

Diseño de implementación de una herramienta de software para el estudio de la epilepsia

María Jesús Angulo¹, ang16508@uvg.edu.gt / María Fernanda Pineda¹, pin16021@uvg.edu.gt / Luis Alberto Rivera¹, larivera@uvg.edu.gt / Héctor Antonio Hurtarte², hahurtarte@uvg.edu.gt

¹Departamento de Ingeniería Electrónica, Mecatrónica y Biomédica, Facultad de Ingeniería, Universidad Del Valle de Guatemala

²Departamento de Ingeniería en Ciencia de la Computación, Facultad de Ingeniería, Universidad Del Valle de Guatemala

RESUMEN: La epilepsia es una enfermedad que afecta a millones de personas en el mundo. En Guatemala, el Centro de Epilepsia y Neurocirugía Funcional, HUMANA, es la única institución que cuenta con especialistas para el tratamiento y estudio de la enfermedad. Mucho del análisis de los datos que se generan en HUMANA se procesan de forma manual, por lo que algoritmos y métodos de reconocimiento de patrones pueden ser de mucha ayuda para automatizar los procesos y facilitar el análisis. El objetivo de este trabajo fue diseñar e implementar una herramienta de software con una base de datos integrada para el procesamiento, análisis y almacenamiento de señales biomédicas de pacientes con epilepsia. Se exploraron redes neuronales artificiales y máquinas de vectores de soporte como algoritmos de clasificación, así como características en el dominio del tiempo y del tiempo-frecuencia de señales electro-encefalográficas, para poder detectar episodios epilépticos. Se validaron los algoritmos usando señales tomadas de repositorios de datos públicos, y logramos porcentajes de clasificación correcta superiores al 96%. La base de datos desarrollada es capaz de almacenar señales e información relevante de pacientes. Se validó, también, su funcionalidad utilizando señales proporcionadas por HUMANA. La herramienta cuenta con una interfaz gráfica que facilita el uso de las funciones de procesamiento y análisis, la interacción con la base de datos, y la visualización de resultados. Este trabajo aporta una herramienta de mucha utilidad para, los especialistas de HUMANA, para la investigación que se lleva a cabo actualmente en la Universidad Del Valle de Guatemala, y para cualquier persona interesada en el estudio de la epilepsia.

PALABRAS CLAVE: epilepsia, señales electro-encefalográficas, aprendizaje automático, base de datos, herramienta de software.

Design and implementation of a software tool for the study of epilepsy

ABSTRACT: Millions of people around the world suffer from epilepsy. The Center for Epilepsy and Functional Neurosurgery, HUMANA, is the only institution with specialists who treat and study the disease in Guatemala. Much of the analysis of data generated at HUMANA is done manually. Therefore, machine learning and pattern recognition methods could be very helpful in automating processes and facilitating the data analysis. The goal of this work was to design and implement a software tool with an integrated database for processing, analyzing, and storing biomedical signals from epilepsy patients. To recognize epileptic episodes, we evaluated artificial neural networks and support vector machines as classification algorithms, and time and time-frequency domain features of electroencephalographic signals. The algorithms using signals from public data repositories were validated, reaching classification accuracies above 96%. The database can store signals and relevant information from patients. Also the database functionality using data provided by HUMANA was validated. The software tool includes a graphical interface, which simplifies the data processing and analysis, the interaction with the database, and results visualization. This work contributes with a useful tool for HUMANA specialists, for undergoing research at Universidad Del Valle de Guatemala, and for anyone interested in studying epilepsy.

KEYWORDS: epilepsy, electroencephalographic signals, machine learning, database, software tool.

Introducción

La epilepsia es un trastorno cerebral crónico no transmisible que provoca desórdenes de la actividad cerebral, causando eventos clínicos denominados crisis epilépticas. Actualmente se estiman más de 70 millones de personas en todo el mundo que padecen esta patología (Thijs *et al.* 2019). Debido a su presencia mundial, este trastorno es objeto de estudio, tanto en el campo de la medicina, como el de ingeniería, con el propósito de mejorar el proceso de diagnóstico, comprender sus posibles causas y brindar tratamientos efectivos para los pacientes.

El electroencefalograma es una prueba ampliamente utilizada en estudios de epilepsia, ya que permite registrar información de la actividad eléctrica del cerebro - señales electroencefalográficas (EEG) (Iriarte *et al.* 2012). En una crisis epiléptica se pueden detectar patrones o registros denominados ictales. La actividad cerebral es altamente variable en el tiempo, pero se pueden distinguir por lo menos cuatro etapas: preictal, ictal, postictal e interictal (Yadollahpour y Jalilifar 2014). Según Gadhomi y colaboradores (2016), el reto principal en el pronóstico de las crisis epilépticas consiste en diferenciar entre las etapas preictal e interictal.

En las últimas dos décadas, algoritmos de aprendizaje automático y reconocimiento de patrones han cobrado relevancia para la detección y la predicción de crisis epilépticas. Dos de los algoritmos más utilizados en la literatura son la Máquina de Vectores de Soporte (*Support Vector Machine* - SVM) y la Red Neuronal Artificial (*Artificial Neural Network* - ANN), y sus variantes. Uno de los primeros estudios que relacionan la detección de epilepsia con las SVM fue presentado por Acir y Güzelis (2004). Shiao y colaboradores (2017) mostraron que sistemas basados en SVM pueden tener muy buenos resultados de detección de segmentos preictales e interictales, tanto en personas como en perros. Chen y colaboradores (2017) usaron la transformada de *wavelet* discreta para obtener características en el dominio del tiempo-frecuencia de señales EEG, y aplicaron SVM para el reconocimiento de patrones epilépticos.

Las ANN también han sido muy exitosas en la detección de crisis epilépticas. Srinivasan y colaboradores (2005) utilizaron un tipo de red recurrente y características en el dominio del tiempo y en el dominio de la frecuencia extraídas de señales EEG. Lograron distinguir entre registros ictales y no ictales con una exactitud mayor al 99%. Redes neuronales convolucionales y redes profundas han sido utilizadas en estudios de predicción de crisis y análisis de patrones interictales (Wang *et al.* 2020; Antoniadis *et al.* 2017).

En muchos estudios de la literatura sobre algoritmos de aprendizaje automático se hace uso de repositorios de datos

públicos. Un ejemplo de estos repositorios es el del *Center for Machine Learning and Intelligent Systems* de la Universidad de California, Irvine¹ El repositorio mantiene unas 588 colecciones de datos de todo tipo (numérico univariado, multivariado, categórico, texto, imágenes, etc.) y proveniente de muchas áreas de conocimiento. Con estos repositorios los investigadores pueden evaluar sus métodos de procesamiento y algoritmos de clasificación, aprovechando los datos generados por otros investigadores y estudios. Esto es particularmente útil, ya que muchas veces no se tiene acceso a equipo para generar datos propios. Además, el uso sistemático de los mismos datos permite una comparación objetiva de los distintos algoritmos propuestos.

Un repositorio muy utilizado para estudios biomédicos y análisis de señales bio-eléctricas es *PhysioNet*, (Goldberger *et al.* 2000), el cual contiene una amplia colección de señales y datos fisiológicos y clínicos². Zhang y Parhi (2016) y Parhi y Zhang (2019) utilizaron colecciones de este repositorio para evaluar un algoritmo de predicción de episodios epilépticos a partir de características de potencia espectral de señales EEG. Kang y colaboradores (2018), y Muñoz y Rivera (2020) usaron polisomnografías obtenidas de la colección *Sleep-EDF Database* para el estudio de patrones de sueño a partir de señales EEG.

En Guatemala, el estudio de la epilepsia está limitado a un porcentaje reducido de neurólogos. El Centro de Epilepsia y Neurocirugía Funcional, HUMANA³, es la única entidad de neurocirujanos especializados para tratamiento de epilepsia y Parkinson. Esta institución cuenta con equipo de adquisición de señales y herramientas para la visualización de estas, pero mucho del análisis de las señales se realiza de forma manual, no automática.

El objetivo del trabajo presente fue desarrollar e implementar una herramienta de *software* con una base de datos integrada para el almacenamiento, procesamiento y análisis de señales biomédicas de pacientes con epilepsia de HUMANA³. Además, se buscó sentar las bases para un repositorio de datos biomédicos.

Materiales y métodos

La figura 1 muestra la arquitectura de la herramienta de *software* que desarrollamos. Esta debía contar con tres partes fundamentales. La primera era un módulo de procesamiento, análisis y clasificación de señales EEG relacionadas a episodios epilépticos. La segunda era una base de datos relacional capaz de almacenar las señales, información relevante de pacientes y pruebas realizadas. Y la tercera era una interfaz gráfica para que los usuarios puedan hacer uso de la funcionalidad de los módulos anteriores, así como visualizar señales y resultados. Desarrollamos la herramienta en el *software* MATLAB®.

¹ UCI Machine Learning Repository es un repositorio de libre acceso con más de 580 conjuntos de datos. <https://archive.ics.uci.edu/ml/index.php>

² *PhysioNet* es un repositorio de datos médicos de acceso libre, manejado por el Laboratorio de Fisiología Computacional del MIT. <https://physionet.org/>, <https://lcp.mit.edu/>

³ <https://humanogt.org/>

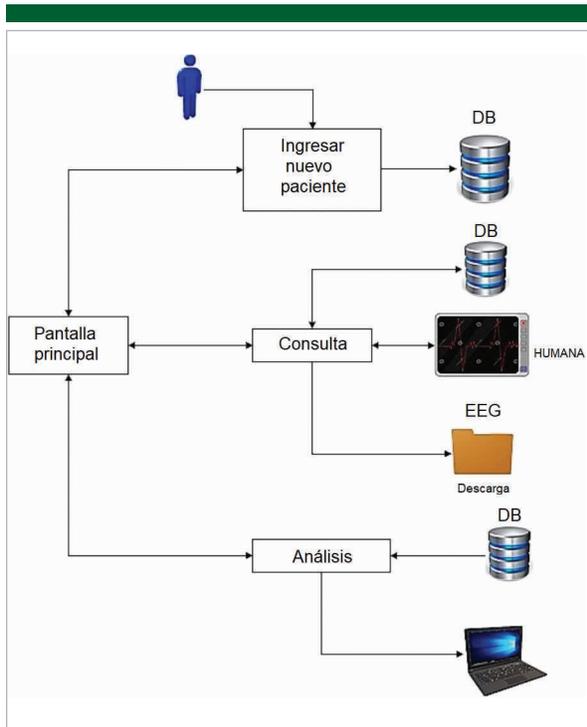


Figura 1. Arquitectura de la herramienta de software.

Módulo de procesamiento, análisis y clasificación

El objetivo primario de este módulo es aplicar métodos de aprendizaje automático para distinguir entre episodios de crisis epiléptica y estados normales (sin crisis) de forma automática, a partir de las señales EEG. La figura 2 ilustra las etapas en que

consiste el módulo. Se requieren colecciones de datos que incluyan señales tomadas en momentos de crisis y no crisis. A las señales se les extraen características, con las cuales se forman vectores de características. Los vectores de las señales de entrenamiento se utilizan para entrenar a los algoritmos clasificadores. Los vectores de las señales de prueba se alimentan a los clasificadores ya entrenados para evaluar qué tan bien logran distinguir entre una y otra clase. Si el rendimiento no es el adecuado, se ajustan los parámetros de los algoritmos y se itera, hasta lograr rendimientos aceptables.

Colecciones de datos

Para validar el módulo utilizamos 2 colecciones de señales. La primera la tomamos del repositorio de datos de la Universidad de Bonn, "UBonn EEG dataset" (Andrzejak *et al.* 2002). Este incluye 5 diferentes conjuntos de señales denotados A, B, C, D, y E; cada uno de estos conjuntos contiene señales EEG de un canal, muestreadas a 173.61 Hz. Estos segmentos fueron seleccionados y recortados de grabaciones continuas de EEG multicanal utilizando el sistema internacional 10-20 de posiciones de electrodos de EEG. Todas las señales presentes en el repositorio fueron filtradas digitalmente usando un filtro pasa altas FIR de orden 48 (ventana *Hamming*) con frecuencia de corte de 0.5 Hz. Los conjuntos A y B contienen segmentos EEG de cinco pacientes sanos (registros normales); las señales presentes en los conjuntos C y D son de 5 pacientes con epilepsia durante un intervalo normal (libre de crisis epilépticas, también considerados como registros normales) y el conjunto E contiene señales de 5 pacientes con epilepsia mientras experimentan crisis epilépticas (registros ictales).

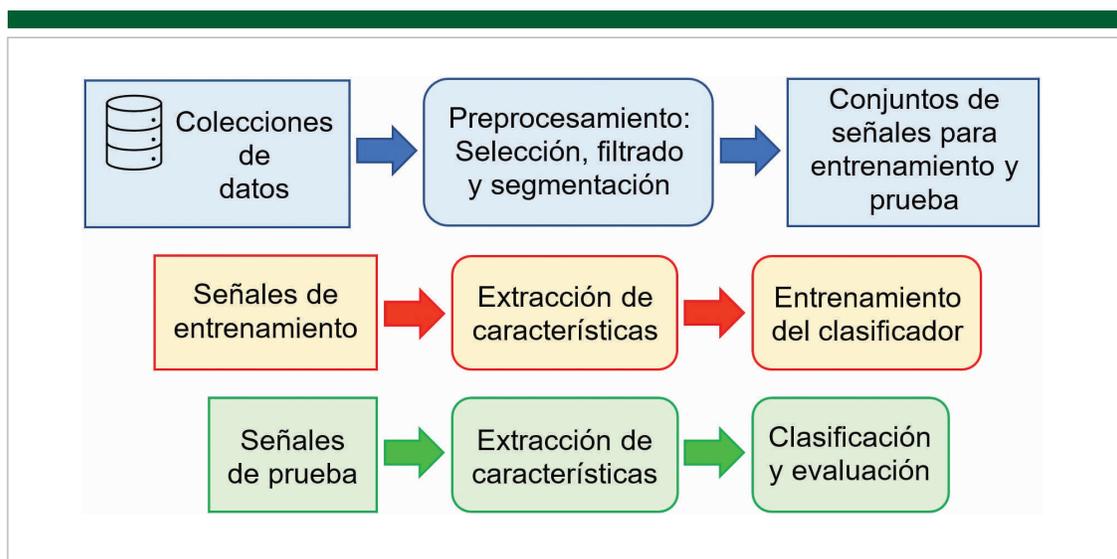


Figura 2. Etapas para el entrenamiento y evaluación de los algoritmos de aprendizaje automático.

La segunda colección de señales la tomamos de *PhysioNet*, del conjunto “CHB-MIT Scalp EEG” (Shoeb, 2009). Este conjunto contiene grabaciones EEG multicanal de pacientes pediátricos con crisis epilépticas intratables, recolectadas en el Hospital Infantil de Boston, Massachusetts. Fue creado por un equipo de investigadores del Hospital Infantil de Boston (CHB) y del Instituto Tecnológico de Massachusetts (MIT) como contribución a *PhysioNet*. Las grabaciones presentes en el conjunto incluyen 23 casos de un total de 22 pacientes (5 hombres, de 3 a 22 años; y 17 mujeres, de 1.5 a 19 años). También se utilizó el sistema internacional 10-20 de posiciones de electrodos de EEG. Las señales fueron muestreadas a 256 Hz. El número total de grabaciones es de 664 (cada uno de los 23 casos contiene entre 9 y 22 grabaciones) de las cuales 129 presentan al menos una crisis epiléptica. Las grabaciones con presencia de crisis contienen anotaciones médicas que indican el inicio y el final de la convulsión (registros ictales).

Los datos de las colecciones de señales corresponden a grabaciones de varios minutos o incluso horas de actividad EEG. Previo a la aplicación de los algoritmos de clasificación, aplicamos filtrado digital adicional a las señales. En primer lugar, un filtro quita banda de 60 Hz para eliminar el ruido introducido por la línea eléctrica. En segundo lugar, un filtro pasa bandas con frecuencias de corte de 0.5 Hz y 70 Hz, para limitar las señales al rango de análisis de interés. Luego del filtrado, necesitamos segmentar las señales completas en ventanas de tiempo más reducidas (subseñales), típicamente de unos cuantos segundos. Por conveniencia, en lo subsecuente nos referimos a estas subseñales simplemente como señales. Para un mejor desempeño de los clasificadores, la cantidad de señales de registro normal y registro ictal se balancea antes de la etapa de entrenamiento. La selección de señales de cada clase se hace de forma aleatoria.

Para la colección UBonn utilizamos los conjuntos de datos A y D como registros normales y el conjunto E como registros ictales. Obtuvimos 300 señales de ~24 s de duración, los cuales representan 40 min de grabación de registros normales de pacientes sanos, 40 min de registros normales de pacientes con epilepsia y 40 min de registros ictales de pacientes con epilepsia. Para la colección CHB-MIT, utilizamos datos de 5 pacientes. La segmentación fue hecha en ventanas de 10 s de duración. En total, obtuvimos 33 min de registros normales y 33 min de registros ictales.

Extracción de características

La herramienta permite extraer características en el dominio del tiempo y del tiempo-frecuencia de las señales EEG. Ambos tipos son ampliamente utilizados en la literatura debido al buen rendimiento de clasificación obtenido, así como el bajo costo computacional y facilidad de implementación. Las características en el dominio del tiempo que utilizamos son: cruces por cero (ZC), valor absoluto promedio (MAV), curtosis, y desviación estándar (STD).

Para el análisis en el dominio del tiempo-frecuencia utilizamos la transformada de wavelet discreta, con funciones madre Daubechies 4 y Daubechies 10, y 7 niveles de descomposición. Las funciones madre Daubechies han dado buenos resultados en la literatura (Feng et al. 2018). Obtenemos las características calculando lo siguiente a cada sub-banda wavelet: potencia, media, curtosis, STD, asimetría estadística, ZC.

Es posible seleccionar una o varias características para formar los vectores de características, tanto para la fase de entrenamiento de los algoritmos, como para la clasificación de las señales.

Algoritmos de clasificación

Como clasificadores, evaluamos los algoritmos SVM y ANN. Para la SVM utilizamos la librería libSVM (Chang y Lin 2011) compatible con MATLAB®. Probamos un *kernel* lineal y uno radial (RBF). La red neuronal fue un perceptrón multicapa de 10 capas ocultas y entrenada con el algoritmo de programación hacia atrás mediante el método de gradiente conjugado escalado. Empleamos funciones del paquete *Deep Learning* de MATLAB®.

En los experimentos que realizamos, los clasificadores son binarios. Las categorías o clases que se deben reconocer son *registro normal* y *registro ictal*. En este estudio no se buscó distinguir entre las etapas preictal, interictal y postictal. Utilizamos la técnica de validación cruzada, dividiendo los vectores de características que representan a las señales en conjuntos de entrenamiento y prueba. El desempeño de los clasificadores se presenta mediante matrices de confusión e indicadores de exactitud (porcentajes de clasificación correcta): $E(\%) = 100 \times ((VN + VP)/N)$, donde VN es verdadero negativo (registro normal clasificado como tal), VP es verdadero positivo (registro ictal clasificado como tal), y N es el total de señales clasificadas.

Un clasificador y una colección de características que pueden funcionar bien para una colección de datos pueden no funcionar tan bien para otra colección de datos. Por lo tanto, fue importante hacer pruebas con distintas combinaciones. Variamos el clasificador (RNA, SVM, con sus variantes), el número y tipo de características (en dominio del tiempo y tiempo-frecuencia), el número y el conjunto de canales EEG usados para extraer las características y el número de particiones de la validación cruzada (k).

Base de datos relacional

Para la creación y administración de la base de datos utilizamos MySQL y la herramienta *phpMyAdmin*. Usamos como servidor una computadora con sistema operativo Windows 10. La base de datos consta de 3 tablas relacionadas (ver figura 3). Consultamos a doctores de HUMANA para determinar la información relevante que debe ser almacenada en la base de datos. Esta incluye edad o fecha de nacimiento, sexo, antecedentes



Figura 3. Diagrama entidad-relación de la base de datos. Cada elemento tiene asociado un tipo de dato. Se observan 7 canales, aunque pueden ser hasta 35.

personales patológicos, diagnóstico de prescripción, si padece de epilepsia y el tipo de crisis. En la tabla principal se incluyeron los campos de dicha información para cada paciente, así como un código o ID como clave primaria. Es importante mencionar que no se incluye el nombre ni cualquier otro dato que permita la identificación de los pacientes por parte de personas externas a HUMANA. En la tabla de pruebas se agregaron campos relevantes para el módulo de análisis, como la frecuencia de muestreo y el número de canales EEG. La tabla pruebas_datos cuenta con un campo de ID de prueba y campos correspondiente a los canales. Un paciente puede tener múltiples pruebas, y una prueba múltiples datos de varios canales.

Para validar la funcionalidad de la base de datos, utilizamos una tercera colección de señales, la cual nos fue proporcionada por HUMANA. Esta colección consistió en 3 grabaciones EEG multicanal de pacientes sanos y pacientes con epilepsia. Las señales fueron muestreadas a 256 Hz y se utilizó el sistema internacional 10-20 de posiciones de electrodos. Las grabaciones fueron de duración aproximada de una hora, y corresponden a pacientes que padecen de epilepsia, que tuvieron alguna crisis durante la prueba o que padecieron de epilepsia y se encuentran en vigilancia. Los datos se obtuvieron en formato EDF, el cual es un estándar para señales biomédicas, y es el mismo formato usado en las otras dos colecciones.

Interfaz gráfica e integración de la herramienta de software

Desarrollamos la interfaz gráfica con el *App Designer* de MATLAB®. La interfaz permite el manejo y análisis de señales EEG, lo que incluye la extracción de características y la aplicación de los algoritmos descritos anteriormente. Además, permite visualizar gráficamente las señales y los resultados del procesamiento. Para integrar la base de datos a la interfaz utilizamos el *driver* ODBC. Es posible hacer consultas de datos

almacenados anteriormente, o ingresar datos nuevos. La herramienta se puede abrir directamente desde MATLAB®, pero también es posible crear un archivo .exe que puede ser ejecutado en computadoras que no cuenten con MATLAB®. Por el momento, esto sólo es posible para el sistema operativo Windows 10.

Una vez completamos la integración de los distintos módulos en la herramienta, los probamos con señales de las tres colecciones descritas anteriormente. Cabe mencionar que no usamos la colección de HUMANA para entrenar y validar los algoritmos clasificación. Esto se debió a que las señales no contaban con anotaciones de clase, por lo que se desconocía a qué tipo de registro correspondían (normal o ictal). Sin embargo, sí las usamos para validar las funciones de filtrado, extracción de características, visualización, ingreso a la base de datos, y recuperación desde la base de datos.

Resultados y discusión

Algoritmos de clasificación

La tabla 1 muestra matrices de confusión y porcentajes de clasificación de 2 casos de prueba que ilustran la aplicación de los algoritmos de clasificación. El caso 1 corresponde al clasificador SVM con un *kernel* lineal, aplicado a la colección CHB-MIT. Utilizamos 2 de los canales EEG, y las características extraídas fueron en el dominio del tiempo (ZC, MAV, curtosis y STD). Usamos el 80% de las muestras disponibles para entrenamiento, y 20% para prueba. El porcentaje de clasificación para este caso fue del 96.3%. El caso 2 corresponde al clasificador ANN, aplicado a los conjuntos D y E de la colección UBonn. Utilizamos características en el dominio del tiempo-frecuencia (potencia, media, curtosis, STD, asimetría estadística, ZC), aplicando la función wavelet Daubechies 10 y los 7 niveles de descomposición. En este caso usamos validación cruzada de $k = 4$ particiones. El porcentaje de clasificación para este caso fue del 96.7%.

Tabla 1. Matrices de confusión y porcentajes de clasificación de dos casos de prueba. Los registros normales se denotan por clase N, y los registros ictales se denotan por clase I. Los cuadros de fondo verde corresponden a verdadero negativo (N, N) y verdadero positivo (I, I), respectivamente.

Caso Ilustrativo 1				Caso Ilustrativo 2			
		Clase pronosticada					
		N	I				
Clase verdadera	N	46.3%	3.7%	92.7%	50.0%	0.0%	100%
	I	0.0%	50.0%	100%	3.3%	46.7%	93.3%
		100%	93.2%	96.3%	93.8%	100%	96.7%
		0.0%	6.8%	3.7%	6.2%	0.0%	3.3%

De todos los experimentos realizados usando las colecciones UBonn y CHB-MIT, obtuvimos porcentajes de clasificación correcta que oscilaron entre el 60% y el 100%. La variación dependió principalmente del tipo y número de características usadas, así como de los canales EEG usados. En general, los resultados fueron ligeramente mejores usando la ANN, aunque la diferencia con la SVM no fue significativa. Las características en el dominio del tiempo resultaron en un mejor rendimiento para estas colecciones.

Aumentar el número de canales, de haber más de uno disponible, y el número de características, puede incrementar la exactitud de los clasificadores. Sin embargo, eso también puede conllevar una mayor carga computacional y tiempo de ejecución al realizar el entrenamiento. Además, implica la necesidad de una mayor cantidad de señales de entrenamiento. En este estudio no realizamos experimentos rigurosos para determinar el balance óptimo entre exactitud y carga/tiempo computacional, ya que el objetivo era demostrar que la herramienta permite al usuario ajustar distintos parámetros, seleccionar distintos clasificadores, y procesar datos. Además, los parámetros óptimos que pudiesen ser encontrados para una colección de datos no necesariamente serán los óptimos para otra colección de datos.

Base de datos, interfaz gráfica e integración de la herramienta de software

La interfaz gráfica de la herramienta cuenta con una ventana de inicio y ventanas adicionales en las que se puede interactuar con la base de datos: conectar, desconectar, añadir información de nuevos pacientes, consultar información de pacientes y pruebas previamente almacenadas. También cuenta con ventanas para usar el módulo de procesamiento y visualización de datos. Por medio de botones, menús, casillas de selección y otros

elementos típicos de interfaces gráficas, se puede navegar entre las ventanas, seleccionar las opciones o parámetros deseados, y aplicar las funciones de procesamiento, entrenamiento, clasificación, gráficas. Las figuras 4 - 9 muestran las ventanas principales.



Figura 4. Ventana de inicio de la interfaz gráfica.

Figura 5. Ventana para añadir información de un nuevo paciente a la base de datos.

En la ventana de inicio (ver figura 4), sólo están habilitados los botones para conectar y analizar pruebas mientras no se haya establecido conexión con la base de datos. Es posible seleccionar archivos locales para procesar y analizar datos contenidos en ellos. No es obligatorio que el usuario utilice datos previamente almacenados en la base de datos. Al presionar el botón *Conectar*, se deben ingresar las credenciales para la base de datos. Si la conexión presenta algún error, se muestran mensajes solicitando verificar la información. Con el botón *Desconectar* se finaliza la sesión y se deshabilitan los botones para agregar pacientes y hacer consultas.

La figura 5 muestra la ventana para ingresar la información de un nuevo paciente, que consiste en un código, sexo, fecha de nacimiento, antecedentes médicos y diagnóstico. Para guardar la información, se valida que estén ingresados los campos requeridos. También se despliegan mensajes indicando si algún código ya existe dentro de la base de datos, y si la información fue guardada exitosamente.

La ventana de consultas, mostrada en la figura 6 tiene las opciones de realizar búsqueda, visualizar información, agregar nuevos archivos y descargar datos existentes en la base de datos. El usuario debe ingresar el código del paciente que desea consultar. En la tabla se muestra información de las pruebas de este paciente, si existen: código de la prueba, fecha del examen, hora de inicio, duración, frecuencia de muestreo y la cantidad de canales de las señales. Mientras no se haya realizado la búsqueda de un paciente en específico, los campos dentro de *Agregar nuevo archivo* y *Descargar pruebas* se encuentran deshabilitados. El botón *Actualizar* tiene la misma validación, ya que sólo puede actualizarse la búsqueda de un paciente previamente seleccionado.

Para seleccionar un nuevo archivo de datos que se quiera almacenar en la base de datos, primero se utiliza el botón *Seleccionar*. Se abre una ventana de navegación en la que el usuario puede revisar carpetas hasta encontrar el archivo EDF correspondiente. Una función especial⁴ se encarga de extraer la información del EDF, que incluye fecha, hora, duración, número de canales, nombre de canales, frecuencia de muestreo, valor máximo y mínimo, y las señales EEG en sí. Las señales se codifican para poder ser almacenadas en el campo prueba de la tabla pruebas, a fin de almacenarlas directamente en la base de datos y no como un archivo externo. Adicionalmente, los registros se almacenan en la tabla pruebas_datos para permitir análisis simultáneo de diferentes pruebas. Con el botón *Guardar* se carga todo a la base de datos. Datos ya almacenados se pueden descargar usando las opciones de la sección *Descargar pruebas*.

Al presionar el botón de *Analizar prueba* en la ventana de inicio, se despliega la ventana para visualización, mostrada en la figura 7. En esta ventana se pueden visualizar canales individuales de las señales EEG, en períodos de tiempo especificados por el usuario.

La ventana para extraer características, ver figura 8, permite generar vectores de características de las señales escogidas en la ventana para visualización, según las opciones escogidas por el usuario.

⁴ edfRead, creada por Brett Shoelson. <https://la.mathworks.com/matlabcentral/fileexchange/31900-edfread>

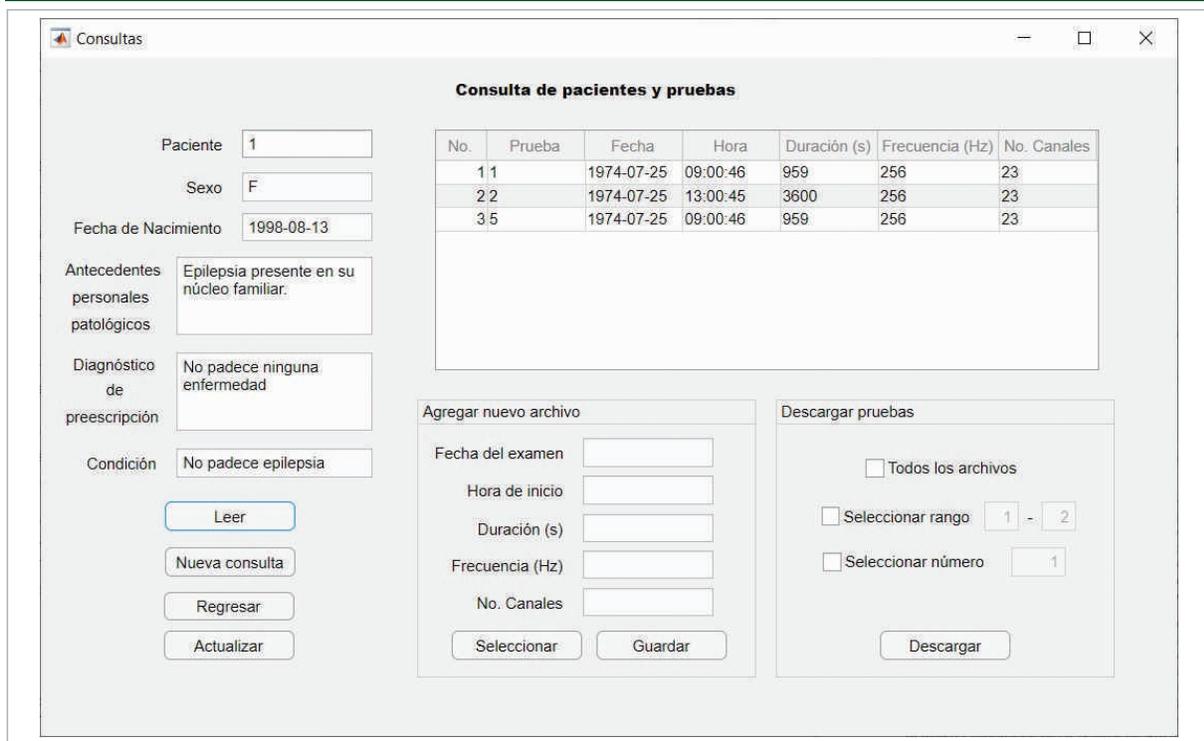


Figura 6. Ventana para consultar de información de pacientes y resultados de análisis previamente almacenados.

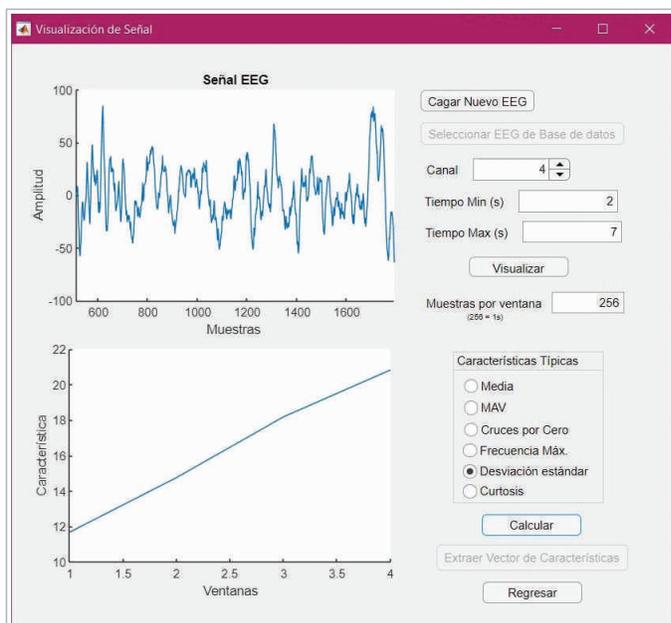


Figura 7. Ventana para visualización de señales y características calculadas.

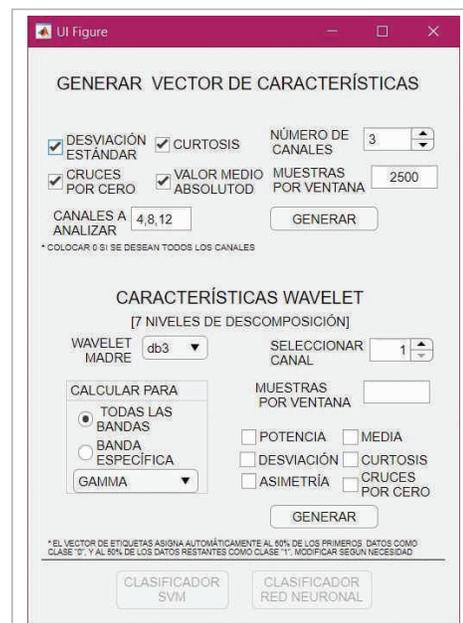


Figura 8. Ventana para selección de características a extraer de las señales EEG.

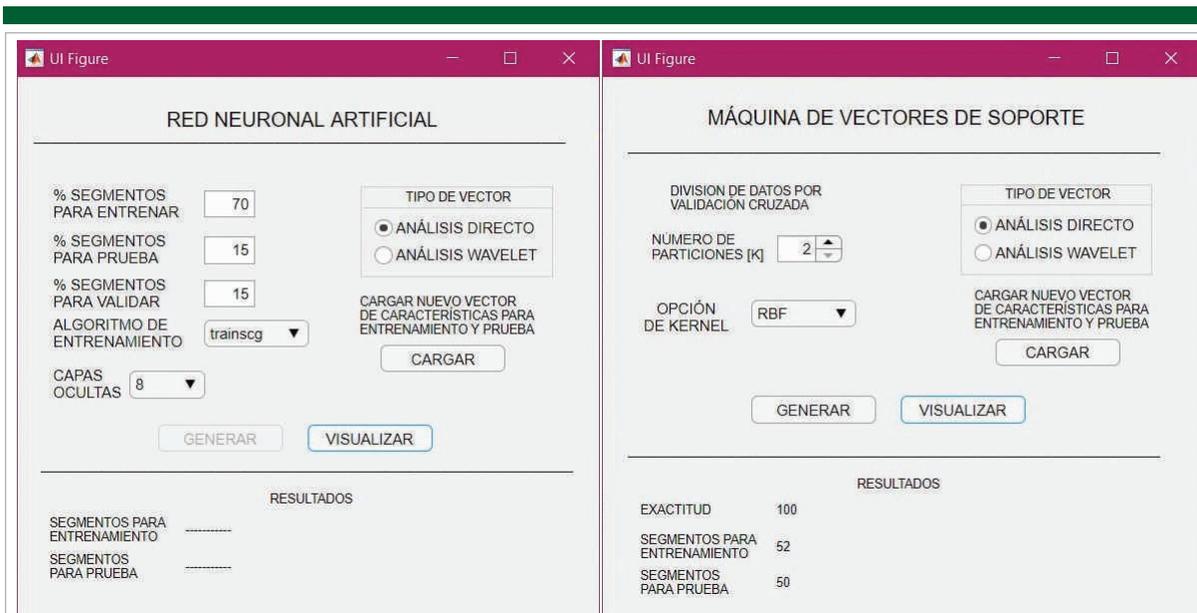


Figura 9. Ventanas para configuración de los clasificadores. La opción de análisis directo se refiere al uso de características en el dominio del tiempo.

Finalmente, la figura 9 muestra las ventanas para la configuración y aplicación de los algoritmos de clasificación. El usuario puede utilizar el botón *Cargar* para ingresar vectores de características previamente calculados, y sus correspondientes etiquetas, necesarios para el entrenamiento de los clasificadores. Es posible visualizar resultados como matrices de confusión, estructuras de la red neuronal, porcentajes de clasificación, y otros. Estos resultados aparecen en ventanas adicionales. Para ambos clasificadores, el usuario tiene la opción de guardar las estructuras de datos correspondientes, y los resultados de entrenamiento y prueba, usando el botón *GENERAR*. Todo se almacena en archivos.mat, que son estándar de MATLAB®. Estos archivos y los datos contenidos pueden ser reutilizados dentro de la herramienta, o fuera de ella, directamente en el entorno MATLAB®.

Conclusiones

La herramienta de *software* que desarrollamos permitió integrar un módulo de procesamiento y análisis de señales EEG asociadas a estudios epilépticos, y una base de datos para el manejo de información relevante, todo dentro de un mismo entorno. Es particularmente útil poder interactuar con la base de datos desde la misma herramienta, sin la necesidad de una aplicación externa. La interfaz gráfica facilitó el uso de las distintas funciones.

Los resultados de clasificación confirman que los algoritmos SVM y ANN y las características en el dominio del tiempo y tiempo-frecuencia seleccionados son adecuados para el análisis de señales EEG asociadas a la epilepsia. Es posible detectar

episodios epilépticos de forma automática con un alto grado de exactitud, siempre que se seleccionen las características adecuadas y se cuente con suficientes datos de entrenamiento. La herramienta desarrollada permite variar los parámetros y las características usadas, con el fin de encontrar la mejor combinación para cada colección de datos que se necesite analizar.

Herramientas de *software* como la presentada pueden ser de mucha utilidad para el estudio de la epilepsia. Actualmente se está utilizando en un proyecto de investigación que busca automatizar procesos de anotación y detección de segmentos de interés en las señales EEG, en HUMANA. Los algoritmos se están refinando para diferenciar entre las etapas preictal e interictal. Además, a la herramienta se le están agregando funciones para hacer análisis de conglomerados, y se está expandiendo la capacidad de almacenamiento de la base de datos. Buscamos generar un repositorio de señales, manejable desde la base de datos, que pueda ser compartido con la comunidad científica.

Aporte de autores

Autor 1 (María Jesús Angulo): Su trabajo de graduación se trató de la parte del procesamiento, análisis y clasificación de los datos. Implementó los algoritmos de aprendizaje automático (características, clasificadores, etc.), realizó los experimentos correspondientes y compiló los resultados. Contribuyó con el desarrollo de la interfaz gráfica y la integración de la herramienta. Revisora del artículo.

Autor 2 (María Fernanda Pineda): Su trabajo de graduación se trató de la parte de la base de datos. Implementó distintas versiones de la base de datos relacional, y realizó los experimentos y validaciones correspondientes. Contribuyó con el desarrollo de la interfaz gráfica y la integración de la herramienta. Revisor del artículo.

Autor 3 (Luis Alberto Rivera): Promotor del concepto de la experimentación. Asesoró a María Jesús y a María Fernanda durante el desarrollo de sus trabajos de graduación, y supervisó el análisis de los datos. Redactor del artículo y autor correspondiente.

Autor 4 (Héctor Antonio Hurtarte): Asesoró a María Fernanda Pineda durante el desarrollo de su trabajo de graduación. Revisor del artículo.

Agradecimiento

Agradecemos al Dr. Abel Sanabria, de HUMANA, por el apoyo para obtener la tercera colección de señales usados en esta investigación, y para resolver dudas sobre la epilepsia y el trabajo que se realiza en HUMANA.

Bibliografía

- Acir, N., Güzelis, C. (2004) *Automatic spike detection in EEG by a two-stage procedure based on support vector machines* Computers in Biology and Medicine 34 (7): 561-575.
- Andrzejak, R., Lehnertz, K., Mormann, F., Rieke, C., David, P., Elger, C. (2001) *Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state* Physical review. E, Statistical, nonlinear, and soft matter physics 64(6 Pt 1):061907.
- Antoniades, A., Spyrou, L., Martin-Lopez, D., Valentin, A., Alarcon, G., and Saneil, S., Cheong Took, C. (2017) *Detection of Interictal Discharges With Convolutional Neural Networks Using Discrete Ordered Multichannel Intracranial EEG* IEEE Transactions on Neural Systems and Rehabilitation Engineering 25 (12): 2285-2294.
- Chang, C.-C., Lin, C.-J. (2011) *LIBSVM: A library for support vector machines* ACM Transactions on Intelligent Systems and Technology 2 (3) 27: 1-27.
- Chen, D., Wan, S., Bao, F.S. (2017) *Epileptic Focus Localization Using Discrete Wavelet Transform Based on Interictal Intracranial EEG* IEEE Transactions on Neural Systems and Rehabilitation Engineering 25 (5): 413-425.
- Feng, L., Li, Z., Wang, Y. (2018) *VLSI Design of SVM-Based Seizure Detection System With On-Chip Learning Capability* IEEE Transactions on Biomedical Circuits and Systems 12(1): 171-181.
- Gadhoumi, K., Lina, J.-M., Mormann, F., Gotman, J. (2016) *Seizure prediction for therapeutic devices: A review.* Journal of Neuroscience Methods 260: 270-282.
- Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P.C., Mark, R., Mietus, J.E., Moody, G.B., Peng, C.K., Stanley, H.E. (2000) *PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals* Circulation [Online]. 101 (23): 215-220.

Iriarte, J., Urrestarazu, E., Alegre, M., Artieda, J. (2012) *Electroencefalografía del Adulto Atlas bilingüe español-inglés basado en casos clínicos* Elsevier España 808 pp.

Kang, D.Y., DeYoung, P.N., Malhotra, A., Owens, R.L., Coleman, T.P. (2018) *A State Space and Density Estimation Framework for Sleep Staging in Obstructive Sleep Apnea* IEEE Transactions on Biomedical Engineering, 65 (6): 1201-1212.

Muñoz, J.P., Rivera, L.A. (2020) *Towards Improving Sleep Quality Using Automatic Sleep Stage Classification and Binaural Beats* Proceedings of the 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) pp. 4982-4985.

Parhi, K.K., Zhang, Z. (2019) *Discriminative Ratio of Spectral Power and Relative Power Features Derived via Frequency-Domain Model Ratio With Application to Seizure Prediction* IEEE Transactions on Biomedical Circuits and Systems 13 (4): 645-657.

Shiao, H.-T., Cherkassky, V., Lee, J., Veber, B., Patterson, E.E., Brinkmann, B.H., Worrell, G.A. (2017) *SVM-Based System for Prediction of Epileptic Seizures From iEEG Signal* IEEE Transactions on Biomedical Engineering, 64 (5): 1011-1022.

Shoeb, A. (2009) *Application of Machine Learning to Epileptic Seizure Onset Detection and Treatment* Tesis de doctorado en Filosofía, Departamento de Ciencia y Tecnología de la Salud, Massachusetts Institute of Technology.

Srinivasan, V., Eswaran, C., Sriraam, N. (2005) *Artificial Neural Network Based Epileptic Detection Using Time-Domain and Frequency-Domain Features* Journal of Medical Systems 29: 647-660.

Thijs, R.D., Surges, R., O'Brien, T.J., Sander, J.W. (2019) *Epilepsy in adults* Lancet 393 (10172): 689-701.

Wang, G., Wang, D., Du, C., Li, K., Zhang, J., Liu, Z., Tao, Y., Wang, M., Cao, Z., Yan, X. (2020) *Seizure Prediction Using Directed Transfer Function and Convolution Neural Network on Intracranial EEG* IEEE Transactions on Neural Systems and Rehabilitation Engineering 28 (12): 2711-2720.

Yadollahpour A, Jalilifar M. (2014) *Seizure Prediction Methods: A Review of the Current Predicting Techniques* Biomedical and Pharmacology Journal 7 (1): 153-162.

Zhang, Z., Parhi, K.K. (2016) *Low-Complexity Seizure Prediction From iEEG/sEEG Using Spectral Power and Ratios of Spectral Power* IEEE Transactions on Biomedical Circuits and Systems 10 (3): 693-706.