Personalized Automated Blood Glucose Forecasting for Type-1 Diabetes Using Machine Learning Algorithms

Avijay Sen,
Franklin High School
Elk Grove, California, United States
avijay.sen12@gmail.com

Dr. Sindhu Ghanta
AIClub Research Institute
Mountain View, CA, USA
sindhu@aiclub.world

Pallavi Bajpai
AIClub Research Institute
Mountain View, CA, USA
pallavi.bajpai@aiclub.world

***Abstract*—Type-1 Diabetes Mellitus (T1DM) is a chronic condition characterized by the pancreas's inability to produce insulin, requiring continuous monitoring and management of blood glucose levels. Accurate prediction of blood glucose levels can significantly improve patient outcomes by reducing hypo- and hyperglycemic events. This study develops a personalized automated blood glucose forecasting system leveraging the past blood glucose levels and insulin pump data. Utilizing the publicly available Diatrend dataset, encompassing thirty-one days of data for five subjects, we evaluated three machine learning algorithms: K-Nearest Neighbors (KNN), Random Forest (RF), and Multilayer Perceptron (MLP). After hyper-parameter tuning, the performance of each algorithm was assessed using Root Mean Squared Error (RMSE), Mean Squared Error (MSE), and the coefficient of determination (R^2), with a particular emphasis on RMSE. The Random Forest model demonstrated superior performance, achieving a test RMSE range of $14.98 - 23.62$ across all subjects. This research highlights the efficacy of supervised machine learning algorithms in predicting blood glucose levels over one-hour intervals for T1DM patients, underscoring the potential of personalized machine learning models to improve diabetes management.**

Keywords- blood glucose prediction; Type-1 Diabetes Mellitus; insulin delivery system

I. INTRODUCTION

Type 1 Diabetes (T1DM) is a chronic condition where the pancreas fails to produce insulin, the hormone needed to control blood sugar levels. People with T1DM face challenges in managing blood sugar, which can be too low (hypoglycemia) or too high (hyperglycemia). Low sugar levels, below 70 mg/dL, can cause symptoms like sweating, hunger, and even serious issues like seizures or coma [1], [2]. High sugar levels, over 140 mg/dL, may lead to problems in the eyes, heart, and nerves [3], [4]. Managing these fluctuations requires careful insulin use, which can be challenging and risky [5].

The eight leading cause of death globally is diabetes [6]. The number of deaths has been increasing since the start of the 21st century [7]. The increasing trend approximates that there will be 13.5-17.4 million people suffering from T1DM by 2040 [8]. Majority of the deaths occur before 70 and are due to high glucose levels [9].

A development of a sophisticated insulin delivery method that combines Continuous Glucose Monitoring (CGM), which utilizes the subcutaneous interstitial fluid to measure glucose levels and insulin pumps which use glycemic data from the monitors to provide temporary insulin formulas like basal or bolus to maintain glucose levels. The device asks the patient for information on physical activity, insulin bolus dosage, meal sizes and carbohydrate content, among other things, in order to obtain more accurate assessments [10].

To further enhance the capabilities of CGMs, Machine Learning (ML) offers a promising avenue. ML can perform human-like tasks through learning from data and being able to adapt to unseen data. There are various types of ML algorithms, such as Supervised, Unsupervised, Semi-supervised, and Reinforcement learning [11]. Supervised learning is typically the task of ML to learn a function that maps an input to an output based on sample input-output pairs [12]. It uses labeled training data and a collection of training examples to infer a function. Supervised learning is carried out when certain goals are identified to be accomplished from a certain set of inputs [13]. There are two different types of common supervised tasks which include "classification" that separates the data or "regression" that fits the data [12]. For the purpose of this study, regression was used consisting of different algorithms [14].

The integration of technology in diabetes management has led to significant advances in the prediction and control of blood glucose levels [15]. CGMs combined with insulin pumps, forms the backbone of artificial pancreas systems, which automate insulin delivery to maintain optimal glucose levels [16]. These closed-loop systems have shown promise in reducing the burden of daily diabetes management and improving overall quality of life for patients [17]. Studies have demonstrated that such systems can significantly improve glycemic control, reduce HbA1c levels, and mitigate the risks associated with long-term diabetes complications [18]. The continuous evolution of these technologies and their integration with machine learning algorithms hold the potential to transform diabetes care, making it more precise, personalized, and effective [19].

Our paper focuses on evaluating closed-loop insulin delivery systems, known as artificial pancreas systems, for their effectiveness and safety in managing T1DM. By analyzing CGM data, we developed a method to fine-tune insulin rates using various ML models. Our personalized approach using the Diatrend dataset demonstrates the strength and flexibility of these models for individual patient needs.

The paper is structured as follows: The Introduction discusses the challenges of T1DM and the role of ML in improving insulin systems. The Related Work section reviews existing models and their limitations. Materials and Methods explain

our dataset, data preparation, and methodology. Results provide an analysis of model performance. Discussion interprets the findings, comparing them with existing methods. Future Work & Limitations suggest improvements and study constraints. The Conclusion summarizes our contributions and highlights the importance of personalized systems in diabetes care.

II. Related Work

Predicting blood glucose levels in patients with T1DM has been the focus of numerous studies employing a variety of machine learning algorithms and models. Machine learning plays a crucial role in predicting blood glucose levels by analyzing vast amounts of data to identify patterns and trends that are not easily discernible by traditional methods. This allows for more accurate and personalized predictions, ultimately improving diabetes management and reducing the incidence of hypo- and hyperglycemic events. Prior research has demonstrated the potential of different methods, yet each approach has limitations that impact the predictability and efficiency of the models.

The emergence of CGMs has introduced different methodologies aimed at forecasting glucose levels. There have been advancements in creating physical models and/or data-driven observational models that attempt to predict glucose levels of patients [20]. A few models that have been used are Proportional-Integral-Derivative (PID) Controllers [21], Artificial Neural Networks (ANNs) [22], Recurrent Neural Networks (RNNs) [23], Long Short-Term Memory (LSTM) Networks [24], Support Vector Machines (SVM) [25], [26], Fuzzy Logic Systems [27], and RFs [28]. Recently, neural network based models are gaining popularity: the use of dilated recurrent neural networks (DRNNs), which have shown promise in improving prediction accuracy by handling sequential data more effectively and overcoming issues like gradient vanishing [29]. Additionally, transfer learning approaches, where models are initially trained on a generalized dataset and then fine-tuned with individual patient data, have demonstrated enhanced prediction accuracy for specific subjects [30].

One notable study, titled “A Machine Learning Approach to Predicting Blood Glucose Levels for Diabetes Management” implemented the Support Vector Regression (SVR) algorithm alongside a physiological model characterized by three compartments: meal absorption dynamics, insulin dynamics, and glucose dynamics. The researchers utilized a small sample size of five T1DM patients to pull different parameters including carbohydrate intake, rapid-acting insulin, bolus and basal rate, body mass, and insulin sensitivity (IS) [31].

Similarly, another research paper compared the efficacy of LSTM networks and Temporal Convolutional Networks (TCNs) for blood glucose level prediction [32]. This study also explored various classification algorithms, including SVM, Naive Bayes, and Decision Tree for comparison. The results indicated that there was little benefit to employing TCN or LSTM over conventional models, pointing to a potential application gap for these cutting-edge neural networks. This emphasizes the necessity of more research to determine the circumstances in which these models could provide meaningful advantages.

Further research evaluating the accuracy of SVM, Naive Bayes, and Decision Tree algorithms in diabetes classification were conducted using the Pima Indian Diabetes Database [33]. One weakness of the dataset was its homogeneity—all of the patients were of the same race. This limited the results’ applicability to more diverse populations with a range of genetic and lifestyle backgrounds.

In order to overcome issues like missing data, research has also been done using RNN algorithms to predict blood glucose levels [34]. The study focused on improving prediction accuracy by utilizing the temporal dependencies in CGM data. The existence of missing data, however, presented a serious problem and might have an effect on the model’s predictability and accuracy. Developing efficient methods to deal with missing data is essential to enhancing RNN models’ resilience in practical applications.

Interestingly, researchers have proposed a hybrid approach combining SVM and Neural Networks (NN) to improve blood glucose level predictions. This method demonstrates enhanced accuracy in glucose forecasting, particularly in reducing prediction errors compared to traditional models [35]. However, the study relies on a relatively small dataset, which may affect the generalizability of the model to broader populations.

Building on these efforts, we adopt a different approach by utilizing the same dataset as prior studies but with distinct model choices and methodology. While deep learning models like LSTM and Encoder-Decoder are commonly used for time-series predictions, as highlighted in "Deep Learning-Based Glucose Prediction Models: A Guide for Practitioners" [36], we opt for simpler machine learning techniques such as KNN, RF, and MLP in order to easily integrate into healthcare systems. Additionally, we focus on hyperparameter optimization for individual subjects rather than complex training strategies like personalized or fine-tuning methods. This allows us to prioritize model simplicity and interpretability while still leveraging the same data.

In contrast to other approaches, our methodology involves the use of three distinct algorithms: KNN, RF, and MLP, to provide personalized solutions for each patient. The dataset we used includes a diverse group of subjects with varied characteristics, such as differences in sex and race, enhancing the representativeness and predictability of our results. By conducting hyperparameter tuning and training multiple models, we selected the best-performing model to ensure the robustness and accuracy of our findings, setting our research apart from previous studies.

This body of related work highlights the ongoing efforts and challenges in predicting blood glucose levels in T1DM patients. Each study contributes uniquely to the field, offering insights and advancements while highlighting areas for further investigation and improvement.

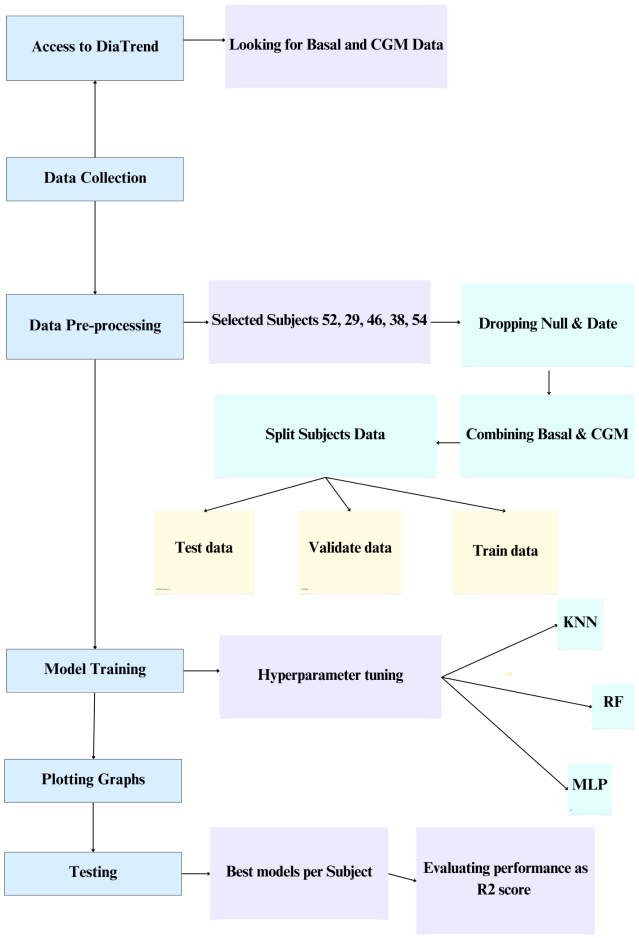


Figure 1. Flowchart of the process.

III. Materials and Methods

This section outlines the materials and methods employed in this study, providing a detailed framework for the research process.

A. Dataset

We utilized the Diatrend dataset [37], which offers extensive continuous data from wearable medical devices. This includes 8,220 days of insulin pump data and 27,561 days of CGM data from 54 diabetic patients. For our analysis, we selected five subjects out of 17 subjects from this dataset who had comprehensive CGM and basal insulin readings available to ensure data completeness.

B. Data Pre-processing

The dataset was initially filtered to identify patients with both CGM and basal insulin readings. To prepare the data for analytical and statistical models, several pre-processing steps were undertaken to ensure data quality and completeness.

Basal insulin data entries, which include fields for "date," "rate" (units per hour), and "duration" (milliseconds), were adjusted to ensure that no single duration exceeded 5 minutes (300,000 milliseconds). Any entries with duration longer than 5 minutes were split into multiple 5-minute segments, and the corresponding timestamps were updated accordingly. This adjustment facilitated accurate alignment with CGM data, ensuring consistent time intervals.

The modified basal insulin data was integrated with the CGM data to create a unified dataset. For each CGM timestamp, the corresponding basal insulin data was merged. If multiple basal insulin entries existed within the interval between two CGM readings, the basal entry that either matched or immediately followed the CGM timestamp that was selected.

Missing values in the CGM data for columns like "mg/dl" (glucose concentration) column were addressed using forward-fill imputation. This method replaces missing values with the last observed value, which is appropriate for maintaining the continuity of time-series data. Both CGM and basal insulin datasets were sorted by date to preserve their temporal sequencing.

To capture both glucose trends and insulin delivery patterns over time, the following features were calculated using a rolling window of 12 data points (equivalent to 1 hour if readings are taken every 5 minutes): Glucose Mean (glucose_mean): The mean glucose level over the window. Glucose Standard Deviation (glucose_std): The standard deviation of glucose levels over the window. Weighted Basal Infusion (basal_infusion): This feature was calculated as the sum of the product of "duration" and "rate" divided by the sum of "duration" over the window, representing the average basal insulin delivery rate weighted by duration.

These features provided a comprehensive view of glucose dynamics and insulin administration, which are critical for predictive modeling in diabetes management. After feature extraction, the dataset for each data point included the following features:

- date: Timestamp of the CGM reading.
- glucose_mean: Mean glucose level over the past hour.
- glucose_std: Standard deviation of glucose levels over the past hour.
- basal_infusion: Weighted average basal insulin infusion rate over the past hour.
- mg/dl: Current glucose reading.

After this, we divided the dataset in the order of time to preserve the temporal order and prevent any mixing of future and past data, hence randomization was not an option. By doing this, we made sure that the model learned from earlier data and was tested on later data, similar to real-world prediction situations, maintaining the quality of our time-based analysis.

- Training Set (70%): The earliest 70% of the data points, used to train the model.
- Validation Set (15%): The subsequent 15% of data points, used for hyper-parameter tuning.
- Test Set (15%): The latest 15% of data points, used to evaluate the model's performance on unseen data.

C. Methodology

For our analysis, we chose KNN, RF, and MLP regression models because of their proven effectiveness in both time-series prediction and glucose level forecasting:

- KNN: Valued for its straightforward approach and ability to capture local data patterns, KNN has successfully been used in glucose prediction, yielding satisfactory outcomes [38].
- RF: This ensemble learning technique improves predictive accuracy and mitigates over fitting. RF models are known for their robust performance in analyzing medical data and offer feature importance metrics, enhancing interpretability [39].
- MLP: As a type of neural network, MLP excels at modeling complex, non-linear relationships, making it highly appropriate for glucose prediction where such intricate patterns are present [40].

In particular, the feature importance scores provided by RF models significantly boost interpretability, which is essential in personalized medicine. Our choices emphasize a balance between achieving high predictive performance and maintaining model interpretability.

The dataset consists of five subjects. The dataset comprises data from five distinct subjects. For each subject, we developed a unique model using data specific to that individual since subject's timestamps were different for all subject's readings. Each of the three algorithms was applied separately to the data from each subject, allowing us to conduct thorough experiments tailored to each subject's dataset.

We conducted hyper-parameter tuning for each model to enhance performance. For KNN, we tested using between 1 and 16 neighbors to find the right balance for understanding both small and large patterns in the data. With RF, we experimented with using between 10 and 100 decision trees and adjusted their depth from 1 to 7 to avoid making the model too complex or too simple. For the MLP, we varied the starting learning rates between 0.00001 and 0.05 and adjusted the number of iterations from 10 to 100 to see how these factors affected the model's learning and improvement speed. Each configuration's performance was assessed using RMSE and R^2 on the validation set. This tuning process was crucial for ensuring generalization and avoiding over fitting. The optimal hyper-parameters differed across subjects, reflecting the unique glucose dynamics of each individual [41]. Furthermore, the best-performing model was employed to evaluate its performance by applying it to the test data of all subjects.

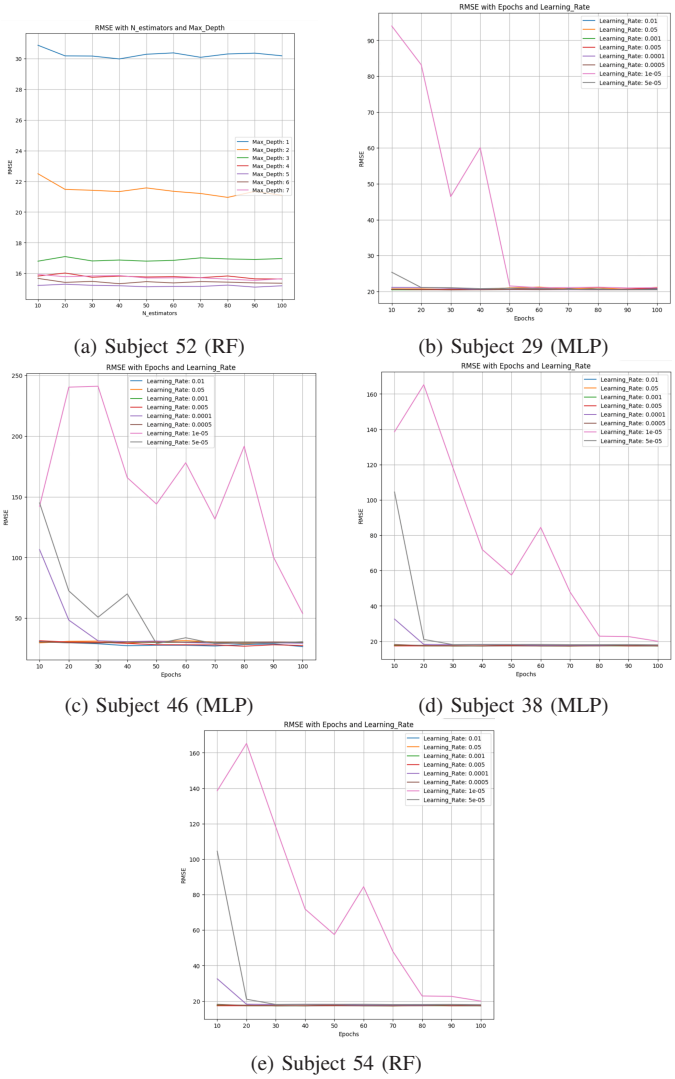
Following an extensive hyper-parameter tuning phase, the models that exhibited the best performance based on validation metrics were selected. These models were then rigorously tested on each of the five subjects' test data to evaluate their reliability. This evaluation involved calculating three key performance metrics: MSE, RMSE, and the R^2. These steps ensured an assessment of the model's predictive capabilities, providing insights into their performance on data that was not used during training and hyper-parameter tuning. Refer to Figure 1 for a visual representation of the Materials and Methods processes.

IV. Results

In this section, we present the performance of three machine learning models— KNN, RF, and MLP— across five different subjects, using the RMSE and R^2 (coefficient of determination) score as the key metric.

Graphs illustrating model performance metrics for each algorithm and subject using validation data are shown in Figure 2. Specifically, we plotted RMSE against the K values for KNN models, RMSE against the number of estimators (n_estimators) for various max_depth configurations in RF models, and RMSE against the number of iterations for different learning rates in MLP regression models.

The graph displays an RMSE range of 14.98 to 23.62 mg/dL. While this is relatively high, it falls within acceptable limits for glucose prediction models. Given that glucose levels can vary significantly and rapid fluctuations are common in Type 1 Diabetes patients, clinical guidelines typically consider deviations within ±30 mg/dL to be acceptable. Therefore, our model's errors are within a clinically relevant range [42].



(a) Subject 52 (RF)

(b) Subject 29 (MLP)

(c) Subject 46 (MLP)

(d) Subject 38 (MLP)

(e) Subject 54 (RF)

Figure 2. Subject's graph of their best performing model

Results of the best model obtained from hyper-parameter tuning on the validation and test datasets is shown in Tables I and II.

TABLE I. TRAINING RESULTS FOR DIFFERENT SUBJECTS AND MODELS

ID	KNN	RF	MLP
52	MSE: 252.947 RMSE: 15.904 R2 Score: 0.912	MSE: 227.535 RMSE: 15.084 R2 Score: 0.921	MSE: 317.137 RMSE: 17.808 R2 Score: 0.890
29	MSE: 438.806 RMSE: 20.947 R2 Score: 0.857	MSE: 425.273 RMSE: 20.622 R2 Score: 0.861	MSE: 420.411 RMSE: 20.503 R2 Score: 0.863
46	MSE: 814.730 RMSE: 28.543 R2 Score: 0.880	MSE: 717.231 RMSE: 26.781 R2 Score: 0.895	MSE: 820.608 RMSE: 28.646 R2 Score: 0.879
38	MSE: 317.209 RMSE: 17.810 R2 Score: 0.866	MSE: 310.727 RMSE: 17.627 R2 Score: 0.869	MSE: 301.532 RMSE: 17.364 R2 Score: 0.873
54	MSE: 342.127 RMSE: 18.496 R2 Score: 0.772	MSE: 299.137 RMSE: 17.295 R2 Score: 0.800	MSE: 375.030 RMSE: 19.365 R2 Score: 0.750

TABLE II. TESTING RESULTS FOR DIFFERENT SUBJECTS AND MODELS

ID	KNN	RF	MLP
52	MSE: 314.087 RMSE: 17.722 R2 Score: 0.926	MSE: 305.725 RMSE: 17.484 R2 Score: 0.928	MSE: 378.007 RMSE: 19.442 R2 Score: 0.911
29	MSE: 414.655 RMSE: 20.363 R2 Score: 0.880	MSE: 391.740 RMSE: 19.792 R2 Score: 0.886	MSE: 385.436 RMSE: 19.632 R2 Score: 0.888
46	MSE: 615.205 RMSE: 24.803 R2 Score: 0.922	MSE: 558.373 RMSE: 23.629 R2 Score: 0.929	MSE: 546.354 RMSE: 23.374 R2 Score: 0.931
38	MSE: 352.870 RMSE: 18.784 R2 Score: 0.800	MSE: 340.414 RMSE: 18.450 R2 Score: 0.807	MSE: 330.102 RMSE: 18.168 R2 Score: 0.813
54	MSE: 235.849 RMSE: 15.357 R2 Score: 0.789	MSE: 224.320 RMSE: 14.977 R2 Score: 0.800	MSE: 293.482 RMSE: 17.131 R2 Score: 0.738

For the training data, the Random Forest model achieved the highest R^2 score of 0.921 for Subject 52, demonstrating better predictive ability compared to the KNN model with an R^2 score of 0.912 and the MLP model with an R^2 score of 0.890. For Subject 29, the MLP model emerged as the best performer with an R^2 score of 0.863, slightly outperforming the RF and KNN models, which had R^2 scores of 0.861 and 0.857, respectively. In the case of Subject 46, the RF model again showed the highest predictive ability with an R^2 score of 0.895, while KNN and MLP had similar performances, with R^2 scores of 0.880 and 0.879, respectively. Additionally, for Subject 38 the MLP model achieved the highest R^2 score of 0.873, indicating better performance than both the RF and KNN models, which had R^2 scores of 0.869 and 0.866, respectively. Finally, for Subject 54, the RF model outperformed the other models with an R^2 score of 0.800, followed by the KNN model with an R^2 score of 0.772, and the MLP model with the lowest performance at an R^2 score of 0.750.

Overall, the RF model consistently executed the highest R^2 scores across the majority of subjects, indicating strong predictive performance. Specifically, the RF model had the highest R^2 scores for Subject 52 (0.921), Subject 46 (0.895), and Subject 54 (0.800). The MLP model performed the best for Subject 29 (0.863) and Subject 38 (0.873). While the KNN model showed strong performance, it did not outperform the RF or MLP models in any subject. These findings imply that the MLP and KNN models are closely followed by the RF model, which is the most reliable option for precise predictions across this dataset.

When evaluating the models on the test data, the RF model obtained the highest R^2 score of 0.928 for Subject 52, closely followed by the KNN model with an R^2 score of 0.926. The MLP model had a slightly lower R^2 score of 0.911. This indicates that both RF and KNN models performed similarly well, slightly outperforming the MLP model for this subject. In the case of Subject 29, the MLP model emerged as the best performer with an R^2 score of 0.888. The RF model also performed well, achieving an R^2 score of 0.886, while the KNN model had a slightly lower R^2 score of 0.880. The differences in performance were minimal, suggesting that all three models were effective for this subject, with the MLP model having a slight edge. For Subject 46, the MLP model demonstrated the highest predictive performance with an R^2 score of 0.931, followed by the Random Forest model with an R^2 score of 0.929. The KNN model also performed strongly with an R^2 score of 0.922, but was slightly surpassed by the other two models. For Subject 38, the MLP model again accomplished the highest R^2 score of 0.813, indicating fitter performance than both the RF model (R^2 score of 0.807) and the KNN model (R^2 score of 0.800). All three models performed well, but the MLP model was the best among them for this subject. Finally, for Subject 54, the RF model exceeded the other models with an R^2 score of 0.800. The KNN model followed with an R^2 score of 0.789, while the MLP model had the lowest performance with an R^2 score of 0.738.

In summary, the performance of each model varied across different subjects, but overall, the RF and MLP models frequently demonstrated superior predictive capabilities. Specifically, the Random Forest model achieved the highest R^2 scores for Subject 52 (0.928) and Subject 54 (0.800), while the MLP model led for Subject 29 (0.888), Subject 46 (0.931), and Subject 38 (0.813). The KNN model showed strong performance but was generally outperformed by the RF and MLP models. These results underscore the value of using multiple models to identify the most effective predictive approach for different datasets.

We subsequently selected the best-performing model, identified by its lowest RMSE score of 14.977, as the most effective approach. To assess the robustness and generalizability of this model, we applied it to the test data of all subjects, evaluating its performance across the entire dataset. This approach allowed us to determine whether the optimized model could maintain its accuracy and reliability when exposed to diverse subject-specific data. Table III presents the model's applicability and performance metrics for each subject.

TABLE III. TESTING RESULTS FOR SUBJECTS ON BEST MODEL

ID	RMSE
52	RMSE: 31.300
29	RMSE: 22.552
46	RMSE: 43.736
38	RMSE: 18.716
54	RMSE: 14.977

V. DISCUSSION

In this study, we developed models to predict blood glucose levels of patients using machine learning algorithms. We tested three different algorithms: KNN, RF, and MLP on five different subject's datasets from the Diatrend dataset. The process for each subject's dataset consisted of training, validation, and testing of the models. The performance of these models was evaluated based on three metrics namely, R^2, MSE, and RMSE. However, for the scope of this study, we narrowed down our analysis for determining best performance to rely more heavily on R^2 and RMSE.

The model hyper-parameters chosen for each patient dataset impacted the RMSE and R^2 values differently for each subject. KNN algorithm did not yield a high performance in any of the subjects.

In case of random forest, for subjects 52 and 54, moderate values of hyper-parameters yielded the highest performance. For Subject 52, the results indicate that increasing the maximum depth and number of estimators generally improves model performance up to a certain point. The lowest MSE and RMSE values are observed at a maximum depth of 5 and 90 estimators, with an MSE of 15.124 and RMSE of 15.085. However, further increases in these hyper-parameters do not lead to substantial improvements and, in some cases, result in slightly worse performance. The results for Subject 54 show a slightly different set of trends. Here, the learning rate and the number of epochs play a crucial role in model performance. It is clear that excessively high or low learning rates lead to poor performance, as evidenced by the extremely high MSE values for learning rates of 0.01 and 0.00001. The most optimal performance is observed at a learning rate of 0.05 with 90 epochs, yielding an MSE of 17.055. While moderate values of hyper-parameters tend to yield better performance generally, the specific sensitivity varies between subjects.

Similarly, in case of subjects 29, 46 and 38, where MLP demonstrated highest validation performance, model values of learning rate and epochs resulted in a better model. For Subject 29, the results indicate that the MSE and RMSE tend to stabilize at lower values when the learning rate is set to 0.001, 0.005, or 0.0005, and the number of epochs ranges from 30 to 60. The best performance is seen with a learning rate of 0.0005 and 30 epochs, achieving the lowest MSE of 20.440. For Subject 46, a learning rate of 0.01 with 100 epochs yielded the lowest MSE of 26.489, suggesting that a higher learning rate combined with a longer training period can enhance performance. Conversely, extremely low learning rates (e.g., 0.0001 and 0.00005) resulted in significantly higher MSE values, highlighting the model could not converge even with a large number of epochs. For Subject 38, the optimal performance is observed with a learning rate of 0.05 and 70 epochs, achieving the lowest MSE of 17.264. Interestingly, very low learning rates such as 0.00001 lead to significantly higher MSE values, indicating poor performance and potentially inadequate learning. This suggests that for this subject, higher learning rates within a moderate range are more effective.

While general trends of moderate hyper-parameter values yielding better results are consistent, specific optimal configurations vary, underscoring the importance of subject-specific tuning for achieving the best predictive accuracy.

Test results evaluated using the best validation models provided a few other insights. For Subject 52, the RF model gave the best result with an R^2 score of 0.928 and RMSE of 17.484. The MLP model worked best for Subject 29 with an R^2 score of 0.888 and RMSE of 19.632 showing the highest performance. Similarly for Subject 46, the MLP model again performed best with the highest R^2 score of 0.931 and RMSE of 23.374. However, Subject 38 also had the MLP model giving the most accurate results with R^2 reaching up to 0.813 and a relatively low RMSE of 18.168. Lastly, for Subject 54, the RF model gave the best performance with an R^2 score of 0.800 and RMSE of 14.977.

Subject-specific performance analysis revealed variability in model performance across individuals. For Subject 29, the MLP model performed best, likely due to its ability to capture the non-linear glucose-insulin relationship. For Subject 46, the Random Forest model excelled, indicating that ensemble methods handled data variability effectively. Subject 54 showed lower R^2 scores across models, suggesting higher data variability or noise, which warrants further investigation.

Additionally, it became evident when the test data for all subjects was run through the best-performing model that creating a uniform, one-size-fits-all model would not be feasible. The results showed significant variability in RMSE scores among different subjects, emphasizing the inherent challenges in developing a single algorithm capable of delivering consistent performance across a diverse population. This variability suggests that subject-specific factors, such as unique glucose dynamics, lifestyle habits, and physiological differences, play a critical role in determining model accuracy. As a result, relying solely on a uniform model could lead to suboptimal outcomes for many individuals, further emphasizing the need to address these differences through tailored approaches.

These findings underscore the importance of adopting personalized modeling techniques rather than a universal solution. By designing models that account for individual characteristics and unique data patterns, it becomes possible to enhance prediction accuracy and optimize clinical outcomes for each subject. The high variability in RMSE scores also suggests that no single algorithm is universally superior for all patients, reaffirming the necessity for a more nuanced approach in algorithm selection and model development.

VI. Limitations & Future Work

The Diatrend dataset provides useful real-world insights, but its size and duration limit the applicability of the findings to a broader group of people. Since the data comes from just 31 days and five subjects, it might not fully capture the range of blood glucose patterns in a larger, more diverse population. This small group of subjects means the models might fit too closely to these individuals, making them less useful for generalization.

Our study focuses on using only past blood glucose levels and insulin pump data, as these two data sources provide direct and continuous indicators of glucose trends relevant to Type-1 Diabetes management. This targeted approach is common in many studies aiming to develop predictive models. While additional variables such as diet and exercise play a vital role in maintaining glucose levels in the human body, including them in our research would increase model complexity and data variability, potentially affecting model accuracy without adequate validation. Therefore, future studies could expand by integrating these broader data types to capture a more holistic picture.

We used simple models for their interpretability and computational efficiency in personalized predictions that analyze data in one-hour chunks. In the future, incorporating advanced models like LSTMs or TCNs could help examine longer temporal patterns, as seen in other studies. Additionally, integrating interpretability methods, such as Shapley Additive exPlanations (SHAP) values for assessing feature importance, could further enhance the clinical applicability of the models.

Future work should involve a larger number of subjects and longer data collection periods to assess model performance across diverse populations. Additionally, incorporating the rest of subjects' data from Diatrend into the modeling process could enhance the algorithms' adaptability and reliability by leveraging existing datasets to refine predictions and optimize performance. It is also important to test these models in actual healthcare settings to evaluate their reliability and usefulness. Integrating them with continuous glucose monitors and insulin pumps could pave the way for clinical trials.

Although the study's small sample size limits its broader applicability, we have optimized the models for the best performance with the given data. Expanding the dataset to include a larger and more diverse population should be a priority for future research.

To summarize, our study provides a strong foundation for further research in the field of blood-glucose level prediction. Future research could focus on additional model fine-tuning and testing other machine learning approaches. Other facets that can be considered include the impact of data quality and volume or understanding and leveraging intra-individual variability to improve accuracy. The ultimate goal lies in the development of an optimal prediction system, one that can adapt and learn from the inputs dynamically while being highly precise and reliable, offering a personalized solution for patients by integrating the prediction system into a closed-loop "artificial pancreas" system.

VII. Conclusion

Our research has established a foundation for an optimal blood glucose prediction system using supervised machine learning, employing three distinct algorithms: KNN, RF, and MLP. The results illustrated the promising potential of this research when further developed. Our models achieved significant predictive performance as indicated by RMSE and R^2 metrics, demonstrated their effectiveness in personalized glucose level prediction. While accuracy in classification tasks was not directly applicable here, the high R^2 values reflected the models' ability to explain a substantial proportion of variance in glucose levels.

References

[1] P. E. Cryer, J. N. Fisher, and H. Shamoon, "Hypoglycemia", *Diabetes care*, vol. 17, no. 7, pp. 734–755, 1994.

[2] P. E. Cryer, S. N. Davis, and H. Shamoon, "Hypoglycemia in diabetes", *Diabetes care*, vol. 26, no. 6, pp. 1902–1912, 2003.

[3] K. Dhatariya, L. Corsino, and G. E. Umpierrez, "Management of diabetes and hyperglycemia in hospitalized patients", 2015.

[4] M. Mouri and M. Badireddy, "Hyperglycemia", in *StatPearls [Internet]*, StatPearls Publishing, 2023.

[5] Y. Marcus *et al.*, "Improving blood glucose level predictability using machine learning", *Diabetes/Metabolism Research and Reviews*, vol. 36, no. 8, e3348, 2020.

[6] A. D. A. ADA, *Statistics about diabetes*, en, Nov. 2023.

[7] W. H. O. WHO, *The top 10 causes of death*, en, Dec. 2020.

[8] I. Ogrotis, T. Koufakis, and K. Kotsa, "Changes in the global epidemiology of type 1 diabetes in an evolving landscape of environmental factors: Causes, challenges, and opportunities", *Medicina*, vol. 59, no. 4, p. 668, 2023.

[9] C. D. Mathers and D. Loncar, "Projections of global mortality and burden of disease from 2002 to 2030", *PLoS medicine*, vol. 3, no. 11, e442, 2006.

[10] S. R. Colberg, R. Laan, E. Dassau, and D. Kerr, "Physical activity and type 1 diabetes: Time for a rewire?", *Journal of diabetes science and technology*, vol. 9, no. 3, pp. 609–618, 2015.

[11] I. H. Sarker, "Machine learning: Algorithms, real-world applications and research directions", *SN computer science*, vol. 2, no. 3, p. 160, 2021.

[12] J. Han, J. Pei, and H. Tong, *Data mining: concepts and techniques*. Morgan kaufmann, 2022.

[13] I. H. Sarker *et al.*, "Cybersecurity data science: An overview from machine learning perspective", *Journal of Big data*, vol. 7, pp. 1–29, 2020.

[14] F. Osisanwo *et al.*, "Supervised machine learning algorithms: Classification and comparison", *International Journal of Computer Trends and Technology (IJCTT)*, vol. 48, no. 3, pp. 128–138, 2017.

[15] M. Tauschmann and R. Hovorka, "Technology in the management of type 1 diabetes mellitus—current status and future prospects", *Nature Reviews Endocrinology*, vol. 14, no. 8, pp. 464–475, 2018.

[16] S. Templer, "Closed-loop insulin delivery systems: Past, present, and future directions", *Frontiers in Endocrinology*, vol. 13, p. 919 942, 2022.

[17] R. Hovorka, "Closed-loop insulin delivery: From bench to clinical practice", *Nature Reviews Endocrinology*, vol. 7, no. 7, pp. 385–395, 2011.

[18] R. M. Bergenstal *et al.*, "Safety of a hybrid closed-loop insulin delivery system in patients with type 1 diabetes", *Jama*, vol. 316, no. 13, pp. 1407–1408, 2016.

[19] T. Zhu, K. Li, P. Herrero, J. Chen, and P. Georgiou, "A deep learning algorithm for personalized blood glucose prediction.", in *KDH@ IJCAI*, 2018, pp. 64–78.

[20] X. Yu *et al.*, "Deep transfer learning: A novel glucose prediction framework for new subjects with type 2 diabetes", *Complex & Intelligent Systems*, pp. 1–13, 2021.

[21] E. M. Watson, M. J. Chappell, F. Ducrozet, S. Poucher, and J. W. Yates, "A new general glucose homeostatic model using a proportional-integral-derivative controller", *Computer methods and programs in biomedicine*, vol. 102, no. 2, pp. 119–129, 2011.

[22] T. Hamdi *et al.*, "Artificial neural network for blood glucose level prediction", in *2017 International Conference on Smart, Monitored and Controlled Cities (SM2C)*, IEEE, 2017, pp. 91–95.

[23] F. Allam, Z. Nossai, H. Gomma, I. Ibrahim, and M. Abdelsalam, "A recurrent neural network approach for predicting glucose concentration in type-1 diabetic patients", in *International Conference on Engineering Applications of Neural Networks*, Springer, 2011, pp. 254–259.

[24] J. Carrillo-Moreno *et al.*, "Long short-term memory neural network for glucose prediction", *Neural Computing and Applications*, vol. 33, pp. 4191–4203, 2021.

[25] R. Bunescu, N. Struble, C. Marling, J. Shubrook, and F. Schwartz, "Blood glucose level prediction using physiological models and support vector regression", in *2013 12th International Conference on Machine Learning and Applications*, IEEE, vol. 1, 2013, pp. 135–140.

[26] W. Yu, T. Liu, R. Valdez, M. Gwinn, and M. J. Khoury, "Application of support vector machine modeling for prediction of common diseases: The case of diabetes and pre-diabetes", *BMC medical informatics and decision making*, vol. 10, pp. 1–7, 2010.

[27] R. Mauseth *et al.*, "Use of a "fuzzy logic" controller in a closed-loop artificial pancreas", *Diabetes technology & therapeutics*, vol. 15, no. 8, pp. 628–633, 2013.

[28] J. I. Hidalgo *et al.*, "Data based prediction of blood glucose concentrations using evolutionary methods", *Journal of medical systems*, vol. 41, pp. 1–20, 2017.

[29] T. Zhu, K. Li, J. Chen, P. Herrero, and P. Georgiou, "Dilated recurrent neural networks for glucose forecasting in type 1 diabetes", *Journal of Healthcare Informatics Research*, vol. 4, pp. 308–324, 2020.

[30] H. Xingsan, Y. Xia, Y. Tao, and L. Hongru, "A deep transfer learning model for personalized blood glucose prediction", in *2021 China Automation Congress (CAC)*, IEEE, 2021, pp. 2045–2049.

[31] K. Plis, R. Bunescu, C. Marling, J. Shubrook, and F. Schwartz, "A machine learning approach to predicting blood glucose levels for diabetes management", in *Workshops at the Twenty-Eighth AAAI conference on artificial intelligence*, 2014.

[32] J. Xie and Q. Wang, "Benchmarking machine learning algorithms on blood glucose prediction for type i diabetes in comparison with classical time-series models", *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 11, pp. 3101–3124, 2020.

[33] D. Sisodia and D. S. Sisodia, "Prediction of diabetes using classification algorithms", *Procedia computer science*, vol. 132, pp. 1578–1585, 2018.

[34] J. Chen, K. Li, P. Herrero, T. Zhu, and P. Georgiou, "Dilated recurrent neural network for short-time prediction of glucose concentration.", in *KDH@ IJCAI*, 2018, pp. 69–73.

[35] A. Aliberti *et al.*, "A multi-patient data-driven approach to blood glucose prediction", *IEEE Access*, vol. 7, pp. 69 311–69 325, 2019.

[36] S. Langarica *et al.*, "Deep learning-based glucose prediction models: A guide for practitioners and a curated dataset for improved diabetes management", *IEEE Open Journal of Engineering in Medicine and Biology*, 2024.

[37] T. Prioleau, A. Bartolome, R. Comi, and C. Stanger, "Diatrend: A dataset from advanced diabetes technology to enable development of novel analytic solutions", *Scientific Data*, vol. 10, no. 1, p. 556, 2023.

[38] M. NirmalaDevi, S. A. alias Balamurugan, and U. V. Swathi, "An amalgam knn to predict diabetes mellitus", in *2013 IEEE International Conference ON Emerging Trends in Computing, Communication and Nanotechnology (ICECCN)*, 2013, pp. 691–695. DOI: 10.1109/ICE-CCN.2013.6528591.

[39] J. J. Khanam and S. Y. Foo, "A comparison of machine learning algorithms for diabetes prediction", *ICT Express*, vol. 7, no. 4, pp. 432–439, 2021, ISSN: 2405-9595. DOI: https://doi.org/10.1016/j.icte.2021.02.004.

[40] S. A. Quchani and E. Tahami, "Comparison of mlp and elman neural network for blood glucose level prediction in type 1 diabetics", in *3rd Kuala Lumpur International Conference on Biomedical Engineering 2006*, F. Ibrahim, N. A. A. Osman, J. Usman, and N. A. Kadri, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 54–58.

[41] J. Wu *et al.*, "Hyperparameter optimization for machine learning models based on bayesian optimizationb", *Journal of Electronic Science and Technology*, vol. 17, no. 1, pp. 26–40, 2019, ISSN: 1674-862X. DOI: https://doi.org/10.11989/JEST.1674-862X.80904120.

[42] 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, 2020. DOI: 10.1109/JBHI.2019.2931842.