ML based Blood Glucose Level Detection using Microwave based sensors

Japleen Kaur [1a], Harshit Manik [1b], Mantra Gupta[1c], Shaurya Punj [1d], Nidhi Upadhyay[1e],Dr. Amanpreet Kaur [1f]

Department of Electronics and Communication Engineering,
Thapar Institute of Engineering and Technology,
Patiala, India

[1a]jkaur_be23@thapar.edu, [1b]hmanik_be23@thapar.edu, [1c]mgupta5_be23@thapar.edu,

[1d]spunj_be23@thapar.edu ,[1e]nupadhyay_phd22@thapar.edu, [1f]amanpreet.kaur@thapar.edu

***Abstract*—This article presents a new, painless way to monitor blood sugar levels without using needles. It uses a microwave based antenna as a sensor to detect changes in blood glucose levels. The changes in the S-parameters of the antenna are then used to train machine learning model like Logistic Regression, Random Forest, XGBoost, and CatBoost to predict the blood sugar level accurately. The study found that some models, especially CatBoost, work very well with an exceptional AUC of 0.97. Overall, this method that uses antenna as sensor for capturing S parameter data for varying blood glucose levels with Machine learning offer a safe, quick, and comfortable alternative for people with diabetes to regularly check their blood sugar levels.**

***Index Terms*—Non invasive Blood Glucose Monitoring, Logistic Regression, Random Forest, XGBoost, CatBoost**

I. INTRODUCTION

Diabetes is a disease affecting the people worldwide, currently there are estimated to be around 212 million people in India with diabetes, out of the 828 million diabetic people worldwide, which roughly makes a quarter of total people suffering from this problem to be from India [1]. Blood glucose level (BGL) spikes in people with diabetes because their bodies are unable to use insulin effectively. Insulin, a hormone generated by the pancreas regulates BGL by allowing glucose from the bloodstream to enter cells within the body. A high level of BGL has a negative impact on our kidneys, hearts and crucial organs [2]. Thus, it is essential for people suffering from diabetes to monitor the glucose level regularly. Currently, the popular way of measuring it is using the invasive method of pricking the finger and the BGL displays on monitor of the device. The measurement of BGL using needle can be uncomfortable as it causes pain, numb sensations, or can even cause shock to young people. Therefore, many noninvasive ways of measuring BGL such as optical technologies like near-infrared spectroscopy, sonophoresis, and optical coherence tomography, as well as other techniques like electromagnetic sensing, reverse iontophoresis, and analysis of biofluids like sweat or saliva are being proposed and researched upon. The major downside of these methods is that there is lack in and inconsistency in comparison with finger stick method, easily affected by environmental factors for reverse iontophoresis, the experimental error ranges are exceptionally high for sonophoresis [3]. Microwave based non-invasive techniques using the S parameter data are gaining attention amongst the researchers. In this method, S parameter data captured from the antennas is taken into account, microwaves may readily pass through the human body and detect blood inside the body without causing any damage. Blood's dielectric property changes with variations in BGL, indicating an in-proportion frequency shift because of the varying Blood glucose levels. This characteristic implies that the microwave sensor is ideally adapted for the non-invasive assessment of biological parameters, such as BGL. When the microwaves are emitted by the antenna it interacts with the blood in the tissues consisting of glucose molecule. These glucose molecules are polar in nature and thus in presence of an electric field they tend to align themselves in the direction of electric field in order to minimize the energy causing the molecules to vibrate or oscillate. Relationship between the electric field during EM wave propagation and polar glucose molecules brings out the alterations in dielectric properties of the material that leads to the modifications in microwave signals. This can be measured and glucose level in specimen can be determined accordingly. Various studies have proposed different antenna designs and sensing techniques to enhance accuracy and reliability. Xiao and Li proposed an ultra-wideband (UWB) microwave sensing technique that analyzes received signals to detect glucose concentration based on dielectric property variations in blood. Their study demonstrated that UWB sensors placed on the earlobe could effectively measure BGL with high sensitivity [4].

A study by Titli and Kumar explored microstrip patch antennas for noninvasive glucose monitoring. They designed a 7 GHz resonant sensor and tested different human tissue models to observe the impact of dielectric constant variations on frequency response [5].

Similarly, another work introduced a narrowband microwave sensor operating at 1.3 GHz. This design measured frequency shifts caused by changes in blood permittivity and achieved a regression model with a 0.75 coefficient of determination (R^2), suggesting a strong correlation between BGL and frequency variation[6].

Sharaf et al. proposed a triple-band monopole antenna that utilized three different frequency bands (2.9 GHz, 4.3 GHz, and 6.5 GHz) for improved glucose sensing. Their reflection-based microwave sensing method demonstrated an impressive sensitivity of 19.43 MHz/mg/dL, which is one of the highest reported in literature [7].

Further advancements in sensor design include complementary split-ring resonator (CSRR) structures, as demonstrated by Omer et al. Their hexagonal CSRR sensor used strong field localization to detect glucose concentration variations efficiently. The Principal Component Analysis (PCA) algorithm was applied to improve accuracy, making this approach promising for practical applications[8].

Wearable solutions have also been investigated, such as the ultra-miniaturized UWB patch antenna developed for on- body health monitoring. This antenna, designed with characteristic mode theory (CMT), achieved a broad bandwidth (3.15 GHz – 10.55 GHz) and maintained robust performance under different human body conditions[9].

Recent advancements in non- invasive blood glucose monitoring have focused on analyzing S-parameters across different frequencies to predict glucose levels. These studies explore the relationship between glucose concentration and the dielectric properties of biological tissues, which can be detected using microwave sensing techniques. innovative approach involves a wearable electromagnetic multi-sensing system designed to replicate vascular anatomy for continuous glucose monitoring. This system measures S- parameter magnitudes and phases within a specific frequency range. The collected data undergo preprocessing steps such as averaging and normalization to extract relevant features. Various machine learning models, including Gaussian Processes, were tested to predict glucose levels, demonstrating high accuracy, particularly for small datasets.

Numerous studies have utilized machine learning (ML) algorithms to predict solar radiation across various global regions, employing models such as XGBoost, Random Forest, and ensemble AI approaches. These models have demonstrated superior performance compared to traditional empirical models, especially when optimized with relevant features. Researchers have emphasized the importance of integrating meteorological and geographical data to improve prediction accuracy. Despite advancements, challenges remain in interpretability, computational efficiency, and scalability of ML models. Hence, recent work focuses on explainable AI techniques like SHAP to enhance model transparency and trust.[10]

Accurate estimation of Global Horizontal Irradiance (GHI) is vital for optimizing solar energy systems. Gupta et al. (2024) employed machine learning models—including Decision Trees, Random Forest, Extreme Gradient Boosting, and Extra Trees—to forecast GHI, utilizing Variance Inflation Factor for feature selection. The Extra Trees model demonstrated superior performance, achieving an MAE of 3.01, RMSE of 1.748, and an R^2 of 0.99. To enhance model interpretability, SHAP (Shapley Additive Explanations) was used, elucidating the impact of each feature on the predictions. This approach underscores the efficacy of combining robust ML algorithms with explainable AI techniques for reliable solar irradiance forecasting[11].

Another technique applied complex-valued neural networks (CVNNs) for glucose sensing in the millimeter-wave frequency band. Researchers analyzed S21 parameters of glucose solutions with varying concentrations and developed a predictive model capable of estimating glucose levels with minimal error. This method highlights the effectiveness of CVNNs in processing intricate S-parameter data for accurate glucose monitoring [12] Additionally, researchers have explored microwave transmission combined with machine learning techniques for non-invasive glucose detection. By modeling the dielectric properties of human tissues, they employed optimization techniques to identify glucose-dependent coefficients. Through the analysis of S-parameters in specific frequency ranges, a linear correlation was established between the measured parameters and glucose levels, demonstrating the potential for real-time, non-invasive monitoring [13].

Overall, these advancements indicate that integrating S-parameter measurements across targeted frequency ranges with machine learning models can significantly improve the accuracy and reliability of non-invasive blood glucose monitoring systems. Existing technologies integrating backend with technologies for IOT based glucose monitoring uses- NO SQL(MongoDB): It stores glucose data efficiently - MQTT Protocol: Latent free data transfer, Edge computing: Raspberry. In order to improve the accuracy of the BGL prediction using antenna-based sensors, this research paper proposes an IoT-based monitoring system using Raspberry Pi implemented for real- time, non-invasive glucose detection. The article presents various ML models for predicting BGL trained using the S parameter datasets while demonstrating model efficacy. Additionally, the results highlight the potential of hybrid deep learning models and microwave antennas for accurate, scalable, and secure diabetic patient monitoring

II. METHODOLOGY

The proposed work is carried out using a flexible antenna designed on a Rogers R5880 substrate with details mentioned in [14]. This antenna is used as a sensor to capture S parameter data from a 3-layer phantom designed in CST MWS V'23 with varying blood sugar levels. This data is then used with the Debye's model that correlates the BGL in human tissues to dielectric properties of the blood[4]. As stated in equation number 1, the Debye model is used to describe the relationship between BGL and the blood plasma's dielectric characteristics [14].

The recorded data is then used to train the ML models namely Logistic Regression, XG Boost, Random Forest and Cat Boost. After training, these ML models predict accurate blood glucose levels in human. Figure 1 shows the methodology followed for carrying out the proposed research work.

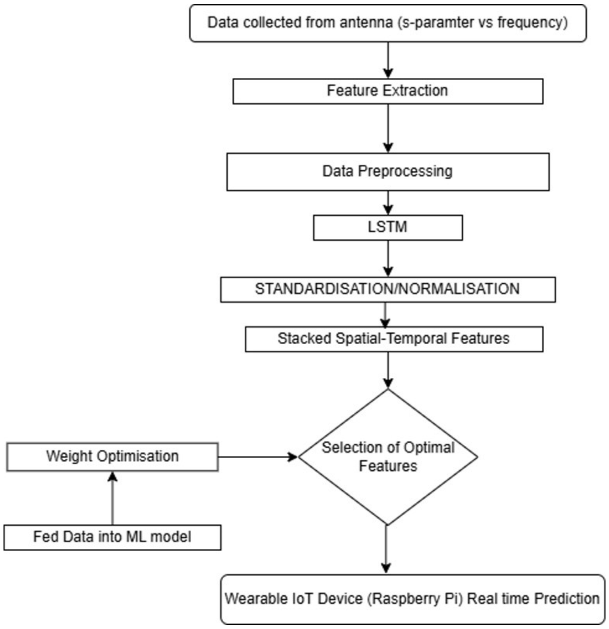

Fig. 1. Flowchart of the proposed methodology. .

A. Dataset Selection and preprocessing

The data set used for BGL analysis is of two types: one that is collected using the antenna and second the publicly available datasets for model training and validation. The dataset includes physiological and biochemical attributes essential for BGL prediction, such as: Patient demographics (age, BMI), Clinical indicators (HbA1c levels, blood glucose levels, insulin dosage, and carbohydrate intake), Sensor data and lifestyle factors (smoking history, physical activity) Preprocessing steps included:

- Missing values in numerical features were handled using K-Nearest Neighbors (KNN) imputation.
- Categorical variables were encoded using Label Encoding.
- Numerical attributes were standardized using Min- Max scaling.
- Feature engineering introduced derived features (BMI category, age group, BMI-age interaction, glucose level classification).
- Data was split into an 80:20 train-test ratio for evaluation.

B. ML Models used for prediction

(i) Logistic Regression

Logistic regression is particularly advantageous for analyzing observational data where adjustments are necessary to mitigate potential bias arising from differences between the groups under comparison. The application of standard linear regression for a binary outcome may yield suboptimal results. The validity of linear regression is based on the homogeneity of outcome variability across all predictor values. Linear regression is insufficient for two-level type of data. Logistic regression addresses this deficiency [15]. Training Logistic Regression on the generated dataset yielded an AUC (Area Under Curve) of 0.93 as shown in fig. 2.

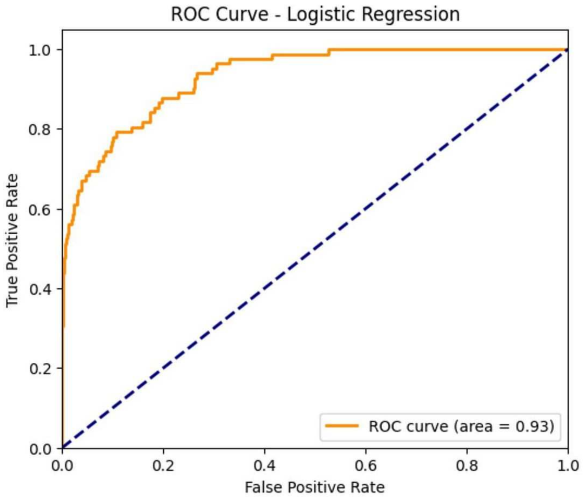

Fig. 2. AUC of Logistic Regression

(ii) Random Forest

Random Forest regression is an ensemble learning method that constructs multiple decision trees and merges their predictions to produce a more accurate and stable estimate. It minimizes the overfitting problem by averaging and selecting the random features, which decorrelates the trees. This ensemble method often yields superior predictive performance and is robust to outliers and noise in the data [16]. Training Random-forest on the generated dataset yielded an AUC (Area Under Curve) of 0.95 as shown in fig. 3.

(iii) eXtreme Gradient Boost

XGBoost is an efficient and scalable implementation of the gradient boosting. The package comprises an efficient linear model solver and a tree learning algorithm. It supports regression, classification, and ranking. The program is designed to be extendable, enabling users to effortlessly set their own aims. It possesses numerous features such as speed, can accept various forms of input data, can accommodate sparse input for both the tree booster and linear booster, and shows superior performance across many datasets [17]. Training XG boost on the generated dataset yielded AUC as 0.96 as shown in fig. 4. It shows a superior classification ability, outperforming both Logistic Regression (AUC = 0.93) and Random Forest (AUC = 0.95).

(iv) CatBoost

CatBoost belongs to the category of Gradient Boosting Decision Tree (GBDT) ensemble machine learning algorithms. CatBoost is a supervised learning system that introduces two innovations: Ordered Target Statistics and Ordered Boosting. CatBoost's initial enhancement to Gradient Boosting is its approach to high cardinality categorical variables. The current iteration of CatBoost defaults to a value of 255 under particular

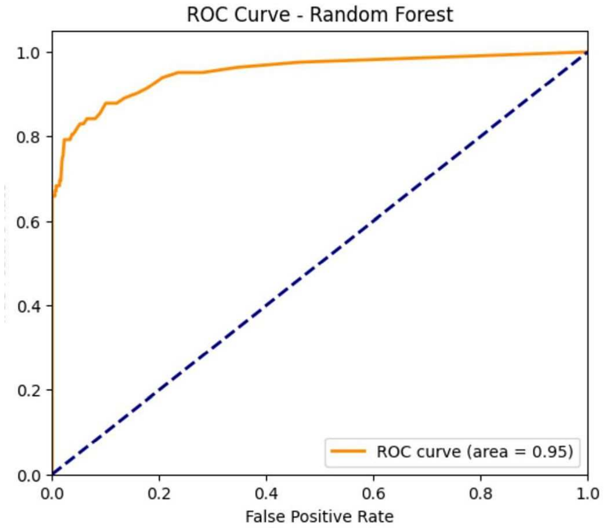

Fig. 3. AUC of Random Forest

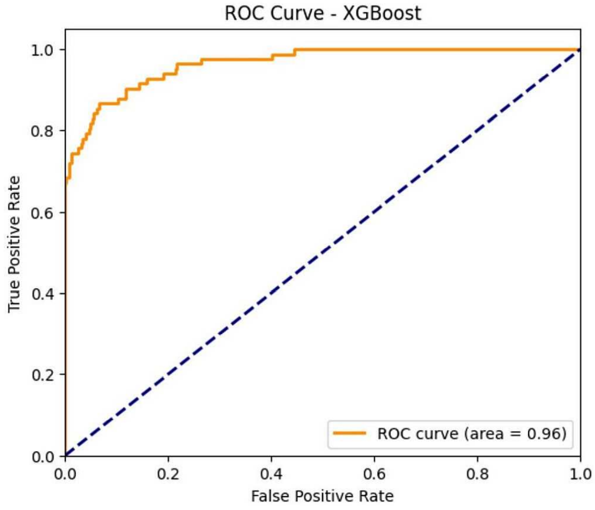

Fig. 4. AUC of XG boost

conditions when executed on GPUs, and 2 when executed on CPUs. This shows CatBoost's sensitivity to hyperparameters [18]. Training Catboost on the generated dataset yielded AUC as 0.97 as shown in fig.5. It shows the highest classification accuracy among all tested models, surpassing Logistic regression, Random Forest and XG Boost.

C. Performance Evaluation of Machine Learning Models

The Receiver Operating Characteristic (ROC) curve effectively visualizes a classifier's performance to determine an appropriate operating point or decision threshold. When the decision threshold is adjusted, several points on the ROC curve are represented as [$P(\mathrm{Fp}) = \alpha, P(\mathrm{Tp}) = 1 - \beta$]. The most straightforward method to compute the area beneath the ROC curve is to utilize trapezoidal integration [19] .

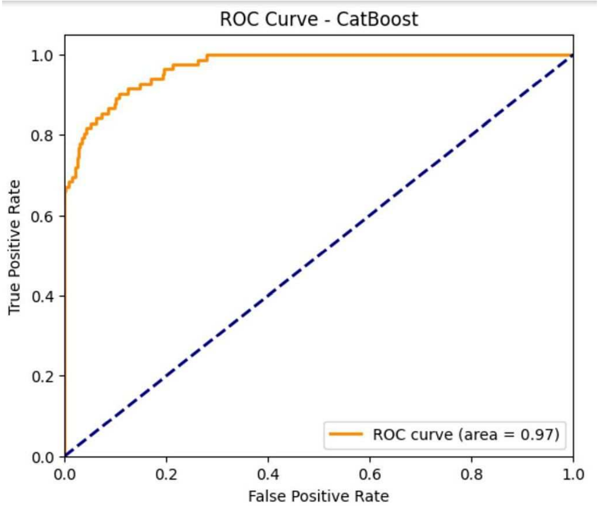

Fig. 5. AUC of Cat boost

$$\mathrm{AUC} = \sum_i (1 - \beta_i \Delta\alpha) + \frac{1}{2}[\Delta(1 - \beta)\Delta\alpha] \qquad (1)$$

where,

$$\Delta(1 - \beta) = (1 - \beta_i) - (1 - \beta_{i-1})$$
$$\Delta\alpha = \alpha_i - \alpha_{i-1}$$

In this work, AUC (ROC) is used as the performance evaluation matrix of different ML models. The results obtained are discussed in table I. CatBoost Algorithm showed the best performance in terms of AUC as 0.97. Fig. 6 shows the comparison plot of the different ML models used in this work. Other performance evaluation such as accuracy, F1, recall and precision score of the models are shown in table II.

TABLE I
AUC OF DIFFERENT MODELS FOR DIABETES PREDICTION.

Model	AUC
Logistic Regression	0.93
Random Forest	0.95
XG Boost	0.96
CatBoost	0.97

TABLE II
PERFORMANCE COMPARISON OF DIFFERENT MODELS FOR DIABETES PREDICTION.

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	95.0%	79.6%	52.4%	63.2%
Random Forest	97.2%	100%	65.8%	79.4%
XGBoost	96.8%	85.7%	73.1%	78.9%
CatBoost	97.1%	94.9%	68.2%	79.4%
TabNet	41.5%	12.2%	100%	21.9%

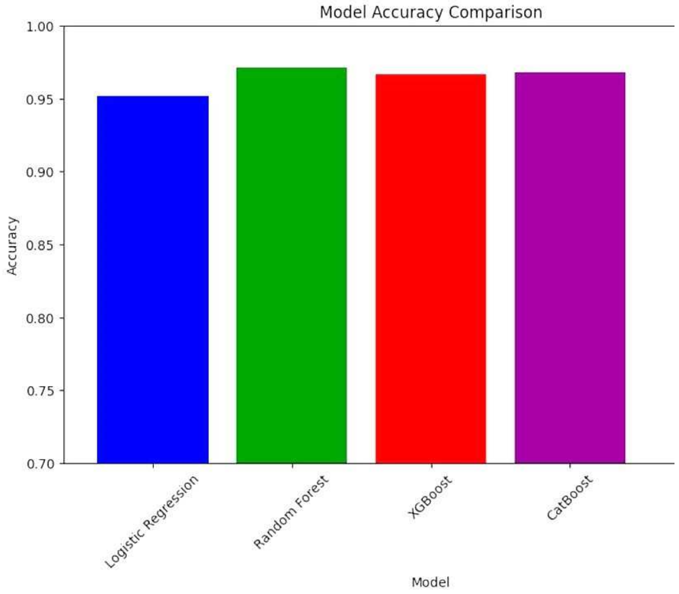

Fig. 6. Model accuracy comparison

III. RESULTS AND DISCUSSIONS

A. Analysis of Evaluation Metrics

The choice of evaluation metrics is crucial in assessing the performance of models for blood glucose level prediction.

- AUC provides an aggregate measure of the model's ability to distinguish between positive and negative classes (e.g., high vs. normal blood glucose) across all possible classification thresholds. A higher AUC indicates better discriminatory power. The study shows that CatBoost achieved the highest AUC of 0.97, suggesting its superior overall ability to differentiate between different blood glucose levels based on the S-parameter data.
- Accuracy measures the overall proportion of correct predictions. Random Forest showed the highest accuracy at 97.2%, indicating that it correctly predicted blood glucose levels most frequently in this dataset.
- Precision focuses on the accuracy of positive predictions (e.g., when the model predicts high blood glucose, how often is it correct). High precision is important to minimize false positives, which could lead to unnecessary interventions.
- Recall (also known as sensitivity) measures the model's ability to find all the positive instances (e.g., correctly identifying all cases of high blood glucose). High recall is critical in medical applications like blood glucose monitoring to avoid missing potentially dangerous high glucose levels (false negatives).
- F1-Score is the harmonic mean of precision and recall, providing a balanced measure of the model's performance, especially useful when there is an uneven class distribution.

B. Comparative Model Performances

While Random Forest achieved the highest accuracy, the discussion highlights that CatBoost showed the best performance in terms of AUC. This suggests that although Random Forest makes slightly more correct predictions overall, CatBoost is better at discriminating between the classes across various thresholds. The results in Table II provide a more nuanced view:

- Logistic Regression, a simpler model, showed an AUC of 0.93. Random Forest, an ensemble method, improved upon Logistic Regression with an AUC of 0.95 and the highest accuracy (97.2%). Its high accuracy and a strong F1-Score of 79.4% indicate a good balance of precision and recall.
- XGBoost, another gradient boosting model, achieved an AUC of 0.96. The discussion notes that XGBoost balanced precision (85.7%) and recall (73.1%), offering reliable predictions with a good trade-off between false positives and false negatives.
- CatBoost, which had the highest AUC (0.97), also demonstrated very high precision (94.9%). However, its recall (68.2%) was lower than XGBoost, suggesting it might have more false negatives than XGBoost, although when it predicts a positive case, it is very likely to be correct. The paper attributes CatBoost's strong performance to its handling of categorical features.
- TabNet, while showing poor accuracy (41.5%) and precision (12.2%), achieved a perfect recall of 100%. The discussion correctly points out that this makes TabNet potentially useful in scenarios where missing a diabetic case (false negative) is critical, even at the cost of a high number of false positives.

C. Analysis of ROC Curves

Figures 2-5 visually support the AUC values presented in Table I. The closer the ROC curve is to the top-left corner, the higher the AUC and the better the model's performance. The ROC curve for CatBoost (Figure 5) is the closest to the top-left, consistent with its highest AUC. The curves for XGBoost (Figure 4) and Random Forest (Figure 3) are progressively further from the top-left, and Logistic Regression (Figure 2) is the furthest, aligning with their respective AUC scores. These curves illustrate the trade-off between the true positive rate (sensitivity) and the false positive rate (1-specificity) at different classification thresholds for each model.

D. Future Work and Practical Implications

The results highlight the potential of using machine learning with microwave-based sensors for non-invasive blood glucose monitoring. CatBoost and Random Forest show promising overall performance. However, the choice of the best model for a real-world application would depend on the specific requirements and the relative costs of false positives and false negatives. For instance, in a screening tool, high recall (like that of TabNet) might be preferred to ensure no cases are missed, even if it means more false alarms. In a monitoring device, a balance of precision and recall (like XGBoost or Random Forest) might be more desirable.

Future work could involve investigating the reasons behind the performance differences in more detail, perhaps through error analysis. Exploring hyperparameter tuning for each model could potentially further enhance their performance. Additionally, testing the models on larger and more diverse datasets, including data from real human subjects, would be crucial to validate these findings and assess the generalizability of the models. Further analysis of the S-parameter data and feature engineering could also potentially lead to improved results.

IV. CONCLUSION

This research paper presents a non-invasive BGL monitoring in humans using antennas as sensors and ML models for decision making process of the data captured. ML models like Random Forest, XG boost, Catboost and TabNet were trained based upon the data captured by the antenna based sensors . While Random-forest achieved the highest accuracy (97.2 percent), making it the most effective traditional ML model, XGBoost balanced precision and recall, ensuring reliable predictions with fewer false positives and negatives. CatBoost outperformed XGBoost in categorical feature handling, demonstrating robustness in structured data. TabNet showed poor accuracy but achieved 100 percent recall, making it useful for risk-averse predictions were missing a diabetic case is critical. Table 1 compares the comprehensive results of different models used for BGL prediction based upon training those models using the S parameter data.

REFERENCES

[1] P. Saeedi et al., “Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition,” Diabetes Res. Clin. Pract., vol. 157, p. 107843, 2019, doi: 10.1016/j.diabres.2019.107843.

[2] A. Losoya-Leal, S. Camacho-Leo´n, G. Dieck-Assad, and S. O. Mart´ınez-Chapa, “State of the art and new perspectives in non-invasive glucose sensors,” Rev. Mex. Ing. Biome´d., vol. 33, no. 1, pp. 41–52, 2012.

[3] A. Kandwal, Y. D. Sharma, R. Jasrotia, C. C. Kit, N. Lakshmaiya, M. Sillanpa¨a¨, L. W. Y. Liu, T. Igbe, A. Kumari, R. Sharma, S. Kumar, and C. Sungoum, “A comprehensive review on electromagnetic wave based non-invasive glucose monitoring in microwave frequencies,” Heliyon, vol. 10, p. e37825, 2024, doi: 10.1016/j.heliyon.2024.e37825.

[4] Xiao and Q. Li, “A noninvasive measurement of blood glucose concentration by UWB microwave spectrum,” IEEE Antennas and Wireless Propagation Letters, vol. 16, pp. 1040–1043, 2017, doi: 10.1109/LAWP.2016.2618946.

[5] S. R. Titli and L. Kumar, “Blood Glucose Monitoring Based on Transmission using Microstrip Patch Antenna,” in Proc. 2024 3rd Int. Conf. for Innovation in Technology (INOCON), Karnataka, India, Mar. 1–3, 2024, pp. 1–6, doi: 10.1109/INOCON60754.2024.10512087.

[6] V. V. Deshmukh and S. S. Chorage, “Non-invasive determination of blood glucose level using narrowband microwave sensor,” J. Ambient Intell. Humaniz. Comput., vol. 12, 2021, doi: 10.1007/s12652-021-03105-z.

[7] F. Sharaf, D. N. Elsheakh, and A. R. Eldamak, “A noninvasive method of monitoring blood glucose levels by using triple-band monopole antenna,” Microw. Opt. Technol. Lett., vol. 67, no. 1, 2025, Art. no. e70065, doi: 10.1002/mop.70065.

[8] A. E. Omer et al., “Low-cost portable microwave sensor for non-invasive monitoring of blood glucose level: novel design utilizing a four-cell CSRR hexagonal configuration,” Scientific Reports, vol. 10, no. 1, p. 15200, 2020, doi: 10.1038/s41598-020-72114-3.

[9] S. Modak et al., “Design and performance measurement of worn-on-body instrumental ultra-miniaturized UWB wearable patch for e-health monitoring,” IEEE Access, early access, 2024, doi: 10.1109/ACCESS.2024.3365938.

[10] R. Gupta, A. K. Yadav, and S. K. Jha, “Prediction of Global Horizontal Irradiance Using an Explainable Data Driven Machine Learning Algorithms,” Electric Power Components and Systems, vol. 0, no. 0, pp. 1–18, Feb. 2024, doi: 10.1080/15325008.2024.2310771.

[11] R. Gupta, A. K. Yadav, and S. K. Jha, “Prediction of Global Horizontal Irradiance Using an Explainable Data Driven Machine Learning Algorithms,” Electric Power Components and Systems, vol. 0, no. 0, pp. 1–18, Feb. 2024, doi: 10.1080/15325008.2024.2310771.

[12] J. Blampey and K. Rahim, ”Hybrid Attention-based Diffusion Network for Blood Glucose Prediction,” arXiv preprint, 2021, arXiv:2111.07455.

[13] Y. Sun et al., ”Sequential LSTM Model for Blood Glucose Level Prediction,” arXiv preprint, 2018, arXiv:1809.03817.

[14] Upadhyay, Nidhi, Amanpreet Kaur, and Arnab Pattanayak. ”Flexible Antenna with AMC Backing for Non-Invasive Blood Glucose Monitoring.”2024 Second International Conference on Microwave, Antenna and Communication (MAC). IEEE, 2024.

[15] LaValley, Michael P. ”Logistic regression.”Circulation117.18 (2008): 2395-2399. vetnik V, Liaw A, Tong C, Culberson JC, Sheridan RP, Feuston BP.

[16] Random forest: a classification and regression tool for compound classification and QSAR modeling. J Chem Inf Comput Sci. 2003;43(6):1947–58.

[17] Chen, Tianqi, et al. ”Xgboost: extreme gradient boosting.”R package version 0.4-21.4 (2015): 1-4.

[18] Y. Sun et al., ”Sequential LSTM Model for Blood Glucose Level Prediction,” arXiv preprint, 2018, arXiv:1809.03817.

[19] Bradley, Andrew P. ”The use of the area under the ROC curve in the evaluation of machine learning algorithms.”Pattern recognition30.7 (1997): 1145-1159.