Accepted Manuscript

A Hybrid Real-time remote monitoring framework with NB-WOA algorithm for patients with chronic diseases

Mohammed K. Hassan, Ali I. El Desouky, Sally M. Elghamrawy, Amany M. Sarhan

PII: S0167-739X(17)32659-6
DOI: https://doi.org/10.1016/j.future.2018.10.021
Reference: FUTURE 4526

To appear in: *Future Generation Computer Systems*

Received date : 15 November 2017
Revised date : 29 August 2018
Accepted date : 16 October 2018

Please cite this article as: M.K. Hassan, et al., A Hybrid Real-time remote monitoring framework with NB-WOA algorithm for patients with chronic diseases, *Future Generation Computer Systems* (2018), https://doi.org/10.1016/j.future.2018.10.021

A Hybrid Real-time Remote Monitoring Framework with NB-WOA algorithm for patients with chronic diseases

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
{eng.mkamal1976, prof.dr.ali.eldosouky}@gmail.com

[b] Department of Computer Engineering, MISR Higher Institute for Engineering and Technology, Egypt
Mansoura Ring Road, Mansoura city 35516, Egypt
sally_elghamrawy@ieee.org

c Department of Computer Engineering and Automatic Control, Faculty of Engineering, Tanta University, Egypt
Sibirbâi, Tanta University, Faculty of Engineering, Tanta city 31527, Egypt
prof.dr.amany.sarhan@gmail.com

Abstract:

The embracing of the Internet of Things (IoT) and Cloud Computing technologies gives excellent opportunities to develop smart healthcare services that have great prediction capabilities. This paper proposes a Hybrid Real-time Remote Monitoring (HRRM) framework, which remote-monitors patients continuously. This smart framework predicts the real health statuses of the patients in real time by using context awareness. The proposed HRRM framework innovates a Patient's Local Module (PLM) that does convergence between IoT sensors and clouds. The HRMM transfers some of the computations to the edge of the network in (PLM) such as converting the low-level data to a higher level of abstraction to speed-up the computations in the cloud portion of the HRMM. The convergence of IoT enables the HRMM to use the enormous cloud power in storing, processing, analyzing big data, building classification models for the category of patients' health status. The local portion of the HRMM uses classification models that have been trained in the cloud to predict the health status of the patient locally in the event of internet interruption or cloud disconnection to save his life in the disconnection periods. Furthermore, this paper proposes a cloud classification technique that is capable of dealing with big imbalanced dataset by minimizing errors especially in the minority class that represents the critical situations. Finally, a hybrid algorithm of Naïve Bayes (NB) and Whale Optimization Algorithm (WOA) has been proposed to select the minimal set of features that achieve the highest accuracy. The (NB-WOA) works as a safe-failure module that decides when to stop the monitoring using HRMM in the case of the failure of influential sensors. Experiments have proved that the HRMM is capable of predicting the health status of the patients suffering from blood pressure disorders accurately. Also, it proved that NB-WOA accelerates the classification process and saves storage space.

Keywords: Smart Healthcare; Internet of Things convergence (IoT); Naïve Bayes (NB); Whale Optimization Algorithm (WOA); Big data; imbalanced dataset.

1. Introduction

Machine learning has many contributions in the medical field such as Remote Patient's Monitoring (RPM) systems that deliver care to the patient suffering from chronic disease especially elderly patients at his home [1]. RPM is defined as using technology to monitor patients remotely (e.g., at his house) to improve patient's quality of life. It tracks the patient continuously without obstruction to the freedom of his movement to prevent possible complications, and all these services should be provided at reasonable cost [2]. Implantable and wearable biomedical sensors have received much attention over the last two decades because of the need to collect sensor data that contains physiological signals, patient's activity during vital signs' measurement, etc. in real time while practicing his daily routine [3]. IoT exploited the progress in ubiquitous sensing which is qualified by Wireless Sensor Network (WSN) technologies to enable actuators and sensors to interact seamlessly with the ambient environment and to share the collected information among different platforms. IoT has made a huge leap by enabling various technologies such as near field communication (NFC), Radio-frequency identification (RFID) and embedded sensor to transform the internet into a fully integrated platform [4,5]. There are many factors that can affect vital signs values of the patient such as patient's activities (current/last), ambient conditions (temperature, humidity, noise, etc.), patient's habits (sleeping, smoking, alcoholic beverages, food, etc.) and many other factors. Context awareness defines the capability of a system to gather information from the surrounding environment at any time to comprehend it and adapt its behavior accordingly. Context-aware RPM model uses this technique to comprehend the current health situation of the patient and provide a personalized health care service accordingly [6]. For example, context-aware RPM refers to an emergency case when the patient's heart rate (HR) increases above normal during sleep while refers to a normal case if the increase in HR occurs during exercise. This technique can be implemented by aggregating all sensor data to the high-level form in one context state for each period. Machine learning is used to understand the health status of the patient and interpret the fluctuations in the patient's vital signs to provide the proper assistive service [7]. The continuous monitoring of patients using RPM models is a source of big data generation because the monitoring period may be extended for years with a fast sampling rate may be in milliseconds resulting in the generation of a huge amount of sensor data. Big data is one of the famous terminologies in the current decade that is used to describe the dataset that fulfills at least one of the characteristics of 4 V's model (Volume- Velocity -Variety - Veracity) according to IBM's formal definition [8]. Therefore, the architecture that contains IoT and cloud components provides scalable data repositories and resilient computation processes on the cloud side for the collected health data by IoT on the local side [9]. Traditional RPMs depend on a standalone application working on a handheld device or local server and always customized for a specific case depending on generic rules [10]. These systems cannot manage big data and cannot be trusted to monitor patients suffering from other diseases. Some researchers tried to address these shortcomings by developing context-aware models to predict the health status of the patient at real-time [11,12]. Furthermore, the recent researches proposed cloud-based frameworks for knowledge extraction from big data using clouds for storing and processing patient's context states to predict the patient's health status at the real time. The weak point of these models that they put the patient at risk when the connection with the cloud is lost or internet is interrupted [13,14]. Also, these models ignored the problem of imbalanced datasets that is always present in this type of data. The primary motivations for this work are:

- The need for developing context-aware RPM, which uses generic and personalized medical rules to build a customized medical assistant that, comprehends the real health status of the patient to minimize false alarms.
- The urgent need for developing an intelligent hybrid classification model that works locally to save the patient's life in the case of cloud system failure or internet interruption.
- The need to address the shortcomings of the previous models in dealing with imbalanced datasets that result in generating false warnings especially in the emergencies that represent the minority class.
- The need to develop an algorithm that identifies the minimum number of attributes required to ensure the continuity of the model's work with highest efficiency and speed.

The rest of this paper is organized as follows: The second section represents related work, which contains Remote Patient Monitor (RPM) models, NB and WOA: The proposed architecture (HRRM) is presented in details in the third Section. The fourth section introduces three case studies for monitoring patients suffering from blood pressure disorders in real-time, also, the sampling methods that will be used to deal with imbalanced datasets. The Proposed Hybrid Knowledge Discovery Classification Model (HKDM) is presented in the fifth section. The sixth section illustrates the proposed NB-WOA. The evaluation of the proposed classifiers and results are outlined in the seventh Section. Our conclusions and future work are drawn in the final section.

2. Related work

2.1. The IoT-Cloud Convergence in Smart healthcare

It is expected that many smart medical services will evolve because of the tremendous development in IoT, cloud, and edge computing domains and the integration among them. This integration helps in developing new medical assistive scenarios and new generations of smart medical services and applications. Recently, this topic has increased interest in both industry and academia aiming to design and implement advanced smart healthcare systems. Most of the currently proposed architectures consist of a set of layers for storing, processing, and analyzing medical data.

Abawajy and Hassan [15] proposed a sustainable Cloud-Based Pervasive Patient Health Monitoring (PPHM) architecture. The PPHM architecture contains three layers as follows: Collection Station, Observation Station, and Data Centre. This architecture has been tested through a case study for real-time monitoring for a patient suffering from congestive heart failure. Chen et al. [16] proposed an Edge-Cognitive-Computing-based (ECC-based) smart-healthcare system for monitoring the physical health of users using cognitive computing. Catarinucci et al. [17] proposed an IoT aware smart hospital system that collects and monitors patients' parameters using the ultra-low-power hybrid sensing network in real time. Manogaran et al. [18] proposed a new architecture to implement the IoT to process scalable sensor data (big data) for healthcare applications. In addition, they provide security services using the integration of fog computing with cloud computing. A healthcare service delivery architecture based on fog computing has been proposed by Andriopoulou et al. [19]. It proposes module between Cloud and IoT devices to enable new types of computing and services. The proposed architecture consists of three main layers, which are: (i) fog servers for storing, processing, and analyzing data, (ii) fog nodes for data aggregation, and (iii) cloud-based module for data storage. Rahmani [20] proposed a Smart e-Health Gateway using the strategic position at the edge of the network. The concept of Fog Computing in Healthcare IoT systems is exploited by forming a Geo-distributed intermediary layer of intelligence between sensor nodes and Cloud. In addition, an IoT-based Early Warning Score (EWS) health monitoring is implemented to address a medical case study. Dimosthenis et al. [21] Proposed an integrated Edge-Fog-Cloud architecture for Healthcare Internet of Things (EFCHIoT) Infrastructure. The EFCHIoT architecture consists of three layers to store medical data, acquire process, and to provide real-time decision-making. The three layers are as follows: the first is the Edge layer that includes portable and wearable computational devices, the second is the Fog layer which is responsible for gathering and processing data from the Edge nodes, and the third is a cloud infrastructure which is responsible for data storage and analysis of the data uploaded from the combination of Fog and Edge levels. Experimental results have proved that EFCHIoT architecture provides real-time decision-making, fast queries' processing, and less power consumption. Our proposed architecture has benefited from these ideas by innovating a hybrid architecture that does the main processing of vital signs in the local portion of the architecture. Moreover, it uses the power of the cloud to store, processes the big imbalanced datasets, and train classification models from a huge number of contexts. Additionally, it transfers the classification model to the cloud portion of the architecture to predict the health status of the patient in the case of internet interruption or cloud disconnection.

2.2. Remote Patient Monitoring (RPM)

RPM has enabled physicians to monitor and observe patients remotely using digital technologies that collect health data, ambient conditions, activities, etc. in any location, such as a patient's home, and to transmit the collected information electronically to healthcare providers for assessment and taking appropriate actions [22,23]. The integration of non-invasive technologies into healthcare management strategies by gathering all possible information from the patient and his ambient environment helps to improve the quality of decision-making [24–26]. RPM is an interdisciplinary field exploiting advancement in many areas such as activity monitoring [27], continuous care [28], personalized care [29], cloud-based healthcare architectures to achieve a breakthrough in this area [13,14,30]. Earlier trials for developing context-aware RPM has some drawbacks, for instance, RPM cannot manage big data because they are based on local architecture, each RPM is designed for a specific disease, and they support a limited number of context awareness services [31,32]. Many researchers struggled to solve the previous shortcomings by developing Context-aware cloud-based models that can extract knowledge from massive data generated from patients' continuous monitoring [33,34]. The most recent researches proposed flexible architectures that facilitates adding or removing contexts easily, and they are suitable for monitoring any patient suffering from any chronic disease [13,14,35]. These architectures wholly depend on clouds in their operation, and this raises many inquiries about what will happen to the monitored patient when the internet connection is interrupted, or failure occurred in the cloud system. Furthermore, the problem of imbalanced datasets has been overlooked, most studies have only focused on accuracy to prove the efficiency of the model, but we believe that this is misleading. The classifier succeeded in predicting patterns that belong to the majority class and failed in the minorities which are more critical in these datasets because they represent the emergency case of the patient [13,14].

2.3. Naïve Bayes

Naive Bayes (NB) is the most popular classifier in the family of probabilistic classifiers. Naive Bayes classifier uses the probabilistic theory to get the correct classification [36,37]. NB has proven its effectiveness in many learning scenarios such as medical diagnosis [38,39], text classification [40], sentiment analysis [41], image processing [42,43] and web mining [44]. Classification using Bayesian network considers the dependency between attributes for obtaining the correct result [45]. NB is a particular case of the Bayesian algorithm assuming that features are independent of each other [46]. This assumption makes training phase simpler and faster with nearly similar results. NB is working as follows [46]:

Let the training set T has some tuples; each one is represented by an n-dimensional vector $X=\{x_1, x_2,\dots, x_n\}$ and each vector describes n attributes $A_1, A_2,\dots, A_n$. Each Sample belongs to one class of m classes: $C_1, C_2,\dots, Cm$.

1. For a given a sample X, the classifier will predict that X belongs to the highest posterior probability of class by recalling Bayes theorem as shown in equation 1

   $$P(C_i \mid X) = (P(X \mid C_i)\, P(C_i))/P(X) \qquad \text{Posterior} = (\text{Likelihood} \times \text{Prior})/\text{Evidence} \tag{1}$$

   X is classified to class C_i (class with highest posterior probability), when $P(C_i|X) > P(C_j|X)$, where $1 \le i, j \le m$. In equation 1, the denominator (evidence P (X)) is the same for all classes, so only the numerator $(P(X|C_i)\ P(C_i))$ is calculated to find the biggest value. The prior probability of class (C_i) can be calculated as in equation 2:

   $$P(C_i) = S_i/S \tag{2}$$

   , where (S_i) is the number of training samples of class (C_i) and (S) is the total number of training samples. If the prior probability term $(P(C_i))$ is unknown, the equal probability is assumed for all classes, then $P(C_1) = P(C_2) = \dots = P(C_m)$, therefore, the target of equation 1 is transformed into maximization for the term $(P(X \mid C_i))$ only.
2. The workload for calculating likelihood $(P(X \mid C_i))$ will be very high especially in multi-dimensional datasets. A simple (naive) assumption solves this problem by stating that individual attribute values are independent of each other under certain conditions and can be calculated as shown in equation 3.

   $$P(X \mid C_i) \approx \prod_{k=1}^{n} P(x_k \mid C_i) \tag{3}$$

3. From given dataset T, $P(x_1 \mid C_i), P(x_2 \mid C_i),\dots, P(x_n \mid C_i)$, can be calculated from the training set. Where X_k refers to k^{th} attribute (A_k) of sample X.
4. From equation 1, the numerator $(P(X \mid C_i)\ P(C_i))$ will be calculated for each class, so that sample X will be predicted as C_i member, if and only if, $P(X \mid C_i)\ P(C_i)$ is the maximum.

2.4. Whale Optimization Algorithm (WOA)

WOA is a nature-inspired meta-heuristic algorithm, which is used to solve optimization problems by mimicking the motion of the whale when hunting the prey.

2.4.1. Whale in nature

Whales are giant predators, which considered as the biggest mammals in the world reaching a length of 30 meters and weight around 180 tons. There are eight primary species of this giant creature such humpback, finback, blue, killer, Minke, sperm, Sei, and southern right. Whales are brilliant animals, and they are emotionally at the same time because they have shared cells in their brains similar to those called spindle cells in human. These cells are responsible for emotions judgment, social behaviors as in humans and this is the cause of the whale smartness [47]. Hence, a whale can socialize more than other animals and live in groups; they can learn, think, judge, communicate and become emotional better than other animals. Humpback whale is one of the biggest baleen whales, and his favorite preys are krill and small fish herds [48,49]. It has a unique hunting technique called bubble-net feeding method. They hunt the victim, which is close to the surface by creating bubbles along a '9'-shaped path. According to the most recent studies, Humpback whales adopt two maneuvers techniques associated with bubbles called ‘upward-spirals’ and ‘double loops.' In the first maneuver, the humpback whales dive in the water around 12 meters down, then start to make a wave of bubbles in a spiral shape encircling the prey. Finally, the whale swims fast toward the surface to hunt the prey as shown in **Fig 1**. The second maneuver includes three different stages: coral loop, lobtail, and capture loop [48]. More details about the whales' behaviors can be found in [48–50].

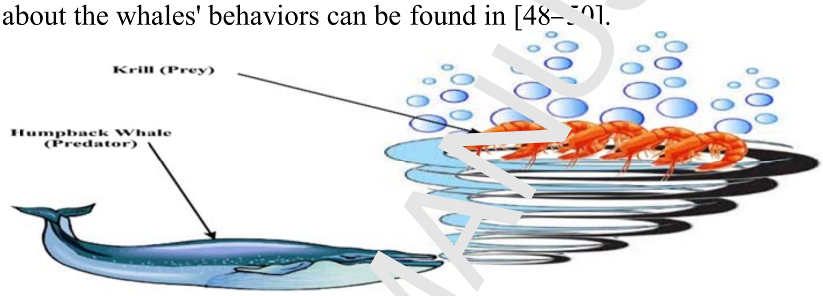

Fig 1. Bubble-net feeding behavior of humpback whales

2.4.2. The Mathematical Mimicking model

The unique actions of a humpback whale in searching for the prey, encircling the prey and spiral bubble-net feeding maneuver, are mathematically modeled as follows:

Encircling prey

Since the target from the movement of the Humpback whales is how to recognize the location of prey to encompass it. WOA assumes that the position of the whales are W_i ∀ i = 1, 2,…, M, where M is the number of whales which is initialised randomly in the search space to search for the position of the optimum solution which is unknown. The best location (resolution) is considered as the position of the target prey or close to the optimum position. After the best search agent is defined, the other search agents will hence try to update their positions towards the best solution or position as in equations 4 and 5

$$\vec{D} = \left| \vec{C}.\ \vec{W}^{*}(t) - \vec{W}(t) \right| \quad (4)$$

$$\vec{W}(t+1) = \vec{W}^*(t) - \vec{A}.\vec{D} \quad (5)$$

Where $\vec{A}$ *and* $\vec{C}$ are coe□cient vectors, which are calculated as in Equations 6 and 7 for (t) that represents the current iteration, $\vec{W}$ is the position vector, $\vec{W^*}$ vector represents the position vector of the best solution until now. $\vec{W^*}$ The vector should be updated in each iteration if there is a better solution.

$$\vec{A} = 2\vec{a}.\vec{r} - \vec{a} \quad (6)$$

$$\vec{C} = 2.\vec{r} \quad (7)$$

Where $\vec{a}$ is linearly decreased from 2 to 0 throughout iterations in exploitation and exploration phases and $\vec{r}$ is a random vector in [0, 1]. For two-dimensional search space as in **Fig 2**, the position of the candidate W_i that is located at (X, Y) can be updated according to the position of the best candidate (X^*, Y^*). The values of $\vec{A}$ *and* $\vec{C}$ vectors control the new position of W_i. Accordingly, any position in the search space can be reached by updating the current position in the neighbourhood of the current best candidate to simulate the method of encircling the prey. This 2-d concept can be extended to n-dimensional search space. The bubble-net feeding behaviour has two phases called exploitation and exploration phases [50,51].

Exploitation phase

In this phase, whales adopt two mechanisms to chase the prey, which can be explained in mathematics as follows:

A. Shrinking encircling mechanism

Decreasing the value of $\vec{a}$ in Equation (6) will control the shrinking mechanism, and then the positions of whales are updated according to Equations (4, 5, 6 and 7). **Fig 2** shows how the current solution (whale) W_i, iteratively converges towards the best solution W^* (the location of the prey), **Fig 2** represents the solution in two dimensional space

B. Spiral updating position

The following steps accomplish the simulation for this behavior:

a) The distance between the current position (solution) W_i and the best solution W^* is calculated.
b) The helix-shaped movement of the humpback whales is mimicked by creating a spiral equation as follows:

$$\vec{W}(t+1) = \vec{D}'.e^{bl}.\cos(2\pi l) + \vec{W}^*(t) \quad (8)$$

$$\vec{D}' = \left|\vec{W}^*(t) - \vec{W}(t)\right| \quad (9)$$

Equation 9 represents the distance between the i^th whale (W_i) and ($\vec{W}^*$) the best solution obtained so far, l is a random number in the interval [-1, 1] and b is a constant that defines the logarithmic spiral's shape [43].
According to the previous equations, the humpback whales move towards the prey with two different kinds of movements simultaneously:

(1) According to equation (4, 5, 6 and 7), the humpback whales move around the victim within a shrinking circle.
(2) According to equation (8 and 9), the humpback whales move towards the prey in a spiral-shaped path.

In WOA algorithm, the whale is switching between these two kinds of movements with equal probability as shown in equation 10 [50].

$$\vec{W}(t+1) = \begin{cases} \vec{W}^*(t) - \vec{A}\vec{D} & if\ p < 0.5 \\ \vec{D}'.e^{bl}.\cos(2\pi l) + \vec{W}^*(t) & if\ p > 0.5 \end{cases} \tag{10}$$

Where p is a random number in the interval [0,1]

Exploration phase (search for pray)

The humpback whale searches for the prey randomly in the exploration phase adopting a different technique by updating his position according to a randomly chosen candidate instead of the best candidate like in the exploitation phase. Mathematically if $|\vec{A}| > 1$, the candidate whale moves far away from the reference whale performing a global search as in equations 11 and 12 [43].

$$\vec{D} = |\vec{C}.\overrightarrow{W_{rand}} - \vec{W}| \tag{11}$$

$$\vec{W}(t+1) = \overrightarrow{W_{rand}} - \vec{A}.\vec{D} \tag{12}$$

$\overrightarrow{W_{rand}}$ is a random position for the randomly chosen whale from the current population of whales.

In WOA, the positions of the search agents are updated at each iteration according to the value of $|\vec{A}|$, if $|\vec{A}| > 1$ the position will be updated randomly according to randomly chosen search agent and if $|\vec{A}| < 1$, the position will be updated according to the best solution. So, the parameter (a) is used to switch smoothly between exploration and exploitation phases. Also, the parameter p controls the switching between the two kinds of whale's movement "spiral or circular motion [50].

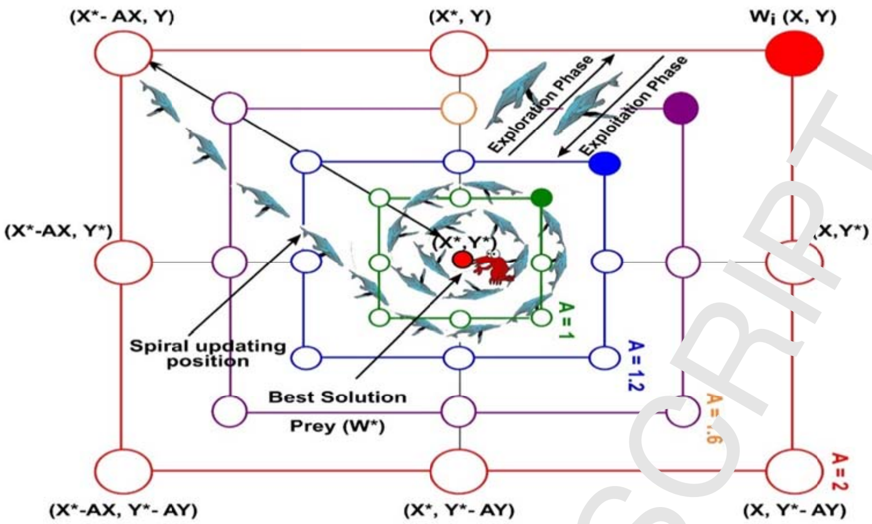

Fig 2. Bubble-net attacking method (exploitation phase) and searching for prey mechanism (exploration phase) implemented in WOA

3. Hybrid Real-time Remote Monitoring (HRRM) Architecture

The proposed (HRRM) facilitates delivering health care to the patient's home. It enables smart hospitals to monitor patients outside of conventional hospital settings and thus increases the number of patients covered by care service and reduce the cost. The smart hospital is a hospital that improves patient care procedures and creates new capabilities by adopting new technologies such as cloud computing, cloud storage, Internet of things (IoT), etc. The proposed architecture is designed over big data model to extract knowledge from gathering a massive amount of medical data, behavioral information and ambient data generated from the continuous monitoring of a significant number of patients in real-time. As depicted in **Fig 3.** HRRM consists of four layers as follows:

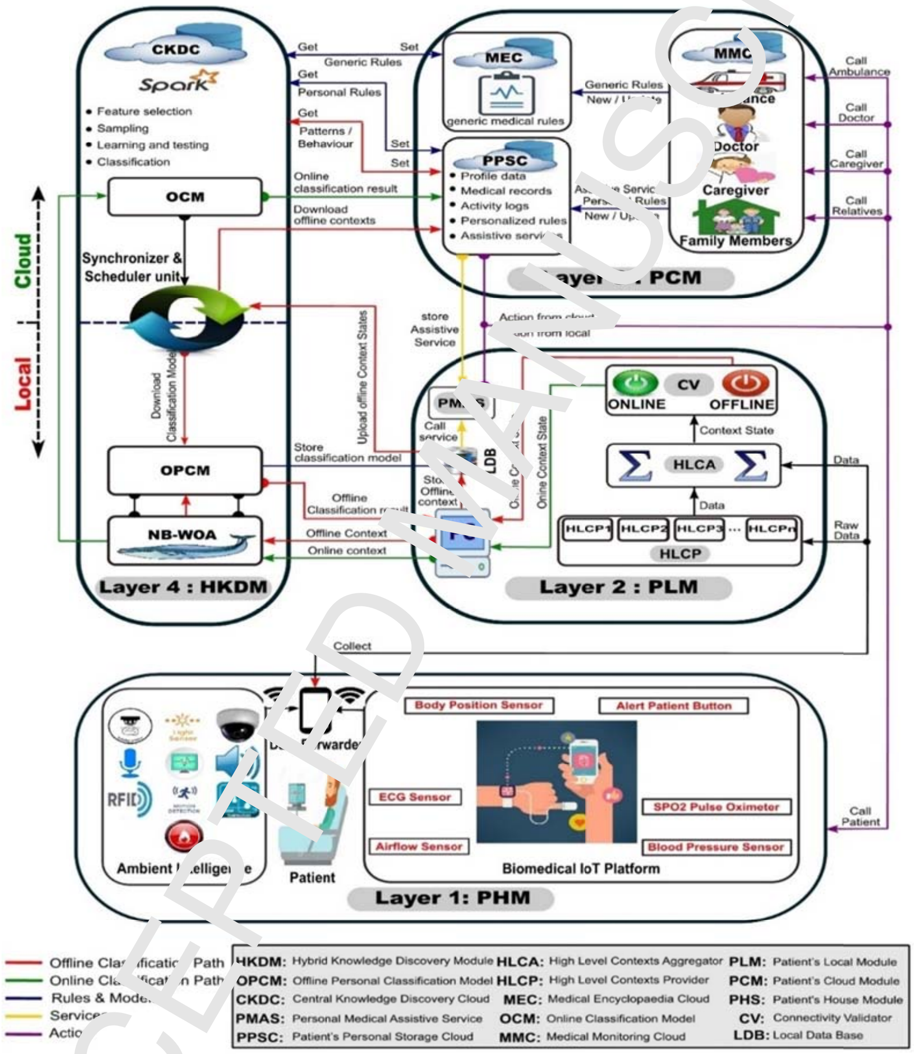

Fig 3. The main components of the proposed HRRM architecture

3.1. Layer 1: Patient's House Module (PHM)

HRRM manages a large number of PHMs for patients monitored by the smart hospital. Each PHM is responsible for gathering sensor data that includes medical data (physiological signals), behavioral patterns (smoking, drinking alcoholics, taking medications, etc.), ambient data (humidity, temperature, noise, etc.), contextual information (location, activity, etc.). The right setup of the RPM system guarantees building reliable supervision system taking into consideration the patient's illness type and his social condition. Each illness type requires selecting suitable actuators, ubiquitous devices, and IoT sensors along with software programs to obtain the necessary sensor data to extract knowledge about the health status of the patient in real-time. Each PHM has a unique identifier in the model to identify the patient in the hybrid architecture. Layer 1 is composed of the following components:

3.1.1 Biomedical IoT Platform

Biomedical sensors are vital instruments in the modern medicine used to collect sensor data that has information on human body and pathology. The continuous development of biomedical sensors provided the market with precise, sensitive, and fast response sensors with competitive price capable of collecting patient's vital signs in real time. Electronic medical (eMedical) kits integrate different types of sensors into one package or add new sensors for building a new medical device as in MySignals eHealth and medical IoT platform [52]. It facilitates measuring more than 20 biometric parameters such as (position, oxygen in the blood, glucose level, blood pressure, pulse, etc.). Furthermore, it supports many connectivity options such as (GPRS, 3G, Bluetooth, Wi-Fi, ZigBee, IEEE 802.15.4, etc.).

3.1.2 Ambient Intelligence devices (AmI)

It refers to electronic environments created using sensors that are sensitive and responsive to the presence of the patient and providing important ambient sensor data needed for the study. Ambient sensors are easily embedded in RPM through different communication media to recognize the patient and his situational context. Furthermore, they can be tailored to the patient's needs and exploit its adaptive and anticipative capabilities. The goal of using ambient sensors in the proposed module is to add smartness, context awareness to the model. Also, it helps in understanding the effect of ambient conditions on patient's physiological signals. Ambient sensors that can be used including, but not limited to, room temperature sensors, humidity sensor, and smoke detector.

3.1.3 Data Forwarder (DF)

It forwards the collected low-level sensor data from ambient devices and high-level sensor data from the eMedical platform over different communication media. DF forwards Low-level sensor data to High-Level Context Provider (HLCP) in Patient Local Module (PLM) converting it to a higher level of abstraction while it forwards high-level data directly to High-Level Context Aggregator (HLCA).

3.2. Layer 2: Patient's Local Module (PLM)

PLM is a central local module responsible for receiving, processing and aggregating the generated sensor data in PHM into one context state. Also, it has a smart unit validating communication with the cloud part of the hybrid model. Furthermore, it acts as a backup module for monitoring the patient by classifying his health status in the case of internet disconnection or when a problem occurres in the cloud system. PLM contains the following components:

3.2.1. High-Level Context Provider (HLCP)

This unit converts raw sensor data to a higher level of abstraction by adopting many techniques such as feature selection, fusion algorithms, and classification algorithms, then the converted sensor data is forwarded to High-Level context Aggregator (HLCA).

3.2.2. High-Level Context Aggregator (HLCA)

This unit is responsible for the aggregation of the output of different HLCPs and sensor data generated from biomedical IoT development platform in one context state. Each context state contains sensor data such as vital signs, ambient data, associated activity, behavioral information, etc. at specific time slot in the form of high-level values. The mining for assembled context states will reveal many mysteries about the fluctuations in patient's vital signs. For example, the increase in HR above the normal range during jogging is interpreted as a normal case, but it is worth investigating if it occurs while the patient is relaxed [5].

3.2.3.Connectivity Validator (CV)

This unit examines the connectivity between the local module and the cloud modules in the proposed hybrid architecture to select the suitable mode accordingly. Internet connection is regularly tested with different measures such as network latency, download speed, upload speed, etc. to switch smoothly between the modes of the model. If CV verifies that the connection is valid and stable an online mode would be selected to classify patient's health status on the cloud side of the architecture while if CV detects any failure in the communication system; offline mode will be chosen to do the same operation on the local part using a backup classification model.

3.2.4.Local Database (LDB)

This unit stores the collected context states reaching to PLM in the offline mode. Also, it contains a backup classification model for classifying patient's health status locally in offline mode when there is a problem in the cloud system. Furthermore, it offers storage space for an updated copy of patient's assistive services, medication time, prescriptions, precautions, prohibitions, radiological investigations, medical reports, etc. which are available also for the cloud part of the hybrid architecture.

3.2.5.Personal Medical Assistive Service (PMAS)

PMAS is a tailored service for every patient according to his illness type and social status. A medical committee composed of a family member, social researcher, physician-in-charge, caregiver and administrative staff puts suitable assistive services for every patient. Assume that classes, which identify patient's health status, are as follows (Normal, Warning, Alert, and Emergency). In Normal cases, the system should work without generating any alerts. In warning cases, a warning message will be sent to caregiver and physician-in-duty. In alert cases, caregivers and physician-in-charge will be called to follow the case in addition to one of his relatives, neighbors or friends. In emergency cases, physician-in-charge and ambulance will be appealed directly to transfer the patient to the hospital. All these notifications can be displayed on monitor or patient's smartphone. Also, a video call can be held between doctor-in-charge and the patient to give him instructions in alert and emergency cases.

3.3. Layer 3: Patient's Cloud Module (PCM)

This module acts as a personal information repository for every patient monitored by HRRM. It is used to classify patient's health status in online mode. PCM consists of the following clouds:

3.3.1. Patient's Personal Storage Cloud (PPSC)

PPSC is a personal cloud storage area; every patient who is monitored by the smart hospital has his own PPSC. This repository retains context states aggregated during system's operation in offline mode until uploaded to the Central Knowledge Discovery's Cloud (CKDC). Also, this cloud keeps patient's profile (e.g., name, age, gender, weight, height, illness history, etc.) and the thresholds of patient's physiological signals, which are taken from Medical Monitoring Cloud (MMC). Furthermore, it keeps medical tests, radiological investigations, medical reports, prescriptions, medicine name, its dose, time, patient's behaviors like smoking and drinking alcoholic beverages, etc. Finally, it retains the last updated version of assistive services approved by the medical committee.

3.3.2. Medical Monitoring Cloud (MMC)

This cloud contains all entities helping in the monitoring operation of the patient both inside and outside the smart hospitals. This cloud has a connection to all persons or services' provider assigned to provide help to the patient when his health is deteriorating. Medical experts transfer their medical knowledge to MEC in the form of generic medical rules while physician-in-charge is responsible for putting personal medical rules and move them to PPSC. Furthermore, physician-in-duty monitors the patient remotely and responses to alarms generated in alert and emergency cases by taking immediate action to save his life.

3.3.3.Medical Encyclopaedia Cloud (MEC)

This medical encyclopedia retains all medical information according to recent researches for every illness type, physiological signals that must be monitored and their ranges and associated symptoms in the form of generic medical rules. The generic medical rules are used with personal rules in building a personalized classification model for every patient which will minimize false alarms. MEC is updated with any new discovered generic rule from knowledge discovery process as will explained in layer 4.

3.4. Layer 4: Hybrid Knowledge Discovery Module (HKDM)

HKDM is a hybrid module that contains components on both local and cloud sides used for knowledge extraction and the classification of patient's health status accordingly. The hybrid architecture aims to exploit merits of both local and cloud architectures and avoid their flaws. The cloud part of the module facilitates working with big data regarding storage and computations, on the other hand, the local part of the module will solve the weakness of the cloud-based models in case of internet interruption or a failure in the cloud system under any circumstances. This module consists of the following components:

3.4.1. Central Knowledge Discovery Cloud (CKDC)

CKDC is one of the core components of the proposed framework, which includes many distributed clouds with large storage capacities to accommodate all context states generated from patient's continuous monitoring. Spark is used to distribute a vast number of contexts, maybe for millions of patients' across different clusters then applying different machine learning techniques in parallel to speed up the knowledge discovery process. As shown in **Fig 3**, the knowledge discovery process is done vertically across the four layers by converting raw data into the first layer into high-level data by HLCP, and then aggregate them with contextual and medical information by HLCA into unified contexts states in the second layer. Generic and personal medical rules are used in the third layer to build a dynamic model customized according to the patient's health status. In the fourth layer, machine-learning techniques are used to extract the knowledge about the patient's health status using a massive number of contexts.

3.4.2. Online Classification Model (OCM)

The learning phase in the proposed classification technique consists of five consecutive stages to build an accurate classifier working in online-mode and capable of dealing with imbalanced datasets. The best-learned model among all clusters will be selected by voting to predict patient's health status in online mode; this technique aims to maximize the accuracy of the classification and minimize the elapsed time. (Will be presented in detail in section 5).

3.4.3. Offline Personal Classification Model (OPCM)

OPCM is a backup copy of OCM that works on the local side (offline mode) when the internet connection interrupted, or failure happens in the cloud system. According to the result of the classification, one of PMAS services, which are stored in LDB, will be called to take appropriate action.

3.4.4. Synchronizer and Scheduler Unit (SSU)

This unit is responsible for information exchange between HKDM, PLM and PCM to ensure that each module has the last version of information required for its proper work. The synchronization of offline context states stored in LDB with PPSC is performed according to a predefined schedule. Moreover, the instant synchronization of the new version of OCM with OPCM to be used in offline classification on the local side is performed. Furthermore, it ensures that PLM has the last version of PMAS.

4. A Case Study on Patients with Blood Pressure disorders

An imbalanced dataset is a dataset that the number of tuples belonging to the majority class outnumbers those belonging to minority classes [53,54]. For example, it is normal in datasets that the patient's health status is classified to one of these classes: (Normal, Warning, Alert, and Emergency) according to patient's context state to be imbalanced. The majority class is a Normal class, while Emergency and Alert classes are the minority classes. As, most classifiers are accuracy-driven that they concentrate on maximizing the overall accuracy and minimizing the overall errors assuming that the distribution of classes is normal and the cost of errors obtained from different classes is same so, they will be biased towards majority classes rather than minority ones [55]. Handling imbalanced class distribution can be classified as following: sampling methods [56,57], cost-sensitive [58,59] and kernel-based methods [60].

A case study is implemented to evaluate the performance of HRRM and to prove its efficiency in classifying patient's health status and its ability to deal with imbalanced datasets. This case study has the following objectives:

a. Verifying that HRRM correctly comprehends the health situation of the patient using context-awareness to achieve more accurate results than traditional systems that adopt generic rules in classification.
b. Verifying that HRRM succeeded in addressing the problem of the imbalanced datasets and its dramatic exacerbation with big data.
c. Validating that the proposed classification technique (HKDDM) succeeded in building a coherent learning model for big data generated from HRRM using a distributed cloud model to speed up classifications and giving instant, accurate results.
d. Validating that the proposed bio-inspired algorithm (NB-WOA) succeeded in selecting the minimum sets of features required for the operation of the model with the highest efficiency and fastest performance.
e. Electing the best classification technique and the best sampling methods that give the best results and suitable to operate with HRRM.

4.1. Case Study Description

As illustrated in **Table 1**, this study has been conducted on three elderly patients suffering from Blood Pressure (BP) disorders and their details are as follows:

Table 1. Patients' records

	Patient 1			Patient 2			Patient 3		
Patient record	a41434			a41466			A40208		
Gender	Female			Male			Female		
Age	71 years			66 years			78 years		
Birth date	27-Aug-1941			11-Mar-1944			11-Feb-1937		
Illness category	Hypertension			Hypotension			Normal + Transient Elevation in BP		
Monitoring start date	31-Aug-2012			22-Jun-2010			15-Mar-2015		
Monitoring duration	Day	Month	Year	Day	Month	Year	Day	Month	Year
	✓	✓	✓	✓	✓	✓	✓	✓	✓

The HRMM has monitored the patients for varying periods as listed in **Table 1** to evaluate its performance in predicting the health status of these patients.

The measuring of physiological signals four or five times daily is not enough to diagnose serious medical illness, especially in medically unstable cases. Thus, the patients were continuously monitored by taking measurements every 15 minutes.

The vital signs vary with the ambient conditions such as humidity, noise, and room temperature. Also, the behaviors of the patient such as smoking, drinking alcoholic beverages, taking medications, physical activity, and stress are the major factors of the fluctuation in his vital signs. Additionally, many additional factors can explain the variations in the patient's physiological signals including age, the degree of the illness, disease's history and the family profile [61,62].

The HRMM has exploited this data to build a context-aware classification model, which is smart enough to distinguish between the different situations and their effect on the physiological signals of the patient. Also, this framework is capable of comprehending the nuances between the different patients. The consideration of these points in the design of the HRMM leads to building a smart, coherent, accurate, and fast framework. Accordingly, this technique will minimize the false alarms that are usually generated from such AALs especially those systems that depend on general medical rules in its operation (traditional AALs).

The learning phase has been performed in the (HKDM) utilizing large historical data from many patients with the same category of illness. The OCM uses the data stored in the KDC to train a classification model for every illness's category. The OCM detects the emergency cases in real time and informs the medical assistive team to take the appropriate action.

Datasets will be distributed among different Hadoop clusters, and the learning phase will be performed in parallel using ensemble vote's classification technique to manage the massive data and speed up the classification process to give results in real time. WEKA (Waikato Environment for Knowledge Analysis) will be used to simulate the proposed classification technique OCM using Spark and evaluate it with different classifiers and sampling methods [63,64].

4.2. The clinical dataset

The clinical data of the elderly patients suffering from blood pressure disorders that are used in these case studies have been taken from PhysioNet MIMIC-II [65].

4.2.1. PhysioBank

The PhysioBank contains four terabytes of digitized vital signs and time series containing over 90,000 recordings organized in more than 80 clinical datasets, and classified according to the types of signals included [66]. The clinical databases include continuous measurements for some vital signs along with, laboratory test results, procedures, medications, caregiver notes, images and imaging reports, and mortality (both in and out of hospital).

4.2.2. The MIMIC II database

The MIMIC-II database is the extension to the first attempt of building a database called MIMIC that contains multi-parameter recordings of ICU patients in the period between 1992 and

1999. The MIMIC II Clinical Database was released in 2011 including over 32,000 subjects for more than 40,000 patients who have been admitted to cardiovascular, medical, surgical, surgical recovery units, and coronary care units at the same hospital [67].

As illustrated in **Table 1**, three patients have been selected to represent the different categories of BP disorders. Moreover, they are used to test the classification models (OCM and OPCM) of the proposed HRMM. Additionally, the OCM uses the data of a large number of patients including these patients to train the classifiers in the learning phase.

4.3. Synthetic Data Generation

As far as we know, there are no real datasets that similar to the data that will be collected by the proposed model. The targeted dataset contains physiological signs, associated activity, ambient conditions, and behavioral information for patients with blood pressure disorders. The continuous monitoring will extend to a year by taking a sample every 15 minutes as shown in **Table 2**. Therefore, datasets will be synthetically generated based on real patients' physiological signals taken from Physionet MIMIC-II database for three patients to mimic data generated from biomedical IoT platform in HRRM [68].

Table 2. A sample of the final dataset

Timestamp	HR	SBP	DBP	MBP	RR	SPO_2	Temp.	Act.	L. Act.	Med.	Sym.	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
23-03-16 4:15	86	146	63	103	17	100	2	2	2	0	0	Normal
24-03-16 0:00	62	181	91	104	23	100	1	1	1	0	1	Emergency

The synthetic dataset takes into consideration the following criteria:

a. The correlation between activities and physiological signals according to **Table 3** that shows the percentage of increase in HR according to different events (e.g., HR is average when the patient is watching TV, but it will be higher when he is on the treadmill) [69].
b. The plausibility of activity time (e.g., eating at 3 p.m. and sleeping at 1 a.m.).
c. The effect of ambient conditions and taking medications on physiological signals.
d. The relationship between patient's health status and symptoms.

The generation of synthetic datasets is performed using MATLAB for hypertensive, hypotensive, and normotensive patients in a year with the same distribution as in real Physionet MIMIC-II datasets. This technique is reliable in the generation of synthetic datasets that are similar to real datasets, which are verified in previous studies of biomedical data analysis [14,70]. Some abnormal physiological signals are added without modifying the remaining attributes of these context states to represent emergency and alert cases. The possible symptoms for each category (Hypertensive patients, Hypotensive patients, and Normotensive patients with transient elevation in blood pressure) are presented in **Table 4**. **Table 5** shows all Types and ranges for all attributes in the generated datasets that are used in the experiment [71]. **Table 6** illustrates the situational classification model used to distinguish the class according to personalized medical rules. **Table 7** shows a set of assistive services that will be executed according to the result of the classification.

Table 3. The percentage of HR increase according to each activity type [69]

Activity	The percentage of HR increase.
Laying	0-10%
Sitting	5-15%
Standing	10-25%
Walking	15-45%
Running/Cycling	35-100%
Stairs	15-80%

Table 4. Symptoms according to patient's category in the experiment

Type	Symptoms	Value (Binary)
Hypertensive	A headache & Anxiety – Fatigue- a severe headache & Anxiety - Pounding in your chest neck, or ears - Vision problems and confusion - Chest pain and difficulty in breathing	6-bit binary (value: 0 to 63)
Hypotensive	Lack of concentration – Fatigue - Blurred vision - Dizziness - Rapid shallow breath - Fainting	6-bit binary (value: 0 to 63)
Normotensive	Uncomfortable- Anxiety – a headache – Fatigue - a severe headache - Dizziness	6-bit binary (value: 0 to 63)

Table 5. Types and ranges for attributes used in the experiment [71]

Name	Attributes	Type	Range value
Vital Signs	Heart Rate (HR)	Numeric	[30-200]
	Systolic Blood Pressure (SBP)	Numeric	[50-230]
	Diastolic Blood Pressure (DBP)	Numeric	[30-140]
	Respiratory Rate (RR)	Numeric	[5- 30]
	Oxygen Saturation (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 6. Situational Classification according medical model

Class	Classification
Normal	HR, SBP, DBP, RR, and SPO$_2$ all values are in expected Range concerning current activity and symptoms = 0
Warning	Any of HR, SBP, DBP, RR and SPO$_2$ increase above the normal range to the warning range or medications not taken or symptoms > 0
Alert	Any of HR, SBP, DBP, RR and SPO$_2$ increase above the warning range to 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 and SPO$_2$ increase above the alert range to the emergency range or more than two vital signs in alert range and (medications not taken or symptoms > 0)

Table 7. Examples for assistive services

Case	Action
Class = Normal	Do nothing
Class = Warning	Warning on patient's mobile or monitor in his home 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 the care giver

4.4. The exploration of the generated datasets

The generated datasets for three patients in a one-year and the distribution of classes are illustrated in **Table 8**, as well as **Fig 4**. It is clear that when general medical rules had been applied to classify the generated contexts to normal and abnormal classes, it failed in classifying most of them, and that, of course, will make the system generates many false alarms. HRRM uses context awareness to comprehend the real situation of patient's health taking into consideration ambient conditions, behavioral information and associated activity to give results that are more accurate. It is clear that the generated datasets are severely imbalanced and (Emergency) class is the minority class while (Normal & Warning) classes are majority classes. Therefore, different sampling methods will be applied to these datasets to address the problem of imbalanced datasets in addition to classification techniques using WEKA and Spark [64,72].

Table 8. Comparison between classifications with HRMM against traditional AALs for three patients in a year

Patient	No. of contexts	Traditional AAL		HRMM			
		Normal	Abnormal	Normal	Warning	Alert	Emergency
P1 (Hypertensive)	35232	2	35230	9307	23347	2404	174
P2 (Hypotensive)	35232	3	35229	19455	14003	1627	147
P3 (Normotensive)	35232	1	35231	12517	21421	1186	108

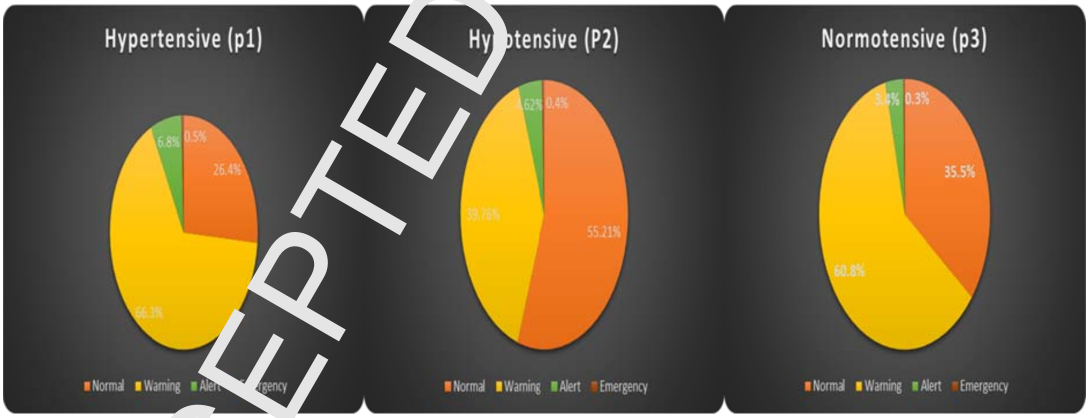

Fig 4. The distribution of classes for three patients' datasets over one year using the HRMM

4.5. The implementation of the case study using Weka and Spark

Table 9 shows hardware specifications of PC used in all experiments, operating system, programming software, and its plugins. The Distributed Weka Base package and Distributed Weka Spark package must be installed after installing the last version of WEKA as shown in Table 8 [73].

Table 9. 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. The Proposed Hybrid Knowledge Discovery Model (HKDM)

The proposed Hybrid Knowledge Discovery Model (HKDM) consists of two classifiers, one of them called Online Classification Model (OCM) that works on the cloud side (online mode). The other one called Offline Personal Classification Model (OPCM) which is a backup copy of OCM that works on the local side (offline mode). The learning phase of the proposed classification process is composed of five stages as depicted in **Fig** 5.

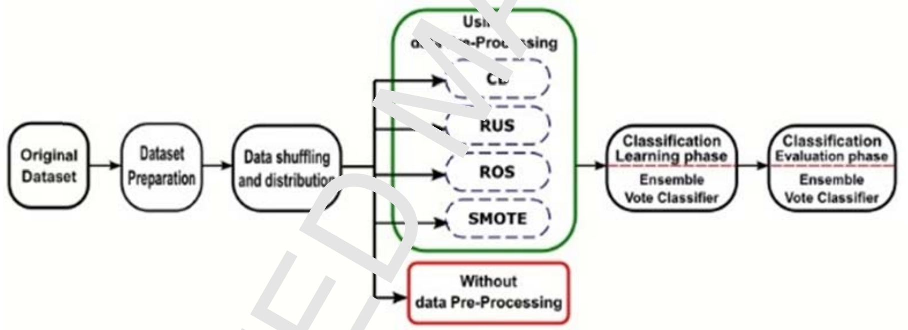

Fig 5. The proposed technique for learning and evaluating CCM

5.1. Online Classification Model (OCM):

Stage 1: Dataset Preparation (Splitting data)

As WEKA and Spark will be used to implement phases of the proposed classification technique that will be used to classify patients' imbalanced contexts, it is important at this stage to prepare datasets to work correctly with them. While '*.arff*' is the default dataset's file extension but to use it with Spark it should split into two files using '*arff header Spark job,'* the first one contains the header section of the dataset with extension '*.names*.' and the second one contains dataset itself with extension '.CSV' (Comma Separated Value). The level of parallelism that will be applied to the dataset is configured in this job by indicating the number of data chunks that will run in parallel on Hadoop Distributed File System (HDFS) or YARN cluster. Also, the same job can be simulated using cores of PC's processor by setting 'master host = local {no of cores}.' In this experiment, datasets will be sliced into four partitions and will be processed using the five cores of the Processor in parallel [63].

Stage 2: Randomly Shuffle dataset

In this stage, "*Randomly Shuffle Data Spark Job*" is used to specify the number of data chunks and the number of instances in each chunk. Furthermore, data chunks are stratified to ensure that each class has almost the same distribution of class values as the original dataset, which helps in getting the best result, especially when using ensemble-voting classifier. If the minority class has samples, less than the number data chunks, these samples will be copied to each data chunk to make sure that each class is represented in every data chunk at least by one sample.

Stage 3: Data Pre-Processing

It is clear that our datasets are suffering from severe imbalance as illustrated in **Table 8**, as well as in **Fig 4**. In this stage, well-known sampling methods such as Class Balancer (CB), Synthetic Minority over Sampling (SMOTE), Random under Sampling (RUS) and Random over Sampling (ROS) are used to process imbalanced datasets. Experiments will be conducted using six well-known classifiers from different classification families over the three patients' datasets with and without these sampling methods.

Stage 4: Learning phase

In this stage, each adopted classifier will be used to build four models, a model for every data chunk on a different cluster and this process will be repeated four more times, using the four sampling methods. The map part of *"Weka Classifier Spark job"* is configured to train the following classifiers: Naïve Bayes (NB), Decision tree C 4.5 (J48), Random Forest (RF), Ripper (JRip), Support Vector Machine (SVM), Nearest Neighbour (IBK) on each data chunk. Then, the reduce job will select the best-generated model by a voted ensemble classifier.

Stage 5: Evaluation Phase

In this stage, the evaluation is performed for every classifier using ten folds cross-validation by configuring "*Weka Classifier Evaluation Spark Job*" to do that. This evaluation module is aggregatable so the overall classification process will be performed through two passes; the first pass makes the classifier training's task is to learn an aggregated classifier over the data and the second pass for evaluation.

5.2. Offline Personal Classification Model (OPCM)

After the evaluation process is completed for all classifiers, the best classification model (OCM) along with the best sampling method will be copied and transferred to the local part of HKDM. This model is called OPCM; it will be used for classifying the incoming contexts into the local side (offline mode).

6. The proposed NB-WOA algorithm for improving HRRM

The failure of any Sensor may affect model's work continuity and its performance. When a sensor stops working, the aggregated context state by HLCA will be incomplete so the dataset will have missing values and thus affects the classification accuracy. This paper proposes a version of Naïve Bayes classifier called NB-WOA, which is used for features' selection. This version adopts a bio-inspired algorithm called Whale Optimization Algorithm (WOA) to optimize its performance as shown in **Fig 6**. This algorithm struggles to find the minimal set of features achieving the best accuracy. The NB-WOA is working in the local portion of the HRMM to achieve the IoT-Cloud convergence by transferring the selection of the features that accelerate classification to the edge of the local portion of the HRMM. The NB-WOA has the following advantages:

- It simplifies the generated model for better interpretation by the domain experts.
- It allows the proposed HRRM model to work in the case of some sensors' failure that doesn't affect the classification accuracy. This classifier works as system's protection module, which determines when the system should stop and when should keep working in the case of any sensor's failure.
- It avoids the curse of dimensionality.
- It reduces overfitting.

The detailed explanation of NB-WOA is as follows:

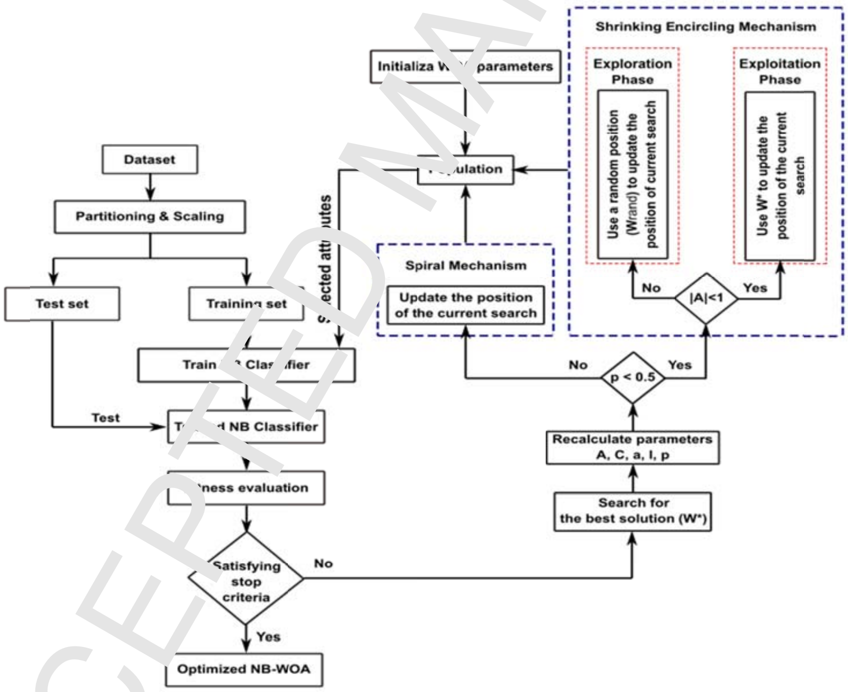

Fig 6. Block diagram of the proposed NB-WOA algorithm

6.1. Parameters' initialization

In the beginning, the parameters of WOA (A, C, a, p, l) were initialized. In the proposed model, the WOA provides the naïve Bayes classifier with some whales, each whale represents a subset of features from the original dataset in (binary form), "1" means that the feature is selected and "0" means not selected. Thus, WOA search to find the best set of features that achieves the highest accuracy with NB. The whales' positions are initialized randomly.

6.2. Fitness evaluation

The optimization operation needed to apply fitness function to assess every whale position is as given by equation 13.

$$Fitness\ (F) = \alpha\gamma_R(D) + \beta\ (|C - R|/|C|) \tag{13}$$

Where $\gamma_R(D)$ is the classification performance of condition feature set R with respect to choice D, R is the length of selected feature subset, C is the aggregate number of features, α and β are two random parameters that depend on each other, α ∈ [0, 1] and β = 1 − α, these parameters are related to the significance of the subset length and the classification performance. This fitness function is used to maximize the classification accuracy; $\gamma_R(D)$ and the proportion of the unselected features to the aggregate number of features; as in the term $|C - R|/|C|$ [74].

6.3. Termination criteria

When the termination criteria are satisfied, the operation ends; otherwise, we proceed with the next generation operation. In the proposed model, the WOA is terminated when a maximum number of iterations are reached or when the best solution (best set of features (BestF) that achieve the highest accuracy (BestAccuracy)) is not modified for a given number of iterations.

6.4. Updating positions

The positions of whales are then updated as in the mathematical model of WOA in section 2.3.2.

6.5. Algorithm NB-WOA

Input:

- *Initialize the whale's population X by a binary code; each whale is composed of a string of feature selection bit.*
- *Initialize parameters (A, C, a, b, l and r)*
- *Initialize BestAccuracy*

Output:

- *BestF according to BestAccuracy*

Algorithm:

```
while (t < maximum number of iterations)
    for each search agent
        Update a, A, C, l and p
            If (p<0.5)                              /*Shrinking Encircling Mechanism*/
                if (|A| < 1)                        /*Exploitation Phase*/
```

```
                The position of the current search agent is updated by Equations (4, 5, 6, 7)
            else if (|A| >1)                                        /*Exploration Phase*/
                Select a random search agent (X_rand)
                The position of the current search agent is updated by Equations (10, 11, 12)
            end if
        else if (p > 0.5)                                           /*Spiral Mechanism*/
            Update the position of the current search by the Equation (8, 9)
        end if
    end for
    Check if any search agent goes beyond the search space and amend it
    Remove attributes, which are not selected from the training sample attribute to
    get the training dataset T/, according to the feature of each whale selection bit.
    Calculate prior probability P (Ci) of each class of training data.
    Calculate likelihood P (x| Ci).
    Calculate the formula = likelihood * prior probability P (x| Ci)* P (Ci)
    Select the maximum prior probability P (x| Ci)* P (Ci) as the predicted class

    Calculate the fitness function (maximization problem) of each search from equation 13.
    Accuracy = f with corresponding feature selection BestF.
    If Accuracy >= BestAccuracy
      Set BestAccuracy = Accuracy with corresponding feature selection BestF
    endif
    Update X* if there is a better solution
    t=t+1
end while
Return X*
```

7. Results and Discussion

Experiments are held to test the proposed HRRM with different six classifiers and four sampling methods. The performance of all classifiers with and without different sampling methods are evaluated regarding Accuracy, overall F-measure, F-measure for emergency class and time elapsed in each experiment. The priority in the selection of the best classifier along with the best sampling method is in following the order: F-measure (emergency class), overall F-measure, accuracy and elapsed time. Ranking ensures selecting the best combination of classifier and sampling methods for HRRM that works with high efficiency and generate minimal false alarms. Accuracy and F-Measures are calculated from equations 14, 15, 16 and 17 [73,75]. Where, TP = True Positives, FP = False Positives, TN = True Negatives and FN = False Negatives

$$\mathrm{Accuracy} = \mathrm{TP} + \mathrm{TN}/\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN} \qquad (14)$$

$$\mathrm{Recall} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FN}) \qquad (15)$$

$$\mathrm{Precision} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FP}) \qquad (16)$$

$$\mathrm{F\ Measure} = 2 * (\mathrm{Precision}\,.\,\mathrm{Recall}/\mathrm{Precision} + \mathrm{Recall}) \qquad (17)$$

Table 10 shows the performance of six classifiers with/without sampling methods for a year using OCM that proposed in section 5.1. The best performance among all versions of the classifier (with/without sampling) is highlighted, and the winning algorithm in each data set is highlighted inside a bold square.

Fig 7, **8** and **9** illustrate a comparison between six classifiers regarding accuracy for the data sets P1, P2, P3, respectively. As shown in **Table 11**, the average rank for these classifiers over datasets indicates that J48, RF, and JRip respectively are the best classifiers that can work with the highest percentage of success in HRRM while the worst are SVM and IBK. It is evident that the elapsed time with SVM classifier increases dramatically with big datasets.

Decision tree and rule-based classifiers have the advantage of generating comprehendible models that can be written in the form of "IF- Then" rules as shown in **Fig 10**. This advantage enables the medical team to investigate the generated model to approve it or to update general or personal medical rules and relearn the classification model again.

Table 10. Comparison between classifiers with/without sampling techniques for patients in one year

Classifier	P1 (Hypertensive Patient)				P2 (Hypotensive Patient)				P3 (Normal Patient)			
	Acc.	F-measure	F-measure (Emergency)	T. (Sec)	Acc.	F-measure	F-measure (Emergency)	T. (Sec)	Acc.	F-Measure.	F-measure (Emergency)	T. (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	[illegible]	[illegible]	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	[illegible]	19	99.9	0.99	0.92	39
JRip +SMOTE	99.9	1	1	10	99.9	**1**	[illegible]	**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	[illegible]	[illegible]	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	[illegible]	[illegible]	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	[illegible]	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	[illegible]	[illegible]	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	[illegible]	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	[illegible]	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	[illegible]	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	[illegible]	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	[illegible]	0.9	0.84	93	91.7	0.92	0.57	92	89	0.89	0.75	107
IBK +SMOTE	**[illegible]**	**0.95**	**0.85**	**47**	92.9	0.93	0.64	43	**92.8**	**0.93**	**0.79**	**50**

Table 11. Classifiers Ranking

Rank	Classifier
1	Decision tree (J48)
2	Random Forest (RF)
3	Ripper (JRip)
4	Naïve Bayes (NB)
5	Nearest Neighbour (IBK)
6	Support Vector Machine (SVM)

As shown in **Table 12**, the best sampling techniques are CB and SMOTE while the worst is RUS. These results proved the ability of HRRM in classifying patients health status with high precision in real-time and recommend using rule-based classifiers (e.g., JRip) or decision trees classifiers (J48) along with SMOTE or CB as sampling methods for the imbalanced dataset.

Table 12. Sampling methods ranking

Rank	Sampling Tech.
1	CB
2	SMOTE
3	ROS
4	RUS

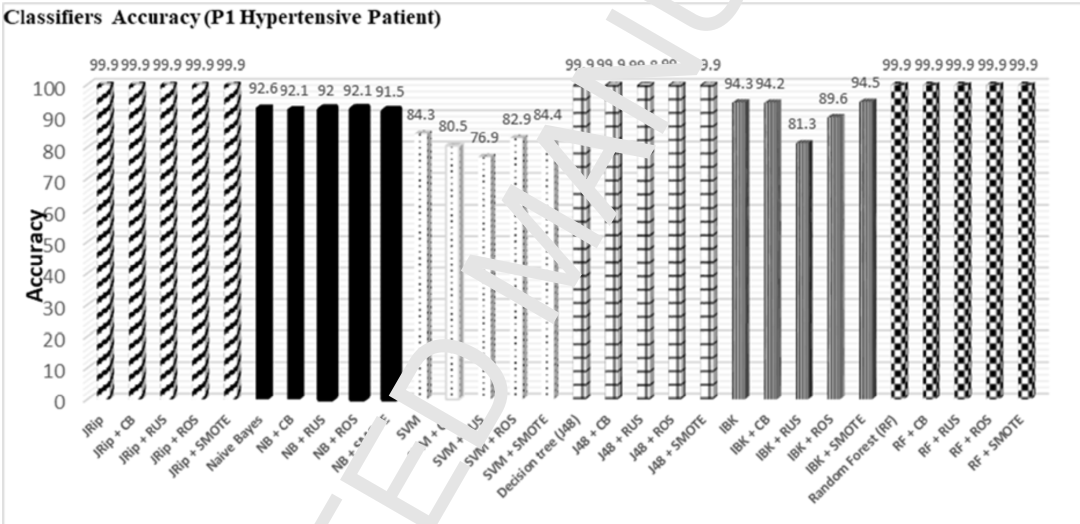

Fig 7. Classifiers Accuracy with/without sampling methods for P1 Hypertensive Patient in one year

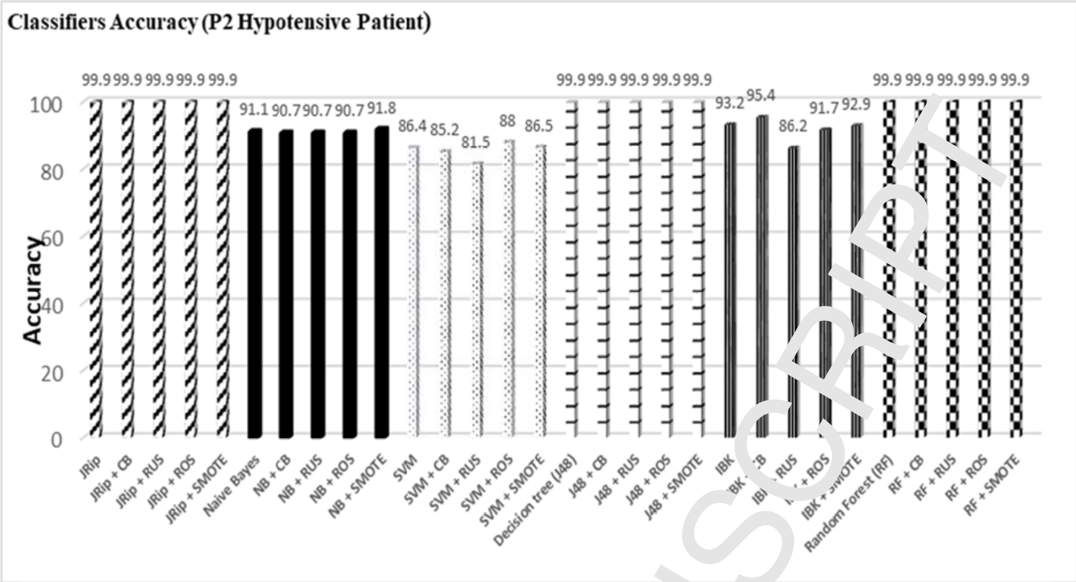

Fig 8. Classifiers Accuracy with/without sampling methods for P2 Hypotensive Patient in one year

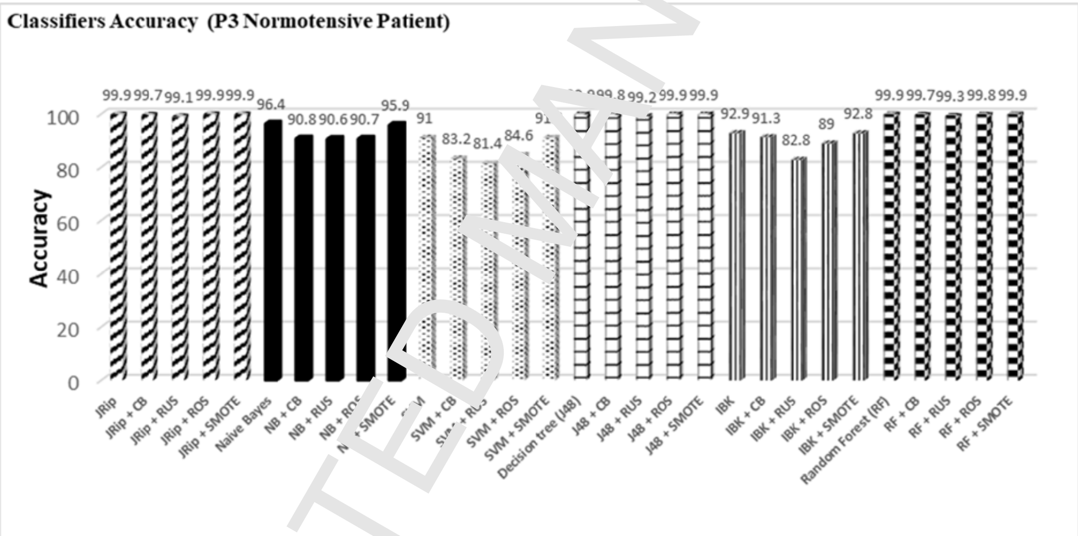

Fig 9. Classifiers Accuracy with/without sampling methods for P3 Normotensive Patient in one year

```
Classifier Model: JRIP rules:
===========
(DBP <= 44) => Class=Emergency (24.0/0.0)
(DBP <= 54) => Class=Alert (36.0/0.0)
(Symptoms >= 1) => Class=Warning (221.0/0.0)
(DBP <= 70) => Class=Warning (24.0/0.0)
(SBP <= 100) => Class=Warning (17.0/0.0)
(SPO2 <= 93) => Class=Warning (12.0/0.0)
(SBP >= 131) => Class=Warning (8.0/0.0)
(RR >= 21) => Class=Warning (11.0/0.0)
(Heartrate <= 59) => Class=Warning (9.0/0.0)
(RR <= 11) => Class=Warning (2.0/0.0)
=> Class=Normal (423.0/0.0)
Number of Rules: 11
```

Fig 10. The sample for one of the generated model using JRIP classifier

As illustrated in **Table 13** and **Fig 11**, the proposed algorithm NB-WOA is tested over the same datasets of the three patients to select the minimal features that achieve the highest accuracy. It needed almost half of the features of the original datasets to achieve the same accuracy or slightly better. Thus, the size of the dataset will shrink by half, speeding up the classification process. As listed in **Table 14**, the next step is to use these features training and evaluating classifiers and to compare the performance of same classifiers over the original datasets that are listed in **Table 10**. The NB-WOA can be used as a safe-fail module to detect when to stop working the model. As listed in **Table 13** for the dataset of the hypertensive patient (P1), the model continues its work if the sensor that registers the room temperature fails. On the contrary, the NB-WOA will stop the model and send alerts to the stakeholders of the model if the sensors that record the HR and BP fail. As illustrated in **Table 14**, as well as, **Figs 12 and 13,** the NB-WOA speeds up classifications and preserves accuracy.

Table 13. The selected features by NB-WOA from each dataset to achieve the highest accuracy

Dataset	No. of total features	Selected features by NB-WOA
P1	11	Five features HR, SBP, DBP, RR, and symptoms
P2	11	Six features HR, SBP, DBP, RR, SPO_2, and symptoms
P3	11	Six features HR, SBP, DBP, RR, SPO_2, and symptoms

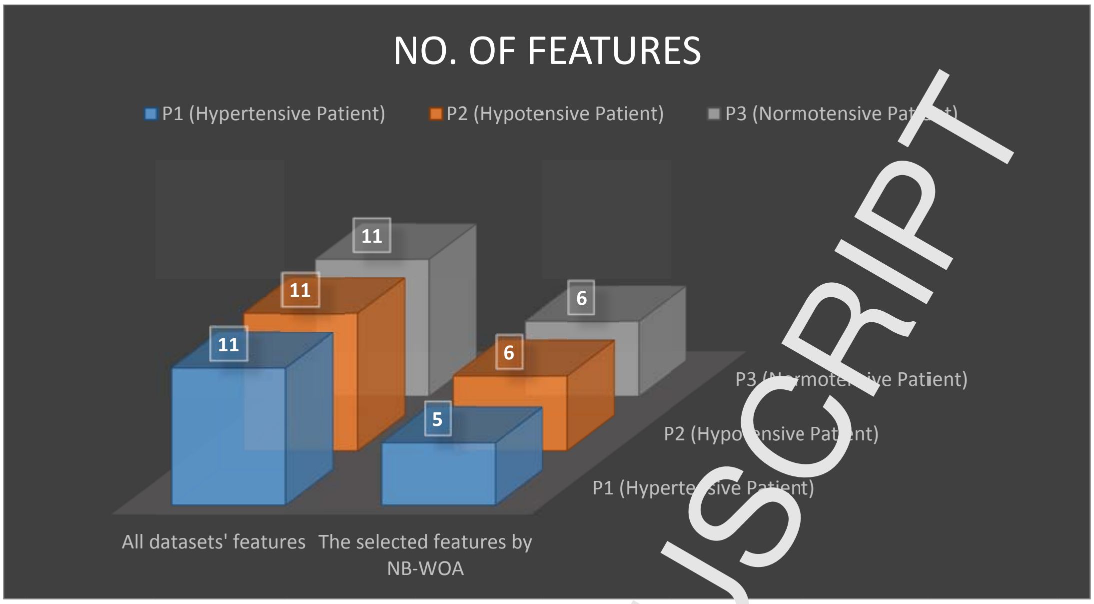

Fig 11. Comparison between number of features in the original datasets and the number of features selected by NB-WOA to achieve the same accuracy

Table 14. Comparison between classifiers' performance using NB-WOA and without

classifier	P1 (Hypertensive)				P2 (Hypotensive)				P3 (Normotensive)			
	Without NB-WOA		With NB-WOA		Without NB-WOA		With NB-WOA		Without NB-WOA		With NB-WOA	
	Accuracy	Time	Accuracy	Time	Accuracy	Time	Accuracy	Time	Accuracy	Time	Accuracy	Time
JRip	99.9	10	99.8	6	99.9	14	99.9	9	99.9	18	99.9	9
NB	92.6	7	92.[illegible]	5	91.1	6	91.4	4	96.4	8	96.8	5
SVM	84.3	38	8[illegible].5	28	86.4	70	86.4	48	91	55	92.1	37
J48	99.9	10	99.[illegible]	7	99.9	20	99.9	13	99.9	8	99.9	6
RF	99.9	16	99.9	8	99.9	18	99.9	8	99.9	18	99.9	10
IBK	94.3	48	95.5	26	93.2	44	93.7	27	92.9	47	93.3	24

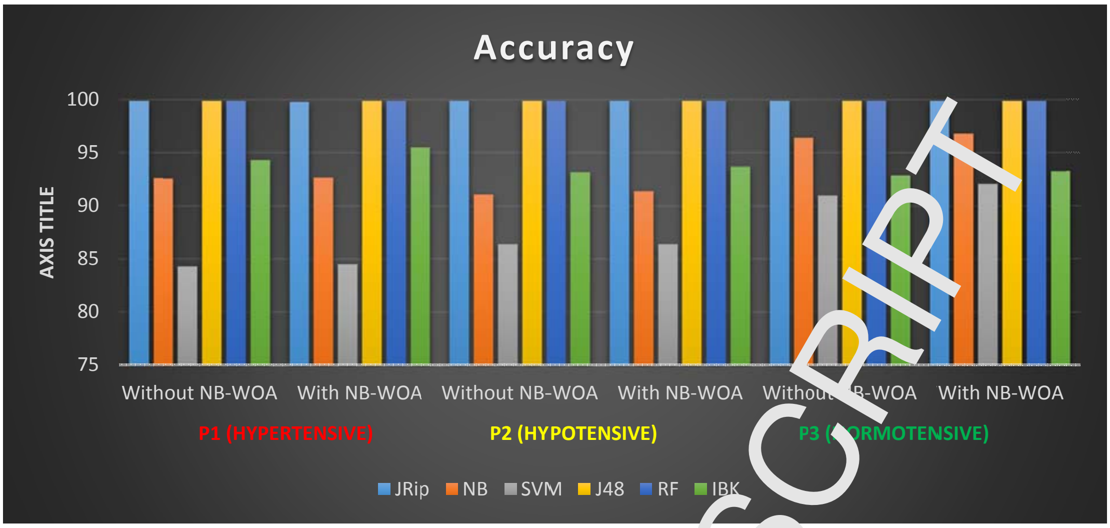

Fig 12. Comparison between the accuracy of the classifiers using NB-WOA and without

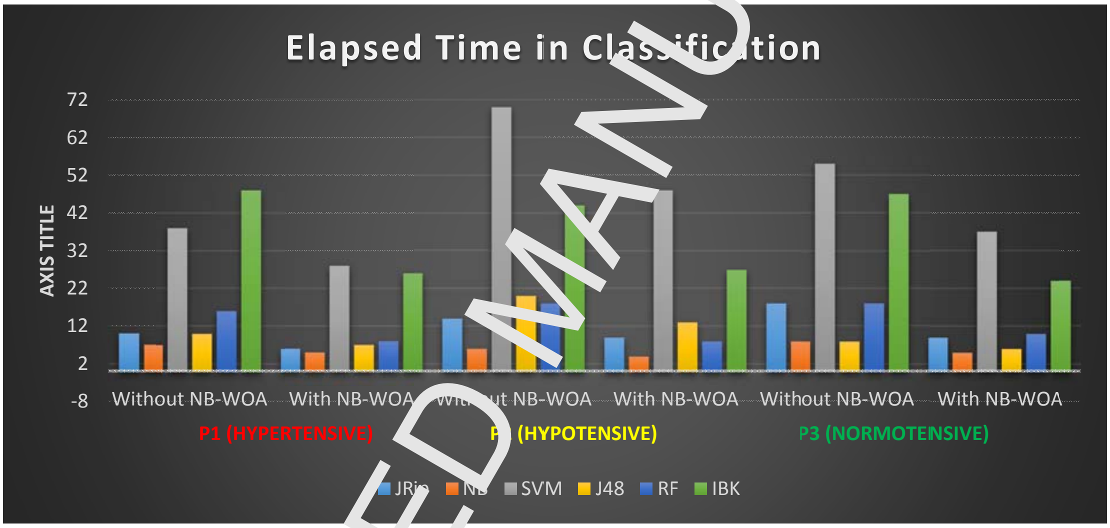

Fig 13. Comparison between the elapsed time in classifications using NB-WOA and without

8. Conclusions and Future work

This Paper proposes a Hybrid Real-time Remote Monitoring (HRRM) framework that monitors the elderly patients suffering from chronic diseases in real time. The proposed framework has addressed the disadvantages of the local and cloud AAL architecture and exploited the advantages of both fusing them into a hybrid architecture that has components in both local and cloud environments. The HRRM has transferred the processing of raw data and the aggregation of high-level data into unified context states to the local portion of the framework instead of clouds as in all cloud-based ALLs. This technique has achieved convergence between IoT and the cloud portion of the framework through the Patient Local Module (PLM). The HRMM has been examined through case studies on patients suffering from different categories of Blood Pressure (BP) disorders. Experimental results have proved that HRMM is a smart healthcare monitoring framework, which is capable of predicting the category of the patient's health status from the current context states accurately. The proposed OCM has succeeded in addressing the problem of big imbalanced datasets by processing data chunks using different sampling methods on different Hadoop clusters using Spark. The findings of this study indicate that the proposed OCM has succeeded in increasing the accuracy of classifications and minimizing error rates, especially for the minority class (emergency class). The OCM has used different sampling methods to preprocess different data chunks across different clusters using Spark in parallel to achieve these results. Our research emphasizes the importance of the PLM not only for the convergence between IoT sensors and clouds but also for the preservation of patients' lives in the case of internet interruption or cloud disconnection. Additionally, experimental results have proved the effectiveness of NB-WOA in selecting the minimal features' set that are mandatory to the proper work of the HRMM without any deterioration in its accuracy. The NB-WOA saves the storage space and accelerates the classifications. Also, it works as a smart safe-failure module that decides to continue the operation of the framework in the case of non-influential sensor failure. If an influential sensor fails, it stops the operation of HRMM to avoid getting wrong classifications' result that put the patient's life at risk. The directions of the future work include the usage of HRMM in monitoring different illnesses, the observation of context domains that may affect patients' vital signs, and the adoption of different bio-inspired algorithms instead of WOA. Additionally, the HRRM framework should be tested from the networking perspective.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] M.J. Coye, A. Haselkorn, S. DeMello, Remote patient management: Technology-enabled innovation and evolving business models for chronic disease care, Health Aff. 28 (2009) 126–135. doi:10.1377/hlthaff.28.1.126.

[2] E.A. Bayliss, J.F. Steiner, D.H. Fernald, L.A. Crane, D.S. Main, Descriptions of barriers to self-care by persons with comorbid chronic diseases, Ann Fam Med. 1 (2003) 15–21. doi:10.1370/afm.4.

[3] S.E. Kern, D. Jaron, Healthcare technology, economics, and policy: An evolving balance, IEEE Eng. Med. Biol. Mag. 22 (2002) 16–19. doi:10.1109/MEMB.2003.1191444.

[4] J. Ana Teresa Aldana, C. Azucena Vallejo, L. Atzori, A. Iera, G. Morabito, D. Bandyopadhyay, J. Sen, F. Breu, S. Guggenbichler, J. Wollmann, I. Brown, J. Gubbi, R. Buyya, S. Marusic, M. Palaniswami, Luis Carlos García, K. M., A.M. Mzahm, M.S. Ahmad, A.Y.C. Tang, S. Peppet, J. Poikolainen, K. Rose, S. Eldridge, C. Lyman, R.R.H. Weber, R.R.H. Weber, a Zanella, N. Bui, a Castellani, L. Vangelista, M. Zorzi, Internet of Things (IoT): A vision, architectural elements, and future directions, Comput. Law Secur. Rev. 29 (2014) 1645–1660. doi:10.1016/j.clsr.2009.11.008.

[5] A.H.T.E. De Silva, W.H.P. Sampath, N.H.L. Sameera, Y.W.R. Amarasinghe, A. Mitani, Development of a wearable tele-monitoring system with IoT for bio-medical applications, in: 2016 IEEE 5th Glob. Conf. Consum. Electron. GCCE 2016, 2016. doi:10.1109/GCCE.2016.7800404.

[6] A. Schmidt, Interactive Context-Aware Systems Interacting with Ambient Intelligence, Ambient Intell. (2005) 159–178.

[7] J. Ye, S. Dobson, S. McKeever, Situation identification techniques in pervasive computing: A review, Pervasive Mob. Comput. 8 (2012) 36–66. doi:10.1016/j.pmcj.2011.01.004.

[8] K. Normandeau, Beyond Volume, Variety and Velocity is the Issue of Big Data Veracity, Insid. Big Data. (2013).

[9] W. Ren, Y. Ren, M.E. Wu, C.J. Lee, 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: pp. 130–133. doi:10.1109/RVSP.2015.38.

[10] J.J. Oresko, Z. Jin, J. Cheng, S. Huang, Y. Sun, H. Duschl, A.C. Cheng, A wearable smartphone-based platform for real-time cardiovascular disease detection via electrocardiogram processing, IEEE Trans. Inf. Technol. Biomed. 14 (2010) 734–740. doi:10.1109/TITB.2010.2047865.

[11] P.L. Emiliani, C. Stephanidis, Universal access to ambient intelligence environments: Opportunities and challenges for people with disabilities, IBM Syst. J. 44 (2005) 605–619. doi:10.1147/sj.443.0605.

[12] V.K. Sarker, M. Jiang, T.N. Gia, A. Anzanpour, A.M. Rahmani, P. Liljeberg, Portable multipurpose bio-signal acquisition and wireless streaming device for wearables, SAS 2017 - 2017 IEEE Sensors Appl. Symp. Proc. (2017) 3–8. doi:10.1109/SAS.2017.7894053.

[13] A. Forkan, I. Khalil, Z. Tari, CoCaMAAL: A cloud-oriented context-aware middleware in ambient assisted living, Futur. Gener. Comput. Syst. 35 (2014) 114–127. doi:10.1016/j.future.2013.07.009.

[14] A. Forkan, I. Khalil, A. Ibaida, Z. Tari, BDCaM: Big Data for Context-aware Monitoring - A Personalized Knowledge Discovery Framework for Assisted Healthcare, IEEE Trans. Cloud Comput. PP (2015) 1–1. doi:10.1109/TCC.2015.2440269.

[15] J.H. Abawajy, M.M. Hassan, Federated Internet of Things and Cloud Computing Pervasive Patient Health Monitoring System, IEEE Commun. Mag. 55 (2017) 48–53. doi:10.1109/MCOM.2017.1600374CM.

[16] M. Chen, W. Li, Y. Hao, Y. Qian, I. Humar, Edge cognitive computing based smart healthcare system, Futur. Gener. Comput. Syst. (2018).

[17] L. Catarinucci, D. De Donno, L. Mainetti, L. Palano, L. Patrono, M.L. Stefanizzi, L. Tarricone, An IoT-Aware Architecture for Smart Healthcare Systems, IEEE Internet Things J. 2 (2015) 515–526. doi:10.1109/JIOT.2015.2417684.

[18] G. Manogaran, R. Varatharajan, D. Lopez, P.M. Kumar, R. Sundarasekar, C. Thota, A new architecture of Internet of Things and big data ecosystem for secured smart healthcare monitoring and alerting system, Futur. Gener. Comput. Syst. 82 (2018) 375–387. doi:10.1016/j.future.2017.10.045.

[19] F. Andriopoulou, T. Dagiuklas, T. Orphanoudakis, Integrating IoT and fog computing for healthcare service delivery, in: Components Serv. IoT Platforms, Springer, 2017: pp. 213–232.

[20] A.M. Rahmani, T.N. Gia, B. Negash, A. Anzanpour, I. Azimi, M. Jiang, P. Liljeberg, Exploiting smart e-Health gateways at the edge of healthcare Internet-of-Things: A fog computing approach, Futur. Gener. Comput. Syst. 78 (2018) 641–658. doi:10.1016/j.future.2017.02.014.

[21] D. Masouros, I. Bakolas, V. Tsoutsouras, K. Siozios, D. Soudris, From edge to cloud: Design and implementation of a healthcare Internet of Things infrastructure, in: 2017 27th Int. Symp. Power Timing Model. Optim. Simul., 2017: pp. 1–6. doi:10.1109/PATMOS.2017.8106984.

[22] Miami University, Telehealth, (2007). http://telehealth.med.miami.edu/what-is-telehealth (accessed September 19, 2017).

[23] U. Himss, Defining Key Health Information Technology Terms, Heal. (San Fr. (2008). http://www.himss.org/defining-key-health-information-technology-terms-onc-nahit (accessed October 5, 2017).

[24] S.R. Steinhubl, E.D. Muse, E.J. Topol, The emerging field of mobile health., Sci. Transl. Med. 7 (2015) 283rv3. doi:10.1126/scitranslmed.aaa3487.

[25] C. Free, G. Phillips, L. Watson, L. Galli, L. Felix, P. Edwards, V. Patel, A. Haines, The Effectiveness of Mobile-Health Technologies to Improve Health Care Service Delivery Processes: A Systematic Review and Meta-Analysis, PLoS Med. 10 (2013). doi:10.1371/journal.pmed.1001363.

[26] S. Hamine, E. Gerth-Guyette, D. Faulx, B.B. Green, A.S. Ginsburg, Impact of mHealth chronic disease management on treatment adherence and patient outcomes: A systematic review, J. Med. Internet Res. 17 (2015). doi:10.2196/jmir.3951.

[27] D. Lymberopoulos, A. Bamis, A. Savvides, Extracting spatiotemporal human activity patterns in assisted living using a home sensor network, Univers. Access Inf. Soc. 10 (2011) 125–138. doi:10.1007/s10209-010-0197-5.

[28] F. Paganelli, E. Spinicci, D. Giuli, ERMHAN: A context-aware service platform to support continuous care

networks for home-based assistance, Int. J. Telemed. Appl. 2008 (2008). doi:10.1155/2008/867639.

[29] E. Baralis, L. Cagliero, T. Cerquitelli, P. Garza, M. Marchetti, CAS-Mine: Providing personalized services in context-aware applications by means of generalized rules, Knowl. Inf. Syst. 28 (2011) 283–310. doi:10.1007/s10115-010-0359-z.

[30] D.B. Hoang, L. Chen, Mobile Cloud for Assistive Healthcare (MoCAsH), in: Proc. - 2010 IEEE Asia-Pacific Serv. Comput. Conf. APSCC 2010, 2010: pp. 325–332. doi:10.1109/APSCC.2010.10[illegible].

[31] P.D. Haghighi, A. Zaslavsky, S. Krishnaswamy, M.M. Gaber, Mobile data mining for intelligent healthcare support, in: Proc. 42nd Annu. Hawaii Int. Conf. Syst. Sci. HICSS, 2009. doi:10.1109/HICSS.2009.309.

[32] T.C. Panagiotakopoulos, D.P. Lyras, M. Livaditis, K.N. Sgarbas, G.C. Anastassopoulos, D.K. Lymberopoulos, A contextual data mining approach toward assisting the treatment of anxiety disorders, IEEE Trans. Inf. Technol. Biomed. 14 (2010) 567–581. doi:10.1109/TITB.2009.20[illegible]905.

[33] E. Badidi, L. Esmahi, A Cloud-based Approach for Context Information Provisioning, Inf. Technol. J. 1 (2011) 8.

[34] E. Ekonomou, L. Fan, W. Buchanan, C. Thüemmler, An integrated cloud-based healthcare infrastructure, in: Proc. - 2011 3rd IEEE Int. Conf. Cloud Comput. Technol. Sci. CloudCom 2011, 2011: pp. 532–536. doi:10.1109/CloudCom.2011.80.

[35] A. Forkan, I. Khalil, Z. Tari, Context-aware Cardiac Monitoring for Early Detection of Heart Diseases, Comput. Cardiol. 2013. 40 (2013) 277–280.

[36] N. Friedman, D. Geiger, M. Goldszmidt, G. Provan, P. Langley, P. Smyth, Bayesian Network Classifiers *, Mach. Learn. 29 (1997) 131–163. doi:10.1023/A:1007465528199.

[37] D. Berend, A. Kontorovich, A finite sample analysis of the Naive Bayes classifier, J. Mach. Learn. Res. 16 (2015).

[38] K.B. Wagholikar, S. Vijayraghavan, A.W. Deshpande, Fuzzy naive bayesian model for medical diagnostic decision support, Conf. Proc. Int. Conf. IEEE Eng. Med. Biol. Soc. 2009 (2009) 3409–3412. doi:10.1109/IEMBS.2009.5332468.

[39] H.H. Inbarani, A.T. Azar, G. Jothi, Supervised hybrid feature selection based on PSO and rough sets for medical diagnosis, Comput. Methods Programs Biomed. 113 (2014) 175–185. doi:10.1016/j.cmpb.2013.10.007.

[40] J. Chen, H. Huang, S. Tian, Y. Qu, Feature selection for text classification with Naïve Bayes, Expert Syst. Appl. 36 (2009) 5432–5435. doi:10.1016/j.eswa.200[illegible].054.

[41] P. Shanmuganathan, C.R. Sakthivel, An efficient naive bayes classification for sentiment analysis on twitter, Data Min. Knowl. Eng. 7 (2015) 180–185.

[42] R. Timofte, T. Tuytelaars, L. Van Gool, Naive bayes image classification: beyond nearest neighbors, in: Asian Conf. Comput. Vis., Springer, 2012: pp. 689–703.

[43] S. McCann, D.G. Lowe, Local Naive Bayes Nearest Neighbor for image classification, in: Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 2012: pp. 3650–3656. doi:10.1109/CVPR.2012.6248111.

[44] C. Zhang, G.-R. Xue, Y. Yu, H. Zhu, Web-scale classification with naive bayes, Proc. 18th Int. Conf. World Wide Web - WWW '09. (2009) 1083. doi:10.1145/1526709.1526867.

[45] J. Hernández-González, I. Inza, J.A. Lozano, Learning Bayesian network classifiers from label proportions, Pattern Recognit. 46 (2013) 3425–3440. doi:10.1016/j.patcog.2013.05.002.

[46] G.I. Webb, Naïve bayes, in: Encycl. Mach. Learn., Springer, 2011: pp. 713–714.

[47] P.R. Hof, E. Van Der Gucht, Structure of the cerebral cortex of the humpback whale, Megaptera novaeangliae (Cetacea, Mysticeti, Balaenopteridae), Anat. Rec. 290 (2007) 1–31. doi:10.1002/ar.20407.

[48] J.A. Goldbogen, A.S. Friedlaender, J. Calambokidis, M.F. McKenna, M. Simon, D.P. Nowacek, Integrative Approaches to the Study of Baleen Whale Diving Behavior, Feeding Performance, and Foraging Ecology, Bioscience. 63 (2013) 90–100. doi:10.1525/bio.2013.63.2.5.

[49] W.A. Watkins, W.E. Schevill, Aerial Observation of Feeding Behavior in Four Baleen Whales: Eubalaena glacialis, Balaenoptera borealis, Megaptera novaeangliae, and Balaenoptera physalus, J. Mammal. 60 (1979) 155–163. doi:10.2307/1379766.

[50] S. Mirjalili, A. Lewis, The Whale Optimization Algorithm, Adv. Eng. Softw. 95 (2016) 51–67. doi:10.1016/j.advengsoft.2016.01.008.

[51] A. Tharwat, Y.S. Moemen, A.E. Hassanien, Classification of toxicity effects of biotransformed hepatic drugs using whale optimized support vector machines, J. Biomed. Inform. 68 (2017) 132–149. doi:10.1016/j.jbi.2017.03.002.

[52] 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).

[53] H. He, E.A. Garcia, Learning from imbalanced data, IEEE Trans. Knowl. Data Eng. 21 (2009) 1263–1284. doi:10.1109/TKDE.2008.239.

[54] Y. Sun, M.S. Kamel, A.K.C. Wong, Y. Wang, Cost-sensitive boosting for classification of imbalanced data, Pattern Recognit. 40 (2007) 3358–3378. doi:10.1016/j.patcog.2007.04.009.

[55] V. López, A. Fernández, S. García, V. Palade, F. Herrera, An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics, Inf. Sci. (Ny). 250 (2013) 113–141. doi:10.1016/j.ins.2013.07.007.

[56] A. Estabrooks, T. Jo, N. Japkowicz, A Multiple Resampling Method for Learning from Imbalanced Data Sets, Comput. Intell. 20 (2004) 18–36. doi:10.1111/j.0824-7935.2004.t01-1-00228.x.

[57] J. Laurikkala, Improving identification of difficult small classes by balancing class distribution, Proc. 8th Conf. AI Med. Eur. Artif. Intell. Med. (2001) 63–66. doi:10.1007/3-540-48229-6_9.

[58] C. Elkan, The foundations of cost-sensitive learning, in: IJCAI Int. Jt. Conf. Artif. Intell., 2001: pp. 973–978. doi:doi=10.1.1.29.514.

[59] K.M. Ting, An instance-weighting method to induce cost-sensitive trees, IEEE Trans. Knowl. Data Eng. 14 (2002) 659–665. doi:10.1109/TKDE.2002.1000348.

[60] R. Akbani, S. Kwek, N. Japkowicz, Applying Support Vector Machines to Imbalanced Datasets, Lnai. 3201 (2004) 39–50. doi:10.1007/978-3-540-30115-8_7.

[61] R.H. Fagard, V.A. Cornelissen, Effect of exercise on blood pressure control in hypertensive patients., Eur. J. Cardiovasc. Prev. Rehabil. 14 (2007) 12–7. doi:10.1097/HJR.0b013e3280128bbb.

[62] P. Jansen, M. Leineweber, T. Thien, The effect of a change in ambient temperature on blood pressure in normotensives, J. Hum. Hypertens. 15 (2001) 113–117. doi:10.1038/sj.jhh.1001134.

[63] I.H. Witten, E. Frank, M. a. Hall, J. Cunningham, M. Humphrey, L. Hunt, B. Mcqueen, L. Smith, S. Inglis, C. Nevill-manning, Data Mining: Practical Machine Learning Tools and Techniques, in: Data Min. Pract. Mach. Learn. Tools Tech., 4th ed., Kaufmann, Morgan, 2016: pp. 539–557. doi:http://dx.doi.org/10.1016/B978-0-12-417295-1.09989-6.

[64] M. Zaharia, M. Chowdhury, M.J. Franklin, S. Shenker, I. Stoica, Spark□: Cluster Computing with Working Sets, HotCloud'10 Proc. 2nd USENIX Conf. Hot Top. Cloud Comput. (2010) 10. doi:10.1007/s00256-009-0861-0.

[65] G.B. Moody, R.G. Mark, A.L. Goldberger, Physionet: A web-based resource for the study of physiologic signals, IEEE Eng. Med. Biol. Mag. 20 (2001) 70–75. doi:10.1109/51.932728.

[66] A.L. Goldberger, L.A.N. Amaral, L. Glass, J.M. Hausdorff, P.C. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C.-K. Peng, H.E. Stanley, PhysioBank, PhysioToolkit, and PhysioNet, Circulation. 101 (2000) E215-20. doi:10.1161/01.CIR.101.23.e215.

[67] D.J. Scott, J. Lee, I. Silva, S. Park, G.B. Moody, L.A. Celi, R.G. Mark, Accessing the public MIMIC-II intensive care relational database for clinical research, BMC Med. Inform. Decis. Mak. 13 (2013). doi:10.1186/1472-6947-13-9.

[68] M. Saeed, M. Villarroel, A.T. Reisner, G. Clifford, L.-W. Lehman, G. Moody, T. Heldt, T.H. Kyaw, B. Moody, R.G. Mark, Multiparameter Intelligent Monitoring in Intensive Care II: A public-access intensive care unit database*, Crit. Care Med. 39 (2011) 952–960. doi:10.1097/CCM.0b013e31820a92c6.

[69] M.A.P. de Castro Ferreira, SHRAM-Smart Heart Rate and Activity Measurement, University of Porto, 2016.

[70] G. Sharma, J. Martin, MATLAB®: a language for parallel computing, Int. J. Parallel Program. 37 (2009) 3–36.

[71] E.C. Jauch, B. Cucchiara, O. Adeoye, W. Meurer, J. Brice, Y.Y.F. Chan, N. Gentile, M.F. Hazinski, Part 11: Adult stroke: 2010 American Heart Association Guidelines for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care, Circulation. 122 (2010). doi:10.1161/CIRCULATIONAHA.110.971044.

[72] M.A. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten, The WEKA data mining software: an update, SIGKDD Explor. 11 (2009) 10–18. doi:10.1145/1656274.1656278.

[73] D.M.W. POWERS, Evaluation: From Precision, Recall and F-Measure To Roc, Informedness, Markedness & Correlation, J. Mach. Learn. Technol. 2 (2011) 37–63. doi:10.1.1.214.9232.

[74] E. Emary, H.M. Zawbaa, Impact of chaos functions on modern swarm optimizers, PLoS One. 11 (2016). doi:10.1371/journal.pone.0158738.

[75] C.E. Metz, Basic principles of ROC analysis, Semin. Nucl. Med. 8 (1978) 283–298. doi:10.1016/S0001-

2998(78)80014-2.

Mohammed K. Hassan received his B.Sc., M.Sc., and Ph.D. degrees in Computer engineering from the Faculty of Engineering, Mansoura University, Egypt. His research interests include (but are not limited to) data mining, classification, big data, cloud computing and Big Data analysis.

Ali I. El Desouky is a full-time Professor at the Computer and Control Department, Mansoura University. He received the B.Sc. degree in electrical engineering from the Faculty of Engineering, Mansoura University, Egypt. He received his M.Sc. and PhD degrees in computer science and control from the Departments of Electronics and Electrical Engineering, University of Glasgow, Glasgow, UK. He also took over many positions of leadership and supervision of many scientific papers. His research interests include (but are not limited to) data mining, classification, Big Data analysis, cloud computing, No-SQL databases, Hadoop and Map Reduce techniques and software engineering. He has published more than 150 articles in well-known international journals.

Sally M. Elghamrawy is an Assistant Professor at the MISR Engineering & Technology Institute and part-time Assistant Professor at computers engineering department -Faculty of Engineering, Mansoura University in Egypt. From the same department, she received a PhD degree in 2012 in Distributed Decision Support Systems Based on Multi Intelligent Agents, received the M SC degree in Automatic Control Systems Engineering in 2006, and received B. Sc. in Computers Engineering and Systems in 2003. She delivering lectures, supervising graduation projects, master's thesis, and doctoral dissertations. Her research focuses on Big Data analysis, No-SQL databases, Hadoop and MapReduce techniques, and software engineering. She is the author of number peer-reviewed publications, receiving best paper awards. She is also an IEEE Member.

Amany M.Sarhan has received the B.Sc. degree in Electronics Engineering, and M.Sc. in Computer Science and Automatic Control from the Faculty of Engineering, Mansoura University, respectively. She awarded the PhD degree as a joint research between Tanta University, Egypt and the University of Connecticut, USA. She is working now as a full Prof. at Computers and Automatic Control Dept., Tanta Univ., Egypt. Her interests are in the area of Network Security, Software restructuring, Object-oriented Database, Fragmentation and allocation of databases and distributed systems and Computations.

Highlights

- A Hybrid Real-time Remote Monitoring (HRRM) framework for patients suffering from chronic diseases is proposed.
- A Hybrid Knowledge Discovery Module (HKDM) is proposed to classify patient’s health status on dual-mode (online - offline).
- The proposed HKDM addresses the problem of imbalanced datasets in big data.
- Naïve Bayes – Whale Optimization Algorithm (NB-WOA) is proposed to select the minimum features’ set required to ensure the continuity of the model's work with highest efficiency and speed.