



Fog-cloud architecture-driven Internet of Medical Things framework for healthcare monitoring

Emre Yıldırım¹ · Murtaza Cicioğlu² · Ali Çalhan³

Received: 24 May 2022 / Accepted: 6 January 2023 / Published online: 21 January 2023
© International Federation for Medical and Biological Engineering 2023

Abstract

The new coronavirus disease (COVID-19) has increased the need for new technologies such as the Internet of Medical Things (IoMT), Wireless Body Area Networks (WBANs), and cloud computing in the health sector as well as in many areas. These technologies have also made it possible for billions of devices to connect to the internet and communicate with each other. In this study, an Internet of Medical Things (IoMT) framework consisting of Wireless Body Area Networks (WBANs) has been designed and the health big data from WBANs have been analyzed using fog and cloud computing technologies. Fog computing is used for fast and easy analysis, and cloud computing is used for time-consuming and complex analysis. The proposed IoMT framework is presented with a diabetes prediction scenario. The diabetes prediction process is carried out on fog with fuzzy logic decision-making and is achieved on cloud with support vector machine (SVM), random forest (RF), and artificial neural network (ANN) as machine learning algorithms. The dataset produced in WBANs is used for big data analysis in the scenario for both fuzzy logic and machine learning algorithm. The fuzzy logic gives 64% accuracy performance in fog and SVM, RF, and ANN have 89.5%, 88.4%, and 87.2% accuracy performance respectively in the cloud for diabetes prediction. In addition, the throughput and delay results of heterogeneous nodes with different priorities in the WBAN scenario created using the IEEE 802.15.6 standard and AODV routing protocol have been also analyzed.

Keywords Cloud computing · Fog computing · IoMT · WBANs · Data analytics · Machine learning

1 Introduction

IoMT is one of the sub-branch of the Internet of Things (IoT) as remote health systems [1]. IoMT comprises wireless communication technologies, wearable health devices, sensor/actuator circuits, remote health monitoring, and IoT software and tools [2]. IoMT aims to construct a smart framework for healthcare. And also, IoMT framework is a complex architecture and connects various components that interact with each other and offers many solutions to end-users [3]. IoMT enhances real-time health monitoring solutions and human–machine interaction that improves patient

participation in decision-making [4]. IoMT applications have significant advantages such as cost reduction in health systems, real-time interventions in emergencies, and remote monitoring in pandemic environments [5].

IoMT concept evolves with the existing medical devices, IoT peripherals, and remote health software and hardware. As an IoT concept, IoMT devices have internet protocol (IP) capabilities to communicate with medical professionals and healthcare personnel via online servers [6]. Another key issue for IoMT is that it generates massive health data. Big health data from the IoMT devices are stored in local or cloud databases [7]. For the analysis of big data, advanced algorithms should be used with the help of machine learning, deep learning, and artificial intelligence for classification, correlation, or prediction [8, 9].

With the aforementioned specifications of the IoMT, the basic components of IoMT can be given in Fig. 1. People under surveillance are equipped with various sensor circuits on/in/around their bodies. The sensor nodes sense the vital signs with the help of their capabilities and they have wired or wireless communication utilities. In addition, the sensors

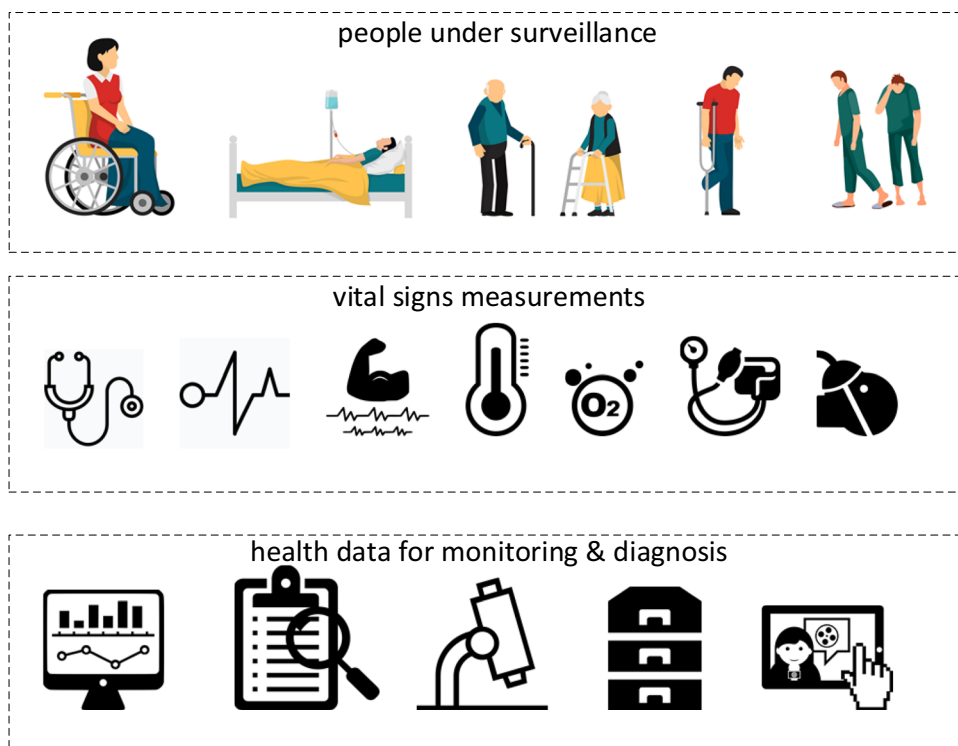
✉ Murtaza Cicioğlu
murtazacicioglu@uludag.edu.tr

¹ Computer Technology Department, Osmaniye Korkut Ata University, Osmaniye, Turkey

² Computer Engineering Department, Bursa Uludağ University, Bursa, Turkey

³ Computer Engineering Department, Düzce University, Düzce, Turkey

Fig. 1 IoMT components



are programmed with continuous or event-driven detection tasks of vital signs. A coordinator node has the control of a body's sensors as a central node. It collects the data from the sensors as a gateway to send to other components of IoMT. Essentially, the coordinator can be a personal computer (PC), smartphone, tablet, or a regular sensor node [10]. IEEE 802.15.6 standard which is the special and significant sensor network component of the IoMT calls the coordinator a HUB node [11]. The details of IEEE 802.15.6 as the WBAN standard will be explained in the following sections. The coordinator has some superior capabilities such as 4G/5G, Wi-Fi, and Bluetooth connections. The most important task of a coordinator is to send collected data to the remote centers. The remote center can be a secondary storage, database, monitoring device, fog, or cloud computing node. At this stage, if necessary, big data analysis begins.

Every day, thousands of people die from late diagnosis or are late for treatment. Real-time IoMT framework that consists of health sensor nodes, data gathering, and data analysis with computing capabilities is a necessity for immediate intervention and early diagnosis. For these purposes, we propose an IoMT framework with the components and functions mentioned above. For this purpose, first of all, Riverbed Modeler simulation software performs intra/inter-WBAN communications with all required design parameters. Next, a gateway with socket programming capabilities forwards the generated health data from Riverbed Modeler to other IoMT peripherals. And then, the diabetes prediction algorithms are performed in fog and cloud nodes. Finally,

the prediction results are given. The contributions of the proposed study can be summarized as follows:

- An IoMT framework is designed with important components and functions such as fog and cloud node capabilities.
- Real-time data has been obtained from WBANs in Riverbed Modeler for a more realistic performance analysis of IoMT.
- Real-time analysis is carried out with fuzzy logic and machine learning models.
- Intra and inter-WBAN communications (IEEE 802.15.6 standard) are modelled as essential components of the IoMT framework with all functions.

The organization of the paper is as follows: The related works are given in Section 2. In Section 3, the proposed IoMT framework with all components is explained in detail. The performance evaluation is examined in Section 4. The conclusions are in the last section.

2 Related works

There are various studies on IoMT, diabetes data analysis, fuzzy logic, and machine learning algorithms in the literature, but these issues have not been discussed together. In this study, we propose an IoMT framework that includes all

the topics mentioned above, as well as the concepts of fog and cloud.

Niswati et al. [12] proposed a fuzzy logic-based decision support system that diagnoses a person's diabetes risk using five variables: glucose, diastolic blood pressure, body mass index, and family history of diabetes 2 h after a meal. The decision support system is programmed in MATLAB according to the Mamdani inference method. In addition, a graphical user interface (GUI) has been developed for the use of the system. Experimental results show that the decision support system works with 100% correct diagnosis. Bressan et al. [13] propose an automatic fuzzy classification system showing the diabetes mellitus type 2 glycemic index level. The fuzzy-based system shows the glycemic index level in four different ways: decrease a lot, decrease, stable, and increase. In addition, the decision tree (DT) algorithm is used as a support in the decision system while fuzzy classification is made. The results show that the fuzzy classification system is promising and helps the decision-making process of doctors. Zou et al. [14] performed analyses predicting diabetes mellitus using hospital physical examination data in Luzhou, China. They used RF, DT, and neural network (NN) classification algorithms to predict diabetes mellitus. In the evaluation using fivefold cross-validation, the RF algorithm has the best prediction result with 80% accuracy.

Muhammad et al. [15] used logistic regression (LR), SVM, K-nearest neighbor (k-NN), RF, naive Bayes (NB), and gradient boosting (GB) to develop a prediction model with supervised machine learning algorithms using type 2 diabetes data from Kano Murtala Mohammed Specialist Hospital in Nigeria. In the predictions made in the study, the RF algorithm has the best performance with an accuracy rate of 88.76%. Yuvaraj and SriPreethaa [16] developed a Hadoop-based application for the classification of diabetes. In practice, they proposed a model using three different machine algorithms for classification: NB, DT, and RF. In the study using Pima Indians diabetes data, the RF algorithm made the best classification with 94% accuracy.

Ramesh et al. [17] proposed an IoT-based framework for the early diagnosis of diabetes by combining health data from personal health devices and smart wearables. In their study, there is a model in which the SVM classification algorithm is used for the prediction of diabetes. In their proposed framework, the results of accuracy, sensitivity, and specificity performance measurements were obtained as 83.20%, 87.20%, and 79%, respectively, with the tenfold cross-validation method. Tan and Halim [18] proposed an IoT-based remote health monitoring system that predicts diabetes and kidney disease in their study. In the architecture of their proposed system, some sensors measure three basic vital values such as body temperature, pulse, and blood pressure. The data received from the sensors are stored in the cloud structure of the system and disease prediction is made. ANN

algorithm is used for disease prediction. The accuracy rates of diabetes and kidney disease predictions were 90.54% and 87.88%, respectively.

Devarajan et al. [19] proposed a fog-assisted real-time health system to predict the user's diabetes risk. In their proposed health system, physiological values are transmitted to the cloud server via IoT smartphones for continuous real-time monitoring. The prediction is made with the J48Graft decision tree algorithm to determine if the blood sugar is at the risk level in the fog layer. When the experimental results are examined, the J48Graft decision tree predicts the risk level of diabetes with a high classification accuracy of 98.56%. In addition, it is seen that the proposed system is successful in terms of energy efficiency, estimation accuracy, computational complexity, and latency. Abdel-Basset et al. [20] proposed a new IoT-based framework for detecting and tracking type 2 diabetes patients. Their proposed framework consists of three main layers: wearable IoT sensors, fog layer, and cloud layer. WBAN and mobile applications are used to monitor the physiological changes of the users and to collect the data changed. In order to detect whether the user has diabetes, a new hybrid decision support system has been developed within the proposed framework and analyses have been carried out with the decision support system. As a result of the analysis, it was seen that the proposed system gave positive results.

Khan et al. [21] proposed an IoMT-based smart monitoring system that can be used by anyone exposed to COVID-19. The system was developed with the MATLAB tool using the Mamdani fuzzy inference method. The proposed system will collect user data directly related to COVID-19, assess the health status, and inform whether there is a need to consult a specialist for quarantine. The accuracy of their system is about 83%. Otoom et al. [22] proposed a real-time COVID-19 detection and monitoring system. The proposed system has been developed to detect suspected COVID-19 cases early, monitor treatment, and understand the disease. In addition, the IoT framework is used to collect real-time symptom data. The system consists of five components: Symptom Data Collection and Uploading, Quarantine/Isolation Center, Data Analysis Center, Health Physicians, and Cloud Infrastructure. In order to detect COVID-19 cases, SVM, NN, NB, k-NN, Decision Table (DeT), Decision Stump (DS), OneR, and ZeroR classification algorithms are used in the system. The results show that five of these eight algorithms achieved an accuracy of over 90%.

Kumar and Gandhi [23] proposed a health monitoring system to store and analyze large volumes of data from wearable IoT sensors. Their proposed system consists of a three-layer architecture: data collection, data storage, and data analytics. The first layer collects the data generated by the IoT sensors. The second layer stores large volumes of sensor data using Apache HBase. Finally, the third layer analyzes heart

diseases by analyzing with LR-based model using Apache Mahout. Receiver-operating characteristic (ROC) analyses are performed in the system to detect heart diseases. Kamarajugadda et al. [24] proposed a new IoMT-based healthcare system using the Biogeography-based Optimization-Support Vector Machine (BBO-SVM) model for heart disease prediction. The proposed model includes parameter settings of the SVM using the BBO algorithm. Statlog Heart Disease dataset is used to evaluate the proposed model. In the detailed analysis, their proposed BBO-SVM model achieved positive results with 89.26% accuracy, 88.33% sensitivity, 87.60% recall, 87.96% *F*-score, and 78.27% Kappa values. Khan and Algarni [25] proposed an IoMT system for the diagnosis of heart disease using Modified Salp Swarm Optimization (MSSO) and Adaptive Neuro-Fuzzy Inference System (ANFIS). In their proposed system, Levy Flight algorithm was used in the data preprocessing stage to make the analyses more effective. According to the analysis results, the proposed system predicts heart disease with 99.45% accuracy and 96.54% precision.

A summary of the comparative analysis of the relevant studies, along with the proposed study, is provided in Table 1. Studies based on diabetes and other diseases are given in [25, 26]. When the related studies are examined, there is no study similar to the framework proposed in our study. The proposed IoMT framework has superior features such as fog and cloud computing capabilities and an expandable application environment by intra-inter-WBAN communications, especially in pandemic situations. Both fuzzy logic and machine learning algorithms are considered for the performance evaluation and also real-time diabetes predictions are carried out in the proposed framework.

3 The proposed IoMT framework

The proposed IoMT framework consists of intra-WBANs, inter-WBAN communications with IoT capabilities, and fog and cloud components with diverse diabetes prediction capabilities as can be seen in Fig. 2. For more realistic performance analysis, intra-WBAN and inter-WBAN structures are built-in Riverbed Modeler. Riverbed Modeler simulation environment generates WBAN data (health big data) for diabetes prediction. Each WBAN has various body sensor nodes and a coordinator node as intra-WBAN. While the intra-WBAN structure consists of a set of body sensors and a coordinator, the intra-WBANs come together to form the inter-WBAN. Also, there are several WBANs in the environment that are capable of communicating with each other as inter-WBAN. In Fig. 2, each human body is equipped with body sensors and the coordinator illustrates the intra-WBAN structure and all these structures which are able to communicate among themselves construct the inter-WBAN. The health data produced in WBANs are transmitted to fog and cloud components for diabetes prediction. The fog component is deployed near the WBAN environment and is capable of predicting diabetes illness with the help of fuzzy logic. The cloud component is far from the WBANs and can be reached with inter-WBAN communications. The cloud component is capable of predicting diabetes illness with the help of machine learning. The fog components provide robust and real-time predictions and the cloud component deliver time-consuming processes with more accurate predictions because of performing complex analyses

Table 1 Comparative analysis of the related work

No	Authors	Type of disease	Machine learning	Fuzzy logic	Fog/cloud	IoT
1	Niswati et al. [26]	Diabetes		√	x	x
2	Bressan et al. [13]	Diabetes	DT	√	x	x
3	Zou et al. [14]	Diabetes	RF, DT, NN	x	x	x
4	Muhammad et al. [15]	Diabetes	LR, SVM, k-NN, RF, NB, GB	x	x	x
5	Yuvaraj and SriPreethaa [16]	Diabetes	NB, DT, RF	x	x	x
6	Ramesh et al. [17]	Diabetes	SVM	x	√	√
7	Tan and Halim [18]	Diabetes and kidney	ANN	x	√	√
8	Devarajan et al. [19]	Diabetes	J48Graft	x	√	√
9	Abdel-Basset et al. [20]	Diabetes		x	√	√
10	Khan et al. [21]	COVID-19		√	√	x
11	Otoom et al. [22]	COVID-19	SVM, NN, NB, k-NN, DeT, DS, OneR, ZeroR	x	√	√
12	Kumar and Gandhi [23]	Heart disease	LR	x	√	√
13	Kamarajugadda et al. [24]	Heart disease	BBO-SVM	x	√	√
14	Khan and Algarni [25]	Heart disease		√	√	x
15	The proposed framework	Diabetes	ANN, SVM, RF	√	√	√

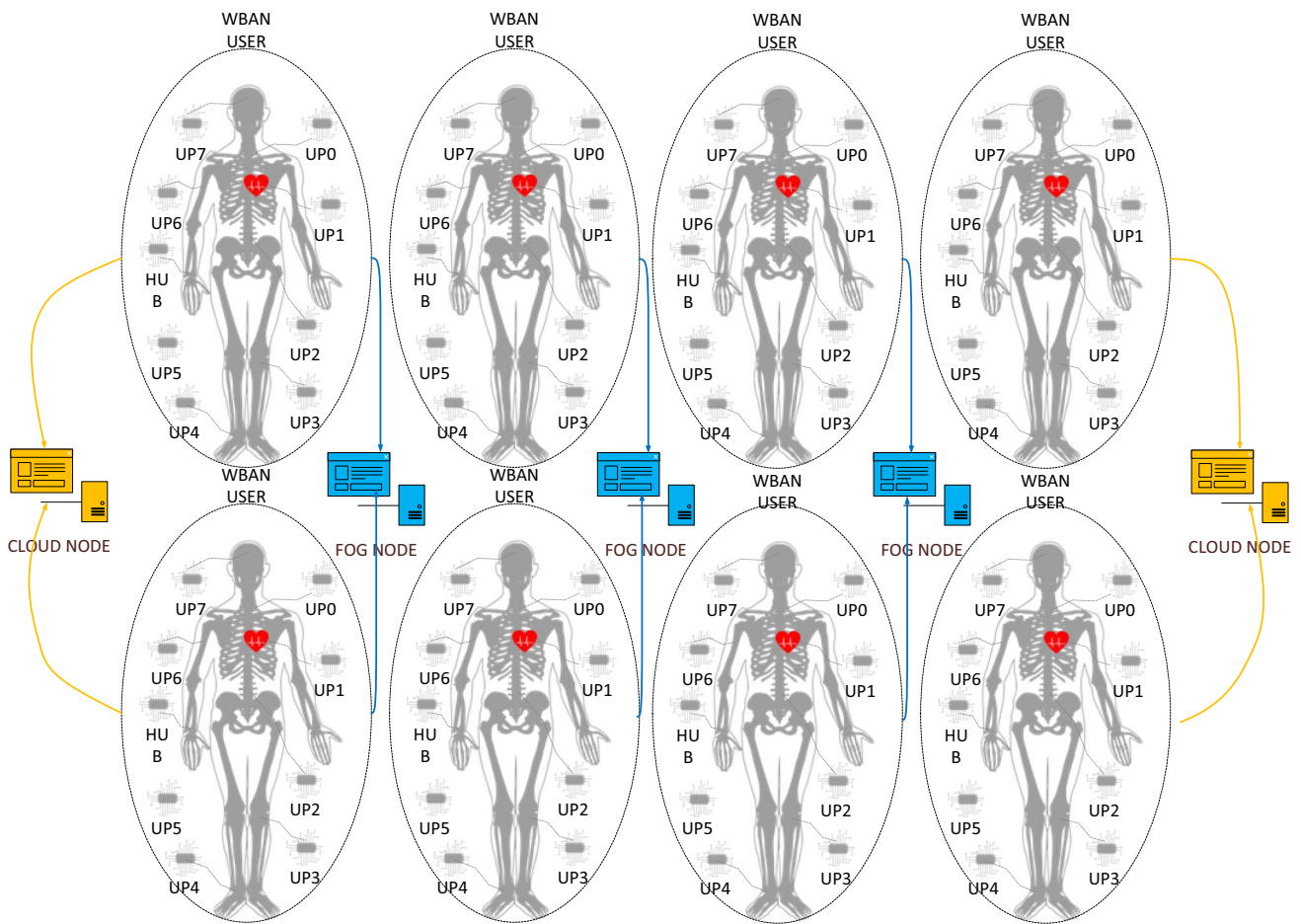


Fig. 2 The proposed IoMT architecture

such as machine learning algorithms. In our simulation scenario, while the fog components are positioned close to the WBANs, the cloud component is located away from the WBANs as can be seen from Fig. 2.

The overall topology of the proposed framework is illustrated in Fig. 3. A BAN (body area network) consists of eight health sensor nodes (N0–N7) and a HUB as a coordinator. Each HUB collects the eight health sensor data and forwards it to the gateway according to the nodes’ priorities. HUBs also communicate with the nearest fog component for fuzzy logic–based diabetes prediction. The BANs can communicate with other BANs named inter-WBAN communications thanks to the Ad hoc On-Demand Distance Vector (AODV) routing algorithm. The cloud component as can be seen from Fig. 3 only communicates with the gateway because of its communication coverage. The cloud component has machine learning capabilities for diabetes prediction.

In Fig. 4, the details of the proposed IoMT framework components are illustrated. The main components of the proposed framework can be summarized as three parts.

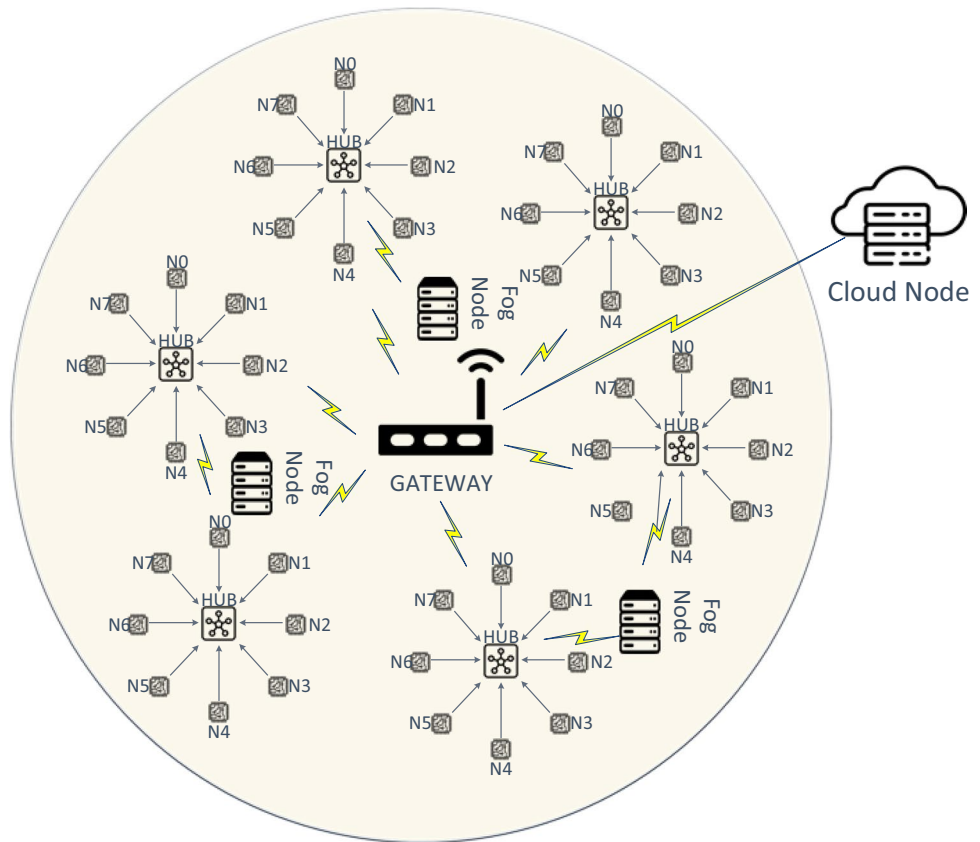
First, the big health data generated in Riverbed Modeler as IEEE 802.15.6 with Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA)–based WBANs. Second, the fog component as a near node to WBANs is capable of fuzzy logic–based diabetes prediction. Third, the cloud component as a far node to WBANs is capable of machine learning–based diabetes prediction. The followed subsection explains the details of the proposed framework components.

3.1 IEEE 802.15.6 WBAN of the IoMT framework

IEEE 802.15.6 standard defines all communication rules for WBAN structures. CSMA/CA-based medium access control mechanism with a priority-based approach on 2.4 GHz Industrial, Scientific, and Medical (ISM) bands is utilized in WBANs. Eight priorities represent the eight health sensor nodes as shown N0...N7 in the simulations and a coordinator node called HUB collects the data from sensors.

Riverbed Modeler is a simulation program that is capable of modelling ad hoc and infrastructure networks with all specifications. For health data generation, IEEE 802.15.6–based

Fig. 3 Topology of the proposed IoMT framework



WBANs are modelled in Riverbed Modeler with Proto-C programming language. Each sensor node in WBANs is capable of producing specific sensor values and the HUB is responsible for gathering health data from sensor nodes.

The priorities that are divided into eight different access categories in the IEEE 802.15.6 standard are shown in Table 2. These different types of traffic are separated as urgent data, high priority medical data or network control, medical data or network control, voice, video, excellent effort, and best effort. Also, these different data traffics perform user preference with minimum and maximum contention window (CW) values for the CSMA/CA mechanism.

3.2 Fog side of the IoMT framework

Node with fog computing capability is located between data source and cloud. So, the fog node is located near the WBANs. Fog computing provides data analysis and storage locally unlike transferring data to a central server. Also, fog computing has limited operations for faster decisions because, in some applications, data may need to be processed as quickly as possible. Fog computing comes into prominence with the IoT concept nowadays.

In light of the above-mentioned advantages of fog computing, we propose a fog computing node structure for fast decisions with limited data in diabetes prediction. Five

different health data as glucose, blood pressure, body mass index, diabetes pedigree function, and pregnancy number are processed for analyzing diabetes as positive or negative. For the decision process of diabetes, fuzzy logic has been utilized in the fog computing node.

Fuzzy logic as a subset of artificial intelligence resembles the human decision-making procedure and deals with vague and imprecise information. The decision-making process is done with taking into account some related quantities as fuzzy system inputs. The inputs are processed with the fuzzy logic system components with the rule base and membership functions. Fuzzy logic is a rule-based system that can trust the practical experience of experts.

The fuzzy logic block diagram is given in Fig. 5 consisting of fuzzifier, knowledge base, inference engine, and defuzzifier units. The fuzzifier is the method of transforming a crisp quantities such as 50, 150, and 200 into fuzzy quantities such as normal, medium, and abnormal. The knowledge base is a collection of fuzzy IF–THEN rules. The inference engine uses the IF–THEN rules along with OR or AND operations for operating decision rules. The defuzzifier is the inverse of fuzzifier and converts the fuzzy results into crisp results.

The proposed fuzzy logic structure has five inputs as plasma glucose concentration, blood pressure, body mass index, diabetes pedigree function, and pregnancy number

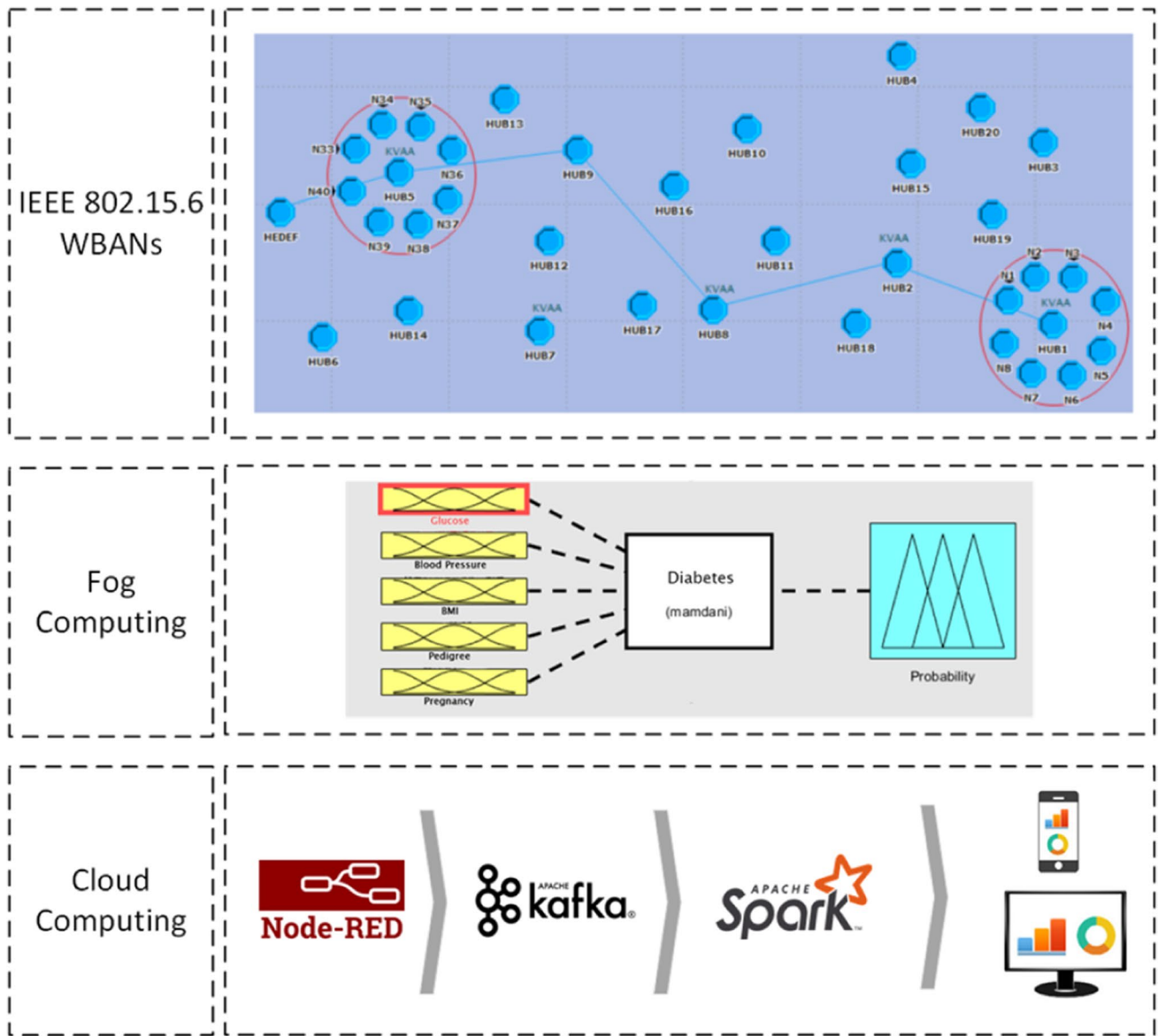


Fig. 4 The main components of the proposed IoMT framework

Table 2 User priority mapping (D, data; M, management)

P	Traffic	Packet type	CSMA/CA	
			CW _{min}	CW _{max}
0	Background	D	16	64
1	Best effort	D	16	32
2	Excellent effort	D	8	32
3	Video	D	8	16
4	Voice	D	4	16
5	Medical data	D/M	4	8
6	High priority medical data	D/M	2	8
7	Emergency	D	1	4

as can be shown in Table 3. These values were chosen by taking reference from the studies in the literature [26]. Each input has three linguistic variables as abnormal, medium, and normal states. In the proposed fuzzy logic system, three membership functions define each variable. All membership functions have been constructed with intervals and their values as stated in Table 3.

In Table 4, some of the proposed fuzzy logic rules are given and there are 243 possible rules in the system. Each row in Table 4 has six columns and the top five columns are the inputs and the last column diagnoses diabetes as output. The proposed fuzzy logic-based diabetes prediction system with the aforementioned specifications can predict diabetes illnesses with 64% probability. This probability has been

Fig. 5 The fuzzy logic block diagram

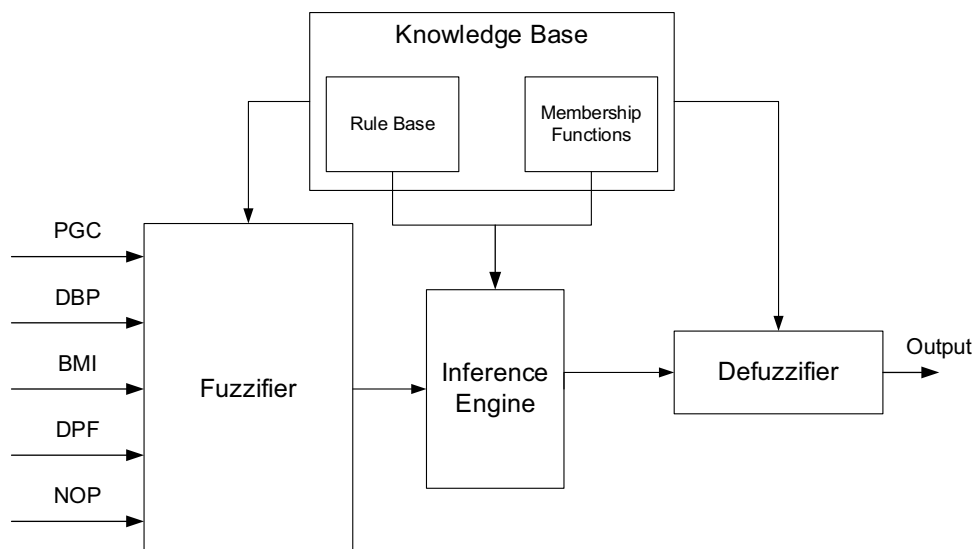


Table 3 The dataset features and value intervals

Data	Linguistic variable	Interval
Plasma glucose concentration at 2h (PGC)	Abnormal	187–232
	Medium	141–186
	Normal	44–140
Diastolic Blood Pressure (mm Hg) (DBP)	Abnormal	92–122
	Medium	81–91
	Normal	30–80
Body Mass Index (BMI)	Abnormal	34–67
	Medium	26–33
	Normal	18–25
Diabetes Pedigree Function (DPF)	Abnormal	0.54–2.29
	Medium	0.50–0.53
	Normal	0.08–0.49
Number of Pregnancy (NOP)	Abnormal	10–17
	Medium	5–9
	Normal	0–4

found after a hundred trials. The fuzzy-based system has been able to predict diabetes correctly in 64 of 100 trials.

3.3 Cloud side of the IoMT framework

Cloud computing gives better facilities than fog computing because its capabilities are much more with high capacity calculations and storage. Cloud computing also provides networking, data analytics, various software, and artificial intelligence/machine learning–based specifications. Besides, cloud computing nodes are far from the data

sources and much more clients can be served in cloud computing nodes.

In the light of the above-mentioned advantages of cloud computing, we propose a cloud computing node structure for high capacity calculations with non-limited data in diabetes prediction. Eight different health data as pregnancy number, glucose, blood pressure, skin thickness, insulin, body mass index, diabetes pedigree function, and age are processed for analyzing diabetes positive or negative. For the decision process of diabetes, SVM, RF, and ANN as machine learning algorithms have been utilized. Apache Spark [27] and Apache Kafka [28] software with machine learning libraries run in the cloud computing node.

Apache Spark is a popular data processing framework built to be reliable, fast, highly fault-tolerant, and consistent with Hadoop [27, 29, 30]. Apache Spark offers a distributed architecture and analysis framework that can handle very large datasets [27]. The most important feature of Apache Spark is that it can perform in-memory calculations. It uses random access memory (RAM) for computation rather than accessing the slow hard disk. It works by distributing the workload in parallel, rather than executing transactions serially when processing big data [31]. Additionally, Apache Spark supports rich libraries, including Spark SQL for SQL, MLlib for machine learning, GraphX for graph processing, and Spark Streaming. Spark’s ecosystem is shown in Fig. 6.

Apache Spark is compatible with Hadoop elements such as Hadoop Distributed File System (HDFS) and Hive and can run on a Hadoop cluster using YARN. Apache Spark is written in Scala and uses the Java Virtual Machine (JVM). However, it also supports languages such as Scala, Java, SQL, Python, and R. However, analyses can be performed using different machine learning algorithms included with MLlib, other ML libraries [32].

Table 4 Fuzzy logic rules with linguistic variables

PGC	DBP	BMI	DPF	NOP	Prediction
NORMAL	NORMAL	NORMAL	NORMAL	NORMAL	NEGATIVE
ABNORMAL	ABNORMAL	ABNORMAL	ABNORMAL	ABNORMAL	POSITIVE
NORMAL	NORMAL	NORMAL	MEDIUM	MEDIUM	NEGATIVE
MEDIUM	NORMAL	ABNORMAL	NORMAL	NORMAL	POSITIVE
MEDIUM	NORMAL	MEDIUM	MEDIUM	NORMAL	NEGATIVE
MEDIUM	NORMAL	MEDIUM	ABNORMAL	ABNORMAL	POSITIVE
MEDIUM	MEDIUM	MEDIUM	NORMAL	NORMAL	NEGATIVE
ABNORMAL	MEDIUM	NORMAL	MEDIUM	MEDIUM	POSITIVE
NORMAL	NORMAL	MEDIUM	MEDIUM	ABNORMAL	NEGATIVE
NORMAL	ABNORMAL	MEDIUM	ABNORMAL	ABNORMAL	POSITIVE

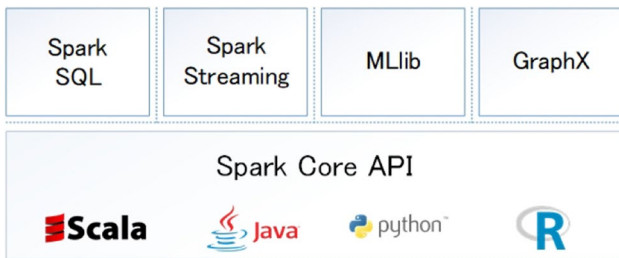


Fig. 6 Apache Spark ecosystem

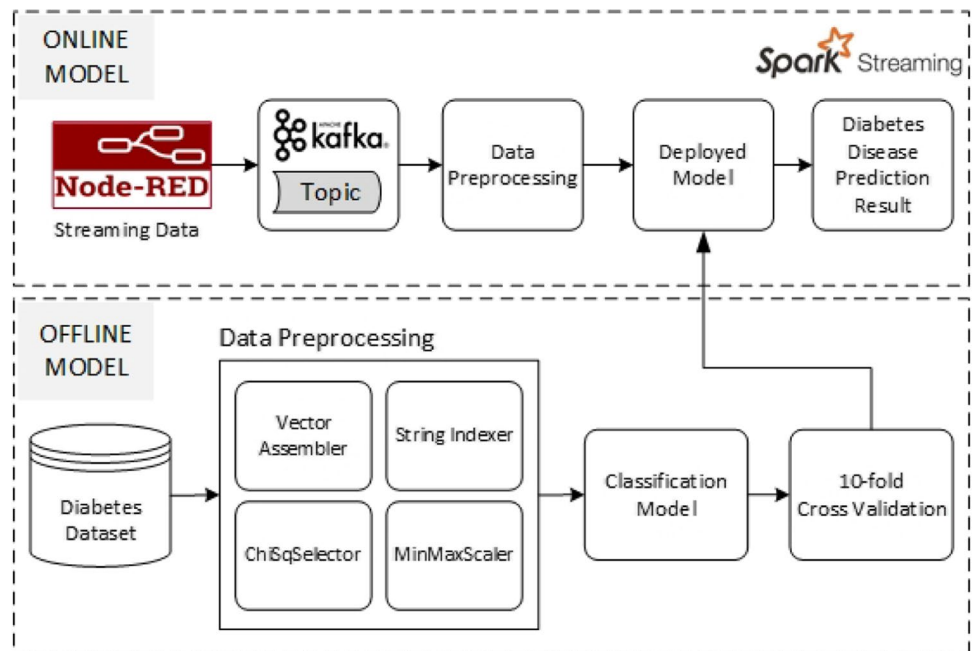
Apache Kafka is a distributed data streaming platform developed by LinkedIn to enable the fast and error-free transfer of data from one system to another. It has a flexible architecture that is fault-tolerant and can be scaled horizontally. The main purpose of using Kafka is

to collect large amounts of data received from the data source and transfer it to other systems with a delay of milliseconds in real time [28]. Cloud computing node structure as seen in Fig. 7 consists of two components as offline and online models. Each component will be explained in detail.

3.3.1 Offline model

The purpose of this component is to develop a model that provides the highest possible accuracy for diabetes prediction. ANN, SVM, and RF classification algorithms are used to create the offline model. The component consists of three stages: data preprocessing, classification algorithms, and evaluation of classification algorithms. Figure 7 gives

Fig. 7 Cloud side of the IoMT framework



the architecture of the components that the construct of the offline model.

a) Data preprocessing

This stage is one of the necessary steps for data to be used effectively in machine learning classification algorithms. For this reason, the values in the diabetes dataset are taken first in the offline model and the data is pre-processed. In the data preprocessing phase of this study, VectorAssembler, StringIndexer, ChiSqSelector, and MinMaxScaler operations are applied.

VectorAssembler is used to vectorally combine features used in training and predicting ML models in a single column. StringIndexer is used to define the properties to be used in ML models as categorical variables or to convert text data to numerical data while preserving the categorical context. ChiSqSelector is used to select the important features from the input features. It is provided by the Spark ML library. ChiSqSelector uses the chi-square independence test to identify the features that are important. In the study, the first 5 features given by the numTopFeatures parameter in ChiSqSelector are selected. Finally, the MinMaxScaler scaler is used to normalize the input features. MinMaxScaler scales data in all features from 0 to 1.

b) Classification algorithm

Machine learning classification models have been used to predict diabetes. Three popular classification algorithms, ANN, SVM, RF, are applied to train and test the diabetes dataset. Also, Apache Spark's MLlib library is used to implement classification algorithms. Here, training data is used to train models and test data is used to evaluate models. ANN is a mathematical model that consists of a large number of simple, highly interconnected processing elements (neurons) to solve a problem. One advantage of a neural network is that it can model various response surfaces when given a sufficient number of hidden nodes. There are several types and architectures of neural networks [33]. In this study, Multilayer Perceptron Classifier (MLPC) model is used. This type of networking has an input layer, two hidden layers, and an output layer.

SVM is a classification method introduced by Boser, Guyon, and Vapnik in 1992 [34]. The SVM classifier is widely used due to its high accuracy and ability to handle high-dimensional data. The SVM training algorithm aims to find a hyperplane that divides the dataset into different numbers of predefined classes consistent with training examples [35]. In this study, linear support vector machine (LinearSVM) model from Apache Spark MLlib library is used. Linear SVM supports binary classification.

RF is a supervised machine learning classification algorithm. There is a direct relationship between the number of trees in the algorithm and the results it can achieve. As the number of trees increases, a precise result is obtained. The difference between the RF algorithm and decision tree algorithm is that finding root node and splitting nodes in RF is random and there is a probabilistic calculation for the process in the DT [36].

In this study, the parameters of the classifiers used from the Apache Spark MLlib library are given in Table 5.

c) Evaluation of classification algorithms

The K -fold cross-validation method has been used to achieve the best accuracy in the classification model. A tenfold cross-validation process has been chosen in the study. In K -fold cross-validation, the dataset splits the dataset into k equal-size folds. The $k-1$ folds are used for training the model and testing the model with the remaining fold. This process (training and testing cycle) is repeated until each fold has been used to test the model. In addition, the accuracy of the model is calculated for each fold. The average accuracy calculated at the end of a tenfold process is used in the evaluation of the model. However, classification models were evaluated with accuracy, recall, precision, and F1-score quality measures. In addition, the receiver-operating characteristic (ROC) curve has been used in the evaluation.

People without diabetes were accepted as a positive class and people with diabetes as a negative class. Here,

True positive (TP)—the number of people who are presumed to have no diabetes but actually do not have diabetes.

False positive (FP)—the number of people who are not presumed to have diabetes but actually have diabetes.

True negative (TN)—the number of people who are predicted to have diabetes but actually have diabetes.

False negative (FN)—the number of people who are predicted to have diabetes but actually do not have diabetes.

Table 5 The parameters of classification algorithm

Classification algorithm	Parameters
ANN	layers = [2, 5–7], seed = 1
SVM	maxIter = 100, regParam = 0.001, standardization = True
RF	maxBins = 10, maxDepth = 5, impurity = "gini", numTrees = 10, seed = 1

The performance metrics are defined below [3]:

Accuracy The ratio of occurrences that the classification algorithms categorize correctly. It means the ratio of correctly estimated samples to the total number of samples.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Recall The ratio of predicted positive samples to total positive samples. Recall is the total number of correct predictions for a given diabetes divided by the total number of samples that should have been predicted as that diabetes.

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

Precision The ratio of true positive to all samples classified as positive. Precision is the total number of correct predictions for given diabetes divided by the total number of predictions for that diabetes.

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

F1-score It is the weighted average of precision and recall.

$$F1 - score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{4}$$

ROC Curve ROC is a probability curve for different classes. A typical ROC curve has false positive rate (FPR) on the X-axis and true positive rate (TPR) on the Y-axis. Using the ROC curve, the curves of different classification models can be directly compared in general or for different thresholds. AUC stands for area under the ROC curve. The larger the AUC area, the better the machine learning models are at distinguishing given classes. The ideal value for AUC is 1.

The time complexity of the ML algorithms used can be viewed as a measure of how fast or slow the model will run for the input size. Usually this analysis is handled

according to the input dimension. The training time complexity of the SVM is given in big O notation as the square of the input size $O(n^2)$ [37]. RF training time complexity is $O(n * \log(n) * \text{dimensional} * \text{number of Decision Trees})$ [38] and time complexity of ANN is $O(\text{epochs} * n * \text{features} * \text{number of neurons})$ [39]. The training times for RF, SVM, and ANN are 8.673, 57.733, and 22.978 respectively.

3.3.2 Online model

This component is used by analyzing real-time WBAN data with machine learning algorithms. Real-time data goes through many stages in this component. The eight different WBAN parameters obtained from the Riverbed Modeler simulation contain the attribute array in the same order as the training data and this data is sent to the Apache Kafka topic with the Node-Red flow development tool. The data coming into Apache Kafka is collected on the relevant topic as JavaScript Object Notation (JSON) data. After this stage, Apache Spark takes the WBAN data from the Kafka topic to analyze it. A machine learning classification model has been developed to predict diabetes. SVM algorithm, which has the best accuracy rate performance, was used in the machine learning model. Finally, by applying the developed model to the real-time WBAN data, estimation results regarding the presence or absence of diabetes are obtained.

4 Performance evaluation

The confusion matrix shows the number of correct and incorrect predictions (also known as the hits and misses) made by classification models based on actual results in the data. The confusion matrix is the most used metric for the performance evaluation of models. In Fig. 8, confusion matrix values resulting from the estimation of ANN, SVM, and RF classification algorithms are shown. TP, TN, FN, and FP values given in the confusion matrix are used to calculate the performance values (accuracy, recall,

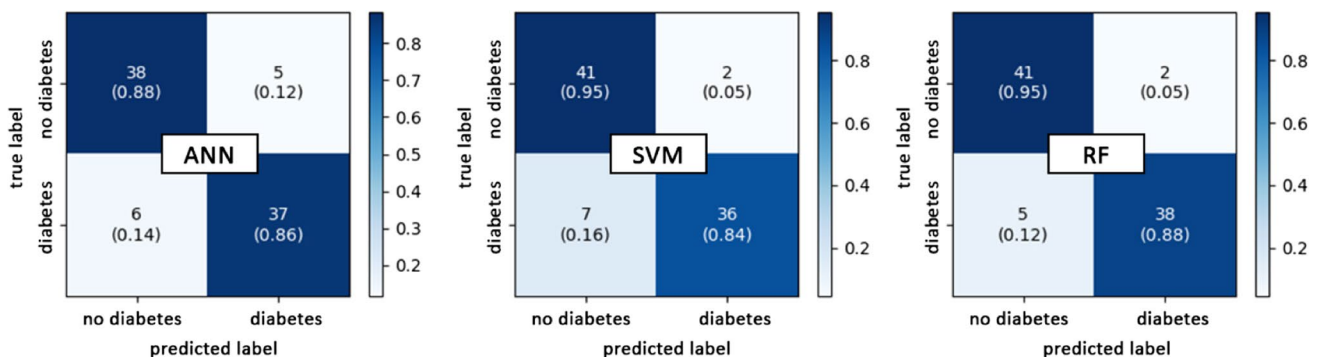
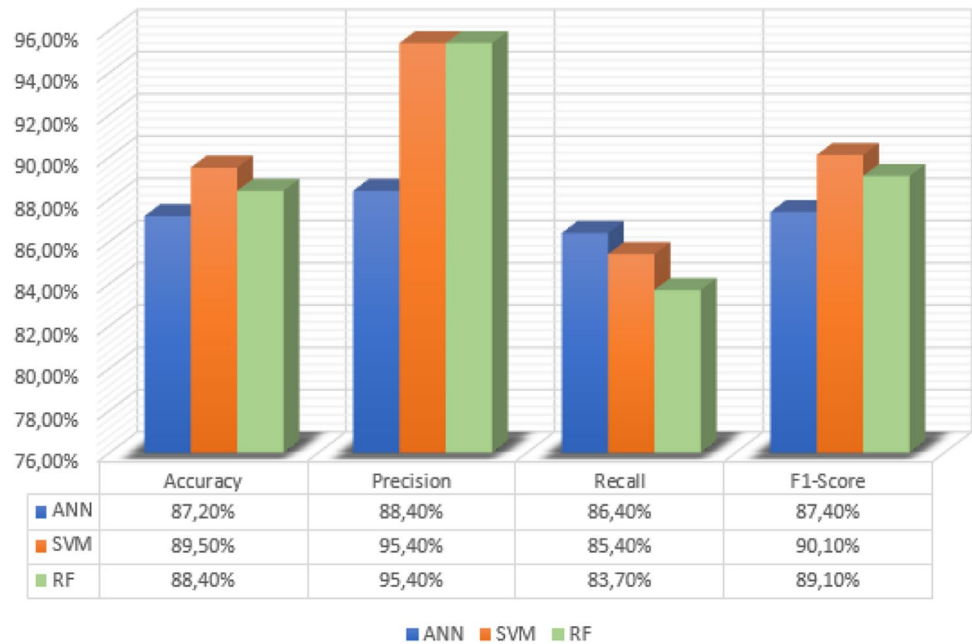


Fig. 8 ANN, SVM, and RF confusion matrix

Fig. 9 Classification algorithms performance values



precision, F1-score) of the ANN, SVM, and RF classification algorithms.

When the performances of the classification algorithms are examined in Fig. 9, it is seen that SVM with 89.5% accuracy, 95.4% precision, and 90.1% F1-score values has the best performance. Also, ANN has the lowest performance with 87.2% accuracy, 88.4% precision, and 87.4% F1-score. However, all these performance values are directly related to the dataset used. Therefore, the values that will be reconstructed with the new values that will be added to the dataset will affect the performance of these algorithms and the algorithm used in the online model may be different.

The graph of ROC curves of different classification models is given in Fig. 10. When the ROC curves were examined, the AUC values of the RF, SVM, and ANN models were obtained as 0.972, 0.963, and 0.952, respectively, and it was observed that these values were very close to the value of 1. It is seen that the classification models used in the detection of diabetes give results close to an ideal ROC curve.

In the proposed simulation model, each sensor node has a different priority. The priorities are divided into eight different categories and are determined as N7, N6, ..., N0 from the highest to the lowest priority. Therefore, each node is assigned to a different priority category. A WBAN consists of eight nodes (each with different priority and different packet inter-arrival time as N7, N6... N0) and one coordinator node (HUB). WBAN is designed according to the contention window values specified in the IEEE 802.15.6 standard and the rules defined for medium access. A heterogeneous structure has been designed due to the task assigned to each node (such as temperature, pulse, and heartbeat) and

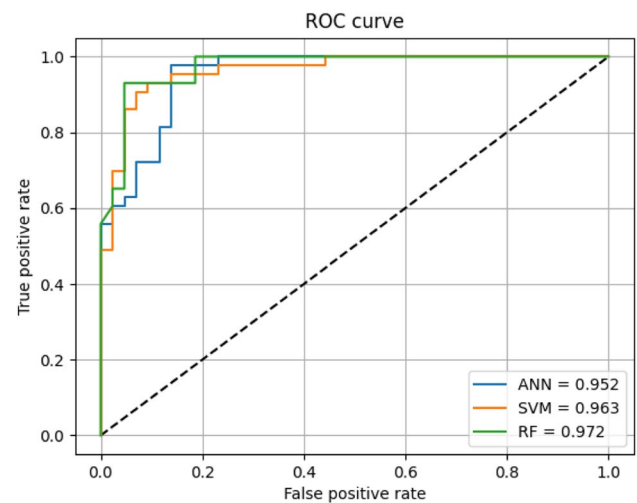


Fig. 10 Classification algorithms ROC curves

packet inter-arrival time. Finally, the source WBAN has to sort its packets according to their priority and transmit them to the gateway or sink (destination). Achieving this goal is subject to coverage limitations in communication technologies. In other words, the source WBAN that is in the coverage area of the target can send its packets directly, while the WBAN outside the coverage area needs a relay or intermediate nodes and routing algorithm. In our study, Ad hoc On-Demand Distance Vector (AODV) routing algorithm, which is frequently used in the literature, has been performed. In this way, even if the source WBAN is out of the target's coverage area is able to deliver the relevant data packets to the destination with the help of the AODV routing algorithm

in a reactive and distributed manner. The parameters used in our simulation scenario are given in detail in Table 6.

In Fig. 11, the throughput results of the packets transmitted to the destination (gateway) by the source WBAN by using CSMA-CA-based IEEE 802.15.6 and AODV routing algorithm are given comparatively according to different priority nodes. According to the obtained results, there is a decrease from N7 to N1 node in throughput. This is because the highest priority node (N7) with a packet inter-arrival time of 0.25 packet/s transmits more packets to the coordinator. On the other hand, the lowest priority node (N0) with 6 packet/s packet inter-arrival time transmits fewer packets to the coordinator. Packet inter-arrival time defined for nodes are given in Table 6. All nodes can transmit their packets to the destination as a successful result of the medium access control technique defined in the CSMA/CA-based IEEE

802.15.6 protocol. In addition, WBANs outside the coverage area successfully have transmitted their packets to the destination with the help of the AODV routing algorithm.

In Fig. 12, the delay results are given comparatively according to different priority nodes. According to the results obtained, nodes with higher priority have a lower delay, while lower priority nodes have a higher delay. The reason for this is the CW ranges are defined for different priorities in the CSMA/CA-based IEEE 802.15.6 protocol. Contention window intervals determine the priorities of the nodes to use the wireless medium to send their packets. As the contention window interval increases, the delays of the nodes also increase. Nodes with the lowest contention window interval can use the wireless medium earlier. Although this approach is based on the backoff algorithm in the traditional CSMA/CA protocol, this algorithm has

Table 6 Simulation parameters

Parameter	Value
Simulation time	3600 s
Node placement	Stationary
Network topology	Star
Physical and MAC models	CSMA-CA-based IEEE 802.15.6
Routing protocol	AODV
Channel frequency and data rate	2.4 GHz and 971.4 kbps
Packet length	100 byte
Number of nodes	16 sensor nodes and 20 coordinator (HUB)
Packet inter-arrival time ^a	N7 = 0.25 s N3 = 3 s N6 = 0.5 s N2 = 4 s N5 = 1 s N1 = 5 s N4 = 2 s N0 = 6 s

^aExponential distribution function

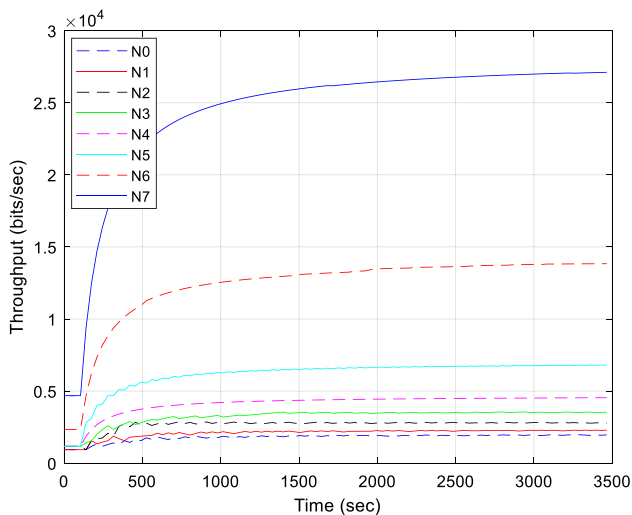


Fig. 11 Comparison of throughput rates of different priority nodes

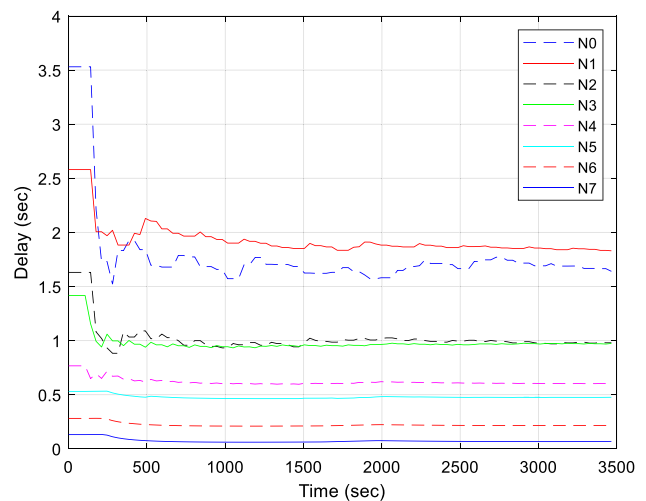


Fig. 12 Comparison of delay results of different priority nodes

been redesigned as an Alternative Binary Exponential Backoff in the IEEE 802.15.6 standard. In this way, delay differences between priorities occurred.

5 Conclusions

Remote health monitoring and diagnosis services are gaining importance day by day. Especially in pandemic situations, increasing health costs and the number of intensive care patients, and decreasing empty hospital beds and the number of healthcare professionals per capita drive the governments to take precautions. IoMT is a promising solution for the aforementioned problems. Accordingly, we proposed an IoMT framework with the help of WBANs, cloud computing, fog computing, fuzzy logic, and machine learning algorithms (SVM, RF, and ANN) in a diabetes prediction scenario. We discussed the proposed IoMT framework with all components. The fuzzy logic gives 64% accuracy performance in fog and SVM, RF, and ANN have 89.5%, 88.4%, and 87.2% accuracy performance respectively in the cloud for diabetes prediction. RF is the more successful for diabetes prediction scenario as a machine learning algorithm. The IoMT approach is built in the Riverbed Modeler, using the CSMA-CA-based IEEE 802.15.6 protocol and the AODV routing algorithm. The end-to-end delay and throughput results of heterogeneous nodes with various priorities are examined comparatively. In light of the results obtained, it has been determined that vital data is transmitted to the destination (sink, gateway, or remote center) with a lower delay and high throughput rate. The Apache Software provides effective and efficient performance, especially in real-time applications such as IoMT frameworks. The developed IoMT framework provides in communities such as hospitals and nursing homes. For future works, various machine learning algorithms and different diseases will be taken into account to build an IoMT framework for several pandemic and healthcare applications.

Declarations

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflict of interest The authors declare no competing interests.

References

- Wazid M, Das AK, Rodrigues JJPC et al (2019) IoMT malware detection approaches: analysis and research challenges. *IEEE Access* 7:182459–182476. <https://doi.org/10.1109/ACCESS.2019.2960412>
- Alsubaei F, Abuhussein A, Shandilya V, Shiva S (2019) IoMT-SAF: internet of medical things security assessment framework. *internet of things (Netherlands)* 8: 1–34. <https://doi.org/10.1016/j.iot.2019.100123>
- Tausif M, Jain A, Khan E, Hasan M (2021) Memory-efficient architecture for FrWF-based DWT of high-resolution images for IoMT applications. *Multimed Tools Appl* 80:11177–11199. <https://doi.org/10.1007/s11042-020-10258-0>
- Haoyu L, Jianxing L, Arunkumar N et al (2019) An IoMT cloud-based real time sleep apnea detection scheme by using the SpO2 estimation supported by heart rate variability. *Futur Gener Comput Syst* 98:69–77. <https://doi.org/10.1016/j.future.2018.12.001>
- De Fazio R, De Vittorio M, Visconti P (2021) Innovative IoT solutions and wearable sensing systems for monitoring human biophysical parameters: a review. *Electronics* 10:1660. <https://doi.org/10.3390/electronics10141660>
- Koutras D, Stergiopoulos G, Dasaklis T et al (2020) Security in IoMT communications: a survey. *Sensors* 20:4828. <https://doi.org/10.3390/s20174828>
- Shi W, Cao J, Zhang Q et al (2016) Edge computing: vision and challenges. *IEEE Internet Things J* 3:637–646. <https://doi.org/10.1109/JIOT.2016.2579198>
- Bibi N, Sikandar M, Ud Din I et al (2020) IoMT-based automated detection and classification of leukemia using deep learning. *J Healthc Eng* 2020:1–12. <https://doi.org/10.1155/2020/6648574>
- Sheeba Rani S, Selvakumar S, Pradeep Mohan Kumar K et al (2021) Internet of medical things (IoMT) with machine learning-based COVID-19 diagnosis model using chest X-ray images. In: *Data Science for COVID-19*. Elsevier, pp 627–641. <https://doi.org/10.1016/C2020-0-01677-4>
- Han T, Zhang L, Pirbhulal S et al (2019) A novel cluster head selection technique for edge-computing based IoMT systems. *Comput Networks* 158:114–122. <https://doi.org/10.1016/j.comnet.2019.04.021>
- SavaşçıŞen S, Cicioğlu M, Çalhan A (2021) IoT-based GPS assisted surveillance system with inter-WBAN geographic routing for pandemic situations. *J Biomed Inform* 116:103731. <https://doi.org/10.1016/j.jbi.2021.103731>
- Niswati Z, Mustika FA, Paramita A (2018) Fuzzy logic implementation for diagnosis of diabetes mellitus disease at puskesmas in East Jakarta. *J Phys Conf Ser* 1114:1–7. <https://doi.org/10.1088/1742-6596/1114/1/012107>
- Bressan GM, Azevedo BCF de, Souza RM de (2020) A fuzzy approach for diabetes mellitus type 2 classification. *Brazilian Arch Biol Technol* 63:1–11. <https://doi.org/10.1590/1678-4324-2020180742>
- Zou Q, Qu K, Luo Y et al (2018) Predicting diabetes mellitus with machine learning techniques. *Front Genet* 9(515):1–10. <https://doi.org/10.3389/fgene.2018.00515>
- Muhammad LJ, Algehyne EA, Usman SS (2020) Predictive supervised machine learning models for diabetes mellitus. *SN Comput Sci* 1:240. <https://doi.org/10.1007/s42979-020-00250-8>
- Yuvaraj N, SriPreethaa KR (2019) Diabetes prediction in healthcare systems using machine learning algorithms on Hadoop cluster. *Cluster Comput* 22:1–9. <https://doi.org/10.1007/s10586-017-1532-x>
- Ramesh J, Aburukba R, Sagahyoon A (2021) A remote healthcare monitoring framework for diabetes prediction using machine learning. *Healthc Technol Lett* 8:45–57. <https://doi.org/10.1049/htl2.12010>
- Tan ET, Halim ZA (2019) Health care monitoring system and analytics based on Internet of Things framework. *IETE J Res* 65:653–660. <https://doi.org/10.1080/03772063.2018.1447402>
- Devarajan M, Subramaniaswamy V, Vijayakumar V, Ravi L (2019) Fog-assisted personalized healthcare-support system for

- remote patients with diabetes. *J Ambient Intell Humaniz Comput* 10:3747–3760. <https://doi.org/10.1007/s12652-019-01291-5>
20. Abdel-Basset M, Manogaran G, Gamal A, Chang V (2020) A novel intelligent medical decision support model based on soft computing and IoT. *IEEE Internet Things J* 7:4160–4170. <https://doi.org/10.1109/JIOT.2019.2931647>
 21. Abbas Khan T, Abbas S, Ditta A, et al. (2020) IoMT-based smart monitoring hierarchical fuzzy inference system for diagnosis of COVID-19. *Comput Mater Contin* 65:2591–2605. <https://doi.org/10.32604/cmc.2020.011892>
 22. Ootom M, Otoum N, Alzubaidi MA et al (2020) An IoT-based framework for early identification and monitoring of COVID-19 cases. *Biomed Signal Process Control* 62:102149. <https://doi.org/10.1016/j.bspc.2020.102149>
 23. Kumar PM, Devi Gandhi U (2018) A novel three-tier Internet of Things architecture with machine learning algorithm for early detection of heart diseases. *Comput Electr Eng* 65:222–235. <https://doi.org/10.1016/j.compeleceng.2017.09.001>
 24. Kamarajugadda KK, Movva P, Raju MN et al (2021) IoMT with cloud-based disease diagnosis healthcare framework for heart disease prediction using simulated annealing with SVM. In: Gupta D Hugo C de Albuquerque V Khanna A Mehta PL (eds) *Smart Sensors for Industrial Internet of Things*. Internet of Things. Springer, Cham, pp 115–126. https://doi.org/10.1007/978-3-030-52624-5_8
 25. Khan MA, Algarni F (2020) A healthcare monitoring system for the diagnosis of heart disease in the IoMT cloud environment using MSSO-ANFIS. *IEEE Access* 8:122259–122269. <https://doi.org/10.1109/ACCESS.2020.3006424>
 26. Niswati Z, Paramita A, Mustika FA (2016) Aplikasi Fuzzy Logic dalam Diagnosa Penyakit Diabetes Mellitus pada PUSKESMAS di Jakarta Timur. *J Nas Teknol dan Sist Inf* 2:21–30. <https://doi.org/10.25077/TEKNOSI.v2i3.2016.21-30>
 27. Zaharia M, Xin RS, Wendell P et al (2016) Apache Spark. *Commun ACM* 59:56–65. <https://doi.org/10.1145/2934664>
 28. Thein KMM (2014) Apache kafka: next generation distributed messaging system. *Int J Sci Eng Technol* 3(47):9478–9483
 29. Hafsa M, Jemili F (2018) Comparative study between big data analysis techniques in intrusion detection. *Big Data Cogn Comput* 3:1. <https://doi.org/10.3390/bdcc3010001>
 30. Misra C, Bhattacharya S, Ghosh SK (2020) A fast scalable distributed kriging algorithm using Spark framework. *Int J Data Sci Anal* 10:249–264. <https://doi.org/10.1007/s41060-020-00215-3>
 31. Meng X, Bradley J, Yavuz B et al (2016) MLlib: machine learning in apache spark. *J Mach Learn Res* 17:1–7
 32. Alber M (2014) Big data and machine learning: a case study with bump boost. Free University of Berlin
 33. Clarke MRB, Duda RO, Hart PE (1974) Pattern classification and scene analysis. *J R Stat Soc Ser A* Wiley 137, pp 442–443. <https://doi.org/10.2307/2344977>
 34. Boser B, Guyon I (1992) A training algorithm for optimal margin classifiers. In: *Proceedings of the fifth annual workshop on Computational learning theory*, pp 44–152. <https://doi.org/10.1145/130385.130401>
 35. Zhu G, Blumberg DG (2002) Classification using ASTER data and SVM algorithms: the case study of Beer Sheva, Israel. *Remote Sens Environ* 80(2):233–240. [https://doi.org/10.1016/S0034-4257\(01\)00305-4](https://doi.org/10.1016/S0034-4257(01)00305-4)
 36. Breiman L (2001) Random forests. *Machine Learning* 45:5–32. <https://doi.org/10.1023/A:1010933404324>
 37. Abdiansah A, Wardoyo R (2015) Time complexity analysis of support vector machines (SVM) in LibSVM. *Int J Comput Appl* 128(3):28–34. <https://doi.org/10.5120/ijca2015906480>
 38. Zheng X, Jia J, Guo S et al (2021) Full parameter time complexity (FPTC): a method to evaluate the running time of machine

- learning classifiers for land use/land cover classification. *IEEE J Sel Top Appl Earth Obs Remote Sens* 14:2222–2235. <https://doi.org/10.1109/JSTARS.2021.3050166>
39. AL-Rousan N, Mat Isa NA, Mat Desa MK, AL-Najjar H (2021) Integration of logistic regression and multilayer perceptron for intelligent single and dual axis solar tracking systems. *Int J Intell Syst* 36(10): 5605–5669. <https://doi.org/10.1002/int.22525>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.



Emre Yildirim is a Ph.D. student at the Electrical-Electronics and Computer Engineering Department of Düzce University, Düzce, Turkey. Since 2019, he has been a member of the Computer Technology Department of Osmaniye Korkut Ata University. His research interests include machine learning, data mining, and big data.



Murtaza Cicioğlu received his Ph.D. degree in Electrical-Electronic and Computer Engineering from Düzce University, Turkey in 2020. He has been a member of the Computer Engineering Department of Bursa Uludağ University, Turkey. His research interests include software-defined networking, wireless communications, 5G, body area networks, cognitive radio.



Ali Çalhan received his M.Sc. and Ph.D. degrees from the University of Kocaeli, Turkey, in 2006 and 2011. Since 2011, he has been a member of the Computer Engineering Department of Düzce University. His research interests are wireless communications, cognitive radio networks, body area networks, and software-defined networks.