



Real-time internet of medical things framework for early detection of Covid-19

Emre Yildirim¹ · Murtaza Cicioğlu² · Ali Çalhan³

Received: 18 February 2022 / Accepted: 28 June 2022 / Published online: 24 July 2022
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2022

Abstract

The Covid-19 pandemic is a deadly epidemic and continues to affect all world. This situation dragged the countries into a global crisis and caused the collapse of some health systems. Therefore, many technologies are needed to slow down the spread of the Covid-19 epidemic and produce solutions. In this context, some developments have been made with artificial intelligence, machine learning and deep learning support systems in order to alleviate the burden on the health system. In this study, a new Internet of Medical Things (IoMT) framework is proposed for the detection and early prevention of Covid-19 infection. In the proposed IoMT framework, a Covid-19 scenario consisting of various numbers of sensors is created in the Riverbed Modeler simulation software. The health data produced in this scenario are analyzed in real time with Apache Spark technology, and disease prediction is made. In order to provide more accurate results for Covid-19 disease prediction, Random Forest and Gradient Boosted Tree (GBT) Ensemble Learning classifiers, which are formed by Decision Tree classifiers, are compared for the performance evaluation. In addition, throughput, end-to-end delay results and Apache Spark data processing performance of heterogeneous nodes with different priorities are analyzed in the Covid-19 scenario. The MongoDB NoSQL database is used in the IoMT framework to store big health data produced in real time and use it in subsequent processes. The proposed IoMT framework experimental results show that the GBTs classifier has the best performance with 95.70% training, 95.30% test accuracy and 0.970 area under the curve (AUC) values. Moreover, the promising real-time performances of wireless body area network (WBAN) simulation scenario and Apache Spark show that they can be used for the early detection of Covid-19 disease.

Keywords Covid-19 diagnosis · Ensemble learning · Real-time analytics · Machine learning · Apache spark

1 Introduction

Covid-19 (SARS-COV2) coronavirus is a new acute respiratory disease that first emerged in Wuhan, China, in the last months of 2019 and then affected the whole world [1, 2]. Covid-19 is transmitted by droplets or contact, causing the death of hundreds of people every day [3]. For this reason, the World Health Organization (WHO)

declared a pandemic on March 11, 2020, because the coronavirus is highly contagious [4]. The uncertainty and anxiety created by the Covid-19 pandemic have led to a global crisis in many sectors, especially in the healthcare field which suffers from overcrowding in hospitals and a shortage of healthcare personnel and equipment. Therefore, technologies with virtual reality (VR) [5], the internet of things (IoT), remote health monitoring systems [6], and artificial intelligence (AI) are used to prevent the spread of the epidemic and to facilitate its resolution. AI with various machine learning and deep learning algorithms can help in the field of health with applications and prevent the spread of the virus or take precautions [7–10].

Advances in IoT have made the processes of detecting, collecting and analyzing data more effective and efficient. The integration of IoT technologies and wireless body area networks (WBANs) and their use in the field of health is

✉ Emre Yildirim
emreyildirim@osmaniye.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

called IoMT. IoMT can be defined as a combination of medical devices and applications that can be connected to healthcare information technology systems using network technologies. IoMT has many benefits such as reducing health costs and increasing the quality of life of individuals by continuously monitoring their health. IoMT with the help of wireless communication technologies such as Wi-Fi, 4G and 5G provides collaboration with various medical devices and circuits such as sensors/actuators designed for physiological data for example respiratory rate, heart rhythm and oxygen content.

IoMT produces big health data using sensors with different characteristics. In order to understand the potential impact of diseases such as Covid-19, the health data produced should be processed in real time with the right techniques to make them meaningful. This situation eases the workload of health personnel and doctors by helping them make decisions. Apache Spark integrates with different systems, receives real-time data, and has the ability to analyze it quickly using different machine learning techniques. In our study, we propose a real-time IoMT framework for the early detection of Covid-19 disease. The IoMT framework has been developed by integrating multiple technologies.

The remainder of this paper is organized as follows. Studies on Covid-19 and ensemble learning are given in Sect. 2. All aspects of the proposed Covid-19 IoMT framework are described in Sect. 3. In Sect. 4, the experimental results are presented. Finally, Sect. 5 describes the discussion and future work.

2 Related works

There are many studies in the literature in which machine learning and deep learning models are used in various areas. For example, Bertolini et al. [11] stated in their studies that machine learning and deep learning are used in many applications in the industry and give positive results. Aafjes-van Doorn et al. [12] examined the effect of machine learning applications in psychotherapy studies. In their study, they stated that machine learning applications have potential opportunities for clinical studies in the field of psychotherapy. In another study, Rafique et al. [13] examined the effect of machine learning on cancer treatment.

In many studies, predictions are usually made using machine learning or deep learning algorithm. Therefore, new methods have emerged in order to give better results for disease prediction. One of these methods is the ensemble learning method. In the ensemble learning method, predictions are not made using a single machine learning or deep learning algorithm. In order to give fairer

and better results in the ensemble learning method, predictions are made by using more than one algorithm at the same time.

Khalaf et al. [14] proposed a new approach to predicting flood severity and water level, using ensemble learning methods to prevent natural disasters such as river flooding. In their proposed approach, long short-term memory (LSTM) and RF classifiers are used together for predictions. Experimental results outperformed other individual models with 81.13% accuracy, 71.4% sensitivity and 85.9% specificity. Cutler et al. [15] used GBTs and RF ensemble models to estimate the occupancy rates of electric vehicle charging stations. The GBTs model showed a 94.8% accuracy and a 0.838 Matthews correlation to be a suitable model for charge-load estimation. Sundareswaran and Lavanya [16] estimated traffic congestion utilizing an ensemble learning method which used multiple deep neural networks. Traffic data from an API is analyzed in Apache Spark to estimate traffic congestion. The results obtained in their study indicated that the ensemble learning method had a positive effect as it improved traffic congestion.

It is seen that the ensemble learning method gives effective results in studies in the field of health. Muhammed et al. [17] proposed an ensemble learning-based system for the detection of preictal status in epileptic seizures. In the study using EEG dataset, predictions were made with convolutional neural networks (CNN), support vector machine (SVM) and LSTM algorithms. In their proposed system, 94.31% accuracy, 94.73% sensitivity and 93.72% precision were obtained. In a similar study, Arora et al. [18] proposed an N-semble-based model in their study to identify genes associated with Parkinson's disease. The model was compared with six different classification methods. The results showed that the proposed model is more effective than other methods with 88.9% precision, 90.9% recall and 89.8 f-score. In other studies, Tuncer et al. [19], Hossein et al. [20] and Ebrahimpour et al. [21] performed ECG arrhythmia beat classification with ensemble learning models they created using ECG signals from the MIT-BIH Arrhythmia database. Sarwar et al. [22] proposed an ensemble model-based expert system for the diagnosis of Type II diabetes in their study. The ensemble model was created with K-nearest neighborhood (KNN), Naive Bayes (NB), artificial neural networks (ANN), and SVM classification models. The proposed expert system estimates the diabetes classification with 98.6% accuracy. Wang et al. [23] and Kadam et al. [24] carried out studies on cancer diagnosis with ensemble learning models created with different algorithms. When the results of the two studies were examined, an accuracy rate of over 97% was obtained. In an IoT-based study, a deep learning-based health framework called DeTrAs has been proposed to help Alzheimer's patients by Sharma et al. [25].

The ensemble learning method has still been used in studies on the Covid-19 pandemic as it gives acceptable results. Ben Yahia et al. [26] presented the ensemble learning method created by LSTM, deep neural networks (DNN) and CNN algorithm to predict the Covid-19 pandemic. They stated that the proposed method can be used with high accuracy rates of 97% and 92%, respectively, in the study using two case studies in China and Tunisia. Tang et al. [27] proposed the EDL-COVID model, and the ensemble deep learning model was used for Covid-19 case detection. Their proposed model consisted of deep learning models with different sensitivities, and these models classify chest X-rays and make predictions by taking the weighted average. It was stated that the prediction results of the EDL-COVID model were promising for Covid-19 case detection with an accuracy rate of 95%. In another study, Biswas et al. [28] developed an ensemble learning method that predicts the diagnosis of Covid-19 using chest images taken with computerized tomography (CT). First of all, CT images were estimated using visual geometry group (VGG-16), ResNet50 and Xception algorithms. Then, predictions were made with the ensemble learning method, which was developed by using these three algorithms together. As a result of the analysis, it was seen that the prediction performance of the model developed with 98.79% accuracy and 99% F1-score was better than other individual algorithms. Kedia et al. [29] aimed to predict Covid-19 disease from X-ray images with the ensemble model they named CovNet-19. In the model they developed, VGG-19, DenseNet-121 and SVM classifiers were used together. The model has shown that it can be used to predict Covid-19 disease with an accuracy rate of 98.33%. In other similar studies, Foysal and Hossain [30] predicted the Covid-19 disease using CT images with 96% accuracy with the ensemble model, in which CNN classifiers were used together. Siswantining and Parlindungan [31] estimated the diagnosis of Covid-19 with 95% accuracy using X-ray images with the ensemble model developed with ANN, CNN and SVM classifiers. Li et al. [32] aimed to predict the Covid-19 disease with a deep ensemble learning method consisting of VGG-16 algorithms. As a result of their study, the proposed model showed that it is a good classification model with 93.57% accuracy, 94.21% sensitivity, 93.93% specificity, and 89.40% precision prediction performance.

A summary of the proposed study and the comparative analysis of related studies is given in Table 1. Studies based on Covid-19 and other diseases are given in [17–32]. As a result, when the studies in the literature are examined, deep learning methods are generally used in ensemble models applied for Covid-19. In addition, the data in the studies are taken from the ready data set. In our study, an IoMT framework consisting of WBANs is proposed for the

prediction of Covid-19 disease. Within the framework of IoMT, a Covid-19 scenario is developed with WBANs in the Riverbed Modeler simulation software. Real-time health data are generated from this simulation scenario. All technologies used have been run simultaneously. Real-time health data are analyzed in Apache Spark data processing engine using Node-Red and Apache Kafka data flow technologies and predictions are made. In addition, in the proposed framework, RF and GBTs algorithms from ensemble learning methods are used to increase the prediction rate of the Covid-19 disease, and their prediction performances are compared. In addition, both WBAN simulation scenario and Apache Spark real-time performance analysis are examined.

The contributions of the proposed study can be summarized as follows:

- GBTs and RF ensemble learning classifiers are used for Covid-19 disease prediction.
- Within the framework of Covid-19 IoMT, real-time health data are obtained from Riverbed Modeler WBAN scenarios for data analysis.
- The throughput and delay results of heterogeneous nodes (with different priorities) in the Covid-19 WBAN scenario are analyzed.
- Open-source Apache Spark technology is used in the analysis of real-time data, Node-Red and Apache Kafka technologies are used in the flow of data, and MongoDB NoSQL database is used in data storage.
- Real-time data processing performance with Apache Spark has been examined.

3 The IoMT proposed framework

In the proposed framework, health data are produced in the Covid-19 WBAN simulation scenario. Obtained health data are sent simultaneously to the Apache Spark data analysis system using Node-Red and Apache Kafka data flow platforms. In the Apache Spark data analysis system, disease predictions are made with the real-time health data produced using the ensemble learning method. In addition, real-time data from the simulation scenario and Covid-19 prediction data are stored in the system using the MongoDB NoSQL database. Figure 1 shows the overall architecture of the proposed IoMT framework.

The concept of real time is considered as transferring the data between cloud and IoMT systems with the lowest delay and accessing instant and accurate data from the users via remote monitoring systems. Therefore, Apache Spark technology is used to meet these real-time QoS expectations in the scenarios. Apache Spark real-time data

Table 1 Comparative analysis of the related work

Studies	Type of disease	IoT	Ensemble model	Apache spark	Real time
Muhammed et al. [17]	Epilepsy	x	✓	x	x
Arora et al. [18]	Parkinson	x	✓	x	x
Tuncer et al. [19]	Heart Disease	x	✓	x	x
Hossain et al. [20]					
Ebrahimpour et al. [21]					
Sarwar et al.[22]	Diabetes	x	✓	x	x
Wang et al. [23]	Cancer	x	✓	x	x
Kadam et al. [24]					
Sharma et al.[25]	Alzheimer	✓	✓	x	x
Ben Yahia et al. [26]	Covid-19	x	✓	x	X
Tang et al. [27]					
Biswas et al. [28]					
Kedia et al. [29]					
Foysal and Hossain [30]					
Siswantining and Parlindungan [31]					
Li et al. [32]					
Our proposed study	Covid-19	✓	✓	✓	✓

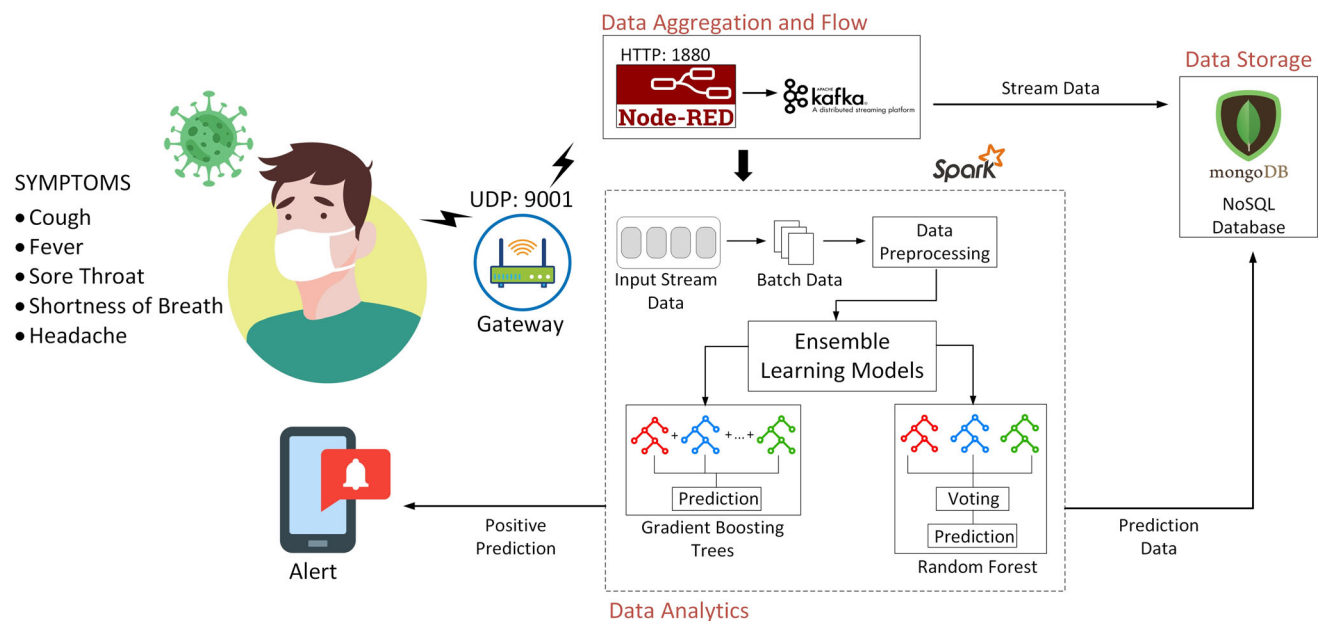


Fig. 1 Overall architecture of the proposed IoMT framework

processing performances has been added to analyze this situation in our study.

In our study, we consider real-time detection of the Covid-19 disease, which has caused the death of thousands of people. To detect this disease, we use the most common symptoms of cough, fever, sore throat, shortness of breath and headache [33]. The various symptoms in the scenario and the value ranges of the data produced for those

symptoms are given in Table 2. Health data of symptoms can be obtained with various sensors and analyzed [34, 35]. Therefore, in our study, the Covid-19 WBAN scenario based on the IEEE 802.15.6 standard is developed using the Riverbed Modeler in order to produce real-time data on these symptoms.

In our study, the data measured by the WBAN sensors was adapted to the dataset from Israel. Detailed

Table 2 Covid-19 WBAN scenario sensor values

Sensor	Symptoms	Interval
N0	Cough	0, 1
N1	Fever	0, 1
N2	Sore throat	0, 1
N3	Shortness of breath	0, 1
N4	Headache	0, 1

information about the database is shown in Table 3. The data set consists of real data and verbal statements from patients who came to the hospital or health institution with the suspicion of Covid-19 [36]. The data measured by each sensor represents the relevant input values in the data set. For example, it is assumed that one of the sensors was measuring breath because it is working in a simulation environment [37].

In the scenario, there are various numbers of end nodes and coordinators (HUBs) where data from these nodes are collected. A representative view of the Covid-19 WBAN scenario developed on Riverbed Modeler is shown in Fig. 2.

Real-time health data produced in the Covid-19 WBAN scenario is collected in the data collection and flow component and sent to the relevant components. The purpose of this component is to transfer data generated in real time as fast as possible and without loss. Two different technologies are used in this component. First of all, in order to receive the data from the WBAN scenario, the flow structure is created as in Fig. 3 on the Node-Red flow platform running on the 1880 port. In this flow structure, the real-time health data produced are taken from the UDP 9001 port and translated into JSON format and sent to the Apache Kafka platform. Incoming data to Apache Kafka is quickly and lossless transferred to data analytics and data storage components. In addition, Apache Kafka can keep data up to the last 48 h for interruptions that may occur in

the system. The detailed structure of the Apache Kafka platform is shown in Fig. 4.

Apache Kafka generally consists of three different components: producer, consumer and topic. “Producers” send data to Apache Kafka. “Topic” keeps the incoming data in different categories. “Consumers” attract and use the data held in different topics to their systems by subscribing to these topics. In addition, unlike these three components, the Apache Zookeeper component manages this flow in Apache Kafka. In this study, the Node-Red streaming platform is used as a producer and sends real-time health data to Apache Kafka. Then, the data in Apache Kafka is kept in the “Covid19Data” topic. Apache Spark as a data analysis component, and MongoDB database, which stores incoming real-time data, subscribe to the “Covid-19Data” topic and receive the data on their systems.

Apache Spark provides a framework for distributed architecture and data analysis and it is reliable, fast, highly fault-tolerant, and capable of handling very large datasets [38]. Spark can perform fast analysis by performing in-memory computation instead of accessing the hard disk. When processing big data, it works by distributing the workload in parallel instead of executing operations serially [39]. In addition, Spark supports rich high-end libraries, including Spark SQL for SQL, Machine Learning Library (MLlib) for machine learning, GraphX for graph processing and Spark Streaming. Spark’s ecosystem is shown in Fig. 5.

In this study, Apache Spark’s Spark Streaming and MLlib libraries are used for predictive analysis of Covid-19 disease. With Spark Streaming, health data are obtained in real time by subscribing from Apache Kafka topic. In the proposed system, data from Apache Kafka are collected and processed every 10 s.

The Covid-19 disease decision-making process in the study is shown in Fig. 6. The Covid-19 database from the Israel state database was used for the ensemble learning model in the decision-making process of the disease. In the

Table 3 Details of Covid-19 data from the Israel state database

Feature	Interval	Total number of records
Test date	Date	278,848
Cough	True, False	
Fever	True, False	
Sore throat	True, False	
Shortness of breath	True, False	
Headache	True, False	
Age 60 and above	None, No, Yes	
Gender	Male, female, none	
Test indication	Abroad, contact with confirmed, other	
Corona result	Negative, positive, other	

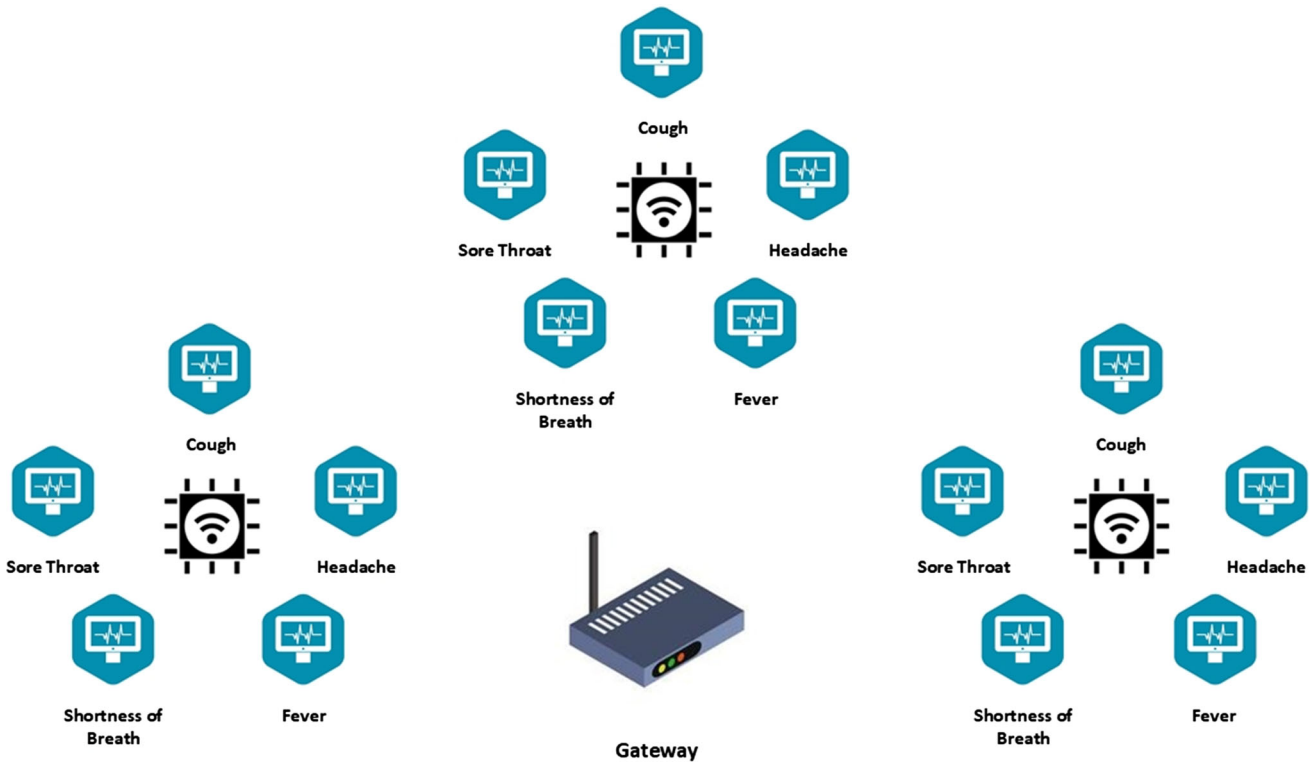


Fig. 2 IoMT representative simulation scenario

Fig. 3 Node-Red data flow structure

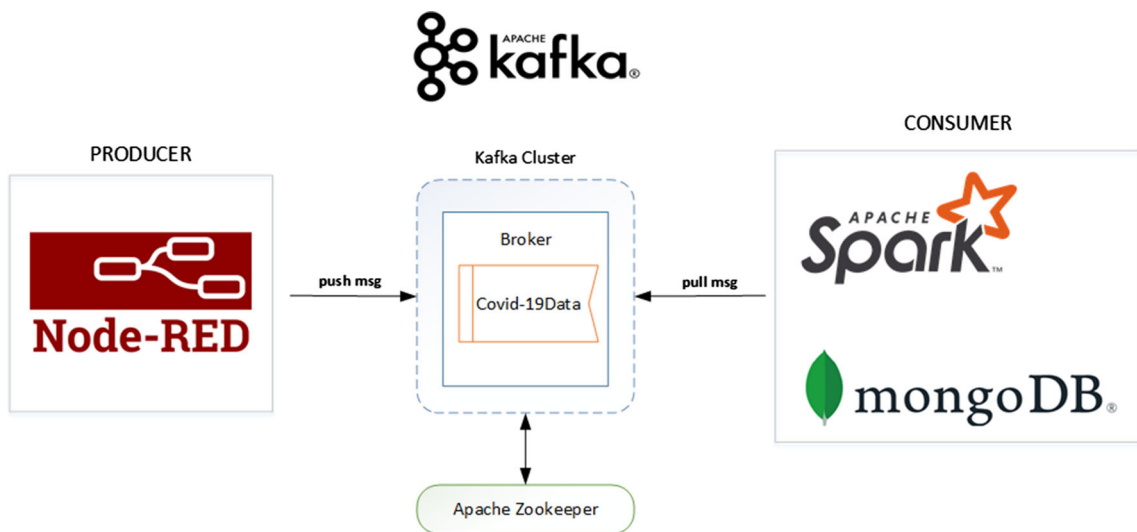
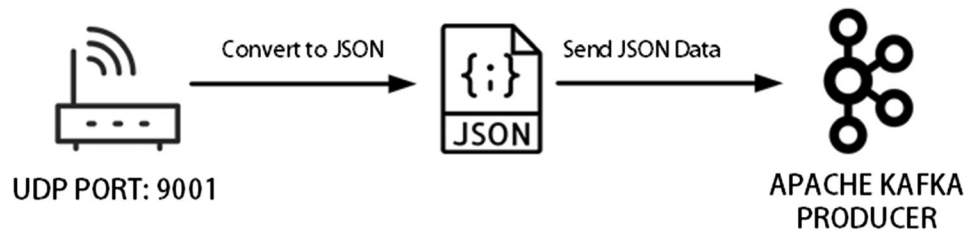


Fig. 4 Apache Kafka data flow structure

Fig. 5 Apache spark ecosystem

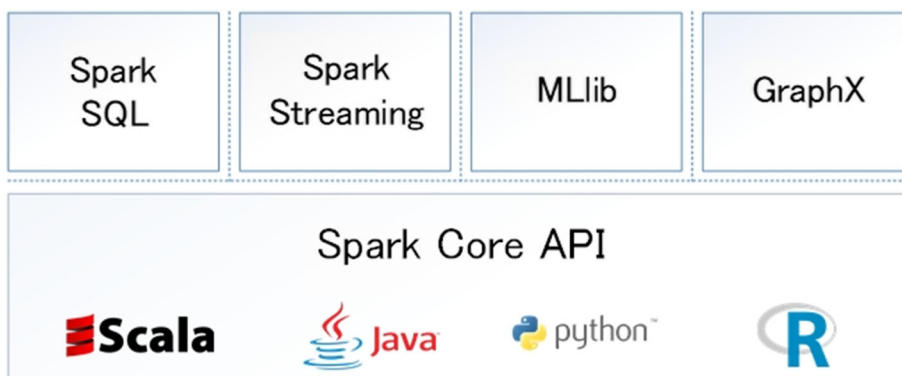
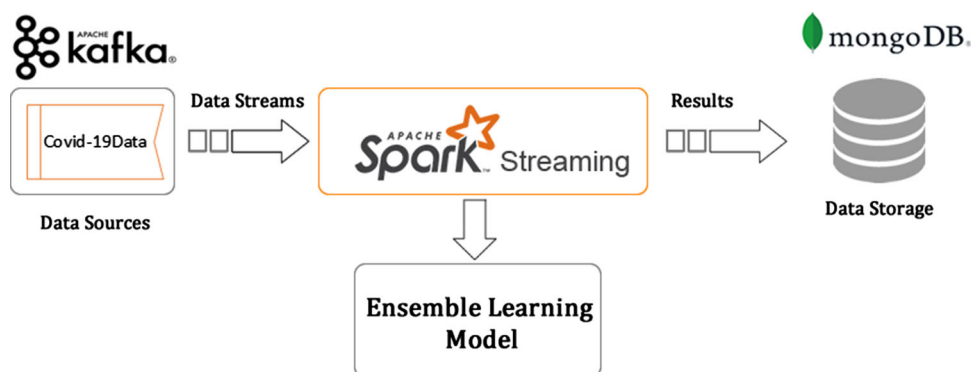


Fig. 6 Apache spark data analytics process



database, there are a total of 278,848 records belonging to two classes: Covid19 patients and normal class. Empty and incorrect data are discarded during the data preprocessing phase, while the prediction model is being created. In addition, the number of data for the two classes is equalized so that the model will give more accurate results. For disease predictions, RF and GBTs ensemble learning classifiers from the MLlib library are used. The stages of constructing ensemble learning prediction models are shown in Fig. 7. In addition, the parameters of the RF and

GBT algorithms that make up the ensemble model in our study are given in Table 4.

After the creation of the ensemble learning model, Covid-19 disease prediction is performed in real time with five different WBAN health data from the simulation scenario. It is clearly stated in studies that identifying symptoms is the most important factor in the determination of the Covid-19 disease [40, 41].

The ensemble learning method is based on the theory that the collective generalization ability of a group will be

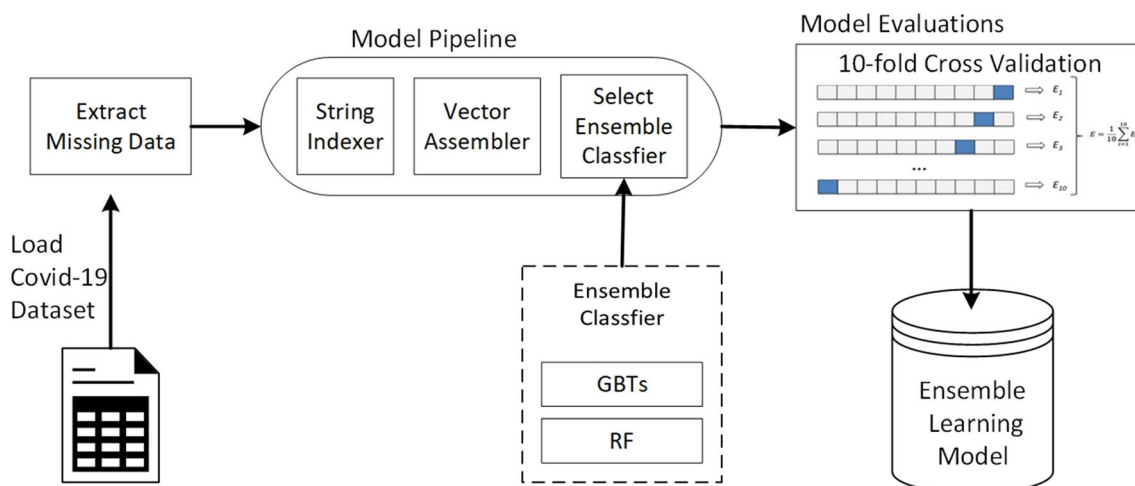


Fig. 7 Stages of ensemble learning prediction models

Table 4 Ensemble model parameters

Random forest	Gradient boosting trees
max_depth = 3	max_depth = 3
max_bins = 10	max_bins = 10
num_trees = 10	num_iter = 10
impurity = “gini”	impurity = “gini”

much stronger than the individual. In other words, ensemble learning is a machine learning paradigm in which multiple students are trained to solve problems [42]. This method provides a more powerful classifier by combining the results of more than one classifier to improve the overall classification results. GBTs are used for classification and regression problems such as decision trees and are one of the ensemble learning boosting methods. GBTs are generated sequentially from many decision trees, and each decision tree is built to minimize and strengthen the error rate of the previous structure [43]. RF classifier is an ensemble learning method based on bagging [44]. The RF classifier creates multiple decision trees and combines them to get a more accurate and stable prediction. Instead of branching each node using the best branch among all the variables, RF branches each node using the best randomly selected variables at each node.

MongoDB NoSQL database can store large data and is used in the Data Storage component. The real-time original health data from the Covid-19 WBAN scenario to the Apache Kafka topic are obtained by subscribing. In addition, the Covid-19 disease prediction results made in the Apache Spark data analysis system are stored in this component. Incoming data in real time are stored in JSON format. The data stored in MongoDB can then be used in visualization applications.

As shown in Fig. 8, the users of the proposed IoMT framework can be patients, healthcare providers, doctors or healthcare professionals. Real-time health data are obtained and analyzed by various sensors in WBAN simulation scenarios on the user, and if there is a suspicion of Covid-19, the patient is notified and asked to go to the nearest health institution. The user is not disturbed in normal situations. On the other hand, the patient health prediction is kept in the relevant databases. If necessary, this data is used by the health institution in accordance with the confidentiality principle.

4 Performance evaluation metrics

Five performance metrics as accuracy, precision, recall, F1-Score and receiver operating characteristic (ROC) curve are used to evaluate the performance of ensemble

learning classifiers [45]. Confusion matrix and cross-validation methods can be used to calculate these performance criteria. Confusion matrix is used to visualize the performance of classification algorithms for binary class by creating a 2×2 matrix. Each row in the matrix shows the number of predicted classes, and each column shows the number of actual classes. As a result of the matrix created, four values are formed: True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN). The formula for the performance criteria calculated using these values is also presented in Eqs. (1)–(4).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (1)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3)$$

$$\text{F1 - score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

Cross-validation is a statistical method that can measure the performance of ensemble learning classifiers. In this method, the data set is divided into parts with a predetermined k value. In this study, the k value is determined as 10 ($k = 10$). In the method, the classification process is performed 10 times, and at each step, one of the divided parts is separated for the test process. The remaining nine are used for training the classifier. The overall result is obtained by taking the average of the classification results obtained after ten steps.

5 The results

For testing the proposed IoMT system, experiments are performed on a computer with AMD Ryzen 5 3500X processor and 16 GB RAM. The data set in the Covid-19 disease decision-making process is used for 80% training and 20% testing purposes. In addition, complexity matrix and tenfold cross-validation are used for training and testing the accuracy of ensemble learning classifiers.

Figure 9 shows the confusion matrices for the test set of ensemble learning classifiers. The GBTs classifier is observed to classify almost all samples in the Covid-19 patient class with 100% accuracy. It correctly classifies 92% of normal class samples. It is observed that the RF classifier correctly classifies 98% of the samples in the Covid-19 patient class and 89% of the samples in the normal class.

Figure 10 shows the training and test accuracy rates of ensemble learning classifiers. In this regard, we can conclude that GBTs have the highest accuracy among the

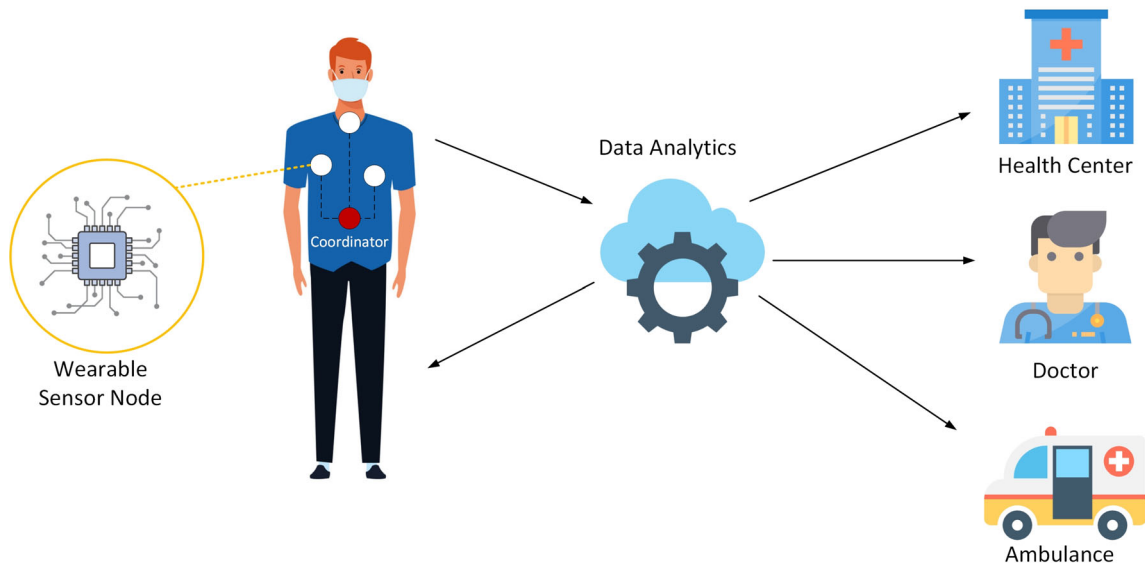


Fig. 8 Use case diagram of the proposed framework

Fig. 9 Complexity matrix of ensemble learning classifiers

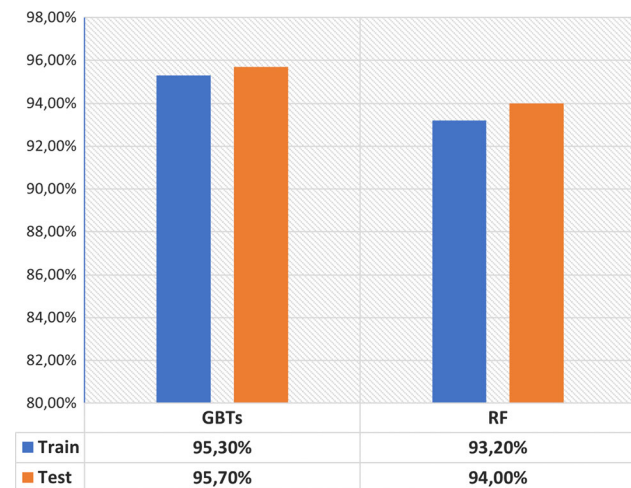
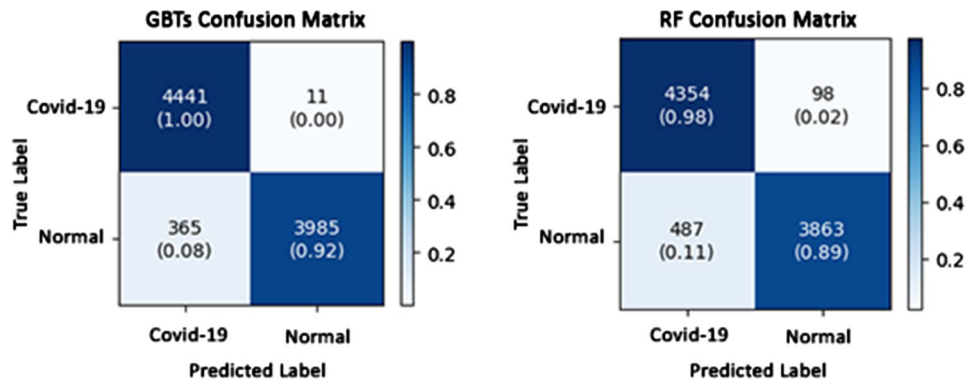


Fig. 10 Training and testing accuracy rates of ensemble learning classifiers

classifiers with 95.30% training and 95.70% testing accuracy rates. The RF classifier has 93.20% training and 94.00% test accuracy, respectively.

Figure 11 shows the precision, recall and F1-score rates of ensemble learning classifiers. For the class with Covid-19 patients, the GBTs classifier has 92% precision, 100% recall and 96% F1-score, respectively. The RF classifier has 90% precision, 100% recall and 94% F1-score, respectively. For the non-ill or normal class, the GBTs classifier has 100% precision, 92% recall and 95% F1-score. The RF classifier has 100% precision, 88% recall and 94% F1-score, respectively. In this context, we can conclude that the GBTs classifier has the highest precision, recall and F1-score values for both the Covid-19 class and the normal class.

Figure 12 shows the ROC curve of ensemble learning classifiers. The ROC curve is developed by plotting the true positive rate (TPR) versus the false positive rate (FPR), which is the area under the curve (AUC). The closer the

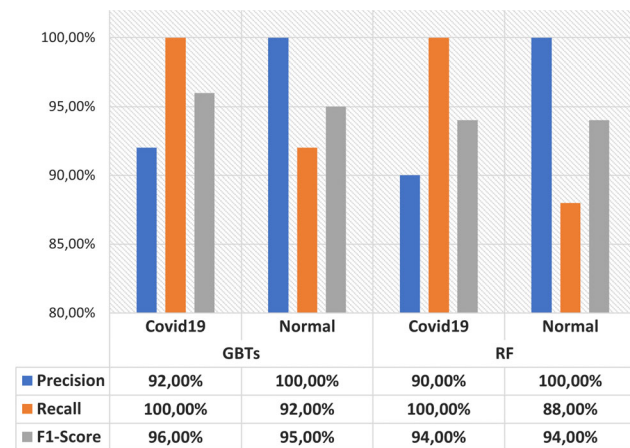


Fig. 11 Precision, recall and F1-score values of ensemble learning classifiers

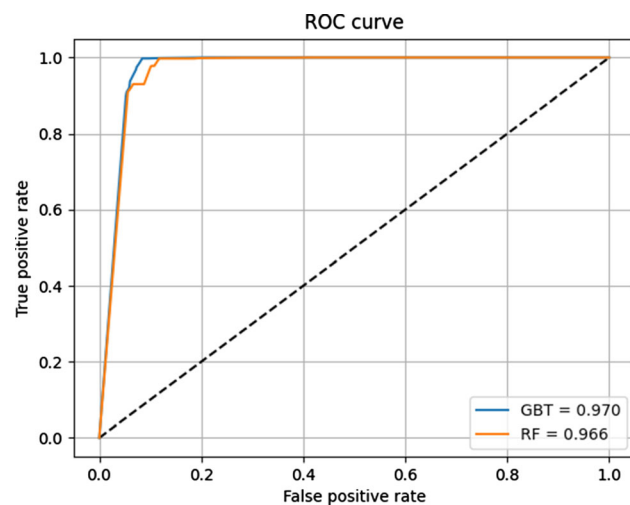


Fig. 12 ROC curve of ensemble learning classifiers

AUC value is to 1, the closer it is to a good classifier. In this context, the AUC value for the GBTs classifier is 0.970, and the AUC value for the RF classifier is 0.966. Although the AUC values for GBTs and RF classifier are very close, we can conclude that the GBTs classifier has a higher score.

A representative view of the simulation scenario designed in the Riverbed Modeler simulation software for the IoMT scenario is given in Fig. 2. Two different scenarios are designed for the performance analysis of the proposed real-time IoMT architecture. In the first scenario, there are five sensors each with different priorities that measure cough, fever, sore throat, shortness of breath and headache levels on three different people and three coordinators (HUB) for coordinating these sensors. The priorities of these sensors are listed as fever, shortness of breath, sore throat, headache and cough, from the highest to lowest. In this way, it is aimed to reach the target with lower

delay and lossless data with high priority. The HUB is responsible for transmitting the packets it receives from the sensors connected to it to the gateway and managing the medium access of sensors. In the gateway, packets, which come with the help of socket programming, are transmitted to cloud software prepared for machine learning. CSMA/CA-based IEEE 802.15.6 standard is used for the communication between the HUB and the sensors. Detailed information about this standard is available in our previous studies [46] and the relevant standard (IEEE technical paper) [47]. There is no communication requirement between HUBs for this scenario.

In the second scenario, the number of nodes and network traffic are increased and priorities of nodes are equalized. The aim is to measure the data processing analysis of Kafka's Spark in heavy network traffic. In this scenario, each sensor has a packet inter-arrival time of 0.1 packet/sec. The approximate packet inter-arrival time for each sensor has been increased 10 times. HUBs transmit all received packets to the gateway. The gateway sends these packets to Node-Red with the help of socket programming. Detailed information about all scenarios is given in Table 5.

Figures 13 and 14 show the performance results in the first scenario. In Fig. 13, the throughput cumulative distribution function (CDF) values are given for the fever, shortness of breath, cough, sore throat and headache sensors in three different individuals. The empirical CDF results of these sensors are calculated for the same simulation scenario. These values allow us to examine the throughput distribution among the sensors. As can be seen in the figure, the throughput CDF value of the sensors with high priority reaches faster. There is no obvious difference due to the small number of sensors in the environment. However, it is observed that the throughput CDF values of vital data of individuals such as fever and shortness of breath reached their maximum more quickly. These results obtained at the gateway are very important for the early detection of patients with Covid-19 disease.

Figure 14 shows the end-to-end delay results between different sensors in the scenario. The delays between the packets sent by the sensors created for five different attributes used in machine learning algorithms to the gateway are discussed. As stated above, different priorities have been established among the sensors. These priorities allow the relevant sensor to send its data to the target more quickly without conflicting with other sensors. When Fig. 14 is examined, it has been observed that the delays of vital data such as fever and shortness of breath, which have high priority, are lower than other data. Subsequently, delays are observed to increase for sore throat, headache and cough. The result only covers the performances in the IoMT architecture. Most important is the early diagnosis of

Table 5 Simulation parameters

Parameter	Value
Simulation time	3600 s
Frequency	2400–2483.5 GHz
Number of nodes	20–60
MAC protocol	CSMA/CA-based IEEE 802.15.6
Priorities	0–5 (low to high)
Bandwidth	1 MHz
Data rate	971.4 kbps
Packet size	100 Byte
<i>Packet inter-arrival time (exponential distribution function)</i>	
Scenario 1 (for low traffic)	(Fever, shortness of breath, cough, sore throat, headache) 1 packet/sec
Scenario 2 (for high traffic)	(Fever, shortness of breath, cough, sore throat, headache) 0.1 packet/sec

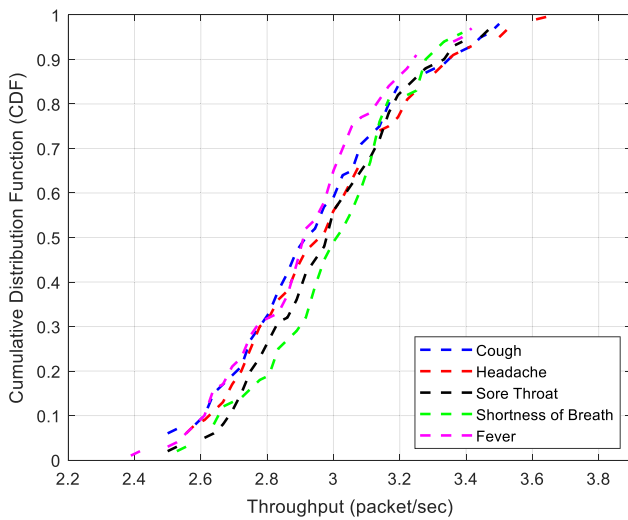


Fig. 13 Throughput cumulative distribution function (CDF)

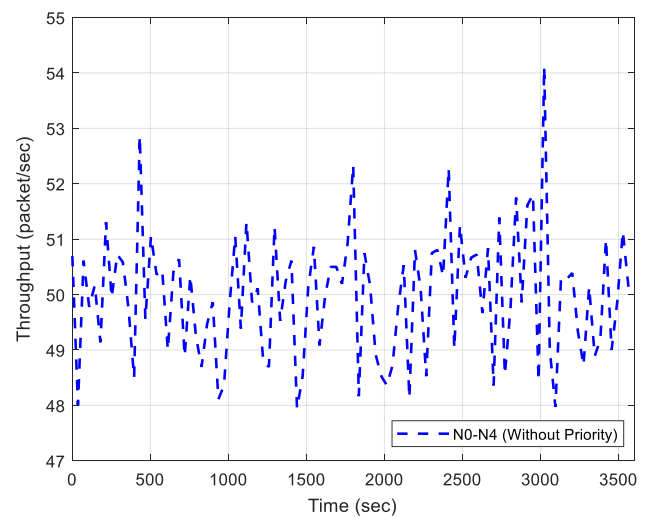


Fig. 15 Throughput results for IoMT scenario

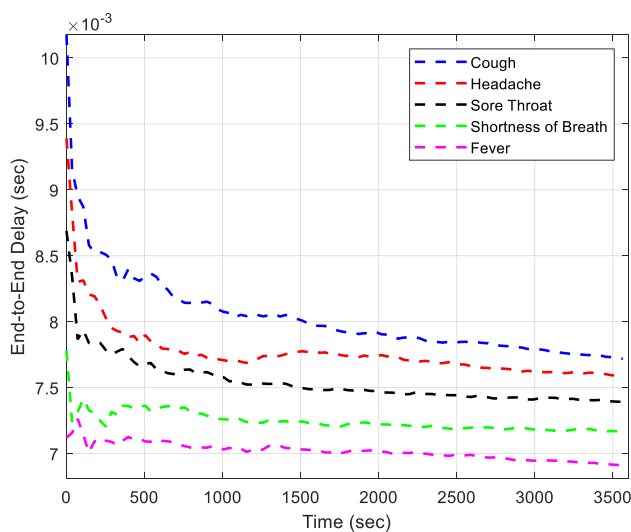


Fig. 14 End-to-end delay results for IoMT scenario

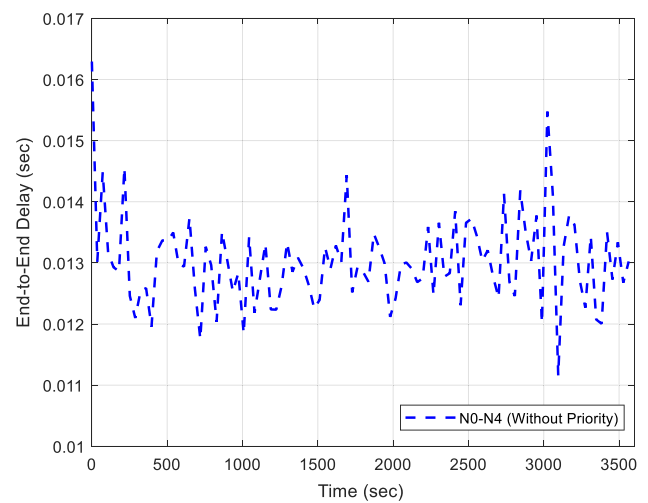


Fig. 16 End-to-end delay results for IoMT scenario

Streaming Statistics

Running batches of 10 seconds for 3 minutes 27 seconds since 2022/02/08 00:59:58 (21 completed batches, 26683 records)

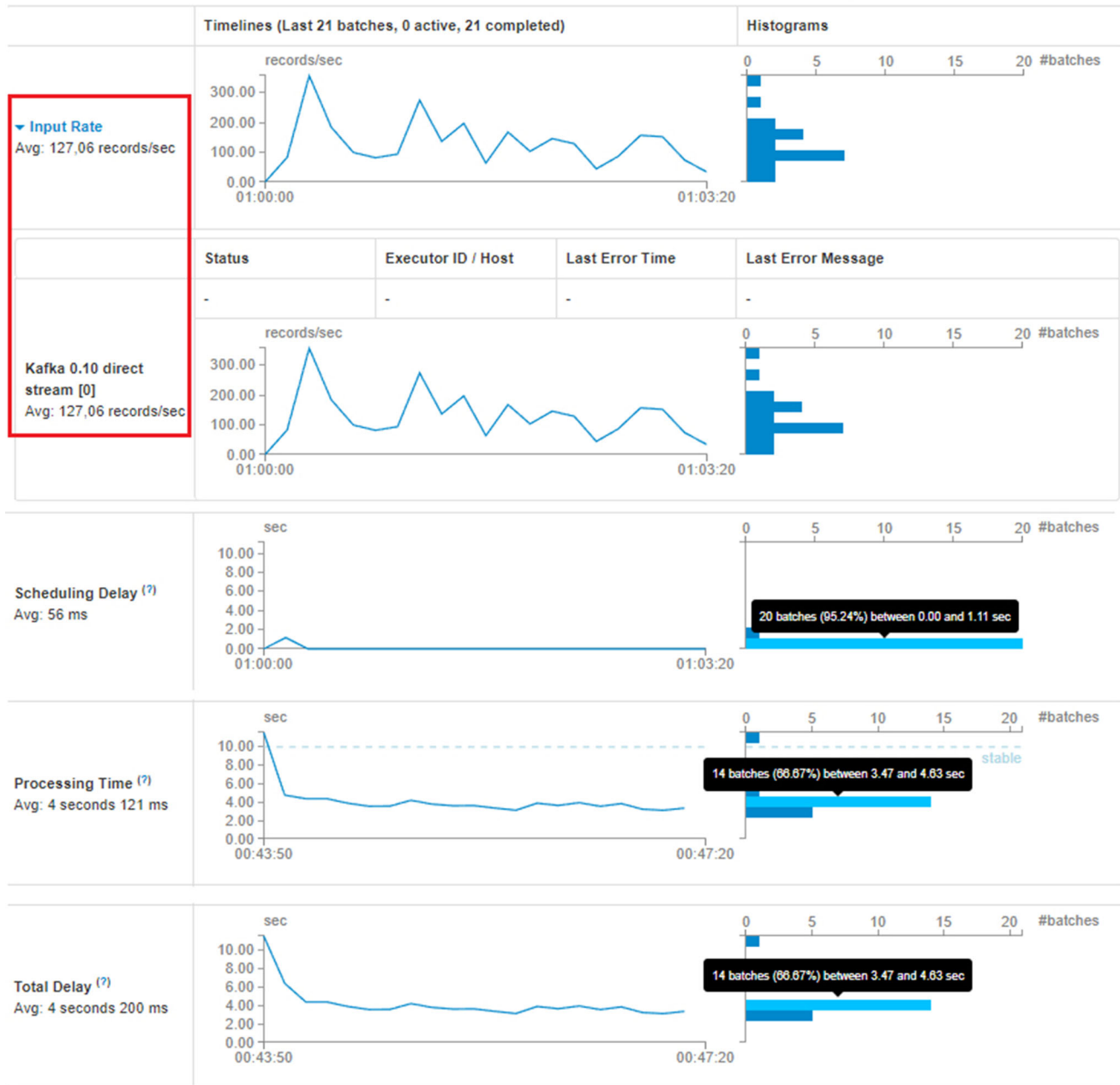


Fig. 17 Apache spark real-time data processing performances

Covid-19 disease which can be achieved quickly using this data obtained with the help of machine learning after the gateway.

Figures 15 and 16 show the performance results in the second scenario where the priorities are equalized, and there is heavy network traffic to increase the contention in medium access. Figure 15 shows the throughput results of this scenario. Since all nodes have equal priorities, the total

throughput results are found to be approximately 50 packets/sec. This result shows that all packets in the environment are successfully forwarded to the destination. In Fig. 16, the end-to-end delays caused by the increase in the packet generation rate and the fixation of the contention window are shared. Delay is about 0.013 s. According to the delay results in the first scenario, the reason for the increase is that the contention window is kept constant for

all nodes, the medium contention is increased and the packet generation rates are increased. The most crucial purpose of this scenario is to evaluate the real-time data processing performance of the proposed architecture within heavy network traffic. These results are explained in Fig. 17.

Apache Spark's real-time data processing performance of the data generated in the WBAN simulation scenario and received from Apache Kafka is shown in Fig. 17. For Apache Spark's real-time performance, data from Apache Kafka is aggregated and analyzed every 10 s by Spark Streaming. Real-time health data are processed 21 times in total for 3 min 27 s in batch mode. Apache Spark processes a total of 26,683 real-time data from the WBAN simulation scenario during data processing. For Apache Spark performance, input rate, scheduling delay, processing time and total delay values are examined. In this context, when the input rate value is examined, an average of 127.06 data per second is obtained from Apache Kafka. When the Kafka stream value is examined, it is seen that it is taken from Apache Kafka in a lossless way. When the data processing times of Apache Spark are examined, scheduling delay shows the delay in data processing, processing time shows the data processing time and total delay shows the sum of these two data. The average of these values is given in Fig. 15, and Apache Spark has an average of 56 ms delay and 4 s 121 ms data processing, with a total of 4 s 200 ms real-time data processing performance. As a result, it has been observed that Apache Spark is a very fast technology for data processing in real time.

6 Conclusion

In this study, an IoT-based real-time IoMT framework has been proposed in order to reduce the impact of deadly diseases such as Covid-19 and to assist healthcare personnel. The proposed IoMT framework consists of real-time data sources with WBANs in the Covid-19 scenario, data analytics consisting of Node-Red and Apache Kafka data flow platform, Apache Spark Streaming and MLlib library (prediction model with GBTs and RF classifier) and MongoDB data storage component. We have discussed in detail the IoMT framework with all its components. The GBTs classifier has the best performance rate with 95.70% and 95.30% accuracy in training and test performance, respectively. In addition, the GBTs classifier has the highest values for Covid-19 patients and normal class with 100% precision, 92% recall and 95% F1-score. Although the AUC values of both classifiers are very close, GBTs has the best AUC value ($AUC_{GBTs} = 0.970$, $AUC_{RF} = 0.966$). Finally, the proposed IoMT framework can forward positive cases to healthcare institutions quickly.

In future studies, it is planned to deal with radiological images in the field of health, and to analyze problems in different fields such as industry, sports and entertainment outside the field of health, with the Apache Spark distributed computing platform developed. However, studies can also be carried out to address the privacy, security and distributed computing problem in IoT applications.

Declarations

Conflict of interest Authors declare that they have no conflict of interest.

References

1. Wu F, Zhao S, Yu B et al (2020) A new coronavirus associated with human respiratory disease in China. *Nature* 579:265–269. <https://doi.org/10.1038/s41586-020-2008-3>
2. Nicogossian A (2012) In the news. *World Med Heal Policy* 4:2020. <https://doi.org/10.1515/1948-4682.1230>
3. Li P, Fu JB, Li KF et al (2020) Transmission of COVID-19 in the terminal stages of the incubation period: a familial cluster. *Int J Infect Dis* 96:452–453. <https://doi.org/10.1016/j.ijid.2020.03.027>
4. Organization WH (2019) WHO director-general's opening remarks at the media briefing on Covid-19. <https://www.who.int/dg/speeches/detail/who-director-generals-opening-remarks-at-the-media-briefing-on-covid-19—11-march-2020>. Accessed 11 Mar 2019
5. Singh RP, Javaid M, Kataria R et al (2020) Significant applications of virtual reality for COVID-19 pandemic. *Diabetes Metab Syndr Clin Res Rev* 14:661–664. <https://doi.org/10.1016/j.dsx.2020.05.011>
6. Bahl S, Singh RP, Javaid M et al (2020) Telemedicine technologies for confronting covid-19 pandemic: a review. *J Ind Integr Manag* 5:547–561. <https://doi.org/10.1142/S2424862220300057>
7. Haleem A, Javaid M, Singh RP, Suman R (2021) Applications of artificial intelligence (AI) for cardiology during COVID-19 pandemic. *Sustain Oper Comput* 2:71–78. <https://doi.org/10.1016/j.susoc.2021.04.003>
8. Sorantin E, Grasser MG, Hemmelmayr A et al (2021) The augmented radiologist: artificial intelligence in the practice of radiology. *Pediatr Radiol*. <https://doi.org/10.1007/s00247-021-05177-7>
9. Holzinger A, Weippl E, Tjoa AM, Kieseberg P (2021) Digital transformation for sustainable development goals (sdgs)-a security, safety and privacy perspective on ai. In: *International cross-domain conference for machine learning and knowledge extraction*. Springer, pp 1–20
10. Ertugrul OF, Emrullah A, Oztekin A, Aldemir E (2021) Detection of Covid-19 from X-ray images via ensemble of features extraction methods employing randomized neural networks. *Eur J Tech* 11:248–254. <https://doi.org/10.36222/ejt.1035007>
11. Bertolini M, Mezzogori D, Neroni M, Zammori F (2021) Machine learning for industrial applications: a comprehensive literature review. *Expert Syst Appl* 175:114820. <https://doi.org/10.1016/j.eswa.2021.114820>
12. Aafjes-van Doorn K, Kamsteeg C, Bate J, Aafjes M (2021) A scoping review of machine learning in psychotherapy research. *Psychother Res* 31:92–116. <https://doi.org/10.1080/10503307.2020.1808729>

13. Rafique R, Islam SMR, Kazi JU (2021) Machine learning in the prediction of cancer therapy. *Comput Struct Biotechnol J* 19:4003–4017. <https://doi.org/10.1016/j.csbj.2021.07.003>
14. Khalaf M, Alaskar H, Hussain AJ et al (2020) IoT-enabled flood severity prediction via ensemble machine learning models. *IEEE Access* 8:70375–70386. <https://doi.org/10.1109/ACCESS.2020.2986090>
15. Hecht C, Figgenger J, Sauer DU (2021) Predicting electric vehicle charging station availability using ensemble machine learning. *Mach Learn* 14:7834. <https://doi.org/10.3390/en14237834>
16. Sundareswaran A, Lavanya K (2020) Real-time vehicle traffic prediction in apache spark using ensemble learning for deep neural networks. *Int J Intell Inf Technol* 16:19–36. <https://doi.org/10.4018/IJIT.2020100102>
17. Muhammad S, Khalid S, Jabbar S, Bashir S (2021) Detection of preictal state in epileptic seizures using ensemble classifier. *Epilepsy Res* 178:106818. <https://doi.org/10.1016/j.eplepsyres.2021.106818>
18. Arora P, Mishra A, Malhi A (2021) N-semble-based method for identifying Parkinson's disease genes. *Neural Comput Appl.* <https://doi.org/10.1007/s00521-021-05974-z>
19. Tuncer T, Dogan S, Pławiak P, Rajendra Acharya U (2019) Automated arrhythmia detection using novel hexadecimal local pattern and multilevel wavelet transform with ECG signals. *Knowl Based Syst* 186:104923. <https://doi.org/10.1016/j.knosys.2019.104923>
20. Hossain MB, Bashar SK, Walkey AJ et al (2019) An accurate QRS complex and P wave detection in ECG signals using complete ensemble empirical mode decomposition with adaptive noise approach. *IEEE Access* 7:128869–128880. <https://doi.org/10.1109/ACCESS.2019.2939943>
21. Ebrahimpour R, Sadeghnejad N, Sajedin A, Mohammadi N (2013) Electrocardiogram beat classification via coupled boosting by filtering and preloaded mixture of experts. *Neural Comput Appl* 23:1169–1178. <https://doi.org/10.1007/s00521-012-1063-6>
22. Sarwar A, Ali M, Manhas J, Sharma V (2020) Diagnosis of diabetes type-II using hybrid machine learning based ensemble model. *Int J Inf Technol* 12:419–428. <https://doi.org/10.1007/s41870-018-0270-5>
23. Wang H, Zheng B, Yoon SW, Ko HS (2018) A support vector machine-based ensemble algorithm for breast cancer diagnosis. *Eur J Oper Res* 267:687–699. <https://doi.org/10.1016/j.ejor.2017.12.001>
24. Kadam VJ, Jadhav SM, Vijayakumar K (2019) Breast cancer diagnosis using feature ensemble learning based on stacked sparse autoencoders and softmax regression. *J Med Syst.* <https://doi.org/10.1007/s10916-019-1397-z>
25. Sharma S, Dudeja RK, Aujla GS et al (2020) DeTrAs: deep learning-based healthcare framework for IoT-based assistance of Alzheimer patients. *Neural Comput Appl.* <https://doi.org/10.1007/s00521-020-05327-2>
26. Ben Yahia N, Dhiaeddine Kandara M, Bellamine BenSaoud N (2022) Integrating models and fusing data in a deep ensemble learning method for predicting epidemic diseases outbreak. *Big Data Res* 27:100286. <https://doi.org/10.1016/j.bdr.2021.100286>
27. Tang S, Wang C, Nie J et al (2021) EDL-Covid: ensemble deep learning for Covid-19 case detection from chest X-ray images. *IEEE Trans Ind Inform* 17:6539–6549. <https://doi.org/10.1109/TII.2021.3057683>
28. Biswas S, Chatterjee S, Majee A et al (2021) Prediction of covid-19 from chest ct images using an ensemble of deep learning models. *Appl Sci.* <https://doi.org/10.3390/app11157004>
29. Kedia P, Anjum KR (2021) CoVNet-19: a deep Learning model for the detection and analysis of Covid-19 patients. *Appl Soft Comput* 104:107184. <https://doi.org/10.1016/j.asoc.2021.107184>
30. Foysal M, Aowlad Hossain ABM (2021) Covid-19 detection from chest CT images using ensemble deep convolutional neural network. In: 2021 2nd international conference for emerging technology INCET, pp. 3–8. <https://doi.org/10.1109/INCET51464.2021.9456387>
31. Siswantining T, Parlindungan R (2021) Covid-19 classification using X-ray imaging with ensemble learning. *J Phys Conf Ser.* <https://doi.org/10.1088/1742-6596/1722/1/012072>
32. Li X, Tan W, Liu P et al (2021) Classification of Covid-19 chest CT images based on ensemble deep learning. *J Healthc Eng.* <https://doi.org/10.1155/2021/5528441>
33. Symptoms (2022). <https://www.covid19.act.gov.au/stay-safe-and-healthy/symptoms-and-getting-tested/symptoms-of-covid-19>
34. 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>
35. Awotunde JB, Ajagbe SA, Idowu IR, Ndunagu JN (2021) An enhanced cloud-iiomt-based and machine learning for effective covid-19 diagnosis system. *Intelligence of things: ai-iiot based critical-applications and innovations.* Springer International Publishing, Cham, pp 55–76. https://doi.org/10.1007/978-3-030-82800-4_3
36. "Covid-19 database" 2022, [Online]. Available: <https://info.data.gov.il/datagov/home/>
37. Liao F, Zhu Z, Yan Z et al (2017) Ultrafast response flexible breath sensor based on vanadium dioxide. *J Breath Res* 11:36002
38. Spark A (2021) Apache spark. <https://spark.apache.org/>
39. 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>
40. ACT Government (2022) Common symptoms of Covid-19. <https://www.covid19.act.gov.au/stay-safe-and-healthy/symptoms-and-getting-tested/symptoms-of-covid-19>
41. CDC 24–7 (2022) Symptoms of Covid-19. <https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html>
42. Polikar R (2006) Ensemble based systems in decision making. *IEEE Circuits Syst Mag* 6:21–44. <https://doi.org/10.1109/MCAS.2006.1688199>
43. Friedman J (2001) Greedy function approximation: a gradient boosting machine author(s): Jerome H. Friedman source: the annals of statistics, Vol. 29, No. 5 (Oct., 2001), pp. 1189–1232 published by: institute of mathematical statistics stable. *Ann Stat* 29: 1189–1232
44. Breiman L (2001) Random forest. *Mach Learn* 45:5–32. <https://doi.org/10.1023/A:1010933404324>
45. Carrington AM, Manuel DG, Fieguth P et al (2022) Deep ROC analysis and AUC as balanced average accuracy, for improved classifier selection, audit and explanation. *IEEE Trans Pattern Anal Mach Intell.* <https://doi.org/10.1109/TPAMI.2022.3145392>
46. Cicioğlu M, Çalhan A (2021) Energy efficiency solutions for IEEE 802.15.6 based wireless body sensor networks. *Wirel Pers Commun* 119:1499–1513. <https://doi.org/10.1007/s11277-021-08292-8>
47. Cicioğlu M, Çalhan A (2020) Energy-efficient and SDN-enabled routing algorithm for wireless body area networks. *Comput Commun* 160:228–239. <https://doi.org/10.1016/j.comcom.2020.06.003>

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