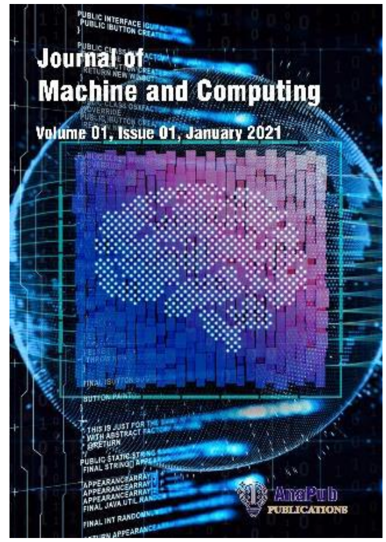
# Journal Pre-proof




Data Visualisation Models for Analytics Use Artificial Intelligence to Predict Diabetes in Women

**Hussein Z. Almngoshi, Mithra C, Abburu Srirama Kanaka Ratnam, Subhashini Shanmugam, Saravanan V and Bhaskar Marapelli**

DOI: 10.53759/7669/jmc202505043

Reference: JMC202505043

Journal: Journal of Machine and Computing.

Received 20 May 2024

Revised form 06 October 2024

Accepted 15 December 2024

**Please cite this article as:** Hussein Z. Almngoshi, Mithra C, Abburu Srirama Kanaka Ratnam, Subhashini Shanmugam, Saravanan V and Bhaskar Marapelli, “Data Visualisation Models for Analytics Use Artificial Intelligence to Predict Diabetes in Women”, Journal of Machine and Computing. (2025). Doi: https://doi.org/10.53759/7669/jmc202505043



AnaPub
PUBLICATIONS

# Data Visualisation Models for Analytics Use Artificial Intelligence to Predict Diabetes in Women

**Hussein Z. Almngoshi[1,2], C. Mithra[3], Abburu Srirama Kanaka Ratnam[4], Subhashini Shanmugam[5,*], V. Saravanan[6], Bhaskar Marapelli[7]**

[1]Department of Computer Science, College of Computer Science and Information Technology, University of Kerbala, Karbala, Iraq.

[2]Department of Information Technology, College of Science, University of Warith Al-Anbiyaa, Karbala, Iraq.

*Email: hussein.z@uokerbala.edu.iq*

[3]Department of Computer Science and Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Avadi, Chennai, 600062, Tamil Naud, India. *Email: mithrac.official@gmail.com*

[4]Departmen of Computer Science and Engineering CSE-Aritifical Intelligence & Machine Learning, Kallam HaranadhaReddy Institute of Technology (KHIT), Dasaripalem, Andhra Pradesh, 522019, India.

*Email: sriram.abburi@gmail.com*

[5]Department of Computer Science and Engineering, B. S. Abdur Rahman Crescent Institute of Science and Technology, Chennai, 600048, Tamil Nadu, India.

***Corresponding Author Email: subhashini[illegible]@gmail.com***

[6]Department of Electronics and Communication Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, 602105, Tamil Nadu, India.

*Email: sarava[illegible]nv.s[illegible]@saveetha.com*

[7]Department of Computer Science & Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, 522302, India. *Email: bhaskarmarapelli@gmail.com*

## Abstract

Identifying and classifying diabetes problems among women can be achieved using several Machine Learning (ML) algorithms. This paper additionally includes a summary of the evaluation of the performance of these MLs with algorithms on many different classification metrics. The AUC-ROC score is the best for Extreme Gradient Boost (XGB) with 85%, followed by SVM and Decision Trees (DT). Logistic Regression (LR) is showing low performance. However, the DT and XGB show promising performance against all the classification metrics. However, the SVM shows a lower support value; hence, it cannot be claimed to be a precious classifier. A study reveals that women are four times more susceptible to diabetic conditions than men. But, the healthcare systems do not give special attention to diabetic conditions in women. This study proposes to predict the probability of diabetes in females based on numerous medical conditions they may have. The ML accurately predicts diabetic complications based on biological conditions such as blood glucose levels, age, Body Mass Index (BMI), numerous pregnant women, and other factors.

***Keywords:*** *Diabetic Predication, SVM, Extreme Gradient Boost, Decision Tree, Logistic Regression*

## 1. Introduction

Diabetes is a non-communicable chronic disease that weakens the functionality of blood sugar flow throughout the body. Diabetes occurs for two primary reasons: one is when the pancreas in the human body loses its ability to secrete adequate amounts of insulin, and the other is when our body cannot utilize the amount of insulin generated by the pancreas. Diabetes can be relatively high in men and women, but the indication of disease has some specific variation in women. The rate of depression is considerably high compared to men; this would gradually increase Diabetes in women as the depression increases. Many unavoidable hormonal factors are contributing to a significant role in increasing depression in women. Some of the hormonal factors include changes in the menstrual cycle, spontaneous abortion, the period of being pregnant, postpartum, and menopause. Hence, this is said to be a lifestyle disease when not predicted at the early stage; this disease can lead to other health issues and even death. Diabetes is the root cause of other chronic diseases such as brain stroke, heart disease, kidney failure, and many more fatal diseases. Since Diabetes is considered the cause of lifestyle habits, prior analysis of Diabetes by experts can assist them in considerable change in food habits and fitness management. Many data mining techniques are used for predicting Diabetes at the initial stage by assessing the hormonal factors, which are immediately alarmed and ascertained. Many of the existing techniques for predicting Diabetes do not consider gestation, pregnancy, and the effects of post-pregnancy as vital factors. It is not fair to treat two anatomically different genders in the same manner. Women are more prone to suffer from hormonal imbalances, especially after motherhood.

The hormones estrogen and progesterone play a significant role in regularizing blood sugar. Since women are more prone to hormonal fluctuations, especially after menopause, hormonal levels play a predominant role in increasing the chances of Diabetes. Hence, it is quintessential that the disease prediction models should pay special mention to these factors. The article focuses on predicting Diabetes by considering biological aspects confined to women rather than generalizing the disease predictions. Also, the statistical analysis reveals the contribution of each factor towards the onset of Diabetes in women.

The paper's organization is as follows: section 2 brings the state of art techniques in Diabetes prediction. Section 3 presents the prediction of Diabetes using various ML and distribution analyses. The Experimental results of ML and validation of algorithms through various assessment measures are shown in Section 4. Section 5 summarizes the research work.

## 2. Related Works

GDM, which is commonly called Gestational Diabetes Mellitus, is a type of disease due to the development of carbohydrate intolerance during pregnancy [1]. ML has been developed in order to analyze and predict gestational diabetes mellitus in the first trimester of pregnancy, which also ensures cost-effectiveness and higher accuracy [2].

A systematic review has been conducted for predicting Diabetes using ML, facilitating more feasible options to drill with the medical data [3]. A regression model is used for preprocessing the missing data, and Logistic Regression (LR) is used for the Feature Selection (FS) [4].

Features were obtained from the different hormonal factors of the model, which have higher performance and accuracy [5]. A comparative study was conducted to analyze diabetes samples of women with different classification algorithms, and the Support Vector Machine (SVM) contributed better prediction with the best outcome in sensitivity and accuracy [6].

The hormonal factor can be predicted separately by Naive Bayes (NB) classifiers because these factors are considered to be the alarming cause for a woman's depression, which in turn increases the chance for the cause of Diabetes [7-10]. A Principal Component Analysis (PCA) is used for FS after data is classified from the dataset [11].

At the same time, they were framing the sample dataset, which includes the waist circumference and a ratio of the waist to the hip area since they are considered symbolic marks for each undiagnosed person [12].

In this context, a PIMA dataset is assumed, and it is applied in which ML overrules diagnosing Diabetes during the training and validation process of the learning model [13]. Analysis of different predictions of the continued performance of various MLs revealed that NB, k-Nearest Neighbour (k-NN), gives a notable output supporting the early diagnosis [14].

A real-time study on persons affected by gestational Diabetes was performed, and various biological tests were performed for 5 years. Hence, the result produced about one-third of women who have Diabetes, and the rest of the women have developed an unpredicted tolerance level for glucose [15].

This type of real-time analysis has proved that the symptoms of Diabetes in women fall entirely different when compared to men. This throws light on the importance of hormonal factors in women. The prediction of Diabetes helps experts to suggest proper intake and adjust lifestyle activities with proper fitness management in women to make Diabetes control [16].

Predicting the diabetic-related risk complications through ML helps to prognose the other comorbidities associated with Diabetes [17]. Special attention in predicting Diabetes in women is done through ML like SVM, k-NN, NB, LR, and Decision Tree (DT) by analyzing women's physiology [18].

Association between Diabetes and metabolic risk factors with suitable sampling techniques is done efficiently by J48 and NB [19]. Another notable work in predicting Diabetes in women is weighing the feature importance [20]. A careful analysis shows that Random Forest (RF) could classify diabetic conditions more efficiently.

**3. Diabetes Prediction using Different ML-ML**

Diabetes has emerged as a lifestyle disease that has affected nearly 199 million women around the globe. Insufficient insulin secretion is the root cause of many other health disorders. They drive home a substantial economic loss to the family and the country, either directly/indirectly. Recent years have seen a surge in the count of women who have Diabetes. The domestic and social responsibilities, along with the physical makeup of women, make them more prone to diabetic conditions. The hormonal imbalance caused by pregnancies, miscarriages, and irregular menstrual cycles hinders the balance in insulin secretion.

The article mainly focuses on deploying various MLs for predicting the onset of Diabetes in females in the age group of above years (Figure 1). The dataset is collected by the National Institute of Diabetes and Digestive and Kidney Diseases and is available in the UCI-ML repository. The rich dataset contains attributes such as pregnancies, values of Oral Glucose Tolerance Test (OGTT), skin thickness, Body Mass Index (BMI), blood pressure, insulin level, pedigree diabetes function, and age of 768 women from the USA. The dependent variable is the outcome of whether the subject is prone to be diabetic or not.

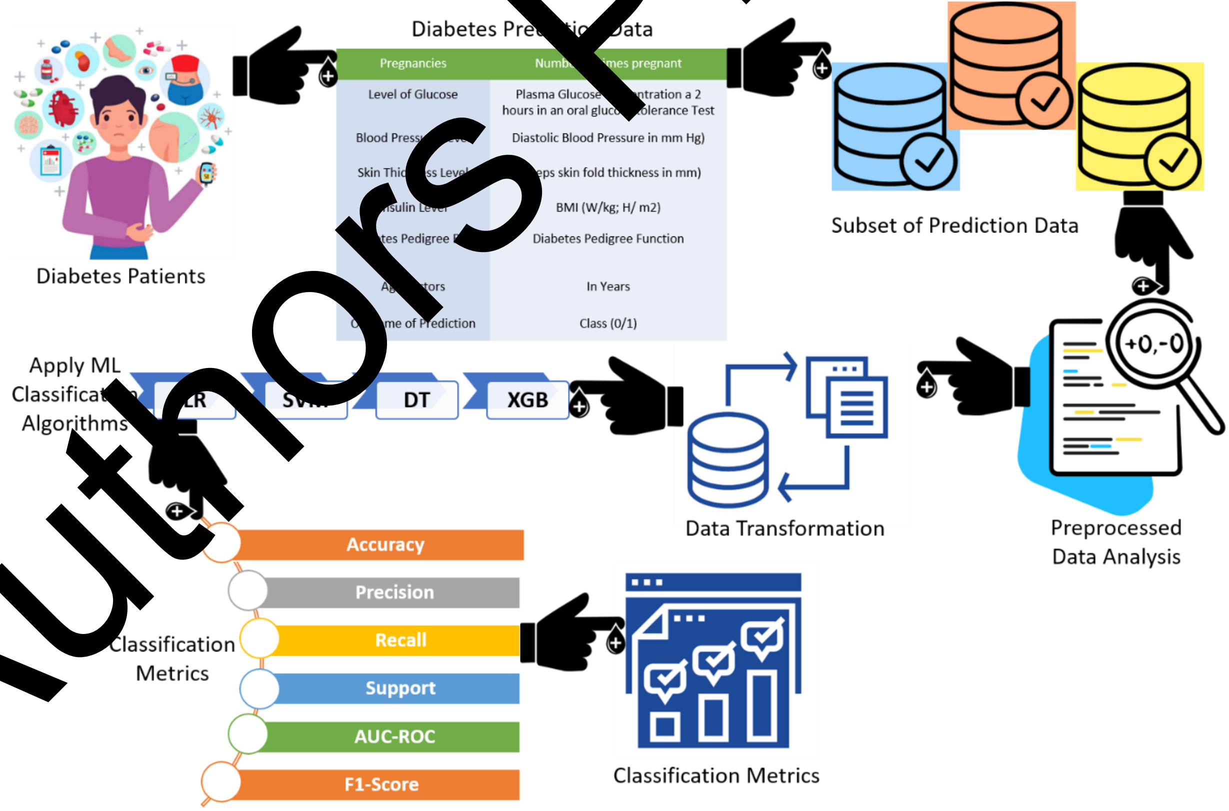
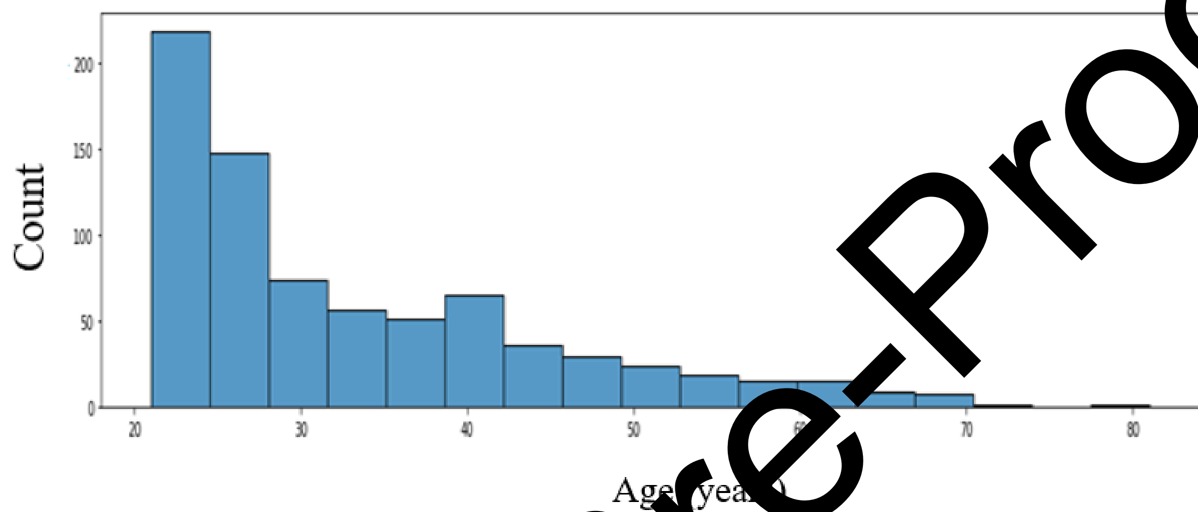


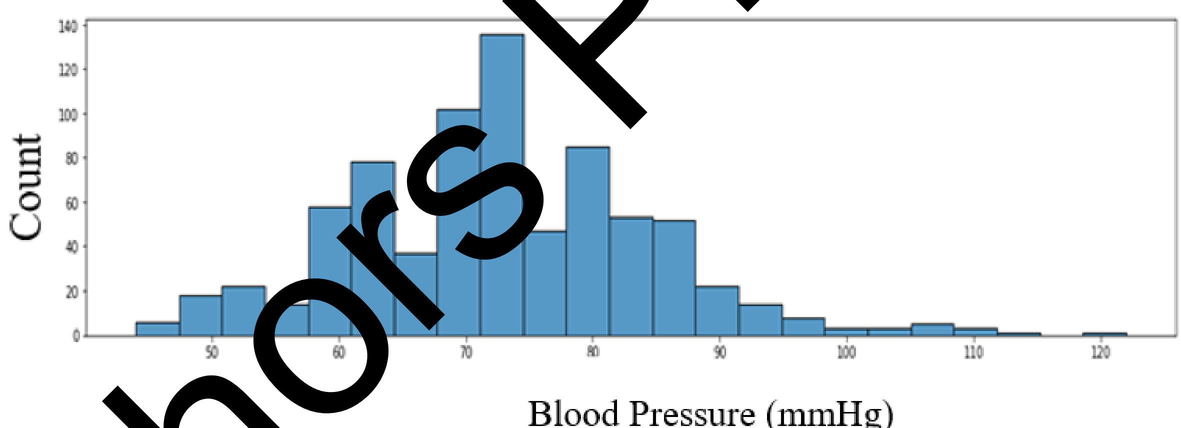

**Figure 1:** Deployment of various MLs for predicting the onset of Diabetes in women

### 3.1 Cleaning the data

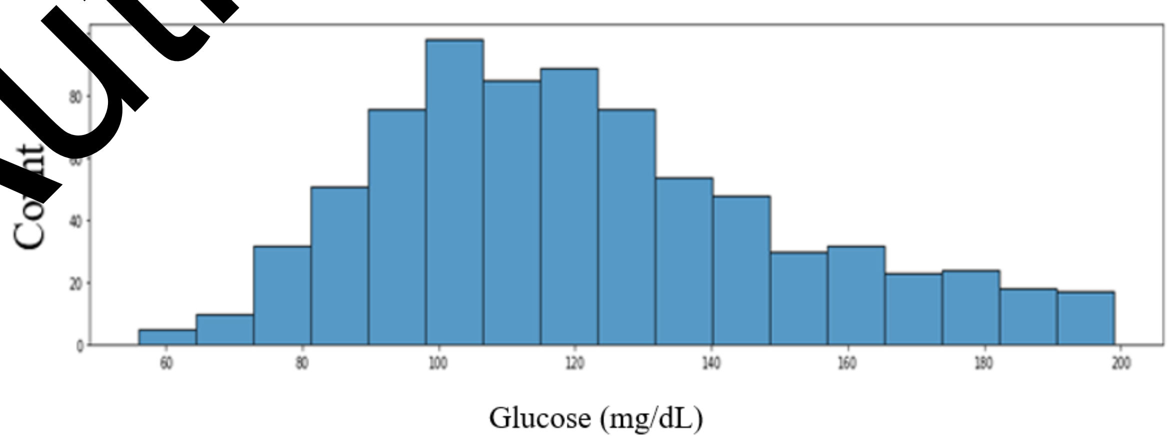
As the dataset is extracted from highly raw data, there is a chance of missing and duplicate values. Treating the data with such imbalances will significantly affect the accuracy of the models. The independent values of the dataset contain numerical data. The missing values are imputed with the median values of each column without loss of generality. The Exploratory Data Analysis (EDA) of the dataset is presented in Figure 2.



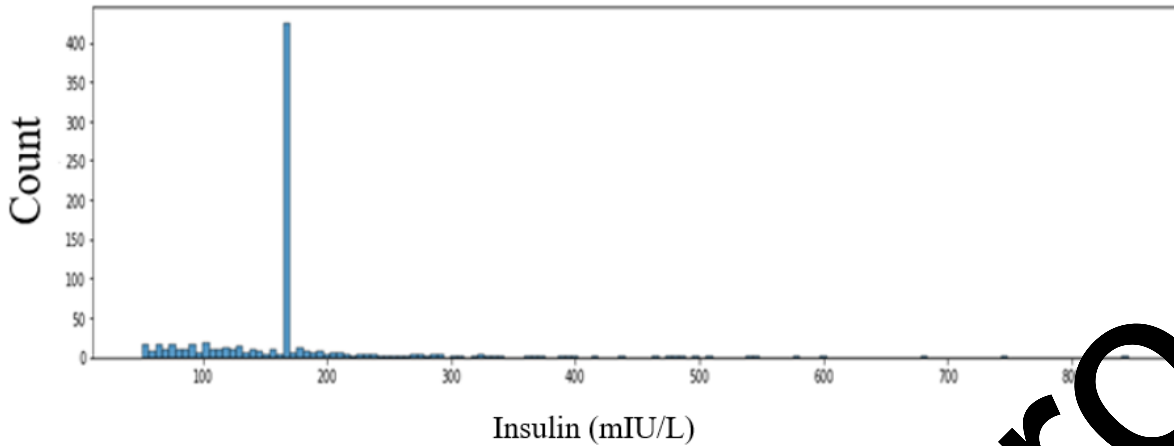

(a)



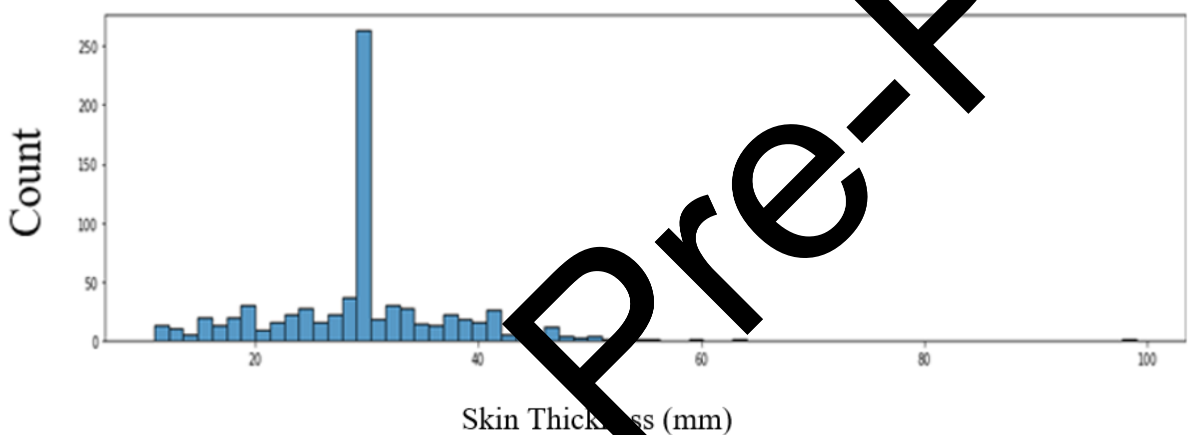

(b)



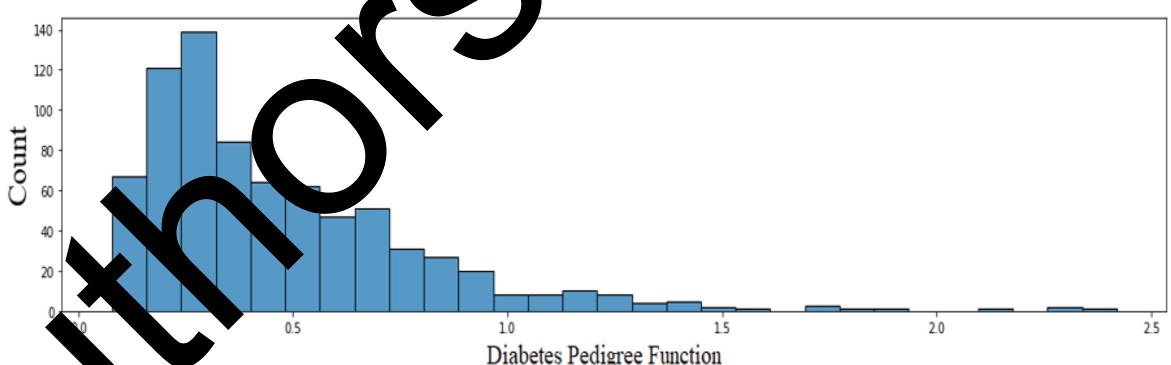
(c)


Count
0
50
100
150
200
250
300
350
400
100
200
300
400
500
600
700
Insulin (mIU/L)

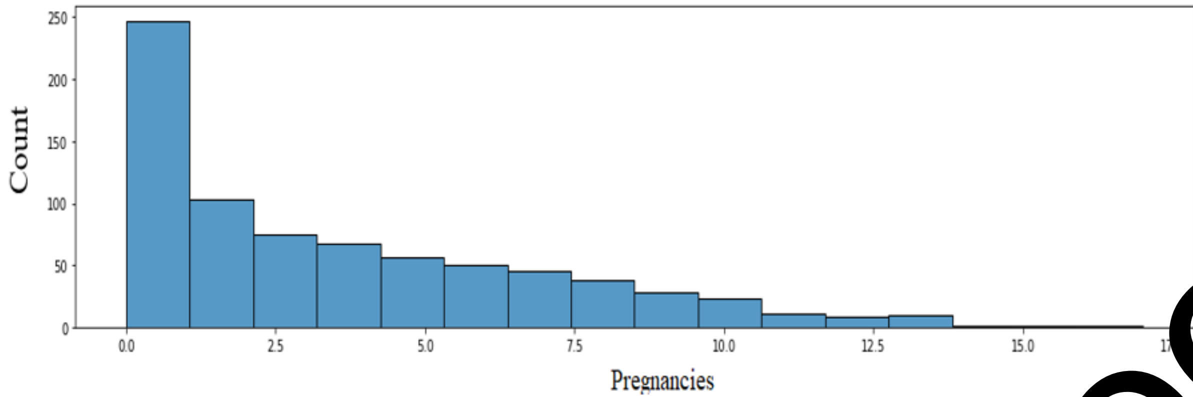
(d)


Count
0
50
100
150
200
250
20
40
60
80
100

(e)


Count
0
20
40
60
80
100
120
140
0.5
1.0
1.5
2.0
2.5
Diabetes Pedigree Function

(f)




**(g)**

**Figure 2:** Histogram of the Features from Pima Indiana Diabetic Dataset (a) Age (b) Blood Pressure, (c) Glucose, (d) BMI, (e) Insulin, (f) Skin Thickness, (g) Diabetes Pedigree Function, (h) Pregnancies

### 3.2 Distribution of Diabetes

It is imperative to assess the distribution of Diabetes across essential attributes. Figure 3 shows the distribution of Diabetes across various age groups. The onset of Diabetes is found to be expected even at the lower age of 21. The widespread presence of diabetic conditions can be seen in middle-aged people.

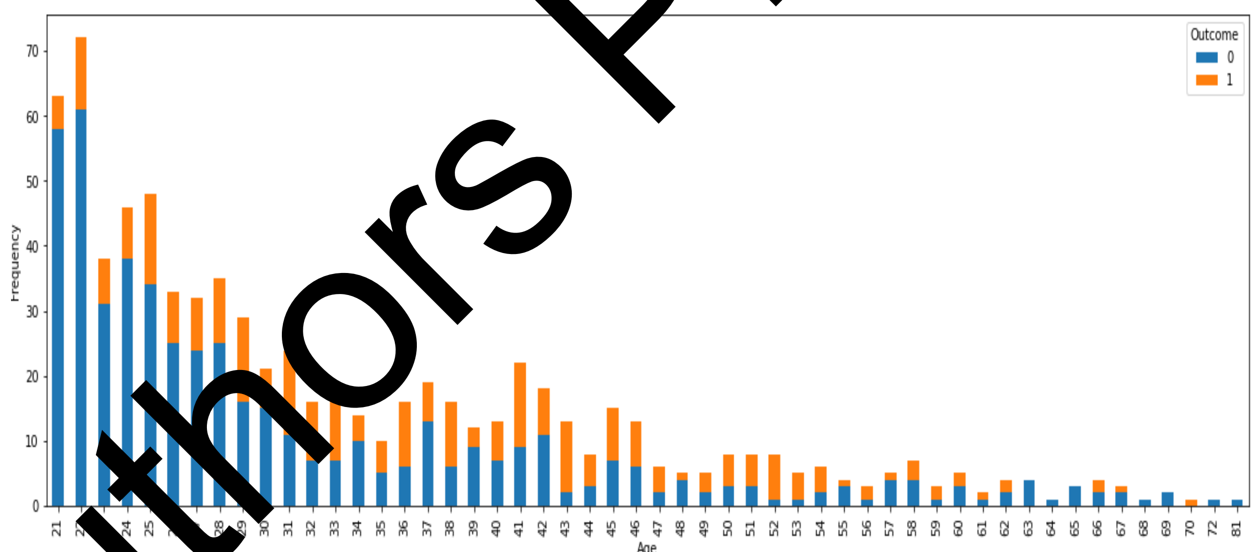



**Figure 3:** Distribution of Diabetes across age

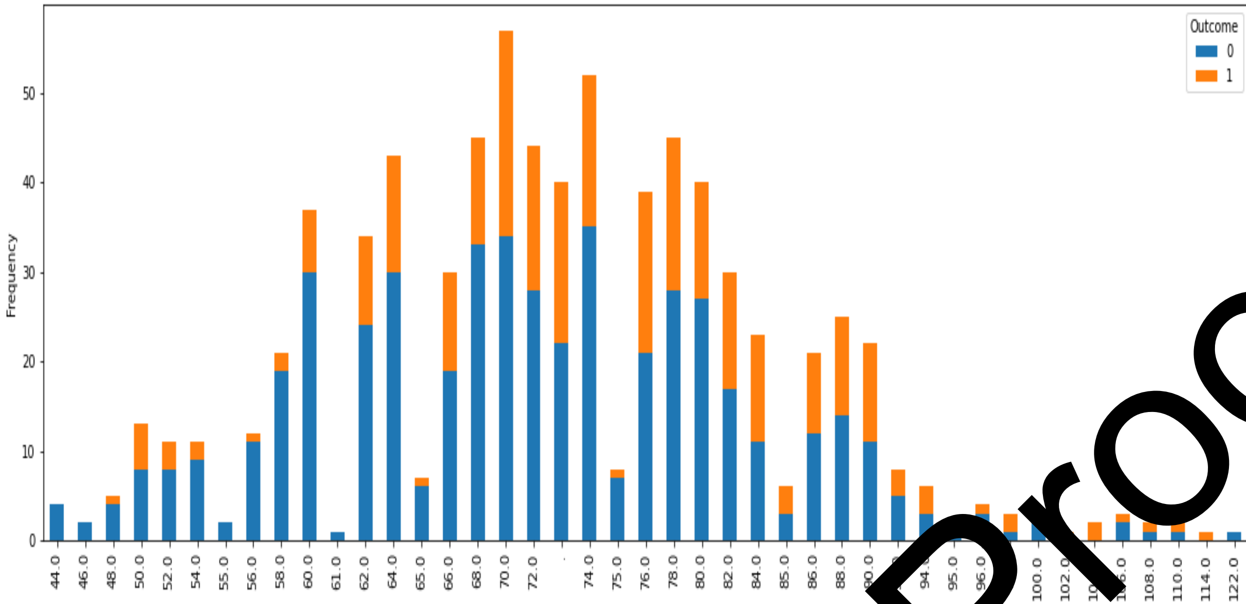



**Figure 4:** Distribution of Diabetes across blood pressure

Figure 4 shows the spread of Diabetes among women with several blood pressure levels. It can be observed that the percentage of women with a high value of Blood pressure is more prone to Diabetes. However, blood pressure contributes to the diabetic condition; even women with lower blood pressure values are subjected to Diabetes. This may be due to other physical conditions like increased BMI. Table 1 summarizes the correlation between various attributes and the onset of Diabetes. Correlation values of various features depict the significance of the attribute in predicting the disease. Glucose, BMI, age, and pregnancies are strongly correlated with Diabetes. The correlation analysis shows that skin thickness, blood pressure, insulin levels, and pedigree function do not profoundly impact the occurrence of Diabetes. Glucose levels and BMI play a significant role in the occurrence of Diabetes. Women with higher glucose levels and BMI should heed their body's insulin secretion.

Medical records face a serious confidentiality threat. Not many countries have released their open medical dataset. The proposed model is very generic; hence, it is suitable for predicting Diabetes from the same set of features across countries. Clinical studies have revealed that people who are obese are more prone to a diabetic condition. Also, a milder form of Diabetes, termed prediabetes, can be diagnosed much earlier from impaired glucose tolerance and resistance of the body to insulin. Gestational Diabetes may increase the chances of acquiring Type 2 diabetes. Apart from these factors, such correlation can be observed from the increase in age and the number of pregnancies since many physiological changes occur.

**Table 1:** Correlation between various attributes and Diabetes

| Attribute | Correlation value | Attribute | Correlation value |
|---|---|---|---|
| Glucose | 0.492591 | BMI | 0.311924 |

| | | | |
|---|---|---|---|
| Age | 0.238356 | Pregnancies | 0.221898 |
| Skin Thickness | 0.209897 | Insulin | 0.187262 |
| Blood Pressure | 0.173955 | Diabetes Pedigree Function | 0.173844 |

Figure 4 shows the histogram analysis of various features that are directly/indirectly associated with the prevalence of Diabetes in the women population. The histogram analysis helps to understand the nature of the population considered for the study. Though this does not involve direct prediction, this analysis gives insight into the analyzed data. The performance tuning of the model can be done only after a detailed analysis of the underlying data. The histogram analysis does not improve the predictive power but helps understand the domain and nature of analyzed data. Figure 2 (a) shows that diabetic conditions are more common among the younger age groups because of gestation in women. Also, most of the population exhibits a moderate rise in glucose levels that are easy to treat. This is evident from Figure 2 (b). In addition, most of the population under study had high insulin levels, thickened skin, and diabetes pedigree functions in the range of 0.5 to 1.

The histogram, correlation, and disease distribution are comprehensive statistical analyses. The distribution of Diabetes across different age groups reveals the impact of the complication even among younger age groups. The correlation analysis portrays the relevance of features contributing to developing Diabetic conditions in women. The proposed work presents a novel ensemble model that is unique among its kind in predicting Diabetes in women.

**3.3 Prediction of Diabetes Using ML**

The healthcare sector has seen the increased deployment of Artificial Intelligence (AI) and ML in disease diagnosis and treatment, and they even play a prominent role in providing recommendations to medical practitioners. This section presents the usage of robust ML in predicting the occurrence of Diabetes among the women population by assessing the attributes listed in Table 2. The algorithms considered for the study are LR, SVM, DT, and XGB.

**3.1. Logistic Regression**

The LR predicts the classes by analyzing the associations among one or more dependent variables. The predictions can fall under finite classes depending on the problem being addressed. This classification algorithm uses a logistic function to fit the output of a linear function into binary classes (0 or 1). This probabilistic approach produces the dichotomic results precisely into two or more classes. This algorithm works better on variables with little correlation between them and estimates their odds log to fit them into an appropriate class. EQU (1) gives the sigmoid function that outputs two classes, as in the case of predicting the onset of Diabetes from the described dataset.

$$Y = \frac{1}{(1 + e^{-k})} \tag{1}$$

The value of 'k' is estimated as the sum of the weights (w) multiplied by the features (x) as

$k = w_0 + w_1 x_1 + \ldots + w_n x_n$. The objective function that minimizes the cost is given in EQU (2).

$$\text{Cost} = (-\frac{1}{n})\sum_{j=1}^{n} Y_j \text{Log}(Y_i) + (1 - Y_i)\text{Log}(1 - Y_i)) \tag{2}$$

**3.2 SVM**

Unlike LR, SVM can be used to perform regression and classification of problems. The best decision boundary, the hyperplane, separates the data points into multiple classes. The extreme cases of these hyperplanes are termed support vectors. The SVM is a small sample learning algorithm that uses structural risk minimization to classify the data. SVM preserves the linearly separable property of the data but handles non-linear data by deploying kernel tricks. EQU (3) explains the classification property of SVM.

$$y[w^T \varphi(x) + \text{bias}] = \begin{Bmatrix} \geq 0, \text{If}_{yes} \\ < 0, \text{If}_{no} \end{Bmatrix} \tag{3}$$

$w^T \varphi(x) + \text{bias}$ maps the virtual hyperplane that separates the classes.

**3.3. Decision Trees**

They are non-parametric supervised algorithms that can be used for regression and classification. The rationale behind learning in DT is to predict the class by learning rules from the underlying data. The DT classifier proceeds to the output class by recursively splitting the input data by assessing its attribute values. These trees can perform well on numerical and definite data. Misclassification costs are the primary metric to evaluate the performance of the tree. The proposed DT uses the Gini Index to partition the data values given in EQU (4).

$$\text{Gini}(X) = 1 - \sum_{j=1} \text{Probability}_j^2 \tag{4}$$

$\text{Prob}_j$ is the probability that the data X belongs to a specified class.

**3.4. Extreme Gradient Boost**

The literature reveals that boosting algorithms best handle noisy data as they are an ensemble of homogeneous trees. Apart from this, the XGB can effectively mitigate the effects of bias and variance in the final prediction. The individual learners are represented as B={$bl_1$, $bl_2$, $bl_3$, $bl_4$}. The final predicted output (Y') is estimated according to EQU (5).

$$Y_i' = \sum_{i=1}^{n} bl(X_i) \tag{5}$$

$X_i$ has indicated the features from the dataset. XGB can produce more accurate approximations from the individual learners because of the usage of the second-order partial derivatives that minimize the loss function, according to EQU (6). The inherent inbuilt L1 and L2 regularizations are generic to the classification.

$$\mathrm{Loss}^t = \sum_{i=1}^{n}[g_i bl_t(X_i) + \frac{1}{2} w_i bl_t^{2}(X_i)] + \Omega(bl_t) \qquad (6)$$

The values of $g_i$ and $w_i$ are the first and second-order gradients of the loss function, respectively. As mentioned above, the gradients are computed in EQU (7) and EQU (8).

$$g_i = \partial_{y'}(t-1) lo(y_i, y'^{t-1}) \qquad (7)$$

$$w_i = \partial^2_{y'^{t-1}} lo(y_i, y'^{t-1}) \qquad (8)$$

The objective function (lo) for the computation of the 1st and 2nd-order gradients is EQU (9).

$$lo(x) \approx lo(x_i) + lo'(x_i) bl_t(X_i) \qquad (9)$$

**4. Experimental Results**

As mentioned earlier, the experimental assessment of the algorithms with their results and inferences are discussed in this section. The algorithms are validated on important classification metrics such as accuracy, F1-score, Precision, Recall, Support, and Area Under the Curve (AUC)-Receiver Operating Characteristics (ROC).

**4.1.Classification Accuracy:** It is the percentage of samples of correctly classified data. It is measured as the ratio between the number of correct classifications against the total classifications made. EQU (10) explains the accuracy measurement of the Diabetes dataset.

$$\text{Accuracy} = \frac{\text{Number of Tests Rightly Classified as Diabetic/Non-Diabetic}}{\text{Complete Classifications}} \qquad (10)$$

**4.2.Precision** is the ratio of exactness of the model's predictions and is explained in EQU (11).

$$\text{Precision} = \frac{\text{Count of Tests Correctly Classified as Diabetic}}{\text{Count of Patients Classified as Diabetic}} \qquad (11)$$

**4.3.Recall:** In a random dataset, recall possesses an inverse relation with precision. However, one cannot always conclude that a decrease in precision will accompany an increase in the recall. This factor is determined by the degree of randomness in the dataset. EQU (12) calculates the recall of the classification results.

$$\text{Precision} = \frac{\text{Count of Tests Correctly Classified to be Diabeteic}}{\text{Number of Correctly Classified Samples}} \qquad (12)$$

**4.4.F1-score:** This measure balances the trade-off between precision and recall. It is the geometric mean of precision and recall. The expression for the F1-score is mentioned in EQU (13).

$$\text{F1-score} = \frac{2*\text{Precision}*\text{Recall}}{\text{Precision} + \text{Recall}} \quad (13)$$

### 4.5.AUC-ROC

Both these are measures for assessing the purity of predicted classes. ROC is a probabilistic measure, while AUC represents the degree of separability between both classes. Higher values indicate that the models can identify diabetic patients from the dataset.

### 4.6. Support

This is the measure of true responses that lie within a class.

**Table 2:** Performance assessment of various ML in identifying diabetic patients

| Algorithm | Accuracy | Precision | Recall | F1-score | AUC-ROC | Support |
|---|---|---|---|---|---|---|
| LR | 76.54 | 75.94 | 76.54 | [illegible] | [illegible] | 69.9 |
| SVM | 86.56 | 86 | 86 | 86 | 84 | 30.07 |
| DT | 99.98 | 99 | 99 | 99 | 84 | 72 |
| XGB | 99.98 | 99 | 99 | 99 | 85 | 75 |

The train-test split is 80-20. The AUC-ROC score is the best for XGB, with 85%, followed by SVM and DT. LR is showing low performance. However, the DT and XGB show promising performance against all the classification metrics. However, the SVM shows a lower support value; hence, it cannot be claimed to be a good classifier. Figure 4. shows the visual representation of the classification scores.

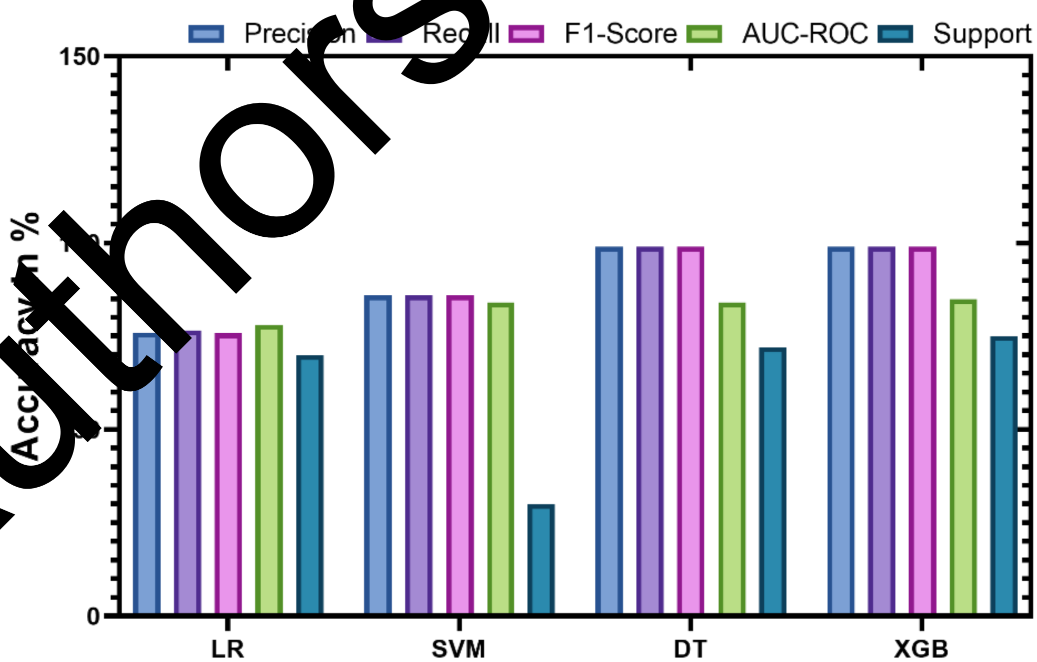



**Figure 5:** Comparative analysis of various ML in the prediction of diabetics

## 5. Conclusion and Future Work

A detailed focus is given on the relationship between various factors of Diabetes in women. Diabetes prediction uses robust classifiers: SVM, LR, DT, and XGB. Finally, a performance assessment of the algorithms is made to study their efficacy. The detailed analysis reveals that DT and XGB perform better than other algorithms in predicting Diabetes. This work can be extended by including FS approaches to include only the predominant features in training the model to obtain a more precise prediction. To summarise, this study elucidates the prevalence of diabetic conditions among women and emphasizes early prediction using various ML.

For the intended use of finding new approaches and effectively resolving specific challenges in diabetic care, researchers and developers are advised to investigate and experiment with an enormous number of AI and methodologies. By constantly improving [illegible] expertise and adopting AI technologies, we can boost the standard of treatment that people with diabetes experience and improve the results.

**Declarations**

**Funding-**Not Applicable

**Conflicts of Interest/Competing Interests-**Not applicable

**Availability of Data and Material-**Not applicable

**Code Availability-**Not Applicable

**References**


[1] S. Campanella, G. Paragliola, V. Cherubini, P. Pierleoni and L. Palma, "Towards Personalized AI-Based Diabetes Therapy: A Review," IEEE Journal of Biomedical and Health Informatics, vol. 28, no. 11, pp. 6944-6957, Nov. 2024, doi 10.1109/JBHI.2024.344[illegible]37.

[2] L. A. DiMeglio, C. Evans-Molina and R. A. Oram, "Type 1 diabetes", *Lancet*, vol. 391, no. 10138, pp. 2449-2462, 2018.

[3] E. A. Pustozerov et al., "Machine Learning Approach for Postprandial Blood Glucose Prediction in Gestational Diabetes Mellitus," IEEE Access, vol. 8, pp. 219308-219321, 2020, doi: 10.1109/ACCESS.2020.3042483.

[4] V. Chandrasekar et al., "Investigating the Use of Machine Learning Models to Understand the Drugs Permeability Across Placenta," IEEE Access, vol. 11, pp. 52726-52739, 2023, doi: 10.1109/ACCESS.2023.3272987.

[5] S. A. Abbas, A. Aslam, A. U. Rehman, W. A. Abbasi, S. Arif and S. Z. H. Kazmi, "K-Means and K-Medoids: Cluster Analysis on Birth Data Collected in City Muzaffarabad, Kashmir," IEEE Access, vol. 8, pp. 151847-151855, 2020, doi: 10.1109/ACCESS.2020.3014021.

[6] M. F. Khan et al., "Reinforcing Synthetic Data for Meticulous Survival Prediction of Patients Suffering From Left Ventricular Systolic Dysfunction," IEEE Access, vol. 9, pp. 72661-72669, 2021, doi: 10.1109/ACCESS.2021.3080617.

[7] A. V. Singh, V. Chandrasekar, P. Laux, A. Luch, S. P. Dakua, P. Zamboni, et al., "Micropatterned neurovascular interface to mimic the blood-brain Barrier's neurophysiology and micromechanical function: A BBB-on-CHIP model", *Cells*, vol. 11, no. 18, pp. 2801, 2022.

[8] V. Chandrasekar, A. V. Singh, R. S. Maharjan, S. P. Dakua, S. Balakrishnan, S. Dash, et al., "Perspectives on the technological aspects and biomedical applications of virus-like particles/nanoparticles in reproductive biology: Insights on the medicinal and toxicological outlook", *Adv. NanoBiomed Res.*, vol. 2, no. 8, 2022.

[9] Y. Akhtar, S. P. Dakua, A. Abdalla, O. M. Aboumarzouk, M. Y. Ansari, J. Abinahed, et al., "Risk assessment of computer-aided diagnostic software for hepatic resection", *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 6, no. 6, pp. 667-677, 2022.

[10] S. Franc et al., "Diabeo system combining a mobile app software with and without telemonitoring versus standard care: A randomized controlled trial in diabetes patients poorly controlled with a basal-bolus insulin regimen", *Diabetes Technol. Therapeutics*, vol. 22, no. 12, pp. 904-911, 2020.

[11] M. K. Hasan, M. A. Alam, D. Das, E. Hossain, and M. Hasan, "Diabetes prediction using ensembling of different machine learning classifiers", *IEEE Access*, vol. 8, pp. 76516-76531, 2020.

[12] V. Jackins, S. Vimal, M. Kaliappan and M. Y. Lee, "AI-based smart prediction of clinical disease using random forest classifier and naive Bayes", *J. Supercomputing*, vol. 77, pp. 5198-5219, 2021.

[13] M. Sevil, M. Rashid, I. Hajizadeh, M. Park, L. Quinn and A. Cinar, "Physical activity and psychological stress detection and assessment of their effects on glucose concentration predictions in diabetes management", *IEEE Trans. Biomed. Eng.*, vol. 68, no. 7, pp. 2251-2260, 2021

[14] F. A. Khan, K. Zeb, M. Al-Rakhami, A. Derhab, and S. A. C. Bukhari, "Detection and prediction of diabetes using data mining: A comprehensive review", *IEEE Access*, vol. 9, pp. 43711-43735, 2021.

[15] L. Gosak, A. Svensek, M. Lorber and G. Stiglic, "Artificial intelligence-based prediction of diabetic foot risk in patients with diabetes: A literature review", *Appl. Sci.*, vol. 13, no. 5, 2023.

[16] H. Nemat, H. Khadem, M. R. Eissa, J. Elliott, and M. Benaissa, "Blood glucose level prediction: Advanced deep-ensemble learning approach", *IEEE J. Biomed. Health Inform.*, vol. 26, no. 6, pp. 2758-2769, 2022.

[17] S.-M. Lee, D.-Y. Kim and J. Woo, "Glucose transformer: Forecasting glucose level and events of hyperglycemia and hypoglycemia", *IEEE J. Biomed. Health Inform.*, vol. 27, no. 3, pp. 1600-1611, 2023.

[18] F. Zaccardi, D. R. Webb, T. Yates, and M. J. Davies, "Pathophysiology of type 1 and type 2 diabetes mellitus: A 90-year perspective", *Postgraduate Med. J.*, vol. 92, no. 1084, pp. 63-69, 2016.

[19] L. A. DiMeglio, C. Evans-Molina, and R. A. Oram, "Type 1 diabetes", *Lancet*, vol. 391, no. 10138, pp. 2449-2462, 2018.

[20] C. K. Boughton and R. Hovorka, "The role of automated insulin delivery technology in diabetes", *Diabetologia*, 2024