Proceedings of the 8th International Conference on Computing Methodologies and Communication (ICCMC-2025)
IEEE Xplore Part Number: CFP25K25-ART; ISBN: 979-8-3315-1211-8

2025 8th International Conference on Computing Methodologies and Communication (ICCMC) | 979-8-3315-1211-8/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICCMC65190.2025.11140629

Interpreting Blood Glucose Effects in Juvenile Diabetes Through XAI

R. Shanthi,
Department of CSE, VISTAS
Chennai-600117, Tamil Nadu, India
Email: shansara2013@gmail.com

R. Anandan,
Department of CSE, VISTAS
Chennai-600117, Tamil Nadu, India
Email: anandan.se@vistas.ac.in

S. Sridevi,
Department of CSE, VISTAS
Chennai-600117, Tamil Nadu, India
Email: sridevis.se@vistas.ac

G. Revathy,
Department of CSE, VISTAS
Chennai-600117, Tamil Nadu, India
Email: revathy.se@vistas.ac.in

S. Meera,
Department of CSE, VISTAS
Chennai-600117, Tamil Nadu, India
Email: meera.se@vistas.ac.in

Abstract

This study reveals that leveraging XAI to assess the impact of nutritional and contextual parameters on blood sugar for type 1 diabetes mellitus in machine learning methods. The method of this study includes Type 1 diabetes mellitus (T1DM) is defined by insulin dependent diabetes as well as difficulty in controlling blood sugar. This project introduces a blood glucose forecasting model that utilizes machine learning methods, namely the Random Forest algorithm, to predict blood glucose levels from past patient data. The model utilizes different input features, such as past glucose levels, insulin doses, and carb intake, to make precise short-term predictions. Data preprocessing methods are used enhance the dataset quality. The model is gauged as effective through its performance measures in terms of Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE), reflecting that the model could effectively predict high variability in blood glucose with utmost precision. Last but not least, the adoption of XAI such as SHAP and LIME adds interpretability to this model and helps users to trace in what manner the given feature contributes towards prediction results. The intended system will enable patients and clinicians which directly contributes patient safety and quality of life. Future developments involve merging real-time data from continuous glucose monitoring, personalization of prediction models for individualized patients. This research establishes the foundation for an extended diabetes management tool that closes the gap between data science and medicine.

Keywords: *Type 1 Diabetes Mellitus, Blood Glucose Prediction, Random Forest Algorithm, SHAP(Shapley Additive explanations), LIME(Local Interpretable Model-agnostic Explanations).*

1.Introduction

Insulin Dependent Diabetes Mellitus or T1DM [1][7]is a persistent condition where the pancreas secretes little or no insulin, and the patients need to control their blood sugar levels by taking external insulin and making life-style modifications. Effective glycaemic control is necessary to avoid complications like hypoglycaemia (hypoglycaemia) or hyperglycaemia (hyperglycaemia), which can result in serious health problems. Continuous Glucose Monitoring (CGM) devices are commonly employed for monitoring blood glucose levels during the day, which gives real-time information that individuals and clinicians can use to resolve and to make correct choice about insulin therapy and meal planning.

Managing blood glucose levels is a complex task due to various factors that affect glucose metabolism, including insulin dosing, food intake, physical activity, and stress. Predicting future blood glucose levels is a challenging problem because it involves nonlinear relationships and dynamic factors that interact in real time. Although CGM systems provide valuable data, there is a need for predictive models to alert patients of impending hypo- or hyperglycaemia.

The current methods aim to develop a more accurate and interpretable predictive model for predicting blood glucose level for every 15 minutes for one hour given past data and patient-specific covariates.

This study suggests a machine learning-driven blood glucose prediction system that uses the Random Forest algorithm to improve prediction accuracy. The solution is organized as follows:

1. Data-Driven Predictive Modeling

2. Machine Learning Model – Random Forest

3. Explainability Using SHAP & LIME

2.Related work

Blood glucose prediction has emerged as an active research domain as a result of increased accessibility of

wearable devices, which yield abundant data on real-time glucose fluctuations. Conventional methods have relied on rule-based systems or basic linear models to forecast blood glucose levels, but these models are incapable of handling the intricate, nonlinear interrelations between factors like insulin dosing, carbohydrate consumption, exercise, and circadian cycles. With advancements in machine learning methods, more advanced models have been proposed for enhanced prediction accuracy such as decision trees, neural networks, and hybrid models that incorporate patient-specific information. This section discusses the current methods of blood glucose prediction, as well as certain research on serial number extraction for time-based prediction intervals, feature extraction methods, and the shortcomings that are present in current systems.

Giovanni annuzzi1 et al,[1] have proposed effect of various factors of food in glycaemic prediction in juvenile diabetes through algorithms in machine learning, this study has proposed glycaemic prediction which involves collecting data from continuous glucose monitors (CGM), dietary intake (carbohydrates, fats, proteins), insulin dosage, physical activity, and other relevant factors. Data preprocessing includes cleaning, handling missing values, normalizing, and creating time-lagged features. Feature selection techniques identify key nutritional factors influencing blood glucose levels. ML models like XGBoost, or LSTMs are applied, followed by model evaluation using metrics such as MAE, RMSE, or R-squared. Explainability tools like SHAP or LIME provide insights into the impact of nutritional factors. However, limitations include the potential for inaccurate dietary data, difficulties in generalizing models across individuals, and challenges in predicting delayed glucose responses from high-fat meals.

Annuzzi, Giovanni, Lutgarda et al,[2] have proposed examine the factors of food that effects on blood glucose anticipating for juvenile diabetes using XAI, the methodology involves collecting real-time data such as continuous glucose monitoring (CGM) readings, detailed dietary intake (carbohydrates, fats, proteins), insulin dosage, activity, and other factors. Data preprocessing includes data scaling, missing data management and feature creation to capture time-dependent patterns. Machine learning models like LSTMs, GRUs, or XGBoost are applied for time-series blood glucose prediction. Explainability techniques are used to interpret model predictions, providing insights into the influence of food on glycemic levels. Limitations include the dependence on accurate dietary data, individual variability in glucose response, and challenges in predicting delayed effects of certain foods.

E. A. Pustogerov, A. S. Tkachuk, E. A. Vasukova,et al,[3] have suggested a method of forecasting blood glucose after via meal in diabetes mellitus during pregnancies called gestational diabetes includes the collection of real-time data such as blood glucose values, the intake of food (particularly the carbohydrate), basal insulin dose, physical workouts, and other parameters of the state of body. The process involves data preprocessing using data scaling, data missing management and feature creation for the generation of time-lagged variables. Model performance which monitored utilizes the metrics such as MAE, RMSE, and R-squared. Explainable AI (XAI) methods like SHAP or LIME offer insights into how various features impact predictions. Some limitations are the need for precise dietary and activity information, variability between individuals in their glucose responses, and inability to capture delayed impacts of fat or protein intake.

Duckworth,C, Guy, M.J.Kumaran, A.O.Kane, et al,[4] [9] have suggested AI methods for real time prediction for high and low glycemia and also to provide personal suggestion. In this the research used the XGBoost machine learning algorithm to forecast hypoglycaemia and hyperglycaemia events up to 60 minutes ahead based on continual blood glucose details from 153 people with juvenile diabetes. Features capturing according to time frames, and population data were included. SHAP (Shapley Additive explanation's) was used to explain individual predictions and determine the top features driving risk predictions for a given user. The research concentrated on one particular age group (young adults) and might not be generalizable to other populations.

Taiyu Zhu, Kezhi Li, Pau Herroro, Pantelis Georgiou, et.al,[5] have suggested Personalized Blood Glucose Prediction Using Machine Learning Techniques. The model incorporated historical continuous glucose monitoring (CGM) data, insulin dosages, carbs intake, and other activity levels to enhance prediction accuracy. Meta-learning was used to adapt the model to individual patient variations, allowing personalized predictions. Data preprocessing involved data scaling, addressing uncertainty, and creating features to enhance model accuracy. Real-time deployment challenges were not fully addressed, impacting the practical application of the system. Finally, the lack of explainability approaches makes healthcare providers to feel difficult to explain and belief the model forecast.

Erico Tjoa, Cuntai Guan,et al,[6] have proposed Explainable AI (XAI) toward medicinal applications, the methodology typically involves reviewing existing XAI techniques used to interpret complex AI models in healthcare. It categorizes these methods into model-specific (e.g., decision trees, linear models) and model-agnostic approaches (e.g., SHAP, LIME) that provide post-hoc explanations. The survey also assesses their applicability in various medical domains like diagnostics, prognosis, and treatment recommendations. Emphasis is placed on evaluating interpretability, transparency, and usability for medical practitioners. However, limitations include the trade-off between model accuracy and explainability, potential biases in

data and model interpretation, and challenges in validating the reliability of explanations.

3. Proposed Methodology

The proposed system aims to develop a robust forecasting model to forecasting blood glucose levels in juvenile Diabetes patients. Here, major elements of the system proposed are:

1. *Dealing with Missing Data:* The system uses missing data imputation methods to make sure that the gaps in the CGM readings do not hamper the model's accuracy.

2. *Enhanced Feature Extraction:* A robust feature engineering process is utilized, extracting meaningful time-based, event-based, and physiological features. This involves rolling windows to extract historical trends and derive features like time of day, insulin timing, and carbohydrate intake.

3. *Machine Learning-Based Prediction:* The model utilizes a Random Forest Regressor to address the intricate, nonlinear relationships between different features that influence blood glucose levels. The model is trained on historical glucose data as well as external variables, including insulin doses, food consumption, and exercise.

4. *Multiple Time Horizons:* The system makes predictions for blood glucose levels at a series of time intervals of every 15 minutes for one hour and provides flexibility in dealing with short-term and long-term glucose excursions.

5. *Model Explainability:* The use of Explainability AI such as SHAP and LIME will allow for greater transparency in the forecast outcomes, helping users, healthcare providers to realize the factors that are contributing most to the predicted glucose levels.

By addressing the shortcomings of existing models and building on the literature around serial number and feature extraction, the proposed system aims to offer a more accurate and personalized approach to blood glucose prediction.

3.1 Architecture

The architecture diagram Fig 1.1 has provided outlines a machine learning workflow that can be closely mapped to the preprocessing steps and subsequent model validation procedure in the code. This system architecture diagram outlines the pipeline for developing and validating a machine learning model for diabetic prediction, such as predicting Type 1 Diabetes Mellitus (T1DM) outcomes. Here's a breakdown of the components:

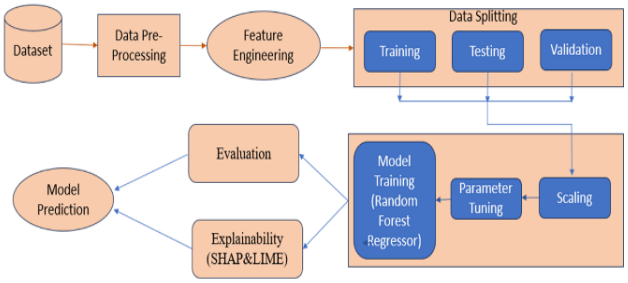

FIG 1.1: Architecture Diagram

3.2 Dataset

A T1DM Data sample is a collection of the data that were gathered from kaggle dataset that has CGM measurements. It comprises data of 100 T1D patients having CGM device, which was meant to calculate the glycemic levels in the scale of 40-400mg/dl. The details of data that contains both genders, between the age from 18 upto 80 years, with a T1D diagnosis. The dataset that includes comprehensive data on dietary factors such as carbohydrate intake, glycemic index, protein, fiber, fat, continuous blood glucose of every 15 minutes for one hour. The dataset is often stored in CSV or other structured formats. The system loads this data using libraries like Pandas in Python, enabling seamless handling of large datasets. Table1.1 and 1.2 provides detailed attribute information for the T1DM diabetes dataset.

Table 1.1 T1DM Diabetes Dataset 1

patie_id	patien age	carbs	prote	fat	fiber	glyce c_inc	contin us_ blood_ ucose	timesta p
0	56	57.660 7	31.237 3	14.39 20	14.311 8	62.43 8	176.477 4	2024-01- 00:00:00
1	69	94.866 1	27.911 7	21.98 96	13.085 1	66.51 6	79.7249	2024-01- 00:05:00
2	46	80.909 0	32.115 4	18.04 40	18.990 9	71.72 2	157.099 2	2024-01- 00:10:00
3	32	52.857 4	33.707 9	18.81 01	18.990 9	62.68 9	134.895 9	2024-01- 00:10:00
4	60	69.931 8	13.234 3	34.54 60	17.609 1	70.26 4	122.805 6	2024-01- 00:20:00

Table 1.2 T1DM Diabetes Dataset 2

missin cbg	basal	bolus	blood_gl ose_15	blood_gl ose_30	blood_gl ose_45	blood_gl ose_60
0	1.961 8	4.099 9	148.7953	210.8996	229.5135	193.7239
0	1.887 5	2.072 1	180.5728	94.45401	105.8751	195.2055
0	0.833 2	2.607 3	128.5257	86.28053	194.8486	135.8440
0	2.257 8	3.312 2	172.5093	92.11231	118.3832	187.1380
0	1.490 7	2.199 7	130.6443	96.95336	222.2455	132.2141

3.3 Data Cleaning and Pre-Processing

Data cleaning and preprocessing module is crucial for preparing the dataset to place it into the model workflow pipeline. The set of data with high quality has better model performance and interpretability.

i. Handling Missing Values:

- Drops rows with missing continuous_blood_glucose.
- Fills missing values in basal and bolus using the mean, which helps quantify the number of missing values in each feature column.
- Missing values are then imputed based on the nature of the data, For continuous features like blood glucose or carbohydrate intake, missing values are filled with the average value of the respective column. For discrete data, such as meal type or insulin type, the method is often used to replace missing values. Visualize the missing data patterns with missingno.

ii. Filtering:

- Filters out patients under the age of 18.
- Filters blood glucose values outside realistic ranges (40–400 mg/dL).

iii. Statistical Calculations:

- Computes basic descriptive statistics for all numeric columns.
- Calculates the standard deviation and coefficient of variation (CV) for continuous blood glucose.
- Computes the mean and standard deviation for future blood glucose predictions (15, 30, 45, 60 minutes).

iv. Timestamp Handling:

- Converts timestamps to date, time and extracts time features like the hour in a day and days in a week.
- Calculation of time difference between consecutive readings.

v. Outlier Removal:

- Optionally removes unrealistic blood glucose levels. Outliers are either removed or capped depending on their significance. Outlier treatment helps stabilize the model and prevents skewed predictions.

3.4 Feature Extraction

Feature engineering is about building new features or reworking the current features to more accurately reflect the hidden patterns in the data. This module contains

- *Time-based features*: Time since last meal, time of insulin administration.
- *Composite features:* Features like total carbohydrate-to-insulin ratio.
- *Encoding Categorical Variables:* Categorical variables, such as meal type, are encoded as numerical values (e.g., one-hot encoding) for model compatibility. Use ordinal encoding for ranked ordered features like severity scale. Use target encoding in cases of high cardinality features wherein the mean target value per category is calculated.

Feature engineering is the process of choosing appropriate variables (features) from the data that can potentially affect blood glucose levels [8]. The chosen features are nutritional intake variables (carbohydrates, protein, fat, Fiber), which affect blood glucose during digestion and absorption. The glycemic_index quantifies how rapidly foods increase blood glucose and is thus an important factor [14]. Insulin management factors such as basal and bolus units are also included since they directly control glucose levels [9]. patient_age is also taken into consideration, because age has an impact on glucose metabolism and insulin sensitivity [10]. These are stored in variable X, the independent variables upon which predictions will be made. The target variable y, assigned continuous_blood_glucose, is the actual blood glucose to be predicted. This systematic selection of features is a crucial first step in feature engineering to enable the model to learn useful patterns and associations to be able to predict glucose levels accurately.

3.4 ML Model Validation

The validation process is performed by utilizing a two-stage splitting of the data to separate the dataset into three sets: 60% training, 20% validation, and 20% test sets. Initially, the train_test_split () method is employed to separate the data into a training set (X_train, y_train) comprising 60% of the data, and a temporary set (X_temp, y_temp) comprising the remaining 40%. Next, the temporary set is split again by train_test_split() into halves: a validation set such as X_valid, y_valid and a test set such as X_test, y_test. The validation set is utilized to adjust model hyperparameters and measure model performance during training to avoid overfitting. The test set is kept unseen until the point of final evaluation, giving an unbiased estimate of the generalization capacity of the model. Random_state=42 ensures reproducibility by stabilizing the random seed for stable results across run repetitions. The well-structured validation process is a necessity to develop robust, reliable predictive models.

3.5 Scaling the Data

Scaling is used to make all features contribute equally to the model, particularly when they have different scales, as is usual in medical data. Various dietary and contextual factors are usually measured in various units and also have different ranges. For instance: Carbohydrate consumption would be quantified in grams and would generally lie between 0 to 100+ grams whereas Blood glucose can be quantified in mg/dL and lie between 40 to 400 mg/dL. In intended work Standard scaler is utilized which works on the basis of standardization to get the mean value as 0 and also have to attain the value 1 for standard deviation for every factors. The change carried out by Standard Scaler has been expressed as:

$$z = \frac{x - \mu}{\sigma}$$

x represents the real factor value, μ represents the mean of the factors, σ is the standard deviation, and z is the standardized factors value.

Feature scaling is done by using the StandardScaler () from the sklearn.preprocessing module. Scaling is one of the important step in data preprocessing especially for models sensitive to feature value magnitudes, like linear models or neural networks. The StandardScaler () scales the factors by subtracting the mean and scaling the unit variance, thus converting the factors in the dataset to get the mean value as 0 and also attain the value 1 for standard deviation. This is done through the formula:

$$X_{scaled} = \frac{X - mean}{Standard\ deviation}$$

First, X_train calculates the mean and the standard deviation using the training set and performs scaling transformation on the same. Then, X_valid and X_test apply the same transformation parameters to the validation and test sets, ensuring consistency. This prevents data leakage, as information from the validation or test sets is not used during the scaling of the training data. Standardization improves model convergence, particularly for algorithms that rely on gradient-based optimization, and make sure that all features have an equal importance on the model's training.

3.6 Training the Model

3.6.1 Random Forest Regressor:

The machine learning algorithms which can be used for regression tasks can build a group of decision trees. Each tree is trained on random subsets of the data and features. The ultimate prediction is usually the average of the predictions from all the individual trees in the forest.

3.6.2 Hyperparameter Tuning:

In this context, hyperparameters such as n_estimators and max_depth are adjusted.

i. n_estimators: More trees can lead to better performance as the model averages more results, but it also increases computational cost.
ii. max_depth: Deeper trees can fit the training data better (leading to low training error) but may overfit if they become too complex.

3.6.2.1 Training Multiple Models:

Different combinations of n_estimators and max_depth values are used to train multiple Random Forest Regressor models. For example, one model might have 100 trees with a max depth of 5, while another might have 200 trees with a max depth of 10, and so on.

3.6.2.2 Validation Set MSE (Mean Squared Error):

- After training, each model is evaluated on a validation **set**, which is a subset of the data not used for training.
- MSE is used as the performance metric for evaluating the validation set's model. MSE estimates the mean squared difference between predicted and actual values, and lower values indicate better performance.

3.6.2.3 Best-Performing Model:

- Once trained and tested on all the models, the best model is the one that has the minimum validation set MSE. This best model should perform better on unseen data since it has been picked on the basis of its accuracy on the validation set (how well the model will actually do in real situations).
- In short, this is a type of hyperparameter tuning in which varying values for the number of trees (n_estimators) and the tree depths (max_depth) are experimented with, and the model that results in the least prediction error (MSE) on a validation set is selected as the best one.

4 Results and Discussions

4.1 Model Evaluation

The ultimate assessment of the best-tuned Random Forest Regressor model is performed based on the unseen test set. Following hyperparameter tuning, the best model (best_model) is utilized to make predictions (y_test_pred) on the scaled test data (X_test_scaled). The performance of the model is subsequently evaluated based on two important assessment metrics:

4.1.1 Mean Squared Error (MSE):

This calculates the mean of the square of differences between predicted (y_test_pred) and actual (y_test) blood glucose levels. MSE is a sensitive measure for large errors, hence it is a good measure to use in pointing out large deviations in predictions.

$$MSE = \frac{1}{n}\Sigma\left(y_{actual} - y_{predicted}\right)^2$$

4.1.2 Mean Absolute Error (MAE):

MAE is the mean absolute difference between the values of actual and predicted, a more understandable metrics of the model prediction accuracy. It is less sensitive to outliers than MSE.

$$MAE = \frac{1}{n}\ \Sigma\left|y_{actual} - y_{predicted}\right|$$

These values are output to give a clear view of the accuracy of the model and the magnitude of the errors. A smaller MSE and MAE show that the model is performing better. Testing on the test set provides a final, unbiased measurement can be applied to new model,

unseen information, making it reliable for real-world blood glucose prediction applications [13].

4.2 Explainability and Visualization Tools

The overall process is implemented to analyse the performance of the best-selected Random Forest Regressor model, interpret its predictions using SHAP and visualize both model insights and blood glucose data trends.

4.2.1 Explainability using SHAP

- The SHAP values are calculated using shap.TreeExplainer(best_model) to explain the predictions by determining the contribution of each attribute to the predicted blood glycemic levels.
- A SHAP summary plot (shap.summary_plot) is generated to provide a global view of the feature importance, indicating which features had the most significant influence on predictions.

The Fig 1.2 and Fig 1.3 provides insights into the impact of individual features like carbohydrates, protein, fat, etc., on the prediction of blood glucose levels.

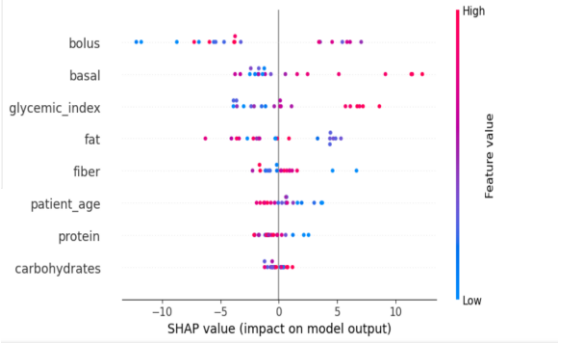

FIG 1.2: Impact of Dietary factors on Blood Glucose

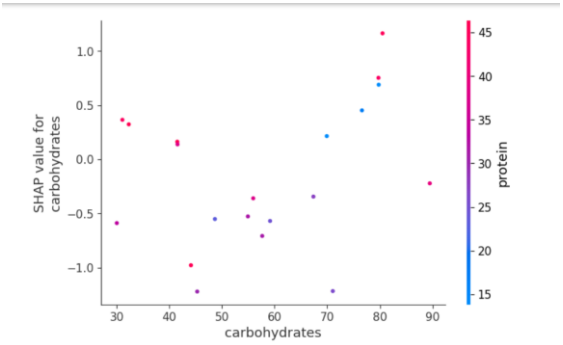

FIG 1.3 SHAP Value Plot for Carbohydrates Showing the Impact on Blood Glucose Prediction with Protein Levels as a Colour Gradient

4.3 Correlation Matrix:

A correlation matrix as shown in Fig 1.4 is generated using to visualize correlations between the features and the blood glucose levels. Strong correlations indicate which factors are most influential in determining glucose levels, aiding in model interpretation and feature selection.

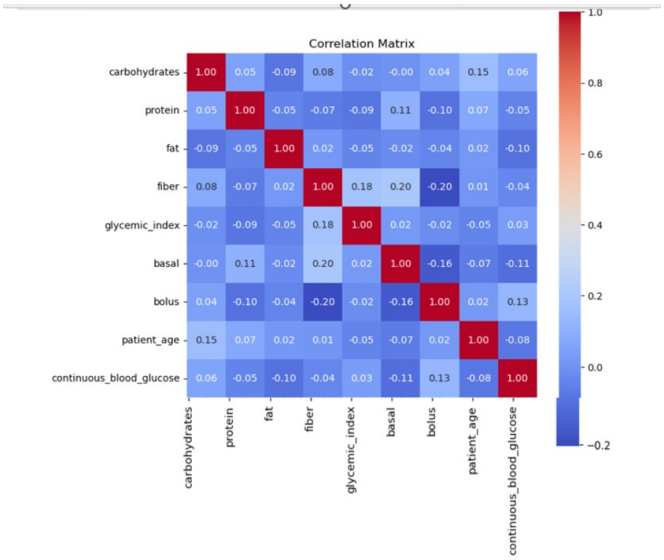

FIG 1.4 Correlation Matrix

4.4 Blood Glucose Trend Analysis

One of the main issues in juvenile diabetes patients is to control blood Glucose occurring after meal, through dosing the insulin bolus to be delivered pre-prandial meals [11][12]. According to these assumptions, an experiment to find the influence of nutritional parameters such as carbohydrates, proteins, fats, fibres and other factors that acts on blood glucose over short and middle time range has been predicted by Machine Learning (ML) methods. The Fig 1.5 shows the prediction of the blood glucose readings for every 15 minutes for one hour after meals using insulin dosing, blood glucose, and nutritional variables in T1D patients on AP systems was used.

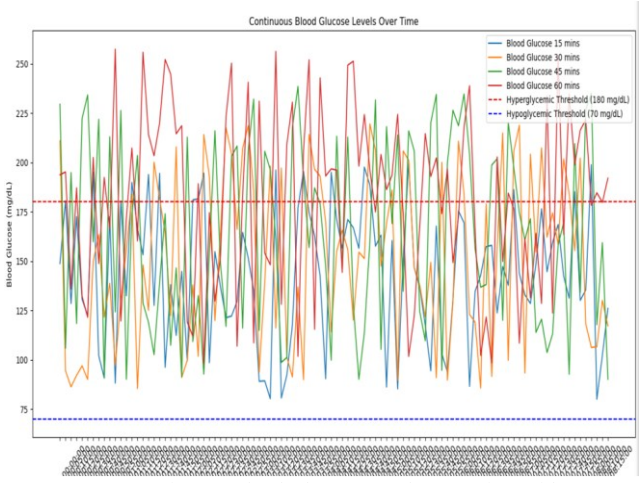

FIG 1.5 Continuous Blood Glucose Levels Over Time with Hyperglycemic and Hypoglycemic Thresholds

4.5 Actual vs Predicted Plot

A scatter plot as shown in Fig1.6 is drawn via plt.scatter() for comparing observed blood glucose measurements and predicted ones. A dotted diagonal red line illustrates the desired case when observed and predicted are the same. Shifts away from this line identify prediction flaws and model inefficiencies.

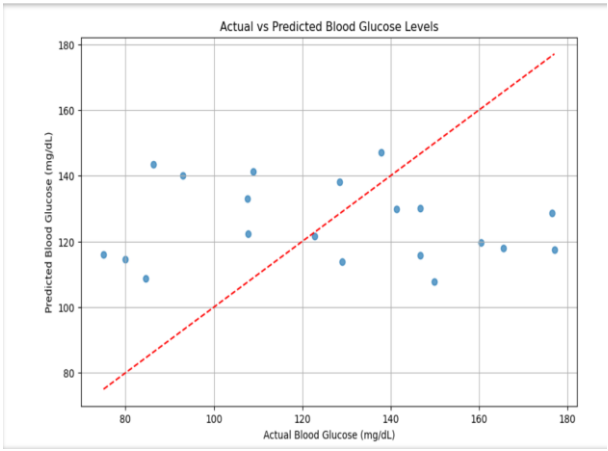

FIG 1.6 Actual vs. Predicted Blood Glucose Levels with Reference Line

4.6 Interpreting Model Prediction using LIME For Explainability

Interpretability is a critical component of the blood glucose prediction system, especially in a healthcare context where model decisions must be explainable.

The LIME framework is used to generate local explanations for individual predictions, making the

Random Forest model more interpretable. LIME (Local Interpretable Model-agnostic Explanations) explains individual predictions by modeling the behavior of the model in the vicinity of each data point using a simple interpretable model, like a linear model. LIME helps to understand features (e.g., carbohydrate consumption, insulin) influencing blood glucose predictions.

4.6.1 Validating LIME Explanations

- To test LIME's validity, explanations from similar instances are compared. Consistent explanations verify that the model's decision-making is in accordance with established physiological factors.
- Clinicians review explanations produced by LIME as shown in Fig 1.7 to verify that they are consistent with medical understanding of glucose control and Fig 1.8 shows the model prediction of blood glycemic values using LIME for the corresponding features.

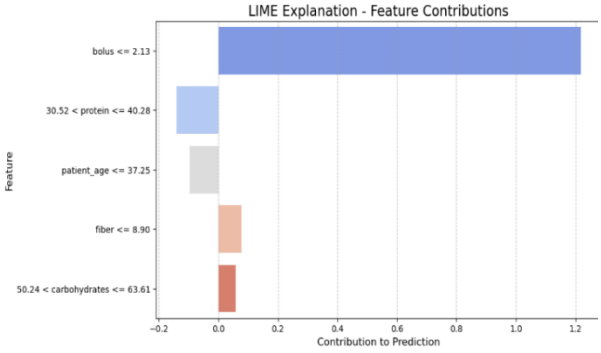

FIG 1.7 Blood Glucose Prediction Using LIME Based On Features

4.6.2 Feature value Table

The "Feature Value" in Table 1.3 contains the list of real values of the respective instance:

Table 1.3 Feature Value Table

bolus	2.03
protein	37.83
basal	2.15
carbohydrates	55.96
Patient_age	29.00

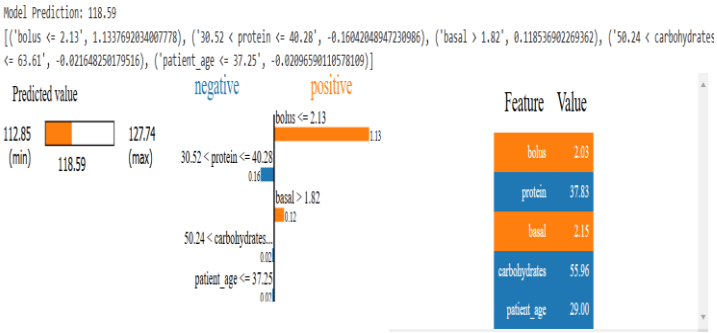

FIG 1.8 Model Prediction Using LIME with Feature Contributions for Blood Glucose

4.7 Model Improvement Strategies

In order to enhance the performance and consistency of the blood glucose prediction model, a number of improvement strategies were investigated:

1) Performance was improved by hyperparameter tuning of Random Forest parameters n_estimators and max_depth, and by sophisticated feature engineering with time-based and physiological features.

2) Accuracy was enhanced through time-aware cross-validation, which minimized model bias and captured true-world trends more accurately. Clinical relevance was leveraged to inform feature selection based on domain expertise.

3) Reliability was enhanced by strong cross-validation, repeated random seed use, and data anomaly management. Regular model updating and drift detection were prioritized for long-term stability.

4) Precision was improved by the addition of weighted loss functions to focus on key glucose ranges and incorporating SHAP and LIME analysis insights to optimize feature importance and model behavior.

5) The model has been improved by employing temporal sequences, includes contextual factors such as physical activities create hybrid predictions and combine the random forest with thin neural network.

6) The tools like SHAP and LIME make clinicians to validate model behaviour, increase transparency and increase patient confidence in automation decision making.

These approaches in combination provide a high-performing, interpretable, and clinically valuable glucose prediction model for the management of juvenile diabetes.

5. Conclusion

The creation of a blood glucose prediction model with the Random Forest algorithm proves a strong, data-

driven methodology for predicting future glucose values for Type 1 Diabetes (T1DM) patients. By leveraging historical patient information, including insulin dosing, blood glucose values, and carbohydrate ingestion, the system can produce short-term forecasts useful for patients and clinicians to inform decisions. The use of SHAP (Shapley Additive explanations) for explainability of the model guarantees not only the accuracy of the predictions but also their interpretability so that users can perceive the reasons behind variations in glucose levels.

The program effectively processes different tasks ranging from data cleaning to missing value management, model optimization, and result visualization. Performance measures such as Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) reflect the model's precision in forecasting glucose levels. Additionally, the system's capability to interpret individual predictions as regards of significance (via SHAP values) provides important perceptions into how different factors, for instance, insulin usage or past glucose readings, impact future glucose levels. This solution provides a possible means for healthcare professionals to better monitor patients and for patients to better control their condition themselves. By anticipating glucose fluctuations, the system can minimize the dangers of hypoglycaemia and hyperglycaemia, enhancing diabetes care overall.

REFERENCES

[1] Annuzzi, Giovanni, Andrea Apicella, Pasquale Arpaia, Lutgarda Bozzetto, Sabatina Criscuolo, Egidio De Benedetto, Marisa Pesola, Roberto Prevete, and Ersilia Vallefuoco. "Impact of nutritional factors in blood glucose prediction in type 1 diabetes through machine learning." *IEEE Access* 11 (2023): 17104-17115.

[2] Annuzzi, Giovanni, Andrea Apicella, Pasquale Arpaia, Lutgarda Bozzetto, Sabatina Criscuolo, Egidio De Benedetto, Marisa Pesola, and Roberto Prevete. "Exploring nutritional influence on blood glucose forecasting for type 1 diabetes using explainable AI." *IEEE journal of biomedical and health informatics* 28, no. 5 (2023): 3123-3133.

[3] Pustozerov, Evgenii A., Aleksandra S. Tkachuk, Elena A. Vasukova, Anna D. Anopova, Maria A. Kokina, Inga V. Gorelova, Tatiana M. Pervunina, Elena N. Grineva, and Polina V. Popova. "Machine learning approach for postprandial blood glucose prediction in gestational diabetes mellitus." *Ieee Access* 8 (2020): 219308-219321.

[4] Duckworth, Christopher, Matthew J. Guy, Anitha Kumaran, Aisling Ann O'Kane, Amid Ayobi, Adriane Chapman, Paul Marshall, and Michael Boniface. "Explainable machine learning for real-time hypoglycemia and hyperglycemia prediction and personalized control recommendations." *Journal of Diabetes Science and Technology* 18, no. 1 (2024): 113-123.

[5] Zhu, Taiyu, Kezhi Li, Pau Herrero, and Pantelis Georgiou. "Personalized blood glucose prediction for type 1 diabetes using evidential deep learning and meta-learning." *IEEE Transactions on Biomedical Engineering* 70, no. 1 (2022): 193-204.

[6] Tjoa, Erico, and Cuntai Guan. "A survey on explainable artificial intelligence (xai): Toward medical xai." *IEEE transactions on neural networks and learning systems* 32, no. 11 (2020): 4793-4813.

[7] A. Katsarou, S. Gudbjornsdottir, A. Rawshani, D. Dabelea, E. Bonifacio, B. J. Anderson, L. M. Jacobsen, D. A. Schatz, and A. Lernmark, "Type 1 diabetes mellitus," Nature Rev. Disease Primers, vol. 3, no. 1, p. 17016, 2017.

[8] S. E. Berry, A. M. Valdes, D. A. Drew, F. Asnicar, M. Mazidi, J. Wolf, J. Capdevila, G. Hadjigeorgiou, R. Davies, H. Al Khatib, and C. Bonnett, "Human postprandial responses to food and potential for precision nutrition," *Nature Med.*, vol. 26, no. 6, pp. 964–973, Jun. 2020.

[9] Amiel, Stephanie A., Nancy Pursey, Bernard Higgins, and Dalia Dawoud. "Diagnosis and management of type 1 diabetes in adults: summary of updated NICE guidance." *Bmj* 351 (2015).

[10] Campbell, Ian W. "Scottish Intercollegiate Guidelines network: management of diabetes (SIGN 55)." *Journal of the Royal College of Physicians of Edinburgh* 32, no. 3 (2002): 162-166.

[11] Hoyos, J. D., María F. Villa-Tamayo, Carlos E. Builes-Montaño, A. Ramirez-Rincón, José Luis Godoy, Jose Garcia-Tirado, and Pablo S. Rivadeneira. "Identifiability of control-oriented glucose-insulin linear models: review and analysis." *IEEE Access* 9 (2021): 69173-69188.

[12] Bekiari, Eleni, Konstantinos Kitsios, Hood Thabit, Martin Tauschmann, Eleni Athanasiadou, Thomas Karagiannis, Anna-Bettina Haidich, Roman Hovorka, and Apostolos Tsapas. "Artificial pancreas treatment for outpatients with type 1 diabetes: systematic review and meta-analysis." *bmj* 361 (2018).

[13] H. N. Mhaskar, S. V. Pereverzyev, and M. D. Van Der Walt, "A deep learning approach to diabetic blood glucose prediction," *Frontiers Appl. Math. Statist.*, vol. 3, p. 14, Jul. 2017.

[14] S. E. Berry, A. M. Valdes, D. A. Drew, F. Asnicar, M. Mazidi, J. Wolf, J. Capdevila, G. Hadjigeorgiou, R. Davies, H. Al Khatib, and C. Bonnett, "Human postprandial responses to food and potential for precision nutrition," *Nature Med.*, vol. 26, no. 6, pp. 964–973, Jun. 2020.

[15] Khodaei, Mohammad Javad, Nicholas Candelino, Amin Mehrvarz, and Nader Jalili. "Physiological closed-loop control (PCLC) systems: Review of a modern frontier in automation." *Ieee Access* 8 (2020): 23965-24005.