International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

IoT-Based Hybrid Ensemble Machine Learning Model for Efficient Diabetes Mellitus Prediction

Rashmi Ashtagi[1], Dr. Pritam Dhumale[2], Deepak Mane[3], H M Naveen[4], Ranjeet Vasant Bidwe[5], Bhushan Zope[6]

Submitted: 24/05/2023 **Revised**: 06/07/2023 **Accepted**: 25/07/2023

Abstract: The widespread chronic ailment known as diabetes affects millions of people worldwide. Early detection and an understanding of the underlying reasons can significantly enhance the outcomes for patients and public health initiatives. We propose a non-invasive self-care system that uses IoT and machine learning (ML) to check blood sugar and other critical markers for early diabetes prediction in response to the growing need for IoT-based mobile healthcare applications to anticipate diseases, including diabetes. Our main objective is to offer cutting-edge diabetes management tools that facilitate patient monitoring and technology-aided decision-making. Our objective was to create a hybrid ensemble ML system that used boosting and bagging methods to anticipate the onset of diabetes. In order to collect data from 13,421 participants and validate the model, an offline survey and an online application based on the Internet of Things were utilized. The fifteen items on the form were all about lifestyle, family history, and health. Our ML model performs better than existing methods, according to the experimental findings from both bases, making it a promising method for better diabetes prediction and management. Our technology has the potential to greatly improve early identification and care for those who are at risk of acquiring diabetes around the world by combining the Internet of Things and machine learning.

Keywords: *Support Vector Machine, Decision Trees, Random Forests, and machine learning for diabetes prediction*

I. Introduction

Diabetes can badly harm several different body organs, including the kidneys, nerves, heart, and eyes [5, 6]. Diabetes is recognized for shortening life expectancy because it is serious and persistent. To address this, machine learning categorization and prediction algorithms have been researched [7]. For all illnesses, no single technique, however, consistently outperforms others in terms of efficacy and accuracy. While one classifier may perform very well on a specific dataset, another method may fare better for a number of disorders [8]. Its occurrence has been steadily increasing as a result of a number of factors, such as changes in dietary practices, lifestyle adjustments, and sedentary behavior. Early detection and accurate diabetes prediction are essential for providing effective diabetes care and preventing the effects of the disease [1]. Machine learning, which has shown to be a powerful tool in the healthcare business, has made it feasible to develop predictive models based on medical data. Massive datasets of patient medical data can be mined for patterns and associations by machine learning algorithms, which can then be used to forecast disease.

One result of people's changing lifestyles on their health is the growth of Diabetes Mellitus, which can affect anyone at any age. Once this syndrome appears, it persists for the rest of one's life as a condition characterized by insufficient pancreatic insulin secretion, which raises blood sugar levels [2]. It is thought that diabetes mellitus, which affects people of all ages, is a global health problem. Diabetes was a 6% contributor to mortality in 2000; over the next 20 years, this number is expected to rise to 45%. Diabetes may be caused by a variety of factors, such as changing eating habits, sedentary lifestyles, smoking, and consuming a lot of high-protein and junk food [23].

[1]*Assistant Professor, Department of Computer Engineering, Vishwakarma Institute of Technology, Bibwewadi, Pune, Maharashtra, India*
[2]*HOD & Associate Professor, Computer Science Engineering Department, Jain College of Engineering and Research, Belagavi, Karnataka, India*
[3]*Associate Professor, Department of Computer Engineering, Vishwakarma Institute of Technology, Bibwewadi, Pune, Maharashtra, India*
[4]*Assistant Professor, Department of Mechanical Engineering, RYM Engineering College, Ballari, Karnataka, India*
[5]*Symbiosis Institute of Technology, Symbiosis International (Deemed University) (SIU), Lavale, Pune, Maharashtra, India*
[6]*Symbiosis Institute of Technology, Symbiosis International (Deemed University) (SIU), Lavale, Pune, Maharashtra, India*
rashmiashtagi@gmail.com1, pritamdhumale.jcer@gmail.com2, dtmane@gmail.com3, naveenhm001@gmail.com4, ranjeetbidwe@hotmail.com5, bhushan.zope@hotmail.com6

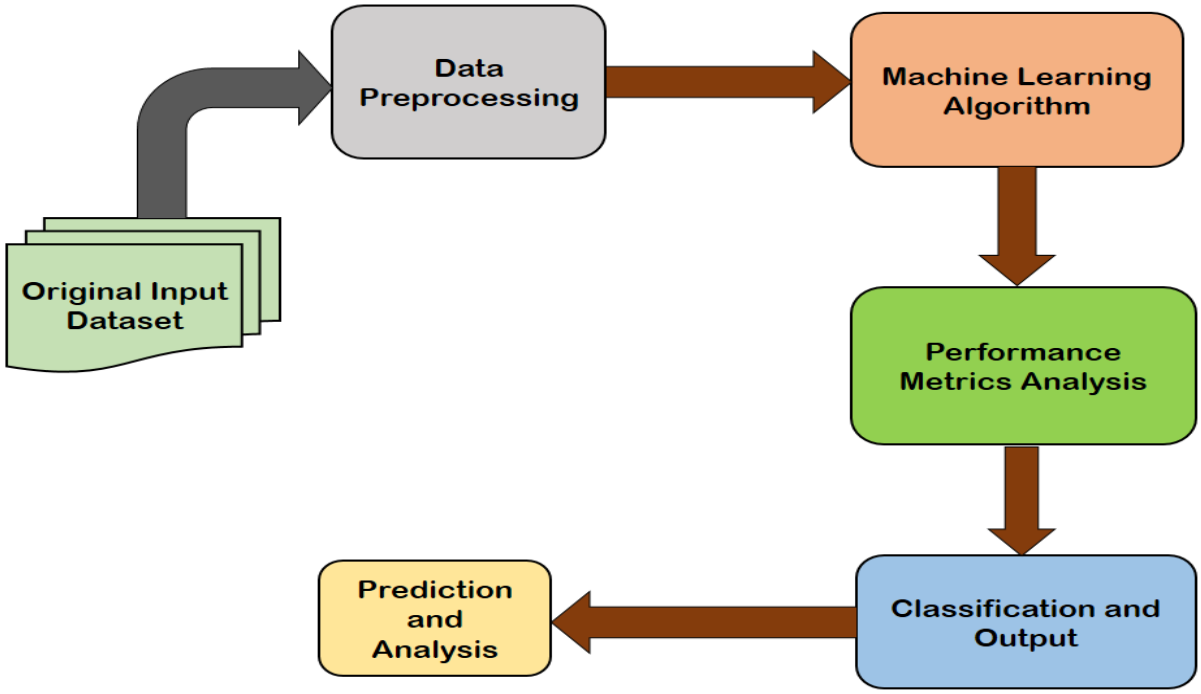

Fig 1: Proposed model for Prediction and analysis

For the condition to be identified, it is crucial to monitor blood sugar levels as well as take other factors like age, physical activity, and lifestyle choices into account. Microvascular and macrovascular difficulties that injure their organs and tissues are considered to affect one-third of diabetics [3]. Medical experts must identify pre-diabetic patients and investigate their glucose tolerance and insulin resistance in order to avert vascular complications.

Patients with diabetes frequently have neuropathy and nephropathy, two significant side effects that can harm peripheral nerves and lead to heart failure. Of diabetics, neuropathy affects more than 50% of them. In diabetic nephropathy, which is characterized by increased albumin levels in the urine, the renal organs are harmed and may fail. Effective predictions of neuropathy and nephropathy require a big amount of data collection from numerous diabetic individuals because [4] the precision of the prediction significantly depends on the quantity of training data taken into account. A diabetic dataset is created by compiling historical data from numerous medical organizations on a range of diabetes patients in order to anticipate sickness. This dataset is then used to train the system so that it can predict outcomes when given a set of input values [5]. One approach to diabetes prediction employs data mining, a technique for extracting useful information from huge databases. Data mining techniques can be effectively modified to solve the diabetic prediction challenge. For this prediction problem, numerous alternative scientific approaches are also accessible [6].

1.1 Symptoms:

Blood sugar levels can affect diabetes symptoms, and some people, particularly those with type 2 diabetes or pre-diabetes, may not exhibit any symptoms at all. Type-1 diabetes symptoms, on the other hand, can manifest more suddenly and can be more severe. The following are typical indicators of both kinds:

- Ketones in the urine Increased thirst
- often urinating
- extreme hunger
- Unaccounted-for weight loss
- Fatigue
- distorted vision
- Slowly heaving wounds
- Infections that come back repeatedly, such gum, skin, or vaginal infections
- BMI of over 25 indicates obesity

Diabetes, which is more common in families, can occur in people whose blood HDL cholesterol levels are less than 40 milligrams per decilitre. A higher risk exists for those with polycystic ovarian syndrome who are older than 45 and come from a certain racial or ethnic background, such as African Americans, Native Americans, Latin Americans, or persons from the Asia-Pacific region. In particular, if they have a sedentary lifestyle, this is true. The term "Internet of Things" (IoT) refers to a network of physically connected objects that may be accessed online.

These "choses" have the ability to gather and disseminate information autonomously and without human involvement via a network because they are outfitted with sensors and IP addresses [9]. The use of technology-based healthcare procedures presents a huge opportunity to improve the effectiveness and calibre of medical care as people become more conscious of and committed to taking care of their own health [10], [13].

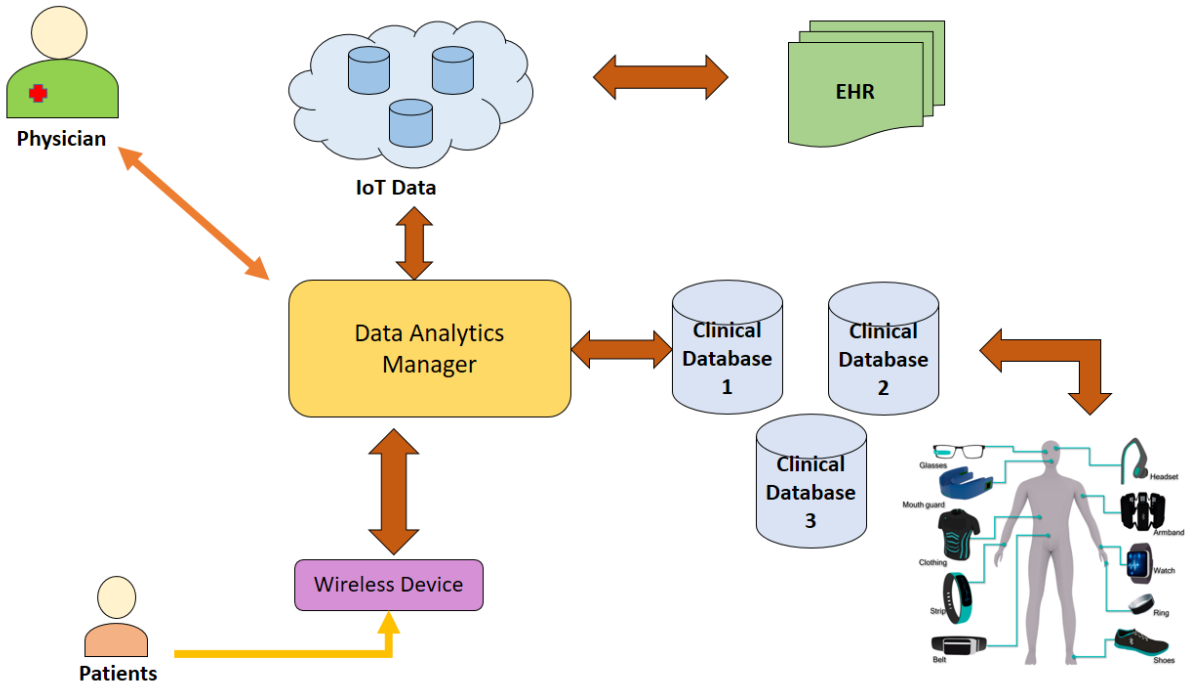

Fig 2: IoT Based health monitoring system

Real-time data collecting and analysis is one benefit of implementing IoT in healthcare. The implementation of this revolutionary strategy in an IoT-enabled hospital is shown in Figure 1. Patients with diabetes will receive ID cards that are connected to a secure cloud that stores their electronic medical records. This information will be simple for doctors and staff to access on a tablet or computer, improving the effectiveness and efficiency of patient treatment.This work offers a diabetes prediction method based on the Disease Influence Measure (DIM). The suggested approach makes use of DIM to evaluate how various parameters affect the process of disease prediction [7]. By adding DIM, the algorithm improves the precision and comprehension of the prediction outcomes, offering insightful knowledge about the significant elements influencing diabetes.

The DIM-based strategy is a viable way to enhance disease prediction and can considerably advance the field of managing diabetic healthcare. With the help of machine learning techniques, we hope to create a diabetes prediction model and investigate the various influences that may contribute to the development of the disease. Various demographic, lifestyle, and clinical factors collected from diabetic patients and non-diabetics will make up the dataset utilized for training and testing the model [8]. Additionally, in order to evaluate the effect of various factors on the prediction process, we will create a disease influence measure (DIM). DIM will offer insightful information about the relative importance of many risk factors influencing the development of diabetes [9].

In the suggested work, we want to categorize and predict diabetic mellitus (DM) using a unique method that combines different classifiers. We seek to improve the early diabetes detection accuracy by overcoming the constraints of individual classifiers, thereby saving lives. The major objective of this study is to create an information system that can predict diabetes more accurately, resulting in better management and treatment for those who are at risk for the condition.

II. Review Of Literature

Chronic metabolic disease known as diabetes mellitus (DM) is characterized by inadequate insulin production [1]. According to the CDC statistics for 2021, 11.3% of the US population has DM, making it the ailment that affects around 1 in 10 people worldwide. Both pharmaceutical and non-pharmacological therapies have been used to lessen the effects of UDM, and new pharmaceutical agents and insulin delivery systems have been developed to enhance glycaemic control. Nevertheless, despite these developments, there is still a substantial variance in glucose control among individuals

with varied characteristics, which raises the risk of complications from diabetes.

To lessen the severity of uncontrolled diabetic mellitus (UDM), a variety of pharmaceutical and non-pharmacological therapies have been used [7]. Glycemic control has been greatly enhanced by the development of new pharmacological agents and insulin delivery systems [10]. Due to the overnight fasting requirement, these routine tests might not always be practical. Additionally, more extensive biological marker data have not been properly included into earlier prediction techniques.

Although they are known to interact with glycemic status, biological indicators such serum electrolytes and haematological indices play a crucial role in the prediction of UDM utilizing machine learning (ML) algorithms. It may be possible to increase the precision and efficacy of UDM prediction by including these biological markers into ML-based prediction models, providing a fresh and all-encompassing method for managing diabetes. Additionally, physical characteristics like weight and height can reveal obesity, which frequently coexists with diabetes [22]. Erythrocyte counts and other blood indicators can also affect haemoglobin levels, which in turn affects HbA1C levels [23]. If a suitable prediction model is created using information from a representative sample, the correlation between these characteristics and glycemic state offers an alternate means of monitoring UDM. We used the All of Us (AoU) research program, which offers a sizable and racially varied sample of the US population, for this study [24].

In comparison to traditional glucose tests, the prediction of UDM utilizing novel features offers a cost-effective approach for glucose monitoring. By identifying those who are at risk, using a predictive model with a wider range of patient characteristics can help decrease complications caused by diabetes and improve the quality of life for diabetes patients. In this situation, our work used a supervised machine learning approach and several patient features to effectively predict UDM. It emphasized the value of patient features and physiological indicators in predicting UDM in the absence of routine glycemic status monitoring.

III. Publically Available Datasets

The Pima Indian Diabetes Dataset (UCI Machine Learning Repository, 1998) is among the best datasets to evaluate machine learning algorithms for diabetes prediction. Based on diagnostic indicators like pregnancy, blood sugar, blood pressure, skin its thickness, diabetic pedigree function, insulin, body mass index, and age, the Pima Indian dataset, which was made public by the National Institute of Diabetes and Digestive and Kidney Diseases in 2013, can be used to determine whether a patient has diabetes.

Table 1: Description of Dataset

Features	Details
Pregnancies	Number of pregnancies that have occurred.
Glucose	After two hours, the plasma glucose concentration (a test of glucose tolerance).
Blood Pressure	Diastolic blood pressure is measured in mm Hg (heartbeats per minute).
Skin Thickness	(mm) The thickness of the triceps' skin folds.
Insulin	Serum insulin concentration (mu U/ml) after two hours.
BMI	Index of body mass.
Diabetes Pedigree Function	Diabetes Family History.
Age	Age expressed in years.
Outcome	As a result, there is a class variable (0: diabetic Negative, 1: Diabetic Positive).

IV. Proposed System

Tables 1 display the data set that was used to forecast diabetes. The dependent variable in this dataset is diabetic parameters, while the independent variables are various variables. The dependent diabetes traits have binary values, with "zero" signifying no diabetes and "one" signifying diabetes. With a ratio of 70:30 for the training

and testing datasets, the total dataset is split into two groups. All four categorization algorithms were applied to provide predictions.

The k-Nearest Neighbours (k-NN) and Support Vector Machine (SVM) classifiers were trained using the training data. These classifiers were used to forecast the results of the test set after training. For the suggested methodology for diabetes prediction, two separate datasets must be pre-processed. To identify relevant variables for diabetes detection in this phase, we look at attribute correlation. After that, training and testing sets of the data are created. The training data is used in a variety of ways to build predictive machine learning models. We use a variety of metrics to evaluate the model's performance. Finally, the best ML model is deployed in a web application using Flask..

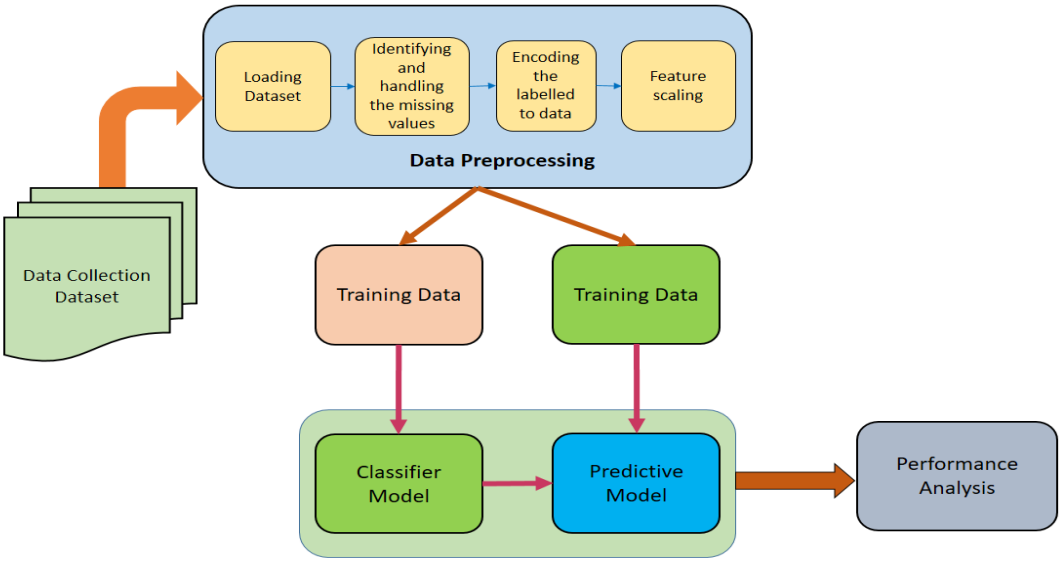

Fig 3: proposed model and flow of proposed work

1. Data Collection:

To make sure that our model was robust, we gathered two distinct datasets, each with a different amount of features or components. These datasets were assembled from a range of sources, including data on diabetes, global health characteristics, and details from several health institutes.

2. Data Preprocessing:

Preparing raw data for analysis and the creation of prediction models requires a critical step called data preparation. It entails preparing the data for machine learning algorithms by cleaning, converting, and organizing it. Enhancing data quality, handling missing values, eliminating inconsistencies, and creating features that more accurately depict the underlying patterns in the data are the main goals of data preprocessing.

2.1 Data Cleaning: Data cleaning entails dealing with outliers, duplicate records, and missing values. Depending on the circumstances, missing data can be eliminated or imputed. To prevent bias in the analysis, duplicate records are often eliminated.

2.2 Data Transformation: Data transformation is sometimes necessary to make sure that the data adheres to a particular distribution. To scale the data inside a certain range, common transformations include logarithmic, square root, or normalization.

2.3 Feature Selection: Selection of Features: The effectiveness and performance of the model can be enhanced by choosing the features that are the most pertinent. To lessen complexity and overfitting, duplicate or irrelevant features may be deleted.

2.4 Feature Engineering: By developing new features from already existing ones, we can gain new knowledge and improve the model's capacity to recognize patterns.

2.5 Data Splitting: Usually, the dataset is used to construct training and testing sets. The testing set is used to evaluate the model's performance on test data after it has been trained using the training set.

2.6 Handling Unbalanced Data: If the classes are unbalanced, the dataset can be balanced using methods like oversampling, undersampling, or artificial data generation.

2.7 Standardization/Normalization: By adjusting the features' scales to a similar range, one feature can be kept from outweighing the others while learning.

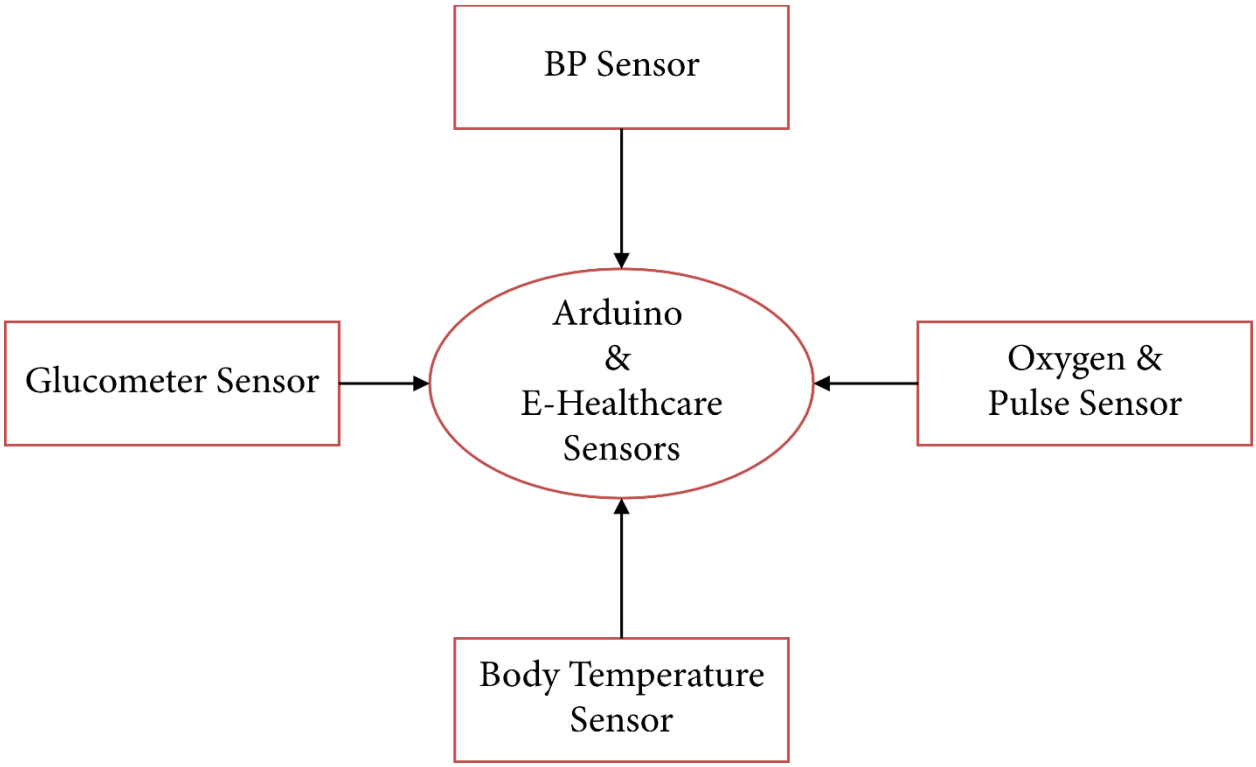

Fig 4: Diabetes monitoring using IoT based sensors

3. Machine Learning Algorithm:

3.1 Naive Bays Algorithm:

Based on the Bayes theorem, the Naive Bayes automated learning method is used to address a number of categorization issues. In this piece, we'll go into greater detail on the Naive Bayes method to clear up any misunderstandings.

Recompile and Prepare Data: Gather an input vector- and class-labeled training dataset.

Calculate each class's preliminary probability, P(Yi), using the occurrences from the formation table. For calculating a class's preliminary probability, use the following formula:

$$P(Yi = yt) = \frac{count(Yi = yt)}{N}$$

Where count(Yi = yt) is the number of instances of the class yt in the series of formation instances, and N is the total number of formation instances.

Calculate the residual probabilities: Calculate the posterior probabilities of the entry vector P(Y|X) for each specific class using the Bayes theorem:

$$P(Yi|Xi) = \frac{\left(P(Xi|Yi) * P(Yi)\right)}{P(Xii)}$$

The class with the highest likelihood of returning must be chosen in order to predict the name of the class for an impending and unknown vector of entry.

3.2 Support Vector Machine:

It is renowned for its capacity to effectively handle big datasets and is particularly excellent in addressing both linear and nonlinear classification issues. To handle numerous concerns like routing, localization, fault detection, congestion control, and communication problems, SVM has also been employed in a variety of fields, including Wireless Sensor Networks (WSNs).

A hyperplane with the greatest margin of separation between the two classes must be found using SVM. You can represent this hyperplane by:

$$W * X + B = 0$$

SVM's decision meaning is described as follows:

$$F(x) = sign(W * X + B)$$

SVM minimises the classification error while maximising the margin between the classes. As a result, the optimisation problem is formulated as follows:

$$\text{minimize: } \tfrac{1}{2}\|w\|^2 + C\Sigma_i \xi_i,$$

$$\text{subject to: } y_i(w \cdot x_i + b) \geq 1 - \xi_i,$$

3.3 Decision Tree:

The decision tree, also known as the classification tree and the regression tree, is a method of guided learning that predicts categorised variables for continuously categorised entry and exit variables. Because of its visual representation, human interpretation is straightforward and supports decision-making.

A splitting rule, also known as feature collection measure, is a heuristic used to choose the optimum criterion for data partitioning in order to produce the most efficient data separation. It aids in identifying the tuple breakpoints at a specific node. Each feature (or attribute) is given a rank or score by the attribute selecting measure based on how well it can describe the provided dataset. The splitting attribute is determined by the attribute with the greatest score. Split points are also chosen for qualities with continuous values in order to define each branch.

1. S is the algorithm's focal point.
2. The algorithm analyses each non-used group attribute, S, and calculates the associated entropy (H) and information gain (IG).

$$Entropy\ E(S) = \sum_{j=1}^{c} - Pi\ log_2 Pi$$

And Information Gain Calculate as:

$$Information\ Gain\ (IG) = Entrophy\ (before) - \sum_{j=1}^{c} Entrophy(j, after)$$

3. The algorithm chooses the property with the highest information gain or lowest entropy.
4. Depending on the chosen characteristic, the given S is divided.
5. The algorithm then iteratively applies to each subassembly, focusing only on characteristics that weren't previously chosen. Ginni Index Calculated as:

$$Ginni\ (Gi) = 1 - \sum_{k=1}^{j} (Pi)^2$$

6. To do this, the algorithm gradually divides the data base into various subgroups according to entropy or information quantity, and does so until a pause requirement is met.
7. Calculating Variance: Steps

$$Variance\ (Vi) = \frac{\sum \sqrt{(X - Xi)}}{n}$$

3.4 K- Nearest Neighbour:

The KNN is the simple algorithm in terms of implementation in machine learning methods is the k-NN (k-Nearest Neighbours) technique. In order to develop the model, the training dataset which acts as the reference data must be stored. The technique locates the closest data points, referred to as the "nearest neighbours," inside the training dataset when producing a estimate for a new statistics point.

The k-NN algorithm can be summed up mathematically as follows:

- Every point of data in the dataset used for training should be measured against the new data point (X).
- Based on the estimated distances, choose the k closest neighbours.
- When performing classification tasks, choose the dominating class among the k closest neighbours and set it as the predicted class for the newly generated data point.
- When completing regression activities, determine the average or weighted mean of the k nearest neighbour's goal values and utilise that quantity as the forecast value for the new data point.

Different proximity metrics, such as the distance from Manhattan or the distance calculated by Euclid, can be used to determine the separation among two data points based on the type of information and the issue at hand.

4. Performance Metrics:

The accuracy (ACC) is calculated as the percentage of correctly classified instances, whether they are normal or attacks, and is determined by the following formula:

$$ACC = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$

The formula for calculating precision (P), which is the proportion of pertinent instances among the identified instances:

$$P = \frac{TP}{(TP + FP)}$$

Recall (R) is calculated as the ratio of the number of relevant instances over the total number of relevant instances discovered:

$$\mathrm{R} = \frac{\mathrm{TP}}{(\mathrm{TP} + \mathrm{FN})}$$

The F1-Score is a metric that combines recall and precision into one number. It can be calculated using the formula below as the weighted average of recall and precision:

$$\mathrm{F1Score} = \frac{(2 * \mathrm{P} * \mathrm{R})}{(\mathrm{P} + \mathrm{R})}$$

In particular, when α = 1, the formula for the F1-Score simplifies. Overall, these formulas allow us to calculate accuracy, precision, recall, and the F1-Score, which are commonly used metrics for evaluating classification performance.

V. Result And Discussion

The framework for predicting diabetes was developed and evaluated in this study using a range of machine learning approaches. Nine parameters were included in the sample, which included 768 records, including age, BMI, skin

thickness, blood pressure, blood glucose levels, and insulin levels. Utilizing criteria including accuracy, precision, recall, and F1-score, the effectiveness of four classifiers Naive Bayes (NB), Decision Tree (DT), k-Nearest Neighbours (KNN), and Support Vector Machine (SVM) was evaluated.

Table 2: Performance metric comparison for different method

Method	Accuracy in (%)	Precision in (%)	Recall in (%)	F1-Score in (%)
NB	98.12	97.23	98.87	98.88
DT	98.87	98.01	99.1	98.72
KNN	99.11	98.01	98.66	98.81
SVM	99.65	97.33	98.44	98.76

One of the classifiers, SVM, demonstrated the best accuracy of 99.65%, proving its ability to accurately predict outcomes for both diabetes and non-diabetic individuals. SVM also managed to achieve a respectable precision of 97.33%, proving its ability to correctly distinguish real positive cases from all predicted positive ones. Additionally, SVM's strong recall rate of 98.44% showed that it could successfully identify a sizeable portion of the dataset's true positive cases.

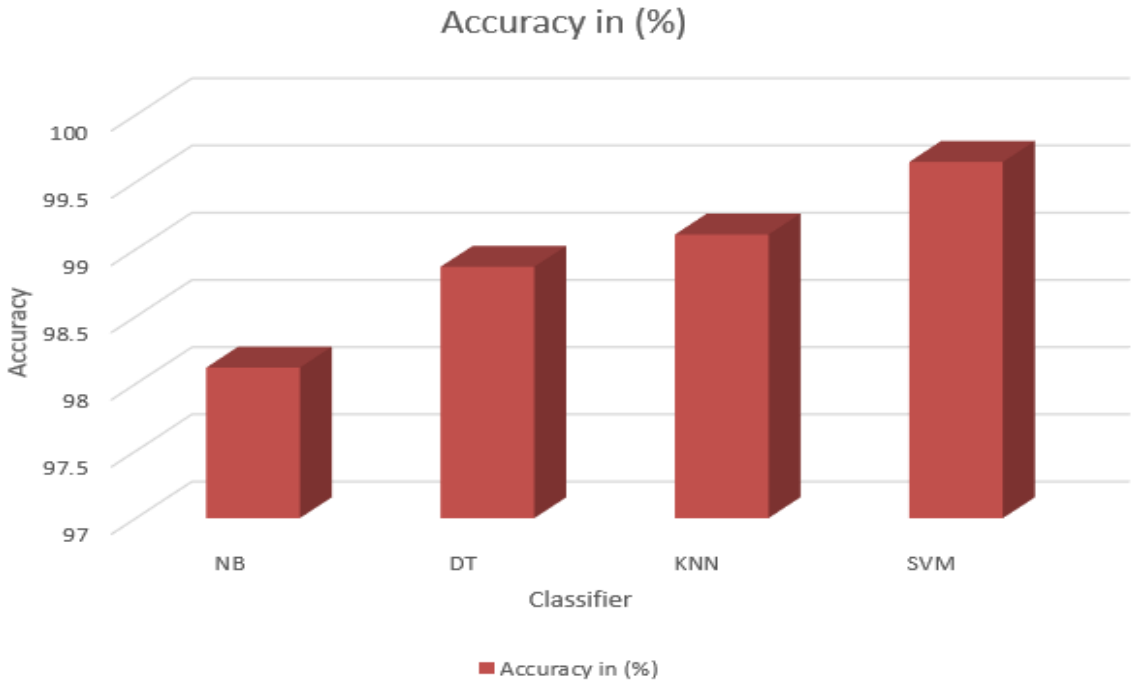

Fig 5: Accuracy comparison of different methods

The Decision Tree classifier, which had an accuracy of 98.87%, came in second place to SVM. It demonstrated a balanced performance with a precision of 98.01%, a recall of 99.10%, and an F1-score of 98.72%. The accuracy, precision, recall, and F1-score of KNN were all excellent, with a combined score of 99.11%, 98.01%, 98.66%, and 98.81%.

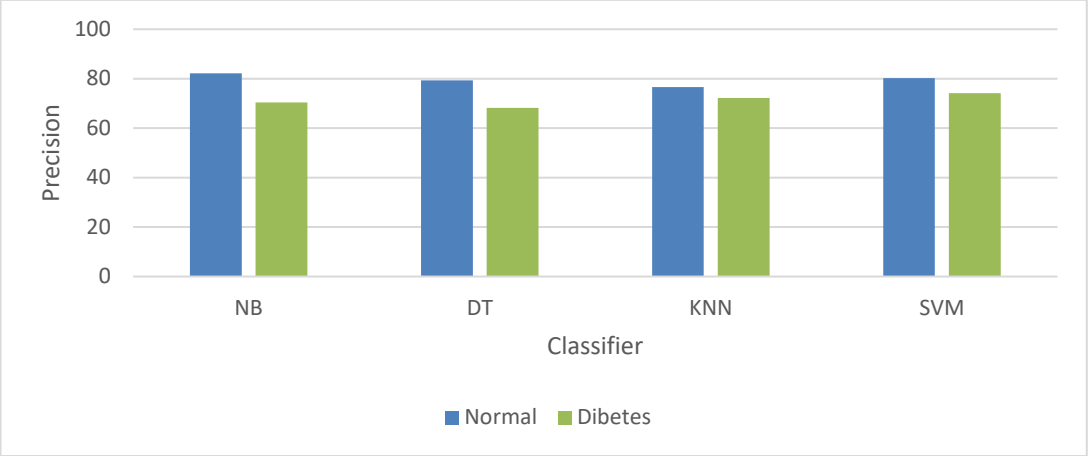

Fig 6: Precision comparison of different method

The Naive Bayes classifier performed well overall, achieving 98.12% accuracy, 97.23% precision, 98.87% recall, and 98.88% F1-score. It showed high predictive powers for diabetes identification although performing somewhat worse than the other classifiers. SVM, one of the classifiers, showed the best accuracy of 99.65%, demonstrating its capacity to predict outcomes correctly for both diabetic and non-diabetic people.

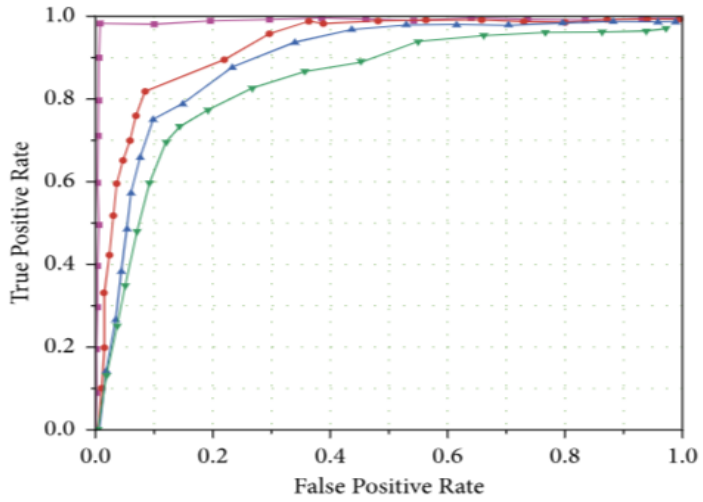

Fig 7: Prima Diabetes ROC curve with AUC

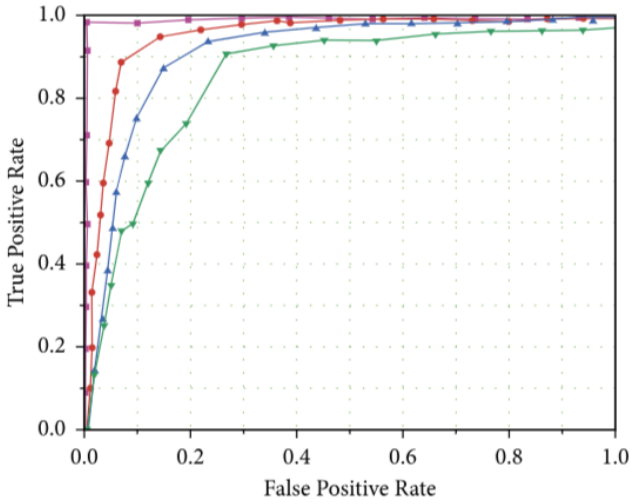

Fig 8: IoT Based collected Diabetes dataset ROC curve with AUC

A reasonable precision of 97.33% was also attained by SVM, demonstrating its capacity to correctly identify genuine positive cases among all predicted positive cases. Additionally, SVM had a high recall rate of 98.44%, demonstrating its capacity to catch a significant fraction of the dataset's genuine positive cases.

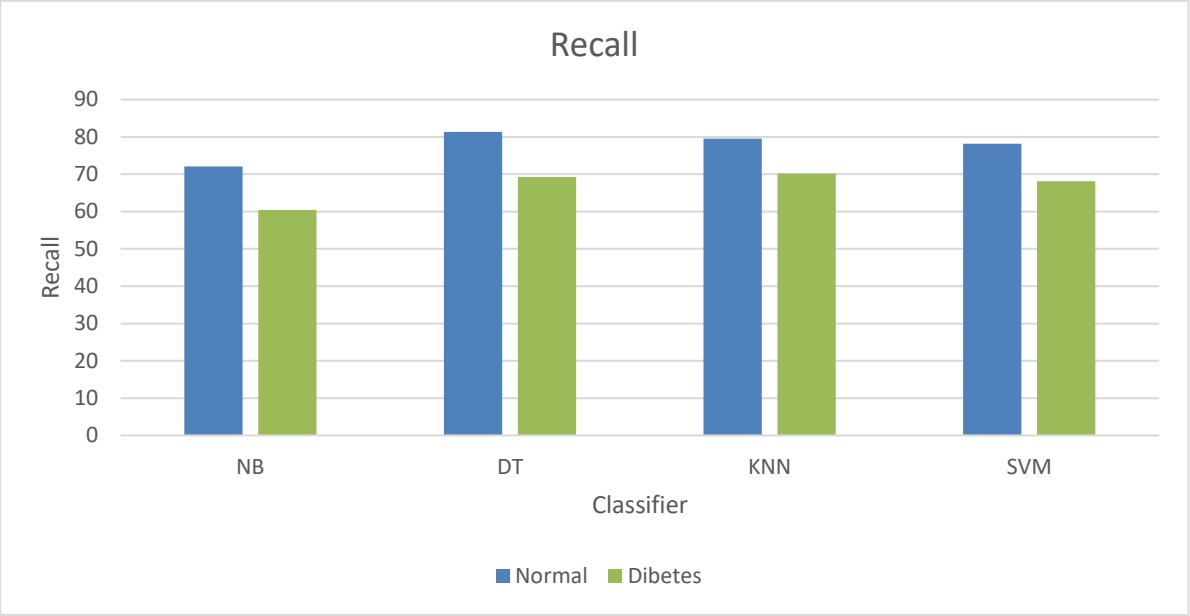

Fig 8: Recall comparison of different method

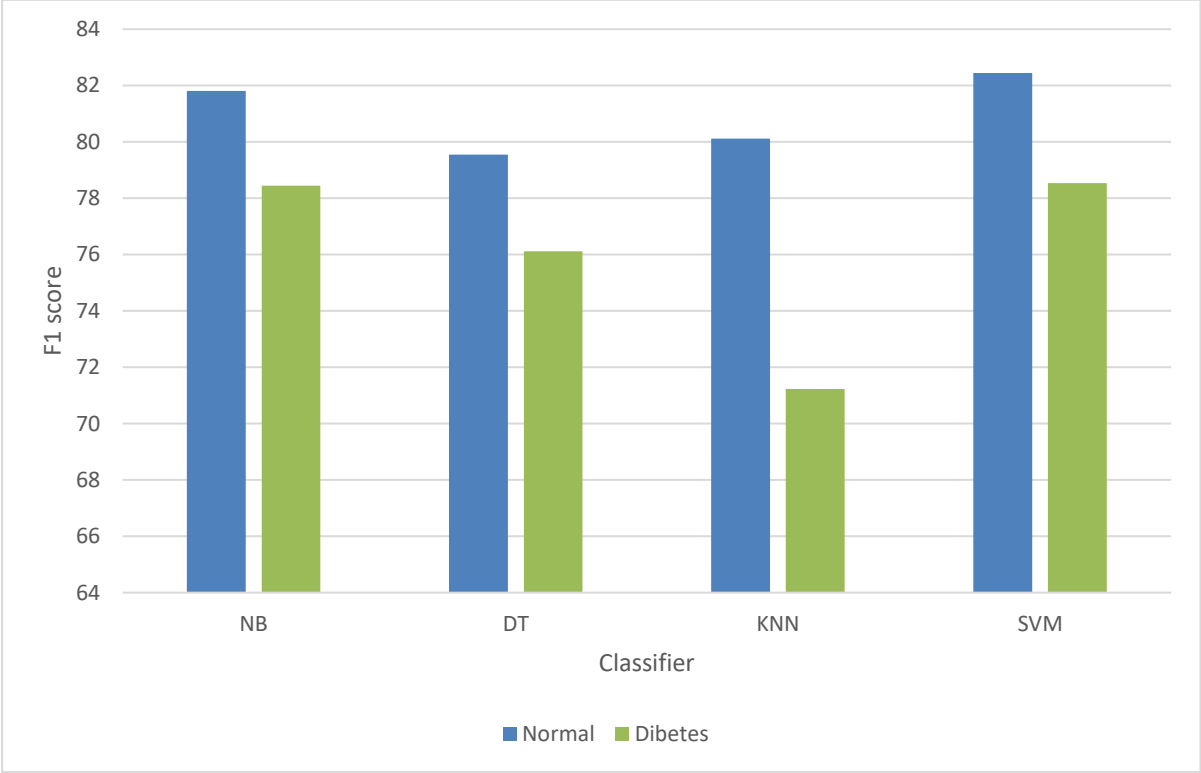

Fig 10: F1 Score comparison of different method

Analysis of the data revealed that the classifiers' accuracy in differentiating between cases of normal and diabetes occurrence varied. In both normal and diabetic situations, SVM demonstrated the highest accuracy, obtaining 82.44% accuracy in normal cases and 78.54% accuracy in diabetic ones. This implies that SVM performed equally well in classifying both classes. Following closely, NB demonstrated competitive results when compared to the other classifiers, f1 score accuracy of 81.81% for normal cases and 78.44% for diabetic cases. Compared to SVM and NB, the Decision Tree classifier had a slightly lower accuracy of 79.55% for normal instances and 76.12% for diabetic cases, suggesting its potential to distinguish between the two classes as shown in figure 10.

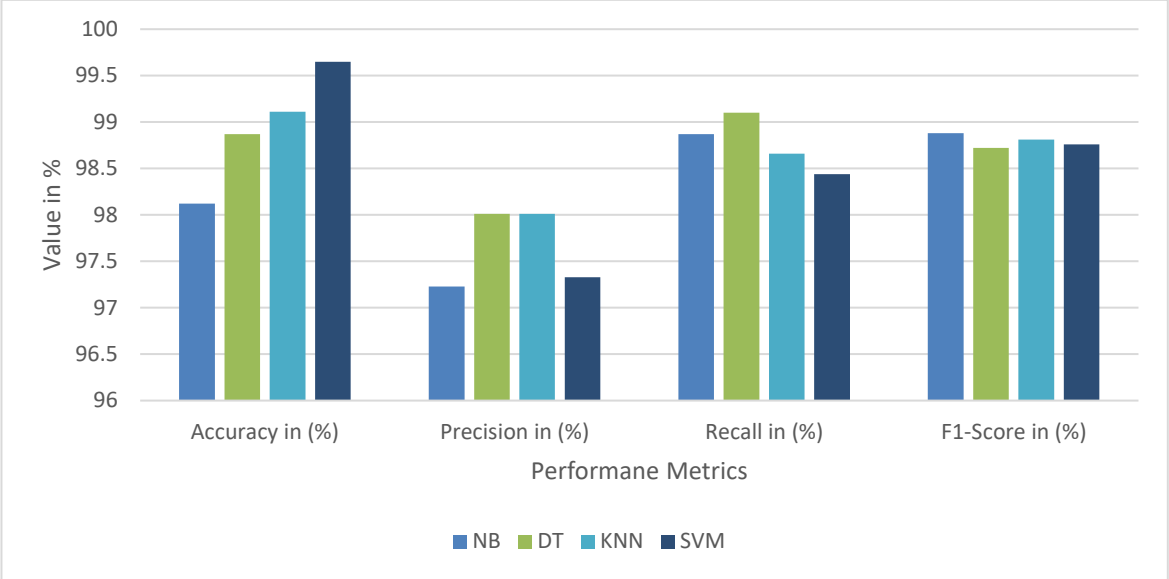

Fig 11: Performance metric comparison for different method

The study was effective in creating a framework for predicting diabetes using machine learning algorithms and assessing its efficacy according to several measures. The findings suggest that SVM, DT, and KNN classifiers, each of which exhibits strengths in distinct performance areas, are particularly good in identifying diabetes. These results offer insightful information for the creation and application of a strong diabetes prediction system.

VI. Conclusion

This study evaluated the potential of machine learning algorithms for diabetes prediction using medical data and disease effect variables. Age, BMI, diabetic pedigree function, skin thickness, blood pressure, blood glucose levels, and insulin levels were among the nine factors in the sample, which contained 768 records. Four classifiers—Naive Bayes (NB), Decision Tree (DT), k-Nearest Neighbors (KNN), and Support Vector Machine (SVM)—were evaluated for accuracy, precision, recall, and F1-score. With a fantastic accuracy of 99.65%, the results demonstrated that SVM was the best-performing classifier. When distinguishing between instances with and without diabetes, it showed exceptional prediction performance. SVM's high precision (97.33%) and recall (98.44%) were also quite high, underscoring its capability to precisely identify true positive cases and capture a substantial portion of actual positive cases in the dataset. Although they each exhibited various strengths in terms of diabetes prediction, NB, DT, and KNN all displayed excellent results. NB demonstrated resilience with an accuracy of 98.12% and excellent recall (98.87%) whereas DT and KNN reached competitive accuracy rates of 98.87% and 99.11%, respectively. This work underlines the value of machine learning in predicting diabetes and provides useful details for developing systems that can accurately diagnose diabetes. The results suggest that SVM, followed by NB, DT, and KNN, is a promising competitor for accurate and reliable diabetes prediction. These findings will assist researchers and medical professionals in selecting the most effective algorithms to aid in the early diagnosis and management of diabetes, which will ultimately improve patient outcomes and public health.

References:

[1] Battelino, T.; Alexander, C.M.; Amiel, S.A.; Arreaza-Rubin, G.; Beck, R.W.; Bergenstal, R.M.; Buckingham, B.A.; Carroll, J.; Ceriello, A.; Chow, E. Continuous glucose monitoring and metrics for clinical trials: An international consensus statement. Lancet Diabetes Endocrinol. 2022, 11, 42–57.

[2] B. Farajollahi, M. Mehmannavaz, H. Mehrjoo, F. Moghbeli, and M. J. Sayadi, "Diabetes diagnosis using machine learning," Frontiers in Health Informatics, vol. 10, no. 1, p. 65, 2021.

[3] K. Ogurtsova, J. D. da Rocha Fernandes, Y. Huang et al., "IDF Diabetes Atlas: global estimates for the prevalence of diabetes for 2015 and 2040," Diabetes Research and Clinical Practice, vol. 128, pp. 40–50, 2017.

[4] H. Qin, Z. Chen, Y. Zhang et al., "Triglyceride to high-density lipoprotein cholesterol ratio is associated with incident diabetes in men: a retrospective study of Chinese individuals," Journal of Diabetes Investigation, vol. 11, no. 1, pp. 192–198, 2020.

[5] J. Xie and Q. Wang, "Benchmarking machine learning algorithms on blood glucose prediction for type I diabetes in comparison with classical time-series models," IEEE Transactions on Biomedical Engineering, vol. 67, no. 11, pp. 3101–3124, 2020.

[6] J. Chaki, S. T. Ganesh, S. K Cidham, and S. AnandaTheertan, "Machine learning, and artificial intelligence-based Diabetes Mellitus detection and self-management: a systematic review," Journal of King Saud University - Computer and Information Sciences, 2020, ISSN 1319-1578.

[7] N. Bhatia and S. Kumar, "Prediction of severity of diabetes mellitus using fuzzy cognitive maps," Advances in Life Science and Technology, vol. 29, pp. 71–78, 2015.

[8] U. Haq, J. P. Li, A. Saboor et al., "Detection of breast cancer through clinical data using supervised and unsupervised feature selection techniques," IEEE Access, vol. 9, pp. 22090–22105, 2021.

[9] Pantalone, K.M.; Misra-Hebert, A.D.; Hobbs, T.M.; Wells, B.J.; Kong, S.X.; Chagin, K.; Dey, T.; Milinovich, A.; Weng, W.; Bauman, J.M. Effect of glycemic control on the Diabetes Complications Severity Index score and development of complications in people with newly diagnosed type 2 diabetes. J. Diabetes 2018, 10, 192–199.

[10] Pettus, J.H.; Zhou, F.L.; Shepherd, L.; Preblick, R.; Hunt, P.R.; Paranjape, S.; Miller, K.M.; Edelman, S.V. Incidences of severe hypoglycemia and diabetic ketoacidosis and prevalence of microvascular complications stratified by age and glycemic control in US adult patients with type 1 diabetes: A real-world study. Diabetes Care 2019, 42, 2220–2227.

[11] Basu, S.; Narayanaswamy, R. A prediction model for uncontrolled type 2 diabetes mellitus incorporating area-level social determinants of health. Med. Care 2019, 57, 592–600.

[12] Chatterjee, R.; Yeh, H.C.; Edelman, D.; Brancati, F. Potassium and risk of Type 2 diabetes. Expert Rev. Endocrinol. Metab. 2011, 6, 665–672.

[13] Jian, Y.; Pasquier, M.; Sagahyroon, A.; Aloul, F. A Machine Learning Approach to Predicting Diabetes Complications. Healthcare 2021, 9, 1712.

[14] Dinh, A.; Miertschin, S.; Young, A.; Mohanty, S.D. A data-driven approach to predicting diabetes and cardiovascular disease with machine learning. BMC Med. Inform. Decis. Mak. 2019, 19, 211.

[15] Zou, Q.; Qu, K.; Luo, Y.; Yin, D.; Ju, Y.; Tang, H. Predicting diabetes mellitus with machine learning techniques. Front. Genet. 2018, 9, 515.

[16] Yang, L.; Gabriel, N.; Hernandez, I.; Winterstein, A.G.; Guo, J. Using machine learning to identify diabetes patients with canagliflozin prescriptions at high-risk of lower extremity amputation using real-world data. Pharmacoepidemiol. Drug Saf. 2021, 30, 644–651.

[17] Del Parigi, A.; Tang, W.; Liu, D.; Lee, C.; Pratley, R. Machine learning to identify predictors of glycemic control in type 2 diabetes: An analysis of target HbA1c reduction using empagliflozin/linagliptin data. Pharm. Med. 2019, 33, 209–217.

[18] Seo, W.; Lee, Y.-B.; Lee, S.; Jin, S.-M.; Park, S.-M. A machine-learning approach to predict postprandial hypoglycemia. BMC Med. Inform. Decis. Mak. 2019, 19, 210.

[19] Hanson, R.L.; Imperatore, G.; Bennett, P.H.; Knowler, W.C. Components of the "metabolic syndrome" and incidence of type 2 diabetes. Diabetes 2002, 51, 3120–3127.

[20] Bhutto, A.R.; Abbasi, A.; Abro, A.H. Correlation of hemoglobinA1c with red cell width distribution and other parameters of red blood cells in type II diabetes mellitus. Cureus 2019, 11, e5533.

[21] All of Us Research Program Investigators. The "All of Us" research program. N. Engl. J. Med. 2019, 381, 668–676.

[22] Ramirez, A.H.; Sulieman, L.; Schlueter, D.J.; Halvorson, A.; Qian, J.; Ratsimbazafy, F.; Loperena, R.; Mayo, K.; Basford, M.; Deflaux, N. The All of Us Research Program: Data quality, utility, and diversity. Patterns 2022, 3, 100570.

[23] R Core Team. R: A Language and Environment for Statistical Computing; R Foundation for Statistical Computing: Vienna, Austria, 2022; Available online: https://www.R-project.org/ (accessed on 13 April 2023).

[24] Menardi, G.; Torelli, N. Training and assessing classification rules with imbalanced data. Data Min. Knowl. Discov. 2014, 28, 92–122.

[25] Fan, Y.; Long, E.; Cai, L.; Cao, Q.; Wu, X.; Tong, R. Machine learning approaches to predict risks of diabetic complications and poor glycemic control in nonadherent type 2 diabetes. Front. Pharmacol. 2021, 12, 1485.

[26] Motaib, I.; Aitlahbib, F.; Fadil, A.; Tlemcani, F.Z.R.; Elamari, S.; Laidi, S.; Chadli, A. Predicting poor glycemic control during Ramadan among non-fasting patients with diabetes using artificial intelligence based machine learning models. Diabetes Res. Clin. Pract. 2022, 190, 109982.

[27] Tao, X.; Jiang, M.; Liu, Y.; Hu, Q.; Zhu, B.; Hu, J.; Guo, W.; Wu, X.; Xiong, Y.; Shi, X. Predicting three-month fasting blood glucose and glycatedhemoglobin of patients with type 2 diabetes based on multiple machine learning algorithms. Research Square. 2022.

[28] Coregliano-Ring, L.; Goia-Nishide, K.; Rangel, É.B. Hypokalemia in Diabetes Mellitus Setting. Medicina 2022, 58, 431.

[29] Mr. Kaustubh Patil. (2013). Optimization of Classified Satellite Images using DWT and Fuzzy Logic. International Journal of New Practices in

Management and Engineering, 2(02), 08 - 12. Retrieved from http://ijnpme.org/index.php/IJNPME/article/view/15

[30] Patil, R. A. ., & Patil, P. D. . (2023). Skewed Evolving Data Streams Classification with Actionable Knowledge Extraction using Data Approximation and Adaptive Classification Framework. International Journal on Recent and Innovation Trends in Computing and Communication, 11(1), 38–52. https://doi.org/10.17762/ijritcc.v11i1.5985

[31] Anupong, W., Yi-Chia, L., Jagdish, M., Kumar, R., Selvam, P. D., Saravanakumar, R., & Dhabliya, D. (2022). Hybrid distributed energy sources providing climate security to the agriculture environment and enhancing the yield. Sustainable Energy Technologies and Assessments, 52 doi:10.1016/j.seta.2022.102142