



TRABAJO FIN DE MASTER

**DETECCIÓN DE LA
ENFERMEDAD DE ALZHEIMER
EN DISTINTAS ETAPAS DE
DEMENCIA MEDIANTE EL USO
DE IMÁGENES DE RESONANCIA
MAGNÉTICA**

JULIO 2025

Julia Martins Guardia
Javier Álvarez Liébana

Índice

1. Introducción	1
2. Objetivos	3
3. Datos	4
4. Extracción de las características de las MRI	6
5. Metodología	9
5.1. Selección de las variables	10
5.2. Balanceo de las clases	13
5.3. Validación cruzada	14
5.4. Algoritmos de clasificación	15
5.4.1. Árbol de decisión	16
5.4.2. Random Forest	17
5.4.3. Regresión Logística	17
5.5. Métricas de evaluación	19
6. Análisis de variables	21
6.1. Estudio de la correlación y colinealidad	21
6.1.1. Matriz de correlación	21
6.1.2. Análisis VIF	22
6.2. Test de Kruskal-Wallis	23
6.3. Test Post-hoc de Dunn	24
6.4. Análisis de componentes principales	25
7. Datos finales	28
8. Resultados	29
8.1. Árbol de decisión	29
8.1.1. Variables seleccionadas	29
8.1.2. PCA dirección radial	30
8.2. Random Forest	32
8.2.1. Variables seleccionadas	32
8.2.2. PCA dirección radial	33
8.3. Regresión Logística	34
8.3.1. Variables seleccionadas	34
8.3.2. PCA dirección radial	35

9. CONCLUSIONES	38
A. Anexos	43
B. Anexos	49

Resumen

El Alzheimer es una enfermedad neurodegenerativa cuya detección temprana es crucial para mejorar el pronóstico de los pacientes. De esta forma, este trabajo amplía el estudio realizado por Elisa Caballero [1], pasando de una clasificación binaria a una clasificación multiclase de *Imágenes de Resonancia Magnética* (IRM), con el objetivo de identificar los distintos estadios de la enfermedad.

En este contexto, el principal reto fue el desequilibrio entre clases, abordado mediante el uso combinado de las técnicas SMOTE y bootstrap con reposición. Además, se exploraron métricas de validación específicas para clasificación multiclase, a fin de evaluar con mayor precisión el rendimiento de los modelos.

La base de datos utilizada se extrae principalmente de *Alzheimer's Disease Neuroimaging Initiative* (ADNI) y contiene 6,400 imágenes MRI en 2D de distintas secciones del cerebro, correspondientes a diferentes etapas de la enfermedad de Alzheimer, lo que proporciona un conjunto de datos satisfactorio para el análisis y la validación de los modelos de clasificación. Contodo, en el trabajo previo, no se utilizaron directamente estas imágenes para la clasificación, sino que se realizó previamente la extracción de sus características mediante la *Matriz de Co-ocurrencia de Niveles de Gris* (GLCM). Por lo tanto, en el presente estudio se optó por seguir la misma línea metodológica y trabajar directamente con los datos ya generados en dicho estudio, centrando el análisis y el desarrollo de los modelos actuales en esas características previamente extraídas.

Por fin, se entrenaran diversos algoritmos de clasificación supervisada, como árboles de decisión, random forest y regresión logística, con el objetivo de clasificar las *imágenes de resonancia magnética* (IRM) en las distintas etapas de la enfermedad.

Palabras clave

Enfermedad de Alzheimer, *Imágenes por Resonancia Magnética* (IRM), Clasificación Multiclase, Aprendizaje Supervisado, *Matriz de Co-ocurrencia de Niveles de Gris* (GLCM).

Abstract

Alzheimer's disease is a neurodegenerative disorder whose early detection is crucial for improving the prognosis of patients. This Master's Thesis builds upon the study conducted by Elisa Caballero [1], shifting from binary classification to a multiclass classification of *Magnetic Resonance Imaging* (MRI) images, with the aim of identifying different stages of the disease.

In this way, the main challenge was the class imbalance, addressed through the combined use of SMOTE and bootstrap with replacement techniques. Additionally, specific validation metrics for multiclass classification were explored to more accurately assess the performance of the models.

The dataset used in this study is primarily sourced from *Alzheimer's Disease Neuroimaging Initiative* (ADNI) and contains 6,400 2D MRI images from different sections of the brain, corresponding to various stages of Alzheimer's disease, which provides a satisfactory dataset for the analysis and validation of the classification models. However, in the previous study, these images were not used directly for classification; instead, feature extraction was performed beforehand using the *Gray-Level Co-occurrence Matrix* (GLCM). Therefore, in the present study, we chose to follow the same methodological approach and work directly with the data already generated in that study, focusing the analysis and model development on those previously extracted features.

Finally, various supervised classification algorithms, such as decision trees, random forest, and logistic regression, were trained with the goal of classifying MRI images into the different stages of the disease.

Keywords

Alzheimer's Disease, *Magnetic Resonance Imaging* (MRI), Multiclass Classification, Supervised Learning, *Gray Level Co-occurrence Matrix* (GLCM).

1. Introducción

La *Enfermedad de Alzheimer* (EA) es un trastorno neurodegenerativo progresivo que afecta principalmente a personas mayores y se manifiesta a través de un deterioro cognitivo gradual, pérdida de memoria y alteraciones en el comportamiento. Se trata de la causa más común de demencia en adultos mayores de 60 años, y, dado que produce un daño neuronal irreversible, la detección temprana es fundamental para ralentizar su avance y mejorar la calidad de vida de los pacientes [2].

Fue descrita por primera vez a principios del siglo XX por el médico alemán Alois Alzheimer, quien identificó en su paciente Auguste Deter síntomas clínicos poco comunes junto con alteraciones cerebrales. Luego, en 1910, Emil Kraepelin creó el término “*Enfermedad de Alzheimer*” al incluirlo en una de sus clasificaciones sobre demencias seniles [3].

Aunque el origen exacto de la enfermedad aún no está completamente claro, la hipótesis más aceptada actualmente es la de la cascada amiloide-tau. Esta teoría sostiene que la acumulación del péptido beta-amiloide forma placas seniles extracelulares que alteran el funcionamiento neuronal, y posteriormente, la acumulación intracelular de la proteína tau en forma de ovillos neurofibrilares desestabilizan el citoesqueleto de las neuronas y contribuyen a su degeneración [4] [5].

Ambas acumulaciones son tóxicas y se diseminan progresivamente por el cerebro, causando la muerte de las neuronas y afectando regiones clave como la corteza entorrinal y el hipocampo, estructuras vinculadas a la memoria y otras funciones cognitivas esenciales [6] [7]. La acumulación de tau define los estadios de la enfermedad según las etapas de Braak [8], comenzando en la corteza entorrinal y expandiéndose hacia otras áreas, mientras que el beta amiloide se deposita inicialmente en el neocórtex y progresa hacia regiones internas como el hipocampo y el cerebelo, conforme a las etapas de Thal [9].

Estos daños alteran el funcionamiento de la memoria, que suele clasificarse en sensorial (localizada en la amígdala), a corto plazo (localizada en el córtex prefrontal) y a largo plazo (localizada en el hipocampo) [10]. En las primeras etapas del Alzheimer el neocórtex y la corteza entorrinal se ven afectados, deteriorando tanto la memoria a corto plazo como el procesamiento de información y desconectando la memoria a largo plazo del hipocampo [11].

Durante décadas, el diagnóstico definitivo del Alzheimer solo era posible mediante la autopsia tras la muerte, en la cual se observaban directamente las estructuras cerebrales dañadas. Con el tiempo, se desarrollaron pruebas psicométricas como el *Mini-Mental State Examination* [12], así como técnicas de neuroimagen como la *Magnetic Resonance Imaging* (IRM) y la *Positron Emission Tomography* (PET), además de la identificación de biomarcadores en el líquido cefalorraquídeo [13]. Estas técnicas, en conjunto, permiten un diagnóstico más temprano y preciso en vida [2].

En este contexto, y considerando el envejecimiento demográfico que afecta a España, donde cada vez hay una mayor proporción de personas mayores, los estudios sobre el Alzheimer se

vuelven particularmente relevantes desde una perspectiva social y sanitaria.

En este marco, la resonancia magnética (IRM) adquiere un papel central, siendo la principal herramienta utilizada en este estudio para la evaluación estructural del cerebro, lo que hace necesario comprender su funcionamiento.

Para ello, es importante entender que la obtención de imágenes mediante resonancia magnética (IRM) se basa en la interacción entre campos magnéticos y los núcleos de hidrógeno del cuerpo humano. Estos núcleos, al ser alineados por un campo magnético y luego estimulados por pulsos de radiofrecuencia, emiten señales al regresar a su estado original, las cuales son captadas por el escáner y transformadas en imágenes detalladas del interior del cuerpo, como el cerebro.

Entre los principales tipos de imágenes obtenidas se encuentran las imágenes T1 y T2, que ofrecen informaciones complementarias sobre los tejidos cerebrales [14]. En particular, las imágenes T1 (utilizadas en este estudio) permiten observar la sustancia gris en tonos de gris, mientras que el aire y los líquidos aparecen en negro, lo que facilita la identificación con mayor precisión de estructuras anatómicas y posibles anomalías.

Estas imágenes se almacenan en formatos estandarizados como *Digital Imaging and Communications in Medicine* (DICOM) o *Neuroimaging Informatics Technology Initiative* (NIFTI), ampliamente adoptados en entornos clínicos y de investigación por su capacidad para conservar información compleja y detallada de los estudios, incluyendo cortes en diferentes planos del cerebro, facilitando su procesamiento posterior [15].

Por fin, una de las grandes ventajas de la IRM es la posibilidad de obtener cortes bidimensionales (*slices*), que permiten examinar con precisión regiones cerebrales específicas. Los más frecuentes en estudios neurológicos son los cortes axiales (utilizadas en este estudio), que atraviesan el cerebro de forma horizontal de la frente hacia la nuca, y los cortes coronales, verticales de arriba hacia abajo. En el contexto del diagnóstico de la enfermedad de Alzheimer, los cortes axiales permiten detectar el aumento de los surcos cerebrales como signo de atrofia, mientras que los cortes coronales permiten evaluar la reducción del volumen del hipocampo, uno de los principales marcadores de esta patología.

No obstante, aunque la IRM ya proporciona información valiosa de forma visual, el avance en el uso de algoritmos automatizados para la clasificación de imágenes ha ganado relevancia en los últimos años. Estos sistemas permiten realizar evaluaciones más rápidas, objetivas y consistentes, apoyando a los profesionales de la salud en la detección precoz de la enfermedad y contribuyendo a mejorar la precisión diagnóstica.

De esta forma, el presente Trabajo de Fin de Máster propone avanzar en el uso de técnicas de clasificación supervisada aplicadas a imágenes de resonancia magnética, con el objetivo de identificar no solo la presencia de la enfermedad, sino también sus diferentes estadios clínicos.

2. Objetivos

El objetivo principal de este trabajo es aplicar técnicas de clasificación supervisada multiclase para detectar los distintos estadios del Alzheimer a partir de *Imágenes de Resonancia Magnética* (MRI) del cerebro. Para ello, se han seleccionado los algoritmos de clasificación: árboles de decisión, Random Forest y regresión logística.

Considerando que la efectividad de los modelos de clasificación depende en gran medida de la calidad de los datos disponibles, la etapa de extracción de características a partir de las imágenes adquiere un papel central como fuente de información relevante. En este sentido, uno de los principales desafíos en este proceso es la falta de alineación entre las imágenes de resonancia magnética, lo que introduce variabilidad en su orientación y en el plano de corte. De esta forma, para mitigar este efecto, un estudio previo empleó la *Matriz de Co-ocurrencia de Niveles de Gris* (GLCM), reconocida por su capacidad para capturar información textural de manera robusta, incluso en ausencia de una alineación precisa [1]. En coherencia con este enfoque, el presente trabajo retoma dicha estrategia metodológica, utilizando directamente las características previamente extraídas de las imágenes.

Asimismo, la elección de modelos clásicos (tanto métodos tradicionales como no tradicionales) frente a redes neuronales profundas se justifica por los siguientes factores:

- **Interpretabilidad:** Estos modelos permiten visualizar cómo las variables influyen en las decisiones relacionadas con el diagnóstico final, lo que garantiza que los resultados sean comprensibles para los profesionales de la salud. No obstante, algunos modelos de machine learning pueden requerir un análisis más profundo para lograr una interpretación clara de sus resultados.
- **Simplicidad y Eficiencia computacional:** Los modelos clásicos son generalmente más fáciles de implementar y demandan menos recursos computacionales en términos de procesamiento y memoria, lo que los convierte en opciones más accesibles, especialmente en entornos con recursos limitados.

Estas características resultan fundamentales en contextos clínicos, donde se valoran herramientas que proporcionen resultados rápidos y comprensibles para respaldar el diagnóstico médico.

Finalmente, con el fin de abordar el fuerte desbalance de clases presente en los datos, se ha optado por una estrategia combinada de sobremuestreo utilizando SMOTE y Bootstrap con reposición. Esta combinación busca mejorar la representatividad de las clases minoritarias y equilibrar el conjunto de entrenamiento, favoreciendo así un desempeño de los modelos de clasificación más estable y justo.

3. Datos

Los datos analizados provienen de la plataforma *Kaggle*, que proporciona acceso a diversas bases de datos, además de facilitar el intercambio de código y recursos educativos [16]. La colección utilizada contiene imágenes de resonancia magnética (MRI) obtenidas de múltiples repositorios, siendo *Alzheimer's Disease Neuroimaging Initiative* (ADNI) la fuente principal.

Fundado en 2004, ADNI es uno de los repositorios más reconocidos en la investigación del Alzheimer, centrado en el estudio de la progresión de la EA mediante diversas tecnologías de imágenes y biomarcadores [17]. Contodo, las imágenes de MRI disponibles en este repositorio se ofrecen sin preprocesamiento, por lo que requieren una selección manual de los slices de interés.

En cambio, las imágenes disponibles en *Kaggle* ya están preprocesadas por profesionales de la salud, por lo que incluyen la selección del slice más representativo y la conversión a formato JPG. Esta preparación resulta especialmente útil para el presente estudio, ya que permite aplicar directamente técnicas de clasificación sin necesidad de un preprocesamiento de los datos.

De esta forma, la base de datos utilizada consta de 6,400 *Imágenes de Resonancia Magnética* (MRI) en dos dimensiones, correspondientes a diferentes secciones del cerebro y a diversas etapas de la enfermedad de Alzheimer. Estas imágenes se distribuyen en cuatro estadios de la enfermedad, en orden creciente de severidad, por lo que se puede observar un aumento progresivo de los surcos cerebrales, como se muestra en la siguiente figura:

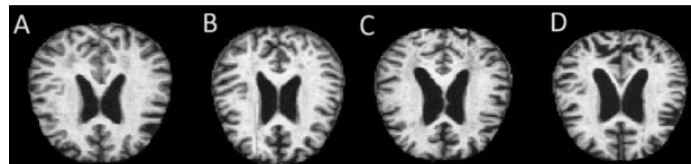


Figura 1: MRI de la categorías del desarrollo de la EA. A) *Non Demented* (3,200 imágenes). B) *Very Mild Demented* (2,240 imágenes). C) *Mild Demented* (896 imágenes). D) *Moderate Demented* (64 imágenes).

Estas imágenes están distribuidas en los distintos estadios de la enfermedad de la siguiente manera:

- ***Non Demented*:** 3,200 imágenes.
- ***Very Mild Demented*:** 2,240 imágenes.
- ***Mild Demented*:** 896 imágenes.
- ***Very Mild Demented*:** 64 imágenes.

Con esto, el primer paso en la extracción de características consiste en obtener las matrices de intensidades de los píxeles de las MRI. Para ello, se considera que cada imagen, en escala de

grises, tiene una resolución de 128×128 píxeles, donde cada píxel posee un valor de intensidad que varía entre 0 (negro) y 256 (blanco). De este modo, es posible construir una matriz de 128×128 que representa los valores de intensidad correspondientes a cada píxel de la imagen [1].

Sin embargo, dado que las imágenes corresponden a cortes obtenidos en distintas alturas del cerebro, no están alineadas espacialmente, lo que complica su comparación directa y requiere el uso de un método más robusto para la extracción de las características relevantes. Por esta razón, se optó por emplear la *Matriz de Co-ocurrencia de Niveles de Gris* (GLCM), una técnica capaz de capturar eficazmente las propiedades texturales de las imágenes, independientemente de su alineación, lo que permite una clasificación más precisa y consistente [1].

4. Extracción de las características de las MRI

Como mencionado anteriormente, la extracción de las características de las imágenes se llevó a cabo mediante el uso de la *Matriz de Co-ocurrencia de Niveles de Gris* (GLCM).

Este metodo permite capturar patrones texturales en las imágenes al contabilizar la frecuencia con la que ciertas combinaciones de niveles de gris aparecen entre píxeles adyacentes en una dirección específica, lo que resulta especialmente útil en contextos como el diagnóstico por imagen, ya que resume de forma cuantitativa la estructura interna del tejido cerebral.

En este sentido, cada matriz GLCM generada es simétrica y cuadrada, de dimensión $(N \times N)$ 256×256 , correspondiente al número de niveles de intensidad de gris (de 0 a 256), donde cada elemento M_{ij} representa cuántas veces un píxel de intensidad i aparece adyacente (en ambos sentidos) a un píxel de intensidad j , según la dirección especificada [1].

Posteriormente, esta matriz es normalizada dividiendo sus elementos por el número total de ocurrencias, lo que da lugar a una matriz P , donde P_{ij} representa la frecuencia relativa de co-ocurrencia entre niveles de gris, es decir, la proporción con la que un píxel con nivel de gris i se encuentra adyacente a otro de nivel j [18].

A respecto de la dirección (*shift*), esta se representa mediante un vector $c_{kl} = (c_{kl,x}, c_{kl,y})$, donde $c_{kl,x}$ indica el desplazamiento en el eje x y $c_{kl,y}$ en el eje y [1] [18]. Se pueden ver un ejemplo de las direcciones consideradas en el calculo de la matriz GLCM en la **Figura 2**.

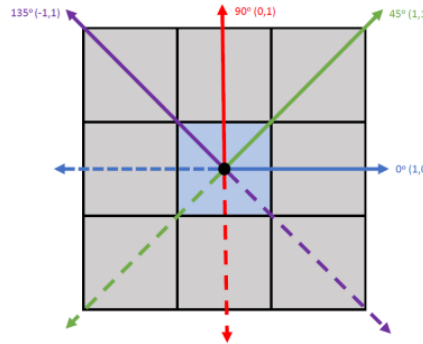


Figura 2: Direcciones (shifts) en las que se pueden contar los píxeles adyacentes para la construcción de la matriz GLCM.

De esta forma, las direcciones consideradas en la construcción de las matrices GLCM a partir de la función `GLCMTextures::make_glcm` han sido las siguientes [18]:

- $c_{kl} = (1, 0)$: el entorno es el píxel de la derecha y su simétrico.
- $c_{kl} = (0, 1)$: el entorno es el píxel de arriba y su simétrico.
- $c_{kl} = (1, 1)$: el entorno es el píxel de la diagonal superior derecha y su simétrico.
- $c_{kl} = (1, -1)$: el entorno es el píxel de la diagonal inferior derecha y su simétrico.

Es fundamental destacar que, en todas las direcciones consideradas para el cálculo de la matriz GLCM, los píxeles adyacentes se contabilizan en ambos sentidos, lo que asegura una representación simétrica y más precisa de los patrones texturales [1] [18].

Además, como complemento a estas direcciones, se empleó una dirección radial, la cual integra simultáneamente todas las orientaciones consideradas y permite capturar de manera más completa y detallada las irregularidades estructurales que podrían estar asociadas a los distintos estadios de la EA [1].

A continuación, una vez construida la matriz GLCM empleando estas direcciones y tras su normalización, se procedió a la extracción de métricas estadísticas mediante la función `glcm_metrics()` del paquete **GLCMTextures**, una herramienta especializada en la construcción de matrices GLCM y en el cálculo automático de sus métricas asociadas [1] [18].

Finalmente, como resultado de este proceso, se obtuvieron nueve métricas, se agrupan según el tipo de característica textural que representan, tal como se muestra a continuación:

1. Métricas de contraste:

- **Contrast:** Mide la variación de niveles de intensidad entre un píxel y sus vecinos, asignando más peso a los pares de píxeles con diferencias de intensidad mayores. El valor de esta variable aumenta cuando hay grandes diferencias de intensidad en la imagen, destacando bordes y detalles finos.

$$\text{Contraste} = \sum_{i,j=0}^{N-1} P_{ij} \cdot (i - j)^2 \quad (1)$$

- **Dissimilarity:** Es similar al contraste pero considera la diferencia absoluta entre los niveles de gris, sin elevarla al cuadrado. Esta variable es sensible a las diferencias de intensidad y valores altos indican un aumento de heterogeneidad en la textura.

$$\text{Dissimilaridade} = \sum_{i,j=0}^{N-1} P_{ij} \cdot |i - j| \quad (2)$$

- **Homogeneity:** Mide la proximidad de los elementos de la matriz GLCM a la diagonal principal. Esta métrica resalta las regiones de la imagen con texturas suaves y uniformes, asignando más peso a los pares de píxeles con diferencias de intensidad menores [19] [18].

$$\text{Homogeneidade} = \sum_{i,j=0}^{N-1} \frac{P_{ij}}{1 + (i - j)^2} \quad (3)$$

2. Métricas de orden:

- **Angular Second Moment (ASM):** También conocida como energía, esta métrica mide la uniformidad o regularidad en la textura de la imagen. Un valor alto indica una textura más homogénea y ordenada, mientras que un valor bajo sugiere una textura más compleja y desordenada [19].

$$ASM = \sum_{i,j=0}^{N-1} P_{ij}^2 \quad (4)$$

- **Entropy:** Mide la aleatoriedad y la cantidad de información en la imagen. Un valor alto indica una mayor complejidad y variabilidad en la textura, mientras que un valor bajo sugiere una textura más uniforme y predecible [18].

$$Entropia = \sum_{i,j=0}^{N-1} P_{ij} \cdot [-\ln(P_{ij})], \quad 0 \cdot \ln(0) = 0 \quad (5)$$

3. Métricas descriptivas:

- **Mean:** Corresponde al promedio de las intensidades ponderadas por su probabilidad relativa de coocurrencia.

$$\mu = \sum_{i,j=0}^{N-1} i \cdot P_{ij} \quad (6)$$

- **Variance:** Corresponde a varianza de las intensidades ponderadas por la probabilidad relativa de co-ocurrencia.

$$\sigma^2 = \sum_{i,j=0}^{N-1} P_{ij} \cdot (i - \mu)^2 \quad (7)$$

- **Correlation:** Evalúa la relación lineal entre los niveles de gris de los píxeles en la imagen. Un valor alto de correlación indica una relación lineal positiva entre los píxeles, lo que sugiere una textura más uniforme y estructurada [18].

$$Correlation = \sum_{i,j=0}^{N-1} P_{ij} \frac{(i - \mu)(j - \mu)}{\sigma^2} \quad (8)$$

Además se calcula una novena métrica llamada suma media:

$$SA = \left(\sum_{k=0}^{2(N-1)} k \cdot kprob \right), \quad \text{com } k = i + j \quad (9)$$

5. Metodología

En este estudio, se ha utilizado el lenguaje de programación R para todas las etapas del proceso, desde el análisis de las variables hasta la clasificación y evaluación de los modelos.

También cabe destacar que la extracción de características descrita en el apartado anterior fue realizada originalmente por Elisa Caballero en un trabajo previo, y los datos derivados de ese proceso han sido reutilizados en este estudio, enmarcándose así como una continuación metodológica directa [1].

Con base en lo mencionado, en el presente estudio se han seguido los siguientes pasos metodológicos:

1. **Selección de las variables:** Se llevó a cabo un análisis exploratorio con el fin de seleccionar las variables predictoras más relevantes, reduciendo la colinealidad entre las métricas GLCM y optimizando así el rendimiento de los algoritmos de clasificación.
2. **Balanceo de las clases:** Dado el fuerte desequilibrio entre las clases en la base de datos, se aplicaron técnicas de remuestreo con el fin de garantizar una representación equitativa. Para eso, se utilizó una combinación de *Synthetic Minority Over-sampling Technique* (SMOTE) y Bootstrap con reposición, lo que permitió aumentar la presencia de las clases minoritarias sin comprometer la diversidad de los datos.
3. **Clasificación:** En la etapa de clasificación, se emplearon diversos algoritmos supervisados con el objetivo de categorizar las imágenes en función de las características extraídas. De este modo, entre los modelos seleccionados se incluyen árboles de decisión, random forest y regresión logística escogidos por su simplicidad, interpretabilidad y eficiencia computacional. Además, en los algoritmos que así lo permiten, se implementó validación cruzada para optimizar la selección de hiperparámetros y mejorar el rendimiento general de los modelos.
4. **Evaluación de modelos:** El mejor modelo seleccionado en la validación cruzada se entrenó utilizando los datos de entrenamiento y se evaluó tanto con los datos de entrenamiento como con los datos de validación y prueba para asegurar su precisión y comprobar su capacidad de generalización.

5.1. Selección de las variables

Para optimizar la construcción de los modelos propuestos, se llevaron a cabo diversos análisis destinados a identificar patrones entre las variables, con el fin de seleccionar las más relevantes. Esta selección, además, contribuyó a mitigar la colinealidad existente entre las variables predictoras (métricas GLCM).

A continuación, se describen los métodos empleados para tal propósito:

1. Matriz de correlación:

Es una forma de analizar la relación lineal entre cada par de variables del conjunto de datos. En esta matriz, cada elemento representa el coeficiente de correlación de Pearson calculado entre dos variables, cuyo valor oscila entre -1 y 1. Un valor cercano a 1 indica una fuerte correlación positiva, un valor próximo a -1 señala una fuerte correlación negativa, y un valor cercano a 0 sugiere la ausencia de correlación lineal entre las variables consideradas [20]. La fórmula del coeficiente de correlación de Pearson entre dos variables X y Y se expresa como [20]:

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (10)$$

Donde:

- r es el coeficiente de correlación de Pearson,
- X_i y Y_i son los valores individuales de las variables X y Y ,
- \bar{X} y \bar{Y} son las medias de X y Y ,
- n es el número total de observaciones.

2. Análisis del *Factor de Inflación de la Varianza* (VIF):

Es una técnica utilizada para identificar la presencia de multicolinealidad entre las variables predictoras en un modelo de regresión. Se calcula ajustando un modelo de regresión lineal en el que cada variable predictora se considera, de forma individual, como variable dependiente, y las restantes como variables explicativas. De esta forma, un valor elevado de VIF indica que una variable puede ser explicada en gran medida por una combinación lineal de las otras, lo que sugiere que podría eliminarse para simplificar el modelo y reducir la varianza de las predicciones [21]. La fórmula para calcular el VIF de una variable X_j es [21]:

$$\text{VIF}_j = \frac{1}{1 - R_j^2} \quad (11)$$

Siendo R_j^2 el coeficiente de determinación obtenido al ajustar la regresión de X_j sobre las demás variables predictoras, el cual se calcula mediante la siguiente fórmula:

$$R_j^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}} \quad (12)$$

En la cual:

$$SS_{\text{res}} = \sum_{i=1}^n (X_{j,i} - \hat{X}_{j,i})^2 \quad \text{y} \quad SS_{\text{tot}} = \sum_{i=1}^n (X_{j,i} - \bar{X}_j)^2 \quad (13)$$

Donde:

- SS_{res} es la suma de los cuadrados de los residuos (errores) del modelo,
- SS_{tot} es la suma total de los cuadrados (variabilidad total de la variable),
- $X_{j,i}$ es el valor observado de la variable X_j en la observación i ,
- $\hat{X}_{j,i}$ es el valor estimado por el modelo de regresión para la observación i ,
- \bar{X}_j es la media de los valores observados de X_j ,
- n es el número total de observaciones.

3. Test de Kruskal-Wallis:

Es una prueba estadística no paramétrica utilizada para comparar tres o más grupos independientes, especialmente cuando los datos no cumplen los supuestos de normalidad o homocedasticidad necesarios para aplicar el análisis de varianza (ANOVA). El test evalúa si existe una diferencia significativa entre las medianas de los grupos, basándose en rangos en lugar de valores absolutos. Por lo tanto, si el resultado del test es significativo, se concluye que al menos uno de los grupos difiere de los demás, aunque no especifica cuáles grupos son los que presentan diferencias [22]. La estadística de este test se calcula de la siguiente manera [22]:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1) \quad (14)$$

Donde:

- k es el número de grupos,
- n_i es el tamaño del grupo i ,
- $N = \sum_{i=1}^k n_i$ es el número total de observaciones,
- R_i es el rangos del grupo i .

4. Test Post-hoc de Dunn:

Es una prueba un poco más específica que el test de Kruskal-Wallis, ya que se utiliza para identificar qué grupos difieren significativamente entre sí. La estadística del test utilizada para comparar dos grupos j y l se calcula de la siguiente manera [23] [22] [24]:

$$z = \frac{R_j - R_l}{\sqrt{\frac{N(N+1)}{12} \left(\frac{1}{n_j} + \frac{1}{n_l} \right)}} \quad (15)$$

Donde:

- R_j y R_l son las medias de los rangos de los grupos j y l ,
- n_j y n_l son los tamaños de muestra de los grupos respectivos,
- N es el tamaño total de la muestra.

5. Análisis de Componentes Principales (PCA):

Es una técnica de reducción de dimensionalidad que transforma un conjunto de variables correlacionadas en un nuevo conjunto de variables no correlacionadas. Estas nuevas variables, denominadas componentes principales, son combinaciones lineales de las variables originales, y permiten representar los datos de manera más compacta, manteniendo la mayor cantidad posible de información.

Para aplicar esta técnica, se calcula la matriz de covarianza, a partir de la cual se obtienen los autovalores y autovectores de la matriz. Los autovectores representan las direcciones de los nuevos ejes, mientras que los autovalores corresponden a la cantidad de varianza que cada componente explica.

Los autovectores se clasifican de acuerdo con sus autovalores, en orden decreciente, lo que permite identificar los componentes principales que retienen la mayor parte de la variabilidad de los datos. Por último, los datos originales se proyectan en una nueva base compuesta por los (k) componentes principales, resultando en un nuevo conjunto de datos con dimensiones reducidas.

5.2. Balanceo de las clases

Después de la evaluación y selección de las variables, y antes de iniciar el desarrollo de los modelos de clasificación, se realizó un balanceo de las clases en todo el conjunto de datos disponible. Esto tuvo como objetivo mitigar los efectos del desbalance en la distribución de categorías, que podrían comprometer la capacidad predictiva de los modelos y generar sesgos hacia las clases mayoritarias.

Para eso, inicialmente se consideró el uso exclusivo del método Bootstrap. Sin embargo, esta opción fue descartada debido al alto desbalance en las muestras. En este contexto, el principal inconveniente de utilizar solo Bootstrap es que, al generar nuevas muestras con reposición, podría haber una repetición excesiva de los mismos datos de la clase minoritaria. Esto podría llevar a un sobreajuste del modelo a esa clase, dando menor peso a las clases mayoritarias, lo que resultaría en una representación inadecuada de la distribución real de los datos y generaría un sesgo que afectaría negativamente la capacidad de generalización del modelo.

Por esta razón, se decidió no proceder exclusivamente con Bootstrap. En su lugar, se empleó inicialmente el método SMOTE, que genera muestras sintéticas para las clases minoritarias, equilibrando de manera más efectiva la distribución de las clases [25] [26] [27]. Sin embargo, dado que este método por sí solo no alcanzó un equilibrio completo de las clases, el balanceo se complementó mediante el uso de Bootstrap con reposición, utilizando como tamaño máximo de muestra el correspondiente a la clase mayoritaria.

Este enfoque facilitó la creación de un entorno más robusto para el entrenamiento de los modelos, ya que permitió asegurar que el modelo tuviera datos representativos de todas las clases, sin introducir sesgo debido a la sobreabundancia de una clase minoritaria.

A continuación, se describe los métodos utilizados con mayor detalle:

1. ***Synthetic Minority Over-sampling Technique (SMOTE)***: Este método consiste en generar nuevas instancias sintéticas de la clase minoritaria mediante la interpolación entre muestras existentes y sus vecinos más cercanos [25] [26] [27].
2. ***Bootstrap***: Este método consiste en generar múltiples subconjuntos a partir de los datos originales mediante muestreo aleatorio con reposición, es decir, seleccionando observaciones al azar que pueden repetirse en cada muestra generada. Esta estrategia permite simular la variabilidad inherente a los datos sin necesidad de recopilar nuevos datos, facilitando la evaluación de la estabilidad y la generalización de los modelos predictivos.

5.3. Validación cruzada

Tras el balanceo de las clases, se procedió al particionamiento de los datos. De este modo, se aplicó un muestreo aleatorio estratificado por clase, dividiendo el conjunto disponible en un 80 % para el entrenamiento del modelo y el 20 % restante para su evaluación en el conjunto de prueba.

A continuación, y antes de la validación cruzada, se definieron grids de hiperparámetros para los algoritmos de clasificación, con el objetivo de optimizar el modelo en función del valor de la accuracy. Esta búsqueda exhaustiva de hiperparámetros se realizó para los algoritmos de Árbol de decisión y Random Forest, mientras que, por su naturaleza más simple, la regresión logística no requirió tal proceso debido a que sus configuraciones suelen ser más limitadas y estables.

En este contexto, se denomina *grid* a una estructura que contiene distintas combinaciones de hiperparámetros específicos para cada algoritmo. Estos hiperparámetros, por ser valores definidos antes del entrenamiento, afectan directamente el comportamiento y rendimiento del modelo. Por lo tanto, probar múltiples combinaciones permite identificar la configuración que logra el mejor equilibrio entre precisión y capacidad de generalización.

Una vez definidos los grids de hiperparámetros, se aplicó la validación cruzada k-fold con el objetivo de garantizar una evaluación robusta y fiable, minimizando el sesgo causado por la partición inicial de los datos, lo que elimina la influencia del azar en la evaluación del modelo.

El procedimiento consiste en dividir el conjunto de datos en k particiones y en cada iteración, una partición se reserva para validación, mientras las k-1 restantes se utilizan para entrenamiento del modelo. En este caso, se empleó una validación cruzada de cinco particiones con tres repeticiones, resultando en 15 modelos distintos por combinación de hiperparámetros [28].

Finalmente, se calcula el promedio de las precisiones obtenidas en las 15 ejecuciones correspondientes a cada configuración evaluada, seleccionando aquella que haya mostrado el mejor rendimiento promedio.

5.4. Algoritmos de clasificación

En esta sección se presentan los algoritmos supervisados empleados para la clasificación de las MRI.

Es importante destacar que, además de los modelos descritos en esta sección, también se probaron los algoritmos *K-Nearest Neighbors* (KNN) y *Extreme Gradient Boosting* (XGBoost). Sin embargo, ambos presentaron dificultades de convergencia y, por este motivo, no fueron incluidos en los resultados finales.

En el caso de KNN, las limitaciones observadas se atribuyen principalmente al incremento en el número de categorías en comparación con el estudio previo [1]. En contraste con el enfoque anterior, en el que la categorización se realizó exclusivamente entre demencia y no demencia, el presente estudio aborda la predicción de los distintos estadios del Alzheimer, clasificados como: *Non Demented*, *Very Mild Demented*, *Mild Demented* y *Moderate Demented*.

En síntesis, el KNN se fundamenta en la comparación de cada observación con sus k vecinos más cercanos mediante el cálculo de la distancia entre puntos, exhibiendo una mayor eficacia en contextos de baja complejidad y con un número reducido de clases, donde las fronteras entre categorías son evidentes y los datos están adecuadamente distribuidos. No obstante, conforme se incrementa la cantidad de clases o la dimensionalidad del conjunto de datos, el algoritmo se ve influenciado por lo que se denomina como *curse of dimensionality* (la maldición de la dimensionalidad). Este fenómeno implica que los datos se vuelven más dispersos y las distancias entre puntos pierden capacidad discriminativa, lo que dificulta la identificación precisa de los vecinos más próximos y la capacidad del KNN para establecer límites claros de decisión, reduciendo significativamente la eficacia del modelo en contextos multiclase más complejos, como el abordado en este estudio.

Por su parte, XGBoost (Extreme Gradient Boosting) a pesar de ser reconocido entre otras técnicas de machine learning por su alta capacidad discriminativa, su robustez y por no verse afectado por problemas de multicolinealidad, suele mostrar un mejor desempeño en bases de datos extensas y con estructuras más complejas.

De esta manera, en conjuntos de datos pequeños o con un número limitado de variables explicativas, como es el caso de este estudio (que dispone únicamente de nueve variables), los beneficios de XGBoost tienden a ser menos significativos, dado que el riesgo de overfitting aumenta considerablemente. En estas circunstancias, el modelo puede ajustarse en exceso a las fluctuaciones aleatorias del reducido conjunto de datos, en lugar de capturar patrones reales y generalizables. Además, la escasez de datos puede comprometer el proceso de generación de divisiones relevantes en los árboles, dificultando la convergencia del algoritmo y resultando en inestabilidad durante el entrenamiento.

5.4.1. Árbol de decisión

Los árboles de decisión son algoritmos basados en un proceso jerárquico de toma de decisiones, en el cual un conjunto de datos se divide en subgrupos según variables relevantes, con el objetivo de llegar a una predicción final. Este método se destaca por ser visual y de fácil interpretación, y aunque se utiliza en este caso para clasificación, también puede aplicarse en problemas de regresión.

Las partes principales de un árbol de decisión son las siguientes:

- **Nodo raíz:** Es el punto inicial del árbol, que representa el conjunto de datos completo.
- **Nodos internos:** Cada uno representa una división de los datos en base a una característica específica.
- **Ramas:** Son las líneas que conectan los nodos y representan las características que determinan cada división.
- **Nodos hoja:** Representan el resultado final de la predicción para cada subgrupo de datos.

El inicio de la clasificación ocurre en el nodo raíz del árbol, donde comienza la separación de los datos según la variable que el algoritmo considera más importante, la cual suele ser la que mejor discrimina las clases. A partir de esta primera división, los datos avanzan por las ramas correspondientes según el valor de la variable evaluada en cada paso. Este proceso continúa hasta alcanzar un nodo hoja, que es donde se asigna la clase más frecuente entre los datos que llegaron a ese nodo.

Cabe destacar que estas divisiones se realizan según un criterio predefinido, lo cual permite al algoritmo seleccionar el punto de corte óptimo para la variable correspondiente, maximizando así la discriminación entre las clases en cada nodo.

En este caso, la métrica empleada para evaluar la calidad de las divisiones del árbol fue la máxima accuracy. En este contexto, una accuracy alta indica un nodo puro, en el que la mayoría de las observaciones pertenecen a la misma clase, mientras que una accuracy baja señala un nodo impuro, con observaciones mezcladas entre diferentes clases [29].

Finalmente, los hiperparámetros que se pueden ajustar dentro de los árboles de decisión los siguientes:

- **Profundidad:** Controla la profundidad máxima del árbol, lo que limita el número de divisiones que puede tener el árbol.
- **Número mínimo:** Establece el número mínimo de observaciones que debe tener cada nodo final (hoja) tras una división.
- **Complejidad:** Controla la poda del modelo, ayudando a equilibrar el error de ajuste (es decir, las clasificaciones incorrectas realizadas por el modelo) y la complejidad del árbol

(la cantidad de nodos que posee). La poda es un proceso en el que los nodos menos importantes son eliminados del árbol para evitar que se vuelva excesivamente complejo. Al ajustar la complejidad, el objetivo es prevenir el sobreajuste, que ocurre cuando el modelo aprende demasiado bien los datos de entrenamiento, pero tiene dificultades para generalizar con datos nuevos. Así, al controlar la complejidad, el modelo puede volverse más simple y más capaz de hacer predicciones precisas en datos que nunca ha visto antes.

5.4.2. Random Forest

El Random Forest es un método de clasificación supervisada que se caracteriza por combinar múltiples árboles de decisión [30].

En términos prácticos, cada árbol que compone el algoritmo se construye a partir de una muestra aleatoria con reposición (bootstrap) extraída del conjunto de datos original. Además, en cada nodo del árbol, la división se realiza considerando únicamente un subconjunto aleatorio de las variables disponibles, lo que incrementa la diversidad entre los árboles y contribuye a una mayor robustez del modelo [29].

Una vez finalizado el entrenamiento de todos los árboles, la predicción final para una nueva observación se determina mediante una votación mayoritaria, en la cual cada árbol emite un “voto” por una clase, y la clase con el mayor número de votos es seleccionada como la predicción del modelo.

Este enfoque mejora la precisión y estabilidad del modelo, superando las limitaciones de los árboles de decisión individuales al reducir el riesgo de sobreajuste y mejorar la capacidad de generalización.

Los hiperparámetros ajustables en este algoritmo son los siguientes:

- **Número de árboles:** La cantidad de árboles de decisión que se construyen en el modelo.
- **Número de predictores aleatorios:** El número de variables aleatorias que se consideran en cada división de cada nodo de un árbol.
- **Tamaño mínimo del nodo:** El tamaño mínimo que cada nodo final (hoja) debe tener.

5.4.3. Regresión Logística

La regresión logística es una técnica estadística habitualmente utilizada para modelar la probabilidad de ocurrencia de un evento binario en función de una combinación lineal de variables predictoras continuas. Sin embargo, cuando la variable dependiente tiene más de dos categorías, como es el caso, se recurre a una extensión del modelo conocida como regresión logística multinomial, la cual permite modelar resultados categóricos con múltiples clases.

De esta forma, dado que el objetivo del modelo es estimar probabilidades, para que estas estimaciones sean válidas, deben estar siempre entre 0 y 1. No obstante, una combinación lineal

simple de las variables independientes, es decir, la suma de estas variables ponderadas por sus coeficientes, puede generar valores que varían desde $-\infty$ hasta $+\infty$, lo cual no es adecuado para representar probabilidades.

Para solucionar este problema, el método aplica una transformación matemática sobre la combinación lineal de las variables independientes, la cual convierte cualquier número real en un valor comprendido entre 0 y 1. Esta transformación, conocida como *logit*, se define como el logaritmo de la razón de probabilidades (*odds*) y se expresa mediante la siguiente fórmula [31]:

$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n = z \quad (16)$$

Donde $\beta_0, \beta_1, \dots, \beta_n$ son los coeficientes del modelo y X_1, X_2, \dots, X_n las variables independientes.

Luego, para obtener la probabilidad p a partir de la combinación lineal z , se aplica la función logística, que es la inversa del logit:

$$p = \frac{1}{1 + e^{-z}} \quad (17)$$

Particularmente, en la regresión logística multinomial, se asume que la variable dependiente sigue una distribución multinomial condicional. En este contexto, el modelo selecciona una de las categorías como referencia y estima, para cada clase k distinta de la de referencia, la siguiente relación:

$$\log\left(\frac{P(Y = k)}{P(Y = \text{ref})}\right) = \beta_{0k} + \beta_{1k}x_1 + \cdots + \beta_{pk}x_p. \quad (18)$$

Donde $P(Y = k)$ representa la probabilidad de que la observación pertenezca a la clase k , y $P(Y = \text{ref})$ es la probabilidad asociada a la categoría de referencia. Los coeficientes β_{jk} capturan el efecto de cada variable independiente x_j sobre la probabilidad relativa de pertenecer a la clase k en comparación con la referencia.

Esta formulación permite analizar el efecto de cada variable sobre la probabilidad relativa de pertenecer a cada categoría frente a la categoría de referencia.

5.5. Métricas de evaluación

Para evaluar de manera rigurosa el desempeño de los modelos de clasificación implementados en este trabajo, fue necesario establecer un protocolo sistemático de validación. Para eso, como se mencionó previamente, el conjunto de datos fue dividido inicialmente en dos muestras, una de entrenamiento y otra de prueba. No obstante, con el fin de enriquecer el análisis y obtener estimaciones más representativas del rendimiento, se incorporó además una etapa de validación cruzada aplicada exclusivamente sobre la muestra de entrenamiento.

En este contexto, cabe destacar que la validación cruzada, aunque previamente utilizada en este trabajo como herramienta para la selección de modelos, fue empleada aquí con un propósito distinto: evaluar con mayor fiabilidad la estabilidad y consistencia del rendimiento del modelo dentro del conjunto de entrenamiento, proporcionando así una visión más completa de su comportamiento. Esta estrategia permitió evaluar su comportamiento en múltiples particiones internas, facilitando no solo el cálculo de métricas promedio, sino también el análisis de su variabilidad.

Con base en esta estrategia, se emplearon diversas métricas de evaluación en todas las muestras consideradas (entrenamiento, validación y prueba), con el fin de cuantificar distintos aspectos del rendimiento de los modelos de clasificación multiclase. A continuación, se describen las métricas utilizadas, junto con su interpretación, aplicabilidad en contextos multinomiales y sus respectivas fórmulas:

- **Accuracy:**

Esta métrica mide la exactitud, o sea, la proporción de observaciones correctamente clasificadas sobre el total de casos evaluados. Particularmente, en el ámbito de variables multiclase, esta métrica se calcula como la proporción total de predicciones correctas sobre el número total de observaciones, sumando los aciertos en todas las clases según la matriz de confusión. A continuación, se presenta la fórmula correspondiente:

$$\text{Accuracy} = \frac{\sum_{i=1}^k C_{ii}}{\sum_{i=1}^k \sum_{j=1}^k C_{ij}} \quad (19)$$

Donde C_{ii} representa los elementos en la diagonal principal de la matriz de confusión (aciertos) y k es el número de clases.

- **Coeficiente Kappa de Cohen:**

El coeficiente Kappa evalúa la concordancia entre las predicciones del modelo y las etiquetas reales, ajustando por el acuerdo esperado por azar, permitiendo así detectar si el modelo asigna correctamente las clases más allá de hacerlo por simple coincidencia. De esta forma, en el contexto de clasificación multiclase, esta métrica resulta particularmente útil, ya que proporciona una medida más robusta del desempeño que la simple exactitud,

especialmente cuando existe un desbalance en la distribución de clases. A continuación, se presenta la fórmula correspondiente:

$$\kappa = \frac{p_o - p_e}{1 - p_e} \quad (20)$$

Donde p_o es la proporción de acuerdo observado y p_e es la proporción de acuerdo esperado por azar. Es importante resaltar que este índice varía entre -1 y 1, en que un valor cercano a 1 indica una concordancia perfecta entre las clasificaciones, uno próximo a 0 refleja el acuerdo esperado por azar, y un valor de -1 representa una discordancia sistemática.

■ **Sensibilidad:**

La sensibilidad mide la capacidad del modelo para identificar correctamente los casos de una clase específica. Particularmente, en el contexto multinomial, la sensibilidad se calcula para cada clase individualmente y luego se obtiene su promedio general. A continuación, se presenta la fórmula correspondiente:

$$\text{Sensibilidad}_i = \frac{TP_i}{TP_i + FN_i} \quad (21)$$

Donde TP_i es el número de verdaderos positivos y FN_i el número de falsos negativos cuando se evalúa la clase i frente a todas las demás.

■ **Especificidad:**

La especificidad evalúa la proporción de verdaderos negativos correctamente identificados. Particularmente, en el contexto multinomial, la especificidad se calcula para cada clase individualmente y luego se obtiene su promedio general. A continuación, se presenta la fórmula correspondiente:

$$\text{Especificidad}_i = \frac{TN_i}{TN_i + FP_i} \quad (22)$$

Donde TN_i es el número de verdaderos negativos y FP_i los falsos positivos cuando se evalúa la clase i frente a todas las demás.

6. Análisis de variables

6.1. Estudio de la correlación y colinealidad

6.1.1. Matriz de correlación

Tal como se mencionó anteriormente, la información obtenida a partir del cálculo de la matriz GLCM en la dirección radial fue seleccionada por su capacidad para representar de manera resumida y eficaz las demás direcciones, lo que facilita una evaluación inicial más ágil y manejable.

Sin embargo, es importante destacar que este resultado fue validado mediante el análisis de la matriz de correlación, considerando inicialmente todas las direcciones y, posteriormente, únicamente la dirección radial. De esta forma, al igual que en el trabajo anterior, se comprobó que el patrón de correlación entre métricas se mantiene constante independientemente de la dirección (*shift*) utilizado para construir las matrices GLCM. Este hallazgo confirma que la dirección radial es representativa del conjunto completo, lo que justifica su elección como base para los análisis posteriores y refuerza la conveniencia de emplear una única dirección como conjunto de datos principal para la clasificación.

De este modo, como primer paso en el análisis de variables, se examinó la matriz de correlación correspondiente a dicha dirección, lo que reveló la existencia de tres grupos de métricas altamente correlacionadas entre sí:

- **Grupo 1:** Contrast, Dissimilarity.
- **Grupo 2:** Homogeneity, ASM, Entropy.
- **Grupo 3:** Mean, Variance, Correlation.

Estos grupos pueden observarse en la **Gráfico 1**. Asimismo, se identificó que las variables SA y Mean presentan una correlación perfecta (coeficiente de correlación igual a 1), lo cual se explica por el hecho de que SA es una transformación lineal de Mean, concretamente el doble de su valor. Esta dependencia lineal exacta respalda la decisión de excluir la variable SA del conjunto de datos, ya que no aporta información adicional y su inclusión podría introducir problemas de colinealidad en los modelos.

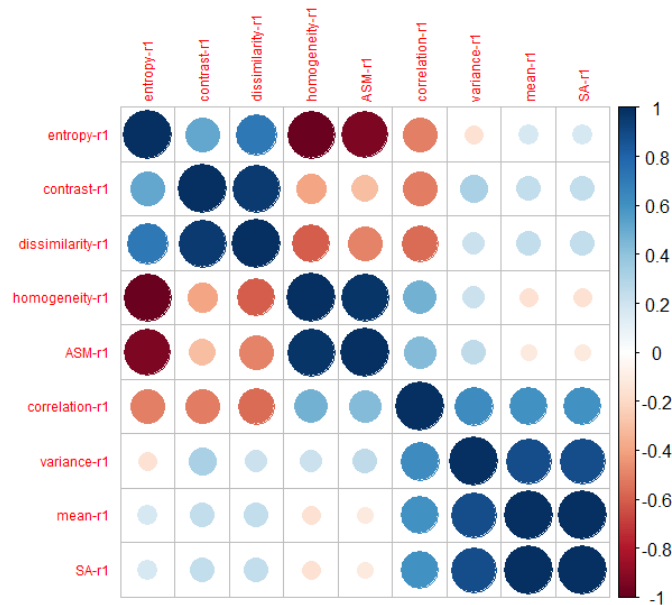


Gráfico 1: Matriz de correlación entre las métricas GLCM de la dirección radial.

Un aspecto adicional a considerar en la figura anterior, es que la variable *Correlation* presenta una alta correlación con todas las demás métricas. Esta situación sugiere que la variable podría ser redundante y, por lo tanto, su inclusión podría complicar los modelos sin aportar información nueva. En consecuencia, se plantea la posibilidad de eliminarla del conjunto de datos para evitar problemas de colinealidad y simplificar el modelo.

6.1.2. Análisis VIF

En el análisis de colinealidad mediante el Factor de Inflación de la Varianza, se eliminaron de forma iterativa aquellas variables con los valores de VIF más altos (ver **Figura 3** para los resultados iniciales), con el fin de minimizar la redundancia entre las variables predictoras.

Check for Multicollinearity

High Correlation

Term	VIF	VIF 95% CI	Increased SE	Tolerance	Tolerance 95% CI
glcm_contrast-r1	58.59	[55.82, 61.50]	7.65	0.02	[0.02, 0.02]
glcm_dissimilarity-r1	94.43	[89.94, 99.13]	9.72	0.01	[0.01, 0.01]
glcm_homogeneity-r1	211.10	[201.05, 221.65]	14.53	4.74e-03	[0.00, 0.00]
glcm_ASM-r1	63.48	[60.47, 66.63]	7.97	0.02	[0.02, 0.02]
glcm_entropy-r1	126.45	[120.44, 132.77]	11.25	7.91e-03	[0.01, 0.01]
glcm_mean-r1	49.86	[47.50, 52.33]	7.06	0.02	[0.02, 0.02]
glcm_variance-r1	72.65	[69.20, 76.26]	8.52	0.01	[0.01, 0.01]
glcm_correlation-r1	31.82	[30.32, 33.39]	5.64	0.03	[0.03, 0.03]

Figura 3: Resultados del análisis VIF inicial sobre las variables de la dirección radial.

```
# Check for Multicollinearity

Low Correlation
```

Term	VIF	VIF 95% CI	Increased SE	Tolerance	Tolerance 95% CI
glcm_contrast-r1	1.16	[1.13, 1.19]	1.08	0.86	[0.84, 0.89]
glcm_ASM-r1	1.11	[1.08, 1.14]	1.05	0.90	[0.88, 0.93]
glcm_mean-r1	1.07	[1.04, 1.10]	1.03	0.94	[0.91, 0.96]

Figura 4: Resultados del análisis VIF final sobre las variables de la dirección radial.

En consecuencia, los resultados presentados en la **Figura 4** permiten observar que el modelo óptimo se alcanza al reducir el conjunto de variables a tres, garantizando una baja colinealidad entre ellas. Además, conviene resaltar que cada una de las variables seleccionadas (*Contrast*, *ASM* y *Mean*) pertenece a uno de los grupos previamente identificados en la matriz de correlación.

Este hallazgo es especialmente relevante, ya que respalda la posibilidad de reducir la dimensionalidad del conjunto original mediante la selección de una variable representativa por grupo, lo que permite simplificar el modelo sin comprometer la calidad de la información.

6.2. Test de Kruskal-Wallis

Además de los análisis paramétricos realizados, se aplicó el test no paramétrico de Kruskal-Wallis a un nivel de significancia del 5 %, con el fin de evaluar si existían diferencias estadísticamente significativas en la distribución de las variables entre los distintos grupos de demencia. Este test resulta adecuado para comparar múltiples grupos sin asumir normalidad en los datos, lo cual es especialmente útil dada la naturaleza potencialmente no paramétrica de las métricas analizadas.

metric <chr>	statistic <dbl>	parameter <int>	p.value <dbl>
contrast	290.088819	3	1.389020e-62
dissimilarity	288.952646	3	2.446690e-62
homogeneity	10.551984	3	1.441237e-02
ASM	3.252663	3	3.542856e-01
entropy	50.092201	3	7.636001e-11
mean	1020.180884	3	7.537712e-221
variance	625.784164	3	2.591434e-135
correlation	1026.005151	3	4.109128e-222

Figura 5: Resultados del test de Kruskal-Wallis sobre las variables de la dirección radial.

De esta forma, y con base en los resultados presentados en la **Figura 5**, siete de ocho variables mostraron valores de p inferiores a 0.05, lo que indica diferencias significativas en sus distribuciones entre las distintas etapas de demencia. En particular, las variables *Mean*, *Correlation* y *Variance* se destacaron por presentar los valores estadísticos más elevados, sugiriendo un alto poder discriminativo.

Este resultado, es coherente con el análisis de colinealidad realizado mediante el VIF, el cual identificó que los datos podrían resumirse a tres variables, cada una perteneciente a uno de los grupos previamente definidos en la matriz de correlación. No obstante, cabe destacar que

la variable *ASM*, a pesar de haber sido seleccionada en dicho análisis, no mostró diferencias significativas entre los grupos en el test de Kruskal-Wallis ($p = 0.354$), lo que pone en duda su utilidad para la clasificación.

Así, al integrar los resultados de los análisis de colinealidad y del test de Kruskal-Wallis, fue posible refinar la selección de variables, priorizando aquellas que combinan un bajo grado de redundancia con una alta capacidad discriminativa. En este proceso, se optó por excluir *Correlation*, debido a su alta colinealidad con el resto de las métricas, y *ASM*, que no presentó diferencias estadísticamente significativas entre los grupos. Esta depuración condujo a una selección final compuesta por *Contrast*, *Entropy* y *Mean*, consideradas las métricas más informativas y adecuadas para la clasificación de las etapas de demencia.

6.3. Test Post-hoc de Dunn

Adicionalmente al análisis previamente realizado, se aplicó el test post-hoc de Dunn con corrección de Bonferroni, considerando un nivel de significancia del 5 %. Esta prueba no paramétrica se emplea para realizar comparaciones múltiples entre grupos una vez obtenido un resultado significativo en el análisis de Kruskal-Wallis. Su propósito es identificar específicamente cuáles pares de grupos presentan diferencias estadísticamente significativas, sin asumir normalidad en los datos y controlando el error tipo I mediante la corrección aplicada.

Es importante destacar que el test se aplicó a todas las variables del conjunto de datos, sin embargo, en esta sección solo se presentan los resultados correspondientes a las tres variables previamente seleccionadas, por considerarse las más relevantes para el análisis.

Comparison <chr>	Z <dbl>	P.unadj <dbl>	P.adj <dbl>
mild_demented - moderate_demented	-1.539876	1.235906e-01	7.415439e-01
mild_demented - non_demented	14.879435	4.482678e-50	2.689607e-49
moderate_demented - non_demented	6.033010	1.609338e-09	9.656025e-09
mild_demented - very_mild_demented	6.519504	7.054016e-11	4.232410e-10
moderate_demented - very_mild_demented	3.604442	3.128247e-04	1.876948e-03
non_demented - very_mild_demented	-11.059840	1.964473e-28	1.178684e-27

Figura 6: Resultados del test Post-hoc de Dunn sobre la variable *Contrast*.

Comparison <chr>	Z <dbl>	P.unadj <dbl>	P.adj <dbl>
mild_demented - moderate_demented	-0.03369721	9.731186e-01	1.000000e+00
mild_demented - non_demented	6.35332734	2.107067e-10	1.264240e-09
moderate_demented - non_demented	1.93667467	5.278512e-02	3.167107e-01
mild_demented - very_mild_demented	2.86715250	4.141834e-03	2.485100e-02
moderate_demented - very_mild_demented	0.92838394	3.532084e-01	1.000000e+00
non_demented - very_mild_demented	-4.60273044	4.169878e-06	2.501927e-05

Figura 7: Resultados del test Post-hoc de Dunn sobre la variable *Entropy*.

Comparison <chr>	Z <dbl>	P.unadj <dbl>	P.adj <dbl>
mild_demented - moderate_demented	-2.326907	1.997020e-02	1.198212e-01
mild_demented - non_demented	-28.625914	3.197998e-180	1.918799e-179
moderate_demented - non_demented	-6.185536	6.189192e-10	3.713515e-09
mild_demented - very_mild_demented	-12.226143	2.253903e-34	1.352342e-33
moderate_demented - very_mild_demented	-1.437280	1.506385e-01	9.038309e-01
non_demented - very_mild_demented	21.731626	1.030875e-104	6.185251e-104

Figura 8: Resultados del test Post-hoc de Dunn sobre la variable *Mean*.

Como reflejan los resultados, en el caso de la variable *Contrast* (**Figura 6**), se observaron diferencias estadísticamente significativas en cinco de los seis pares de grupos clínicos analizados. Estos resultados sugieren que esta métrica posee un buen poder discriminativo para diferenciar sujetos en distintos grados de deterioro cognitivo, con excepción únicamente del par *mild_demented* y *moderate_demented*.

En relación con los resultados de la variable *Entropy* (**Figura 7**), se observaron diferencias significativas principalmente al comparar el grupo *non_demented* con los grupos que presentan algún grado de demencia, lo que indica que esta métrica permite una buena discriminación entre sujetos sanos y aquellos con deterioro cognitivo en distintos niveles. No obstante, cabe señalar que la variable no distingue adecuadamente entre los grupos *moderate_demented* y *non_demented*, aunque sí logra diferenciar bien entre *mild_demented* y *very_mild_demented*, lo que refuerza su utilidad para detectar etapas tempranas del deterioro.

Por último, en lo que respecta a la variable *Mean* (**Figura 8**), se evidenció un alto poder discriminativo, al mostrar diferencias altamente significativas en la mayoría de los pares de grupos comparados. Este resultado refuerza su valor como métrica clave para la clasificación entre los distintos estadios de la enfermedad de Alzheimer.

Estos resultados respaldan la elección de las tres variables como representaciones relevantes de los patrones texturales presentes en las imágenes.

6.4. Análisis de componentes principales

En base a las conclusiones previas, a partir de este momento se decidió emplear exclusivamente las métricas derivadas de la dirección radial, descartando la variable *SA* debido a que no aportaba información adicional relevante y presentaba un riesgo considerable de multicolinealidad.

Además, dado que la variable *Correlation* presentaba riesgo de multicolinealidad y el test de Kruskal-Wallis cuestionó la utilidad de la variable *ASM* por no mostrar diferencias significativas entre los grupos, se evaluó individualmente la inclusión de ambas variables en el PCA, comparando los resultados obtenidos con y sin su incorporación.

En este sentido, se comprobó que la exclusión de la variable *Correlation* no afecta los resultados, en contraste con la exclusión de la variable *ASM*, que sí genera un impacto significativo. Por esta razón, se decidió eliminar *Correlation* y conservar *ASM* en el conjunto de datos para el

Análisis de Componentes Principales (PCA).

Así, a partir de esta selección, se llevó a cabo un análisis de la varianza explicada por las componentes, cuyo resultado se muestra en el **Gráfico 2**, donde se observa que las tres primeras componentes explican aproximadamente el 98 % de la varianza total. Este resultado es consistente con los análisis previos de correlación, lo que reafirma que los datos pueden resumirse en tres variables y, más específicamente, confirma la conveniencia de emplear un PCA basado en la selección de tres componentes principales.

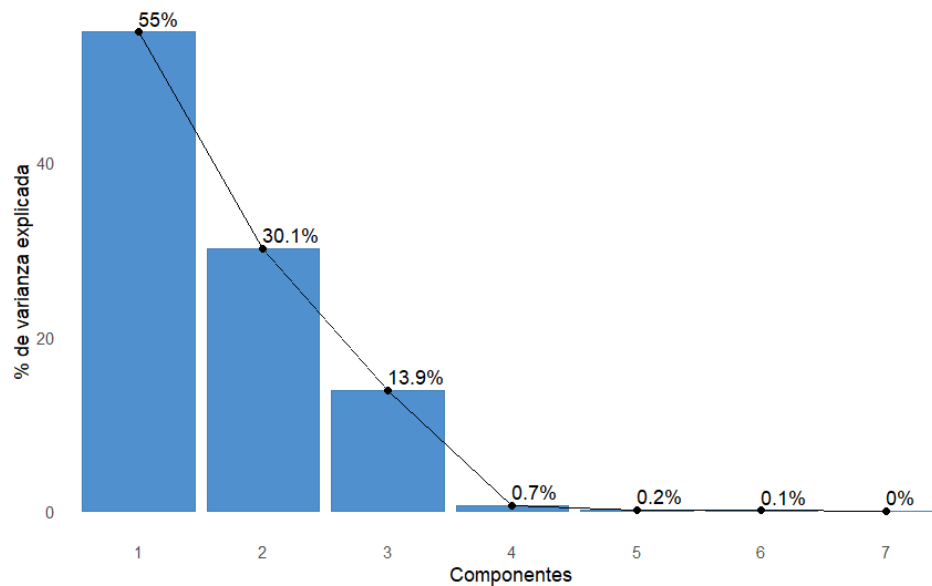


Gráfico 2: Varianza explicada por las componentes principales del PCA de la dirección radial.

Por consiguiente, se analizaron de forma visual las métricas con mayor importancia para cada componente principal, es decir, aquellas que muestran mayor correlación con cada una de ellas. Este análisis tiene como finalidad comprobar si las variables que más influyen en cada componente coinciden con los grupos previamente identificados mediante el estudio de correlaciones y el VIF.

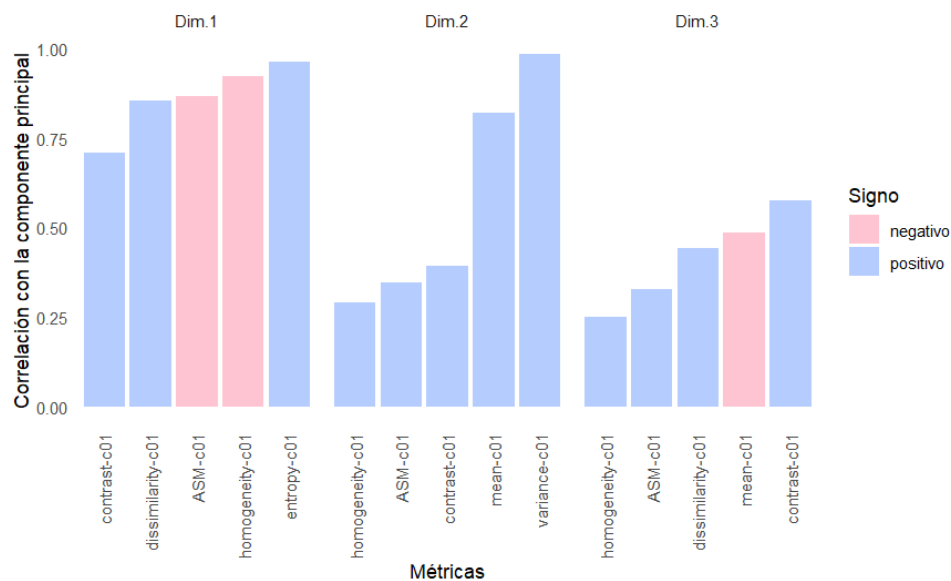


Gráfico 3: Las 5 métricas que más contribuyen a los componentes principales de la dirección radial.

De este modo, tal como se muestra en el **Gráfico 3**, las variables con mayor correlación con las componentes principales son *Entropy*, *Variance* y *Contrast*. Este resultado valida que la estructura revelada por el PCA está en consonancia con los patrones observados previamente, lo que refuerza la estrategia de reducción de variables y la selección de métricas representativas.

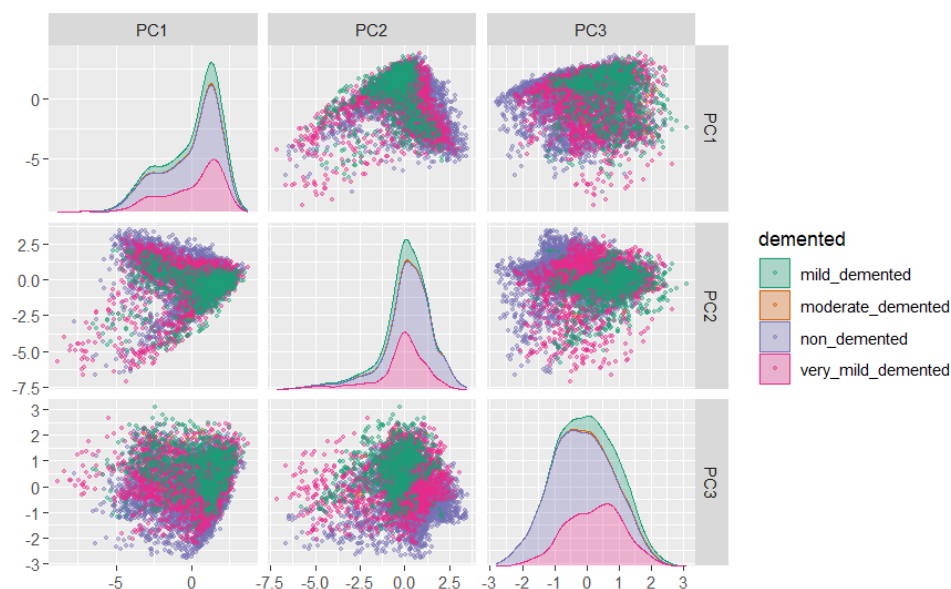


Gráfico 4: Distribución de las componentes principales del PCA del conjunto de datos de la dirección radial.

Por último, se analiza la relación entre las componentes principales obtenidas mediante PCA y la variable objetivo, con el propósito de evaluar su capacidad para discriminar entre clases. En este sentido, se observa en el **Gráfico 4** que las tres componentes principales (PC1, PC2 y PC3)

no permiten una separación visual clara, dado que las clases se superponen considerablemente en los gráficos de dispersión.

De esta forma, ante la falta de una diferenciación visual clara entre las clases mediante las componentes principales obtenidas por PCA, resulta razonable considerar su aplicación en modelos de clasificación más complejos, como árboles de decisión o random forest, ya que podrían capturar relaciones no lineales entre las componentes principales y la variable objetivo.

7. Datos finales

A partir de los análisis realizados en el apartado de análisis de variables, se identifican tres hallazgos principales: que la dirección radial es la más adecuada para continuar con el proceso de clasificación; que los datos pueden resumirse en tres variables o componentes principales; y que la métrica *Correlation*, al igual que la variable *SA*, no resulta relevante en este contexto, mientras que *ASM* es considerada relevante únicamente en el conjunto de datos correspondiente al análisis mediante PCA.

En consecuencia, se identifican dos estrategias posibles para la selección de variables:

1. **Uso de componentes principales:** Realizar la clasificación utilizando las tres componentes principales obtenidas mediante PCA, que explican el 98 % de la varianza presente en los datos.
2. **Selección de variables representativas:** Realizar la clasificación por medio de la selección de una métrica representativa de cada grupo identificado.

Grupo	PCA	VIF
1	Contrast	Contrast
2	Entropy	Entropy
3	Variance	Mean

Cuadro 1: Resumen de las variables más representativas por grupo según los análisis del PCA y VIF.

De esta forma, se procedió al desarrollo de los modelos mediante los diferentes algoritmos de clasificación explorados en este trabajo, aplicando en cada uno de ellos las dos estrategias de selección de variables disponibles. Como parte de este proceso, los datos finales utilizados en los modelos de clasificación se presentan en el **Cuadro 1**.

8. Resultados

8.1. Árbol de decisión

8.1.1. Variables seleccionadas

Para llevar a cabo el primer método abordado para la clasificación, se utiliza el algoritmo de árboles de decisión aplicado sobre las variables *Contrast*, *Entropy* y *Mean* correspondientes a la dirección radial. A continuación, se presenta el conjunto de hiperparámetros empleados en la construcción del grid para la optimización del modelo.

Complejidad	Profundidad	Tamaño mínimo
0.001, 0.01, 0.1, 1	1, 2, 3, 4	10, 25, 50, 100

Cuadro 2: Grid de hiperparámetros del árbol de decisión.

Tras la selección de los mejores hiperparámetros para el modelo mediante la validación cruzada, se definió el modelo óptimo como un árbol con una complejidad mínima de poda de 0.001, una profundidad máxima de 4 niveles y un mínimo de 10 observaciones por nodo terminal, el cual puede observarse en el **Gráfico 5**.

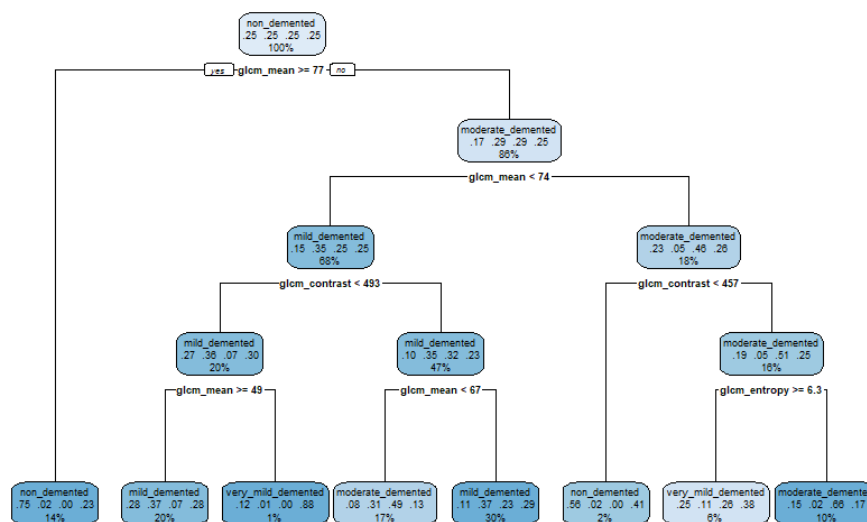


Gráfico 5: Árbol de decisión sobre las variables *Contrast*, *Entropy* y *Mean* de la dirección radial.

A continuación, se presentan los resultados de desempeño del árbol de decisión obtenidos en los conjuntos de entrenamiento, validación y prueba:

Muestra	Accuracy	Kappa	Sensibilidad	Especificidad
Entrenamiento	0.4810	0.3080	0.4810	0.8270
Validación (media)	0.4801	0.3065	0.4798	0.8266
Error estándar	0.0062	0.0074	0.0058	0.0018
Prueba	0.4821	0.3094	0.4821	0.8274

Cuadro 3: Resultados de desempeño del árbol de decisión sobre las variables *Contrast*, *Entropy* y *Mean* de la dirección radial.

A partir de estos resultados, se observa que el modelo de árbol de decisión, construido sobre las variables *Contrast*, *Entropy* y *Mean* de la dirección radial, presenta en general un bajo desempeño, aunque consistente entre las diferentes particiones del conjunto de datos.

En particular, en el conjunto de entrenamiento, la exactitud alcanzó un 46.7 %, con un coeficiente Kappa de 0.2899, lo que sugiere una capacidad limitada del modelo para clasificar correctamente. Este desempeño se refleja también en los datos de validación, donde la exactitud media fue de 46.1 %, acompañada de una desviación estándar relativamente baja (0.71 %), la cual indica cierta estabilidad en los resultados. Asimismo, en el conjunto de prueba, el modelo obtuvo una exactitud del 46.7 %, lo que confirma la consistencia de su bajo rendimiento fuera de la muestra.

En cuanto a las métricas de sensibilidad y especificidad, se destaca que la sensibilidad se sitúa en torno al 46.7 %, indicando que el modelo tiene dificultades para identificar correctamente todas las clases. En contraste, la especificidad es considerablemente más alta, aproximadamente 82.2 %, lo que sugiere que el modelo es más efectivo para reconocer correctamente las instancias que no pertenecen a una determinada clase.

En resumen, estos resultados evidencian que, aunque el modelo muestra una cierta estabilidad, su capacidad predictiva es limitada. Por tanto, sería recomendable considerar otras técnicas de modelado que permitan mejorar su desempeño en la identificación de los distintos estadios de la demencia.

8.1.2. PCA dirección radial

Para dar continuidad al análisis del algoritmo de clasificación mediante árboles de decisión, se desarrolló un modelo utilizando las componentes principales correspondientes a la dirección radial, previamente presentadas en apartados anteriores. En esta etapa, se empleó el mismo conjunto de hiperparámetros definido anteriormente para la selección del modelo óptimo.

Tras la selección de los mejores hiperparámetros para el modelo mediante la validación cruzada, se definió el modelo óptimo como un árbol con una complejidad mínima de poda de 0.001, una profundidad máxima de 4 niveles y un mínimo de 10 observaciones por nodo terminal, el cual puede observarse en el **Gráfico 6**.



Muestra	Accuracy	Kappa	Sensibilidad	Especificidad
Entrenamiento	0.4824	0.3099	0.4824	0.8275
Validación (media)	0.4811	0.3079	0.4808	0.8270
Error estándar	0.0079	0.0093	0.0066	0.0023
Prueba	0.4785	0.3047	0.4785	0.8262

Cuadro 4: Resultados de desempeño del árbol de decisión sobre el PCA de la dirección radial.

A partir de estos resultados, se observa que el modelo de árbol de decisión, construido sobre las variables de componentes principales (PCA) de la dirección radial, presenta en general un bajo desempeño, aunque consistente entre las diferentes particiones del conjunto de datos.

En particular, con respecto a la exactitud calculada mediante el cálculo de la precisión (*accuracy*), en todas las muestras el nivel de acierto es bajo para un problema de clasificación multiclase. Esto se evidencia al evaluar el coeficiente Kappa, que fue cercano a 0.31 para dichas muestras, indicando que el modelo tiene una capacidad limitada para clasificar correctamente, aunque ligeramente superior a una asignación aleatoria. Cabe destacar que, en la muestra de validación, aunque la exactitud media fue igualmente baja (aproximadamente 48.1 %), el error estándar asociado fue reducido (0.79 %), lo que evidencia un rendimiento relativamente consistente entre los distintos subconjuntos evaluados.

En cuanto a las métricas de sensibilidad y especificidad, se destaca que la sensibilidad se sitúa en torno al 48 %, lo que indica una capacidad limitada del modelo para identificar correctamente las instancias pertenecientes a cada clase. En contraste, la especificidad es considerablemente

mayor, aproximadamente 83 %, lo que sugiere que el modelo es más efectivo para reconocer correctamente las instancias que no pertenecen a una clase determinada.

En resumen, estos resultados evidencian que, aunque el modelo muestra una cierta estabilidad, su capacidad predictiva es limitada. Esto resalta la necesidad de considerar técnicas de modelado alternativas, con el objetivo de mejorar la correcta identificación de todas las clases.

8.2. Random Forest

8.2.1. Variables seleccionadas

Para llevar a cabo el segundo método abordado para la clasificación, se emplea el algoritmo de Random Forest sobre las variables *Contrast*, *Entropy* y *Mean*, correspondientes a la dirección radial. A continuación, se presenta el conjunto de hiperparámetros utilizados en la construcción del grid para la optimización del modelo.

Número de variables al azar	Número de árboles	Mínimo de observaciones por nodo
1, 3, 6	5, 10, 25, 50	10, 25, 50, 100

Cuadro 5: Grid de hiperparámetros del Random Forest.

Tras la selección de los mejores hiperparámetros mediante validación cruzada, se definió el modelo óptimo como aquel configurado con un número de variables aleatorias igual a 1, un total de 50 árboles y un mínimo de 10 observaciones por nodo terminal.

A continuación, se presentan los resultados de desempeño del modelo de Random Forest obtenidos en los conjuntos de entrenamiento, validación y prueba:

Muestra	Accuracy	Kappa	Sensibilidad	Especificidad
Entrenamiento	0.9926	0.9901	0.9926	0.9975
Validación (Media)	0.9202	0.8935	0.9202	0.9734
Error estándar	0.0047	0.0062	0.0045	0.0016
Prueba	0.9430	0.9240	0.9430	0.9810

Cuadro 6: Resultados de desempeño del Random Forest sobre las variables *Contrast*, *Entropy* y *Mean* de la dirección radial.

A partir de estos resultados, puede observarse que el modelo de Random Forest, construido a partir de las variables *Contrast*, *Entropy* y *Mean* correspondientes a la dirección radial, exhibe un desempeño altamente satisfactorio en todas las etapas de evaluación.

En particular, en el conjunto de entrenamiento, el modelo alcanzó una exactitud del 99.3 %, una sensibilidad del 99.3 % y una especificidad del 99.8 %, acompañadas de un coeficiente Kappa de 0.99. Estos valores indican una clasificación prácticamente perfecta dentro de la muestra, evidenciando una excelente capacidad del modelo para aprender los patrones subyacentes en

los datos. Este desempeño se refleja también en los datos de validación, donde se observó una ligera disminución en el rendimiento, con una exactitud promedio de aproximadamente 91.9 %, acompañada de una desviación estándar relativamente baja (0.56 %), la cual indica una cierta estabilidad en los resultados.

Por su parte, en el conjunto de prueba, el modelo mostró un comportamiento coherente con los conjuntos de entrenamiento y validación, lo que confirma que no solo es eficaz dentro de la muestra utilizada para el entrenamiento, sino que también mantiene su capacidad predictiva en datos no vistos, discriminando adecuadamente entre las distintas clases.

En resumen, estos resultados evidencian que el modelo de Random Forest, construido sobre las variables seleccionadas, constituye una herramienta robusta, estable y confiable para la clasificación en este contexto.

8.2.2. PCA dirección radial

Para dar continuidad al análisis del algoritmo Random Forest, se desarrolló un modelo utilizando las componentes principales correspondientes a la dirección radial, previamente presentadas en apartados anteriores. En esta etapa, se empleó el mismo conjunto de hiperparámetros definido anteriormente para la selección del modelo óptimo.

Tras la selección de los mejores hiperparámetros mediante validación cruzada, se definió el modelo óptimo como aquel configurado con un número de variables aleatorias igual a 1, un total de 50 árboles y un mínimo de 10 observaciones por nodo terminal.

A continuación, se presentan los resultados de desempeño del modelo de Random Forest obtenidos en los conjuntos de entrenamiento, validación y prueba:

Muestra	Accuracy	Kappa	Sensibilidad	Especificidad
Entrenamiento	0.9939	0.9918	0.9939	0.9980
Validación (Media)	0.9303	0.9071	0.9304	0.9768
Error estándar	0.0050	0.0066	0.0046	0.0017
Prueba	0.9422	0.9230	0.9422	0.9807

Cuadro 7: Resultados de desempeño del Random Forest sobre el PCA de la dirección radial.

A partir de estos resultados, se observa que el modelo de Random Forest, construido a partir de las componentes principales correspondientes a la dirección radial, presenta un desempeño altamente satisfactorio en todas las métricas de evaluación.

Además, al evaluar el modelo en los datos de validación y prueba, se obtuvieron resultados similares con una desviación estándar consideravelmente baja, lo que confirma que el modelo mantiene un alto nivel de desempeño fuera de la muestra.

En resumen, estos resultados evidencian que el modelo de Random Forest construido sobre el PCA demuestra una sólida capacidad para generalizar y discriminar adecuadamente entre las

distintas clases, consolidándose como una herramienta robusta y confiable para la clasificación en este estudio, al minimizar eficazmente el riesgo de sobreajuste.

8.3. Regresión Logística

8.3.1. Variables seleccionadas

Para llevar a cabo el último enfoque de clasificación abordado en este trabajo, se utilizó un modelo de regresión logística multinomial sobre las variables *Contrast*, *Entropy* y *Mean*, correspondientes a la dirección radial. Sin embargo, a diferencia de métodos como los árboles de decisión o el modelo Random Forest, la regresión logística no requiere un proceso de ajuste de hiperparámetros, lo que simplifica su implementación.

En la figura siguiente se presenta el modelo de regresión logística multinomial obtenido:

```
# weights: 20 (12 variable)
initial value 28102.959289
iter 10 value 25997.616236
final value 25163.523695
converged
Call:
multinom(formula = class ~ ., data = train_data)

Coefficients:
              (Intercept) glcm_contrast glcm_entropy glcm_mean
mild_demented      10.247298    0.014189368   -0.2776091  -0.2209207
moderate_demented  11.651699    0.017919016   -1.5807451  -0.1533513
very_mild_demented  7.571914    0.008017613   -0.3715583  -0.1277769

Std. Errors:
              (Intercept) glcm_contrast glcm_entropy glcm_mean
mild_demented    0.007751673  0.0003734441  0.04734935  0.003935559
moderate_demented 0.008519884  0.0004068778  0.05389633  0.003838467
very_mild_demented 0.006782260  0.0003532559  0.04475512  0.003430884

Residual Deviance: 50327.05
AIC: 50351.05
```

Figura 9: Resultado del modelo de regresión logística multinomial sobre las variables *Contrast*, *Entropy* y *Mean* de la dirección radial.

Este modelo posibilita estimar la probabilidad de que una imagen se clasifique en alguno de los diferentes estadios de la enfermedad de Alzheimer (*mild_demented*, *moderate_demented* o *very_mild_demented*), en comparación con la categoría de referencia *non_demented*.

De esta forma, según los coeficientes obtenidos, de forma general, se observa que la variable *glcm_contrast* presenta valores positivos en las tres clases, lo que indica que, cuanto mayor sea el contraste de la imagen, mayor será la probabilidad de que esta sea clasificada en alguno de los estadios de demencia en comparación con la categoría *non_demented*.

Por otro lado, las variables *glcm_entropy* y *glcm_mean* presentan coeficientes negativos, lo cual sugiere que valores más elevados de entropía o de media se asocian con una menor probabilidad de que la imagen pertenezca a esas clases, en comparación con la categoría de ausencia de demencia.

En conjunto, estos resultados señalan al contraste como un factor asociado a un mayor riesgo

de demencia, mientras que la entropía y el valor medio podrían actuar como indicadores de un menor riesgo.

A continuación, se presentan los resultados de desempeño del modelo de regresión logística obtenidos en los conjuntos de entrenamiento, validación y prueba:

Muestra	Accuracy	Kappa	Sensibilidad	Especificidad
Entrenamiento	0.3678	0.1571	0.3678	0.7893
Validación (Media)	0.3642	0.1525	0.3646	0.7881
Error estándar	0.0020	0.0027	0.0019	0.0007
Prueba	0.3596	0.1462	0.3596	0.7865

Cuadro 8: Resultados de desempeño del modelo de regresión logística multinomial sobre las variables *Contrast*, *Entropy* y *Mean* de la dirección radial.

A partir de estos resultados, se observa que el modelo de regresión logística, construido a partir de las variables *Contrast*, *Entropy* y *Mean* correspondientes a la dirección radial, presenta en general un bajo desempeño, aunque consistente entre las diferentes particiones del conjunto de datos.

En particular, en el conjunto de entrenamiento, la exactitud alcanzó aproximadamente un 37 %, con un coeficiente Kappa de 0.16, lo que indica una capacidad limitada del modelo para identificar correctamente las distintas clases dentro de esta muestra. Este desempeño se mantiene en el conjunto de validación, donde la exactitud media fue del 36.9 %, acompañada de una baja desviación estándar (0.26 %), lo que sugiere estabilidad en los resultados. De manera similar, en el conjunto de prueba, el modelo obtuvo una exactitud del 36.2 %, confirmando la consistencia de su bajo rendimiento fuera de la muestra de entrenamiento.

En cuanto a las métricas de sensibilidad y especificidad, se destaca que la sensibilidad se sitúa en torno al 37 %, indicando que el modelo tiene dificultades para identificar correctamente todas las clases. En contraste, la especificidad es considerablemente más alta, aproximadamente 79 %, lo que sugiere que el modelo es más efectivo para reconocer correctamente las instancias que no pertenecen a una determinada clase.

En resumen, estos resultados evidencian que, si bien el modelo presenta cierta estabilidad, su capacidad de discriminación entre las diferentes clases es limitada. Por tanto, resulta aconsejable explorar otras técnicas de modelado que puedan mejorar su desempeño en la identificación de los distintos estadios de la demencia.

8.3.2. PCA dirección radial

Con el objetivo de profundizar en el análisis del modelo de regresión logística, se desarrolló una versión basada en las componentes principales correspondientes a la dirección radial, descritas en secciones anteriores.

En la figura siguiente se presenta el modelo de regresión logística multinomial obtenido:

```
# weights: 20 (12 variable)
initial value 28102.959289
iter 10 value 26318.463639
final value 25027.593063
converged
Call:
multinom(formula = class ~ ., data = train_data)

Coefficients:
              (Intercept)          Dim.1          Dim.2          Dim.3
mild_demented    -0.30724685  0.10465547 -0.5198360  1.4736356
moderate_demented -0.28735509  0.18641336 -0.1385483  1.5365198
very_mild_demented 0.07131468  0.04348453 -0.2957077  0.8624377

Std. Errors:
              (Intercept)          Dim.1          Dim.2          Dim.3
mild_demented    0.02539280  0.01198927  0.01664666  0.02800590
moderate_demented 0.02520721  0.01239709  0.01935785  0.02833585
very_mild_demented 0.02166400  0.01121919  0.01519876  0.02516294

Residual Deviance: 50055.19
AIC: 50079.19
```

Figura 10: Resultado del modelo de regresión logística multinomial sobre el PCA de la dirección radial.

Este modelo posibilita estimar la probabilidad de que una imagen se clasifique en alguno de los diferentes estadios de la enfermedad de Alzheimer (*mild_demented*, *moderate_demented* o *very_mild_demented*), en comparación con la categoría de referencia *non_demented*.

De esta forma, según los coeficientes obtenidos, en términos generales se observa que la primera componente principal (*Dim1*) y la tercera componente principal (*Dim3*) presentan valores positivos en las tres clases, lo que indica que, cuanto mayores sean los valores en estas dimensiones, mayor será la probabilidad de que la imagen sea clasificada en alguno de los estadios de demencia, en comparación con la categoría *non_demented*.

Por otro lado, la segunda componente principal (*Dim2*) presenta coeficientes negativos, lo que sugiere que valores más altos en esta componente se asocian con una menor probabilidad de que una imagen pertenezca a esas clases, en comparación con la categoría de ausencia de demencia.

En conjunto, estos resultados indican que valores elevados en (*Dim1*) y (*Dim3*) parecen asociarse con un mayor riesgo de demencia, mientras que valores altos en (*Dim2*) podrían estar relacionados con un menor riesgo.

A continuación, se presentan los resultados de desempeño del modelo de regresión logística obtenidos en los conjuntos de entrenamiento, validación y prueba:

Muestra	Accuracy	Kappa	Sensibilidad	Especificidad
Entrenamiento	0.3881	0.1841	0.3881	0.7960
Validación (Media)	0.3852	0.1805	0.3856	0.7952
Error estándar	0.0017	0.0024	0.0019	0.0006
Prueba	0.3898	0.1864	0.3898	0.7966

Cuadro 9: Resultados de desempeño del modelo de regresión logística multinomial sobre el PCA de la dirección radial.

A partir de estos resultados, se observa que el modelo de regresión logística, construido a partir del PCA correspondiente a la dirección radial, presenta en general un bajo desempeño, aunque consistente entre las diferentes particiones del conjunto de datos.

En particular, en el conjunto de entrenamiento, la exactitud alcanzó aproximadamente un 38,8 %, con un coeficiente Kappa de 0,16, lo que refleja una capacidad limitada del modelo para identificar correctamente las distintas clases en esta muestra. Este desempeño se mantiene en el conjunto de validación, donde la exactitud media fue del 38,5 %, acompañada de una baja desviación estándar (0,17 %), lo que indica estabilidad en los resultados. De manera similar, en el conjunto de prueba se observan resultados coherentes con los de entrenamiento y validación, confirmando que, aunque el modelo es estable, su capacidad predictiva general fuera de la muestra es limitada.

En cuanto a las métricas de sensibilidad y especificidad, se destaca que la sensibilidad se sitúa en torno al 38 %, indicando que el modelo tiene dificultades para identificar correctamente todas las clases. En contraste, la especificidad es considerablemente más alta, aproximadamente 79 %, lo que sugiere que el modelo es más efectivo para reconocer correctamente las instancias que no pertenecen a una determinada clase.

En resumen, estos resultados evidencian que, si bien el modelo presenta cierta estabilidad, su capacidad de discriminación entre las diferentes clases es limitada. Por tanto, resulta aconsejable explorar otras técnicas de modelado que puedan mejorar su desempeño en la identificación de los distintos estadios de la demencia.

9. CONCLUSIONES

Como conclusión, los resultados obtenidos muestran que, en general, para todos los modelos evaluados, existe una consistencia entre las muestras de entrenamiento, validación y prueba, lo que refleja la estabilidad de los resultados, independientemente del nivel de desempeño alcanzado por cada modelo.

Sin embargo, los modelos basados en árboles de decisión y regresión logística presentan una capacidad limitada para clasificar correctamente las distintas clases. Además, muestran una sensibilidad relativamente baja, en contraste con una especificidad elevada, lo que sugiere que estos modelos son más eficaces para identificar correctamente las instancias que no pertenecen a una determinada clase.

Por otro lado, el modelo de Random Forest demostró un rendimiento notablemente superior, con niveles de exactitud, sensibilidad y especificidad consistentemente altos en todas las particiones evaluadas. Estos resultados evidencian su capacidad para aprender patrones complejos y generalizar de manera eficaz a nuevos datos, lo que lo consolida como la técnica más robusta y confiable en el contexto analizado. Adicionalmente, su arquitectura contribuyó a mantener la estabilidad del modelo y a mitigar el riesgo de sobreajuste.

En paralelo, cabe señalar, que la aplicación del *Análisis de Componentes Principales* (PCA) no condujo a mejoras significativas en el rendimiento de ninguno de los modelos evaluados.

En este contexto, se considera que este trabajo ha realizado aportes relevantes al campo del análisis de imágenes para el diagnóstico de enfermedades neurodegenerativas. Específicamente, la extracción de características texturales radiales a partir de matrices GLCM, la agrupación de dichas métricas en tres grupos, el uso del SMOTE como técnica de balanceo y los enfoques de clasificación propuestos constituyen avances significativos que no solo mejoran la interpretabilidad y el análisis de los datos, sino que también optimizan los tiempos de ejecución.

Técnica	Accuracy (%)
Random Forest variables seleccionadas	99,26
Random Forest PCA	99,39
LSTM	99,014
CNN	95,0
SVM	84,87

Cuadro 10: Comparativa de los resultados de rendimiento frente a la literatura existente.

Además, de manera complementaria, como es posible observar en la **Cuadro 10**, los resultados obtenidos con el modelo de Random Forest se aproximan a los reportados en estudios previos [32], [33] y [34], los cuales han recurrido a algoritmos más complejos (como las *Redes Neuronales Convolucionales* (CNN, de Convolutional Neural Network), *Máquinas de Vectores de Soporte* (SVM, de Support Vector Machine) y *Memorias a Largo Plazo* (LSTM, de Long Short-Term Memory)). Estes algoritmos, si bien son usualmente preferidos por su alta precisión

y su capacidad para manejar datos complejos y de alta dimensionalidad, también conllevan tiempos de ejecución elevados y una interpretabilidad limitada [35]. En este sentido, los hallazgos obtenidos y las metodologías propuestas en este estudio abren nuevas vías para investigaciones futuras y aplicaciones prácticas que pueden contribuir a facilitar el diagnóstico de la enfermedad de Alzheimer.

Es importante resaltar que los resultados estuvieron condicionados por las imágenes disponibles, las cuales no presentaban una alineación consistente respecto a los cortes (*slices*). En este sentido, contar con una base de datos de resonancia magnética (MRI) que incluya una selección más precisa y homogénea de los cortes podría contribuir a mejorar aún más la precisión del diagnóstico automático.

Por último, como proyección para trabajos futuros, se recomienda explorar la aplicación de técnicas de sobremuestreo, como el SMOTE, directamente sobre las imágenes como estrategia de balanceo, en lugar de aplicarlas sobre las características texturales derivadas. Asimismo, sería pertinente explorar métodos de visualización e interpretación orientados a desentrañar el comportamiento del modelo de Random Forest, permitiendo identificar con mayor claridad el impacto de cada variable en las decisiones del modelo.

Bibliografía

- [1] Caballero Testón E, Álvarez Liébana J, Calviño Martínez A. Trabajo Fin de Máster en Bioestadística: Clasificación Supervisada de Imágenes de Resonancia Magnética para la Detección de Enfermedad de Alzheimer; 2024. Trabajo Fin de Máster, Universidad Complutense de Madrid.
- [2] MayoClinic. Enfermedad de Alzheimer; 2024. Accedido: 2024-05-23. <https://www.mayoclinic.org/es/diseases-conditions/alzheimers-disease/symptoms-causes/syc-20350447>.
- [3] Sadurní J. Alois Alzheimer, el neuropsiquiatra que descubrió la demencia que lleva su nombre; 2023. Accedido: 2024-05-14. https://historia.nationalgeographic.com.es/a/alois-alzheimer-neuropsiquiatra-que-descubrio-demencia-que-lleva-su-nombre_16883.
- [4] Karran E, Mercken M, De Strooper B. The amyloid cascade hypothesis for Alzheimer's disease: an appraisal for the development of therapeutics. *Nature Reviews Drug Discovery*. 2011;10:698-712.
- [5] Rius-Pérez S, Tormos A, Pérez S, Taléns-Visconti R. Patología vascular: ¿causa o efecto en la enfermedad de Alzheimer? *Neurología*. 2018;33:112-20.
- [6] Picazo MCB, Alcañiz AL, Valero , Mora N, Salvador MDA, Ribes JV. Estudio de toxicidad del péptido beta-amiloide en linfocitos humanos. *Revista Española de Geriatria y Gerontología*. 2006;41:178-82.
- [7] Breijyeh Z, Karaman R. Comprehensive Review on Alzheimer's Disease: Causes and Treatment. *Molecules*. 2020;25:5789.
- [8] Braak H, Braak E. Neuropathological staging of Alzheimer-related changes. *Acta Neuropathologica*. 1991;82:239-59.
- [9] Thal DR, Rüb U, Orantes M, Braak H. Phases of A-deposition in the human brain and its relevance for the development of AD. *Neurology*. 2002;58:1791-800.
- [10] Mayfield. Mayfield brain & spine;. Accedido: 2024-05-23. <https://mayfieldclinic.com/pe-anatbrain.htm>.
- [11] Sobreviela T, Insausti R, Salinas A, Gonzalo LM, Insausti R. La corteza entorrinal (formación del hipocampo) en el envejecimiento y en la enfermedad de Alzheimer. Interpretación neuroanatómica. *Revista de Medicina de la Universidad de Navarra*. 2016:19-27.

- [12] Folstein MF, Folstein SE, McHugh PR. “Mini-mental state”. A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*. 1975;12:189-98.
- [13] Haan Jd, Csinscik L, Parker T, Paterson RW, Slattery CF, Foulkes A, et al. Retinal thickness as potential biomarker in posterior cortical atrophy and typical Alzheimer’s disease. *Alzheimer’s Research Therapy*. 2019;11.
- [14] Abdulla S. MR imaging; 2017. Accedido: 2024-05-23. <https://www.radiologycafe.com/frcr-physics-notes/mr-imaging/>.
- [15] Li X, Morgan PS, Ashburner J, Smith J, Rorden C. The first step for neuroimaging data analysis: DICOM to NIfTI conversion. *Journal of Neuroscience Methods*. 2016;264:47-56.
- [16] Kumar S, Shastri S. Alzheimer MRI Preprocessed Dataset; 2022. <https://www.kaggle.com/dsv/3364939>.
- [17] Saykin AJ, ADNI. ADNI;. Accedido: 2024-05-24. <https://adni.loni.usc.edu/about/>.
- [18] Ilich AR. GLCMTextures; 2020. <https://github.com/ailich/GLCMTextures>.
- [19] Ajmera G. Feature extraction of images using GLCM (gray level cooccurrence matrix); 2024. Accedido: 2024-05-25. <https://medium.com/@girishajmera/feature-extraction-of-images-using-glcmm-gray-level-cooccurrence-matrix-e4bda>.
- [20] Cruz P. Bioestatística - Correlação de Pearson;. <https://sweet.ua.pt/pedrocruz/bioestatistica/cor-pearson.html#gsc.tab=0>.
- [21] DataCamp. Variance Inflation Factor (VIF): O que é e Como Calcular;. <https://www.datacamp.com/pt/tutorial/variance-inflation-factor>.
- [22] Aragón J. Test Kruskal-Wallis (e Diagnóstico de Multicolinearidade via VIF);. https://rpubs.com/Joaquin_AR/219504.
- [23] Pohlert T. posthoc.kruskal.dunn.test: Pairwise Test for Multiple Comparisons of Mean Rank Sums (Kruskal-Dunn);. <https://www.rdocumentation.org/packages/PMCMR/versions/4.3/topics/posthoc.kruskal.dunn.test>.
- [24] Dunn OJ. Multiple Comparisons Using Rank Sums. *Technometrics*. 1964;6(3):241-52. Accessed June 29, 2025. Available from: <https://www.stat.cmu.edu/technometrics/59-69/VOL-06-03/v0603241.pdf>.
- [25] Torgo L. Data Mining using R: Learning with Case Studies. CRC Press; 2010.

- [26] Torgo L. SMOTE: Synthetic Minority Over-sampling Technique;. <https://www.rdocumentation.org/packages/DMwR/versions/0.4.1/topics/SMOTE>.
- [27] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-Sampling Technique. *Journal of Artificial Intelligence Research*. 2002;16.
- [28] DataCamp. K-Fold Cross-Validation: O que é e Como Funciona;. <https://www.datacamp.com/pt/tutorial/k-fold-cross-validation>.
- [29] Rodrigo JA. RPubS - Árboles de predicción: bagging, random forest, boosting y C5.0;. Accedido: 2024-07-15. https://rpubs.com/Joaquin_AR/255596.
- [30] IBM. What is random forest?; 2021. Accedido: 2024-06-04. <https://www.ibm.com/topics/random-forest>.
- [31] Agresti A. *Categorical Data Analysis*. 3rd ed. John Wiley & Sons; 2013. Chapters on Multinomial Logistic Regression.
- [32] Wu JCH, Chien TC, Chang CC, Chang HI, Tsai HJ. Learning-Based Progression Detection of Alzheimer's Disease Using 3D MRI Images. *International Journal of Intelligent Systems*. 2025. Article Open Access.
- [33] Olatunde OO, Oyetunde KS, Han J, Khasawneh MT, Yoon H. Multiclass classification of Alzheimer's disease prodromal stages using sequential feature embeddings and regularized multikernel support vector machine. *NeuroImage*. 2024. Open Access.
- [34] Alp S, Akan T, Bhuiyan MS, Disbrow EA, Conrad SA. Joint transformer architecture in brain 3D MRI classification: its application in Alzheimer's disease classification. *Scientific Reports*. 2024. Open Access.
- [35] Asanza WR, Olivo BM. *Redes neuronales artificiales aplicadas al reconocimiento de patrones*. vol. 1. Editorial UTMACH; 2018.

A. Anexos

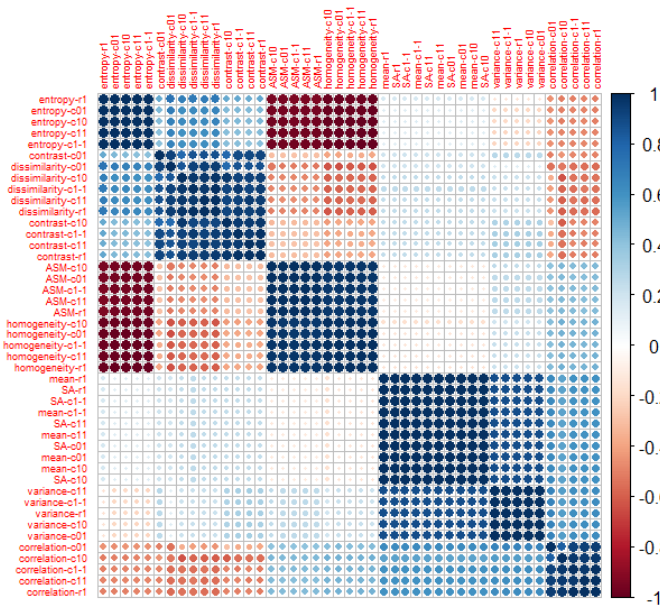


Gráfico A.1: Matriz de correlación entre las métricas GLCM de todas las direcciones.

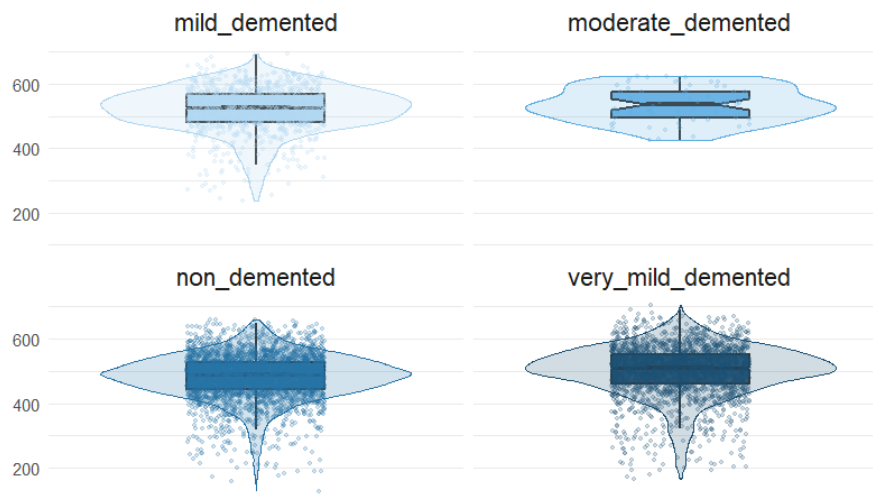


Gráfico A.2: Variable Contrast desagregada por niveles de deterioro cognitivo

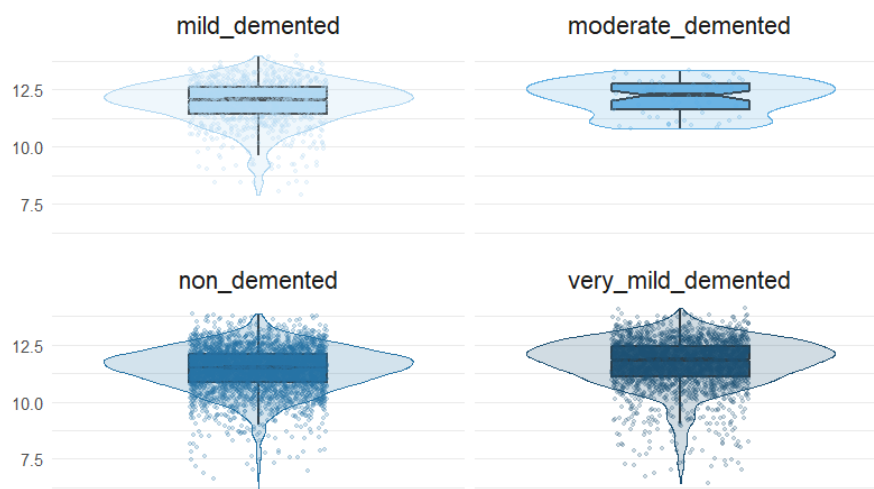


Gráfico A.3: Variable Dissimilarity desagregada por niveles de deterioro cognitivo

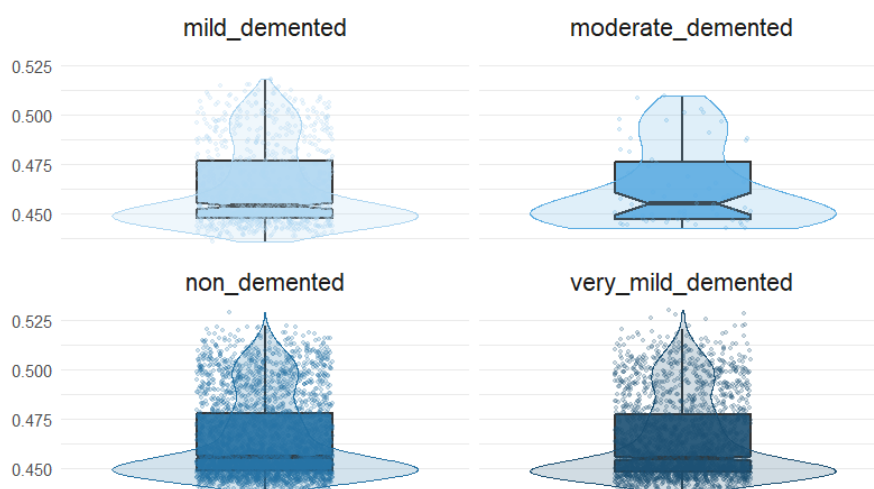


Gráfico A.4: Variable Homogeneity desagregada por niveles de deterioro cognitivo

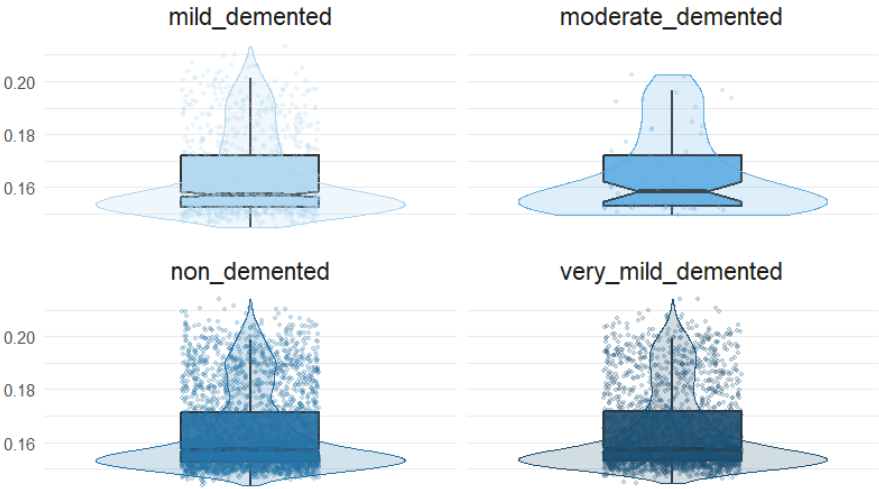


Gráfico A.5: Variable ASM desagregada por niveles de deterioro cognitivo

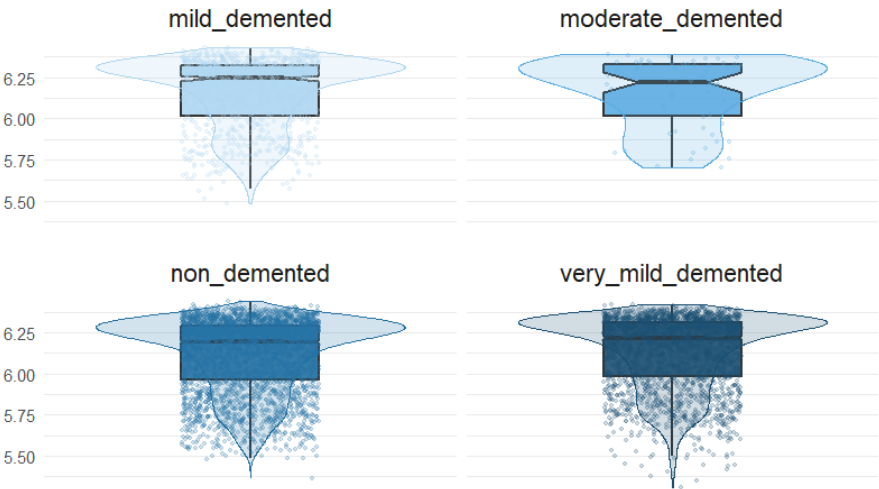


Gráfico A.6: Variable Entropy desagregada por niveles de deterioro cognitivo

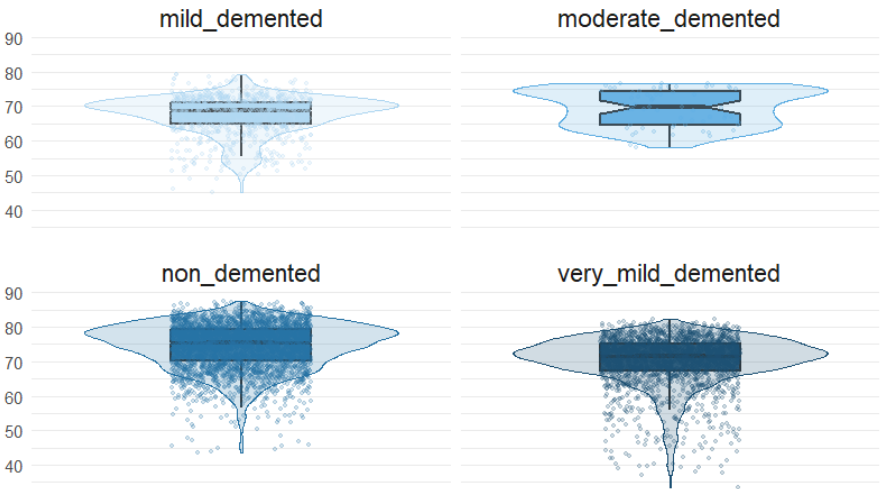


Gráfico A.7: Variable Mean desagregada por niveles de deterioro cognitivo

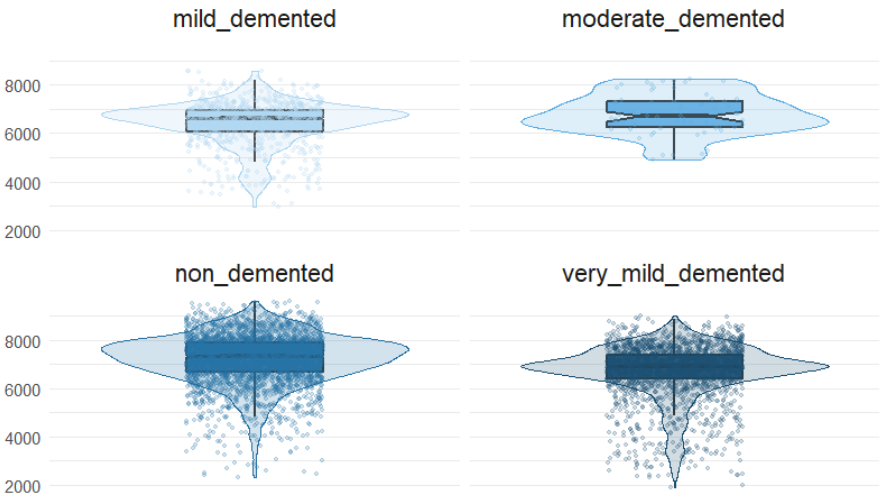


Gráfico A.8: Variable Variance desagregada por niveles de deterioro cognitivo

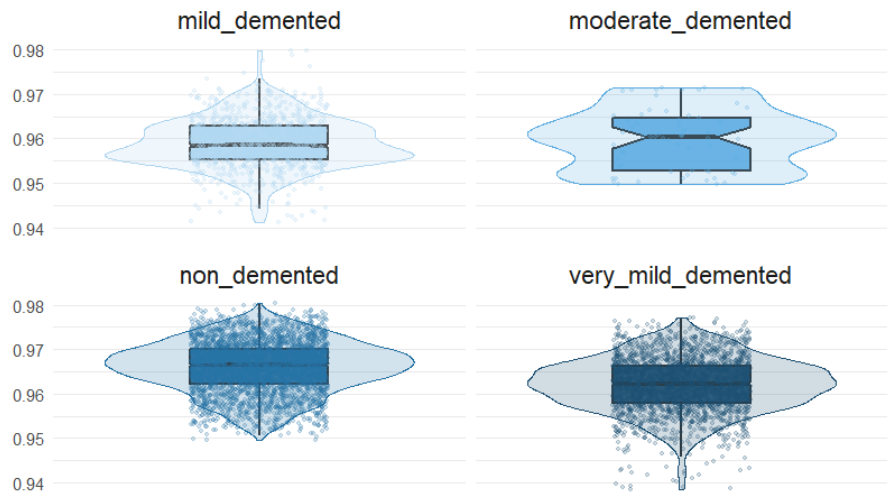


Gráfico A.9: Variable Correlation desagregada por niveles de deterioro cognitivo

```
# weights: 20 (12 variable)
initial value 28102.959289
iter 10 value 25997.616236
final value 25163.523695
converged
Call:
multinom(formula = class ~ ., data = train_data)

Coefficients:
              (Intercept) glcm_contrast glcm_entropy glcm_mean
mild_demented      10.247298    0.014189368   -0.2776091  -0.2209207
moderate_demented   11.651699    0.017919016   -1.5807451  -0.1533513
very_mild_demented   7.571914    0.008017613   -0.3715583  -0.1277769

Std. Errors:
              (Intercept) glcm_contrast glcm_entropy glcm_mean
mild_demented   0.007751673    0.0003734441    0.04734935    0.003935559
moderate_demented 0.008519884    0.0004068778    0.05389633    0.003838467
very_mild_demented 0.006782260    0.0003532559    0.04475512    0.003430884

Residual Deviance: 50327.05
AIC: 50351.05
```

Figura 11: Resultado del modelo de regresión logística multinomial sobre las variables *Contrast*, *Entropy* y *Mean* de la dirección radial.

```
# weights: 20 (12 variable)
initial value 28102.959289
iter 10 value 26318.463639
final value 25027.593063
converged
Call:
multinom(formula = class ~ ., data = train_data)

Coefficients:
              (Intercept)          Dim.1          Dim.2          Dim.3
mild_demented    -0.30724685  0.10465547 -0.5198360  1.4736356
moderate_demented -0.28735509  0.18641336 -0.1385483  1.5365198
very_mild_demented 0.07131468  0.04348453 -0.2957077  0.8624377

Std. Errors:
              (Intercept)          Dim.1          Dim.2          Dim.3
mild_demented    0.02539280  0.01198927  0.01664666  0.02800590
moderate_demented 0.02520721  0.01239709  0.01935785  0.02833585
very_mild_demented 0.02166400  0.01121919  0.01519876  0.02516294

Residual Deviance: 50055.19
AIC: 50079.19
```

Figura 12: Resultado del modelo de regresión logística multinomial sobre el PCA de la dirección radial.

B. Anexos

El código utilizado en este trabajo puede encontrarse en el siguiente enlace: https://github.com/juguardia/TFM_ALZHEIMER

**Declaración Responsable sobre Autoría y Uso Ético de
Herramientas de Inteligencia Artificial (IA)**

Yo, Julia Martins

Con DNI/NIE/PASAPORTE:

GF057903

declaro de manera responsable que el/la presente:

- ☐ Trabajo de Fin de Grado (TFG)
- ☒ Trabajo de Fin de Máster (TFM)
- ☐ Tesis Doctoral

Titulado/a

CLASIFICACIÓN SUPERVISADA DE IMÁGENES DE RESONANCIA MAGNÉTICA PARA LA DETECCIÓN DE ALZHEIMER

es el resultado de mi trabajo intelectual personal y creativo, y ha sido elaborado de acuerdo con los principios éticos y las normas de integridad vigentes en la comunidad académica y, más específicamente, en la Universidad Complutense de Madrid.

Soy, pues, autor del material aquí incluido y, cuando no ha sido así y he tomado el material de otra fuente, lo he citado o bien he declarado su procedencia de forma clara -incluidas, en su caso, herramientas de inteligencia artificial-. Las ideas y aportaciones principales incluidas en este trabajo, y que acreditan la adquisición de competencias, son mías y no proceden de otras fuentes o han sido reescritas usando material de otras fuentes.

Asimismo, aseguro que los datos y recursos utilizados son legítimos, verificables y han sido obtenidos de fuentes confiables y autorizadas. Además, he tomado medidas para garantizar la confidencialidad y privacidad de los datos utilizados, evitando cualquier tipo de sesgo o discriminación injusta en el tratamiento de la información.

En Madrid a 30 de Junio de 2025.

Julia Mantimo Guardia