



Predicción de la fase pre-ictal de convulsiones en pacientes con epilepsia a partir de señales electroencefalográficas y electrocardiográficas

John Jairo Martinez Saiz

Universidad Antonio Nariño

Facultad de ingeniería mecánica, electrónica y biomédica

Bogotá, Colombia

2021

Predicción de la fase pre-ictal de convulsiones en pacientes con epilepsia a partir de señales electroencefalográficas y electrocardiográficas

John Jairo Martinez Saiz

Proyecto de grado presentado como requisito parcial para optar al título de:

Ingeniero Biomédico

Director:

PhD. Mario E. Duarte

Línea de investigación:

Procesamiento de señales Biomédicas

Universidad Antonio Nariño

Facultad de ingeniería mecánica, electrónica y biomédica

Bogotá, Colombia

2021

Scientists study the world as it is; engineers create the world that has
never been.

Theodore Von Karman

Agradecimientos

Tengo tanto por agradecer a esas personas, compañeros, amigos, familia como primo Cristian Mendoza, por el apoyo incondicional, constante, de principio a fin. Sin ellos difícilmente este proyecto se hubiera culminado satisfactoriamente.

Por razones como esas y más... es que les digo de manera más sincera, ¡muchas gracias!

Mario Duarte, gracias por el apoyo, la constancia y estar hasta el final con este proyecto. Más que un tutor, un amigo constante.

Mi familia, siempre con la fé intacta y con el mejor ánimo impulsaron terminar mi proyecto para seguir creciendo como persona íntegra.

Natalia Villamizar, una compañera y amiga, el apoyo en diferentes etapas del proyecto finalizado.

A mi madre, por su constante ayuda y esfuerzo mutuo para la culminación de este proyecto.

Resumen

Las convulsiones son perjudiciales para los pacientes, los cuales, sin la predicción oportuna, pueden llegar a la muerte. Por lo tanto, tener algoritmos que indiquen cuándo va a ocurrir un episodio epiléptico brinda seguridad y tiempo de acción para actuar.

El presente trabajo, se enfoca en la predicción de convulsiones en pacientes con epilepsia a partir de señales de electroencefalografía (EEG) y Electrocardiografía (ECG). El estudio se realizó en pacientes que sufrieron convulsiones y se implementan algoritmos de **Machine Learning** para la predicción de la fase pre-ictal de convulsiones usando la herramienta “classification learner” de Matlab.

Para el desarrollo del trabajo, se utilizó la metodología CRISP-DM, con la cual se logró extraer características de 10 pacientes con la finalidad de entrenar diferentes algoritmos de clasificación.

Se consideró las señales EEG y EKG en conjunto y por separado para mostrar cuál de las dos obtuvo un mejor rendimiento según las métricas computadas a partir de la matriz de confusión. Se mostró que la mejor sensibilidad fue obtenida cuando se trabajó con las características extraídas de las EEG y EKG en conjunto.

Palabras Clave: Machine Learning, electroencefalografía, electrocardiografía, metodología CRISP-DM.

Abstract

Seizures are harmful to patients, who, without timely prediction, can lead to death. Therefore, having algorithms that indicate when an epileptic episode is going to occur provides security and action time to act.

The present work focuses on the prediction of seizures in patients with epilepsy from electroencephalography (EEG) and Electrocardiography (ECG) signals. The study was carried out in patients who suffered seizures and Machine Learning algorithms were implemented for the prediction of the pre-ictal phase of seizures using the "Class Learner" tool from Matlab.

For the development of the work, the CRISP-DM methodology was used, with which characteristics of 10 patients can be extracted in order to train different classification algorithms.

The EEG and EKG signals were considered together and separately to show which of the two obtained better performance according to the metrics computed from the confusion matrix. It was shown that the best sensitivity was obtained when the characteristics extracted from the EEG and EKG were worked together.

Key Words: Machine Learning, electroencephalography, electrocardiography, CRISP-DM methodology.

Contenido

Resumen.....	9
Abstract.....	10
Lista de figuras.....	14
Lista de tablas.....	15
1 Introducción.....	17
1.1 Estado del Arte.....	19
1.1.1 Modelos usando EEG.....	20
1.1.2 Modelos usando ECG.....	21
1.1.3 Modelos usando EEG y ECG.....	21
1.2 Planteamiento Del Problema.....	22
1.3 Justificación.....	23
1.4 Objetivos.....	24
1.4.1 Objetivo General.....	24
1.4.2 Objetivos Específicos.....	24
2 Marco Teórico.....	27
2.1 Sistema Interfaz cerebro-computadora (BCI).....	27
2.1.1 Componentes del sistema BCI.....	27
2.2 Electrocardiografía.....	29
2.2.1 Importancia de la EKG en la epilepsia.....	31
2.3 Electroencefalografía.....	32
2.3.1 Importancia del EEG en la epilepsia.....	36
2.4 Sistema internacional 10-20 para la adquisición de señales EEG.....	36
2.5 Metodología Crisp-DM.....	40
2.5.1 Fases de la metodología Crisp-DM.....	42
2.5.1.1 Entendimiento del negocio.....	42
2.5.1.2 Entendimiento de los datos.....	43
2.5.1.3 Preparación de datos.....	43
2.5.1.4 Modelado.....	44
2.5.1.5 Evaluación.....	45
2.6 Modelos entrenados.....	46
2.6.1 Árbol de decisiones.....	46
2.6.2 Análisis discriminante (AD).....	46
2.6.3 Clasificadores de Regresión logística.....	47
2.6.4 Clasificador Naïve Bayes (NBC).....	47
2.6.5 Maquina de soporte Vectorial (MSV).....	47

2.6.6 Clasificador K-Vecino más Cercano (KNN).....	48
2.6.7 Combinación de clasificadores.....	48
3 Metodología.....	51
2.6 Business Understanding	51
2.7 Data Understanding.....	51
2.7.3 Información de la base de datos.	51
3.3 Data Preparation	54
3.3.1 Selección de Datos	54
3.3.2 Selección de tiempos.....	54
3.3.3 Filtros usados.....	56
3.3.3.1 Common average reference.....	57
3.3.4 Extracción de características.....	57
2.8 Modeling	58
3.3.1 Configuración de Clasification Learner	58
3.3.2 Modelos Usados.....	59
3.4 Evaluación.....	60
3.4.1 Matrices de confusión.....	60
4 Resultados	63
4.1 Resultados obtenidos.....	63
5. Conclusiones y recomendaciones.....	72
5.1 Conclusiones	72
5.2 Recomendaciones.....	72

Lista de figuras

Figura 1. Componentes del sistema BCI.....	29
Figura 2. Ondas de electrocardiografía.....	30
Figura 3. Sistema de conducción del corazón.....	31
Figura 4. Onda Delta.....	34
Figura 5. Onda Theta.....	34
Figura 6. Onda Alfa.....	34
Figura 7. Onda Beta.....	35
Figura 8. Onda Gamma.....	35
Figura 9. Sistema internacional 10-20.....	37
Figura 10. A) Vista de perfil. B) Vista superior. FP, punto polar; O, punto occipital.....	37
Figura 11. A) Vista de perfil. B) Vista superior. Fz, punto frontal; Cz, punto central; Pz, Punto parietal.....	38
Figura 12. Medición coronal lateral. A) Vista de frontal. B) Vista superior. Situación de los electrodos T3 y T4.....	38
Figura 13. A) Vista de perfil. B) Vista superior. Situación de los electrodos C3 y C4; F3 y F4; P3 y P4.....	39
Figura 14. A) Vista de perfil. B) Vista superior. Situación de los electrodos FP1 y FP2, O1 y O2, F7 y F8, T5 y T6, A1 y A2.....	40
Figura 15. Grado de utilización de las distintas metodologías de minería de datos.....	41
Figura 16. Secuencia de proceso Crisp-DM.....	41
Figura 17. Señal EKG del paciente PN11 con delimitaciones de fase ictal, pre-ictal y no pre-ictal.....	55
Figura 18. Señal EEG del paciente PN11 sin filtro de bandas.....	55
Figura 19 correlograma de todas las señales EEG y EKG.....	70

Lista de tablas

Tabla 1. Algoritmos con su exactitud.....	19
Tabla 2. Rango de frecuencia de las diferentes ondas	33
Tabla 3. Zonas Craneales	40
Tabla 4. señales almacenadas en los archivos. Edf.....	52
Tabla 5. Señales EKG	53
Tabla 6. Filtros usados para las bandas.....	56
Tabla 7. Valor de matrices de confusión.	64
Tabla 8. Métricas de los modelos de la Tabla 9.	67
Tabla 9. PCA junto al modelo RusBoosted Trees	69
Tabla 10. Correlograma de Variables.....	70
Tabla 11. Matriz de confusión de EEG y EEG con filtro de banda en conjunto, y EKG sola.	71

1. Introducción

Las señales electroencefalográficas (EEG) son oscilaciones que aparecen como potenciales eléctricos en la superficie cerebral, estas pueden ser adquiridas por medio de electrodos. En 1895, Hans Berger adquirió por primera vez las señales EEG con la finalidad de detectar la epilepsia, daños cerebrales y algún tipo de tumor en el cráneo (Koertge, N. 2008). En la actualidad, se usan sistemas BCI (Brain Computer Interface) para digitalizar señales electroencefalográficas, con el fin de procesarlas y analizarlas en un computador (Moreno, I., Batista, E., Serracin, S., Moreno, R., Gómez, L., Serracin, J., ... & Boya, C. 2019).

La electrocardiografía es la actividad eléctrica del corazón, la cual sirve para detectar problemas cardiacos. Einthoven observó las ondas que denominó como P, Q, R, S, T y U. Donde P a nivel del corazón mide la activación auricular, QRS es el procedido de la onda P la cual es excitación de la musculatura de ambos ventrículos y al tabique interventricular. S y T indican la despolarización de la onda y U aparece después de la onda T donde es la repolarización de los músculos capilares (Zavala-Villeda & José Alfredo 2017).

Cuando ocurre una convulsión de manera espontánea se considera que el paciente en ese momento está en fase ictal. El conocimiento acerca de cómo se generan las convulsiones permite comprender algunos factores específicos e inespecíficos que afectan los patrones de la aparición de convulsiones, información que puede ser útil para el manejo del paciente (Engel, J. 2013). Cuando un paciente que está en estado epiléptico refractario (EER), no es tratado a tiempo son propensos a muerte por convulsiones (Nariño, D. 2003)

La epilepsia se define como un trastorno cerebral causada por descargas desordenadas de las neuronas cerebrales (Sánchez, L. P. 2020). Las crisis epilépticas se clasifican de la siguiente manera:

- Crisis parcial: las descargas desordenadas son generadas en sólo un hemisferio o lóbulo cerebral.

La crisis parcial se puede clasificar en:

- Simples: La consciencia de los pacientes no es afectada.
- Complejas: La consciencia de los pacientes es afectada.
- Crisis generalizada: las descargas desordenadas son generadas en ambos hemisferios.

Se presentan los siguientes síntomas en simples y complejas: Signos vegetativos, sensitivos, visuales, psíquicos, alucinaciones olfativas o auditivas, automatismos y alteración del lenguaje

La crisis generalizada se puede clasificar en (Nieto, M., Nieto, M., & Nieto, E. 2008).

- Convulsivas.
- Mioclonías: sacudidas musculares breves, que dura de 20 a 120 minutos.
- Tónicas: Contracción de todos los músculos y extremidades lo cual genera la caída del paciente.
- Clónicas: Contracciones musculares de todo el cuerpo violentamente.
- Tónico-Clónicas: Consiste en la unión de los dos, causa caída del paciente seguido de movimientos violentos
- Atónicas: Generan pérdida del tono postural lo que puede generar caída del paciente.
- No convulsivas: El paciente pierde consciencia, queda ausente con la mirada perdida y luego recupera la consciencia, el paciente puede presentar mioclonías.

En (Yamada, M. 2013). se determinó la asociación entre el estado convulsivo y el metabolismo óseo, relacionando el estado convulsivo y el historial de fracturas. Se encontró que el estado de convulsiones incontroladas es un factor de riesgo significativo para la alteración del metabolismo óseo.

Las personas epilépticas se enfrentan a labores cotidianas donde predecir las convulsiones de estos pacientes sería de gran ayuda para evitar traumas. Por ejemplo, riesgo de ahogo en la ducha, accidentes de tránsito, traumas en neonato o abortos en mujeres en embarazo, entre otros (Larson, D. E. 1990).

1.1 Estado del Arte

En (Forney, E. M. 2011).se propuso un algoritmo para la predicción de desorden en las neuronas a partir de señales EEG con la interfaz BCI, basado en Red neuronal recurrente de Elman (ERNN), donde la exactitud fue definida en diferentes modelos. El modelo 1 fue evaluado con una y dos capas, obteniendo una exactitud de 71% y 74%, respectivamente.

En (Feltane, A. 2016), se desarrolló y probó algoritmos para la detección de ataques epilépticos que están en la Tabla 1. Para comparar la eficacia de los algoritmos en los datos registrados con electrodos de disco convencionales (EEG) versus nuevos electrodos de disco de anillo concéntrico tripolar (tEEG).

Tabla 1. Algoritmos con su exactitud

	Sensitivity (%)	Specificity (%)	Accuracy (%)
tEEG data from rats			
AdaBoost	90.91	78.72	84.81
SVM	93.02	100	96.51
Average	91.96	89.36	90.66
EEG data from humans [2]			
AdaBoost	100	96.88	98.44
SVM	100	100	100
Average	100	98.44	99.22

Nota Feltane, A. 2016.

En Idiáquez, J., Rivera, E., Santin, J., & Godoy, J. (2009), se demostró que las alteraciones cardiacas asociadas en pacientes epilépticos ocurren aproximadamente en 40% de pacientes con este trastorno, donde la mayoría son taquicardias sinusales (90%), también ocurren despolarizaciones auriculares prematuras, complejos ventriculares prematuros, bradicardias, pausas sinusales y asistolias. Donde demuestra que estas alteraciones en pacientes en EER tienen mayor mortalidad que pacientes en EE.

1.1.1 Modelos usando EEG

En (Olivares, V. M. N. 2014), extrajeron de la señal EEG las características propias de una persona sana, una persona epiléptica o una persona que está sufriendo un ataque epiléptico; observaron cómo la amplitud de la señal EEG en un ataque epiléptico es mayor con respecto a una persona sana. Clasificaron todas las sub-bandas y la desviación estándar, identificando que en las bandas de frecuencias delta y alfa se tiene una mayor variabilidad durante un ataque epiléptico.

En (Usman, S. M., Usman, M., & Fong, S. 2017), se propone un modelo que proporciona métodos confiables tanto de preprocesamiento como de extracción de características para la predicción de la fase pre-ictal de las convulsiones. Se extrajeron características del dominio de tiempo y frecuencia para entrenar un modelo de predicción, con un tiempo máximo de anticipación de 33 minutos. Concluyeron que el modelo propuesto logró tener una sensibilidad y especificidad promedio de 92,7% y 90,8% respectivamente. Por lo anterior se espera que incluir señales EKG mejoren estos valores.

En (García López, B. 2015), se tomaron 175 casos de alteraciones intercríticas de las señales EEG del lóbulo frontal, donde dividieron las señales en tres apariencias morfológicas de estudio de la actividad intercrítica: su morfología (puntas, ondas agudas simples, ondas agudas difásicas, polipuntas, ondas lentas irregulares), su voltaje (menor de 50 μV , entre 50-100 μV , 100-200 μV y más de 200 μV) y la actividad. Concluyen que el ritmo reclutante, el cual es la característica morfológica que más se observa en la señal EEG, se observa al inicio de las crisis epilépticas. Adicionalmente, en el lóbulo frontal se evidencia una disminución en la amplitud de las señales EEG en la fase proscritica y, por lo tanto, fue el principal apoyo para localizar el foco de estas descargas desordenadas. Por último, se reporta que las señales ECG sufren cambios al momento de un caso de epilepsia.

1.1.2 Modelos usando ECG

En (Cassater, D. 2017), el autor, Cassater Daiana, muestra el uso de señales ECG para la predicción de epilepsia en pediatría, debido a que el sistema cardiovascular y nervioso comparten unas estructuras moleculares llamadas: canales iónicos. El autor asegura que se puede identificar, en los canales iónicos, los casos de epilepsia en el paciente donde se concluye que las señales EEG y ECG están correlacionados para la detección de la epilepsia.

Por otra parte, en (Ghosh, A., Sarkar, A., Das, T., & Basak, P. 2017), tomaron señales ECG de pacientes sin antecedentes de ataques epilépticos y pacientes con antecedentes de ataques epilépticos con el fin de detectar la QRS y extraer los picos R para luego calcular los intervalos entre cada R. De las señales ECG seleccionaron la fase pre-ictal 5 minutos antes de la fase ictal para el intervalo RR extrajeron la entrada para el dominio del tiempo, de la frecuencia y análisis de índice no lineal con el algoritmo Pan-Tompkins; Concluyeron que el sistema nervioso autónomo SNA influye en el sistema cardiovascular y especialmente en la frecuencia cardiaca.

Teniendo en cuenta el trabajo de (Ghosh, A., Sarkar, A., Das, T., & Basak, P. 2017) nos indican como el corazón es alterado en respuesta a una convulsión que sufre una persona epiléptica y (Cassater, D. 2017) como las señales EEG y EKG están unidas por canales iónicos por esto se decide utilizar las señales EKG para observar si los predictores mejoran.

1.1.3 Modelos usando EEG y ECG

En (Meisel, C., & Bailey, K. A. 2019), usan algoritmos de aprendizaje profundo para la detección de convulsiones en pacientes epilépticos a partir de señales EEG, ECG y ECoG. Concluyeron que para la información pre-ictal, un único canal de ECG contiene una cantidad comparable de información que los 21 canales de EEG; y que el algoritmo de aprendizaje profundo superó de manera general algoritmos como: k-NN, SGD, and AdaBoost donde la exactitud de clasificación promedio de todos los sujetos fue del 77,1%. Donde el autor uso señales de ECoG las cuales son invasivas. En el anterior trabajo se realiza el

aprendizaje profundo comparando solo tres algoritmos, y debido a esto surge la pregunta si este mismo comportamiento se puede observar al usar otros algoritmos de Inteligencia artificial.

En (Detti, P., Vatti, G., & Zabalo Manrique de Lara, G. 2020), se presenta una base de datos que consta de registros de señales EEG de 14 pacientes adquiridos en la Unidad de Neurología y Neurofisiología de la Universidad de Siena, con una frecuencia de muestreo de 512 Hz y 1 o 2 señales ECG considerando que los sistemas de predicción de la fase pre-ictal aún son tema de investigación se realizó un estudio con algoritmos de Machine Learning.

1.2 Planteamiento Del Problema

En las variaciones de la onda de las señales EEG y ECG es difícil identificar cambios significativos en la fase pre-ictal de las convulsiones. Por lo anterior, los autores, Acharya, U Rajendra and Sree, S Vinitha and Swapna, G and Martis, Roshan Joy and Suri, Jasjit S, recomiendan el uso de algoritmos como: Máquina de vectores de soporte (SVM), impulso adaptativo (AdaBoost), entre otros, para tener una predicción eficaz y con menor margen de error.

En (Meisel, C., & Bailey, K. A. 2019) hacen uso del aprendizaje profundo comparándolo con de tres algoritmos de aprendizaje en Machine Learning, se decidió verificar si el hecho se mantiene para otros algoritmos de aprendizaje o cambia, esto es debido al tiempo de trabajo y complejidad que tiene el aprendizaje profundo.

En (Cassater, D. 2017), nos indican como las señales EKG pueden servir como predictores para las convulsiones, por esta razón se decidió tomar las señales EKG y EEG en conjunto para conocer si el uso de los algoritmos de aprendizaje mejoraba esto con el fin de predecir casos de convulsiones en pacientes con epilepsia.

Según (Rosas-Romero, R., Guevara, E., Peng, K., Nguyen, D. K., Lesage, F., Pouliot, P., & Lima-Saad, W. E. 2019). la epilepsia tiene cuatro estados, los cuales son:

- El ictal que es cuando el paciente está sufriendo indicios de convulsiones.
- El pre-ictal es la fase que precede a las convulsiones.
- El inter-ictal corresponde al periodo en que el paciente está terminando de tener su crisis convulsiva
- El post-ictal corresponde al periodo cuando el paciente supera la crisis.

Adicionalmente, un paciente es diagnosticado con estado epiléptico refractario (EER) cuando los medicamentos habituales para las convulsiones no funcionan, por ende, identificar oportunamente el estado pre-ictal de las convulsiones sería un proceso eficaz para poder proporcionar el sedante al paciente y que éste no sufra una crisis involuntaria que podría generarle daños. Los medicamentos dejan de hacer efecto entre el 10% al 40% de los pacientes (Hernández, O. H., Guerra, A., Arango, J. C., Ciro, J. D., Delgado Ch, H., Uribe, C. S., ... & Jiménez, M. E. 2011). Generalmente, el personal médico no logra evidenciar la fase pre-ictal, inclusive cuando los pacientes están hospitalizados. Por lo anterior, se han usado diferentes métodos a partir de señales EEG y ECG para la predicción de la fase pre-ictal de las convulsiones debido a la mortalidad elevada en los pacientes EER.

Por lo anterior, en este proyecto se plantea la predicción de la fase pre-ictal en pacientes con epilepsia utilizando algoritmos de Machine Learning a partir de señales Electroencefalográficas y Electrocardiográficas con base a la siguiente base de datos (Detti, P., Vatti, G., & Zabalo Manrique de Lara, G. 2020)

1.3 Justificación

En (Rufo Campos, M. 2007), el Estado Epiléptico (EE) es considerado una emergencia en el ambiente médico por su alta morbilidad y mortalidad, por esto requiere un manejo oportuno y atención inmediata. La mortalidad del EE varía entre 17 y 26%. Para los pacientes que sobreviven, el 10% a 23% presentan alguna alteración neurológica.

Las convulsiones son perjudiciales para los pacientes EER (Hernández, O. H., Guerra, A., Arango, J. C., Ciro, J. D., Delgado Ch, H., Uribe, C. S., ... & Jiménez, M. E. 2011)., los cuales, sin la predicción oportuna, pueden llegar a la muerte.

Tener algoritmos que indiquen cuándo va a ocurrir un episodio epiléptico brinda seguridad y tiempo de acción para actuar. Adicionalmente, de acuerdo a (Mason, E. R. 2020), cada día hay más de 250 mutaciones en los tejidos, donde la hipótesis, apoyada por datos electrofisiológicos, es que están apareciendo cada día más pacientes EER. Los resultados sugieren que las mutaciones generan epilepsia a través de distintos efectos biofísicos que pueden requerir farmacoterapias novedosas o adaptadas para un tratamiento eficaz.

En (Téllez-Zenteno, J. F., Nguyen, R., & Hernández-Ronquillo, L. 2010)., se identifica que los pacientes que ya han tenido crisis epilépticas, y no volvieron a presentar estas crisis tienen un 7 % de probabilidad de sufrir un accidente automovilístico. Por otra parte, los pacientes que padecen de crisis epilépticas tienen un 26% de probabilidad de sufrir un accidente automovilístico, porque el tiempo de reacción es muy corto.

Existen mujeres embarazadas que también sufren crisis epilépticas, en las cuales pueden causar a la madre y al feto hipoxia y acidosis, e inclusive, pueden llevar a la muerte. En (Nariño, D. 2003), se reporta la aparición de malformaciones congénitas, correspondiente al 12.3% de neonatos de madres que padecieron crisis.

1.4 Objetivos

1.4.1 Objetivo General

Predecir la fase pre-ictal en pacientes con epilepsia utilizando algoritmos de **Machine Learning** a partir de señales Electroencefalográficas y Electrocardiográficas.

1.4.2 Objetivos Específicos

- Construir una base de datos de las señales EEG y ECG que corresponden y no corresponden a la fase pre-ictal de convulsiones.

- Obtener un conjunto de características de las señales EEG y ECG relevantes para la predicción de la fase pre-ictal de convulsiones.
- Usar algoritmos de **Machine Learning** para la predicción de la fase pre-ictal de convulsiones.
- Evaluar el desempeño de los algoritmos para la predicción de la fase pre-ictal de convulsiones.

2 Marco Teórico

2.1 Sistema Interfaz cerebro-computadora (BCI)

La interfaz BCI el autor (Detti, P., Vatti, G., & Zabalo Manrique de Lara, G. 2020). La utilizo para crear una base de datos la cual fue usada en este proyecto donde ellos utilizaron esta interfaz para poder extraer la información del cerebro de cada paciente, con el fin de transformar esa señal de análogo a digital para luego poder entrenar algoritmos para la predicción de la fase pre-ictal.

El sistema cerebro-computadora con sus siglas en ingles BCI o (Brain Computer Interface) es el sistema que se encarga de capturar manifestaciones eléctricas, magnéticas o de otro tipo de actividad cerebral, las cuales son interpretadas y ejecutadas por una computadora u otro dispositivo (Gutiérrez-Martínez, J. 2013).

El sistema es considerado como una herramienta de gran utilidad para colocar la comunicación, restablecer funciones y ofrecer procesos de rehabilitación a pacientes con discapacidad motora (Gutiérrez-Martínez, J. 2013).

2.1.1 Componentes del sistema BCI

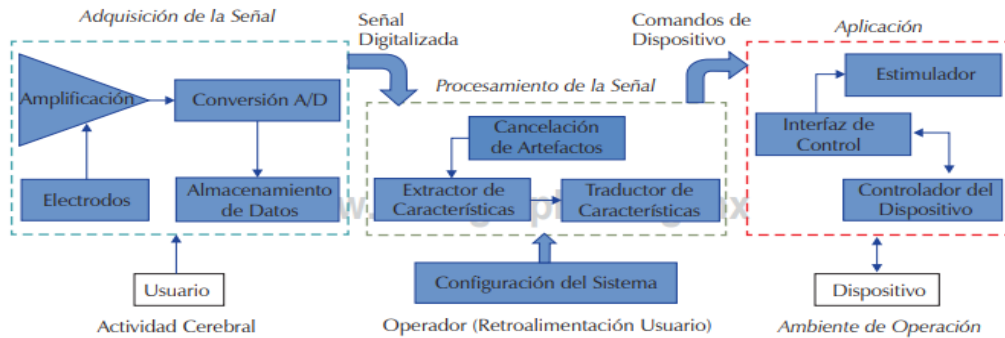
La figura 1 muestra el modelo funcional del sistema BCI (Gutiérrez-Martínez, J. 2013)."

- Adquisición de la señal y acondicionamiento de la señal. Captura la actividad eléctrica generada de forma analógica por medio de electrodos para su luego transformación a digital. La información suministrada por la

- base de datos es una señal análoga que posteriormente fue convertida a señal digital para el tratamiento de las señales.
- Procesamiento de la señal. Recibe la señal eléctrica cerebral digitalizada y la transforma en comandos que entiende el dispositivo y que el usuario desea controlar con el fin de poder manipularla con el programa llamado Matlab. Este bloque se divide en tres etapas que actúan de forma secuencial.
 - A. Cancelación de artefactos. Se encargan de eliminar artefactos debidos a otros tipos de actividades eléctricas producto del movimiento ocular, muscular, el producido por la línea eléctrica o las señales de electrodos especialmente correlacionados.
 - B. Extracción de características. En esta etapa se traduce la señal cerebral de entrada en un conjunto de características correlacionado con el fenómeno asociado a la señal, se extrae las características de la base de datos con el fin de trabajar con estas en los algoritmos de Machine Learning.
- Interfaz de control. Donde recoge las instrucciones de control y realiza las acciones correspondientes en el dispositivo. Puede incorporar una pantalla que proporcione retroalimentación al usuario, donde en esta interfaz se realiza el procesamiento de las señales de la base de datos para extraer datos que se utilizaron para la finalidad.
- Configuración. Permite definir y ajustar parámetros del sistema. El operador no tiene que ser una persona que sepa del tema. Sino que puede ser el propio usuario o en el caso más deseable, algoritmos automáticos que ajustan el comportamiento del sistema en función de los resultados obtenidos y la retroalimentación del usuario, esto con el fin de escoger los tiempos a trabajar con la base de datos.
- Dispositivo. Existe un rango limitado de dispositivos que puedan ser usados en un sistema BCI, como computadoras, sintetizadores de voz, neuroprótesis y otros objetos en el ambiente del usuario como la televisión o encender o apagar la luz de la habitación. Estos dispositivos usualmente

- son físicos, pero también pueden ser virtuales, donde los dispositivos usados fueron computador, y Matlab como dispositivo virtual.
- Ambiente de operación. Se refiere al ambiente físico esta información puede alterar la información que se encuentra la base de datos con la que se va trabajar.”

Figura 1. Componentes del sistema BCI.



Nota Gutiérrez-Martínez, J. 2013

2.2 Electrocardiografía

Las señales de electrocardiografía en este proyecto son usadas debido a que la literatura nos informa que las señales EEG y EKG tienen alteraciones cuando un paciente con epilepsia sufre una convulsión y con la finalidad de revisar si estas señales logran aportar entre ellas.

La electrocardiografía proporciona información de las señales eléctricas generadas por el corazón denominadas señales ECG o EKG, la sigla EKG está siendo utilizada recientemente debido a los trabajos relacionados con el idioma alemán, esta información es básica y fundamental debido a que no se puede obtener a través de otra exploración.

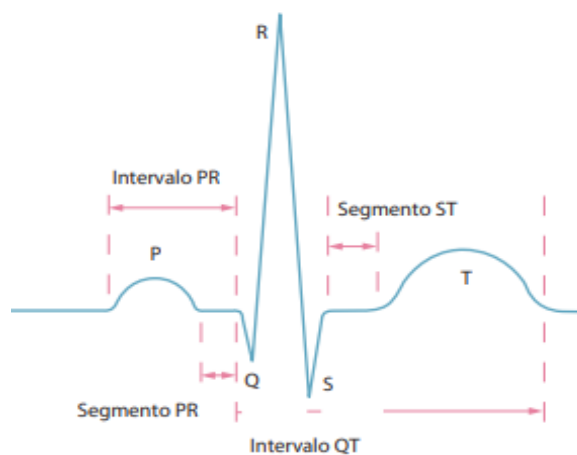
Las señales EKG tienen en común canales iónicos con las señales EEG por ende se tuvo en cuenta las señales EKG para este proyecto con el fin determinar cómo se comportaba el corazón con los ataques de epilepsia y ver si las matrices de algoritmos de Machine Learning mejoraban incluyendo estas señales.

Realizar la EKG no es doloroso, es rápida, sencilla y segura, debido a que se realiza con electrodos superficiales los cuales captan las señales P, Q, R, S, T y U; los electrodos pueden causar daño serio en pieles sensibles con esta prueba.

Las pruebas de EKG se utilizan para conocer si el corazón está teniendo alguna anomalía, confirmar anomalías que los médicos detectan por medios mecánicos como el tensiómetro, estetoscopio y entre otros (Azcona, L. 2009).

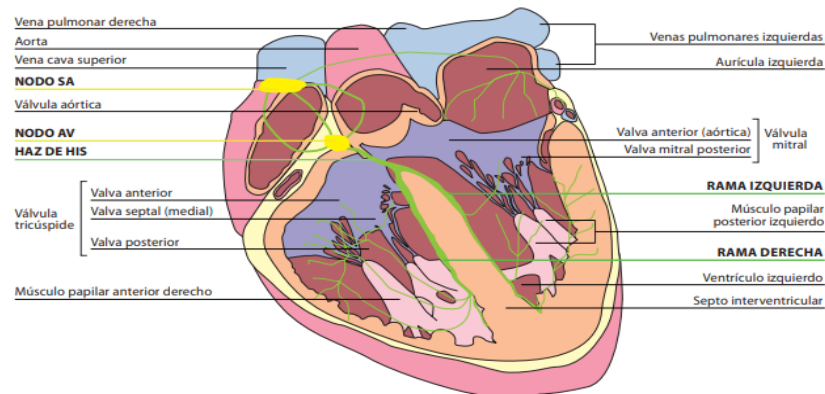
La realización de una electrocardiografía consta de un paciente que se coloque en un lugar acostado se hace la consulta al paciente de si tiene alguna patología respecto al corazón el paciente es ubicado boca arriba en la camilla, donde se le coloca un electrodo en la parte del pulmón derecho el cuarto espacio intercostal, el segundo es puesto en el lado del pulmón derecho en el cuarto espacio intercostal y el tercer electrodo se coloca en el pie izquierdo como punto de referencia el ECG hace descargas según el triángulo de Einthoven lo cual se empieza a imprimir estos registros en un papel o en una memoria del dispositivo, para su lectura con el medico especializado en estas señales.

Figura 2. Ondas de electrocardiografía.



Nota Azcona, L. 2009

Figura 3. Sistema de conducción del corazón.



Nota Azcona, L. 2009.

El corazón tiene un sistema de conducción el cual es un tejido especializado que se encarga de hacer el inicio de los impulsos eléctricos del corazón y conducen estos por el corazón. Lo cual genera el movimiento del corazón (Azcona, L. 2009).

La frecuencia cardíaca esta regula por los siguientes sistemas de conducción:

- Nodo sinoauricular (nodo SA).
- Nodo auriculoventricular (nodo AV).
- Sistema de His-Purkinje.

Las ondas que como P, Q, R, S, T y U. Donde P a nivel del corazón mide la activación auricular, QRS es el procedido de la onda P la cual es excitación de la musculatura de ambos ventrículos y al tabique interventricular. S y T indican la despolarización de la onda y U aparece después de la onda T de los cuales dice que puede ser la repolarización de los músculos capilares (Zavala-Villeda & José Alfredo 2017).

2.2.1 Importancia de la EKG en la epilepsia

En (OPHERK, Christian; COROMILAS, James; HIRSCH, Lawrence,2002), el trabajo realizado por ellos muestra como aumenta la frecuencia cardiaca en 101

pacientes epilépticos de 102. En las siguientes investigaciones encontraron que en 22/102 sufrieron diferentes taquicardias sinusales cuando sufrían una convulsión, 6/102 sufrieron cambios potencialmente graves. El 35% de las anomalías ictales fueron observadas en las convulsiones generalizadas y el 13% fueron potencialmente graves donde la frecuencia cardíaca en la fase ictal fue mayor ($P < 0,03$) y las anomalías en el electrocardiograma fueron más frecuentes ($P < 0,05$), las respuestas de ellos fueron comparadas con convulsiones no generalizadas y generalizadas.

En (REJDAK Konrad, 2011), muestra que las alteraciones cardíacas como: las bradiarritmias y asistolia, así como las taquiarritmias y las alteraciones de la repolarización cardíaca en pacientes con epilepsia tienen mayor posibilidad de sufrir muerte súbita inesperada, el autor afirma que llega a representar aproximadamente la mitad de las muertes relacionadas con convulsiones en este grupo de pacientes, la no supervisión de estas convulsiones por medio del sistema EKG es el principal factor de riesgo de muertes súbitas. Y la principal dificultad para identificar los pacientes en riesgo es la presencia de arritmias cardíacas en el momento ictal de las convulsiones debido a que es similar en pacientes sanos. En pacientes con EER es más peligroso la no supervisión por esta razón surge la necesidad de métodos más sensibles para detectar la disfunción cardíaca con el fin de prevenir la mortalidad en pacientes epilépticos.

2.3 Electroencefalografía

Las señales EEG son usados en este proyecto debido a la literatura reportada en los siguientes párrafos donde se reporta como la más usada porque no es invasiva y tiene menor costo que otras técnicas para la captación de señales cerebrales y usada para casos de epilepsia e identificación de los tipos de epilepsia

La electroencefalografía (EEG) se encarga del estudio de las señales eléctricas producidas por el sistema nervioso central (SNC) el cual al captar las señales puede observar cómo se comporta el cerebro con diferentes órdenes dadas por

el SNC hacia las partes de movimiento voluntario del cuerpo (Ramos-Argüelles,2009).

La medición y almacenamiento de estas señales se ha utilizado para encontrar trastornos del sistema nervioso central (SNC), para la captación de las señales se han utilizado invasivas y no invasivas, donde las invasivas corresponden a dos tipos: por medio electrodos de aguja que van al tejido nervioso del cerebro y electrodos en la superficie cerebral, la no invasiva es la colocación de electrodos en el cráneo (Gómez Figueroa, L. J. 2016).

Tabla 2. Rango de frecuencia de las diferentes ondas

Ondas	Frecuencia (HZ)
Delta	0.5-3.5
Theta	3-8.5
Alfa	7-13.5
Beta	11-33.5
Gamma	24-100.5

Nota Peirotén López de Arbina, N. 2018.

En las señales EEG la frecuencia es el criterio más importante y usado en la evaluación de anomalías y comportamientos en la investigación cognitiva donde encontramos las siguientes ondas que cada una de ellas cuentan con diferente frecuencia lo que permite la detección de ellas las frecuencias para la detección.

La onda Delta poseen una amplitud mayor de onda y corresponde a el sueño profundo (donde la persona no presenta sueños), esta onda es visible en los niños y bebés, pero a medida que el humano crece produce menos ondas de este tipo debido que poco a poco se va perdiendo la capacidad de descansar a medida que va pasando el tiempo y por ende esta onda es escasa en adultos (Peirotén López de Arbina, N. 2018).

Figura 4. Onda Delta



Nota Peirotén López de Arbina, N. 2018.

La onda está relacionada con actividades corporales de las que no somos conscientes, como la regulación del ritmo cardíaco o la digestión.

La onda theta corresponde a las capacidades imaginativas, con la reflexión y el sueño. Esta onda suele ser elevada cuando las emociones son muy profundas.

Figura 5. Onda Theta



Nota Peirotén López de Arbina, N. 2018

La onda alfa se encuentra donde hay calma, pero no sueño, relajación y meditación. Estudios han demostrado que se puede tener estas ondas cuando en la sala viendo la televisión o en la cama descansando, pero sin dormir.

Figura 6. Onda Alfa



Nota Peirotén López de Arbina, N. 2018

El nivel de esta onda se ve más en los niños y adolescentes que adultos, perime la reducción de la presión arterial e incremento del flujo sanguíneo al cerebro.

La onda beta se observa cuando la actividad neuronal es intensa debido a la frecuencia tan alta, la encontramos en acciones cotidianas en que llaman toda nuestra atención..

Las ondas proporcionan que la persona sea más receptiva, enfocada a mejorar incluso la capacidad para resolver problemas.

Figura 7. Onda Beta



Nota Peirotén López de Arbina, N. 2018

Cuando se genera la máxima activación, consecuente conducir, realizar un parcial, hacer una ponencia, etc. puede provocar ansiedad o estrés.

La onda gamma está relacionada con tareas que requieren un procesamiento cognitivo alto, donde está catalogado nuestro estilo de aprendizaje.

Figura 8. Onda Gamma



Nota Peirotén López de Arbina, N. 2018

La actividad de la onda gamma es menor en pacientes con problemas de aprendizaje y/o mentales.

2.3.1 Importancia del EEG en la epilepsia

En (SMITH, Shelagh JM 2005), nos dice como este sistema es utilizado para poder determinar el tipo de convulsión y el síndrome de epilepsia en pacientes con epilepsia y con estos datos poder elegir el medicamento antiepiléptico. Estos hallazgos permiten el diagnóstico en términos de si es trastorno convulsivo generalizado o focal, o si es parte de un síndrome en específico. El EEG puede ayudar a distinguir entre una crisis parcial compleja con una descarga epileptiforme interictal (DEI) focal y una crisis de tipo ausencia con una DEI generalizado.

El autor (SMITH, Shelagh JM 2005), nos indica que el 50% de los pacientes con epilepsia en la prueba de EEG muestran una descarga de DEI, también nos dice que se debe realizar un estudio a profundidad cuando esto no ocurre, esto se hace por medio de la captación de las señales del cerebro cuando el paciente este dormido naturalmente o por fármacos debido a que estos fármacos no alteran el rendimiento de la captación de estas señales.

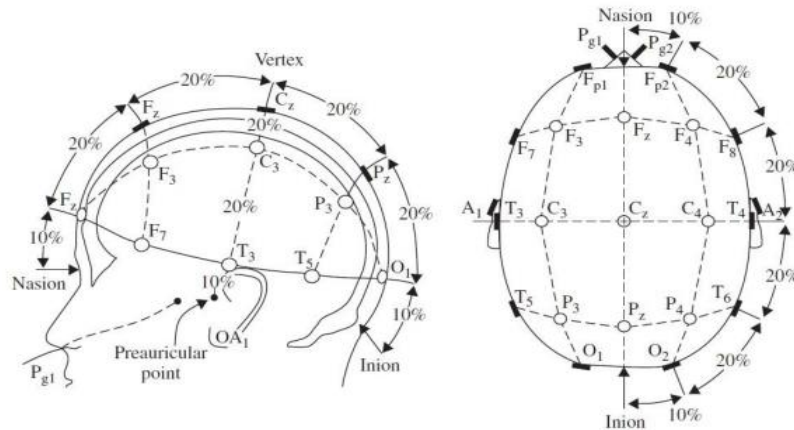
2.4 Sistema internacional 10-20 para la adquisición de señales EEG.

El autor (Detti, P., Vatti, G., & Zabalo Manrique de Lara, G. 2020), de la base de datos reporta el uso de este sistema para la captación de las señales EEG con el fin de crear la base de datos, esta base de datos fue utilizada en el presente proyecto.

Es catalogado como el método de posicionamiento de electrodos estándar, donde en este sistema el 10 hace referencia a que las distancias reales entre electrodos adyacentes son el 10%, mientras que el 20 se refiere al 20% de la distancia total delantera/trasera o derecha/izquierda, de manera que la zona que registre cada electrodo sea comparable entre los registros realizados a diferentes individuos (Sanei & Chambers, 2007) y con eso se logra una división del cráneo con puntos de referencia para cubrir todas las regiones del cerebro.

En la figura 7 se puede ver el diagrama del sistema 10-20. 20 donde el cerebro es mapeado por cuatro puntos: Nasal (Nasion), Occipital (inion) y dividido en derecha (números pares) e izquierda (números impares) por pre auriculares (orejas):

Figura 9. Sistema internacional 10-20

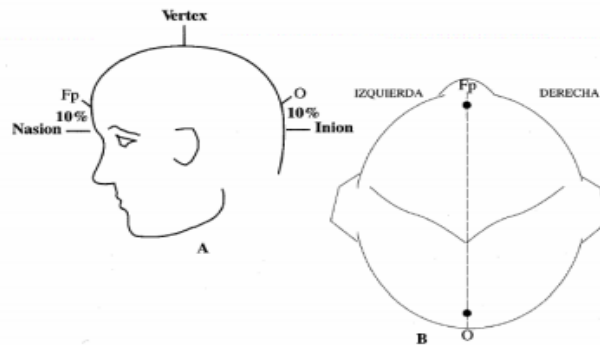


Nota Sanei y Chambers, 2007

Los electrodos según este sistema se proceden de la siguiente manera (Avila Perona, 2006):

- A. Se mide la distancia entre el nasion y el inion pasando por el vertex. El 10% de esta distancia sobre el nasion señala el punto Fp (Frontal Polar). El 10% de esta distancia sobre el inion señala el punto o (occipital).

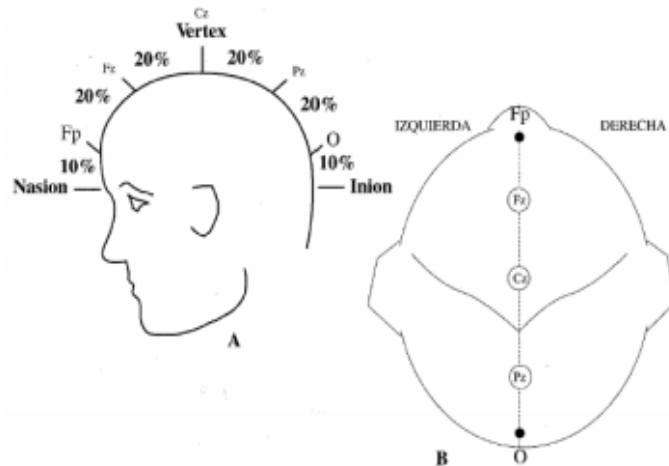
Figura 10. A) Vista de perfil. B) Vista superior. FP, punto polar; O, punto occipital



Nota Avila Perona, 2006

- B Entre los puntos FP y O se sitúan otros tres puntos espaciados a intervalos iguales (entre cada dos el 20% de la distancia nasion-inion). Estos tres puntos son, de adelante hacia atrás, el Fz (Frontal) el Cz (Central o Vertex) y el Pz (parietal).

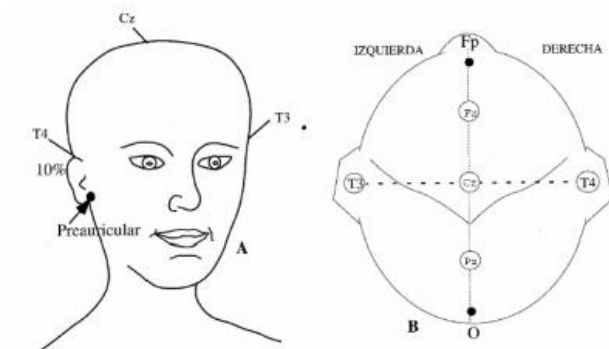
Figura 11. A) Vista de perfil. B) Vista superior. Fz, punto frontal; Cz, punto central; Pz, Punto parietal.



Nota Avila Perona, 2006

- C Se mide la distancia entre los puntos pre auriculares (situados por delante del pabellón auditivo) pasando por el vertex (Cz). El 10% de esta distancia marca la posición de los puntos temporales mediales, T3 (izquierdo) y T4 (derecho).

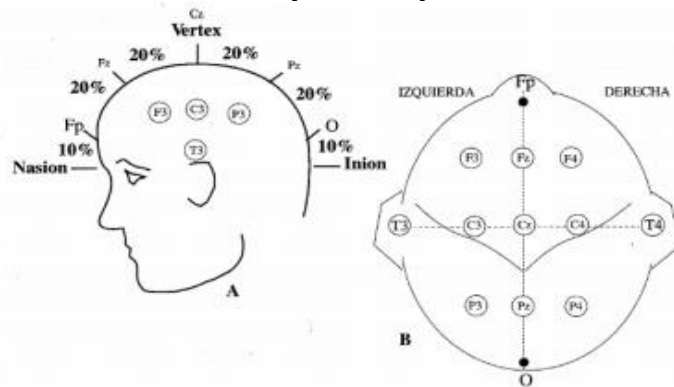
Figura 12. Medición coronal lateral. A) Vista de frontal. B) Vista superior. Situación de los electrodos T3 y T4.



Nota Avila Perona, 2006

- D Un 20% de la medida por encima de los puntos temporales medios se colocan los electrodos C3 (izquierda) y C4 (derecha). El vertex es ahora el punto de intersección entre la línea anteroposterior y la línea coronal lateral.
- E Los electrodos F3 y F4 (Izquierda y derecha, respectivamente) están situados de forma equidistante entre el punto frontal medio (Fz) y la línea de electrodos temporales.
- B. f. Los electrodos P3 y P4 equidistan entre el punto P medio y la línea de los electrodos temporales.

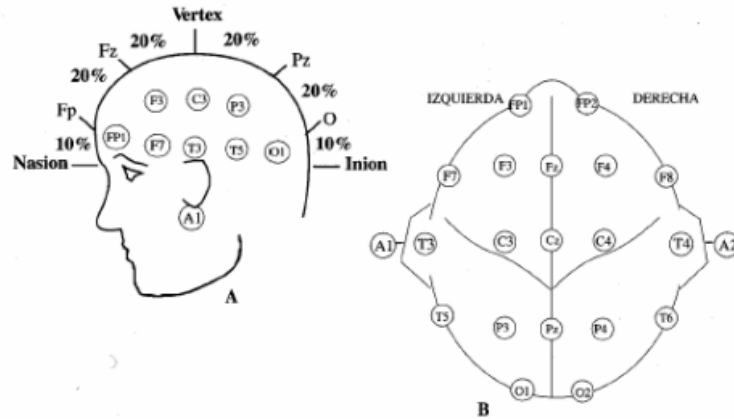
Figura 13. A) Vista de perfil. B) Vista superior. Situación de los electrodos C3 y C4; F3 y F4; P3 y P4.



Nota Avila Perona, 2006

- C. g. Se mide la distancia entre el punto medio Fp y el punto medio O a través de T3. El 10% de esta distancia a través de Fp corresponde a los electrodos FP1 y FP2. El 10% de esta distancia a través de O corresponde a los electrodos O1 y O2.
- D. El electrodo F7 se sitúa equidistante entre los puntos FP1 y T3. Para el F8 se procede de igual manera, pero entre FP2 y T4.
- E. El electrodo T5 se sitúa en la línea media entre T3 y O1. Para el T6 se procede de igual manera, pero entre T4 y O2.
- F. A un 10% de los temporales T3 y T4 se sitúan los electrodos auriculares A1 y A2 respectivamente.

Figura 14. A) Vista de perfil. B) Vista superior. Situación de los electrodos FP1 y FP2, O1 y O2, F7 y F8, T5 y T6, A1 y A2.



Nota Avila Perona, 2006

Como regla los electrodos izquierdos llevan numeración impar y los de la derecha numeración par, y los del medio reciben el subíndice de z lo cual quiere decir que es cero, el cráneo presenta diferentes tamaños en bebés, niños, adolescentes y adultos, por ende, la cantidad de electros posibles a colocar no son los de un bebe a los de un adulto debido a que estos quedarían muy pegados y podría existir choque entre cada electrodo (Avila Perona, 2006).

Tabla 3. Zonas Craneales

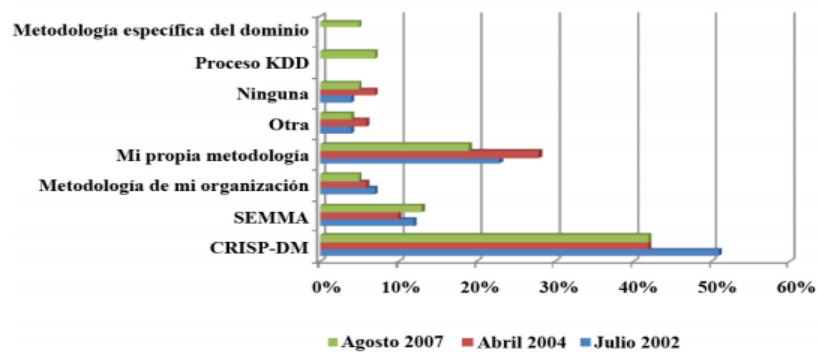
F	Frontal
P	Parietal
T	Temporal
O	Occipital
A	Auricular

2.5 Metodología Crisp-DM

CRISP-DM (Cross Industry Standard Process for Data Mining), Muestra el proceso de como los expertos abordan el tema minería de datos.

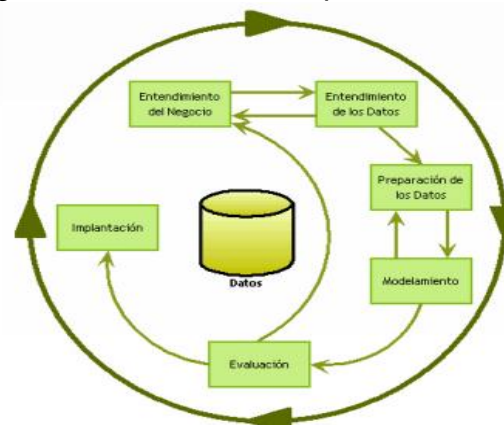
Los métodos creados por lo general vienen de las experiencias propias y también de los procedimientos estándar más conocidos. Como se logra observar en la figura 15, la metodología Crisp-DM en el caso de los proyectos de implementación de minería de datos es una de las metodologías que ha sido más usado por los organismos públicos y empresas ha sido esta metodología, estos datos son obtenidos de (Galán, V. 2015) y por el autor sacados de kdnuggets.com se muestra en la figura 15, donde es compartido el grado de utilización a lo largo del 2002 hasta el 2007 en las principales guías de desarrollo de proyectos de minería de datos según las encuestas realizadas. siendo la más empleada de las distintas metodologías. (Galán, V. 2015).

Figura 15. Grado de utilización de las distintas metodologías de minería de datos.



Nota Galán, V. 2015.

Figura 16. Secuencia de proceso Crisp-DM



Nota Galán, V. 2015

CRISP-DM, tiene 6 fases en su estructura como se observa en la figura 16. Las fases son I. Entendimiento del Negocio II. Entendimiento de los datos III. Preparación de los Datos IV. Modelamiento V. Evaluación e VI. Implementación, la metodología se destaca por ser bidireccional con lo cual en casi cada fase se puede devolver a la anterior fase y reevaluar en caso de necesitar los datos con la anterior fase.

2.5.1 Fases de la metodología Crisp-DM

2.5.1.1 Entendimiento del negocio (Galán, V. 2015).

“Esta fase busca la comprensión de los objetivos y requisitos del proyecto, esto para convertir los anteriores objetivos en algo técnico y un plan de proyecto. Para lograr la mejor utilidad de la minería de datos, lo que permitirá recolectar los datos correctos e interpretar correctamente los resultados. y en un plan preliminar cuya meta sea el alcanzar los objetivos del negocio, Figura 16.

En esta fase se trabaja:

- Determinar los objetivos del negocio. Esta es la primera tarea a desarrollar y tiene como metas determinar cuál es el problema que se desea resolver, por qué la necesidad de utilizar la minería de datos y definir los criterios de éxito. Los problemas pueden ser diversos.
- Evaluación de la situación. Se analiza la situación antes de iniciar el proceso de minería de datos, donde se generan las siguientes preguntas: ¿cuál es el conocimiento que se tiene del problema antes de ejecutarlo?, ¿la base de datos es suficiente para resolver el problema?, entre otras. Se delimitan los requisitos del problema.
- Realizar el plan del proyecto. tiene como objetivo desarrollar un plan para el proyecto, que describa los pasos a seguir, las técnicas a emplear en cada uno de ellos y la pregunta de investigación.

2.5.1.2 Entendimiento de los datos.

Esta fase consiste en la recolección de la información inicial de los datos que van a trabajar, con el fin de lograr saber qué tipo de datos se está tratando, como se pueden trabajar los datos, su calidad y establecer que mecanismos de trabajo sirven para esta base de datos. En esta fase como las dos siguientes son las fases en las que más se demanda esfuerzo debido a que son la base para poder realizar con éxito la metodología, figura 15.

- Recolectar los datos iniciales. En esta fase se hace la lista de que datos podemos encontrar en la base de datos original esto con el fin de elaborar una lista para hacer su localización, técnicas de recolección, los problemas y soluciones a algún problema que sucede en el proceso.
- Descripción de los datos. Entender que se encuentra en la base de datos esto implica en la extracción de características a leer para saber que se encuentra en la base de datos para poder identificarla, el significado de cada campo y la descripción del formato inicial.
- Exploración de los datos. En esta se procede a la toma de datos de forma global para entender qué tipo de datos son y que se necesita para poder leer los datos, con esto se genera información de exploración para así poder trabajar los datos.

2.5.1.3 Preparación de datos.

Se procede a su acondicionamiento para adaptarlos a las técnicas de minería de datos, las cuales las posibles son visualización de los datos, medidas de exploración u otras variables. En la preparación la selección de datos, modelado, limpieza de datos, generación de variables. La preparación de datos está arraigada con la fase de modelado, debido a que la técnica de modelado al ser elegida puede alterar la fase de preparación de datos de una manera o de otra, por ello las fases de preparación y de modelado interactúan de forma permanente.

Seleccionar los datos. Se selecciona los datos según criterios que se deseen aplicar para la clasificación y posterior trabajo de estos para la minería de datos.

- Limpiar los datos. Es cuando se necesita hacer filtros a las señales porque contienen ruido y eso se puede con normalización de las señales, filtros debido a ruidos, etc.
- Construir los datos. Se una nueva información la cual consta de los datos que se adquirieron usando la anterior fase, esto con el fin de lograr una base de datos más sólida.
- Integrar los datos. Esta fase se conecta con la anterior debido a que en esta fase se puede hacer extracción de características las cuales van generar una construcción de datos los cuales van generar para poder se modelados en la siguiente fase.

2.5.1.4 Modelado

Se eligieron diferentes algoritmos que se utilizaron a lo largo del proyecto para poder trabajar los datos y mejor modelo para la minería de datos. Se elijen bajo el concepto de ser la técnica apropiada para resolver el problema o el tiempo adecuado para obtener un modelo que de una precisión que sea alta o dependiendo las métricas con las que se van a evaluar los modelos. Después de concluir estas tareas genéricas se procede a la generación y evaluación del modelo. Los parámetros usados en el modelo dependen de las características extraídas de los datos.

Escoger la técnica de modelado. En esta área se hace la elección de que algoritmo es el mejor, cual algoritmo tuvo el mejor rendimiento para lo propuesto en el proyecto esto con el fin de hacer la minería de datos de con las técnicas el cómo K- nearest neighbors (K-NN), árboles de decisiones, lineal discriminante, regresión logística, naive bayes o RusBoosted Trees, regresión o redes neuronales, en segmentación, redes neuronales, técnicas de visualización, etc.

- Generar el plan de prueba. Toda la información que se toma es guardada para hacer pruebas posteriormente con ellas con el fin de hacer diferentes pruebas con la minería de datos la clasificación de datos, para poder construir el modelo basado en el conjunto de entrenamiento y medir el desempeño que se generó con el conjunto de prueba.
- Construir el modelo. Con el fin de poder hacer la minería de datos se crea un modelo con un conjunto de características y predictores los cuales van aportar para entrenar algoritmos de predicción. Para poder seleccionar los mejores parámetros a trabajar y su rendimiento justificado
- Evaluar el modelo. Los modelos se evalúan con diferentes técnicas que podemos encontrar, como lo serian con matrices de confusión, curvas Roc y entre otras técnicas para evaluar el rendimiento de los algoritmos entrenados.

2.5.1.5 Evaluación

La fiabilidad es calculada para los diferentes modelos y es aplicada solamente para los datos sobre los análisis realizados, se deben tener en cuenta los resultados que se obtuvieron con el fin de poder repetir algún paso anterior en caso que en algún paso se cometiera algún error. En caso que el modelo formado según los criterios de función establecidos sea válido, ahora este modelo explorara los algoritmos para tener una explotación de los datos obtenidos teniendo en cuenta los anterior se genera la siguiente exploración del modelo.

- Evaluar los resultados. En esta exploración se realiza la evaluación de como cumple con relación a los objetivos del proyecto propuesto por esta razón esta relación con el entendimiento del negocio, y se busca determinar si los algoritmos entrenados son deficientes por alguna razón según los objetivos o si es mejor utilizar los datos con problemas en tiempo real.

Los resultados deben estar de acuerdo a los objetivos dichos en el entendimiento del negocio con, lo que permite generar preguntar que el en la mayor de los casos los objetivos se cambian y dan inicio a una nueva metodología Crisp- DM

- Revisar el proceso. Se verifica si el modelo se puede mejorar de alguna forma con el fin de mejorar el rendimiento según las métricas que lo evalúan”.

2.6 Modelos entrenados

Los algoritmos de aprendizaje de machine learning que fueron utilizados en este proyecto como clasificadores para la predicción de la fase pre-ictal de pacientes con epilepsia. Son diferentes en cuanto a su método de clasificación y por eso se hace una investigación de ellos.

2.6.1 Árbol de decisiones

En (BLANCO, Isvani Frías, 2010) nos proporciona información acerca de los árboles de decisión, donde nos dice que son similar a un diagrama de flujo el cual está conformado el nodo raíz la cual es la característica más significativa seguido de los nodos de pruebas donde están todas las condiciones y pruebas que se deben realizar para llegar a un resultado y los nodos de decisión los cuales serían la predicción que está haciendo el algoritmo.

2.6.2 Análisis discriminante (AD)

En (FERNÁNDEZ, Edgar, Acuña, 2000) da a entender que el AD se utiliza para predecir la pertenencia a partir de los predictores con la finalidad de entender las diferencias los grupos y lograr predecir la verosimilitud basándose en los valores de los predictores para encanillarlos en una clase o grupo. Ejemplos de análisis discriminante son distinguir entre la fase pre-ictal y la fase no pre-ictal de acuerdo sus características a sus predictores a través de sus predictores.

2.6.3 Clasificadores de Regresión logística

En (AGUILERA, Ana; SUBERO, Alberto 2011) los autores nos dicen que este clasificador implementa las técnicas de regresión logística lineal y arboles de decisiones donde crea un modelo lineal de los datos introducidos y un proceso de ajuste estable. El primer modelo nos muestra una varianza baja con un sesgo potencialmente alto y los árboles de decisiones el sesgo es bajo lo que permite la captura de patrones no lineales, pero el algoritmo se vuelve menos estable y se puede sobre ajustar. Y concluye que la combinación de los dos se han obtenidos buenos resultados.

2.6.4 Clasificador Naïve Bayes (NBC)

En (MOSQUERA, Rodolfo; CASTRILLÓN, Omar D.; PARRA 2018) los autores nos hablan del algoritmo NBC, el cual es considerado como un clasificador que es probabilístico simple, el problema con el algoritmo es que es de suposición de la independencia y por lo general los atributos son violados para los conjuntos de datos verdaderos, el clasificar puede trabajar con una cantidad pequeña de datos de entrenamiento pero su rendimiento es mejor con datos en tiempo real lo que le permite predecir la clase de la instancia que se ejecuta. Y para datos dimensionales altos es mejor debido a que se estima independientemente.

2.6.5 Maquina de soporte Vectorial (MSV)

En (GALINDO, Eiber A.; PERDOMO, Jairo A.; FIGUEROA-GARCÍA, Juan C,2020), nos dicen que la MSV es un modelo de aprendizaje supervisado. Donde el objetivo de este clasificador es crear un hiperplano el cual se va encargar de separar las clases una de las otras con el fin de maximizar su distancia entre ellas. El principal objetivo de las funciones kernel (Lineal MSV, Cuadrática MSV, Cubica MSV, Fino Gaussian MSV, Medio Gaussiano MSV, Coarse Gaussian MSV) es maximizar el margen entre hiperplanos.

2.6.6 Clasificador K-Vecino más Cercano (KNN)

En (MJAHAD, Azeddine,2017), las nuevas instancias son clasificadas a modo de clase mayoritaria entre los k vecinos más cercanos a los datos del entrenamiento KNN, lo consideran como un algoritmo perezoso y si tiene muchos datos el algoritmo puede ser demorado, en el momento que se realiza el entrenamiento el no construye modelos si no que solo guarda instancias, la clasificación se da por terminada cuando llega la instancia de test y da por entendido que asume la clase de un dato que dependa solo de los k vecinos más cercanos, funcionan muy bien con dos clases y se le puede realizar la media a los k vecinos en caso de requerir una regresión

2.6.7 Combinación de clasificadores

En (MAYHUA LÓPEZ, Efraín Tito 2013) nos dicen como este algoritmo ha sido propuesta como una herramienta para mejorar el rendimiento de los clasificadores individuales. El objetivo de la combinación de clasificadores individuales es el ser más certeros, precisos y exactos. Los métodos más conocidos son el Bagging y Boosting.

El método Bagging. La operativa del método es la siguiente:

- El genera muestras aleatorias las cuales se organizarán en conjuntos de entrenamiento con un muestreo aleatorio de las muestras.
- Cada subconjunto de entrenamiento aprende un modelo.
- Se predice la clase para cada clasificador para luego ser clasificado por la clase con mayor voto.

El método Boosting, se basado en asignarle un peso para cada uno de los conjuntos de entrenamiento, cuando se ejecuta varios procedimientos se aprende un modelo por ende se minimiza la suma de los pesos de los clasificadores que han sido erróneamente entrenados lo que permite actualizar cada entrenamiento incrementando el peso de los mal clasificados y disminuyendo el peso de los mejores clasificados, donde la decisión final se toma

por una votación que el mayor ejecuta la acción la cual es ponderada entre todos los conjuntos entrenados

El algoritmo de Clasificación Rusbooted Trees, es la combinación del Rusbooted con el Decision Trees el cual pertenece al grupo de clasificadores en conjunto el cual se destaca por su rendimiento para la predicción en este proyecto, donde:

RusBoosted: Combina el muestreo y el aumento de datos los cuales al realizar esto permiten mejorar el rendimiento para la clasificación y este algoritmo se usa cuando los datos son muy desequilibrados

Decision Trees: En español Árbol de Decisión se destaca por ser un método de representación esquemática de las alternativas le permite una fácil toma de decisión para la predicción, el cual es presentado por ramas las cuales dan paso a decisiones y soluciones en de diferentes resultados.

3 Metodología

Para el desarrollo del proyecto, se utilizará la metodología Crisp-DM para la minería de datos y extracción de características relevantes para la predicción de la fase pre-ictal de las convulsiones de la base de datos (Detti, P., Vatti, G., & Zabalo Manrique de Lara, G. 2020).

2.6 Business Understanding

En esta fase se concibe los problemas del proyecto, se genera la pregunta de investigación, la definición de los objetivos. Esta fase del proyecto es entender desde el principio el proyecto y observar el trabajo cómo va a influir en el resultado final del mismo. Esta fase se adelantó en el planteamiento del problema del presente documento.

2.7 Data Understanding

En esta etapa se diseña el código en Matlab para entender y comprender los datos. Según la metodología, se empezará con un entendimiento de la información registrada en la base de datos reportada en (Detti, P., Vatti, G., & Zabalo Manrique de Lara, G. 2020). Para realizar lo anterior, se importarán los datos en Matlab, se identificarán los problemas de calidad y se visualizarán las señales.

2.7.3 Información de la base de datos.

En (Detti, P., Vatti, G., & Zabalo Manrique de Lara, G. 2020).se presenta una base de datos que consta de registros de EEG de 14 pacientes adquiridos en la Unidad de Neurología y Neurofisiología de la Universidad de Siena. En el registro solo se encuentra reportada la fase-ictal en un archivo .txt.

El formato de las señales está en formato de datos europeos y sus siglas (.edf) el cual contiene las señales de los pacientes que sufrieron convulsiones en diferentes tiempos este archivo contiene dos archivos los cuales son las señales

EEG y EKG, y la información de las señales donde nos dice: la frecuencia de muestreo que fue tomada las señales, las unidades, filtros usados, lugar de los electrodos en la señal EEG, transductores usados, cantidad de señales EEG y EKG adquiridas.

En la Tabla 4 muestra cómo está organizada la información para solo un paciente, en el archivo. Edf donde HP: Filtro pasa alto, LP: Filtro pasa bajo, N: Filtro Notch, μV : Micro Voltio:

Tabla 4. señales almacenadas en los archivos. Edf

Transductores	Lugar	Lugar	Filtros	Unidad	Frecuencia de Muestreo(Hz)
Transductor AC	'EEGFp1'	Frontal	HP:1.591549Hz LP:30Hz N:50Hz	μV	512
Transductor AC	'EEGC3'	Central	HP:1.591549Hz LP:30Hz N:50Hz	μV	512
Transductor AC	'EEGP3'	Parietal	HP:1.591549Hz LP:30Hz N:50Hz	μV	512
Transductor AC	'EEGO1'	Occipital	HP:1.591549Hz LP:30Hz N:50Hz	μV	512
Transductor AC	'EEGF7'	Frontal	HP:1.591549Hz LP:30Hz N:50Hz	μV	512
Transductor AC	'EEGT3'	Temporal	HP:1.591549Hz LP:30Hz N:50Hz	μV	512
Transductor AC	'EEGT5'	Temporal	HP:1.591549Hz LP:30Hz N:50Hz	μV	512
Transductor AC	'EEGFc1'	Central	HP:1.591549Hz LP:30Hz N:50Hz	μV	512
Transductor AC	'EEGFc5'	Frontal	HP:1.591549Hz LP:30Hz N:50Hz	μV	512
Transductor AC	'EEGCp5'	Central	HP:1.591549Hz LP:30Hz N:50Hz	μV	512
Transductor AC	'EEGF9'	Frontal	HP:1.591549Hz LP:30Hz N:50Hz	μV	512
Transductor AC	'EEGCz'	Central	HP:1.591549Hz LP:30Hz N:50Hz	μV	512
Transductor AC	'EEGPz'	Parietal	HP:1.591549Hz LP:30Hz N:50Hz	μV	512
Transductor AC	'EEGFp2'	Frontal	HP:1.591549Hz LP:30Hz N:50Hz	μV	512
Transductor AC	'EEGF4'	Frontal	HP:1.591549Hz LP:30Hz N:50Hz	μV	512
Transductor AC	'EEGC4'	Central	HP:1.591549Hz LP:30Hz N:50Hz	μV	512
Transductor AC	'EEGP4'	Parietal	HP:1.591549Hz LP:30Hz N:50Hz	μV	512
Transductor AC	'EEGO2'	Occipital	HP:1.591549Hz LP:30Hz N:50Hz	μV	512
Transductor AC	'EEGF8'	Frontal	HP:1.591549Hz LP:30Hz N:50Hz	μV	512

Transductores	Lugar	Lugar	Filtros	Unidades	Frecuencia de Muestreo(Hz)
Transductor AC	'EEGT6'	Temporal	HP:1.591549Hz LP:30Hz N:50Hz	μV	512
Transductor AC	'EEGFc6'	Frontal	HP:1.591549Hz LP:30Hz N:50Hz	μV	512
Transductor AC	'EEGCp2'	Central	HP:1.591549Hz LP:30Hz N:50Hz	μV	512
Transductor AC	'EEGCp6'	Central	HP:1.591549Hz LP:30Hz N:50Hz	μV	512
Transductor AC	'EEGF10'	Frontal	HP:1.591549Hz LP:30Hz N:50Hz	μV	512
Transductor DC	'EKGEKG'	Electrocardiografía	HP:1.591549Hz LP:30Hz N:50Hz	μV	512

Nota Detti, P., Vatti, G., & Zabalo Manrique de Lara, G. 2020

Según la información obtenida de la base de datos, de los 14 pacientes incluidos en el experimento, solamente, 10 de ellos tienen registro de la señal EKG. En la tabla 5, se muestran los pacientes con registro de información electrocardiográfica.

Tabla 5. Señales EKG

PN00	Tiene señales EKG
PN01	Tiene señales EKG
PN03	Tiene señales EKG
PN05	Tiene señales EKG
PN06	Tiene señales EKG
PN07	Tiene señales EKG
PN09	Tiene señales EKG
PN10	Tiene señales EKG
PN11	Tiene señales EKG
PN12	Tiene señales EKG
PN13	Tiene señales EKG
PN14	No tiene señales EKG
PN16	No tiene señales EKG
PN17	No tiene señales EKG

3.3 Data Preparation

En esta sección se realizará la extracción de las características y el análisis de la información. Se seleccionan todos los pacientes de la base de datos (Detti, P., Vatti, G., & Zabalo Manrique de Lara, G. 2020) de señales EEG y señales EKG. Para su posterior modelado se realizó un filtro y se extrajo características.

3.3.1 Selección de Datos

La base de datos tiene pacientes no tienen señales EKG por ende se discriminó de la base de datos pacientes que no tuvieran esta señal, también se discriminó señales que pesaban más de 2Gb y debido a falta de recursos computacionales para poder procesar esta información en Matlab.

Pacientes usados en este estudio

- PN00
- PN05
- PN06
- PN09
- PN10
- PN11
- PN12
- PN13

3.3.2 Selección de tiempos

Las señales originales se dividieron en cuatro partes, la primera denominada fase no pre-ictal, la segunda fase pre-ictal, tercera fase ictal y la cuarta fase post ictal; la literatura encontrada nos dice que la fase pre-ictal es antes de la fase ictal con un rango de 2 minutos, 5 minutos, 7 minutos y hasta 30 minutos (Das, A., Cash, S. S., & Sejnowski, T. J. 2020), (Usman, S. M., Usman, M., & Fong, S. 2017).

En el presente trabajo, se etiquetó como fase pre-ictal, a partir de los 7 minutos antes que suceda la fase ictal, y hasta el momento que inicia la fase ictal.

Por su parte, la fase no pre-ictal fue considerada a partir de los 5 minutos en que inicia la grabación hasta que inicia la fase pre-ictal, es decir, 7 minutos antes del inicio de la fase ictal. En la figura 17 se muestra esto para señales EKG y la 18 para señales EEG.

Figura 17. Señal EKG del paciente PN11 con delimitaciones de fase ictal, pre-ictal y no pre-ictal.

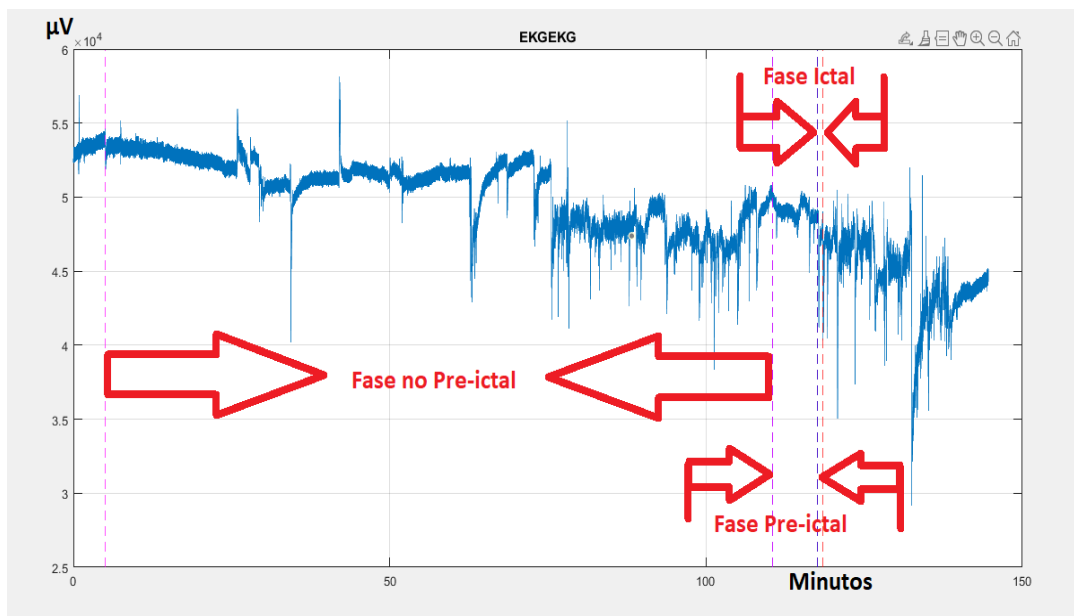
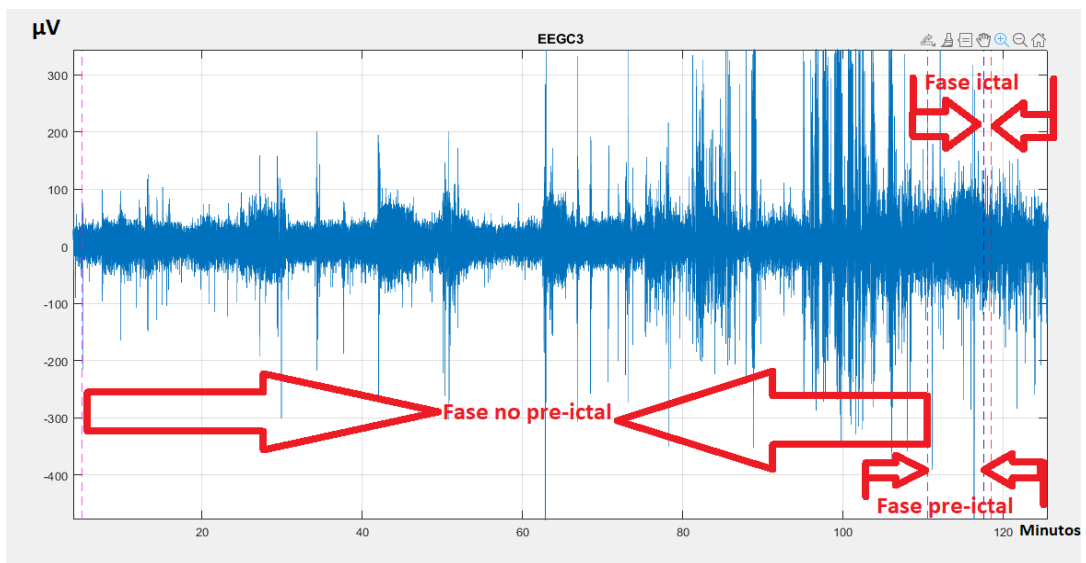


Figura 18. Señal EEG del paciente PN11 sin filtro de bandas



La fase pre-ictal y no pre-ictal se le hizo una segmentación de 2 minutos con un solapamiento del 50% del tiempo de la segmentación para su posterior almacenamiento. Se dejaron programables estos tiempos en el código para próximos experimentos.

3.3.3 Filtros usados

Para realizar la extracción de características, se filtró la señal de la fase pre-ictal y no pre-ictal en las bandas Delta, Theta, Alfa, Beta y Gamma. Se usó la app que tiene Matlab llamada "Filter Designer", los datos usados para el filtro los encontramos en la literatura (Peirotén López de Arbina, N. 2018). y se muestran en la tabla 6.

La frecuencia de muestreo usada corresponde con la que viene registrada en la base de datos, se escoge el orden del filtro (N) de tamaño 250 para tener una mayor atenuación y se justifica debido que hay un número de muestras de señales EEG suficientes por cada ventana de tiempo (61.440 muestras), el tipo de filtro es FIR debido que es tiene fase lineal y se usa para evitar distorsión en la fase de las señales, y las frecuencias de corte se usan dependiendo en qué frecuencia podemos encontrar cada banda, según literatura (Peirotén López de Arbina, N. 2018).

Tabla 6. Filtros usados para las bandas

Banda	Tipo de filtro	Método de diseño	Orden del filtro (N)	Frecuencia de muestreo (Hz)	Fc1	Fc2
Delta	Pasa banda	FIR	250	512	0,5	3,5
Theta	Pasa banda	FIR	250	512	3	8,5
Alfa	Pasa banda	FIR	250	512	7	13,5
Beta	Pasa banda	FIR	250	512	11	33,5
Gamma	Pasa banda	FIR	250	512	24	100,5

3.3.3.1 Common average reference

En una EEG el potencial de campo local a menudo se corrompe por artefactos espacialmente correlacionados en los electrodos, más visto en sujetos despiertos. Basado en el concepto de cancelación de ruido adaptativo que utiliza las características correlativas de fuentes de ruido comunes e implementos con referencia de promedio común (y sus siglas en inglés CAR). Para eliminar los artefactos espacialmente correlacionados (Xinyu, L., Hong, W., Shan, L., Yan, C., & Li, S. 2017), con el fin de tener una mejor extracción de características se realiza el Common average reference (CAR) y así discriminar señales vecinas que puedan afectar los resultados.

Este preprocesamiento consiste en normalizar los datos donde se realizó en el entorno de Matlab, y consiste en promediar las señales y sustraer la media de las señales, esto solo se aplicó a las señales EEG.

3.3.4 Extracción de características.

Las señales obtenidas fueron EEG (todos los canales), EKG y EEG (filtradas en las bandas: Delta, Theta, Alfa, Beta y Gamma), estas señales fueron segmentadas como se habla en la sección de selección de tiempos y organizadas en una matriz de la siguiente forma: [EEG, EEG Delta, EEG Theta, EEG Alfa, EEG Beta, EEG Gamma, EKG], con el fin de una fácil comparación y extracción de características. Para cada una de las ventanas de tiempos extraídas a cada una de estas señales, se le realizó el cálculo de las siguientes características:

- Varianza
- Máximo absoluto
- Mínimo absoluto
- Valor medio
- Media geométrica
- Media Armónica
- Rango
- Desviación absoluta media
- Rango Inter cuartil
- Asimetría (tercer momento central)

- Curtosis (cuarto momento central)

El código para la extracción de características fue almacenada en dos partes, con una columna extra identificando las fases: Uno para Fase pre-ictal y cero para Fase no pre-ictal. Luego las características fueron guardadas en un .mat esto es posible porque las características usan menos recursos en memoria que las señales originales y permite su almacenamiento con facilidad, esto se hace para su posterior modelado .

2.8 Modeling

En esta etapa se usarán diferentes modelos, basado en machine learning, para la predicción de la fase pre-ictal de las convulsiones en pacientes con epilepsia.

Para el modelado se utilizaron las características extraídas en la preparación de datos de la metodología Crisp-DM, en Matlab se usó la App “Clasificación Learner” para entrenar modelos con las características.

3.3.1 Configuración de Clasification Learner

Validación cruzada: Se realiza una validación cruzada de 5 folds, debido a que es un valor común usado para medir el desempeño de las máquinas de aprendizaje, y si se colocan mayores valores se requería mayor costo computacional.

Se obtuvo una matriz de 2295x2256 donde la columna 2255 es el número de predictores del sistema, los cuales indicarán si la señal fue fase pre-ictal o no pre-ictal. De manera explícita, los 2.255 predictores corresponden a:

$$(Nc_eeg \times Ch) + (Nc_eeg \times Ch \times B) + (Nc_ekg \times Ch),$$

donde Nc_eeg corresponde a 34 canales, B a 5 bandas de frecuencia EEG, Ch a 11 características extraídas y Nc_ekg a 1 señal de EKG.

Se obtuvo un total de 171 muestras correspondientes a la fase pre-ictal y las restantes de 2.124 de fase no pre-ictal.

Se recuerda que se tiene la matriz de información va con todas las características extraídas y que tiene una columna con las respuestas que indican si son pre-ictales y no pre-ictal, con el fin de clasificar si hay fase pre-ictal.

3.3.2 Modelos Usados

Se entrenaron modelos en la app de Matlab “Classification Learner” para las características Extraídas

- Decision Trees
 - Fine Tree
 - Medium Tree
 - Coarse Tree
- Discriminant Analysis
 - Linear Discriminat
 - Quadratic
- Logistic Regression Classifiers
 - Logistic Regression
- Naive Bayes Classifiers
 - Gaussian Naive Bayes
 - Kernel Naive Bayes
- Support Vector Machines
 - Linear SVM
 - Quadratic SVM
 - Cubic SVM
 - Fine Gaussian SVM
 - Medium Gaussian SVM
 - Coarse Gaussian SVM
- Nearest Neighbor Classifiers
 - Fine KNN
 - Medium KNN
 - Coarse KNN
 - Cosine KNN
 - Cubic KNN

- Weighted KNN
- Ensemble Classifiers
 - Boosted Trees
 - Bagged Trees
 - Subspace Discriminant
 - Subspace KNN
 - RusBoosted Trees

3.4 Evaluación

En esta etapa se evalúan los modelos utilizados para identificar las características más influyentes en la predicción de la fase pre-ictal de convulsiones en pacientes con epilepsia (Arancibia, J. A. G. 2009).

Para evaluar el desempeño de los algoritmos de Machine Learning, se propuso usar la matriz de confusión para la evolución de los algoritmos anteriormente entrenados en el entorno de clasification Learner.

3.4.1 Matrices de confusión.

Métricas de evaluación

Basado en la literatura (Muñoz, J. M. S. 2016), la exactitud es una medida válida pero no es válida cuando se tiene un conjunto de datos desbalanceados. Por lo tanto, en este proyecto se observa la exactitud, pero no se tiene en cuenta para la toma de decisiones. En estos casos se usa la matriz de confusión la cual, para este caso de clasificación binaria, cuenta la cantidad de Verdaderos Positivos (VP), Verdaderos Negativos (VN), Falsos Positivos (FP) y Falsos Negativos (FN). Usando la matriz de confusión, se puede computar la métrica exactitud:

Ecuación de Exactitud

$$(VP+VN) / (VP+FP+FN+VN)$$

Por su parte, la sensibilidad muestra qué proporción de todas las fases pre-ictal calculó el algoritmo como fase pre-ictal, si la sensibilidad es alta significa que la fase pre-ictal no se encuentra en mayor cantidad en los falsos negativos. Por ende se busca algoritmos que tengan mejor sensibilidad para el proyecto puesto que lo primordial es poder predecir la fase pre-ictal lo mejor posible y evitar los falsos negativos

Ecuación de la sensibilidad

$$S = VP/(VP+FN)$$

La precisión da a conocer que cantidad de los verdaderos positivos son realmente fase pre-ictal y no fase no pre-ictal. Lo que significa que entre mayor sea la precisión menor será la cantidad de fase no pre-ictal catalogada como fase pre-ictal.

Ecuación de la precisión

$$P = VP/(VP+FP)$$

La especificidad muestra qué proporción de todas las fases no pre-ictal calculó el algoritmo como fase pre-ictal, si la sensibilidad es alta significa que la fase no pre-ictal no se encuentra en mayor cantidad en los falsos positivos, como resultado el algoritmo tuvo mayor cantidad de verdaderos negativos.

Ecuación de la especificidad

$$E = VN/(VN+FP),$$

F1 promedio ponderado de la sensibilidad y precisión.

Ecuación de promedio ponderado

$$F1 = 2 * (S * P) / (S + P)$$

Con los anteriores cálculos se evalúan los 24 modelos entrenados en el Classification Learner y teniendo en consideración que la fase más importante es la fase pre-ictal, se optará por escoger la mejor sensibilidad de todos los modelos.

4 Resultados

4.1 Resultados obtenidos

Haciendo hincapié en los objetivos específicos del presente trabajo, se tomaron las señales de los pacientes de la base de datos original, en esas señales algunas no tienen señales EKG o el equipo no pudo procesar señales tan grandes, debido a que se utilizaron herramientas de Matlab y no propias

La implementación de este proyecto permite mejorar la calidad de vida los médicos que tratan con pacientes en estado epiléptico refractario (EER), asimismo permite disminuir la mortalidad de los pacientes EER, también ayuda en la predicción de la fase pre-ictal de la mujeres embarazadas con el fin de la sedación y no suministrar medicamento que pueda ser nocivo para el bebé, las dos señales en conjunto como la EEG y EKG resultaron ser una buena herramienta para predicción de la fase pre-ictal y aportar conocimiento con el fin ayudar al área de la salud por medio de instrumentos de predicción.

Se construyó una base de datos tomado de la base original, la cual se dividió en dos partes: fase pre-ictal y fase no pre-ictal en ambas señales se encuentran señales EEG, EEG (filtrado por bandas Alfa, beta, delta, theta y gamma) y EKG, donde la fase pre-ictal se hace 7 minutos antes de la fase ictal y la fase no pre-ictal 5 minutos después de la fase inicial de la grabación con fin en el inicio de la fase pre-ictal.

Se obtuvo un conjunto de características de la fase pre-ictal y fase no pre-ictal las cuales fueron la varianza, máximo absoluto, mínimo absoluto, valor medio, media geométrica, media armónica, rango, desviación absoluta media, rango Inter cuartil, asimetría (tercer momento central), curtosis (cuarto momento central) y en esta etapa se adiciono a la fase pre-ictal 1 al final de cada extracción y 0 para fase no pre-ictal, para su posterior almacenamiento en conjunto de las fases, y para su posterior análisis en algoritmos de machine learning.

Se logró hacer el entrenamiento de algoritmos de machine learning donde se utilizó el programa de matlab. Se logró acceder a estos algoritmos ejecutando la app classification learner con el conjunto de datos de características extraídos anteriormente, y fueron entrenados estos algoritmos para su posterior evaluación.

Mediante la matriz de confusión se logró evaluar el desempeño de los algoritmos de machine learning, 24 algoritmos de machine learning fueron entrenados, donde solo se lograron tener en cuenta 2 de estos algoritmos por su tasa de sensibilidad debido a la importancia en que el sistema no tome en cuenta los falsos positivos, debido que cuando hablamos de falsos negativos estamos hablando que la fase pre-ictal que fue clasificada como fase no pre-ictal, en caso hipotético el paciente está teniendo una fase pre-ictal y el sistema no lo detecta por ende que estas señales están siendo pasadas por el algoritmo ocasiona que cuando el paciente sufre la fase pre-ictal el sistema no pueda detectar y sufra una convulsión sin que se dé cuenta el modelo entrenado.

Tabla 7. Valor de matrices de confusión.

Modelo	Verdadero Positivo (VP)	Falso Negativo (FN)	Falso Positivo (FP)	Verdadero Negativo (VN)
Fine Tree	63	108	116	2008
Medium Tree	44	127	42	2082
Coarse Tree	18	153	20	2104
Linear Discriminant	94	77	589	1535
Logistic Regression	85	86	866	1258
Gaussian Naive	123	48	1298	826

Bayes				
Kernel Naive Bayes	2	169	24	2100
Linear SVM	6	165	20	2104
Quadratic SVM	50	121	54	2070
Cubic SVM	60	111	72	2052

Modelo	Verdadero Positivo (VP)	Falso Negativo (FN)	Falso Positivo (FP)	Verdadero Negativo (VN)
Fine Gaussian SVM	3	168	4	2120
Medium Gaussian SVM	-	171	-	2124
Coarse Gaussian SVM	-	171	-	2124
Fine KNN	91	80	64	2060
Medium KNN	7	164	20	2104
Coarse KNN	-	171	-	2124
Cosine KNN	7	164	22	2102
Cubic KNN	7	164	15	2109
Weighted KNN	30	141	12	2112
Boosted Trees	29	142	12	2112
Bagged Trees	20	151	3	2121
Subspace Discriminant	109	62	107	2017
Subspace KNN	94	77	43	2081
RusBoosted Trees	143	28	471	1653

En el modelo RusBoosted Trees está detectando 143 (83.6%) casos de fase pre-ictal de 171 casos verdaderos, estos 28 casos están pasando como fase no pre-ictal, y de la fase no pre-ictal 1653 (77.8%) de 2124 casos reales categorizando 471 casos como fase pre-ictal.

Tomando en cuenta que estos 28 casos, del modelo RusBoosted Trees, son posibles convulsiones a los pacientes se tiene que este sistema no es el perfecto para lo que se quiere lograr, pero si es el mejor de los modelos. Teniendo en cuenta que existen falsos positivos los cuales generarán una alarma cada vez que el sistema clasifique que es fase no pre-ictal por ende tampoco resultan ser lo mejor, debido a que van existir alarmas cuando no exista un potencial caso de epilepsia.

Tabla 8. Métricas de los modelos de la Tabla 9.

Modelo	Exactitud	sensibilidad	Precisión	F1	Especificidad
Fine Tree	90,23965 14	36,84210 53	35,19553 07	36	94,538606 4
Medium Tree	92,63616 56	25,73099 42	51,16279 07	34,24124 51	98,022598 9
Coarse Tree	92,46187 36	10,52631 58	47,36842 11	17,22488 04	99,058380 4
Linear Discriminant	70,98039 22	54,97076 02	13,76281 11	22,01405 15	72,269303 2
Logistic Regression	58,51851 85	49,70760 23	8,937960 04	15,15151 52	59,227871 9
Gaussian Naive Bayes	41,35076 25	71,92982 46	8,655876 14	15,45226 13	38,888888 9
Kernel Naive Bayes	91,59041 39	1,169590 64	7,692307 69	2,030456 85	98,870056 5
Linear SVM	91,93899 78	3,508771 93	23,07692 31	6,091370 56	99,058380 4
Quadratic SVM	92,37472 77	29,23976 61	48,07692 31	36,36363 64	97,457627 1
Cubic SVM	92,02614 38	35,08771 93	45,45454 55	39,60396 04	96,610169 5
Fine Gaussian SVM	92,50544 66	1,754385 97	42,85714 29	3,370786 52	99,811676 1

Modelo	Exactitud	sensibilidad	Precisión	F1	Especificidad
Medium Gaussian SVM	92,5490196	0	N/A	N/A	100
Coarse Gaussian SVM	92,5490196	0	N/A	N/A	100
Fine KNN	93,7254902	53,2163743	58,7096774	55,8282209	96,9868173
Medium KNN	91,9825708	4,09356725	25,9259259	7,07070707	99,0583804
Coarse KNN	92,5490196	0	N/A	N/A	100
Cosine KNN	91,8954248	4,09356725	24,137931	7	98,9642185
Cubic KNN	92,2004357	4,09356725	31,8181818	7,25388601	99,2937853
Weighted KNN	93,3333333	17,5438597	71,4285714	28,1690141	99,4350283
Boosted Trees	93,2897604	16,9590643	70,7317073	27,3584906	99,4350283
Bagged Trees	93,2897604	11,6959064	86,9565217	20,6185567	99,8587571
Subspace Discriminant	73,6519071	63,5802469	15,9689923	25,5266419	74,4218971
Subspace KNN	94,7712418	54,9707602	68,6131387	61,038961	97,9755179
RusBoosted Trees	78,2570806	83,625731	23,2899023	36,433121	77,8248588

El modelo Subspace KNN presenta el valor más alto para el parámetro F1 , de donde se concluye que, a comparación del anterior modelo, este genera tantos reportes de falsos positivos, donde la fase no pre-ictal tiene un índice de 97% de especificidad y sólo el 3% se fue a Falso positivo y el sistema no va a dar una alarma tan seguido como lo sería con el modelo RusBoosted Trees. En contraparte el modelo genera más falsos negativos, lo cual resulta malo para la

finalidad del proyecto, debido que la probabilidad que el sistema detecte una fase pre-ictal es baja (baja sensibilidad)

En vista que los resultados no son perfectos, se busca maximizar la sensibilidad donde el mejor modelo con la mejor sensibilidad es el RusBoosted Trees, y el Subspace KNN es el mejor F1, pero este modelo tiene mala sensibilidad. Los indicadores de estos modelos muestran que la exactitud no es una buena métrica debido al desbalanceamiento que existe entre los datos de la fase pre-ictal y la no pre-ictal.

Considerando que existe una gran cantidad de predictores, se usa el algoritmo de Análisis de componente principales (PCA), donde el vector de características fue reducido a: 30, 50, 80 y 100, para reducir la dependencia entre las características. , Los resultados se muestran en la Tabla 9:

Tabla 9. PCA junto al modelo RusBoosted Trees

PCA	Verdadero Positivo (VP)	Falso Negativo (FN)	Falso Positivo (FP)	Verdadero Negativo (VN)	Exactitud (%)	Sensibilidad (%)
30	127	44	624	1500	70,8932462	74,2690058
40	119	52	631	1493	70,2396514	69,5906433
50	117	54	564	1560	73,0718954	68,4210526
80	107	64	553	1571	73,1154684	62,5730994
100	116	55	567	1557	72,8976035	67,8362573
Sin PCA	Verdadero Positivo (VP)	Falso Negativo (FN)	Falso Positivo (FP)	Verdadero Negativo (VN)	Exactitud (%)	Sensibilidad (%)
2255	143	28	471	1653	78,2570806	83,625731

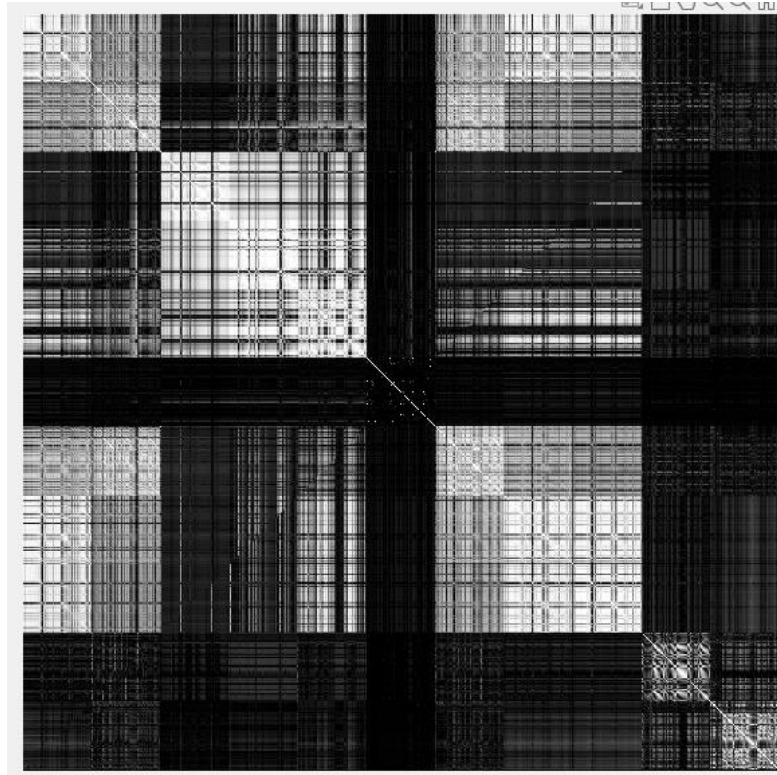
Se logra evidenciar que la sensibilidad no es la mejor al reducir la cantidad de características. Un estudio más detallado se puede realizar para encontrar un número óptimo de características..

Para revisar la dependencia entre los predictores, se realizó el correlograma que se muestra en la figura 19; donde, los colores claros indican una alta correlación y los colores oscuros una baja correlación. En la tabla 10, se observan los predictores que tienen mayor correlación..

Tabla 10. Correlograma de Variables

Variable 1	Correlación	Variable 2
EEGF3' Varianza	0,99631825	EEGFz' Varianza
EEGC3' Varianza	0,98643073	EEGFz' Varianza
EEGF3' Varianza	0,99300942	EEGC3' desviación absoluta media
EEGF4' Varianza	0,99091955	EEGC3' desviación absoluta media
EEGFp1' Varianza Banda Delta	1	EEGFp1' Varianza Banda Alfa
EEGF3' Banda Theta	0,99874154	EEGF3' Varianza Banda Alfa
Media Armónica EKG	0,99586893	Valor medio EKG
Media Armónica EKG	0,99902158	Media Armónica EKG
Media Geométrica Banda Alfa EEGFp1	1	Media Geométrica EEG sin procesar EEGFp1
Varianza Banda Beta EEGCp5	0,9953563	Varianza Banda Theta EEGCp1
Varianza Banda Beta EEGCp6	0,99560788	Varianza Banda Theta EEGCp5
Varianza Banda Gamma EEGFp1	1	Varianza EEG sin procesar EEGFp1
Varianza EEG sin procesar EEGFp2	0,99743609	Varianza EEG sin procesar EEGF3
Varianza EEG sin procesar EEGFc5	0,99773662	Varianza EEG sin procesar EEGFc2
Varianza Banda Delta EEGFp1	1	Varianza Banda Beta EEGFp1
Varianza Banda Delta EEGF3'	0,9997382	Varianza Banda Beta EEGF3
Varianza Banda Delta EEGC3'	0,99135479	Varianza Banda Beta EEGC3

Figura 19 correlograma de todas las señales EEG y EKG



Para validar el aporte que realizan cada una de las señales EEG y EKG en la predicción de la fase pre-ictal, se entrena y se valida el modelo RusBoosted Trees con (1) predictores de solamente las señales EEG, (2) el mismo modelo solamente con características de las señales de EKG, y (3) el entrenamiento características de las señales EEG y EKG. En la tabla 11, se muestra el resultado de la exactitud y sensibilidad de los tres modelos donde se observa que incluir las señales EKG mejora el desempeño del algoritmo.

Tabla 11. Matriz de confusión de EEG y EEG con filtro de banda en conjunto, y EKG sola.

Señales	Verdaderos positivos (VP)	Falsos Negativos (FN)	Falsos Positivos (FP)	Verdaderos Negativos (VN)	Exactitud (%)	Sensibilidad (%)
EEG y EEG con filtro de banda	126	36	435	1684	79,3511618	77,7777778
EKG	102	60	544	1575	73,5203858	62,962963
EEG y EEG con filtro de banda y EKG	143	28	471	1653	78,2570806	83,625731

5. Conclusiones y recomendaciones

5.1 Conclusiones

En este proyecto se desarrolló un algoritmo para predecir la fase pre-ictal (7 minutos antes de la fase ictal) de las señales EEG y EKG de pacientes con epilepsia que sufrieron convulsiones.

Se creó una base de datos con señales EEG, EEG con filtro (banda Alfa, Beta, Theta, Delta, Gamma) y EKG de pacientes con epilepsia que fueron divididas y almacenadas en fase pre-ictal y no pre-ictal, a esta base de datos luego se le extrajeron 11 características y después estas características se almacenaron, con las anteriores características se logró entrenar 24 algoritmos de clasificación de Machine Learning para la predicción de la fase pre-ictal de convulsiones donde estos algoritmos fueron evaluados por matrices de confusión con la finalidad de reconocer que algoritmo es el mejor para la clasificación.

De acuerdo a las matrices de confusión el modelo RusBoosted Trees fue el que tuvo el mejor desempeño con sensibilidad de 83,625731% y exactitud 78,2570806%, bajo el criterio que la sensibilidad es la prioridad para identificar la fase pre-ictal y así saber cuándo el paciente está en esta fase para que él sea tratado a tiempo.

Se logró demostrar que al trabajar las señales de EEG y EKG en conjunto la matriz de confusión mejoraba utilizando el modelo RusBoosted Trees que fue el que mejor sensibilidad obtuvo en este trabajo.

5.2 Recomendaciones

Recomendaciones basadas en los resultados del presente trabajo:

- El trabajo se realizó extrayendo 11 características, se pueden extraer más características para luego entrenar los modelos.

- Se recomienda el uso aprendizaje profundo como algoritmos de uso, por falta de tiempo y la complejidad de este aprendizaje no se realizó en este proyecto.
- Se recomienda revisar la correlación que existe entre los predictores para encontrar un conjunto óptimo de predictores.
- Se llevó a cabo la predicción de la fase pre-ictal pero las matrices de confusiones no fueron las mejores por esto se propone, un estudio de la fase no pre-ictal que no termine en la parte inicial de la fase pre-ictal, si no que tengan un espacio considerable entre estas fases.
- Se recomienda utilizar las características morfológicas de las señales EKG(P, PR, QRS, ST, T y U) y no la señal completa EKG .
- En el trabajo se usaron las señales EEG y EKG, se logró evidenciar las fases ictal, pre-ictal, no pre-ictal y post-ictal, sin embargo se recomienda el uso de marcadores de características morfológicas en el dominio temporal o estimadores estadísticos revisar el comportamiento de los modelos de Machine Learning.

Bibliografía

1. ZAVALA-VILLEDA, José Alfredo. Descripción del electrocardiograma normal y lectura del electrocardiograma. *Revista Mexicana de Anestesiología*, 2017, vol. 40, no S1, p. 210-213. Koertge, N. (2008). Complete Dictionary of Scientific Biography.
2. MORENO, Iveth, et al. Los sistemas de interfaz cerebro-computadora basado en EEG: características y aplicaciones. *I+ D Tecnológico*, 2019, vol. 15, no 2, p. 13-26. Engel, J. (2013). *Seizures and epilepsy* (Vol. 83). Oxford University Press.
3. NARIÑO, Daniel. Epilepsia y embarazo. *Asociación Colombiana de Neurología [en línea][consultado el 13/04/2018]*. Disponible en <http://www.acnweb.org/guia/g2c11i.pdf>, 2003.
4. SÁNCHEZ, Leonardo Palacios. EPILEPSIA: ALGUNAS DENOMINACIONES DADAS A TRAVÉS DE LOS SIGLOS. *Medicina*, 2020, vol. 42, no 3, p. 486-490.
5. BARRERA, Manuel Nieto; JIMÉNEZ, M. Nieto; JIMÉNEZ, E. Nieto. Epilepsias y síndromes epilépticos del preescolar y del escolar. *Protocolos de Neurolo*, p. 28.
6. YAMADA, Mikiko. *UNCONTROLLED SEIZURES AND BONE HEALTH AMONG ADULT EPILEPSY PATIENTS*. 2013. Tesis Doctoral. University of Kansas.
7. Larson, D. E. (1990). *Mayo Clinic family health book*. W. Morrow.
8. FORNEY, Elliott M. *Electroencephalogram classification by forecasting with recurrent neural networks*. 2011. Tesis Doctoral. Colorado State University.
9. FELTANE, Amal. Time-frequency based methods for non-stationary signal analysis with application to EEG signals. 2016.
10. IDIÁQUEZ, Juan, et al. Bradicardia asociada a crisis epilépticas: Casos clínicos. *Revista médica de Chile*, 2009, vol. 137, no 3, p. 401-404.
11. ROSAS-ROMERO, Roberto, et al. Prediction of epileptic seizures with convolutional neural networks and functional near-infrared spectroscopy signals. *Computers in biology and medicine*, 2019, vol. 111, p. 103355.
12. HERNÁNDEZ, Olga Helena, et al. Estado epiléptico refractario en la unidad de cuidados intensivos: revisión de la literatura y diagrama de flujo de manejo en el Instituto Neurológico de Antioquia: [revisión]. *Acta neurol. colomb*, 2011, p. 33-45.
13. RAMOS-ARGÜELLES, F., et al. Técnicas básicas de electroencefalografía: principios y aplicaciones clínicas. En *Anales del sistema sanitario de Navarra*. Gobierno de Navarra. Departamento de Salud, 2009. p. 69-82. Gómez Figueroa, L. J. (2016). *Análisis de señales EEG para detección de eventos oculares, musculares y cognitivos* (Doctoral dissertation, Industriales).
14. PEIROTÉN LÓPEZ DE ARBINA, Nagore. Diseño de una red neuronal en Matlab para análisis de señales de electroencefalograma. 2018.
15. AZCONA, Luis. El electrocardiograma. López Farré A, Macaya Miguel C, directores. *Libro de la salud cardiovascular del Hospital Clínico San Carlos y la fundación BBVA*. 1ª ed. Bilbao: Fundación BBVA, 2009, p. 49-56.
16. TEPLAN, Michal, et al. Fundamentals of EEG measurement. *Measurement science review*, 2002, vol. 2, no 2, p. 1-11.
17. SÁNCHEZ, J. Técnicas de toma de datos y análisis de Electroencefalografía. *Sevilla: Universidad de Sevilla*, 2014.
18. ARMAN, Sabbir Ibn; AHMED, Arif; SYED, Anas. Cost-effective eeg signal acquisition and recording system. *International Journal of Bioscience, Biochemistry and Bioinformatics*, 2012, vol. 2, no 5, p. 301.
19. KO, Wen H. Active electrodes for EEG and evoked potential. En *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Vol. 20 Biomedical Engineering Towards the Year 2000 and Beyond* (Cat. No. 98CH36286). IEEE, 1998. p. 2221-2224.
20. ROWAN, James; TOLUNSKY, Eugene. *Conceptos básicos sobre EEG: con mini atlas*. 2004.

21. Barea Navarro, R. 2009. Electroencefalografía. Instrumentación Biomédica. Madrid, España. Universidad de Alcalá, 2-24.
22. Taywade S.A.; Raut R.D. 2012. A Review: EEG signal analysis with different methodologies. Proceeding on a National Conference on Innovative Paradigms in Engineering and Techonology, 6, 29-31.
23. Sanei, S.; y Chambers, J.A. 2007. EEG Signal Processing, Ltd, 1-267, Cardiff, UK.
24. GUTIÉRREZ-MARTÍNEZ, Josefina, et al. Los sistemas de interfaz cerebro-computadora: una herramienta para apoyar la rehabilitación de pacientes con discapacidad motora. Investigación en discapacidad, 2013, vol. 2, no 2, p. 62-69.
25. DETTI, Paolo; VATTI, Giampaolo; ZABALO MANRIQUE DE LARA, Garazi. EEG Synchronization Analysis for Seizure Prediction: A Study on Data of Noninvasive Recordings. *Processes*, 2020, vol. 8, no 7, p. 846.
26. GALÁN, V. Aplicación de la Metodología CRISP-DM a un proyecto de Minería de Datos en el entorno Universitario. Recuperado: https://e-archivo.uc3m.es/bitstream/handle/10016/22198/PFC_Victor_Galan_Cortina.pdf, 2015.
27. MOINE, Juan Miguel; HAEDO, Ana Silvia; GORDILLO, Silvia Ethel. Estudio comparativo de metodologías para minería de datos. En *XIII Workshop de Investigadores en Ciencias de la Computación*. 2011.
28. DAS, Anup; CASH, Sydney S.; SEJNOWSKI, Terrence J. Heterogeneity of preictal dynamics in human epileptic seizures. *IEEE Access*, 2020, vol. 8, p. 52738-52748.
29. USMAN, Syed Muhammad; USMAN, Muhammad; FONG, Simon. Epileptic seizures prediction using machine learning methods. *Computational and mathematical methods in medicine*, 2017, vol. 2017.
30. Tsuchimoto, S., Shibusawa, S., Iwama, S., Hayashi, M., Okuyama, K., Mizuguchi, N., ... & Ushiba, J. (2021). Use of common average reference and large-Laplacian spatial-filters enhances EEG signal-to-noise ratios in intrinsic sensorimotor activity. *Journal of Neuroscience Methods*, 353, 109089.
31. GALLARDO ARANCIBIA, José Alberto. *Metodología para la definición de requisitos en proyectos de data mining*. 2009. Tesis Doctoral. Informatica.
32. OLIVARES, Victor Miguel Nieves. Analisis estadístico de senales EEG para la deteccion de ataques epilepticos. 2014.
33. GARCÍA LÓPEZ, Beatriz, et al. *Frontal lobe epilepsy and EEG: Neurophysiological approach*. Universitat Autònoma de Barcelona, 2015.
34. CASSATER, Daiana. *Electrocardiographic characteristics of children with seizures and syncope*. 2017. Tesis Doctoral. Universidad Autónoma de Madrid.
35. GHOSH, Arijit, et al. Pre-ictal epileptic seizure prediction based on ECG signal analysis. En *2017 2nd International Conference for Convergence in Technology (I2CT)*. IEEE, 2017. p. 920-925.
36. MEISEL, Christian; BAILEY, Kimberlyn A. Identifying signal-dependent information about the preictal state: a comparison across ECoG, EEG and EKG using deep learning. *EBioMedicine*, 2019, vol. 45, p. 422-431.
37. RUFO CAMPOS, Miguel, et al. Cómo afrontar la epilepsia: Una guía para pacientes y familiares. 2007.
38. HERNÁNDEZ, Olga Helena, et al. Estado epiléptico refractario en la unidad de cuidados intensivos: revisión de la literatura y diagrama de flujo de manejo en el Instituto Neurológico de Antioquia:[revisión]. *Acta neurol. colomb*, 2011, p. 33-45.
39. MASON, Emily R. *Epilepsy Mutations in Different Regions of the Nav1. 2 Channel Cause Distinct Biophysical Effects*. 2020. Tesis Doctoral.
40. TÉLLEZ-ZENTENO, José Francisco; NGUYEN, Rita; HERNÁNDEZ-RONQUILLO, Lizbeth. Lesiones, accidentes y mortalidad en epilepsia: Una revisión de sus prevalencias, factores de riesgo y prevención. 2010.
41. NARIÑO, Daniel. Epilepsia y embarazo. *Asociación Colombiana de Neurología [en línea][consultado el 13/04/2018]*. Disponible en <http://www.acnweb.org/guia/g2c11i.pdf>, 2003.

42. XINYU, Liu, et al. Adaptive common average reference for in vivo multichannel local field potentials. *Biomedical engineering letters*, 2017, vol. 7, no 1, p. 7-15.
43. MUÑOZ, José Manuel Sánchez. Análisis de calidad cartográfica mediante el estudio de la matriz de confusión. *Pensamiento Matemático*, 2016, vol. 6, no 2, p. 9-26.
44. ACHARYA, U. Rajendra, et al. Automated EEG analysis of epilepsy: a review. *Knowledge-Based Systems*, 2013, vol. 45, p. 147-165.
45. REJDAK, Konrad, et al. Analysis of ventricular late potentials in signal-averaged ECG of people with epilepsy. *Epilepsia*, 2011, vol. 52, no 11, p. 2118-2124.
46. OPPERK, Christian; COROMILAS, James; HIRSCH, Lawrence J. Heart rate and EKG changes in 102 seizures: analysis of influencing factors. *Epilepsy research*, 2002, vol. 52, no 2, p. 117-127.
47. SMITH, Shelagh JM. EEG in the diagnosis, classification, and management of patients with epilepsy. *Journal of Neurology, Neurosurgery & Psychiatry*, 2005, vol. 76, no suppl 2, p. ii2-ii7.
48. BLANCO, Isvani Frías, et al. Clasificadores y multclasificadores con cambio de concepto basados en árboles de decisión. *Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial*, 2010, vol. 14, no 45, p. 32-43.
49. FERNÁNDEZ, Edgar Acuña. Notas de Análisis Discriminante. 2000.
50. AGUILERA, Ana; SUBERO, Alberto. Modelos de clasificación en marcha patológica usando árboles de regresión logística. *Multiciencias*, 2011, vol. 11, no 3, p. 310-318.
51. MOSQUERA, Rodolfo; CASTRILLÓN, Omar D.; PARRA, Liliana. Máquinas de Soporte Vectorial, Clasificador Naïve Bayes y Algoritmos Genéticos para la Predicción de Riesgos Psicosociales en Docentes de Colegios Públicos Colombianos. *Información tecnológica*, 2018, vol. 29, no 6, p. 153-162.
52. GALINDO, Eiber A.; PERDOMO, Jairo A.; FIGUEROA-GARCÍA, Juan C. Estudio comparativo entre máquinas de soporte vectorial multiclase, redes neuronales artificiales y sistema de inferencia neuro-difuso auto organizado para problemas de clasificación. *Información tecnológica*, 2020, vol. 31, no 1, p. 273-286.
53. MJAHAD, Azeddine, et al. Detección de fibrilación ventricular mediante tiempo-frecuencia y clasificador knn sin extracción de parámetros. *Revista Iberoamericana de Automática e Informática industrial*, 2017, vol. 15, no 1, p. 124-132.
54. MAYHUA LÓPEZ, Efraín Tito. Elementos locales en conjuntos de clasificadores diseñados por "Boosting". 2013.