

UNIVERSIDAD TECNICA FEDERICO SANTA MARIA

Memoria de Título

Estimación de la edad cerebral mediante registros de electroencefalografía (EEG) utilizando técnicas de aprendizaje de máquinas

Tesis para optar al título de
Ingeniero Civil Electrónico

Alumno
Felipe Ignacio Carriel Rubilar

Guía de Tesis/Profesor Supervisor
María José Escobar

Revisores/Comisión Co-Referente
Matías Zañartu
Mónica Otero
Pavel Prado

Julio 26 2024, Valparaíso, Chile

*Hay quienes luchan un día y son buenos. Hay otros que luchan un
año y son mejores. Hay quienes luchan muchos años y son muy buenos.
Pero hay los que luchan toda la vida: esos son los imprescindibles.
- Bertolt Brecht*

*Dedicado a mi familia, Marjorie, Carlos, Nicolas, Valentina y
Alicia. Dedicado también a mis profesores y todas las personas que me
han acompañado en este proceso.*

AGRADECIMIENTOS

QUISIERA agradecer a todas las personas quienes me han acompañado en este proceso, principalmente a mi familia, Marjorie, Carlos, Nicolas, Valentina y Alicia.

Agradecer a mis amigas y amigos que han estado acompañandome incluso en los momentos más difíciles, a quienes aún están y a quienes no.

Agradezco a mis profesores Mónica y Pavel quienes me otorgaron todo su apoyo y fueron pieza fundamental en este trabajo.

A mis amigos de la universidad quienes día a día nos brindamos el apoyo constante para sobrellevar todo esto, Felipe, Gabriel, Sebastian y Agustín.

A mis compañeros de música, mi grupo Espora, quienes colateralmente estuvieron acompañandome en todo esto durante varios años, Pablo, Andrés y Vicente.

Al huerto Quillay, amiga Mayra y su apoyo incondicional durante todos estos años donde las semillas nos juntaron.

A mi compañera Amanda quien me brindó su amor durante todo este periodo.

Agradecer a mis amigos de vida los grieta, Dante, Ramiro, Pablo y todos los toys.

Y finalmente agradecer a todas las personas que han sido parte de mi vida y a quienes nos hemos apoyado mutuamente en esta.

RESUMEN

LA EDAD BIOLÓGICA cerebral en contraste con la edad cronológica tiene un gran potencial en cuanto a la evaluación del envejecimiento y deterioro del cerebro, posiblemente asociado a diversas patologías cerebrales. Bajo este escenario el electroencefalograma (EEG) podría tener el potencial de generalizar y caracterizar el deterioro cerebral. Se evaluaron distintos algoritmos de aprendizaje de máquina para realizar la comparativa entre cuales de estos algoritmos son los más óptimos para realizar este tipo de tareas, entre ellos algoritmos clásicos de regresión tales como las máquinas de vectores de soporte (SVM) y el de K Vecinos más Cercanos (KNN), también se utilizaron modelos basados en árboles como Bosques Aleatorios, XGBoost y CatBoost, para finalmente evaluar el rendimiento para este tipo de tareas de regresión con Redes Neuronales de perceptrón Multicapa (MLP). Para cada algoritmo se analizaron las contribuciones que cada característica de EEG aportaba a la predicción del modelo con el fin de posteriormente obtener las regiones cerebrales más influyentes en la predicción de la edad. En esta memoria de título se ha generado un flujo de trabajo basado en los algoritmos de aprendizaje de máquina que permite optimizar cada modelo y ajustarlos de la mejor manera a los datos de entrenamiento para realizar la estimación de edad cerebral y la caracterización de cada región cerebral en las predicciones realizadas, esto con el fin de identificar qué regiones son las más importantes al momento de predecir la edad cerebral. Las regiones mas prominentes en las predicciones de Edad Cerebral fueron en la región occipital, y los modelos que mejor rendimiento tuvieron fueron los basados en árboles de decisión, aún quedando camino por recorrer en cuanto a la interpretabilidad final de los modelos y en el potencial que las herramientas de aprendizaje automático contribuyen a la caracterización del deterioro cerebral.

Palabras Claves

Aprendizaje automático, Edad Cerebral, Electroencefalografía, SHAP, Python

ABSTRACT

BIOLOGICAL BRAIN AGE , in contrast to chronological age, holds significant potential for assessing brain aging and deterioration, which may be associated with various brain pathologies. In this context, electroencephalography (EEG) could potentially generalize and characterize brain deterioration. Various machine learning algorithms were evaluated to determine which of these algorithms are most optimal for such tasks, including classical regression algorithms such as Support Vector Machines (SVM) and K-Nearest Neighbors (KNN). Tree-based models like Random Forests, XGBoost, and CatBoost were also used, culminating in the evaluation of Multilayer Perceptron (MLP) neural networks for their performance in regression tasks. For each algorithm, the contributions of each EEG feature to the model's prediction were analyzed to subsequently identify the most influential brain regions in age prediction.

In this thesis, a workflow based on machine learning algorithms has been developed to optimize each model and fit them to the training data as accurately as possible for estimating brain age and characterizing each brain region in the predictions made. This aims to identify which regions are most important when predicting brain age. The most prominent regions in Brain Age predictions were found in the occipital region, and the models that performed best were tree-based models. However, there is still progress to be made regarding the final interpretability of the models and the potential that machine learning tools contribute to characterizing brain deterioration.

Keywords

Machine Learning, Brain Age, Electroencephalography, SHAP, Python

ÍNDICE

AGRADECIMIENTOS	I
RESUMEN	II
ABSTRACT	III
ÍNDICE DE FIGURAS	VII
ABREVIACIONES	XIII
1. INTRODUCCIÓN	1
1.1. Resumen de los Capítulos	1
2. OBJETIVOS	3
2.1. Objetivo General	3
2.2. Objetivos Específicos	3
2.3. Alcances y Limitaciones	3
2.4. Contribuciones	4
3. ESTADO DEL ARTE	5
4. FUNDAMENTOS TEÓRICOS	8
4.1. Edad Biológica Cerebral	8
4.2. Electroencefalograma	9
4.3. Preprocesamiento	10
4.4. Extracción de Características	11
4.4.1. Características Frecuenciales Periódicas	11
4.4.2. Características Frecuenciales Aperiódicas	12
4.5. Aprendizaje Automático	12
4.5.1. Regresión Lineal Simple	12
4.5.2. Regresión Lineal Múltiple	13
4.5.3. Support Vector Regression	15

4.5.4.	KNN	16
4.5.5.	Random Forest	17
4.5.6.	XGBoost	18
4.5.7.	CatBoost	20
4.5.8.	Red Neuronal	20
4.5.9.	Evaluación del Modelo	22
4.5.10.	Valores SHAP	23
5.	METODOLOGÍA	26
5.1.	Base de Datos	26
5.2.	Preprocesamiento	26
5.3.	Preprocesamiento Adicional para Extracción de Características	27
5.4.	Extracción de Características	27
5.5.	Estimación de Edad Cerebral	29
5.5.1.	Modelos de Aprendizaje de Máquina	32
5.5.2.	Búsqueda de Hiperparámetros	33
5.5.3.	Entrenamiento del Modelo	34
5.5.4.	Evaluación del Modelo	35
5.5.5.	Importancia de Características Mediante Valores SHAP	36
5.5.6.	Incremento de Características	37
6.	IMPLEMENTACIÓN	38
6.1.	Entorno Virtual	38
6.2.	Librerías	39
6.3.	Análisis de Datos	40
6.4.	Clase BaseRegressor	44
6.4.1.	Métodos	44
6.5.	Clases de Algoritmos	45
6.5.1.	Support Vector Regresion	45
6.5.2.	K Vecinos más Cercanos	45
6.5.3.	Random Forest	46
6.5.4.	XGBoost	46
6.5.5.	CatBoost	47
6.5.6.	Red Neuronal Multicapa	47
6.6.	Implementación General	50
7.	RESULTADOS	51
7.1.	Support Vector Regression	53
7.1.1.	Entrenamiento del Modelo	54
7.1.2.	Importancia de Características y Valores Shap	55
7.1.3.	Importancia por Región	55
7.2.	K Vecinos Más Cercanos	56
7.2.1.	Entrenamiento del Modelo	57

7.2.2.	Importancia de Características y valores Shap	58
7.2.3.	Importancia por región	58
7.3.	Random Forest	59
7.3.1.	Entrenamiento del Modelo	60
7.3.2.	Importancia de Características y Valores SHAP	61
7.3.3.	Importancia por región	61
7.4.	XGBoost	62
7.4.1.	Entrenamiento del Modelo	63
7.4.2.	Importancia de Características y Valores Shap	64
7.4.3.	Importancia por Región	64
7.5.	CatBoost	65
7.5.1.	Entrenamiento del Modelo	66
7.5.2.	Importancia de Características y Valores SHAP	67
7.5.3.	Importancia por Región	67
7.6.	Red Neuronal MLP	68
7.6.1.	Entrenamiento del Modelo	69
7.6.2.	Importancia de Características y valores Shap	70
7.6.3.	Importancia por Región	70
7.7.	Comparación entre algoritmos	71
7.8.	Comparación de Importancia por Región	72
7.9.	Incremento de Características	73
7.10.	Mejores características	74
7.10.1.	Entrenamiento del modelo	74
7.10.2.	Importancia de Características y valores SHAP	75
8.	CONCLUSIONES	76
8.1.	Discusión	77
8.2.	Trabajos Futuros	77
	BIBLIOGRAFÍA	78

Índice de figuras

4.1. Ondas Cerebrales: Principales ondas cerebrales registradas en el electroencefalograma	10
4.2. Modelo de Regresión lineal simple	13
4.3. Modelo de regresión lineal múltiple	15
4.4. Modelo de máquina de vectores de soporte para regresión, la banda ϵ representa la banda de error permitida en el modelo, mientras que los puntos ξ representan los valores de holgura para los vectores de soporte.	16
4.5. Modelo de los K vecinos más cercanos, la predicción (punto rojo) viene dado por el promedio entre los vecinos más cercanos(puntos verdes)	17
4.6. Modelo de bosques aleatorios para regresión, donde la predicción es el promedio de las predicciones de todos los árboles del bosque. . . .	18
4.7. Modelo de XGBoost, se genera cada árbol de forma consecutiva y la predicción final se obtiene como la suma de la predicción de todos los árboles, en este caso cada árbol no predice la variable y como tal, si no que va mejorando el resultado del árbol anterior.	19
4.8. Modelo de Red Neuronal Multicapa, en este caso cuenta con una capa de entrada de 3 neuronas que representan las características del modelo a predecir, cuenta con 2 capas ocultas densas que se encargan de procesar la información de entrada, para finalmente pasar por una capa lineal que convierte la información procesada en un valor continuo y	21
4.9. Activación de una neurona, las características se combinan con sus pesos correspondientes para luego entrar a la neurona que realiza la sumatoria de los valores de entrada para posteriormente pasarlos a una función de activación.	22

4.10. Análisis de contribuciones, la contribución de la característica X_3 para una instancia determinada se calcula como la diferencia entre la predicción cuando la característica se encuentra presente en la coalición y cuando esta característica no está presente en la coalición.	25
4.11. Las contribuciones totales de X_3 se calculan analizando todas las posibles combinaciones de la instancia de coalición.	25
5.1. Metodología para estimación de la edad cerebral, a partir de los registros EEG primero se preprocesan los datos para eliminar ruido y artefactos indeseados, luego se normalizan los datos para obtener una estandarización y que todos tengan la misma estructura, se realiza una transformación de fuente para pasar de las señales EEG obtenida de cada electrodo a señales EEG focalizadas por regiones cerebrales, luego se extraen diversas características de estas señales por cada región cerebral para finalmente realizar una regresión mediante modelos de aprendizaje de máquina para lograr predecir la edad cerebral.	27
5.2. Regiones cerebrales resultantes de la transformación de fuente utilizadas para el análisis de estimación de la edad cerebral.	28
5.3. (a) Espectro de frecuencia canónico, se define la banda alfa entre 8.5 y 12 Hz, alfa 1 se define desde 8.5 a 10 Hz y alfa 2 desde 10.5 a 12 HZ. (b) Espectro de frecuencia individual, se define alfa inferior desde la transición entre la banda teta y alfa hasta el máximo valor de alfa individual y se define alfa superior desde alfa individual más 2 Hz.	28
5.4. Extracción de características para registros EEG, se presentan características de tipo frecuencial en la banda alfa, EEG corresponde a la densidad de potencia normalizada con respecto a la propia banda y RPD la densidad de potencia normalizada con respecto al espectro completo, además se tienen características de valor de frecuencia donde ocurre el alfa individual y la transición de frecuencia de teta a alfa.	29
5.5. (a) Regresión lineal basada en IAF_OCC_left, se nota un decaimiento en la en la frecuencia de la banda alfa individual en función de un aumento en la edad. (b) Predicciones hechas en base al modelo de regresión lineal utilizando solo la característica IAF_OCC_left.	30
5.6. (a) Regresión lineal múltiple, se genera un hiperplano que modela la tendencia en la banda alfa para las características IAF_OCC_left e IAF_OCC_right. (b) Predicciones hechas en base al modelo de regresión lineal múltiple utilizando solo la característica IAF_OCC_left e IAF_OCC_right.	31
5.7. Entrenamiento del modelo	32

5.8. División de los datos en datos de entrenamiento y datos de prueba, los datos de entrenamiento se utilizan posteriormente como entrenamiento y validación mediante la validación cruzada, los datos de prueba se utilizan solo para evaluar el modelo.	32
5.9. Búsqueda de Hiperparámetros utilizando validación cruzada de 10 pliegues mediante el algoritmo de Bayessearchcv, el modelo óptimo es aquel que consigue una puntuación promedio lo más alta posible, considerando que el algoritmo de Bayes lo que hace es maximizar la puntuación es necesario resolver el problema dual para lo cual se utiliza el error absoluto medio negativo.	33
5.10. Entrenamiento del modelo utilizando validación cruzada y los mejores hiperparámetros obtenidos en la búsqueda bayesiana, de los 10 modelos obtenidos mediante validación cruzada se escoge el modelo del cual se obtenga un menor MAE	34
5.11. Evaluación del modelo utilizando los datos de prueba y el mejor modelo obtenido en la validación cruzada, se realiza la regresión de la edad cerebral utilizando el MAE, MSE, RMSE y R^2 como métricas de evaluación.	35
5.12. (a) Valores SHAP evaluados para cada instancia de datos, un valor negativo indica que al aumentar dicha característica contribuye negativamente al modelo y un valor positivo indica que al aumentar dicha característica contribuye positivamente al modelo. (b) Suma de valores SHAP y su correlación con respecto a la variable que se quiere predecir.	36
5.13. Incremento de características aplicadas al mejor modelo con el fin de obtener el mínimo de características necesarias para obtener un rendimiento máximo, se entrena el modelo primero con la característica que más contribuye en la explicación SHAP, luego con las 2 mejores características y así hasta completar las 72 características estudiadas.	37
6.1. Creación de entorno virtual en Python que permite empaquetar las librerías utilizadas para mantener la compatibilidad entre las versiones.	38
6.2. Librerías utilizadas para el proceso de estimación de edad cerebral	39
6.3. Distribución de densidad de edades para los datos a analizar, se aprecia la falta de datos alrededor de los 50 años.	40
6.4. correlaciones de características con respecto a la edad, las características de potencia espectral y el valor de alfa canónica presentan correlaciones negativas, es decir, disminuyen cuando la edad aumenta, las frecuencias de transición entre alfa y theta presentan correlación npositiva, es decir, aumentan cuando la edad aumenta.	41

6.5. Distribución de las características EEG en su forma natural, escaladas con MinMaxScaler y escaladas con StandarScaler. Para el algoritmo de Red Neuroanl Multicapa se utiliza StandardScaler debido a que se necesita que las características tengan rangos de valores similares de lo contrario los perceptrones podrian generar pesos infinitos o nulos, para el resto de los algoritmos se utilizó MinMaxScaler.	43
6.6. Diagrama de clases, se presenta una clase principal BaseRegressor que contiene los principales métodos del modelo, además se presentan las clases de cada algoritmo los cuales heredan de la clase principal BaseRegressor.	49
6.7. Implementación del flujo de trabajo, se presentan los metodos de la clase BaseRegressor y el flujo de trabajo general que permite obtener la estimación de edad cerebral a partir de características extraídas de EEG.	50
7.1. Comparación de las métricas de evaluación para el mejor modelo en cada uno de los algoritmos, el mejor rendimiento, basandose en obtener un menor error absoluto medio (MAE) y un mayor R^2 lo obtiene XGBoost y CatBoost respectivamente.	51
7.2. Búsqueda intensiva de hiperparámetros para SVR, el mejor conjunto de hiperparámetros se evalua mediante el Error Absoluto Medio Negativo, se presentan los 10 mejores conjuntos de hiperparámetros obtenidos identificando la iteración en la cual ocurrió.	54
7.4. Valores SHAP calculados para las predicciones.	55
7.5. Suma de valores Shap y correlación de cada caracterísitca con la Edad.	55
7.6. Suma de las 20 mejores SHAP por región cerebral, se normaliza entre 0 y 1 para analizar la importancia de cada región a la predicción.	55
7.7. Regiones cerebrales mas influyentes en la predicción de Estimación de Edad Cerebral para SVR, las regiones más importantes fueron la Occipital izquierda y derecha.	55
7.8. Búsqueda intensiva de hiperparámetros para KNN, el mejor conjunto de hiperparámetros se evalua mediante el Error Absoluto Medio Negativo, se presentan los 10 mejores conjuntos de hiperparámetros obtenidos identificando la iteración en la cual ocurrió.	57
7.10. Valores SHAP calculados para las predicciones.	58
7.11. Suma de valores Shap y correlación de cada caracterísitca con la Edad.	58
7.12. Suma de las 20 mejores SHAP por región cerebral, se normaliza entre 0 y 1 para analizar la importancia de cada región a la predicción.	58
7.13. Regiones cerebrales mas influyentes en la predicción de Estimación de Edad Cerebral para KNN, las regiones más importantes fueron la Frontal Media izquierda y Occipital derecha.	58

7.14. Búsqueda intensiva de hiperparámetros para Bosques Aleatorios, el mejor conjunto de hiperparámetros se evalúa mediante el Error Absoluto Medio Negativo, se presentan los 10 mejores conjuntos de hiperparámetros obtenidos identificando la iteración en la cual ocurrió.	60
7.16. Valores SHAP calculados para las predicciones.	61
7.17. Suma de valores Shap y correlación de cada característica con la Edad.	61
7.18. Suma de las 20 mejores SHAP por región cerebral, se normaliza entre 0 y 1 para analizar la importancia de cada región a la predicción.	61
7.19. Regiones cerebrales más influyentes en la predicción de Estimación de Edad Cerebral para Bosques Aleatorios, las regiones más importantes fueron la Occipital derecha y Cingular derecha.	61
7.20. Búsqueda intensiva de hiperparámetros para XGBoost, el mejor conjunto de hiperparámetros se evalúa mediante el Error Absoluto Medio Negativo, se presentan los 10 mejores conjuntos de hiperparámetros obtenidos identificando la iteración en la cual ocurrió.	63
7.22. Valores SHAP calculados para las predicciones.	64
7.23. Suma de valores Shap y correlación de cada característica con la Edad.	64
7.24. Suma de las 20 mejores SHAP por región cerebral, se normaliza entre 0 y 1 para analizar la importancia de cada región a la predicción.	64
7.25. Regiones cerebrales más influyentes en la predicción de Estimación de Edad Cerebral para XGBoost, las regiones más importantes fueron la Occipital derecha y Cingular derecha.	64
7.26. Búsqueda intensiva de hiperparámetros para CatBoost, el mejor conjunto de hiperparámetros se evalúa mediante el Error Absoluto Medio Negativo, se presentan los 10 mejores conjuntos de hiperparámetros obtenidos identificando la iteración en la cual ocurrió.	66
7.28. Valores SHAP calculados para las predicciones.	67
7.29. Suma de valores Shap y correlación de cada característica con la Edad.	67
7.30. Suma de las 20 mejores SHAP por región cerebral, se normaliza entre 0 y 1 para analizar la importancia de cada región a la predicción.	67
7.31. Regiones cerebrales más influyentes en la predicción de Estimación de Edad Cerebral para CatBoost, las regiones más importantes fueron la Occipital derecha y Cingular derecha.	67
7.32. Búsqueda intensiva de hiperparámetros para MLP, el mejor conjunto de hiperparámetros se evalúa mediante el Error Absoluto Medio Negativo, se presentan los 10 mejores conjuntos de hiperparámetros obtenidos identificando la iteración en la cual ocurrió.	69
7.34. Valores SHAP calculados para las predicciones.	70
7.35. Suma de valores Shap y correlación de cada característica con la Edad.	70
7.36. Suma de las 20 mejores SHAP por región cerebral, se normaliza entre 0 y 1 para analizar la importancia de cada región a la predicción.	70

7.37. Regiones cerebrales mas influyentes en la predicción de Estimación de Edad Cerebral para el modelo de Red Neuronal, las regiones más importantes fueron la Occipital derecha e izquierda.	70
7.38. Predicciones	71
7.39. Comparación de regiones cerebrales	72
7.40. Evaluación del modelo incrementando las características que entrega SHAP una a una para el algoritmo de XGBoost, es posible notar que a partir de las 20 caracterpisticas las métricas de evaluación comienzan a tomar un valor poco variable, es decir, bastan al menos 20 características para obtener un rendimiento óptimo del modelo. . .	73
7.41. Búsqueda intensiva de hiperparámetros para XGBoost utilizando las 20 mejores características entregada por su evaluación de valores SHAP.	74
7.42. Predicciones del a Edad Cerebral utilizando XGBoost y las 20 características que más contribuyen al modelo.	74
7.43. Valores SHAP	75
7.44. Suma de valores SHAP y correlación	75

ABREVIACIONES

Mayúsculas

EEG	: Electroencefalografía
MEG	: Magnetoencefalografía
MAE	: Error medio absoluto (mid absolute error)
MSE	: Error cuadrático medio (mid square error)
SVR	: Support vector regression
RF	: Random Forest
XGB	: XGBoost
MLP	: Mluti Layer Perceptron

Minúsculas

cv	: cross validation
----	--------------------

INTRODUCCIÓN

La estimación de la edad cerebral a partir de registros de electroencefalogramas (EEG) y técnicas de aprendizaje automático ha emergido como un campo de investigación prometedor en la neurociencia computacional. La edad cerebral, una medida del envejecimiento del cerebro, puede diferir de la edad cronológica y proporcionar información valiosa sobre la salud neurológica y el riesgo de enfermedades neurodegenerativas.

A través del EEG se puede visualizar la actividad eléctrica cerebral de manera directa, capturando patrones dinámicos y complejos que reflejan el funcionamiento cerebral en tiempo real. Esta técnica no invasiva es invaluable para el estudio del envejecimiento cerebral, ya que permite analizar cómo cambian los patrones de actividad a lo largo de la vida de un individuo.

Explorando varios enfoques de aprendizaje automático para estimar la edad cerebral utilizando datos de EEG, se implementan y comparan modelos de regresión avanzados, incluyendo modelos clásicos de aprendizaje de máquina tales como las máquinas de vectores, así como modelos más complejos tales como XGBoost o modelos basados en redes neuronales. La optimización de hiperparámetros mediante técnicas de búsqueda bayesiana se emplea para mejorar el rendimiento de los modelos y asegurar que las predicciones sean lo más precisas posible. El objetivo principal de este trabajo es desarrollar un marco metodológico sólido que no solo permita estimar con precisión la edad cerebral, sino también identificar biomarcadores cerebrales asociados con el envejecimiento.

1.1. Resumen de los Capítulos

Este documento está dividido en siete capítulos, el **Capítulo 1** es una introducción y primer acercamiento al tema de investigación. En el **Capítulo 2** se presentan los objetivos principales del tema a tratar, además se presentan los

alcances, limitaciones y contribuciones del tema. En el **Capítulo 3** se expone el estado del arte con la bibliografía detallada sobre otros estudios existentes con respecto a la estimación de la edad cerebral. En el **Capítulo 4** se presentan los principales fundamentos teóricos. En el **Capítulo 5** se detallan los pasos a seguir para la estimación final de la edad cerebral. En el **Capítulo 6** se describe la implementación de los algoritmos en Python. En el **Capítulo 7** se presentan y describen los resultados para los distintos algoritmos y una comparativa mediante métricas de evaluación. Y finalmente en el **Capítulo 8** se realizan observaciones sobre todo el trabajo realizado, resumiendo los principales resultados obtenidos dentro de un marco tanto cualitativo como cuantitativo.

OBJETIVOS

2.1. Objetivo General

El objetivo principal de este trabajo es evaluar el rendimiento de distintos modelos de aprendizaje de máquina para la estimación de la edad biológica cerebral en base a registros de electroencefalografía de 1377 personas de estado saludable de un estudio realizado por el consorcio de EuroLaD-EEG [1]. Esto se realizará utilizando características extraídas de los registros EEG focalizada en distintas regiones cerebrales con el fin de determinar que regiones y características son las que más influyen en el envejecimiento cerebral.

2.2. Objetivos Específicos

Para este propósito, se llevarán a cabo las siguientes etapas:

1. Búsqueda de los mejores hiperparámetros para los distintos algoritmos de aprendizaje de máquina que se ajusten a los datos EEG utilizados.
2. Evaluación de los distintos algoritmos de aprendizaje de máquina para estimación de la edad biológica cerebral.
3. Identificar las características y regiones cerebrales más relevantes en la estimación de la edad cerebral.

2.3. Alcances y Limitaciones

Los modelos de aprendizaje automático tienen una amplia gama de alcances y limitaciones, especialmente en aplicaciones específicas como la estimación de la edad cerebral mediante características de EEG, pudiendo procesar grandes volúmenes de datos y detectar patrones complejos que no son evidentes para los humanos,

así pueden identificar signos tempranos de envejecimiento cerebral anormal, lo que es crucial para intervenciones tempranas en condiciones neurodegenerativas, sin embargo estos modelos suelen estar limitados puesto que requieren grandes cantidades de datos para entrenarse adecuadamente. Un modelo entrenado en un conjunto de datos específicos puede no generalizar bien a otros conjuntos de datos con diferentes características demográficas, puesto que los modelos pueden estar sesgados si los datos de entrenamiento no son representativos de la población general, finalmente los modelos avanzados, especialmente aquellos que utilizan técnicas de aumento de gradientes o redes neuronales, pueden ser computacionalmente costosos y requerir hardware especializado para entrenarse eficientemente.

2.4. Contribuciones

En este documento se generó un flujo de trabajo para la estimación de la edad cerebral mediante registros de EEG, además de comparar los distintos métodos y algoritmos para realizar la estimaciones, para así poder descubrir cuales son los algoritmos más óptimos para realizar este tipo de tareas. Finalmente se logran obtener cuales son las características y regiones más importantes en la estimación de la edad cerebral.

ESTADO DEL ARTE

LA edad biológica en contraste de la edad cronológica que una persona pueda tener surge como un potente bioindicador del estado de salud de dicha persona, lo cual ha sido estudiado durante varias décadas [2] [3] [4] abordando diversos biomarcadores que funcionarían como "*relojes biológicos*", dando cuenta del estado de envejecimiento de cada individuo. Dentro de ese marco surge la edad biológica cerebral como un posible bioindicador para caracterizar los diversos trastornos neurodegenerativos asociados al envejecimiento del sistema nervioso central que afecta a numerosas personas alrededor del mundo. Mientras la edad cronológica está sujeta a la fecha de nacimiento de cada individuo la "*edad cerebral*" puede diferir de esta última, representando un cerebro más o menos joven con respecto a una muestra de sujetos [5].

Hasta el momento los métodos más comunes para la estimación del deterioro cognitivo han sido empleados mediante Resonancia magnética o Resonancia magnética funcional, trabajos como los de Cole y Franke, muestran que mediante imágenes de Resonancia magnética y modelos basados en redes neuronales convolucionales permitieron obtener con alta precisión un estimador de la edad biológica cerebral y asociaron esta edad cerebral al riesgo de mortalidad [6][7][8][9] y trabajos como los de Besheti [10] muestran cómo esta edad cerebral está directamente relacionada a enfermedades neurológicas como lo es la enfermedad de Alzheimer, donde sujetos que presentaban dicha enfermedad a la vez presentaban un cerebro más envejecido.

Aunque estos modelos basados en neuroimagen de Resonancia magnética han logrado dar buenos resultados con respecto a la caracterización del deterioro cerebral al igual que otros modelos basados en Magnetoencefalografía (MEG) tienen el problema de ser poco comunes en los centros de salud pública llevando dichos resultados a un sesgo poblacional, provocando que hoy en día sea necesario buscar alternativas más accesibles para la población en general, y es en este ámbito donde la

Electroencefalografía (EEG) juega un papel fundamental al momento de caracterizar los trastornos neurodegenerativos siendo capaz de marcar una diferencia en el campo de la neurociencia [11].

Los enfoques actuales han aprovechado los algoritmos de aprendizaje de máquina para lograr una mejor caracterización de los registros de EEG, logrando predecir diversos tipos de trastornos y patologías, tales como la enfermedad de Alzheimer o la Demencia frontotemporal temprana[12]. Los trabajos de Gemein se han enfocado en implementar y comparar 2 enfoques clásicos en el ámbito del aprendizaje de máquina y el análisis de registros de EEG[13], en primer lugar utiliza métodos de aprendizaje profundo como redes neuronales convolucionales para tratar los datos de EEG como si fuera una imagen y donde a través de las capas ocultas es posible extraer características complejas de manera implícita, para finalmente realizar una clasificación de patología en el caso de variables categóricas[14] o bien estimar la edad cerebral como una variable continua[15]. Y un segundo enfoque es realizado a través de una extracción manual de características, que pueden ser espectrales periódicas, aperiódicas, de complejidad, entre otras, para posteriormente aplicar los algoritmos clásicos de aprendizaje de máquina de regresión y clasificación [16] tales como regresión lineal, máquinas de soporte, bosques aleatorios, XGBoost entre otros[17][18].

A pesar de los buenos resultados obtenidos por los métodos basados en aprendizaje profundo en cuanto a las predicciones que pueden realizar, superando en la mayoría de los casos a cualquier otro enfoque de aprendizaje de máquina, al momento de requerir obtener información específica de biomarcadores que logren caracterizar el deterioro cerebral estos métodos debido a sus capas ocultas y extracción implícita de características no son capaces de dar cuenta con facilidad qué características biológicas son los que influyen principalmente en la determinación de este deterioro cerebral, por lo tanto los enfoques basados en la extracción manual de características podrían determinar de mejor manera cuales son estas características biológicas que logran identificar el deterioro cerebral.

Un trabajo de Zoubi [19] publicado en 2018 muestra un enfoque clásico para estimación de edad cerebral basado en EEG, que incluye un preprocesamiento inicial, un extracción de características espectrales y una posterior regresión realizada con los algoritmos de aprendizaje de máquina, logrando un valor de $R^2 = 0,34$ con el algoritmo de máquina de soportes. Luego Engemann en 2022[20], basado en los trabajos anteriores de Gemein realiza una comparación de diversos enfoques de aprendizaje de máquina, entre ellos el de aprendizaje profundo y la extracción de características manuales para predecir la edad cerebral, además, los enfoques multimétricos dan cuenta de la necesidad de abordar estos problemas desde más de una arista para lograr obtener un mejor resultado.

En la siguiente memoria se propone utilizar una base de datos que involucra a 1298 participantes de estado saludable de un estudio realizado por el consorcio de EuroLaD-EEG [1] que abarca regiones de América Latina (BrainLat) [21] y Europa

y seguir el flujo de trabajo publicado por Prado para armonización y tratamiento de datos EEG para la caracterización de la enfermedad de Alzheimer, que en una primera instancia incluye un preprocesamiento, normalización y transformación de fuente mediante sLORETA[22], luego, a partir de estas regiones cerebrales obtenidas con sLORETA se extraen distintas características espectrales periódicas[23] y aperiódicas en la banda de frecuencia alpha[24], dividiendo los datos en una partición de 8:2 para datos de entrenamiento y prueba respectivamente[25], utilizando los algoritmos clásicos de aprendizaje de máquina tales como regresión lineal múltiple, máquina de vectores de soporte, bosques aleatorios, redes neuronales y XGBoost. Se propone una validación cruzada de 10 pliegues [26][27] y métricas de evaluación el Error Cuadrático Medio (MSE) y el Error Absoluto Medio (MAE) lo más bajo posible y un Coeficiente de Determinación (R^2) lo más alto posible.

FUNDAMENTOS TEÓRICOS

EL CONCEPTO de la edad biológica cerebral abarca la idea de que el cerebro puede envejecer de manera distinta a otros órganos del cuerpo, lo que puede reflejarse en cambios en la actividad eléctrica y en la estructura cerebral. Entender estos procesos es crucial para desarrollar métodos precisos de estimación de la edad cerebral.

Teniendo al EEG como herramienta invaluable en este trabajo, ya que permite registrar la actividad eléctrica generada por el cerebro en tiempo real y con una alta resolución temporal. Las distintas frecuencias de las ondas cerebrales capturadas por el EEG proporcionan información sobre el estado funcional del cerebro, desde el reposo hasta la realización de tareas cognitivas específicas.

4.1. Edad Biológica Cerebral

En el estudio del envejecimiento y la salud cognitiva, se emplean dos conceptos fundamentales: la edad cronológica y la edad biológica. Mientras que la edad cronológica hace referencia al tiempo transcurrido desde el nacimiento de un individuo, la edad biológica representa una medida más compleja, que evalúa el estado de salud y funcionamiento de los tejidos, órganos y sistemas del cuerpo en relación con la edad cronológica. Dentro de este marco, emerge la noción de la edad cerebral, una perspectiva específica que se concentra en el desarrollo y el envejecimiento del órgano central del sistema nervioso: el cerebro.

La edad cerebral constituye una evaluación del estado y funcionamiento del cerebro en comparación con la edad cronológica de un individuo. Este concepto no solo abarca la estructura física del cerebro, sino también su funcionamiento cognitivo y emocional. Es decir, la edad cerebral busca comprender cómo el cerebro envejece o se desarrolla en comparación con el resto del cuerpo a lo largo del tiempo.

La estimación de la edad cerebral se vuelve esencial en diversos campos

de la neurociencia y para llevar a cabo esta estimación, se recurre a una variedad de técnicas, entre las mas comunes destacan el fMRI (imagen por resonancia magnética funcional), el MRI (imagen por resonancia magnética), el MEG (magnetoencefalografía) y el EEG (electroencefalografía). Estas herramientas permiten examinar la estructura, la función y la conectividad del cerebro, proporcionando así una evaluación más precisa de su edad biológica.

4.2. Electroencefalograma

El EEG registra la actividad eléctrica del cerebro mediante electrodos colocados en el cuero cabelludo. Esta técnica proporciona una excelente resolución temporal, permitiendo el registro de la actividad cerebral en tiempo real con alta precisión temporal. El EEG es especialmente útil para estudiar eventos cerebrales rápidos y transitorios, como los relacionados con el procesamiento sensorial, la atención y la memoria.

La actividad eléctrica cerebral se genera por la interacción compleja de las neuronas en el cerebro. Cuando las neuronas se activan, producen corrientes eléctricas que pueden ser detectadas en la superficie del cuero cabelludo mediante electrodos. Estas corrientes se generan principalmente en las sinapsis neuronales, donde las señales químicas entre las neuronas se traducen en señales eléctricas. La actividad eléctrica neuronal puede variar en amplitud, frecuencia y sincronización, lo que refleja diferentes procesos neuronales y estados mentales.

El EEG registra principalmente cinco tipos de ondas cerebrales como se muestra en la Figura 4.1, cada una asociada con diferentes estados de conciencia y funciones cognitivas:

- Ondas Delta (δ): Son las ondas de menor frecuencia, oscilando entre 0.5 y 4 hercios (Hz). Se observan principalmente durante el sueño profundo (fase de sueño de onda lenta) y están relacionadas con la restauración y el mantenimiento del cuerpo.
- Ondas Theta (Θ): Tienen una frecuencia de alrededor de 4 a 8 Hz. Se presentan en mayor medida durante la transición entre la vigilia y el sueño, así como durante ciertos estados de relajación profunda. Las ondas theta están asociadas con la creatividad, la imaginación y la memoria episódica.
- Ondas Alfa (α): Con una frecuencia de aproximadamente 8 a 12 Hz, las ondas alfa son prominentes en el estado de vigilia relajada, especialmente cuando los ojos están cerrados. Se cree que reflejan una actividad neuronal sincronizada en la corteza cerebral y están asociadas con la relajación y la atención selectiva.
- Ondas Beta (β): Tienen una frecuencia de alrededor de 12 a 30 Hz y se observan predominantemente durante la vigilia activa y el estado de alerta. Las ondas beta están relacionadas con la actividad mental enfocada, el pensamiento analítico y la resolución de problemas.

- Ondas Gamma (γ): Son las ondas de mayor frecuencia, con una frecuencia de más de 30 Hz. Las ondas gamma están vinculadas con procesos cognitivos superiores, como la percepción sensorial, la atención, la memoria de trabajo y el aprendizaje.

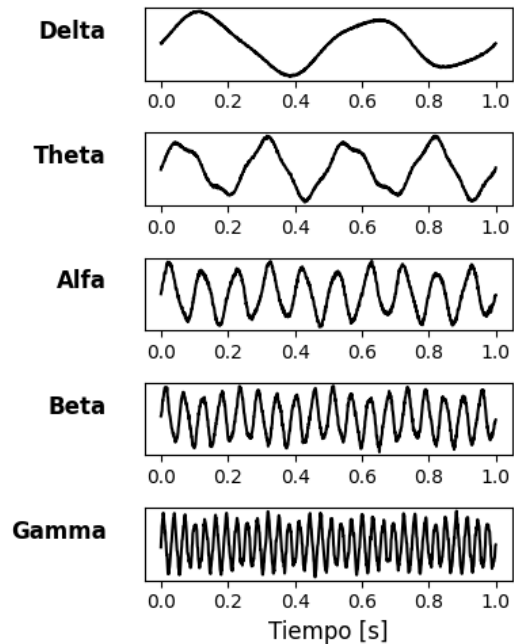


Figura 4.1: **Ondas Cerebrales:** Principales ondas cerebrales registradas en el electroencefalograma

Estas ondas cerebrales se pueden observar en diferentes combinaciones y patrones en el EEG, proporcionando información valiosa sobre el estado funcional del cerebro en distintas condiciones y estados mentales. El análisis de estas ondas, junto con otros parámetros, como la amplitud y la sincronización, permite obtener una comprensión más profunda de la actividad cerebral y sus correlatos con la función cognitiva y el estado de salud mental.

4.3. Preprocesamiento

El preprocesamiento clásico de una señal EEG es esencial para garantizar la calidad de los datos y eliminar artefactos no deseados que puedan interferir con el análisis posterior. Este proceso generalmente incluye una serie de pasos que varían según las características específicas de los datos y los objetivos del estudio. Primero, se realiza el filtrado para eliminar componentes de frecuencia no deseados como el ruido de línea eléctrica, artefactos musculares y de movimiento, utilizando filtros pasa bajos, pasa altos o pasa banda. Luego, se procede a la remoción

de artefactos provenientes de movimientos oculares, movimientos musculares e interferencias eléctricas externas, utilizando técnicas como el rechazo de épocas y métodos de regresión. El análisis de componentes independientes (ICA) es una técnica avanzada que separa las fuentes de actividad cerebral de los artefactos en la señal EEG, permitiendo la identificación y eliminación de los componentes de artefactos. En ocasiones, es necesario ajustar la frecuencia de muestreo de la señal EEG mediante el resampleo, utilizando técnicas de interpolación o decimación para facilitar el procesamiento o la comparación con otros conjuntos de datos. La referencia promedio es una técnica común para eliminar el sesgo introducido por la referencia original, calculando la media de todas las señales de EEG y sustrayéndola de cada canal individual. Estos pasos de preprocesamiento son fundamentales para asegurar que la señal EEG sea de alta calidad y esté libre de contaminaciones, lo que permite una interpretación y análisis más precisos de la actividad cerebral.

4.4. Extracción de Características

La extracción de características de señales EEG es un paso fundamental que logra dar interpretabilidad a estas señales temporales. Entre las principales características que se pueden extraer de una señal EEG se encuentran las características de tipo frecuencial periódico, frecuencial aperiódico y métricas de complejidad.

4.4.1. Características Frecuenciales Periódicas

Las métricas espectrales periódicas se refieren a los componentes de la señal EEG que tienen una naturaleza rítmica o cíclica, estas métricas permiten analizar cómo la energía de la señal se distribuye a lo largo de diferentes bandas de frecuencia definidas de la señal EEG. Un ejemplo es la Densidad de Potencia Espectral (PSD) que mide la potencia de las señales EEG distribuidas a través de distintas frecuencias.

$$\text{PSD}(f) = \frac{1}{T} |X(f)|^2, \quad (4.1)$$

donde $X(f)$ es la Transformada de Fourier de la señal $x(t)$.

Además se tiene la densidad de potencia espectral normalizada (nPSD), que se obtiene dividiendo la PSD por la suma total de la PSD en todas las frecuencias, normalizando así la potencia espectral.

$$\text{nPSD}(f) = \frac{\text{PSD}(f)}{\sum_f \text{PSD}(f)} \quad (4.2)$$

Esto puede también expresarse como la proporción de la potencia en una banda específica respecto a la potencia total de la señal.

4.4.2. Características Frecuenciales Aperiódicas

Las métricas espectrales aperiódicas se refieren a los componentes del espectro de potencia de una señal EEG que no muestran periodicidad o ritmicidad clara. Se puede modelar de la siguiente manera:

$$A(f) = b - \log(k + f^x), \quad (4.3)$$

donde A es el valor del componente aperiódico en la frecuencia f , b es el desplazamiento (offset), k el punto de inflexión (knee), x es el exponente de la frecuencia que controla la pendiente de la caída en el espectro.

4.5. Aprendizaje Automático

Para lograr realizar la estimación de la edad cerebral mediante las características extraídas de los registros de electroencefalografía se tienen diversos algoritmos de aprendizaje de máquina para realizar una regresión de forma lineal y no lineal, pudiendo, mediante un conjunto de características X predecir una variable dependiente Y , en este caso la edad cerebral.

Los algoritmos a utilizar son regresión lineal, support vector regression, random forest, redes neuronales y XGBoost [20] [19] [17].

4.5.1. Regresión Lineal Simple

En la regresión lineal simple se tiene una única variable independiente x para modelar una variable dependiente y , teniendo la siguiente relación lineal entre las 2 variables:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \forall i, \quad (4.4)$$

donde y_i y x_i representan una instancia específica i . Además el parámetro β_0 representa el intercepto de la recta, el parámetro β_1 es la pendiente y los valores ϵ_i corresponde a los errores asociados al modelo, esto se modela como una linea recta como se aprecia en la Figura 4.2.

Mediante el método de mínimos cuadrados ordinarios se puede estimar los valores de β_0 y β_1 . Se calcula la suma de los errores cuadráticos (SSE) o E :

$$E = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2 \quad (4.5)$$

Se minimiza el error calculando las derivadas parciales de E con respecto a β_0 y β_1 , se igualan las derivadas a cero y se resuelve el sistema para encontrar los parámetros:

$$\frac{\partial E}{\partial \beta_0} = -2 \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) = 0 \quad (4.6)$$

$$\frac{\partial E}{\partial \beta_1} = -2 \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) x_i = 0 \quad (4.7)$$

Obteniendo las siguientes soluciones:

$$\beta_1 = \frac{\sum_i^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_i^n (x_i - \bar{x})^2} \quad (4.8)$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x} \quad (4.9)$$

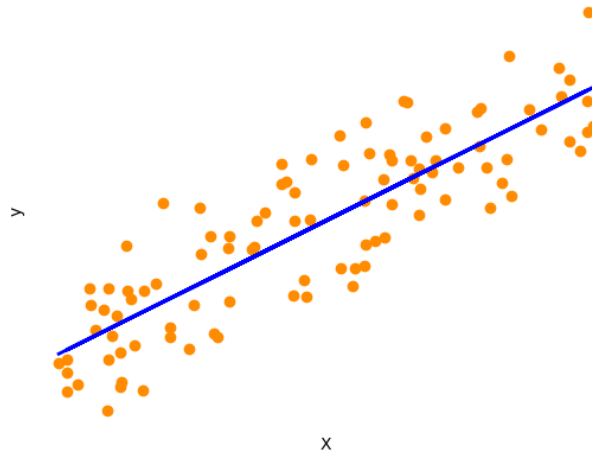


Figura 4.2: Modelo de Regresión lineal simple

4.5.2. Regresión Lineal Múltiple

Ahora, para un modelo multimétrico se deben usar modelos de regresión múltiple, donde se tienen m variables independientes $x_1, x_2, \dots, x_m$ que modelan una variable independiente y , definiendo el modelo lineal multivariado siguiente:

$$y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \dots + \beta_m x_{i,m} + \epsilon_i, \forall i \in \{1, n\} \quad (4.10)$$

Este modelo genera un hiperplano que modela las variables de interés x , figura 4.3, Visto de forma matricial:

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon, \quad (4.11)$$

donde $\mathbf{Y}$ es un vector de tamaño $n \times 1$ y $\mathbf{X}$ es una matriz de tamaño $n \times (m + 1)$ correspondiente a las distintas métricas explicativas de la forma:

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad (4.12)$$

$$\mathbf{X} = \begin{pmatrix} x_{1,0} & x_{1,1} & x_{1,2} & \cdots & x_{1,m} \\ x_{2,0} & x_{2,1} & x_{2,2} & \cdots & x_{2,m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n,0} & x_{n,1} & x_{n,2} & \cdots & x_{n,m} \end{pmatrix} \quad (4.13)$$

donde n representa el número de instancias y m corresponde al número de características, además la primera columna es constante igual a 1 para incluir la variable β_0 del intercepto.

Luego se tiene β que es un vector de tamaño $(n+1) \times 1$:

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_n \end{pmatrix} \quad (4.14)$$

$$SSE = (\mathbf{Y} - \mathbf{X}\beta)^T (\mathbf{Y} - \mathbf{X}\beta) \quad (4.15)$$

$$\beta = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \quad (4.16)$$

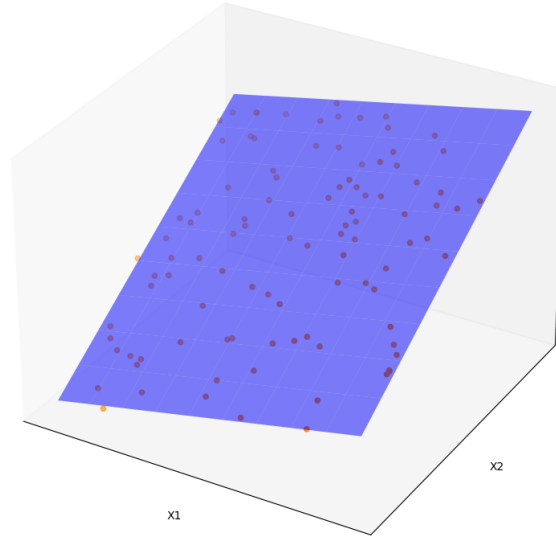


Figura 4.3: Modelo de regresión lineal múltiple

4.5.3. Support Vector Regression

La Regresión de Máquinas de Vectores de Soporte (SVR) [28] es una extensión del algoritmo de Máquinas de Vectores de Soporte (SVM) utilizado para problemas de regresión en lugar de clasificación. En SVR, la idea principal es encontrar una función que tenga como máximo una desviación ϵ de los valores objetivo reales para todas las muestras de entrenamiento, Figura 4.4. Esto se traduce en la búsqueda de una función $f(x)$ tal que:

$$f(x) = \langle w, x \rangle + b, \quad (4.17)$$

donde $\langle w, x \rangle$ es el producto punto entre el vector de pesos w y el vector de características x , y b el intercepto. La tolerancia ϵ define una "banda" alrededor de la función $f(x)$ donde los errores no se penalizan. Los puntos de datos que caen dentro de esta banda no contribuyen al error.

La función objetivo de SVR es la siguiente:

$$\min_{\mathbf{w}, b} \left(\frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \right) \quad (4.18)$$

Sujeto a las restricciones:

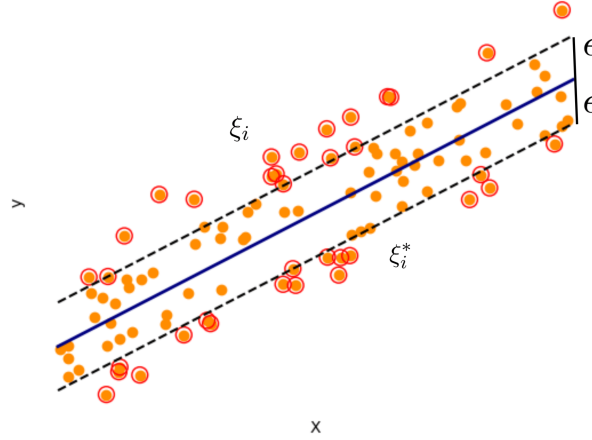


Figura 4.4: Modelo de máquina de vectores de soporte para regresión, la banda ϵ representa la banda de error permitida en el modelo, mientras que los puntos ξ representan los valores de holgura para los vectores de soporte.

$$\begin{aligned}
 y_i - \mathbf{w}\mathbf{x}_i - b &\leq \epsilon + \xi_i \quad \text{para } i = 1, 2, \dots, n \\
 -y_i + \mathbf{w}\mathbf{x}_i + b &\leq \epsilon + \xi_i^* \quad \text{para } i = 1, 2, \dots, n \\
 \xi_i, \xi_i^* &\geq 0 \quad \text{para } i = 1, 2, \dots, n
 \end{aligned} \tag{4.19}$$

Donde C es el parámetro de regularización que controla el equilibrio entre la amplitud del margen y el error de entrenamiento, ξ_i y ξ_i^* son variables de holgura que permiten manejar puntos que están fuera del margen ϵ , donde ϵ es el parámetro de insensibilidad que define un margen dentro del cual los errores no se penalizan.

Las predicciones se obtienen como:

$$\hat{y} = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \langle x_i, x \rangle + b \tag{4.20}$$

Donde α_i y α_i^* corresponden a los multiplicadores de Lagrange obtenidos en la optimización del modelo.

4.5.4. KNN

El algoritmo de los vecinos más cercanos (KNN) se basa en predecir un valor en base al valor de los datos que estén más cercanos al valor a predecir, en este caso, K representa el número de vecinos más cercanos a considerar al hacer una predicción, Figura 4.5.

Para determinar qué puntos son los "más cercanos", se utiliza una medida de distancia. La más común es la Distancia Euclidiana:

$$\text{Distancia Euclidiana}(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{\sum_{k=1}^d (x_{ik} - x_{jk})^2} \quad (4.21)$$

$$\text{Distancia Manhattan}(\mathbf{x}_i, \mathbf{x}_j) = \sum_{k=1}^d |(x_{ik} - x_{jk})| \quad (4.22)$$

Donde $\mathbf{x}_i$ y $\mathbf{x}_j$ son dos puntos de datos y d es la dimensión (número de características) de los datos.

Finalmente la predicción está dada por el promedio entre los vecinos más cercanos a ese punto.

$$\hat{y} = \frac{1}{K} \sum_{i=1}^K y_i \quad (4.23)$$

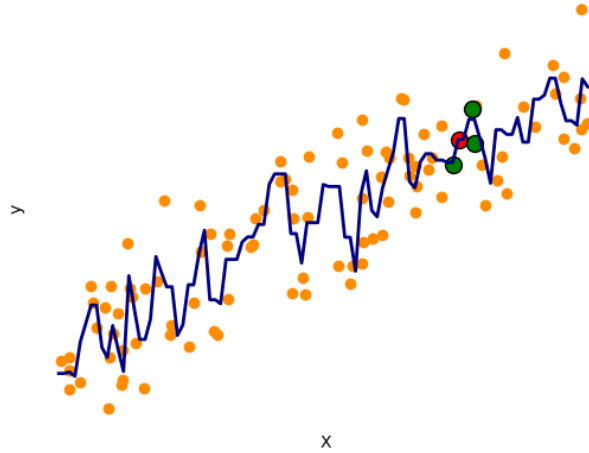


Figura 4.5: Modelo de los K vecinos más cercanos, la predicción (punto rojo) viene dado por el promedio entre los vecinos más cercanos (puntos verdes)

4.5.5. Random Forest

El algoritmo de bosques aleatorios (Random Forest) es un método de aprendizaje automático basado en el conjunto de árboles de decisión Figura 4.6, estos árboles de decisión corresponden a un modelo predictivo que divide el espacio de características en regiones homogéneas mediante reglas de decisión.

Se crea un número fijo de árboles de decisión y para cada uno de estos, se genera un subconjunto de datos mediante muestreo con reemplazo (bootstrap). Esto significa que algunos datos pueden repetirse en el subconjunto y algunos pueden

omitirse, además, en cada nodo de cada árbol se incluye solo un subconjunto de características aleatorias, esto introduce una mayor variabilidad en los árboles y reduce la correlación entre ellos.

Para cada nodo del árbol, selecciona la mejor característica x_i y el punto de división que minimicen el error de regresión como función objetivo.

$$E = \frac{1}{n} \sum_{i=1}^n l(y_i, \hat{y}_i) \quad (4.24)$$

Donde n es el total de muestras y l la función error objetivo, que puede ser alguna métrica de evaluación como el MSE o el MAE.

Finalmente la predicción final del modelo corresponde al promedio entre las predicciones de todos los árboles.

$$\hat{y} = \frac{1}{K} \sum_{i=1}^K \hat{y}_i \quad (4.25)$$

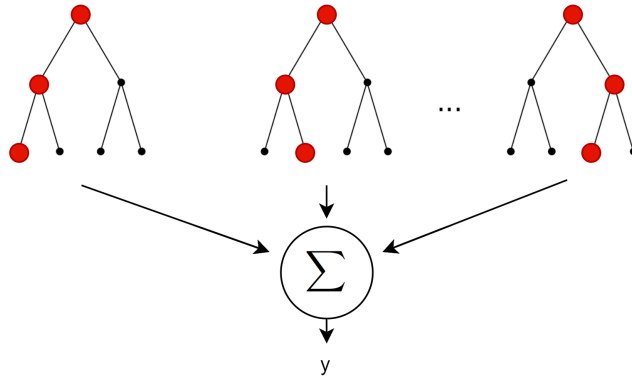


Figura 4.6: Modelo de bosques aleatorios para regresión, donde la predicción es el promedio de las predicciones de todos los árboles del bosque.

4.5.6. XGBoost

XGBoost o Extreme Gradient Boosting es una implementación de modelos de árboles de decisión ensamblados [29], el proceso se realiza de manera secuencial, a diferencia de Random Forest donde todos los árboles se creaban de forma paralela, además, en XGBoost cada nuevo modelo intenta corregir los errores cometidos por el conjunto anterior Figura 4.7.

La predicción se consigue como la suma de todas las predicciones de los árboles.

$$\hat{y}_i = \phi(x_i) = \sum_{k=1}^K f_k(x_i) \quad (4.26)$$

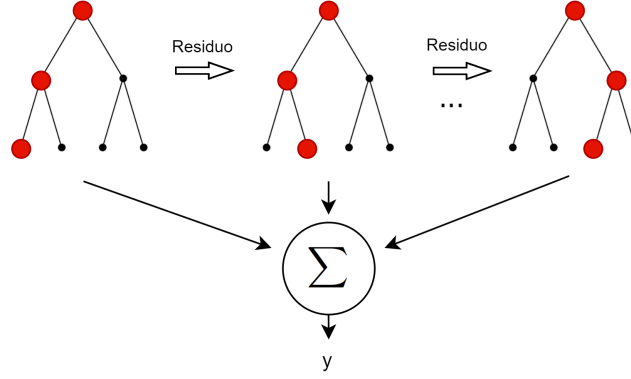


Figura 4.7: Modelo de XGBoost, se genera cada árbol de forma consecutiva y la predicción final se obtiene como la suma de la predicción de todos los árboles, en este caso cada árbol no predice la variable y como tal, si no que va mejorando el resultado del árbol anterior.

La función objetivo de XGBoost se define como la suma entre el error del modelo más un factor de regularización.

$$\mathcal{L}^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) \quad (4.27)$$

Donde el factor de regularización está dado por:

$$\Omega(f_t) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (4.28)$$

donde T es el número de hojas en el árbol, γ es el parámetro de regularización para el número de hojas, λ es el parámetro de regularización L2

Para realizar la actualización de la función objetivo se realiza la expansión por serie de Taylor:

$$\ell(y, \hat{y} + f(x)) \approx \ell(y, \hat{y}) + g f(x) + \frac{1}{2} h f(x)^2 \quad (4.29)$$

- $g = \frac{\partial \ell(y, \hat{y})}{\partial \hat{y}}$: corresponde al gradiente de la función
- $h = \frac{\partial^2 \ell(y, \hat{y})}{\partial \hat{y}^2}$: corresponde al hessiano de la función

$$\mathcal{L}^{(t)} \approx \sum_{i=1}^n \left[\ell(y_i, \hat{y}_i^{(t-1)}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t(x_i)^2 \right] + \Omega(f_t) \quad (4.30)$$

4.5.7. CatBoost

CatBoost (Categorical Boosting) [30] al igual que XGBoost, se basa en el método de aumento de gradiente sobre árboles de decisión siendo optimizado para manejar características categóricas de manera eficiente. Presenta una función objetivo similar a XGBoost pero no se expone de forma explícita el parámetro de regularización, aunque el algoritmo como tal si implementa dicha regularización interna.

$$\mathcal{L}(\theta) = \sum_{i=1}^n \ell(y_i, \hat{y}_i) \quad (4.31)$$

Las predicciones del modelo se calculan como:

$$\hat{y} = F_T(x) = \sum_{t=0}^T \alpha h_t(x) \quad (4.32)$$

donde α es un parámetro que regula la importancia de las predicciones como tamaño de paso. Si α es demasiado grande, el modelo puede sobreajustarse a los datos de entrenamiento rápidamente y no generalizar bien a datos nuevos. Si es demasiado pequeño, el modelo puede necesitar muchas iteraciones para converger, haciendo el proceso de entrenamiento muy lento.

4.5.8. Red Neuronal

Una red neuronal es un modelo computacional inspirado en el funcionamiento del cerebro humano, diseñado para aprender patrones y realizar tareas específicas mediante la interconexión de unidades básicas llamadas neuronas. Estas redes están compuestas por capas de neuronas interconectadas, cada una de las cuales procesa y transmite información a través de conexiones ponderadas.

Un tipo específico de red neuronal es la Red Neuronal Multicapa (MLP) Figura 4.8, que es un tipo de red neuronal con propagación hacia adelante que consiste en múltiples capas de neuronas, donde cada capa está completamente conectada a la siguiente.

Estructura de una red MLP:

- Capa de Entrada: Recibe el vector de características $X = (x_1, x_2, x_3)$
- Capas Ocultas: Procesan la información recibida de la capa anterior. Cada neurona en una capa oculta aplica una transformación lineal seguida de una función de activación no lineal.

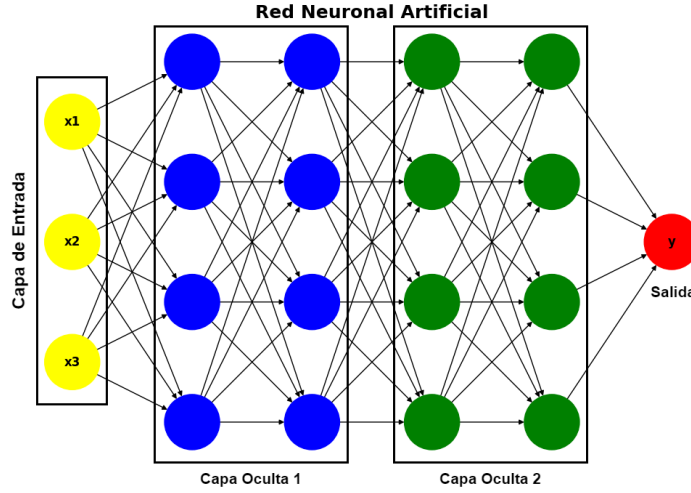


Figura 4.8: Modelo de Red Neuronal Multicapa, en este caso cuenta con una capa de entrada de 3 neuronas que representan las características del modelo a predecir, cuenta con 2 capas ocultas densas que se encargan de procesar la información de entrada, para finalmente pasar por una capa lineal que convierte la información procesada en un valor continuo y .

- Capa de Salida: Genera la salida de la red. En el caso de regresión, suele ser una función lineal.

Para una neurona j en una capa oculta o de salida, la combinación lineal de sus entradas z_j se calcula como:

$$z_j = \sum_{i=1}^k w_{ij}x_i + b_j \quad (4.33)$$

Donde w_{ij} es el peso que conecta la entrada x_i con la neurona j y b_j es el sesgo de la neurona j

La salida de la neurona j , a_j , se obtiene aplicando una función de activación ϕ a la combinación lineal z_j como se muestra en la Figura 4.9

$$a_j = \phi(z_j) \quad (4.34)$$

Funciones de activación:

- Sigmoides: $\phi(z) = \frac{1}{1+e^{-z}}$
- Tangente hiperbólica ($\tanh$) : $\phi(z) = \frac{e^z + e^{-z}}{e^z - e^{-z}}$
- ReLU: $\phi(z) = \max(0, z)$

funcion objetivo:

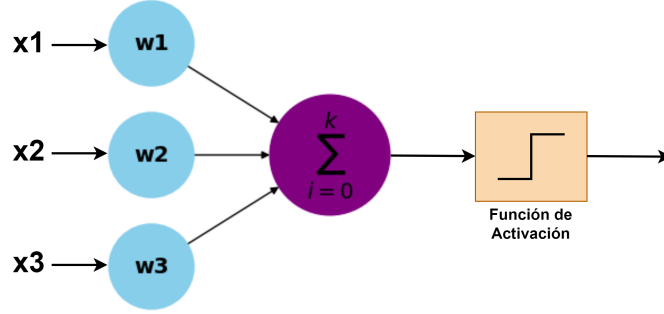


Figura 4.9: Activación de una neurona, las características se combinan con sus pesos correspondientes para luego entrar a la neurona que realiza la sumatoria de los valores de entrada para posteriormente pasarlos a una función de activación.

$$L(\mathbf{y}, \hat{\mathbf{y}}) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i) \quad (4.35)$$

Derivada de la función objetivo:

$$\frac{\partial L}{\partial w_{ij}} = \frac{2}{n} (y_i - \hat{y}_i) \frac{\partial \hat{y}_i}{\partial w_{ij}} \quad (4.36)$$

Actualización de Pesos:

$$w_{ij} \leftarrow w_{ij} - \eta \frac{\partial L}{\partial w_{ij}} \quad (4.37)$$

4.5.9. Evaluación del Modelo

Para saber qué tan bueno es el rendimiento de un modelo de aprendizaje automático, se utilizan distintas métricas de evaluación. A continuación, se describen algunas de las más comunes y su importancia en la evaluación de modelos:

1. Error Cuadrático Medio (MSE)

El Error Cuadrático Medio es una métrica que mide el promedio de los cuadrados de los errores, es decir, la diferencia promedio al cuadrado entre los valores observados y_i y los valores predichos $\hat{y}_i$. Esta métrica penaliza fuertemente los errores grandes, lo que la hace útil en situaciones donde es crucial minimizar grandes desviaciones.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4.38)$$

2. Error Absoluto Medio (MAE)

El Error Absoluto Medio mide el promedio de los errores absolutos, es decir, la diferencia promedio en valor absoluto entre los valores observados y_i y los valores predichos $\hat{y}_i$. A diferencia del MSE, el MAE no penaliza tanto los errores grandes, por lo que puede ser más robusto frente a valores atípicos.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (4.39)$$

3. Coeficiente de Determinación (R^2)

El Coeficiente de Determinación (R^2) es una métrica que indica la proporción de la varianza en los datos observados que es explicada por el modelo. Un R^2 cercano a 1 indica que el modelo explica bien la variabilidad de los datos, mientras que un R^2 cercano a 0 sugiere que el modelo no captura adecuadamente la información contenida en los datos.

$$R^2 = 1 - \frac{MSE}{\text{Varianza total de los datos}} = 1 - \frac{MSE}{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2} \quad (4.40)$$

4.5.10. Valores SHAP

Los valores Shapley son una herramienta basada en la teoría de juegos utilizada en para explicar la contribución de cada característica a la predicción de un modelo.

■ Predicción en un Modelo Lineal:

Suponiendo un modelo lineal de la siguiente manera:

$$\hat{f}(x) = \beta_0 + \beta_1 x_1 + \dots + \beta_m x_m, \quad (4.41)$$

donde x viene siendo la instancia a la cual se le calculará las contribuciones, cada x_j corresponde a cada característica, con $j = 1, \dots, p$, y cada β_j el peso correspondiente a cada atributo j .

La contribución ϕ_j de la función j -ésima en la predicción $\hat{f}(x)$ es la siguiente:

$$\phi_j(\hat{f}) = \beta_j x_j - E(\beta_j X_j) = \beta_j x_j - \beta_j E(X_j), \quad (4.42)$$

donde $E(\beta_j X_j)$ corresponde a la estimación del efecto medio para la característica j . La contribución es la diferencia entre el efecto de la característica menos el efecto promedio.

Al sumar todas las contribuciones de características para una instancia específica se obtiene lo siguiente:

$$\sum_{j=1}^m \phi_j(\hat{f}) = \sum_{j=1}^m (\beta_j x_j - E(\beta_j X_j)) \quad (4.43)$$

$$= (\beta_0 + \sum_{j=1}^m \beta_j x_j) - (\beta_0 + \sum_{j=1}^m E(\beta_j x_j)) \quad (4.44)$$

$$= \hat{f}(x) - E(\hat{f}(x)) \quad (4.45)$$

Esto representa el valor predicho para el punto de datos x menos el valor promedio predicho.

Para modelos más complejos, como los no lineales, no tenemos pesos β_j fáciles de interpretar. Aquí es donde el Valor de Shapley entra en juego. El Valor de Shapley para una característica j se define como su contribución a la predicción ponderada y sumada sobre todas las combinaciones posibles de características:

■ **Valor de Shapley:**

El valor de Shapley $\phi_i(v)$ se calcula como:

$$\phi_i(v) = \sum_{S \subseteq M \setminus \{i\}} \frac{|S|! \cdot (|M| - |S| - 1)!}{|M|!} (v(S \cup \{i\}) - v(S)), \quad (4.46)$$

donde v es la función característica que asigna un valor a cada coalición de características, v es la función característica que asigna un valor a cada coalición de características y S es una coalición de características que no incluye a la característica i .

La contribución de una característica i se calcula observando cómo cambia la predicción del modelo cuando se agrega la característica i a la coalición S .

$$(v(S \cup \{i\}) - v(S)) \quad (4.47)$$

Este término mide la diferencia en la predicción del modelo cuando la característica i está incluida en comparación con cuando no lo está tal como se muestra en la Figura 4.10. Básicamente, esta diferencia representa la contribución marginal de la característica i a la coalición S .

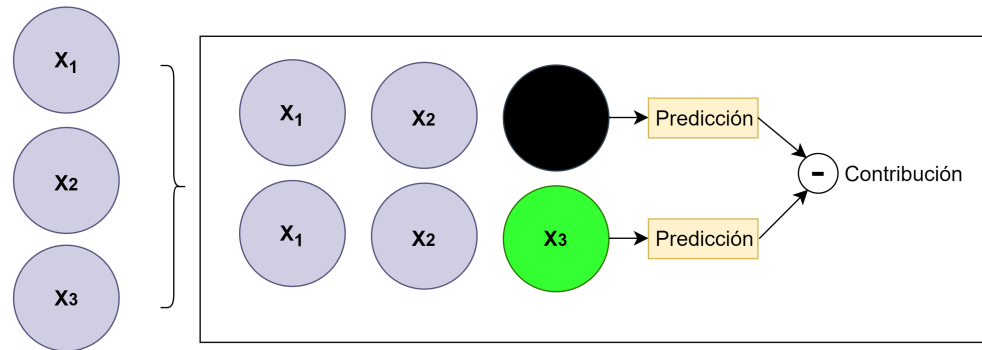
Características

Figura 4.10: Análisis de contribuciones, la contribución de la característica X_3 para una instancia determinada se calcula como la diferencia entre la predicción cuando la característica se encuentra presente en la coalición y cuando esta característica no está presente en la coalición.

De esta manera se calculan las contribuciones de la características i para todas las posibles combinaciones de la coalición, como se muestra en la Figura 4.11.

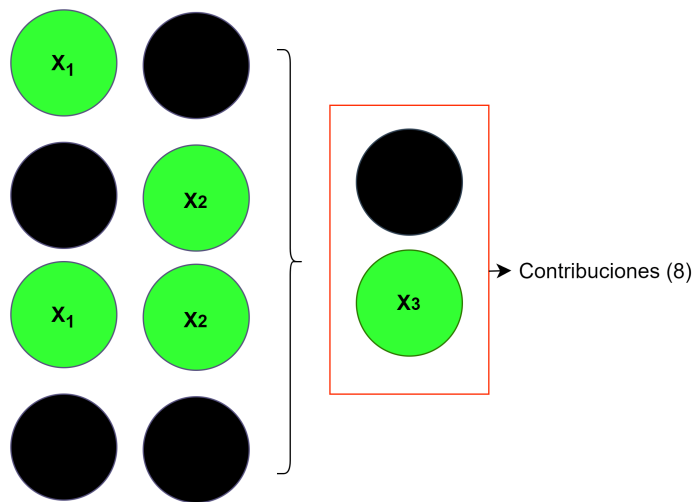


Figura 4.11: Las contribuciones totales de X_3 se calculan analizando todas las posibles combinaciones de la instancia de coalición.

METODOLOGÍA

LOS MÉTODOS para el análisis de EEG basados en la extracción de características de las señales cerebrales han sido eficaces en tareas de clasificación y regresión del deterioro cerebral[20][19][13]. Estos métodos cuentan principalmente con una fase de i) preprocesamiento, la cual se encarga de limpiar y normalizar las señales para mejorar la calidad de los datos; ii) extracción de características, la cual identifica patrones relevantes en las señales, estas características pueden ser de carácter frecuencial periódico, frecuencial aperiódico o temporales; y iii) aprendizaje de máquina, el cual utiliza las características extraídas para entrenar modelos que puedan clasificar o predecir el estado del cerebro en base a la información contenida en las señales EEG, Figura 5.1. Estos métodos, al extraer manualmente las características de las señales permiten identificar qué características biológicas son las más influyentes en la caracterización del deterioro cerebral, pudiendo otorgar mejor interpretabilidad de los datos.

5.1. Base de Datos

Se cuenta con una base de datos que involucra a 1337 participante de estado saludable de un estudio realizado por el consorcio de EuroLaD-EEG [1] que abarca regiones de América Latina (BrainLat) [21] y Europa.

5.2. Preprocesamiento

Para el trabajo de los datos de EEG se utilizará el flujo de trabajo propuesto por [31] que presenta una serie de pasos a seguir para el preprocesamiento de los datos que en primera instancia incluye filtrado, resampleo y remoción de artefactos.

En la parte de preprocesamiento se realiza un filtrado de 0.5 a 40 Hz y un remuestreo de 512 Hz, los datos se referencian utilizando la referencia promedio

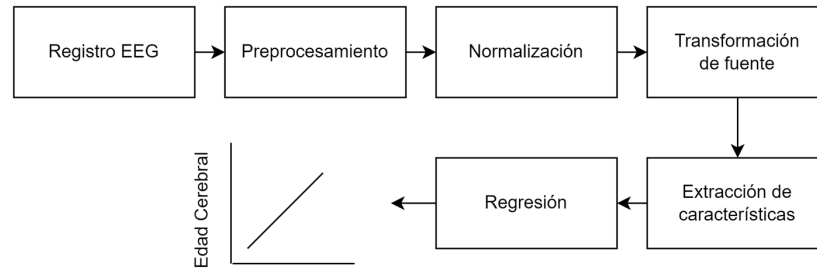


Figura 5.1: Metodología para estimación de la edad cerebral, a partir de los registros EEG primero se preprocesan los datos para eliminar ruido y artefactos indeseados, luego se normalizan los datos para obtener una estandarización y que todos tengan la misma estructura, se realiza una transformación de fuente para pasar de las señales EEG obtenida de cada electrodo a señales EEG focalizadas por regiones cerebrales, luego se extraen diversas características de estas señales por cada región cerebral para finalmente realizar una regresión mediante modelos de aprendizaje de máquina para lograr predecir la edad cerebral.

utilizando REST (reference electrode standardization technique). La remoción de artefactos se realiza utilizando ICA.

5.3. Preprocesamiento Adicional para Extracción de Características

Antes de realizar la extracción de características se realiza un preprocesamiento y modelado adicional, siguiendo el flujo de trabajo propuesto por [31] [12] se realizará una normalización de los datos y una transformación de fuente, en este caso se utiliza sLORETA [22], utilizando solamente una región cerebral que corresponde a un conjunto de los vóxel generados por el algoritmo sLORETA, de manera de evaluar el rendimiento que puede entregar.

5.4. Extracción de Características

Se utilizarán características de la señal de tipo frecuencial periódico [23], las cuales corresponden a características espectrales en la banda de frecuencia alpha (α), donde los límites de la banda se definen de la siguiente manera[24].

1. alfa canónica: 8.5-12 Hz, corresponde a la banda alpha estandarizada, definida de igual manera para todos los individuos
2. alfa inferior canónica: 8.5-10 Hz
3. alfa superior canónica: 10.5-12 Hz
4. alfa individual: definido como la frecuencia en la que se alcanza la máxima amplitud de la banda de frecuencia alpha de manera específica para cada individuo, de igual manera se define la transición de frecuencia theta/alfa (TF),

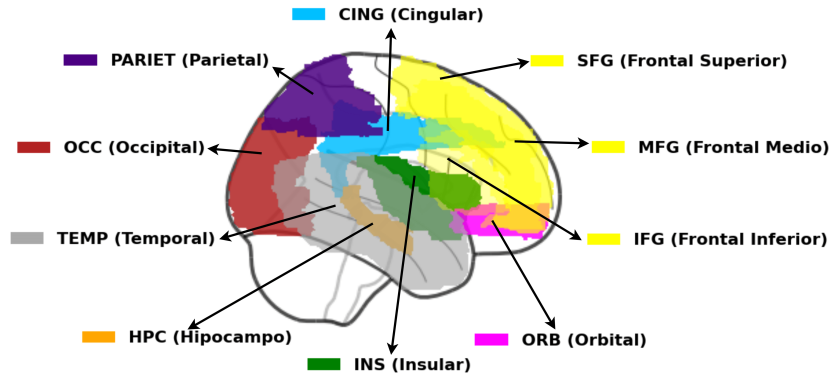


Figura 5.2: Regiones cerebrales resultantes de la transformación de fuente utilizadas para el análisis de estimación de la edad cerebral.

definida como la potencia mínima en el rango de frecuencia theta calculada en la segunda mitad del rango de frecuencia theta.

5. alfa individual inferior: α_{baja} (TF a IAF)
6. alfa individual superior: α_{alta} (IAF a IAF+2)

Estas características espectrales son calculadas por región cerebral, se utilizarán características provenientes de las regiones parietal, occipital, temporal, hipocampo, insular, orbital, frontal inferior, frontal medio, frontal superior y cingular.

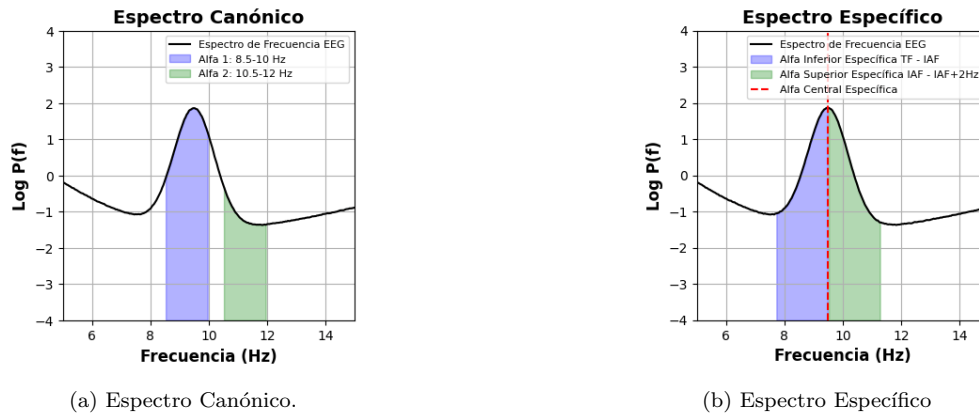


Figura 5.3: (a) Espectro de frecuencia canónico, se define la banda alfa entre 8.5 y 12 Hz, alfa 1 se define desde 8.5 a 10 Hz y alfa 2 desde 10.5 a 12 Hz. (b) Espectro de frecuencia individual, se define alfa inferior desde la transición entre la banda teta y alfa hasta el máximo valor de alfa individual y se define alfa superior desde alfa individual más 2 Hz.

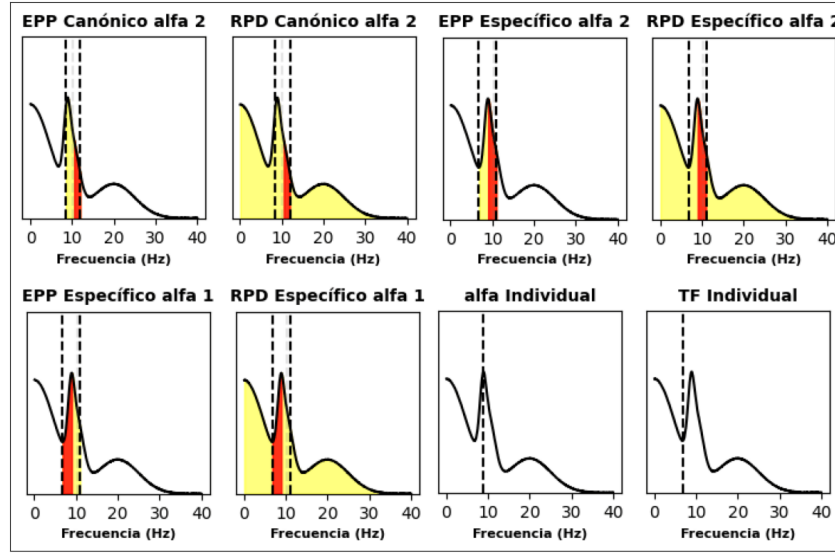


Figura 5.4: Extracción de características para registros EEG, se presentan características de tipo frecuencial en la banda alfa, EPP corresponde a la densidad de potencia normalizada con respecto a la propia banda y RPD la densidad de potencia normalizada con respecto al espectro completo, además se tienen características de valor de frecuencia donde ocurre el alfa individual y la transición de frecuencia de tetha a alfa.

5.5. Estimación de Edad Cerebral

Una vez teniendo las características extraídas desde los registros EEG la idea es poder predecir mediante estas características la edad de un sujeto específico y en su defecto en un conjunto de muestras. Suponiendo que se intenta predecir la edad mediante una sola característica, se podría generar un modelo de regresión lineal para ver la tendencia de esta característica con respecto a la edad de los sujetos como se muestra en la Figura 5.6.

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \forall i \quad (5.1)$$

Donde y_i corresponde a la edad sujeto i y x_i una característica específica de EEG para cada uno de los sujetos, de esta forma se ajustarían los parámetros β_0 y β_1 del modelo.

$$\mathbf{Y} = \begin{matrix} \text{edad} \\ \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} \end{matrix} \quad \mathbf{X} = \begin{matrix} \text{IAF_OCC_left} \\ \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} \end{matrix} \quad (5.2)$$

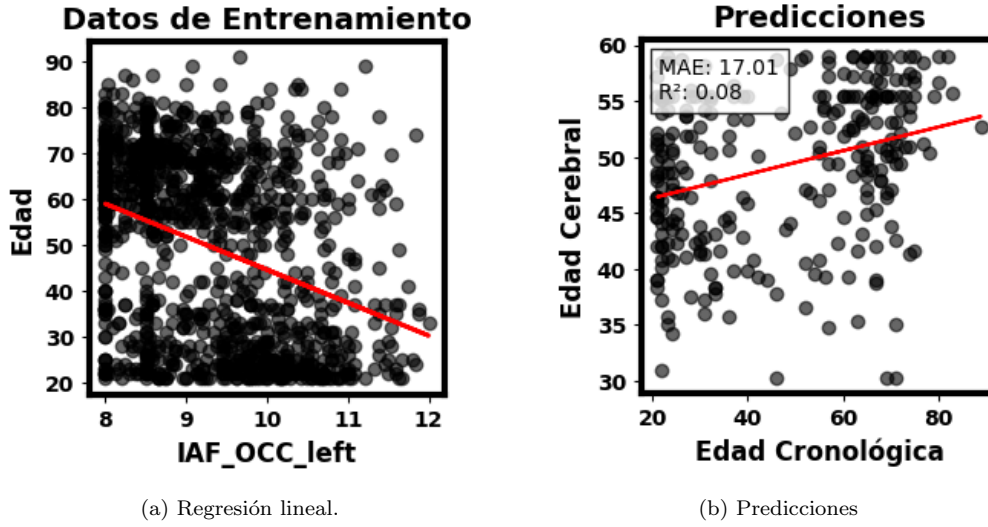


Figura 5.5: (a) Regresión lineal basada en IAF_OCC_left, se nota un decaimiento en la en la frecuencia de la banda alfa individual en función de un aumento en la edad. (b) Predicciones hechas en base al modelo de regresión lineal utilizando solo la característica IAF_OCC_left.

Ahora, si utilizamos 2 características se podría generar un hiperplano donde la combinación entre estas características podrían dar una mejor predicción que en un modelo lineal simple.

$$y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \epsilon_i, \quad (5.3)$$

$$\mathbf{Y} = \begin{pmatrix} \text{edad} \\ y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} \quad \mathbf{X} = \begin{pmatrix} \text{IAF_OCC_left} & \text{IAF_OCC_right} \\ x_{1,1} & x_{1,2} \\ x_{2,1} & x_{2,2} \\ \vdots & \vdots \\ x_{m,1} & x_{m,2} \end{pmatrix} \quad (5.4)$$

Para un modelo de regresión lineal simple se obtiene un MAE de 17.01 y un R^2 de 0.08, para un modelo de regresión lineal múltiple utilizando 2 características se obtiene un MAE de 16.83 y un R^2 de 0.12, finalmente para un modelo lineal múltiple utilizando 72 características se obtiene un MAE de 14.57 y un R^2 de 0.24. Estos resultados dan cuenta que no basta con un modelo lineal para realizar la estimación de la edad cerebral por lo que es necesario utilizar algoritmos de aprendizaje de máquina para encontrar patrones no lineales para así mejorar las predicciones en el modelo.

Se propone la siguiente estrategia de trabajo definida en la Figura 5.7, se divide el total de los datos en datos de prueba y datos de entrenamiento (Figura 5.8), donde

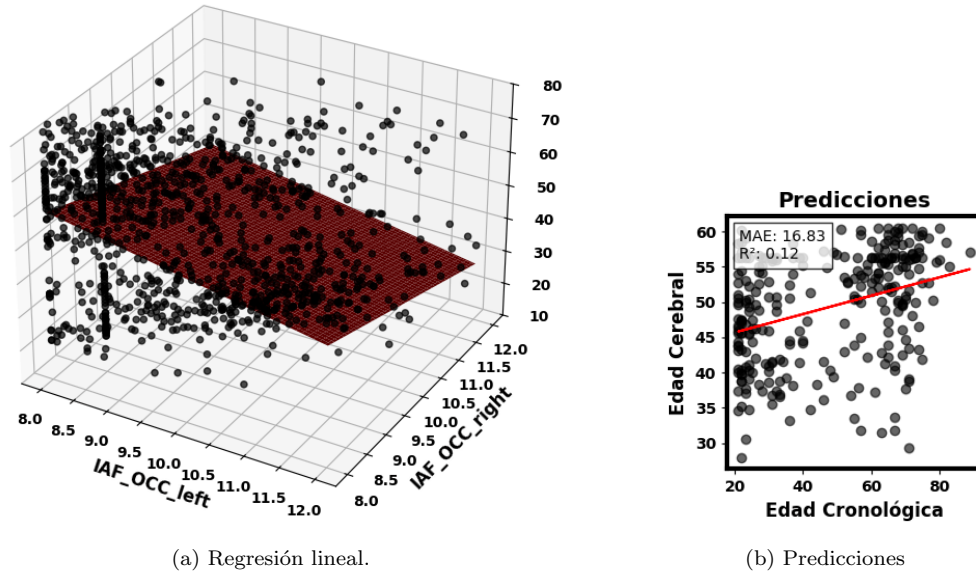
Modelo Ajustado a IAF_OCC_left y IAF_OCC_right

Figura 5.6: (a) Regresión lineal múltiple, se genera un hiperplano que modela la tendencia en la banda alfa para las características IAF_OCC_left e IAF_OCC_right. (b) Predicciones hechas en base al modelo de regresión lineal múltiple utilizando solo la característica IAF_OCC_left e IAF_OCC_right.

estos últimos son los que utilizarán para entrenar los modelos, y mediante técnicas de validación cruzada se utilizarán como datos de entrenamiento y validación. Posterior a eso se instancia un modelo de aprendizaje automático para regresión, para el cual se realizará una búsqueda intensiva de hiperparámetros para obtener los modelos más óptimos para nuestra instancia de datos de entrenamiento. Finalmente con los mejores hiperparámetros obtenidos para el modelo se entrena y evalúa el modelo final.

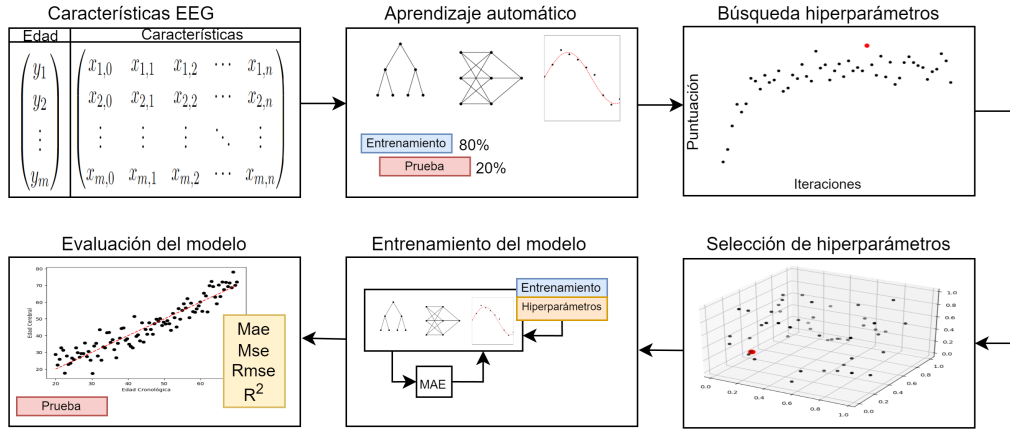


Figura 5.7: Entrenamiento del modelo

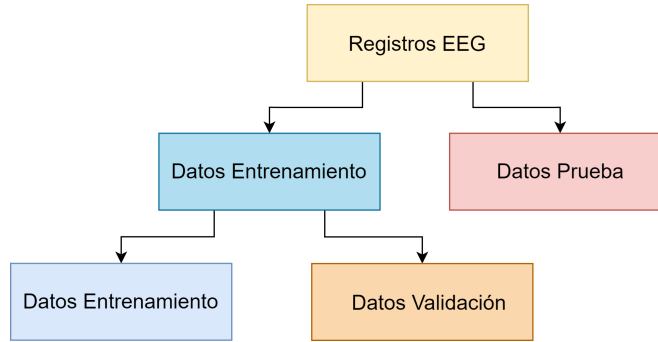


Figura 5.8: División de los datos en datos de entrenamiento y datos de prueba, los datos de entrenamiento se utilizan posteriormente como entrenamiento y validación mediante la validación cruzada, los datos de prueba se utilizan solo para evaluar el modelo.

5.5.1. Modelos de Aprendizaje de Máquina

Para realizar la estimación cerebral se utilizarán algoritmos de aprendizaje de máquina, los cuales permiten reconocer e identificar patrones en las características de EEG mediante el entrenamiento y aprendizaje automático, logrando así generar un modelo que nos permite dar interpretabilidad y caracterización a los datos de EEG. Los algoritmos a utilizar son Support Vector Regression, KNN, Random Forest, Xgboost, CatBoost y red neuronal multicapa descritos anteriormente en el capítulo 4. Para cada algoritmo, al igual que en los modelos de regresión lineal, el vector $\mathbf{Y}$ contiene las edades y_i de cada sujeto en el set de datos, y la matriz $\mathbf{X}$ contiene las características de EEG, donde cada sujeto posee un vector de características $x_{i,m} = (x_{1,1}, x_{1,2}, \dots, x_{1,m})$.

5.5.2. Búsqueda de Hiperparámetros

Para cada modelo se realiza una búsqueda intensiva de hiperparámetros utilizando la librería BayesSearchCV de Scikit-Optimize. Para el entrenamiento de cada modelo se utilizará la técnica de validación cruzada con 10 pliegues, es decir, para el conjunto de datos de entrenamiento, se divide en 10 conjuntos de datos aleatoriamente, y se realizan 10 entrenamientos distintos, utilizando de manera consecutiva, uno de esos conjuntos como prueba y los otros 9 conjuntos como entrenamiento. Posterior a eso, para cada conjunto de hiperparámetros escogidos por el algoritmo de Bayes se entrenan 10 modelos distintos y se promedia el error medio absoluto (MAE) entregado por cada modelo. Se itera el mismo procedimiento la cantidad suficiente de veces que permita encontrar el mejor conjunto de hiperparámetros como el conjunto que entrega el menor MAE promedio (Figura 5.9).

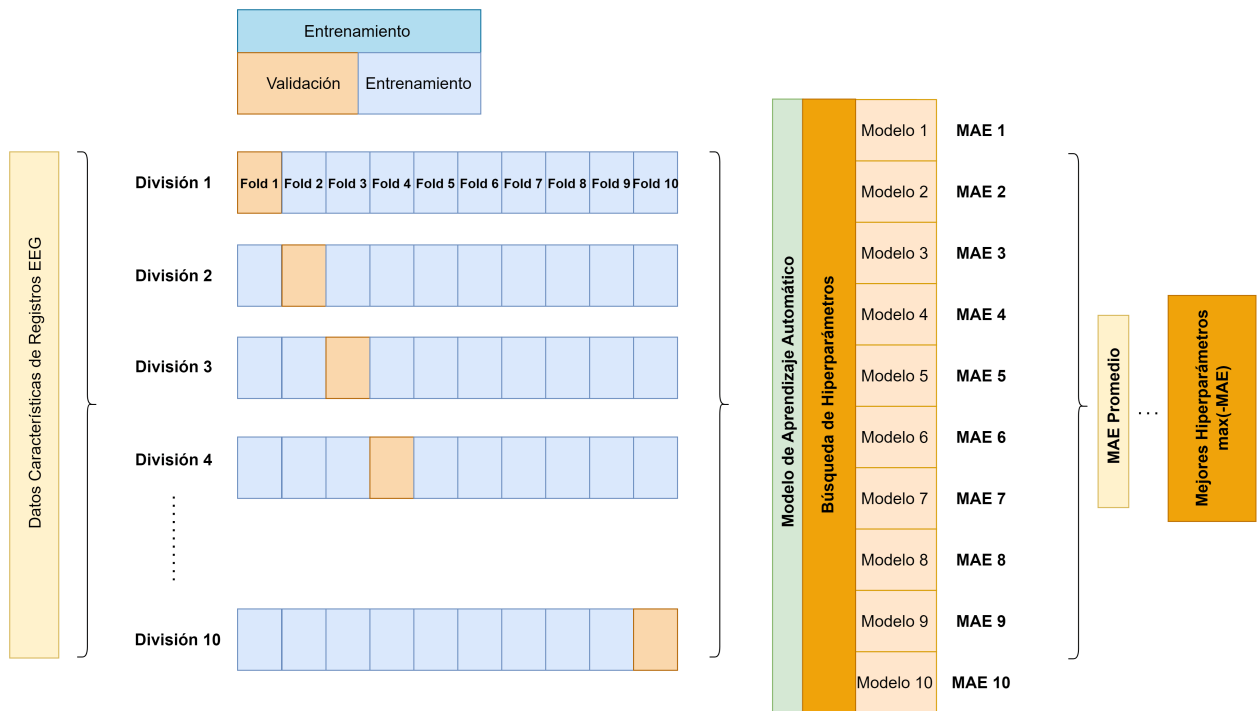


Figura 5.9: Búsqueda de Hiperparámetros utilizando validación cruzada de 10 pliegues mediante el algoritmo de Bayessearchcv, el modelo óptimo es aquel que consigue una puntuación promedio lo más alta posible, considerando que el algoritmo de Bayes lo que hace es maximizar la puntuación es necesario resolver el problema dual para lo cual se utiliza el error absoluto medio negativo.

5.5.3. Entrenamiento del Modelo

Una vez realizada la búsqueda intensiva de hiperparámetros, se entrenan 10 modelos distintos mediante la técnica de validación cruzada utilizando los mejores hiperparámetros, donde finalmente, se escoge el mejor de estos 10 modelos, es decir el que entregue un menor MAE en validación (Figura 5.10).



Figura 5.10: Entrenamiento del modelo utilizando validación cruzada y los mejores hiperparámetros obtenidos en la búsqueda bayesiana, de los 10 modelos obtenidos mediante validación cruzada se escoge el modelo del cual se obtenga un menor MAE

5.5.4. Evaluación del Modelo

El rendimiento de cada modelo será considerado mediante las distintas métricas de evaluación, estas son el MAE, MSE, RMSE y R^2 (Figura 5.11). La idea es obtener un error medio absoluto (MAE) lo más bajo posible considerando los datos de prueba definidos inicialmente.

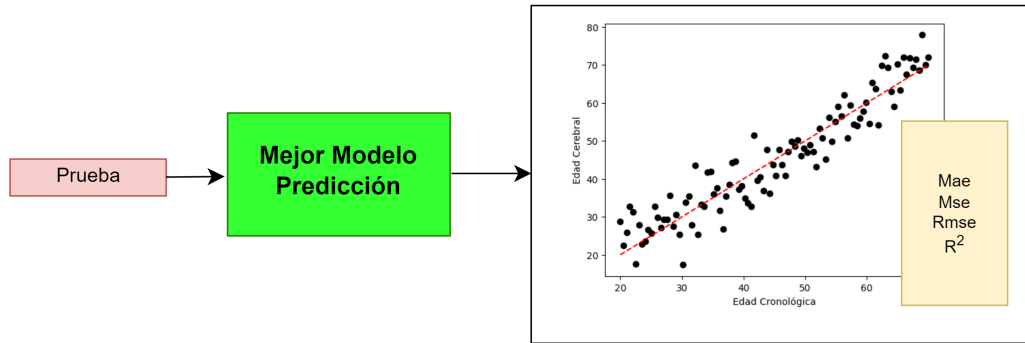


Figura 5.11: Evaluación del modelo utilizando los datos de prueba y el mejor modelo obtenido en la validación cruzada, se realiza la regresión de la edad cerebral utilizando el MAE, MSE, RMSE y R^2 como métricas de evaluación.

5.5.5. Importancia de Características Mediante Valores SHAP

Para estimar el valor de contribución de cada característica en los modelos de aprendizaje de máquina se utilizará el método de SHAP (explicaciones aditivas Shapley), el cual explica las contribuciones individuales de cada característica basado en los valores de Shapley. SHAP proporciona una forma práctica y eficiente de calcular los valores de Shapley ya que calcular estos valores directamente puede ser computacionalmente costoso porque implica considerar todas las posibles combinaciones de características, lo cual es exponencial en el número de características. SHAP implementa algoritmos eficientes para aproximar los valores de Shapley, haciendo uso de métodos específicos del modelo como TreeSHAP para modelos de árboles y KernelSHAP para modelos generales.

$$g(z') = \phi_0 + \sum_{j=1}^M \phi_j z'_j, \quad (5.5)$$

donde g es el modelo de explicación que descompone la predicción del modelo en sus componentes básicos, $z' \in 0,1^M$ corresponde al vector de coalición, M el tamaño máximo de la coalición y ϕ_j la contribución de cada característica. De esta manera es posible tomar un conjunto de instancias o muestras de datos, y calcular la contribución ϕ_j que cada característica aporta a la predicción de la edad de cada una de estas muestras, para finalmente sumar las contribuciones totales de cada característica (Figura 5.12).

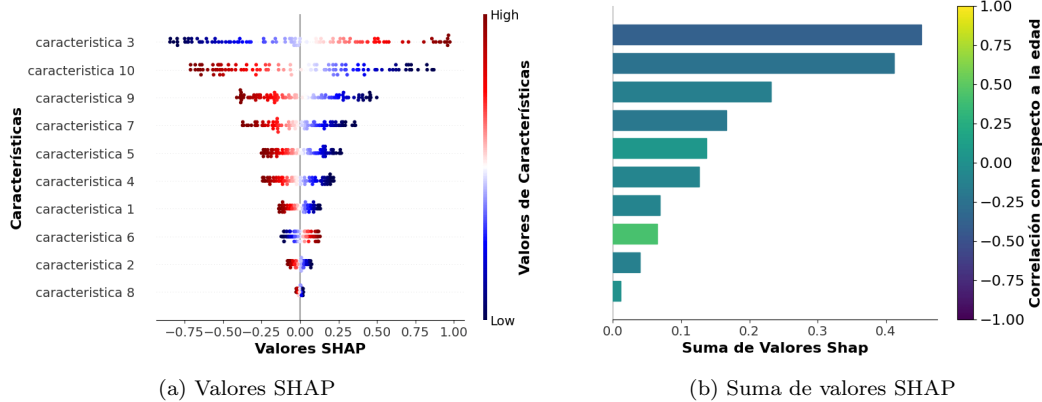


Figura 5.12: (a) Valores SHAP evaluados para cada instancia de datos, un valor negativo indica que al aumentar dicha característica contribuye negativamente al modelo y un valor positivo indica que al aumentar dicha característica contribuye positivamente al modelo. (b) Suma de valores SHAP y su correlación con respecto a la variable que se quiere predecir.

5.5.6. Incremento de Características

Para el algoritmo que de un mejor rendimiento se aplicará la misma metodología descrita anteriormente pero realizando un incremento de características de manera secuencial, basándose en el valor de contribución obtenido de los valores SHAP, de esta manera se obtendrá el mínimo conjunto de características necesarias para lograr el máximo rendimiento del modelo (Figura 5.13).

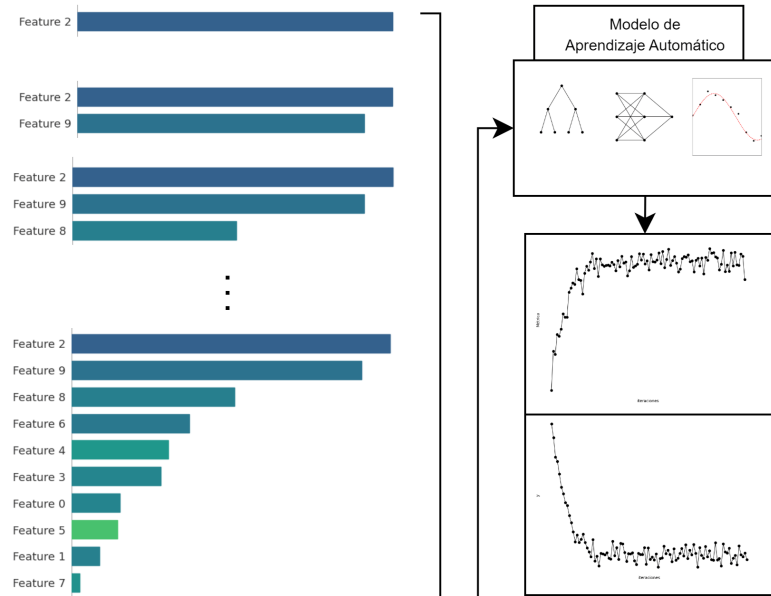


Figura 5.13: Incremento de características aplicadas al mejor modelo con el fin de obtener el mínimo de características necesarias para obtener un rendimiento máximo, se entrena el modelo primero con la característica que más contribuye en la explicación SHAP, luego con las 2 mejores características y así hasta completar las 72 características estudiadas.

IMPLEMENTACIÓN

La implementación del análisis de las características de EEG se realiza en el lenguaje de programación Python, implementando una clase principal **BaseRegressor** y 6 clases secundarias para cada algoritmo las cuales heredan a la clase **BaseRegressor**, estas son **SVRRegressor(BaseRegressor)**, **KNNRegressor(BaseRegressor)**, **RFRRegressor(BaseRegressor)**, **XGBoostRegressor(BaseRegressor)**, **CatBoostRegressor(BaseRegressor)**, **MLPRegressorCustom(BaseRegressor)**.

Toda la implementación se encuentra disponible en el siguiente repositorio: github.com/carrielfelipe/Memoria_Edad_Cerebral

6.1. Entorno Virtual

Se crea un entorno virtual de Python para empaquetar las librerías utilizadas, de tal manera que se tenga un control de las versiones de cada una de ellas para así tener compatibilidad entre los distintos equipos donde se ejecutan los modelos.

```
1 % Creacion de entorno virtual en python
2 python -m venv brain_age_env
3 brain_age_env/Scripts/activate
```

Figura 6.1: Creación de entorno virtual en Python que permite empaquetar las librerías utilizadas para mantener la compatibilidad entre las versiones.

6.2. Librerías

A continuación se presentan las librerías utilizadas y sus versiones:

```
1 catboost==1.2.5
2 matplotlib==3.8.4
3 nilearn==0.10.4
4 numpy==1.26.4
5 openpyxl==3.1.2
6 pandas==2.2.2
7 scikit-learn==1.4.2
8 scikit-optimize==0.10.1
9 scipy==1.13.0
10 seaborn==0.13.2
11 shap==0.45.0
12 xgboost==2.0.3
```

Figura 6.2: Librerías utilizadas para el proceso de estimación de edad cerebral

6.3. Análisis de Datos

Se presenta la distribución de las edades para los datos a utilizar, que comprende a 1377 personas sanas entre los 20 y 91 años de edad (Figura 6.3).

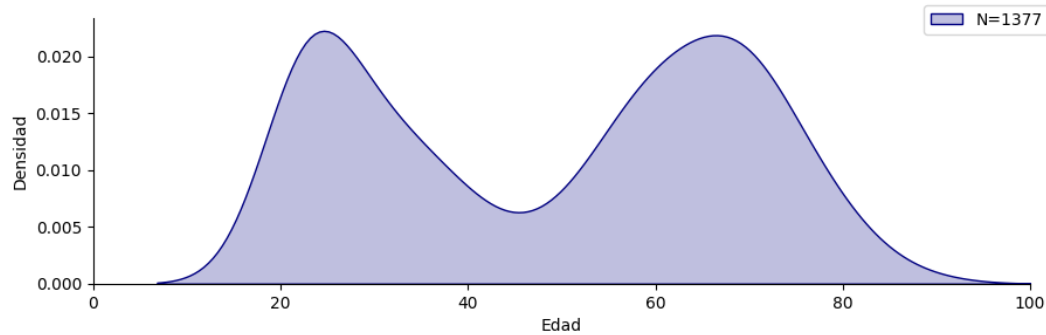


Figura 6.3: Distribución de densidad de edades para los datos a analizar, se aprecia la falta de datos alrededor de los 50 años.

Además se presentan las correlaciones correspondientes a cada una de las características con respecto a la edad, es posible notar que en general las correlaciones son negativas, tanto para métricas de potencia como para las bandas de frecuencia de alfa canónica, esto quiere decir que las densidades de potencia espectral y la frecuencia donde se encuentra el máximo valor de alfa disminuyen a medida que la edad aumenta. Por otra parte las frecuencias de transición theta alfa aumentan a medida que la edad aumenta, esto quiere decir que la banda alfa se hace más estrecha.

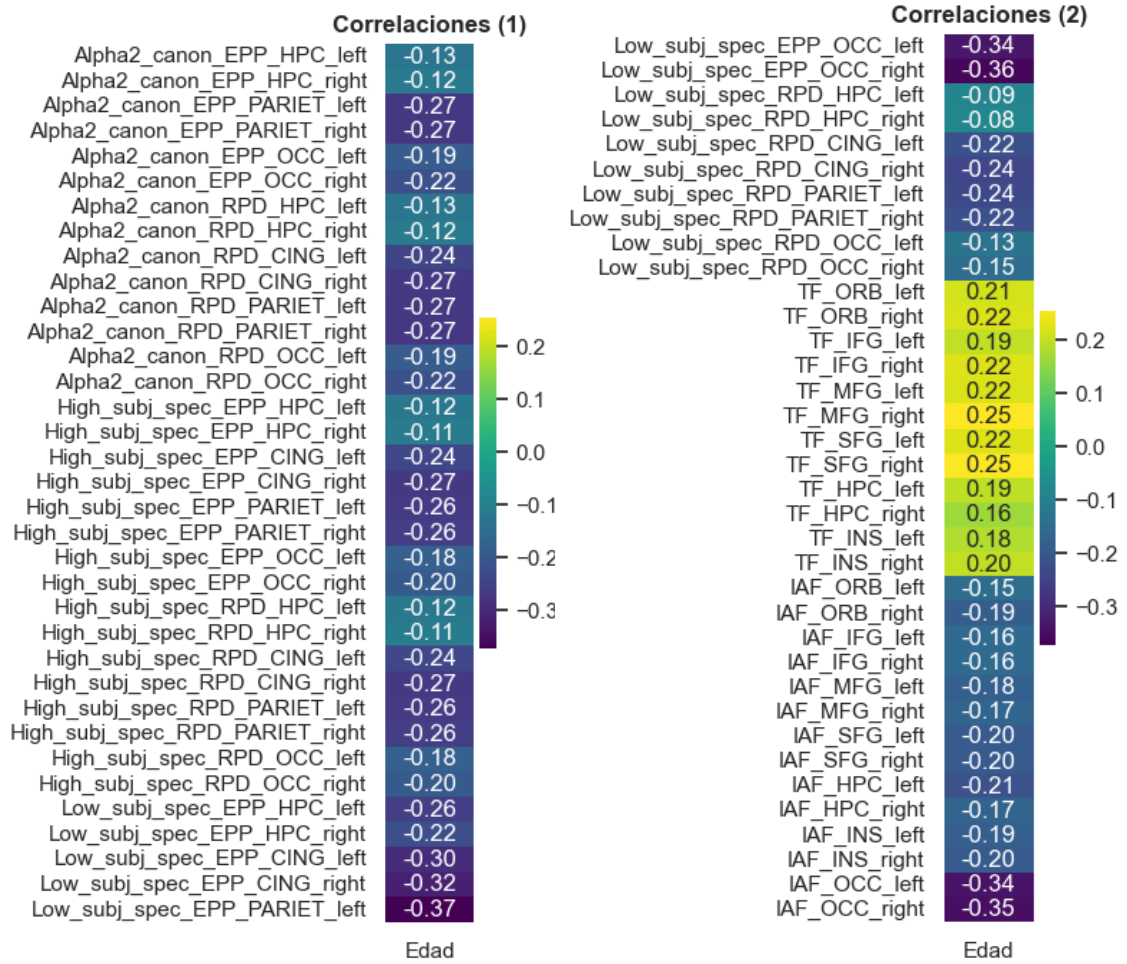


Figura 6.4: correlaciones de características con respecto a la edad, las características de potencia espectral y el valor de alfa canónica presentan correlaciones negativas, es decir, disminuyen cuando la edad aumenta, las frecuencias de transición entre alfa y theta presentan correlación positiva, es decir, aumentan cuando la edad aumenta.

Se realiza un escalamiento de las características para que todas las características se encuentren en una escala similar, para esto se utiliza MinMaxScaler y StandardScaler.

- MinMaxScaler transforma las características escalando cada una de ellas a un rango específico, típicamente entre 0 y 1.

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (6.1)$$

- StandardScaler transforma las características para que tengan una media de 0 y una desviación estándar de 1.

$$X' = \frac{X - \mu}{\sigma} \quad (6.2)$$

Para los algoritmos en general se utiliza MinMaxScaler, sin embargo para redes neuronales es necesario utilizar StandardScaler de lo contrario se obtienen pesos infinitos o nulos para la gran diferencia de escala en los datos.

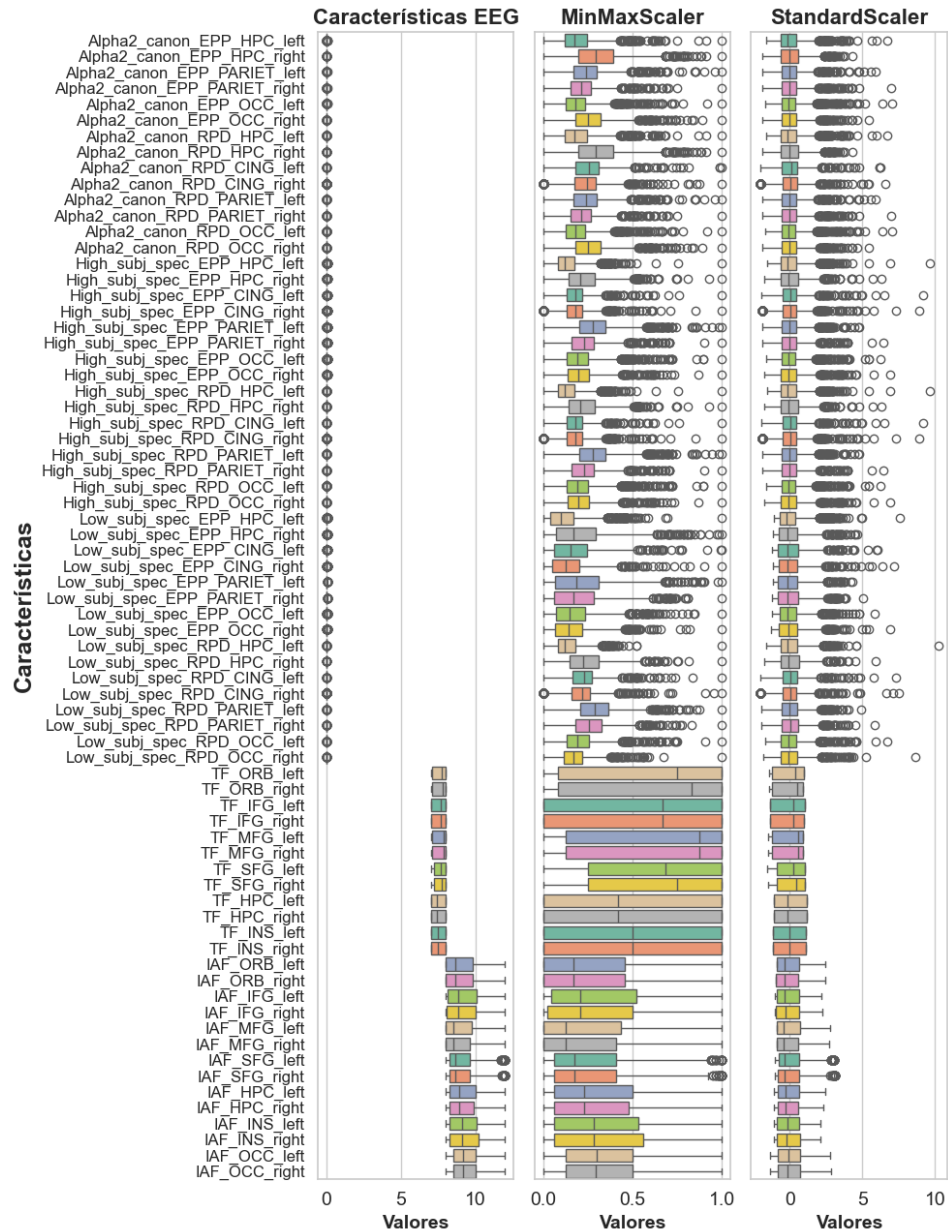


Figura 6.5: Distribución de las características EEG en su forma natural, escaladas con MinMaxScaler y escaladas con StandardScaler. Para el algoritmo de Red Neuroanl Multicapa se utiliza StandardScaler debido a que se necesita que las características tengan rangos de valores similares de lo contrario los perceptrones podrian generar pesos infinitos o nulos, para el resto de los algoritmos se utilizó MinMaxScaler.

6.4. Clase BaseRegressor

```
class BaseRegressor(scaler=None, params=None, params_space=None,  
fit_param_search=None, model_params_search=None, fit_params_train=None,  
model_params_train=None, name_model=None, save_path=None)
```

Es la clase principal y contiene los principales métodos de los distintos modelos.

6.4.1. Métodos

- **__init__:** Constructor de la clase BaseRegressor. Inicializa los parámetros y configuraciones básicas del modelo, incluyendo el escalador, los parámetros, los espacios de parámetros para la búsqueda de hiperparámetros, y otros parámetros relacionados con el entrenamiento y la predicción.
- **preprocess_data:** Escala las características utilizando MinMaxScaler. Devuelve un DataFrame con las características escaladas.
- **set_data:** Divide los datos en conjuntos de entrenamiento y prueba después de escalarlos.
- **search_best_model:** Realiza la búsqueda de hiperparámetros utilizando BayesSearchCV con validación cruzada. Devuelve el mejor modelo encontrado y sus mejores parámetros.
- **trainer:** Entrena un modelo de regresión utilizando validación cruzada. Devuelve los resultados del entrenamiento y el mejor modelo encontrado.
- **predicter:** Realiza predicciones utilizando el modelo entrenado en el conjunto de prueba.
- **regression_metrics:** Calcula las métricas de evaluación (MAE, MSE, RMSE y R^2) entre las predicciones y los valores verdaderos.
- **best_hyper:** Obtiene los mejores hiperparámetros para las mejores puntuaciones en validación cruzada.
- **plot_search_best_model:** Imprime los mejores parámetros encontrados y la evolución del error absoluto medio negativo. También ajusta una curva exponencial a los errores y resalta los 10 mejores resultados.
- **plot_regresion:** Grafica los resultados de la regresión, incluyendo las métricas de rendimiento para el conjunto de prueba.
- **plot_metricas_evaluacion:** Grafica las métricas de evaluación (MAE, MSE, RMSE y R^2) para cada conjunto (entrenamiento, validación y prueba).
- **feature_importance_shap:** Calcula y devuelve los valores SHAP y la importancia de las características utilizando shap.Explainer o shap.KernelExplainer.

- **plot_feature_importance:** Grafica la importancia de las características basada en los valores SHAP.

6.5. Clases de Algoritmos

Se genera para cada algoritmo de aprendizaje automático una clase que hereda a la clase **BaseRegressor**

6.5.1. Support Vector Regresion

```
class SVRegressor(BaseRegressor)
(save_path=None, scaler=None, params=None, params_space=None, fit_params_search=None,
model_params_search=None, fit_params_train=None, model_params_train=None,
name_model="SVR")
```

Se utiliza la librería SVR de Sickit Learn

Espacio de búsqueda:

- 'C': corresponde al parámetro de penalización de los errores. Un valor más alto de C intentará minimizar el error de entrenamiento, pero puede llevar a un modelo menos generalizable (overfitting). Un valor más bajo de C permitirá un margen más amplio, pero con más errores de entrenamiento, lo que puede mejorar la generalización. Se probarán valores entre (0.01, 100.0)
- 'gamma' : Es un parámetro que define la influencia de un solo punto de entrenamiento. En núcleos no lineales, como RBF, un valor alto de gamma hará que cada punto de soporte tenga un impacto más significativo y de menor alcance, lo que puede llevar a un sobreajuste (overfitting). Un valor bajo de gamma hará que los puntos de soporte tengan un impacto más amplio y suave, lo que puede llevar a un subajuste (underfitting). Se probarán valores entre (0.01, 10.0).
- 'epsilon' : Es un parámetro de insensibilidad que define un margen de insensibilidad dentro del cual los errores no son penalizados, se buscarán entre valores de (0.001, 1.0)
- 'kernel': Es una función que define la forma en que se transforman los datos de entrada a un espacio de características de mayor dimensión, se utilizarán las funciones lineal, radial y polinómica.

6.5.2. K Vecinos más Cercanos

```
class KNNRegressor(BaseRegressor):
(save_path=None, scaler=None, params=None, params_space=None, fit_params_search=None,
model_params_search=None, fit_params_train=None, model_params_train=None,
name_model="KNN"):
```

Espacio de búsqueda:

- `'n_neighbors'`: Número de vecinos más cercanos que se utilizarán para realizar la predicción. Se utilizarán valores de (1,50).
- `'weights'`: Este parámetro determina cómo se ponderan las contribuciones de los vecinos más cercanos en la predicción, estas contribuciones se pueden hacer de forma uniformes o que los vecinos mas cercanos contribuyan más.
- `'p'`: Determina la métrica de distancia que se utilizará. $p = 1$: Distancia de Manhattan, $p = 2$: Distancia Euclidiana, $p \geq 2$: Distancia de Minkowski, donde p representa el orden de la distancia.

6.5.3. Random Forest

```
class RFRegressor(BaseRegressor):
    (save_path=None, scaler=None, params=None, params_space=None, fit_params_search=None,
    model_params_search=None, fit_params_train=None, model_params_train=None,
    name_model=RF):
```

Espacio de búsqueda:

- `n_estimator`: especifica el número de árboles que se deben construir en el bosque aleatorio.
- `max_depth`: Es la profundidad máxima permitida para cada árbol en el bosque

6.5.4. XGBoost

```
class XGBoostRegressor(BaseRegressor):
    (save_path=None, scaler=None, params=None, params_space=None, fit_params_search=None,
    model_params_search=None, fit_params_train=None, model_params_train=None,
    name_model="XGBoost"):
```

Espacio de búsqueda:

- `n_estimator`: especifica el número de árboles que se deben construir en el modelo.
- `max_depth`: Es la profundidad máxima permitida para cada árbol del modelo
- `learning_rate`: Controla la tasa de aprendizaje del modelo. Es la cantidad con la que los pesos de los árboles se actualizan en cada paso del entrenamiento. Un valor más bajo generalmente conduce a un aprendizaje más lento pero más preciso.

6.5.5. CatBoost

```
class CatBoostRegressor(BaseRegressor):  
    (save_path=None, scaler=None, params=None, params_space=None, fit_params_search=None,  
model_params_search=None, fit_params_train=None, model_params_train=None,  
name_model= "CatBoost"):
```

Espacio de búsqueda:

- `n_estimator`: especifica el número de árboles que se deben construir en el modelo.
- `max_depth`: Es la profundidad máxima permitida para cada árbol del modelo
- `learning_rate`: Controla la tasa de aprendizaje del modelo. Es la cantidad con la que los pesos de los árboles se actualizan en cada paso del entrenamiento. Un valor más bajo generalmente conduce a un aprendizaje más lento pero más preciso.

6.5.6. Red Neuronal Multicapa

```
class MLPRegressorCustom(BaseRegressor):  
    (save_path=None, scaler=None, params=None, params_space=None, fit_params_search=None,  
model_params_search=None, fit_params_train=None, model_params_train=None,  
name_model= "Mlp"):
```

Espacio de búsqueda:

- `'layer1'`: Número de neuronas en la primera capa oculta. Define el tamaño de la primera capa oculta en la red neuronal.
- `'layer2'`: Número de neuronas en la segunda capa oculta. Define el tamaño de la segunda capa oculta en la red neuronal.
- `'num_hl'`: Número de capas ocultas. Especifica cuántas capas ocultas tendrá la red neuronal. Este parámetro puede ser usado para decidir dinámicamente la profundidad de la red.
- `'batch_size'`: Tamaño del lote para la optimización por mini-lotes. Especifica cuántas muestras se procesan antes de actualizar los parámetros del modelo. Un valor más alto puede conducir a una convergencia más estable, mientras que un valor más bajo puede resultar en una convergencia más rápida pero menos estable.
- `'activation'`: Función de activación para las capas ocultas. Puede ser `'identity'`, `'logistic'`, `'tanh'`, `'relu'`, etc. Define cómo las neuronas en las capas ocultas transforman la entrada. Por ejemplo, `'relu'` (Rectified Linear Unit) es una elección común debido a su buen rendimiento en redes profundas.

- 'solver': Algoritmo para la optimización. Puede ser 'lbfgs', 'sgd', 'adam', etc. 'lbfgs' es un optimizador basado en cuasi-Newton, 'sgd' es el gradiente descendente estocástico y 'adam' es un método de optimización que es robusto y eficiente en el ajuste de modelos complejos.

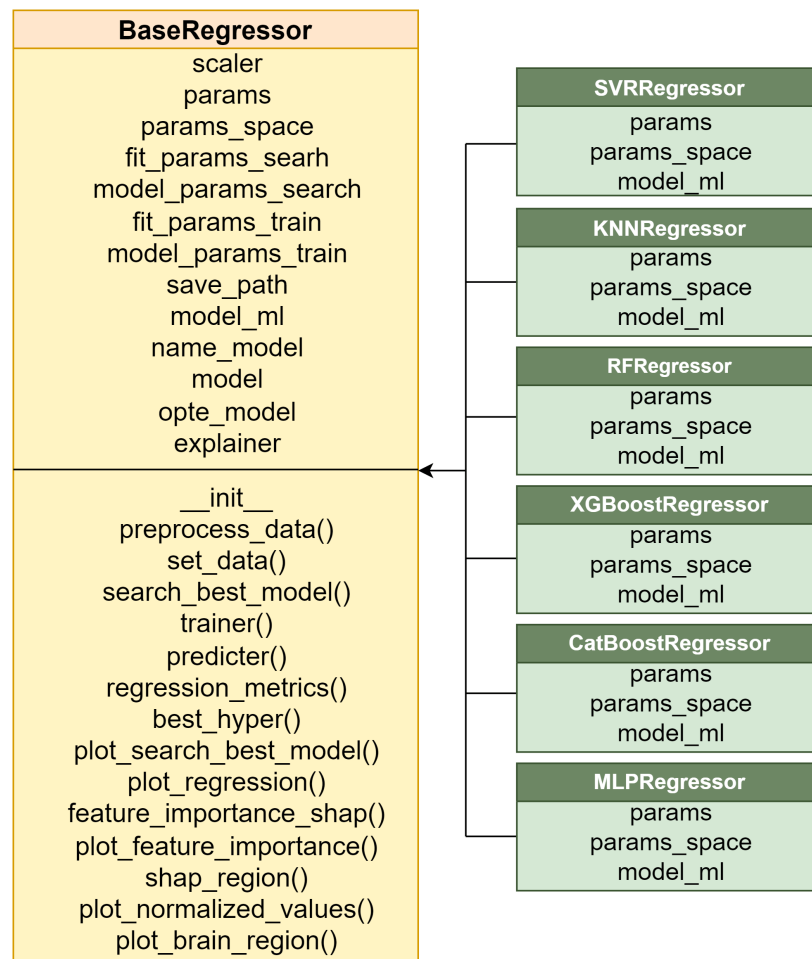


Figura 6.6: Diagrama de clases, se presenta una clase principal **BaseRegressor** que contiene los principales métodos del modelo, además se presentan las clases de cada algoritmo los cuales heredan de la clase principal **BaseRegressor**.

6.6. Implementación General

Se presenta a continuación la implementación general de los métodos de clase generados, que incluye en una (A) primera instancia la inicialización de los datos y del modelo a utilizar, (B) seguido de la búsqueda intensiva de hiperparámetros, *search_best_model()* realiza la búsqueda intensiva utilizando el algoritmo de *BayesSearchCV* de la biblioteca de *sklearn* identificando los 10 mejores rendimientos entregados en validación cruzada, (C) luego está el ajuste, entrenamiento y evaluación del modelo para el mejor conjunto de hiperparámetros encontrado anteriormente, el método *trainer()* se encarga de entrenar el modelo utilizando validación cruzada y quedándose con el modelo que entregue un menor MAE en los datos de validación, se grafican las predicciones y el ajuste de la recta para visualizar el desvío que tiene con respecto a la recta ideal. (D) Para el mejor modelo, el método *feature_importance_shap()* obtiene los valores SHAP para la instancia de datos de prueba con el fin de darle interpretabilidad al modelo mediante la importancia de características, las que posteriormente en (E) son agrupadas por region cerebral y normalizadas entre 0 y 1, estas regiones son pintadas en (F) para dar una interpretabilidad final de qué regiones son las más influyentes en la predicción de edad cerebral.

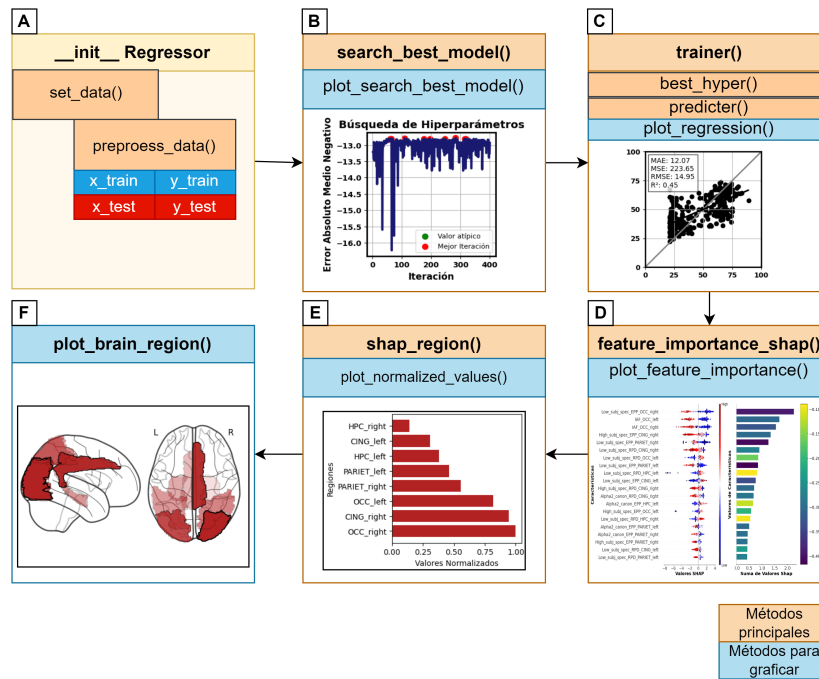


Figura 6.7: Implementación del flujo de trabajo, se presentan los metodos de la clase BaseRegressor y el flujo de trabajo general que permite obtener la estimación de edad cerebral a partir de características extraídas de EEG.

RESULTADOS

Se presentan a continuación los resultados obtenidos para cada modelo de aprendizaje de máquina, los mejores resultados los obtienen los modelos basados en árboles de decisión, en especial el algoritmo de Xgboost con un $R^2 = 0,45$ y un $MAE = 12,07$, seguido de los modelos de árboles se encuentra el modelo basado en red neuronal multicapa con un $R^2 = 0,35$ y un $MAE = 12,80$, modelos clásicos de aprendizaje de máquina como las máquinas de vectores y los vecinos más cercanos fueron los peor evaluados, SVR con un $R^2 = 0,26$ y un $MAE = 17,26$ y KNN con un $R^2 = 0,26$ y un $MAE = 17,26$.

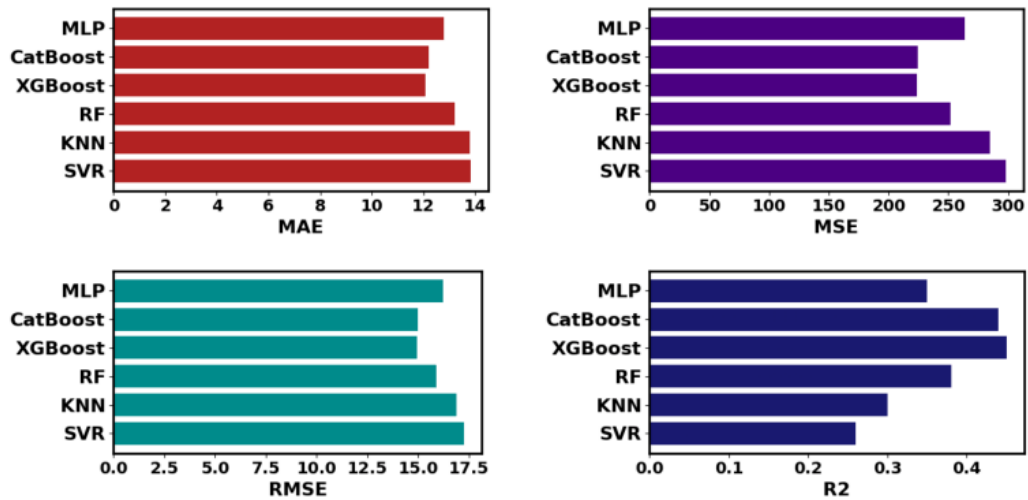


Figura 7.1: Comparación de las métricas de evaluación para el mejor modelo en cada uno de los algoritmos, el mejor rendimiento, basandose en obtener un menor error absoluto medio (MAE) y un mayor R^2 lo obtiene XGBoost y CatBoost respectivamente.

Se presentan además los informes para cada algoritmo, que incluye la búsqueda intensiva de hiperparámetros, la evaluación del mejor modelo, el rendimiento de cada pliegue en la validación cruzada, la importancia de características mediante valores SHAP y la importancia por región cerebral.

La búsqueda intensiva de hiperparámetros se realizó con 400 iteraciones para cada algoritmo identificando las 10 mejores iteraciones obtenidas en validación cruzada, estos máximos fueron marcados con color rojo, además los valores atípicos, como el obtenido en la red neuronal fueron reemplazados por el menor valor obtenido y marcados con color verde. Se considera el Error Absoluto Medio Negativo como puntaje de evaluación, esto es debido a que se está buscando obtener un Error Medio Absoluto (MAE) lo más bajo posible, sin embargo el algoritmo de bayes está diseñado para maximizar el puntaje obtenido, por lo tanto se maximiza el Error Absoluto Medio Negativo.

Para cada algoritmo se utilizaron los hiperparámetros obtenidos en la mejor iteración evaluando el rendimiento de cada uno de los pliegues mediante las métricas de evaluación de MAE, MSE, RMSE y R^2 . Se utilizó el modelo que de un menor MAE entre los 10 pliegues generados en validación cruzada. Una vez teniendo el mejor modelo se realizaron las predicciones en los datos de prueba obteniendo la edad cerebral para cada una de las muestras, se graficó la relación entre la edad cronológica y la edad cerebral para cada una de las muestras y se realizó una regresión lineal a estos datos para obtener una recta ajustada y comparar la desviación con la recta de edad ideal.

Una vez teniendo el mejor modelo entrenado se calcularon los valores explicativos SHAP para los datos de prueba, así obtener la contribución que cada característica generó en la predicción de edad cerebral, generando un orden por importancia de características. En el gráfico de SHAP el color rojo implica que dicha característica se encuentra sobre el promedio de los datos y el color azul que dicha característica se encuentra bajo el promedio de los datos. Un valor negativo implica que la característica contribuye negativamente y un valor positivo que esta contribuye positivamente, esto quiere decir que si la variable presenta contribuciones negativas al disminuir dicha característica su valor la edad cerebral aumentaría. Se sumaron los valores absolutos SHAP para cada característica generando un gráfico de barras con valores acumulados de las contribuciones en cada instancia de datos, obteniendo además la correlación que tiene esa característica con respecto a la edad.

Posterior a la obtención de la contribución de cada característica en el modelo se tomaron las 20 mejores características de cada uno de estos y se sumaron la totalidad de los valores absolutos de SHAP por región cerebral, los cuales son posteriormente normalizados entre 0 y 1, identificando así cuales fueron las regiones cerebrales que más influyeron en la predicción de la edad.

7.1. Support Vector Regression

Para el algoritmo de máquinas de vectores para regresión los mejores puntajes en la búsqueda intensiva de hiperparámetros se obtuvieron en las últimas iteraciones (Figura 7.2), denotando un estancamiento en la búsqueda puesto que ya ha realizado todas las combinaciones posibles y comienza a iterar con valores cercanos entre si, los mejores hiperparámetros encontrados fueron de $C = 100$, $epsilon = 1,0$, $gamma = 0,05465$ y un $kernel = rbf$, es decir se utilizó un valor de C lo más grande posible, una banda $epsilon$ tambien lo mas ancha posible se prefirió un $kernel$ de tipo radial en vez de lineal o polinómico.

Para el mejor modelo seleccionado se tiene el rendimiento obtenido para cada uno de los pliegues en validación cruzada, notando además que el modelo no logra ajustarse de buena manera a los datos de entrenamiento, obteniendo un valor de R^2 por debajo de 0,4 en cada uno de los pliegues, esto posiblemente ocurrido por características en el modelo que no logran ser significativas para la predicción. Se utiliza el modelo que de un menor MAE entre los 10 pliegues generados por la validación cruzada, en este caso correspondiente al quinto pliegue, obteniendo un $MAE = 13,82$ y un $R^2 = 0,26$.

Las características más influyentes en las predicciones del algoritmo de máquina de vectores de soporte fueron en primer y segundo lugar las de IAF_OCC_right e IAF_OCC_left respectivamente las cuales corresponden a características que representan la ubicación del peak de frecuencia alfa visto en la región occipital, posterior a eso le siguen características que representan la frecuencia específica en donde ocurre la transición de frecuencia entre la banda alfa y theta (TF) para las regiones frontal e insular.

Al sumar la totalidad de las contribuciones SHAP por regiones cerebrales se obtuvo que las principales regiones que contribuyeron en las predicciones fueron la región occipital izquierda y la región occipital derecha, seguidos de la región insular derecha y región frontal media.

7.1.1. Entrenamiento del Modelo

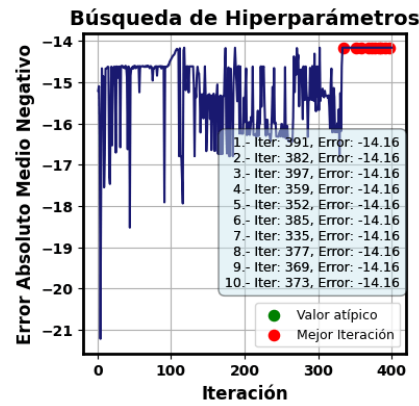
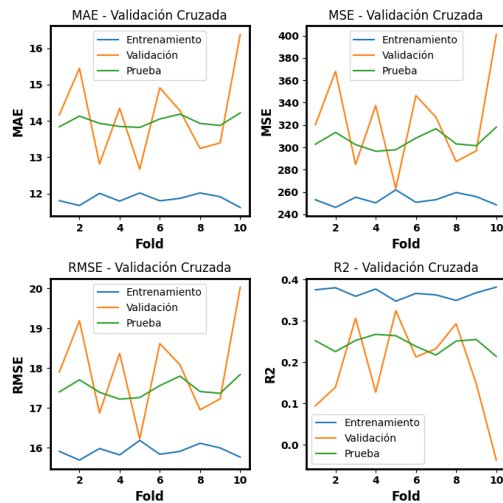
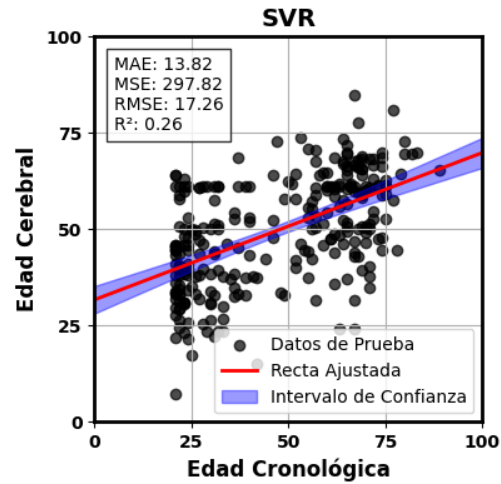


Figura 7.2: Búsqueda intensiva de hiperparámetros para SVR, el mejor conjunto de hiperparámetros se evalúa mediante el Error Absoluto Medio Negativo, se presentan los 10 mejores conjuntos de hiperparámetros obtenidos identificando la iteración en la cual ocurrió.



(a) Evaluación del modelo en validación cruzada para SVR utilizando el mejor conjunto de hiperparámetros encontrados en la búsqueda bayesiana.



(b) Predicciones de los datos de prueba para el mejor modelo, se gráfica la relación de la Edad Cronológica con la Edad Cerebral, se presentan las métricas de evaluación para mejor modelo, estas son el MAE, MSE, RMSE y R^2 . Se realiza un modelo de regresión lineal que relacione la Edad Cronológica con la Edad Cerebral de manera ver como esta línea de tendencia difiere con la recta de edad ideal.

7.1.2. Importancia de Características y Valores Shap

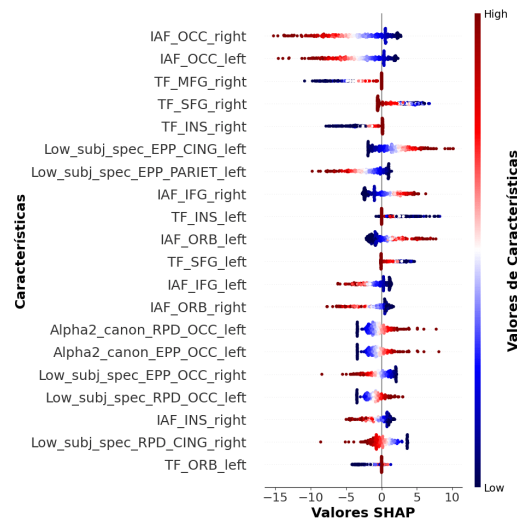


Figura 7.4: Valores SHAP calculados para las predicciones.

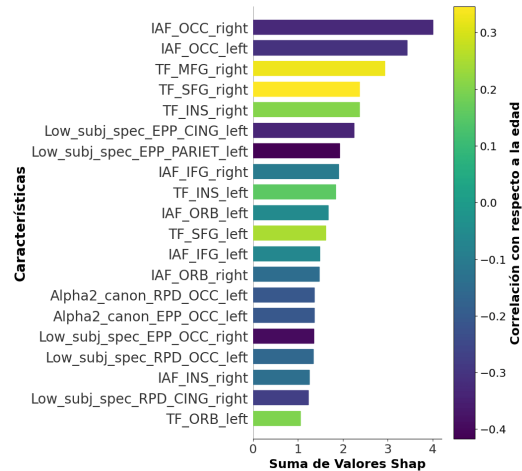


Figura 7.5: Suma de valores Shap y correlación de cada característica con la Edad.

7.1.3. Importancia por Región

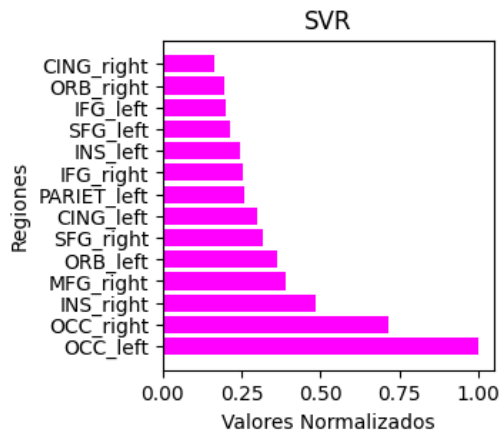


Figura 7.6: Suma de las 20 mejores SHAP por región cerebral, se normaliza entre 0 y 1 para analizar la importancia de cada región a la predicción.

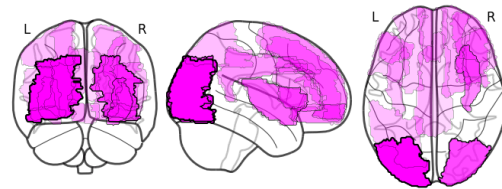


Figura 7.7: Regiones cerebrales más influyentes en la predicción de Estimación de Edad Cerebral para SVR, las regiones más importantes fueron la Occipital izquierda y derecha.

7.2. K Vecinos Más Cercanos

Para el algoritmo de los K Vecino Más Cercanos los mejores puntajes en la búsqueda intensiva de hiperparámetros se comenzaron a obtener a partir de la iteración 27 (Figura 7.2), denotando un rápido ajuste de hiperparámetros, al ser un algoritmo que no realiza un entrenamiento como tal puesto que las predicciones las basa en la totalidad de los datos de entrenamiento de manera directa identificando los vecinos más cercanos presenta un R^2 definido como 1 y un error cercano a 0 en entrenamiento para términos prácticos, los mejores hiperparámetros encontrados fueron de $n_{neighbors} = 14$, $p = 1$ y $weights = 'distance'$, es decir el número óptimo de vecinos fue de 14 para un máximo de 50, se prefirió una distancia de Manhattan ($p = 1$) y que los pesos a los vecinos fueran asignados dando prioridad a los vecinos más cercanos.

Para el mejor modelo seleccionado se tiene el rendimiento obtenido para cada uno de los pliegues en validación cruzada. Esto podría significar que el conjunto de datos que se utilicen para realizar las predicciones puede ser significativo al momento de medir el rendimiento del modelo. Se utiliza el modelo que entregue un menor MAE entre los 10 pliegues generados por la validación cruzada, en este caso correspondiente al quinto pliegue, obteniendo un $MAE = 13,80$ y un $R^2 = 0,30$.

Las características más influyentes en las predicciones del algoritmo de los K vecinos más cercanos fueron en primer y segundo lugar las de IAF_OCC_right e IAF_OCC_left respectivamente las cuales corresponden a características que representan la ubicación del peak de frecuencia alfa visto en la región occipital, posterior a eso le siguen características que representan la frecuencia en donde ocurre la transición de frecuencia entre la banda alfa y theta (TF) en la region frontal media (MFG), posterior a esto le siguen características de alfa individual en diversas regiones del hemisferio izquierdo del cerebro.

Al sumar la totalidad de las contribuciones SHAP por regiones cerebrales se obtuvo que las principales regiones que contribuyeron en las predicciones fueron la región frontal media izquierda (MFG_left) y la región occipital derecha (OCC_right), seguidos de la region insular izquierda (INS_left) e hipocampo medio izquierdo (HPC_left).

7.2.1. Entrenamiento del Modelo

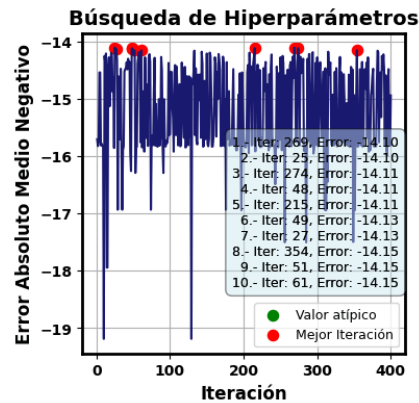
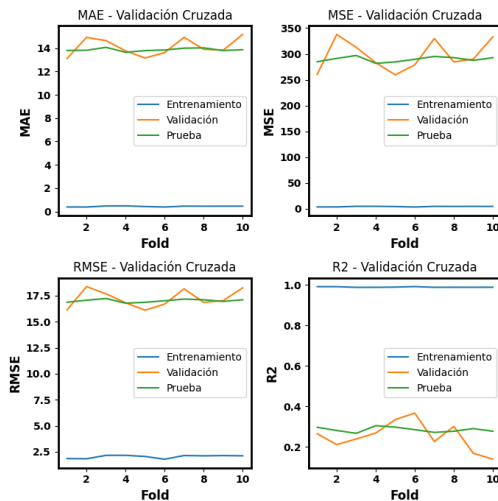
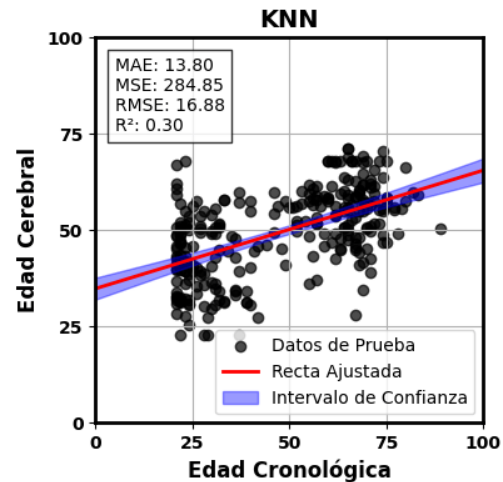


Figura 7.8: Búsqueda intensiva de hiperparámetros para KNN, el mejor conjunto de hiperparámetros se evalúa mediante el Error Absoluto Medio Negativo, se presentan los 10 mejores conjuntos de hiperparámetros obtenidos identificando la iteración en la cual ocurrió.



(a) Evaluación del modelo en validación cruzada para KNN utilizando el mejor conjunto de hiperparámetros encontrados en la búsqueda bayesiana.



(b) Predicciones de los datos de prueba para el mejor modelo, se gráfica la relación de la Edad Cronológica con la Edad Cerebral, se presentan las métricas de evaluación para mejor modelo, estas son el MAE, MSE, RMSE y R^2 . Se realiza un modelo de regresión lineal que relacione la Edad Cronológica con la Edad Cerebral de manera ver como esta línea de tendencia difiere con la recta de edad ideal.

7.2.2. Importancia de Características y valores Shap

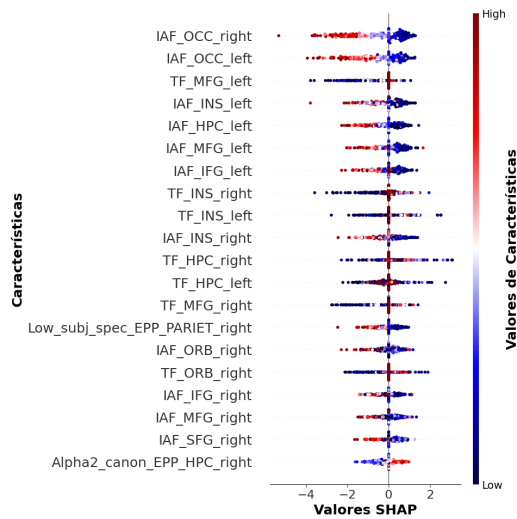


Figura 7.10: Valores SHAP calculados para las predicciones.

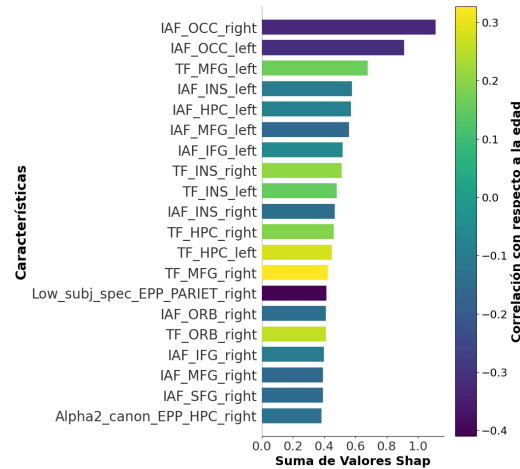


Figura 7.11: Suma de valores Shap y correlación de cada caracterísitca con la Edad.

7.2.3. Importancia por región

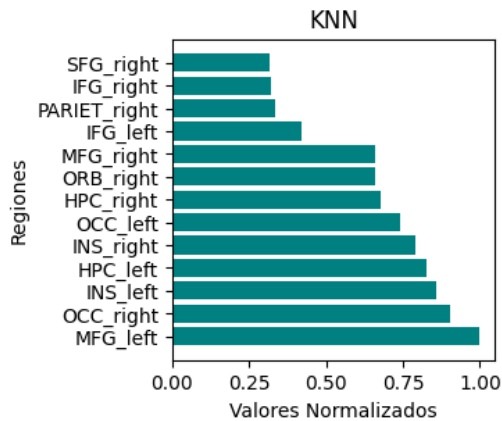


Figura 7.12: Suma de las 20 mejores SHAP por región cerebral, se normaliza entre 0 y 1 para analizar la importancia de cada región a la predicción.

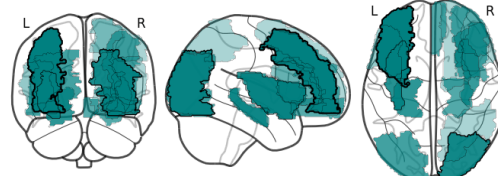


Figura 7.13: Regiones cerebrales mas influyentes en la predicción de Estimación de Edad Cerebral para KNN, las regiones más importantes fueron la Frontal Media izquierda y Occipital derecha.

7.3. Random Forest

Para el algoritmo de bosques aleatorios el mejor puntaje se obtuvo en la iteración 394 (Figura 7.14) y el segundo mejor puntaje en la iteración 62 y, lo cual implicaría que el modelo obtiene de manera temprana un conjunto óptimo de hiperparámetros generando una especie de estacionalidad creciente, los mejores hiperparámetros encontrados fueron de $max_depth = 46$, $n_estimators = 898$, es decir se utilizaron 898 árboles de decisión con una profundidad de 46.

Para el mejor modelo seleccionado se tiene el rendimiento obtenido para cada uno de los pliegues en validación cruzada, notando que el modelo logra ajustarse bastante bien a los datos de entrenamiento. Se utiliza el modelo que de un menor MAE entre los 10 pliegues generados por la validación cruzada, en este caso correspondiente al quinto pliegue, obteniendo un $MAE = 13,20$ y un $R^2 = 0,38$.

Las características más influyentes en las predicciones del algoritmo de bosques aleatorios fueron en primer lugar la potencia espectral normalizada en la banda alfa baja específica de cada sujeto (Low_subj_spec_EPP_OCC_right), segundo y tercer lugar las de IAF_OCC_left e IAF_OCC_right respectivamente las cuales corresponden a características que representan la ubicación del peak de frecuencia alfa visto en la región occipital, posterior a eso le siguen características que representan la densidad de potencia para las regiones cingular (CING) y parietal (PARIET).

Al sumar la totalidad de las contribuciones SHAP por regiones cerebrales se obtuvo que las principales regiones que contribuyeron en las predicciones fueron la región occipital derecha (OCC_right), seguidos de la región cingular derecha (CING_right) y región occipital izquierda (OCC_left).

7.3.1. Entrenamiento del Modelo

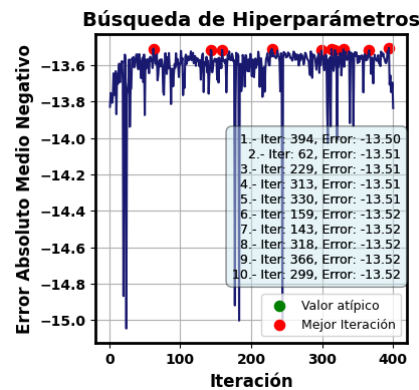
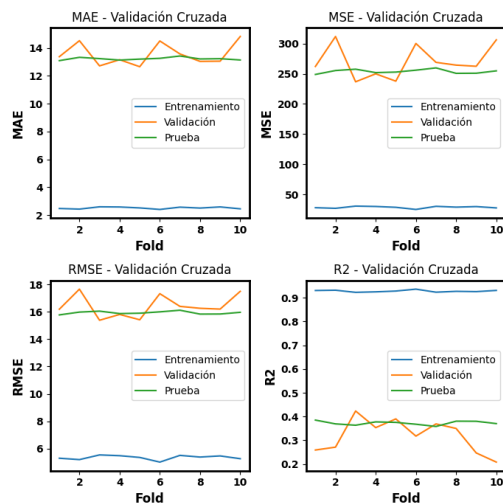
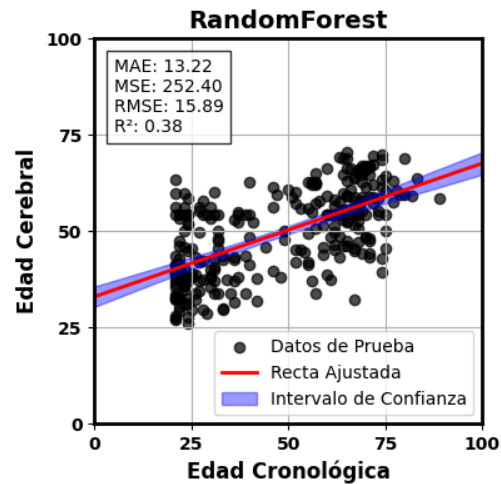


Figura 7.14: Búsqueda intensiva de hiperparámetros para Bosques Aleatorios, el mejor conjunto de hiperparámetros se evalúa mediante el Error Absoluto Medio Negativo, se presentan los 10 mejores conjuntos de hiperparámetros obtenidos identificando la iteración en la cual ocurrió.



(a) Evaluación del modelo en validación cruzada para Bosques Aleatorios utilizando el mejor conjunto de hiperparámetros encontrados en la búsqueda bayesiana.



(b) Predicciones de los datos de prueba para el mejor modelo, se gráfica la relación de la Edad Cronológica con la Edad Cerebral, se presentan las métricas de evaluación para mejor modelo, estas son el MAE, MSE, RMSE y R^2 . Se realiza un modelo de regresión lineal que relacione la Edad Cronológica con la Edad Cerebral de manera ver como esta línea de tendencia difiere con la recta de edad ideal.

7.3.2. Importancia de Características y Valores SHAP

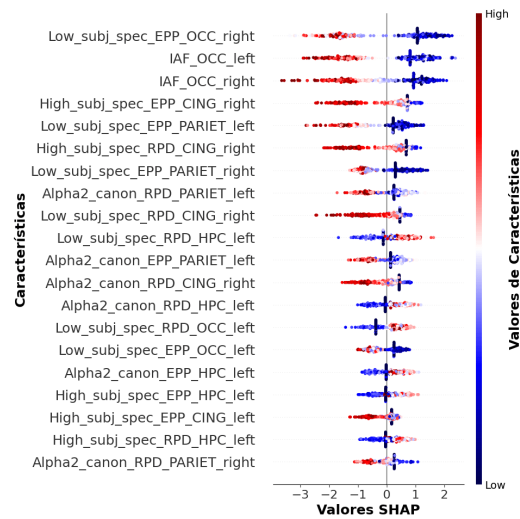


Figura 7.16: Valores SHAP calculados para las predicciones.

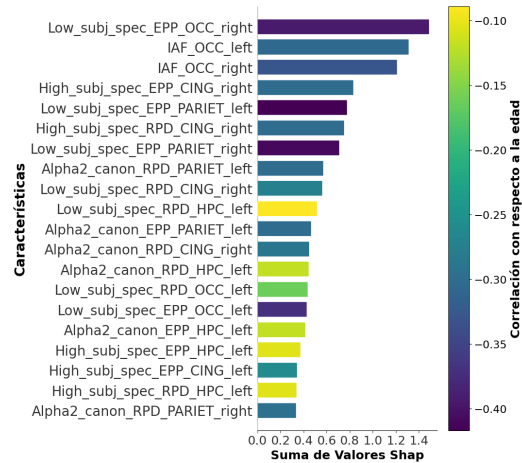


Figura 7.17: Suma de valores Shap y correlación de cada caracterísitca con la Edad.

7.3.3. Importancia por región

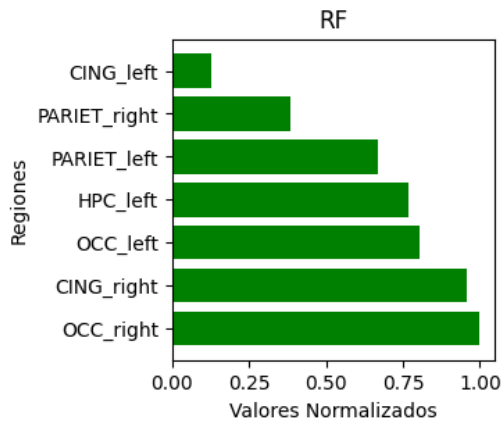


Figura 7.18: Suma de las 20 mejores SHAP por región cerebral, se normaliza entre 0 y 1 para analizar la importancia de cada región a la predicción.

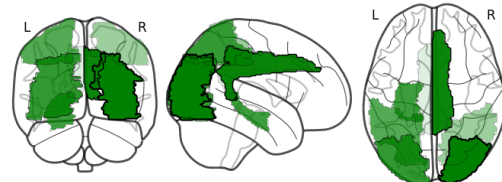


Figura 7.19: Regiones cerebrales mas influyentes en la predicción de Estimación de Edad Cerebral para Bosques Aleatorios, las regiones más importantes fueron la Occipital derecha y Cingular derecha.

7.4. XGBoost

Para el algoritmo de XGBoost el mejor puntaje se obtuvo en la iteración 283 (Figura 7.20) y el segundo mejor puntaje en la iteración 109, a diferencia del algoritmo de bosques aleatorios, el algoritmo de XGBoost demora un poco más en encontrar los mejores hiperparámetros, sin embargo considerando el poco margen de error entre las 10 iteraciones se podría considerar que a partir de las 59 iteraciones de búsqueda el algoritmo de Bayes ya es capaz de obtener los mejores hiperparámetros para el modelo. Los mejores hiperparámetros encontrados fueron $max_depth = 7$, $n_estimators = 550$ y un $learning_rate = 0,0167$, es decir se utilizaron 550 árboles de decisión con una profundidad de 7 y el valor que cada árbol contribuye al modelo lleva un escalamiento de 0,0167.

Para el mejor modelo seleccionado se tiene el rendimiento obtenido para cada uno de los pliegues en validación cruzada, notando que el modelo logra ajustarse bastante bien a los datos de entrenamiento, logrando además el mejor rendimiento en prueba entre todos los algoritmos. Se utiliza el modelo que de un menor MAE entre los 10 pliegues generados por la validación cruzada, en este caso correspondiente al quinto pliegue, obteniendo un $MAE = 12,07$ y un $R^2 = 0,45$.

Las características más influyentes en las predicciones del algoritmo de XGBoost fueron en primer lugar la potencia espectral normalizada en la banda alfa baja específica de cada sujeto (Low_subj_spec_EPP_OCC_right), segundo y tercer lugar las de IAF_OCC_left e IAF_OCC_right respectivamente las cuales corresponden a características que representan la ubicación del peak de frecuencia alfa visto en la región occipital, posterior a eso le siguen características que representan la densidad de potencia para las regiones cingular (CING) y parietal (PARIET).

Al sumar la totalidad de las contribuciones SHAP por regiones cerebrales se obtuvo que las principales regiones que contribuyeron en las predicciones fueron la región occipital derecha (OCC_right), seguidos de la región cingular derecha (CING_right) y región occipital izquierda (OCC_left).

7.4.1. Entrenamiento del Modelo

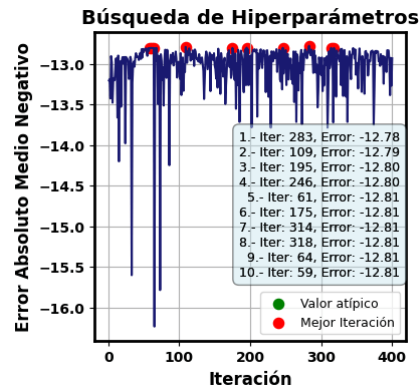
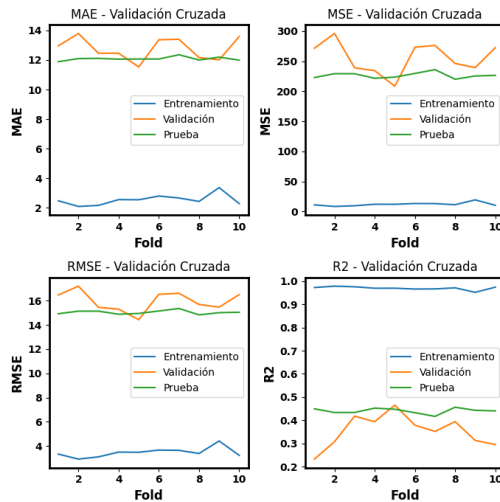
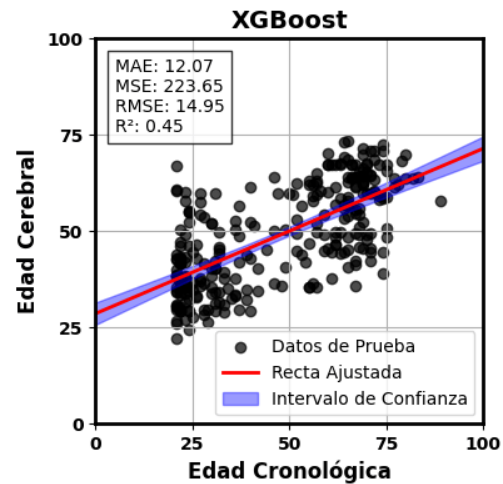


Figura 7.20: Búsqueda intensiva de hiperparámetros para XGBoost, el mejor conjunto de hiperparámetros se evalúa mediante el Error Absoluto Medio Negativo, se presentan los 10 mejores conjuntos de hiperparámetros obtenidos identificando la iteración en la cual ocurrió



(a) Evaluación del modelo en validación cruzada para XGBoost utilizando el mejor conjunto de hiperparámetros encontrados en la búsqueda bayesiana.



(b) Predicciones de los datos de prueba para el mejor modelo, se grafica la relación de la Edad Cronológica con la Edad Cerebral, se presentan las métricas de evaluación para mejor modelo, estas son el MAE, MSE, RMSE y R^2 . Se realiza un modelo de regresión lineal que relacione la Edad Cronológica con la Edad Cerebral de manera ver como esta línea de tendencia difiere con la recta de edad ideal.

7.4.2. Importancia de Características y Valores Shap

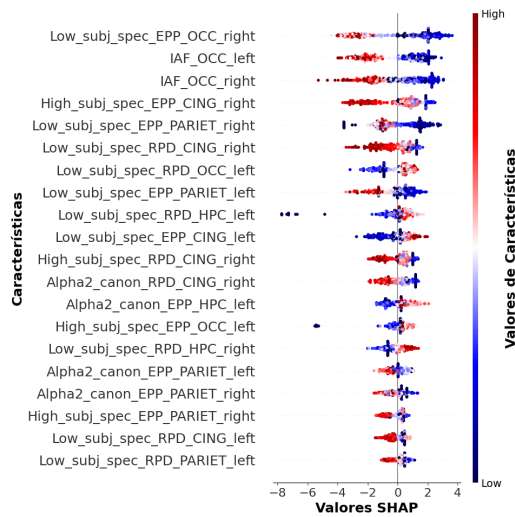


Figura 7.22: Valores SHAP calculados para las predicciones.

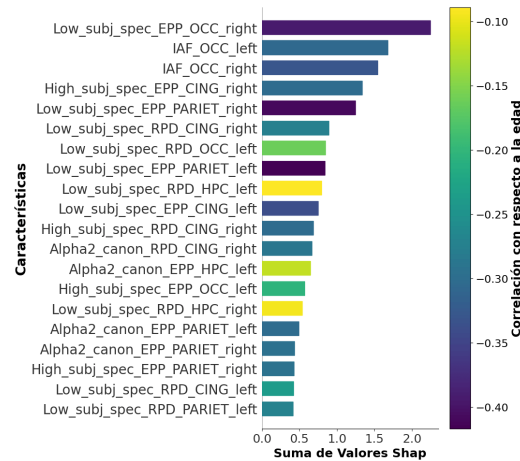


Figura 7.23: Suma de valores Shap y correlación de cada caracterísitca con la Edad.

7.4.3. Importancia por Región

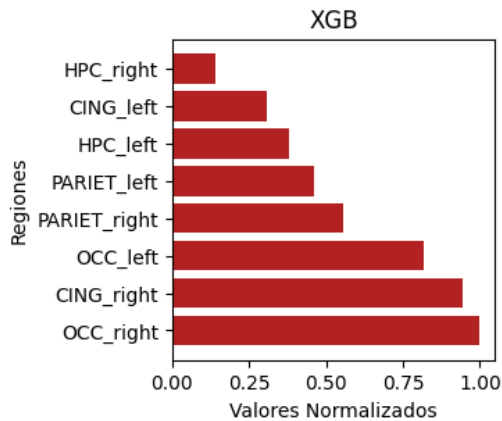


Figura 7.24: Suma de las 20 mejores SHAP por región cerebral, se normaliza entre 0 y 1 para analizar la importancia de cada región a la predicción.

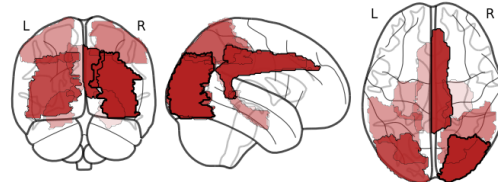


Figura 7.25: Regiones cerebrales mas influyentes en la predicción de Estimación de Edad Cerebral para XGBoost, las regiones más importantes fueron la Occipital derecha y Cingular derecha.

7.5. CatBoost

Para el algoritmo de CatBoost el mejor puntaje se obtuvo en la iteración 284 (Figura 7.26) y el segundo mejor puntaje en la iteración 386, siendo un algoritmo algo más lento en comparación a los anteriores modelos basados en árboles, cabe considerar que para el algoritmo de CatBoost el hardware que se utilizó fue distinto a los modelos anteriores, ya que este modelo fue ejecutado en un procesador Intel Core i7-1355U y el resto de los modelos se ejecutaron con el procesador Ryzen 5 5500u presentando búsquedas distintas entre ambos hardware. Los mejores hiperparámetros encontrados fueron $depth = 7$, $n_estimators = 1334$ y un $learning_rate = 0,02178$, es decir se utilizaron 1334 árboles de decisión con una profundidad de 7 y el valor que cada árbol contribuye al modelo lleva un escalamiento de 0,02178.

Para el mejor modelo seleccionado se tiene el rendimiento obtenido para cada uno de los pliegues en validación cruzada, notando que el modelo logra ajustarse bien a los datos de entrenamiento pero no tanto como los modelos de árboles anteriores. Se utiliza el modelo que de un menor MAE entre los 10 pliegues generados por la validación cruzada, en este caso correspondiente al noveno pliegue, obteniendo un $MAE = 12,20$ y un $R^2 = 0,45$.

Las características más influyentes en las predicciones del algoritmo de CatBoost fueron en primer lugar las características IAF_OCC_left e IAF_OCC_right respectivamente las cuales corresponden a características que representan la ubicación del peak de frecuencia alfa visto en la región occipital, seguido en tercer lugar de la potencia espectral normalizada en la banda alfa baja específica de cada sujeto (Low_subj_spec_EPP_OCC_right), posterior a eso le siguen características que representan la densidad de potencia para las regiones cingular (CING) izquierda y derecha.

Al sumar la totalidad de las contribuciones SHAP por regiones cerebrales se obtuvo que las principales regiones que contribuyeron en las predicciones fueron la región occipital derecha (OCC_right), seguidos de la región cingular derecha (CING_right) y región occipital izquierda (OCC_left).

7.5.1. Entrenamiento del Modelo

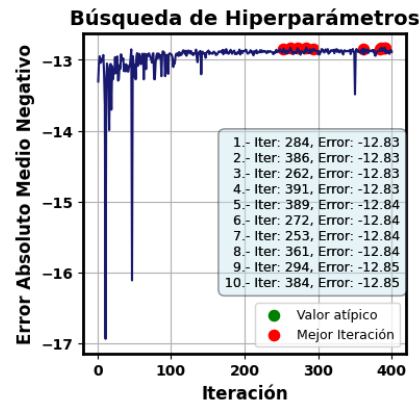
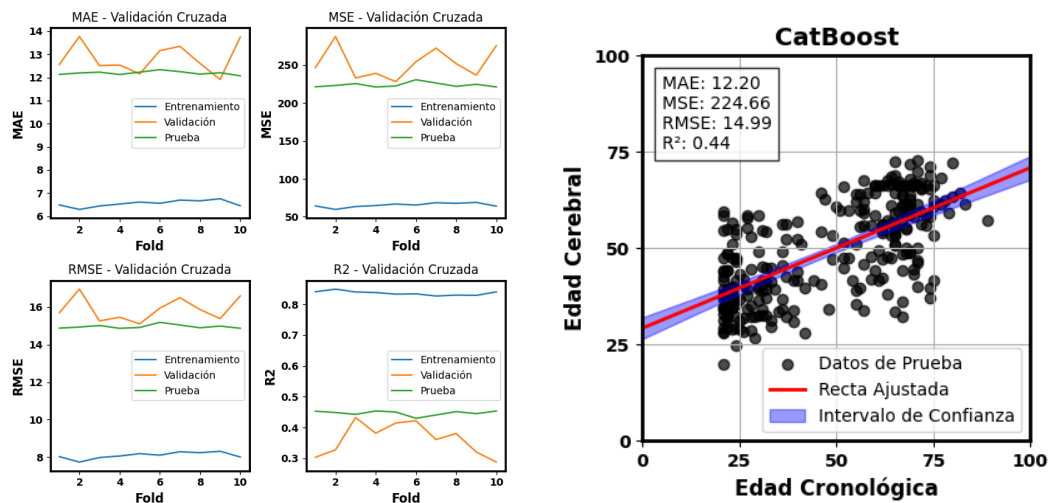


Figura 7.26: Búsqueda intensiva de hiperparámetros para CatBoost, el mejor conjunto de hiperparámetros se evalúa mediante el Error Absoluto Medio Negativo, se presentan los 10 mejores conjuntos de hiperparámetros obtenidos identificando la iteración en la cual ocurrió



(a) Evaluación del modelo en validación cruzada para CatBoost utilizando el mejor conjunto de hiperparámetros encontrados en la búsqueda bayesiana.

(b) Predicciones de los datos de prueba para el mejor modelo, se gráfica la relación de la Edad Real con la Edad Predicha, se presentan las métricas de evaluación para mejor modelo, estas son el MAE, MSE, RMSE y R^2 . Se realiza un modelo de regresión lineal que relacione la Edad Cronológica con la Edad Cerebral de manera ver como esta línea de tendencia difiere con la recta de edad ideal.

7.5.2. Importancia de Características y Valores SHAP

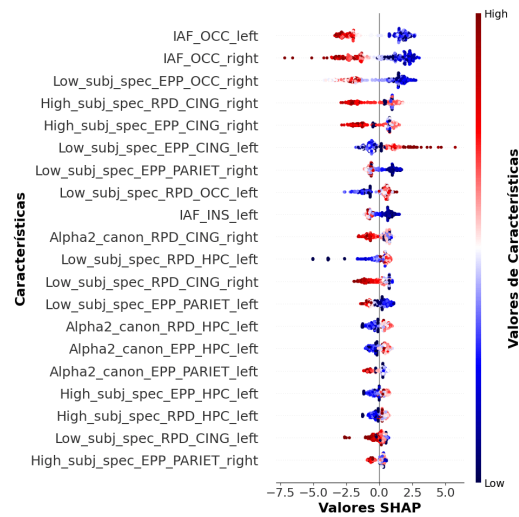


Figura 7.28: Valores SHAP calculados para las predicciones.

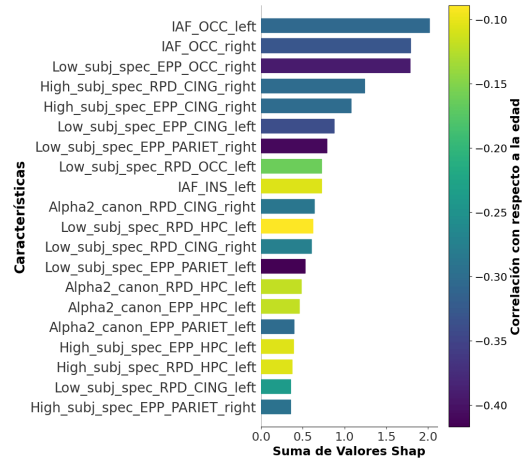


Figura 7.29: Suma de valores Shap y correlación de cada caracterísitca con la Edad.

7.5.3. Importancia por Región

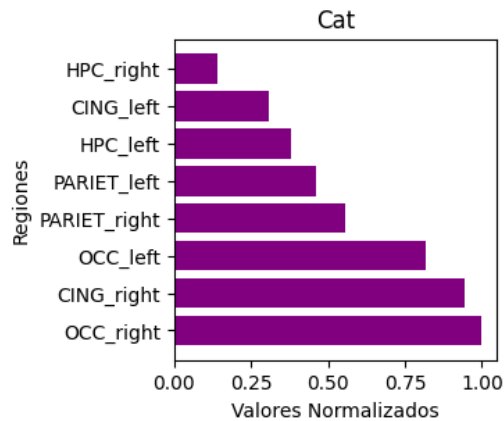


Figura 7.30: Suma de las 20 mejores SHAP por región cerebral, se normaliza entre 0 y 1 para analizar la importancia de cada región a la predicción.

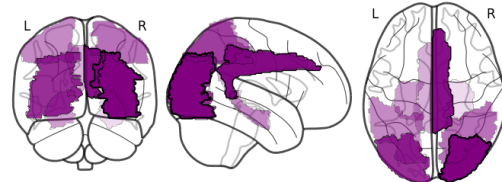


Figura 7.31: Regiones cerebrales mas influyentes en la predicción de Estimación de Edad Cerebral para CAtBoost, las regiones más importantes fueron la Occipital derecha y Cingular derecha.

7.6. Red Neuronal MLP

Para el algoritmo de Red Neuronal Multicapa el mejor puntaje se obtuvo en la iteración 125 (Figura 7.32) y el segundo mejor puntaje en la iteración 50, siendo el algoritmo mas lento en entrenar y realizar la búsqueda intensiva de hiperparámetros, además a diferencia de los otros algoritmos que utilizaron escalamiento de datos con MinMaxScaler que escala los datos entre 0 y 1 se utilizó StandardScaler que escala las características para que tengan media cero y varianza unitaria, esto con el propósito de tener todas las características en valores similar, evitando que la red neuronal genere pesos infinitos o nulos por la diferencia de escala entre las características. Los mejores hiperparámetros encontrados fueron '*activation*' = '*logistic*', *batch_size* = 128, *layer_1* = 1584, *layer_2* = 644 y un '*solver*' = '*sgd*', es decir se utilizaron 2 capas densas de 1584 y 644 capas respectivamente, como función de activación la función logística y el optimizador 'sgd'.

Para el mejor modelo seleccionado se tiene el rendimiento obtenido para cada uno de los pliegues en validación cruzada, notando que el modelo no logra ajustarse bien a los datos de entrenamiento, lo cual podría suponer que se podría poner el foco en realizar más iteraciones dentro del modelo para que logre aprender de mejor manera los datos, que no sea un buen modelo para este tipo de tareas o bien existen características redundantes que empeoran el modelo final. Se utiliza el modelo que de un menor *MAE* entre los 10 pliegues generados por la validación cruzada, en este caso correspondiente al quinto pliegue, obteniendo un $MAE = 12,80$ y un $R^2 = 0,35$.

Las características más influyentes en las predicciones del algoritmo de Red Neuronal Multicapa fueron en general de densidad de potencia para la banda alfa baja específica, primero *Low_subj_spec_RPD_OCC_left*, seguido de *Low_subj_spec_EPP_CING_left*, posterior a eso le siguen características que representan la densidad de potencia en la banda alfa baja específica para las regiones del hemisferio derecho.

Al sumar la totalidad de las contribuciones SHAP por regiones cerebrales se obtuvo que las principales regiones que contribuyeron en las predicciones fueron la región occipital derecha (*OCC_right*), seguidos de la región occipital izquierda (*OCC_left*), y la región cingular derecha (*CING_right*).

7.6.1. Entrenamiento del Modelo

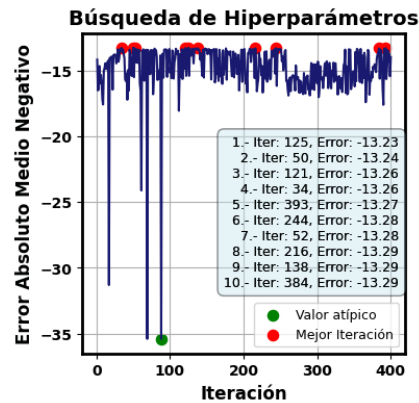
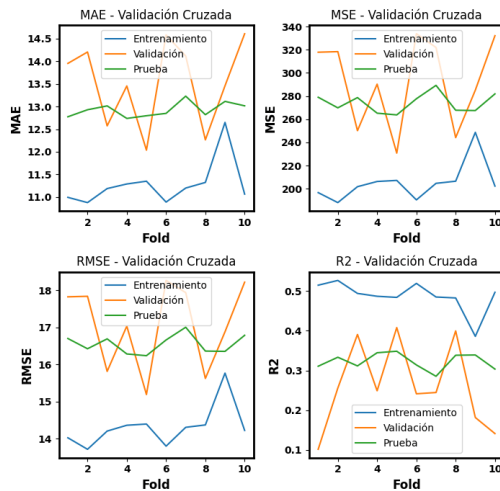
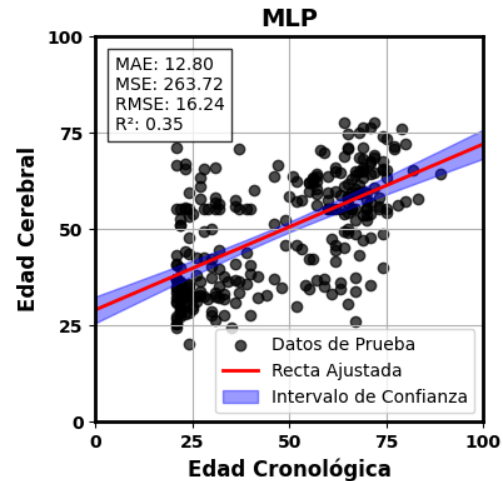


Figura 7.32: Búsqueda intensiva de hiperparámetros para MLP, el mejor conjunto de hiperparámetros se evalúa mediante el Error Absoluto Medio Negativo, se presentan los 10 mejores conjuntos de hiperparámetros obtenidos identificando la iteración en la cual ocurrió.



(a) Evaluación del modelo en validación cruzada para el modelo de Red Neuronal Multicapa utilizando el mejor conjunto de hiperparámetros encontrados en la búsqueda bayesiana.



(b) Predicciones de los datos de prueba para el mejor modelo, se grafica la relación de la Edad Cronológica con la Edad Cerebral, se presentan las métricas de evaluación para mejor modelo, estas son el MAE, MSE, RMSE y R^2 . Se realiza un modelo de regresión lineal que relacione la Edad Cronológica con la Edad Cerebral de manera ver como esta línea de tendencia difiere con la recta de edad ideal.

7.6.2. Importancia de Características y valores Shap

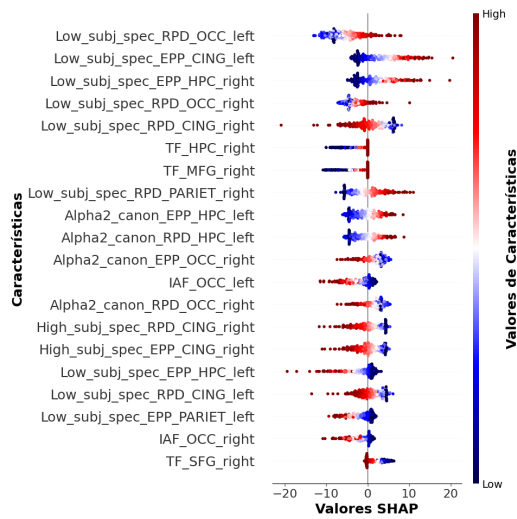


Figura 7.34: Valores SHAP calculados para las predicciones.

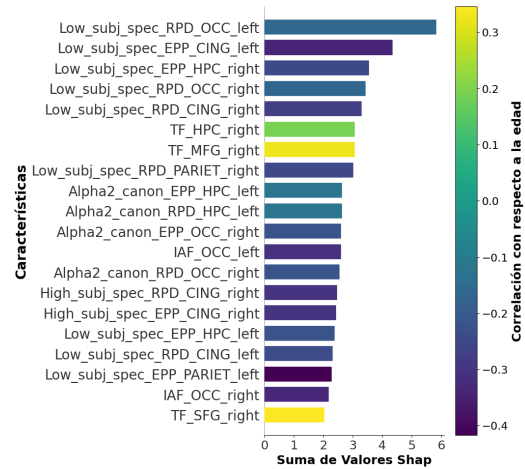


Figura 7.35: Suma de valores Shap y correlación de cada caracterísitca con la Edad.

7.6.3. Importancia por Región

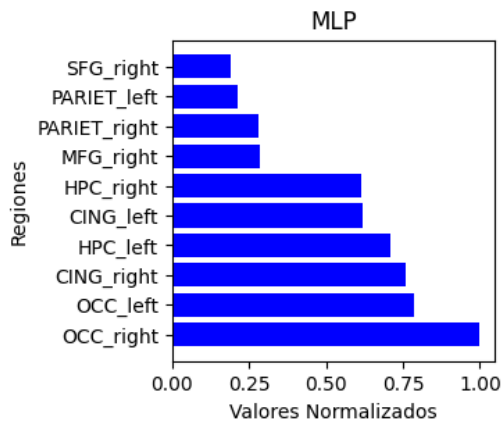


Figura 7.36: Suma de las 20 mejores SHAP por región cerebral, se normaliza entre 0 y 1 para analizar la importancia de cada región a la predicción.

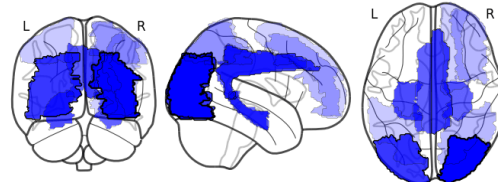


Figura 7.37: Regiones cerebrales mas influyentes en la predicción de Estimación de Edad Cerebral para el modelo de Red Neuronal, las regiones más importantes fueron la Occipital derecha e izquierda.

7.7. Comparación entre algoritmos

A continuación se presenta un cuadro comparativo para las predicciones hechas por los distintos algoritmos para el mismo set de datos, se presenta en color negro el modelo de regresión lineal ajustado a las predicciones y en color gris la línea de tendencia ideal de *edad cronológica = edad predicha*.

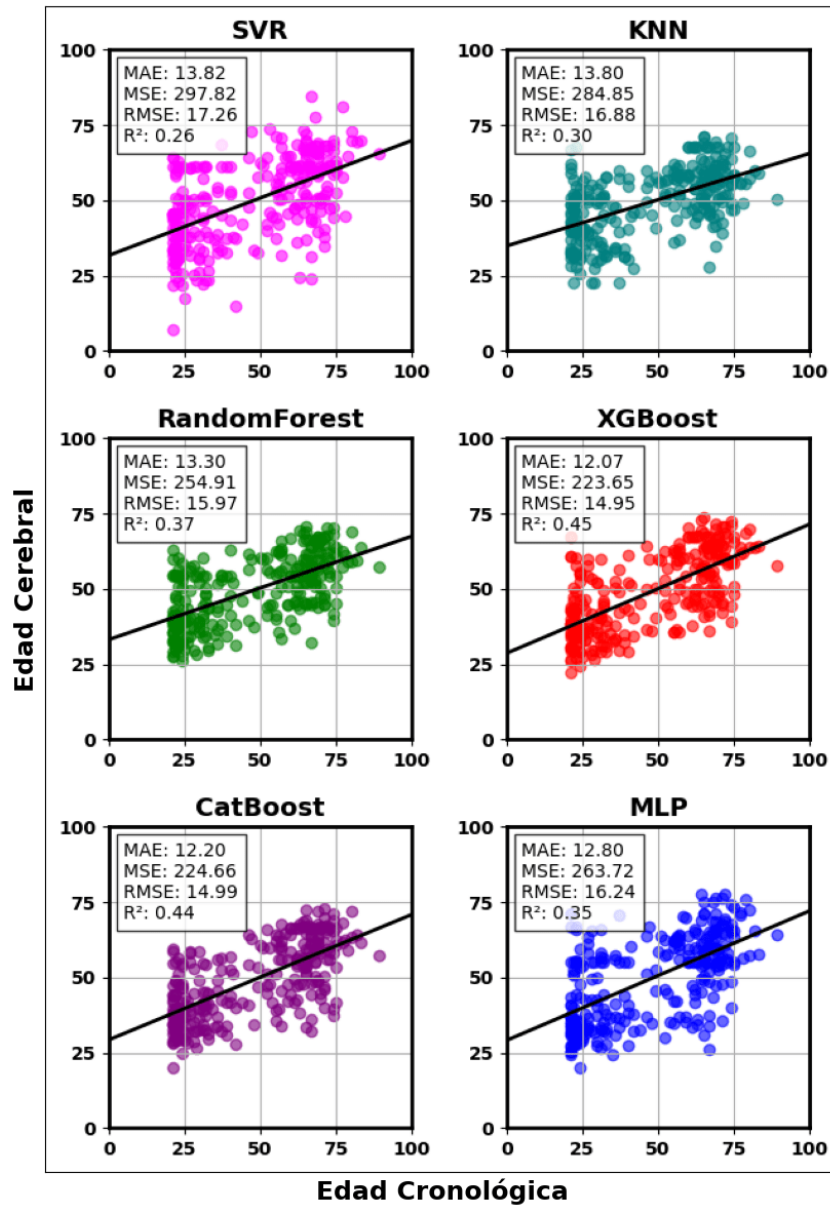


Figura 7.38: Predicciones

7.8. Comparación de Importancia por Región

Se presenta a continuación un cuadro comparativo de las regiones cerebrales que más influyeron en las predicciones de edad cerebral para cada uno de los algoritmos, notando además que para modelos basados en árboles las regiones fueron similares, ya que en los 3 modelos la principal región fue en la zona occipital derecha, cingular derecha y occipital izquierda. La región que más se repitió entre los 6 modelos fue la región occipital.

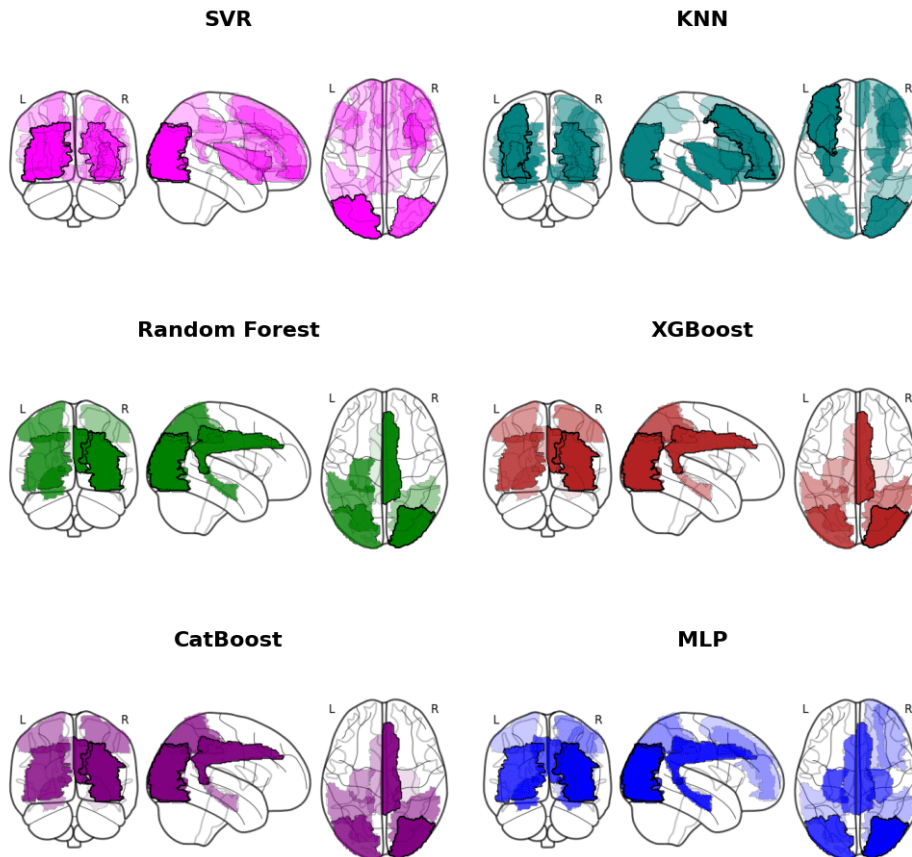


Figura 7.39: Comparación de regiones cerebrales

7.9. Incremento de Características

Para el mejor modelo, en este caso el algoritmo de XGBoost se evaluó iterativamente el modelo a medida que se fueron agregando características 1 a 1 en el orden de mayor contribución en los valores SHAP hasta agregar la última característica en esta lista de contribuciones con el fin de obtener como aumenta el rendimiento a medida que se agregan características y cual es el mínimo número de características que se necesitan para obtener un rendimiento óptimo. En el caso de este algoritmo de aumento de gradiente el número óptimo de características se definió como 20.

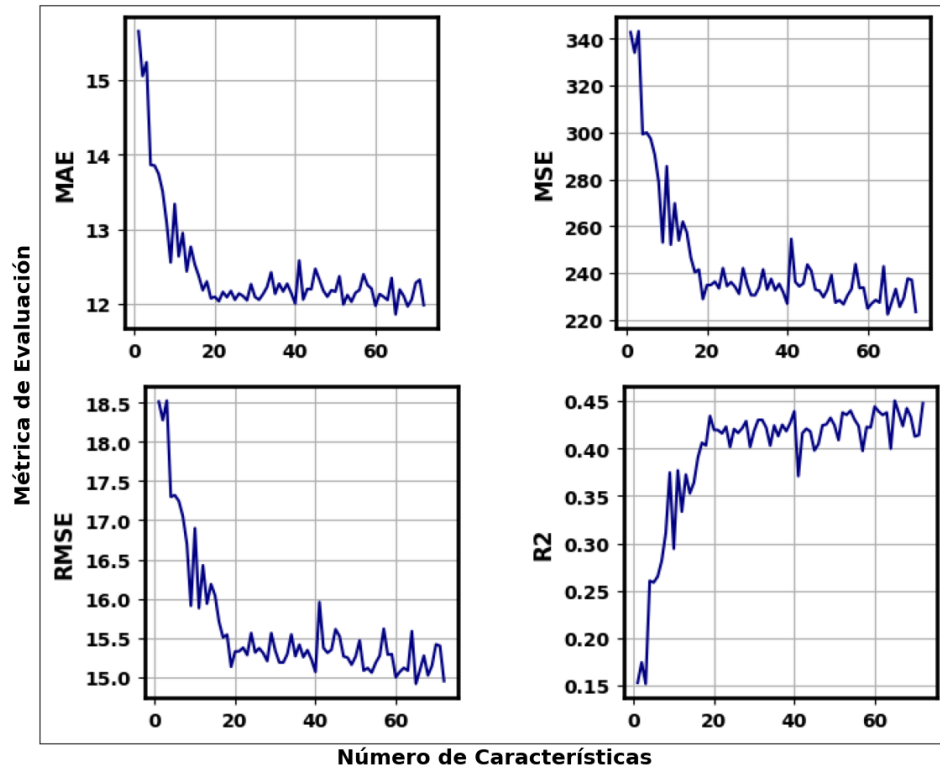


Figura 7.40: Evaluación del modelo incrementando las características que entrega SHAP una a una para el algoritmo de XGBoost, es posible notar que a partir de las 20 características las métricas de evaluación comienzan a tomar un valor poco variable, es decir, bastan al menos 20 características para obtener un rendimiento óptimo del modelo.

7.10. Mejores características

Para el mejor modelo, el algoritmo de XGBoost, se entrena utilizando las 20 mejores características evaluadas por valores SHAP las cuales según la evaluación de incremento de características se obtiene como un número óptimo de características, dando un resultado de $MAE = 12,14$ y $R^2 = 0,43$, que es solo un poco peor que el modelo evlauado con 72 características, con lo cual es posible deducir que el algoritmo de XGBoost resulta robusto a la presencia de características poco relevantes para el modelo.

7.10.1. Entrenamiento del modelo

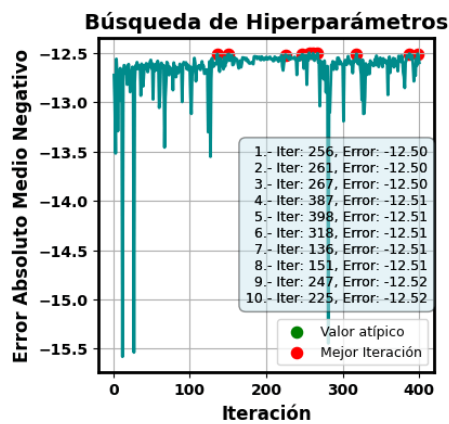


Figura 7.41: Búsqueda intensiva de hiperparámetros para XGBoost utilizando las 20 mejores características entregada por su evaluación de valores SHAP.

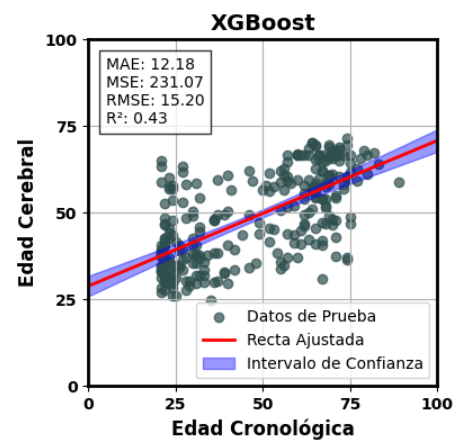


Figura 7.42: Predicciones del a Edad Cerebral utilizando XGBoost y las 20 características que más contribuyen al modelo.

7.10.2. Importancia de Características y valores SHAP

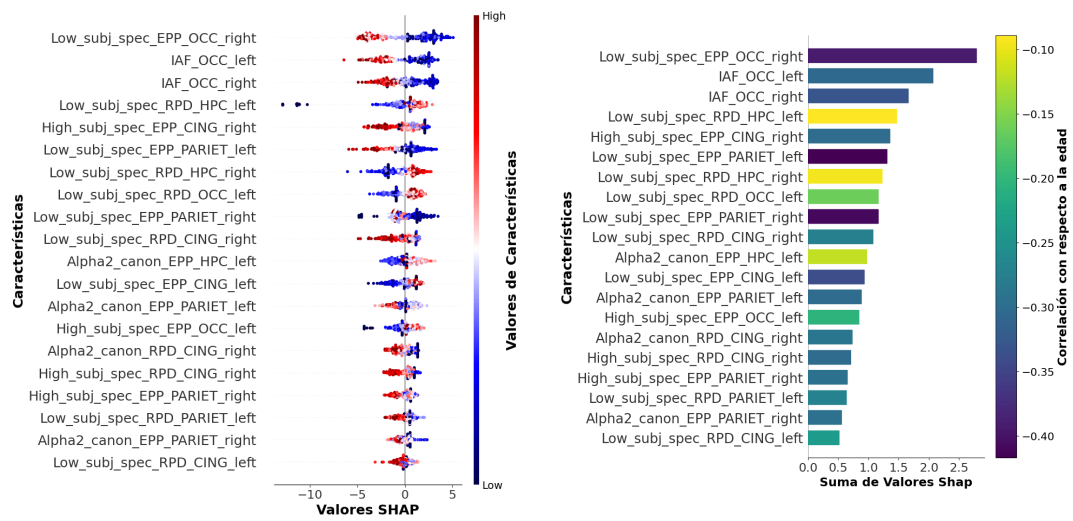


Figura 7.43: Valores SHAP

Figura 7.44: Suma de valores SHAP y correlación

CONCLUSIONES

FINALMENTE se concluye que los algoritmos más óptimos para realizar este tipo de tareas son los modelos basados en árboles de decisión, en especial los que utilizan aumento de gradiente como XGBoost y CatBoost, obteniendo un $MAE = 12,07$ y un $R^2 = 0,4$ para el algoritmo de XGBoost, siendo este el mejor rendimiento obtenido entre los modelos. Para todos los algoritmos en general se repitió que la región occipital fue la más influyente al momento de realizar las predicciones de edad cerebral, y las características que más se repitieron fueron las de IAF_OCC_right y IAF_OCC_left, correspondiente al valor de frecuencia alfa peak individual de cada individuo, pese a esto cada algoritmo difirió en qué características fueron más importantes, no así los modelos basados en arboles que fueron casi idénticos en la contribución de sus características a la predicción. Se repitió en la mayoría de los algoritmos que el quinto pliegue de la validación cruzada obtuvo el mejor modelo. Para el mejor modelo se determinó y evaluó cuál era el mínimo de características necesarias para obtener un rendimiento en las predicciones, obteniendo un número de 20 características.

Pese a que los resultados no logran superar en términos de predicción a los metodos de EEG basados en extracción de características mediante aprendizaje profundo o métodos más estudiados como los basados en imagenes de resonancia magnética, logra mantenerse dentro de los resultados presentados de la literatura para metodologías basadas en extracción manual de características.

A diferencia de los métodos basados en extracción de características con aprendizaje profundo, los métodos basados en extracción manual de características han permitido identificar biomarcadores específicos del cerebro, los cuales se ven alterados en el envejecimiento acelerado, permitiendo dar una descripción más detallada de qué lo que está sucediendo mas allá de obtener una predicción más certera.

8.1. Discusión

Al realizar análisis interpretativos en la predicción de la edad cerebral mediante aprendizaje automático y características de EEG, es esencial considerar diversas aristas. Aunque es posible desarrollar buenos modelos para nuestros análisis, la interpretación final de lo visualizado es fundamental. El hecho de que una característica contribuya más que otra no significa que la menos destacada carezca de importancia; simplemente, en esa instancia de datos, la primera tiene mayor influencia. Por esta razón, es crucial realizar múltiples pruebas para evaluar el comportamiento de los modelos en diferentes situaciones.

El hardware utilizado es un aspecto que no debe pasarse por alto, ya que los modelos de aprendizaje automático son computacionalmente costosos. Sin hardware especializado, la búsqueda intensiva de hiperparámetros puede demorar días. Además, la cantidad de datos disponibles es una limitante significativa si buscamos obtener modelos robustos.

En respuesta a estas limitaciones, es necesario adaptar los modelos a nuestras necesidades, considerando el tiempo de entrenamiento y el rendimiento que cada uno ofrece. Evaluar los modelos con diferentes conjuntos de muestras y características y realizar una interpretación profunda de los resultados obtenidos nos permitirá tomar decisiones informadas sobre las directrices a seguir.

8.2. Trabajos Futuros

Dada la amplia gama de posibilidades para abordar este tipo de problemas, es necesario cubrir la mayor cantidad de aristas posible. Para mejorar la generalidad del algoritmo, se podría evaluar su rendimiento utilizando únicamente las características más relevantes, descartando aquellas que no contribuyen a la predicción.

En el caso de las redes neuronales, se podría investigar el impacto de aumentar el número de iteraciones internas del algoritmo en el rendimiento del modelo, tanto en los datos de entrenamiento como en los de prueba. Además, sería valioso explorar cómo el envejecimiento cerebral influye en la predicción de patologías, asociando el envejecimiento a enfermedades como la demencia o el Alzheimer.

Bibliografía

- [1] Mario A Parra-Rodriguez et al. “The EuroLaD-EEG consortium: towards a global EEG platform for dementia, for seeking to reduce the regional impact of dementia”. En: *Alzheimer’s Dementia* 18 (S6 dic. de 2022). ISSN: 1552-5260. DOI: 10.1002/alz.059944.
- [2] Petr Klemra y Stanislav Doubal. “A new approach to the concept and computation of biological age”. En: *Mechanisms of Ageing and Development* 127 (3 mar. de 2006), págs. 240-248. ISSN: 00476374. DOI: 10.1016/j.mad.2005.10.004.
- [3] Toshiyuki Furukawa et al. *Assessment of Biological Age by Multiple Regression Analysis*. 1975. URL: <http://geronj.oxfordjournals.org/>.
- [4] Harry Benjamin. *Biologic Versus Chronologic Age*. 1947. URL: <http://geronj.oxfordjournals.org/>.
- [5] James H. Cole et al. *Brain age and other bodily ‘ages’: implications for neuropsychiatry*. Feb. de 2019. DOI: 10.1038/s41380-018-0098-1.
- [6] J. H. Cole et al. “Brain age predicts mortality”. En: *Molecular Psychiatry* 23 (5 mayo de 2018), págs. 1385-1392. ISSN: 14765578. DOI: 10.1038/mp.2017.62.
- [7] James H. Cole y Katja Franke. *Predicting Age Using Neuroimaging: Innovative Brain Ageing Biomarkers*. Dic. de 2017. DOI: 10.1016/j.tins.2017.10.001.
- [8] James H. Cole et al. “Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker”. En: *NeuroImage* 163 (dic. de 2017), págs. 115-124. ISSN: 10959572. DOI: 10.1016/j.neuroimage.2017.07.059.
- [9] Katja Franke et al. “Estimating the age of healthy subjects from T1-weighted MRI scans using kernel methods: Exploring the influence of various parameters”. En: *NeuroImage* 50 (3 abr. de 2010), págs. 883-892. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2010.01.005.

- [10] Iman Beheshti, Norihide Maikusa e Hiroshi Matsuda. “The association between “Brain-Age Score” (BAS) and traditional neuropsychological screening tools in Alzheimer’s disease”. En: *Brain and Behavior* 8 (8 ago. de 2018). ISSN: 21623279. DOI: 10.1002/brb3.1020.
- [11] Fernando Lopes da Silva. *EEG and MEG: Relevance to neuroscience*. Dic. de 2013. DOI: 10.1016/j.neuron.2013.10.017.
- [12] Pavel Prado et al. “Source space connectomics of neurodegeneration: One-metric approach does not fit all”. En: *Neurobiology of Disease* 179 (abr. de 2023). ISSN: 1095953X. DOI: 10.1016/j.nbd.2023.106047.
- [13] Lukas A.W. Gemein et al. “Machine-learning-based diagnostics of EEG pathology”. En: *NeuroImage* 220 (oct. de 2020). ISSN: 10959572. DOI: 10.1016/j.neuroimage.2020.117021.
- [14] Robin Tibor Schirrmeister et al. “Deep learning with convolutional neural networks for decoding and visualization of EEG pathology”. En: (ago. de 2017). URL: <http://arxiv.org/abs/1708.08012>.
- [15] Mariam Khayretdinova et al. “Predicting age from resting-state scalp EEG signals with deep convolutional neural networks on TD-brain dataset”. En: *Frontiers in Aging Neuroscience* 14 (dic. de 2022). ISSN: 16634365. DOI: 10.3389/fnagi.2022.1019869.
- [16] Mohammad Parsa Hosseini, Amin Hosseini y Kiarash Ahi. “A Review on Machine Learning for EEG Signal Processing in Bioengineering”. En: *IEEE Reviews in Biomedical Engineering* 14 (2021), págs. 204-218. ISSN: 19411189. DOI: 10.1109/RBME.2020.2969915.
- [17] Lea Baecker et al. *Machine learning for brain age prediction: Introduction to methods and clinical applications*. Oct. de 2021. DOI: 10.1016/j.ebiom.2021.103600.
- [18] Marjolein M.L.J.Z. Vandenbosch et al. “EEG-based age-prediction models as stable and heritable indicators of brain maturational level in children and adolescents”. En: *Human Brain Mapping* 40 (6 abr. de 2019), págs. 1919-1926. ISSN: 10970193. DOI: 10.1002/hbm.24501.
- [19] Obada Al Zoubi et al. “Predicting age from brain EEG signals-a machine learning approach”. En: *Frontiers in Aging Neuroscience* 10 (JUL jul. de 2018). ISSN: 16634365. DOI: 10.3389/fnagi.2018.00184.
- [20] Denis A. Engemann et al. “A reusable benchmark of brain-age prediction from M/EEG resting-state signals”. En: *NeuroImage* 262 (nov. de 2022). algoritmo ML: Random Forest para extraccion de características. ISSN: 10959572. DOI: 10.1016/j.neuroimage.2022.119521.

- [21] Pavel Prado et al. “The BrainLat project, a multimodal neuroimaging dataset of neurodegeneration from underrepresented backgrounds”. En: *Scientific Data* 10 (1 dic. de 2023), pág. 889. ISSN: 2052-4463. DOI: 10.1038/s41597-023-02806-8. URL: <https://www.nature.com/articles/s41597-023-02806-8>.
- [22] R D Pascual-Marqui. *Standardized low resolution brain electromagnetic tomography (sLORETA): technical details Running title: sLORETA*. URL: <http://www.keyinst.unizh.ch/loreta.htm>.
- [23] Aron T. Hill et al. “Periodic and aperiodic neural activity displays age-dependent changes across early-to-middle childhood”. En: *Developmental Cognitive Neuroscience* 54 (abr. de 2022). ISSN: 18789307. DOI: 10.1016/j.dcn.2022.101076.
- [24] Claudio Babiloni et al. *International Federation of Clinical Neurophysiology (IFCN) – EEG research workgroup: Recommendations on frequency and topographic analysis of resting state EEG rhythms. Part 1: Applications in clinical research studies*. Ene. de 2020. DOI: 10.1016/j.clinph.2019.06.234.
- [25] Kevin K. Dobbin y Richard M. Simon. “Optimally splitting cases for training and testing high dimensional classifiers”. En: *BMC Medical Genomics* 4 (2011). ISSN: 17558794. DOI: 10.1186/1755-8794-4-31.
- [26] Pradyumna Lanka et al. “Supervised machine learning for diagnostic classification from large-scale neuroimaging datasets”. En: *Brain Imaging and Behavior* 14 (6 dic. de 2020), págs. 2378-2416. ISSN: 19317565. DOI: 10.1007/s11682-019-00191-8.
- [27] Sebastian Moguilner et al. “Multi-feature computational framework for combined signatures of dementia in underrepresented settings”. En: *Journal of Neural Engineering* 19 (4 ago. de 2022). ISSN: 17412552. DOI: 10.1088/1741-2552/ac87d0.
- [28] Corinna Cortes, Vladimir Vapnik y Lorenza Saitta. *Support-Vector Networks Editor*. 1995.
- [29] Tianqi Chen y Carlos Guestrin. *XGBoost: A Scalable Tree Boosting System*. URL: <https://github.com/dmlc/xgboost>.
- [30] Liudmila Prokhorenkova et al. *CatBoost: unbiased boosting with categorical features*. URL: <https://github.com/catboost/catboost>.
- [31] Pavel Prado et al. “Harmonized multi-metric and multi-centric assessment of EEG source space connectivity for dementia characterization”. En: *Alzheimer’s and Dementia: Diagnosis, Assessment and Disease Monitoring* 15 (3 jul. de 2023). ISSN: 23528729. DOI: 10.1002/dad2.12455.

