Computers and Electrical Engineering 70 (2018) 1034–1048

Contents lists available at ScienceDirect

Computers and Electrical Engineering

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

Intelligent hybrid remote patient-monitoring model with cloud-based framework for knowledge discovery☆

Mohammed K. Hassan[a], Ali I. El Desouky[a], Sally M. Elghamrawy[b,*], Amany M. Sarhan[c]

[a] *Department of Computer Engineering and Systems, Faculty of Engineering, Mansoura University, Egypt, 60 Elgomhouria St., Mansoura city 35516, Egypt*
[b] *Department of Computer Engineering, MISR Higher Institute for Engineering and Technology, Egypt, Mansoura Ring Road, Mansoura city 35516, Egypt*
[c] *Computer and Control Engineering Department, Faculty of Engineering, Tanta University, Egypt*

ARTICLE INFO

Article history:
Received 26 August 2017
Revised 21 February 2018
Accepted 22 February 2018
Available online 2 March 2018

Keywords:
Ambient assisted living (AAL)
Continuous remote monitoring
Context awareness
Internet of things (IoT)
Machine learning
Imbalanced datasets
Big data
Cloud computing

ABSTRACT

This paper proposes an intelligent hybrid context-aware model for patients under supervision at home that adopts a hybrid architecture with both local and cloud-based components. The cloud-based portion of the model facilitates storing and processing the big data generated by ambient assisted living systems that are used to monitor patients suffering from chronic diseases in their homes, particularly the elderly. The local portion of the model monitors patients in the event of internet disconnections or any other failure in the cloud system. The proposed model utilises context-aware techniques by monitoring physiological signals, ambient conditions, and patient activities simultaneously to derive the real-time health status of the patient. Experimental results demonstrate the effectiveness of our proposed model for monitoring patients and accurately detecting emergencies in imbalanced datasets through a case study on patients suffering from blood-pressure disorders.

1. Introduction

Machine learning has many applications in the medical field, such as ambient assisted living (AAL) systems that are responsible for supporting patients who are suffering from chronic diseases [1]. AAL is defined by its use of all concepts, methods, products, services, and medical assistance systems that enable elderly patients to live independently for as long as possible, as safely as possible. The average human lifespan has increased by five years between 2000 and 2016 to reach 71.4 years (73.8 years for females and 69.1 years for males) [2]. The direct consequences of this increase in average lifespan are a lack of caregivers to serve an increasing number of elderly patients and an increase in healthcare budgets. These problems are more severe in developing countries with high populations than in other countries because of low health budgets, poor healthcare infrastructure, and a lack of well-trained medical staff.

☆ Reviews processed and recommended for publication to the Editor-in-Chief by Guest Editor Dr. J. D. Peter.

* Corresponding author.
E-mail addresses: m_kamal1976@hotmail.com (M.K. Hassan), prof.dr.ali.eldosouky@gmail.com (A.I. El Desouky), sally_elghamrawy@ieee.org, sally@mans.edu.eg (S.M. Elghamrawy), amany_m_sarhan@yahoo.com (A.M. Sarhan).

https://doi.org/10.1016/j.compeleceng.2018.02.032
0045-7906/

In the last decade, wearable and implantable biomedical sensors have attracted significant interest because of the need to monitor vital signs, activities, or any health parameters in real time, without obstructing the movement of a patient or restricting their freedom. Wireless sensor network (WSN) technologies are the primary drivers of the rapid development in ubiquitous sensing, which enables sensors and actuators to interact seamlessly with the ambient environment and share collected information between various platforms. Internet of things (IoT) sensors have achieved breakthroughs using various technologies, such as radio-frequency identification (RFID), near field communication (NFC), and non-invasive sensors to transform the internet into a fully integrated platform [3,4].

Patient vital signs vary based on activities, temperature, humidity, smoking, sleeping patterns, and many other factors [5]. For example, when a patient's heart rate (HR) increases significantly during sleep, it is considered to be an abnormal case. In contrast, it is considered to be normal if the patient is performing exercises. Additional factors, such as patient age, gender, daily personal habits, or personal differences for the same medical issues, may affect their vital signs [6]. All of these factors emphasise the importance of context awareness, which is defined as the ability to comprehend the situational context of collected data and provide customised personal service based on patient health status [7]. The aggregation of data generated by AAL systems with high levels of abstraction during predefined time windows into a single contextual state can provide detailed information regarding the health conditions of the patient and reveal insights regarding fluctuations in the values of patient vital signs using machine learning [8].

Big data is a buzzword in current research. The characteristics of big data are described by the 'Four Vs' model (volume, variety, velocity, and veracity) according to IBM's formal definition. Data must satisfy at least one of these characteristics to be called big data [9]. The continuous monitoring of patients using AAL systems is a source of big data because monitoring duration varies from hours to years with a sampling rate may be on the order of milliseconds. Therefore, any architecture that utilises IoT sensors for continuous patient monitoring depends on cloud-based components to provide scalable data repositories and resilient computation processes for big data [10].

Traditional health care assistant (HCA) systems depend on standalone applications on a local server or handheld device, which are always customised for a specific patient suffering from a particular disease based on a set of generic rules [11]. These systems do not have flexible generic architectures to facilitate the monitoring of different patients suffering from various diseases. Therefore, some studies have proposed context-aware AAL systems for processing big data to extract knowledge regarding patient health status in real time [12].

Recently proposed cloud-based architectures depend entirely on the cloud for classifying patient health status. This means the patient is at risk when the internet connection is interrupted or a problem occurs in the cloud system [13,14]. Additionally, the performance of these models has been evaluated over imbalanced datasets using only accuracy as a metric, which yields misleading results. Furthermore, previous studies did not adopt sampling methods designed to process imbalanced datasets.

The primary motivations for this work are:

- The need to integrate generic and personalised medical rules, which are designed by medical experts to be exploited during the implementation of intelligent personal medical assistants that are aware of the individual differences between patients, which helps to minimise false alarms.
- The need for developing an intelligent hybrid model that utilises two modes to classify patient health status if the connection to the cloud is lost or the internet is disconnected. A hybrid model benefits from the advantages of both local and cloud-based architectures, while avoiding their disadvantages.
- The need for building a model that addresses the shortcomings in previous models in terms of handling imbalanced datasets.

Considering these points when developing the proposed intelligent hybrid context-aware model for patients under supervision at home (IHCAM-PUSH) helped to produce a fast, accurate, and fault tolerant-model.

The remainder of this paper is organised as follows. Section 2 summarises the background of remote patient monitoring (RPM) models and imbalanced datasets. The architecture of the proposed IHCAM-PUSH is discussed in detail in Section 3. Section 4 presents the proposed hybrid classification model (HCM). Section 5 provides a case study on monitoring patients with blood pressure (BP) disorders using IHCAM-PUSH. Section 6 presents experimental results and Section 7 discusses our final conclusions.

2. Background

AAL is an RPM model that helps physicians, nurses, and caregivers to monitor elderly patients remotely using digital technologies. RPM gathers physiological signals, ambient data, associated activities, etc. in a remote location, such as a patient's home, then transfers this data to a central healthcare provider (e.g., a smart hospital) for investigation and assessment to determine preventive actions. The inclusion of non-invasive wearable sensors in healthcare management strategies facilitates the gathering of patient physiological signals and other ambient data to be shared among different platforms to make accurate decisions regarding patient health [15].

In the last decade, there have been many studies on patient-monitoring models using mobile devices, but these models have encountered many obstacles, such as their inability to handle big datasets and a reliance on general medical rules. Models that rely on generic medical rules and ignore the ambient circumstances of vital sign readings are prone to failure and often produce false alarms [11]. The contributions of researchers in many fields, such as personalised care, continuous care, cloud-based healthcare architectures, and activity monitoring, have led to tremendous advancement in this field [13,14]. Previously proposed context-aware models suffer from deficiencies in architectural generality because each model is designed for a particular disease supports a limited number of context-aware services, and does not support big data [16]. Many attempts have been made to propose RPMs that handle big data by adopting cloud-based architectures to manage data generated by the continuous monitoring of patients [17]. Recently proposed architectures have become more comprehensive by offering context-aware cloud-based models, but they adopt entirely cloud-based architectures [13,14]. One of the significant drawbacks of these methods is the inability to monitor elderly patients when the connection to the cloud system is lost or in the internet connection is interrupted. Additionally, there are significant issues in these models when dealing with imbalanced datasets. Results demonstrating that these models have achieved high accuracy may be misleading because the classification model may only succeed in predicting patterns that belong to the majority class and fail for minority classes [13,14]. This paper proposes a novel hybrid context-aware architecture called IHCAM-PUSH to address the deficiencies in previous architectures that put the patient at risk in the event of internet failure. IHCAM-PUSH exploits a cloud architecture for storing, processing, and extracting knowledge from big datasets. It also implements a local module to monitor patient health status in the event of internet disconnection or a failure in the cloud.

3. The proposed architecture: IHCAM-PUSH

The proposed IHCAM-PUSH enables smart hospitals to monitor patients in their homes instead of sacrificing limited hospital beds. Smart hospitals are hospitals that have adopted new technologies such as IoT, cloud computing, and cloud storage, to improve patient care procedures and facilitate new capabilities, such as monitoring patients remotely in their homes [18]. The proposed model facilitates the acquisition, storage, processing, analysis, and visualisation of big data collected by AAL systems and the contextual information that is associated with it. As illustrated in Fig. 1, IHCAM-PUSH consists of the following four layers:

3.1. Layer 1: AAL

IHCAM-PUSH manages a large number of AAL systems that monitor and record vital signs, activity, medication, location, and ambient conditions (temperature, humidity, etc.). The selection of sensors, actuators, ubiquitous devices, and software services is performed based on a patient's illness. The overall performance of the model depends on having the proper setup for each patient according to their illness category. Each AAL system has a unique identifier in the model to identify a patient in the cloud architecture. Layer 1 contains the following components:

3.1.1. MySignals platform

The availability of open source electronic health (e-health) platforms on the market facilitates the collection of patient physiological signals in real time. E-health kits have different types of sensors for measuring different signals (BP, pulse, body temperature, patient position, electrocardiogram (ECG), etc.), and the ability to add custom medical sensors to create novel medical devices. These kits support many connectivity options, such as Wi-Fi, 3G, general packet radio service (GPRS), Bluetooth, IEEE 802.15.4, and ZigBee, as well as the MySignals platform [19].

3.1.2. Ambient sensors and devices

Ambient sensors, such as smoke detectors, light detectors, room temperature sensors, and humidity sensors, are embedded in the AAL layer to record readings from the ambient environment in combination with patient vital signs to study the effects of ambient circumstances on patient vital signs.

3.1.3. Data collector (DC)

This layer collects low-level data from ambient sensors and devices, and high-level data from the MySignals platform through different communication media. The DC passes low-level data to the high-level feature provider (HLFP) in the outpatient local monitoring module (OLMM) for conversion to a higher level of abstraction, while high-level data is passed directly to the high-level features aggregator (HLFA) to be aggregated with the converted data.

3.2. Layer 2: OLMM

The OLMM is a central unit responsible for collecting, processing, and aggregating the data generated by AAL systems into a single contextual state and storing it. Additionally, it monitors patient health status at home in the event of internet interruption or a failure in the cloud system. The OLMM contains the following components:

3.2.1. HLFP

The HLFP utilises different techniques for features selection, classification, and fusion to convert raw data from low-level data into a higher level of abstraction. The converted data is forwarded to the HLFA [13].

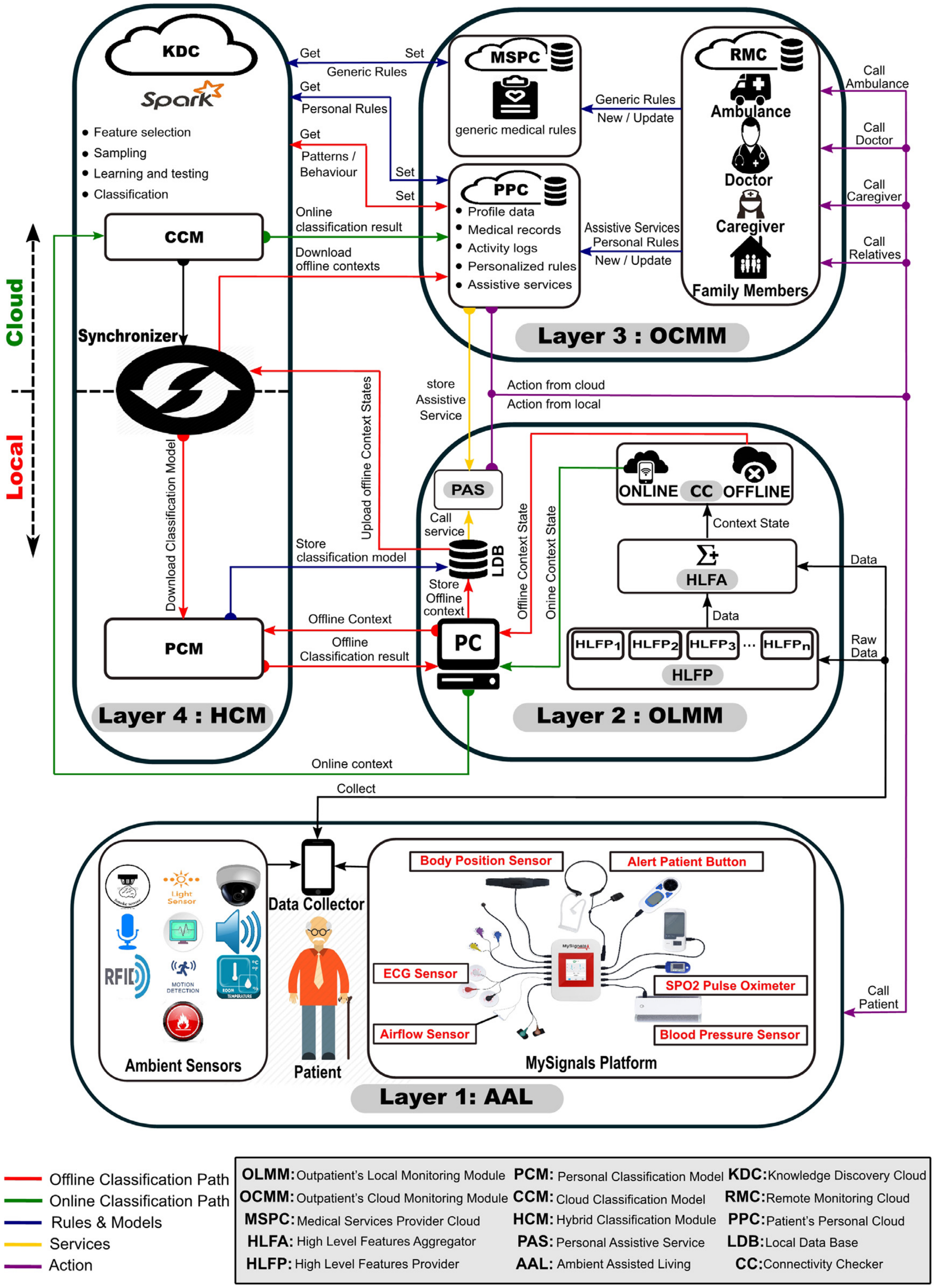

Fig. 1. The proposed IHCAM-PUSH architecture and main components of the model.

3.2.2. HLFA

The HLFA is responsible for the aggregation of converted data from different HLFP and high-level values generated by the MySignals kit into a single contextual state that contains patient physiological signals, ambient conditions, activity, etc. at a specific time. This technique provides a better understanding of vital sign fluctuations. For example, an increase in HR value while resting is abnormal, but is normal while jogging [8].

3.2.3. Connectivity checker (CC)

The CC is the unit that determines the mode that will be used for the classification of patient health status, either cloud-based or local (online/offline). The CC verifies the connection to the cloud and examines the speed of the internet connection using network traffic statistics, download speed, and upload speed. If the CC determines that the connection is valid (online), the patient's current contextual state will be transferred to the cloud to classify their current medical status. In contrast, if the CC determines that the connection is invalid (offline), the same operations will be performed locally.

3.2.4. Local database (LDB)

The LDB stores context states that arrive at the OLMM in offline mode. Furthermore, it stores copies from the classification model, which is trained and tested on the cloud to perform classification locally during periods of internet interruption. Additionally, it stores a list of predefined assistive services that can be provided to the patient based on the results of classification, which indicates the category of a patient's health status.

3.2.5. Personal assistive services (PAS)

A committee consists of social researchers, doctors, family members, caregivers, and administrative staff defines a set of assistive services for every patient. PAS are a collection of assistive services for each patient according to their health status and social circumstances, as determined by the committee. For example, in normal cases, the system will work in silent mode, but in alert cases, the system will send a message to caregivers and notify the patient. In alert cases, the system will communicate with caregivers, neighbours, relatives, or a nearby friend via phone call or short message service (SMS). In emergency cases, the assistive service will call an ambulance and the doctor in charge directly and communicate with the patient or caregivers using a microphone, speaker, and webcam to monitor the patient. All notifications can be displayed on a monitor in AAL to provide the patient with immediate instructions from doctors or general announcements from the system according to their health status.

3.3. Layer 3: outpatient cloud monitoring module (OCMM)

The OCMM acts as a personalised knowledge discovery module for every patient and is monitored by IHCAM-PUSH when the connection to the cloud system is valid according to the status of the CC (Online). The OCMM contains the following clouds:

3.3.1. Patient personal cloud (PPC)

The PPC is a personal storage area in the cloud for every patient that is supervised by the smart hospital. This repository contains context states, comprehensive information regarding the patient's personal profile (e.g., age, sex, height, weight, race), and individual rules for vital sign thresholds, which are derived from a remote monitoring cloud (RMC). Furthermore, the PPC monitors activities or behaviours, such as drinking alcohol or smoking, for every patient. Additionally, it maintains medical reports, radiological investigations, treatment plans, prescriptions, medication times, and a set of assistive services, which are recommended by the medical committee.

3.3.2. Medical services provider cloud (MSPC)

The MSPC plays the role of a medical encyclopaedia, where medical experts and doctors transfer their medical knowledge according to recent studies in the form of medical rules for every disease, including its symptoms, vital signs that must be monitored, and their ranges as generic rules. It also contains personal rules for patients if necessary. These rules are transferred to the knowledge discovery cloud (KDC) for pattern recognition and classification. Any newly discovered generic rule in the KDC will be shared with MSPC to add to or update rules that already exist. Any newly created personal rules will be synchronised with the PPC.

3.3.3. RMC

The RMC contains persons or entities that are assigned by the smart hospital to help and monitor the patient in the event of health deterioration. Doctors and medical experts transfer their medical knowledge to the PPC in the form of personal medical rules or the MSPC in the form of generic medical rules. Furthermore, doctors monitor the patient remotely and respond to alerts that are generated by the system to take immediate action.

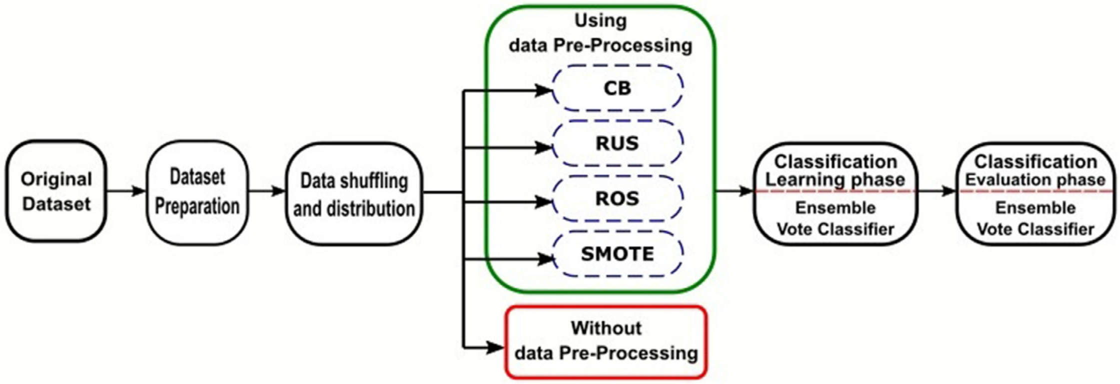

Fig. 2. The proposed technique for learning and evaluating using the CCM.

3.4. Layer 4: HCM

The HCM is a hybrid module that contains components on the cloud and locally. This architecture aims to achieve superior classification results by minimising false alerts by considering the advantages of cloud and local architectures and avoiding their disadvantages [13,14]. The cloud-based module will facilitate the accommodation of the massive number of patterns required for the classification of patient health status. The local module will enable the system to preserve patient life in the event of internet disconnection or loss of connection to the cloud. This module contains the following components:

3.4.1. KDC

The KDC is the mastermind our framework and consists of various clouds with large storage capacity, which can accommodate all contexts gathered from continuous monitoring in AAL. The knowledge discovery process is accomplished in this architecture vertically by converting low-level data into a higher level of abstraction via the HLFP. The KDC then aggregates these data with contextual and medical information to extract knowledge regarding patient health status using machine learning. This cloud uses Spark to distribute processing and storage across different clusters to speed up the classification process.

3.4.2. Cloud classification model (CCM)

The CCM classifies incoming context states from the patient on the cloud (online mode) to predict their health status. The proposed classification technique aims to maximise classification efficiency according to various measures and minimise the elapsed time (this will be presented in details in Section 4).

3.4.3. Personal classification model (PCM)

The PCM classifies the context states of the patient on the local side (offline mode according to CC) when the internet connection is interrupted or contact with the cloud is lost. It sends the results of classification to the PAS to take appropriate action in AAL and/or RMC.

4. The proposed HCM

The HCM utilises two classifiers, one of them is the CCM that works on the cloud side (online mode) and the other is the PCM that works on the local side (offline mode).

4.1. The proposed CCM

The proposed CCM is composed of five stages, as depicted in Fig. 2. The five stages are implemented using the Waikato environment for knowledge analysis (Weka) and its packages, as well as Spark-Weka.

- **Implementation using Weka**
  1. **Weka**

  Weka is an environment containing a collection of machine learning algorithms that can perform many data mining tasks, such as association rules, classification, clustering, regression, data pre-processing, and visualisation [20].

  2. **The Distributed Weka Packages**

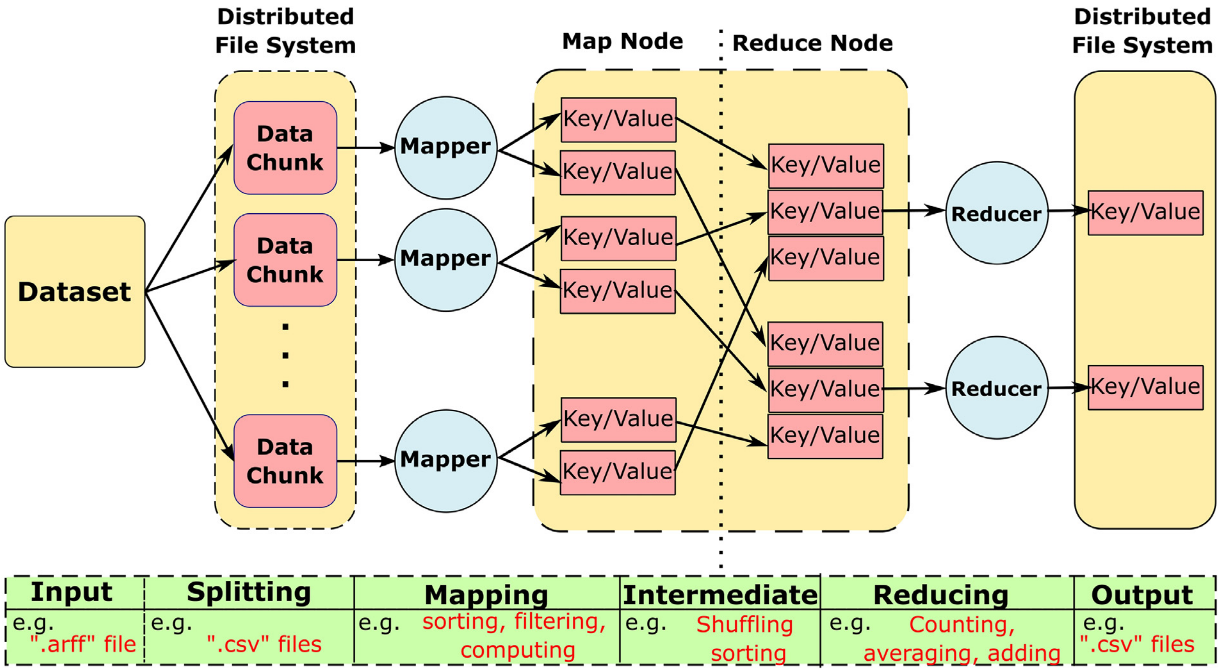

Fig. 3. The MapReduce programming model used by the Distributed Weka package.

A plugin is utilised to allow the Weka algorithms to run on a cluster of machines. Data is divided over nodes into multiple processing clusters, such as Hadoop clusters, and each chunk of the dataset is processed independently of the other pieces. Distributed Weka contains the following two packages [21]:

- **Distributed Weka Base Package**

  This package provides general MapReduce tasks for machine learning that are not tied to any particular map-reduce framework implementation, as shown in Fig. 3. It includes tasks for computing summary statistics, training classifiers, and clustering.

- **Distributed Weka Spark package**

  This package is a wrapper for the base tasks that works on the Spark platform using a particular implementation of the MapReduce programming model. It contains Spark classes that are required for running Spark in local mode and sourcing data from a local file system. This mode launches workers in separate threads to exploit all the cores in a desktop machine.

4.1.1. Stage 1: dataset preparation (Splitting data)

In this stage, the '*arff header Spark job*' split datasets into two files: one that contains a header section in a file with the extension '.names' and another with the extension '.csv' (comma separated value) that contains the dataset itself without a header. The dataset in the '.*csv*' file will be partitioned into n data chunks by configuring the '*arff header Spark job*' for n partitions, where $n = \{1, 2, 3\ldots\}$. Additionally, the '*arff header Spark job*' sets the number of cores of a computer's processor that are used to process data chunks in parallel or utilises real Hadoop clusters [20].

4.1.2. Stage 2: dataset shuffling

In this stage, the '*Randomly Shuffle Data Spark Job*' is used to randomly shuffle and stratify input data with handling for minority classes. If a minority class has samples less than the number of data chunks (e.g., the dataset has two records with minority classes and will be partitioned into four partitions), then these samples will be copied into others data chunks to ensure that each class is represented in every data chunk at least by one sample.

4.1.3. Stage 3: data pre-processing

In this stage, data is processed using various well-known sampling methods, such as random over-sampling (ROS), random under-sampling (RUS), synthetic minority over-sampling technique (SMOTE), and class balancer (CB). Data pre-processing is performed in parallel, where every sampling method is applied to every data chunk.

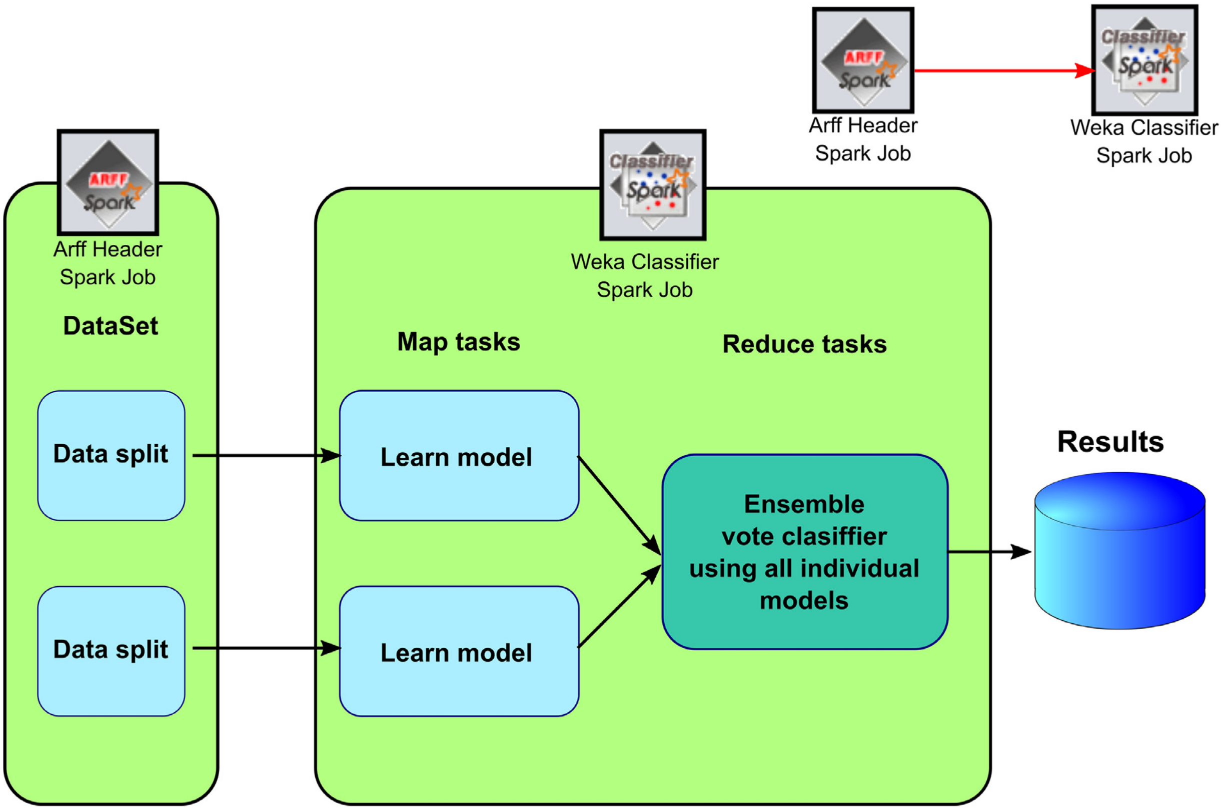

Fig. 4. Stage 4: learning phase of the CCM classifier using the '*Weka Classifier Spark job*'.

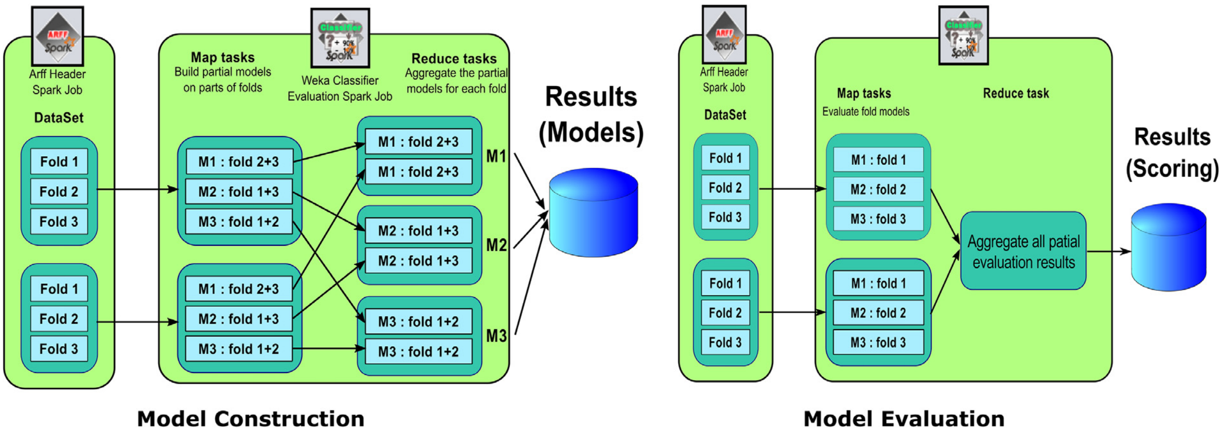

Fig. 5. Stage 5: evaluation phase of the CCM classifier using the '*Weka Classifier Evaluation Spark job*'.

4.1.4. Stage 4: learning phase

In this stage, the '*Weka Classifier Spark job*' is used to implement the learning phase of the proposed CCM classifier. As shown in Fig. 4, the map portion of the '*Weka Classifier Spark job*' is used to train the classifier on each data chunk, which produces a number of classification models equal to the number of data chunks. The reduce portion will aggregate the individual models using a voted ensemble classifier (taking votes to select the best model among all data chunks or using direct aggregation for classifiers that can be aggregated, such as Naïve Bayes' .)

4.1.5. Stage 5: evaluation phase

In this stage, each classifier is evaluated using ten-fold cross-validation from the '*Classifier Evaluation Spark Job*' . As illustrated in Fig. 5, this phase is implemented in two steps: one for construction of the model and a second for evaluation of the model.

Table 1
Sample patient dataset.

Timestamp	HR	SBP	DBP	MBP	RR	SPO_2	Temperature	Activity	Last Activity	Medication	Symptoms	Class
23-03-16 0:00	78	159	91	106	19	100	0	1	1	1	3	Warning
23-03-16 2:45	102	144	61	111	10	100	0	2	2	0	17	Alert
23-03-16 3:00	60	146	81	96	18	100	0	2	2	0	0	Normal
24-03-16 0:00	62	181	91	104	23	100	1	1	1	0	1	Emergency

4.2. The proposed PCM

After the evaluation process is completed for all classifiers, the best CCM and best sampling method are copied and transferred to the local portion of the HCM and labelled as the PCM. The PCM is used for classifying incoming contexts on the local side in the event of internet interruption (offline mode).

5. Case study

This case study was performed to determine the validity of IHCAM-PUSH and its components for monitoring patients suffering from chronic diseases, in addition to the following objectives:

5.1. Case study description and objectives

- IHCAM-PUSH will be used to monitor three patients suffering from BP disorders (hypertensive patient P1, hypotensive patient P2, and normotensive patient P3) with a measurement every 15 min for one year for each patient as a source of big data. The generated big data will be used verify the ability of IHCAM-PUSH to build learning models using a distributed cloud model (Weka-Spark) to speed up classification and provide instant, accurate results.
- IHCAM-PUSH collects patient context states, including vital signs, associated activity, and ambient conditions, to create a context-aware model. We then classify the health statuses of the patients (normal, warning, alert, or emergency) to verify the ability of the system to predict a correct health status.
- The proposed HCM will be tested on these big datasets using different classifiers and sampling methods to verify its efficiency in monitoring patients in dual mode (online/offline) and its ability to handle imbalanced datasets.

5.2. The initial setup

IHCAM-PUSH is a general framework that can be used to monitor any patient with any disease using IoT sensors in AAL. Because this study applies to patients with chronic BP disorders, the following points must be taken into consideration:

- Patient vital signs indicate the general health status of the patient. The vital signs that must to be monitored for a specific patient with a particular disease are determined by doctors and medical experts in the RMC in layer three.
- In this case study, the vital signs that were monitored for the patients are HR, systolic blood pressure (SBP), diastolic blood pressure (DBP), mean blood pressure (MBP), respiratory rate (RR), and saturation of peripheral oxygen (SPO_2).
- Because IHCAM-PUSH is intended to be a context-aware model, vital signs will be collected along with the activity of the patient, ambient conditions (e.g., room temperature), and personal behaviour (e.g., taking medications), which are examples of different factors that can affect vital signs. This technique will enable IHCAM-PUSH to comprehend the health status of the patient.
- The setup of AAL in layer one will be performed according to the vital signs and ambient conditions that must be monitored by adopting suitable IoT sensors.
- The MSPC contains generic medical rules, such as BP ranges, HR ranges, the relationship between age and maximum HR, relationship between HR and the intensity of activity and age [22], and relationship between age and the normal ranges of vital signs [23]. All of these rules can be stored in the medical encyclopaedia on the MSPC and categorised according to each illness type.
- The PPC contains the personal and medical data of the patient, such as the patient profile, medications, x-rays, medical tests, personal medical rules, personal contacts, and set of assistive services.

5.3. Synthetic data generation

As far as we know, no real dataset that contains sensor data from the long-term monitoring of elderly patients who suffer from DP disorders exists with the required format for the case study, as shown in Table 1. Therefore, synthetic datasets will be generated based on real vital signs produced from eMedical IoT kits (MySignals) in AAL. Vital signs are taken from the Physionet MIMIC-II database for three real patients over one or more days (hypertensive, hypotensive, and normotensive) [24]. The synthetic dataset contains vital signs, ambient conditions, activities associated with vital sign readings, and

Table 2
Symptoms according to patient categories in the experiment.

Type	Symptoms	Value (binary)
Hypertensive	Headache and anxiety; fatigue, severe headache and anxiety; pounding in your chest, neck, or ears; vision problems and confusion; chest pain and difficulty in breathing	6 bit binary (value: 0–63)
Hypotensive	Lack of concentration; fatigue; blurred vision; dizziness; rapid shallow breathing; fainting	6 bit binary (value: 0–63)
Normotensive	Uncomfortable; anxiety; headache; fatigue; severe headache; dizziness	6 bit binary (value: 0 –63)

Table 3
Types and ranges of attributes used in the experiment [23].

Name	Attributes	Type	Range value
Vital signs	HR	Numeric	[30–200]
	SBP	Numeric	[50–230]
	DBP	Numeric	[30–140]
	RR	Numeric	[5–30]
	SPO_2	Numeric	[40–100]
Activity	Current activity/ last activity	Resting	1
		Sleeping	2
		Walking	3
		Eating	4
		Exercising	5
		Household	6
Ambient conditions	Room temperature	Normal	0
		Hot	1
		Cold	2
Medication	Taken or not	Boolean	0 or 1
Symptoms	Symptoms	Boolean	[0–63]

Table 4
Situational classification according medical model.

Class	Classification
Normal	HR, SBP, DBP, RR, and SPO_2 values all in expected range based on current activity and symptoms = 0
Warning	Any of HR, SBP, DBP, RR, or SPO_2 in the warning range, or medications not taken, or symptoms > 0
Alert	Any of HR, SBP, DBP, RR, or SPO_2 in the alert range or more than two vital signs in warning range and (medications not taken or symptoms > 0)
Emergency	Any of HR, SBP, DBP, RR, or SPO_2 in the emergency range or more than two vital signs in alert range and (medications not taken or symptoms > 0)

behavioural information. This dataset will simulate long-term monitoring with a uniform sampling rate (one year with sampling every 15 min).

The synthetic dataset takes the following criteria into consideration:

- The correlation between vital signals and activities are taken into account based on the timing of each activity (e.g., HR is higher when the patient plays sports and lower during relaxation, eating at 3 p.m., and sleeping at 2 a.m.) [22].
- The effects of ambient conditions and medications on vital signs.
- The relationship between patient physiological signs and symptoms of their illness.

MATLAB was used to generate datasets for three different patients (hypotensive, hypertensive, and normotensive) with the same distributions as the real datasets in Physionet MIMIC-II. The synthetic generation of datasets is reliable for generating big data that is similar to real data to simulate monitoring for extended periods (e.g., one year or more). This reliability has been proven in previous studies on biomedical data analysis [14]. Some abnormal vital signs were added without modifying other context values to represent emergency cases.

The possible symptoms of each category of BP disorder are listed in Table 2. Table 3 lists all the types and ranges of all attributes in the synthetic datasets that were used in this experiment [23]. The situational classification model that was used to predict classes according to personalised medical rules is defined in Table 4. Table 5 lists the set of assistive services that will be provided to the patients according to their health statuses (the predicted class).

Table 5
Examples of assistive services.

Case	Action
Class = Normal	Do nothing
Class = Warning	Warn patient via mobile phone, monitor, or SMS
Class = Alert	SMS or phone call to the physician-in-charge to review the case
Class = Emergency	Call ambulance directly or after confirmation from physician-in-charge
Medication = 0	Alert the patient or caregiver

Table 6
Comparison between classifications with IHCAM-PUSH versus generic rules for three patients over one year.

Patient	No. of contexts	Generic rules		IHCAM-PUSH			
		Normal	Abnormal	Normal	Warning	Alert	Emergency
P1 (Hypertensive)	35,232	2	35,230	9307	23,347	2404	174
P2 (Hypotensive)	35,232	3	35,229	19,455	14,003	1627	147
P3 (Normotensive)	35,232	1	35,231	12,517	21,421	1186	108

Table 7
Hardware and software specifications.

Name	Detailed settings	Name	Detailed settings
Hardware		Software	
CPU	Intel ® Core ™ I5 3317U	Operating system	Windows 10 64 bit
Frequency	1.7 GHz	Software	MATLAB R2016b (9.1) 64 bit
RAM	6 GB		Weka 3.8.1
			Plugins:
			DistributedWekaBase version (1.0.17)
			DistributedWekaSpark version (1.0.9)
Hard drive	1 TB		*SMOTE* version (1.0.3)

5.4. Dataset exploration

As shown in Table 6, it is evident that when general medical rules are applied to classify context states into normal and abnormal classes, they fail to classify many cases, which puts the patient at risk and generates many false alarms. IHCAM-PUSH has succeeded in classifying patients into different categories with the aid of context awareness, which helps the system to understand the real health situation by considering associated activities and ambient conditions to provide results that are more accurate. It is clear that datasets that are imbalanced must be processed using various sampling methods. After the completion of dataset generation, different data-processing methods and data-mining techniques will be applied using Weka and Spark [20].

5.5. Tools and configurations

All experiments were performed on the same PC with the specifications listed in Table 7. The last version of Weka must be installed before installing the Distributed Weka Base package and Distributed Weka Spark package [20].

6. Results and discussion

Experiments were performed to analyse the performance of the proposed IHCAM-PUSH with different types of classifiers. Accuracy, overall F-measure, F-measure for the emergency class, and time elapsed for each experiment will be used to compare these classifiers and select the best classifier for IHCAM-PUSH. The F-measure for the emergency class has the priority in evaluation for detecting abnormal cases and minimising false alerts. Next, F-measure, accuracy, and time elapsed are the most important, in descending order. The accuracy and F-Measures are calculated using Eqs. (1)–(4) [25].

$$\text{Accuracy} = (\text{TP} + \text{TN})/(\text{TP} + \text{TN} + \text{FP} + \text{FN}) \tag{1}$$

$$\text{Recall} = \text{TP}/(\text{TP} + \text{FN}) \tag{2}$$

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP}) \tag{3}$$

Table 8
Comparison of classifiers with and without sampling techniques for patients over one year.

Classifier	P1 (Hypertensive Patient)				P2 (Hypotensive Patient)				P3 (Normal Patient)			
	Accuracy	F-measure	F (Emergency)	Time (Sec)	Accuracy	F-measure	F-(Emergency)	Time (Sec)	Accuracy	F-measure	F-(Emergency)	Time(Sec)
JRip	**99.9**	**1**	**1**	**10**	99.9	1	1	14	**99.9**	**0.99**	**0.96**	**18**
JRip +CB	99.9	1	1	10	99.9	1	1	10	99.7	0.99	0.85	10
JRip +RUS	99.9	0.99	1	8	**99.9**	**1**	**1**	**9**	99.1	0.99	0.7	9
JRip +ROS	99.9	1	1	17	99.9	1	1	19	99.9	0.99	0.92	39
JRip +SMOTE	99.9	1	1	10	99.9	1	1	12	99.9	0.99	0.92	23
NB	92.6	0.93	0.92	7	91.1	0.91	0.97	6	**96.4**	**0.96**	**0.74**	**8**
NB +CB	**92.1**	**0.92**	**0.98**	**7**	90.7	0.91	1	7	90.8	0.92	0.28	7
NB +RUS	92	0.92	0.98	7	90.7	0.91	1	7	90.6	0.92	0.27	7
NB +ROS	92.1	0.92	0.97	7	90.7	0.91	1	7	90.7	0.92	0.28	8
NB +SMOTE	91.5	0.92	0.96	7	**91.8**	**0.92**	**1**	**7**	95.9	0.96	0.68	11
SVM	84.3	0.84	0.79	38	86.4	0.86	0.26	70	91	0.91	0.65	55
SVM +CB	80.5	0.82	0.97	34	85.2	0.85	0.6	34	83.2	0.84	0.44	34
SVM +RUS	76.9	0.78	0.98	11	81.5	0.81	0.62	16	81.4	0.82	0.4	16
SVM +ROS	**82.9**	**0.84**	**0.98**	**88**	**88**	**0.88**	**0.79**	**169**	84.6	0.86	0.49	174
SVM +SMOTE	84.4	0.84	0.94	38	86.5	0.86	0.71	68	**91.1**	**0.91**	**0.73**	**51**
J48	99.9	1	0.98	10	99.9	1	0.99	20	99.9	0.99	0.95	8
J48 +CB	***99.9***	***1***	***1***	***8***	99.9	1	0.99	8	99.8	0.99	0.9	8
J48 +RUS	99.8	0.99	1	6	***99.9***	***1***	***1***	***8***	99.2	0.99	0.76	8
J48 +ROS	99.9	1	1	9	99.9	1	1	10	99.9	0.99	0.93	10
J48 +SMOTE	99.9	1	0.99	8	99.9	1	1	9	**99.9**	**1**	**0.95**	**13**
RF	99.9	0.99	0.99	16	99.9	1	0.99	18	99.9	0.99	0.90	18
RF +CB	**99.9**	**1**	**1**	**12**	99.9	1	1	12	***99.7***	***0.99***	***0.98***	***12***
RF +RUS	99.9	0.99	1	9	**99.9**	**1**	**1**	**10**	99.3	0.99	0.75	11
RF +ROS	99.9	1	1	20	99.9	1	1	21	99.8	0.99	0.97	23
RF +SMOTE	99.9	1	1	15	99.9	1	0.99	21	99.9	0.99	0.94	22
IBK	94.3	0.94	0.78	48	93.2	0.93	0.52	44	92.9	0.93	0.72	47
IBK +CB	94.2	0.94	0.84	47	**95.4**	**0.95**	**0.58**	**47**	91.3	0.91	0.74	47
IBK +RUS	81.3	0.82	0.85	28	86.2	0.86	0.57	27	82.8	0.83	0.74	33
IBK +ROS	89.6	0.9	0.84	93	91.7	0.92	0.57	92	89	0.89	0.75	107
IBK +SMOTE	**94.5**	**0.95**	**0.85**	**47**	92.9	0.93	0.64	43	**92.8**	**0.93**	**0.79**	**50**

$$\mathrm{F-Measure} = 2^{*}(\mathrm{Precision}^{*}\,\mathrm{Recall})/(\mathrm{Precision} + \mathrm{Recall}) \tag{4}$$

where TP = true positives, FP = false positives, TN = true negatives, and FN = false negatives.

Table 8 details the performance of six classifiers with and without sampling methods for a one-year time period using the proposed CCM. The best performance among all versions of every classifier (with and without sampling) is indicated in bold font and the winning algorithm for each dataset is indicated in italic bold font. Figs. 6, 7, and 8 present a comparison between six classifiers in terms of accuracy for datasets P1, P2, and P3, respectively. As shown in Table 9, the average rank of the classifiers over the different datasets shows that decision tree (J48), random forest (RF), ripper classifier (JRip), and naïve Bayes' (NB) are the best classifiers that yield the highest success rates in IHCAM-PUSH. It is evident that the elapsed time with the support vector machine (SVM) classifier increases dramatically with big datasets. The worst classifiers are the SVM and nearest neighbour (IBK) classifiers. It is clear that the SVM classifier consumes more time than any other classifier.

Rule-based classifiers and decision tree classifiers have another advantage in that the generated models are readable and can be written in the form of "If-Then" rules. This advantage enables the medical team to comprehend the model and approve it or update generic or specific medical rules, and then retrain the classification model. As shown in Table 9, the best sampling techniques are CB and SMOTE, while the worst is RUS. These results verify the ability of IHCAM to classify patient health status with high precision. These results indicate that we should use J48 or JRip in IHCAM and adopt SMOTE or CB as sampling methods for imbalanced datasets.

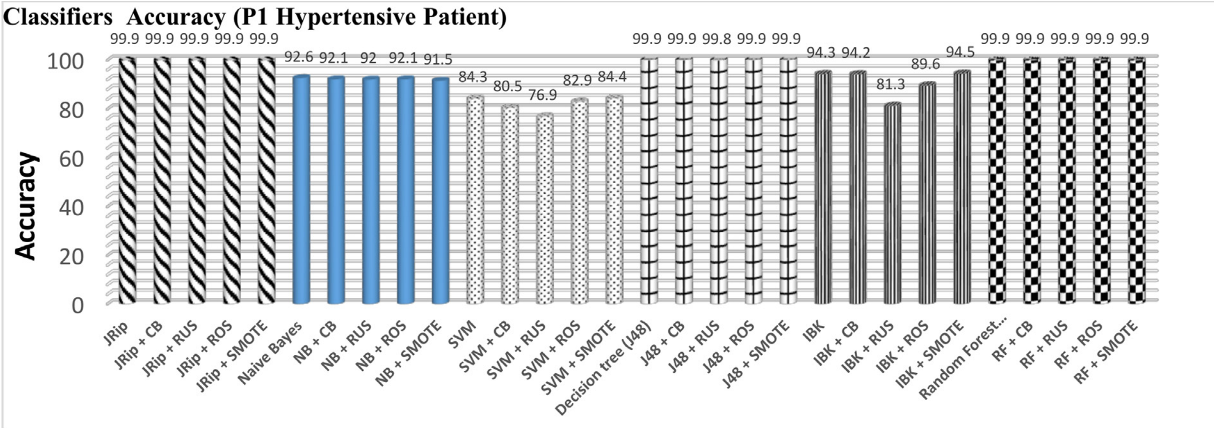

Fig. 6. Classifier accuracy with and without sampling methods for P1 hypertensive patient over one year.

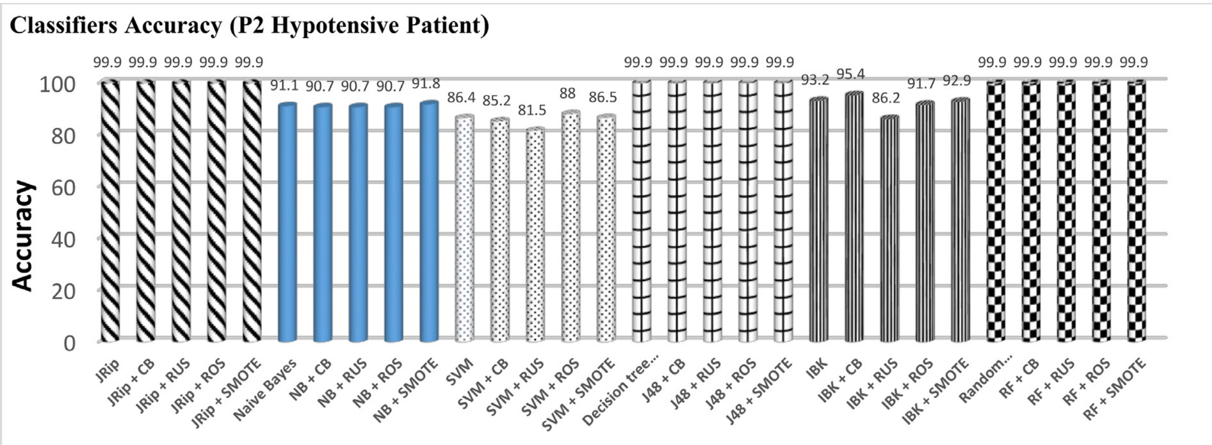

Fig. 7. Classifier accuracy with and without sampling methods for P2 hypotensive patient over one year.

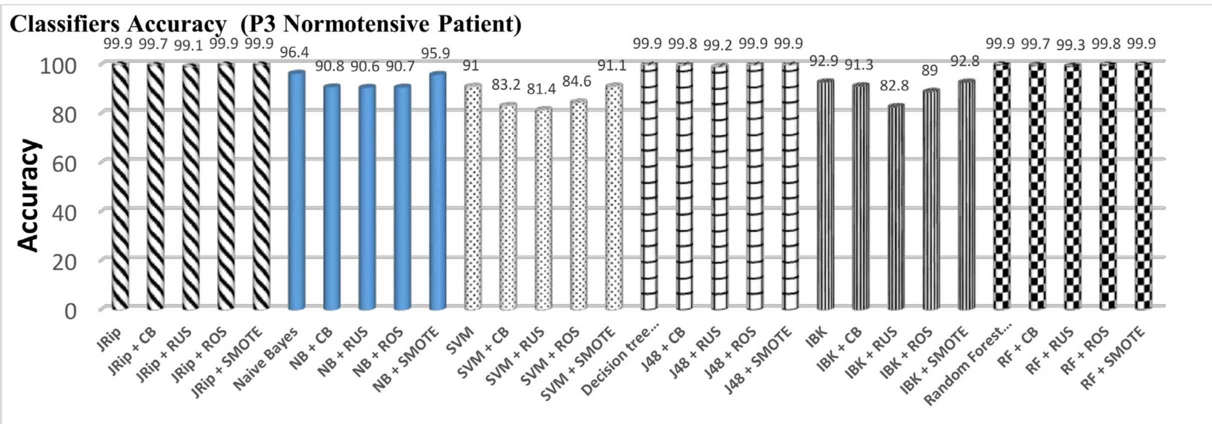

Fig. 8. Classifier accuracy with and without sampling methods for P3 normotensive patient over one year.

7. Conclusions

The proposed framework has succeeded in accommodating and analysing big data generated by continuous patient monitoring in real time. Smart hospitals can adopt the proposed architecture to monitor their patients in their homes in real time. The proposed model utilises the power of clouds for storage and computation. Additionally, a local module that mon-

Table 9
Ranking of classifiers and sampling methods.

Rank	Classifier	Sampling techniques
1	J48	CB
2	RF	SMOTE
3	JRip	ROS
4	NB	RUS
5	IBK	
6	SVM	

itors the patient with the same efficiency in the event of internet interruption or a failure in the cloud system has been implemented. Our work has led us to conclude that the proposed model is fast, accurate, and fault-tolerant when it comes to monitoring patients with BP disorders and predicting their health status with high precision and minimal false alerts. The findings of this study indicate that using sampling techniques in parallel via Spark increases the accuracy of classification and minimises overall error rate, particularly for minority classes (emergency class). The proposed technique achieved higher speed, better accuracy, and higher F-measure than previously proposed systems that depend on MapReduce and association rules without handling imbalanced datasets. The importance of developing a local module that simulates the capabilities of a cloud module to preserve patient life in the event of internet interruption or any defect in the cloud system has been proved. Future work will concentrate on the implementation of this framework for different diseases and additional context domains.

References

[1] Rashidi P, Mihailidis A. A survey on ambient-assisted living tools for older adults. IEEE J Biomed Heal Inf 2013;17:579–90. doi:10.1109/JBHI.2012.2234129.
[2] World Health Organization World health statistics 2016: monitoring health for the SDGs, sustainable development goals. Rep Monit Heal Sustain Dev Goals 2016:1–41 doi:ISBN 978 92 4 156526 4.
[3] Gubbi J, Buyya R, Marusic S, Palaniswami M. Internet of things (IoT): a vision, architectural elements, and future directions. Future Gener Comput Syst 2013;29:1645–60. doi:10.1016/j.future.2013.01.010.
[4] Darwish A, Hassanien AE, Elhoseny M, Sangaiah AK, Muhammad K. The impact of the hybrid platform of internet of things and cloud computing on healthcare systems: opportunities, challenges, and open problems. J Ambient Intell Hum Comput 2017:1–16.
[5] Das SK, Cook DJ. Designing smart environments: A paradigm based on learning and prediction. Int Conf Pattern Recognit Mach Intell. Springer; 2005. p. 80–90.
[6] Schriger DL. Approach to the patient with abnormal vital signs. In: Goldman's Cecil Medicine, 1; 2011. p. 27–30. doi:10.1016/B978-1-4377-1604-7.00007-5.
[7] Schmidt A. Interactive context-aware systems interacting with ambient intelligence. Ambient Intell 2005:159–78.
[8] Ye J, Dobson S, McKeever S. Situation identification techniques in pervasive computing: a review. Pervasive Mob Comput 2012;8:36–66. doi:10.1016/j.pmcj.2011.01.004.
[9] Normandeau K. Beyond volume, variety and velocity is the issue of big data veracity. Inside Big Data 2013.
[10] Ren W, Ren Y, Wu ME, Lee CJ. A robust and flexible access control scheme for cloud-IoT paradigm with application to remote mobile medical monitoring. In: Proc. - 2015 3rd Int. Conf. Robot. Vis. Signal Process. RVSP 2015; 2016. p. 130–3. doi:10.1109/RVSP.2015.38.
[11] Oresko JJ, Jin Z, Cheng J, Huang S, Sun Y, Duschl H, Cheng AC. A wearable smartphone-based platform for real-time cardiovascular disease detection via electrocardiogram processing. IEEE Trans Inf Technol Biomed 2010;14:734–40. doi:10.1109/TITB.2010.2047865.
[12] Sarker VK, Jiang M, Gia TN, Anzanpour A, Rahmani AM, Liljeberg P. Portable multipurpose bio-signal acquisition and wireless streaming device for wearables. In: SAS 2017 - 2017 IEEE Sensors Appl Symp Proc; 2017. p. 3–8. doi:10.1109/SAS.2017.7894053.
[13] Forkan A, Khalil I, Tari Z. CoCaMAAL: a cloud-oriented context-aware middleware in ambient assisted living. Future Gener Comput Syst 2014;35:114–27. doi:10.1016/j.future.2013.07.009.
[14] Forkan A, Khalil I, Ibaida A, Tari Z. BDCaM: big data for context-aware monitoring - a personalized knowledge discovery framework for assisted healthcare. IEEE Trans Cloud Comput 2015:1 –1.
[15] Steinhubl SR, Muse ED, Topol EJ. The emerging field of mobile health. Sci Transl Med 2015;7 283rv3.
[16] Panagiotakopoulos TC, Lyras DP, Livaditis M, Sgarbas KN, Anastassopoulos GC, Lymberopoulos DK. A contextual data mining approach toward assisting the treatment of anxiety disorders. IEEE Trans Inf Technol Biomed 2010;14:567–81. doi:10.1109/TITB.2009.2038905.
[17] Abdelaziz A, Elhoseny M, Salama AS, Riad AM. A machine learning model for improving healthcare services on cloud computing environment. Measurement 2018.
[18] Holzinger A, Röcker C, Ziefle M. From smart health to smart hospitals. In: From smart health to smart hospitals; 2015. p. 8700. (Including Subser Lect Notes Artif Intell Lect Notes Bioinformatics). doi:10.1007/978-3-319-16226-3_1.
[19] Libelium Comunicaciones Distribuidas S.L. Mysignals hw - ehealth and medical IoT development platform for arduino 2017. http://www.my-signals.com/#what-is-mysignals (accessed January 1, 2017).
[20] Witten IH, Frank E, Hall Ma, Cunningham J, Humphrey M, Hunt L, Mcqueen B, Smith L, Inglis S, Nevill-manning C. Data mining: practical machine learning tools and techniques. Data Min Pract Mach Learn Tools Tech. 4th ed. Morgan: Kaufmann; 2016. p. 539–57. doi: http://dx.doi.org/10.1016/B978-0-12-417295-1.09989-6.
[21] Koliopoulos AK, Yiapanis P, Tekiner F, Nenadic G, Keane J. A parallel distributed weka framework for big data mining using spark. In: Proc. - 2015 IEEE Int. Congr. Big Data, BigData Congr. 2015; 2015. p. 9–16.
[22] de Castro Ferreira MAP. SHRAM-smart heart rate and activity measurement. 2016.
[23] Jauch EC, Cucchiara B, Adeoye O, Meurer W, Brice J, Chan YYF, Gentile N, Hazinski MF. Part 11: adult stroke: 2010 American Heart Association guidelines for cardiopulmonary resuscitation and emergency cardiovascular care. Circulation 2010;122. doi:10.1161/CIRCULATIONAHA.110.971044.
[24] Saeed M, Villarroel M, Reisner AT, Clifford G, Lehman L-W, Moody G, Heldt T, Kyaw TH, Moody B, Mark RG. Multiparameter intelligent monitoring in intensive care II: a public-access intensive care unit database*. Crit Care Med 2011;39:952–60. doi:10.1097/CCM.0b013e31820a92c6.
[25] POWERS DMW. Evaluation: from precision, recall and f-measure to ROC, informedness, markedness & correlation. J Mach Learn Technol 2011;2:37–63 doi:10.1.1.214.9232.

Mohammed K. Hassan received an M.Sc. degree in Computer Engineering and Systems from Mansoura University, Egypt. He is currently a PhD student at the Computer Engineering and Systems Department, Faculty of Engineering, Mansoura University, Egypt. His current interests include data mining, classification, cloud computing, and big data.

Ali I. El Desouky received an M.Sc. degree and PhD in Computer Science and Control from the Electronics and Electrical Engineering Department, University of Glasgow, Glasgow, UK. He is currently a full-time professor at the Computer Engineering and Systems Department, Mansoura University. His current research focuses on data mining, classification, big data analysis, cloud computing, Hadoop, and MapReduce techniques.

Sally M. Elghamrawy received a PhD in Computer Engineering and Systems from Mansoura University, Egypt. She is currently an assistant professor at the MISR Higher Institute for Engineering and Technology, Egypt. She is the author of number peer-reviewed publications, receiving best paper awards. She is also a member of IEEE. Her research interests include big data analytics, Cloud computing and NoSQL databases.

Amany M. Sarhan received a PhD in Computer Engineering and Systems from a joint study program between Tanta University, Egypt and the University of Connecticut, USA. She is currently a professor at the Computers and Automatic Control Department, Tanta University, Egypt. Her research interests include network security, big data, object-oriented databases, fragmentation and allocation of databases, and distributed systems.