

UNIVERSIDAD DEL VALLE DE GUATEMALA
Facultad de Ingeniería



**Herramienta de Software para el Estudio de la Epilepsia -
Fase II**

Protocolo de trabajo de graduación presentado por Jorge Diego
Manrique Sáenz, estudiante de Ingeniería Mecatrónica

Guatemala,

2021

Resumen

Este protocolo presenta una propuesta para mejorar la herramienta para el estudio de la epilepsia desarrollada en la fase I por María Fernanda Pineda, tesis “Diseño e Implementación de una Base de Datos de Señales Biomédicas de Pacientes con Epilepsia” [1]. Se busca realizar modificaciones y mejoras a la herramienta para que el Centro de Epilepsia y Neurocirugía Funcional HUMANA cuente con un programa que les resuelva ciertas dificultades que se les presentan y además a reducir tiempos en los procesos internos como lo es la evaluación de las señales electroencefalográficas de los pacientes.

Es necesario evaluar los programas que utilizan actualmente en HUMANA para la captura de las encefalogramas ya que para poder presentar una herramienta funcional y útil también es necesario eliminar tareas repetidas que al final cuestan tiempo importante y pueden causar errores involuntarios.

Con las modificaciones a la herramienta de la fase I se busca incluir funcionalidades que puedan aportar valor a futuros proyectos que sigan esta línea de investigación.

Por último, después de haber realizado las modificaciones y mejoras correspondientes se propondrán modos para el uso de la herramienta de forma remota. Esto debido a que la presente situación mundial obliga a buscar alternativas para poder trabajar de forma remota por razones de seguridad.

Antecedentes

La epilepsia es una de las condiciones neurológicas más comunes, sin embargo, aún no logramos comprenderla ni se cuenta con un tratamiento específico. La epilepsia es una condición que causa convulsiones recurrentes sin provocar. Una convulsión epiléptica es causada por señales neurológicas anormales. La epilepsia es una de las condiciones neurológica más comunes, aproximadamente 50 casos nuevos por cada 100,000 habitantes anualmente. Aproximadamente el 1 % de la población mundial sufre de epilepsia y un tercio de los pacientes sufre de epilepsia refractaria. La epilepsia refractaria se refiere a que las convulsiones no pueden ser controladas por los medicamentos antiepilépticos mas recomendados o utilizados. El 75 % de los pacientes con epilepsia lo empiezan a padecer en la niñez [2].

Las convulsiones epilépticas son clasificadas dentro de 3 categorías: Generalizada, focal o parcial y espasmo epiléptico. Las convulsiones generalizadas empiezan en las redes neuronales bilaterales. Las convulsiones focales o parciales se originan en las redes neuronales de un hemisferio del cerebro. Una convulsión puede iniciar como focal o parcial y luego convertirse en generalizada. Las convulsiones pueden originarse en la corteza o en estructuras subcorticales. Utilizado historia detallada, información clave obtenida por EEG y otra información relevante generalmente un médico puede clasificar las convulsiones/tipo de epilepsia y generar un diagnostico y un plan para el tratamiento [2].

Una base de datos relacional es un tipo de base de datos que almacena y provee acceso a datos que están relacionados a otros. Estas bases de datos se basan en un modelo, intuitivo, y fácil para almacenar la información en tablas. En estas tablas cada fila es un registro con un número de identificación único llamado “llave”. Cada columna de la tabla representa un

atributo con información específica del registro [3].

El aprendizaje automático o *machine learning* es utilizado ampliamente en las áreas de la salud y la biología. Investigadores y científicos lo utilizan en distintas áreas ya que permite el análisis de mucha información y permite visualizar alternativas para diferentes aplicaciones. El aprendizaje automático se puede utilizar para la clasificación y detección de epilepsia. También se ha utilizado para análisis de señales EEG ya que de esta forma se puede descubrir información importante de la señal [4].

El Centro de Epilepsia y Neurocirugía Funcional HUMANA, es una organización formada por profesionales en Neurociencias que trabajan en beneficio de los pacientes que padecen problemas Neurológicos de difícil control, Epilepsia, Parkinson, Tumores Cerebrales, Columna Vertebral, Movimientos Anormales entre otros. Humana es el Centro de Referencia en Neurociencias para Guatemala y Centro América donde los médicos cuentan con los mejores recursos para el diagnóstico y tratamiento de enfermedades cerebrales [5].

Las tesis desarrolladas por María Fernanda Pineda [1] y María Jesús Angulo [6] consistieron en el desarrollo de una herramienta para análisis de señales electroencefalográficas y diseño de una base de datos para almacenar las señales grabadas por HUMANA. Esta base de datos almacena información sobre los pacientes como edad, datos de la señal como frecuencia y número de canales, pero no nombres de los pacientes ya que busca conservar la confidencialidad de los pacientes.

La base de datos considera hasta 35 canales de los cuales 20 son obligatorios y almacena la información en 3 tablas, una para datos del paciente, otra para datos descriptivos de la prueba y la última con las grabaciones de los diferentes canales de la señal. La aplicación para grabar las señales fue desarrollada en Matlab por lo que es necesario tener instalado este programa para poder ejecutarlo o generar un ejecutable previamente.

El proceso para importar las señales a la base de datos es el siguiente: En primer lugar, se exporta la señal en el programa utilizado en HUMANA para la captura de la señal encefalográfica. En segundo lugar, se llena el formulario con los datos del paciente en el programa desarrollado en Matlab por Maria Fernanda Pineda. En tercer lugar, se exporta el archivo .edf.

Actualmente el flujo no es automático y posee pasos duplicados ya que los datos del paciente se ingresan 2 veces. Primero en el programa utilizado para capturar las señales EEG. Y segundo en la herramienta desarrollada en Matlab. Esto incrementa el tiempo utilizado y da lugar a errores no voluntarios a la hora de ingresar los datos.

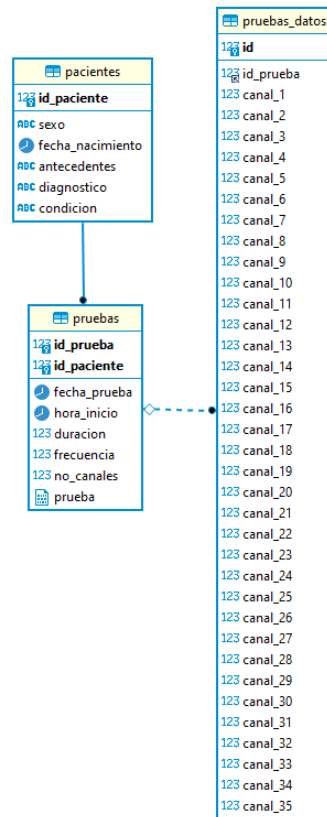


Figura 1. Estructura de la base de datos desarrollada en la fase I [1].



Figura 2. Aplicación desarrollada en Matlab para carga de las señales EEG [1].

Justificación

Actualmente HUMANA realiza el análisis de las señales EEG manualmente. Como es un proceso manual y en situaciones hay señales muy largas los tiempos de análisis pueden variar significativamente. Ya que es un proceso manual es necesario tomar en cuenta que se pueden cometer errores y obviar información importante de la señal. Por lo tanto, automatizar el proceso es de suma importancia y la calidad del análisis esta directamente relacionado con la calidad de la información.

La herramienta actual desarrollada por Maria Fernanda Pineda [6] permite el almacenamiento de las señales en una base de datos, pero no considera el almacenamiento ni visualización de los datos resultantes de la anotación de las señales EEG ni tampoco existen usuarios con diferentes permisos para mantener la confidencialidad de cierta información. Esto es importante ya que los resultados deben almacenarse en algún lugar. También es de suma importancia poder visualizar las anotaciones realizadas a las señales y *los resultados de los análisis efectuados* Además no podemos obviar la responsabilidad de mantener la confidencialidad de los datos de los pacientes.

Se busca dejar un programa funcional que permita administrar usuarios y sus permisos y que sea de mucha utilidad para automatizar las tareas lo más posible en el día a día de HUMANA. También es necesario realizar modificaciones a las tablas definidas y a la aplicación para integrar el uso de diferentes usuarios. Por ejemplo, existe una tabla en la base de datos que no cuenta con una llave primaria y la forma de autenticar credenciales para la conexión a la base de datos es por medio de un archivo .csv lo cual no es practico ni lo correcto.

Objetivos

Objetivo General

Mejorar la herramienta de *software* desarrollada en la fase anterior del proyecto de estudio de la epilepsia, y adaptarla para su uso en el Centro de Epilepsia y Neurocirugía Funcional (HUMANA).

Objetivos Específicos

- Expandir la funcionalidad de la herramienta de *software* desarrollada en la fase anterior.
- Optimizar el modelo y funcionalidad de la base de datos con que cuenta la herramienta.
- Desarrollar una versión de la herramienta compatible con las estaciones de trabajo de HUMANA.
- Integrar en la aplicación un módulo de control de usuarios y sus permisos.
- Proponer un modo para la operación remota de la herramienta.

Marco teórico

Epilepsia

La epilepsia es una condición que causa convulsiones recurrentes sin provocar. Una convulsión epiléptica es causada por señales neurológicas anormales. La epilepsia es una de las condiciones neurológica más comunes, aproximadamente 50 casos nuevos por cada 100,000 habitantes anualmente. Aproximadamente el 1 % de la población mundial sufre de epilepsia y un tercio de los pacientes sufre de epilepsia refractaria. El 75 % de los pacientes con epilepsia lo empiezan a padecer en la niñez [2].

La epilepsia afecta tanto a hombres como a mujeres de todas las razas, orígenes étnicos y edades [7].

Los síntomas de las convulsiones pueden variar ampliamente. Algunas personas con epilepsia simplemente miran de manera fija por unos segundos durante una convulsión, mientras que otras mueven repetidamente los brazos o las piernas. Tener una sola convulsión no significa que padezcas epilepsia. Por lo general, se requieren al menos dos convulsiones no provocadas para determinar un diagnóstico de epilepsia [7].

Tipos de Epilepsia

Las epilepsias se clasificaban según su sitio de inicio, si la causa era conocida o no lo era. El sistema de clasificación actual toma en cuenta las causas estructurales y genéticas e incluye el tipo de convulsión, el diagnóstico del síndrome y el grado de deterioro funcional. La forma de clasificar seguirá evolucionando a medida que se conozca mas de la epilepsia y de la genética [2].

Tipos de convulsiones

Las convulsiones epilépticas son clasificadas dentro de 3 categorías: Generalizada, focal o parcial y espasmo epiléptico. Las convulsiones generalizadas empiezan en las redes neuronales bilaterales. Las convulsiones focales o parciales se originan en las redes neuronales de un hemisferio del cerebro. Una convulsión puede iniciar como focal o parcial y luego convertirse en generalizada. Las convulsiones pueden originarse en la corteza o en estructuras subcorticales [2].

Electroencefalograma

El electroencefalograma (EEG) es una grabación de la actividad cerebral. Para obtener estas grabaciones se utilizan pequeños sensores pegados a la cabeza en diferentes posiciones los cuales captan las señales eléctricas generadas por las células en el cerebro cuando estas se comunican entre sí [8].

Estas señales son grabadas por una maquina para luego ser examinadas por un medico

para determinar si existe alguna anormalidad en ellas. La grabación y evaluación de estas señales se lleva a cabo por especialistas altamente entrenados llamados neurofisiólogos clínicos [8].

Una EEG puede ser utilizada para diagnosticar y monitorear varias condiciones que afectan el cerebro. Puede ayudar a identificar la causa de ciertos síntomas como por ejemplo convulsiones o problemas para recordar, pero también puede ayudar a saber más sobre alguna condición ya diagnosticada [8].

El uso principal para los electroencefalogramas consiste en la detección e investigación de la epilepsia. Los EEG ayudan al médico a identificar el tipo de epilepsia y el motivo por el cual las convulsiones son causadas como también el mejor método de tratamiento [8].

Además, los EEG también pueden ser utilizados para investigar otros problemas o condiciones como por ejemplo: demencia, golpes en la cabeza, tumores cerebrales, inflamación cerebral y desordenes de sueño [8].

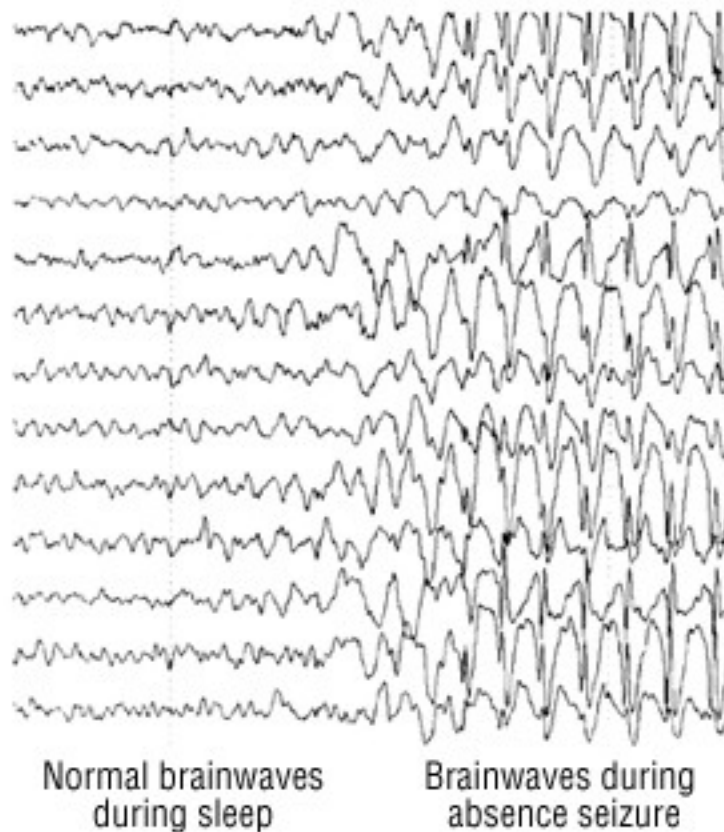


Figura 3. Señal EEG normal vs señal EEG durante una convulsión. [9].

Bases de datos

Una base de datos relacional es un tipo de base de datos que almacena y provee acceso a datos que están relacionados a otros. Estas bases de datos se basan en un modelo, intuitivo,

y fácil para almacenar la información en tablas. En estas tablas cada fila es un registro con un número de identificación único llamado “llave”. Cada columna de la tabla representa un atributo con información específica del registro [3].

El objetivo de una base de datos es recolectar y mantener información de tal forma que esté disponible para referencia de operaciones o análisis para la toma de decisiones. Las bases de datos deben cumplir con un número de requerimientos técnicos. Por ejemplo: La data debe ser accesible para varios usuarios, la información debe ser recolectada desde que se genera, esta información puede tenerse almacenada por mucho tiempo sin necesidad de utilizarse inmediatamente, la lectura y escritura de información debe de poder realizarse constantemente y la información debe almacenarse de tal forma que este relacionada y pueda buscarse fácilmente [10].

Organización

El modelo de entidad-relación para bases de datos fue diseñado y desarrollado por Peter Chen del MIT en 1976. Es una forma intuitiva y conceptual de representar el modelo de datos en una base de datos. Normalmente se representa con un diagrama de entidad-relación. El diagrama de entidad-relación permite visualizar las tablas, sus atributos, llaves primarias y foráneas, así como las relaciones que existen entre tablas. En una base de datos relacional es necesario vincular las tablas con información utilizando llaves foráneas para consultar fácilmente información en las diferentes tablas. Las bases de datos relacionales son ampliamente utilizadas debido a su simplicidad para modelar, representar y consultar la información [11].

Conceptos básicos

Schemas: es una colección de objetos dentro de una base de datos. El *schema* tiene un dueño y puede contener varios objetos dentro de la base de datos como, por ejemplo: Tablas, *triggers*, secuencias, vistas, *procedures*, etc [12].

Tablas: Unidad básica de almacenamiento en una base de datos. Consiste en columnas que pueden ser de diferente tipo de dato y de diferentes tamaños y de filas o registros con la información que se desea almacenar [13].

Tipos de datos: Los tipos de datos determinan los valores y tipo de información que se desea almacenar en la columna indicada en la tabla. Pueden ser por ejemplo: INT (numero entero), Float (número real), String (campo alfanumérico), etc [14].

Llaves primarias: la llave primaria consiste en la columna o combinación de columnas que identifican cada registro de la tabla para que este registro pueda ser consultado o modificado fácilmente por separado o en conjunto con otros registros [15].

Llaves foráneas: las llaves foráneas permiten relacionar las tablas de la base de datos entre sí. La llave foránea se define en la tabla “hija” y define las columnas que se relacionan con las columnas en la tabla “padre” [16].

Aprendizaje automático

El aprendizaje automático o *machine learning* es utilizado ampliamente en las áreas de la salud y la biología. Investigadores y científicos lo utilizan en distintas áreas ya que permite el análisis de mucha información. Dependiendo del tipo de modelo utilizado y la aplicación es posible responder distintas dudas que puedan surgir. Un modelo de clasificación permite determinar en base a datos históricos, por ejemplo, si algo va a suceder o no. Un modelo de series de tiempo permite predecir valores de una variable que depende del tiempo tomando en cuenta la estacionalidad, tendencia, ciclos, etc. Un modelo de *clustering* permite agrupar hechos, objetos, personas en base a sus características. Las aplicaciones son muy variadas así como los modelos disponibles [4].

Aplicación en la medicina y epilepsia

Machine learning se puede utilizar para la clasificación y detección de epilepsia. También se ha utilizado para análisis de señales EEG ya que de esta forma se puede descubrir información importante de la señal [1].

Metodología

- Investigar sobre los programas utilizados en HUMANA para la toma de encefalogramas. Determinar que base de datos utilizan y si es posible obtener las credenciales de conexión.
- Determinar si es posible conectar directamente la base de datos a los programas utilizados en HUMANA para eliminar tareas duplicadas para los usuarios. De no ser posible continuar con el procedimiento actual y extraer la mayor cantidad de datos de los archivos .edf del pacientes.
- Evaluar la herramienta desarrollada en la fase I y proponer alternativas o mejoras. Listar características y puntos de mejora.
- Diseñar prototipos independientes a la herramienta de la fase I para la interfaz de administración de usuarios y sus permisos y de integración de nuevas entidades en la base de datos para el almacenamiento de resultados del proceso de anotación y del entrenamiento de los modelos, iterando hasta obtener un diseño que se adapte a las necesidades y requerimientos.
- Integrar las mejoras y modificaciones correspondientes a la herramienta de la fase I según lo evaluado anteriormente.
- Realizar pruebas de control de calidad tratando de abarcar todas las posibles variaciones de utilización de la herramienta para asegurar su correcto funcionamiento cuando la herramienta ya se encuentre instalada en la computadora/servidor de HUMANA.
- Exportar la herramienta para su uso directamente en Windows para no requerir que Matlab esté instalado en las máquinas de HUMANA.
- Investigar y proponer alternativas para el uso de la herramienta de forma remota.

Cronograma de actividades

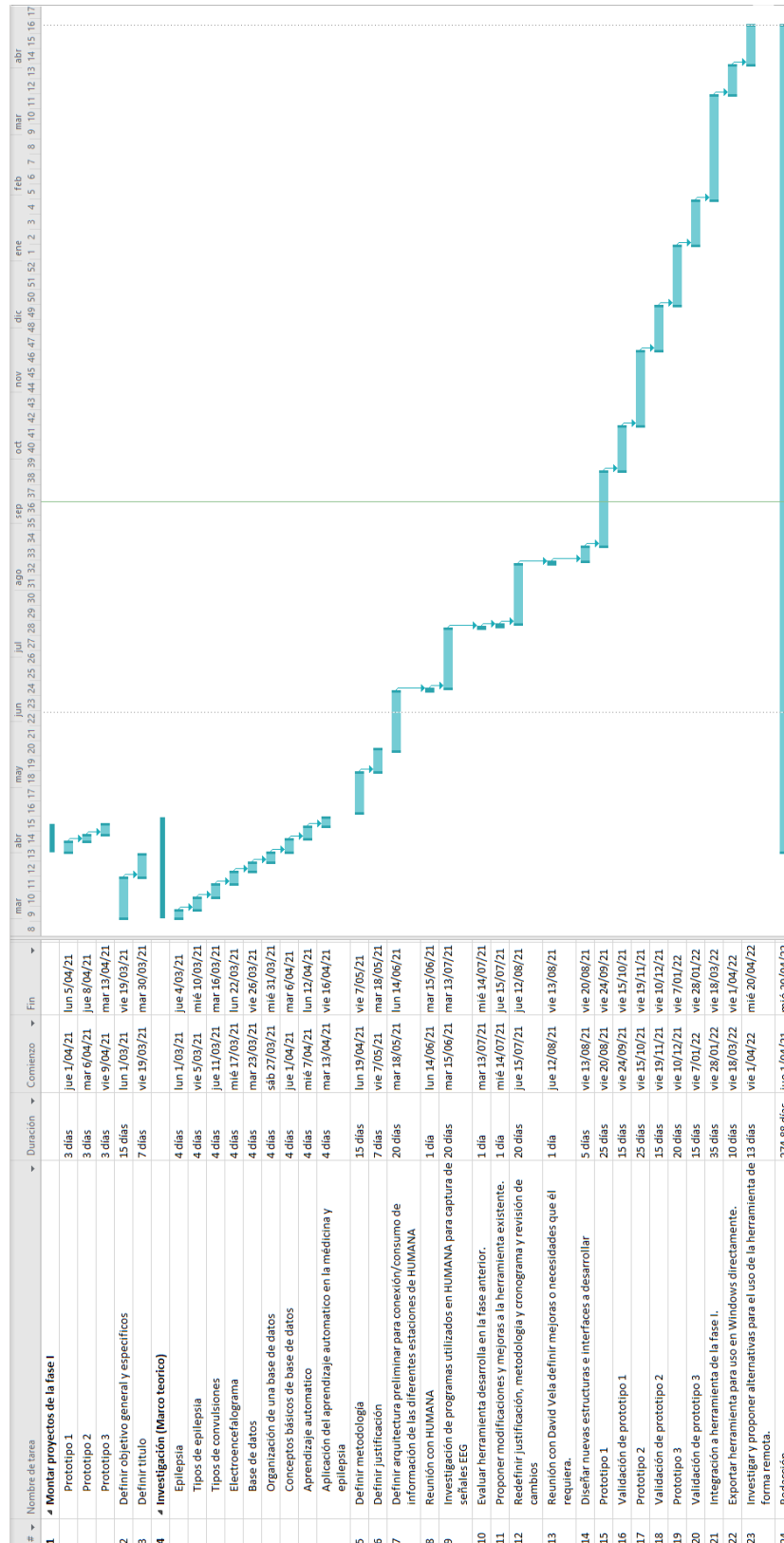


Figura 4. Cronograma de actividades (2021-2022)

#	Actividades	Descripción
1	Montar proyectos de la fase I	Poner en funcionamiento los prototipo de la fase anterior para familiarizarse con lo ya desarrollado.
	Prototipo 1	
	Prototipo 2	
	Prototipo 3	
2	Definir objetivo general y específicos	Redactar objetivos en base a los que se quiere lograr para esta fase.
3	Definir titulo	Redactar el titulo de la tesis tomando en consideración los objetivos definidos.
4	Investigación (Marco teorico)	Investigación de temas relacionados con la tesis e información que puede ser de utilidad para el lector para que pueda comprender lo que se trabajo.
	Epilepsia	
	Tipos de epilepsia	
	Tipos de convulsiones	
	Electroencefalograma	
	Base de datos	
	Organización de una base de datos	
	Conceptos básicos de base de datos	
	Aprendizaje automatico	
	Aplicación del aprendizaje automatico en la medicina y epilepsia	
5	Definir metodología	Redactar los pasos necesarios para poder cumplir los objetivos definidos.
6	Definir justificación	Indicar el ¿Por qué? de tesis e indicar algunos puntos de mejora identificados.
7	Definir arquitectura preliminar para conexión/consumo de información de las diferentes estaciones de HUMANA	Definir una arquitectura preliminar de las herramientas a utilizar y como se planea conectarlas.
8	Reunión con HUMANA	Reunión o enviar correo a HUMANA para poder tener la lista completa de plrogramas que ellos utilizan para la toma de EEGs.

Cuadro 1. Descripción de actividades (Primera parte)

#	Actividades	Descripción
9	Investigación de programas utilizados en HUMANA para captura de señales EEG	Es necesario investigar que base de datos utilizan estos programas y si es posible conectarse y extraer la información directamente. Puede ser necesario contactar a la empresa que desarrolló el programa.
10	Evaluar herramienta desarrollada en la fase anterior.	Listar pros y contras de la herramienta desarrollada en la fase I para identificar otros posibles puntos de mejora.
11	Proponer modificaciones y mejoras a la herramienta existente.	Lluvia de ideas para identificar características que sería bueno incluir en la herramienta.
12	Redefinir justificación, metodología y cronograma y revisión de cambios.	Modificar las secciones del protocolo ya que en la última reunión con HUMANA se descartó la instalación de cualquier programa en los computadores con el software de captura de señales EEG.
13	Reunión con David Vela definir mejoras o necesidades que él requiera.	Listar características que David Vela requiera sean incluidas en la herramienta.
14	Diseñar nuevas estructuras e interfaces a desarrollar.	Hacer un <i>sketch</i> de la nueva estructura a incluir en la base de datos y de la interfaz para visualizar los resultados de la anotación.
15	Prototipo 1	Realizar prototipo 1 de estructura para almacenamiento de resultados e interfaz para visualización de resultados de la anotación de las señales EEG.
16	Validación de prototipo 1	Pruebas para determinar posibles errores e identificar características que valdría la pena incluir.
17	Prototipo 2	Realizar prototipo 2 aplicando mejoras y solución de posibles defectos identificados en el prototipo 1.

Cuadro 2. Descripción de actividades (Segunda parte)

#	Actividades	Descripción
18	Validación de prototipo 2	Pruebas para determinar posibles errores e identificar características que valdría la pena incluir.
19	Prototipo 3	Realizar prototipo 3 aplicando mejoras y solución de posibles defectos identificados en el prototipo 2.
20	Validación de prototipo 3	Pruebas para determinar posibles errores.
21	Integración a herramienta fase I.	Integración de último prototipo a herramienta de la fase I para poder acceder a las nuevas características desde un menú centralizado.
22	Exportar herramienta para uso en Windows directamente.	De ser necesarios exportar la herramienta de matlab a un ejecutable de windows para poder utilizarlo sin necesidad de contar con matlab.
23	Investigar y proponer alternativas para el uso de la herramienta de forma remota.	Listar posibles formas de utilizar la herramienta de forma remota y determinar en conjunto con HUMANA la mejor alternativa.
24	Redacción	Redacción de la tesis conforme se vayan trabajando las actividades.

Cuadro 3. Descripción de actividades (Tercera parte)

Índice preliminar

- Prefacio
- Lista de figuras
- Lista de cuadros
- Resumen
- Abstract

1. Introducción

2. Antecedentes

3. Justificación

4. Objetivos

5. Alcance

6. Marco teórico

- a)* Epilepsia
- b)* Tipos de epilepsia
- c)* Tipos de convulsiones
- d)* Electroencefalograma
- e)* Base de datos
- f)* Organización de una base de datos
- g)* Conceptos básicos
- h)* Aprendizaje automático
- i)* Aplicación en la medicina y epilepsia del aprendizaje automático

7. Diseño experimental

8. Experimentos

- a)* Conexión a base de datos de los programas utilizados en HUMANA
- b)* Levantar ambiente de desarrollo utilizando los programas utilizados en HUMANA
- c)* Crear la base de datos definida en la fase I
- d)* Cargar señales de prueba en bases de datos de origen
- e)* Planificar *dataflows*
- f)* Implementar *dataflows*
- g)* Crear *job* para ejecución de ETL periódicamente
- h)* Validación de datos y medición de tiempos de ejecución
- i)* Optimizar procesos

- j) Despliegue/Implementación en producción
- 9. Resultados
 - a) *Dataflows* utilizados
 - b) Particularidades y descripción de *dataflows* utilizados
 - c) Tiempos de ejecución
 - d) Posibles limitantes
- 10. Conclusiones
- 11. Recomendaciones
- 12. Bibliografía
- 13. Anexos
- 14. Glosario

Referencias

- [1] M. F. P. E., “Diseño e Implementación de una Base de Datos de Señales Biomédicas de Pacientes con Epilepsia,” Tesis de licenciatura, Universidad Del Valle de Guatemala, 2021.
- [2] C. E. Stafstrom, “Seizures and epilepsy: an overview for neuroscientists,” *PubMed*, vol. 5, n.º 6, a022426, 2015.
- [3] Oracle, “What is a relational database?,” 2020, Consultado el: 13/05/2021. dirección: <https://www.oracle.com/database/what-is-a-relational-database/>.
- [4] M. K. Siddiqui, “A review of epileptic seizure detection using machine learning classifiers,” *PubMed*, vol. 7, n.º 1, pág. 5, 2020.
- [5] H. Gt, “Especialistas en enfermedades neurológicas de difícil control,” *humanagt.org*, 2021.
- [6] M. J. A. T., “Análisis y Reconocimiento de Patrones de Señales Biomédicas de Pacientes con Epilepsia,” Tesis de licenciatura, Universidad Del Valle de Guatemala, 2020.
- [7] M. F. for Medical Education y R. (MFMER), “Epilepsia,” *mayoclinic.org*, 2018.
- [8] N. Uk, “Electroencephalogram (EEG),” *nhs.uk*, 2018.
- [9] H. H. Publishing, “Electroencephalogram (EEG),” 2019, Consultado el: 28/06/2021. dirección: <https://www.health.harvard.edu/diseases-and-conditions/electroencephalogram-eeg>.
- [10] Wiederhold, “Databases,” *Computer*, vol. 17, n.º 10, págs. 211-223, 1984. DOI: 10.1109/MC.1984.1658971.
- [11] B. R. Sinha, P. P. Dey, M. N. Amin y G. W. Romney, “Database modeling with Object Relationship Schema,” en *2013 12th International Conference on Information Technology Based Higher Education and Training (ITHET)*, 2013, págs. 1-7. DOI: 10.1109/ITHET.2013.6671029.

- [12] Oracle, “Introduction to Schema Objects,” 2021, Consultado el: 13/05/2021. dirección: https://docs.oracle.com/cd/B19306_01/server.102/b14220/schema.htm.
- [13] —, “Managing Tables,” 2021, Consultado el: 13/05/2021. dirección: https://docs.oracle.com/cd/B19306_01/server.102/b14231/tables.htm.
- [14] —, “Selecting a Datatype,” 2021, Consultado el: 13/05/2021. dirección: https://docs.oracle.com/cd/A58617_01/server.804/a58241/ch5.htm.
- [15] —, “Primary Keys,” 2021, Consultado el: 13/05/2021. dirección: <https://docs.oracle.com/en/database/other-databases/nosql-database/12.2.4.5/java-driver-table/primary-keys.html#GUID-6A063474-4A3A-4981-B1D8-71D0D95BE8DF>.
- [16] —, “FOREIGN KEY Constraints,” 2021, Consultado el: 13/05/2021. dirección: https://docs.oracle.com/cd/E17952_01/mysql-5.6-en/create-table-foreign-keys.html.