

Priorización Automática de Neuroimágenes en Resonancia Magnética.

Automatic Prioritization of Neuroimaging in Magnetic Resonance

**Servicios de Salud San Vicente Fundación
Universidad EAFIT**

Estudiante

Danny Styvens Cardona Pineda
Estadístico, MSc Epidemiología

Trabajo para optar al título
Magister en Ciencia de Datos y Analítica

Asesor

Henry Laniado Rodas
Doctor en Ingeniería Matemática

Co - asesor

Juan Gabriel Castrillón Guzmán
Doctor en *Medical Life Science and Technology*

Maestría en Ciencia de los Datos y Analítica
Facultades de Ciencias, Ingeniería y Economía
Universidad EAFIT
Medellín
2020

CONTENIDO

1	INTRODUCCIÓN	7
2	MARCO TEÓRICO	8
2.1	Resonancia magnética y neuroimágenes	8
2.2	Simetría entre hemisferios del cerebro humano	8
2.3	Coeficiente de correlacion cruzada (NCC).....	10
2.4	BIANCA (<i>Brain Intensity Abnormality Classification Algorithm</i>)	11
2.5	Permutaciones entre hemisferios (HCP).....	12
2.6	Entropía de valores locales permutados entre hemisferios (HPLE).....	13
2.7	Eigenfaces	14
2.8	Modelos de clasificación supervisados.....	16
3	OBJETIVOS	18
3.1	General.....	18
3.2	Específicos.....	18
4	MÉTODOS	19
4.1	Tipo de estudio.....	19
4.2	Población.....	19
4.3	Muestreo	19
4.4	Tamaño de la muestra	19
4.5	Selección de la muestra.....	19
4.5.1	Criterios de inclusión.....	19
4.5.2	Criterios de exclusión.....	19
4.6	Fuentes de información.....	19
4.7	Control de sesgos	20
4.8	Limitaciones del estudio.....	20
4.9	Procesamiento de imágenes.....	20
4.10	Variables del estudio	21
5	Consideraciones éticas	21
6	Resultados	22

6.1	HCP.....	23
6.1.1	HCP con valores de intensidades en el intervalo $[0,1]$	23
6.1.2	HCP con mapas de probabilidad.....	26
6.2	HPLE.....	28
6.3	Eigenfaces	31
6.3.1	Eigenfaces por hemisferios y grupo (sanos, patológicos) con valores de intensidades en $[0,1]$ (enfoque A).....	32
6.3.2	Eigenfaces con NCC (enfoque B)	33
6.3.3	Eigenfaces con mapa de probabilidades Bianca (enfoque B)	35
7	Discusión	37
8	Conclusiones	40
9	Código.....	40
10	Propuesta de otras líneas de trabajo	40
11	Referencias	41

ÍNDICE DE TABLAS

Tabla 1 Variables en el estudio	21
Tabla 2. Medidas de resumen del test HCP de sanos y patológicos.....	25
Tabla 3. Matriz de confusión según la distribución del test HCP	26
Tabla 4. Medidas de resumen del test HCP de sanos y patológicos - Bianca.....	28
Tabla 5. Matriz de confusión según la distribución del test HCP - Bianca.....	28
Tabla 6. Capacidad predictiva del HPLE según niveles de significancia, aplicando prueba U de Mann-Whitney.	29
Tabla 7. Capacidad predictiva del HPLE según niveles de significancia, aplicando la prueba t de student.	30
Tabla 8. Desempeño del modelo SVM sobre datos de prueba, Eigenfaces enfoque A.	33
Tabla 9. Desempeño de modelos sobre datos de prueba, Eigenfaces enfoque B.....	34
Tabla 10. Desempeño de modelos sobre datos de prueba, Eigenfaces-Bianca enfoque B.	36
Tabla 11. Desempeño de modelo SVM sobre datos de prueba, reteniendo el 90% de variabilidad en Eigenfaces.....	39

ÍNDICE DE FIGURAS

Figura 1. Mapa de probabilidades por hemisferio obtenidos con Bianca, en una imagen de un paciente diagnosticado con Oligodendroglioma.....	11
Figura 2. Selección de regiones en ambos hemisferios para calculo del HPLE, sin hacer permutaciones.	13
Figura 3. Selección de regiones en ambos hemisferios para calculo del HPLE, haciendo permutaciones.	13
Figura 4. Procesamiento de neuroimágenes de MRI.....	20
Figura 5. Distribución de las diferencias entre hemisferios ajustadas, según el grupo.	23
Figura 6. Distribución del test HCP en 20 neuroimágenes de pacientes sanos.	24
Figura 7. Distribución del test HCP en 15 neuroimágenes de pacientes patológicos.....	24
Figura 8. Distribución del test HCP en 20 neuroimágenes de pacientes sanos - Bianca.....	27
Figura 9. Distribución del test HCP en 15 neuroimágenes de pacientes patológicas - Bianca.	27
Figura 10. Distribución de HPLE entre hemisferios, según el grupo.	29
Figura 11. Tendencia de indicadores de desempeño del HPLE según el nivel de significancia, aplicando prueba U de Mann-Whitney.	30
Figura 12. Tendencia de indicadores de desempeño del HPLE según el nivel de significancia, aplicando prueba t de student.....	31
Figura 13. Diferencia entre hemisferios de las distancia al subespacio.....	32
Figura 14. Validación cruzada de modelos de clasificaicón, Eigenfaces enfoque A.....	32
Figura 15. Validación cruzada de modelos de clasificaicón, Eigenfaces enfoque B.....	34
Figura 16. Capacidad predictiva del modelo SVM, Eigenfaces enfoque B.....	35
Figura 17. Validación cruzada de modelos de clasificaicón, Eigenfaces-Bianca enfoque B.	35
Figura 18. Capacidad predictiva de modelo SVM y NB, Eigenfaces-Bianca enfoque B.	37
Figura 19. Capacidad predictiva de modelo SVM, reteniendo un 90% de variabilidad con Eigenfaces.	39

RESUMEN

Introducción

Partiendo de la hipótesis de simetría entre los hemisferios del cerebro, se proponen medidas de similaridad para la detección entre un cerebro sano y uno patológico a partir de neuroimágenes. Se exploran tres medidas, la primera es un índice de correlación cruzada (NCC), el segundo es un test basado en permutaciones entre hemisferios (HCP) y el tercero se trata de la medición de la entropía de valores locales permutados entre hemisferios (HPLE). También se calcularon mapas de probabilidades con Bianca para usarlo como input de las medidas de similaridad. Además, se aplicó la técnica Eigenfaces sobre el mapa NCC y el mapa de probabilidades, con este método se obtuvieron subespacios con las principales características del conjunto de imágenes. Finalmente se ajustaron modelos de clasificación para la predicción de las categorías sano/patológico.

Resultados

El HCP no logró capturar la suficiente información de las imágenes como para distinguir entre sanos y patológicos, con una baja detección de patológicos y sanos (61% y 70% respectivamente). Con el HPLE se observaron mejores resultados comparado con el HCP, con un f1-score igual a 64% y un *recall* igual al 78% con un nivel de significancia del 5%, sin embargo, no es suficiente aun para sugerir que se puede usar en la práctica clínica a menos que se le hagan mejoras.

Con el método de Eigenfaces se obtuvieron los mejores resultados, especialmente cuando se trabajó con el mapa de correlaciones NCC en lugar del mapa de probabilidades. El modelo con mejor rendimiento y poder predictivo fue el SVM con un 98% de capacidad predictiva.

Conclusiones

El uso del mapa de correlaciones NCC obtenido de una secuencia FLAIR para medir similaridad entre hemisferios, resultó ser útil para la predicción de estudios sanos/patológicos, después de generar un subespacio con el método Eigenfaces y aplicar un modelo de clasificación de máquinas de soporte vectorial. Estos métodos aplicados en conjunto permitieron hacer una priorización adecuada con un 98% de capacidad predictiva.

Palabras claves: Mapa de correlaciones, Bianca, permutaciones, entropía, Eigenfaces, modelo de clasificación.

SUMMARY

Introduction

According to the hypothesis of symmetry between brain hemispheres, similarity measures are proposed for the detection of healthy and pathological brain, with neuroimaging. Three measures are explored, the first is a cross-correlation index (NCC), the second is a test based on permutations between hemispheres (HCP) and the third is the measurement of entropy of local values permuted between hemispheres (HPLE) . Probability maps with Bianca were also calculated to use as an input of similarity measures. In addition, the Eigenfaces technique was applied on the NCC map and the probability map, with this method subspaces were obtained with the main characteristics of the image set. Finally, classification models were adjusted for the prediction of the healthy/pathological categories.

Results

The HCP was unable to capture enough information from the images to distinguish between healthy and pathological, with low detection of pathological and healthy (61% and 70% respectively). With HPLE, better results were observed compared to HCP, with an f1-score equal to 64% and a recall equal to 78% with a significance level of 5%, however, it is not enough yet to suggest that it can be used in clinical practice unless improvements are made.

With the Eigenfaces method, the best results were obtained, especially when working with the NCC correlation map instead of the probability map. The model with the best performance and predictive power was the SVM with 98% predictive capacity.

Conclusions

The use of the NCC correlation map obtained from a FLAIR sequence to measure the similarity between hemispheres is useful for the prediction of healthy/pathological images, after generating a subspace with the Eigenfaces method and applying a vector support machine classification model. These methods applied together allowed adequate prioritization with 98% predictive capacity.

Keywords: Correlation map, Bianca, permutations, entropy, Eigenfaces, classification model.

1 INTRODUCCIÓN

El cerebro humano es un universo fascinante que esta siendo explorado hace décadas, en esta exploración, personas con una gran visión y duro trabajo han logrado descifrar aspectos importante sobre funciones relacionadas con regiones específicas y detalles anatómicos de gran impacto para la ciencia (1,2). Aun así, todavía falta mucho por descubrir y aprender sobre este excepcional órgano.

Gracias al desarrollo de tecnologías como la resonancia magnética (RM) se tienen nuevas opciones para la investigación y el diagnóstico en el área de neurología. Una de las ventajas de RM comparados con otras tecnologías como la tomografía computada, es que no se expone a los pacientes a radiación ionizante, además de su utilidad clínica como ayuda diagnóstica.

Este trabajo se enfoca en la aplicación de conceptos basados en conocimiento anatómico del cerebro, específicamente, se parte de un concepto de simetría aproximada entre hemisferios del cerebro humano. Bajo esta “sencilla” pero potente premisa se emprende una búsqueda de características cuantitativos en las imágenes que permitan obtener una ayuda para la priorización clínica de manera automática de neuroimágenes en resonancia magnética. Se incluyen imágenes de pacientes sanos y enfermos (tumores, eventos cerebrovasculares, enfermedad de sustancia blanca).

Partiendo de la hipótesis de simetría entre los hemisferios del cerebro, se proponen algunas mediadas de similaridad para la detección entre un cerebro sano y uno patológico a partir de imágenes. Se exploran tres medias, la primera es un índice de correlación cruzada (NCC) con el que se construye un mapa de correlaciones, el segundo es un test basado en permutaciones entre hemisferios (HCP) y el tercero se trata de la medición de la entropía de valores locales permutados entre hemisferios (HPLE). También se calcularon mapas de probabilidades con un método llamado Bianca (3) para usarlo como input de las medidas de similaridad, y se compararon con los resultados obtenidos con las intensidades extraídas de la secuencia de recuperación de la inversión atenuada de fluido (FLAIR).

Con el fin de probar si la extracción de características adicionales a las consideradas con el HCP y HPLE, permiten obtener mejores resultados, se aplico la técnica Eigenfaces (4) sobre el mapa NCC y el mapa de probabilidades. Con este método se obtuvieron subespacios con las principales características del conjunto de imágenes. Finalmente se ajustaron modelos de clasificación para la predicción de las categorías sano/patológico.

Previo a la aplicación de métodos y pruebas que se acaban de mencionar, se hizo un procesamiento de las imágenes DICOM. El procesamiento de las imagines consistió en la extracción del cerebro, corrección por movimientos y normalización

al espacio estándar MNI 152 (5,6). Adicionalmente se extrajo cada hemisferio del cerebro.

2 MARCO TEÓRICO

2.1 Resonancia magnética y neuroimágenes

Las imágenes de resonancia magnética (MRI por sus siglas en ingles) tiene sus fundamentos en la resonancia magnética nuclear; en términos generales, lo que pasa en el proceso es que, a partir de la generación de un campo magnético de gran intensidad y señales externas al campo, se excitan ciertos átomos en el organismo, los cuales absorben y emiten energía, luego las señales de excitación pueden ser procesadas para obtener “mapas” de la estructura del cuerpo estudiada. Hay dos conceptos importantes en resonancia magnética (RM), el tiempo de relajación característico (T1) y el tiempo característico constante (T2; a grandes rasgos se podría decir que la secuencia T1 es el tiempo de recuperación de una componente de la magnetización, y la secuencia T2 el tiempo de recuperación (7).

Las neuroimágenes son especialmente relevantes dentro de la RM porque han permitido estudiar de manera más amplia el cerebro humano; estas se dividen en dos tipos, las estructurales y las funcionales; con las estructurales se busca analizar el cerebro más desde un punto de vista anatómico, mientras que en el segundo tipo el foco es la funcionalidad (áreas o redes de activación y flujos)(8).

Dentro de las patologías o hallazgos en las que se usan las neuroimágenes estructurales como ayuda diagnóstica se encuentran los tumores, epilepsia, hemorragia cerebral, esquizofrenia, entre otros (8). Las resonancias magnéticas funcionales (fMRI) por su lado han permitido detectar posibles áreas afectadas por una patología o un trauma, esto es de mucha ayuda para la práctica clínica ya que permite identificar el correcto funcionamiento o chequeo de funciones de algunas de las áreas o redes del cerebro (9,10).

2.2 Simetría entre hemisferios del cerebro humano

Algunos estudios han abordado la detección de tumores y otros hallazgos mediante segmentación y predicción automática (11). En este trabajo se incluyen conocimientos anatómico y clínicos en los análisis (12), con lo que se busca construir modelos matemáticos para la detección de imágenes sospechosas, una vez detectadas, se podría hacer una priorización automática dentro de los sistemas de información para que sea vista rápidamente por el personal médico.

Se propone medir la simetría entre los hemisferios derecho e izquierdo como indicador de cambios anatómicos que pueden sugerir lesiones y/o trastornos neurológicos (12–14).

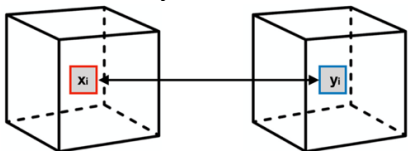
Algunos autores analizan la asimetría entre estructuras de hemisferios del cerebro en términos de dimensiones (largo, ancho, alto y radios), encontrando posibles diferencias, aunque estos resultados no logran ser consistentes aun (15). Sin embargo, en este trabajo se propone abordar el problema como una relación entre los píxeles de cada hemisferio, o en su defecto, entre regiones locales, pero en ambos casos manteniendo la relación espacial como factor de comparación, es decir que se hace necesario hacer un procesamiento de las imágenes que permita sobreponer un hemisferio sobre otro (*flip*) para asegurar que se está en un mismo espacio. Por supuesto, llevando antes las imágenes a lo que se conoce en neuroimágenes como un espacio estándar, de manera que permita luego comparar resultados entre imágenes de cerebros que provienen de distintos sujetos; de esta manera se evitan ambigüedades que puedan resultar de basarse solo en la forma del cerebro (16).

Como se menciona, dadas las diferencias anatómicas entre cerebros es necesario realizar una extracción del cerebro de la imagen original (5) (se excluye el tejido no cerebral) y una normalización que permita hacer comparaciones entre cerebros de varios individuos (6). Después de realizar esta normalización las relaciones de simetría entre hemisferios podrían variar entre cerebros. Por este motivo se propone una medida “intrínseca” de cada cerebro. Básicamente se busca conocer cual es la relación entre valores de píxeles entre ambos hemisferios de cada individuo para tratar de identificar anomalías; note que bajo este enfoque podrían existir algunas variaciones de ambas estructuras que serían biológicamente plausibles a partir de las mediciones que se le hagan a los sujetos sanos, aun así, las variaciones en esta medida no es un motivo de preocupación en esta investigación ya que lo que interesa es la relación intrínseca, en resumen, lo que se propone es tratar la relación entre hemisferios como una variable aleatoria que puede ser modelada y cuyos diferencias por individuo dependen exclusivamente de las características propias de este, una vez halla sido normalizado.

Otro aspecto importante y que es fundamental son las secuencias usadas para obtener las imágenes, ya que de esto dependen los valores de los píxeles obtenidas. En este estudio se usan dos secuencias, la Flair y la T1. Con la Flair se busca obtener valores de hiperintensidades que puedan indicar algún tipo de anomalía, y la secuencia T1 se uso para delimitar con mayor precisión el espacio descrito por cada cerebro para generar mapas de probabilidad con el método BIANCA (*Brain Intensity Abnormality Classification Algorithm*), el cual es descrito en la sección 2.4

2.3 Coeficiente de correlacion cruzada (NCC)

El coeficiente de correlación cruzada (NCC - *Normalized Cross Correlation*) ha sido ampliamente usado como una métrica para medir la similaridad entre imágenes (17); esta medida aplica básicamente el mismo concepto del coeficiente de correlación de Pearson. Para este proyecto se busca crear un mapa de correlaciones entre hemisferios, de manera que se compara píxel a píxel, donde las estimaciones de tendencia central y de dispersión son calculados a partir de los valores de los pixeles de cada hemisferio respectivamente

$$NCC = \frac{1}{N} \sum_i \frac{(x_i - \bar{x})(y_i - \bar{y})}{\sigma_x \sigma_y}$$


Donde x_i, y_i son voxels de los volúmenes de ambos hemisferios y $\bar{x}, \bar{y}, \sigma_x, \sigma_y$ son sus respectivas medias y varianzas.

El NCC puede interpretarse como la relación lineal entre la intensidad de voxels de dos volúmenes, en este caso, se trata de volúmenes que corresponden al hemisferio derecho e izquierdo

Es posible obtener una versión robusta del NCC usando la mediana en lugar del promedio (18); en este trabajo no se hace la prueba, pero en futuros trabajo puede resultar de interés calcular esta versión robusta del NCC y comparar los resultados, además de evaluar costo computacional.

Lo que se busca con el NCC es calcular un mapa de correlaciones en el mismo espacio del cerebro, para luego construir una imagen de cada hemisferio que puedan ser comparadas en magnitud-ubicación (píxel-ubicación). Para hacer esto se plantean los siguientes pasos:

1. Generación de un espacio común de comparación para ambos hemisferios con el cual puedan ser comparados luego cada uno por separado. En este caso el espacio en común es la correlación cruzada directa entre ambos hemisferios
2. Estimación de mapa de correlación del hemisferio derecho, comparado con el espacio base obtenido en 1.
3. Estimación de mapa de correlación del hemisferio izquierdo, comparado con el espacio base obtenido en 1.

Lo que se espera de esta medida es que entre más simetría existe entre los valores de las intensidades obtenidas con la secuencia Flair, más parecidas deben ser los mapas de ambos hemisferios. Es importante mencionar que es posible considerar el mapa NCC completo del cerebro al unir los mapas de cada hemisferio, de echo, este ejercicio se realiza con el fin de averiguar si el espacio NCC completo permite detectar entre un cerebro sano y uno enfermo, después de aplicar métodos estadísticos.

2.4 BIANCA (*Brain Intensity Abnormality Classification Algorithm*)

Lo que se ha mencionado hasta ahora solo considera trabajar con intensidades obtenidas de las neuroimágenes, sin embargo, resulta interesante observar cuales serían los resultados si se usan mapas de probabilidad en lugar de las intensidades crudas. Con este enfoque se quiere evaluar si los mapas de probabilidad proporcionan datos más precisos al momento de comparar hemisferios de imágenes de pacientes sanos y patológicos.

Con este algoritmo se busca detectar cuan probable es que un valor de un píxel corresponda a una hiperintensidad sospechosa, basado en sus k-vecinos más próximos. Para esto se hace un entrenamiento con imágenes marcadas manualmente (se marcan las regiones patológicas de una neuroimagen slide por slide. Se trata de un algoritmo supervisado), para este entrenamiento se requiere entre 10 y 20 imágenes con marcación. Este algoritmo esta implementado en la biblioteca FSL (3). En la siguiente figura se puede observar un mapa de probabilidades obtenido con Bianca, después de aplicar el algoritmo a una imagen de un paciente diagnosticado con Oligodendroglioma.

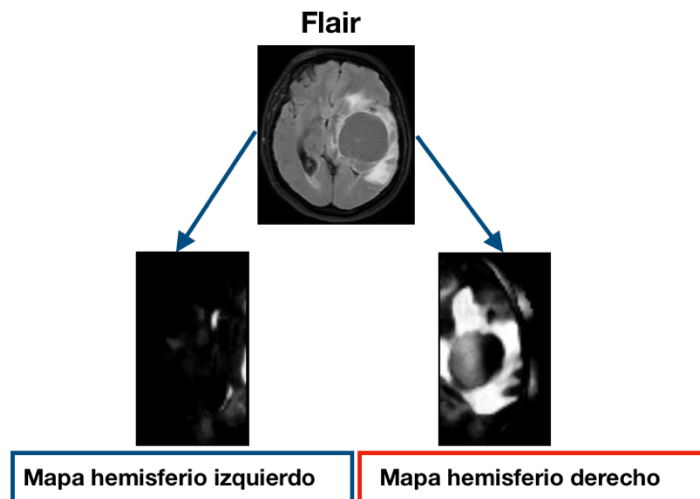


Figura 1. Mapa de probabilidades por hemisferio obtenidos con Bianca, en una imagen de un paciente diagnosticado con Oligodendroglioma.

En este trabajo se hizo un entrenamiento con 20 imágenes marcas por una médica radióloga con 9 años de experiencia en neuroimágenes.

2.5 Permutaciones entre hemisferios (HCP)

Se construyo una medida para comparar ambos hemisferios, en este caso se probó con los pixeles en una escala entre cero y uno, y también se prueba con el mapa de probabilidades obtenido con Bianca. En ambos casos se construyó de la siguiente manera:

1. Sea n el número de pixeles contenido en cada hemisferio, luego se genera una muestra aleatoria de números en el intervalo $[1, n]$ con reemplazo, donde finalmente solo permanecen los valores únicos.
2. Se seleccionan en las imágenes de ambos hemisferios las posiciones generadas en el paso 1, después de haber colapsado las imágenes en un vector n -dimensional de $x*y*z$, donde x, y, z corresponde a las dimensiones originales de las imágenes de cada hemisferio.
3. Los pixeles seleccionados para cada hemisferio en el paso 2, son intercambiados, obteniendo un par de vectores H_R, H_L
4. Se calcula la norma L a los vectores obtenidos en 3, y se construye la razón $r(H_R, H_L) = L(H_R)/L(H_L)$
5. Se Repite los pasos 1 a 5, k veces.

Se espera que $r(H_R, H_L) \approx 1$ bajo la hipótesis de que ambos hemisferios son aproximadamente simétricos.

Lo que se busca es construir un test que se vale de permutaciones (19) para encontrar la distribución bajo la hipótesis nula $r(H_R, H_L) = 1$, para esto se hace un intercambio aproximadamente del 50% de los datos entre hemisferios, k -veces; evitando así que las diferencias que se observen en el test sean debidas al azar, es decir, es menos probable que hiperintensidades de la misma magnitud ubicadas en diferentes regiones de los hemisferios den como resultado que son simétricos, cuando en realidad no lo son. Este paso de intercambio entre valores de hemisferios que coincidan en las mismas posiciones le permite integrar al test el concepto de ubicación espacial de manera implícita.

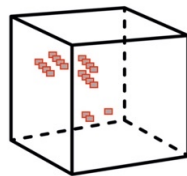
Aunque es relativamente sencillo el calculo del test, su costo computacional es alto si se considera hacer el test para una muestra alta de imágenes, a pesar que el algoritmo descrito podría ser escrita para que parte de los cálculos se hagan de manera paralela para mejorar su performance (20).

Para tratar de mejorar la detección de diferencias entre hemisferios, también se aplico el test solo en aquellos pixeles que eran diferentes de cero.

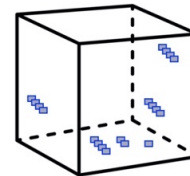
2.6 Entropía de valores locales permutados entre hemisferios (HPLE)

En el HCP no se consideran efectos locales o de vecindario (se hizo pixel a pixel); non el HPLE se busca mejora este aspecto al comparar los valores de intensidad de manera local por medio de los niveles de entropía después de haber permutado pixeles locales de cada hemisferio. La hipótesis bajo la que se construye es que en un cerebro sano, de manera local, la distribución de los valores observados de pixeles en cada hemisferio deben tener valores de entropía similares, mientras que en un cerebro enfermo puede no ocurrir esto. Un aspecto importante de esta prueba son las permutaciones que se realizan previo al calculo de la entropía; su importancia radica en que se le da mayor capacidad de encontrar diferencias al HPLE.

Considere que se seleccionan valores en cada hemisferio como se muestra en la figura 2, y suponga que se quiere calcular el valor de la entropía sin hacer permutaciones



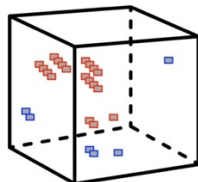
Hemisferio Derecho



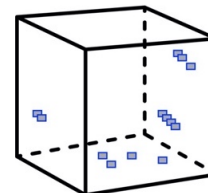
Hemisferio Izquierdo

Figura 2. Selección de regiones en ambos hemisferios para calculo del HPLE, sin hacer permutaciones.

Observe que la entropía calculada puede ser igual, aunque la distribución de pixeles sea espacialmente diferente. Ahora, supongo que se permutan los valores y se vuelve a calcular la entropía, en este caso la entropía daría valores diferentes (figura 3).



Hemisferio Derecho



Hemisferio Izquierdo

Figura 3. Selección de regiones en ambos hemisferios para calculo del HPLE, haciendo permutaciones.

Se propone un procedimiento sencillo para determinar las probabilidades de manera discreta:

- a. Calcule el valor de percentiles de los pixeles (por ejemplo, los percentiles 25, 50 y 75).
- b. Considere estos valores como valores referencia de su distribución.
- c. Repite $100/n_{percentiles}$ cada valor de referencia.
- d. Finalmente, calcule la probabilidad de cada valor referencia con los valores obtenidos en c.

Es importante mencionar que en este trabajo se evitó incluir valores extremos como valores de referencia, por lo que se usaron los percentiles 25, 50 y 75.

Después de describir la forma en que se calcularon las probabilidades, a continuación, se listan los pasos seguidos para la estimación de la entropía:

1. Identificación de pixeles diferentes de cero en alguno de los dos hemisferios.
2. Selección los valores alrededor de los pixeles identificados en el paso 1 con una extensión igual en cada dimensión, en este caso se tienen 3 dimensiones y las subregiones que se seleccionaron son de tamaño 27 (3x3x3).
3. Realizar las permutaciones de valores entre cada subregión (hay que recordar que se están comparando subregiones entre hemisferios que están en la misma ubicación espacial).
4. Calcular las probabilidades como se describió previamente
5. Finalmente se estima la entropía propuesta por Shannnon (21)

$$H(X) = - \sum_i p(x_i) \log_2 p(x_i)$$

Algo interesante del HPLE como indicador de similaridad, es que usa la incertidumbre como medida de comparación de hiperintensidades entre hemisferios en regiones locales, lo que puede arrojar información adicional si se compara con el test HCP. Además, los valores obtenidos en cada hemisferio pueden ser tratado como variables aleatorias con el fin de realizar pruebas estadísticas para detectar diferencias.

2.7 Eigenfaces

Eigenfaces es un método que fue propuesto para tratar de reconocer imágenes de rostros humanos, es esencialmente una variación del método de componentes principales (PCA por sus siglas en ingles) con el que se busca encontrar un

subespacio que resuma de la mejor manera el espacio generado por el conjunto de imágenes de la muestra (22,23), en este caso se trata de neuroimágenes.

Uno de los problemas de usar PCA sobre las imágenes, es que este trabaja directamente sobre la matriz de covarianzas generada por los datos, la cual en el procesamiento de imágenes puede alcanzar dimensiones bastante altas. Considere una imagen en 2 dimensiones de tamaño $N \times N$, por ejemplo, 250×250 píxeles, luego en un conjunto de imágenes del mismo tamaño cada imagen está definida en un espacio de dimensión 62.500, por lo que la manipulación de la matriz de covarianzas requerida para los análisis de PCA se hace computacionalmente inviable con dimensiones 62.500×62.500 .

Como alternativa, con Eigenfaces se propone extraer información relevante de un conjunto de imágenes, pero en este caso se hace a partir de las M imágenes que se tenga y no del número de dimensiones descrito por la imagen. Este es un salto importante ya que reduce los cálculos necesarios en el caso que $M < N$ (que es en la mayoría de las veces). A continuación se describe de manera breve los principios y pasos a seguir para obtener los subespacios mencionados (4) :

1. Calcular la imagen media o mediana de las neuroimágenes, ψ
2. Representar cada imagen en un vector n -dimensional Γ_i , donde n es el producto de las dimensiones x, y, z (colapsar las imágenes)
3. Centralizar cada imagen $\Phi_i = \Gamma_i - \psi$
4. Calcular la matriz de covarianzas:
 Sea $A = [\Phi_1 \ \Phi_2 \ \Phi_3 \ \dots \ \Phi_M]$, luego
 $C = A^T A$
5. Encontrar los valores y vectores propios asociados a C , tal que
 $A^T A v_i = \mu_i v_i$
6. Seleccionar los vectores propios asociados a los k valores propios más altos.

Puede decirse que las imágenes que forman el subespacio son las “características” más representativas construidas a partir de las imágenes originales, con las que se puede reconstruir las imágenes originales con un alto grado de precisión; entendiendo la precisión como el grado de variabilidad explicado por el subespacio (mismo concepto usado en PCA) (23). Igual que en PCA se busca encontrar unos vectores y valores propios para encontrar el mejor subespacio, en este caso reciben el nombre de *eigenfaces* y *eigenvalues*, como es habitual los *eigenvalues* indican el grado de variabilidad capturado por cada *eigenfaces*, donde cada *eigenfaces* corresponde a una imagen que es combinación lineal de las imágenes originales. Observe que el número posible de vectores propios es igual al número de imágenes en la muestra y no el número de dimensiones definido por las imágenes.

En el presente trabajo se uso el método de eigenfaces desde dos perspectivas para tratar de evaluar en que casos es más apropiado usarlo para determinar si una imagen proviene de un cerebro sano o enfermo:

- A. Se estima un subespacio en cada hemisferio para sanos y otro para enfermos, luego se calculan las distancias de las proyecciones de cada hemisferio a su respectivo subespacio y finalmente se calculan las diferencias de las distancias. Bajo este enfoque se espera que las diferencias de las distancias de los sanos sean menores que la de los enfermos. Aquí se trabajo directamente con los valores de pixeles de cada hemisferio, pero también es posible hacerlo con los mapas de probabilidad obtenidos con Bianca.
- B. También se trabaja con el mapa de correlaciones generado con el NCC, observe que el espacio del NCC esta en el mismo espacio de la neuroimagen original, pero esta fue construida por hemisferios (ver sección 2.32.3). Con estos valores se construye un subespacio para las imágenes de pacientes sanos, luego se proyectan ambos conjuntos de datos al subespacio, y con estas proyecciones se construye un modelo de clasificación para predecir si se trata de una imagen con signos de patología o no. Este mismo ejercicio se hizo también construyendo el subespacio con los patológicas.

En la literatura se han propuesto otros métodos lineales y no lineales para la reducción de dimensionalidad en imágenes (24), sin embargo en este estudio solo se uso el método de *eigenfaces*.

2.8 Modelos de clasificación supervisados

Los modelos de clasificación han sido usados en problemas de predicción en diversos campos de estudio (25); con la introducción de nuevos y mejorados algoritmos de optimización y más capacidad de computo, se tienen disponibles varios tipos de modelos en paquetes como R y Python. Entre ellos se encuentra el modelo logístico, arboles de decisión, random forest, maquinas de soporte vectorial, Naive Bayes, redes neuronales, entre otros.

La filosofía detrás de los modelos de clasificación supervisados es que, al tener un conjunto de datos etiquetados con la clasificación verdadera, se trata de construir un modelo que sea capaz de detectar patrones en los datos que le permitan distinguir entre los grupos definidos por las etiquetas. Para la detección de estos patrones o características, en cada tipo de modelo en particular se construyen funciones que deben ser optimizadas para estimar los parámetros propuestos (26).

Lo que se conoce hoy como *Machine Learning*, se podría decir que es una extensión de los modelos estadísticos convencionales, pero con variaciones y mejoras en los algoritmos para resolver las funciones de pérdida, por ejemplo, el algoritmo de gradiente descendiente (27). Otro de los grandes hitos en *Machine Learning* unido a la capacidad computacional es el desarrollo de redes neuronales como herramienta de predicción, consiguiendo construir topologías bastante complejas con millones de parámetros que permiten resolver problemas que antes eran intratables (28).

Es claro que la era del Machine Learning es la era de las predicciones, pero poco se ha desarrollado para resolver problemas de causalidad (¿cuáles son los factores causales y cuáles son sus pesos), por lo que aun queda mucho por trabajar en esta línea, además de mejorar los resultados obtenidos hasta ahora con los modelos no supervisados.

En el presente trabajo se usan los siguientes modelos de clasificación para tratar de predecir si una neuroimagen proviene de un cerebro sano o uno patológico: Logístico, análisis discriminante, árboles de decisión, random forest, máquinas de soporte vectorial, Adaboosting, Stochastic Boosting y Naive Bayes.

Para evaluar la capacidad predictiva se usaron cinco indicadores:

- **Accuracy:** Media global de lo correctamente predicho, se puede presentar como proporción o porcentaje.
- **Precisión:** Es la razón entre los casos positivos predichos correctamente y el total de casos predichos como positivos.
- **Recall:** Razón entre los casos positivos predichos correctamente y el total de casos positivos en la muestra.
- **F1-score:** Es un promedio ponderado del recall y la precisión.
- **Área bajo la curva roc:** indicador que mide la capacidad predictiva de un modelo.

En todos los casos se dividieron los datos en dos conjuntos, uno para el entrenamiento (70%) y otro de prueba (30%), también se estandarizaron las variables que se ingresaron a los modelos.

3 OBJETIVOS

3.1 General

Predecir la prioridad clínica de neuroimágenes en resonancia magnética mediante modelos estadísticos que puedan ser implementados en un ambiente de producción.

3.2 Específicos

- ◆ Adecuar un servidor dentro de la red del hospital donde se hace el estudio.
- ◆ Establecer una conexión a las bases de datos de imágenes
- ◆ Establecer flujo de trabajo para la extracción y procesamiento de imágenes.
- ◆ Realizar marcación de las imágenes para entrenamiento con Bianca.
- ◆ Procesar imágenes para la aplicación de Bianca.
- ◆ Calcular NCC con mapa de probabilidades Bianca
- ◆ Aplicar HCP, HPLE y Eigenfaces
- ◆ Ajustar y seleccionar modelo con mejor poder predictivo.

4 MÉTODOS

4.1 Tipo de estudio

Es un estudio de corte transversal.

4.2 Población

La población son las imágenes de resonancia magnética de la especialidad de neurología que han sido adquiridas en Servicios de Salud - San Vicente Fundación entre los años 2015 y 2019. Se cuenta con aproximadamente 8000 neuroimágenes.

4.3 Muestreo

Se toma un conjunto de imágenes para las que se tenía disponible una clasificación hecha por radiólogos expertos en neurología, por lo que se trata de un muestreo por conveniencia.

4.4 Tamaño de la muestra

En total se cuenta con 369 neuroimágenes (212 sanos, 157 patológicas). Para los mapas de probabilidades calculado con Bianca, de las 369 imágenes se trabajó con 271 (196 sanos y 75 patológicos).

4.5 Selección de la muestra

La selección de la muestra se hizo de manera aleatoria, aunque el tamaño de la muestra es obtenido de manera no probabilística.

4.5.1 Criterios de inclusión

Se incluyen estudios de neuroimágenes de cráneo que cuentan con las secuencias T1 y Flair.

4.5.2 Criterios de exclusión

Se excluyen aquellos estudios que han tenido intervenciones quirúrgicas previas a la toma de las imágenes, tampoco se consideran aquellas imágenes que presentan artefactos.

4.6 Fuentes de información

Se usa una fuente secundaria de datos la cual consiste en imágenes alojadas en un servidor de Servicios de Salud – San Vicente Fundación en un PACS en formato DICOM.

4.7 Control de sesgos

Para evitar sesgos de selección se incluyen estudios con diferentes tipos de tumores, enfermedad cerebrovascular y enfermedad de sustancia blanca, buscando que el modelo de predicción aprenda a detectar anomalías que provengan de diferentes patrones.

También se incluye un 42% de imágenes patológicas en la muestra para evitar desbalanceo en la muestra.

4.8 Limitaciones del estudio

Este estudio solo contempla la clasificación en dos categorías, sano o patológicas, no se trabaja con escalas ordinales para la priorización, lo cual puede ser importante abordarlo en otros estudios.

Tampoco se cuenta con un tamaño de muestra estimado bajo principios probabilísticos, pero para efectos de predicción de imágenes patológicas (si/no) el tamaño de la muestra es adecuado si se considera que un buen modelo debería por lo menos tener una capacidad predictiva del 80%, con lo que se obtendría un tamaño de muestra igual a 239 imágenes ($N=8.000$, $p=0.8$, nivel de confianza=95% y un error esperado del 5%); por lo que este no es un problema de fondo, aunque en futuros estudios se recomienda aumentar el tamaño de muestra y evaluar de nuevo los métodos propuestos.

4.9 Procesamiento de imágenes

Las neuroimágenes fueron tomadas en resonadores Philips Ingenia de 1.5 y 3.0 Tesla, el procesamiento o preparación de las imágenes fueron realizadas con la biblioteca FSL. El procesamiento de las imágenes consiste en la extracción del cerebro (excluyendo tejido que no es parte del cerebro y hueso), una corrección por movimientos y una normalización al espacio estándar MNI 152 (5,6). Adicionalmente se extra cada hemisferio del cerebro.

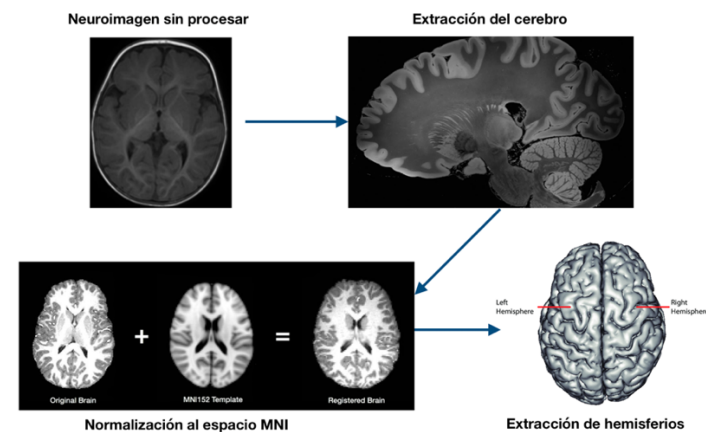


Figura 4. Procesamiento de neuroimágenes de MRI.

4.10 Variables del estudio

Tabla 1 Variables en el estudio

Clasificación	Variable	Tipo	Nivel de medición	Valores validos	Unidades
Clínicas	Grupo prioridad	Cualitativa	Nominal	Normal patológica	NA
Volúmenes	Hemisferio derecho	Cuantitativa	Escala	[0,1000]	mm3
	Hemisferio izquierdo	Cuantitativa	Escala	[0,1000]	
	Mapa de probabilidades-Bianca	Cuantitativa	Escala	[0,1]	
	Mapa de correlaciones - NCC	Cuantitativa	Escala	[0,1]	
	HCP	Cuantitativa	Escala	[0,∞)	NA
	HPLC	Cuantitativa	Escala	[0,1]	NA

5 Consideraciones éticas

El proyecto considera las normas internacionales y nacionales para la investigación con seres humanos, entre ellos el Código del Tribunal Internacional de Núremberg (29), en el que se exponen los principios básicos para investigación en seres humanos tales como: consentimiento informado, mínimos riesgos, máximos y derecho al retiro voluntario. El informe Belmont (30) que considera los principios de respeto, beneficencia y justicia. La declaración de Helsinki de la Asociación Médica Mundial la cual afirma que el bienestar de la persona prevalece sobre otros intereses y rescata el deber de proteger la vida, la salud, la dignidad, la integridad, el derecho a la autodeterminación, la intimidad y la confidencialidad de la información. De acuerdo con la resolución número 8430 de 1993 de la República de Colombia expedida por el Ministerio de Salud y tal como dicta en el Título II, capítulo 1, artículo 11, esta es una investigación con riesgo mínimo ya que se emplea el registro de datos existentes (fuente de dato secundaria).

Nota: Este estudio se centra en realizar procesamiento de imágenes, aplicar métodos estadísticos y matemáticos en relación con el manejo de los datos obtenidos, para los cuales se siguen respetando los criterios de confidencialidad y de uso científico y académico de la información.

6 Resultados

A continuación se presentan los resultados del proyecto, en primer lugar se muestran los resultados obtenidos con el test de permutaciones entre hemisferios (HCP), luego se presentan lo observa con los valores de entropía de valores locales permutados entre hemisferios (HPLE) y finalmente son presentados los resultados obtenidos con Eigenfaces.

Antes de entrar en detalles, se realiza una exploración inicial de las diferencias en valor absoluto de los hemisferios para cada grupo por separado, en sanos y patológicos. Este primer acercamiento permite dar una idea global de las distribuciones de las diferencias entre hemisferios en cada grupo. Para cada imagen se repitió 1000 veces el siguiente ejercicio:

1. Se genera una muestra aleatoria de n valores con reemplazo, donde solo permanecen los valores únicos, los cuales corresponden a índices de las posiciones de valores en los vectores de intensidades de las imágenes colapsadas.
2. Los índices generados en 1 se seleccionan en cada hemisferio y se calculan el valor absoluto de sus diferencias $D(H_r, H_l)$ (D es un vector de tamaño k , donde k corresponde al tamaño del vector generado en 1)
3. Se calcula la proporción de valores en D que son mayores a cero

$$\hat{p} = \frac{\sum_i I(D_i)}{k}, \quad i = 1, 2, 3, \dots, k$$

donde

$$I(x) = \begin{cases} 1, & x > 0 \\ 0, & x = 0 \end{cases}$$

4. Se calcula el promedio de los valores mayores a cero en D

$$\bar{d} = \frac{\sum_j D_j}{l}, \quad j = 1, 2, 3, \dots, l$$

donde l es la cantidad de elementos tal que $D > 0$

5. Estime $\overline{dp} = \bar{d} + \hat{p}$

Los resultados de estas estimaciones se resumen en la figura 5

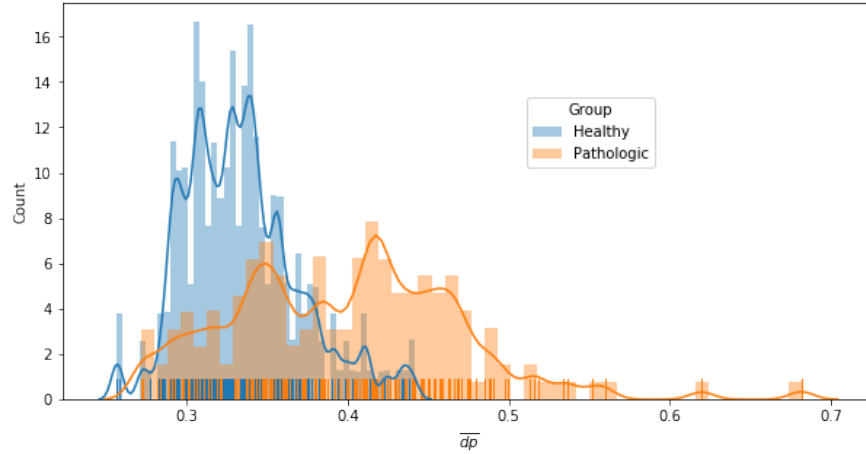


Figura 5. Distribución de las diferencias entre hemisferios ajustadas, según el grupo.

Se observa que los valores de \overline{dp} en el caso de los patológicos tiene una distribución con un rango más amplio con una cola a la derecha (figura 5). Al definir el umbral $tr = \max(\overline{dp}_{healthy})$, se encontró que el 26% de los valores de $\overline{dp}_{pathologic}$ superaron este umbral, es decir que con esta medida se logra diferenciar solo un 26% de los patológicos, el 74% restante se confunde con los valores de los sanos. Este es un resultado importante porque sugiere que se deben encontrar características en los datos que permitan hacer una buena “separación” de ambos grupos, sabiendo que el calculo de una diferencia directa no es suficientes, aunque sea ajustada y/o escalada (al escalar \overline{dp} se obtuvieron resultados similares).

6.1 HCP

EL HCP se realizo con valores de intensidades normalizados en el intervalo $[0,1]$ y con el mapa de probabilidades obtenido con Bianca.

6.1.1 HCP con valores de intensidades en el intervalo $[0,1]$

Para cada imagen se obtuvieron 500 valores del test $r(H_R, H_L) = L(H_R)/L(H_L)$ como se menciona en la sección 2.5, usando la norma L_1

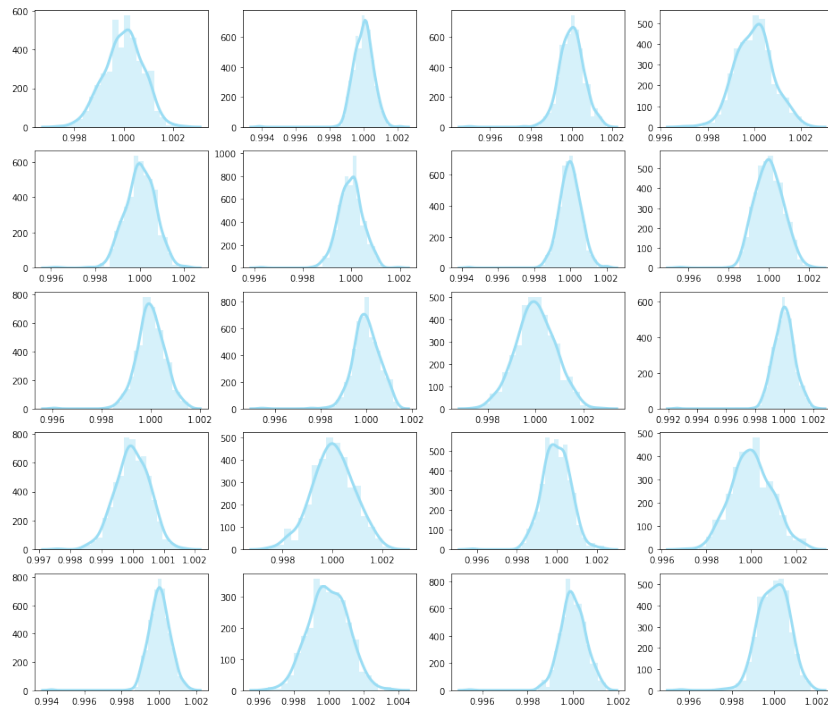


Figura 6. Distribución del test HCP en 20 neuroimágenes de pacientes sanos.

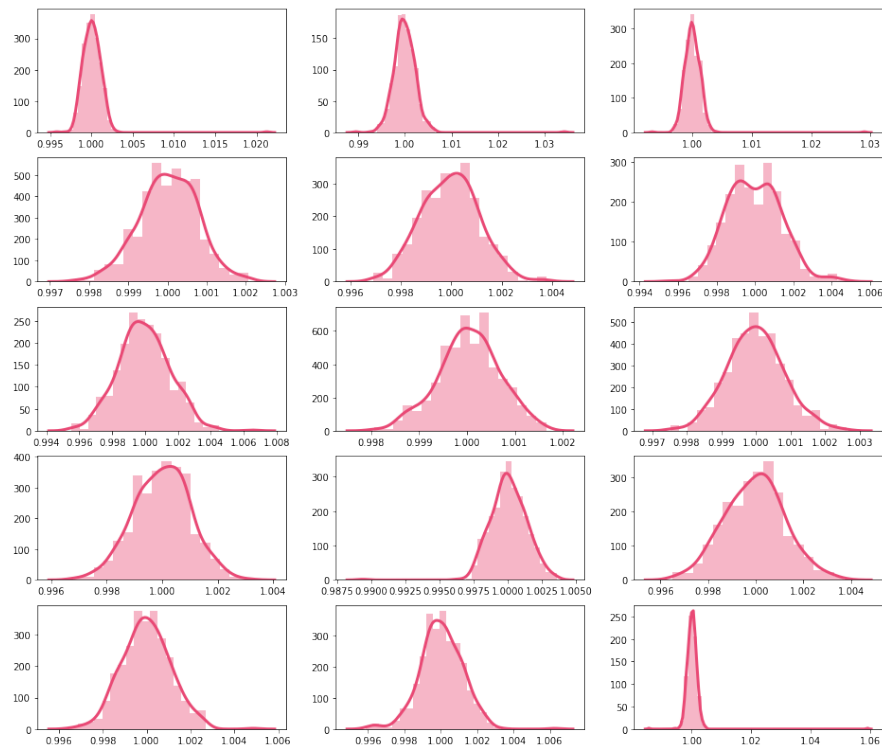


Figura 7. Distribución del test HCP en 15 neuroimágenes de pacientes patológicos.

En el caso de los pacientes sanos la distribución observada para el test es coherente con lo que se planteo en la hipótesis (valores que están alrededor de 1); sin embargo, para los patológicos se esperaba que su distribución no estuviera centrada en 1 (figuras 6 y 7). Un aspecto que se debe resaltar es que, aunque en efecto para ambos tipos de imágenes el test HCP esta alrededor de 1, para los patológicos se observa un rango de valores más amplia y una mayor varianza (tabla 2)

Tabla 2. Medidas de resumen del test HCP de sanos y patológicos.

Medida de resumen	Sanos	Patológicos	Patológicos/sanos
Media	1.0	1.0	1.0
Mediana	1.0005	1.0006	1.0
Desviación estándar	0.0021	0.0169	8.0476
Rango	0.0062	0.0142	2.2903
Rango intercuartílico	0.0009	0.0014	1.5555
Mínimo	0.996	0.9923	0.9963
Máximo	1.0022	1.0066	1.0044

De la tabla 2 se observa que la desviación estándar de los patológicos es 8 veces la de los sanos, su rango 2 veces mayor y el rango intercuartílico 1.5.

Dado este comportamiento se podría tratar de plantear una prueba que se base en la distribución del mínimo, el máximo y la desviación, como ejercicio se establecieron de los sanos, el valor del percentil 90 del mínimo de HCP (min), el percentil 10 del máximo (max) y el percentil 10 de la desviación (min_{std}). Luego se establecieron las siguientes reglas de clasificación:

$$\hat{p}_{min} = \frac{\sum_i I_{min}(HCP)}{n}, \hat{p}_{max} = \frac{\sum_i I_{max}(HCP)}{n}, \quad i = 1, 2, 3, \dots, 500$$

donde

$$I_{min}(x) = \begin{cases} 1, & x \leq min \\ 0, & x > 0 \end{cases}$$

$$I_{max}(x) = \begin{cases} 1, & x \geq max \\ 0, & x < 0 \end{cases}$$

$$Zonas = \begin{cases} \text{patológica: } \hat{p}_{min} > 0.01 \vee \hat{p}_{min} > 0.01 \wedge \hat{\sigma}_{HCP} > min_{std} \\ \text{indecisión: } \hat{p}_{min} > 0.01 \vee \hat{p}_{min} > 0.01 \wedge \hat{\sigma}_{HCP} \leq min_{std} \\ \text{Sanos: Otro caso} \end{cases}$$

De esta manera se clasifican los estudios de la siguiente manera:

Tabla 3. Matriz de confusión según la distribución del test HCP

Grupo	Zonas		
	Sanos	Patológica	Indecisión
Sanos	149 (70%)	22 (11%)	41(19%)
Patológicos	34 (22%)	96 (61%)	27 (17%)

De la tabla 3 se observa que la clasificación correcta de patológicos y sanos es baja (61% y 70% respectivamente), aunque si se calculan estos valores sin tener en cuenta la zona de indecisión, los resultados mejoran (73% y 87% respectivamente), aun así, sigue siendo un resultado poco satisfactorio para usarlo en la practica clínica.

Los resultados obtenidos con el HCP evidencian que este test tal como fue propuesto, tiene dificultades para separar de manera satisfactoria ambos grupos, aunque se logran resultados mucho mejores comparado con las diferencias ajustadas que se presentaron al principio del capitulo. Es posible que se deban agregar otras características a la prueba en busca de que capture más diferencias.

Se realizó el mismo ejercicio, pero solo con las intensidades mayores que cero, el comportamiento fue similar (no se presentan los resultados).

6.1.2 HCP con mapas de probabilidad

En este caso también se observan valores que oscilan al redor de 1 (figuras 8 y 9), sin embargo, no se observan tantas diferencias en el rango de valores del HCP, pero si en la varianza (tabla 4). Por tanto la matriz de confusión se construye teniendo en cuenta solo la desviación (tabla 5).

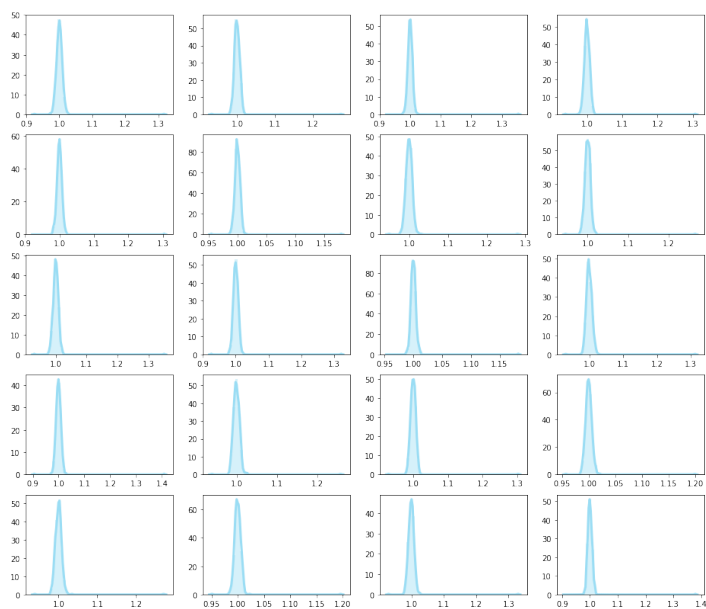


Figura 8. Distribución del test HCP en 20 neuroimágenes de pacientes sanos - Blanca.

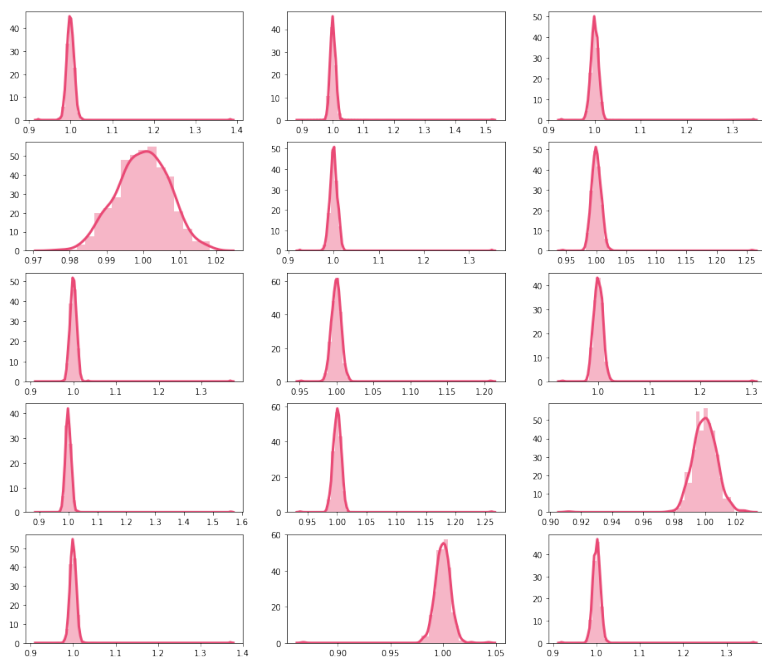


Figura 9. Distribución del test HCP en 15 neuroimágenes de pacientes patológicas - Blanca.

Tabla 4. Medidas de resumen del test HCP de sanos y patológicos - Bianca.

Medida de resumen	Sanos	Patológicos	Patológicos/sanos
Media	1.0005	1.0005	1.0
Mediana	1.0048	1.0048	1.0
Desviación estándar	0.0804	0.1451	1.8047
Rango	0.3524	0.3339	0.9475
Rango intercuartílico	0.0096	0.0095	0.9896
Mínimo	0.9361	0.9352	0.9990
Máximo	1.2885	1.2691	0.9849

De la tabla 4 se observa que la desviación es la única variable en la que el HCP de los patológicos muestra más diferencias, siendo este 1.8 veces la desviación de los sanos.

Tabla 5. Matriz de confusión según la distribución del test HCP - Bianca.

	Zonas	
Grupo	Sanos	Patológica
Sanos	20 (10%)	176 (90%)
Patológicos	16 (33%)	59 (77%)

La matriz de confusión se construyó con lo mismo criterios que la obtenida en la tabla 4, salvo que en este caso el único parámetro considerado fue la desviación. En este caso los resultados fueron malos en la detección de sanos, para los patológicos se observan mejores resultados (tabla 5). Estos valores sugieren que el mapa de probabilidades obtenido con Bianca no es una buena opción para calcular el test HCP.

6.2 HPLE

El test HPLE fue calculado con los valores de intensidades en $[0,1]$. En la figura 10 se muestra la distribución de las diferencias entre hemisferios, se observa que de nuevo las neuroimágenes patológicas son las que mayor variabilidad presentan.

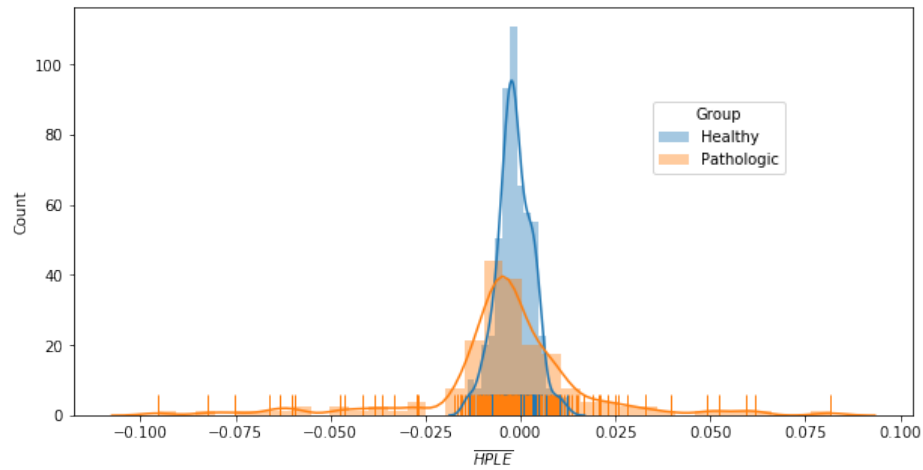


Figura 10. Distribución de HPLE entre hemisferios, según el grupo.

A diferencia del HCP, en el HPLE se tienen valores para ambos hemisferios, por lo que es posible aplicar pruebas estadísticas paramétricas y no paramétricas para evaluar si hay diferencias significativas entre hemisferios en cada imagen. Para evaluar si hay diferencias, se aplicó el estadístico t con varianzas diferentes y el test U de Mann-Whitney, con el primero se busca comparar las medias del HPLE de ambos hemisferios y con el segundo se trata de averiguar si ambas muestras provienen de la misma distribución (31).

Con ambas pruebas se observa una capacidad predictiva similar del HPLE (tablas 6 y 7), en general se puede decir que el HPLE tiene un bajo rendimiento para los diferentes niveles de significación (figuras 11 y 12).

En ambos casos, a medida que aumenta el nivel de significancia, también lo hace el Recall, pero disminuyen los demás indicadores (figuras 11 y 12).

Tabla 6. Capacidad predictiva del HPLE según niveles de significancia, aplicando prueba U de Mann-Whitney.

α	VP	VN	FP	FN	Accuraccy	Precision	Recall	F1-score
0.001	86	181	31	71	0.72	0.74	0.55	0.63
0.01	106	158	54	51	0.72	0.66	0.68	0.67
0.02	112	148	64	45	0.70	0.64	0.71	0.67
0.03	115	138	74	42	0.69	0.61	0.73	0.66
0.04	116	135	77	41	0.68	0.60	0.74	0.66
0.05	116	126	86	41	0.66	0.57	0.74	0.64
0.06	117	122	90	40	0.65	0.57	0.75	0.65
0.07	119	119	93	38	0.64	0.56	0.76	0.64
0.08	120	117	95	37	0.64	0.56	0.76	0.64
0.09	120	112	100	37	0.63	0.55	0.76	0.64

0.10	123	110	102	34	0.63	0.55	0.78	0.65
0.11	125	107	105	32	0.63	0.54	0.80	0.64
0.12	125	104	108	32	0.62	0.54	0.80	0.64
0.13	126	100	112	31	0.61	0.53	0.80	0.64
0.14	129	96	116	28	0.61	0.53	0.82	0.64
0.15	129	91	121	28	0.60	0.52	0.82	0.64
0.16	130	89	123	27	0.59	0.51	0.83	0.63
0.17	132	83	129	25	0.58	0.51	0.84	0.63
0.18	134	83	129	23	0.59	0.51	0.85	0.64
0.19	134	81	131	23	0.58	0.51	0.85	0.64
0.2	135	76	136	22	0.57	0.5	0.86	0.63

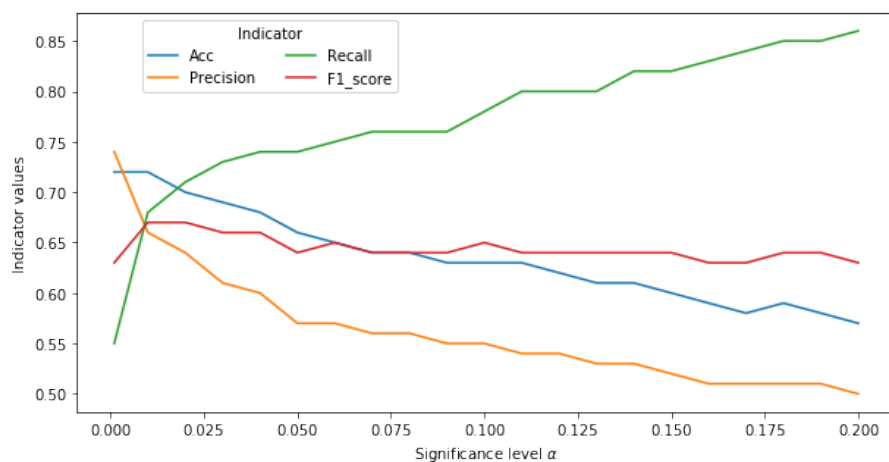


Figura 11. Tendencia de indicadores de desempeño del HPLE según el nivel de significancia, aplicando prueba U de Mann-Whitney.

Tabla 7. Capacidad predictiva del HPLE según niveles de significancia, aplicando la prueba t de student.

α	VP	VN	FP	FN	Accuraccy	Precision	Recall	F1-score
0.001	91	176	36	66	0.72	0.72	0.58	0.64
0.01	111	148	64	46	0.70	0.63	0.71	0.67
0.02	115	134	78	42	0.67	0.60	0.73	0.66
0.03	116	124	88	41	0.65	0.57	0.74	0.64
0.04	120	117	95	37	0.64	0.56	0.76	0.64
0.05	123	109	103	34	0.63	0.54	0.78	0.64
0.06	127	101	111	30	0.62	0.53	0.81	0.64
0.07	130	94	118	27	0.61	0.52	0.83	0.64
0.08	131	89	123	26	0.60	0.52	0.83	0.64
0.09	131	84	128	26	0.58	0.51	0.83	0.63
0.1	133	77	135	24	0.57	0.50	0.85	0.63
0.11	134	72	140	23	0.56	0.49	0.85	0.62

0.12	136	70	142	21	0.56	0.49	0.87	0.63
0.13	137	67	145	20	0.55	0.49	0.87	0.63
0.14	139	65	147	18	0.55	0.49	0.89	0.63
0.15	140	62	150	17	0.55	0.48	0.89	0.62
0.16	141	61	151	16	0.55	0.48	0.90	0.63
0.17	141	59	153	16	0.54	0.48	0.90	0.63
0.18	142	57	155	15	0.54	0.48	0.90	0.63
0.19	145	55	157	12	0.54	0.48	0.92	0.63
0.2	145	53	159	12	0.54	0.48	0.92	0.63

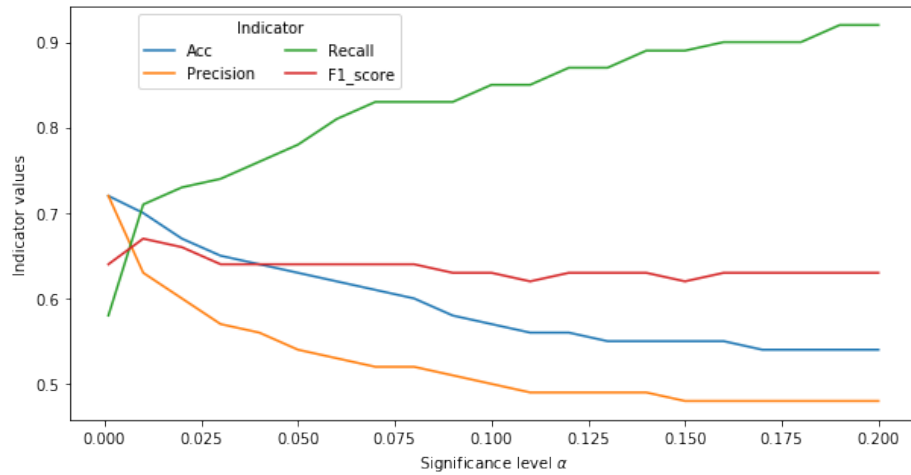


Figura 12. Tendencia de indicadores de desempeño del HPLE según el nivel de significancia, aplicando prueba t de student.

6.3 Eigenfaces

Como se menciona, se realizaron dos tipos de análisis con Eigenfaces, el primero es construyendo subespacios por hemisferios y grupo, el segundo es construyendo un subespacio según el grupo pero sobre todo el cerebro.

Para cada caso, la selección de los Eigenfaces o subespacio se hizo reteniendo los eigenvalues de mayor valor que retuvieran el 99% de la variabilidad. El valor de 99% se elige ya que el interés principal no es la reducción de dimensionalidad, si no tratar de encontrar un subespacio capaz de capturar la información más relevante, para que luego las proyecciones de cualquier neuroimagen sobre este indiquen que tanta información relevante (con respecto al subespacio) contiene esa imagen proyectada. Sin embargo, al hacer el ejercicio capturando el 90% de la variabilidad, se siguieron observando valores consistentes de los modelos, con una disminución de solo un 1% en su poder predictivo (tabla 11 y figura 19).

6.3.1 Eigenfaces por hemisferios y grupo (sanos, patológicos) con valores de intensidades en [0,1] (enfoque A)

Los subespacios encontrados en las imágenes se contruyeron con 119 y 202 componentes para el hemisferio derecho y el izquierdo respectivamente, y para los patológicos fueron 110 y 148.

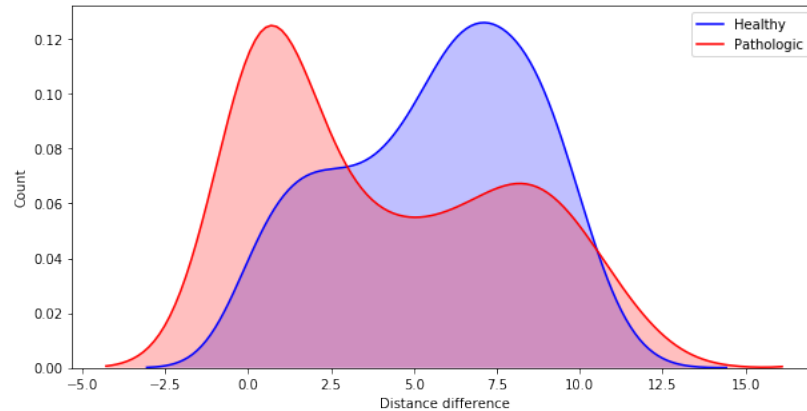


Figura 13. Diferencia entre hemisferios de las distancia al subespacio.

Las diferencias entre hemisferios de las distancias al respectivo subespacio construido muestran distribuciones con formas diferentes, pero con su rango de valores similar, lo que sugiere que la diferencia calculada puede no ser suficiente para distinguir entre ambos grupos (figura 11). Para comprobarlo, se ajustan modelos con las diferencias de las distancias obtenidas.

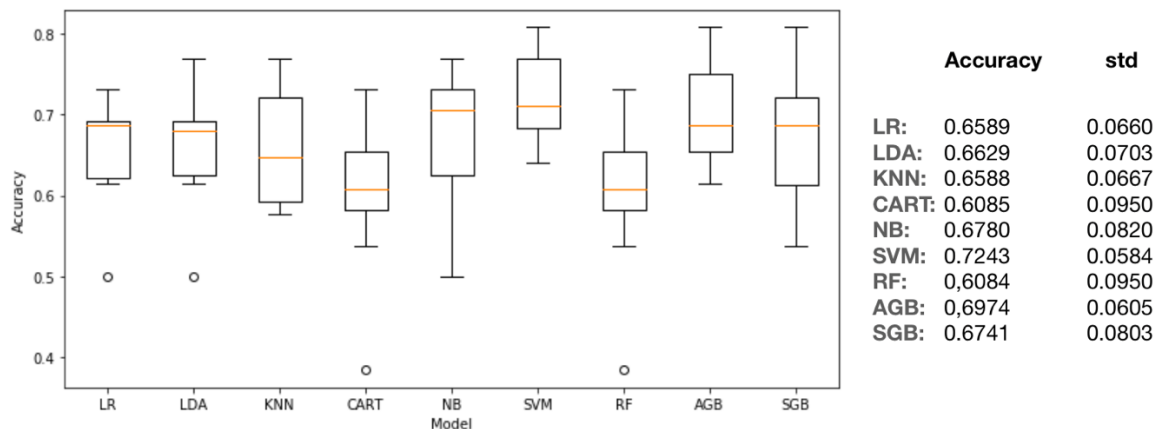


Figura 14. Validación cruzada de modelos de clasificación, Eigenfaces enfoque A.

* LR:logístico, LDA: Discriminante lineal, KNN: *k-nearest neighbors*, CART: Árboles de decisión, NB: *Gaussian naïve Bayes*, SVM: Máquina de Soporte Vectorial, RF: Random Forest, AGB: AdaBoost, SGB: Gradient Boosted Decision Trees.

Después de comparar diferentes modelos de clasificación, las maquinas de soporte vectorial son las que mejor comportamiento tuvieron (figura 12). Una vez se hacen predicciones sobre los datos de prueba, se observa un accuracy igual al 61% (tabla 6), lo que en términos generales es un resultado poco satisfactorio para la practica clínica.

Tabla 8. Desempeño del modelo SVM sobre datos de prueba, Eigenfaces enfoque A.

Categoría	Valores predichos		Indicadores			
	Sanos	Enfermos	Precisión	Recall	F1-score	n
Sanos	51	13	0.63	0.80	0.70	64
Enfermos	30	17	0.57	0.36	0.44	47
Total / Promedio	81	30	0.60	0.61	0.59	111

* Valores predichos sobre el 30% de los datos (datos de prueba)

Observe que el desempeño del modelo fue peor en los enfermos, lo que hace que este modelo sea poco viable.

6.3.2 Eigenfaces con NCC (enfoque B)

Hasta el momento, uno de los grandes problemas en las pruebas que se han propuesto, es que la construcción y comparación directa entre hemisferios con distancias a subespacios o con valores de intensidades “crudos”, no logran captar las características necesaria para hallar diferencias significativas.

Por este motivo se propone trabajar con el NCC, el cual esta en el espacio original de la imagen. Es importante recordar que el espacio NCC fueron reconstruidos a partir de supuestos de simetría entre hemisferios, lo que quiere decir que, aunque el método se aplique al espacio del cerebro completo, los valores con los que trabaja provienen de medidas de similaridad entre hemisferios.

A continuación se muestran los resultados de aplicar Eigenfaces sobre los mapas NCC reteniendo el 99% de variabilidad (ver sección 6.3), obteniendo 200 componentes.

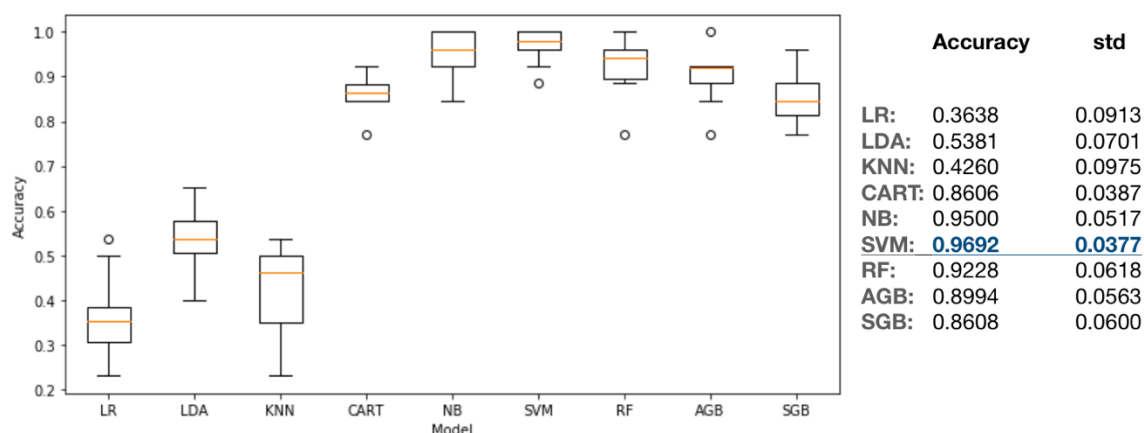


Figura 15. Validación cruzada de modelos de clasificaicón, Eigenfaces enfoque B.

* LR:logístico, LDA: Discriminante lineal, KNN: *k-nearest neighbors*, CART: Arboles de decisión, NB: *Gaussian naïve Bayes*, SVM: Maquina de Soporte Vectorial, RF: Random Forest, AGB: AdaBoost, SGB: Gradient Boosted Decision Trees.

Se observa que el modelo SVM tiene un mejor desempeño en la validación cruzada, además de ser el más consistentes con una desviación estándar igual a 0.0377 (figura 13).

Ahora se evalúa el desempeño de los modelos y para el mejor modelo de verificara su capacidad predictiva con el área bajo la curva roc. Observe que hasta el momento el método usado en esta sección es el de mejores resultados.

Tabla 9. Desempeño de modelos sobre datos de prueba, Eigenfaces enfoque B.

Modelo	Categoría	Valores predichos		Indicadores			
		Sanos	Enfermos	Precisión	Recall	F1-score	n
LR	Sanos	7	57	0.26	0.11	0.15	64
	Enfermos	20	27	0.32	0.57	0.41	47
	Total/promedio	27	84	0.29	.031	0.26	111
RF	Sanos	62	2	0.94	0.97	0.95	64
	Enfermos	4	43	0.96	0.91	0.93	47
	Total/promedio	66	45	0.95	0.95	0.95	11
SVM	Sanos	63	1	0.98	0.98	0.98	64
	Enfermos	1	46	0.98	0.98	0.98	47
	Total/promedio	64	47	0.98	0.98	0.98	111
AGB	Sanos	55	9	0.95	0.86	0.90	64
	Enfermos	3	44	0.83	0.94	0.88	47
	Total/promedio	58	53	0.90	0.89	0.89	111
SGB	Sanos	52	12	1.00	0.81	0.90	64
	Enfermos	0	47	0.80	1.00	0.89	47
	Total/promedio	52	59	0.91	0.89	0.89	111

* LR:logístico, SVM: Maquina de Soporte Vectorial, RF: Random Forest,AGB: AdaBoost, SGB: Gradient *Boosted Decision Trees*. Valores predichos sobre el 30% de los datos (datos de prueba)

En este caso, modelo SVM tiene un mejor desempeño para predecir comparado con los demás modelos, anqué el modelo de RF también obtuvo buenos resultados (tabla 7). La capacidad predictiva del modelo de maquinas de soporte vectorial es bastante bueno con un 98% (figura 14).

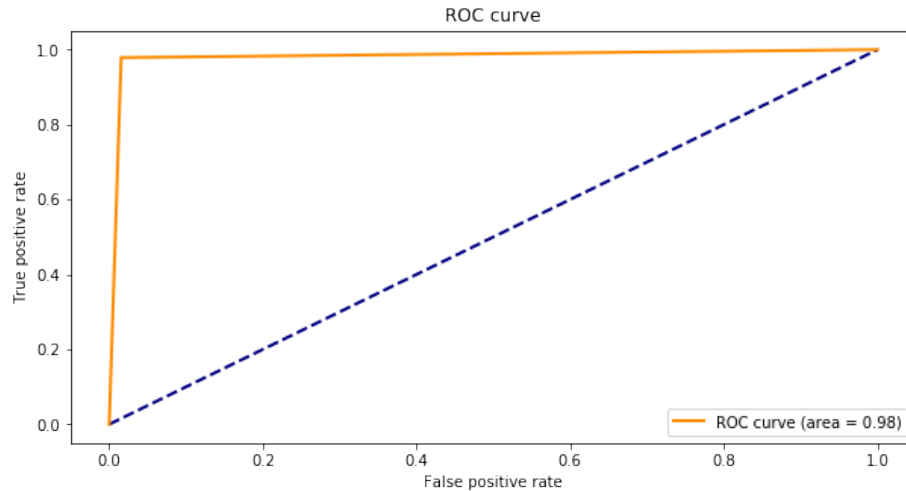


Figura 16. Capacidad predictiva del modelo SVM, Eigenfaces enfoque B.

6.3.3 Eigenfaces con mapa de probabilidades Bianca (enfoque B)

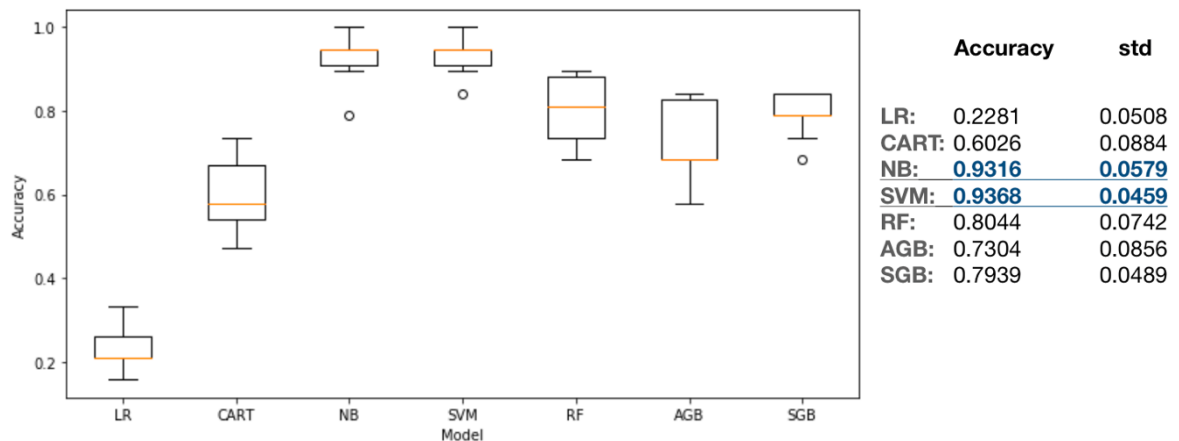


Figura 17. Validación cruzada de modelos de clasificaión, Eigenfaces-Bianca enfoque B.

* LR:logístico, CART: Arboles de decisión, NB: *Gaussian naïve Bayes*, SVM: Maquina de Soporte Vectorial, RF: Random Forest, AGB: AdaBoost, SGB: Gradient *Boosted Decision Trees*.

Después de aplicar Eigenfaces a los mas de probabilidad obtenidos con Bianca, se observa que este método también arroja buenos resultados, especialmente con los modelos maquinas de soporte vectorial y Naive Bayes (figura 15).

Tabla 10. Desempeño de modelos sobre datos de prueba, Eigenfaces-Bianca enfoque B.

Modelo	Categoría	Valores predichos		Indicadores			
		Sanos	Enfermos	Precisión	Recall	F1-score	n
LR	Sanos	7	52	0.29	0.12	0.17	59
	Enfermos	17	6	0.10	0.26	0.15	23
	Total/promedio	24	58	0.24	0.16	0.16	82
RF	Sanos	58	1	0.78	0.98	0.87	59
	Enfermos	16	7	0.88	0.30	0.45	23
	Total/promedio	74	8	0.81	0.79	0.75	82
SVM	Sanos	59	0	0.88	1.00	0.94	59
	Enfermos	8	15	1.00	0.65	0.79	23
	Total/promedio	67	15	0.91	0.90	0.90	82
NB	Sanos	58	1	0.88	0.98	0.93	59
	Enfermos	8	15	0.94	0.65	0.77	23
	Total/promedio	66	16	0.90	0.89	0.88	82
AGB	Sanos	53	6	0.88	0.90	0.89	59
	Enfermos	7	16	0.73	0.70	0.71	23
	Total/promedio	60	22	0.84	0.84	0.84	82
SGB	Sanos	57	2	0.81	0.97	0.88	59
	Enfermos	13	10	0.83	0.43	0.57	23
	Total/promedio	70	12	0.82	0.82	0.80	82

Bajo este enfoque, los modelos con mejor desempeño fueron SVM y NB como se esperaba, sin embargo, sus recall es bajo con un 65%, se observa entonces que estos dos modelos predicen muy bien si clasifican como patológicos, pero no logran predecir todos los patológicos (cuando predicen un enfermo lo hace muy bien, pero deja de predecir una proporción considerable de ellos) (tabla 8).

La capacidad de los modelos SVM y NB es similar con un 83% y 82% respectivamente (figura 16).

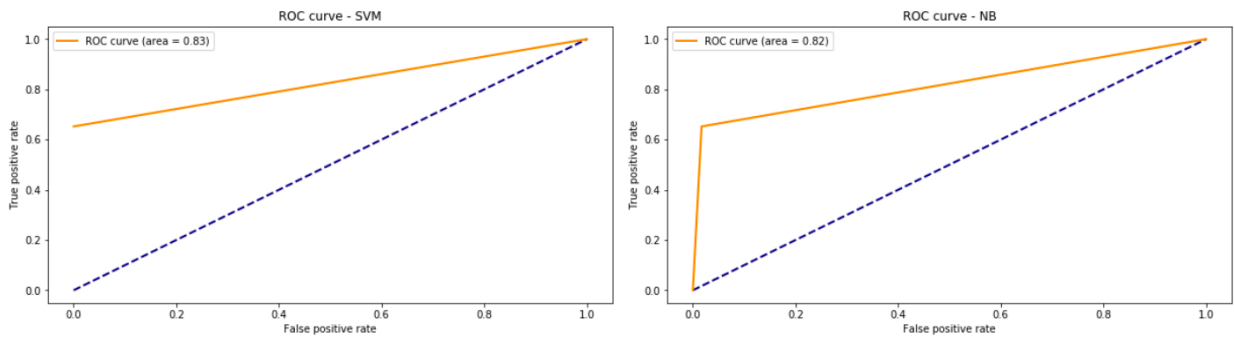


Figura 18. Capacidad predictiva de modelo SVM y NB, Eigenfaces-Bianca enfoque B.

7 Discusión

El HCP no logra capturar la suficiente información de las imágenes como para distinguir entre sanos y patológicos, con una detección de patológicos y sanos baja (61% y 70% respectivamente) (tabla 3), esto se relaciona con el hecho que los valores del HCP de estos dos grupos solo se diferencia por su varianza y rango (figuras 6 y 7, tabla 2). Es posible que se pueda mejorar estos resultados complementándolo con pruebas de asimetría y curtosis (tercer y cuarto momento), ya que estos dan información adicional sobre la distribución de los datos. Al aplicar el HCP sobre el mapa de probabilidades calculados con Bianca, se obtuvieron resultados más pobres que al usar las intensidades en $[0, 1]$, con un 10% clasificados correctamente para sanos y 77% para patológicos (tabla 5), esto tiene relación con la posible “perdida” de información al aplicar Bianca (tabla 4) ya que la probabilidad con la que se trabaja, esta basado en los k-vecinos más próximos, lo que se podría ver como un tipo de suavizado, aunque no es así técnicamente.

El hecho de que el mapa de probabilidades no funcione muy bien con el HCP, no quiere decir que el método Bianca arroje pobres resultados, de hecho, se observo un buen comportamiento en la predicción de zonas de hiperintensidades en cerebros patológicos con este método (figura 1). En futuros estudios se podría abordar con más profundidad, tratando de identificar para cuales patologías es más viable usarlo.

Con el HPLE se observaron mejores resultados comparado con el HCP, con un f1-score igual a 64% y un *recall* igual al 78% con un nivel de significancia del 5% (tabla 7), sin embargo, no es suficiente aun para sugerir que se puede usar en la practica clínica. Una ventaja importante del HPLE sobre el HCP es que se tienen distribuciones por hemisferios y se puede complementar con pruebas paramétricas y no paramétricas para determinar si hay diferencias estadísticas (tablas 6 y 7). Con valores más altos del nivel de significancia se mejora la capacidad para detectar

neuroimágenes patológicas del HPLE, sin embargo, los demás indicadores de rendimiento medidos disminuyen drásticamente, es decir que disminuye su capacidad para detectar verdaderos sanos, lo que puede resultar poco ventajoso en la práctica. Un nivel de significancia del 5% puede ser un valor razonable si se quiere usar el HPLE como medida de detección (figuras 11 y 12).

Aunque los resultados del HPLE no fueron del todo satisfactorias, es un método interesante que podría ser explorado en futuros trabajos. Lo que lo hace llamativo es que logro detectar algunas diferencias usando dos conceptos útiles, el primero es con el uso de permutaciones para aumentar la capacidad de detección de diferencias (disminución del error tipo I), y el segundo tiene que ver con el valor de la entropía como medida de incertidumbre. En otros trabajos se puede probar usando parámetros diferentes para la selección de áreas y proponiendo otras formas de calcular las probabilidades de ocurrencia de intensidades en dichas áreas, también es posible tratar de integrar más información a este indicador o incluso hacer ponderaciones basadas en el grado de intensidad de las áreas seleccionadas.

Finalmente, al aplicar el método de Eigenfaces se obtuvieron los mejores resultados, especialmente cuando se trabajo con el mapa de correlaciones NCC en lugar del mapa de probabilidades, aunque con los mapas de probabilidades también se observaron buenos resultados (figuras 15 y 17). El modelo con mejor rendimiento y poder predictivo fue el SVM con un 98% de capacidad predictiva (tabla 9 y figura 16).

Un aspecto interesante al observar los resultados obtenidos con Eigenfaces y compararlo con los otros métodos, evidencia que la información adicional capturado por los subespacios (que el HCP y HPLE no lo consideran), le permite a los modelos de clasificación tener un buen desempeño, además de ser un métodos computacionalmente viable al hacer una variación en la matriz de covarianzas usada en PCA (22,23).

Como se menciona en la sección 6.3, los resultados que se obtuvieron con las predicciones después de capturar un 90% de la variabilidad con el subespacio construido, mostraron ser consistentes. Con un modelo de maquinas de soporte vectorial se obtuvo un accuracy del 96.40% (tabla 11), y un área bajo la curva roc igual al 97% (figura 19).

Tabla 11. Desempeño de modelo SVM sobre datos de prueba, reteniendo el 90% de variabilidad en Eigenfaces.

Categoría	Valores predichos		Indicadores			
	Sanos	Enfermos	Precisión	Recall	F1-score	n
Sanos	61	3	0.98	0.95	0.97	64
Enfermos	1	46	0.94	0.98	0.96	47
Total/promedio	62	49	0.96	0.96	0.96	111

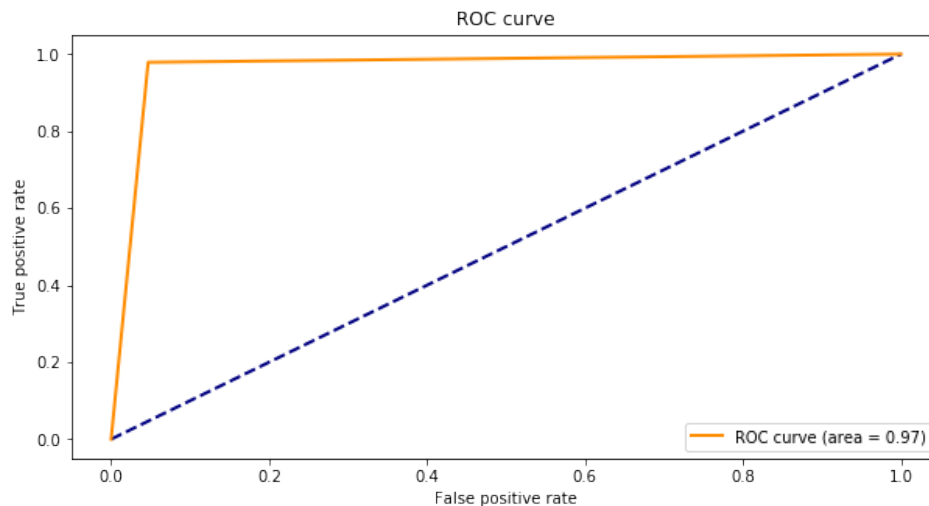


Figura 19. Capacidad predictiva de modelo SVM, reteniendo un 90% de variabilidad con Eigenfaces.

Otro factor fundamental y que resulta de mucha utilidad, es que el resultado final resulta de un subespacio que fue construido con el mapa NCC, la que se interpreta como una medida de similitud entre hemisferios, por lo que variaciones en parámetros de adquisición o la implementación de nuevos protocolos en la adquisición de neuroimágenes no debería afectar el NCC (se trata de una medida intrínseca del cerebro), a menos que se vea comprometida la calidad de la imagen con artefactos o el estudio en análisis tenga cirugías previas. Sin olvidar por supuesto que el NCC es calculado con los valores de intensidades de la secuencia Flair.

Es cierto que los resultados del NCC, Eigenfaces y SVM combinados fueron buenos; aun así, es necesario hacer pruebas en un ambiente de producción y hacer seguimiento por algún tiempo con el fin de monitorear su desempeño.

8 Conclusiones

El uso del mapa de correlaciones NCC obtenido de una secuencia flair para medir similitud entre hemisferios resultó ser útil para la predicción de estudios sanos/patológicos, después de generar un subespacio con el método Eigenfaces y aplicar un modelo de clasificación de máquinas de soporte vectorial. Estos métodos aplicados en conjunto permiten hacer una priorización adecuada con un 98% de capacidad predictiva.

La priorización de la que se habla puede ser vista en el ámbito clínico como un triaje inicial que ayude a optimizar los flujos de trabajo de los médicos radiólogos en neuroimágenes. En próximos trabajos se podría tratar de replicar esta metodología en el triaje de más de dos categorías.

El HPLE es una medida interesante que puede ser estudiada con más detalle integrando otros conceptos y haciendo algunas variaciones tratando de mejorar su capacidad de detección, por ejemplo, asignando pesos a las áreas seleccionadas una vez se han calculado las medidas de incertidumbre.

9 Código

En https://github.com/styvens/priorizacion_neuroimagenes_RM se encuentra el código usada en el proyecto, incluyendo este documento.

10 Propuesta de otras líneas de trabajo

Se propone trabajar con cuatro medidas que desde el punto de vista clínico y anatómico cobran relevancia y es posible que permitan detectar patologías específicas como lo son las enfermedades neurodegenerativas, estas medidas son: volumen de sustancia gris, volumen de sustancia blanca, volumen del hipocampo e integrarlo con los resultados del volumen de hemisferios presentados en este trabajo.

Las variables propuestas son volúmenes con las que se podrían medir alteraciones anatómicas importantes que están relacionadas con algunas patologías, por ejemplo se sabe que la atrofia de sustancia gris, sustancia blanca y del hipocampo está asociado a ciertos trastornos neurológicos (32–34)

Existe evidencia que afirma que la cantidad de sustancia gris y sustancia blanca en el cerebro disminuye con la edad (35,36), sin embargo una pérdida excesiva de estas se puede asociar a trastornos neurológicos como el Alzheimer, Parkinson, entre otros (34,37–39). Como un criterio o indicador de los volúmenes de estas

sustancias es posible establecer valores de referencia según la edad con el fin de detectar posibles anomalías (40).

Resulta interesante incluir un área específica como el hipocampo, ya que esta estructura tiene un papel importante en el cerebro por su relación con la memoria y el aprendizaje (41). Se sabe que alteraciones en esta estructura puede relacionarse también con algunos trastornos neurológicos (42,43).

11 Referencias

1. Spota BB. Histórico y evolución de la neurología. Archivos de Neuro-Psiquiatria. septiembre de 1946;4(3):295–303.
2. Aguilera JFC. Charcot y su Legado a la Medicina. Gac Med Mex. 2012;148(3):321–6.
3. Griffanti L, Zamboni G, Khan A, Li L, Bonifacio G, Sundaresan V, et al. BIANCA (Brain Intensity AbNormality Classification Algorithm): A new tool for automated segmentation of white matter hyperintensities. NeuroImage. 1 de noviembre de 2016;141:191–205.
4. Turk M, Pentland A. Eigenfaces for recognition. J Cognitive Neuroscience. 1 de enero de 1991;3(1):71–86.
5. Smith SM. Fast robust automated brain extraction. Hum Brain Mapp. noviembre de 2002;17(3):143–55.
6. Brett M, Johnsrude IS, Owen AM. The problem of functional localization in the human brain. Nat Rev Neurosci. marzo de 2002;3(3):243–9.
7. Vadim Kuperman. Introduction. En: Magnetic Resonance Imaging. United States of America: Academic Press; 2000. p. 1–7.
8. Symms M, Jäger HR, Schmierer K, Yousry TA. A review of structural magnetic resonance neuroimaging. Journal of Neurology, Neurosurgery & Psychiatry. 1 de septiembre de 2004;75(9):1235–44.
9. Stinnett TJ, Zabel MK. Neuroanatomy, Broca Area. En: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2019 [citado 29 de octubre de 2019]. Disponible en: <http://www.ncbi.nlm.nih.gov/books/NBK526096/>

10. Rottschy C, Langner R, Dogan I, Reetz K, Laird AR, Schulz JB, et al. Modelling neural correlates of working memory: A coordinate-based meta-analysis. *NeuroImage*. 1 de marzo de 2012;60(1):830–46.
11. Sun L, Zhang S, Chen H, Luo L. Brain Tumor Segmentation and Survival Prediction Using Multimodal MRI Scans With Deep Learning. *Front Neurosci* [Internet]. 2019 [citado 29 de octubre de 2019];13. Disponible en: <https://www.frontiersin.org/articles/10.3389/fnins.2019.00810/full>
12. Liu SX. Symmetry and asymmetry analysis and its implications to computer-aided diagnosis: A review of the literature. *Journal of Biomedical Informatics*. 1 de diciembre de 2009;42(6):1056–64.
13. Kalavathi P, Senthamilselvi M, Prasath VBS. Review of Computational Methods on Brain Symmetric and Asymmetric Analysis from Neuroimaging Techniques. *Technologies*. junio de 2017;5(2):16.
14. Toga AW, Thompson PM. Mapping brain asymmetry. *Nat Rev Neurosci*. enero de 2003;4(1):37–48.
15. Xiang L, Crow T, Roberts N. Cerebral torque is human specific and unrelated to brain size. *Brain Struct Funct*. 1 de abril de 2019;224(3):1141–50.
16. Filipek PA, Richelme C, Kennedy DN, Caviness VS. The Young Adult Human Brain: An MRI-based Morphometric Analysis. *Cereb Cortex*. 1 de julio de 1994;4(4):344–60.
17. Gupta M, Pande A, Rao RS, Tripathi AK. Design Pattern Detection by normalized cross correlation. En: 2010 International Conference on Methods and Models in Computer Science (ICM2CS-2010). 2010. p. 81–4.
18. Falk M. On Mad and Comedians. *Annals of the Institute of Statistical Mathematics*. 1 de diciembre de 1997;49(4):615–44.
19. Oden A, Wedel H. Arguments for Fisher's Permutation Test. *Ann Statist*. marzo de 1975;3(2):518–20.
20. Bacher A, Bodini O, Hwang H-K, Tsai T-H. Generating Random Permutations by Coin Tossing: Classical Algorithms, New Analysis, and Modern Implementation. *ACM Trans Algorithms*. 7 de febrero de 2017;13(2):24:1–24:43.
21. Shannon CE. A mathematical theory of communication. *SIGMOBILE Mob Comput Commun Rev*. 1 de enero de 2001;5(1):3–55.
22. Sirovich L, Kirby M. Low-dimensional procedure for the characterization of human faces. *J Opt Soc Am A, JOSAA*. 1 de marzo de 1987;4(3):519–24.

23. Turk MA, Pentland AP. Face recognition using eigenfaces. En: 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Proceedings. 1991. p. 586–91.
24. Anissa B, Naouar B, Arsalane Z, Jamal K. Face recognition: Comparative study between linear and non linear dimensionality reduction methods. En: 2015 International Conference on Electrical and Information Technologies (ICEIT). 2015. p. 224–8.
25. Kuhn M, Johnson K. Classification Models. En: Applied predictive modeling. Springer; 2013. p. 247–457.
26. Igual L, Seguí S. Supervised Learning. En: Introduction to Data Science. Springer; 2017. p. 67–96.
27. Skiena SS. Machine Learning. En: The data science design manual. Springer; 2017. p. 351–71.
28. Aggarwal CC. Neural networks and deep learning. Springer; 2018.
29. de Núremberg C. Tribunal Internacional de Núremberg. Recuperado de http://www.sld.cu/galerias/pdf/sitios/prevemi/codigo_nuremberg.pdf. 1947;
30. Beauchamp TL. The belmont report. The Oxford textbook of clinical research ethics. 2008;149–55.
31. Conover W. Two Independent Samples. En: Practical Nonparametric Statistics. New York: John Wiley&Sons Inc.; 1999. p. 271–88.
32. Hashimoto M, Kazui H, Matsumoto K, Nakano Y, Yasuda M, Mori E. Does donepezil treatment slow the progression of hippocampal atrophy in patients with Alzheimer's disease? Am J Psychiatry. abril de 2005;162(4):676–82.
33. Amiri H, de Sitter A, Bendfeldt K, Battaglini M, Gandini Wheeler-Kingshott CAM, Calabrese M, et al. Urgent challenges in quantification and interpretation of brain grey matter atrophy in individual MS patients using MRI. Neuroimage Clin. 26 de abril de 2018;19:466–75.
34. Palotai M, Nazeri A, Cavallari M, Healy BC, Glanz B, Gold SM, et al. History of fatigue in multiple sclerosis is associated with grey matter atrophy. Sci Rep. 14 de octubre de 2019;9(1):14781.
35. Ge Y, Grossman RI, Babb JS, Rabin ML, Mannon LJ, Kolson DL. Age-Related Total Gray Matter and White Matter Changes in Normal Adult Brain. Part I: Volumetric MR Imaging Analysis. American Journal of Neuroradiology. 1 de septiembre de 2002;23(8):1327–33.

36. Guttmann CRG, Jolesz FA, Kikinis R, Killiany RJ, Moss MB, Sandor T, et al. White matter changes with normal aging. *Neurology*. 1 de abril de 1998;50(4):972.
37. Chen F-X, Kang D-Z, Chen F-Y, Liu Y, Wu G, Li X, et al. Gray matter atrophy associated with mild cognitive impairment in Parkinson's disease. *Neurosci Lett*. 23 de marzo de 2016;617:160–5.
38. Ting WK-C, Fischer CE, Millikin CP, Ismail Z, Chow TW, Schweizer TA. Grey matter atrophy in mild cognitive impairment / early Alzheimer disease associated with delusions: a voxel-based morphometry study. *Curr Alzheimer Res*. 2015;12(2):165–72.
39. Kunst J, Marecek R, Klobusiakova P, Balazova Z, Anderkova L, Nemcova-Elfmarkova N, et al. Patterns of Grey Matter Atrophy at Different Stages of Parkinson's and Alzheimer's Diseases and Relation to Cognition. *Brain Topogr*. 2019;32(1):142–60.
40. Farid N, Girard HM, Kemmotsu N, Smith ME, Magda SW, Lim WY, et al. Temporal Lobe Epilepsy: Quantitative MR Volumetry in Detection of Hippocampal Atrophy. *Radiology*. 1 de agosto de 2012;264(2):542–50.
41. Johnston D, Amaral DG. Hippocampus. En: *The Synaptic Organization of the Brain*. Fifth. Oxford University Press; 2004. p. 455–98.
42. Horn R, Ostertun B, Fric M, Solymosi L, Steudel A, Möller H-J. Atrophy of Hippocampus in Patients with Alzheimer's Disease and Other Diseases with Memory Impairment. *DEM*. 1996;7(4):182–6.
43. Lee K, Lee YM, Park J-M, Lee B-D, Moon E, Jeong H-J, et al. Right hippocampus atrophy is independently associated with Alzheimer's disease with psychosis. *Psychogeriatrics*. marzo de 2019;19(2):105–10.