Database of Biomedical Signals of Patients with Epilepsy

María Fernanda Pineda, Héctor Hurtarte, Luis Alberto Rivera
Department of Electronics, Mechatronics and Biomedical Engineering
Universidad del Valle de Guatemala
Guatemala, Guatemala
{pin16021, hahurtarte, larivera}@uvg.edu.gt

***Abstract*—Databases are a useful resource in terms of management and classification of data, they allow to store them in a single place in an ordered way. Their advantages range from quick access to stored information to redundancy management of data and security. The databases, not only of electroencephalographic (EEG) signals but biomedical signals in general, have allowed advances in research, since they allow for analysis without having to go to the hospitals to obtain them. The goal of this work was to develop of a relational database of biomedical signals from patients with epilepsy from the Centro de Epilepsia y Neurocirugía Funcional, HUMANA. The pieces of information stored in the database can be useful for researchers in analysis and classification tasks. Furthermore, the results of tests performed on the signals can be stored back as binary files and other formats. A link between the database and Matlab was also developed in a toolbox. The toolbox includes several interfaces for interacting with the database, and for processing signals and applying machine learning algorithms to those signals. It is possible to add new patients to the database, retrieve information from the patients, add analysis results, visualize those results, and more.**

***Index Terms*—database, eeg, relational database**

I. Introduction

The functioning of the human body is frequently associated with electrical, chemical or acoustic signals, and they are carriers of information that describe brain, cardiac and muscular activity. Each type of signals can be interpreted differently, since they have characteristics and patterns that allow clinic diagnosis. Biomedical signals have played an important role in continuous research on the human body [1]. As an example, an electroencephalogram (EEG) is a test that allows studying the electrical activity of the brain and allows the diagnosis of diseases such as epilepsy.

Epilepsy is one of the oldest known diseases, surrounded by ignorance, fears and social esteem and influences the quality of life of the patient and their relatives. Around the world, some 50 million people have epilepsy, making it, not only one of the most common neurological disorders, but also a common condition such as diabetes or high blood pressure [2].

Databases are a set of information belonging to the same context, ordered in a systematic way for later retrieval and analysis [3]. These arise from the need to be able to store information and go to it later. Biomedical signal databases are a useful key when it comes to research, since they allow the testing and validation of algorithms performed for signal processing and obtaining relevant characteristics from them [4].

The goal of this work is the development of a relational database of biomedical signals from patients with epilepsy from the Centro de Epilepsia y Neurocirugía Funcional, HUMANA, as well as its interaction with the Matlab for the development of a software tool or toolbox in collaboration with María Angulo. The toolbox includes several interfaces for interacting with the database, and for processing signals and applying machine learning algorithms to those signals. It is possible to add new patients to the database, retrieve information from the patients, add analysis results, visualize those results, and more.

The development of the toolbox aims to present the importance of implementation of the databases in the research, since they support in the validation of processing algorithms and in obtaining features for anomaly detection tin the functioning of the human body.

II. Background

A. Relational Databases

A relational database is a type of database that stores and provides access to data entries that are related to one another. Relational databases are based on the relational model, an intuitive and straightforward way of representing data in tables [5]. The standard user and application program interface (API) of a relational database is the Structured Query Language (SQL). SQL statements are used both for interactive queries for information from a relational database and for gathering data for reports [6].

The relational model is best for maintaining data consistency across applications and database copies. Relational databases excel at this type of data consistency, ensuring that multiple instances of a database have the same data all the time [5]. Defined data integrity rules must be followed to ensure the relational database is accurate and accessible [6]. Data integrity is the completeness, accuracy and overall consistency of the data [7]. Further benefits of relational databases include [8]:

- Scalability. New data may be added independent of existing records.
- Simplicity. Complex queries are easy for users to perform with SQL.

- Data Accuracy. Normalization procedures eliminate design anomalies.
- Data Integrity. Strong data typing and validity checks ensure accuracy and consistency.
- Security. Data in tables within a RDBMS can limit access to specific users.
- Collaboration. Multiple users can access the same database concurrently.

In relational databases, each table has a unique identifier, or PRIMARY KEY, which identifies the information in the table, and each row contains a unique instance of data for the categories defined by the columns [8]. The logical connection between different tables can then be established with the use of FOREIGN KEYS - a field in a table that connects to the primary key data of another table [8].

1) Entity-Relationship Diagrams: An entity-relationship diagram (ERD) is a type of flowchart that illustrates how "entities", such as people, objects, or concepts, relate to each other within a system. ER diagrams are often used to design or debug relational databases in the engineering fields of text-software, business information systems, education, research, etc.; and employ a defined set of symbols, such as rectangles, diamonds, ovals, and connecting lines to represent the inter-connection of entities, relationships, and their attributes [9].

B. Some Biomedical Signals Databases

PhysioNet is an alliance of U.S. industry partners established in 1999 that seeks to provide free access to large collections of physiological signals and corresponding open source software. They have financial support from the National Institute of Health *NIBIB* and *NIGMS*. In addition, about 800 academics articles are published each year with support of the data from PhysioNet [10].

PhysioBank is a large and growing archive of well-characterized digital records of physiological signals and related data for use by the biomedical research community. PhysioBank currently includes multi-parameter cardiopulmonary, neuronal, and biomedical signal databases of healthy subjects and patients with variety of conditions with important public health implications. It now contains more than 75 databases that can be freely downloaded [11].

PsycTherapy is an American Psychologial Association database containing approximately 300 psychological therapy demonstration vidios showing clinical work with individuals, couples and families. It contains spontaneous and impromptu therapy sessions recorded over the past 10 years. It offers the possibility of tagging or commenting on segments in each client's therapy demonstration, saving or sharing personal playslists [12].

A research developed in the neuroscience center of Havana, Cuba, called "Development of a database system for electrophisiological signals", was published in the PubMed database of the United States. It consists on a database system to store recorded electrophisiological signals with different evoked potentials. The system has several functions such as the storage of the recording parameters, the recovery and backup of information and the transference from one database to another. All this set allows to facilitate the analysis of evoked potentials either for clinical or research purposes or as a clinical information management tool [13].

III. Database Implementation

The *phpMyAdmin* software was used for the creation of the database within the *MySQL* environment. The structure of the database is illustrated in the EDR in Figure 1, which consists of three tables: a main table with patient information, one with medical test information and another with signal data. The biomedical signals were obtained from patients with epilepsy from HUMANA. In addition to the signals, qualitative and quantitative information was obtained from each patient, which allowed proper classification and organization of the data. In order to protect the patients' privacy, no names or personal information was collected.

The patient information stored in the main table includes an ID, date of birth, gender, personal pathology history, prescription diagnosis, whether you have epilepsy and seizure types. This information is important for doctors to analyze an EEG. The table with the medical test information includes patient ID, test ID, test date, start time, duration, frequency, number of channels and the test in a binary file. The stored information, such as frequency, number of channels and duration, is used for signal processing. Finally, the table with the signal data includes the test ID and 35 fields corresponding to the channels of the stored test. It is important to mention that each test has a different number of channels.

It can be seen that in the test table there is a field that stores the EEGs. The test field is a *longblob* type, which can store a certain amount of data as binary strings [14]. Within the database, this field stores the data vector of the signal corresponding to the test.

The tables are related by using foreign keys. The patient ID is the primary key in the patient table while it is the foreign key in the test table; likewise the test ID is the primary key in the test table and the foreign key in the test data table.

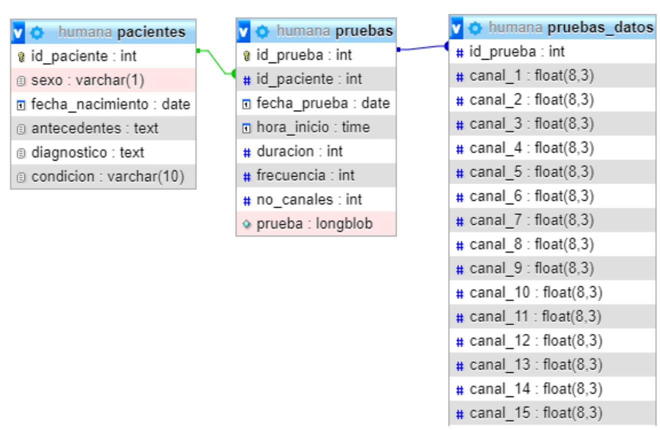

Fig. 1. Entity-Relationship Diagram of the Database.

This was done to store multiple tests that correspond to their respective patients.

IV. Toolbox Implementation

In order to achieve the goal of the work, the general architecture of the toolbox was defined as shown in Figure 2, which contemplates the visualization and writing between the user and the database. The toolbox was developed in *App Designer* from Matlab, making a connection called "humana" through an ODBC driver to acces the database and the are related by means of querys. The toolbox can be installed on any computer with *MySQL* as database management system. The analysis and processing part was developed by María Angulo in [15].

To achieve a better interaction with the toolbox, a configuration file was implemented. The configuration file contains information on the name of the database connection and the user name and password for the use of *MySQL*. Also, corresponding validations were made to guarantee the connection to the database. The validations include having a connection with the name "humana" and that the tables described in the previous section are created.

Different interfaces were designed: the main page, adding a new patient, consulting patients and their tests, visualizing tests and analysis. The database connection is made on the home page and then the other interfaces can be accessed as shown in Figure 3 and thus achieving the connection in each of them to the database, with the exception of the analysis interface, since a local file can be selected for proper processing. If an error occurs in the connection, messages are displayed to the user and the other interfaces cannot be accessed. It is important to mention that all windows can return to the home page and the user can continue to use the toolbox.

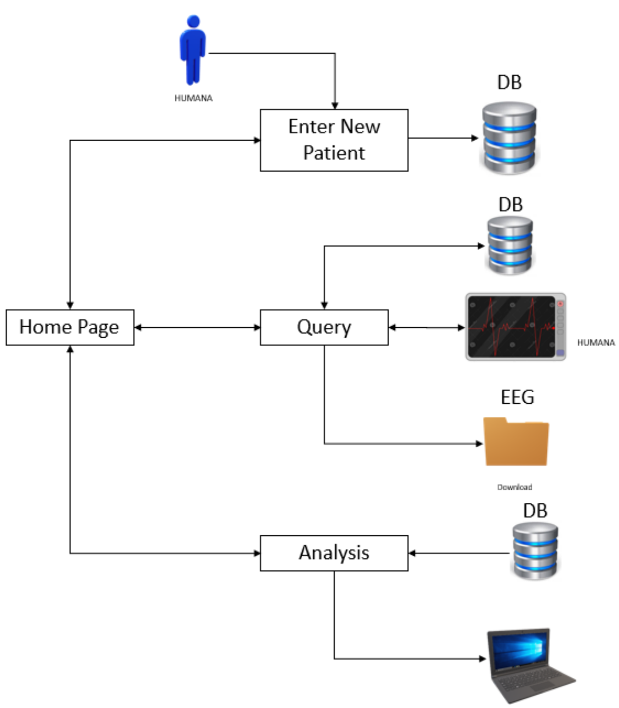

Fig. 2. General Architecture of the Toolbox.

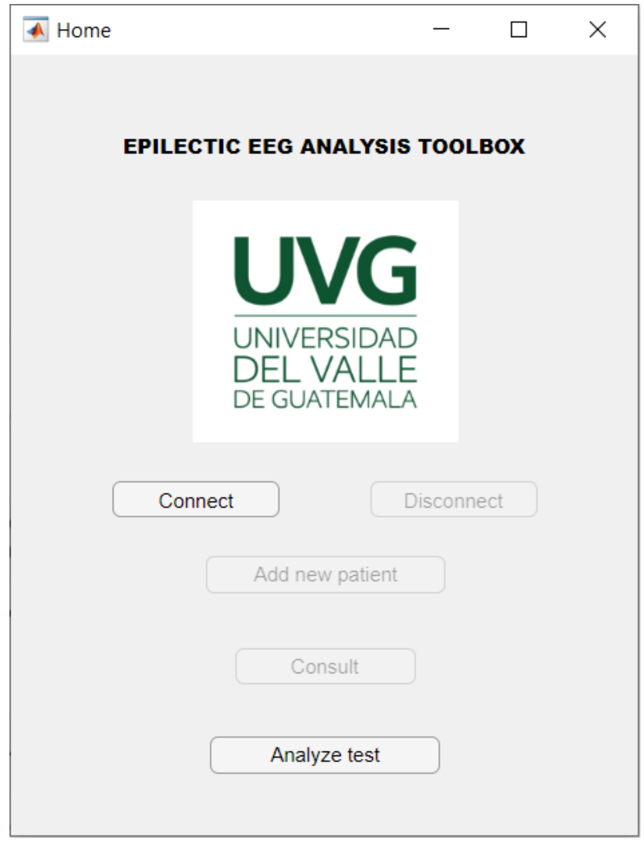

Fig. 3. Homepage of the Toolbox.

The New Patient Entry window asks for the patient's ID, gender and date of birth, as well as the patient's history and diagnostic information. The fields were validated in such a way that the formats match those accepted within the database. In order to take advantage of the *App Designer* and its variety of objects, the design is based on text fields and a calendar to select the patient's date of birth. Buttons were included to save information, new entry and return to the main screen, as well as validations that must be entered for all required fields. In addition, messages of "patient code already existing within the database" and "patient successfully saved" were included.

The query window can be considered as the most complete, since it has the options to search, display information, add a new file and download existing tests in the database. This window is linked to all the tables of the database, since all its fields correspond to the existing columns within the database. The design of the window consist of text fields and tables for displaying information. The user must enter the ID of the patient he wants to consult and, if it exists, his data and the tests of this patient that are already stored are shown.

The display table lists the existing tests for the consulted patient and their respective information. If the user wishes to store a new file, the select button is used, which opens a navigation window within the computer in which the user can review folders until the EDF corresponding to the patient is found. This button has the validation that only that type of file can be stored within the database. Additionally, there is the option of downloading the stored tests. There is the option to download all files, a range or a specific file, this with the purpose of contributing to development of algorithms that want to use the stored information.

In the analysis window, a display and selection window is shown, in which the user can select a local file or one stored in the database. The display button allows to observe by channel the signal of the selected file. Additionally, the most common characteristics of the signals.

From the display window, you can redirect to the feature extraction window to perform the respective signal analysis. In this window it is possible to generate a feature vector according to options chosen by the user and then select the classifier to be used. The results of the classifiers can be displayed or stored in "mat" files and new feature vectors can be loaded for further analysis. It is important to mention that all these algorithms were developed by María Angulo in [15].

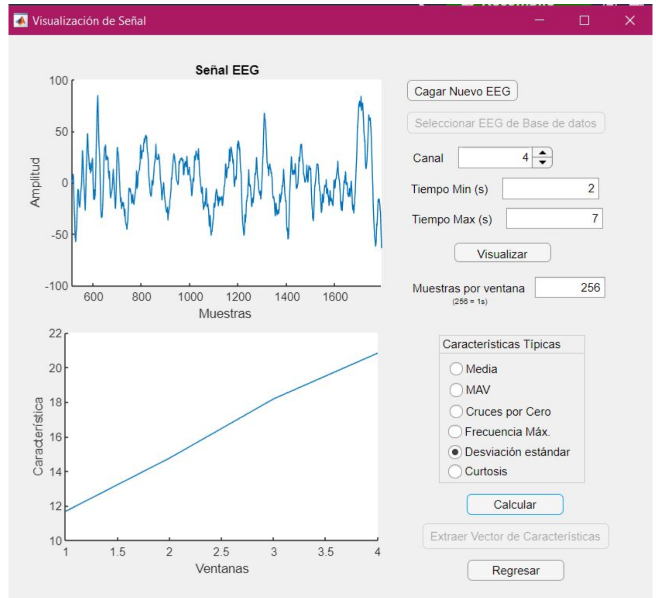

Fig. 4. Query and Visualization

V. Discussion and Conclusions

The Centro de Epilepsia y Neurocirugía Funcional, HUMANA, provided the tests used in this work for patients with epilepsy who are admitted. The signals were obtained in EDF format, of which the recordings were approximately one hour long. Additionally, HUMANA indicated that the tests correspond to patients who suffer from epilepsy, who had some seizures during the test or who suffered from epilepsy and are now under surveillance.

In order to read the EDFs from Mtlab, the *edfread* function [16] was used. This function opens EDF files and returns two variables with test data. One variable corresponds to a structure, which has information from the test performed on the patients and information from the signals. The other variable corresponds to an $n \times m$ matrix that stores the signals, where n corresponds to the number of channels and m to the number of samples in the file.

To store the signal within the database, it was necessary to use the *serialize* function [17]. This function is responsible for encoding a vector of any type to a sequence of unsigned 8-bit integers. This is done in order to store the data in the binary field. To download the tests already stored was used the *deserialize* function [17], which is responsible for decoding the information.

As one of the results of the implementation of the toolbox, figure 4 was obtained. In the figure it can be seen that the integration of data query and analysis and processing algorithms was achieved. From the toolbox you can select which test you want to analyze after a query of the existing tests. Similarly, a local EEG can be selected.

One of the obstacles encountered during storage to the database was the size of the data. To solve this it was necessary to change the parameter *max_allowed_packet*, which refers to the maximum data size to be stored in a transaction in *MySQ*L [18]. By default, 4 MB are allowed per transaction, so the parameter was set to 750 MB. Another obstacle was the time of storage data by data to the database, since being very large files the time of sending data increases.

Finally, it was possible to implement a relational database capable of storing relevant information on patients with epilepsy and EEG. The type of database developed allowed relating a patient to multiple tests and a test to multiple data and channels. The toolbox was able to send information to the database and retrieve it for further processing and analysis.

Of the obstacles encountered, for future work it is recommended that data storage be worked on in the background of the toolbox. Additionally, include notifications and warnings to the user in case the maximum file size of maximum transaction size is exceeded. Another recommendation is to include, within the database, a column that identifies the type of biomedical signal.

References

[1] A. B. Valiente, "Detección de crisis epilépticas en el dominio temporal a partir del eeg mediante svms," Universidad Carlos III de Madrid, Tech. Rep., 2014.
[2] (2019, jun) Epilepsia. [Online]. Available: https://www.who.int/es/news-room/fact-sheets/detail/epilepsy
[3] D. P. Valdés. (2007, oct) ¿qué son las bases de datos? [Online]. Available: http://www.maestrosdelweb.com/que-son-las-bases-de-datos/
[4] Y. V. P. Gutiérrez, "Base de datos de señales electrofisiológicas." Escuela de Ingeniería de Antioquia, Tech. Rep., 2013.
[5] "¿qué es una base de datos relacional?" [Online]. Available: https://www.oracle.com/ar/database/what-is-a-relational-database/
[6] (2020, april) Relational database. [Online]. Available: https://searchdatamanagement.techtarget.com/definition/relational-database
[7] A. Emaz and P.-R. Una, "¿qué es una base de datos relacional?" 1998. [Online]. Available: https://aws.amazon.com/es/relational-database/
[8] "What is a relational database? definition and faqs." [Online]. Available: https://www.omnisci.com/technical-glossary/relational-database

[9] Lucidchart. Qué es un diagrama entidad-relación. [Online]. Available: https://www.lucidchart.com/pages/es/que-es-un-diagrama-entidad-relacion
[10] PhysioNet. (1999) Physionet - the research resourse for complex physiologic signals. [Online]. Available: http://www.physionet.org/
[11] (2016, aug) Physiobank. [Online]. Available: https://archive.physionet.org/physiobank/
[12] A. P. Association. (2012) Psyctherapy. [Online]. Available: http://www.apa.org/pubs/databases/psyctherapy/index.aspx
[13] P. A. De Armas JL, "Development of a database system for electrophysiological signals," 1995, pubMed, 1-2.
[14] "Mysql 8.0 reference manual :: 11.3.4 the blob and text types." [Online]. Available: https://dev.mysql.com/doc/refman/8.0/en/blob.html
[15] M. J. Angulo, "Análisis y reconocimiento de patrones de señales biomédicas de pacientes con epilepsia." Universidad del Valle de Guatemala, Tech. Rep., 2020.
[16] B. Shoelson. (2020) edfread. [Online]. Available: https://la.mathworks.com/matlabcentral/fileexchange/31900-edfread
[17] Tim. (2020) Serialize/deserialize. [Online]. Available: https://la.mathworks.com/matlabcentral/fileexchange/29457-serialize-deserialize
[18] "Mysql 8.0 reference manual :: B.3.2.8 packet too large." [Online]. Available: https://dev.mysql.com/doc/refman/8.0/en/packet-too-large.html: :text=The largest possible packet that,server or client is 1GB.