



**HAL**  
open science

# Collaborative Internet of Things (C-IoT) Data Analysis for Enhancing Activity Recognition and Preventing Serious Health Problems

Imad Belkacem, Aymen Gammoudi, Nasredine Cheniki, Yacine Sam, Nizar  
Messai

► **To cite this version:**

Imad Belkacem, Aymen Gammoudi, Nasredine Cheniki, Yacine Sam, Nizar Messai. Collaborative Internet of Things (C-IoT) Data Analysis for Enhancing Activity Recognition and Preventing Serious Health Problems. 30th IEEE International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises WETICE 2021, Oct 2021, Bayonne (en ligne), France. hal-03468091

**HAL Id: hal-03468091**

**<https://hal.science/hal-03468091>**

Submitted on 6 Dec 2021

**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.

# Collaborative Internet of Things (C-IoT) Data Analysis for Enhancing Activity Recognition and Preventing Serious Health Problems

Imad Belkacem\*, Aymen Gammoudi†, Nasredine Cheniki‡, Yacine Sam§ and Nizar Messai¶

University of Tours, LIFAT Laboratory

Tours, France

Email: {firstname.lastname}@univ-tours.fr

**Abstract**—Collaborative Internet of Things (C-IoT) is an emerging paradigm that has generated a large amount of accessible and usable data in real-time, constituting an interesting tool for decision-making problems. The paper presents a healthcare data analytics approach based on collecting and analyzing data from C-IoT. We first provide an architectural plan involving several distributed sites collaborating to collect necessary data for patient’s health conditions monitoring. We use for that well-known IoT technologies like Radio Frequency Identification (RFID), Near Field Communication(NFC), Beacons and Ambient Assisted Living sensors. We then learn on such data to create adequate patient profiles and use the Pearson Correlation Coefficient (PCC) over a long period of time data to detect potential health risks. Finally, decisions are made regarding the patient’s condition according to the symptoms/diseases detected.

**Index Terms**—C-IoT, RFID, NFC, Beacon, Healthcare, Activity Recognition, Data Analysis, Monitoring

## I. INTRODUCTION

The Internet of Things has all the assets to allow collaborative working when the various geographically distant sites (Internet connected) cooperate and share information collected from different readers and sensors to achieve a common goal. This constitutes the basics of the emerging Collaborative IoT paradigm (C-IoT) that contributes to considerably increase the amount of data being able to help decision-makers [1]. In the context of C-IoT, data can be collected, integrated and analyzed allowing decision-makers to choose appropriate actions and provide analysis allowing the explanation of the decided actions [2]. Healthcare is one of the important areas where C-IoT can constitute a powerful tool to improve human health and living conditions. In a collaborative health system, multiple information can be gathered and shared between several actors (e.g., doctor, pharmacist, emergency doctor, etc.) in order to remotely cooperate for the patient’s health monitoring purposes (e.g., blood pressure, heart, and respiratory rates, etc...). Taking advantage of the multiple data sources available in the context of C-IoT paradigm, we present here an approach to collect and analyze great amounts of data to predict serious health problems such as emergencies (seizures or attacks). In fact, an unusual observation, such as the sudden rise of a patient’s temperature, does not automatically mean danger for the patient, and paradoxically a perturbation over a period of time may go unnoticed and a potential emergency can occur

at any time if no action is taken. That’s what motivated us to exploit patients’ data stored in the different databases/sensors deployed in different geographical sites, in a C-IoT context. We proceed by recovering different dated identifications and measurements taken by different sensors of vital parameters like blood pressure, temperature, respiratory frequency, etc., to draw the patient’s profile. Such a profile is then compared to real-time data regarding the patient’s health state. The analysis of collected data based on RFID and AAL (Ambient Assisted Living) sensors makes it possible to make the right decisions (e.g., trigger alerts) and appropriate recommendations (avoid dangerous activities or advice activities) in case of problems. Moreover, it improves non-visual recognition of activities to alert the patient using beacons with personalized warning notifications in case of non-serious problems and caregiver (notify medical staff, doctor, emergency physician, pharmacist, patient’s relatives) to react according to the patient’s health situation and the proximity distance between the caregiver and the patient.

The rest of the paper is organized as follows: Section II presents recent work tackling C-IoT in Healthcare and activity recognition. Section III details the proposed contribution, specifically the architecture and the methodology allowing to ensure the enhancement of non-visual recognition of activities and the prevention of serious health problems and presents then the statistical techniques for collected data analysis and Section IV concludes the paper and gives prospects.

## II. RELATED WORK

This section presents recent work regarding activity recognition and Healthcare in the context of C-IoT. The authors in [3] present an approach to refine visual activity recognition based on semantic rules reasoning. The information is collected from sensors in a Smart Home context for elderly people monitoring and can easily be generalized to healthcare. In the approach, data collection and activities detection are mainly based on visual supervisions of the patient, which can be invasive, since the tendency nowadays is the use of non-invasive and non-visual data to not disturb the patient, especially with cameras records. Moreover, video and image processing methods are more expensive and time-consuming. Compared to [3], our approach, based on non-visual sensors, is more suitable for

patients wishing not to be supervised with images and video records. Moreover, our approach is time computationally more efficient since the analyzed data recovered from sensors is smaller, so easier to be processed.

The authors in [4] identify the challenges facing health sectors in dealing with the COVID-19 epidemic and propose potential solutions in terms of popular technologies such as AI and IoT by conducting a critical literature review. In [5] authors propose online activity recognition where data is collected from three patients' sensors, the accelerometer, the gyroscope, and the magnetometer. The approach is completely smartphone-based in the patient's health monitoring process, that is, it can be useless if the patient forgets or lose her smartphone; a fact that can be probable in the case of elderly people. In our approach, we however use AAL sensors embedded in RFID/NFC tags and online collect and analyze data from different remote geographic sites since in case of emergency the connection should be maintained active between the server and the patient. We then apply statistical tests on the collected data to detect dangerous behaviors such as disturbances in heart rate, respiratory rate, etc. Such detection can prevent more dangerous situations (heart attack, apnea, etc).

### III. A C-IOT BASED ARCHITECTURE TO COLLECT AND ANALYZE DATA

The global architecture of our approach, depicted in Figure 1, can be deployed in an individual house as well as in one shared between several people such as a Nursing Home. The components of our system consist from one hand on  $n$  geographically distant sites where a patient can be observed, and a central server that controls identification and observation processes from the other. In such a C-IoT context with a great volume of collected/analyzed data, this distribution of the architecture over multiple geographical sites, facilitated by the advent of service-oriented technology [6], is a necessity to make efficient data processing.

The communication between the different sites and the server is realized with *Reader Web Service (RWS)* module that allows data recovery from the different sites' readers/sensors. Technically, the central server, located in the principal node, sends periodically SOA Web services requests to each station managing a particular site (see Figure. 1). Stations, equipped with RFID readers and other sensors, respond then with another Web service application containing information about inventory and different environmental observations. Each site  $S_i$  contains in fact  $n_i$  readers allowing to detect and inventory a set of tags containing  $m_i$  sensors (with  $m_i \leq n_i$ : a tag carries at most one sensor). In every single site  $S_i$ , we have patients with *RFID/NFC* tags worn or attached to. Some of them are equipped with AAL sensors (sensors of heat, heart rate,...). The sensors, important elements of the architecture, make it possible to read the physical attributes of the patient carrying the tag in addition to her identifications. Here, each *RFID/NFC* sensor tag provides an elegant, precise and relatively easy way to implement the solution [7].

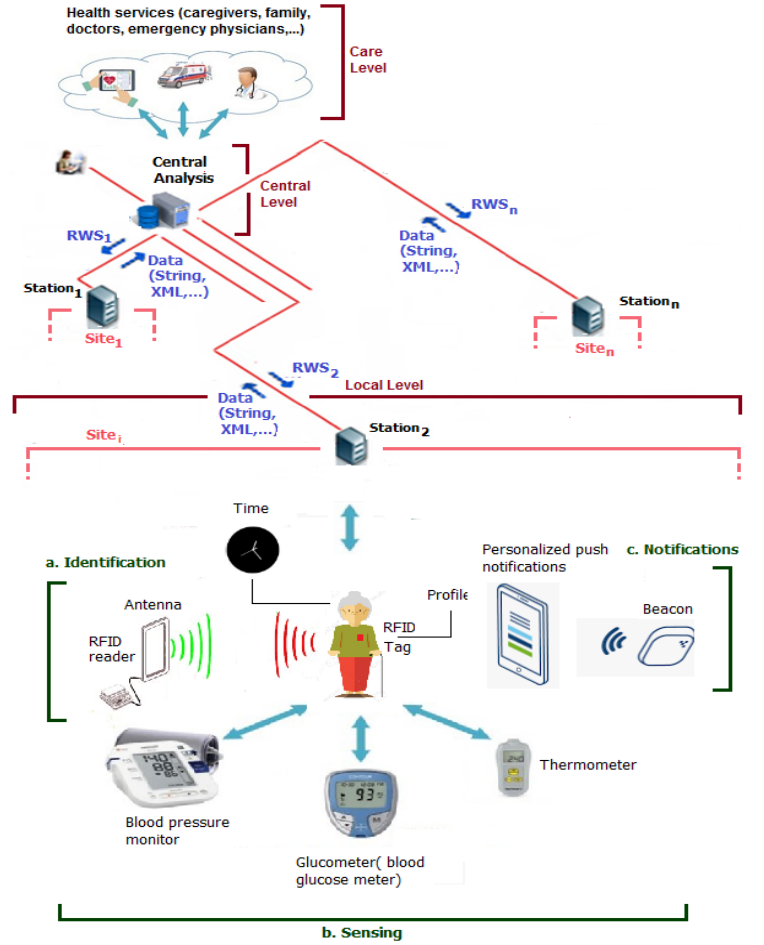


Fig. 1. The C-IoT Network architecture : a distributed C-IoT system observing  $n$  geographical distant sites.

#### Algorithm 1 : Data Collection From A Particular Site

---

**Inputs:**  $code\_site, url$   
**Outputs:**  $V, M$

```

1 Begin
2   Function  $get\_Data\_String(url, code\_site)$ 
3     call  $\leftarrow$  new call of the reader web service (RWS)
4     /* Invoke the service */
5     Response  $resp = call.invoke(url, code\_site)$ 
6     If ( $resp.generatedFault() \parallel$  timestamp expired) then
7       /* Error */
8     Else
9       /* Call was successful */
10      String result =  $resp.getString()$ 
11    End If
12    /* Data Recuperation */
13    String str =  $getDataString(url, code\_site)$ 
14    /* Data Filtering */
15     $I \leftarrow Extract\_Vector(str)$ 
16     $M \leftarrow Extract\_Matrix(str)$ 
17    /* Analysis will be based on I and M */
18  End

```

---

Algorithm 1 consists in a technique for data recovering from a given station using as input its code and Uniform Resource Locator (URL). The function *get\_String\_data(url,site\_code)* allows the invocation of a new RWS using URL and site code. An exception code is sent to the server in case of failure or timeout response. The function *Extract\_Vector (String)* generates the vector of inventories  $V$  from the data string returned by the RWS, where *Extract\_Matrix (String)* is a function that generates the matrix of environmental data  $M$ . The vector of inventories is defined as follows :

$$V=[I_1, I_2, \dots, I_n]$$

where  $I_i$  is the number of detected tags of the inventory done by the  $i^{th}$  reader. The matrix of environmental data is defined by:

$$M_{m,n} = \begin{pmatrix} M_{1,1} & M_{1,2} & \dots & M_{1,n} \\ M_{2,1} & M_{2,2} & \dots & M_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ M_{m,1} & M_{m,2} & \dots & M_{m,n} \end{pmatrix}$$

where  $M_{i,j}$  presents the observation of the sensor  $C_i$  read by the reader  $R_j$ . The analysis will be based on this data to monitor the patient's health. Algorithm 1 concerns data collection from a single site. Its complexity could be  $O(n \times m)$  where  $n$  is the maximum number of RFID readers and  $m$  the maximum number of sensors deployed in the site.

The global architecture is composed of three levels. (i) The **Care Level** consists on people taking charge of monitoring the patient's health: caregivers, doctors, nurses, pharmacists, emergency doctors, or patients relatives. At the (ii) **Central Level**, data collected from the different sites where the patients are monitored is analyzed. This level makes moreover it possible to make the necessary decisions after analyzing the recovered data. In the (iii) **Local Level**, three kind of operations are performed. The **Identification** of the patient using RFID and NFC technologies. The **Capture** allows vital parameters retrieval based on observations from AAL sensors incorporated in the various RFID and NFC tags. **Notification** operation consists to send information using beacons located in the various places where the observed person may be located.

Algorithm 2, implemented in the Central server allows to ensure the monitoring of the patients in different sites of the C-IoT, takes as input a list of couples  $\{\text{code\_site}, \text{url}\}$  and produces as output a list of events  $\{ev_1, ev_2, \dots, ev_n\}$  and decisions  $\{D_1, D_2, \dots, D_n\}$ .

Algorithm 2's complexity is the same as that of Algorithm 1 seen above, since it consists to repeat Algorithm 1 data collection function on a fixed number of C-IoT geographical sites. So the decision taking process is also realized in a reasonable way.

We present now our two phases methodology to detect patient's activity and potential health problems. The **Learning phase** allows to draw the health profile of the patient from the collected information: (i) sleeping time at night, and nap time. (ii) AAL medical parameters; temperature, respiratory

---

## Algorithm 2 : Decision Taking Algorithm

---

**Inputs:**  $L$  : List of (code\_site, url)

**Outputs:**  $Ev$  : List of Events,  $D$  : List of Decisions

```

1 Begin
2 Regulate the various sites on the same clock
3 Repeat /* Periodically */
4   Foreach (code_site in L) do
5     /* Data Recovery*/
6     String str ← get_Data_String(url, code_site)
7     /* Data Filtering */
8     V ← Extract_Vector(str)
9     M ← Extract_Matrix(str)
10  End Foreach
11  /* Data stotage, selection and analysis*/
12  Deduce a list of Events Ev
13  /* Decision Taking */
14  Taking a list of decisions D based on Ev
15 End Repeat
16 End

```

---

rate, heart rate, blood pressure ; (iii) frequency of going to the sanitary and length of time spent there. This information completes the known diseases about the patient. The second phase consists in **Data analysis and patient monitoring**. After the Learning phase, the health state of the patient being monitored is obtained. Therefore, its health state is periodically (daily, weekly, monthly,...) analyzed and comparing to those drawn up in the learning phase. This phase goes through several stages that are: data streams retrieval and storage; data filtering; building data structures ready for analysis; selection of data required for analysis; analysis of the selected data; detection of simple events  $\{ev_1, ev_2, \dots, ev_n\}$ ; composition to have complex events and taking decisions  $\{D_1, D_2, \dots, D_n\}$  according to detected events and sending alerts.

We present the statistical methodology used in our approach for health problems detection based on the data collected in the context of our architecture using C-IoT technologies. The statistical hypothesis tests [8]–[11] are used for monitoring the patient's health. A statistical hypothesis, called confirmatory data analysis, is an hypothesis that can be tested based on the observation of a process modeled using a set of random variables [10]. To detect a potential danger regarding the patient's health, his/her current health state data is compared to his/her profile data. We notice here that both the profile and the current health state data are represented by homogeneous data of the identification vector  $V$  and a matrix  $M$  data related to a given vital parameter.

Pearson's Correlation Coefficient (PCC) is useful for potential health problems detection. It is an interesting index reflecting a linear relationship between two continuous variables. The correlation coefficient (in absolute value) varies between 0 and 1, with 0 reflecting a zero relationship between the two variables. In the health context, a change in the health state generates more correlated data to profile data and can mean danger on the patient's health. The PCC ( $r$ ) between the patient

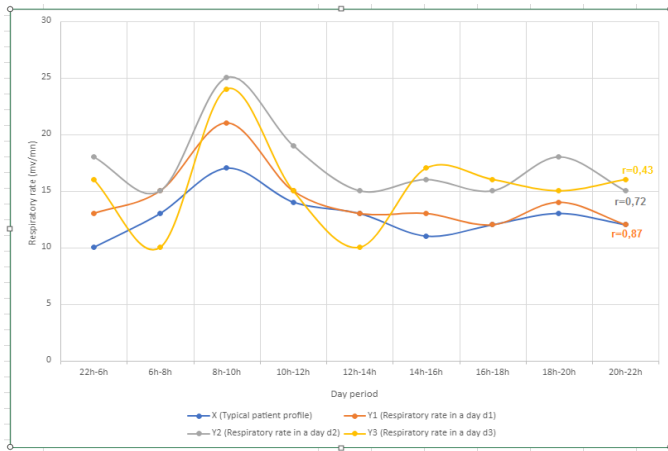


Fig. 2. Scatter plot between the typical patient profile data ( $X$ ) and (a) the correlated state ( $Y_1$  in blue with  $r = 0,87$ ); (b) the first uncorrelated State ( $Y_2$  in grey with  $r = 0,72$ ); (c) the second uncorrelated State ( $Y_3$  in yellow with  $r = 0,43$ ).

profile data ( $X$ ) and current state data ( $Y$ ) is defined as:

$$r = \left| \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 (Y_i - \bar{Y})^2}} \right| \quad (1)$$

where  $X_i$  is the  $i^{th}$  value of the profile data;  $Y_i$  is the  $i^{th}$  data value of the vector of values to compare; and  $\bar{X}$ ,  $\bar{Y}$  are respectively the sample mean of the two vectors. Depending on the value of  $r$ , the Central Server detects either an improvement in the patient's health state or a deterioration that needs to be verified by a doctor. Thus, in order to consider no change in the health state, the correlation value  $r$  needs to be close to 1. We have fixed the threshold  $r = 0.8$  and consider that the profile and current state data are correlated if  $r \leq 0.8$ . For example, one can track the respiration rate that represents the number of respiratory cycles (an inspiration and an expiration) per minute; and compare it to patient profile data or reference data that is between 12 and 20 cycles for the adult. For example, based on a synthetic dataset that we have generated, Figure 2 shows the variation of the PCC between her typical profile  $X$  for a patient and his health conditions in different days ( $Y_1$ ,  $Y_2$  and  $Y_3$ ) regarding the breathing frequency vital parameter.

We give hereafter the interpretation of results given in the scatter graphs in Figure 2 in days  $Y_1$ ,  $Y_2$  and  $Y_3$ . For the first day ( $Y_1$ ), the PCC ( $r$ ) is evaluated to 0.87, that is qualified by a *largely positive*. This result can be interpreted by a stability of the patient's health state of health. In  $Y_2$ , the PCC is evaluated to 0.72, that is qualified by *positive*; and this can be interpreted by an instability in the patient's health state. Finally, and for the day  $Y_3$ , the PCC is evaluated to 0.43, that is qualified by *moderately positive* and this can be interpreted as a disturbance in the patient's breathing.

An aspect to consider when interpreting the correlation coefficient is sensitiveness to extreme data values. A single value being very different from the others can significantly

change the value of the coefficient. For all the extreme values, the causes should be determined to prevent the danger of this singular suspicious behaviors. To do, the generalized test Extreme Studentized Deviate (ESD test) ([11]–[14]) is used.

Regarding the sample size, the larger the sample size, the correlation coefficient provides more information about the linear correlation provided by the data. However, this does not invalidate the test, because the test remains valid for small samples too.

#### IV. CONCLUSION & PERSPECTIVES

We have presented a collaborative Internet of Things (C-IoT) architecture based on RFID enriched with the use of innovative technologies as AAL sensors and Beacons. The data recovering algorithm allows data collection from remote geographical sites to be periodical analysed in order to detect abnormal health situations. The proposed analysis algorithm compares the current health state of the patient to her data profile based on Pearson's Correlation Coefficient (PCC). That real-time monitoring process can predict crises that can be fatal, detect new diseases, provide recommendations or alert patients and caregivers. The approach is validated on synthetic data and we plan as a future work to test it in realistic situations with real patients.

#### REFERENCES

- [1] J. B. Borges Neto, T. H. Silva, R. M. Assunção, R. A. Mini, and A. A. Loureiro, "Sensing in the collaborative internet of things," *Sensors*, vol. 15, no. 3, pp. 6607–6632, 2015.
- [2] F. Behmann and K. Wu, "Collaborative internet of things (C-IoT): For future smart connected life and business". John Wiley & Sons, 2015.
- [3] N. Ramoly, V. Vassout, A. Bouzeghoub, M. A. El Yacoubi, and M. Hariz, "Refining visual activity recognition with semantic reasoning," in *2017 IEEE 31st International Conference on Advanced Information Networking and Applications (AINA)*. IEEE, 2017, pp. 720–727.
- [4] S. Kumar, R. D. Raut, and B. E. Narkhede, "A proposed collaborative framework by using artificial intelligence-internet of things (ai-iot) in covid-19 pandemic situation for healthcare workers," *International Journal of Healthcare Management*, vol. 13, no. 4, pp. 337–345, 2020.
- [5] A. R. Javed, M. U. Sarwar, M. O. Beg, M. Asim, T. Baker, and H. Tawfik, "A collaborative healthcare framework for shared healthcare plan with ambient intelligence," *Human-centric Computing and Information Sciences*, vol. 10, no. 1, pp. 1–21, 2020.
- [6] Y. Natis and R. Schulte, "Introduction to service-oriented architecture," Gartner Research Group, Tech. Rep. AV-19-6751, April 2003.
- [7] A. Ruhanen, M. Hanhikorpi, F. Bertuccelli, A. Colonna, W. Malik, D. Ranasinghe, S. L. Tomas, N. Yan, and M. Tavilampi, "Sensor-enabled rfid tag handbook," Building Radio frequency IDentification for the Global Environment Project, Tech. Rep. MSU-CSE-00-2, 2008.
- [8] M. Kendall and A. Stuart, *The Advanced Theory of Statistics, Vol 2: Inference and Relationship*, London, 1973.
- [9] G. Saporta, "Probabilités, analyse des données et statistique," *Editions Technip, Paris*, 1990.
- [10] A. Stuart, J. K. Ord, and S. Arnold, *Kendall's advanced theory of statistics. Vol. 2A: Classical inference and the linear model*. London: Hodder, 1999, vol. 2.
- [11] I. Belkacem, S. Nait-Bahloul, and D. Sauveron, "Enhancing dependability through profiling in the collaborative internet of things," *Multimedia Tools and Applications*, vol. 78, no. 3, pp. 2983–3007, 2019.
- [12] F. E. Grubbs, "Procedures for detecting outlying observations in samples," *Technometrics*, vol. 11, no. 1, pp. 1–21, 1969.
- [13] F. E. Grubbs and G. Beck, "Extension of sample sizes and percentage points for significance tests of outlying observations," *Technometrics*, vol. 14, no. 4, pp. 847–854, 1972.
- [14] B. Rosner, "Percentage points for a generalized ESD many-outlier procedure," *Technometrics*, vol. 25, no. 2, pp. 165–172, 1983.