
Epileptic EEG Analysis Toolbox: Manual de Usuario

María Angulo Tijerino (ang16508@uvg.edu.gt)
Fernanda Pineda Esmieu (pin16021@uvg.edu.gt)

Noviembre, 2020

Descripción

El toolbox Epileptic EEG Analysis es una herramienta diseñada para facilitar el análisis e implementación de algoritmos de aprendizaje automático en señales EEG con el fin de caracterizar las señales EEG de pacientes con epilepsia y detectar la presencia de crisis en las mismas. Epileptic EEG Analysis Toolbox esta dirigido a estudiantes que buscan validar sistemas de detección automática de crisis epilépticas. Busca optimizar y facilitar el proceso de diagnóstico del personal médico, mediante una interfaz gráfica que permita estudiar detalladamente las señales EEG y generar resultados. Puede descargarse de forma gratuita en el siguiente link, <https://github.com/larivera-UVG/Datos-Epilepsia>. Para su uso, es necesaria la instalación del software MATLAB.

Contents

I	Epileptic EEG Analysis Toolbox: Contenido	2
1.1	Visualización de la señal EEG	2
1.2	Extracción de Características	2
1.3	Algoritmos de Clasificación	2
1.4	Conexión con Base de Datos HUMANA	2
2	Instalación del Toolbox Epileptic EEG Analysis	3
2.1	Requerimientos	3
3	Ejemplos de uso	3
3.1	Selección de Data	4
3.2	Extraer Vector de Características	4
3.3	Implementar Clasificadores	8
3.4	Exportación y Generación de Resultados	9

Guía Rápida para Uso de Toolbox

I Epileptic EEG Analysis Toolbox: Contenido

I.1 Visualización de la señal EEG

La interfaz de visualización permite graficar la señal EEG seleccionada en formato EDF. El usuario tiene opción de cargar la señal directamente de un archivo local de su computador, o seleccionar una señal existente en la base de datos HUMANA. La ventana de visualización incluye una gráfica de amplitud vs tiempo donde se muestra el canal especificado durante el período de tiempo seleccionado. Una gráfica adicional para observar características típicas del segmento de señal visualizado y un menú de opciones de las características disponibles en el toolbox.

I.2 Extracción de Características

En esta ventana el usuario puede seleccionar del menú de opciones las características que desea extraer de la señal previamente cargada. Adicionalmente, se debe especificar el tamaño de ventana, debe escogerse el número de canales a analizar y especificar los mismos y finalmente el tipo de análisis a realizar: análisis directo y análisis wavelet. Ambos análisis siguen la metodología propuesta en la tesis de investigación "Análisis y Reconocimiento de Patrones de Señales Biomédicas de Pacientes con Epilepsia" [1]. El análisis directo consiste en extraer características en el dominio del tiempo directamente de los segmentos de la señal EEG. El análisis wavelet, como su nombre lo indica, utiliza la transformada discreta de wavelet para descomponer la señal EEG en sub-bandas de coeficientes wavelets y de estos coeficientes se extraen las características (dominio tiempo-escala).

Como resultado, la herramienta entrega un archivo de formato MAT con el vector de características y un vector de clases generado automáticamente con "ceros" y "unos" balanceados según el tamaño del vector de característica resultante. El vector de clases se genera tomando en consideración que la señal EEG cargada se encuentra ordenada de forma que la primera mitad de las muestras de la señal corresponden a un tipo de registro y la siguiente mitad de las muestras corresponden a otro tipo de registro. Es decir que la cantidad de "ceros" presentes en el vector de clase es igual a la primera mitad de las muestras de la señal y la cantidad de "unos" equivalen a la mitad restante de las muestras de la señal.

En caso que la señal EEG no se encuentre ordenada de esta forma, el usuario debe generar manualmente el vector de clases y tendrá opción de cargarlo posteriormente en la ventana de clasificación del toolbox. El botón generar permite guardar el archivo de formato MAT que contiene el vector de características y de clases en la ubicación que el usuario seleccione.

I.3 Algoritmos de Clasificación

La herramienta utiliza los resultados generados en la ventana extracción de características para implementar el clasificador seleccionado por el usuario: máquinas de vectores de soporte (SVM) y redes neuronales artificiales (ANN). Si se desean cargar vectores de características y clases externos el usuario debe cargar un archivo MAT que contenga el vector de características y el vector de clases.

Para el clasificador SVM, el toolbox utiliza validación cruzada. El usuario debe seleccionar el número de particiones previo a cargar los datos para dividir el vector de características en muestras para entrenamiento y prueba del clasificador. El clasificador de redes neuronales permite seleccionar los porcentajes de entrenamiento, validación y testeo. Adicionalmente permite escoger el algoritmo de entrenamiento de la red neuronal y el número de capas ocultas.

Para generar los resultados es necesario indicar el tipo de análisis utilizado para la extracción de características (wavelet o directo). Los resultados de la clasificación se presentan mediante el toolbox deep learning de MATLAB, donde el usuario puede visualizar la estructura de la red, las matrices de confusión e indicadores típicos para evaluar desempeño del algoritmo. Para el caso del clasificador SVM, se presenta un resumen en la parte inferior de los resultados obtenidos y se despliega la matriz de confusión. En ambos casos puede guardarse los resultados en un archivo MAT en la ruta especificada por el usuario.

I.4 Conexión con Base de Datos HUMANA

El toolbox EEG Analysis se incluye en la interfaz de la base de datos HUMANA (figura 5), creada por María Fernanda Pineda. Esta base de datos incluye señales biomédicas de pacientes con epilepsia del Centro de Epilepsia y Neurocirugía Funcional HUMANA, Guatemala. De cada paciente se obtuvieron variables cualitativas y cuantitativas que permitieron la clasificación y organización de datos, así como también las señales biomédicas de cada uno de los pacientes. Es importante mencionar que, para resguardar la privacidad de los pacientes, no se obtuvieron nombres u otra información que pudiera revelar su identificación.

La base de datos phpMyAdmin se utilizó para la creación de la base de datos relacional dentro del entorno de MySQL. La base de datos desarrollada consta de diferentes tablas en las que se puede almacenar información importante del paciente que un investigador necesita para clasificar y utilizar las señales almacenadas, información de las pruebas para realizar el análisis correspondiente para cada una de ellas, así como los datos de las pruebas almacenadas como archivos binarios y dato por dato de cada prueba.

2 Instalación del Toolbox Epileptic EEG Analysis

2.1 Requerimientos

Para la correcta instalación del toolbox Epileptic EEG Analysis es necesario contar con la plataforma MATLAB. Es compatible a partir de las versiones R2018. Adicionalmente debe descargarse la carpeta “Epileptic EEG Analysis” contenida en el repositorio Datos-Epilepsia de dominio público. Los archivos contenidos en la carpeta indicada deben permanecer juntos para el correcto funcionamiento de la herramienta.

El toolbox está diseñado para clasificación binaria, registros con crisis y sin crisis epilépticas. Tomando en consideración señales EEG balanceadas (mismo número de muestras con crisis y sin crisis). Por lo cual, se aconseja seleccionar y ordenar la data previamente al realizar la clasificación. Debe asegurarse el formato EDF para la correcta lectura de la señal EEG.

3 Ejemplos de uso

A continuación se describen ejemplos de como interactuar con el toolbox Epileptic EEG Analysis. El menú principal, Figura [1], permite la conexión y consulta de la base de datos HUMANA. En caso no se desee utilizar esta base de datos debe seleccionarse la opción analizar prueba y seleccionar un archivo EDF con la señal EEG que se desea analizar.



Figure 1: Menú principal toolbox Epileptic EEG Analysis

3.1 Selección de Data

Al seleccionar la opción analizar data, Figura 2, el usuario tendrá opción de cargar la señal EEG de un archivo local de su computador o seleccionar una prueba existente en la base de datos HUMANA. Como se menciona previamente, el formato de las señales debe ser de tipo EDF.



Figure 2: Ventana visualización de señal

Posterior a seleccionar la data, el usuario debe seleccionar el canal y el período de tiempo en segundos previo a seleccionar la opción visualizar. El período máx en segundos esta dado según la longitud de la prueba seleccionada, si el toolbox detecta que el usuario ingresa una cantidad mayor a los segundos contenidos en la prueba autorellenará la casilla tiempo máx y mostrará únicamente la longitud total de la prueba ingresada.

Si no se desea visualizar la señal, puede seleccionarse directamente la opción extraer vector de características lo cual permite al usuario generar un vector de características y de clases como se especificará en la siguiente sección.

El usuario debe presionar el botón visualizar siempre que realice un cambio en las opciones: cambio de canal o cambio de período de tiempo. Adicionalmente esta ventana incluye un menú de características típicas para señales EEG de pacientes con epilepsia. El usuario puede seleccionar una opción a la vez y el botón calcular extraerá la característica seleccionada en el período de tiempo seleccionado y se desplegará en la gráfica inferior.

Esta gráfica necesita ser actualizada presionando el botón visualizar siempre que se realice un cambio de canal o período de tiempo y posteriormente el botón calcular, si se realiza un cambio de característica,

En la Figura 3, se describe un ejemplo del canal 4 de una señal EEG cargada de un archivo local, en el período de tiempo de 2-7 segundos y se calculan la desviación estándar en ventanas de 1 segundo (cada 256 muestras). Ambas gráficas pueden ser exportadas como imagen utilizando la opción exportar al colocar el mouse sobre cualquiera de ellas.

Se recomienda utilizar períodos de tiempos cortos en esta ventana para disminuir el tiempo de procesamiento. El tamaño de muestras esta relacionado a la frecuencia de muestreo de la señal cargada, por defecto el toolbox considera una frecuencia de 256 Hz. Sin embargo esto puede modificarse en caso que se desee extraer las características en ventanas menores a 1s o cuando la señal cargada es muestreada a una frecuencia diferente. Para calcular el número de muestras por ventana puede utilizarse la siguiente ecuación:

$$Muestras\ por\ ventana = T_{segundos} \times F_{muestreo} \quad (1)$$

$$Muestras\ por\ ventana = 1(s) \times 256(Hz) = 256$$

3.2 Extraer Vector de Características

La ventana de extracción de características, Figura 4, contiene dos tipos de análisis para extraer características de la señal cargada, como se describe en la sección 1.2. El análisis directo (parte superior) incluye 4 opciones de características del dominio del tiempo: Desviación estándar, curtosis, cruces por cero y valor medio absoluto.

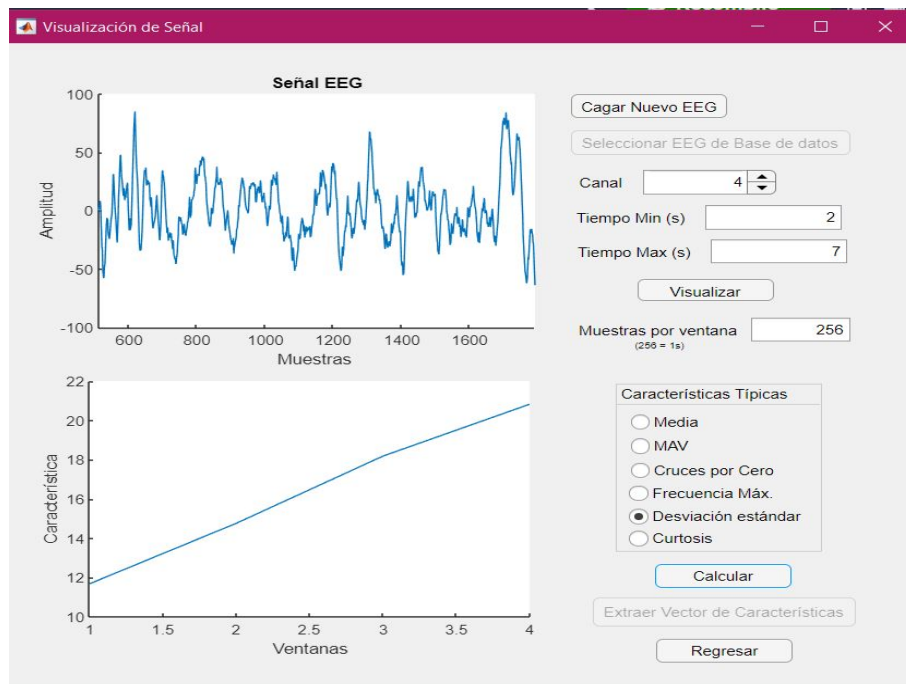


Figure 3: Ejemplo de uso, ventana visualización

Adicionalmente permite analizar 1- 4 canales simultáneamente e incluye la opción de analizar todos los canales simultáneamente. Sin embargo, el usuario debe tomar en consideración el aumento de tiempo en generar un resultado debido a la gran carga computacional.

El toolbox utiliza por defecto un tamaño de ventana de 2500 muestras. Tomando en consideración una frecuencia de muestreo típica en señales EEG, 2500 muestras representan una ventana de aproximadamente 10 segundos. El tamaño de ventana puede ser modificado por el usuario. A menor tamaño de ventana aumentará el tiempo de procesamiento y la carga computacional. Los botones generar, extraen las características seleccionadas según el análisis

The screenshot shows a window titled "GENERAR VECTOR DE CARACTERÍSTICAS". It has two main sections: "ANÁLISIS DIRECTO" and "ANÁLISIS WAVELET". In the "ANÁLISIS DIRECTO" section, there are checkboxes for "DESVIACIÓN ESTÁNDAR", "CURTOSIS", "CRUCES POR CERO", and "VALOR MEDIO ABSOLUTO", a "NÚMERO DE CANALES" dropdown set to 2, a "MUESTRAS POR VENTANA" input set to 2500, and a "CANALES A ANALIZAR" input set to 1,2. A "GENERAR" button is at the bottom right of this section. The "ANÁLISIS WAVELET" section is titled "CARACTERÍSTICAS WAVELET" and includes "[7 NIVELES DE DESCOMPOSICIÓN]", a "WAVELET MADRE" dropdown set to "db3", a "SELECCIONAR CANAL" dropdown set to 1, and a "MUESTRAS POR VENTANA" input set to 2500. There are checkboxes for "POTENCIA", "MEDIA", "DESVIACIÓN", "CURTOSIS", "ASIMETRÍA", and "CRUCES POR CERO". A "CALCULAR PARA" section has radio buttons for "TODAS LAS BANDAS" (selected) and "BANDA ESPECÍFICA", with a "GAMMA" dropdown. A "GENERAR" button is at the bottom right. At the very bottom, there are buttons for "CLASIFICADOR SVM" and "CLASIFICADOR RED NEURONAL". A small note at the bottom states: "* EL VECTOR DE ETIQUETAS ASIGNA AUTOMÁTICAMENTE AL 50% DE LOS PRIMEROS DATOS COMO CLASE '0', Y AL 50% DE LOS DATOS RESTANTES COMO CLASE '1'. MODIFICAR SEGUN NECESIDAD".

Figure 4: Ventana extracción de características

escogido y guarda los resultados en un archivo M compatible con MATLAB. Este archivo contiene dos variables, un

vector de clases y una estructura que contiene resultados de las características y otros aspectos importantes para el usuario. Ambos archivos se describen en las figura 5, 6 y se detallan en las tablas 3.2 y 3.2.

Workspace	
Name ^	Value
Etiquetas	368x1 double
Features_result	1x1 struct

Variables - Features_result	
+2	Features_result x Features_result.Ca
1x1 struct with 6 fields	
Field ^	Value
Vector_Cara...	368x4 double
Caracteristic...	','Curtosis','CrucesPorCero','
canales	[1,2]
data	921600x2 double
FMuestreo	256
Ventana	2500

Figure 5: Ejemplo Resultado Análisis Directo

Workspace	
Name ^	Value
Etiquetas	368x1 double
Features_Wavelet_result	1x1 struct

Variables - Features_Wavelet_result	
Features_Wavelet_result x	
1x1 struct with 7 fields	
Field ^	Value
VectorCarac...	368x7 double
Caracteristic...	','Media','
Canal	1
fbanda	'Theta, Theta/Beta, Beta/Alpha, Alpha/Gamma, Gamma/Delta, Delta, Aprox'
Coeficientes...	1277x1 double
FMuestreo	256
Ventana	2500

Figure 6: Ejemplo Resultado Análisis Wavelet

Nombre	Tipo	Contenido
Etiquetas	vector	Vector de clases, 0 y 1 representan tipos de registro "Con crisis" y "Sin crisis". La dimensión del vector quivale al número total de muestras de la señal dividido el tamaño de ventana escogido.
Vector_Caracteristicas	vector	Vector resultado de características escogidas. Las columnas representan las características según el número de canales escogido. El orden de las características esta dado por la variable "características". Por ejemplo, Si se escogieron 2 características y 3 canales diferentes la dimensión de las columnas será igual a 6. Las primeras dos representan las 2 características del primer canal. Las columnas 3-4 las características del segundo canal, y las últimas 2, representan las características del tercer canal.
Caracteristicas	char	El texto representa el orden y el nombre de las características extraídas. Tomádo el ejemplo de 2 características y 3 canales, 'Curtosis, Cruces por cero', indica que la columnas 1,3 y 5 son vectores resultado de extraer curtosis en los 3 canales escogidos, y las columnas 2,4, y 6 son vectores resultado de extraer cruces por cero en los 3 canales escogidos.
Canales	vector	Vector que indica los canales y el orden en que se analizaron para extraer las características.
Data	vector	Vector que incluye la señal EEG analizada sin ser seccionada en ventanas. La dimensión de las columnas es igual al número de canales procesados y cumple el orden del vector "canales". Por ejemplo, si el vector canales es igual a [2,4,6], la 1 columna del vector "data" representa las muestras totales del canal 2, la 2 columna del vector "data" representa las muestras del canal 4 y la última columna representará las muestras totales del canal 6.
FMuestreo	int	Incluye la frecuencia de muestreo encontrada en el archivo EDF del EEG analizado aproximada a un entero, en caso sea un número decimal.
Ventana	int	Representa el tamaño de ventana utilizado para la extracción de características. Es decir, el número de muestras por ventanas que se utilizó para calcular iterativamente cada característica.

Table 1: Análisis Directo: Descripción de Variables

Nombre	Tipo	Contenido
Etiquetas	vector	Vector de clases, 0 y 1 representan tipos de registro "Con crisis" y "Sin crisis". La dimensión del vector quivale al número total de muestras de la señal dividido el tamaño de ventana escogido.
Vector_Caracteristicas	vector	Vector resultado de características escogidas. Las columnas representan las características según las sub-bandas escogidas. Debido a que el toolbox utiliza 7 niveles de descomposición la dimensión máxima de columnas posibles es igual a 42. Tomándo en cuenta que el usuario escoge extraer las 6 características disponibles para todas las 7 sub-bandas.
Caracteristicas	char	El texto representa el orden y el nombre de las características extraídas.
Canal	int	Indica el canal analizado
Fbanda	char	El texto indica que sub-bandas escogio el usuario para extraer las características. Las bandas se acotan en rangos de frecuencia que coinciden con los 5 ritmos cerebrales. El nivel 1 representa Theta, los niveles 2-3 Theta/Beta, los niveles 3-4 Beta/Alpha, niveles 4-5 Alpha/Gamma, los niveles 5-6 Gamma/Delta y el nivel 7 representa el ritmo Delta.
CoeficientesWavelet	vector	El vector incluye las bandas de coeficientes utilizadas para extraer las características seleccionadas.
FMuestreo	int	Incluye la frecuencia de muestreo encontrada en el archivo EDF del EEG analizado aproximada a un entero, en caso sea un número decimal.
Ventana	int	Representa el tamño de ventana utilizado para la extracción de características. Es decir, el número de muestras por ventanas que se utilizó para calcular iterativamente cada característica.

Table 2: Análisis Wavelet: Descripción de Variables

3.3 Implementar Clasificadores

Posterior a extraer el vector de características se habilitan las opciones de clasificar la señal utilizando algoritmos de machine learning: SVM y ANN. Las ventanas de ambos clasificadores se detallan en las figuras 7 y 8.

RED NEURONAL ARTIFICIAL

% SEGMENTOS PARA ENTRENAR

70

% SEGMENTOS PARA PRUEBA

15

% SEGMENTOS PARA VALIDAR

15

ALGORITMO DE ENTRENAMIENTO

trainscg

CARGAR NUEVO VECTOR DE CARACTERÍSTICAS PARA ENTRENAMIENTO Y PRUEBA

CARGAR

CAPAS OCULTAS

8

TIPO DE VECTOR

☒ ANÁLISIS DIRECTO
☐ ANÁLISIS WAVELET

GENERAR

VISUALIZAR

RESULTADOS

SEGMENTOS PARA ENTRENAMIENTO

SEGMENTOS PARA PRUEBA

Figure 7: Ventana Clasificador ANN

Figure 8: Ventana Clasificador SVM

Ambas ventanas requieren especificar el tipo de análisis que se utilizó para extraer el vector de características. Además, permiten que el usuario ingrese un nuevo archivo *M* con el vector de características y el vector de clases deseado. Se incluye esta opción debido a que el toolbox genera automáticamente el vector de clases con el supuesto que la señal ingresada esta balanceada y ordenada de forma que la mitad de sus muestras pertenecen a la clase 0 y la mitad restante a la clase 1. El archivo ingresado debe contener específicamente las siguientes variables:

- data: vector de características de dimensión $n \times m$, donde n representa el número de ventanas seleccionadas para entrenamiento del clasificador y m la multiplicación del número de características por el número de canales analizados.
- labels: vector de dimensión $n \times 1$, que contiene "ceros" en las ventanas que representan la clase 1, y "unos" en las ventanas que representan la clase 2. El número de filas debe corresponder al número de ventanas que contiene el vector de características "data".

Para la implementación de la red neuronal artificial, el toolbox utiliza un red perceptrón multicapa. El usuario puede seleccionar el algoritmo de entrenamiento, el número de capas ocultas y los porcentajes de división de datos para entrenamiento, prueba y validación. En el caso del clasificador SVM, el toolbox utiliza validación cruzada "*K-fold*", donde el usuario tiene opción de modificar el número de particiones utilizadas para encontrar el grupo de entrenamiento y prueba. En ambos clasificadores, el botón visualizar despliega los resultados de la clasificación, matriz de confusión y exactitud. Adicionalmente, se detalla un resumen de resultados en la parte inferior de ambas ventanas.

3.4 Exportación y Generación de Resultados

El botón generar extrae los resultados de clasificación según el algoritmo seleccionado en un archivo *M* compatible con MATLAB, la estructura de ambos archivos se describen en las figuras 9 y 10. Las variables de resultado para cada clasificador se detallan en la tabla 3.4.

Workspace	
Name ^	Value
Etiquetas	368x1 double
RN_resultados	1x1 struct
VectorCaracteristicas	368x4 double

Variables - RN_resultados	
RN_resultados	
1x1 struct with 2 fields	
Field ^	Value
net	1x1 struct
salidas	1x368 double

Figure 9: Resultados clasificador ANN

Workspace	
Name ^	Value
data_entrenamiento	197x4 double
data_prueba	66x4 double
etiquetas_entrenamiento	197x1 double
etiquetas_prueba	66x1 double
svm_resultados	1x1 struct

Variables - svm_resultados	
svm_resultados	
1x1 struct with 3 fields	
Field ^	Value
Predicted_la...	66x1 double
test_features	66x2 double
Accuracy	100

Figure 10: Resultados clasificador SVM

ANN Descripción de Variables		
Etiquetas	vector	Vector de clases reales, 0 y 1 representan tipos de registro "Con crisis" y "Sin crisis". La dimensión del vector equivale al número total de muestras de la señal dividido el tamaño de ventana escogido.
Vector_Caracteristicas	vector	Vector resultado de características escogidas. Las columnas representan las características según el análisis escogido. Con el análisis directo las columnas representan las características según el número de canales. Con Análisis Wavelet, las columnas representan las características según las subbandas escogidas.
net	struct	Resultado del toolbox deep learning para redes neuronales de MATLAB. Incluye 29 variables relacionadas al diseño, estructura, entrenamiento y resultados de la red neuronal creada.
Salidas	vector	Vector de pesos con etiquetas de salida predichas por el clasificador. Valores cercanos a 1 (mayores a 0.5) pertenecen a la clase 1, de lo contrario pertenecen a la clase 0
SVM Descripción de Variables		
Data_entrenamiento	vector	El vector incluye el porcentaje de datos para entrenamiento del vector característica según la validación cruzada realizada.
Data_prueba	vector	El vector incluye el porcentaje de datos para prueba del vector característica según la validación cruzada realizada
Etiquetas_entrenamiento	vector	El vector incluye las clases que representan al vector Data_entrenamiento
Etiquetas_prueba	vector	El vector incluye las clases que representan al vector Data_prueba.
SVM_resultados	struct	La estructura incluye el vector de características de prueba y el vector de clases predicho por el clasificador. Adicionalmente, se incluye el porcentaje de exactitud.

Table 3: Resultados de Clasificadores: Descripción de Variables

References

- [1] María Jesús Angulo. Análisis y reconocimiento de patrones de señales biomédicas de pacientes con epilepsia. Technical report, Universidad del Valle de Guatemala, 2020.