Abstract: The increase in the number of devices equipped with physiological sensors and their low price mean that they can be used in many fields. One of these fields is health-care and home-care for the elderly or people with disabilities. The development of such devices makes it possible to monitor their condition continuously and at all times. A continuous monitoring not only establishes an image of the user’s status, but also detects possible anomalies. Therefore, it is necessary to develop a distributed architecture that allows expert analysts to access the data provided by the sensors at all times and from anywhere. This paper introduces the development and implementation of the concept of distributed architecture, focusing on the minimum requirements needed to carry it out. All the necessary modules are described for different stages: acquisition, communication and processing of physiological signals. The last stage is carried out by a machine learning system. The complete reporting and storage system is also described. Finally, the most important conclusions that have emerged during the development are reported.

Keywords: NoSQL; physiological signals; distributed architecture

1. Introduction

The rapid increase in ubiquity, mobility, big data, data analytics and cloud computing, as well as the advances in wearable devices, are transforming our society. At the same time, the rapid development of novel wearable devices and their lower cost are expanding the number of connected devices. This growth opens up numerous fields that would benefit from applying this technology [1–3]. The health and domestic care sector also takes advantage from this increase in research. Unfortunately, the number of elderly people with disabilities or chronic illnesses increases too day by day. The use of sensor technology has the potential to provide a significant impact on their daily lives. A continuous monitoring of their physiological variables within an intelligent environment will help them improve their quality of life and independence [4,5].

Thus, the use of wearable devices for continuous monitoring will be crucial in health care in the near future. They would allow for close monitoring of changes in an individual’s vital signs and provide feedback to help maintain optimal health. When being embedded in a telemedicine system, these devices may also be used to alert medical personnel when potentially life-threatening situations occur [6,7]. For example, they can be used as part of a diagnostic procedure, a supervised recovery
from a surgical procedure, the optimal management of a chronic illness, psychiatric therapies, and also to monitor the compliance with treatment guidelines, among many other possibilities.

The utilisation of this type of devices offers many advantages such as low weight, low price and also the fact that they are minimally invasive. However, the integration of different devices poses some difficulties as well. In general, the relatively small size and heterogeneity of the different capture devices typically results in small storage and processing capacity [8–10]. This is even worse if we take into consideration that these devices demand a large storage and processing capacity, a high-speed connection to the network, as well as the possibility to make decisions in real time [11].

Fortunately, the rise of systems based on big data (BD) and machine learning (ML) help facing the problem of storage and processing capacity [12–14]. The power of calculation and the large storage capacity of ML and BD provides a way of solving this issue [15]. For this reason, it is necessary to create tools that allow the adaptation, integration and analysis of all physiological data obtained [7,16]. Therefore, the main objective of this study is to establish the foundations of a distributed system based on non-relational databases (NoSQL) for the acquisition, processing and visualisation of physiological signals. The proposed system also integrates a classification system (“Machine Learning System”) that determines the status of the users as well as the occurrence of an incident.

The rest of the paper is structured as follows. A brief review of previous works on the topic is provided in Section 2. Section 3 details the design and implementation of the proposed architecture. Finally, Section 4 presents the most outstanding conclusions of this study.

2. Background and Related Works

In recent years, the development of systems capable of integrating several biomedical sensors in order to measure an individual’s physiological signs has received a great attention from the research community. These physiological sensors include electroencephalography (EEG), blood volume pressure (BVP), electrodermal activity (EDA), accelerometer (ACC), and temperature (TEMP). Until recently, this type of data acquisition required the use of bulky systems, usually wired, which made them invasive, expensive and impractical [17]. The traditional approach to signal acquisition was based on the development of tools that allowed the local storage of the data [18].

Recently, a new method has been proposed to transmit physiological signals over a wireless network [19]. The signals are converted into a series of data that are transmitted to a computer via the Bluetooth protocol. As can be seen in Figure 1, the acquisition procedure is one-way. The acquisition devices are placed on the user’s body, the data are then sent wireless to a computer using the Bluetooth protocol. The computer stores the data using comma-separated value (.csv) files. After this, the signals are processed in order to extract information, which is usually not done in real time in other approaches. This procedure requires time to perform the corresponding analysis, which normally causes the acquisition system to be separated from the analysis system [9,20]. However, the research on distributed systems is changing the linear paradigm of acquisition.

![Figure 1. Classic model for acquisition of physiological signals.](image)

Therefore, a distributed system with high availability, processing and data storage capacity seems more appropriate for use in health care [8,21,22]. Indeed, a new set of health services has been
developed, such as flow-based decision support services, data mining and pattern-based visualisation, and monitoring services [23]. These services enable managers to control the quality of health systems.

2.1. Distributed Systems and Databases

A distributed system is defined as a set of standalone computers connected to each other by a network, and supported by a distributed software. This allows computers to coordinate their activities, sharing hardware, software and data resources in such a way that the end user perceives it as if it were a single computer, even when the machines are in different locations [24].

All distributed systems must be based on six characteristics: resource sharing, extensibility, concurrency, scalability, fault tolerance and transparency. These characteristics are aligned with the point of view of this study, as they make it possible to give a distributed approach to the data collected. This proposal allows having a significant number of devices sending data simultaneously without affecting the service.

The use of distributed databases is becoming more and more common nowadays due to the large amount of data handled by web sites, applications and services [25,26]. The large volume of data and the low storage capacity of the devices, together with the need of having many users sending data at the same time, makes the use of distributed databases suitable for the purpose of this work.

2.2. NoSQL Database

Despite the lack of a formal definition, a NoSQL database refers to a wide class of data management systems that differ in important aspects from the classical model of relational database management systems. The main difference is that they do not use SQL as the main data query language. NoSQL databases utilise a variety of data models to access and manage data such as documents, graphs, key-value, in-memory, and search. These database types are specifically optimised for applications that require large volumes of data, low latency, and flexible data models [25,27]. Therefore, NoSQL databases are structures that are capable of storing information in situations in which relational databases present certain scalability and performance problems.

An analysis of the different NoSQL technologies available lead us to use the ecosystem provided by MongoDB against other alternatives. MongoDB has been defined as a document type NoSQL database based on a JSON document model. The use of this type of document means that there is no need to follow a rigid scheme in the insertion of data. This laxity allows having data from different sensors in a same document. In this way, if the connection with any of the sensors is dropped, no information is lost [28]. Being a distributed system, MongoDB provides developers with four essential features: availability, workload isolation, horizontal scalability and data location [28,29].

MongoDB offers the ability to keep several copies of the data stored using different data sets (nodes). This database makes use of a sharding technique, which is a method that allows dividing the total set of data between the different nodes of the distributed system. It enables to have the data protected in different replica sets within the database. This type of writing guarantees the availability of the data without sacrificing their consistency. A unique feature of this database is the figure of the arbiter. When one (or more) of the nodes in the database is no longer available, the remaining nodes determine (through a consensus protocol leaded by the arbiter) which of them acts as the main database, while the rest of nodes act as a replica set. This procedure minimises data loss by enacting an autonomous recovery in case of failure [28,29].

The MongoDB database has many advantages over other databases such as Cassandra, CouchDB and Redis [30,31]. The system replication and sharding, the arbitration system and the automatic restoration of the data in case of the disconnection of any of the nodes makes it adequate for the task at hand. These are the main reasons that lead us to propose it for the implementation of our system.
3. Modelling an Architecture for Physiological Signals

3.1. Architecture Description

The proposed architecture is based on four facets: Sensing, Data Management, Machine Learning and a Reporting System. As it can be guessed from the context, Sensing refers to all devices that measure physiological variables in the user’s environment. Data Management is aimed at configuring and managing the system so that the produced data is correctly stored and used. Machine Learning carries out the analysis of the data and Reporting System provides a report of the analyses performed. Next, a more detailed description of the system will be made.

As can be seen in Figure 2, the information flow starts from the user. The raw signals captured by the sensors are transmitted to the local server using the Bluetooth protocol. When they reach the local server, these data are stored in a local NoSQL database. The first replication of the data takes place within this local database. We use the sharding system to organise the data in the different nodes of the local database. In the next step, if the system is connected to the Internet, the data are replicated in the cloud server where all data are stored and processed. In the case that an Internet connection is not available, the local server stores the data and uploads it to the cloud as soon as possible. Once the data are stored in the “Cloud Storage” database, they are replicated once again and stored in two different NoSQL databases. There are two main reasons for this. On the one hand, it provides a backup of the data and, on the other, the second database can be used by the “Machine Learning Service”.

The data stored in “Machine Learning Storage” are transferred to the “Machine Learning Service” (see Section 3.3), where they are processed by the several algorithms associated to each type of physiological signal. On the other hand, “Cloud Storage” is only dedicated to storing data and displaying the associated reports. Once the different signals have been processed, the obtained data are transferred again to the “Cloud Storage” database where are added to the user’s data.
The main purpose is to have the development system separated from the production system. In this way, new processing and classification models can be developed, preventing the system from failing. This is one of the advantages of using a distributed database. The two environments can be separated while improving the consistency, availability and robustness of the system [21,28].

3.2. Data Management System

One of the most important parts of the architecture is the “Data Management System” (DMS). Due to the heterogeneity and complexity of the data and the different sampling rates by which they were acquired, it is necessary to build a subsystem to effectively manage the data. For this purpose, the architecture relies on a previous development that synchronises the data when stored locally [32]. Figure 3 depicts the functioning of the DMS. The signals are obtained from the different acquisition systems. Then, they are forwarded to the “Data Parsing” module where they are synchronised with the time stamp of the system. Finally, they are stored at the same time both in the local NoSQL database and in the “Cloud Storage”, if it is available.

![Figure 3. Diagram of “Data Management System”](image)

3.3. Machine Learning Service

This module is in charge of processing all the data acquired from each user. As can be seen in Figure 4, it is replicated in the “Machine Learning Storage” database. When performing the analysis, the user’s data are extracted and sent to different modules. Although there are physiological signals that do not need to be processed such as TEMP, ACC data and IB, there are others like BVP, EDA and EEG signals that need to be operated. Our proposal is to create a system that is capable of processing autonomously and obtaining different markers to establish the psycho-physical state of the user.

Different classification and analysis systems are needed. In the case of the EDA signal, this marker is a very good indicator of stress [33], allowing, for instance, the use of a support vector machine based approach [34]. EEG signals are another good indicator of the user’s emotional state [35–37]. For this purpose, non-linear and other approaches based on responses in the different frequency bands can be used [38–40]. On the other hand, the TEMP, IB, BVP and ACC signals can be used to determine the user’s state of agitation and stress. Therefore, our main aim is to create an active monitoring system to help in decision making, therefore, integrating the technologies that are available.
Figure 4. Diagram of “Machine Learning Service”.

3.4. Reporting System

The Reporting System consists of a web-based monitor, which enables an expert analyst to observe data acquired from a subject. In this system, data is accessible at any time, anywhere and from any type of device. This Reporting System does not only feed on the raw data obtained from each of the users, but it also uses the developed “Machine Learning Service” to have more information on the variables obtained. It also allows setting different kinds of alarms in an automatic way, alerting of important changes in the state of the user. Figure 5 depicts a diagram of the different modules of the proposed Reporting System.
The information is extracted from the “Cloud Storage” database, which passes it through a module called “Physiological Monitor”. This module is in charge of managing all the information that has to be shown to the analyst. On the other hand, the Web Interface access the following items:

- **Real-Time Dynamics**: Real-time query of each of the signals obtained from the different sensors.
- **Status Overview**: Checks the current status of each user as well as the most relevant information.
- **Device Distribution**: Lists the most relevant information of the connected devices.
- **User Record**: Allows consulting the history of the stored data as well as the processed signals.

The data displayed on the screen has been divided into three elements:

- **Real-Time Monitor**: Displays in real time the signals obtained from the different sensors.
- **Real-Time Query**: Checks in real time the current status of each user.
- **Anomaly Notification**: This module is responsible for continuous sampling on the status of the user and, in the event of detecting some kind of anomaly, warning both the analyst as well as the user.

4. Conclusions

This article has presented a proposal of a distributed architecture for the acquisition, treatment and storage of physiological signals. This approach is based on distributed architectures and NoSQL databases. The architecture allows the connection of several devices such as a wearables devices and a brain-computer interface. It can be consulted at any time and in any place.

This system is aimed at helping vulnerable population groups, such as elderly people or people with disabilities. Therefore, the devices used must be as non-invasive as possible, allowing the monitoring and tracking of the user at all times. The system must issue an alert if there is an anomaly in the behaviour of the users. It is necessary to develop the tools that provide all the markers needed to detect these situations. The aim is the creation of an intelligent system that will automatically perform the tasks currently performed by an analyst. This decision-making system must be based on two main technologies: big data and machine learning. These technologies are being used more and more nowadays, providing many opportunities in different fields. The use and adaptation of these techniques can significantly boost the fields of health-care and home-care.

As a preliminary work, some functionalities have already been developed. The real-time acquisition system, which is a very important part of the architecture, has already been implemented. However, there is still work left in development of the “Machine Learning Service”. As a proof of concept, support vector machines have been used so far for stress detection taking advantage of electrodermal activity as a marker. Therefore, it remains as future work to implement the rest of the markers.

To sum up, this paper has described the architecture of a distributed and intelligent monitoring system to be used in health and home care. It is only the first step towards the full development of the system. As our research develops and new features are added to the system, it will cover all the functionality described in the near future.

**Author Contributions**: R.S.-R., B.G.-M., M.A.V.-Q. and A.F.-C. conceived and designed the architecture; R.S.-R., A.S.G. and A.F.-C. wrote the paper.

**Acknowledgments**: This work was partially supported by Spanish Ministerio de Ciencia, Innovación y Universidades, Agencia Estatal de Investigación (AEI)/European Regional Development Fund (FEDER, UE) under DPI2016-80894-R grant, and by Biomedical Research Networking Center in Mental Health (CIBERSAM) of the Instituto de Salud Carlos III. R.S.-R. holds BES-2017-081958 scholarship from Spanish Ministerio de Educación y Formación Profesional. B.G.-M. holds FPU16/03740 scholarship from Spanish Ministerio de Educación y Formación Profesional.

**Conflicts of Interest**: The authors declare no conflict of interest. The funding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.
Abbreviations

The following abbreviations are used in this manuscript:

- ACC: Accelerometer
- BCI: Brain-Computer Interface
- BD: Big Data
- BVP: Blood Volume Pressure
- DMS: Data Management System
- EDA: Electrodermal Activity
- EEG: Electroencephalogram
- IB: Intervals between Heartbeats
- ML: Machine Learning
- MongoDB: Humongous Database
- NoSQL: Not only SQL
- SQL: Standardised Query Language
- TEMP: Temperature

References


© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).