Computers and Electrical Engineering 65 (2018) 222–235

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

A novel three-tier Internet of Things architecture with machine learning algorithm for early detection of heart diseases☆

Priyan Malarvizhi Kumar*, Usha Devi Gandhi

School of Information Technology and Engineering, VIT University, Vellore, Tamil Nadu, India

ARTICLE INFO

Article history:
Received 18 January 2017
Revised 2 September 2017
Accepted 4 September 2017
Available online 8 September 2017

Keywords:
Internet of Things
Machine learning
Wearable IoT devices
Big data
ROC analysis

ABSTRACT

Among the applications enabled by the Internet of Things (IoT), continuous health monitoring system is a particularly important one. Wearable sensor devices used in IoT health monitoring system have been generating an enormous amount of data on a continuous basis. The data generation speed of IoT sensor devices is very high. Hence, the volume of data generated from the IoT-based health monitoring system is also very high. In order to overcome this issue, this paper proposes a scalable three-tier architecture to store and process such huge volume of wearable sensor data. Tier-1 focuses on collection of data from IoT wearable sensor devices. Tier-2 uses Apache HBase for storing the large volume of wearable IoT sensor data in cloud computing. In addition, Tier-3 uses Apache Mahout for developing the logistic regression-based prediction model for heart diseases. Finally, ROC analysis is performed to identify the most significant clinical parameters to get heart disease.

1. Introduction

In recent years, there has been a perceptible increase the number of wearable devices for monitoring the patients' health, fitness and activities on a continues basis [1]. This has a long term impact on the recording of health, administration and clinical service to patient's physiological information. This advancement also helps the provision of more details relating to the daily routine and physical examination. During the health monitoring period, IoT wearable devices are attached with the human body to track the various health metrics that include blood pressure, heart rate, body temperature, respiratory rate, blood circulation level, body pain and blood glucose level [2]. The data collected from the IoT-based wearable devices are stored in a clinical database for necessary action when the patients' health condition deteriorates.

In general, traditional structured query language based databases are used in IoT health monitoring system to store clinical data. There has been an increase in the variety and quantity of IoT-based health monitoring devices in recent times. Hence, the traditional data processing methods and tools are not being used to store sensor data of huge volume generated by various IoT devices [3]. Scalable NOSQL (non structured query language) databases have to be used in the IoT-based health monitoring system. Researchers have started the use of big data and NOSQL technologies in various IoT applications. For example, Hassanalieragh et al. have used cloud computing with big data technologies to store the clinical data generated

☆ Reviews processed and recommended for publication to the Editor-in-Chief by Associate Editor Dr. R. Varatharajan.

* Corresponding author.
E-mail addresses: priyanm.k2016@vitstudent.ac.in (P.M. Kumar), ushadevi.g@vit.ac.in (U. Devi Gandhi).

http://dx.doi.org/10.1016/j.compeleceng.2017.09.001
0045-7906/© 2017 Published by Elsevier Ltd.

by various IoT devices [4]. In this application, the proposed health monitoring system continuously observes the individual's health condition. When, the health metrics such as ECG, respiratory rate, heart rate, sweating, skin temperature, blood pressure and heart sound go beyond standard values, the IoT devices send an alert message with the observed health measures to the doctor and other care holders.

Sun et al. have developed the IoT-based tailings dam monitoring system to monitor emergency situations in a tailings dam [5]. In this approach the cloud computing based scalable approach is used for taking the necessary action when situations of emergency arise. Rohokale et al. have developed IoT-based health monitoring system to observe health parameters such as hemoglobin (HB), blood pressure (BP), blood sugar and abnormal cellular growth [6]. The existing approaches use only traditional databases and tools to process the huge volume of sensor data generated from IoT devices. Hence, there is a need to develop an efficient and scalable architecture that stores as well as analyzes the huge volume of clinical data. This paper proposes a scalable big data based IoT health monitoring system for addressing this issue.

The proposed IoT-based framework is interconnected with cloud computing technology to increase scalability and availability. In addition, the proposed architecture uses Apache HBase to store the huge volume of the sensor data in the cloud. The individuals' health data is collected with the help of RFID and 5G mobile networks. In addition, Apache Mahout is used in the proposed health monitoring system for building the logistic regression-based prediction model for heart diseases. Finally, the performance of the prediction model is comparatively analyzed with the help of various performance evaluation metrics. The computed results such as throughput, sensitivity, accuracy and f-measure are used for demonstrating the efficiency and performance of the proposed IoT-based continuous health monitoring system.

The proposed IoT-based continuous health monitoring system is explained as follows: Section 1 describes the introduction to IoT-based health monitoring system. Section 2 reviews the recent works done in IoT-based healthcare systems. The proposed IoT-based continuous health monitoring system is explained in Section 3. Result and discussion, and performance evaluation are described in Sections 4 and 5 respectively. Section 6 concludes the paper.

2. Related work

The Internet of Things (IoT) is an interconnection of various physical objects for observing the physical events on a continues basis. The connected IoT devices communicate with each other with the help of advanced wireless networks and sensors [7]. IPv4 Internet was used in the last decade to transfer data at high speed. Advancements in network connectivity have helped enhancement of IPv4 Internet to IPv6 Internet with reduced delay and response time. IoT-based frameworks follow the layered architecture for transfer of the signal and communication between the devices. The layers that play an important role in network connectivity include Application Layer, Communication Layer, Security Layer, Embedded Layer, Hardware Layer, Integration Layer and DB Layer. RFID tags, sensors and actuators are used widely in IoT-based frameworks. Unique addressing schemes are used in IoT technology for mutual interaction between IoT devices [8].

The use of IoT technology in various fields has been on the increase. For example, Bäumer et al. have discussed the potential opportunities in using Internet of Things in a business organization [9]. CodeBlue project is the healthcare project developed by Harvard University. The essential role of the CodeBlue project is in the monitoring of the individuals' health parameters such as ECG, EKG, EMG, SpO2, pulse oximeter and Mica2 motes. Various electronic devices such as PDAs, laptops and personal computers are used in the CodeBlue project for necessary action from doctors and care holders when the patients' health condition deteriorates [10]. Published and subscription architecture are used in the CodeBlue project to deliver the health status of the patients in continues manner [11–13]. Researchers from the University of Virginia have developed the Alarm-Net framework to monitor the patients' health on a continues basis. The three-tier network architecture is used in the Alarm-Net project to sense the physiological parameters. IoT devices such as ECG, accelerometer and SpO2 sensors are attached to the human body in the first tier phase. The second tier focuses on observing the environmental parameters such as heat, moisture, movement and brightness [14]. Environmental sensors are attached to living things to observe environmental parameters. Tier-3 architecture is used for providing the network connectivity between the gateways. Tier three phase uses the internet protocol (IP)-based network to enable the wireless connection between source and destination [15,16].

The first tier of Alarm-net is used for sensing the physiological parameters from a patient and transferring the clinical data from the single-hop to the second tier phase. The second tier focuses on sending the clinical data from tier two to the third tier using the shortest-path-first routing protocol. This project is widely used for predicting the emergency conditions of the patients on the basis of the prior health records. Similarly, MobiCare is another healthcare project developed by Chakravorty et al. [17]. The project finds extensive use in the monitoring of patients' health over a wide-area. This project observes the clinical measures of the patients meticulously and sends the physiological values to the doctor and the care holder with the help of fog and cloud computing. The IoT wearable sensor devices used in the MobiCare project include SpO2, ECG and blood oxygen [18,19].

The MobiCare project senses the individuals' physiological information efficiently and sends it to the doctor and the care holder through a mobile phone and PDA. CDMA or GPRS/UMTS wireless technologies are used for transfer of the clinical data collected from the sensors to the doctor. The project uses HTTP POST protocol for sending the physiological data between the source and the destination. Similarly, PAM project developed by Blum et al. help observation of the mental health condition of the patients [20]. The essential goal of the PAM project is to identify the bipolar disorder (BP) in advance. Infrastructure and architecture based technologies are used in the PAM project for developing the PAM-I and PAM-A blocks respectively. PAM-I based system uses electronic equipments like PDA, cellular phones, wearable IoT devices and PC. PAM project uses environmental sensors in the IoT-based health monitoring system to observe the noise and dust pollution in the atmosphere [21]. Bluetooth with 5G mobile networks is used for connecting the various IoT sensors in the health monitoring system [22,23]

The intermediate servers make connection observation of the clinical data and store them into the database for clinical data analytics. The existing IoT-based health monitoring systems have not used big data technologies to store and process such huge volume of clinical data. In order to fulfill this gap, the proposed framework uses Apache HBase to store the large volume of clinical data in distributed manner [24]. Once the data is stored into the HBase, machine learning algorithms are used for processing such huge volume of data. This paper encourages the use of Apache Mahout based machine learning algorithms to develop the prediction model. MapReduce based logistic regression is identified to model the early stage of heart disease. A comparative analysis of the performance of the proposed prediction model is made using other existing machine learning approaches. The experimental results prove the effectiveness of the proposed model.

3. Proposed framework for IoT-based Health Monitoring System

The proposed IoT-based Health Monitoring System consists of a three-tier architecture to store and process a huge volume of wearable sensor data. Tier-1 focuses on collecting data from IoT wearable sensor devices. Tier-2 uses Apache HBase to store the huge volume of wearable IoT sensor data in the cloud. In addition, Tier-3 uses the Apache Mahout to develop the logistic regression-based prediction model for heart diseases. Fig. 1 represents the proposed framework for IoT-based Health Monitoring System. Fig. 2 represents the workflow for the proposed framework.

3.1. Tier 1: data collection

The proposed IoT-based health monitoring framework consists of three blocks, namely, data collection block, data storage block and data analytics block. Data collection block is used for collecting the individuals' physiological data using wearable IoT sensor devices. IoT wearable devices attached to the human body collect the patient's clinical data in a continuous manner. When the clinical measure of the individuals exceeds its normal value, the devices send an alert massage with clinical value to the doctor and the care holder. The alert messages and clinical values are collected and stored in the database in continues basis. The proposed framework uses 5G mobile networks for transfer of the clinical data into the clinical database for enabling the necessary action in an emergency situation. Algorithm 1 represents the IoT device initialization and continuous monitoring procedure and indicates the necessary steps to move the clinical data (*clinical_data.csv*) observed from IoT devices into the Amazon S3 bucket (*health_data)*. The proposed IoT-based health monitoring system uses 's3cmd' method for transfer of the clinical data from the local disk into the cloud database.

3.2. Tier 2: data storage

In general, IoT devices have the objective of sending clinical data continuously. It is difficult to store and process such huge amount data by traditional data processing tools and techniques. The proposed framework uses big data technologies to store the clinical data in a distributed manner. Apache HBase plays an important role in data storage in a distributed manner. Physical devices and personal computers are not sufficient to store the huge volume of data generated from the IoT wearable devices. The proposed IoT-based health monitoring system solves the problem using cloud computing technologies for scalability and elasticity. An account is created with Amazon to get the virtual machines with Apache HBase database access. Initially, clinical data collected from the human body is transferred to the Amazon simple storage service S3 with the help of the 's3cmd utility' method. Hence, when the clinical measure of the individuals' exceeds its normal value, the IoT devices send the clinical measures to the Amazon S3. Compared to Amazon S3 storage, Apache HBase provides a scalable data storage system in a distributed manner. Hence, the clinical data is transferred from the Amazon S3 to Apache HBase. The proposed health monitoring system uses Apache Pig to transfer the clinical data from Amazon S3 to Apache HBase.

Apache Pig is widely used to extract, load and transform the huge semi-structured, unstructured and structured data. PigLatin is a language used in Apache Pig for performing the data transformation. Algorithm 2 represents the PigLatin program for data transformation of the clinical data between the Amazon S3 bucket and Apache HBase. Fig. 3 is the data flow diagram for the proposed framework. Wearable IoT devices fixed to the human body to collect the individual's physiological information. These devices generate a large volume of data that cannot be stored in the traditional databases. Hence, hadoop distributed file system is used for storing a large volume of data in a scalable manner. Algorithm 3 is used for transfer of the data between Amazon S3 and Apache HBase.

3.3. Tier 3: data analytics

Data analytics block used for the development of the prediction model using logistic regression. The proposed framework uses Apache Mahout based machine learning libraries for implementing the logistic regression, which is one of the

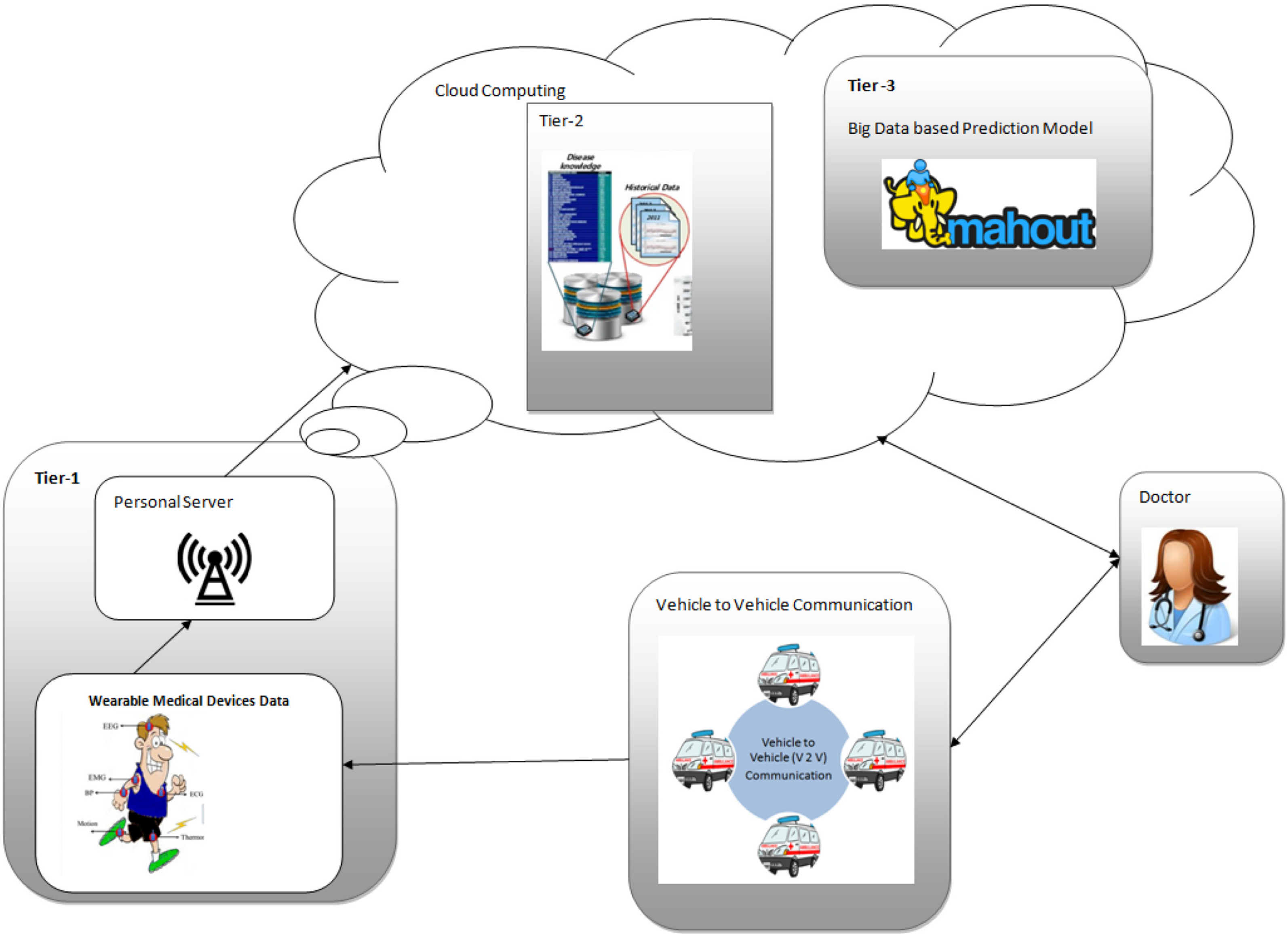

Fig. 1. Proposed framework for IoT-based health monitoring system.

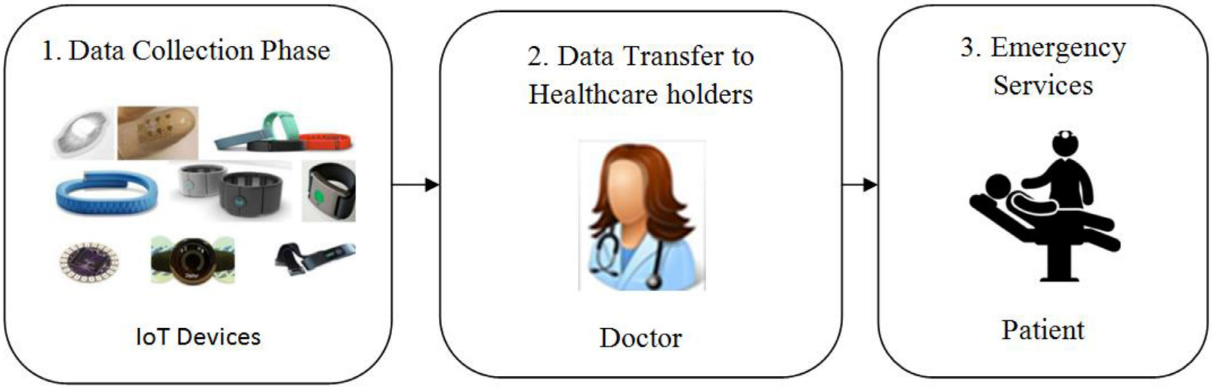

Fig. 2. Workflow for the proposed framework.

machine learning methods for predicting the coefficient between a dependent variable and one or more independent variables. Logistic regression works in the same manner as linear regression, with the difference of a dependent variable. In linear regression, dependent variable Y's and independent variable X's are in the form of numbers, whereas, in logistic regression, the independent variable X's may also be categorical values and the dependent variable Y's are coded as 1 in most cases (for those who present) 0 (for those who absent). Multiple logistic regression is defined by,

$$logit(p) = b_0 + b_1X_1 + b_2X_2 + \ldots + b_kX_k \tag{1}$$

Where,

p = denotes the probability of presence of the dependent variable (heart disease 0 or 1)
b = denotes the coefficients,
X_i = denotes independent variables,
i = Number of clinical parameters
X1 = Respiratory Rate (RP),
X2 = Heart rate (HR),
X3 = Blood Pressure-Systolic Range (BP-SR),
X4 = Blood Pressure - Diastolic Range (BP-DR),
X5 = Body Temperature (BT),
X6 = Blood Sugar - Fasting (BS-F),
X7 = Blood Sugar - Post Meal (BS-PM).

Logged odds are used for defining logistic transformation. The general formula for calculating of the odds ratio is defined by,

$$odds = \frac{p}{1-p} = \frac{probability\ of\ presence\ of\ the\ variable}{probability\ of\ absence\ of\ the\ variable} \tag{2}$$

Whereas logged odds is defined by,

$$logit(p) = \ln\left(\frac{p}{1-p}\right) \tag{3}$$

Logged odds are converted into probabilities for calculating the probability of the presence of the particular event. The conversion between the logged odds and probabilities is defined by,

$$p = \frac{e^{b_0+b_1X_{1+}b_2X_2+\ldots+b_kX_k}}{1+e^{b_0+b_1X_{1+}b_2X_2+\ldots+b_kX_k}} \tag{4}$$

The proposed IoT-based health monitoring framework uses logistic regression to develop the prediction model for monitoring heart diseases at the early stage. The Apache Mahout is used with Elastic MapReduce (EMR) framework for the development of the prediction model in cloud computing. Once the clinical data is stored in the Apache HBase, the Mahout-based logistic regression uses the prior clinical records for developing a prediction model. Pseudo code 1 represents the

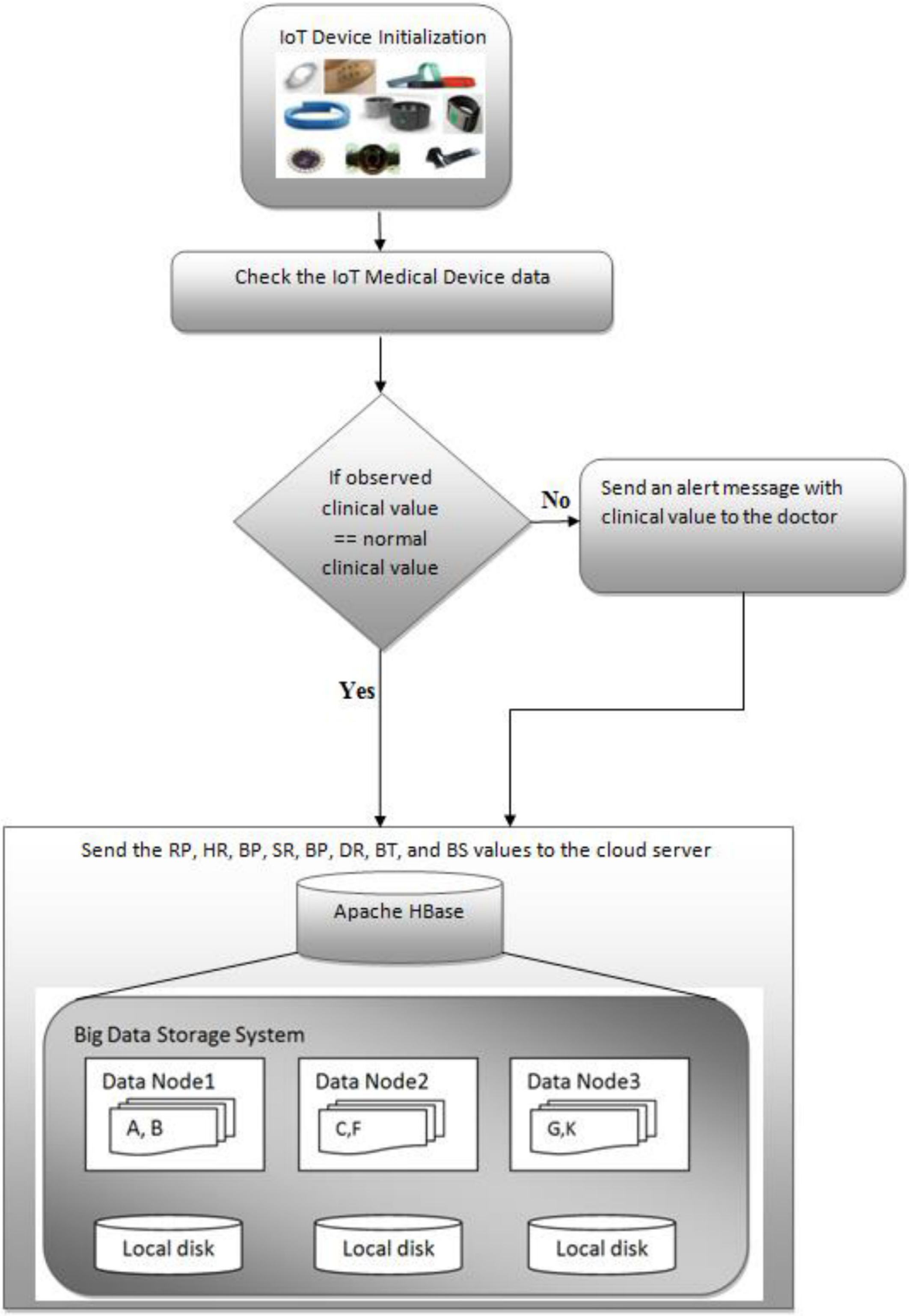

Fig. 3. Data flow diagram for the proposed framework.

implementation steps for the Apache Mahout in the Hadoop Distributed File System (HDFS). The implementation is done in the Linux based distributed environment.

Pseudo code 1: Mahout Prediction Model Development in the Linux Environment
Step1: *Install Hadoop and Java*
Step2: *Install maven 3*
sudo apt-get install maven
Step3: *mvn installation verification*
1. *xyz@chameera-VirtualBox:~$ mvn -v*
2. *Apache Maven 3.0.4*
3. *Maven home: /usr/share/maven*
4. *Java version: 1.7.0_45, vendor: Oracle Corporation*
5. *Java home: /usr/local/lib/jdk1.7.0_45/jre*
6. *Default locale: en_US, platform encoding: UTF-8*
7. *OS name: "linux", version: "3.11.0–15-generic", arch: "amd64", family: "unix"*
8. *xyz@xyz-VirtualBox:~$*

Step4: *Install svn*
1. *Open terminal*
   (Ctrl + Alt + T)
2. *sudo apt-get install subversion*

Step5: *Verify that the installation*
1. *xyz@xyz-VirtualBox:~$ svn –version*
   svn, version 1.7.9 (r1462340)
   compiled Oct 15 2013, 12:40:34
   Copyright (C) 2013 The Apache Software Foundation.

Step6: *Install Mahout*
1. *Select the directory to install Apache Mahout*
   cd /home/xyz/home
2. *Make new directory*
   mkdir mahout
3. *Find the new directory*
   cd mahout
4. *Use Subversion to check out the code*
   svn co http://svn.apache.org/repos/asf/mahout/trunk
5. *Find the trunk directory.*
   cd trunk
6. *Build the Apache mahout*
   mvn install
   mvn -DskipTests

Pseudo code 2 represents the steps to train the logistic regression in Linux based distributed environment. Pseudo code 3 represents the performance of the logistic regression-based prediction model to identify the heart disease.

Pseudo code 2: Mahout Prediction Model Training Phase
Step1: Find the Mahout Path
cd /home/xyz/home
cd mahout
Step2: Prediction Model Training
mahout org.apache.mahout.classifier.sgd.TrainLogistic –passes 1 –rate 1 –lambda 0.5 –input donut.csv –features 21 –output donut.model –target color –categories 2 – predictors x y xx xy yy a b c –types n n

Pseudo code 3: Mahout Prediction Model Prediction Phase
Step1: Find the Mahout Path
cd /home/xyz/home
cd mahout
Step2: Prediction Model Training
mahout org.apache.mahout.classifier.sgd.RunLogistic –input donut.csv –model donut.model –auc –scores –confusion

4. Result and discussion

Table 1 depicts the predicted coefficient values for the logistic regression-based prediction model. Respiratory Rate (RP), Heart rate (HR), Blood Pressure (BP)-Systolic Range (SR) and Blood Pressure (BP)-Diastolic Range (DR) are identified as significant variables for heart disease on the basis of p-values. Body Temperature (BT), Blood Sugar (BS)-Fasting and Blood Sugar (BS)-Post Meal are considered as not significant variables for heart disease based on the p-values. In this study, seven attributes are used for performing the experiment [25].

Table 1
Logistic regression results.

Variable	Beta	Std error	Z	P value	Odds ratio
Respiratory Rate (RP)	−0.17	0.0356	−0.2764	0.0023	3.41
Heart rate (HR)	−0.04	0.0164	−0.0923	0.0018	4.32
Blood Pressure (BP): Systolic Range (SR)	−0.02	0.0873	−0.0537	0.0145	2.11
Blood Pressure (BP):Diastolic Range (DR)	−0.47	0.1023	−0.2894	0.0334	1.01
Body Temperature (BT)	−0.01	0.3421	−0.0323	0.1543	0.75
Blood Sugar (BS):Fasting	0.45	0.3873	0.8732	0.3451	0.05
Blood Sugar (BS):Post Meal	0.23	0.0461	0.3984	0.0174	1.14

Table 2
Confusion matrix.

Variable	Level	Yes	No
Respiratory Rate (RP)	Low (<12)	229	86
	Medium (12–50)	61	254
	High (>50)	25	290
Heart rate (HR)	Low (<60)	92	223
	Medium (60–160)	136	179
	High (>160)	87	228
Blood Pressure (BP): Systolic Range (SR)	Low (<75)	2	313
	Medium (75–140)	268	47
	High (>140)	45	360
Blood Pressure (BP):Diastolic Range (DR)	Low (<50)	59	256
	Medium (50–90)	93	222
	High (>90)	163	478
Body Temperature (BT)	Low (<36.6)	59	256
	Medium (36.6–37)	201	114
	High (>37)	55	260
Blood Sugar (BS):Fasting	Low (<70)	146	169
	Medium (70–100)	45	270
	High (>100)	124	191
Blood Sugar (BS): Post Meal	Low (<70)	43	358
	Medium (70–140)	36	279
	High (>140)	236	79

5. Performnace evaluation

Sensitivity, specificity, positive likelihood ratio (PLR), negative likelihood ratio (NLR), disease prevalence (DP), positive predicted value (PPV) and negative predicted value (NPV) are calculated for evaluating of the prediction model. The validations metric are defined by,

$$Specificity \ = \frac{True\ Negative\ (TN)}{False\ Positive(FP) + True\ Negative\ (TN)} \tag{5}$$

$$Positive\ Liklihood\ Ratio\ (PLR) = \frac{Sensitivity}{100 - Specificity} \tag{6}$$

$$Negative\ Liklihood\ Ratio\ (NLR) = \frac{100 - Sensitivity}{Specificity} \tag{7}$$

$$Positive\ Predicted\ Value\ (PPV) \ = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP) + False\ Positive(FP)} \tag{8}$$

$$Negative\ Predicted\ Value\ (NPV) \ = \frac{True\ Negative\ (TN)}{True\ Negative\ (TN) + False\ Negative(FN)} \tag{9}$$

$$Sensitivity = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP) + False\ Negative(FN)} \tag{10}$$

Table 3
ROC analysis.

Variable	Cut point	True positive rate	False positive rate
Respiratory Rate (RP)	12	0.7270	0.2730
	50	0.9206	0.0794
Heart rate (HR)	60	0.2921	0.7079
	160	0.7238	0.2762
Blood Pressure (BP): Systolic Range	75	0.0630	0.9370
	140	0.8571	0.1429
Blood Pressure (BP):Diastolic Range	50	0.1873	0.8217
	90	0.4825	0.5175
Body Temperature (BT)	36.6	0.1873	0.8217
	37	0.8254	0.1746
Blood Sugar: Fasting	70	0.4635	0.5365
	100	0.6063	0.3937
Blood Sugar: Post Meal	70	0.1365	0.8635
	140	0.2508	0.7492

Table 4
Validation metrics.

Variable	Cut point	Sensitivity	Specificity	Positive likelihood ratio	Negative likelihood ratio	Disease prevalence	Positive predicted value	Negative predicted value
Respiratory Rate (RP)	12	72.7%	86.25%	5.33	0.32	33.33%	72.7%	86.35%
	50	92.06%	46.03%	1.71	0.17	33.33%	46.03%	92.06%
Heart rate (HR)	60	29.21%	64.60%	0.83	1.10	33.33%	29.21%	64.66%
	160	72.38%	36.19%	1.13	0.76	33.33%	36.19%	72.38%
Blood Pressure: SR	75	0.63%	56.53%	0.01	1.76	30.43%	0.63%	56.53%
	140	85.71%	50.00%	1.71	0.29	30.43%	42.86%	88.89%
Blood Pressure: DR	50	18.73%	73.22%	0.70	1.11	24.78%	18.73%	73.22%
	90	48.25%	50.00%	0.97	1.03	24.78%	24.13%	74.57%
Body Temperature	36.6	18.73%	59.37%	0.46	1.37	33.33%	18.73%	59.37%
	37	82.54%	41.27%	1.41	0.42	33.33%	41.27%	82.54%
Blood Sugar: Fasting	70	46.35%	72.03%	1.66	0.74	32.98%	44.92%	73.17%
	100	60.63%	30.32%	0.87	1.30	33.33%	30.32%	60.63%
Blood Sugar: Post Meal	70	13.65%	50.00%	0.27	1.73	30.55%	10.72%	56.83%
	140	25.08%	11.03%	0.28	6.79	30.55%	11.03%	25.08%

$$Disease\ \Pr evalence\ (DP) = \frac{True\ Positive\ \ (TP) + False\ Negative(FN)}{True\ Positive\ \ (TP) + False\ Positive(FP) + True\ Negative\ \ (TN) + False\ Negative(FN)} \tag{11}$$

Table 2 depicts the confusion matrix generated from the prediction model for various levels of clinical parameters such as Respiratory Rate (RP), Heart rate (HR), Blood Pressure-Systolic Range (BP-SR), Blood Pressure - Diastolic Range (BP-DR), Body Temperature (BT), Blood Sugar -Fasting (BS-F) and Blood Sugar - Post Meal (BS-PM). Receiver operating characteristic (ROC) analysis is performed for evaluating of the most significant clinical values. Tables 3 and 4 show that Respiratory Rate (RP) at the rate of 50 and 12 as highly significant in causing the heart disease. In addition, heart rate 160 is highly significant indicating the heart disease. Similarly, Blood Pressure (BP): Systolic Range and Body Temperature (BT) at 140 and 37 is also considered as a highly significant variable for heart disease. Figs. 4 and 5 represents the true positive and false positive rates, and performance evaluation metrics for various significant clinical parametric values. Fig. 4 is a graphical visualization of Table 3 while Fig. 5 is a graphical visualization of Table 4.

Algorithm 1
IoT device initialization and continuous monitoring.

Step1: *Fix the IoT medical devices in patients' body Start Device Initialization*
Step2: *Continuously monitoring the patient health condition based on the following metrics*

1. *if(Approximate Age==Newborn)*
2. *{*
3. *Respiratory Rate (RP)==30-50 && Heart rate (HR)==100-160&&Blood Pressure (BP):Systolic Range (SR)==75-100&&Blood Pressure (BP):Diastolic Range (DR)==50-70&&Body Temperature (BT)==36.6 -37&&Blood Sugar (BS):Fasting==70-100&&Blood Sugar (BS):Post Meal==70-140*
4. *Send the RP, HR, BP, SR, BP, DR, BT, and BS values to the Amazon S3 data store*
5. *}*
6. *elseif(Approximate Age==0-5 months)*
7. *{*
8. *Respiratory Rate (RP)== 25-40&& Heart rate (HR)== 90-150&&Blood Pressure (BP):Systolic Range (SR)==75-100&&Blood Pressure (BP):Diastolic Range (DR)==50-70&&Body Temperature (BT)==36.6 -37&&Blood Sugar (BS):Fasting==70-100&&Blood Sugar (BS):Post Meal==70-140*
9. *Send the RP, HR, BP, SR, BP, DR, BT, and BS values to the Amazon S3 data store*
10. *}*
11. *elseif(Approximate Age==6-12 months)*
12. *{*
13. *Respiratory Rate (RP)== 20-30&& Heart rate (HR)== 80-140&&Blood Pressure (BP):Systolic Range (SR)==75-100&&Blood Pressure (BP):Diastolic Range (DR)==50-70&&Body Temperature (BT)==36.6 -37&&Blood Sugar (BS):Fasting==70-100&&Blood Sugar (BS):Post Meal==70-140*
14. *Send the RP, HR, BP, SR, BP, DR, BT, and BS values to the Amazon S3 data store*
15. *}*
16. *elseif(Approximate Age==1-3 years)*
17. *{*
18. *Respiratory Rate (RP)== 20-30&& Heart rate (HR)== 80-130&&Blood Pressure (BP):Systolic Range (SR)== 80-110&&Blood Pressure (BP):Diastolic Range (DR)== 50-80&&Body Temperature (BT)==36.6 -37&&Blood Sugar (BS):Fasting==70-100&&Blood Sugar (BS):Post Meal==70-140*
19. *Send the RP, HR, BP, SR, BP, DR, BT, and BS values to the Amazon S3 data store*
20. *}*
21. *elseif(Approximate Age==3-5 years)*
22. *{*
23. *Respiratory Rate (RP)== 20-30&& Heart rate (HR)== 80-120&&Blood Pressure (BP):Systolic Range (SR)== 80-110&&Blood Pressure (BP):Diastolic Range (DR)== 50-80&&Body Temperature (BT)==36.6 -37&&Blood Sugar (BS):Fasting==70-100&&Blood Sugar (BS):Post Meal==70-140*
24. *Send the RP, HR, BP, SR, BP, DR, BT, and BS values to the Amazon S3 data store*
25. *}*
26. *elseif(Approximate Age==6-10 years)*
27. *{*
28. *Respiratory Rate (RP)== 15-30&& Heart rate (HR)== 70-110&&Blood Pressure (BP):Systolic Range (SR)== 85-120&&Blood Pressure (BP):Diastolic Range (DR)== 55-80&&Body Temperature (BT)==36.6 -37&&Blood Sugar (BS):Fasting==70-100&&Blood Sugar (BS):Post Meal==70-140*
29. *Send the RP, HR, BP, SR, BP, DR, BT, and BS values to the Amazon S3 data store*
30. *}*
31. *elseif(Approximate Age==11-34 years)*
32. *{*
33. *Respiratory Rate (RP)== 12-20&& Heart rate (HR)== 60-105&&Blood Pressure (BP):Systolic Range (SR)== 95-140&&Blood Pressure (BP):Diastolic Range (DR)== 60-90&&Body Temperature (BT)==36.6 -37&&Blood Sugar (BS):Fasting==70-100&&Blood Sugar (BS):Post Meal==70-140*
34. *Send the RP, HR, BP, SR, BP, DR, BT, and BS values to the Amazon S3 data store*
35. *}*
36. *elseif(Approximate Age==35-100 years)*
37. *{*
38. *Respiratory Rate (RP)== 12-30&& Heart rate (HR)== 60-100&&Blood Pressure (BP):Systolic Range (SR)== 95-140&&Blood Pressure (BP):Diastolic Range (DR)== 60-90&&Body Temperature (BT)==36.6 -37&&Blood Sugar (BS):Fasting==70-100&&Blood Sugar (BS):Post Meal==70-140*
39. *Send the RP, HR, BP, SR, BP, DR, BT, and BS values to the Amazon S3 data store*
40. *}*
41. *else*
42. *{*
43. *If health parameters is not regular*
44. *Send the voice alert "Patient is abnormal" with Clinical value the RP, HR, BP, SR, BP, DR, BT, and BS values to the Doctor as well as Amazon S3 data store*
45. *}*

Algorithm 2
Store the IoT wearable sensor devices data into Amazon S3 data store.

1. ***Step1:*** *Identify the System Name in the Amazon S3*
2. ***Step 2:****Create the directory in the Amazon S3*
   Name_of_the_System $ s3cmd mb s3://health_data
   Bucket 's3:// health_data/' created
3. ***Step 3:*** *Use put method to store the clinical data into the Amazon S3 directory*
   Name_of_the_System $ s3cmd put clinical_data.csv. s3:// health_data
4. ***Step 4:*** *Visualize the log file of the clinical data*
   sample-syslog.log -> s3:// health_data/clinical_data.csv

Algorithm 3
Data transformation between the Amazon S3 and Apache HBase.

1. ***Data:*** *Clinical data collected from IoT sensor devices which is stored in Amazon S3*
2. ***Input:*** *Amazon S3 health data Table*
3. ***Output:*** *Pig Data Load Table query*
4. ***Step-1: Hadoopdistcp s3n:****health_data/clinical_data.csv/user/patient1*
5. ***Step-2: if*** *(Table_Name ≠ NULL)* ***then***
6. ***Step-3:load*** *'/user/patient/ clinical_data.csv'* ***using PigStorage(',')*** ***as***
   *(**add** field name, field data type);*
7. ***stored into*** *'hbase://clinical_data Table'* ***usingorg.apache.pig.backend.hadoop.hbase.HBaseStorage***
   *(**add** column family name patient ID: field column family data type Integer);*
   *(**add** column family name Respiratory Rate (RP):field column family data type Integer);*
   *(**add** column family name Heart rate (HR):field column family data type Integer);*
   *(**add** column family name Blood Pressure (BP):Systolic Range (SR):field column family data type Integer);*
   *(**add** column family name Blood Pressure (BP):Diastolic Range (DR):field column family data type Integer);*
   *(**add** column family name Body Temperature (BT):field column family data type Double);*
   *(**add** column family name Blood Sugar (BS):field column family data type Integer);*
   *(**add** column family name Blood Sugar (BS):Post Meal: field column family data type Integer);*
8. ***Step-4:*** *Return the Pig Data Load Table query*

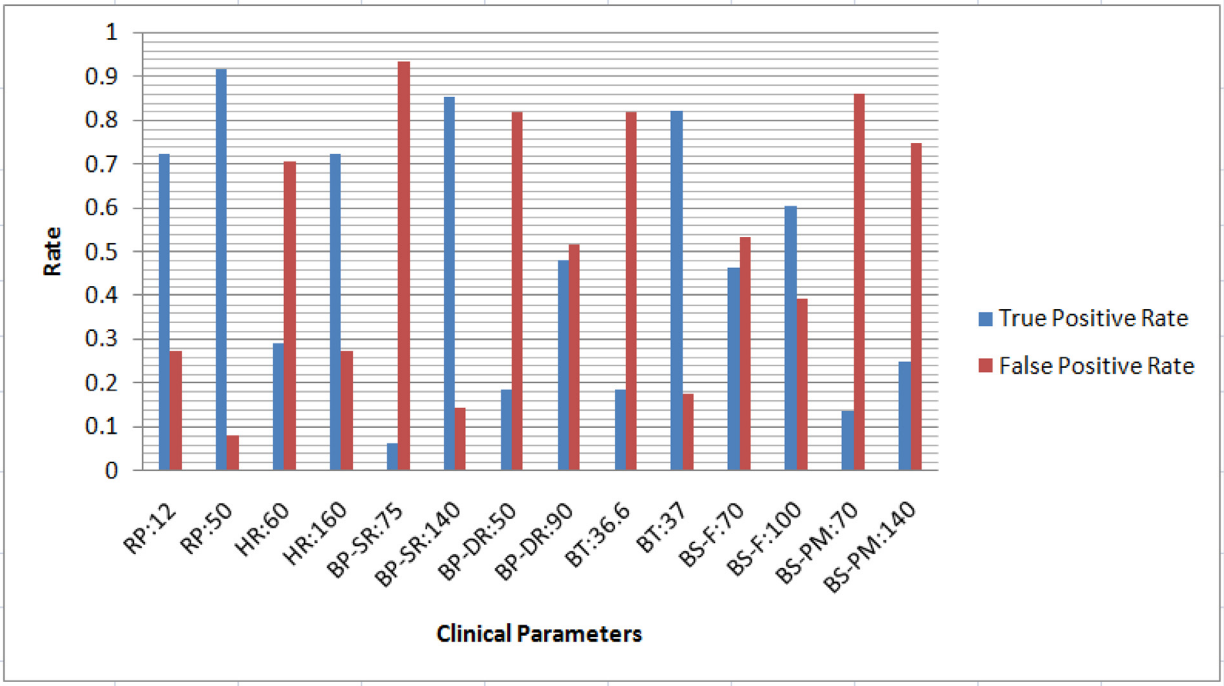

Fig. 4. True positive and false positive rate.

6. Conclusion

This paper proposes a scalable IoT based three-tier architecture to process the sensor data and identify the most significant clinical parameters to get heart disease. The most significant clinical parameters that indicate impending heart disease are identified with the help of ROC analysis. Blood Sugar (BS)-Fasting and Blood Sugar (BS)-Post Meal are found to have positive correlation with heart disease. However, Respiratory Rate (RP), Heart rate (HR), Blood Pressure (BP): Systolic Range (SR), Blood Pressure (BP): Diastolic Range (DR) and Body Temperature (BT) are found to be negatively correlated with heart disease. The experimental results prove that Respiratory Rate (RP) at around 50 and 12 is highly significant in the indication of the heart disease. Heart rate 160 is also indication of heart disease. Similarly, Blood Pressure (BP): Systolic Range and Body Temperature (BT) at 140 and 37 are also considered as a highly significant variable for heart problem indication. The future work of this study is to propose a continuous health monitoring system with a doctor on the move with an ambulance. An energy efficient node selection algorithm is identified for future work to choose the best mobile ambulance.

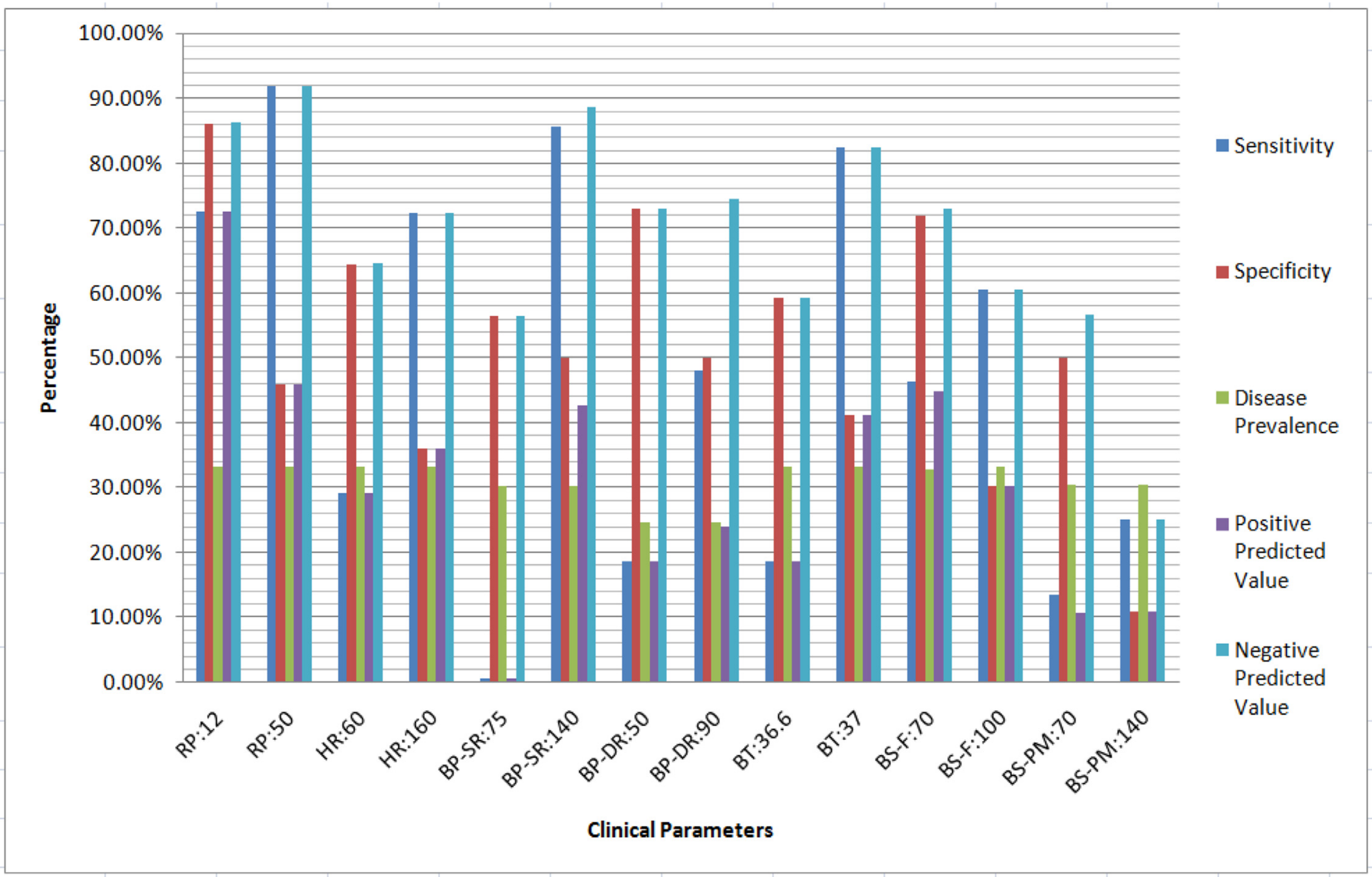

Fig. 5. Performance evaluation metrics.

References

[1] Lorincz K, Malan DJ, Fulford-Jones TR, Nawoj A, Clavel A, Shnayder V, et al. Sensor networks for emergency response: challenges and opportunities. Pervasive Comput, IEEE 2004;3(4):16–23.
[2] Shafiq MZ, Ji L, Liu AX, Pang J, Wang J. A first look at cellular machine-to-machine traffic: large scale measurement and characterization. ACM SIGMETRICS Perform Eval Rev 2012;40(1):65–76.
[3] Verma PK, Verma R, Prakash A, Agrawal A, Naik K, Tripathi R, et al. Machine-to-Machine (M2M) communications: a survey. J Network Comput Appl 2016;66:83–105.
[4] Hassanalieragh M, Page A, Soyata T, Sharma G, Aktas M, Mateos G, et al. Health monitoring and management using internet-of-things (IoT) sensing with cloud-based processing: opportunities and challenges. In: Proceedings of 2015 IEEE international conference on services computing (SCC). IEEE; 2015 June, p. 285–292.
[5] Sun E, Zhang X, Li Z. The Internet of Things (IOT) and cloud computing (CC) based tailings dam monitoring and pre-alarm system in mines. Saf Sci 2012;50(4):811–15.
[6] Rohokale VM, Prasad NR, Prasad R. A cooperative Internet of Things (IoT) for rural healthcare monitoring and control. In: Wireless communication, vehicular technology, information theory and aerospace & electronic systems technology (wireless VITAE), 2011 2nd international conference on. IEEE; February 2011. p. 1–6.
[7] Baoyun W. Review on Internet of Things. J Electr Measur Instrum 2009;23(12):1–7.
[8] Doppler K, Rinne M, Wijting C, Ribeiro CB, Hugl K. Device-to-device communication as an underlay to LTE-advanced networks. IEEE Commun Mag 2009;47(12):42–9.
[9] Manogaran G, Lopez D. A Gaussian process based big data processing framework in cluster computing environment. Cluster Comput 2017:1–16.
[10] Ishaq I, Carels D, Teklemariam GK, Hoebeke J, Abeele FVD, Poorter ED, et al. IETF standardization in the field of the Internet of Things (IoT): a survey. J Sens Actuator Networks 2013;2(2):235–87.
[11] Kumar P, Lee HJ. Security issues in healthcare applications using wireless medical sensor networks: a survey. Sensors 2011;12(1):55–91.
[12] Manogaran G, Thota C, Kumar MV. MetaCloudDataStorage architecture for Big Data security in cloud computing. Procedia Comput Sci 2016 Dec 31;87:128–33.
[13] Manogaran G, Thota C, Lopez D, Vijayakumar V, Abbas KM, Sundarsekar R. Big data knowledge system in healthcare. In: Internet of Things and big data technologies for next generation healthcare; 2017. p. 133.
[14] Gubbi J, Buyya R, Marusic S, Palaniswami M. Internet of Things (IoT): a vision, architectural elements, and future directions. Future Gener Comput Syst 2013;29(7):1645–60.
[15] Whitmore A, Agarwal A, Da Xu L. The Internet of Things—a survey of topics and trends. Inf Syst Front 2015;17(2):261–74.
[16] Manogaran G, Thota C, Lopez D, Vijayakumar V, Abbas KM, Sundarsekar R. Big data knowledge system in healthcare. In: Internet of Things and big data technologies for next generation healthcare. Springer International Publishing; 2017. p. 133–57.
[17] Varatharajan R, Manogaran G, Priyan MK, Balaş VE, Barna C. Visual analysis of geospatial habitat suitability model based on inverse distance weighting with paired comparison analysis. Multimed Tools Appl 2017:1–21.
[18] Atzori L, Iera A, Morabito G. The Internet of Things: a survey. Comput Networks 2010;54(15):2787–805.
[19] Manogaran G, Thota C, Lopez D, Vijayakumar V, Abbas KM, Sundarsekar R. Big data knowledge system in healthcare. In: Internet of things and big data technologies for next generation healthcare. Springer International Publishing; 2017. p. 133–57.
[20] Blum JM, Magill EH. The design and evaluation of personalised ambient mental health monitors. In: Proceedings of 7th consumer communications and networking conference (CCNC). IEEE; 2010. p. 1–5.
[21] Bonomi F, Milito R, Zhu J, Addepalli S. Fog computing and its role in the internet of things. In: Proceedings of the first edition of the MCC workshop on Mobile cloud computing ACM, August; 2012. p. 13–16.
[22] Manogaran G, Lopez D. Spatial cumulative sum algorithm with big data analytics for climate change detection. Comput Electr Eng 2017. http://dx.doi.org/10.1016/j.compeleceng.2017.04.006.
[23] Priyan MK, Devi GU. Energy efficient node selection algorithm based on node performance index and random waypoint mobility model in internet of vehicles. Cluster Comput 2017:1–15.
[24] Varatharajan R, Manogaran G, Priyan MK, Sundarasekar R. Wearable sensor devices for early detection of Alzheimer disease using dynamic time warping algorithm. Cluster Comput 2017:1–10.
[25] UCI Machine Learning Repository: Heart Disease Data Set Archive.ics.uci.edu. Retrieved 22 July 2017, from http://archive.ics.uci.edu/ml/datasets/heart+Disease.

Priyan Malarvizhi Kumar is currently pursuing a Ph.D. in the Vellore Institute of Technology University. He received my Bachelor of Engineering and Master of Engineering degree from Anna University and Vellore Institute of Technology University, respectively. His current research interests include Big Data Analytics, Internet of Things, Internet of Everything, Internet of Vehicles in Healthcare.

G. Usha Devi is working as an Associate Professor in the School of Information Technology and Engineering, Vellore Institute of Technology University. He received his Bachelor of Engineering and Master of Engineering degree from the Anna University. His current research interests include big data analytics and wireless networks. He has published number of international journals and conferences.