INTERNATIONAL JOURNAL OF INTEGRATED ENGINEERING
ISSN: 2229-838X e-ISSN: 2600-7916

IJIE

Vol. 17 No. 6 (2025) 292-303
https://publisher.uthm.edu.my/ojs/index.php/ijie

Exploratory Anomaly Detection with Blood Glucose Level Time Series Prediction

Ade Anggian Hakim[1], Sharlini Singa Durai[1], Farhanahani Mahmud[1,2]*, Chin Fhong Soon[1,2], Zarina Tukiran[1], Rudi Setiawan[3]

[1] *Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia (UTHM), 86400 Parit Raja, Batu Pahat, Johor, MALAYSIA*
[2] *Microelectronics and Nanotechnology Shamsuddin Research Centre, Institute of Integrated Engineering, UTHM, 86400 Parit Raja, Batu Pahat, Johor, MALAYSIA*
[3] *Faculty of Industrial Technology, Institut Teknologi Sumatera, Terusan Ryacudu Street, Way Huwi, South Lampung Regency, Lampung, 35365, INDONESIA*

*Corresponding Author: farhanah@uthm.edu.my
DOI: https://doi.org/10.30880/ijie.2025.17.06.024

Article Info

Received: 30 May 2025
Accepted: 13 September 2025
Available online: 30 December 2025

Keywords

Blood glucose anomaly detection, LSTM, rule-based, Z-score, isolation forest

Abstract

Diabetes patients need effective blood glucose (BG) management to avoid developing serious health complications. Real-time BG prediction and anomaly detection through deep learning techniques improve diabetes care in this project. This study utilized the ShanghaiT1DM dataset to train Long Short-Term Memory networks for blood glucose prediction with a dataset split of 70% training and 30% testing aimed at optimizing a 30-minute prediction horizon. The study imputed missing data before validating stationarity through both the Augmented Dickey-Fuller and the Kwiatkowski-Phillips-Schmidt-Shin tests. The evaluation of anomaly detection methods included the rule-based approach alongside the statistical technique of Z-score and the machine learning algorithm of the isolation forest method. The highest accuracy for the detection of hypoglycemia was attained by the isolation forest (0.948), followed by the rule-based (0.887) and Z-score (0.791) methods. In the detection of hyperglycemia, the most effective method was the rule-based (0.847), followed by lower accuracies from the Z-score (0.715) and isolation forest (0.550) methods. Furthermore, the rule-based method exhibited superior performance in both the detection of hypoglycemia (accuracy = 0.887) and hyperglycemia (accuracy = 0.847), exhibiting high precision, recall, and F1-scores throughout, hence established as the strongest method for the detection of anomalies. The results from this research attest that the combination of LSTM-based prediction of blood glucose and rule-based detection of anomaly yields the most accurate method for the detection of hypoglycemia and hyperglycemia from the dataset analyzed. Though the rule-based method proved superior over statistical and machine learning methods, the Z-score and isolation forest methods retain potential for improvement.

1. Introduction

Diabetes, a prevalent metabolic disorder affecting millions globally, necessitates effective blood glucose (BG) management to mitigate complications and long-term health risks [1]. Traditional methods, requiring frequent

manual interventions, underscore the need for automated, accurate predictive systems [1]. Type 1 Diabetes Mellitus (T1DM), characterized by beta cell degradation, highlights the urgency for advanced monitoring and intervention strategies [2].

Blood glucose anomaly detection is critical for managing diabetes, where abnormal glucose levels can lead to severe health consequences. Various approaches have been proposed and refined, incorporating advancements in continuous glucose monitoring (CGM) systems; machine learning, deep learning and hybrid models. Rule-based and statistical approaches have been foundational in early anomaly detection methods [3,4]. Rule-based methods have been commonly used in clinical practice to trigger alerts when glucose levels cross predefined limits. In contrast, statistical methods like Z-score, autoregressive models and moving averages have been employed for glucose trend analysis. Machine learning and deep learning methods have significantly advanced blood glucose anomaly detection with supervised learning models like decision trees, support vector machines (SVM), random forests, long short-term memory networks (LSTM) and convolutional neural networks (CNN) have been widely used to detect anomalies in labelled glucose data and unsupervised learning like k-means clustering, isolation forests, and one-class SVM are used in detecting anomalies without needing predefined labels [5-8]. Additionally, hybrid models such as model predictive control and Bayesian and probabilistic models combine physiological knowledge with data-driven approaches for more robust anomaly detection [9-11].

The identification of the anomalies in predicted time-series blood glucose levels is established based on forecasting models such as LSTM for the prediction of short-term glucose dynamics and thereafter identifying the abnormal deviations based on residual analysis or machine learning methods such as Z-score, isolation forest, or autoencoders. By integrating this method with edge computing, real-time detection of hypoglycemia, hyperglycemia, and inconsistencies in data is achieved that will enable timely alerting independent of cloud connectivity. This has the potential for effective personalization in the management of diabetes, particularly for portable or wearable monitoring systems. Nonetheless, implementation in resource-constrained edge devices gives rise to concerns of limited memory, less processing power, and energy efficiency. Models therefore require optimization in direction toward lightweight computation while maintaining accuracy, robustness, and reliability in a wide range of real-world conditions. Addressing these concerns shall be a prerequisite for furthering effective real-time glucose monitoring solutions. In this direction, use of lightweight or less complex models becomes a prerequisite since not only is the computation overhead less, but a higher feasibility for constant execution by wearable and portable devices is induced, while maintaining nonetheless clinically useful accuracy [12-14].

This paper investigated BG anomaly detection by integrating LSTM-based time-series prediction and rule-based, Z-score, and isolation forest methods and tested them on the ShanghaiT1DM dataset. It is hoped that it offers comparative analysis of statistical and machine learning approaches used for predicted glucose levels for the detection of hypoglycemia and hyperglycemia as a reference for exploration of less computationally intensive models that are able to achieve clinically significant levels of accuracy. This connection is particularly significant for edge device implementation, whereby memory, compute power, and energy are limited. By connecting accurate computation of anomaly detection and the requirement for efficient computation, this work can serve as a springboard for further development of useful, portable, and real-time glucose systems that are suited for individualized diabetic therapy. Our contributions can be summarized as follows. 1) Demonstrates that integrating time-series blood glucose prediction using LSTM with anomaly detection techniques can effectively identify hypoglycemia and hyperglycemia events. 2) Provides a comparative evaluation of statistical and machine learning approaches in hypoglycemia and hyperglycemia detection, offering insights into the accuracy and robustness of the methods.

2. Related Work

In this section, we mainly review previous research related to anomaly detection in predicted time-series blood glucose levels.

2.1 Anomaly Detection from Model-predicted Glucose

Blood glucose (BG) prediction has increasingly relied on deep learning, particularly Long Short-Term Memory (LSTM) networks, in extracting temporal associations in glucose dynamics and enabling short-term prediction for active management of diabetes [15-17]. Accurate prediction, however, is not quite enough for applications, since the identification of anomalies is needed for early indication of hypoglycemia and hyperglycemia and also for the identification of sensor faults. Classical statistical approaches, such as Z-score residual analysis and adaptive filters, have been used for the identification of irregular deviations from predicted glucose levels [18], whereas machine learning approaches, such as isolation forests and autoencoders, offer enhanced functionalities for the identification of infusion-set malfunction and inherent complex irregularities [19-21].

Open-source databases such as ShanghaiT1DM [22] and T1DiabetesGranada [23] have made model building and model validation for those models feasible. However, it is challenging to implement those approaches in wearable or portable systems due to constraints in energy, computation, and memory. Recent research identifies

that model pruning, quantization, and lightweight models play a crucial role in edge-level computation-efficiency vs. prediction-robustness trade-off [25, 24]. Hence, available research emphasizes not only accurate BG prediction and anomaly detection but also resource-aware model construction that makes real-time, continuous execution feasible for wearable and edge systems.

3. Methodology

This study used the Chinese Diabetes Dataset ShanghaiT1DM [22], an open-access dataset. This dataset contains 16 data of continuous glucose monitoring (CGM) with 15-minute time intervals on data spanning 3 to 14 days for 12 individuals diagnosed with T1DM. The methodologies employed in this study are to develop and implement anomaly detection in predicted blood glucose levels for individuals with type 1 diabetes mellitus (T1DM). Python and its libraries (e.g., pandas, NumPy, matplotlib) are extensively used for data manipulation, visualization, and model development. The study leverages Google Colaboratory for its integrated Jupyter Notebook environment, facilitating efficient development and execution of Python code, which is particularly advantageous for machine learning tasks.

Fig. 1 shows the overall process to evaluate anomaly detection methods using LSTM-predicted blood glucose levels. It begins with data preparation, which includes preprocessing for missing data imputation and train/test split, and analysis for seasonality identification and stationarity testing. Following this, the LSTM neural network is developed for blood glucose prediction with the performance evaluation using Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) metrics. The procedure then diverges into three paths for anomaly detection: one employing a rule method, another employing the Z-score statistical method, and the last employing the Isolation Forest algorithm. Performance evaluation involves assessing prediction accuracy and comparing the effectiveness of the outlier detection methods. Finally, a confusion matrix was used to quantify the effectiveness of anomaly detection techniques, and accuracy, precision, recall, and F1 score were used to assess their performance.

3.1 Data Preparation

The methodology begins with the meticulous handling of the dataset. Only one dataset, named Dataset 3, from 16 datasets was specifically chosen for the result presentation due to a limited number of pages, and a missing timestamp was from Dataset 3. The missing value in the continuous glucose monitoring (CGM) variable was addressed using imputation techniques, calculating the average of the values immediately before and after the missing timestamp and using that average as the filled-in value tailored for univariate time series data. Python libraries such as Matplotlib, Pandas, NumPy, and PyTorch are employed for data manipulation, visualization, and deep learning model implementation.

The dimensionality and temporality of the dataset were verified to understand its structure and time-related aspects, followed by decomposing the time series into trend, seasonal, and residual components. Stationarity tests were conducted using the Augmented Dickey-Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests to assess the data's stability over time. In contrast to the ADF test, the KPSS test assumes stationarity as the null hypothesis. A p-value exceeding a specified significance level (e.g., 0.05) indicates stationarity in the series. Conversely, a p-value below the significance level suggests non-stationarity [26]. Moreover, ground truth labels for hypoglycemia, hyperglycemia, and normal glucose levels were generated based on predefined thresholds, ensuring the dataset's readiness for subsequent analysis.

3.2 LSTM Model Development for BG Time Series Prediction

Long Short-Term Memory (LSTM) deep learning model was employed to predict blood glucose levels, leveraging its capability to capture temporal dependencies in sequential data. The LSTM architecture consists of an input layer tailored for univariate time series data and a hidden layer comprising 50 LSTM units to extract temporal features. The output layer maps LSTM outputs to predicted glucose values, with a horizon of 2 and 4 representing 30-minute and 60-minute prediction intervals, respectively, calculated based on the dataset's time interval per data point. The dataset was split into 70% training and 30% testing sets, with sequences generated using a lookback of specified time steps. Training involves 200 epochs selected based on experimental results using Adam optimizer and Mean Squared Error (MSE) loss function, managed by PyTorch DataLoader for efficient batch processing. Performance evaluation encompasses tracking RMSE and MAE metrics.

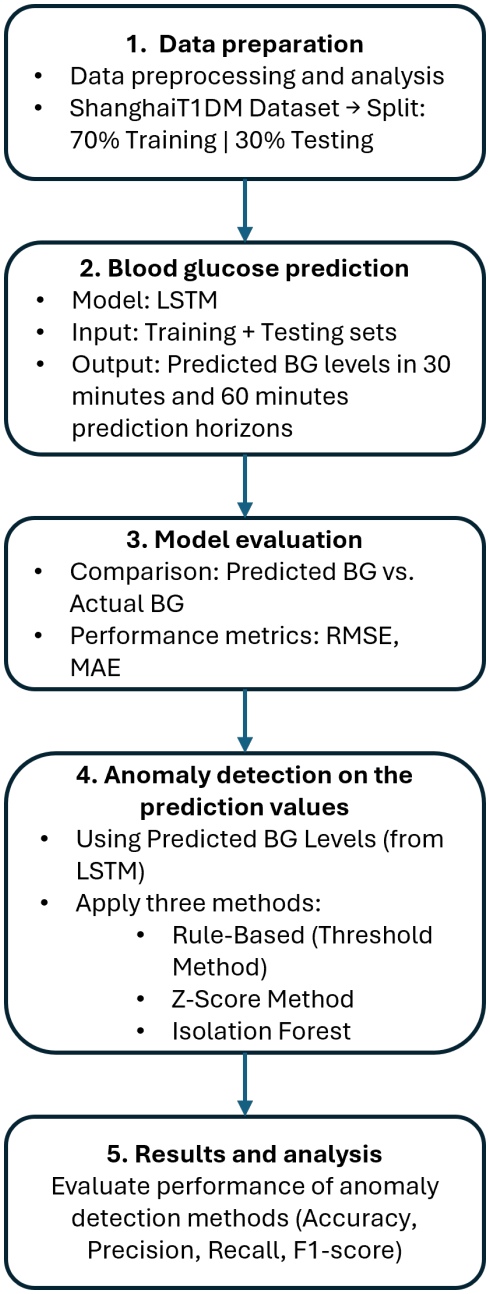


Fig. 1 *The overall evaluation process of anomaly detection using LSTM-based blood glucose predictions*

3.3 Anomaly Detection of Hypoglycemic and Hyperglycemic Events

The anomaly detection in this study employed three distinct techniques: rule-based, statistical-based, and machine learning-based detection. The rule-based method defines hypoglycemia and hyperglycemia thresholds at 70 mg/dL and 180 mg/dL [27], respectively, labelling anomalies as -1 for hypoglycemia and 1 for hyperglycemia. Statistical-based detection uses the Z-score method to identify outliers, with the Z-score threshold greater than 2 or less than -2 to detect moderate outliers [28]. On the other hand, a stricter criterion of Z-scores greater than the threshold of 3 or less than -3 identifies data points as highly unusual or extreme outliers. The choice of threshold depends on the sensitivity needed for the analysis: using a threshold of 3 (or -3) detects fewer, more extreme outliers, while a threshold of 2 (or -2) or less flags more, moderately unusual points. The Z-score was calculated using Equation 1, and it was then compared to the set thresholds. Where Z is the Z-score, X is the specific CGM data point, μ is the average of the CGM data, and σ is the standard deviation. In this work, the Z-score statistical-based anomaly was detected using three thresholds, which are 1.5, 2.0 and 2.5, to determine the best threshold. If the Z-score exceeded or under the defined threshold, the data points were classified as anomalies.

$$Z = (X - \mu) / \sigma \quad (1)$$

While the machine learning-based detection applies the isolation forest algorithm to classify data points as inliers or outliers, with an isolation contamination parameter set to 'auto'. This parameter determines the proportion of the dataset that the algorithm should consider anomalies [28, 29]. This automatic adjustment helps to adaptively identify a suitable contamination rate, which can be particularly useful when the exact proportion of anomalies in the data is unknown. This approach allows the Isolation Forest to effectively distinguish between normal and anomalous data points without requiring a predefined contamination fraction. Then the labels are assigned to anomalies such as hypoglycemia or hyperglycemia based on the blood glucose thresholds.

Dataset 3 was specifically selected for the result presentation for its missing timestamps, imputed with the mean glucose values before and after gaps [30]. Visualization of actual versus predicted glucose levels was performed using Python libraries, highlighting anomalies with scatter plots. Performance evaluation included comparing results against ground truth labels, calculating the accuracy, precision, recall, and F1-score for each detection method, to ensure a comprehensive assessment of the anomaly detection.

4. Results and Discussion

The ShanghaiT1DM dataset, which contains 16 datasets of time-series glucose data, was analyzed using the seasonal decompose function, revealing a clear trend component but no detectable seasonal pattern or significant residual noise. Stationarity tests, including the Augmented Dickey-Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests, confirmed the data's stationarity. The missing CGM value on the timestamp 2021-05-22 02:47:00 was imputed using the average of the values of the CGM immediately before and after the missing timestamp, which is 39.6 mg/dL. Fig. 2 shows the dataset 3 stationarity test result in (a) and was imputed with the average value as shown in (b). Additionally, baseline data of BG anomalies were created based on thresholds at 70 mg/dL and 180 mg/dL for all 16 datasets to classify events as hypoglycemia and hyperglycemia, respectively, for performance evaluation of the anomaly detection methods.

```
KPSS Test Results:
KPSS Statistic: 0.22909942985465473
p-value: 0.1
Critical Values:
10%: 0.347
5%: 0.463
2.5%: 0.574
1%: 0.739
KPSS Test Result: Data is stationary

ADF Test Results:
ADF Statistic: -5.865352558924688
p-value: 3.335542959868294e-07
Critical Values:
1%: -3.437439232284603
5%: -2.8646696476686477
10%: -2.5684364181154127
ADF Test Result: Data is stationary
<ipython-input-6-17791c9d368c>:57: InterpolationWarning:
```

(a)

	Date	glucose_value
0	2021-05-21 13:32:00	336.6
1	2021-05-21 13:47:00	343.8
2	2021-05-21 14:02:00	340.2
3	2021-05-21 14:17:00	331.2
4	2021-05-21 14:32:00	318.6
..	...	...
929	2021-05-31 05:47:00	72.0
930	2021-05-31 06:02:00	73.8
931	2021-05-31 06:17:00	81.0
932	2021-05-31 06:32:00	86.4
933	2021-05-31 06:47:00	88.2

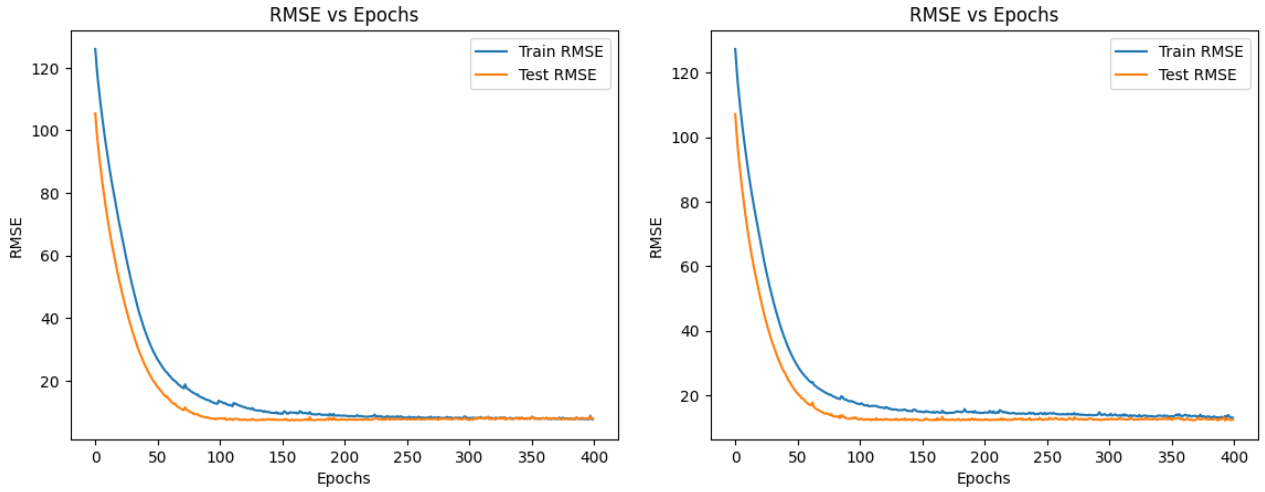
[934 rows x 2 columns]

Missing timestamps and their imputed values:
Timestamp: 2021-05-22 02:47:00, Imputed Value: 39.6

(b)

Fig. 2 *Data preprocessing results from Dataset 3. (a) Stationarity test results; (b) Imputed value using the average*

The LSTM model has been trained on 16 datasets over 200 epochs, with a lookback period of 10 time steps and prediction horizons (PHs) of 2- and 4-time steps, each 15 minutes apart. By epoch 200, the model achieved balanced performance between the model training and testing without overfitting. Fig. 3 shows the train and test learning curves representing RMSE vs the number of epochs for PHs 2 and 4 from Dataset 3. The trend for RMSE is generally decreasing over epochs as the model learns the patterns in the data. For a shorter PH, which is 2, the model can often capture the immediate patterns more effectively, resulting in a lower RMSE. The RMSE stabilises after fewer epochs compared to a larger PH, indicating that the model has learned the short-term dependencies well. Meanwhile, for PH 4, the trend for RMSE decreases a bit slowly and is higher overall compared to the horizon 2 case, as predicting further into the future is inherently more challenging.



(a) **(b)**

Fig. 3 *Train and test learning curves for prediction horizons (PHs) of (a) 2; and (b) 4-time steps from Dataset 3*

Table 1 shows the average RMSE and MAE values for PHs 2 and 4 obtained from 16 datasets. These results highlight that PH 2 balances accuracy and efficiency for short-term forecasting, making it optimal for anomaly detection.

Table 1 *Average RMSE and MAE values of blood glucose prediction for prediction horizons 2 and 4*

Prediction Horizon	RMSE (mg/dl)		MAE (mg/dl)	
	Train	Test	Train	Test
2 (30 min)	10.285	14.116	7.159	9.752
4 (60 min)	17.429	22.380	12.118	16.009

Anomalies were detected using three methods, which are the rule-based, Z-score and isolation forest machine learning model for all 16 datasets. In the Z-score statistical-based anomaly, it is found that threshold 1.5 detects anomalies better than thresholds 2.0 and 2.5, as more data points were relevantly detected as anomalies because more Z-scores of the data points exceeded or fell below the threshold of 1.5 compared to the thresholds of 2 and 2.5.

Table 2 shows the overall performance evaluation using accuracy, precision, recall, and F1-score key performance metrics for anomaly detection in hypoglycemia and hyperglycemia based on the rule, Z-score, with the threshold of 1.5 and isolation forest. The results indicate the average values of the performance metrics from the testing sets of the 16 datasets. Evaluating the anomaly detection methods for hypoglycemia and hyperglycemia reveals nuanced performance across the three different anomaly detection approaches. For hypoglycemia detection, the highest accuracy was 0.948, achieved by the isolation forest method. This was followed by the rule-based method with an accuracy of 0.887, and the Z-score method with 0.791. For hyperglycemia detection, the rule-based method achieved the highest accuracy of 0.847. The Z-score method recorded a lower accuracy of 0.715, and the isolation forest method achieved 0.550.

The rule-based method provides the best balance for hypoglycemia anomaly detection, with higher precision and recall. Meanwhile, the Z-score has the lowest recall and F1-score, indicating poorer performance in identifying true cases. The isolation forest shows the highest accuracy but relatively lower recall and F1-score, suggesting it might be overly conservative in its predictions. Moreover, the rule-based method significantly outperforms the other methods for hyperglycemia detection, showing the highest accuracy, precision, recall, and F1-score. The Z-score offers a reasonable balance but with lower performance metrics, and the isolation forest has the lowest accuracy and precision, indicating the weakest overall performance for hyperglycemia detection.

However, it is also interesting to note that the baseline data of the BG anomalies derived are from the actual data according to the threshold of hypoglycemia and hyperglycemia; thus, although the rule-based anomaly detection method performed better, it cannot be said that the performance of the Z-score and isolation forest methods is not good at all. Their performance may improve by using ready-labeled datasets to evaluate their performance, or for the machine learning model, its performance can vary based on whether they are trained on shared dataset or personalized for individual users.

Int. Journal of Integrated Engineering Vol. 17 No. 6 (2025) p. 292-303

Table 2 *Overall performance evaluation for hypoglycemia and hyperglycemia in rule-based, Z-score and isolation forest anomaly detection methods*

Anomaly detection method	Hypoglycemia				Hyperglycemia			
	Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score
Rule-based	0.887	0.626	0.552	0.525	0.948	0.847	0.8553	0.849
Z-score	0.791	0.475	0.255	0.281	0.715	0.532	0.443	0.428
Isolation forest	0.948	0.545	0.364	0.417	0.550	0.302	0.490	0.315

Fig. 4, Fig. 5 and Fig. 6 show the actual versus predicted plots of BG values in PH 2 with anomaly detection data points of Dataset 3 in rule-based, Z-score with the threshold of 1.5 and isolation forest anomaly detection methods, respectively. From the results of the rule-based approach shown in Fig. 4, the red line represents actual glucose levels, while the blue and green lines show predictions for the training and testing using Dataset 3. Magenta and yellow markers on the blue line indicate instances of predicted hyperglycemia and hypoglycemia in the training set, whereas cyan and purple markers on the green line highlight these anomalies in the testing set. Gray markers denote normal glucose levels for both datasets. Meanwhile, Fig. 5 displays the anomalies in Dataset 3, detected using a Z-score threshold of 1.5. The magenta marker indicates both hyperglycemia and hypoglycemia in the training set, while the purple marker represents hypoglycemia and hyperglycemia in the testing set. Whilst, according to the results of the isolation forest method shown in Fig. 6, hyperglycemia anomalies are marked in magenta and hypoglycemia anomalies in cyan for the training set, while hyperglycemia anomalies are purple and hypoglycemia anomalies are yellow for the testing set. These visual representations demonstrate each of the approaches' ability to identify abnormal glucose levels, showcasing their potential for glucose monitoring and management.

Table 3 presents the confusion matrix analysis for anomaly detection using the three methods for hypoglycemia and hyperglycemia in the testing set of 270 data points, with 74 positive and 196 negative cases for hypoglycemia and 18 positive and 252 negative cases for hyperglycemia, of Dataset 3. From the rule-based anomaly detection for hypoglycemia, the confusion matrix values indicate that the model correctly identified 67 true positives (TP) and 187 true negatives (TN), while it incorrectly classified 9 false positives (FP) and 7 false negatives (FN). The accuracy of the model, which measures the proportion of correctly classified instances, is high at 0.9407, indicating reliable overall performance. Precision, which indicates the proportion of true positives among all positive predictions, is 0.8816, suggesting that most positive predictions are correct. Recall, reflecting the model's ability to identify actual positives, is 0.9054, demonstrating good sensitivity. The F1-score, a harmonic mean of precision and recall, is 0.8933, showing a balanced performance between precision and recall. Overall, these metrics indicate that the method performs well in detecting outliers for hypoglycemia. Meanwhile, for hyperglycemia, the confusion matrix shows that the model correctly identified 17 TP and 248 TN, with 4 FP and 1 FN. The accuracy, indicating the proportion of correctly classified instances, is very high at 0.9815, reflecting strong overall performance. The precision, which measures the proportion of true positives among all positive predictions, is 0.8095, suggesting that most positive predictions are correct. The recall, indicating the model's ability to detect actual positives, is 0.9444, demonstrating high sensitivity. The F1-score, a harmonic mean of precision and recall, is 0.8718, indicating a balanced performance between precision and recall. Overall, these metrics indicate that the model is highly effective in detecting hyperglycemic levels using the rule-based method.

From the Z-score threshold 1.5 anomaly detection for hypoglycemia, the confusion matrix shows that the model did not identify any true positive and also did not falsely identify any cases as hypoglycemia (FP: 0). However, it failed to detect 74 cases of actual hypoglycemia, classifying them all as false negatives (FN: 74), while correctly identifying 196 instances as not hypoglycemic (TN: 196). The accuracy, which represents the proportion of correct classifications, is 0.7259. Precision and F1-score are not defined (nan) because there are no true positive predictions, rendering these metrics meaningless in this context. The recall is 0.0, indicating that the model did not detect any instances of hypoglycemia. These results suggest that the model is ineffective for hypoglycemia detection, essentially indicating that the test set does not contain any hypoglycemic instances according to the model's predictions. Whereas for hyperglycemia, the model correctly identified all 18 hyperglycemia cases (TP: 18) without missing any (FN: 0) but also falsely flagged 15 non-hyperglycemia cases (FP: 15). It correctly identified 237 non-hyperglycemia cases (TN: 237). The accuracy is high at 0.9444. Precision is moderate at 0.5455 due to the false positives, but recall is perfect at 1.0, meaning all hyperglycemia cases were detected, and the F1-score is 0.7059, indicating moderate overall performance in Z-score.

According to the isolation forest model for hypoglycemia, the model achieved high accuracy (0.9185) with 57 TP, 191 TN, 5 FP, and 17 FN. Precision is 0.9194, indicating most identified hypoglycemia cases are correct. The recall, reflecting the model's ability to identify actual hypoglycemia cases, is 0.7703, indicating some missed detections. The F1-score is 0.8382, showing a good balance between precision and recall. Overall, the model effectively detects hypoglycemia with high precision and good overall performance. For hyperglycemia, the

confusion matrix in Table 3 shows that the model correctly identified 17 TP and 180 TN, but it also produced 72 FP and missed 1 actual hyperglycemia case (FN). The model's accuracy is 0.7296, indicating a moderate proportion of correct classifications. Precision is low at 0.191, suggesting that many identified hyperglycemia cases are incorrect. However, recall is very high at 0.9444, meaning the model successfully detected nearly all actual hyperglycemia cases. The F1-score is 0.3178, reflecting the imbalance between precision and recall.

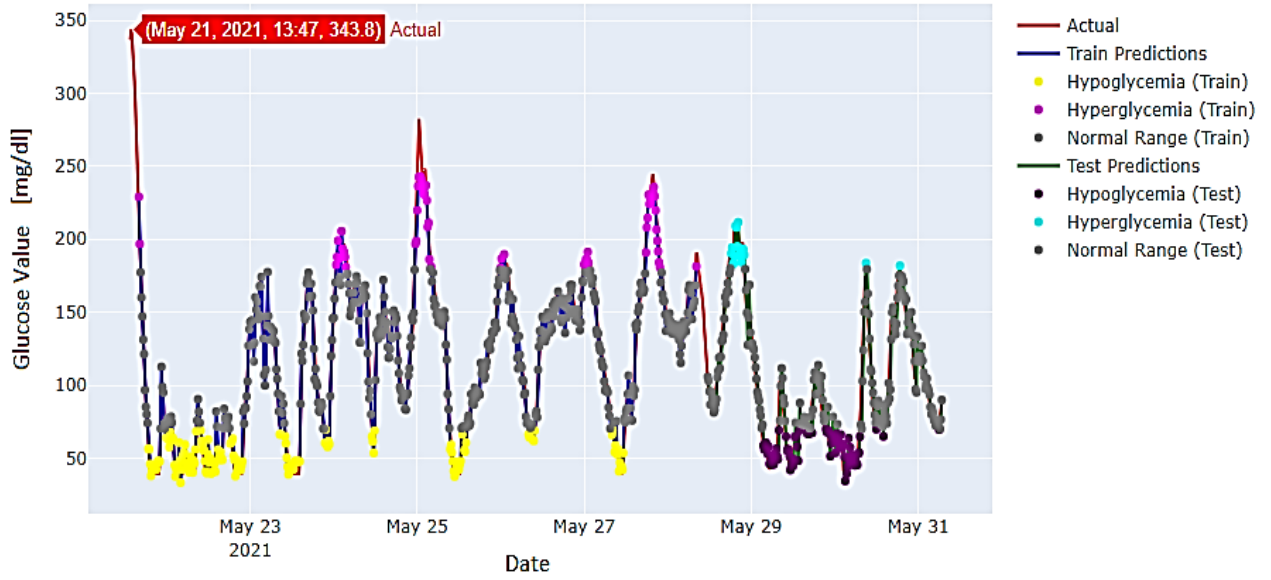


Fig. 4 *Actual versus predicted plots with anomaly detection data points of Dataset 3 in the rule-based anomaly detection method*

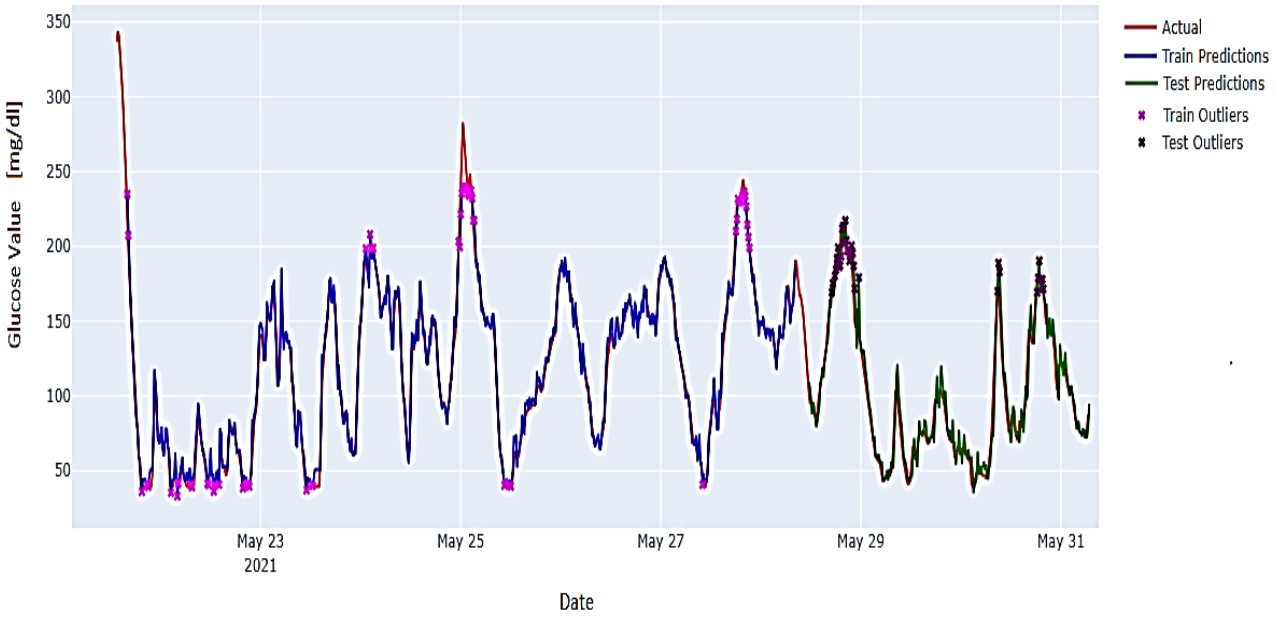


Fig. 5 *Actual versus predicted plots with anomaly detection data points of Dataset 3 in Z-score anomaly detection method*

Int. Journal of Integrated Engineering Vol. 17 No. 6 (2025) p. 292-303

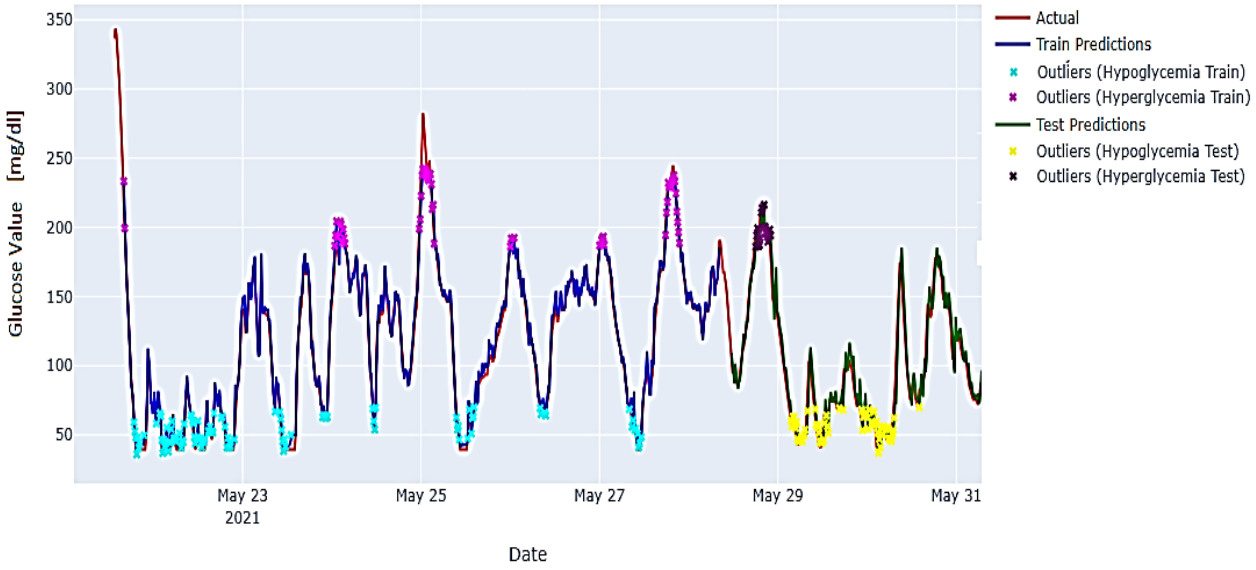


Fig. 6 *Actual versus predicted plots with anomaly detection data points of Dataset 3 in isolation forest anomaly detection method*

Table 3 *Confusion matrix analysis for anomaly detection using the three methods for hypoglycemia and hyperglycemia in the testing set of Dataset 3*

Anomaly detection method	Anomaly type	Confusion Matrix	Accuracy	Precision	Recall	F1-score
Rule-based	Hypoglycemia	TP: 67 FP: 9 FN: 7 TN: 187	0.9407	0.8816	0.9054	0.8933
	Hyperglycemia	TP: 17 FP: 4 FN: 1 TN: 248	0.9815	0.8095	0.9444	0.8718
Z-score	Hypoglycemia	TP: 0 FP: 0 FN: 74 TN: 196	0.7259	nan	0.0	nan
	Hyperglycemia	TP: 18 FP: 15 FN: 0 TN: 237	0.9444	0.5455	1.0	0.7059
Isolation forest	Hypoglycemia	TP: 57 FP: 5 FN: 17 TN: 191	0.9185	0.9194	0.7703	0.8382
	Hyperglycemia	TP: 17 FP: 72 FN: 1 TN: 180	0.7296	0.191	0.9444	0.3178

5. Conclusion

This work aims to analyze time-series blood glucose data, develop a deep learning Long Short-Term Memory (LSTM) model for the BG prediction, and explore the rule-based, Z-score statistical and isolation forest machine learning anomaly detection methods on the BG prediction using the ShanghaiT1DM dataset. The dataset analysis confirmed stationarity and addressed missing values in dataset 3 through imputation. LSTM model showed promise in predicting glucose values, with a forecasting horizon of two (30 minutes) proving optimal for accurate short-term predictions. The rule-based method performed better for both hypoglycemia and hyperglycemia anomaly detection, while the Z-score method showed averaged performance outcomes in both hypoglycemia and hyperglycemia detections and the isolation forest was particularly limited in detecting hyperglycemia. While the rule-based method showed superior performance in this study, the results do not preclude the potential of the Z-score and isolation forest methods. Their accuracy could improve when evaluated on pre-labeled datasets, and machine learning models may yield better results when tailored to individual users rather than relying solely on shared datasets. Thus, it is important to understand the strengths and limits of these models to use them effectively in managing glucose level issues. Future work will enhance LSTM models with autoencoders and diabetes-specific features, and optimize them for edge deployment to enable real-time, personalized blood glucose monitoring.

Acknowledgement

The communication of this research was made possible with the support of the TIER 1 Grant No. Q872 and the Institute of Integrated Engineering, Universiti Tun Hussein Onn Malaysia, through the Microelectronics and Nanotechnology – Shamsuddin Research Centre Fund (E15225). The authors would also like to thank Microelectronics & Nanotechnology-Shamsuddin Research Centre, Universiti Tun Hussein Onn Malaysia, for the related facilities.

Conflict of Interest

The manuscript has not been published elsewhere and is not under consideration by other journals. All authors have approved the review, agree with its submission and declare no conflict of interest on the manuscript

Author Contribution

The authors confirm contribution to the paper as follows: ***study conception and design:*** *Sharlini Singa Durai, Farhanahani Mahmud, Ade Anggian Hakim;* ***data collection:*** *Sharlini Singa Durai;* ***analysis and interpretation of results:*** *Sharlini Singa Durai, Farhanahani Mahmud, Ade Anggian Hakim;* ***draft manuscript preparation:*** *Ade Anggian Hakim, Farhanahani Mahmud, Sharlini Singa Durai;* ***draft manuscript review:*** *Chin Fhong Soon, Zarina Tukiran, Rudi Setiawan. All authors reviewed the results and approved the final version of the manuscript.*

References

[1] Woldaregay, A. Z., Årsand, E., Botsis, T., Albers, D., Mamykina, L., & Hartvigsen, G. (2019). Data-driven blood glucose pattern classification and anomalies detection: machine-learning applications in type 1 diabetes. Journal of medical Internet research, 21(5), e11030.

https://doi.org/10.2196/11030

[2] R. Khardori, "Type 1 Diabetes Mellitus: Practice Essentials, Background, Pathophysiology," Medccape, Dec. 13, 2023. [Online]. Available: https://emedicine.medscape.com/article/117739-overview

[3] Li, G., & Jung, J. J. (2023). Deep learning for anomaly detection in multivariate time series: Approaches, applications, and challenges. Information Fusion, 91, 93-102.

https://doi.org/10.1016/j.inffus.2022.10.008

[4] Yang, M., & Zhang, J. (2023). Data anomaly detection in the internet of things: A review of current trends and research challenges. International Journal of Advanced Computer Science and Applications, 14(9).

https://doi.org/10.14569/IJACSA.2023.0140901

[5] Meneghetti, L., Terzi, M., Del Favero, S., Susto, G. A., & Cobelli, C. (2018). Data-driven anomaly recognition for unsupervised model-free fault detection in artificial pancreas. *IEEE Transactions on Control Systems Technology*, *28*(1), 33-47.

https://doi.org/10.1109/TCST.2018.2885963

[6] Hidalgo, J. I., Colmenar, J. M., Kronberger, G., Winkler, S. M., Garnica, O., & Lanchares, J. (2017). Data based prediction of blood glucose concentrations using evolutionary methods. *Journal of medical systems*, *41*(9), 142.

https://doi.org/10.1007/s10916-017-0788-2

[7] Rajeswari, A. M., Yalini, S. K., Janani, R., Rajeswari, N., & Deisy, C. (2018, April). A comparative evaluation of supervised and unsupervised methods for detecting outliers. In *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)* (pp. 1068-1073). IEEE.

https://doi.org/10.1109/ICICCT.2018.8473123

[8] Yang, X., Qi, X., & Zhou, X. (2023). Deep learning technologies for time series anomaly detection in healthcare: A review. *IEEE Access*, *11*, 117788-117799.

https://doi.org/10.1109/ACCESS.2023.3325896

[9] Kang, S. L., Hwang, Y. N., Kwon, J. Y., & Kim, S. M. (2022). Effectiveness and safety of a model predictive control (MPC) algorithm for an artificial pancreas system in outpatients with type 1 diabetes (T1D): systematic review and meta-analysis. Diabetology & metabolic syndrome, 14(1), 187. https://doi.org/10.1186/s13098-022-00962-2

[10] Sun, Y., & Kosmas, P. (2024). Integrating Bayesian approaches and expert knowledge for forecasting continuous glucose monitoring values in type 2 diabetes mellitus. *IEEE Journal of Biomedical and Health Informatics*. https://10.1109/JBHI.2024.3472077

[11] Patharkar, A., Cai, F., Al-Hindawi, F., & Wu, T. (2024). Predictive modeling of biomedical temporal data in healthcare applications: review and future directions. *Frontiers in Physiology*, *15*, 1386760. https://doi.org/10.3389/fphys.2024.1386760

[12] Barbato, M. P., Rigamonti, G., Marelli, D., & Napoletano, P. (2025). Lightweight Sequential Transformers for Blood Glucose Level Prediction in Type-1 Diabetes. arXiv preprint arXiv:2506.07864. https://doi.org/10.48550/arXiv.2506.07864

[13] Farahmand, E., Soumma, S. B., Chatrudi, N. T., & Ghasemzadeh, H. (2024). Hybrid attention model using feature decomposition and knowledge distillation for glucose forecasting. *arXiv preprint arXiv:2411.10703*. https://doi.org/10.48550/arXiv.2411.10703

[14] Gragnaniello, M., Marrazzo, V. R., Borghese, A., Maresca, L., Breglio, G., & Riccio, M. (2024). Edge-AI Enabled Wearable Device for Non-Invasive Type 1 Diabetes Detection Using ECG Signals. Bioengineering, 12(1), 4. https://doi.org/10.3390/bioengineering12010004

[15] Hakim, A. A., Mahmud, F., & Morsin, M. (2024). Assessment of Deep Learning Model System for Blood Glucose Time-Series Prediction. *Journal of Science and Technology*, *16*(1), 65-75. https://publisher.uthm.edu.my/ojs/index.php/JST/article/view/16370

Butunoi, B. P., Stolojescu-Crisan, C., & Negru, V. (2024, September). Blood Glucose Prediction in Type 1 Diabetes Based on Long Short-Term Memory. In International Conference on Computational Collective Intelligence (pp. 458-469). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-70259-4_35

[16] Jaloli, M., & Cescon, M. (2023). 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, 17(6), 1590-1601. https://doi.org/10.1177/19322968221092785

[17] Manzoni, E., Rampazzo, M., & Del Favero, S. (2021). Detection of Glucose Sensor Faults in an Artificial Pancreas Via Whiteness Test on Kalman Filter Residuals. IFAC-PapersOnLine, 54(7), 274-279. https://doi.org/10.1016/j.ifacol.2021.08.371

[18] Idi, E., Prendin, F., Sparacino, G., & Del Favero, S. (2024). Autoencoder-based Detection of Insulin Pump Faults in Type 1 Diabetes Treatment. *IEEE Journal of Biomedical and Health Informatics*. https://doi.org/10.1109/JBHI.2024.3518233

Idi, E., Facchinetti, A., Sparacino, G., & Del Favero, S. (2024). Supervised and Unsupervised Approaches for the Real-time Detection of Undesired Insulin Suspension Caused by Malfunctions. *Journal of Diabetes Science and Technology*, 19322968241248402. https://doi.org/10.1177/19322968241248402

[19] Meneghetti, L., Dassau, E., Doyle III, F. J., & Del Favero, S. (2022). Machine Learning-based Anomaly Detection Algorithms to Alert Patients Using Sensor Augmented Pump of Infusion Site Failures. *Journal of Diabetes Science and Technology*, *16*(3), 641-648.

https://doi.org/10.1177/1932296821997854

[20] Zhao, Q., Zhu, J., Shen, X., Lin, C., Zhang, Y., Liang, Y., Cao, B., Li, J., Liu, X., Rao, W., & Wang, C. (2023). Chinese diabetes datasets for data-driven machine learning. Scientific Data, 10(1), 35.

https://doi.org/10.1038/s41597-023-01940-7

[21] Rodriguez-Leon, C., Aviles-Perez, M. D., Banos, O., Quesada-Charneco, M., Lopez-Ibarra Lozano, P. J., Villalonga, C., & Munoz-Torres, M. (2023). T1DiabetesGranada: a longitudinal multi-modal dataset of type 1 diabetes mellitus. Scientific Data, 10(1), 916. https://doi.org/10.1038/s41597-023-02737-4

[22] Alzoubi, Y. I., Topcu, A. E., & Elbasi, E. (2025). A Systematic Review and Evaluation of Sustainable AI Algorithms and Techniques in Healthcare. IEEE Access. https://doi.org/10.1109/ACCESS.2025.3596189

[23] Rourke, C., & Leclair, M. (2025). Scalable and Secure Edge AI: Foundations, Applications, and Open Research Issues. Transactions on Computational and Scientific Methods, 5(6). https://doi.org/10.5281/zenodo.15703866

[24] Kumar, G. V. (2024, May 5). Statistical Tests to Check Stationarity in Time Series. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2021/06/statistical-tests-to-check-stationarity-in-time-series-part-1/"Blood glucose monitoring," Diabetes Australia, June 1, 2023. [Online]. Available: https://www.diabetesaustralia.com.au/managing-diabetes/blood-glucose-monitoring/

[25] Mensi, A., & Bicego, M. (2021). Enhanced anomaly scores for isolation forests. Pattern Recognition, 120, 108115. https://doi.org/10.1016/j.patcog.2021.108115

[26] Dhiraj, K. (2021, April 9). Anomaly Detection Using Isolation Forest in Python. *Paperspace*. https://blog.paperspace.com/anomaly-detection-isolation-forest/

[27] Secherla, S. (2021, January 13). Different Imputation Methods to Handle Missing Data. *Medium*. https://towardsdatascience.com/different-imputation-methods-to-handle-missing-data-8dd5bce97583V.