Journal of Intelligent Systems and Internet of Things
Vol. 17, No. 01, PP. 106-117, 2025
DOI: https://doi.org/10.54216/JISIoT.170108

Internet of Medical Things Powered by Machine Learning for Real-Time Diabetes Prediction

Qusay Saihood[1,*], Inas H Kareem[2], Omar Ayad Ismael[3], Saad I. Mohammed[4], Ahmed NO Algburi[5]
[1]University of Al Maarif, Al Anbar, 31001, Iraq
[2]College of Dentistry, Ashur University, Baghdad, Iraq
[3]College of Islamic Sciences, University of Diyala, Diyala, Iraq
[4]Department of Anesthesiology Techniques, Al-Hadi University College, Baghdad-10011, Iraq
[5]Communication Technical Engineering, Al-Farahidi University, Baghdad, Iraq
Emails: qusayallahibi95@gmail.com; enas.hadi@au.edu.iq; omarismail@uodiyala.edu.iq; dr.saad65@gmail.com; ahmednoori@uoalfarahidi.edu.iq

Abstract

Diabetes is a common chronic illness that requires ongoing patient monitoring to diagnose the condition in a timely manner. With the significant advancements of the Internet of Medical Things (IoMT) sector in recent years, it is feasible now to monitor the patient's information continuously. There are many studies that used IoMT and machine learning (ML) techniques to diagnose diabetes but so far, the accuracy of the performance is still below the required level. Therefore, this study proposes a common framework for IoMT, cloud, and ML techniques to diagnose diabetes in real-time. IoMT devices continuously collect vital information of diabetic patients such as glucose and insulin levels. Then, this data is transmitted using various communication technologies to be stored in the cloud for diagnosis. Finally, to improve diagnostic accuracy, voting ensemble strategy-based method has been proposed that combines predictions from three base ML techniques (Support Vector Machine (SVM), Decision Tree (DT), and Random Forest (RF)). The proposed voting model achieved promising results in diagnosing diabetes with an accurate rate of up to 98.0%, outperforming the base classifiers in this and previous studies.

Received: January 30, 2025 Revised: February 27, 2025 Accepted: April 02, 2025

Keywords: Internet of Medical Things; Cloud; Machine Learning; Voting Ensemble Method; Diabetes; Diagnosis

1. Introduction

Diabetes is one of the most common chronic illnesses that develops when the pancreas is unable to produce enough insulin that is called type 1 diabetes (T1D). Alternatively, when the body's insulin is ineffectively used that is called type 2 diabetes (T2D) [1]. Insulin is a hormone that helps to keep blood sugar levels under control. Uncontrolled diabetes causes hyperglycemia, which damages many of the body systems, particularly the nerves and blood vessels over time [2]. According to reports published by the World Health Organization (WHO), diabetes is among the top ten causes of death worldwide. Studying indicates that there are about (422 million) people with diabetes in the world. Diabetes also causes an increase in mortality, causing (1.5 million) deaths annually. There has been a noticeable increase in the number of diabetics in recent years. Both the number of cases and the prevalence of diabetes have steadily increased over the past few decades [3]. Diabetes can lead to serious complications such as heart disease, high blood pressure, and stroke [4]. Regular glucose monitoring, on the other hand, is critical for reducing and avoiding diabetes complications [5]. The great development in IoT technologies provides essential solutions in monitoring patient health and diagnosing chronic diseases [6]. IoT is a network environment in which each connected node can communicate with other nodes to transmit critical data for important data to make precise and real-time decisions [7]. These characteristics enable IoT to provide effective solutions in critical scenarios such as medical applications. Therefore, using IoMT technology, diabetes disease can be monitored, and the medical parameters of the patient transmitted to health care centers or doctors in real-time [8].

IoMT consists of many types of sensors and medical devices connected to a healthcare provider's computer system over the Internet. Remote patient monitoring devices, wearables, sensors, and health trackers are examples of IoMT products [9]. These devices monitor the patient's health and continuously collect vital information such as glucose level, Insulin level, heart rate, pulse, blood pressure, body temperature, etc. [10]. IoMT generates a massive amount of data that can be carried from one end to another by using wireless communication technologies such as (ZigBee, WI-Fi, Bluetooth, NFC, RFID (Radio Frequency Identification), 4G, 5G, etc.) [11]. through these technologies, the data generated by IoMT is continuously carried to storage in the cloud. Cloud servers are gaining popularity as a means of storing and processing enormous amounts of data and giving access to real-time data from a distance. That is allowing healthcare monitors to monitor patient health directly and continuously [12]. Analyzing such huge and continuous data is a big challenge, as it is difficult for a person to monitor and analyze such a huge amount of data. Artificial intelligence (AI) techniques, especially ML techniques, provide the best solutions for analyzing this data to diagnose and identify diseases in real-time [2].

One of the practical applications of ML in healthcare is the prediction and diagnosis of diseases. Therefore, ML techniques provide the most appropriate solution to deal with the huge data generated by IoMT devices [13]. In recent years, there are many studies addressing the diagnosis of diabetes using IoMT-ML but so far, the accuracy of the performance is still under the required level [1, 13]. To build a remote monitoring system for diagnosing diabetes, some challenges exist. One of the most important of these challenges is choosing an appropriate communication technology for real-time data transmission [14]. In addition, gathering and storing data from IoMT devices is a big challenge [13]. While the biggest challenge is to use an effective ML technique that can deal well with this continuous and large data to diagnose diabetes in real-time and with providing high classification accuracy [12].

Therefore, this study proposes a common framework for the IoMT, cloud, and ML techniques for real-time diabetes diagnosis. IoMT devices collect vital information from the patient and transmit this data using communication technologies to the cloud, where this data is stored in the cloud and analyzed using ML techniques. To diagnose diabetes with high accuracy and efficiency, data preprocessing was effectively implemented by filling in missing values, dealing with outliers, and balancing data classes. In addition, a voting strategy was proposed to build an ensemble model using three base techniques (DT, RF, and SVM). The proposed model in this study was trained and tested using a Pima Indians Diabetes Database [15], and performance was evaluated using a set of performance measures such as (Accuracy, Precision, Sensitivity, and F1-score). In addition, the performance of the proposed voting model was compared with five base ML techniques (SVM, DT, RF, K-Nearest Neighbors (KNN), and Logistic Regression (LR)) used in this study. The main contributions of this study are:

- Simulating a complete framework based on the IoMT, cloud, and ML techniques to diagnose diabetes in real-time.
- Performing efficient data preprocessing to prepare the data well.
- Exploring the performance of five base ML classifiers for diabetes diagnosis.
- Fine-tuning the hyperparameters of base classifiers using Random Search (RS) method.
- A voting strategy was proposed to build an ensemble model using three base ML techniques (SVM, DT, and RF).

After the introduction, related works are discussed in Section 2 especially works dealing with IoMT-ML for predicting diabetes. Section 3 explains the stages of the proposed framework (IoMT devices, communication techniques, cloud, and ML). Section 4 combines and shows the results. Section 5 contains the conclusion and future work.

2. Related Works

Diabetes disease is among the most common diseases in the world and leads to serious complications that affect human life if it is not diagnosed early. There is a lot of literature that uses ML techniques to diagnose diabetes, but there are still limited studies that combine the advantages of the IoMT and ML to diagnose diabetes in real-time.

In [16], Chatrati, Saiteja Prasad, et al., introduced a system that works on monitoring both the patient's glucose level and blood pressure level at home and alerts the health care provider if an abnormality is detected. This system receives both input glucose and blood pressure readings and uses ML techniques for early diagnosis for each diabetes and blood pressure. This system was trained and tested on Pima Indians Diabetes Database where the SVM technique has achieved the highest accuracy with 75%.

In [17] Alfian, Ganjar, et al., suggested diabetes monitoring system monitor patient health using a Bluetooth Low Energy (BLE)-based sensor, real-time data processing, and ML techniques to help diabetics better manage their chronic condition. This system uses sensors to collect vital sign data from users, which is then transmitted via BLE for storage. MongoDB was used to store patients' data and Apache Kafka was used as a streaming platform. Finally, the Multilayer perceptron technique (MLP) is used to predict the presence or likelihood of developing diabetes in the future, whereas long short-term memory (LSTM) is used to predict future blood sugar levels.

In [18] Rghioui, Amine, et al., suggested the use of 5G in diabetes monitoring systems. A set of sensors and wearable devices, a smartphone application, and a server with a database make up the system architecture. The proposed system is prepared to gather data on diabetes patients' blood glucose levels, physical activity, and temperature using a combination of wearable devices and sensors. Wi-Fi is used to connect various sensors to the smartphone, where data is transmitted from the smartphone via a 5G connection to a database. Next, ML techniques are used to predict diabetes. Among the several ML techniques used in this study, the Sequential Minimal Optimization (SMO) technique gave a higher classification accuracy than the other techniques used.

In [19], Parampreet, et al., suggested using a framework based on the use of IoT devices, Cloud, and ML techniques for predicting diabetes. Smart wearables devices continuously monitor and collect blood glucose data that is transmitted to store in the cloud where the ensemble model is used to diagnose diabetes. When the model was tested on "Pima Indians Diabetes," the ensemble model combining both DT technique and neural network technique got the best accuracy of 94.5%.

[20] Pavleen Kaur, et al., proposed a framework that allows remote E-Health monitoring based on IoT equipment and cloud computing assistance while embedding ML features in the cloud. Several ML techniques, namely Linear-SVM, KNN, DT, MLP, and RF were applied on datasets of various diseases inclusive (Diabetes, Breast Cancer, Heart Disease, surgery data, thyroid, dermatology, and liver disorder). The suggested RF technique has achieved promising results in the prediction of various diseases.

In [21] Khanam, Jobeda Jamal, and Simon Y. Foo., applied 7 ML techniques to diagnose diabetes by using the Pima Indian Diabetes dataset. NN, with two hidden layers, achieved the highest accuracy of 88.6%. Also, in [22] Sisodia, Deepti, and Dilip Singh Sisodia used three different techniques (SVM, NB, and DT) for early diagnosis of diabetes where the NB technique achieved the highest result with an accuracy of 76.30%. While in [23] ANN technique provided the best accuracy of 75.7% superior to both RF and K-mean techniques. In [24] Tigga, Neha Prerna, and Shruti Garg used the LR algorithm to improve diabetes prediction as the system achieved good accuracy results.

In [25] Larabi-Marie-Sainte, Souad, et al., experimented with the use of new and rare methods of ML techniques to predict diabetes. However, with the lack of sufficient knowledge of the disease and deep interpretation of the data, these techniques did not achieve better results than previous studies.

In [26] Ashiquzzaman, Akm, et al., presented a system for predicting diabetes using the dropout method to reduce the problem of overfitting that most ML techniques suffer from it. The proposed deep neural network achieved superior performance with an accuracy of 88.41%. Table 1 shows a complete summary of studies related to this study.

In [27] Saihood, Qusay, and Emrullah Sonuç proposed a comprehensive framework to improve the accuracy of diabetes diagnosis based on three stages: In the first stage, missing values were filled, data was normalized, outliers were handled, and important features were selected. In the second stage, the hyperparameters of individual models used were tuned. In the last stage, these individual classifiers were combined using the ensemble method.

Table 1: Summary of studies mentioned in related works.

Info. / No.	Authors	Year	IoMT	Communication Techniques	Cloud	Data	ML Techniques
1.	Chatrati, Saiteja Prasad, et al. [16]	2020	Yes	No	NO	Pima Indians Diabetes Database	SVM, KNN, DT, LR, and Linear discriminant analysis (LDA)
2.	Alfian, Ganjar, et al. [[17]	2018	Yes	Yes	Yes	Pima Indians Diabetes Database	RF, NB, SVM, LR, and MLP

3.	Rghioui, Amine, et al. [18]	2020	Yes	Yes	No	Collected by sensers	NB, J48, SMO, ZeroR, OneR, Simple Logistic, and RF
4.	Kaur, Parampreet, et al. [19]	2018	Yes	Yes	Yes	Pima Indians Diabetes Database	Ensemble model, Neural Network,RF, NB, SVM, and DT
5.	Pavleen Kaur, et al. [20]	2019	Yes	Yes	Yes	Pima Indians Diabetes Database	kNN, Linear-SVM, DT, MLP, and RF
6.	Khanam, Jobeda Jamal, and Simon Y. Foo. [21]	2021	NO	NO	NO	Pima Indians Diabetes Database	DT, RF, NB, LR, ANN, SVM, and Adaboost (AB)
7.	Sisodia, Deepti, and Dilip Singh Sisodia. [22]	2018	NO	NO	NO	Pima Indians Diabetes Database	NB, SVM, and DT
8.	Alam, Talha Mahboob, et al. [23]	2019	NO	NO	NO	Pima Indians Diabetes Database	ANN, RF, and K-mean
9.	Tigga, Neha Prerna, and Shruti Garg.[24]	2021	NO	NO	NO	Pima Indians Diabetes Database	LR
10.	Larabi-Marie-Sainte, Souad, et al. [25]	2019	NO	NO	NO	Pima Indians Diabetes Database	REPTree, KStar, oneR, PART, SMO, and BayesNet
11.	Ashiquzzaman, Akm, et al. [26]	2018	NO	NO	NO	Pima Indians Diabetes Database	DNN
12.	Saihood, Qusay, and Emrullah Sonuç [27]	2023	NO	NO	NO	Pima Indians Diabetes Database	Ensemble Methods (Bagging, Boosting, and Stacking)

3. Proposed Methodology

Early prediction and diagnosis of diabetes is especially important to avoid serious complications that threaten the patient's life. Diabetes can be diagnosed by monitoring some vital information of a person such as (blood glucose level, insulin level, and blood pressure level etc.). Therefore, this study proposes a combined framework IoMT devices, Cloud, and ML for real-time diabetes diagnosis. IoMT devices consist of a group of sensors and wearable devices that collect vital information about diabetic patients. The data generated by the IoMT devices is carried to storage in the cloud using communication technologies. Then, diabetes is diagnosed using the voting ensemble method. The proposed system architecture framework consists of two stages; the first stage collects data from the patient, and the second stage diagnoses diabetes. Figure 1 shows the proposed framework for this study.

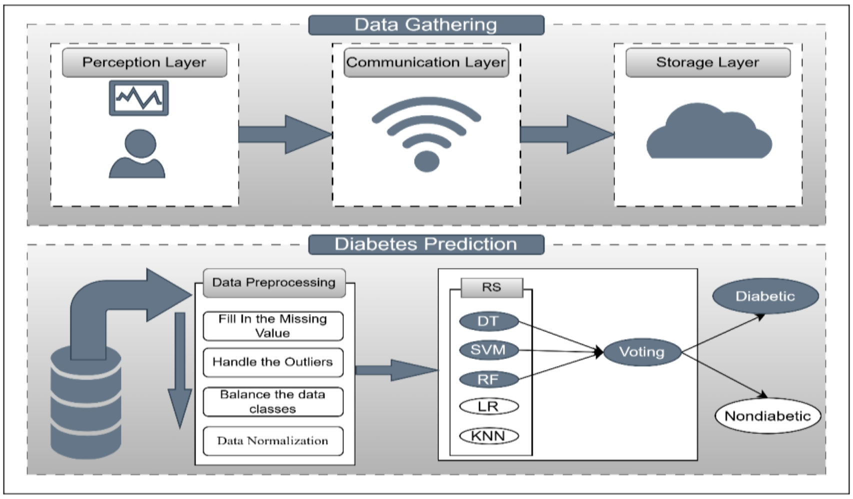

Figure 1. Proposed framework for diabetes diagnosis.

The following steps explain the workflow of the proposed framework:

3.1. Data Gathering

The proposed model framework for data acquisition from diabetic patients consists of three layers (perception layer, communication layer, and storage layer).

3.3.1. Perception layer

This layer is the closest to the patient and consists of IoMT technologies that include both sensors and wearable devices. This layer is responsible for collecting vital information from diabetic patients. The sensors and wearable devices collect health parameters of diabetic patients such as glucose level, insulin level, blood pressure level, body temperature, etc. [16, 18].

3.1.2. Communication Layer

Vital information collected from different IoMT devices is transmitted to the cloud through various communication technologies. The choice of the type of communication technologies depends on the type of sensor, which includes ZigBee, WI-Fi, Bluetooth, NFC, RFID, 4G, 5G, etc. [11].

3.1.3. Storage Layer

The data generated by the IoMT is stored in the cloud for further analysis. Cloud storage provides a convenient solution to accommodate this huge amount of data generated continuously. The cloud environment also provides a secure environment for remote data storage and access, which provides the opportunity to effectively monitor the vital information of diabetic patients [19].

3.2. Diabetes Prediction

After obtaining data of diabetic patients using IoMT devices and storing it in the cloud. The data is analysed using ML techniques for the early diagnosis of diabetes. The voting strategy is used for a more accurate diagnosis process. In addition, the performance of five base ML techniques is individually tested for performance comparison. Before implementing the diabetes diagnosis process, data preprocessing is performed, which consists of filling in missing values, detecting and handling outliers, handling the imbalance issue, and normalizing numerical values.

The process of diagnosing diabetes using ML techniques goes through the following stages:

3.2.1. Data Preprocessing

First, as indicated in Table 2, the dataset utilized in this work is the Pima Indians Diabetes Dataset [15], which includes (768) samples and (9) characteristics.

Table 2: Description of Pima Indians Diabetes Dataset.

Info. / No.	Features	Types
1.	Target (Diabetic and Non-Diabetic)	binary
2.	Diabetes Pedigree Function	Numeric
3.	BMI	Numeric
4.	Insulin Level	Numeric
5.	Skin Thickness	Numeric
6.	Blood Pressure	Numeric
7.	Glucose Level	Numeric
8.	Pregnancies	binary
9.	Age	Numeric

Data preprocessing is a crucial step in ML approaches since raw data frequently contains duplicate, irrelevant, and inconsistent information. The caliber of the data fed into ML techniques determines much of their effectiveness [28]. Thus, proper preprocessing is essential for creating a model with acceptable accuracy and high performance. The following stages provide a summary of the preprocessing of the diabetes data used in this study:

3.2.1.1. Fill In the Missing Value

The diabetes dataset used in this study contains some missing values, which are preferably treated to improve diagnostic accuracy. In this work, the features with missing values were filled by calculating the meaning of the feature column containing the missing value according to the target class it belongs to, as shown in Table 3.

Table 3: Mean Values of Features with Missing Values According to The Target Class.

Features / Class Target	Insulin	Glucose	Blood Pressure	Skin Thickness	BMI
0 (non-Diabetic)	102.5	107.0	70.0	27.0	30.1
1 (Diabetic)	169.5	140.0	74.5	32.0	34.3

3.2.1.2. Handle the Outliers

Outliers have a significant negative impact on the diagnostic process because they differ significantly from the rest of the data. Therefore, detecting and handling outliers is one of the most important steps in data preprocessing that can greatly improve diagnostic accuracy. In this work, the interquartile range (IQR) method was applied to detect and handle the outliers within Pima Indians Diabetes Dataset. It is one of the commonly used statistical methods for detecting outliers within a dataset. IQR used to measure the distance between the first quartile (Q1) and the third quartile (Q3), to identify outliers. It is a good metric of prevalence that is used to determine the range that represents the middle half of the data [27].

Outliers are detected and handled by the IQR method through the following steps:

First, Q1 is determined.

Second, Q3 is determined.

Third, IQR is Calculated by:

$$IQR = (Q3 - Q1) \qquad (1)$$

Fourth, the lower bounds of the normal data range are determined by:

$$\text{lower whisker} = (\text{Q1} - (1.5 * \text{IQR})) \qquad (2)$$

Fifth, the upper limits of the normal data range are determined by:

$$\text{Upper whisker} = (\text{Q3} + (1.5 * \text{IQR})) \qquad (3)$$

Sixth, the values below (lower longitudinal) and above (upper longitudinal) are replaced by the median value.

3.2.1.3. Balance the data classes

The Synthetic Minority Oversampling Technique (SMOTE) was used to balance the classes, where the data consisted of two classes with diabetes, which numbered 500, and those who did not have diabetes numbered 268. Table 4 shows the number of samples for each class after using over-sampling.

Table 4: The number of per class before and after oversampling

Sampling / Classes of Diabetes	Before Over Sampling	After Over Sampling
Class 0 has diabetes	500	500
Class 1 has no diabetes	268	500

3.2.1.4. Data normalization

To normalize the values of numerical column to a uniform scale, the Min-Max normalization method was applied to convert the numerical column values in the Pima Indians Diabetes Dataset to a common scale between 0 and 1 without distorting the value ranges, using the following equation [29]:

$$Y = \frac{X - Min(X)}{Max(X) - Min(X)} \qquad (4)$$

Where Y represents the normalized value, X represents the original value.

3.2.2. Classification Techniques

Brief description of the ML techniques used:

3.2.2.1. Decision Trees (DT) Technique

A non-parametric supervised learning approach is well known for its effectiveness in classification applications. This flexible method can be used for both regression and classification issues. With the highest node acting as the root, the decision tree's topology resembles a hierarchical tree. A feature test is associated with each internal node, branches represent test results, and leaf nodes [30] represent class labels. When diagnosing diabetes, a Classification and Regression Tree (CART) is utilized. To pick features for classification issues, CART generates binary trees and uses the Gini index function.

$$\text{Gini Index} = 1 - \int_{J=1}^{c} Pi^2 \qquad (5)$$

where Pi is the probability that a tuple in dataset belongs to class c.

3.2.2.2. Support Vector Machine (SVM) Technique

It is a powerful supervised Model that may be used for both regression problems; however, it is mostly used for classification tasks. Predicting a hyperplane or linear border that divides two classes in a dataset is the basic goal of support vector machines (SVM) classification. Computing the distances between data points that belong to different classes is a step in the SVM process. These are called support vectors, and they are the locations closest to each class's hyperplane. By choosing the one with the biggest margin between the hyperplane and the support vectors, the SVM algorithm aims to determine the ideal hyperplane [30].

3.2.2.3. Logistic Regression (LR) Technique

A logistic curve is used in LR, a supervised learning classification model, to model data and predict the probability of an event occurring using a statistical framework. Establishing the odds connected to a certain occurrence is a step in the logistic regression method. This function produces an S-shaped curve by projecting discrete values, usually binary outcomes (0/1 or yes/no), based on a particular collection of independent factors [58]. Most notably, LR is a multivariable technique that seeks to determine a functional connection between one or more outcome (dependent) variables and two or more predictor (independent) variables. In the current investigation, Binary LR was utilized to forecast the presence of only two categories of results: "Diabetic" or "Not Diabetic" [24].

3.2.2.4. Random Forest (RF) Technique

It is a supervised ML method that may be used for problems involving regression and classification. Differentiating itself from the traditional single decision tree model, RF builds many decision trees. This is due to its ensemble learning approach. This ensemble method gives the classifier more resilience and effectiveness. RF creates a large number of Classification and Regression Trees (CART), each of which is trained on a randomly selected portion of the original dataset. A majority vote from the group determines the categorization choice once the collective decisions of all the trees in the forest are combined [31].

3.2.2.5. K-Nearest Neighbors (KNN) Technique

It is a simple ML method that may be used for regression and classification. The basic tenet of KNN is that comparable items are equal, which implies that things that are like one other are often closer to each other. KNN determines the distance between each item in the training data and the item under consideration during the classification process. The ideal value for K, which represents the number of elements to be classified as closest neighbors, is then determined. Usually, a range of values for K is investigated to identify the best choice. A majority vote among the nearby data points determines the categorization result. The metric used to measure the spatial separation between two locations is called the Euclidean distance [30, 31].

$$\text{Euclidean distance (I, J)} = \sqrt[2]{(XI1 - XJ1)^2 + \cdots + (XIN - XJN)^2} \quad (6)$$

3.2.3. Random Search (RS) Method

RS is one of the commonly used techniques to automatically tune hyperparameters for ML techniques. Unlike grid search, which tries all possible possibilities for hyperparameters, which is computationally expensive, RS selects hyperparameters randomly. Random search has proven to be more efficient and superior to the Grid search technique due to its ability to explore the hyperparameter space more widely and identify the optimal hyperparameters with less computational cost compared to grid search [32].

3.2.4. Voting Method

Voting is an ensemble method in ML field that is used to improve the performance of individual classifiers by combining the predictions of more than one classifier. The voting method relies on voting to determine the final prediction result, which can be done through voting strategies: majority voting, weighted voting, or soft voting [33]. The weighted voting strategy was used to combine three classifiers (SVM, DT, and RF) to predict whether a patient has diabetes or not.

3.3. Classification Performance

The performance of ML techniques was evaluated using four commonly used performance metrics (accuracy, precision, sensitivity, and F1 score), in addition to calculating the confusion matrix for the final voting ensemble model. The positive samples were those that showed diabetes, and they were represented by the number '1,' whereas the negative samples were those that showed health, and they were represented by the number '0'[27].

- True Positives (TP): Diabetic patients who have already been diagnosed do have diabetes.
- True Negatives (TN): Non-diabetic patients who have already been diagnosed do not have diabetes.
- False Positives (FP): Non-diabetic patients who have been diagnosed do have diabetes.
- False Negatives (FN): Diabetic patients who have been diagnosed do not have diabetes.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{TP + TN + FP + FN} \quad (7)$$

$$\text{Precision} = \frac{\text{TP}}{TP + FP} \quad (8)$$

$$\text{Sensitivity} = \frac{\text{TP}}{TP + FN} \quad (9)$$

$$\text{Specificity} = \frac{\text{TN}}{TN + FP} \quad (10)$$

$$\text{F1 Score} = \frac{2 * \text{Precision} * \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \quad (11)$$

4. Results and Discussion

In this study, five base ML techniques and the proposed voting (SVM, DT, and RF) model were trained and tested for diabetes diagnosis. The dataset used in this study was separated into 80% for training and 20 for testing. Accuracy, precision, sensitivity, and F1 score were used as performance measures to evaluate the performance of the used ML models. As shown in Table 5, the voting (SVM, DT, and RF) model achieved the best result with an accuracy of 98.0%, superior to each of the individual classifiers in this study. Among the individual techniques used in this work, the RF technique achieved the highest result with an accuracy of 97.5%. The DT technique also achieved a good result in predicting diabetes with an accuracy of 95.5%, followed by all SVM, KNN and LR with 93.0%, 92.0% and 86.5% accuracy, respectively.

Table 5: Performance evaluation of base ML techniques and a stacking classifier.

Evaluation Metrics / ML Techniques	Accuracy	Precision	Sensitivity	F1-Score
DT	96.5%	94.9%	97.9%	96.3%
SVM	93.0%	92.63%	92.6%	92.6%
LR	86.5%	82.7%	90.5%	86.4%
RF	97.5%	95.0%	100.0%	97.4%
KNN	92.0%	89.1%	94.7%	91.8%
Voting Model	98.0%	96.9%	98.9%	97.9%

The proposed voting model, which combines basically three classifiers (DT, RF, And SVM), has achieved promising results during the comparison with the previous studies [16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27], as shown in Table 6.

Table 6: Comparing the performance of the proposed model with previous studies.

Info. / No.	Study	ML Techniques	Accuracy
1.	[16]	SVM	75%
2.	[17]	MLP	77.1%
3.	[19]	Ensemble model	94.5%
4.	[20]	RF	81.16%
5.	[21]	ANN	88.6%.
6.	[22]	NB	76.30%
7.	[23]	ANN	75.7%
8.	[24]	LR	75.32%
9.	[25]	REPTree	74.48%
10.	[26]	DNN	88.41
11.	[27]	Stacking Method	97.50%
12.	Our Study	Voting Model	98.0%

In most cases, the voting model performs better than the base ML classifiers because it combines the predictions of a set of weak classifiers to build a strong classifier and improve performance. In this study, the predictions of the top 3 classifiers (SVM, DT, and RF) out of the five base classifiers (LR, KNN, SVM, DT, and RF) used in this study were combined using a weighted voting strategy.

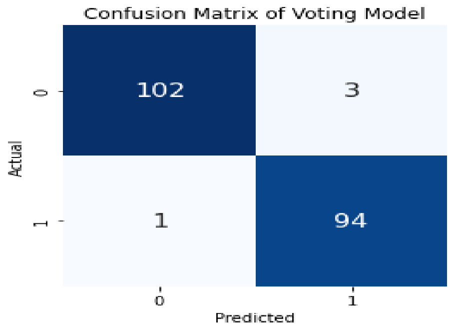

Figure 2. Confusion matrix of the proposed voting model.

The proposed voting model demonstrated exceptional performance in diagnosing all cases, correctly classifying 102 positive cases and 94 negative cases, indicating high classification accuracy. However, the model misclassified only 3 negative cases as positive and 1 positive case as negative, indicating a low error rate as shown in figure 2.

In addition, to achieve an effective diabetes diagnosis, data preprocessing steps must be well implemented, because the performance of the ML technique largely depends on the quality of the data presented. In addition, adjusting the hyperparameters for each base ML technique is important to achieve better performance. Therefore, in this study, after performing data preprocessing and adjusting the hyperparameters of each ML technique, the ML technique made a strong improvement in the prediction outcome, superior to the results of the previous study.

The superiority of the proposed voting model in this study is not limited to the accuracy of performance compared to the ensemble methods mentioned in previous works [19, 27], but it goes beyond that to build a ML model with less computational cost and less complexity. The proposed voting model combines the predictions of three fine-tuned classifiers with low computational cost, RS method was used to tune the models to restrict the depth of the calculations and reduce computational consumption.

5. Conclusion

This study presented a common framework based on the IoMT, cloud, and ML techniques for effective diabetes diagnosis. IoMT devices monitor diabetes patients and collect vital information continuously. This data is transmitted using various communication technologies to be stored in the cloud. After that, ML techniques are used to diagnose the disease in real-time. For a more accurate diagnosis, a voting ensemble method, which combines three base classifiers (SWM, DT, and RF), is conducted. The proposed voting model achieved promising results with an accuracy of 98.0% for diabetes diagnosis, superior to the base techniques (DT, SVM, LR, RF, and KNN) used in this study as well as the results of previous studies. The proposed system has proven highly efficient in the effective diagnosis of diabetes, which aids physicians in making the right decision. Determining the appropriate ML technique for the diagnostic process is very important for the efficiency of the diagnosis. In addition, preprocessing the data as filling in missing values and dealing with outliers is very important to achieve high diagnostic accuracy. To obtain a more accurate diagnosis in future work, it is recommended to build ensemble models that combine DL techniques. In addition, using the real data generated by IoMT devices will give huge data, which will help in improving the accuracy of the diagnosis.

Funding: “This research received no external funding”

Conflicts of Interest: “The authors declare no conflict of interest.”

References

[1] M. Eghbali-Zarch and S. Masoud, “Application of machine learning in affordable and accessible insulin management for type 1 and 2 diabetes: A comprehensive review,” *Artif Intell Med*, vol. 151, p. 102868, May 2024, doi: 10.1016/J.ARTMED.2024.102868.

[2] K. R. Tan et al., “Evaluation of Machine Learning Methods Developed for Prediction of Diabetes Complications: A Systematic Review,” *J. Diabetes Sci. Technol.*, vol. 17, no. 2, pp. 474–489, Mar. 2023, doi: 10.1177/19322968211056917/SUPPL_FILE/SJ-DOCX-1-DST-10.1177_19322968211056917.DOCX.

[3] “Diabetes,” World Health Organization. Accessed: Jan. 18, 2025. [Online]. Available: https://www.who.int/health-topics/diabetes#tab=tab_1.

[4] R. Raghow et al., "Diabetes mellitus and stroke: A clinical update," *World J. Diabetes*, vol. 8, no. 6, p. 235, 2017, doi: 10.4239/WJD.V8.I6.235.

[5] American Diabetes Association, "Standards of Medical Care in Diabetes—2022 Abridged for Primary Care Providers," *Clinical Diabetes*, vol. 40, no. 1, pp. 10–38, Jan. 2022, doi: 10.2337/CD22-AS01.

[6] N. S. Sworna, A. K. M. M. Islam, S. Shatabda, and S. Islam, "Towards development of IoT-ML driven healthcare systems: A survey," *Journal of Network and Computer Applications*, vol. 196, p. 103244, Dec. 2021, doi: 10.1016/J.JNCA.2021.103244.

[7] X. Liu et al., "Smart Contact Lenses for Healthcare Monitoring and Therapy," *ACS Nano*, vol. 18, no. 9, pp. 6817–6844, Mar. 2024, doi: 10.1021/ACSNANO.3C12072/ASSET/IMAGES/MEDIUM/NN3C12072_0017.GIF.

[8] V. Kumar et al., "Development of Medical IOT System for the Prediction of Heart Disease," in *2024 4th International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, 2024, pp. 951–956, doi: 10.1109/ICACITE60783.2024.10617362.

[9] Y. A. A. S. Aldeen, M. Salleh, and M. A. Razzaque, "State of the art survey on security issue in cloud computing architectures, approaches and methods," *Journal of Theoretical and Applied Information Technology*, vol. 75, no. 1, pp. 53-61, 2015.

[10] B. A. Muthu et al., "IOT based wearable sensor for diseases prediction and symptom analysis in healthcare sector," *Peer-to-Peer Networking and Applications*, vol. 13, no. 6, pp. 2123–2134, Nov. 2020, doi: 10.1007/S12083-019-00823-2.

[11] M. L. Umashankar et al., "A Survey on IoT Protocol in Real-Time Applications and Its Architectures," in *Lecture Notes in Electrical Engineering*, vol. 947, pp. 119–130, 2023, doi: 10.1007/978-981-19-5936-3_12.

[12] T. C. T. Chen, H. C. Wu, and M. C. Chiu, "A deep neural network with modified random forest incremental interpretation approach for diagnosing diabetes in smart healthcare," *Appl. Soft Comput.*, vol. 152, p. 111183, Feb. 2024, doi: 10.1016/J.ASOC.2023.111183.

[13] K. S. L. Kazi and M. A. Mahant, "Machine Learning-Driven Internet of Things (MLIoT)-Based Healthcare Monitoring System," IGI Global, 2023, doi: 10.4018/979-8-3693-9641-4.ch007.

[14] A. Yadav et al., "Fog enabling technologies in healthcare: A review," in *Artificial Intelligence, Blockchain, Computing and Security - Proceedings of the International Conference on Artificial Intelligence, Blockchain, Computing and Security (ICABCS)*, vol. 1, pp. 531–538, Dec. 2024, doi: 10.1201/9781003393580-81.

[15] "Pima Indians Diabetes Database," Kaggle. Accessed: Jan. 18, 2025. [Online]. Available: https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database.

[16] S. P. Chatrati et al., "Smart home health monitoring system for predicting type 2 diabetes and hypertension," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 3, pp. 862–870, Mar. 2022, doi: 10.1016/J.JKSUCI.2020.01.010.

[17] G. Alfian et al., "A Personalized Healthcare Monitoring System for Diabetic Patients by Utilizing BLE-Based Sensors and Real-Time Data Processing," *Sensors*, vol. 18, no. 7, p. 2183, Jul. 2018, doi: 10.3390/S18072183.

[18] A. Rghioui et al., "A Smart Architecture for Diabetic Patient Monitoring Using Machine Learning Algorithms," *Healthcare*, vol. 8, no. 3, p. 348, Sep. 2020, doi: 10.3390/HEALTHCARE8030348.

[19] P. Kaur et al., "CI-DPF: A Cloud IoT based Framework for Diabetes Prediction," in *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, pp. 654–660, Jul. 2018, doi: 10.1109/IEMCON.2018.8614775.

[20] P. Kaur, R. Kumar, and M. Kumar, "A healthcare monitoring system using random forest and internet of things (IoT)," *Multimedia Tools and Applications*, vol. 78, no. 14, pp. 19905–19916, Jul. 2019, doi: 10.1007/S11042-019-7327-8.

[21] J. J. Khanam and S. Y. Foo, "A comparison of machine learning algorithms for diabetes prediction," *ICT Express*, vol. 7, no. 4, pp. 432–439, Dec. 2021, doi: 10.1016/J.ICTE.2021.02.004.

[22] D. Sisodia and D. S. Sisodia, "Prediction of Diabetes using Classification Algorithms," *Procedia Comput. Sci.*, vol. 132, pp. 1578–1585, Jan. 2018, doi: 10.1016/J.PROCS.2018.05.122.

[23] T. Mahboob Alam et al., “A model for early prediction of diabetes,” *Inform. Med. Unlocked*, vol. 16, p. 100204, Jan. 2019, doi: 10.1016/J.IMU.2019.100204.

[24] N. P. Tigga and S. Garg, “Predicting Type 2 Diabetes Using Logistic Regression,” in *Lecture Notes in Electrical Engineering*, vol. 673, pp. 491–500, 2021, doi: 10.1007/978-981-15-5546-6_42.

[25] S. Larabi-Marie-Sainte et al., “Current Techniques for Diabetes Prediction: Review and Case Study,” *Applied Sciences*, vol. 9, no. 21, p. 4604, Oct. 2019, doi: 10.3390/APP9214604.

[26] A. Ashiquzzaman et al., “Reduction of Overfitting in Diabetes Prediction Using Deep Learning Neural Network,” in *Lecture Notes in Electrical Engineering*, vol. 449, pp. 35–43, 2018, doi: 10.1007/978-981-10-6451-7_5.

[27] Q. Saihood and E. Sonuç, “A practical framework for early detection of diabetes using ensemble machine learning models,” *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 31, no. 4, pp. 722–738, Jul. 2023, doi: 10.55730/1300-0632.4013.

[28] B. L. Ortiz et al., “Data Preprocessing Techniques for Artificial Intelligence (AI)/Machine Learning (ML)-Readiness: Systematic Review of Wearable Sensor Data in Cancer Care (Preprint),” *JMIR Mhealth Uhealth*, vol. 12, no. 1, p. e59587, Sep. 2024, doi: 10.2196/59587.

[29] L. Rahmad Ramadhan and Y. Anne Mudya, “A Comparative Study of Z-Score and Min-Max Normalization for Rainfall Classification in Pekanbaru,” May 2024. Accessed: Jan. 18, 2025. [Online]. Available: http://ipublishing.intimal.edu.my/jods.html.

[30] Q. Saihood and E. Sonuç, “The Efficiency of Classification Techniques in Predicting Anemia Among Children: A Comparative Study,” in *Communications in Computer and Information Science*, vol. 1548, pp. 167–181, 2022, doi: 10.1007/978-3-030-97255-4_12.

[31] N. Jalel Ibrahim, “Software Defect Prediction System Based on Well-Tuned Random Forest Technique,” doi: 10.24412/2413-2527-2023-335-51-59.

[32] B. Bischl et al., “Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges,” *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 13, no. 2, p. e1484, Mar. 2023, doi: 10.1002/WIDM.1484.

[33] S. Chatterjee and Y. C. Byun, “Voting Ensemble Approach for Enhancing Alzheimer’s Disease Classification,” *Sensors*, vol. 22, no. 19, p. 7661, Oct. 2022, doi: 10.3390/S22197661.

[34] D. Sisodia and D. S. Sisodia, “Prediction of Diabetes using Classification Algorithms,” *Procedia Comput. Sci.*, vol. 132, pp. 1578–1585, Jan. 2018, doi: 10.1016/J.PROCS.2018.05.122.