




# Machine Learning Based Prediction and Modeling in Healthcare Secured Internet of Things

Charafeddine E. Aitzaoui<sup>1</sup> · Adnane Latif<sup>1</sup> · Abderrahim Benslimane<sup>2</sup> · Hui-Hsin Chin<sup>3</sup> 

Accepted: 26 November 2020 / Published online: 15 January 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC part of Springer Nature 2021

## Abstract

In this paper, we present the concept and the prototype implementation of our novel “Smart Observatory of Involuntary Medical Seizures (SOIMS)”. SOIMS merges Wireless Body Area Networks (WBAN), Internet of Things (IoT) and Machine Learning (ML) as an intelligent platform for the prediction and modelling of involuntary seizures. The prediction process is elaborated with our proposed algorithms, namely, Qualifying Linear Regression Algorithm (QuLRA), Selective Clustering Algorithm (SeCA) and Real Time Clusters Correlation Algorithm (RT2CA). The assessment of the proposed system is validated based on the Physionet ECG patients’ dataset. The implementation of the prototype involves an IoT/WEB proxy security embedded for translation between nodes CoAP/DTLS protocol and Hospital Information System (HIS) HTTP/TLS protocol. Our proposed solution outperforms existing schemes in the literature at different levels, namely: a) it uses a hierarchical combination of machine learning and prediction algorithms; b) it is open-source, interoperable and user friendly; c) it is a secured prototype implementation; and d) it reaches a higher rate of accuracy according to the correlation criterion.

**Keywords** WBAN · Machine learning · IoT · ECG · Tachycardia · Prediction

## 1 Introduction

Epilepsy, asthma, cardiac syncope, and other involuntary medical crises are the most important causes of humanity’s suffering, and even with the relentless evolution of medicine; the random behaviour of these seizures is a major drawback in the process of diagnosis and treatment of these diseases.

Therefore, the reduction of these risk factors remains the best therapeutic management to reduce the number of patient deaths.

In the medical field, doctors are regularly required to make several decisions, such as the prevention of pathologies, or the prediction of seizures for chronic diseases with random behaviour.

In this work, we aim to introduce the design and implementation of an intelligent platform for the prediction of

involuntary seizures that we called “Smart Observatory of Involuntary Medical Seizures (SOIMS)”. SOIMS is the real-time predictive medical platform, which combines three technologies:

- **Wireless Body Area Network:** a wireless network of human body sensors. Each sensor node is generally able to detect one or more physiological signals from the human body or its environment.
- **The Internet of Things (IoT):** a technology that provides advanced services by interconnecting objects (physical or virtual), through information and communication technologies;
- **Artificial Intelligence:** a system integrating developed algorithms of machine learning, filtering, data adaptation, and modeling of patients’ profiles.

The primary purpose of SOIMS scheme is to help doctors and medical practitioners to make effective decisions based on artificial intelligence and internet of things capabilities.

Among several, the most important technical contribution of the proposed SOIMS scheme lies in the possibility to predict medical seizures nesting multiple machine learning algorithms into a single prediction chain to prevent, prematurely, patients from possible damages or bad consequences. With

✉ Hui-Hsin Chin  
hh.chin@ocu.edu.tw

<sup>1</sup> National School of Applied Sciences, University of Cadi Ayyad, Marrakesh, Morocco

<sup>2</sup> LIA/CERI, University of Avignon, Avignon, France

<sup>3</sup> Department of Information Technology, Overseas Chinese University, Taichung, Taiwan

this aim, our proposed method for seizures prediction consists of three algorithms:

- A qualifying algorithm based on linear regression which detects if patients' medical dataset is normal and then whether a suspect crisis is coming.
- A selective algorithm that identifies the pre-crisis phase in the whole patient's dataset.
- A decisive algorithm that uses the correlation metric to decide if a seizure will occur.

Thereafter, to evaluate the SOIMS scheme, we implemented a prototype, using the *Raspberry Pi* middleware IoT proxy. This implementation allows the collection of ECG signal of the patient and notifies the patient in the case of a predicted crisis. Throughout this paper, the general scheme of SOIMS, the prediction process, and an implemented prototype will be presented. The paper is organized as follows: Section 1 discusses related works and some motivations. Section 2 introduces an overview on Machine Learning, IoT and WBANs. Section 3 examines our proposed algorithm for seizures prediction. Section 4 presents the implementation of the proposed algorithm on our proposed IoT proxy. The summaries of this work and future work will be stated in the last section.

## 2 Related works

Recently, the integration of IoT and WBAN in the medical field has received great interest from researchers. Consequently, related to IoT, many studies have been conducted such as:

- Tyagi et al in [1] highlighted the gain of IoT in healthcare and investigated its technical features to make it a reality and identify the opportunities.
- Xu et al in [2] presented a data model to record and use the IoT data. They also designed and developed a resource-based ubiquitous data accessing method to collect and publish IoT data globally so that it can be accessed anywhere and anytime.

Related to ML many studies on the application of machine learning to the prediction of diseases and seizures have been published, for instance:

- Kiral-Kornek et al in [3] explored deep learning algorithm in order to extract features from Electroencephalogram (EEG) to predict epilepsy seizures.
- Hijazi et al in [4] proposed a method to filter patients' electrocardiograms (ECGs) and applied machine-learning classifiers to identify cardiac health risks and estimate severity.

These studies altogether support the idea that machine learning and IoT are the future of medical research in many steps starting with data collection and finishing with decision making (e.g. prediction). Considered independently, each of these technologies is already impressive. However, the challenge is to merge them into one ecosystem. This convergence became today a necessity. In the literature, many publications have discussed the topics of IoT and ML convergence for many applications, Chunxiao Jiang et al. in [5] reviewed the benefits of artificial intelligence that aided 5G wireless systems equipped with machine learning. Khalim Meerja et al. in [6], discussed the issues of handling data analysis in the new cloud based IoT network architecture, and the incurred price and overall efficiency for storing and analyzing data for these networks.

In our turn, through this paper, we suggest a realistic application where we merge IoT and ML to benefit healthcare in the context of SOIMS, the smart predicting platform that ensures:

- Monitoring of patient in mobile and real-time mode using our proposed IoT proxy;
- Prediction of involuntary seizures before it occurs based on our proposed algorithms;
- Notification to the patient once the crisis is expected; and
- Alerting the medical monitoring team in case of predicted crisis.

Meanwhile, cardiovascular diseases are of the great cause of death in the world with a rate of 17.9 million per year according to the WHO estimate (2016 statistics) [7]. In this paper, the concept of SOIMS was applied to Junctional Tachycardia seizure [8].

## 3 Machine learning and IoT in medical field

### 3.1 Machine learning in medical field

The main objective of machine learning starts with observations that we provide, in order to look for models that help us make better decisions in the future without human intervention and adapt actions.

Therefore, machine learning is the science that gives systems the ability to learn and improve automatically. It is based on the development of computer programs that can use real-world observations and interactions and learn for themselves.

Thus, in a medical application, machine learning can facilitate accurate medical diagnosis by taking into account clinical trials of patient data, studies and researches, see [9, 10].

In literature, there are two main families of machine learning, namely, supervised and unsupervised machine learning.

### 3.1.1 Supervised machine learning

Each supervised process has two phases, see [11]: the first is the so-called “learning phase” during which the “labeled data” model is to be determined. Then, this knowledge will be developed in the second phase which is called “test phase”, which consists of predicting the class or the label of a new datum from the model that has been learned in the first phase.

There are many supervised learning algorithms, but in this paper, we will limit our interest to simple linear regression.

In Simple linear regression, a variable  $Y$  is modeled by an affine function of another variable  $X$ . It may be just an exploratory approach, to describe the influence of the variable  $X$  over the variable  $Y$ . If we denote  $Y$  the real random variable to be explained and  $X$  the explanatory variable. The model is to assume that, on average,  $E(Y)$  is an affine function of  $X$ . The formula of the model implicitly assumes a prior notion of causality, in the sense that  $Y$  depends on  $X$ , because the model is not symmetrical, see [12].

$$E(Y) = f(X) = \beta_0 + \beta_1 X \quad (1)$$

Where

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (2)$$

$Y$  is the explained variable,

$X$  is the explanatory variable,

$\beta_0$  and  $\beta_1$  are the regression coefficients.

### 3.1.2 Unsupervised machine learning

This type of technique aims to find the similarities and paths between data where all the variables used are inputs, see [13]. There are many unsupervised machine learning algorithms, but in this paper, we singled out the K-Means clustering based on the correlation coefficient as a discriminating criterion between clusters.

- Clustering (K-Means)

K-means clustering is a type of unsupervised learning used to classify unlabeled data. The goal of this algorithm is to discriminate data into groups; the number of groups is represented by the variable  $K$ . This method aims to assign data to classes (clusters) based on the similarities of individuals.

The K-means clustering algorithm uses iterative refinement to produce the result. First, we provide the  $K$ , which means the number of clusters, so that the algorithm separates the data into  $K$  different groups. K-Means calculates the distance

between each instance to express similarities. The algorithm then iterates between two steps:

- Data assignment step:

In this step, each data point is assigned to its closest centroid using Euclidean squared minimum distance

$$\text{Arg min}_{ci \in C} \text{dist}(Ci, X) \quad (3)$$

Where  $ci$  is the collection of centroids in the set  $C$ ,  $X$  is a data point and  $\text{dist}(\cdot)$  is the Euclidean distance, see [14].

- Centroid update step:

In this step, the centroids are recalculated by averaging all data points assigned to their clusters. The cluster changes its center after each iteration until it reaches the stagnation phase.

- Correlation

Studying the correlation between two or more random variables is accomplished by studying the intensity of the connection that can exist between these variables.

The correlation coefficient between two real random variables  $X$  and  $Y$ , denoted by  $\text{Cor}(X, Y)$ , is defined by:

$$\text{Cor}(X, Y) = \frac{E(XY) - E(X)E(Y)}{\sigma_X \sigma_Y} \quad (4)$$

Where  $E(X)$  is the mathematical expectation and  $\sigma_X$  and  $\sigma_Y$  are their respective variances [15].

## 3.2 IoT and WBAN integration in E-health

The Internet of Things is a technical concretization of ubiquitous computing systems, where technology is naturally integrated with our daily objects, see [9, 16, 17]. This very promising scheme paves the way for a multitude of scenarios, based on the interconnection between the physical and the virtual world: e-health, smart cities, logistics, and security.

It is becoming increasingly important to use IoT technology in different healthcare sectors, see [18].

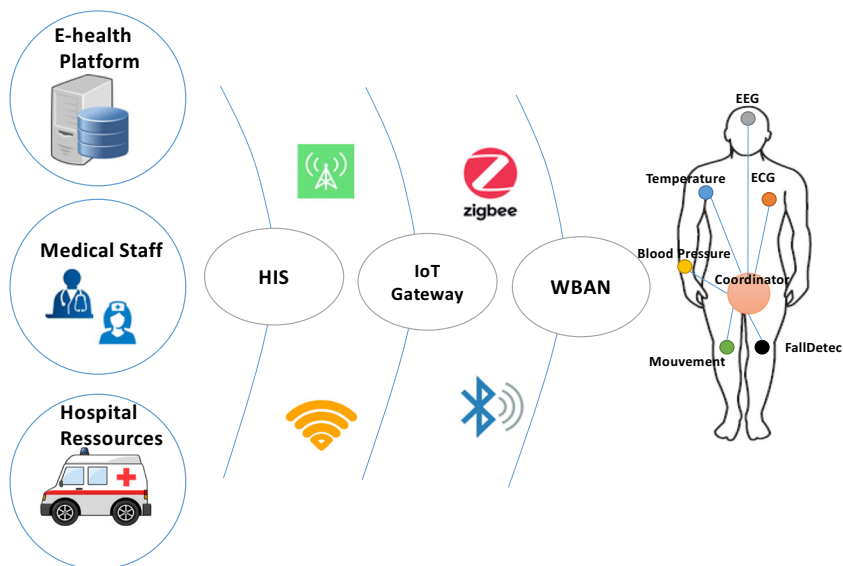
The authors in [9], the authors presented a survey of related works on IoT and ML in e-healthcare during the last five years in its general architecture and application.

As shown in Fig. 1, IoT compatible devices allow doctors to collect information that they would not know otherwise.

With the innovation of wearable sensors and wireless communication technologies, the wireless body area network is gaining attraction from the IoT healthcare systems.

A WBAN is either a wearable and/or implanted human body network of sensors. Each sensor node is generally able to detect one or more physiological signals from the human body or its environment. The sensor node stores and then

Fig. 1 General IoT healthcare architecture



transmits the measured data via a wireless network protocol. Among the major radio standards that has been used in WBANs, we quote Bluetooth, IEEE 802.15.4 and IEEE 802.15.6, see [19, 20].

Having discussed Medical IoT architectures and ML algorithms, the following sections of this paper addresses new schemes of merging these two technologies into one intelligent platform. To implement SOIMS, we propose:

- a hierarchical real-time ML-based prediction scheme using linear regression, clustering and correlation algorithms.
- an IoT proxy that acts as an **intermediary** between our ECG node, implementing an IoT CoAP/DTLS layer stack, and the Hospital Information System, implementing HTTP/TLS layer stack (see Fig. 2).

### 4 Hierarchical real time ML based prediction algorithm

In order to implement and evaluate the SOIMS scheme applied to Junctional Tachycardia seizures, we first propose our prediction methods and algorithms, then we use Physionet ECG data sets (MIT-BIH Arrhythmia Database [21]) for the purpose of validation. This database comprises 48 half-hour selections of two-channel ambulatory ECG recordings obtained from 47 patients with several arrhythmia diseases.

The procedure of seizure prediction is elaborated in three hierarchical stages, as shown in Fig. 3. The detail of used algorithms and criteria choice is given in the next sections.

#### 4.1 Qualifying linear regression algorithm (QuLRA)

As illustrated in Fig. 4, the process consists of applying the linear regression algorithm to both of real-time and learning datasets of each patient. The choice of linear regression for this stage is justified by the fact that cardiologists consider that the ECG exam is normal, when the baseline is identical or parallel to the isoelectric line ( $\beta_1 = \beta'_1$ ), see [22].

The isoelectric line given by the real-time dataset is described as:

$$f(X) = \beta_1 X + \beta_0 \tag{5}$$

And the baseline given by the learning dataset is described as:

$$f(X) = \beta_1' X + \beta_0' \tag{6}$$

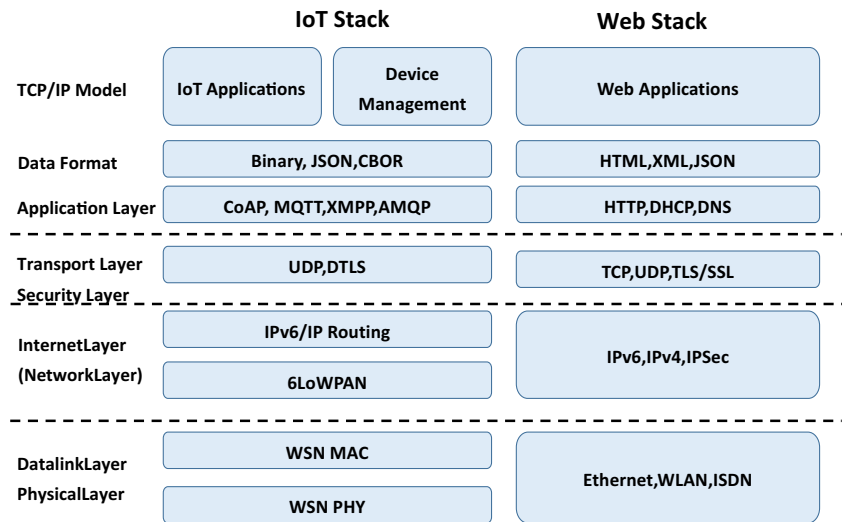
This preliminary stage that we propose is very important for the prediction algorithm for two reasons:

- Firstly, it verifies if the ECG real-time dataset is normal or not, and then if a suspect crisis is coming.
- Secondly, this preliminary algorithm optimizes the system resources allocation and exempts it from additional treatment and energy consumption since the algorithm’s second step begins only after a positive result of the first step.

#### 4.2 Selective clustering algorithm (SeCA)

When this stage is triggered it means that a positive preliminary check is satisfied. In this step, SOIMS uses another

Fig. 2 Difference between IoT stack and Web stack



patient’s training dataset (for the same patient) which contains some already recorded seizures.

Thereafter we apply our proposed (SeCA) based K-means algorithm in order to classify the learning dataset into four clusters as it is observed in Fig. 5.

The consideration of non-supervised algorithm is justified by the instability of morphological and temporal characteristics of ECG for different patients and under different temporal and physical conditions, see [23].

The choice of K-means for this algorithm is motivated by two factors:

- ECGs produce a huge dataset and
- K-means is the fastest clustering algorithm, see [24].

In K-means, the parameter K is equal to 4 in our scheme because each crisis is characterized by 4 phases, namely, normal phase, pre-crisis phase, crisis phase, and post-crisis phase.

Therefore, the first cluster is the cluster of normal values where no crisis is detected. The second cluster is the pre-crisis

cluster, which is also the most important cluster that contains values before crises that will be used thereafter for the prediction process. The third and the fourth clusters concerns respectively the crisis and the post-crisis phases.

Hence, the selection of the pre-crisis cluster is very important for the following steps. Our experimental researches demonstrate that the centroid of the pre-crisis cluster (contains most of pre-crisis pulses values) is the nearest centroid to the linear regression model in the sense of Euclidean distance in Eq. (8).

$$\min_{ci \in C} \text{dist}(ci, f(x)) \tag{7}$$

Where  $f(x)$  is the model given in Eq. 1 and  $ci$  is the collection of centroids in the set C.

$$\text{dist}(Ci, f(x)) = \frac{|\beta_1 Ci + \beta_0|}{\sqrt{\beta_1^2 + \beta_0^2}} \tag{8}$$

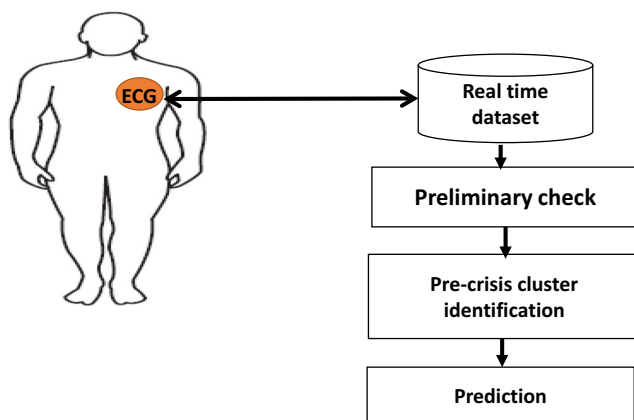


Fig. 3 SOIMS: three steps prediction algorithm

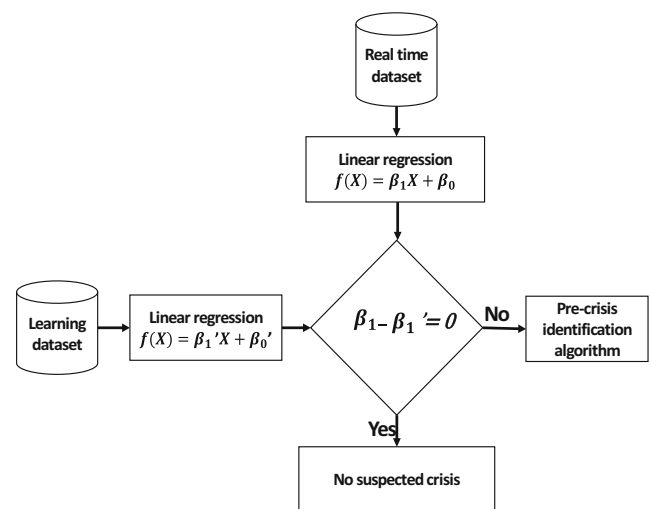


Fig. 4 Preliminary check algorithm

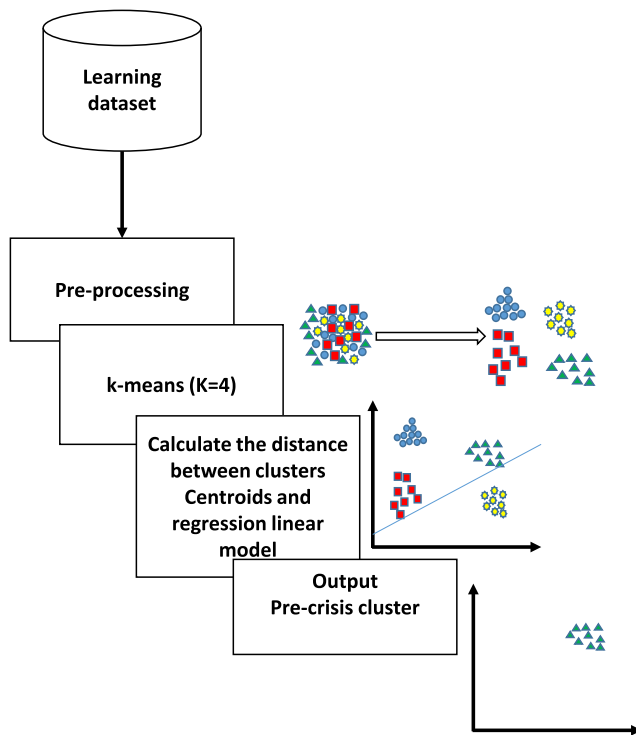


Fig. 5 Pre-crisis cluster identification process

### 4.3 Real time clusters correlation algorithm (RT2CA)

In this step, a cluster correlation algorithm is used between real time dataset clusters and learning dataset clusters. As for the learning dataset, the RT2CA algorithm that we suggest implements the K-means algorithm to classify the real time dataset into four clusters. Then, we use the linear correlation as a discrepancy measure between real time clusters and the pre-crisis cluster identified by (SeCA) algorithm in order to detect whether pre-crisis’s symptoms are detected. The algorithm process is illustrated in Fig. 6. The decision rule is set according to the correlation ratio (CR) given by Eq.(9).

$$CR(X, Y) = \frac{E(XY) - E(X)E(Y)}{\sigma_X - \sigma_Y} \tag{9}$$

Where X represents real-time dataset cluster and Y represents learning dataset pre-crisis cluster,  $E(X)$  is the mathematical expectation and  $\sigma_X$  and  $\sigma_Y$  their respective variances.

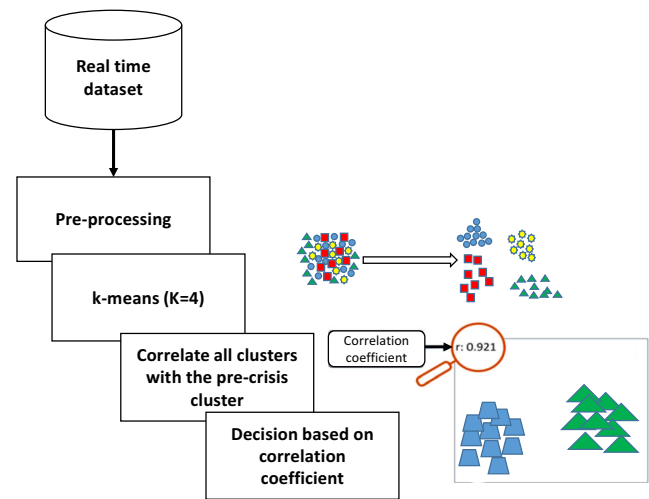


Fig. 6 Pre-crisis correlation with real time dataset clusters and decision-making based on CR

Two clusters have strong dependency when their correlation coefficient value is close to 1. Experimentally, it was observed the presence of crisis (in dataset) when  $CR > 0.9$ . In our proposed algorithm, the decision metrics for the prediction algorithm are summarized in Table 1.

## 5 Mobile WBAN IoT implementation for JT prediction

The prototype implementation of SOIMS for JT context considers both of hardware and software implementation. Figure 10 illustrates the prototype implementation of the proposed design.

### 5.1 Design and technical specifications

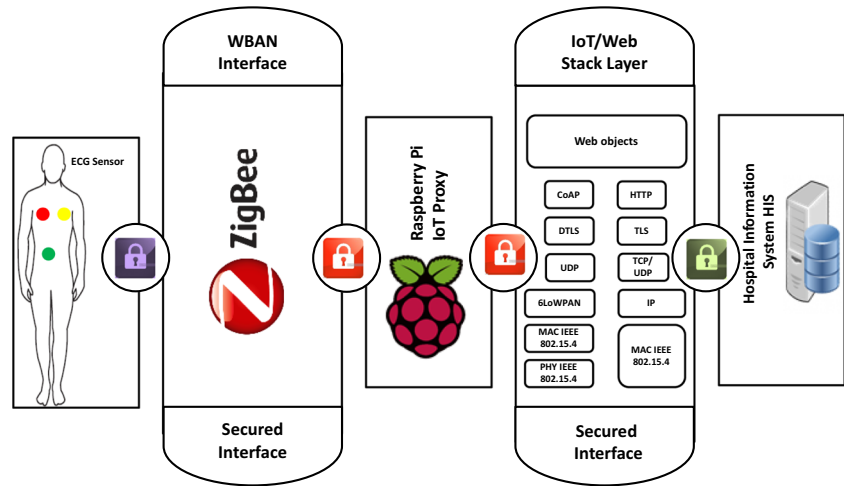
The IoT design for SOIMS that we propose in Fig. 7 is a high-level architecture respecting the Continua Design Guidelines (CDG) [25].

On the patient side, we use an ECG sensor to deliver the electrical activity of the patient’s heart. The communication between the ECG sensor and the *Raspberry Pi* proxy is Constrained Application Protocol (CoAP) based. The choice of CoAP as node applicative layer communication is justified by its high performances in terms of low energy consumption and low resources allocation (RAM and CPU). Thus, our proposed prototype implements the CoAP in both of ECG node and *Raspberry Pi* middleware proxy.

Table 1 Prediction decision in function of CR

Correlation Ratio	CR < 0.6	0.6 ≤ CR ≤ 0.9	0.9 < CR
Decision	Normal state	Warning	Predicted seizure is coming

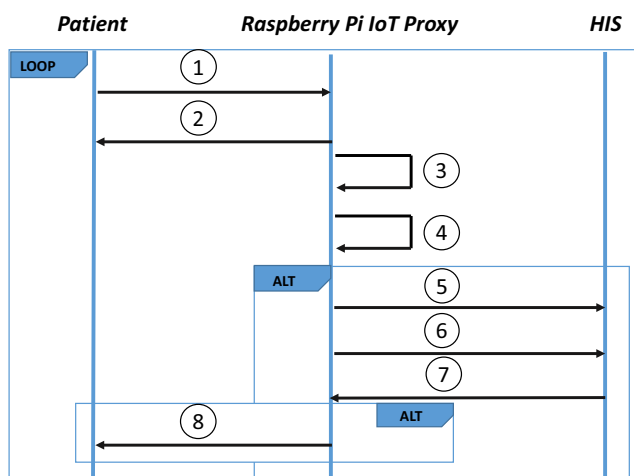
**Fig. 7** High Level IoT Architecture for SOIMS



Because the general Hospital Information Systems are web service HTTPS/HTTP based, we implement a *Raspberry Pi* proxy to translate protocols between the CoAP client (ECG sensor) and the HTTPS server (HIS), thus our IoT proxy implements also the HTTPS/TCP web stack layer.

Figure 8 illustrates the changes of request and response messages between the two sides:

1. The CoAP client posts periodically the ECG data of the patient to the IoT proxy;
2. The IoT proxy (CoAP agent) acknowledges the received message;
3. The IoT proxy decapsulates the posted CoAP message content and stores the data in the real-time data set;
4. Then the implemented Hierarchical Real-Time ML-based Prediction Algorithm presented in the previous section is applied;
5. When a positive crisis is predicted a notification is sent to the HIS using HTTPS;



**Fig. 8** Exchanged request/response messages between IoT proxy and HIS

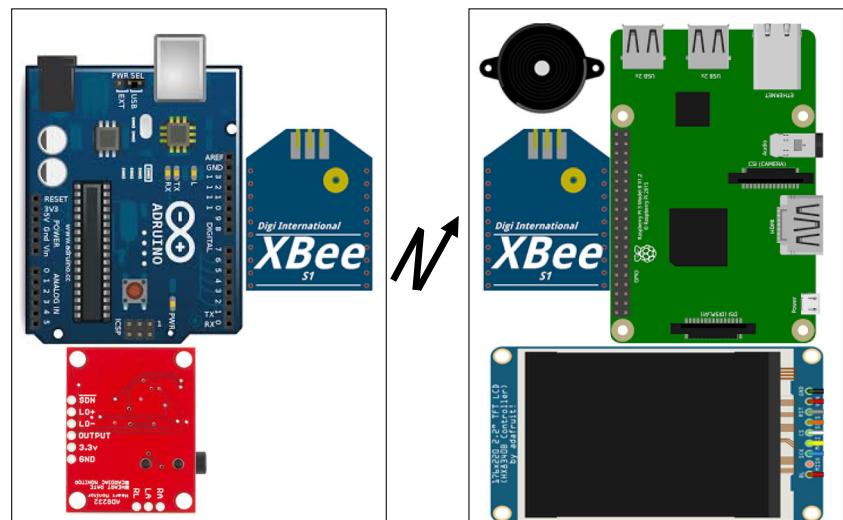
6. The IoT proxy sends the real-time ECG of the patient to HIS;
7. If the crisis is confirmed by the medical team, an alert is sent to the proxy; and
8. The IoT proxy display/play the alert to the patient

## 5.2 Hardware prototype design

In this work, the data acquisition is done in real-time on physiological monitoring data of patients; the collection is done by attaching to each patient a biomedical IoT prototype. As illustrated in Fig. 9 the prototype is constituted of:

- A heart rate sensor (AD8232) which is an integrated signal-conditioning block for ECGs and other bio-potential measurement applications. It is designed to extract, amplify and filter weak bio-potential signals in noisy conditions. This design makes it possible to use an analog-digital converter (ADC) or an integrated microcontroller in order to easily acquire the output signal, see [26].
- An Arduino data acquisition board which is an electronic board equipped with a microcontroller. The microcontroller allows the events detected by sensors to program and control actuators, see, [27].
- An IoT semantic proxy which is based on the *Raspberry Pi* B3 card for the data processing and the execution of the prediction algorithm. The *Raspberry Pi* card is equipped with a powerful Quad-Core ARM 1.2 GHz processor, 1024 MB of RAM, a Dual-Core GPU, and natively equipped with Wi-Fi b / g / n and Bluetooth 4.1, see [28, 29].
- WBAN communication interface which uses an Xbee module, see [30].

**Fig. 9** Hardware used in the prototype implementation



### 5.3 Software implementation

The *Raspberry Pi* proxy is used in the context of this implementation, for three main functions:

- Local real-time running of our prediction algorithm of tachycardia junction seizures,
- Remote communication with the observatory unit (Telemonitoring Team) in case of a predicted crisis, as well as the transfer of the daily ECG routine of a supervised patient, and
- Protocol translation between the ECG node implementing an IoT layer stack and HIS implementing a WEB layer stack.

With this aim, we have built a software using Python framework embedded with the *Raspbian* operating system. The developed software is the implementation of the aforementioned algorithms, in addition to the communication with the observatory unit (doctors' team), since even if the *Raspberry Pi* system can predict a coming crisis, the medical team still must confirm the final decision. In all cases, the system sends real-time ECG data to the observatory unit using the internet. In case of a high likely crisis, an alert is also sent.

The effective demonstration of our designed SOIMS prototype is given in Fig. 10.

## 6 Results and discussions

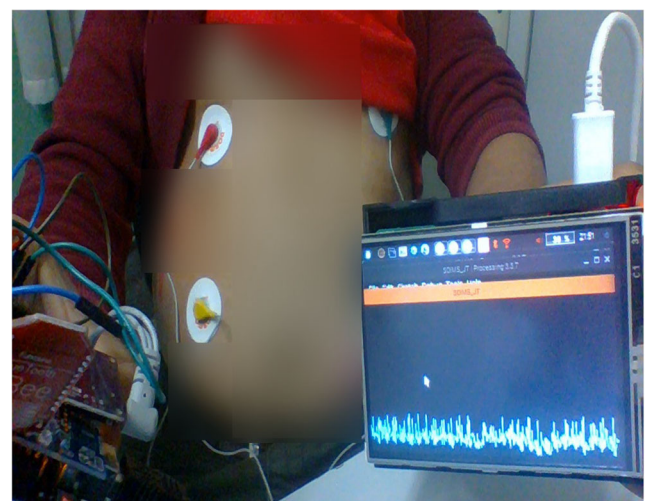
In this section, we present and analyze the obtained results. To this propose, the process described in the previous sections will be applied systematically on several patients' data set.

Starting with the QuLRA algorithm, the linear regression method allows us to detect the appearance of ECG signal

abnormality. The comparison between a healthy and a suffering patient is illustrated in Fig. 11 and Fig. 12. Figure 11 shows that the baseline is identical to the isoelectric line, as a result, the ECG is normal and no seizure has been detected (negative QuLRA test). Whereas, Fig. 12 shows that the baseline is inclined (i.e. not identical to the isoelectric line), which means that the ECG signal is not normal (positive QuLRA test) and a suspected seizure is potentially coming.

As the patient's ECG in Fig. 12 presents a positive QuLRA test, the implemented system starts the SeCA algorithm, which is based on the K-means clustering algorithm. As mentioned in the previous sections, the dataset is clustered into four clusters, and the most important cluster for us is the pre-crisis cluster.

Practically, in all our tests, it was observed that the pre-crisis cluster is the one where the centroid is the nearest to



**Fig. 10** Prototype demonstration of SOIMS



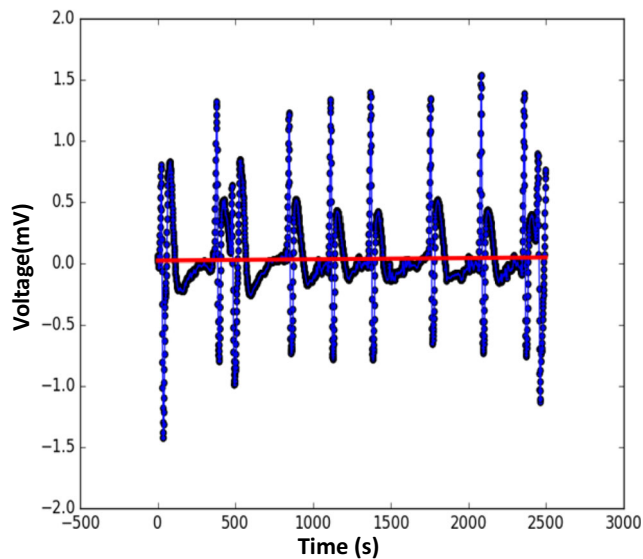


Fig. 11 Linear Regression Model of patient where no crisis is suspected

the regression linear model (slanting line in case of crisis), which helps to readily choose the best pre-crisis cluster.

Figure 13 illustrates the four ECG clusters (learning dataset), where the red cluster is the pre-crisis cluster and the green one is the crisis cluster. Furthermore, we can also observe that the distance between the centers (bleu marks) of the two clusters is about  $(1670 - 920 = 750 \text{ s})$  12.5 min. Thus, the proposed system is able to forecast the occurrence of the crisis 12.5 min in advance.

Thereafter our system launches the RT2CA process which consists of clustering the real-time collected patient dataset into four clusters and correlates it with the pre-crisis cluster identified by SeCA. When the correlation ratio corresponds to the criterion of the targeted class ( $CR > 0.9$ ), an alert is sent to

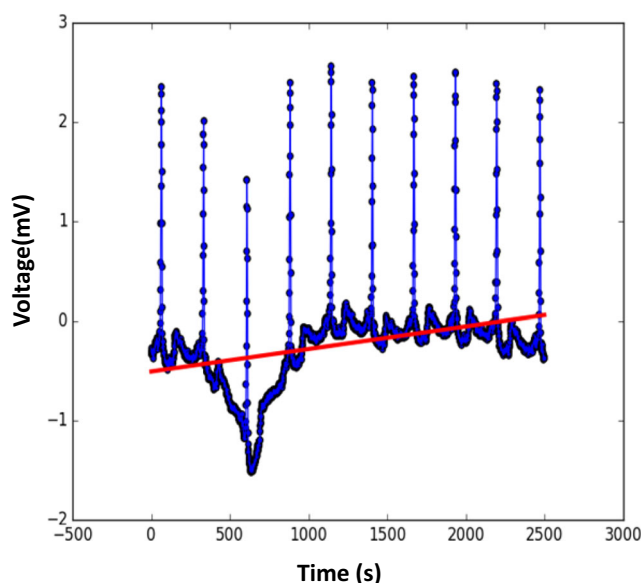


Fig. 12 Linear Regression Model of patient where crisis is suspected

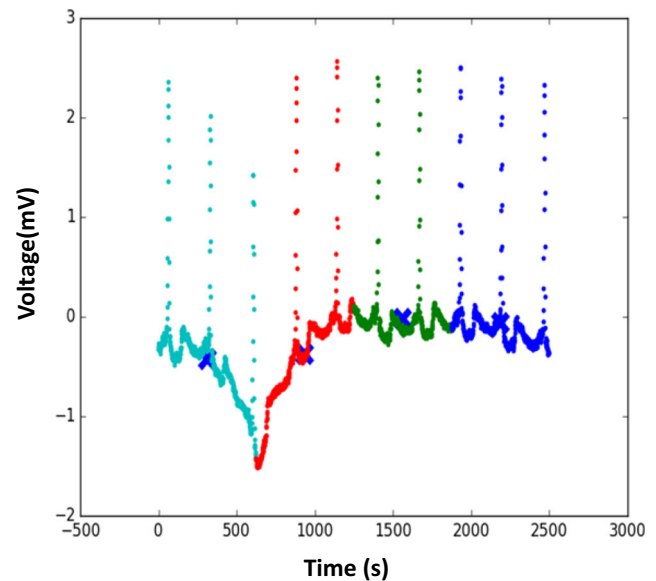


Fig. 13 Abnormal patient ECG dataset clusters

the observatory unit to decide if the ECG signal is referring to a high likely coming crisis. In Fig. 14, we observe that the first warning is received at the time 14:57 min, then the prediction algorithm detects at 15:00 that a real crisis is coming.

Once the observatory unit confirms an alert, our proposed Raspberry Pi proxy notifies the patient that a crisis is coming and advises him, for example, to prevent from driving.

To the best of our knowledge, compared to the existing works that investigate other heart diseases prediction, our proposed approach outperforms all the existing works without any exception either in terms of:

- The nature of the adopted solution and also its corresponding method;
- The way of data acquisition, by collecting data in real-time;
- The user-friendly point of view, by offering the mobility advantage, thanks to the WBAN and IoT technologies; and
- In terms of accuracy, the Rand Index ( $RI$ ) in Eq. (10) measures the percentage of decisions that are accurate.

$$RI = \frac{TP + TN}{TP + TN + FP + FN} \quad (10)$$

Where true positive (TP), true negative (TN), false positive (FP) and false negative (FN) are used to calculate the accuracy.

The proposed solution provides a remarkable accuracy rate which is equal to 94.58%, improving upon the existing works cited in Table 2, namely, the method in [31], using three

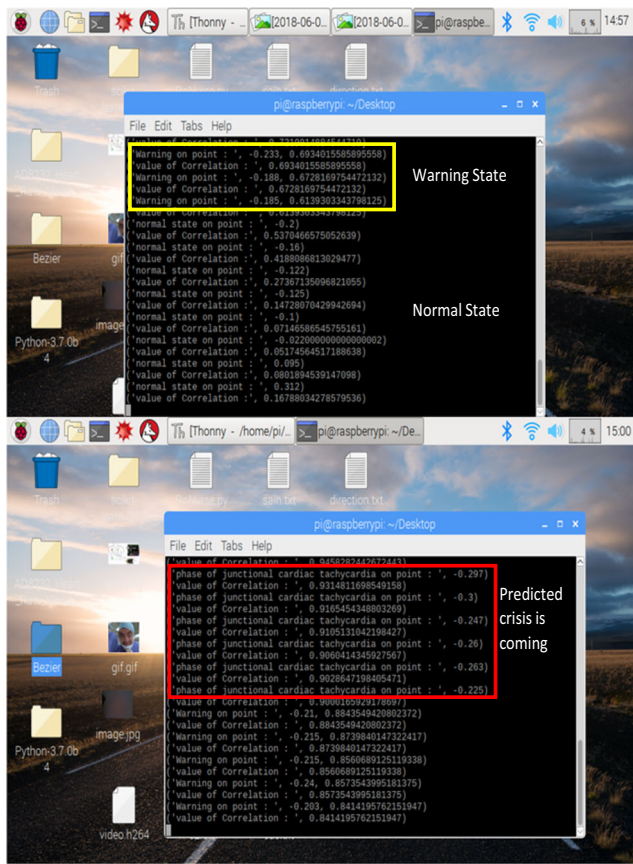


Fig. 14 Prediction results screen

different machine learning algorithms, where the maximal system accuracy is about 49.43%. Besides, methods in [32] and in [33], used respectively Weighted Associative Classifier (WAC) and Probabilistic Neural Network algorithms, the accuracies are respectively 81.51% and 92.10%. At their turn, Purushottam et al. in [34] used an efficient heart disease prediction system based on decision tree algorithm where they gain 86.7% of accuracy. Ultimately, In 2018 authors in [35]

used a gini index decision tree method with neural network classifiers, where they performed an accuracy of 78.3%. We also observe that not all the previously proposed systems support real-time mode in data collection and system response; additionally, the prediction time is undefined in the previous works.

### 7 Conclusion AND perspectives

This paper presents an innovative scheme, Smart Observatory of Involuntary Medical Seizures (SOIMS), which is applied to the prediction of a particular case of heart diseases, namely, junctional Tachycardia. Through this paper, we present the general architecture of SOIMS and both the technical and the practical contributions.

We have proposed our hierarchical real-time ML-based prediction scheme which consists of three algorithms, namely, Qualifying Linear Regression Algorithm (QuLRA), Selective Clustering Algorithm (SeCA) and Real-Time Clusters Correlation Algorithm (RT2CA), which can predict seizure 12.5 min before it occurs with an accuracy of 94.58%.

Then we presented the high-level SOIMS secured IoT architecture and our proposed IoT proxy that ensure the communication between the patient ECG node (CoAP/DTLS based) and the Hospital Information System (HTTP/TLS based).

Thereafter, we implement a realistic SOIMS prototype in order to evaluate the proposed prediction algorithm and the designed IoT proxy.

Thus, this paper has opened the door for a new generation of smart healthcare systems; combining WBAN low power, IoT connectivity and mobility, and machine learning, allowing prediction with high accuracy.

Nevertheless, during the design of SOIMS prototype, we faced many limitations. To measure the performance of our

Table 2 Works on heart diseases prediction

Paper	Disease	Technique Used	Method	Real-time Mode	Accuracy
R. Awang et al. (2008)	Heart Disease	Decision Trees, Naïve Bayes, Neural Network	Simulation	N/A	49.34%, 47.58%, 41.85%
J. Soni et al. (2011)	Heart Disease	Weighted Associative Classifier (WAC)	Simulation	N/A	81.51%
I.S.F Dessai et al. (2013)	Heart Disease	Probabilistic Neural Network	Simulation	N/A	92.10%
Purushottam et al. (2016)	Heart Disease	decision tree	Simulation	N/A	86.7%
K. Mathan et al. (2018)	Heart Disease	Gini index decision tree method with neural network classifiers	Simulation	N/A	78.3%
SOIMS	Heart Disease	IoT Linear regression K-means and Correlation	Implementation	YES	94.58%

algorithm, like the accuracy, we should wait till the detection of a seizure's outbreak. This time is not well known. Moreover, it is difficult to find volunteer patients to test our prototype. Finally, dataset history of volunteer patients having the JT disease doesn't exist.

Yet, the proposed solution still can be improved upon its current version, for instance, by reducing the size of the presented prototype to be more user-friendly to the patient, e.g. a wearable smartwatch.

As a perspective, we intend to extend SOIMS with the followings:

- Introducing other improved protocols like the IEEE 802.15.6 and IEEE 802.11ba protocol [31] which outperforms the Xbee protocol with higher data rate, lower power consumption, and longer battery lifetime.
- Comparing with other machine learning methods like SVM (Support Vector Machine), neural network and decision tree.
- Extending to other involuntary seizure such as epilepsy and asthma.

**Acknowledgments** The authors are grateful to Dr. L. CHAKIRI, the regional director of the Ministry of Health in the Marrakech-Safi region for her cooperation and her commitment for the success of this work.

## References

1. Tyagi S, Agarwal A, Maheshwari P (2016) A conceptual framework for iot-based healthcare system using cloud computing, International Conference - Cloud System and Big Data Engineering Confluence, 503–507
2. Xu B, Xu LD, Cai H, Xie C, Hu J, Bu F (2014) Ubiquitous data accessing method in iot-based information system for emergency medical services. *IEEE Transactions on Industrial Informatics* 10(2):1578–1586
3. Kiral-Kornek I, Roy S, Nurse E, Mashford B, Karoly P, Carroll T, Grayden D (2018) Epileptic seizure prediction using big data and deep learning: toward a mobile system. *EBioMedicine* 27:103–111
4. Hijazi S, Page A, Kantarci B, Soyata T (2016) Machine learning in cardiac health monitoring and decision support. *Computer* 49(11):38–48
5. Jiang C, Zhang H, Ren Y, Han Z, Chen K, Hanz L (2017) Machine learning paradigms for next-generation wireless networks. *IEEE Wireless Communications* 24(2):98–105
6. Meerja KA, Naidu PV, Kalva KSR (2019) Price Versus Performance of Big Data Analysis for Cloud Based Internet of Things Networks. *Mobile Networks and Applications* 24(3):1078–1094
7. <https://www.who.int> accessed on April 12, 2018
8. Perry JC, Garson A (1993) Complexities of junctional tachycardias. *J. Cardiovasc Electro-physiol* 4:224–238
9. Ait Zaoui CE, Latif A (2017) Internet of Things and Machine Learning Convergence: The E-healthcare Revolution, 2nd International Conference on Computing and Wireless Communication Systems, ICCWCS'17, Nov. 14–16, Larache, Morocco
10. Akay A, Hess H (2019) Deep learning: current and emerging applications in medicine and technology, *IEEE journal of biomedical and health informatics*
11. Kotsiantis SB, Zaharakis I, Pintelas P (2007) Supervised machine learning: a review of classification techniques. *Emerging artificial intelligence applications in computer engineering* 160:3–24
12. Montgomery DC, Peck EA, Vining GG (2012) Introduction to linear regression analysis, John Wiley & Sons
13. Hastie T, Tibshirani R, Friedman J (2009) Unsupervised learning. In: The elements of statistical learning, Springer, New York, NY pages 485–585
14. Huang Z (1998) Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data mining and knowledge discovery* 2(3):283–304
15. Hall MA (2000) Correlation-based feature selection of discrete and numeric class machine learning
16. Deng DJ, Lin YP, Yang X, Zhu J, Li YB, Chen KC (2017) IEEE 802.11ax: highly efficient WLANs for intelligent information infrastructure. *IEEE Communications Magazine*, volume 55(12):52–59
17. Deng DJ, Gan M, Guo YC, Lin JYYP, Lien SY, Chen KC (2019) IEEE 802.11ba: Low-Power Wake-Up Radio for Green IoT. *IEEE Communications Magazine* 57(7):106–112
18. Eddabbah M, Moussaoui M, Laaziz Y (2018) Performance evaluation of a smart remote patient monitoring system based heterogeneous WSN, *International Journal of Advanced Computer Science and Applications(IJACSA)*, volume 9(8)
19. Zaouiat CEA, Latif A (2019) Improvement of WBAN Performances by a Hybrid Model: Design and Evaluation of a Novel Inter-MAC Layer Exploited in Medical Applications. *International Journal of Internet Protocol Technology* 12(1):26–34
20. Lin CC, Deng DJ, Chen ZY, Chen KC (2016) Key Design of Driving Industry 4.0: joint energy-efficient deployment and scheduling in group-based industrial wireless sensor networks. *IEEE Communications Magazine* 54(10):46–52
21. Goldberger, L. Ary, Amaral, A. Lui, Glass, Leon, “PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals”, *Circulation*, 101(23), 215–220, (2000)
22. Dubois R (2004) Application des nouvelles méthodes d'apprentissage à la détection précoce d'anomalies cardiaques en électrocardiographie, Thèse de doctorat. Université Pierre et Marie Curie-Paris VI
23. Hoekema R, Uijen GJH, Oosterom AV (2001) Geometrical aspects of the interindividual variability of multilead ECG recordings. *IEEE Trans. Biomed. Eng.* 48(5):551–559
24. Kiranyaz S, Ince T, Pulkkinen J, Gabbouj M (2011) Personalized long-term ECG classification: a systematic approach. *Expert Systems with Applications* 38(4):3220–3226
25. <https://www.pchalliance.org/continua-design-guidelines> accessed on July 1, (2018)
26. <https://www.sparkfun.com> accessed on July 1, (2018)
27. <https://www.arduino.cc> accessed on July 1, (2018)
28. <https://www.raspberrypi.org/> accessed on July 1, (2018)
29. Deng DJ, Yen HC (2005) Quality-of-service provision system for multimedia transmission in IEEE 802.11 wireless LANs. *IEEE Journal on Selected Areas in Communications* 23(6):1240–1252
30. <https://www.seeedstudio.com> accessed on July 1, (2018)
31. Awang R, Palaniappan S (2008) Intelligent heart disease prediction system using data mining technique, *International Journal of Computer Science and Network Security*, volume 8(8)
32. Soni J, Ansari U, Sharma D, Soni S (2011) Predictive data mining for medical diagnosis: an overview of heart disease prediction. *International Journal of Computer Applications* 17(8):43–48

33. Isf. Dessai, “Intelligent heart disease prediction system using probabilistic neural network”, *International Journal on Advanced Computer Theory and Engineering*, volume 2(3), pages 2319–2526, (2013)
34. Saxena K, Sharma R (2016) Efficient heart disease prediction system. *Procedia Computer Science* 85:962–969
35. Mathan K, Kumar PM, Panchatcharam P, Manogaran G, Varadharajan R (2018) A novel Gini index decision tree data mining method with neural network classifiers for prediction of heart disease. *Design Automation for Embedded Systems* 22(3): 225–242

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