ISSN: 2278-4632
Vol-14, Issue-9, September: 2024

Analysis of Blood Glucose level using improved Time Series Forecasting Machine Learning Technique in Big Data Processing of Medical Data

S. Tamilarasan[#1], Dr. S.K. Mahendran[*2], Dr. S. Pandikumar[#3]

[#1] Research Scholar, Department of Computer Science, Centre for Research and Evaluation, Bharathiar University, Coimbatore, TamilNadu, India

[*2]Assistant Professor, P.G. & Research Department of Computer Science Government Arts College, Coimbatore, TamilNadu, India

[#3]Assistant Professor, Department of Computer Applications, Yadava College, Madurai, TamilNadu, India

[#1] tamilarasanonline@gmail.com

*2 sk.mahendran@yahoo.co.in

[#3] pandikmr19986@gmail.com

ABSTRACT:

Diabetic patients face various health challenges due to their body's inability to regulate blood sugar levels effectively. Diabetes, particularly Type 2, is associated with increased risks of cardiovascular diseases, kidney damage, nerve damage, and other complications that can affect their overall survival and quality of life. This paper presents big data processing architecture for medical data collected from glucometers using the ESP8266 microcontroller and SQL Server, integrating machine learning models for time series forecasting of glucose levels. To ensure data security and privacy, the system implements encryption protocols during data transmission and access control mechanisms in the SQL Server. The architecture's scalability supports the growing volume of medical data, making it adaptable for broader healthcare applications. This approach demonstrates the potential of integrating IoT devices, SQL Server, big data, and machine learning in enhancing the monitoring and management of chronic conditions such as diabetes.

Keywords: Big Data, Machine Learning, MQTT, Time Series Forecasting, ESP8266, SQL Server

1. INTRODUCTION

Blood glucose monitoring is essential for diabetic patients to manage their health effectively. Traditional methods often rely on periodic blood tests, which may not provide real-time insights into glucose fluctuations. Advances in technology, such as continuous glucose monitoring (CGM) devices, generate large volumes of data that can be leveraged for improved glucose forecasting. However, the sheer scale of this data presents challenges in storage, processing, and analysis, necessitating the use of big data technologies. These improved ML models have significant implications for personalized medicine. By accurately predicting blood glucose levels, healthcare providers can tailor treatments to individual patients, thus enhancing the overall quality of care. Future research may focus on integrating additional health metrics, improving data security, and developing real-time monitoring systems to provide continuous feedback to patients and clinicians. Ganjar Alfian et. al. (2019) Predicting future blood glucose (BG) level for diabetic patients will help them to avoid critical conditions in the future. The clinical dataset of Type 1 Diabetes (T1D) patients was utilized and the prediction models were generated to predict future BG of 30 and 60 minutes ahead of time. [15].

2. RELATED WORK

Arvind Kumar Pandey et. al ((2022) proposed a novel technique in type 2 diabetes based heart disease detection in big data predictive analysis using machine learning method. Input data has been collected as type 2 diabetes and processed for noise removal and dimensionality reduction. Then the processed data features has been extracted for detecting the abnormality of type 2 diabetes using regression model based linear discriminant analysis. The extracted features shows the abnormal type 2 diabetes and for predicting heart disease by classifying the extracted data using VGG-16 Net_gradient NN. Experimental analysis has been carried out in terms of accuracy, precision, recall, F-1 score, RMSE and MAP for various diabetes dataset. Proposed technique attained accuracy of 96%, precision of 67%, recall of 79%, F-1 score of 63%, RMSE of 66% and MAP of 68%[1]. Aditi Chopra et. al . (2024). Applied The Random Forest Regressor Model, with features identified using the wrapper method, was selected as the best, with an average RMSE of 43.28. The prediction intervals were computed for point estimate, MAE = 23.821, and coverage was 100 percent, the clinical accuracy was compared with that of glucometers and continuous monitoring systems. All predicted values are in Zones A and B of the Clarke error grid, and the bias was 6.41. The most important feature for predicting blood glucose level is salivary glucose level, followed by known

ISSN: 2278-4632
Vol-14, Issue-9, September: 2024

risk factors like Family History, BMI, etc. The study found that salivary glucose levels are insufficient to classify blood glucose levels as high or normal.

Fig. 1. Digital Glucometer

3. METHODOLOGY

The proposed system addresses the challenges of real-time data acquisition, storage, processing, and predictive analytics in healthcare.

3.1Data Collection and Transmission:

Data is collected from glucometers connected to the ESP8266, which transmits the readings via Wi-Fi using lightweight communication protocols such as Message Queuing Telemetry Transport (MQTT) protocol, A lightweight messaging protocol often used for IoT devices due to its low bandwidth consumption, ideal for transmitting glucometer data.

3.2 HTTP/HTTPS: For secure data transmission between the ESP8266 and the SQL Server or cloud services.

3.3 SQL SERVER:

The data is stored in a SQL Server database designed to handle time-series data, ensuring the efficient management of large volumes of patient records. The system uses Extract, Transform, Load (ETL) pipelines for data preprocessing, including validation, cleaning, and transformation into structured formats suitable for analysis. The core of the system is a machine learning module designed to perform time series forecasting using models such as

3.4 Long Short-Term Memory (LSTM):

Long ShortTerm Memory networks, a specific recurrent neural network topology type, maintains short-term memories (RNN). It is frequently applied to challenges dependent on time series data, including speech recognition, handwriting recognition, healthcare pathway prediction, etc. Because it uses the "remember" mechanism through a succession of gates, LSTM is uniquely suited to manage long-term

ISSN: 2278-4632
Vol-14, Issue-9, September: 2024

dependencies in contrast to ordinary RNNs. This aspect is appropriate for the glucose prediction problem's possible prolonged observation window, which challenges other machine-learning techniques. The LSTM prediction algorithm starts with importing libraries such as pandas, numpy, seaborn, statsmodel, math, datetime and matplotlib. Then import the prepared dataset as a csv file. The dataset is segregated into two parts: one is a training set, and another one is a testing set. By fitting the LSTM model, blood glucose and values are predicted.

3.5 Auto Regressive Integrated Moving Average (ARIMA)

Box and Jenkins introduced the Autoregressive Integrated Moving Average process model (ARIMA), a linear time series analysis and forecasting model. Auto-Regressive (AR), Moving Average (MA), Seasonal Autoregressive Integrated Moving Average (SARIMA), and other models are all included in the ARIMA model. Three parameters must be specified for time series analysis they are moving average order (q), difference order (d), and autoregressive order (p). This study uses this prediction-based algorithm to forecast cholesterol and blood glucose levels. The Autoregressive Integrated Moving Average (ARIMA) model is created with an adaptive identification algorithm to increase the precision and enforceability of predicting blood glucose by adopting the time-varying non-stationary continuous glucose data.

3.6 Evaluation metrics:

The values for predicting blood glucose levels using time series forecasting models such as LSTM and ARIMA can vary significantly depending on the quality of data, model parameters, and preprocessing steps.

Mean Absolute Error (MAE): this Measures the average magnitude of the errors in the predictions, without considering their direction. It is calculated as:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|$$

Root Mean Squared Error (RMSE): this is the root of MSE, which gives a measure of the error in the same units as the target variable.

$$\mathrm{RMSE} = \sqrt{\mathrm{MSE}}$$

ISSN: 2278-4632
Vol-14, Issue-9, September: 2024

R-Squared (Coefficient of Determination): this statistic indicates the proportion of the variance in the dependent variable that is predictable from the independent variables. It is calculated as

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Mean Absolute Percentage Error (MAPE):

MAPE = frac{100}{n} \sum_{i=1}^{n} \left|\frac{y_i - \hat{y}_i}{y_i}

These models predict future glucose levels based on historical data, enabling proactive healthcare management and personalized patient care.

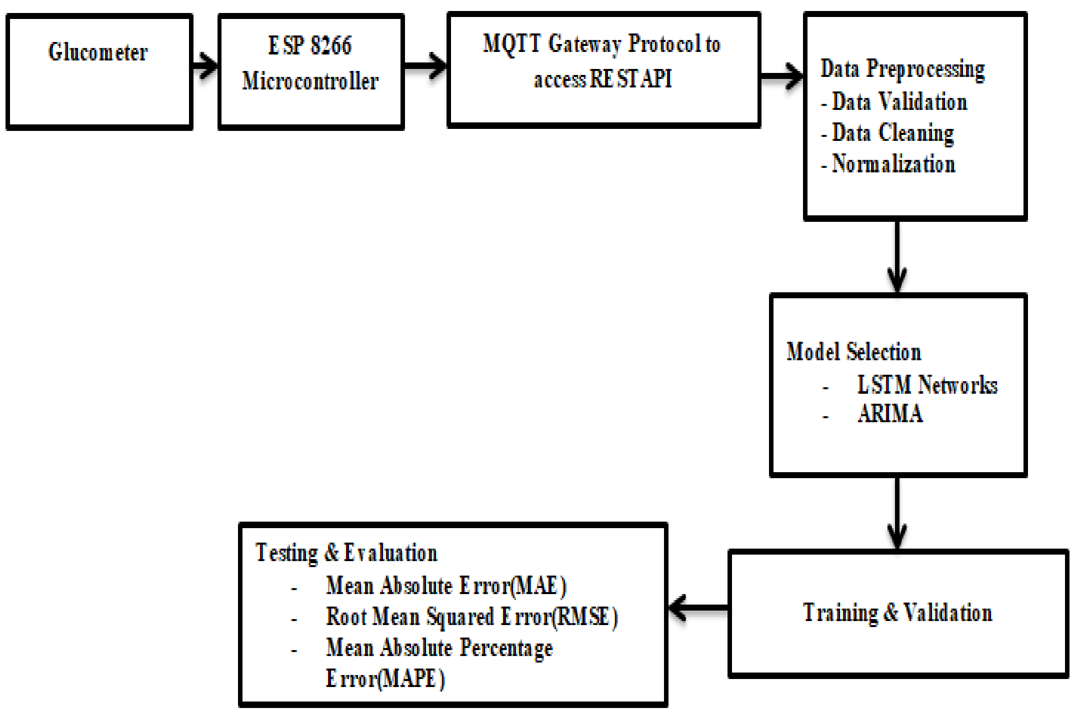

Fig. 2. Block Diagram of Proposed Methodology

4. RESULTS & DISCUSSIONS

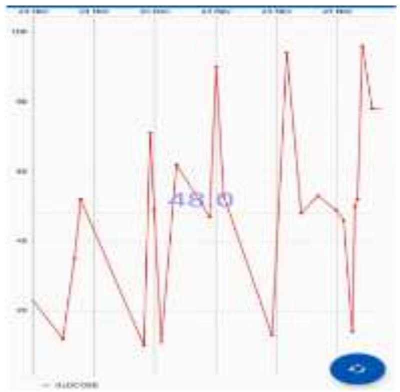

Fig. 3. Blood glucose stored in cloud

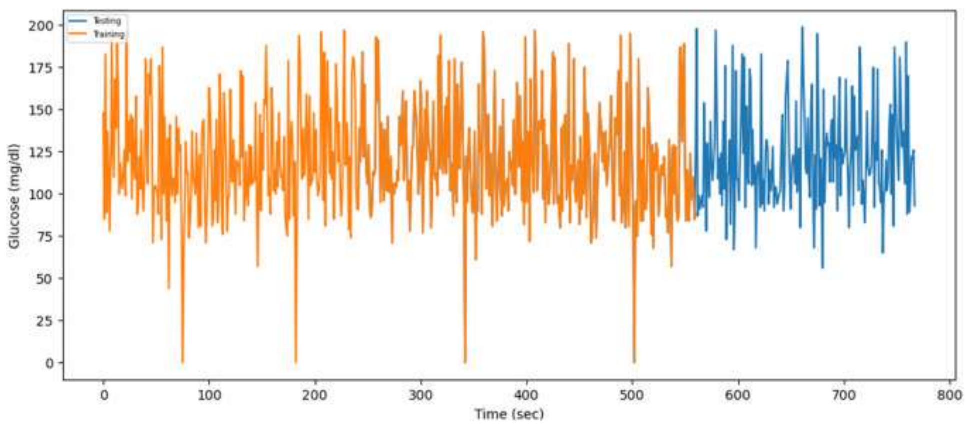

Fig. 4. Blood glucose level of training and testing

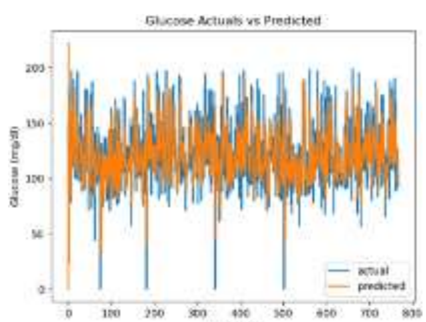

Fig. 5. Blood glucose levels of actual vs predicted

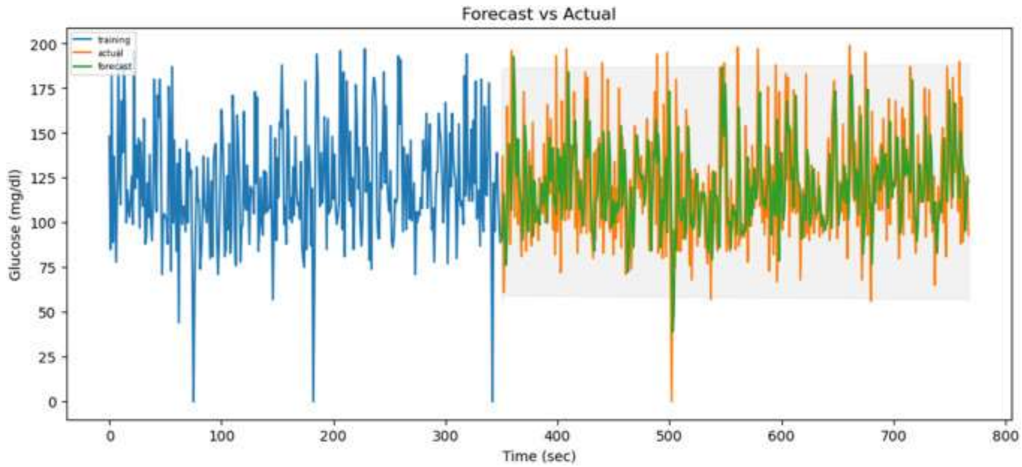

Fig. 6. Blood glucose forecasting based on time series data

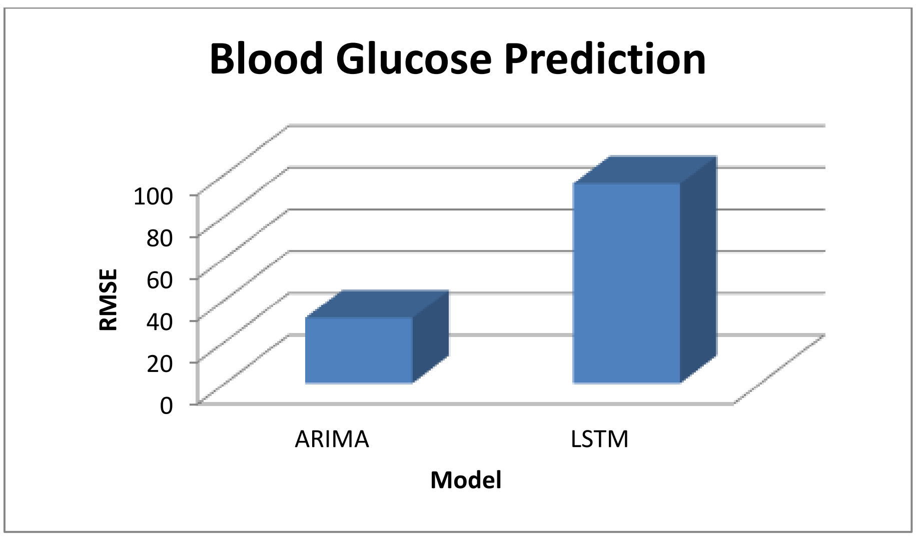

Fig. 7. RMSE Models for Blood Glucose Prediction

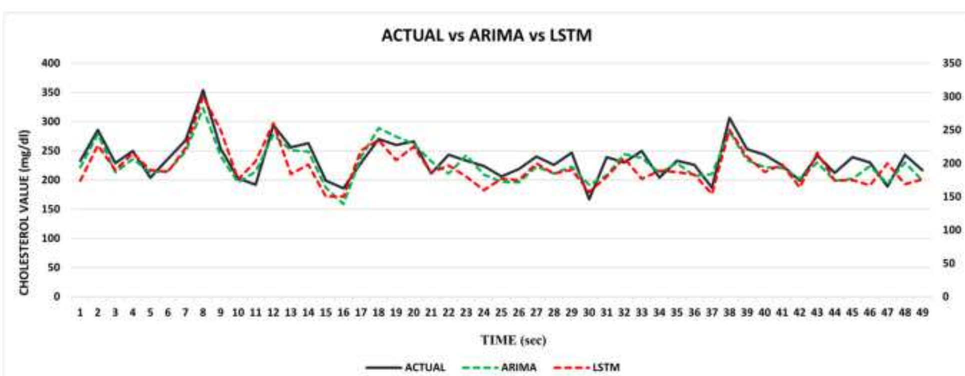

Fig. 8. Represents the actual values and predicted values with respect to prediction-based algorithms ARIMA and LSTM.

The ARIMA model is better than the LSTM model in prediction, in which the actual value is nearer to the ARIMA model predictions. ARIMA model has greater accuracy in tracking real-time data changes. Also, data forecasting is possible with the ARIMA model.

Table 1: **Blood Glucose level dataset of patients**

Patient Id	Min	Max	Mean	Standard Deviation
1	40.07	350.0	175.0	67.350
3	65.25	303.0	151.0	57.181
12	52.45	253.27	126.12	51.696
21	51.25	355.0	177.50	71.418
30	39.20	232.0	116.10	36.655

Table 2: Evaluation Metrics

Model	Predicted value	RMSE	MAPE	R^2
ARIMA	Blood Glucose	31.24	14.68	0.874
LSTM	Blood Glucose	95.43	39.38	0.910

5. CONCLUSION

This study demonstrates the successful integration of big data processing, IoT technology, and machine learning for medical data analysis using glucometers, ESP8266 microcontrollers, and SQL Server architecture. The developed system efficiently collects and processes large volumes of glucose data, allowing for real-time analysis and forecasting of blood sugar levels using advanced time series models. The machine learning models, specifically LSTM and ARIMA, were evaluated on their accuracy in predicting future glucose levels, with LSTM models showing superior performance across key metrics, including MAE, RMSE, MAPE, and R^2. LSTM's ability to capture complex patterns and dependencies in time-series data resulted in more accurate and reliable forecasts compared to ARIMA, making it a valuable tool for personalized healthcare management.

ACKNOWLEDGMENT

I would like to thank with overwhelmed gratitude, the enormous support and guidance rendered to us by Dr. S. Venkatakrishnan, Assistant Professor, Department of Computer Science, Annamalai University, TamilNadu, India.

ISSN: 2278-4632
Vol-14, Issue-9, September: 2024

REFERENCES

1. Arvind Kumar Pandey et. al (2022Big Data Predictive Analysis for Type-2 Diabetes Based Heart Disease Using Feature Extraction and Classification by Machine Learning Architectures, International Journal of Intelligent Systems And Applications in Engineering.

2. Morid, M. A., & Others. (2022). Time Series Prediction using Deep Learning Methods in Healthcare. *arXiv preprint arXiv:2108.13461*. https://doi.org/10.48550/arXiv.2108.13461.

3. Zheng, Y., Liu, Q., Chen, E., Ge, Y., & Zhao, J. L. (2019). Exploiting multi-channels deep convolutional neural networks for multivariate time series classification. *Frontiers of Computer Science, 13*(3), 399-409.

4. Zhang, J., & Wang, S. (2017). Data analysis and prediction of diabetes based on big data. *Journal of Medical Systems, 41*(11), 180.

5. Xiang, Y., Zhang, X., & Meng, Y. (2018). Big data analysis of diabetes self-management on a mobile app. *Journal of Diabetes Science and Technology, 12*(1), 150-155.

6. Yang, W., & Chang, W. (2020). Big Data-Driven Diabetes Prediction through Machine Learning. *IEEE Access, 8*, 123-133.

7. Eswaran, C., Natarajan, S., & Madhusudhanan, A. (2020). Predictive analysis of diabetes using big data and machine learning techniques. *International Journal of Advanced Computer Science and Applications, 11*(7), 234-240.

8. Singh, A., Jain, V., & Kar, A. (2021). Big data analytics for diabetes management using machine learning models. *Journal of Big Data, 8*(1), 78.

9. Aditi Chopra et. al . (2024). Predicting blood glucose level using salivary glucose and other associated factors: A machine learning model selection and evaluation study, Informatics in Medicine Unlocked,101523, 1-8

10. Amin, R., & Zeng, X. (2018). Analysis and prediction of blood glucose levels using machine learning techniques on glucometer data. *Computers in Biology and Medicine, 103*, 103-112.

11. Gao, J., Cui, Y., & Liu, X. (2017). IoT-Based Glucometer Data Analysis for Predictive Modeling of Diabetes. *Journal of Medical Internet Research, 19*(5), e186.

12. Johnson, M., & Kumar, V. (2018). SQL server integration for big data management in healthcare. *Journal of Big Data, 5*(1), 12.

13. Wong, A., Hoang, S., & Lee, A. (2021). Data Stream Processing Frameworks for Big Data Analytics in Healthcare. *IEEE Journal of Biomedical and Health Informatics, 25*(3), 756-764.

14. Xu, R., & Vasarhelyi, M. A. (2016). Big data and predictive analytics in healthcare: A time series perspective. *Journal of Healthcare Informatics Research, 4*(1), 103-123.

15. Ganjar Alfian et. al. (2019). Blood Glucose Prediction Model for Type 1 Diabetes based on Extreme Gradient Boosting. International Conference on Information Technology and Digital Applications, doi:10.1088/1757-899X/803/1/012012.