Hadoop-based Intelligent Care System (HICS): Analytical Approach for Big Data in IoT

M. MAZHAR RATHORE, Kyungpook National University
ANAND PAUL, Kyungpook National University
AWAIS AHMAD, Yeungnam National University
MARCO ANISETTI, Università degli Studi di Milano
GWANGGIL JEON, Incheon National University

The Internet of Things (IoT) is increasingly becoming a worldwide network of interconnected things that are uniquely addressable, via standard communication protocols. The use of IoT for continuous monitoring of public health is being rapidly adopted by various countries while generating a massive volume of heterogeneous, multisource, dynamic, and sparse high-velocity data. To handle such an enormous amount of high-speed medical data while integrating, collecting, processing, analyzing, and extracting knowledge constitutes a challenging task. On the other hand, most of the existing IoT devices do not cooperate with one another by using the same medium of communicationaf. For this reason, it is a challenging task to develop healthcare applications for IoT that fulfill all user needs through real-time monitoring of health parameters. Therefore, to address such issues, this paper proposed a Hadoop-based intelligent care system (HICS) that demonstrates IoT-based collaborative contextual Big Data sharing among all of the devices in a healthcare system. In particular, the proposed system involves a network architecture with enhanced processing features for data collection generated by millions of connected devices. In the proposed system, various sensors, such as wearable devices, are attached to the human body that measure health parameters and transmit them to a primary mobile device (PMD). The collected data are then forwarded to intelligent building (IB) using the Internet where the data are thoroughly analyzed to identify abnormal and serious health conditions. Intelligent building comprises: 1) a Big Data collection unit (used for data collection, filtration, and load balancing); 2) a Hadoop processing unit (HPU) (comprised of HDFS and MapReduce); and 3) an analysis and decision unit. The HPU, analysis, and decision unit are equipped with a medical expert system, which reads the sensor data and performs actions in the case of an emergency situation. To demonstrate the feasibility and efficiency of the proposed system, we use publically available medical sensory data sets and real-time sensor traffic while identifying the serious health conditions of patients by using thresholds, statistical methods, and machine learning techniques. The results show that the proposed system is very efficient and able to process high-speed WBAN sensory data in real-time.

Categories and Subject Descriptors: **C.2.1 [Computer-Communication Networks]**: Network Architecture and Design

General Terms: design; algorithms; performance.

Additional Keywords and Phrases: IoT; healthcare; Big Data; PMD; intelligent building.

This study was supported by the Brain Korea 21 Plus project (SW Human Resource Development Program for Supporting Smart Life) funded by the Ministry of Education, School of Computer Science and Engineering, Kyungpook National University, Korea (21A20131600005). This work was also supported by the IT R&D Program of MSIP/IITP. [10041145, Self-Organized Software Platform (SoSp) for Welfare Devices].

Author's addresses: M. Mazhar Rathore, School of Computer Science and Engineering, Kyungpook National University Daegu, Korea, 702-701. Email: rathoremazhar@gmail.com. Anand Paul, School of Computer Science and Engineering, Kyungpook National University Daegu, Korea, 702-701. Email: paul.editor@gmail.com. Awais Ahmad, Department of Information and Communication Engineering, Yeungnam University, Gyeongbuk, Korea, 38541. Email: aahmad.marwat@gmail.com. Marco Anisetti, Dipartimento di Informatica, Università degli Studi di Milano, Italy. Email: marco.anisetti@unimi.it. Gwanggil Jeon, Incheon National University, Korea. Email: gjeon@inu.ac.kr.

DOI:http://dx.doi.org/10.1145/0000000.0000000

1. INTRODUCTION

Since rapid development began in the beginning of the 21st century, healthcare systems in the Internet of Things (IoT) have been recognized as constituting a revolution in ICT [Ashton 2009, Xu and Wang 2013]. Mainly, IoT connects physical devices, such as sensors, actuators, embedded sensors, and radio frequency identification (RFID) into a single medium, called the Internet. It is considered to be an empowering technology to achieve a global infrastructure of interconnected physical components [Welbourne et al. 2009]. Furthermore, IoT extends the usage of the Internet into our daily lives by connecting billions of smart devices [Kortuem et al. 2010], which introduces momentous changes in the way that we live and interact with other devices [Jara et al. 2012; Yang et al. 2014].

In healthcare system applications, the wireless body area network (WBAN) offers a novel archetype for wireless sensor networks (WSNs) in monitoring bio-medical sensors, such as biochips. These sensors can be attached to the human body or clothes [Xing and Zhu 2009; Cavallari et al. 2014; Alam and Hamida 2014] in order to measure parameters associated with the human body. The measured values can be collected and relayed to a main server using Internet Protocol Version 6 (IPv6) over a Low-Power Wireless Personal Area Network (6LoWPAN) [Kushalnagar et al. 2007; Montenegro et al. 2007]. This helps in connecting these nodes to an IPv6 network, which plays a fundamental role in health diagnostic issues. In order to analyze the amount of data collected, a need exists to transmit the data to a server via a gateway node and the Internet to generate results. For such application, ZigBee technology can be employed, which uses IEEE 802.15.4 PHY along with MAC standards [Kiran et al. 2014]. Since the health sensors that collect data are wireless, maintaining battery power and battery replacement present critical challenges.

In IoT, intelligent components, RFID tags [Lee et al. 2010], sensors, actuators, etc., have been rapidly developed. As a result, dramatic growth has been achieved in the field of IoT applications [Li et al. 2011; Broll et al. 2009]. These technologies have enabled great advancements in healthcare systems [European Commission Information Society 2008; Hande and Cem 2008; National Information Council 2008; Li et al. 2013]. Some approaches aim to integrate wearable devices to attain better IoT for e-healthcare systems [Castillejo et al. 2013]. A mobile-based telemonitoring system for chronic diseases is proposed in [Morak et al. 2012]. In such application, the amount of data collected from the sensor nodes should be accessible anytime and anywhere, which requires continuous network connectivity, and generates a massive volume of data (referred to as 'Big Data'). In addition, according to the GSMA, the total number of devices connected to each other will reach 15 billion in 2015, and increase to 24 billion in 2020[1] [GSMA 2013]. The processing of such high-speed sensory traffic is also an essential obstacle that must be overcome. In such circumstances, healthcare systems will face a critical challenge in real-time processing, analysis, and decision-making. Existing standard techniques are incapable of handling such high volume and high velocity of data with heterogeneous information to generate real-time action in cases of emergencies. On the other hand, an increasing number of Big Data solutions exist that are, at least theoretically, capable of addressing volume and velocity requirements. However, in many cases, these solutions are in preliminary phases or not fully suitable for IoT scenarios (e.g., H2020 Toreador[2] and H2020 Evotion[3]). Others are not fully capable to fulfill all of the requirements of a generic and

[1] http://gigaom.com/2011/10/13/internet-of-things-will-have-24-billiondevices-by-2020/
The Mobile Economy, GSMA, 2013.
[2] http://www.toreador-project.eu/
[3] http://h2020evotion.eu/

complete healthcare system [Fang et al. 2016; Muin et al. 2014]. Therefore, a system is required that is capable of taking care of human beings (patients in the home, people outside of the home, people who are exercising or driving, etc.), which not only addresses required medications but can also perform continuous monitoring of patients. The mentioned challenges can only be overcome through rapid collection and aggregation, as well as parallel and efficient processing of incoming high-speed medical, sensory data.

For this reason, the Hadoop-based intelligent care system (HICS) using a Big Data analytical approach is proposed in this paper, which is based on parallel-processing and multiple Hadoop HDFS data nodes. In the proposed system, a human body uses wearable devices and other physical body sensors that measure blood pressure (BP), pulse rate, signs of diabetes, etc., and sends this information to an attached coordinator. The measured data are then transmitted to the primary medical device (PMD) using Bluetooth or Zigbee IEEE 802.15.4 technology. The PMD is connected to the Internet through 3G/LTE/WiFi via gateways. Each gateway is responsible for collecting the measured data from various PMDs, which are then sent to intelligent building (IB). IB constitutes the backbone of the proposed HICS network architecture, which processes massive volumes of incoming data streams by using high-speed capture devices, such as RF_RING and TNAPI [F. Fusco and L. Deri 2010], and aggregates the results in its collection unit. The collected data are then sent to the Hadoop processing unit (HPU) for further processing. The HPU performs analysis algorithms, including statistical calculations, comparisons and other operations, and generates results. Finally, the analysis and decision unit responds to the system (in the case that a patient requires a remote physician or an ambulance) based on the results generated by the HPU. The proposed IoT-based HICS system has a strong network architecture that comprises five networking layers, namely: 1) a data collection layer; 2) a communication layer; 3) a processing layer; 4) a management layer; and 5) a service layer. The main contribution of this work is summarized as follows. A network architecture with a five-layered structure is proposed and implemented using the Hadoop ecosystem that performs a collaborative task to process and analyze data. To the best of our knowledge, the proposed network architecture is the first architecture specifically designed for healthcare systems, in which the processing of real-time, as well as offline, data is natively based on Hadoop. Secondly, our system can handle the massive volume of data generated by connected healthcare devices by dividing them into components and performing analyses using Hadoop. Moreover, the intelligent building concept is introduced, which is mainly responsible for managing, processing and analyzing incoming sensor data, and finally make decisions intelligently. This enables not only efficient handling of the massive volume of data, but also provides feedback to users “anytime-anywhere-anyhow.” The whole system is implemented in a real environment using Hadoop on UBUNTU 14.04. Finally, the medical sensor datasets are replayed to check the feasibility, accuracy, and efficiency of the proposed system by identifying serious health conditions of patients by using thresholds, statistical methods, and machine learning.

2. RELATED WORK

In the development of an intelligent IoT environment, it is essential to understand that the ultimate goal of connected things can be seriously limited in the contexts of power, computation, and efficiency. In this section, we focus mainly on IoT applications and their drawbacks.

Various approaches have been developed for the smart environment, such as mobile intelligent IoT applications, in which they assume that each of the u-things have the capability of decision-making based on environmental conditions [Roderic and Hanson 2009; Tangab et al. 2012; Liao et al. 2012]. In such methodologies, performance degradation is at a peak since the majority of the u-things have limited capabilities to handle such expensive computations.

An iHome Health-IoT platform is proposed, which involves an open-platform-based intelligent medicine box (iMedPack) with heightened connectivity and interchangeability for the integration of devices, as well as services [Yang et al. 2014]. The mentioned platform, however, lacks efficient communication between the IoT cloud and the user. A smart kitchen platform is proposed, in which the sensors and actuators are controlled by an Intel Galileo board [Wu et al. 2014]. In this approach, the sensors deployed at the doors' entrance do not provide the actual reading of an object, and it may be an animal entering the room instead of the intended disabled person. In addition, various issues related to IoT architecture, implementation, as well as the identification of promising IoT applications and the management of data streams from connected things are described [Zhou 2013]. It is worth noting that cloud services, as well as Big Data approaches, could be utilized to store, analyze, and visualize the data streams being generated by IoT devices, which improve various important aspects, such as scalability, availability, flexibility, and adaptability [Clayman and Galis 2011]. An Intelligent power management system for such M2M/IoT devices are already studies by [Paul 2014]. Dramatic improvement in these aspects is critical for the increasing millions of interconnected things predicted by the IoT perspective. Many studies are conducted using various evolutionary algorithm including particle swarm optimization [Paul et al. 2003]. Furthermore, various database tools are available that have severe limitations regarding querying, indexing, processing, modeling, and transaction handling of various types of sensory data [Cooper 2009]. All of these factors are currently absent in existing healthcare systems.

Overall, the techniques mentioned above either follow any existing IoT architecture or design a novel architecture for a particular problem, which only covers a limited scope of healthcare. Existing systems rarely integrate new manufactured devices, which constitute an essential element in enhancing the basic components of a healthcare system in IoT. However, a desirable system should be capable of taking care of human beings (patients inside of the home, people outside of the home) from the perspective of all aspects, which not only addresses required medications, but can also perform continuous monitoring of patients and can take appropriate action in real time. The mentioned challenges can only be accomplished through rapid collection and aggregation, as well as parallel and efficient processing of incoming high-speed medical and sensory data. With the above requirements in mind, the proposed approach needs to have a communication link with only one of the control modules, such as the PMD, of its application environment. In additon, it is necessary that the u-things are not required to perform extensive computations, as most of their processing is done on the Internet or through intelligent building.

Concerning data processing, there are some solutions based on Big Data analytic engines for healthcare and medical applications [Fang et al. 2016; Muin et al. 2014; Hermon et al. 2014]. Toreador [Ardagna et al. 2016; Ardagna et al. 2017] is a pioneering H2020 project focusing on Big Data as a service that can be useful in the context of streaming processing even in a healthcare scenario. However, it is still under development and its customization for our scenario with different sources of heterogeneous sensor data, and in the context of user-centric evaluation, remains unfeasible. Evotion [Prasinos et al. 2017] is an H2020 project focused on Big Data for healthcare, and specifically hearing loss scenarios. More similar to our system, it is focused on helping users in understanding diseases according to the analytical processing of data coming from sensors, including wearable sensors. It includes a decision support system that takes advantage of Big Data analytics. However, it is still in a very early stage. [Muin et al. 2014] focused on Big Data for public health that emphasizes initial requirements and advantages from a processing perspective. [Hermon et al. 2014] used a systemic review approach to create a categorisation of Big Data use in healthcare. More recently, [Fang et al. 2016] presents a comprehensive survey on the existing challenges, techniques, and future directions for computational health informatics in the Big Data age. This paper, however, is not focused specifically on processing capabilities, but rather networking architecture and functionality. We will address analytics and processing more in detail in our future work.

3. HADOOP BASED INTELLIGENT CARE SYSTEM (HICS)

In this section, we discuss the proposed scheme, including a sensor deployment scenario, which describes the functionalities of the medical and activity sensors, followed by the communication model for Hadoop-based HICS, as well as the network architecture, including intelligent building and its algorithm.

3.1 Sensor Deployment Scenario

Healthcare systems comprise dynamic processes, including pre-treatment processing, in-treatment processing, and post-treatment processing.

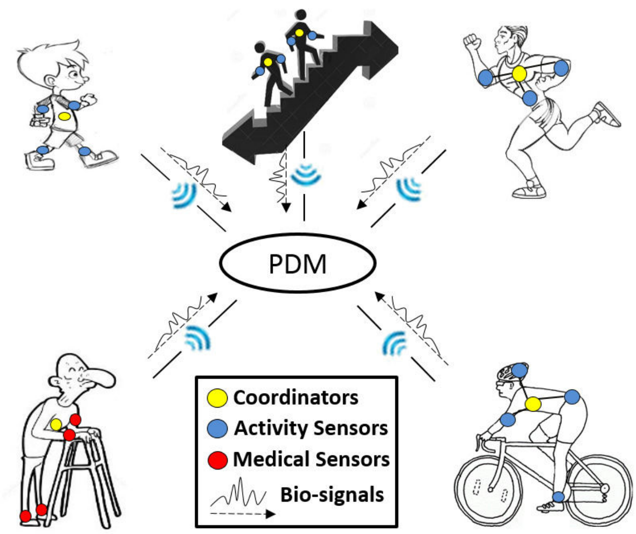

Fig. 1. Deployment scenario for the proposed system model.

In Figure 1, various sensors, such as activity sensors, medical sensors, and coordinator sensors are attached to human body parts. The human body parts include the wrist, ankle, heart, chest, head/helmet (while cycling), and other body parts. These sensors collect heterogeneous data, such as diabetic information, heart rate, and blood pressure data. In addtion, various activities are recorded, such as physical exercises, including sitting, walking, walking upstairs or downstairs, and cycling. During such activities, the sensors read the body temperature, heart beat, blood pressure, sweat level, and glucose level. In order to collect and aggregate data from various sensors and make the system energy-efficient, we have utilized a coordinator, which works as a sink node. A coordinator node acts as a relay node that collects the data and transmits them to the primary medical device (PMD). The PMD could be a cell phone application or a small wireless device/access point connected to the Internet. The communication between the PMD and the coordinator is achieved through Bluetooth or Zigbee technology.

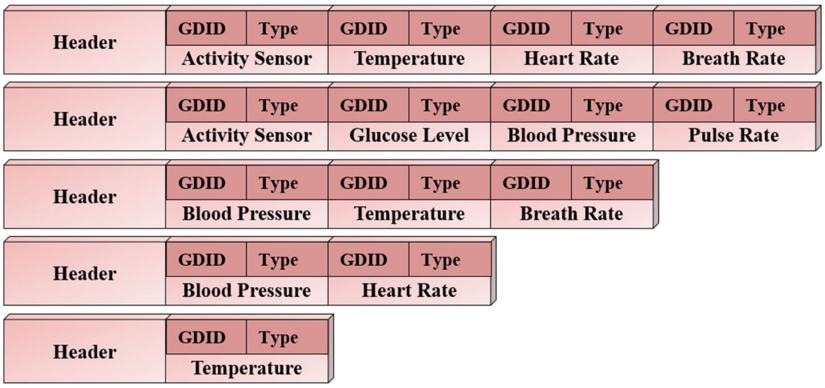

Fig. 2. Various block structures of sensor readings sent by the coordinator.

The coordinator receives health readings from various sensors attached to the body. Each sensor's frame contains three main information types: 1) GDID (globally unique device identifier), which uniquely identifies the sensor; 2) type, which identifies the type of sensor, such as glucose, blood pressure, pulse rate, etc.; and 3) value, which comprises reading of the sensor. The coordinator aggregates and then encapsulates all of the sensors' readings into a single packet/block in a proper format and sequence. It initially adds the packet header, which mainly contains the U_ID (user id), which uniquely identifies the sensor and the user. Later, it enlarges it by adding various sensor readings in the following sequence, depending on the available readings: 1) glucose level reading; 2) blood pressure reading; 3) pulse rate; 4) temperature; 5) heart rate; and 6) breath rate. Moreover, all of the readings from the various sensors are added for a particular time in a single packet. Since every sensor is not required to transmit data directly to the PMD, energy is conserved. Moreover, a number of sensors on the patient's body depend on the patient's requirement. Therefore, it is not mandatory that all of the sensors are attached to the user's body. Figure 2 shows the various block structures at the coordinator node, depending on the number of sensors attached to the body. In intelligent building, the readings in a single block are classified by the type of field.

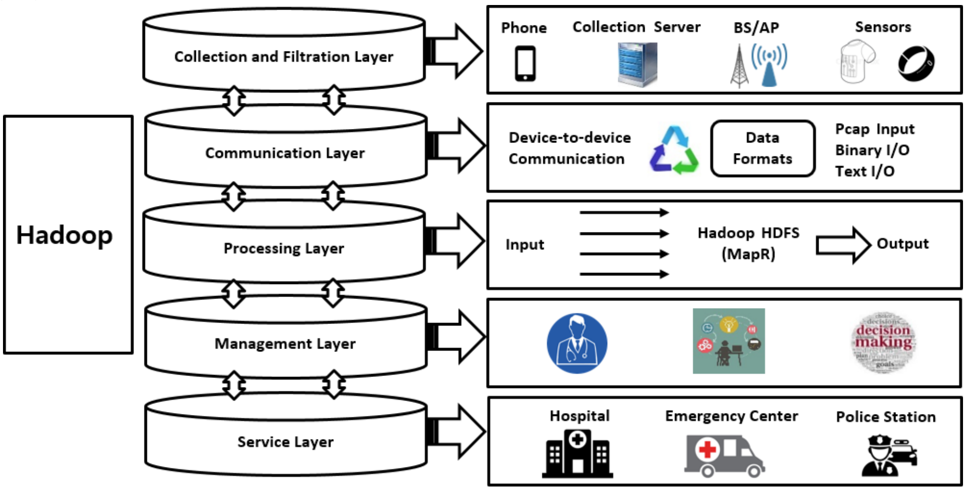

Fig. 3. Five-layer structure of the proposed healthcare IoT system.

3.2 Layered Structure of HICS

A smart solution for healthcare systems in IoT is presented in Figure 3. It consists of five functional layers: 1) a data collection layer; 2) a communication layer; 3) a processing layer; 4) a management layer; and 5) a service layer.

In the proposed layered architecture, the data collection layer (layer 1) provides data collection functionality, including data-sensing using medical sensors, data acquisition, and data buffering. It also provides the filtration service that reduces the overall volume of the data to be processed by discarding all of the unnecessary metadata and repeated readings. The communication layer (layer 2) offers end-to-end connectivity to the various devices involved in the healthcare applications. It is responsible for the transmission of data from sensors to the coordinator and to the PMD using Zigbee and Bluetooth technology, from the PMD to IB using the Internet (cellular or wired network), and from the server to the server in IB via an Ethernet. Morevoer, this layer tranforms data into a suitable format for Hadoop processing. It also translates the BAN data into various sequence files using Pcap Input, Text I/O, and Binary I/O.

The processing layer (layer 3) is a fundamental component of the IB that receives sequence files from the collection unit. It processes the data while performing necessary statistical calculations based on the nature of the data. For Big Data, to accomplish high efficiency, the overall data are disassembled into small pieces, and each of the pieces is separately processed in parallel using Hadoop HDFS and MapReduce. Afterwards, the results from each piece are acquired again and kept in the result storage area for future analysis.

The management layer (layer 4) is the smart layer of the system, comprising various medical expert systems that examine the processing layer's results and recommends corresponding actions. For example, if the patient has either high blood pressure or a low sugar level, this layer then directs the emergency division for feedback generation to the user. For such emergency actions, in which the user does experience any delay, these layers should be sufficiently efficient to rapidly generate appropriate action.

Finally, the service layer (layer 5), delivers connectivity to the end user to access various facilities, such as hospitals (e.g., remote physician support, routine medical examinations), emergency treatment (e.g., high blood pressure), ambulances (e.g., blood pressure under a predefined threshold, which can cause fatal problems), and police stations. Furthermore, doctors can also monitor the patient by continuous analysis of his or her medical history. These services enable a doctor to connect to a facility to obtain a patient's present health status.

3.3 Proposed HICS Architecture

The concept of the proposed healthcare IoT system is shown in Figure 4. Various kinds of sensors are attached to the human body that are utilized to measure blood pressure, pulse rate, human motion, diabetic status, to name but a few. To achieve a better understanding of the proposed system, we assume that all of the users are equipped with smart devices. In our example scenario, we consider a patient in a home, an elderly user, a user doing physical exercise, and others. In these scenarios, if a patient's blood pressure or other disease crosses a defined threshold, or if an elderly man has a heart attack or a user doing physical exercise requires first aid, the sensors transmit the measured health parameter readings to the agent (e.g., raspberry-pi). Raspberry-pi is a device used to convert sensor data to mobile readable data. After the conversion, the mobile readable data are forwarded to the primary mobile device (PMD) using ZigBee, Bluetooth, or IEEE 8.2.15.4. The PMD can be a smart phone or any ZigBee, Bluetooth, or IEEE 8.2.15.4 device with an Internet connection. The PMD is attached to intelligent building via the Internet (LTE/3G/WiFi). Intelligent building is a smart block used for storing and processing sensor data, and it executes certain actions depending on the context of the data. Therefore, for the above cases, intelligent building executes different actions for individual patients (to record the history of a patient in a hospital, or a change in doctor prescriptions), for elderly users (requiring a remote physician or ambulance), for people engaged in physical exercise (to provide first aid from a nearby first aid hospital), for an automobile transporting children and adults (to alert a police station), and others. A detailed explanation of intelligent building, i.e., the core component of the system, is described in the following section.

3.3.1 Intelligent Building

Intelligent building is the central component of the proposed system. The building is a complete intelligent system that handles incoming high-speed Big Data from a large number of body area sensors. It utilizes the parallel-processing paradigm of the Hadoop ecosystem. The system is responsible for collecting, processing, and analyzing health sensor data from a large number of people having sensors on their bodies by continuously monitoring health parameters. Intelligent building, which can be considered as an intelligent healthcare system, is mainly composed of a collection unit, a processing unit having a Hadoop system equipped with an intelligent medical expert system, a sensor health measurement patient database, an aggregation result unit, and application layer services.

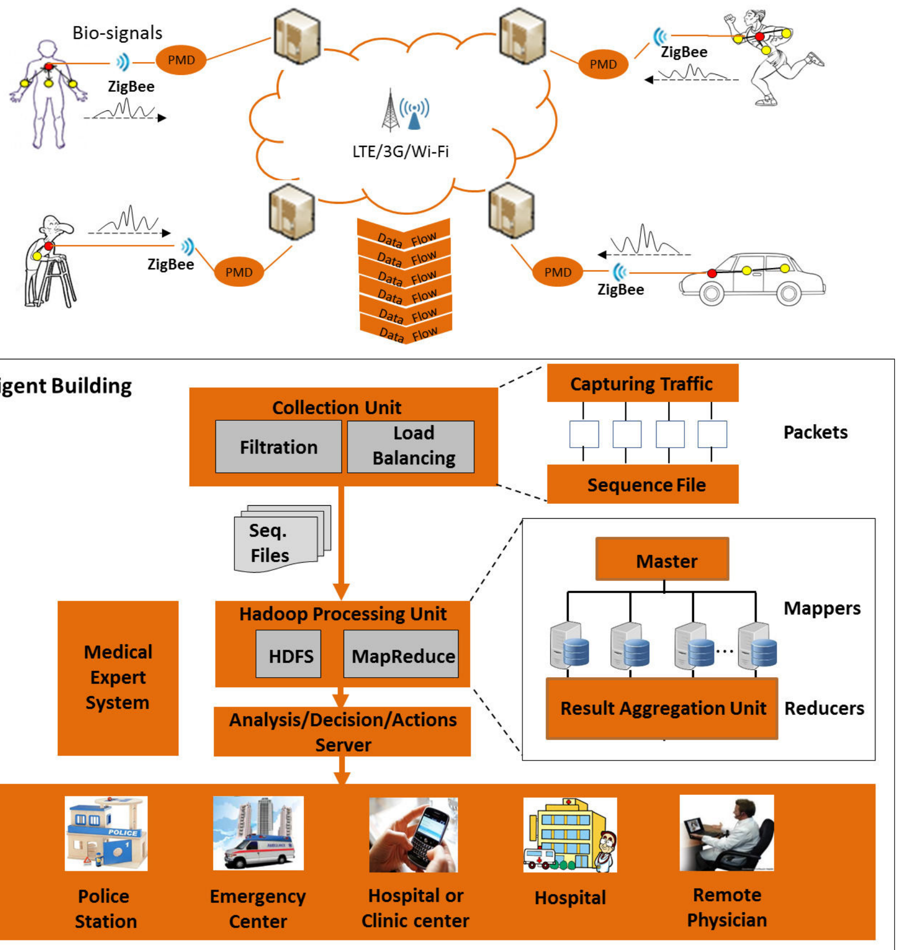

Fig. 4. Application scenario for the proposed healthcare IoT system.

The collection unit is the entry point for the incoming health data from the body area networks. It continuously collects data from each registered person in the BAN network. It might have a single server performing all functionalities of the collection, filtration and load balancing, or have multiple servers for each functionality, depending on the complexity of the system. We considered multiple servers for the collection unit, in which the collection servers collect high-speed incoming WBAN sensors. They extract the required information from each packet, such as GDID (globally unique device identifier) which uniquely identifies the sensors, U_ID (user id), and all sensor measurements contained in a single packet. Moreover, it filters all of the necessary data by discarding all unnecessary repeated information. It employs Hadoop libraries, such as *Hadoop-pcap-lib, Hadoop-pcap-serde*, and *Hadoop-pcap-input* to process the incoming packet and generate a sequence file from sensor readings encapsulated in the packet. For each distinct U_ID, one sequence file is generated. All of the readings from that user are added to the corresponding sequence file. When the sequence file has reached its size or time threshold, it is sent to the Hadoop processing unit to process the sequence file by analyzing and calculating statistical parameters. The load balancer, which is sometimes the master node in the Hadoop setup, decides which data nodes will process the sequence file. Each processing server or data node has a GDID range for which it processes the sequence file. The processing server processes and analyzes patient data that are within its GDID range.

The Hadoop processing unit comprises various master nodes and many data nodes. It uses the Hadoop distributed file system (HDFS) on various parallel data nodes to store the data in blocks. Each data node is equipped with the proposed algorithm using the MapReduce implementation to process the sequence files either by calculating statistical parameters or by analyzing sensors' health readings in a sequence file to generate intermediate results for decision-making. When the data node receives any sequence file from the collection unit, it executes the algorithm on the sequence file and generates intermediate results. Various mapper and reducer functions work in parallel to achieve high efficiency. The mapper function initially decides whether each sensor reading is normal and does not need to be analyzed deeply, or has some abnormal values that require analysis and emergency action. It compares sensor values with their corresponding normal threshold values and, if satisfied, just stores them in the database without further time-consuming analysis. When any sensor value from WBAN satisfies its corresponding serious threshold value, it generates an alert directly to the application layer for a rapid response. The application layer performs a quick action depending on the sensor, its value and patient, such as call the police in the case of an accident, or call a doctor or an ambulance in the case of a heart attack, serious diabetes, BP, or dangerous pulse reading. Apparently, when the sensor's reading is neither normal nor too serious, analysis is required. Such values are processed by calculating statistical parameters or performing other calculations, depending on the algorithm, to generate intermediate results for a final decision. Finally, the aggregation unit of the Hadoop processing system aggregates the results using the reducer from various parallel-processing servers (data nodes) and send thems to the final decision server.

Decision servers are equipped with an intelligent medical expert system, machine learning (ML) classifiers, and other medical problem- detection algorithms for further analysis and decision-making. It analyzes the current results received from the processing unit, depending on the previous history of the patient using complex medical expert systems, machine ML classifiers, to name but a few. The concrete details of the medical expert systems are beyond the scope of the current paper due to their vastness, but it is evident that a Big Data system, such as the proposed one, constitutes their fundamental basis. For instance, in [Anisetti et al. 2017] some examples of complex analytic processing suitable for medical evaluations have been proposed in the field of diabetic pathology detection and prevention using a Big Data analytic engine. In our scenarios, for the sake of completeness, we use simple machine learning classifiers, such as REPTree, in the decision server, since they are sufficiently efficient and accurate for normal, straightforward, and simple disease detection.

3.3.1.1 Proposed Algorithm

An algorithm is designed for the proposed HICS intelligent building system to process high-speed WBAN sensor data. The notations used in the proposed algorithm are described in Table 1. SAVE_DB ( ), Analyze_Data ( ), and Emergency ( ) are functions that represent some actions. SAVE_DB is responsible for storing sensor data into the database corresponding to the GDID and U_ID. The Analyze_Data function implements expert systems, ML classification algorithms, calculations of statistical parameters, as well as other complex medical problem-detection algorithms. The emergency function alerts the application and service layer to perform a rapid action in the case that a serious reading is received from patient sensors, or generates alerts based on the analysis results generated by the processing servers. Algorithm 1 shows the pseudo-code of the proposed system. Initially, the collection unit receives a packet from the PMD with all sensors' readings encapsulated in a block. The collection unit extracts all medical readings corresponding to the GDID, U_ID, and activity performed (if an activity reading exists). In the next step, these readings are added to a sequence file as a one record corresponding to the GDID and U_ID in the particular sequence file. Therefore, for each packet, there is one record that is added to a sequence file. When the sequence file reaches its particular time or size threshold, it is sent to the Hadoop processing system, i.e., the Hadoop_Processing function. Hadoop_Processing initially checks whether the medical readings are within a normal range, such as: 1) BP, not less than 70 and not greater than 140; 2) diabetes, not less than 70 and not more than 200; and 3) pulse rate, from 70-90. The normal range values are saved in the database without any further processing. On the other hand, if any one of the values from these sensors resides in a serious range, such as: 1) BP, below 40 and above 200; 2) diabetes, below 40 and above 400; 3) pulse rate, above 100 for adults and above 130 for children, as well as a limit of 0-240 bpm heart rate, 1-120 bpm breath rate, and 33-42 °C temperature. In all such serious cases, emergency action must be taken. Normal and serious threshold values for each type of sensor vary depending on the unit and intensity of their effects, as well as the activity performed while the readings are taken. Moreover, if the sensor reading is neither too serious nor too normal, then statistical calculations or other medical measurements are performed on them to generate intermediate results for final analysis and decision-making. These results are then sent to the decision server for final decisions or disease classification.

Table 1. Notations used in Algorithm 1.

Notations	**Description**
GDID	Global unique device identifier
U_ID	User identifier
Pi	Packets 'i' containing sensor measurements received from the patient
N	Number of sensors for a single patient
ʃ_t	Time duration of a sequence file
$S_K M$	Reading from a sensor K attached to a human body, where 1≤K≤N
T_{LN}^{K}	Normal lower threshold of sensor K
T_{UN}^{K}	Normal upper threshold of sensor K
T_{LS}^{K}	Serious and emergency lower threshold for sensor K
T_{US}^{K}	Serious and emergency upper threshold for sensor K
Activity_M	Activity sensor measurements
Seq_File	Sequence file generated for Hadoop processing
SAVE_DB( )	Function to save sensor data into the database
Analyze_Data( )	A generic function contains algorithms, such as a medical expert system algorithms, disease detection, BP, diabetes reading analyses, etc.
Emergency( )	Call an emergency system, such as an ambulance, emergency hospital department, or the police.

ALGORITHM 1.

INPUT: BAN medical and activity sensor readings in packets P_i where *i r*epresent the packet number.
OUTPUT: Alerts, emergency action, medication, suggestions, and prescriptions.

STEPS:

```
1.  ForEach Pi  Do
2.  |   Record := Extract (GDID, U_ID, Activity_M, S1M, S2M, S3M,…., SKM),  where 1≤K≤N,
3.  |   Seq_File(GDID, U_ID) := Seq_File(GDID, U_ID) +  Record.
4.  |   IF   time(Seq_File(GDID, U_ID)) ≥ ʆ_t || size(Seq_File(GDID, U_ID)) ≥  HDFS_Block_Size     Do
5.  |   |       Results := Hadoop_Processing(Seq_File(GDID, U_ID)).
6.  |   |       Send_Decision_Server(Results).
7.  |   END IF
8.  END ForEach
9.  FUNCTION Hadoop_Processing(Seq_File(GDID, U_ID))
10. |   START
11. |   ForEach Record in Seq_File(GDID, U_ID) Do
12. |   |       Dectect_Activity(Activity_M) && initialize → T^K_LN , T^K_UN, T^K_UN , T^K_LS , T^K_US
13. |   |       ForEach SKM in Record Do
14. |   |       |       IF (T^K_LN  ≤  SkM  ≤ T^K_UN )  DO // sensor reading is normal then just save data
    |   |       |       |       SAVE_DB().
15. |   |       |       |       Return→next record.
16. |   |       |       ELSEIF (T^K_LS  ≥   SkM ||SkM ≥   T^K_US ) Do // sensor reading is abnormal
    |   |       |       |       then call emergency services
17. |   |       |       |       Emergency ( ).
18. |   |       |       ELSE //otherwise analyze data then decide
19. |   |       |       |       Analyze_Data (SkM).
20. |   |       |       ENDIF
21. |   |       END ForEach
22. |   End ForEach
23. END FUNCTION
24. END
```

4. IMPLEMENTATION AND EVALUATION

The proposed algorithm is implemented using the Hadoop single node setup on an UBUNTU 14.04 LTS core TMi5 machine with a 3.2 GHz processor and 4 GB RAM. This section presents all of the implementation details of the proposed system. This section also provides an evaluation of the system regarding system efficiency and response time.

MapReduce is used as front-end programming with *Hadoop-pcap-lib*, *Hadoop-pcap-serde*, and *Hadoop-pcap-input* libraries for network packet processing and generating sequence files at the collection unit. The map function of our implementation maps U_ID, type, and corresponding sensor values. It compares the values with thresholds and generates action or alerts when required. Moreover, it generates intermediate results and sends them to the reduce function as U_ID as a key, and results as a value at the aggregation unit. The reducer aggregates the results, and then sorts and organizes them. Finally, a decision is made based on the results. The Hadoop MapReduce implementation of the whole system on various data nodes enables the efficient processing of the algorithm in a parallel environment. In addition, the algorithm is also implemented using simple Java programming to achieve a comparative analysis of Hadoop implementation.

Datasets are collected from the UCI [Online: UCI Machine Learning Repository: Diabetes Data Set 2015; UCI Machine Learning Repository: ICU Data Set 2015] repository and the WISDM Lab [Online: WISDM Lab: Dataset 2015; Kwapisz et al. 2010] for evaluating the efficiency of the proposed system. The UCI diabetes dataset [Online: UCI Machine Learning Repository: Diabetes Data Set 2015] comprises diabetes outdoor patient records at various time slots, such as pre-breakfast, post-breakfast, pre-lunch, post-lunch, pre-snack, pre-supper, and post-supper. The value parameter shows the diabetes value of a patient. The dataset contains 13437 records from various patients at different time points. The UCI ICU dataset [Online: UCI Machine Learning Repository: ICU Data Set 2015] holds the ICU sensed data of an 8.5-month-old, 5 kg female child patient having a biliary atresia problem, i.e., liver failure with coagulopathy. The dataset includes various types of medical measurements from different devices, such as heart rate (bpm), respiration rate (breath/ min), arterial pressure – mean (mm Hg), arterial pressure diastolic (mm Hg), arterial pressure systolic (mm Hg), arterial O2 saturation (%), tidal volume, PIP (cm H20), etc. It comprises 7931 records with various continuous timestamps. The WISDM Lab dataset [Kwapisz et al. 2010] has two files: 1) WISDM_raw, which covers more than 1048576 records from three sensors only (x-acceleration, y-acceleration, z-acceleration), and 2) WISMDM_Transformed, which is the tranformed form of WISDM_raw file with no missing value and contains 5418 records with 46 parmeters. The WISMDM dataset contains sensors' readings while performing various activities, such as *jogging*, *walking*, *walking upstairs*, and *walking downstairs*. In addition to all of these datasets, we also used other datasets containing data from different numbers of sensors to evaluate system efficiency. Moreover, health care dataset files are replayed to the system as real-time network traffic to check the real-time efficiency of the system.

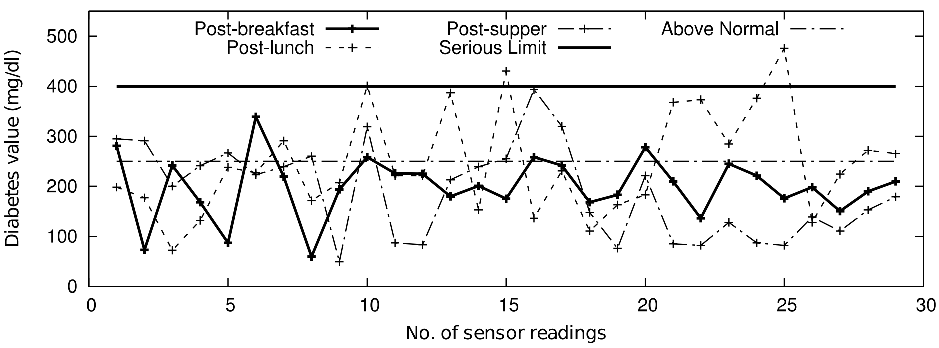

Fig. 5. Post-meal diabetes measurements of a patient.

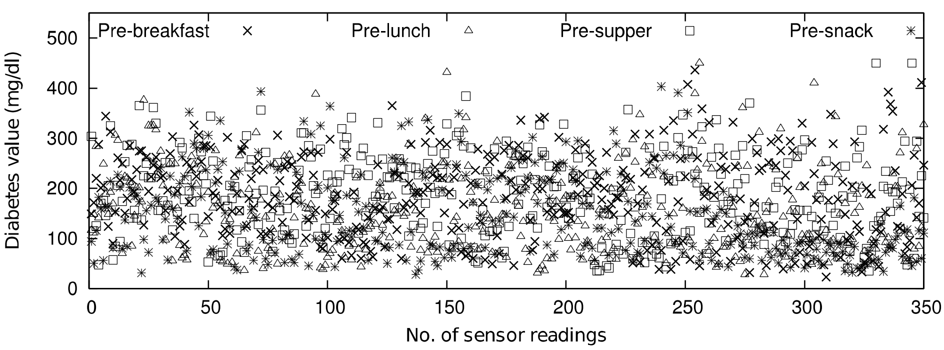

Fig. 6. Pre-meal diabetes measurements of a patient.

The aim of the following evaluations is to demonstrate the utility of the proposed system in concrete scenarios using literature databases. More specifically, we evaluate our system with respect to homogeneous data flows with a quick reaction in the case of an emergency, with: 1) contextual, but direct, rules on the data flow; 2) contextual, but complex, rules on the data flow; and 3) contextual inference from different data flows. Decisions on multiple heterogeneous flows are demanded of the decision server and to the complex correlation-based analytics, which constitute topics of our future work.

Considering the diabetes patient, blood glucose measurements are taken for analysis. A patient is monitored by considering his or her regular post-breakfast, post-lunch, post-supper, pre-breakfast, pre-lunch, pre-supper, and pre-snack glucose level in milligrams of glucose per deciliter of blood (mg/dl). The system saves data into the database, analyzes data as batch processing, or takes an emergency action based on the values received from the user/patient. Figure 5 presents a graph of the first 30 blood glucose measurements of the patient at post-breakfast, post-lunch, and post-supper. In this case, the patient is an insulin-dependent diabetes patient. For this reason, the system adapts to the context and only stores the value when it is less than 250 mg/dl by considering it as an average threshold. However, when the received measurement crosses the serious threshold, i.e., 400 mg/dl as at the 25th and 15th readings, an emergency action is taken either by admitting the patient to a hospital or requesting that he or she increase the insulin dose. Diabetes values that are in between the normal and serious threshold are analyzed, and medicine is prescribed or insulin treatment is suggested depending on the analysis results. While analyzing the post-meal patient readings, most of the post-supper glucose levels cross the average threshold, but rarely after breakfast, as shown in Figure 5. Consequently, the patient is asked to change food for dinner and is also recommended to increase insulin dosage.

The pre-meal diabetes level is also considered, as shown in Figure 6, at the pre-meal, pre-breakfast, pre-lunch, pre-super, and pre-snack diabetes level. In this case, the analysis window is greater to cover the objective of the evaluation. More specifically, the glucose test crosses the serious threshold nine times. Thus, a quick action is taken, in which the patient is recommended to increase the insulin dose immediately. Most of the abnormal actions are observed at pre-lunch and pre-supper, which shows the rise in glucose level at PM times.

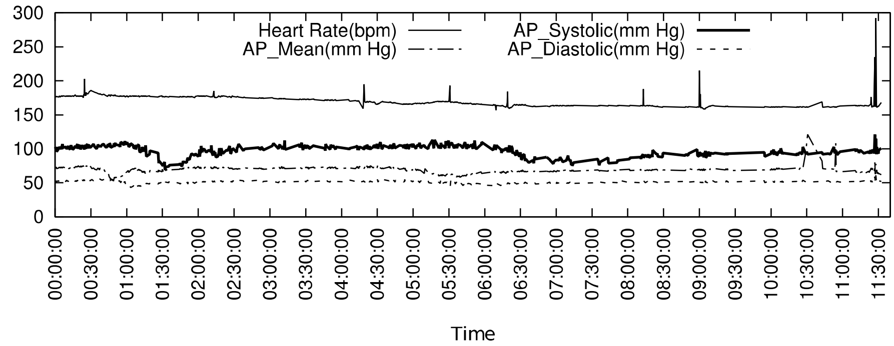

Fig. 7. ICU patient measurements.

The system also analyzed heart patient measurements related to heart rate (bpm), arterial pressure – mean (AP_mean, mm Hg), arterial pressure diastolic (AP_ Diastolic, mm Hg), and arterial pressure systolic (AP_Systolic, mm Hg). The analysis report is presented in Figure 7. The considered patient is a small child aged 8.5 months. For this reason, the average heart rate threshold is 190. Sometimes, the heart rate crosses the average threshold, such as at 0:24:41, 04:19:04, and 09:00:09. However, at that time, other sub-ordinary measures, such as AP_mean, AP_Systolic, and AP_Diastolic are normal. For this reason, emergency action is not taken. Moreover, at the time of 11:26:54 and 11:27:54, the heart rate is dramatically increased, and AP_mean, AP_Systolic, and AP_Diastolic also cross the normal threshold. Therefore, an alert is generated by the system, and an emergency call is made.

The proposed system considers various medical parameters monitored by different medical sensors while performing varied daily activities, including blood pressure and heart rate measurements while *jogging, walking, sitting, standing, walking upstairs*, and *walking downstairs*. In order to identify the user's activity, the proposed system uses decision tree based machine learning classifiers by taking user data through cell phone accelerometer as well as sensors on body parts, such as the ankle, chest, wrist, etc. More specifically, the accelerations, i.e., x-acceleration, y-acceleration, and z-acceleration are the gravitation accelerations towards the center of the Earth and can be valued as 10=1g=9.8 m/s^2 and 0=no acceleration at the x, y, and z directions. The value of x, y, and z-acceleration always lies between -20 to 20, and is shown in Figures 8, 9, 10, and 11 for a particular user while *jogging, walking, walking upstairs*, and *walking downstairs*, respectively. Every activity has distinct values for x, y, and z-accelerations. Y values are always greater than x and z in any activity. At the time of jogging, mostly x and z values overlap. The y value lies between 10 to 20, and z fluctuates from -10 to 15, and is rarely less than -10. While walking upstairs and walking downstairs, y acceleration is always greater than 0. Moreover, y is always distinct to x and z values in the case of walking. Additionally, z values are nearer to 0 but mostly fluctuate from -5 to 10. While walking upstairs, z is always nearer to 0, but mainly differs from y. In addition, x and z are always greater than -5. Although the variation between x and z is minuscule while walking downstairs, we found some abnormalities when the user is walking downstairs, as shown at the start values of x, y, and z in Figure 11.

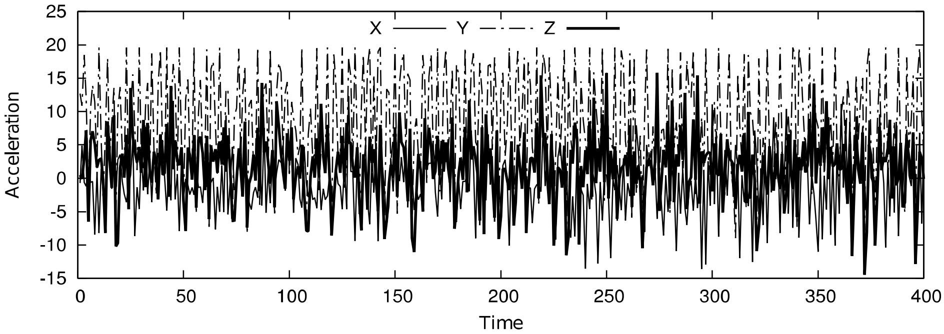

Fig. 8. Acceleration chart of a patient while jogging.

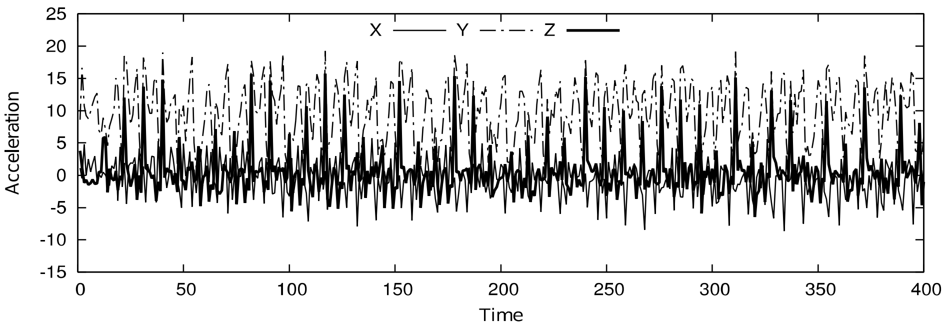

Fig. 9. Acceleration chart of a patient while walking.

For activity detection corresponding to a medical reading of a patient, we used decision-based machine learning classifiers, i.e., REPTree, Random Forest [Breiman 2001], J48 graft version [Geoffrey 1999], and simple Classification And Regression Tree (CART) [Breiman 1984]. For the training of classifiers, we took the WISMDM_Transformed dataset having x, y, and z-accelerations values from different users with 2081 samples of walking (38.4%), 1625 samples of jogging (30%), 632 samples of walking upstairs (11.7%), 528 samples of walking downstairs (9.8%), 306 samples of sitting (5.7%), and 246 samples of standing

(4.6%). Later, the x, y, and z-accelerations values are transformed using simple statistical measures into forty-three parameters [Kwapisz et al. 2010] to build a feature set for all the classifiers. The feature set contains thirty bin parameters, i.e., X0, X1, X2,….., X9, Y0, Y1, Y2, …., Y9, Z0, Z1, Z2,…….., Z9, whose values are calculated by counting the number of the fraction of accelerometer samples that lie within that particular bin. Next nine parameters are calculated by taking the 1) mean value (XAVG, YAVG,and ZAVG), 2) standard deviation (XSTANDDEV, YSTANDDEV, and ZSTANDDEV), and 3) average absolute deviation values (XABSOLDEV, YABSOLDEV, and ZABSOLDEV) of x, y, and z-accelerations. We also performed a peak value measure to obtain three more parameters, i.e., XPEAK, YPEAK, and ZPEAK by discovering the maximum value and local maxima with 10% amplitude in all three accelerations. Later, the mean value of the time elapsed between two consecutive peaks are considered.

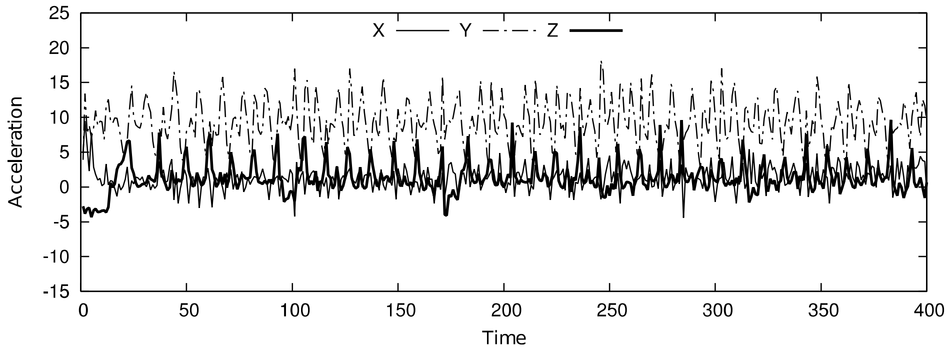

Fig. 10. Acceleration chart of a patient while walking upstairs.

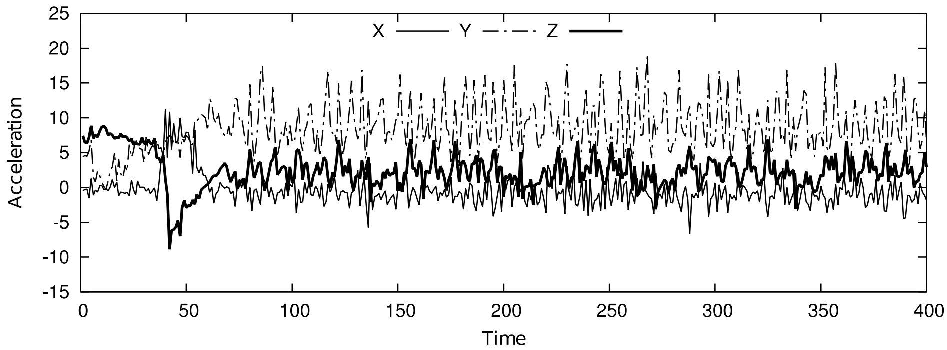

Fig. 11. Acceleration chart of a patient while going downstairs.

Finally, the result parameter is measured as: $\mathrm{AVG}(\sqrt{\sum(xi^2 + yi^2 + zi^2)})$. Using these forty-three parameters on four decision tree models, i.e., REPTree, Random Forest, J48 graft, and CART, we identified six user activities (jogging, walking, sitting, standing, walking upstairs, and walking downstairs) corresponding to users' medical readings. Out of 5418 activity records, REPTree correctly identified 4591 activities with 84.7% true positive (TP), Random Forest correctly identified 4832 activities with 89.2% true positive (TP), J48 correctly identified 4651 activities with 85.8% true positive (TP), and CART correctly identified 4661 activities with 86% true positive (TP). The overall accuracy rate of the activity detection with respect to TP is presented in Figure 12 and with respect to false positive (FP) in Figure 13. All the decision tree classifiers have very low detection rate for upstairs and downstairs and higher false positive rate for walking. Since both of the activities i.e. walking downstairs and upstairs have a fewer number of samples, so it affects the detection rate of both activities. Also, most of these two activities are also detected as walking, which increases the false positive rate of walking and decreases the TP rate of walking upstairs and downstairs. Overall, the Random Forest decision model performs outstandingly with highest TP and lowest FP. However, the Random Forest takes more time, i.e., 3610 milliseconds while building the model. Whereas, REPTree takes very short time, i.e., 140 seconds to build the model. The performance comparison of all the decision models is shown in Figure 14.

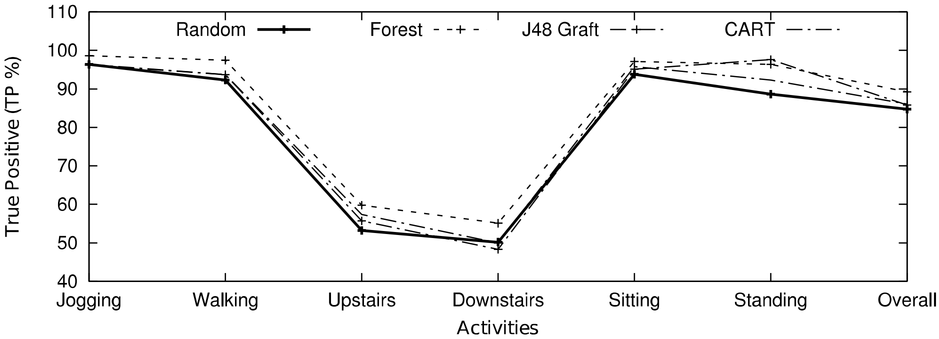

Fig. 12. Activity Detection Accuracy in terms of true positive using four decision tree models

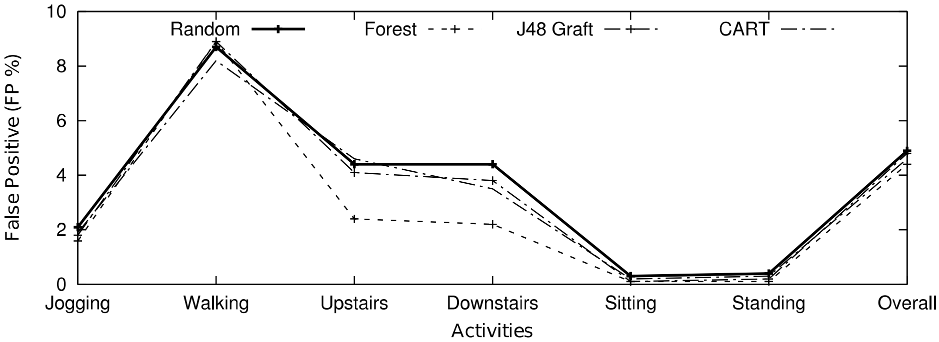

Fig. 13. Activity Detection Accuracy in terms of false positive using four decision tree models.

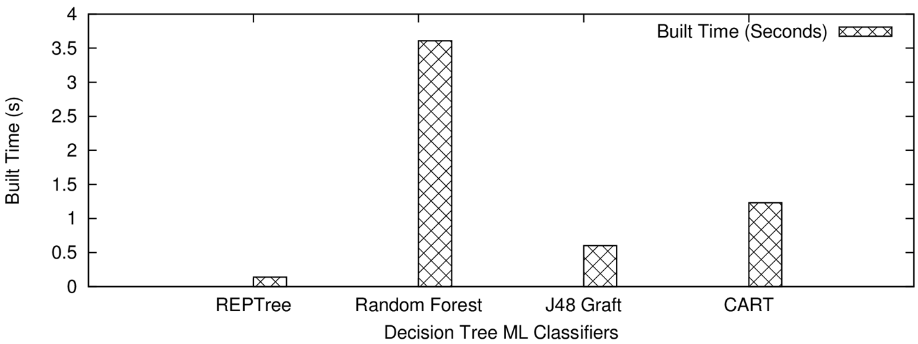

Fig. 14. Model building time for the considered Decision Tree classifiers.

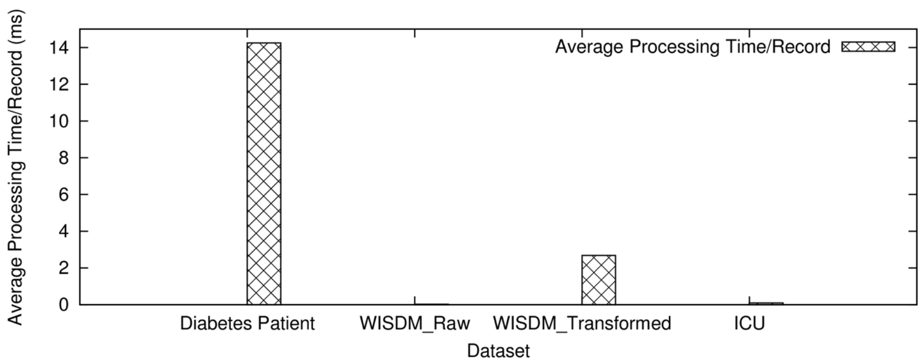

Fig. 155. Average processing time of the proposed system on various datasets.

Considering performance, the proposed system mainly focuses on the high-speed processing of a large amount of WBAN data. Consequently, we considered the average processing time taken to process one record for various datasets to evaluate the system. Figure 15 shows that the average processing time is less than 15 ms, even on a single Hadoop node setup. Most of the datasets require less than a 3 ms average processing time. This is because the diabetes datasets contain records from serious diabetes patients and the size of the record file is too small, as compared to other files. Therefore, a large amount of input, output, and switching is done because of the MapReduce function. For this reason, the processing time is greater, nearly 15 ms per record. Furthermore, we also observed that when we increased the number of sensors per record (per person or packet), the processing time was also increased, as illustrated in Figure 16. The increase in the average processing time is due to the comparative increase in overall thresholds due to the expansion of the number of sensors. The increase in average processing time corresponding to the number of sensors is presented in Figure 17.

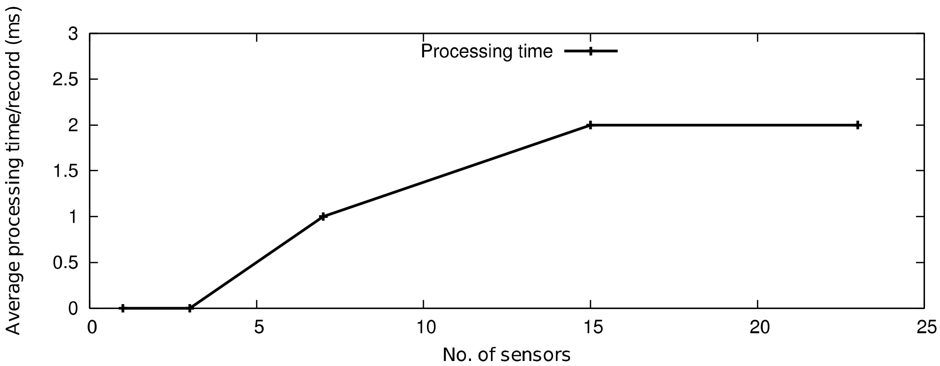

Fig. 16. Average processing time of the proposed system depending on the number of sensors per record.

We also generated a dataset with various numbers of serious readings for blood pressure, diabetes, heart rate, and temperature for several patients at different locations in a city. We then tested the corresponding system response time to generate alerts and emergency actions. Under our assumptions, we considered 100 to 1000 readings that cross serious thresholds, with location and time information, and then measured the response time of the proposed system. Figure 17 shows the number of serious readings in the dataset and the corresponding response time. For 100 serious readings within a second, the system only takes approximately 500 ms to respond and generate alerts. Moreover, with the increase of serious readings per second in the hundreds, there is a very slight increase in response time. Finally, we evaluated the Hadoop system implementation of the proposed architecture by comparing it with a simple programming implementation offering the same functionalities, but without the distribution and capabilities of the Java programming language with respect to the average processing time to process 1 MB of data. Initially for small datasets (approximately less than 100 MB), Java implementation outperforms the Hadoop MapReduce implementation. The MapReduce implementation performs parallel tasking on a single dataset by dividing it into blocks of 64 MB. On the other hand, it is not beneficial to use the MapReduce implementation for smaller size datasets, as it requires a large number of input, output, and switching operations due to the operating nature of the MapReduce functions. However, for large datasets, i.e., Big Data, it is more convenient and beneficial to use the Hadoop implementation. Figure 18 shows a comparison between both of the implementations, considering average processing time per MB. As the size of the dataset is enlarged, the Hadoop implementation requires less time to process 1 MB of data. On the other hand, the Java implementation requires more time when the file is larger, i.e., larger than 100 MB. Therefore, the average processing time is reduced while increasing the dataset size, and the processing time is increased while the dataset grows in the case of the simple programming environment.

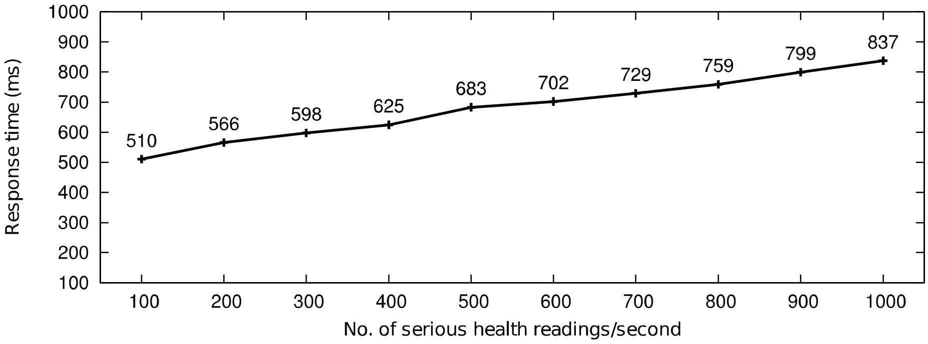

Fig. 17. Number of serious health readings and corresponding response times.

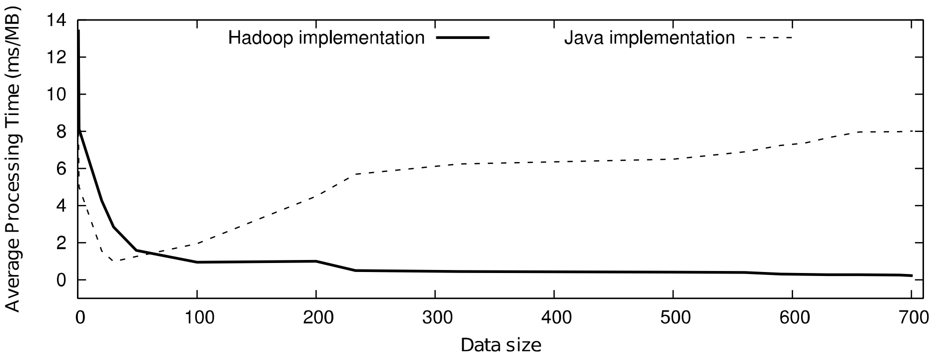

Fig. 18. Efficiency comparison between two implementations: Hadoop-based and Java-based.

5. COMPARATIVE ANALYSIS WITH EXISTING SCHEMES

This section summarizes the challenges presented in the literature review. Theoretical comparisons with existing techniques are also provided in the current section. In Table 2, a comparison is made based on the parameters that are involved in it. IoT has the capability to introduce numerous advantages over various fields. However, many other challenges are also presented by IoT in healthcare fields due to limitations in energy, reliability, security, aggregation, bandwidth, processing, memory capabilities, etc. The rest of the parameters are given in Table 2. User access to the web application and mobile application is represented by O.

Table 2. Summary and comparison of the existing systems vs. the proposed system.

Mechanisms		Issues	User Access		References
			Web	**Mob. App.**	
mHealth	Mobility	Minimizing network lifetime.		O	[Yang et al. 2014]
	Nodes interaction				
Semantic Knowledge	Learning	Security and intelligent mechanisms are required which is as same as for human communications.	O	--	[Miori and Russo 2012]
	Artificial intelligence				
m-Health	Security	Adaptability towards the needs and capabilities of the people in need of care.	O	--	[Doukas et al. 2012]
	Authentication				

VIRTUS IoT	Real-time communication Reliability	Unable to enhance the efficiency of the network due to event-driven middleware.	O	--	[Bazzani et al. 2012]
*i*Health	Fuzzy assisted data gathering Processing Energy efficient	Power constrained and with limited communication capability.	O	--	[Bhunia et al. 2014]
Personal Health Environmen t	Knowledge-based systems Aggregation Bandwidth Frame size Power consumption	A critical issue for knowledge mining, analysis, and trending.	O	--	[Jara et al. 2012]
Emergency Medical Services	Accessibility Collect Integrate flexibly	It causes the heterogeneity problem of the data format in the IoT platform.	O	--	[Xu and Da Xu 2012]
Urban planning and building smart cities	Data Collection Data aggregation Real-time analysis	It causes delay in the processing since real-time data is queued in the processor and wait to get into the processor.	O	--	[Rathore et al. 2016]
Real-time Medical Emergency Response System	Medical sensors Intelligent building Hadoop processing unit	It causes fragmentation in the data packet since heterogeneous data is generated by various sensors.	O	O	[Rathore et al. 2016]
Smartbuddy	Social Internet of Things Human behavior Real-time data analysis	A critical issue in the design of smartbuddy is to distinguish heterogeneous data, i.e., social data and network data.	O	O	[Paul et al. 2016]
Graph-based M2M	Graph based decision Energy efficiency Heterogeneous devices	It causes additional delay in recognizing heterogeneous devices	O	O	[Paul, 2013]
Smart Cyber Society	Social Internet of things Heterogeneous devices	Major issue in the design of smart society is the delay, i.e., heterogeneous devices failed to react on real-time scenario.	O	O	[Ahmad et al. 2016]
Real-time big data analysis	Feature detection Using statistical methods	It fails to extract hidden features in the data when it comes to real-time data analysis	O		[Rathore et al. 2016]
Divide-and-conquer based data analysis	Divide-and – conquer mechanism Hidden features detection River and land detection	This scheme also inefficient to detect hidden features in the continuous real-time data analysis	O	`	[Ahmad et al. 2016]`
	Energy efficient				

HICS: Hadoop-based Health Care System	Delay tolerant Data processing Data actuator Data collection Data storage Big Data processing Result aggregation	Complete solution that will benefit advanced analytic capabilities provided in future evolutions.	O	O	Proposed system

Most of the recent research in healthcare systems in IoT has aimed at improving efficiency in resource usage. Consequently, additional investigations are necessary that focus on semantic knowledge of IoT that will help to improve learning techniques, performance, energy consumption, and minimize bandwidth usage. Secondly, special techniques are required to aggregate, process, and store (if needed) data for the devices, which will not only improve network efficiency, but also increase data processing at intermediate devices. This technique is not yet employed in healthcare systems, since it depends on various characteristics of IoT communication, such as data traffic. In addition, IoT-based mobile applications also have the problem substantial overhead due to inefficient capacity regarding mobility management, and these obstacles must be overcome. Whenever a device needs to be connected to any one of various networks, such as WiFi or 3G, the device exploits the whole scenario for its advantages with respect to multiple paths and/or economic networks in terms of the usage of resources. Such technique can be employed in m-Health [Doukas et al. 2012], VIRTUS IoT [Bazzani et al. 2012], iHealth [Bhunia et al. 2014], and personal health environments [Jara et al. 2012].

In order to address the above mentioned issues, there is a need to focus on different data aggregating and exploration techniques for overall performance of optimization and efficiency. The aggregation methods should be combined with the transmission scheduling techniques because the exploration techniques focus on the reduction of size and amount of data necessary to be transmitted, such as data compression and data concatenation schemes. Furthermore, a need exists to develop a system architecture for healthcare systems in IoT that is amenable to increase or decrease the number of devices (scalability) and handle network topology (self-reconfiguration) [Xu and Da Xu 2012]. In addition, the architecture should be able to comfortably handle the enormous amount of data generated by devices used in healthcare systems since all of the devices are interconnected. Moreover, there is a need to be able to generate quick responses in the case of emergencies. Secondly, such a massive volume of data requires efficient processing, as well as enough storage devices to could hold the continuous stream of data. Similarly, smart city concept is also used to assist human beings [Rathore et al. 2016]. However, quick response to the user is not considered by the proposed scheme. Also, emergency response system based on big data is used to help elderly age people [Rathore et al. 2016]. In the proposed scheme, wireless body area network is considered, which failed to integrate various data generated by each sensor. Smartbuddy and smart city concept are based on Social Internet of Things, which is also one of the important commodity to facilitate users in the city [Paul et al. 2016, Ahmad et al. 2016]. In a similar fashion, group based M2M, real-time big data analysis, and divide-and-conquer based big data analysis is used to extract features in massive data sets [Paul 2013, Rathore et al. 2016, Ahmad et al. 2016]. However, processing efficiency is still to take on priority.To the best of our knowledge, none of the schemes presented in Table 2 provides such types of features. Therefore, we have proposed a system that can aggregate the data, efficiently process the aggregated data, and store them for future use. The proposed system also efficiently handles the delay caused by the millions of devices involved in the communication. Such delay has been overcome by using 3GPP or Wi-Fi technology, mobile applications, and timing control mechanisms.

6. CONCLUSIONS

In this paper, we proposed an intelligent healthcare IoT system using WBAN applications on Hadoop-based processing servers. The proposed system involves the different aspects of hospitals, emergency services, first aid, and police stations. The proposed network architecture comprises five network layers: 1) the data collection layer; 2) the communication layer; 3) a processing layer; 4) the management layer; and 5) the service layer. The network layers constitute the backbone of the healthcare IoT systems, and provide end-to-end connectivity to all of the connected smart devices. The applications of the proposed network architecture leverage sensors, coordinators, PMD, and intelligent building, which offers a promising solution by automatically reminding users of their prescriptions, as well as helping them in various circumstances (e.g., first aid, remote physicians, police stations, etc.). With the aim to develop continuous follow-up and monitor users' vital signs ("anytime-anywhere-anyhow"), a flexible system has been developed, which is based on intelligent building. Intelligent building receives data from various users and processes, and analyzes them using Hadoop and generates output for decision-making. Based on the output, the machine executes individual actions (e.g., first aid, remote physicians, reminding the patient about prescriptions, etc.). The performance of the system is tested and compared with the performance of a simple Java-based implementation. The final evaluations demonstrate that the performance of the proposed network architecture fulfills the needs of the users connected to it, whether the input data are real-time or offline. In the future, we are planning to provide: 1) security features for the system, such as confidentiality and user authentication; 2) privacy features for patients' data; and 3) improved analytic capabilities to support complex medical diagnoses using complex medical classifiers.

ACKNOWLEDGEMENT

This research work is supported by National Research Fund of Korea(NRF) grant funded by the Korean Government (NRF-2017R1C1B5017464)-

REFERENCES

Dan Xie Roderic A. Grupen Allen Hanson, "Context-Aware Search using Cooperative Agents in a Smart Environment", Workshop on Applications of Computer Vision (WACV), 2009, pp: 1 – 6, 10.1109/WACV.2009.5403047

Lei Tangab, Xingshe Zhoub, Christian Beckerc, Zhiwen Yub, Gregor Schielec, "Situation-based Design: a Rapid Approach for Pervasive Application Development", (2012) 9th International Conference on Ubiquitous Intelligence and Computing and 9th International Conference on Autonomic and Trusted Computing, pp 128-135, DOI:10.1109/UIC-ATC.2012.58

W. P. Liao, Y. Z. Ou, Chu, E. T. H., C. S. Shih, J. W. S. Liu, "Ubiquitous Smart Devices and Applications for Disaster Preparedness", (2012) 9th International Conference on Ubiquitous Intelligence and Computing and 9th International Conference on Autonomic and Trusted Computing, pp 764- 770, DOI: 10.1109/UICATC. 2012.98

Claudio Faris, Luci Pirmez, Falvia C. Delicato, Henrique Soares, Igor L. dos Santos, Luiz F. R. C. Carmo, "A control and decision system for smart buildings", (2013) IEEE 10th International Conference on Ubiquitous Intelligence, pp: 256-261.

Geng Yang, Li Xie, Matti Mäntysalo, Xiaolin Zhou, Zhibo Pang, Li Da Xu, Sharon Kao-Walter, Qiang Chen, and Li-Rong Zheng," A Health-IoT Platform Based on the Integration of Intelligent Packaging, Unobtrusive Bio-Sensor, and Intelligent Medicine Box", IEEE Transactions on Industrial Informatics, Vol. 10, No. 4, November 2014.

Chao-Lin Wu, Chuang-Wen You, Chun-Yuan Chen, Ching-Chi Chuang, Tsung-Chi Chiang, "Exploring the collaborative context reasoning in IoT based Intelligent are Service", IEEE 7th International Conference on Service-Oriented Computing and Applications, 978-1-4799-6833-6/14, 2014 IEEE DOI 10.1109/soca.2014.18.

S. Clayman and A. Galis, "Inox: A managed service platform for inter-connected smart objects," in Proceedings of the Workshop on Internet of Things and Service Platforms, ser. IoTSP '11, NY, USA, 2011, pp. 2:1–2:8.

M. Zhou, "Internet of things: Recent advances and applications," in Computer Supported Cooperative Work in Design (CSCWD), 2013 IEEE 17th International Conference on, June 2013, pp. 1–1.

J. A. Cooper J, "Challenges for database management in the internet of things." IETE Tech Rev, vol. 26, pp. 320–9, 2009

M. P. R. Sai Kiran, P. Rajalakshmi, Krishna Bharadwaj, Amit Acharyya, "Adaptive Rule Engine Based IoT Enabled Remote Healthcare Data Acquisition and Smart Transmission System" 2014 IEEE World Forum on Internet of Things (WF-IoT).

K. Ashton, (Jun. 2009). That 'Internet of Things' Thing, RFID J. [Online]. Available: http://www.rfidjournal.com/articles/view?4986.

S. Li, L. Xu, and X. Wang, "Compressed sensing signal and data acquisition in wireless sensor networks and Internet of Things," IEEE Trans. Ind.Informat., vol. 9, no. 4, pp. 2177–2186, Nov. 2013.

E. Welbourne et al., "Building the Internet of Things using RFID: The RFID ecosystem experience," IEEE Internet Comput., vol. 13, no. 3, pp. 48–55, Jun. 2009.

G. Kortuem, F. Kawsar, D. Fitton, and V. Sundramoorthy, "Smart objects as building blocks for the Internet of things," IEEE Internet Comput., vol. 14, no. 1, pp. 44–51, Feb. 2010.

S.-Y. Lee, L.-H. Wang, and Q. Fang, "A low-power RFID integrated circuits for intelligent healthcare systems," IEEE Trans. Inf. Technol. Biomed., vol. 14, no. 6, pp. 1387–1396, Nov. 2010.

X. Li, R. Lu, X. Liang, X. Shen, J. Chen, and X. Lin, "Smart community: An internet of things application," IEEE Commun. Mag., vol. 49, no. 11, pp. 68–75, Nov. 2011.

G. Broll, E. Rukzio, M. Paolucci, M. Wagner, A. Schmidt, and H. Hussmann, "Perci: Pervasive service interaction with the Internet of Things," IEEE Internet Comput., vol. 13, no. 6, pp. 74–81, Nov. 2009.

European Commission Information Society, (2008). Internet of Things in 2020: A Roadmap for the Future [Online]. Available: http://www.iot-visitthefuture.eu

A. Hande and E. Cem, "Wireless sensor networks for healthcare: A survey," Comput. Netw. Vol. 54, no. 15, pp. 2688–2710, Oct. 2010.

National Information Council, 2008). Global Trends 2025: A Transformed World. US Government Printing Office [Online]. Available: http://www.acus.org/publication/global-trends-2025-transformed-world

S. Li, L. Xu, and X. Wang, "A continuous biomedical signal acquisition system based on compressed sensing in body sensor networks," IEEE Trans. Ind. Informat., vol. 9, no. 3, pp. 1764–1771, Aug. 2013.

P. Castillejo, J.-F. Martinez, J. Rodriguez-Molina, and A. Cuerva, "Integration of wearable devices in a wireless sensor network for an E-health application," IEEE Wireless Commun., vol. 20, no. 4, pp. 38–49, Aug. 2013.

J. Morak, H. Kumpusch, D. Hayn, R. Modre-Osprian, and G. Schreier, "Design and evaluation of a telemonitoring concept based on NFC-enabled mobile phones and sensor devices," IEEE Trans. Inf. Technol. Biomed., vol. 16, no. 1, pp. 17–23, Jan. 2012.

J. Xing and Y. Zhu, "A survey on body area network," in Proc. WiCOM, 2009.

Riccardo Cavallari, Flavia Martelli, Ramona Rosini, Chiara Buratti, Roberto Verdone, "A Survey on Wireless Body Area Networks: Technologies and Design Challenges," IEEE Communications Surveys & Tutorials, Vol. 16, No. 3, Third Quarter 2014.

Muhammad Mahtab Alam and Elyes Ben Hamida, "Surveying Wearable Human Assistive Technology for Life and Safety Critical Applications: Standards, Challenges, and Opportunities," Sensors, Vol 14, 2014.

N. Kushalnagar, G. Montenegro, and C. Schumacher, "IPv6 over Low-Power Wireless Personal Area Networks (6LoWPANs): Overview, Assumptions, Problem Statement, and Goals," IETF RFC 4919, Aug. 2007.

G. Montenegro, N. Kushalnagar, J. Hui, and D. Culler, "Transmission of IPv6 Packets over IEEE 802.15.4 Networks," IETF RFC4944, Sept. 2007.

[Online]. Available: UCI Machine Learning Repository: Diabetes Data Set, "https://archive.ics.uci.edu/ml/datasets/Diabetes," Accessed on 31 January 2015.

[Online]. Available: UCI Machine Learning Repository: ICU Data Set, https://archive.ics.uci.edu/ml/datasets/ICU," Accessed on 31 January 2015.

[Online]. Available: WISDM Lab: Dataset, "www.cis.fordham.edu/wisdm/dataset.php," Accessed on 31 January 2015.

Lin Yang, Yanhong Ge, Wenfeng Li, Wenbi Rao, "A Home Mobile Healthcare System for Wheelchair Users", Proceedings of the IEEE 18th International Conference on Computer Supported Cooperative Work in Design, DOI: 10.1109/CSCWD.2014.6846914, Page(s): 609 – 614, 2014.

Vittorio Miori, Dario Russo, "Anticipating health hazards through an ontology-based, IoT domotic environment" Sixth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing, Page(s): 745 - 750 2012.

Charalampos Doukas, Ilias Maglogiannis, Vassiliki Koufi, Flora Malamateniou, George Vassilacopoulos, "Enabling Data Protection through PKI encryption in IoT m-Health Devices" Proceedings of the 2012 IEEE 12th International Conference on Bioinformatics & Bioengineering (BIBE), Larnaca, Cyprus, 11-13 November 2012.

Marco Bazzani, Davide Conzon, Andrea Scalera, Maurizio A. Spirito, Claudia Irene Trainito, "Enabling the IoT paradigm in e-health solutions through the VIRTUS middleware" IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications, DOI: 10.1109/TrustCom.2012.144, Page(s): 1954 – 1959, 2012.

Suman Sankar Bhunia, Sourav Kumar Dhar, Nandini Mukherjee, "iHealth: A Fuzzy approach for provisioning Intelligent Health-care system in Smart City" e-Health Pervasive Wireless Applications and Services (eHPWAS'14), DOI: 10.1109/WiMOB.2014.6962169, Page(s): 187 – 193, 2014.

Antonio J. Jara, Miguel A. Zamora, and Antonio F. Skarmeta, "Knowledge acquisition and management architecture for mobile and personal Health environments based on the Internet of Things" IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications, DOI: 10.1109/TrustCom.2012.194, Page(s): 1811 – 1818, 2012.

Boyi Xu, Li Da Xu, Senior Member, IEEE, Hongming Cai, Cheng Xie, Jingyuan Hu, and Fenglin Bu, "Ubiquitous Data Accessing Method in IoT-Based Information System for Emergency Medical Services" IEEE Transactions on Industrial Informatics, Vol. 10, No. 2, DOI: 10.1109/TII.2014.2306382, Page(s): 1578 – 1586, May 2014.

Rathore, M. Mazhar, Awais Ahmad, Anand Paul, and Seungmin Rho. "Urban planning and building smart cities based on the internet of things using big data analytics." Computer Networks 101 (2016): 63-80.

Rathore, M. Mazhar, Awais Ahmad, Anand Paul, Jiafu Wan, and Daqiang Zhang. "Real-time Medical Emergency Response System: Exploiting IoT and Big Data for Public Health." Journal of medical systems 40, no. 12 (2016): 283.

Paul, Anand, Awais Ahmad, M. Mazhar Rathore, and Sohail Jabbar. "Smartbuddy: defining human behaviors using big data analytics in social internet of things." IEEE Wireless Communications 23, no. 5 (2016): 68-74.

Paul, Anand. "Graph based M2M optimization in an IoT environment." In Proceedings of the 2013 Research in Adaptive and Convergent Systems, pp. 45-46. ACM, 2013.

Ahmad, Awais, Anand Paul, M. Mazhar Rathore, and Hangbae Chang. "Smart cyber society: Integration of capillary devices with high usability based on Cyber–Physical System." Future Generation Computer Systems 56 (2016): 493-503.

Rathore, Muhammad Mazhar Ullah, Anand Paul, Awais Ahmad, Bo-Wei Chen, Bormin Huang, and Wen Ji. "Real-time big data analytical architecture for remote sensing application." IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 8, no. 10 (2015): 4610-4621.

Ahmad, Awais, Anand Paul, and M. Mazhar Rathore. "An efficient divide-and-conquer approach for big data analytics in machine-to-machine communication." Neurocomputing 174 (2016): 439-453.

Jennifer R. Kwapisz, Gary M. Weiss and Samuel A. Moore, "Activity Recognition using Cell Phone Accelerometers", Proceedings of the Fourth International Workshop on Knowledge Discovery from Sensor Data (at KDD-10), Washington DC. http://www.cis.fordham.edu/wisdm/public_files/sensorKDD-2010.pdf

Paul A., "Real-time power management for embedded M2M using Intelligent learning methods", ACM Transaction on Embedded Computing Systems, Vol.13, Issue 5S Article 148, 2014

Paul A, TAA Victorie, AE Jeyakumar, "Particle Swarm Approach for retiming in VLSI", Circuits and Systems, 2003 IEEE 46th Mid-West Symposium on 3, pages 1532-1535.

F. Fusco and L. Deri, "High Speed Network Traffic Analysis with Commodity Multi-core Systems", ACM IMC 2010, Nov. 2010.

Fang, R., Pouyanfar, S., Yang, Y., Chen, S. C., & Iyengar, S. S, "Computational health informatics in the big data age: a survey". ACM Computing Surveys (CSUR), 49(1), 12, 2016.

Khoury, Muin J., and John PA Ioannidis. "Big data meets public health." Science 346.6213: 1054-1055. 2014,

Hermon, R., & Williams, P. A., "Big data in healthcare: What is it used for?". Australian eHealth Informatics and Security Conference, 2014.

C.A. Ardagna and P. Ceravolo and E. Damiani, "Big Data Analytics as-a-Service: Issues and challenges," in Proc. of the 3rd International Workshop on Privacy and Security of Big Data (PSBD 2016), Washington, VA, USA, December 2016.

Claudio A. Ardagna, Valerio Bellandi, Michele Bezzi, Paolo Ceravolo, Ernesto Damiani, Cedric Hebert, "A Model-Driven Methodology for Big Data Analytics-as-a-Service," in Proc. of the 6th IEEE International Congress on Big Data, Honolulu, HI, USA, June 2017.

Prasinos M., Spanoudakis G., Koutsouris D., "Towards a Model-Driven Platform for Evidence based Public Health Policy Making", SEKE 2017. Proceedings. 29th International Conference on Software Engineering and Knowledge Engineering 2017.

Marco Anisetti, Valerio Bellandi, Marco Cremonini, Ernesto Damiani, Jonatan Maggesi, "Big Data Platform for Public Health Policies". IEEE SWC 2017.

Breiman, Leo. "Random forests." Machine learning 45.1 (2001): 5-32.

Webb, Geoffrey I. "Decision tree grafting from the all-tests-but-one partition." IJCAI. Vol. 2. 1999.

Breiman, Leo, et al. Classification and regression trees. CRC press, 1984.