807–814, 2010.

[58] A. Apicella, F. Isgrò, and R. Prevete, "A simple and efficient architecture for trainable activation functions," *Neurocomputing*, vol. 370, pp. 1–15, 2019.

[59] S. M. Pappada, B. D. Cameron, P. M. Rosman, R. E. Bourey, T. J. Papadimos, W. Olorunto, and M. J. Borst, "Neural network-based real-time prediction of glucose in patients with insulin-dependent diabetes," *Diabetes technology & therapeutics*, vol. 13, no. 2, pp. 135–141, 2011.

[60] C. Zecchin, A. Facchinetti, G. Sparacino, G. De Nicolao, and C. Cobelli, "Neural network incorporating meal information improves accuracy of short-time prediction of glucose concentration," *IEEE transactions on biomedical engineering*, vol. 59, no. 6, pp. 1550–1560, 2012.

[61] L. Bozzetto, D. Pacella, L. Cavagnuolo, M. Capuano, A. Corrado, G. Scidà, G. Costabile, A. Rivellese, and G. Annuzzi, "Postprandial glucose variability in type 1 diabetes: The individual matters beyond the meal," *Diabetes Research and Clinical Practice*, vol. 192, p. 110089, 2022.

[62] M. S. Boyne, D. M. Silver, J. Kaplan, and C. D. Saudek, "Timing of changes in interstitial and venous blood glucose measured with a continuous subcutaneous glucose sensor," *Diabetes*, vol. 52, no. 11, pp. 2790–2794, 2003.

[63] D. Slattery, S. Amiel, and P. Choudhary, "Optimal prandial timing of bolus insulin in diabetes management: a review," *Diabetic Medicine*, vol. 35, no. 3, pp. 306–316, 2018.

[64] R. Basu, C. Dalla Man, M. Campioni, A. Basu, G. Klee, G. Toffolo, C. Cobelli, and R. A. Rizza, "Effects of age and sex on postprandial glucose metabolism: differences in glucose turnover, insulin secretion, insulin action, and hepatic insulin extraction," *Diabetes*, vol. 55, no. 7, pp. 2001–2014, 2006.

[65] L. Bozzetto, M. Giorgini, A. Alderisio, L. Costagliola, A. Giacco, G. Riccardi, A. A. Rivellese, and G. Annuzzi, "Glycaemic load versus carbohydrate counting for insulin bolus calculation in patients with type 1 diabetes on insulin pump," *Acta diabetologica*, vol. 52, pp. 865–871, 2015.

[66] L. Bozzetto, A. Alderisio, M. Giorgini, F. Barone, A. Giacco, G. Riccardi, A. A. Rivellese, and G. Annuzzi, "Extra-virgin olive oil reduces glycemic response to a high–glycemic index meal in patients with type 1 diabetes: a randomized controlled trial," *Diabetes Care*, vol. 39, no. 4, pp. 518–524, 2016.

[67] V. Felizardo, N. M. Garcia, N. Pombo, and I. Megdiche, "Data-based algorithms and models using diabetics real data for blood glucose and hypoglycaemia prediction–a systematic literature review," *Artificial Intelligence in Medicine*, vol. 118, p. 102120, 2021.

[68] "UVA/Padova simulator." `https://tegvirginia.com/software/t1dms-2014/`, 2014. [Online; accessed 11-October-2023].

[69] "Ohio dataset." `http://smarthealth.cs.ohio.edu/OhioT1DM-dataset.htm`. [Online; accessed 11-October-2023].

[70] "DirectNet dataset." `https://public.jaeb.org/direcnet/stdy/167`. [Online; accessed 11-October-2023].

[71] G. Ferrara, J. Kim, S. Lin, J. Hua, E. Seto, *et al.*, "A focused review of smartphone diet-tracking apps: usability, functionality, coherence with behavior change theory, and comparative validity of nutrient intake and energy estimates," *JMIR mHealth and uHealth*, vol. 7, no. 5, p. e9232, 2019.

[72] S. Gioia, I. M. Vlasac, D. Babazadeh, N. L. Fryou, E. Do, J. Love, R. Robbins, H. S. Dashti, and J. M. Lane, "Mobile apps for dietary and food timing assessment: Evaluation for use in clinical research," *JMIR Formative Research*, vol. 7, p. e35858, 2023.

[73] M. Sevil, M. Rashid, I. Hajizadeh, M. Park, L. Quinn, and A. Cinar, "Physical activity and psychological stress detection and assessment of their effects on glucose concentration predictions in diabetes management," *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 7, pp. 2251–2260, 2021.

[74] M. R. Askari, M. Rashid, X. Sun, M. Sevil, A. Shahidehpour, K. Kawaji, and A. Cinar, "Detection of meals and physical activity events from free-living data of people with diabetes," *Journal of Diabetes Science and Technology*, p. 19322968221102183, 2022.

[75] H. Wang, E. Doumard, C. Soulé-Dupuy, P. Kémoun, J. Aligon, and P. Monsarrat, "Explanations as a new metric for feature selection: a systematic approach," *IEEE Journal of Biomedical and Health Informatics*, 2023.

[76] J. Zacharias, M. von Zahn, J. Chen, and O. Hinz, "Designing a feature selection method based on explainable artificial intelligence," *Electronic Markets*, vol. 32, no. 4, pp. 2159–2184, 2022.

[77] M. Vijayan, S. Sridhar, and D. Vijayalakshmi, "A deep learning regression model for photonic crystal fiber sensor with xai feature selection and analysis," *IEEE Transactions on NanoBioscience*, 2022.

[78] M. Parillo, G. Annuzzi, A. A. Rivellese, L. Bozzetto, R. Alessandrini, G. Riccardi, and B. Capaldo, "Effects of meals with different glycaemic index on postprandial blood glucose response in patients with type 1 diabetes treated with continuous subcutaneous insulin infusion," *Diabetic medicine*, vol. 28, no. 2, pp. 227–229, 2011.