Asian Journal of Pharmaceutical and Clinical Research

Special Issue (April)

Online - 2455-3891
Print - 0974-2441

Full Proceeding Paper

A STUDY ON CLOUD BASED BIOSIGNALS MANAGEMENT FRAMEWORK

SIDDHARTH ARUN PANIGRAHI, UMAMAHESWARI V

Department of CSE, School of Electronics Engineering, VIT Chennai Campus, Chennai, Tamil Nadu, India. Email: asha.s@vit.ac.in

Received: 23 January 2017, Revised and Accepted: 03 March 2017

ABSTRACT

An analytical study of the complete framework for the management of biosignals is done. The framework provides for the acquisition, and storage of the biosignals, along with the associated metadata. It also provides solutions for validation, synchronization of acquired signals, thus allowing error-free signal inputs for further statistical analysis. The model comprises primarily of four layers, namely, acquisition, validation, post-processing, and statistical analysis layers. In addition, a presentation layer is also provided, wherein the appropriate end-user can use a suitable client or web service to access the results of the statistical analysis. The raw data are deliberately spilt into two: Internal data (actual signal data) and external data (metadata) and they interact only when necessary (e.g., Identifying the biosignal's origin). Microservices are used to compartmentalize the functionalities required in the system. Additional solutions to problems plaguing the present models (like cloud-upload bottleneck) are also discussed.

Keywords: Biosignals, Statistical analysis, Data management.

 DOI: http://dx.doi.org/10.22159/ajpcr.2017.v10s1.19557

INTRODUCTION

With the rapid development of sensing systems and the improvement of the quality of biosignal acquisition, the need for having a proper framework for the management and analysis of the acquire biosignals gains the utmost importance in the field of modern medical analysis. Certainly, systems have already been established for addressing the above problem, but they have their own problems nonetheless. Even if their individual problems are overlooked, new problems are encountered when cross-platform interaction occurs. This is due to the fact that every single system uses its own set of techniques and protocols and can rarely work in coherence with other systems. For example, different biosignals such as electrocardiogram (ECG) and electroencephalograph (EEG) use different storage formats, which are further classified into so many types. There has not been a single standard or protocol that has yet been defined for the storage and processing of widely used signals like ECG, EEG, let alone the minor signals that have a very specialized and limited use.

Inhomogeneity also arises when the file formats used of the biosignals are taken into consideration [2]. Ad-hoc solutions do exist, most common of which are MATLAB, XML, European data standard (EDF), etc. However, these solutions are more generic in nature, hence they cannot address to the problems of sustainability of data.

This creates lots of problems when the data have to be shared among peers and researchers when needed, thus affecting the concept of reproducibility and reusability.

Beyond the storage inhomogeneity, the creation of a standardized algorithm for the processing of the data that has been acquired a priori is also needed.

Certain systems have already been developed, including a framework that discusses about the aforementioned problems, although not without its faults.

Hence, in this paper, a discussion is made about a new, multi-layered framework, where each layer performs the following respective functions:

- Layer 1: Acquisition of the required biosignal
- Layer 2: Validation of the acquired biosignal for removal of artifacts and synchronization issues
- Layer 3: Post-processing of the validated signal wherein pre-designed algorithms are applied to extract the required information based on the input
- Layer 4: Statistical analysis of the post-processed results obtained in the previous layer to provide meaningful information to the end-user.

Digital research data are set to be strategically divided into two parts, namely, internal data information (IDI) which is the actual biosignal, and external data information (EDI), which consists of "metadata," i.e., the tags associated with the biosignal containing extra attributes such as, origin, sampling rate, and sensor used. The combination of these two, used at the right time, will provide a flexible yet not excessively redundant method to manage and analyze the biosignal. The prototype is then initialized by using a cloud and localized file server combination to provide the required amount of redundancy.

EXISTING WORK

As mentioned before, there have been few models that have been designed and currently been implemented in actual medical practice. The basis for all the models mentioned below is the open systems interconnection layer model, which provides a clear and detailed understanding of how each layer interacts with the other and what processes are to occur in the system.

Systems that presently address to the problems discussed above [13] are either purely cloud-based (distributed) in nature [12] or are based completely on a localized file-server system. The former provides remote accessibility and reduces the cost of installing an entire file server in a physical data center. The latter provides faster access of data and more data security as the data never leaves the physical system into an external client. Regardless of which approach is used, the two fundamental end-points of the system are the multi-faceted sensor modules used for biosignal acquisition and a web-based client for the end-user to access the processed data.

The attributes of the systems aforementioned, as well as their advantages and their potential drawbacks are discussed [3] as follows:

Cloudwave

Cloudwave is a bio-medical analysis system that is used specifically designed to record various patterns of ECG signals [10], analyze

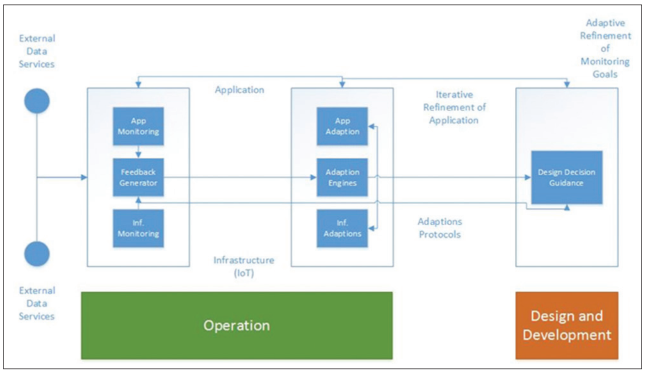

Fig. 1: Framework depiction of cloudwave system. Figure clearly mentions how the various modules of data analytics are used for data retrieved from external sources and an adaptive analytics system developed for the monitoring of the input data

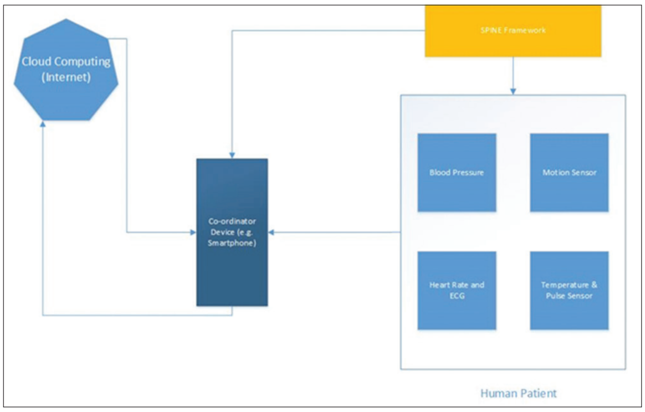

Fig. 2: Framework depiction of signal processing in node environment system. Figure clearly mentions how the various sensors are deployed on the patient as wearables and how the modules are connected to the cloud computing platform via a coordinator device, which in the most common case would be a modern smartphone, connected to the internet and logged into the cloud platform with the appropriate authentication

them in real-time using parallelized algorithms and provide a signal visualization and further querying functionalities [6]. This system is used for the identification of epileptic patients and helps in the prevention of sudden unexplained death in epilepsy.

Big-data analytics (along with Hadoop) is used in the deployment of the system. This system relies on terabytes of biosignal data and can accurately detect features and classify events based on this aforementioned data.

The results of this data analysis are stored in the EDF format, which henceforth uses an adaption of the PhysioNet technology for data storage.

To allow users or researchers to access the data a web-based portal is used, wherein users can send a signal query to the server, which in turn "serves" user accordingly, based on the level of authentication he/she possesses.

In practical scenarios, this system is extensively used in the processing of ECG signals, allowing little to moderate amount of data interfacing with mobile devices.

The system is quite specialized in detecting epileptic symptoms, yet the system fails to deliver when a broad array of biosignals are brought into the picture.

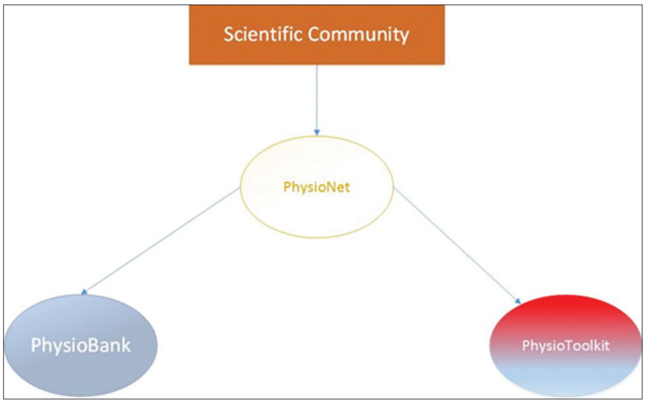

Fig. 3: Framework depiction of PhysioNet framework, which mentions the way in which the researchers are connected to the PhysioBank archive and can use the PhysioToolkit, all through a common gateway called PhysioNet

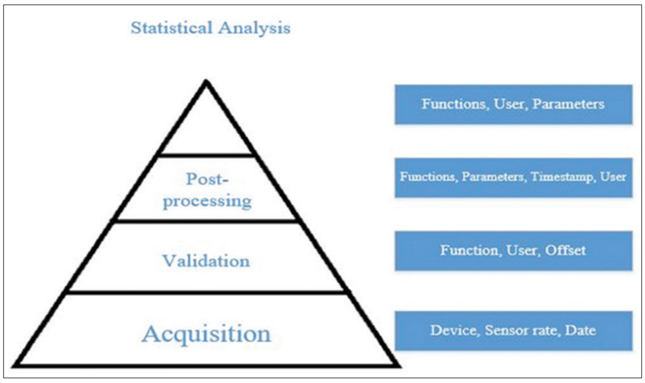

Fig. 4: Proposed data management model for internal data information and external data information in bio-signals. Each layer here represents separate challenges throughout the entire management process. After collecting the data (data acquisition), it has to be cleared of artefacts and other irregularities, and synchronized accordingly (validation) before the associated parameters can be derived after applying appropriate algorithms (post-processing) and statistically analyzed (statistics). All layers are framed by a vertical presentation layer, which enables users to interact with data in different layers and to assign the associated external data information

Signal processing in node environment (SPINE)

SPINE has been developed keeping in mind, the shortcomings of specialized systems (like Cloudwave) which fail to address to the management of data acquired from multiple sensors at a time. SPINE provides a framework of acquiring data from multiple wearable biosignal sensors and records and stores them in real-time on a smartphone. Then again, it loses the edge when providing thorough analysis of any acquired biosignal as it is primarily designed only for real-time data recording and storage.

SPINE adopts the usage of signal measuring/recording instruments to create a wireless body sensor network based monitoring application which can extend very good levels of flexibility in implementing signal processing algorithms that are used in sensor data analysis and their respective classification.

In recent developments, SPINE has been successfully collaborated with BodyCloud, which is an open source platform developed for the servicing of smart health.

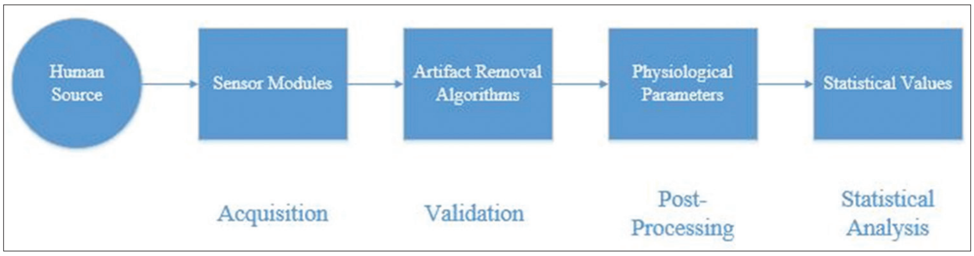

Fig. 5: Processing steps of biosignals according to the proposed layer model. Specific sensors are applied to a physiological system and their response to changes within this system is recorded and digitalized, yielding digital biosignals (in the acquisition layer). After cleaning signals from artifacts and synchronizing (validation layer), the validated signals can be used for the calculation of physiological parameters

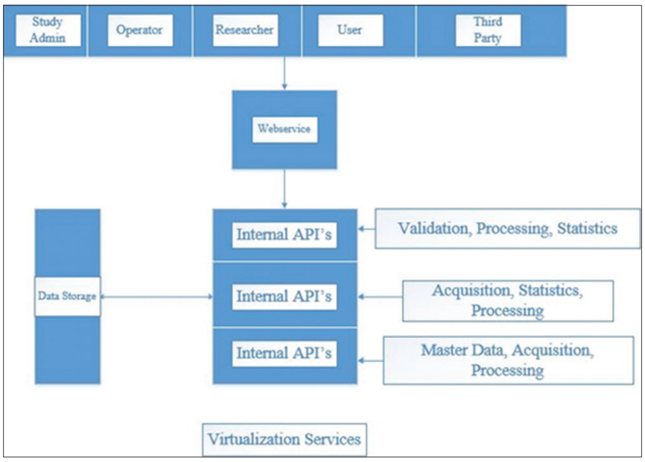

Fig. 6: System architecture [7] design showing the usage of microservices for performing the basic functions. Also shows how users can access the data from the storage, to which they have access

The latter is designed as a SaaS architecture which allows the efficient storage and management of sensor data streams, thereby allowing all its front-end applications to be cross-disciplinary and specialized in nature.

PhysioNet

PhysioNet [14] was initially a generalized biosignal analysis framework which is used to further develop specialized analysis applications upon. PhysioNet allows the integration of the storage and analysis platforms on the cloud by using virtualized databases, hence improving accessibility and flexibility. Furthermore, it provides a very accurate analysis of the biosignal as needed.

As of now, PhysioNet stands as an online forum for disseminating and exchanging recorded biosignals and facilitating cooperative analysis of the same biosignal by allowing integration of various biomedical algorithms applied worldwide.

In addition, PhysioNet has extended its online capabilities by providing what is called is PhysioBank.

PhysioBank is an extremely large and continually growing archive of the biosignals and their analysis results which obtained from PhysioNet.

The primary drawback of the approach of PhysioNet is that it requires voluminous amount of data (in range of terabytes) before it can provide any accurate result.

DISSOCIATION OF BIOSIGNAL DATA

Biosignals are often taken as a single entity when it comes to recording, storage and analysis [13]. This causes fundamental issues in accurate analysis of the same, like identification of the source of the processed signal, time of the signal acquisition, etc., Thus, the biosignal is split into two major parts, the first being IDI, which comprises of the actual signal information that is to undergo the analysis procedures and the second being EDI, which consists of the multitudes of "metadata" information associated with the biosignal such as, sensor identification, signal origin, patient details, timestamps, functions pertaining to the current biosignal.

IDI is the key area of interest as it contains vital information pertaining to the physiological event under scrutiny. For instance, IDI can be extracted from a given biosignal by applying suitable signal processing algorithms.

On the contrary, EDI contains information regarding the metadata of the previously acquired biosignal. This can be obtained directly but has to be, implicitly or explicitly, fed into the system as tags.

As much appealing it is to disregard EDI and concentrate only on IDI for further analysis, it is extremely difficult or in some cases, even impossible to correctly interpret IDI without the presence of EDI. Based on real-life encounters, it is found that a flexible and modular technique of managing EDI becomes quite important. EDI is also of utmost importance not only in interpreting the biosignal accurately but it also enables researchers to reuse the data in future projects.

Unfortunately, in present systems, IDI assessment is given primary focus and EDI management and integration with IDI is given very less attention.

To efficiently manage a system where there is a segregation of IDI and EDI, an efficient framework becomes mandatory, where there is strategic interaction of both types of data such that it can produce meaningful analysis results.

To do this, a five-layered framework [1] is proposed where each layer standardizes the state of processing that the biosignals undergo.

Acquisition layer

Measurement of the biosignal counts as the fundamental step for the processing and calculation of IDI. Depending on what type of biosignal is in the field of interest, a variety of sensors, amplifiers, transducers, and other such hardware which have to be interfaced adequately to create the desired sensing system may be used.

To maintain homogeneity of the input biosignals, any inconsistency with respect to the same should be addressed a priori [5].

EDI, in this layer, refers to the boundary conditions of the measurements.

Validation layer

The biosignal that is acquired from the sensors often contains unwanted artifacts. Artifacts are generated from either abnormal sensor movements or due to the introduction of external noise.

These artifacts must be removed from the acquired biosignal using some algorithm of preprocessing, to obtain "clean" biosignals that can be further passed on to the next stage of post-processing.

EDI in this region comprised the algorithms applied for validation and the identified noisy regions (from where artifacts are irremovable).

Post-processing layer
The core features of the validated biosignal can be obtained using appropriate algorithms, and thus the key physiological parameters can be extracted, which can safely be denoted as the IDI.

The EDI comprises the algorithms used in extracting the physiological parameters and the arguments of using the same in the given scenario.

Statistics layer
The main challenge encountered in this layer is the combination of IDI and EDI collected in the previous layers and provide the resulting meaningful analysis of the biosignal which the end-user can understand. To justify or negate the argument, statistical characteristics are evaluated.

A specialized software architecture needs to be designed that can service the layers and address to the functions as mentioned above [11], to finally allow for nearly seamless and comprehensive interaction of IDI and EDI.

The system is split primarily into two major parts, a standalone client (a local server) and a microservices server. The stand-alone client addresses to pre-processing and validation of the biosignals that are acquired and also stores the same to provide redundancy to the data.

The microservices are designed to function remotely on the cloud [8].

Now choosing microservices over a single module deployment is more feasible as it reduces the taxation on the remote cloud server. The following are the advantages of using microservices architecture against the traditional single module architecture:
a. Reduces the overall overhead generated on the cloud system by transmitting the EDI or IDI on a need-to-know basis, instead of dumping everything on the cloud which would simply slow the cloud's performance or possibly even cause a crash.
b. Allows, to a certain degree, the introduction of data abstraction into the entire model, which inherently provides certain level of privacy and data security.

The microservice architecture becomes almost mandatory as design the model to work in an EDI-IDI separation fashion is made. As data are stored at a different location and are only accessed via a linking fashion only when needed, it quite efficiently addresses to the problems encountered in the traditional approach.

This also allows various users to particularly use only those microservices that they have permission for usage of. For example, the patient needs to only view the final reports of his statistical analysis and nothing more. The researcher needs to see all the sensor data but needs not know the EDI pertaining to the identity of the patient. The doctor needs to know the statistical reports and the patient's clinical history but needs not know the details of the sensor's acquisition, properties, etc.

POTENTIAL GAPS AND FUTURE WORK

Despite the elaborate and efficient design and description of the system, there certainly are a few flaws in the approach which needs to be addressed to make design even more efficient. The system requires data to be uploaded onto a cloud platform, or at least the part of it that is required for the post-processing and statistical analysis of the biosignal acquired. This can be done either on platforms like ThingSpeak or Microsoft Azure or Amazon AWS (although in this case, the usage of Python along with MATLAB functions makes ThingSpeak the feasible option). As these cloud services have their own (i.e. external) database management solutions, uploading the EDI and IDI of a patient can cause privacy issues and might seem unethical too. Hence, the data transmitted should be, at least in some form, encrypted before transmission.

Along with the privacy risks that are generated while uploading data onto the cloud, when uploading the IDI specifically, the process cannot be seamless as there will always be low data intake bandwidth when compare to the upload rate. This will cause some serious bottlenecking and possible data loss. To avoid this, the traditional approach can be extended to use the TCP protocol to work along with user datagram protocol (UDP) protocol [9]. The advantage of doing this is that TCP has a property of retrying the transmission of data packets until acknowledged, which can address the error-free transmission, and UDP protocol has a fire-and-forget property which can be used to improve the overall transmission speed.

ACKNOWLEDGMENT

I would like to acknowledge and appreciate the guidance of my mentors and supervisors Prof. Umamaheshwari V and Prof. Venkatsubramanian K of School of Electronics Engineering, Vellore Institute of Technology, Chennai, also for their constant support and vigilance which pushed me further into exploring into the problem at hand.

REFERENCES

1. Thürk F, Kampusch S, Kaniusas E. Management framework for biosignals in biomedical studies: From study design to data statistics. IEEE Trans Instrum Meas 2016;65(4):776-82.
2. Reilly S, Schallier W, Schrimpf S, Smit E, Wilkinson M. Report on Integration of Data and Publications. Opportunities Data Exchange, Technical Report; 2011.
3. Anderson NR, Lee ES, Brockenbrough JS, Minie ME, Fuller S, Brinkley J, *et al*. Issues in biomedical research data management and analysis: Needs and barriers. J Am Med Inform Assoc 2007;14(4):478-88.
4. Hey AJ, Trefethen AE. The data deluge: An e-science perspective. In: Grid Computing: Making the Global Infrastructure a Reality. New York, NY, USA: Wiley; 2003.
5. Kaniusas E. Biomedical Signals and Sensors I: Linking Physiological Phenomena and Biosignals. Berlin, Germany: Springer-Verlag; 2012.
6. Trigo JD, Alesanco A, Martínez I, García J. A review on digital ECG formats and the relationships between them. IEEE Trans Inf Technol Biomed 2012;16(3):432-44.
7. Thürk F, Kampusch S, Kaniusas E. Strategic framework for management of hybrid biosignals from study design to statistics. In: Proceeding Sensors Applications Symposium; 2015. p. 85-9.
8. Newman S. Building Microservices. Sebastopol, CA, USA: O'ReillyMedia, Inc.; 2015.
9. Zimmermann H. OSI reference model-the ISO model of architecture for open systems interconnection. IEEE Trans Commun 1980;28(4):425-32.
10. Pandey S, Voorsluys W, Niu S, Khandoker A, Buyya R. An autonomic cloud environment for hosting ECG data analysis services. Future Gener Comput Syst 2012;28(1):147-54.
11. Fielding RT. Architectural Styles and the Design of Network-based Software Architectures, Ph.D. Dissertation. Irvine, CA, USA: University California, Department of Information and Computer Science; 2000.
12. Rosenthal A, Mork P, Li MH, Stanford J, Koester D, Reynolds P. Cloud computing: A new business paradigm for biomedical information sharing. J Biomed Inform 2010;43(2):342-53.
13. Chen J, Qian F, Yan W, Shen B. Translational biomedical infor-matics in the cloud: Present and future. BioMed Res Int 2013;2013:8.
14. Moody GB, Mark RG, Goldberger AL. PhysioNet: A webbased resource for the study of physiologic signals. IEEE Eng Med Biol Mag 2001;20(3):70-5.