



**HAL**  
open science

## A Sensor-Based Data Analytics for Patient Monitoring in Connected Healthcare Applications

Hassan Harb, Abbass Nasser, Ali Mansour, Eduardo Motta Cruz, Isabel de La  
Torre Diez

► **To cite this version:**

Hassan Harb, Abbass Nasser, Ali Mansour, Eduardo Motta Cruz, Isabel de La Torre Diez. A Sensor-Based Data Analytics for Patient Monitoring in Connected Healthcare Applications. IEEE Sensors Journal, 2021, 21 (2), pp.974-984. 10.1109/JSEN.2020.2977352 . hal-02948489

**HAL Id: hal-02948489**

**<https://univ-rennes.hal.science/hal-02948489>**

Submitted on 25 Sep 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A Sensor-Based Data Analytics for Patient Monitoring in Connected Healthcare Applications

Hassan Harb, Ali Mansour, Abbass Nasser, Eduardo Motta Cruz and Isabel de la Torre Díez

**Abstract**—Nowadays, keeping a strong and good health is one of the main concern of the general public or governments. The Internet of Things (IoT) has been emerged as an efficient solution to build smart healthcare systems deployed either at hospitals or in-home. Such networks rely on biomedical sensors which are used in electronics-based medical equipment to remotely collect vital signs of patients (pressure, temperature, hart rate, oxygen saturation etc.). Generally, these biosensors are implemented on or inside the patient’s body and take three types of record data such as numerical, images and videos. However, the big data collected by various biomedical sensors along with the need of emergency detection, the limited sensor energies, and the prediction of the progress of patient situation are the major challenges for heath-based IoT applications. In order to overcome these challenges, we propose, in this paper, an efficient sensor-based data analytics for real-time patient monitoring and assessment to help both hospital and medical staff. The proposed mechanism consists in three phases: Emergency detection, adapting sensing frequency and real time prediction of patient situation. Through simulations on real health data, we show the effectiveness of our mechanism compared to other exiting techniques.

**Keywords**—Internet of Things (IoT), Connected heathcare, Emergency detection, Adaptive sampling rate, Data prediction, Energy efficiency.

## I. INTRODUCTION

**T**ODAY, our world is facing three major challenges regarding the public health: the increase number of aging or elderly persons, the population growth, and the increasing prevalence of severe diseases. This leads to complicate the missions of both hospitals and medical staff (e.g. nurses and physicians). Hence, the healthcare has taken, over the last decade, more attention from governments, individuals and companies which invest and spend more money in providing health care services and applications (the market is expected to reach USD \$10 trillion in 2022 [1]). Recently, the emergence of the IoT technology and wireless sensors plays a key role in the exponential increase of healthcare systems, e.g. known as *connected healthcare* [2], [3]. This technology can provide a low cost solution for patient monitoring and tracking at anytime and anywhere, e.g. at hospital or in-home, then allows a fast doctors access to patient data.

H. Harb and A. Nasser are with Lab-STICC, CNRS UMR 6285, ENSTA-Bretagne, Brest, France. They are also with the Computer Science department, American University of Culture and Education (AUCE), Beirut, Lebanon, e-mails: firstnamelastname@auce.edu.lb.

A. Mansour is with Lab-STICC, CNRS UMR 6285, ENSTA-Bretagne, Brest, France, e-mail: mansour@ieee.org.

E. Motta Cruz is with Bretagne Loire and Nantes Universities, UMR 6164, IETR Polytech,Nantes, France, e-mail: eduardo.mottacruz@univ-nantes.fr.

I. de la Torre Díez is with Department of Signal Theory and Communications, and Telematics Engineering, University of Valladolid, Paseo de Belén, 15,47011 Valladolid, Spain, e-mail: isator@tel.uva.es.

In connected healthcare, we divide sensors into three main categories depending on the type of collected data: numerical (for vital signs monitoring), images (like X-ray, dental imaging, etc.) or videos (for surgery operations like cardiology). Fig. 1 shows the most used types of biosensors in connected healthcare [4] which can be classified as follows:

- *Epidermal biosensors*: They aim to monitor the blood glucose (Fig. 1 (a)), the lactate, the uric acid and the urea (Fig. 1 (b)), the sweat alcohol (Fig. 1 (c)), and the glucose monitoring and the therapy applied to human forearm (Fig. 1 (d)).
- *Tear-based biosensors*: This type of sensors is a small spring-like electrochemical devices designed in a form of wearable contact lens tear applied to an artificial eye (Fig. 1 (e)(f)(g)).
- *Saliva-based biosensors*: They are dielectric detectors that are mostly mounted onto tooth and dedicated to monitor the glucose level (Fig. 1 (h)), the salivary uric acid concentrations (Fig. 1 (i)), and to detect floods and fluids during ingestion like sugars and alcohol salinity (Fig. 1 (j)).

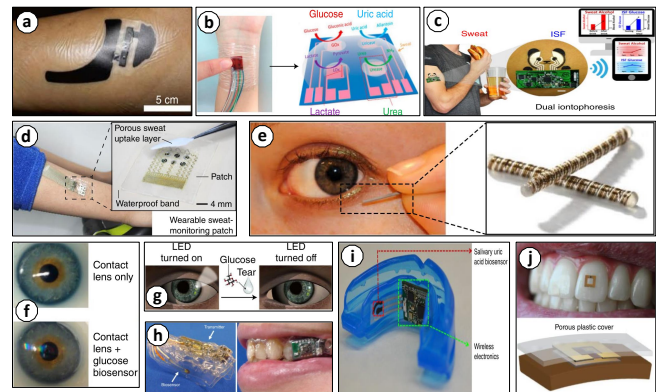


Fig. 1. Various types of biomedical sensors (regenerated from [4]).

Connected healthcare applications provide several challenges: First, reducing the energy consumption in the biosensors in order to save its limited power supply and ensure a long time patient monitoring [5]–[7]. Second, detecting emergency situations of the patient as fast as possible and quickly reported to the medical staff in order to take a suitable action [8]–[10]. Third, studying the progress of the patient behavior and predicting his future situation are important challenges in healthcare. By doing that, patient can be given the adequate treatment before entering a critical situation.

In order to overcome the above challenges, we propose an efficient data analytics mechanism for real-time connected

healthcare applications. The proposed mechanism uses data analytics techniques for prediction and reduction. In addition, our mechanism consists in three phases: Emergency detection, adapting sensing frequency and predicting patient situation progress. We conducted a set of simulations on real health data in order to show the efficiency of our mechanism while comparing the results to other existing techniques.

The remainder of this paper is organized as follows: Section II presents an overview on various data analytics techniques related to connected healthcare existing in the literature. In Sections III, IV and V, we detail the three phases proposed in our mechanism respectively. Section VI describes the implementation of our mechanism and explains the obtained results. Finally, Section VII concludes the paper and gives directions for future work.

## II. RELATED WORK

Connected healthcare is becoming an essential solution for hospitals that allows to record and analyze patient data, take right decisions and finally save lives and money. In the literature, we can find a lot of healthcare techniques ranging from diagnosis to treatment and prevention [11]–[13]. The authors of [14], [15] present an overview on data analytics algorithms proposed in the literature for healthcare-based IoT applications.

Some works in connected healthcare are focused on reducing big data collection using aggregation, compression and prediction methods [16]–[23]. In [16], the authors propose a Priority-based Compressed Data Aggregation (PCDA) technique in order to reduce the amount of health data transmitted. PCDA uses compressed sensing approach followed by a cryptographic hash algorithm to save information accuracy before sending data for diagnosis. The authors of [17] propose a two-level anomaly detection technique to remove false alarms and detect an emergency situation of a patient. The first level uses a game-theoretic approach in order to search the spatiotemporal correlation among observed features. In the second level, the Mahalanobis distance has been used to ensure a general view for multivariate analysis. In [18], a cloud-based connected healthcare system, called BigReduce, is proposed. The objective of BigReduce is to minimize the data processing cost at the base station according to two schemes applied locally at the IoT sensors: reduction and decision schemes. In [24], the authors propose a classification technique based on a defined threshold where sensor's readings are classified into three types: urgent (above threshold), semiurgent (close to threshold) and nonurgent (less than threshold). Furthermore, the authors introduce a routing protocol for a medical sensor that enables transmitting packets during gateway failure. Finally, the authors of [19] aim to extend the lifetime of the wireless body network by selecting the optimized paths between the nodes and the remote healthcare sink. The proposed technique is based on a cluster routing protocol combined with Q-learning approach.

Recently, the authors of [5], [8]–[10] open a new trend in connected healthcare by proposing several frameworks for a real-time patient monitoring and assessment. In [8], a

framework for a stress detection and evaluation has been proposed. The framework works by detecting first stress signals according to skin conductance parameter, then the stress level is evaluated through fuzzy inference system based on patient vital signs, particularly heart rate, respiration rate, and average blood pressure. In a recent work [9], the authors propose a data management framework for data collection and decision making in connected healthcare. The framework relies on three algorithms: first, an emergency detection algorithm aims to send critical records directly to the coordinator; second, an adaptive sampling rate algorithm based on ANOVA and Fisher test in order to allow each sensor to adapt its sampling frequency to the variation of the patient situation; third, a data fusion and decision making model is proposed at the coordinator and it is based on a decision matrix and the fuzzy set theory. Although its great advantages for the patient monitoring and assessment, the proposed framework suffers from several disadvantages: 1) sensors only send critical records to the coordinator thus medical staff cannot avoid patient enters a dangerous situation; 2) in case of low critical patient none of the data will be archived in the hospital thus, revising patient archive to check patient progress by doctors is not possible; 3) ANOVA and Fisher test are only sensitive to a significant variation in data, thus adapting sensor rate for patients with medium criticality is not possible.

Unfortunately, although the proposed mechanisms and techniques carry many advantages and tackle several problems in connected healthcare, but they mainly suffer from several drawbacks: 1) complexity which is not suitable to limited resources of biosensors; 2) they propose separately data aggregation, prediction or emergency detection; 3) they may solve at most two problems from the above mentioned ones in connected healthcare. However, our proposed mechanism consists of a sequence of phases aiming to tackle problems ranging from emergency detection to patient situation prediction.

## III. PATIENT EMERGENCY DETECTION

Monitoring patient and rapidly acting after a critical situation constitute major tasks of medical staff, otherwise patient may be in critical situation leading to his death. In the first phase of our mechanism, we propose an emergency detection algorithm allowing each biosensor to inform medical staff about any abnormal situation. To verify abnormal situations, let us first define the periodic patient monitoring model and the Early Warning Score (EWS) guide.

### A. Periodic Patient Monitoring

In health applications, data collection is usually done through small biosensors which are implemented on the patient body in order to collect its vital signs (such as Heart Rate (HR), Systolic Blood Pressure (SBP), Respiration Rate (RR), Oxygen Saturations (OS), Body Temperature (BT), etc). Assume we have a set  $P$  of  $\gamma$  patients as follows:  $P = [P_1, P_2, \dots, P_\gamma]$  where each patient  $p \in P$  is assigned various types of biosensors to collect the set of vital signs, e.g.  $V = [HR, SBP, RR, OS, BT, \dots]$ , of the patient. For analysis purposes, we assumed that each biosensor  $B_v^p$ , assigned

to the patient  $p$ , periodically monitors one vital sign,  $v \in V$ , then it sends the collected records toward the sink. Thus, each  $B_v^p$  collects a vector  ${}_tR_v^p$  of  $\tau$  records during a period time  $t$  as follows:  ${}_tR_v^p = [r_1, r_2, \dots, r_\tau]$ .

### B. Early Warning Score (EWS) Guide

EWS is a guide based on vital signs, e.g.  $V$ , and used by medical staff within hospital in order to track the criticality level of a patient. For each vital sign  $v \in V$ , the collected record  $r_i \in {}_tR_v^p$  is compared to a normal range in order to calculate a score  $s_i$  between 0 and 3; 0 means normal record where other values indicate abnormal situation with increasing of severity when the score increases. Hence, a set of record scores  ${}_tS_v^p = [s_1, s_2, \dots, s_\tau]$  is calculated for each  ${}_tR_v^p$ . Meanwhile, EWS is used to determine the appropriate response of the medical staff where low score yields to minimize the frequency of patient's monitoring while high score requires an immediate response from the emergency team. Fig. 2 shows one of the most used EWS guides developed in UK and widely distributed, called National EWS (NEWS) [25].

| PHYSIOLOGICAL PARAMETERS | 3     | 2        | 1           | 0           | 1           | 2         | 3          |
|--------------------------|-------|----------|-------------|-------------|-------------|-----------|------------|
| Respiration Rate         | ≤8    |          | 9 - 11      | 12 - 20     |             | 21 - 24   | ≥25        |
| Oxygen Saturations       | ≤91   | 92 - 93  | 94 - 95     | ≥96         |             |           |            |
| Any Supplemental Oxygen  |       | Yes      |             | No          |             |           |            |
| Temperature              | ≤35.0 |          | 35.1 - 36.0 | 36.1 - 38.0 | 38.1 - 39.0 | ≥39.1     |            |
| Systolic BP              | ≤90   | 91 - 100 | 101 - 110   | 111 - 219   |             |           | ≥220       |
| Heart Rate               | ≤40   |          | 41 - 50     | 51 - 90     | 91 - 110    | 111 - 130 | ≥131       |
| Level of Consciousness   |       |          |             | A           |             |           | V, P, or U |

Fig. 2. National Early Warning Score (NEWS).

### C. Emergency Detection Algorithm

As mentioned previously, biosensors collect vital signs in a periodic way. However, this data collection model produces a huge amount of data which leads to quickly deplete the available energy of sensors as well as to complicate the data analysis. Obviously, a native intuition is to enforce each biosensor to send only critical records, e.g. those have scores  $> 0$ , during a period to the medical staff, while ignoring the transmission of normal records. Although this approach reduces the data transmission but prevent the monitoring of patient's progress. Hence, in order to overcome this problem, we propose to find relationships among records collected at each period before sending them toward the sink. In this paper, we are interested in the Simple Linear Regression (SLR) method to find data relationships thanks to two reasons: first, it is less complex and suitable to limited biosensor resources; second, health data are naturally linearly related.

Mathematically, given two vectors of data with equal size,  $X = [x_1, x_2, \dots, x_\tau]$  and  $Y = [y_1, y_2, \dots, y_\tau]$ , then the SLR equation can be calculated according to equation 1. In our case,  $X$  represents the records collected during a period time  $t$  while  $Y$  indicates their order numbers.

$$y = a + bx, \quad (1)$$

where

$$a = \frac{\sum_{i=1}^{\tau} y_i \times \sum_{i=1}^{\tau} x_i^2 - \sum_{i=1}^{\tau} x_i \times \sum_{i=1}^{\tau} x_i y_i}{\tau \times \sum_{i=1}^{\tau} x_i^2 - (\sum_{i=1}^{\tau} x_i)^2}$$

and

$$b = \frac{\tau \times \sum_{i=1}^{\tau} x_i y_i - \sum_{i=1}^{\tau} x_i \times \sum_{i=1}^{\tau} y_i}{\tau \times \sum_{i=1}^{\tau} x_i^2 - (\sum_{i=1}^{\tau} y_i)^2}$$

Thus, based on the EWS guide and the SLR equation, Algorithm 1 describes the emergency detection process which is applied at each biosensor itself. The algorithm takes as input the vector of records collected by a biosensor during a period time  $t$ . Then, it calculates the score for each record and it only sends it to the medical staff in case of critical record, e.g. having score  $> 0$  (lines 1-6). At the end of the period, the biosensor calculates the coefficients ( $a$  and  $b$ ) of SLR equation based on the whole records in  ${}_tR_v^p$ , and sends them to the sink for archive purpose (lines 7-8).

---

#### Algorithm 1 Emergency Detection Algorithm.

**Require:** A patient:  $p$ , A biosensor:  $B_v^p$ , A period time:  $t$ , Records collected during  $t$ :  ${}_tR_v^p = [r_1, r_2, \dots, r_\tau]$ .

**Ensure:** A set of critical records.

- 1: **for** each record  $r_i \in {}_tR_v^p$  **do**
  - 2:     calculate score  $s_i$  of  $r_i$  according to EWS
  - 3:     **if**  $s_i > 0$  **then**
  - 4:         send  $r_i$
  - 5:     **end if**
  - 6: **end for**
  - 7: calculate  $a$  and  $b$  for  ${}_tR_v^p$  based on equation 1
  - 8: send  $a$  and  $b$
- 

In order to clarify the process of Algorithm 1, let us consider an illustrative example. Suppose a HR biosensor collects a vector of 10 readings during a period  $t$ :  ${}_tR_{HR}^p = [88, 90, 91, 91, 90, 90, 93, 100, 112, 135]$ . Consequently, based on equation 1, the biosensor first calculates the SLR coefficients  $a = 77$  and  $b = 3.8$  then, according to EWS, it computes the vector of scores corresponding to records in  ${}_tR_{HR}^p$  as follows:  ${}_tS_{HR}^p = [0, 0, 1, 1, 0, 0, 1, 1, 2, 3]$ . Therefore, the biosensor sends the SLR coefficients followed by the critical records, e.g.  $[77, 3.8, 91, 91, 93, 100, 112, 135]$ , to the medical staff.

#### IV. ADAPTING SENSING FREQUENCY (ASF)

In health applications, data collected by each biosensor are highly correlated to the situation of the monitored patient; more the situation is stable more the data redundancy is. Hence, we categorize the situation of a patient into 3 levels:

- *Low risk*: in which patient is in a normal situation where all the collected records are in the normal range. This situation requires minimum patient observations from the medical staff.

- *Medium risk*: it indicates that the patient is in unstable situation where the collected records vary between normal and critical scores. In this level, the patient needs periodic attention from the medical staff.
- *High risk*: In this level, all the collected records are critical and the patient enters a dangerous situation that requires a continuous monitoring.

Obviously, low and high criticality situations yield to a significant redundancy level among the collected data. This leads to consume the biosensor energy and overloading the medical staff by sending redundant information. To reduce data collection, we adapt the sensing frequency of each biosensor. The basic idea of our approach is to dynamically adapt the sensing frequency of each biosensor to fit with the dynamic variation of the monitored vital sign.

Admitting that records collected in consecutive periods are generally similar, our algorithm works on rounds where each round equals to  $\beta$  periods. Subsequently, we aim to determine the stability level of a patient during a round then to dynamically adapt the sensing frequency of a biosensor in the next round, accordingly to reach the stability level; more the patient situation is unstable more the sensing frequency is increased and vice versa. Derived from  ${}_tR_v^p$ , we define the subset  ${}_t^sR_v^p$  which only contains the records with a score  $s$  in  ${}_tR_v^p$ , where  $s \in [0, 3]$ . Accordingly,  ${}_t^sS_v^p$  is the subset with scores  $s$  of  ${}_tS_v^p$  for the set of records  ${}_tR_v^p$ . Thus,  ${}_tR_v^p = \cup_{i=0}^3 {}_t^iR_v^p$  while  ${}_tS_v^p = \cup_{i=0}^3 {}_t^iS_v^p$ . Let  $|X|$  be the norm zero, e.g. the number of elements of a set  $X$ . For the sake of simplicity, let assume a round consists of 2 periods  $t$  and  $w$  with equal size  $\tau$  where two record vectors  ${}_tR_v^p$  and  ${}_wR_v^p$  are collected in  $t$  and  $w$  respectively. Then, the stability level, noted as *stab*, of the vital sign  $v$  of a patient  $p$  can be calculated as the overlap between the number of similar records (having the same scores) in  ${}_tR_v^p$  and  ${}_wR_v^p$  as shown in the following equation:

$$stab({}_tR_v^p, {}_wR_v^p) = \frac{\sum_{k=0}^3 \min(|{}_t^kR_v^p|, |{}_w^kR_v^p|)}{\min(|{}_tR_v^p|, |{}_wR_v^p|)} \times 100 \quad (2)$$

where  $|{}_t^kR_v^p|$  means the number of recods with score  $k$  in  ${}_tR_v^p$ . Therefore, *stab* will range between 0 and 100 while 0 means that patient is in a fully stable situation and 100 indicates a severe unstable situation of the patient.

In order to calculate its new sensing frequency (indicated as  $\tau'$ ) for the next round based on the obtained stability level, the biosensor uses the following equation:

$$\tau' = \frac{100 - stab({}_tR_v^p, {}_wR_v^p)}{100} \times \tau \quad (3)$$

The idea behind the above equation is that the biosensor decreases its sensing frequency when a low/high criticality situation is detected and increases it when a medium criticality is noticed. This allows the biosensor to avoid collecting records with similar scores either in low or critical situation and to collect more records when patient situation is unstable.

Finally, in order to make our algorithm more efficient, we propose to adjust a minimum sensing frequency, e.g.  $\tau'_{min} \in [0, 100]$ , when calculating the new sampling one.

The optimal value of  $\tau'_{min}$  is determined by the medical staff based on the criticality level of the patient situation and the observed vital sign. Therefore, if the new sensing frequency of a biosensor ( $\tau'$ ) is less than the minimum assigned ( $\tau'_{min}$ ), then the biosensor adapts its new sensing frequency to  $\tau'_{min}$  and not to  $\tau'$  in order to save the integrity of the monitored condition.

Algorithm 2 describes the adaptive sensing frequency process which is applied at each biosensor  $B_v^p$ , at the end of each round. For each period in the round, the algorithm calculates the scores of records collected during the period while counting the number of records having same scores (lines 1-6). Then, the overlap between the scores of periods within the round is calculated according to equation (2) (line 7). After that, the biosensor computes its new sampling and adapts its sensing frequency to the new one only if it is greater than the minimum frequency assigned ( $\tau'_{min}$ ) or to the  $\tau'_{min}$  otherwise (lines 8-10).

---

**Algorithm 2** Adapting Sensing Frequency Algorithm.

---

**Require:** A patient  $p$ , A biosensor:  $B_v^p$ , A round of  $\beta$  periods, Initial period size:  $\tau$ , A minimum sensing frequency:  $\tau'_{min}$ .

**Ensure:** New sensing frequency:  $\tau'$ .

- 1: **for** each period  $i$  in the round, where  $i \in [1, \beta]$  **do**
  - 2:   calculate the vector of scores  ${}_iS_v^p$  for  ${}_iR_v^p$  according to EWS
  - 3:   **for**  $k = 0$  to 3 **do**
  - 4:     calculate  $|{}_i^kR_v^p|$
  - 5:   **end for**
  - 6: **end for**
  - 7: calculate *stab*( ${}_1R_v^p, {}_2R_v^p, \dots, {}_\beta R_v^p$ )
  - 8: calculate  $\tau'$  based on equation (3)
  - 9:  $\tau' = \max(\tau', \tau'_{min})$
  - 10: **return**  $\tau'$
- 

Let us consider an illustrative example for Algorithm 2. Suppose a round consists of 2 periods  $t$  and  $w$  of size 10 records, e.g.  $\tau = 10$ . Then, assume the calculated scores for both periods are  ${}_tS_v^p = [0, 0, 0, 1, 1, 1, 1, 2, 2, 3]$  and  ${}_wS_v^p = [0, 0, 1, 1, 2, 2, 2, 2, 3, 3]$  respectively. Therefore, the similar scores in  ${}_tS_v^p$  can be aggregated as  ${}_tS_v^p = [(0, 3)(1, 4)(2, 2)(3, 1)]$  and for  ${}_wS_v^p$  as  ${}_wS_v^p = [(0, 2)(1, 2)(2, 4)(3, 2)]$ . After that,  $stab({}_tR_v^p, {}_wR_v^p) = \frac{\min(3, 2) + \min(4, 2) + \min(2, 4) + \min(1, 2)}{\min(3, 2) + \min(4, 2) + \min(2, 4) + \min(1, 2)} \times 100 = 70$ . Consequently,  $\tau' = \frac{100 - 70}{100} \times 10 = 3$ . So, if we take a minimum frequency  $\tau'_{min} = 2$  then, the biosensor adapts its sensing frequency for the next round to 3 while if we take  $\tau'_{min} = 4$  then the new sampling is adapted to 4.

## V. PATIENT SITUATION PROGRESS PREDICTION

As mentioned before, the patient may enter in a critical situation leading to his death at any time. Hence, predicting the progress situation of a patient is very important for both doctors and nurses; from one hand, doctors can avoid a critical situation using pretreatment and, on the other hand, it helps nurses to estimate the needed attention for the patient.

Therefore, after generating data for each biosensor, we propose a prediction algorithm in order to allow the medical staff to be updated on the situation of a patient and takes the appropriate action. Indeed, several prediction techniques are proposed by researchers like linear and logistic regression, decision tree, random forecast, neural network, etc. These techniques are introduced in various domains such as stock production, scientific studies, sport monitoring, financial sector, psychology, etc. In this paper, we are focusing on the Long Short-Term Memory (LSTM) which is a famous prediction method and widely adapted in various applications. Let first recall the LSTM prediction method, then we adapt it to the connected healthcare case.

### A. Recall of LSTM

Generally, the Long Short-Term Memory (LSTM) [26] is an artificial and enhanced version of the recurrent neural network (RNN) model. It is well-suited to classify, process and make predictions based on time series data. Compared to RNN, LSTM enjoys several advantages: first, remembering data for a *long* time period even with huge size; second, classifying data into *important* or not; third, supporting *multiple* layers to process data instead of single one.

Formally, the architecture of LSTM consists of a network of cells and has three gates for the flow of information: input, output and forget gates (Fig. 3). Each cell has three inputs: Previous cell state ( $u_{t-1}$ ), previous hidden state ( $h_{t-1}$ ) and the input at the current time ( $x_t$ ). Furthermore, each cell contains four neural networks where each of them has three parameters: first, the number of blocks ( $\mathcal{O}$ ) indicating the capacity of the neuron; second, the number of time steps ( $\mathcal{E}$ ) which represents the size of the input vector ( $x_t$ ) used to predict the next time step ( $y_t$ ); third, the number of features ( $\mathcal{F}$ ) which indicates the dimensions feed at each time step.

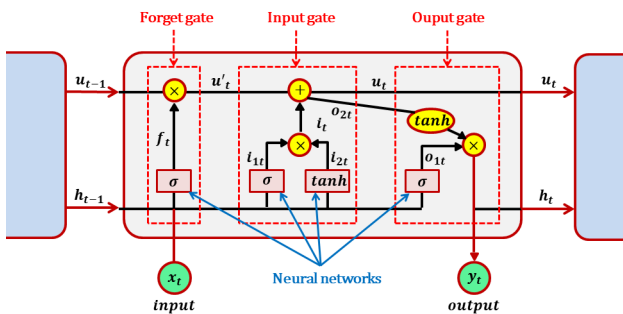


Fig. 3. LSTM cell architecture.

### B. Applying LSTM in our Mechanism

As biosensors collect data in a sequential manner; Sequencing prediction shows its efficiency and performance regarding the analysis of the health data over time. This makes connected healthcare a suitable domains to apply LSTM model. Indeed, like adapting sensing frequency, our prediction algorithm proposed for tracking a patient situation progress works in rounds. Assume each round equals to  $\beta$  periods, then we

divide periods within each round into two parts:  $\alpha$  periods for training data sets and the remaining  $\beta - \alpha$  periods for data testing and prediction. Thus, after regenerating data for  $\alpha$  periods according to the SLR, we run the patient situation algorithm in order to predict data for the remaining periods and store them in the sink node. Thus, medical staff can access the data and predict the patient severity in the next periods.

Formally, our patient progress algorithm based on the LSTM model can be applied according to the following steps:

- 1) *Receiving training data*: At each round, the algorithm generates record vectors of the first  $\alpha$  periods according to SLR equation, e.g.  $R_v^p = [{}_1R_v^p, {}_2R_v^p, \dots, {}_\alpha R_v^p]$ , for each biosensor  $B_v^p$  which acts as the training data for this biosensor.
- 2) *Data normalization*: Usually, the normalization is used to scale up data with different types into the same range  $[0, 1]$ . In our simulations we used the MinMaxScaler algorithm in order to normalize data received from the biosensors as a preprocessing step before doing any prediction process. For each  ${}_tR_v^p \in R_v^p$ , a Min-Max scaling is typically done according to the following equation:

$$r_i'^n = \frac{r_i^t - ({}_tR_v^p)_{min}}{({}_tR_v^p)_{max} - ({}_tR_v^p)_{min}}. \quad (4)$$

where  $r_i'^n$  is the normalized value for the record  $r_i^t \in {}_tR_v^p$ ;  $({}_tR_v^p)_{min}$  and  $({}_tR_v^p)_{max}$  are the minimum and maximum records in  ${}_tR_v^p$  respectively.

- 3) *Determining neural network parameters*: It indicates the selection of the values of the number of blocks ( $\mathcal{O}$ ), the number of time steps ( $\mathcal{E}$ ) and the number of features ( $\mathcal{F}$ ). Indeed, in order to evaluate their impacts, the values of  $\mathcal{O}$  and  $\mathcal{E}$  are varied in our simulations where  $\mathcal{F}$  is fixed to 1, because we predict a recorded data for each biosensor separately.
- 4) *Training the LSTM*: The objective of this step is to allow the LSTM to study the variation between the training data in order to predict the future data. This step is usually based on two concepts: loss function and optimizer. The loss function allows to calculate the variation between the training data where the optimizer is an iterative method that randomly uses a sample of data in order to optimize the error in the loss function. In this paper, we used the Mean Square Error (MSR) (see equation (5)) as a loss function accompanied with the Adam optimizer [27].

$$MSE(R_v^p) = \frac{\sum_{q=1}^{\alpha} \sum_{k=1}^{\tau} (r_k - ({}_qR_v^p)_{mean})}{\alpha \times \tau}. \quad (5)$$

where  $r_k \in {}_qR_v^p$  and  $({}_qR_v^p)_{mean}$  is the mean value of  ${}_qR_v^p$ .

- 5) *Predicting data*: It represents the final step of LSTM which allows the medical staff to predict the data of the remaining  $\beta - \alpha$  periods in the round. Subsequently, when receiving the data of the first  $\alpha$  periods the algorithm uses the last  $\mathcal{E}$  values of the periods in order to predict all values of the remaining periods.

## VI. PERFORMANCE EVALUATION

In order to evaluate the performance of our mechanism, we used real health data collected from MultipleIntelligent Monitoring in Intensive Care (MIMIC) database of PhysioNet [28]. MIMIC contains data for about 72 patients where recorded on vital signs including Heart Rate (HR), Systolic Blood Pressure (SBP), Respiration Rate (RR) and Oxygen Saturation (OS). Every second, the biosensor collects new reading for each vital sign then it sends toward the coordinator for archive purpose. In our simulation, we used a file that includes a log of about 100000 readings for each patient. We assume that each biosensor reads the data from its corresponding file for a period of time, then it sends them toward a coordinator placed at 50 meters after applying our mechanism. We implemented the algorithms used in our mechanism based on Java simulator and we compared the obtained results to those obtained in the technique proposed in [9], e.g. modified local emergency detection (MLED).

Table I summarizes the parameters used in our simulation with their tested values.

| Parameter                 | Symbol        | Values                  |
|---------------------------|---------------|-------------------------|
| number of patients        | $\gamma$      | 72                      |
| number of features        | $\mathcal{F}$ | [HR, SBP, RR, OS]       |
| period size               | $\tau$        | 3600 records            |
| round size                | $\beta$       | 2, 3, 5 periods         |
| minimum sensing frequency | $\tau'_{min}$ | 10%, 20%, 30% of $\tau$ |
| number of blocks          | $\mathcal{O}$ | 5                       |
| number of time steps      | $\mathcal{E}$ | 600, 3600 records       |
| training data size        | $\alpha$      | 1 period                |

TABLE I  
SIMULATION ENVIRONMENT.

### A. Sensing Frequency Adaptation Study

In Fig. 4, we show how each biosensor adapts its sensing frequency after applying our algorithm (see Algorithm 2, referred as ASF in upper graph in each subfigure) and that proposed in [9] (referred as MLED in lower graph in each subfigure) for various types of patient situations. We fixed the period size ( $\tau$ ) to 3600 records and the round size  $\beta$  to 2 periods. Indeed, three observations are eminent according to the obtained results: first, both algorithms allow each biosensor to dynamically adapt its sensing frequency after each round. However, the sensing frequency using ASF is more adapted than that using MLED. This is because, ASF relies on adapting sensor frequency according to the stability of patient where MLED only adapts sensor frequency if the variation among collected records meets the Fisher threshold. Second, sensing adaptation is almost fix for biosensors of the same patient in both algorithms. This confirms the correlation between vital signs of the same patient. Third, sensing frequency of each biosensor is more adapted with low and high patient situations compared to patient with medium criticality. This is because the variation among records collected in low and high situations is less significant than that with of medium patients situation.

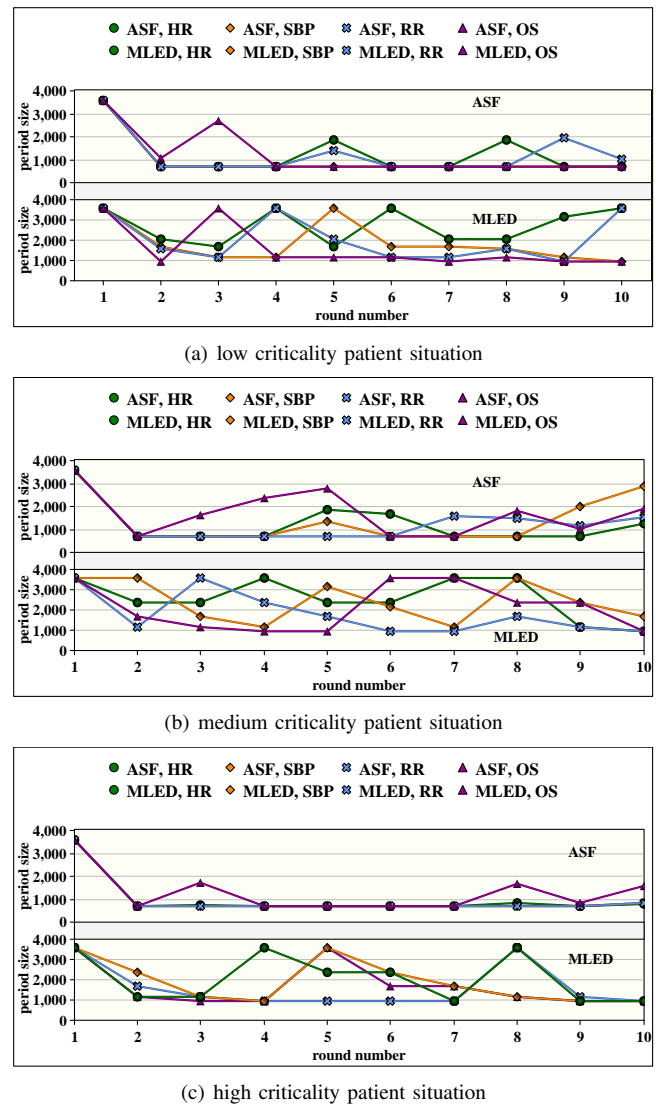


Fig. 4. Variation of sensing sensor frequency during rounds,  $\tau = 3600$ ,  $\beta = 2$ .

### B. Data Transmission Study from Biosensors

Fig. 5 shows the percentage of data sent from each biosensor to the sink node after applying the emergency detection and adapting sensing frequency algorithms (referred as EDASF). The simulations are made with respect to three parameters: the period size ( $\tau$ ), the round size ( $\beta$ ) and the minimum sensing frequency threshold ( $\tau'_{min} = 20\%$ ). The obtained results show that the data transmission is highly reduced using our algorithms compared to MLED algorithm. This is due to the sensing frequency which is more adapted using our algorithms (see results of Fig. 4). Indeed, we observe that HR, SBP, RR and OS biosensors reduce up to 94%, 52%, 39% and 40% of its data transmission respectively compared to MLED. In addition, using our algorithm, we can show that:

- The percentage of data transmission is almost fix for each biosensor when increasing the period size (Fig. 5(a)). This is because, the stability of patient situation will not highly changed within 1 to 2 hours, 3600 and 7200 records

respectively.

- By varying the round size from 2 to 3 periods, each biosensor increases its data transmission to the master node (Fig. 5(b)). This is due to the variation among records which will increase when the round size increases.
- The data transmission from each biosensor is slowly increased when increasing the minimum sensing frequency threshold (Fig. 5(c)).

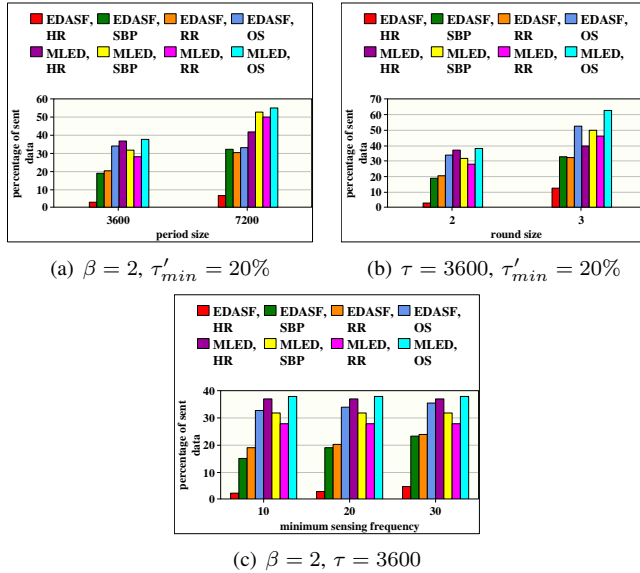


Fig. 5. Percentage of records sent from biosensors for a medium patient criticality level.

### C. Comparison Study: Raw Data vs Adaptive Data vs Sent Data vs Regenerated Data

Fig. 6 shows an illustrative example for comparison between raw data, adaptive data after applying ASF algorithm, sent data according to Algorithm 1 and regenerated data after applying the SLR equation at the sink. In each subfigure (a, b and c), the upper graph shows the difference between raw and adaptive data; while, in the lower graph, we show the comparison between raw, sent and regenerated data. The comparison is made according to three patient situations: low, medium and high risks. The obtained results allow several observations: 1) ASF algorithm eliminates redundancy among collected records by reducing data collection in all cases (curves blue and green in Figs. 6(a) to 6(c)). Also, we confirm the behavior of ASF by collecting more data in the medium critical situation due to the instability of the monitored vital sign. Second, the emergency detection algorithm allows to reduce the data transmission from each biosensor to the sink by sending only the critical records with SLR coefficients. Thus, by comparing raw and sent curves, we show that in low risk, only the SLR coefficients are sent to the sink without any records (Fig. 6(a)) while, in medium patient situation, a few of records are sent (Fig. 6(b)) and, finally, in high risk situation, all the records are sent to the medical staff (Fig. 6(c)). Third, by comparing the raw and the regenerated data, we show that data regenerated at the sink

allow to save the situation of each patient before storing data in the sink.

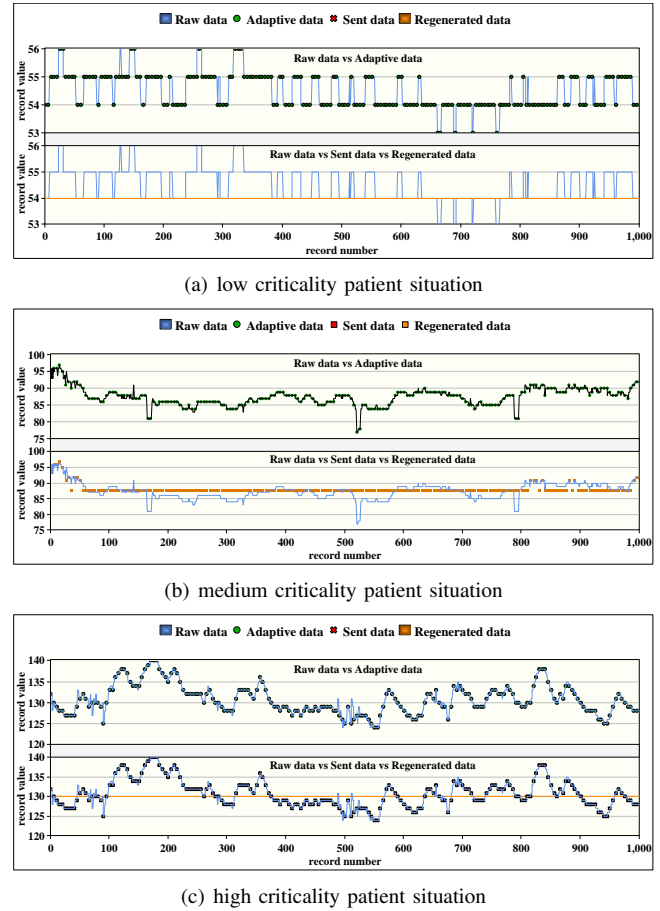


Fig. 6. Comparison between raw vs adaptive vs sent vs regenerated data for a portion of 1000 collected HR records.

### D. Energy Consumption Study in Biosensor

In this section, we study the energy consumed in each biosensor depending on the patient situation (Fig. 7). Since the energy consumption is highly related to the amount of data transmission, thus our algorithms will highly outperform MLED in terms of saving biosensor energies. However, in order to evaluate our algorithms in real world scenarios, we compared our algorithms to the naïve approach where all collected data are sent to the medical staff. We fixed the initial energy to 13000 energy unit and we assumed that each collected record consumes 0.1 unit and each sent record requires 0.7 unit. The obtained results show that our algorithms can prolong the lifetimes of HR, SBP, RR and OS biosensors up to 6, 6, 5 and 4 times respectively compared to the naïve approach. We can also observe that the best biosensor, in terms of its lifetime, is not the same in all cases. For instance, the lifetime of HR biosensor is largely extended with low and medium patient situations while SBP acts as the best biosensor with high criticality situation. This is because, the variation among collected records is not the same for all biosensors of the same patient. Finally, we notice that the lifetime of



biosensors in medium risk situation is almost less extended compared to low and high risk situations which confirms the behavior of our algorithms.

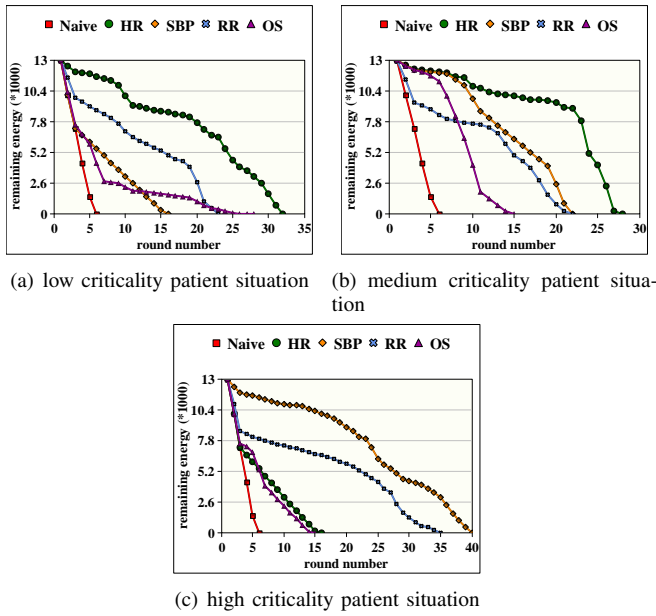


Fig. 7. Energy consumption in each biosensor during rounds,  $\tau = 3600$ ,  $\beta = 2$ ,  $\tau'_{min} = 20\%$ .

### E. Patient Situation Progress Study using LSTM

In this section, we show the efficiency of LSTM algorithm in terms of predicting the situation progress of a patient. In our simulations, we take a portion of regenerated data, stored in the sink, of a patient in medium criticality situation collected during 10 rounds. Each round consists of 2 periods of 3600 records (1 hour) of each one (e.g.  $\beta = 2$ ). Then, in order to make the scenario more realistic, we divided the rounds into two equal parts of 5 rounds for each; the first round in each part acts as the training data (e.g.  $\alpha = 1$ ) while the 4 remaining rounds in the part represent the tested data that should be predicted by the medical staff for the patient. Fig. 8 shows the variation between the raw data (blue curve) and the predicted data (orange curve) obtained after applying LSTM over the training data (green curve). We fixed the number of steps  $\mathcal{E}$  to 600 records which means that the medical staff predicts the future situation of a patient based on the previous records collected during the last 10 minutes. The obtained results show a good accuracy of LSTM regarding various monitored vital signs of a patient. Thus, the medical staff will have an accurate information about the patient situation progress for the next 8 hours (e.g. 4 rounds) thus, they can prevent patient to enter in a critical situation and the nurses can determine the appropriate observations needed for that patient. We can also observe that the accuracy of the predicted data is conserved for HR and RR biosensors more than for SBP and OS. This is due to the small variation of the records collected for HR and RR compared to that among records for SBP and OS.

Fig. 9 shows the accuracy of LSTM with respect to the number of steps ( $\mathcal{E}$ ) between 600 (e.g. 10 minutes) and 3600

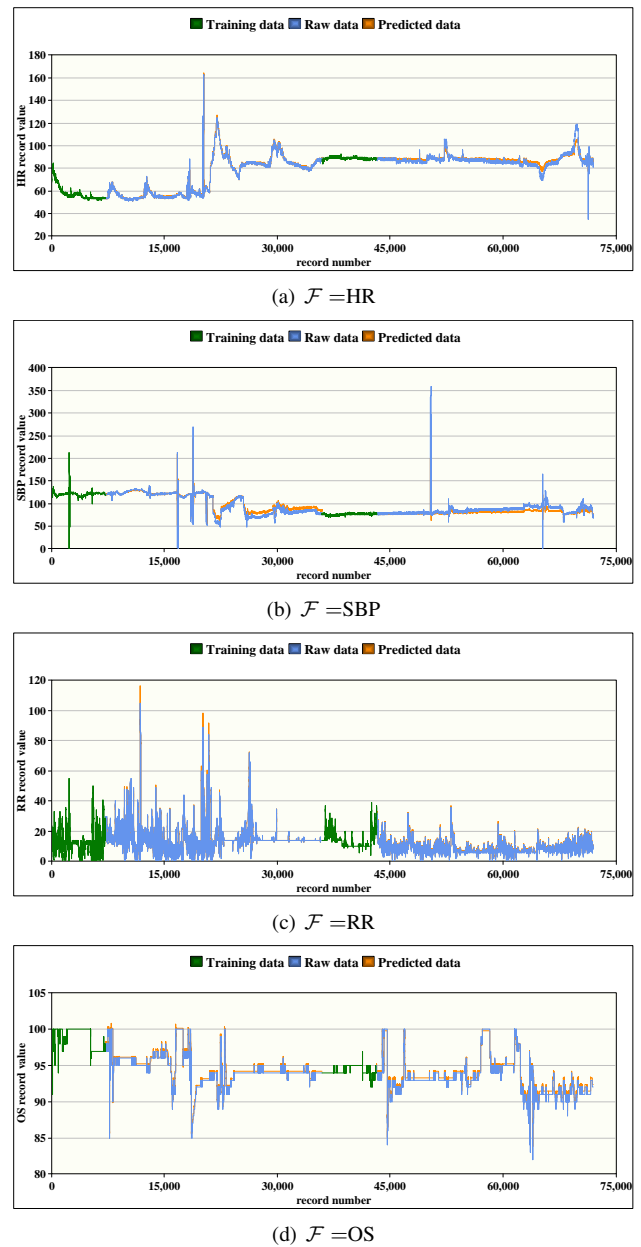


Fig. 8. Variation between raw and predicted data generated using LSTM method for a medium patient criticality situation,  $\beta = 2$ ,  $\alpha = 1$  round,  $\mathcal{E} = 600$ .

(e.g. 1 hour) records. Obviously, the decreasing value of  $\mathcal{E}$  will lead to increase the accuracy of LSTM. This is because a future record is mostly correlated with its nearest previous records than the far ones. Therefore, the obtained results show more accurate level with a number of steps equals to 600 records.

### F. Further Discussions

In this section, we give further consideration to our proposed mechanism while summarizing the obtained results of ASF and MLED in terms of parameters shown in Table II.

From the sensing frequency adaptation point of view, both mechanisms adapt dynamically the sensing frequency of the

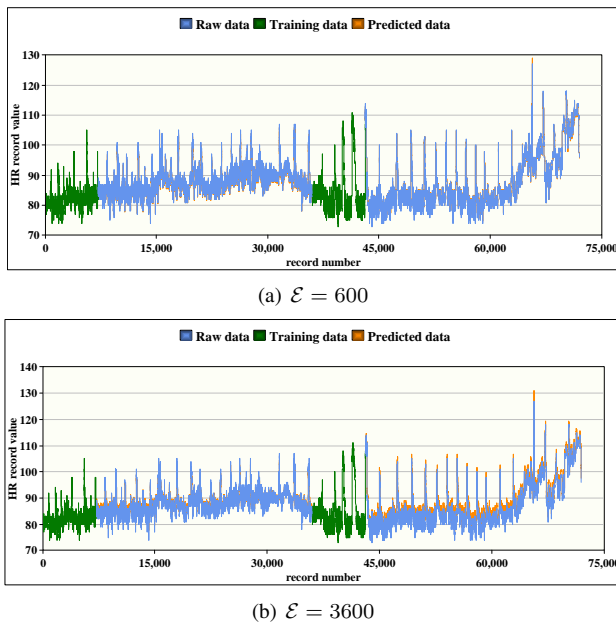


Fig. 9. Variation between raw and predicted data generated using LSTM method for a low patient criticality situation,  $\beta = 2$ ,  $\alpha = 1$  round,  $\mathcal{F} = \text{HR}$ .

sensor after each round according to the variation of the monitored condition. However, the stability level used in ASF algorithm allows each sensor to adapt more their frequency sensing than the Fisher test used in MLED. Therefore, in the applications, where a dynamic adaptation for the sensing frequency of sensors without respecting any threshold, the ASF algorithm becomes more suitable.

From the data transmission reduction point of view, EDASF algorithm outperforms the MLED in terms of reducing the amount of data collection and transmission compared to MLED. Subsequently, EDASF can reduce from 39% to 94% compared to MLED depending on the variation of the monitored vital sign. Consequently, when the priority for the application is the reduction of the amount of data transmission in order to make less complex the analysis of data at the sink side, the EDASF algorithm becomes more suitable.

From the energy consumption point of view, our ASF algorithm can largely reduce the energy consumption due the huge reduction in the data collection and transmission compared to the naïve mechanism. Subsequently, it can extend the sensor lifetime by 4 to 6 times depending on the chosen values of the used parameters.

From the data accuracy point of view, both algorithms EDASF and MLED ensure a high level of data integrity without any loss of information. This is because the LSTM method used in our mechanism can accurately regenerate the raw data collected by the sensors while, in MLED, the integrity of data is ensured through the Fisher threshold assigned by the expert.

Indeed, the selection of the threshold values and the processing complexity of the LSTM are two main challenges facing our mechanism. From one hand, selecting the appropriate values of thresholds is very essential in our mechanism which highly affects the results. Indeed, we believe that threshold

|      | Sensing frequency adaptation | Data transmission reduction | Energy consumption | Data accuracy |
|------|------------------------------|-----------------------------|--------------------|---------------|
| ASF  | very dynamic                 | high                        | high               | high          |
| MLED | less dynamic                 | medium                      | medium             | high          |

TABLE II  
COMPARISON BETWEEN ASF AND MLED MECHANISMS.

values should be determined by the decision makers or experts depending on the monitored features (e.g. vital signs). On the other hand, the processing complexity of the LSTM is highly dependent on the desired level of accuracy needed to predict the patient situation progress; more we need to increase the accuracy of LSTM more the processing complexity is, and vice versa. In order to overcome this problem, we can simply increase the resources at the sink node or to decrease the number of blocks used in LSTM.

## VII. CONCLUSION AND FUTURE WORK

Connected healthcare will continue to rise in order to improve the quality of care and provide low cost solutions for hospitals. In this paper, we have proposed an energy-efficient data analytics mechanism for a real-time patient monitoring and assessment. The proposed mechanism is mainly based on three phases: Emergency detection, adapting sensing frequency and patient situation progress prediction. Through simulations on real health data, we demonstrated the effectiveness of our mechanism in terms of saving the energy of each biosensor and reducing data redundancy, compared to other existing techniques.

As future work, we have three main directions to enhance our mechanism. First, we seek to adapt our proposed approach to take into account the correlation between neighboring nodes when sending the data to the sink. As the sensor nodes send their data at the same time (at the end of each period), collisions between packets are likely to happen repeatedly. Then it is essential for sensor nodes to be able to detect this repeated collision and introduce a phase shift between the two transmission sequences in order to avoid further collisions. Second, we plan to allow our mechanism to adjust the sensing frequency on the basis of the available energy beside the redundancies between readings collected in different periods. Third, we seek to adapt our platform to take into account various biomedical data like images for organs, video for operations, etc.

## REFERENCES

- [1] Deloitte, "2019 global health care outlook: Shaping the future," <https://www2.deloitte.com/global/en/pages/life-sciences-and-healthcare/articles/global-health-care-sector-outlook.html#report>, pp. 1–41, 2019.
- [2] N. Dey, A. S. Ashour, and C. Bhatt, "Internet of things driven connected healthcare," in *Internet of things and big data technologies for next generation healthcare*. Springer, 2017, pp. 3–12.
- [3] P. Singh, "Internet of things based health monitoring system: opportunities and challenges," *International Journal of Advanced Research in Computer Science*, vol. 9, no. 1, 2018.
- [4] J. Kim, A. S. Campbell, B. E.-F. de Ávila, and J. Wang, "Wearable biosensors for healthcare monitoring," *Nature biotechnology*, p. 1, 2019.

- [5] C. Habib, A. Makhoul, R. Darazi, and R. Couturier, "Real-time sampling rate adaptation based on continuous risk level evaluation in wireless body sensor networks," in *2017 IEEE 13th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*. IEEE, 2017, pp. 1–8.
- [6] M. Ibrahim, H. Harb, A. Nasser, A. Mansour, and C. Osswald, "On-in: An on-node and in-node based mechanism for big data collection in large-scale sensor networks," in *2019 27th European Signal Processing Conference (EUSIPCO)*. IEEE, 2019, pp. 1–5.
- [7] —, "Adaptive strategy and decision making model for sensing-based network applications," in *2019 19th International Symposium on Communications and Information Technologies (ISCIT)*. IEEE, 2019, pp. 96–101.
- [8] M. Koussaifi, C. Habib, and A. Makhoul, "Real-time stress evaluation using wireless body sensor networks," in *2018 Wireless Days (WD)*. IEEE, 2018, pp. 37–39.
- [9] C. Habib, A. Makhoul, R. Darazi, and C. Salim, "Self-adaptive data collection and fusion for health monitoring based on body sensor networks," *IEEE transactions on Industrial Informatics*, vol. 12, no. 6, pp. 2342–2352, 2016.
- [10] C. Habib, A. Makhoul, R. Darazi, and R. Couturier, "Health risk assessment and decision-making for patient monitoring and decision-support using wireless body sensor networks," *Information Fusion*, vol. 47, pp. 10–22, 2019.
- [11] G. Giorgi, "A combined approach for real-time data compression in wireless body sensor networks," *IEEE Sensors Journal*, vol. 17, no. 18, pp. 6129–6135, 2017.
- [12] S. Vitabile, M. Marks, D. Stojanovic, S. Pillana, J. M. Molina, M. Krzyszton, A. Sikora, A. Jarynowski, F. Hosseinpour, A. Jakobik *et al.*, "Medical data processing and analysis for remote health and activities monitoring," in *High-Performance Modelling and Simulation for Big Data Applications*. Springer, 2019, pp. 186–220.
- [13] N. Boudargham, J. B. Abdo, J. Demerjian, C. Guyeux, and T. Atechian, "Efficient cluster-based routing algorithm for body sensor networks," in *2018 IEEE Middle East and North Africa Communications Conference (MENACOMM)*. IEEE, 2018, pp. 1–6.
- [14] L. Syed, S. Jabeen, S. Manimala, and H. A. Elsayed, "Data science algorithms and techniques for smart healthcare using iot and big data analytics," in *Smart Techniques for a Smarter Planet*. Springer, 2019, pp. 211–241.
- [15] M. Mohammadi, A. Al-Fuqaha, S. Sorour, and M. Guizani, "Deep learning for iot big data and streaming analytics: A survey," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 4, pp. 2923–2960, 2018.
- [16] B. O. Soufiene, A. A. Bahattab, A. Trad, and H. Youssef, "Lightweight and confidential data aggregation in healthcare wireless sensor networks," *Transactions on Emerging Telecommunications Technologies*, vol. 27, no. 4, pp. 576–588, 2016.
- [17] A. Arfaoui, A. Kribeche, S. M. Senouci, and M. Hamdi, "Game-based adaptive anomaly detection in wireless body area networks," *Computer Networks*, vol. 163, p. 106870, 2019.
- [18] T. Wang, M. Z. A. Bhuiyan, G. Wang, M. A. Rahman, J. Wu, and J. Cao, "Big data reduction for a smart city's critical infrastructural health monitoring," *IEEE Communications Magazine*, vol. 56, no. 3, pp. 128–133, 2018.
- [19] F. Kiani, "Reinforcement learning based routing protocol for wireless body sensor networks," in *2017 IEEE 7th International Symposium on Cloud and Service Computing (SC2)*. IEEE, 2017, pp. 71–78.
- [20] S. Begum, S. Barua, and M. Ahmed, "Physiological sensor signals classification for healthcare using sensor data fusion and case-based reasoning," *Sensors*, vol. 14, no. 7, pp. 11 770–11 785, 2014.
- [21] S. M. Martínez Chávez, M. E. Rivero-Angeles, L. I. Garay-Jiménez, and I. C. Romero Ibarra, "Priority schemes for life extension and data delivery in body area wireless sensor networks with cognitive radio capabilities," *Wireless Communications and Mobile Computing*, vol. 2019, 2019.
- [22] R. Ward, "Energy-efficient data reduction techniques for eeg wireless body sensor networks," in *Qatar Foundation Annual Research Forum Volume 2013 Issue 1*, vol. 2013, no. 1. Hamad bin Khalifa University Press (HBKU Press), 2013, pp. BIOP-041.
- [23] M. Ayaz, M. Ammad-Uddin, Z. Sharif, A. Mansour, and E.-H. M. Aggoune, "Internet-of-things (iot)-based smart agriculture: Toward making the fields talk," *IEEE Access*, vol. 7, pp. 129 551–129 583, 2019.
- [24] A. S. Abiodun, M. H. Anisi, I. Ali, A. Akhuzada, and M. K. Khan, "Reducing power consumption in wireless body area networks: a novel data segregation and classification technique," *IEEE Consumer Electronics Magazine*, vol. 6, no. 4, pp. 38–47, 2017.
- [25] National Early Warning Score (NEWS), Royal College of Physicians, London, U.K., <http://www.rcplondon.ac.uk/resources/national-early-warning-score-news>, 2015.
- [26] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [27] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, pp. 1–15, 2014.
- [28] "Mimic database on physionet," <https://www.physionet.org/>, accessed: 2016-10-4, 2000-present.