AMERICAN JOURNAL OF MANAGEMENT AND IOT MEDICAL COMPUTING
Peer Reviewed, Referred & Indexed Journal
E-ISSN: 3069-0110
Vol.5, No.2(2) (2026)
www.ajmimc.com

Synergistic Modeling of PPG Morphophysiological Patterns and Clinical Parameters for Fine-Grained Glycemic State Inference

C. Jyothi Sree[1], Chitte Geethanjali[1], K. Ramana[1], Pasham Nandini[1], Kavali Niranjan[1]

[1]Department of Computer Science and Engineering, [1]Sree Dattha Institute of Engineering and Science, Nagarjuna Sagar Road, Sheriguda, Ibrahimpatnam, Rangareddy Dist, 501510, Telangana, India.

ABSTRACT

The increasing prevalence of diabetes and related metabolic disorders has created a strong demand for intelligent systems capable of accurately predicting blood glucose levels using non-invasive physiological signals. Continuous monitoring and timely prediction are essential to prevent complications and support effective clinical decision-making. However, many earlier approaches rely on basic statistical analysis and manual interpretation, which struggle to handle complex relationships among multiple health parameters such as Photoplethysmography (PPG) signal characteristics, heart rate, and patient demographics. These methods often lack scalability, adaptability, and predictive accuracy when applied to real-world healthcare datasets. Additionally, the absence of automated systems limits real-time analysis and delays preventive interventions. To address these challenges, this project proposes a machine learning-based predictive framework that leverages multiple regression models, including Linear Regression (LR), Ridge Regression, Least Absolute Shrinkage and Selection Operator (Lasso) Regression, Decision Tree Regressor (DTR), and an ensemble-based approach using Natural Gradient Boosting (NGB). The system processes physiological signals such as PPG along with attributes like age, gender, height, and weight to estimate glucose levels. Data preprocessing techniques such as missing value handling, feature scaling, and train-test splitting are applied to ensure robust model performance. The framework is deployed through a Flask based web application that enables model training, performance evaluation using metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R^2) score, and real-time prediction for both single and batch inputs. The proposed system enhances prediction accuracy, reduces manual dependency, and provides a scalable solution for healthcare monitoring, enabling early detection and improved patient outcomes.

Key words: glucose level prediction, PPG signal, non-invasive monitoring, machine learning, regression models, healthcare analytics, Flask Framework.

1. INTRODUCTION

Diabetes is a progressive metabolic disorder characterized by impaired insulin secretion or reduced cellular responsiveness to insulin, resulting in abnormal fluctuations in blood glucose levels. Sustained hyperglycemia over time significantly increases the risk of complications such as cardiovascular disease, nephropathy, neuropathy, and retinopathy, making continuous monitoring essential for effective disease management. According to the International Diabetes Federation (IDF), an estimated 463 million individuals aged between 20 and 79 were living with diabetes in 2019, and this number is projected to escalate to nearly 700 million by 2045, highlighting the growing global burden of the disease [1]. Moreover, diabetes-related complications accounted for over 4.2 million deaths worldwide in the same year. Despite advancements in healthcare, the most commonly adopted glucose monitoring techniques remain invasive, primarily involving finger-prick blood sampling using glucometers, test strips, and lancets, along with auxiliary materials such as alcohol swabs and dressings to reduce infection risk. While these methods provide clinically accurate readings, they are associated with discomfort, inconvenience, and reduced patient compliance, particularly for individuals requiring

frequent monitoring, thereby emphasizing the need for non-invasive and continuous monitoring solutions.

In response to these limitations, extensive research has been conducted to develop alternative sensing technologies for glucose estimation. Material-based sensing approaches, including carbon nanotube-integrated sensors, fluorescence-based detection, and plasmonic sensing techniques, have been explored for their high sensitivity; however, they often involve complex calibration procedures, susceptibility to environmental interference such as autofluorescence, and challenges related to scalability and portability [2]. Electrical sensing methods, such as electromagnetic wave-based detection, millimeter-wave radar systems, and bioimpedance analysis, aim to estimate glucose variations by analyzing changes in tissue electrical properties. Although promising, these techniques are often affected by external conditions like temperature, hydration levels, and skin properties, leading to variability in measurements and reduced reliability [3]. Optical sensing techniques, including Optical Coherence Tomography (OCT), near-infrared spectroscopy, and mid-infrared spectroscopy, have demonstrated the capability to detect glucose concentration changes under controlled experimental setups; however, their practical deployment is limited due to high equipment costs, longer response times—sometimes extending up to several minutes—and sensitivity to ambient conditions. Other innovative approaches involve the use of Galvanic Skin Response (GSR) sensors and analysis of alternative biological fluids such as saliva, sweat, and tears to estimate glucose levels. Although these methods reduce the need for blood sampling, they still encounter challenges related to signal consistency, calibration complexity, and variability across individuals [4].

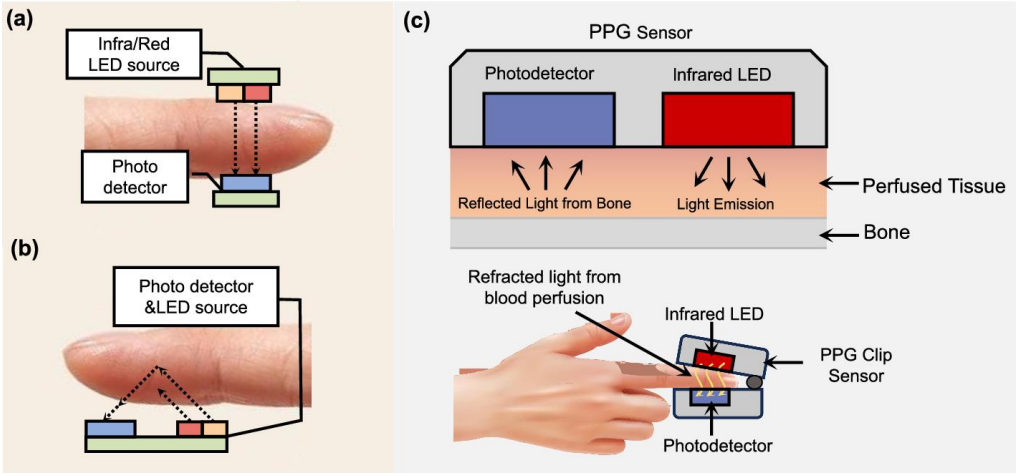


Fig.1: Predicting glucose levels using PPG sensor

Among the emerging techniques, Photoplethysmography (PPG) has gained significant attention as a non-invasive, cost-effective, and easily deployable solution. PPG measures volumetric changes in blood circulation within microvascular tissues using optical sensors, and it is already widely integrated into wearable devices for monitoring physiological parameters such as heart rate and oxygen saturation. The waveform characteristics of PPG signals, including amplitude, peak variations, and pulse morphology, have been found to correlate with underlying physiological changes, including blood glucose fluctuations. Recent studies suggest that by combining PPG signal analysis with machine learning algorithms, it is possible to model complex relationships between physiological features and glucose levels, enabling accurate prediction without the need for invasive procedures [5]. This approach not only improves patient comfort and compliance but also supports real-time, continuous monitoring, making it highly suitable for next-generation healthcare systems focused on preventive care and early intervention.

2. LITERATURE SURVEY

Owess et al. [6] developed supervised machine learning frameworks to identify and predict elevated blood glucose levels using population health data. Their study utilized risk-related attributes such as age, body mass index, dietary behavior, physical activity, comorbid conditions, and fasting glucose values obtained from the STEPwise approach dataset. Multiple classification techniques, including Random Forest, Decision Tree, AdaBoost, XGBoost, Bagging-based Decision Trees, and Multi-Layer Perceptron, were trained and evaluated. Among these, the Random Forest model demonstrated the highest predictive capability, achieving an accuracy of 98.4%, followed closely by other ensemble-based approaches, indicating the effectiveness of combining clinical risk factors with machine learning for glucose prediction.

Gudiño-Ochoa et al. [7] introduced an innovative framework that integrates Tiny Machine Learning (TinyML) with an electronic nose (e-nose) system to enable diabetes detection and blood glucose prediction through breath analysis. To overcome data scarcity, the study employed Conditional Tabular Generative Adversarial Networks (CTGAN) to generate synthetic data from a limited real-world dataset comprising both diabetic and non-diabetic individuals. The expanded dataset was then used to train machine learning models, where Light Gradient Boosting Machine achieved 86% accuracy in glucose prediction, while Random Forest attained 94.14% accuracy in diabetes classification, demonstrating the potential of combining synthetic data generation with embedded systems for real-time healthcare applications.

Ali et al. [8] conducted a comparative study of machine learning techniques for predicting interstitial glucose levels using data collected from wearable devices. The dataset included signals from a wrist-worn smart watch, continuous glucose monitoring readings, and dietary intake logs. Various machine learning models such as Decision Tree, Support Vector Machine, Random Forest, Linear Discriminant Analysis, K-Nearest Neighbors, Gaussian Naïve Bayes, Lasso, Ridge, Elastic Net, and XGBoost were evaluated. The study analyzed performance using multiple metrics, including accuracy, precision, recall, F1-score, and ROC curves, with results indicating varying effectiveness across models depending on data characteristics and feature representation.

Gunathilaka et al. [9] proposed a non-invasive diagnostic methodology for identifying diabetic conditions using pulse wave analysis combined with deep learning. The approach involved collecting pulse waveform data from both healthy and diabetic individuals, followed by preprocessing and feature extraction. The processed data were used to train deep learning architectures such as Convolutional Neural Networks, Visual Geometry Group networks, and Residual Networks. Among these, deeper models demonstrated superior performance, with VGG-based architectures achieving higher classification accuracy, highlighting the importance of deep feature extraction in physiological signal analysis.

Leung et al. [10] presented a comprehensive study on non-invasive glucose sensing technologies, emphasizing the importance of developing reliable alternatives to traditional invasive monitoring methods. The work reviewed various sensing modalities and highlighted the challenges associated with achieving accurate, real-time glucose estimation without direct blood sampling. Despite advancements in minimally invasive continuous glucose monitoring systems, the study concluded that achieving consistent, non-invasive measurement in real-world environments remains a significant challenge due to physiological variability and external influences.

Chauhan et al. [11] focused on predicting the progression of diabetes mellitus using supervised machine learning techniques. The study analyzed patient-related data to model disease progression patterns and employed multiple machine learning algorithms to improve predictive performance. The results demonstrated that supervised learning approaches can effectively capture complex relationships within clinical datasets, enabling early detection and better management of disease progression, thereby supporting proactive healthcare interventions.

Chowdhury et al. [13] proposed a machine learning-based framework for estimating blood pressure using PPG signals combined with demographic information. The system involved extracting time-domain, frequency-domain, and time-frequency features from PPG waveforms and their derivatives. Feature selection techniques were applied to reduce redundancy and improve model efficiency. Several regression models were trained and evaluated, with Gaussian Process Regression delivering the most accurate results. Although the study focused on blood pressure estimation, it highlighted the strong potential of PPG signals for modeling physiological parameters relevant to glucose prediction.

Zhang et al. [14] developed a smartphone-based system for non-invasive blood glucose monitoring using PPG signal processing and machine learning techniques. The approach included capturing PPG signals through mobile camera videos, applying signal correction methods to remove baseline drift, and segmenting the waveform into individual cycles. Feature extraction was performed using Gaussian-based representations, and machine learning models were applied to classify glucose levels into predefined categories. The system achieved high accuracy in distinguishing valid and invalid signals, demonstrating the feasibility of smartphone-based health monitoring solutions.

Chu et al. [15] presented a deep learning-driven method for predicting blood glucose levels using PPG signals. The study introduced a Deep Sparse Capsule Network designed to capture complex patterns within physiological signals. Data preprocessing involved noise reduction using filtering techniques to preserve waveform characteristics. The proposed model was further enhanced with attention mechanisms to improve prediction performance. Experimental results indicated that deep learning architectures can significantly improve the accuracy and robustness of non-invasive glucose estimation.

3. PROPOSED SYSTEM

The system architecture is designed as a comprehensive machine learning framework integrated with a web-based application to predict glucose levels using physiological and demographic data. It begins with data acquisition from a structured dataset containing features such as PPG signal, heart rate, systolic peak, diastolic peak, pulse area, gender, age, height, and weight, with glucose level serving as the target variable for prediction. The data preprocessing stage ensures data quality by handling missing values through mean imputation, followed by splitting the dataset into training and testing subsets in an 80–20 ratio to enable unbiased model evaluation. Feature scaling is applied using StandardScaler specifically for linear models to normalize feature distributions and improve convergence, while tree-based models are trained on unscaled data to retain inherent feature relationships and avoid distortion. The system incorporates multiple regression models including LR, RR, Lasso, DTR, and NGB, allowing the capture of both linear dependencies and complex nonlinear interactions within the data. Each model is trained independently using the prepared dataset, and predictions are generated on unseen test data to evaluate generalization capability.

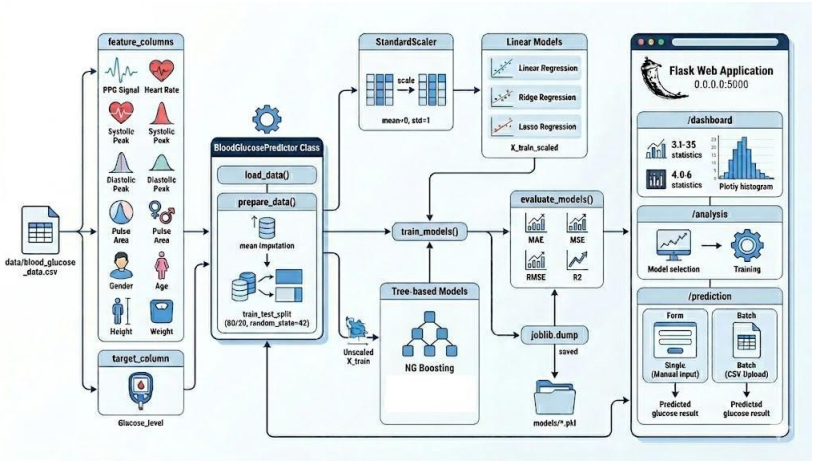


Fig. 2: Blood Glucose Prediction System Architecture.

The performance of each model is assessed using metrics such as MAE, MSE, RMSE, and R^2 score, providing a comprehensive understanding of prediction errors and model accuracy. Based on these evaluations, trained models are serialized and stored, enabling efficient reuse without repeated training, and associated scalers are also saved to maintain consistency during inference. As illustrated in Fig. 2, the architecture ensures a smooth data flow from preprocessing to prediction, integrating all components within a Flask-based web application that provides an interactive dashboard for dataset visualization, statistical insights, and model comparison. The application supports both single prediction through manual input and batch prediction via file upload, where incoming data is processed, transformed as required, and passed to the selected model to generate outputs. Additionally, the system dynamically loads saved models during runtime, reducing computational overhead and improving response time. This architecture is modular, allowing easy extension with additional models or features, and scalable for handling larger datasets or real-time inputs, ultimately delivering an efficient, accurate, and user-friendly solution for glucose level prediction.

3.1 NGB Regressor

NGB builds probabilistic models using boosting. Steps: (1) input features, (2) train sequential base learners to reduce error, (3) combine predictions from all learners using weighted sum, and (4) output predicted glucose value with uncertainty. NGB is ideal for modeling uncertainty in predictions, which is crucial in biomedical applications. It provides not only accurate predictions but also confidence intervals as shown in fig 3.

Internal operations:

- Load input features.
- Train sequential base learners on residuals.
- Combine predictions from all learners using weighted sum.
- Generate predicted glucose level and uncertainty.
- Evaluate model performance on test data.

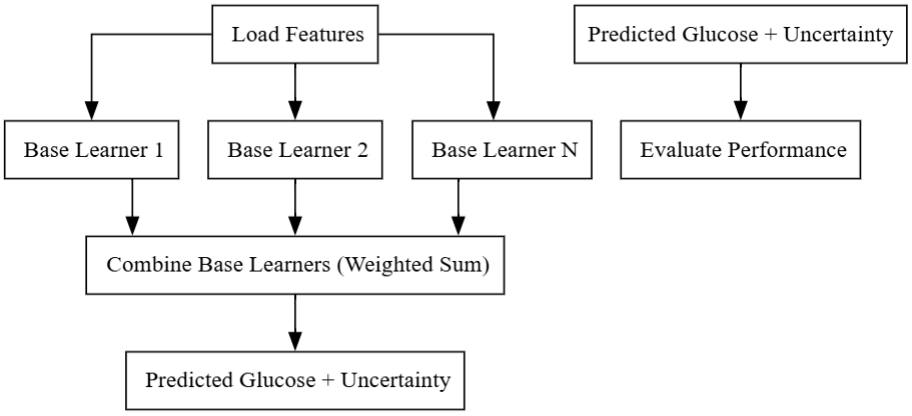


Fig. 3: NGB Internal Operations Architecture.

4. RESULTS DESCRIPTION

Fig. 4 shows the Dataset Dashboard, which provides a comprehensive statistical overview of the data used for the prediction models. At the top, key metrics such as Total Samples (844,946), number of features, mean glucose, and standard deviation are displayed. The main body of the dashboard presents glucose level statistics, including minimum, maximum, mean, and standard deviation values. A histogram visualizes the distribution of blood glucose levels across the dataset. Finally, the page details all the dataset features, categorized into PPG Signal Features, Demographic Features, and the Target Variable (glucose level).

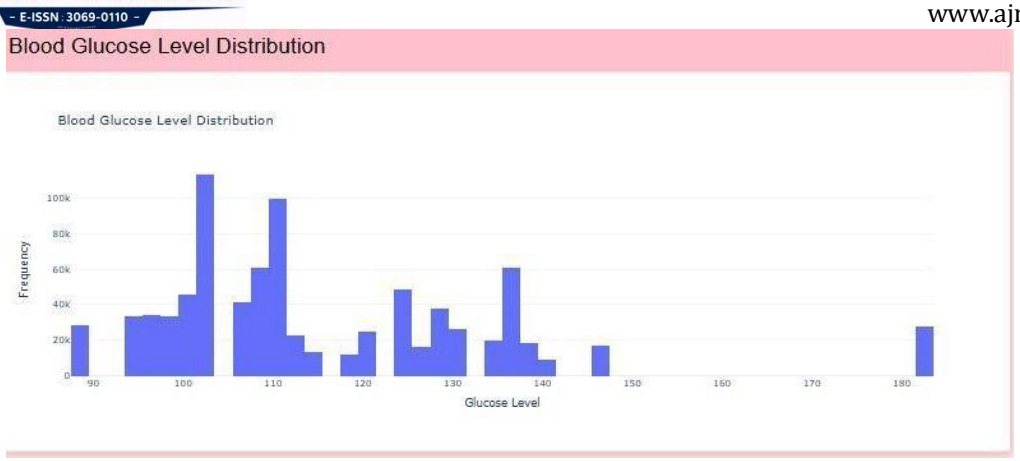


Fig. 4. Dataset dashboard with glucose level statistics and features

Fig. 5. shows the analysis page after training an NG Boosting model, a powerful ensemble technique. The model performance results for NG Boosting are exceptionally strong, with an MAE of only 0.4686 and an almost perfect R^2 Score of 1.0000. This suggests the model has learned the training data extremely well, potentially to the point of overfitting. The interface in this image also shows the results for the Decision Tree model just above it, allowing for a direct and striking comparison between the two tree-based methods, highlighting the significant performance gain from using the boosting algorithm.



Fig. 5. Analysis page displaying NG boosting model performance

Fig. 6 demonstrates the prediction capability of the NG Boosting model, which is one of the top-performing algorithms in the analysis phase. Using the identical input data as the previous example, the user has selected "NG Boosting" as the prediction model. The system calculates and displays the output, resulting in a predicted glucose level of 118.77 mg/dL. This value is remarkably close to the Decision Tree's prediction, yet slightly different, illustrating the subtle variations between models. This interface allows users to easily compare predictions from different algorithms for the same data sample.

AMERICAN JOURNAL OF MANAGEMENT AND IOT MEDICAL COMPUTING
Peer Reviewed, Referred & Indexed Journal
E-ISSN: 3069-0110
Vol.5, No.2(2) (2026)
www.ajmimc.com

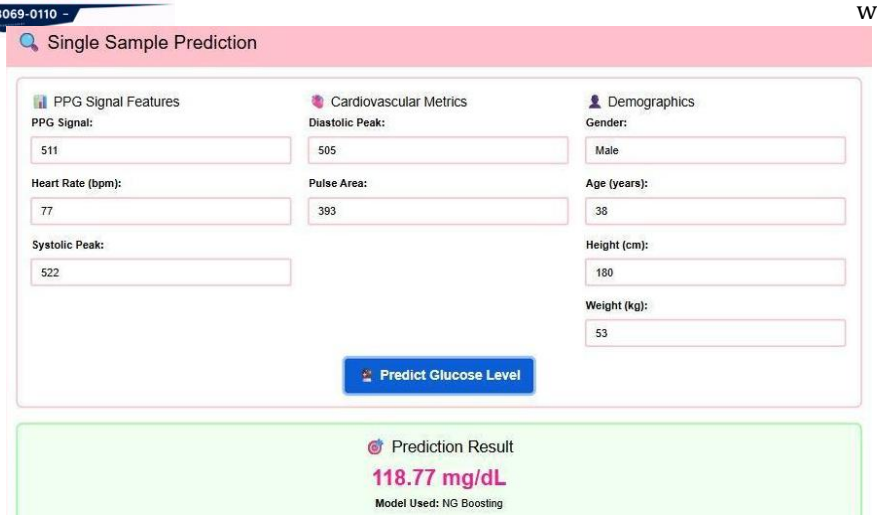


Fig. 6. Glucose level prediction using the NG boosting model

This section provides a detailed comparative analysis of the machine learning models developed for both regression (predicting interaction duration and blood glucose levels) and classification (predicting user satisfaction). The performance of each model was evaluated using standard metrics to identify the most effective and reliable algorithms for the given tasks. The analysis also examines the practical application of these models in single-sample predictions.

Table 1: comparative analysis of regression for blood glucose levels.

Model	MAE	RMSE	MSE	R^2 Score
LR	14.5925	18.8059	324.2124	0.1092
RR	14.5925	18.8059	324.2124	0.1092
Lasso	14.5972	18.1508	329.4505	0.0948
Decision Tree	11.5330	14.8742	221.2420	0.3921
NGB	0.0099	0.1321	0.0172	1.0000

The regression models were tasked with predicting a continuous value. Their performance was compared using Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the R^2 Score. A lower value for error metrics (MAE, MSE, RMSE) and a higher R^2 Score (closer to 1.0) indicate better model performance

5. CONCLUSION

The developed system effectively presents a non-invasive solution for predicting glucose levels by utilizing PPG signal characteristics along with demographic information within a machine learning-based framework. Several regression techniques, including LR, RR, Lasso, DT, and NG Boosting, were applied and comparatively analyzed to determine the most suitable model for accurate prediction. The evaluation results demonstrate enhanced performance across metrics such as MAE, MSE, RMSE, and R^2, with NG Boosting exhibiting a stronger ability to model complex relationships and deliver more consistent predictions. The deployment of these models within a Flask-based application allows both real-time individual predictions and batch processing, improving usability and accessibility. In contrast to invasive monitoring techniques, this approach provides a more scalable, efficient, and cost-effective alternative for continuous glucose level estimation.

REFERENCES

[1] DeFronzo, R.A.; Bonadonna, R.C.; Ferrannini, E. Pathogenesis of NIDDM: A Balanced Overview. *Diabetes Care* **1992**, *15*, 318–368.

[2] Bandeiras, C. Advancing Technology for a Healthier Humanity. *IEEE Potentials* **2020**, *39*, 6–8.

[3] International Diabetes Federation. Facts & Figures. Available online: https://www.idf.org/aboutdiabetes/what-is-diabetes/facts-figures.html

[4] Vashist, S.K.; Zheng, D.; Al-Rubeaan, K.; Luong, J.H.; Sheu, F.-S. Technology behind commercial devices for blood glucose monitoring in diabetes management: A review. *Anal. Chim. Acta* **2011**, *703*, 124–136.

[5] Akbulut, F.P.; Ikitimur, B.; Akan, A. Wearable sensor-based evaluation of psychosocial stress in patients with metabolic syndrome. *Artif. Intell. Med.* **2020**, *104*, 101824.

[6] Owess, M.M.; Owda, A.Y.; Owda, M.; Massad, S. Supervised Machine Learning-Based Models for Predicting Raised Blood Sugar. *Int. J. Environ. Res. Public Health* **2024**, *21*, 840. https://doi.org/10.3390/ijerph21070840

[7] Gudiño-Ochoa, A.; García-Rodríguez, J.A.; Cuevas-Chávez, J.I.; Ochoa-Ornelas, R.; Navarrete-Guzmán, A.; Vidrios-Serrano, C.; Sánchez-Arias, D.A. Enhanced Diabetes Detection and Blood Glucose Prediction Using TinyML-Integrated E-Nose and Breath Analysis: A Novel Approach Combining Synthetic and Real-World Data. *Bioengineering* **2024**, *11*, 1065. https://doi.org/10.3390/bioengineering11111065

[8] Ali, H.; Niazi, I.K.; White, D.; Akhter, M.N.; Madanian, S. Comparison of Machine Learning Models for Predicting Interstitial Glucose Using Smart Watch and Food Log. *Electronics* **2024**, *13*, 3192. https://doi.org/10.3390/electronics13163192.

[9] Gunathilaka, H.; Rajapaksha, R.; Kumarika, T.; Perera, D.; Herath, U.; Jayathilaka, C.; Liyanage, J.; Kalingamudali, S. Non-Invasive Diagnostic Approach for Diabetes Using Pulse Wave Analysis and Deep Learning. *Informatics* **2024**, *11*, 51. https://doi.org/10.3390/informatics11030051

[10] Leung, H.M.C.; Forlenza, G.P.; Prioleau, T.O.; Zhou, X. Noninvasive Glucose Sensing In Vivo. *Sensors* **2023**, *23*, 7057. https://doi.org/10.3390/s23167057

[11] Chauhan, A.S.; Varre, M.S.; Izuora, K.; Trabia, M.B.; Dufek, J.S. Prediction of Diabetes Mellitus Progression Using Supervised Machine Learning. *Sensors* **2023**, *23*, 4658. https://doi.org/10.3390/s23104658

[12] Islam, T.T.; Ahmed, M.S.; Hassanuzzaman, M.; Bin Amir, S.A.; Rahman, T. Blood Glucose Level Regression for Smartphone PPG Signals Using Machine Learning. *Appl. Sci.* **2021**, *11*, 618. https://doi.org/10.3390/app11020618.

[13] Chowdhury, M.H.; Shuzan, M.N.I.; Chowdhury, M.E.H.; Mahbub, Z.B.; Uddin, M.M.; Khandakar, A.; Reaz, M.B.I. Estimating Blood Pressure from the Photoplethysmogram Signal and Demographic Features Using Machine Learning Techniques. *Sensors* **2020**, *20*, 3127. https://doi.org/10.3390/s20113127

[14] G. Zhang *et al.*, "A Noninvasive Blood Glucose Monitoring System Based on Smartphone PPG Signal Processing and Machine Learning," in *IEEE Transactions on Industrial Informatics*, vol. 16, no. 11, pp. 7209-7218, Nov. 2020, doi: 10.1109/TII.2020.2975222.

[15] Chu, J.; Yang, W.-T.; Lu, W.-R.; Chang, Y.-T.; Hsieh, T.-H.; Yang, F.-L. 90% Accuracy for Photoplethysmography-Based Non-Invasive Blood Glucose Prediction by Deep Learning with Cohort Arrangement and Quarterly Measured HbA1c. *Sensors* **2021**, *21*, 7815. https://doi.org/10.3390/s21237815