

Clasificación de electroencefalogramas (EEG).

Introducción.

El estudio realizado para hacer la clasificación de electroencefalogramas consiste en una evaluación de respuesta inhibitoria en trastorno por déficit de atención con hiperactividad y problemas de aprendizaje.

Se realizó a cierto grupo de sujetos y estos tenían que hacerlo de dos formas:

1. Ojos cerrados.
2. Ojos abiertos.

Ambos estando en reposo.

Así mismo, la colocación de electrodos fue puesta de acuerdo a las áreas en las que se divide el cerebro.

Objetivo.

El objetivo de esto es crear un algoritmo capaz de clasificar los trastornos antes mencionados y así nos permita distinguir las condiciones para otros sujetos.

Desarrollo.

En las reuniones el doctor Gregorio nos fue asesorando sobre ciertos filtros y las áreas del cerebro.

Las áreas seleccionadas para las pruebas se tomaron de acuerdo a lo explicado en las reuniones.

Para todas las pruebas se seleccionó el área frontal ya que es el área donde se registran mejor las actividades del cerebro.

Para algunas pruebas se seleccionó el área parietal para poder comparar las diferencias con la frontal.

La transformada de wavelet se aplicó ya que al investigar (tesis, artículos, videos), se llegó a la conclusión de que éstas sirven mejor para filtrar datos de electroencefalogramas.

Algunas actividades se registran mejor con una frecuencia de 4 a 8 hz. Es por esa razón que se usaron distintas frecuencias en los filtros de parabanda y pasabanda. Se comparan estos dos últimos ya que uno filtra las ondas entre las frecuencias seleccionadas y el otros corta las frecuencias y usa todas menos las que se encuentran en ese rango.

Los datos contenidos en cada archivo son demasiados, por lo tanto se optó por utilizar el promedio de cada conjunto de electrodos (divididos por área). También se utilizó la medida de dispersión de

los datos (desviación estándar) para ver si el comportamiento de los modelos variaba y la mediana.

Las 3 primeras pruebas se realizaron como exploración para ver el comportamiento de los datos con un algoritmo de clasificación binaria es por eso que solo se utilizan 2 sujetos por prueba y solo 2 carpetas.

Prueba 1.

La primera prueba consiste en una clasificación binaria entre PELF y TDAH. Por esa razón seleccionamos a un sujeto de la carpeta PELF y a otro de la carpeta TDAH.

Después de esto seleccionamos los 10 electrodos correspondientes al área frontal y los extraemos de cada uno de los sujetos y sacamos el promedio de cada uno de estos.

Una vez teniendo nuestras matrices de cada sujeto, aplicamos un filtro biortogonal que pertenece a la transformada wavelet, este a su vez nos regresa dos matrices, una de coeficientes de alta frecuencia y otra de coeficientes de baja frecuencia (de aproximación y detallados).

Se trabajó con ambas matrices para comparar los resultados obtenidos después de aplicar la regresión.

Como se utiliza una regresión logística, agregamos una "columna objetivo" la cual nos va a ayudar a distinguir entre PELF y TDAH, por lo tanto a esa columna le asignamos los valores de 1 a todos los datos que pertenecen a nuestra carpeta TDAH y con 0 a todos aquellos que son de la carpeta PELF.

Ya que aplicamos lo anterior, procedemos a unir la matriz de coeficientes de aproximación de PELF con los coeficientes aproximación de TDAH.

Ya que tenemos la matriz final con la que vamos a trabajar, estandarizamos nuestros datos. Una vez estandarizados segmentamos nuestros datos para entrenar nuestro modelo en un 80% y el restante se usa para evaluarlo.

De igual forma se aplica el mismo proceso con las matrices de coeficientes detallados.

Una vez que se corre el modelo, los datos obtenidos son:

	precision	recall	f1-score	support
0	0.82	0.82	0.82	34
1	0.76	0.76	0.76	25
accuracy			0.80	59
macro avg	0.79	0.79	0.79	59
weighted avg	0.80	0.80	0.80	59

Los resultados obtenidos no son malos pero como solo es una exploración, se decidió realizar más pruebas y ver el comportamiento de acuerdo a los estadísticos que se iban usando.

Prueba 2.

En esta segunda prueba la selección de sujetos, datos y forma de extraer los archivos fue la misma. La única diferencia fue que en lugar de obtener el promedio de estos, se aplicó una desviación estándar.

Una vez aplicada, los pasos siguientes fueron los mismos que en la prueba 1.

Las matrices finales que se trabajaron fueron de igual forma las de coeficientes de aproximación y detallados, obteniendo como resultado:

Coeficientes de aproximación

	precision	recall	f1-score	support
0	0.85	0.97	0.90	34
1	0.95	0.76	0.84	25
accuracy			0.88	59
macro avg	0.90	0.87	0.87	59
weighted avg	0.89	0.88	0.88	59

Coeficientes detallados

	precision	recall	f1-score	support
0	0.97	1.00	0.99	34
1	1.00	0.96	0.98	25
accuracy			0.98	59
macro avg	0.99	0.98	0.98	59
weighted avg	0.98	0.98	0.98	59

Los resultados que se obtuvieron en esta prueba son demasiado buenos y es muy raro que eso llegue a pasar, puede que exista un sobreajuste con los datos.

Prueba 3.

Para la prueba 3, los sujetos con los que se trabajaron fueron los mismos, se seleccionó a cada uno de las carpetas TDAH y PELF.

La selección de electrodos fue diferente, aquí se seleccionaron los electrodos correspondientes al área temporal.

De igual forma a cada grupo de electrodos por carpeta se les sacó el promedio y se unieron para así crear una matriz de TDAH y otra correspondiente a PELF.

Una vez teniendo nuestras matrices de cada sujeto, aplicamos un filtro biortogonal que pertenece a la transformada wavelet, este a su vez nos regresa dos matrices, una de coeficientes de alta frecuencia y otra de coeficientes de baja frecuencia.

Se trabajó con ambas matrices para comparar los resultados obtenidos después de aplicar la regresión.

Como se utiliza una regresión logística, agregamos una “columna objetivo” la cual nos va a ayudar a distinguir entre PELF y TDAH, por lo tanto a esa columna le asignamos los valores de 1 a todos los datos que pertenecen a nuestra carpeta TDAH y con 0 a todos aquellos que son de la carpeta PELF.

Ya que aplicamos lo anterior, procedemos a unir la matriz de coeficientes de alta frecuencia de PELF con los coeficientes de alta frecuencia de TDAH.

Ya que tenemos la matriz final con la que vamos a trabajar, estandarizamos nuestros datos. Una vez estandarizados segmentamos nuestros datos para entrenar nuestro modelo en un 80% y el restante se usa para evaluarlo.

De igual forma se aplica el mismo proceso con las matrices de coeficientes de baja frecuencia.

Una vez que se corre el modelo, los datos obtenidos son:

	precision	recall	f1-score	support
0	0.88	0.82	0.85	34
1	0.78	0.84	0.81	25
accuracy			0.83	59
macro avg	0.83	0.83	0.83	59
weighted avg	0.83	0.83	0.83	59

	precision	recall	f1-score	support
0	0.90	0.82	0.86	34
1	0.79	0.88	0.83	25
accuracy			0.85	59
macro avg	0.84	0.85	0.85	59
weighted avg	0.85	0.85	0.85	59

A partir de estas pruebas se decidió tomar todos los sujetos para así poder observar si los algoritmos desarrollados cumplían con su objetivo de clasificar a cada sujeto en su etiqueta correspondiente.

Para las pruebas 4,5 y 6 se utilizó una frecuencia de 10 a 40 Hz ya que entre estas frecuencias se procesan mejor las ondas.

Prueba 4.

Para esta prueba se utilizaron todos los sujetos de prueba en cada uno de los archivos de las carpetas (PELF, TDAH y Controles).

Al abrir la carpeta PELF, seleccionaba el grupo de electrodos que pertenecen al área frontal, juntaba todos los electrodos seleccionados en una sola matriz para al final obtener un promedio de todos estos datos.

De igual forma se aplicó esto para las carpetas TDAH y Controles.

Por lo tanto nuestro resultado fueron 3 matrices.

A estas 3 matrices se les aplicó un filtro de parada de banda de 10 a 40 Hz.

Después de esto se le agregó la "columna objetivo". Para los sujetos pertenecientes a Controles se les asignó el número 1, para los del grupo TDAH el 2 y para los otros el 3.

Una vez que ya las matrices tuvieran la columna de objetivo asignada, se unen estas 3 para así obtener una sola matriz.

Después se estandarizaron los datos con los que se iban a trabajar para aplicar los distintos modelos.

KNN

Este algoritmo nos ayuda a clasificar basándose en las observaciones más cercanas que se están tratando de predecir para así clasificarlos.

Para esta prueba primero se utilizó un $n=5$ (el que viene por default), donde n corresponde al número de vecinos cercanos.

Después se usó un $n=3$, para que identificara a sus 3 vecinos más cercanos, ya que nuestros sujetos están clasificados en 3 distintas carpetas.

Los resultados con las pruebas anteriores fueron muy bajos.

Al usar un $n=9$ y $n=10$, los resultados subieron considerablemente y con un $n>10$, la exactitud comienza a bajar.

Estas mismas modificaciones se hicieron en todas las pruebas que incluyen el algoritmo knn.

Árbol de decisión.

En este algoritmo se utilizó el árbol de decisión como clasificador y sus valores que vienen por default.

SVM.

Para la máquina de soporte vectorial utilizamos el módulo Clasificador de soportes vectoriales (SVC).

Prueba 5.

Para esta prueba se utilizaron todos los sujetos de prueba en cada uno de los archivos de las carpetas (PELF, TDAH y Controles).

Al abrir la carpeta PELF, seleccionaba el grupo de electrodos que pertenecen al área frontal, juntaba todos los electrodos seleccionados en una sola matriz para al final obtener la desviación estándar de todos estos datos.

De igual forma se aplicó esto para las carpetas TDAH y Controles.

Por lo tanto nuestro resultado fueron 3 matrices.

A estas 3 matrices se les aplicó un filtro de parada de banda de 10 a 40 hz.

Después de esto se le agregó la “columna objetivo”. Para los sujetos pertenecientes a Controles se les asignó el número 1, para los del grupo TDAH el 2 y para los otros el 3.

Una vez que ya las matrices tuvieran la columna de objetivo asignada, se unen estas 3 para así obtener una sola matriz.

Después se estandarizaron los datos con los que se iban a trabajar para aplicar los distintos modelos.

Prueba 6.

Para esta prueba se utilizaron todos los sujetos de prueba en cada uno de los archivos de las carpetas (PELF, TDAH y Controles).

Al abrir la carpeta PELF, seleccionaba el grupo de electrodos que pertenecen al área frontal, juntaba todos los electrodos seleccionados en una sola matriz para al final obtener la mediana de todos estos datos.

De igual forma se aplicó esto para las carpetas TDAH y Controles.

Por lo tanto nuestro resultado fueron 3 matrices.

A estas 3 matrices se les aplicó un filtro de parada de banda de 10 a 40 hz.

Después de esto se le agregó la “columna objetivo”. Para los sujetos pertenecientes a Controles se les asignó el número 1, para los del grupo TDAH el 2 y para los otros el 3.

Una vez que ya las matrices tuvieran la columna de objetivo asignada, se unen estas 3 para así obtener una sola matriz.

Después se estandarizaron los datos con los que se iban a trabajar para aplicar los distintos modelos.

Resultados de
y 6.

	KNN	Árbol de decisión	SVM
Promedio	44.44%	63.63%	54.54%
Std	66.66%	37.50%	63.63%
Median	44.44%	36.36%	18.18%

las pruebas 4,5

Los resultados de las pruebas 4,5 y 6 no fueron los mejores, por esa razón se hicieron otras pruebas con distintas frecuencias.

Para las pruebas 7,8 y 9 se utilizó una frecuencia de 4 a 8 Hz ya que entre estas frecuencias se procesan mejor las ondas para el área temporal.

En estas tres pruebas, solo se aplicaron los modelos de KNN y SVM.

Prueba 7.

Para esta prueba se utilizaron todos los sujetos de prueba en cada uno de los archivos de las carpetas (PELF, TDAH y Controles).

Al abrir la carpeta PELF, seleccionaba el grupo de electrodos que pertenecen al área frontal, juntaba todos los electrodos seleccionados en una sola matriz para al final obtener un promedio de todos estos datos.

De igual forma se aplicó esto para las carpetas TDAH y Controles.

Por lo tanto nuestro resultado fueron 3 matrices.

A estas 3 matrices se les aplicó un filtro de parada de banda de 4 a 8 hz.

Después de esto se le agregó la “columna objetivo”. Para los sujetos pertenecientes a Controles se les asignó el número 1, para los del grupo TDAH el 2 y para los otros el 3.

Una vez que ya las matrices tuvieran la columna de objetivo asignada, se unen estas 3 para así obtener una sola matriz.

Después se estandarizaron los datos con los que se iban a trabajar para aplicar los distintos modelos.

Prueba 8.

Para esta prueba se utilizaron todos los sujetos de prueba en cada uno de los archivos de las carpetas (PELF, TDAH y Controles).

Al abrir la carpeta PELF, seleccionaba el grupo de electrodos que pertenecen al área frontal, juntaba todos los electrodos seleccionados en una sola matriz para al final obtener la desviación estándar de todos estos datos.

De igual forma se aplicó esto para las carpetas TDAH y Controles.

Por lo tanto nuestro resultado fueron 3 matrices.

A estas 3 matrices se les aplicó un filtro de parada de banda de 4 a 8 hz.

Después de esto se le agregó la “columna objetivo”. Para los sujetos pertenecientes a Controles se les asignó el número 1, para los del grupo TDAH el 2 y para los otros el 3.

Una vez que ya las matrices tuvieran la columna de objetivo asignada, se unen estas 3 para así obtener una sola matriz.

Después se estandarizaron los datos con los que se iban a trabajar para aplicar los distintos modelos.

Prueba 9.

Para esta prueba se utilizaron todos los sujetos de prueba en cada uno de los archivos de las carpetas (PELF, TDAH y Controles).

Al abrir la carpeta PELF, seleccionaba el grupo de electrodos que pertenecen al área frontal, juntaba todos los electrodos seleccionados en una sola matriz para al final obtener la mediana de todos estos datos.

De igual forma se aplicó esto para las carpetas TDAH y Controles.

Por lo tanto nuestro resultado fueron 3 matrices.

A estas 3 matrices se les aplicó un filtro de parada de banda de 4 a 8 hz.

Después de esto se le agregó la “columna objetivo”. Para los sujetos pertenecientes a Controles se les asignó el número 1, para los del grupo TDAH el 2 y para los otros el 3.

Una vez que ya las matrices tuvieran la columna de objetivo asignada, se unen estas 3 para así obtener una sola matriz.

Después se estandarizaron los datos con los que se iban a trabajar para aplicar los distintos modelos.

Resultados de las

	KNN	SVM
Promedio	22.22%	27.27%
Std	27.27%	27.27%
Median	36.36%	45.45%

pruebas 7,8 y 9.

Prueba 10.

Para esta prueba se utilizaron todos los sujetos de prueba en cada uno de los archivos de las carpetas (PELF, TDAH y Controles).

Al abrir la carpeta PELF, seleccionaba el grupo de electrodos que pertenecen al área frontal, juntaba todos los electrodos seleccionados en una sola matriz para al final obtener un promedio de todos estos datos.

De igual forma se aplicó esto para las carpetas TDAH y Controles.

Por lo tanto nuestro resultado fueron 3 matrices.

A estas 3 matrices se les aplicó un filtro de paso de banda de 10 a 40 hz.

Después de esto se le agregó la “columna objetivo”. Para los sujetos pertenecientes a Controles se les asignó el número 1, para los del grupo TDAH el 2 y para los otros el 3.

Una vez que ya las matrices tuvieran la columna de objetivo asignada, se unen estas 3 para así obtener una sola matriz.

Después se estandarizaron los datos con los que se iban a trabajar para aplicar los distintos modelos.

Prueba 11.

Para esta prueba se utilizaron todos los sujetos de prueba en cada uno de los archivos de las carpetas (PELF, TDAH y Controles).

Al abrir la carpeta PELF, seleccionaba el grupo de electrodos que pertenecen al área frontal, juntaba todos los electrodos seleccionados en una sola matriz para al final obtener la desviación estándar de todos estos datos.

De igual forma se aplicó esto para las carpetas TDAH y Controles.

Por lo tanto nuestro resultado fueron 3 matrices.

A estas 3 matrices se les aplicó un filtro de pasabanda de 10 a 40 hz.

Después de esto se le agregó la “columna objetivo”. Para los sujetos pertenecientes a Controles se les asignó el número 1, para los del grupo TDAH el 2 y para los otros el 3.

Una vez que ya las matrices tuvieran la columna de objetivo asignada, se unen estas 3 para así obtener una sola matriz.

Después se estandarizaron los datos con los que se iban a trabajar para aplicar los distintos modelos.

Prueba 12.

Para esta prueba se utilizaron todos los sujetos de prueba en cada uno de los archivos de las carpetas (PELF, TDAH y Controles).

Al abrir la carpeta PELF, seleccionaba el grupo de electrodos que pertenecen al área frontal, juntaba todos los electrodos seleccionados en una sola matriz para al final obtener la mediana de todos estos datos.

De igual forma se aplicó esto para las carpetas TDAH y Controles.

Por lo tanto nuestro resultado fueron 3 matrices.

A estas 3 matrices se les aplicó un filtro de pasabanda de 10 a 40 hz.

Después de esto se le agregó la “columna objetivo”. Para los sujetos pertenecientes a Controles se les asignó el número 1, para los del grupo TDAH el 2 y para los otros el 3.

Una vez que ya las matrices tuvieran la columna de objetivo asignada, se unen estas 3 para así obtener una sola matriz.

Después se estandarizaron los datos con los que se iban a trabajar para aplicar los distintos modelos.

Resultados de
10,11 y 12.

	KNN	Árbol de decisión	SVM
Promedio	33.33%	18.18%	36.36%
Std	44.44%	18.18%	27.27%
Median	33.33%	45.45%	45.45%

las pruebas

Prueba 13.

Para esta prueba se utilizaron todos los sujetos de prueba en cada uno de los archivos de las carpetas (PELF, TDAH y Controles).

Al abrir la carpeta PELF, seleccionaba el grupo de electrodos que pertenecen al área frontal, juntaba todos los electrodos seleccionados en una sola matriz para al final obtener un promedio de todos estos datos.

De igual forma se aplicó esto para las carpetas TDAH y Controles.

Por lo tanto nuestro resultado fueron 3 matrices.

A estas 3 matrices se les aplicó un filtro de pasabanda de 4 a 8 hz.

Después de esto se le agregó la “columna objetivo”. Para los sujetos pertenecientes a Controles se les asignó el número 1, para los del grupo TDAH el 2 y para los otros el 3.

Una vez que ya las matrices tuvieran la columna de objetivo asignada, se unen estas 3 para así obtener una sola matriz.

Después se estandarizaron los datos con los que se iban a trabajar para aplicar los distintos modelos.

Prueba 14.

Para esta prueba se utilizaron todos los sujetos de prueba en cada uno de los archivos de las carpetas (PELF, TDAH y Controles).

Al abrir la carpeta PELF, seleccionaba el grupo de electrodos que pertenecen al área frontal, juntaba todos los electrodos seleccionados en una sola matriz para al final obtener la desviación estándar de todos estos datos.

De igual forma se aplicó esto para las carpetas TDAH y Controles.

Por lo tanto nuestro resultado fueron 3 matrices.

A estas 3 matrices se les aplicó un filtro de pasabanda de 4 a 8 hz.

Después de esto se le agregó la “columna objetivo”. Para los sujetos pertenecientes a Controles se les asignó el número 1, para los del grupo TDAH el 2 y para los otros el 3.

Una vez que ya las matrices tuvieran la columna de objetivo asignada, se unen estas 3 para así obtener una sola matriz.

Después se estandarizaron los datos con los que se iban a trabajar para aplicar los distintos modelos.

Prueba 15.

Para esta prueba se utilizaron todos los sujetos de prueba en cada uno de los archivos de las carpetas (PELF, TDAH y Controles).

Al abrir la carpeta PELF, seleccionaba el grupo de electrodos que pertenecen al área frontal, juntaba todos los electrodos seleccionados en una sola matriz para al final obtener la mediana de todos estos datos.

De igual forma se aplicó esto para las carpetas TDAH y Controles.

Por lo tanto nuestro resultado fueron 3 matrices.

A estas 3 matrices se les aplicó un filtro de pasabanda de 4 a 8 hz.

Después de esto se le agregó la “columna objetivo”. Para los sujetos pertenecientes a Controles se les asignó el número 1, para los del grupo TDAH el 2 y para los otros el 3.

Una vez que ya las matrices tuvieran la columna de objetivo asignada, se unen estas 3 para así obtener una sola matriz.

Después se estandarizaron los datos con los que se iban a trabajar para aplicar los distintos modelos.

Resultados de las

	KNN	SVM
Promedio	33.33%	36.36%
Std	33.33%	36.36%
Median	33.33%	45.45%

pruebas 13,14 y 15.

Procedimiento por pasos.

1. Selección de sujetos
2. Selección de área:
Frontal/temporal/parietal
3. Filtro o transformada:
Pasabanda/ Parabanda/ Wavelet
4. Agregar columna objetivo a cada matriz
5. Unir las matrices para crear una sola
6. Segmentar los datos de prueba y entrenamiento
7. Estandarizar los datos
8. Aplicación del modelo:
Logit/KNN/SVM/Árbol de decisión