REVISTA CUBANA
DE TRANSFORMACIÓN DIGITAL

RNPS 2487 • ISSN 2708-3411
Vol. 6 • 2026 • e299

ARTÍCULO ORIGINAL

Integración multimodal de datos cognitivos y morfométricos para el diagnóstico de la etapa de progresión de la enfermedad de Alzheimer

Multimodal Integration of Cognitive and Morphometric Data for Staging Alzheimer's Disease Progression

Sheyla León Pino
sheylapino28@gmail.com • *https://orcid.org/0009-0007-0194-9436*

Eduardo Garea Llanos
egarea@gmail.com • *https://orcid.org/0000-0001-9101-1589*

CENTRO DE NEUROCIENCIAS DE CUBA

Recibido: 2026-02-03 • Aceptado: 2026-06-23

RESUMEN

Este estudio aborda el diagnóstico temprano y preciso de la enfermedad de Alzheimer, un desafío clave en neurología clínica. Investigamos un enfoque que combina biomarcadores morfométricos de resonancia magnética con evaluaciones cognitivas, buscando superar las limitaciones de los métodos diagnósticos que utilizan estas fuentes de información de forma aislada. Analizamos datos de 1,382 participantes de la base ADNI, clasificados en controles cognitivos, deterioro cognitivo leve (MCI) y Alzheimer. Implementamos algoritmos de aprendizaje automático, destacando XGBoost con una precisión equilibrada de 83.72%, superior a los modelos basados solo en variables cognitivas. Los resultados indican que la integración multimodal mejoró la detección del estado MCI, con un aumento en sensibilidad. Los biomarcadores con mayor poder predictivo fueron los volúmenes hipocampales, de la corteza entorrinal y las medidas de sustancia gris total. Este enfoque ofrece una herramienta diagnóstica más precisa y se alinea con el marco de investigación NIA-AA, que enfatiza la definición biológica del Alzheimer a través de biomarcadores. Un diagnóstico más temprano y preciso permitiría intervenciones más oportunas y una mejor selección de pacientes para ensayos clínicos. Este trabajo demuestra el potencial de la inteligencia artificial para integrar diferentes fuentes de información en un modelo predictivo, con posibles aplicaciones en la práctica clínica.

Palabras clave: Alzheimer; biomarcadores; cognición; inteligencia artificial; resonancia magnética.

ABSTRACT

This study addresses the early and accurate diagnosis of Alzheimer's disease, a key challenge in clinical neurology. We investigate an approach that combines morphometric biomarkers from magnetic resonance imaging with cognitive assessments, seeking to overcome the limitations of diagnostic methods that use these information sources in isolation. We analyzed data from 1,382 participants from the ADNI database, classified into cognitive controls, mild cognitive impairment (MCI), and Alzheimer's. We implemented machine learning algorithms, highlighting XGBoost with a balanced accuracy of 83.72%, superior to models based only on cognitive variables. The results indicate that multimodal integration improved the detection of MCI status, with an increase in sensitivity. The biomarkers with the greatest predictive power were hippocampal volumes, entorhinal cortex, and total gray matter measurements. This approach offers a more accurate diagnostic tool and aligns with the NIA-AA research framework, which emphasizes the biological definition of Alzheimer's through biomarkers. An earlier and more precise diagnosis would allow for more timely interventions and better selection of patients for clinical trials. This work demonstrates the potential of artificial intelligence to integrate different sources of information into a predictive model, with possible applications in clinical practice.

Keywords: Alzheimer's disease; biomarkers; cognition; artificial intelligence; magnetic resonance imaging.

INTRODUCCIÓN

La enfermedad de Alzheimer (EA) es la principal causa de demencia a nivel mundial y representa un desafío sanitario y socioeconómico de gran magnitud. Según la Organización Mundial de la Salud, se estima que para 2050 el número de personas afectadas podría triplicarse, alcanzando más de 150 millones de casos globalmente. Este incremento exponencial está asociado al envejecimiento poblacional, lo que subraya la urgencia de mejorar las estrategias de diagnóstico y manejo clínico (World Health Organization, 2022).

El diagnóstico temprano es fundamental, ya que la neurodegeneración comienza años antes de la manifestación clínica evidente. Sin embargo, la heterogeneidad en la presentación clínica y la superposición con otros trastornos neurodegenerativos dificultan la identificación precisa de las etapas iniciales de la EA (Dubois et al., 2021). Las limitaciones de las pruebas neuropsicológicas tradicionales incluyen su susceptibilidad a factores culturales, educativos y motivacionales, lo que puede llevar a sub o sobrediagnóstico, especialmente en poblaciones diversas (Barnes et al., 2009).

En este contexto, la neuroimagen ha revolucionado la investigación y la práctica clínica, permitiendo la visualización in vivo de los cambios estructurales cerebrales asociados a la EA. Los avances en resonancia magnética (RM) han posibilitado la cuantificación precisa de volúmenes cerebrales y el análisis de patrones de atrofia

característicos, como la afectación temprana del hipocampo y la corteza entorrinal (Braak & Braak, 1991; Jack et al., 1992).

No obstante, la interpretación aislada de biomarcadores morfométricos puede ser insuficiente debido a la variabilidad interindividual y la influencia de factores comórbidos, como enfermedades vasculares o depresión. Por ello, la integración de múltiples fuentes de datos —cognitivos, clínicos y de neuroimagen— mediante enfoques de inteligencia artificial representa una solución innovadora y prometedora para mejorar la precisión diagnóstica.

La integración multimodal se fundamenta en el marco de investigación NIA-AA, que propone una definición biológica de la EA basada en biomarcadores, desplazando el enfoque exclusivamente clínico (Jack et al., 2018). Este paradigma reconoce que la acumulación de amiloide, la taupatía y la neurodegeneración pueden detectarse mediante técnicas de neuroimagen y análisis de líquidos biológicos antes de la aparición de síntomas cognitivos.

Estudios recientes han demostrado que los modelos predictivos que combinan datos de neuroimagen con evaluaciones cognitivas superan en sensibilidad y especificidad a los métodos tradicionales (Falahati et al., 2014; Risacher et al., 2019). Particularmente, el uso de algoritmos de aprendizaje automático permite identificar patrones complejos y no lineales en grandes conjuntos de datos, facilitando la clasificación precisa de las distintas etapas de la EA.

La base de datos ADNI ha sido esencial para el desarrollo y validación de estos modelos, proporcionando un repositorio estandarizado de datos clínicos, neuropsicológicos y de neuroimagen de miles de participantes (Weiner et al., 2013). El acceso a estos datos ha impulsado la reproducibilidad y la comparación de resultados entre diferentes grupos de investigación.

El objetivo principal de este estudio es desarrollar y validar un modelo predictivo que integre de manera simultánea datos cognitivos (provenientes de evaluaciones neuropsicológicas) y biomarcadores morfométricos obtenidos mediante resonancia magnética, con el fin de mejorar la precisión y sensibilidad en la clasificación de las etapas de progresión de la enfermedad de Alzheimer. Se busca, específicamente, superar las limitaciones de los enfoques diagnósticos tradicionales que emplean estas fuentes de información de forma aislada, logrando así una detección más temprana y precisa de la enfermedad, especialmente en la etapa de deterioro cognitivo leve (MCI, por sus siglas en inglés Mild Cognitive Impairment).

Este trabajo propone una metodología innovadora basada en la integración multimodal de datos, utilizando algoritmos avanzados de aprendizaje automático (como XGBoost y LightGBM), para construir un modelo capaz de distinguir entre controles cognitivos, pacientes con MCI y pacientes con Alzheimer. Esta integración permite aprovechar la complementariedad entre los cambios estructurales cerebrales detectados por neuroimagen y el desempeño cognitivo, aumentando significativamente la capacidad discriminativa respecto a modelos que solo consideran una de estas dimensiones.

Además, el estudio plantea que este enfoque puede convertirse en una herramienta diagnóstica aplicable en la práctica clínica, alineándose con el marco de investigación NIA-AA que prioriza la definición biológica de la enfermedad mediante biomarcadores. Se espera que la implementación de este modelo facilite intervenciones más oportunas y una mejor selección de pacientes para ensayos clínicos, contribuyendo así a una gestión clínica más eficiente y personalizada de la enfermedad de Alzheimer.

Aunque numerosos estudios han empleado la base ADNI y algoritmos de aprendizaje automático para clasificar sujetos a lo largo del continuo NC–MCI–EA, muchos se han centrado en un único dominio de datos (por ejemplo, solo

neuroimagen o solo pruebas cognitivas) o en tamaños muestrales más pequeños. En este trabajo se propone un enfoque multimodal que integra un conjunto amplio de biomarcadores morfométricos con evaluaciones neuropsicológicas en una cohorte de 1 382 participantes, utilizando algoritmos de última generación (XGBoost, LightGBM) con optimización sistemática de hiperparámetros. Este enfoque permite evaluar de forma cuantitativa el beneficio real de la integración multimodal frente a modelos unimodales y proporciona evidencia específica sobre la mejora en la detección del MCI, etapa crítica para la intervención temprana.

METODOLOGÍA

Diseño del estudio y población

Se realizó un estudio observacional retrospectivo utilizando datos provenientes de las cohortes ADNI-1, ADNI-GO y ADNI-2 del estudio Alzheimer's Disease Neuroimaging Initiative (ADNI), un proyecto longitudinal y multicéntrico diseñado para caracterizar la progresión de la enfermedad de Alzheimer, proporcionando datos estandarizados de neuroimagen, biomarcadores y evaluaciones cognitivas (Thompson et al., 2017; Weiner et al., 2013). La muestra incluyó 1,382 participantes, clasificados en tres grupos diagnósticos: controles cognitivos (NC), deterioro cognitivo leve (MCI) y enfermedad de Alzheimer (EA), de acuerdo con los criterios estandarizados del DSM-IV y el National Institute on Aging. La clasificación diagnóstica se basó en evaluaciones clínicas, neuropsicológicas y funcionales realizadas por expertos.

Variables y fuentes de datos

Se analizaron tres dominios principales de variables:

- Evaluaciones cognitivas: Incluyeron pruebas estandarizadas como el Mini-Mental State Examination (MMSCORE), escalas de funciones ejecutivas y funcionales (por ejemplo, FAQ), y otras pruebas neuropsicológicas relevantes para la detección de deterioro cognitivo.

- Variables demográficas y clínicas: Edad, nivel educativo, género, estado civil y severidad clínica, que permiten ajustar los modelos y controlar posibles factores de confusión.

- Biomarcadores de neuroimagen: Volúmenes hipocampales, volumen de sustancia blanca, volúmenes amigdalinos, volumen de la corteza entorrinal, medidas de sustancia gris y volumen cerebral total, obtenidos mediante resonancia magnética estructural y procesados con herramientas automatizadas como FreeSurfer.

Preprocesamiento de datos

El flujo de análisis se implementó en Python 3.10 usando Jupyter Notebooks, con control de versiones en Git. Se utilizaron las bibliotecas pandas, numpy y scikit-learn. Los datos crudos se cargaron desde un archivo CSV delimitado por punto y coma, y se filtraron las columnas de interés mediante un script personalizado. La variable dependiente (DXCURREN) se codificó numéricamente (0 = NC, 1 = MCI, 2 = EA) para su procesamiento algorítmico.

Para manejar valores faltantes, se aplicó el método de imputación k-nearest neighbors (KNNImputer) para variables con menos del 15% missingness, seleccionándose k=5 como un compromiso entre estabilidad de la imputación y complejidad computacional, en concordancia con recomendaciones habituales en la literatura, preservando la mayor cantidad posible de información. Los datos se dividieron en un conjunto de entrenamiento

(80%) y otro de prueba (20%) mediante partición estratificada (train_test_split), garantizando la representación proporcional de cada grupo diagnóstico. Las variables continuas se estandarizaron (StandardScaler) o normalizaron (RobustScaler, MinMaxScaler) según su distribución, utilizando umbrales de asimetría (skewness) calculados mediante análisis exploratorio. Esto permitió que todas las variables tuvieran un peso comparable en los modelos.

- StandardScaler para edad (skewness=0.12)
- RobustScaler para volúmenes cerebrales (skewness >1.5)
- MinMaxScaler para variables ordinales

Análisis estadístico y modelización

El análisis exploratorio incluyó visualizaciones interactivas usando matplotlib/seaborn: distribuciones de edad/educación con histogramas y medidas de asimetría, diagramas de caja comparando biomarcadores por grupo diagnóstico y matriz de correlación de 21x21 variables (Figura 1). Para la clasificación diagnóstica, se evaluaron seis algoritmos de aprendizaje automático mediante validación cruzada estratificada (5-fold) para evaluar la estabilidad y generalización de los modelos, minimizando el riesgo de sobreajuste:

- Random Forest (RF),
- Support Vector Machine (SVM),
- XGBoost (XGB),
- LightGBM (LGBM),
- Regresión logística multinomial,
- k-nearest neighbors (KNN).

La selección de estos algoritmos se basó en su capacidad para manejar datos heterogéneos, su robustez ante variables correlacionadas y su uso frecuente en estudios biomédicos.

La optimización de hiperparámetros se realizó mediante Optuna, una herramienta de optimización bayesiana, con 30 iteraciones por modelo. La búsqueda se llevó a cabo dentro de cada fold de la validación cruzada estratificada de 5 particiones, asegurando que la selección de hiperparámetros no se beneficiara de información del conjunto de prueba y evitando así la fuga de datos. Para XGBoost, se ajustaron parámetros como learning_rate (0.01–1), max_depth (5–30) y subsample (0.5–1). En LightGBM, se optimizaron lambda_l1 (0.01–0.1) y min_child_samples (20–70). El rendimiento de los modelos se evaluó mediante precisión equilibrada (balanced accuracy), puntaje F1 ponderado (weighted F1-score), curvas ROC multiclase y matriz de confusión normalizada. El rendimiento de los modelos se midió mediante:

- Precisión equilibrada (balanced accuracy): Evalúa la capacidad del modelo para clasificar correctamente cada grupo, teniendo en cuenta posibles desbalances de clase. Se seleccionó porque, en contextos con clases desbalanceadas, proporciona una estimación más representativa que la precisión simple al calcular el promedio del recall por clase y reducir el sesgo hacia la clase mayoritaria, lo cual resulta adecuado para estudios como ADNI, donde los grupos diagnósticos suelen ser desiguales.

- Puntaje F1 ponderado (weighted F1-score): Considera tanto la precisión como la exhaustividad, ponderando por el tamaño de cada clase.
- Curvas ROC multiclase: Permiten visualizar la capacidad discriminativa global del modelo.
- Matriz de confusión normalizada: Facilita la interpretación de los aciertos y errores en cada grupo diagnóstico.

Además, se comparó el rendimiento de los distintos modelos mediante dos pruebas estadísticas complementarias sobre los resultados de la validación cruzada estratificada y el conjunto de prueba, considerando diferencias significativas aquellas con un valor de $p < 0.05$:

Test de Wilcoxon signed-rank: Se aplicó para comparar los scores de accuracy por fold (n=5) entre pares de modelos durante la validación cruzada. Esta prueba no paramétrica es adecuada para datos ordinales/continuos pequeños y emparejados, evaluando si las diferencias sistemáticas en rendimiento entre algoritmos son estadísticamente significativas sin asumir normalidad de las diferencias entre folds.

Test de McNemar: Se utilizó para analizar las tablas de contingencia de predicciones en el conjunto de prueba independiente, enfocándose en errores discordantes (casos donde un modelo acierta y otro falla). Es específicamente sensible a diferencias prácticas en clasificación binaria/multiclase, evaluando directamente si un clasificador es superior al otro en términos de decisiones correctas vs. incorrectas.

La combinación de ambas pruebas proporcionó una evaluación robusta: Wilcoxon analizó la consistencia del rendimiento a través de folds (estabilidad), mientras McNemar evaluó la superioridad práctica en predicciones reales (utilidad clínica). Esta estrategia dual es estándar en machine learning biomédico para evitar conclusiones sesgadas por usar una sola métrica estadística.

Implementación del pipeline

Se construyó un pipeline de procesamiento integrado que incluía:

- Imputación de datos faltantes
- Escalado de características
- Clasificación mediante el modelo optimizado

Este pipeline garantizó la reproducibilidad y consistencia del proceso, desde la carga de datos hasta la predicción final. El desarrollo se realizó en Jupyter Notebooks y se versionó mediante Git para asegurar trazabilidad y control de cambios.

Análisis de importancia de características

Se aplicaron técnicas de interpretación de modelos (feature importance) para identificar las variables más relevantes en la clasificación diagnóstica. Se observó que los volúmenes hipocampales, el volumen de la corteza entorrinal y las medidas de sustancia gris total fueron los biomarcadores con mayor poder predictivo, mientras que las pruebas cognitivas como MMSCORE y FAQTOTAL (puntaje total de la Functional Activities Questionnaire) mantuvieron alta relevancia, especialmente al integrarse con los datos morfométricos.

RESULTADOS Y DISCUSIÓN

El análisis comparativo de los seis algoritmos de aprendizaje automático demostró que los modelos basados en árboles de decisión, particularmente XGBoost y LightGBM, ofrecieron el mejor rendimiento para la clasificación de las tres etapas diagnósticas: controles cognitivos (NC), deterioro cognitivo leve (MCI) y enfermedad de Alzheimer (AD) (Tabla1). Tras la optimización de hiperparámetros, el modelo XGBoost alcanzó una precisión equilibrada del 83.72% y un puntaje F1 ponderado de 83.69% en el conjunto de validación, superando significativamente a los modelos basados únicamente en variables cognitivas, que lograron una precisión equilibrada de 78% (Tabla2).

Tabla1

Modelo Original	Accuracy	Recall	Precision	F1 Score
XGB_CLF	83.44%	83.44%	83.69%	83.41%
RF_CLF	82.18%	82.18%	82.58%	82.17%
KNN_CLF	73.32%	73.32%	73.63%	73.37%
LGBM_CLF	83.17%	83.17%	83.44%	83.16%
LR	58.03%	58.03%	59.35%	57.30%
SVC_CLF	56.27%	56.27%	71.61%	46.12%

Tabla2

Modelo Optimizado (Optuna)	Accuracy	Recall	Precision	F1 Score
TUNED_XGB_CLF	83.72%	83.72%	84.00%	83.69%
TUNED_LGBM_CLF	83.66%	83.66%	83.68%	83.65%
TUNED_RF_CLF	82.51%	82.51%	82.92%	82.49%

Las diferencias en precisión equilibrada entre el modelo XGBoost y los demás algoritmos fueron estadísticamente significativas en la validación cruzada estratificada ($p < 0.05$), lo que respalda que la superioridad observada no se debe al azar.

Evaluación del Modelo Final mediante Curva ROC y Matriz de Confusión

Para evaluar en profundidad el rendimiento del modelo final optimizado, se generaron la curva ROC multiclase y la matriz de confusión correspondiente. La Curva ROC (Figura 1) confirma la excelente capacidad discriminatoria del modelo para distinguir entre las tres clases. Los valores del Área Bajo la Curva (AUC) fueron de 0.95 para la clase NC, 0.92 para MCI y 0.95 para AD. Estos resultados indican un rendimiento excepcional, cercano al clasificador

perfecto (AUC=1.0), siendo la clase MCI, como era de esperar dada su naturaleza de transición, la que presenta un AUC ligeramente inferior, aunque aún sobresaliente.

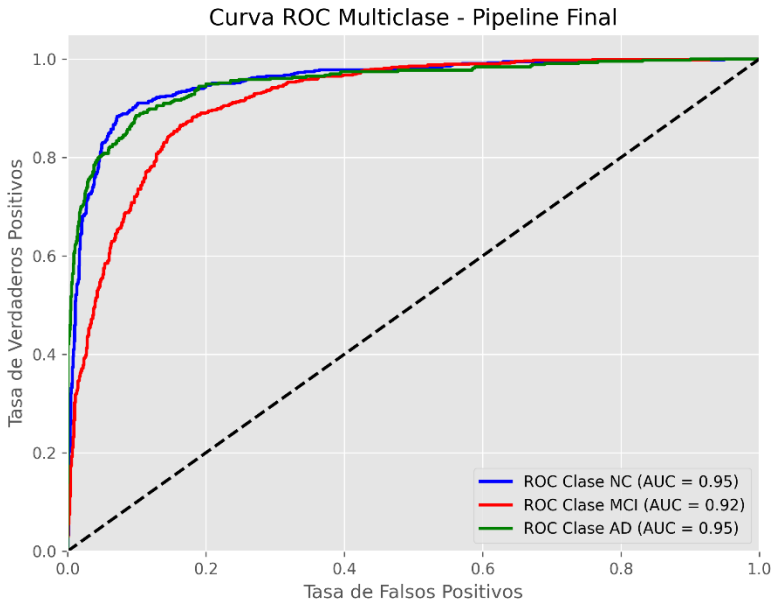


Figura 1. Curvas ROC multiclase del modelo XGBoost para la clasificación de controles cognitivos (NC), deterioro cognitivo leve (MCI) y enfermedad de Alzheimer (EA) en el conjunto de prueba, con los correspondientes valores de área bajo la curva (AUC) para cada clase.

La Matriz de Confusión (Figura 2) proporciona una visión detallada de los aciertos y errores del clasificador. Se observa que la mayoría de las predicciones se concentran en la diagonal principal, lo que indica clasificaciones correctas. Por ejemplo, de 549 sujetos NC, 464 fueron correctamente identificados. Los errores más significativos se produjeron en la confusión entre NC y MCI, con 80 sujetos NC clasificados erróneamente como MCI, y 66 sujetos MCI como NC. Asimismo, existe una superposición discernible entre MCI y AD, con 93 casos de AD siendo predichos como MCI. Este patrón de confusión es clínicamente plausible, ya que refleja la continuidad sintomática y patológica entre estas etapas.

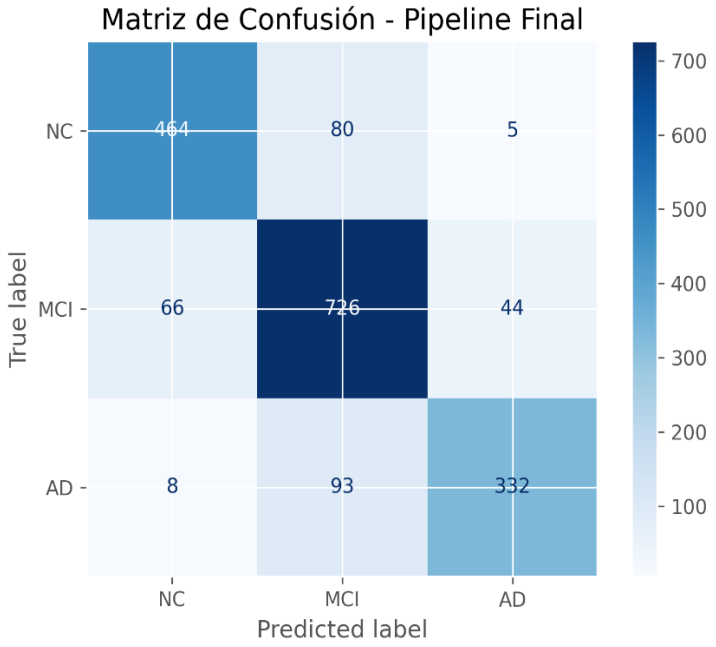


Figura 2. Matriz de confusión normalizada del modelo XGBoost en el conjunto de prueba, que muestra las proporciones de clasificaciones correctas en la diagonal principal y los casos de confusión entre controles cognitivos (NC), deterioro cognitivo leve (MCI) y enfermedad de Alzheimer (AD).

Métricas de Clasificación por Clase

El Reporte de Clasificación (Figura 3) desglosa las métricas de precisión, exhaustividad (recall) y puntuación F1 para cada clase. El modelo logró una precisión balanceada entre las tres clases, con valores de puntuación F1 de 0.85 (NC), 0.84 (MCI) y 0.82 (AD). La exactitud (accuracy) global del modelo fue del 85%. Es notable que la clase MCI, a menudo la más difícil de clasificar, alcanzó la mayor exhaustividad (recall = 0.87), lo que significa que el modelo es particularmente eficaz para identificar correctamente a la mayoría de los individuos con esta condición. Por su parte, las clases NC y AD mostraron una mayor precisión (0.86 y 0.87, respectivamente), indicando que cuando el modelo predice estas clases, es muy probable que sea correcto.

Reporte de Clasificación:

	precision	recall	f1-score	support
NC	0.86	0.84	0.85	549
MCI	0.81	0.87	0.84	836
AD	0.87	0.77	0.82	433
accuracy			0.84	1818
macro avg	0.85	0.83	0.84	1818
weighted avg	0.84	0.84	0.84	1818

Figura 3. Métricas de precisión, exhaustividad (recall) y F1-score por clase para el modelo XGBoost, junto con la exactitud global, en la tarea de clasificación de controles cognitivos (NC), deterioro cognitivo leve (MCI) y enfermedad de Alzheimer (AD).

Importancia de las variables predictoras

El análisis de importancia de características reveló que los volúmenes hipocampales (TOTAL_HIPPO), el volumen de la corteza entorrinal (ENTORHINAL_VOLUME) y las medidas de sustancia gris total (TOTAL_GRAY) fueron los biomarcadores con mayor poder predictivo en el modelo final. Entre las variables cognitivas, el Mini-Mental State Examination (MMSCORE) y la escala funcional FAQTOTAL mantuvieron alta relevancia, pero su capacidad discriminativa se potenció notablemente al combinarse con los datos morfométricos, especialmente para diferenciar entre controles y sujetos con MCI.

El mapa de calor de correlación entre características (Figura 4) mostró una asociación significativa entre los volúmenes hipocampales y el diagnóstico, reforzando la evidencia previa sobre la importancia de la atrofia mesial temporal en la EA.

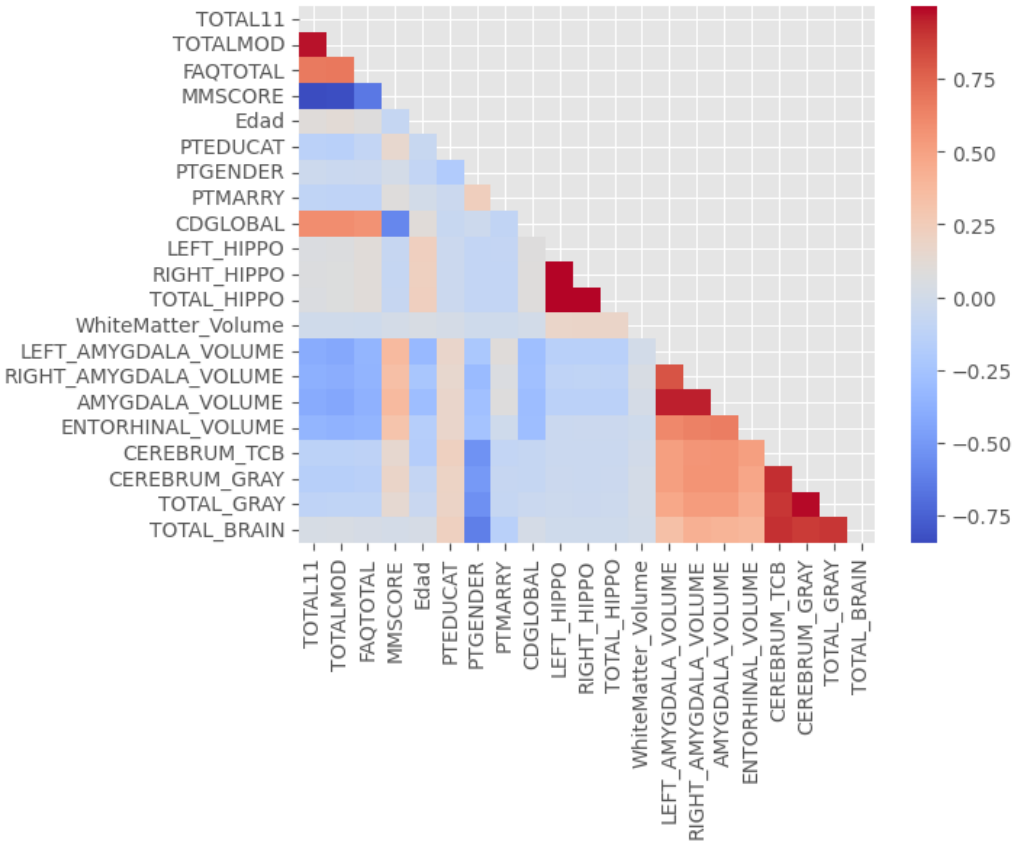


Figura 4: Figura 4. Mapa de calor de correlación entre las características cognitivas y morfométricas incluidas en el modelo, destacando la asociación entre los volúmenes hipocampales y el diagnóstico a lo largo del continuo NC–MCI–AD.

El análisis de importancia de características reveló que las variables cognitivas MMSCORE y FAQTOTAL mantuvieron una alta relevancia predictiva incluso en el modelo integrado, pero su poder discriminativo aumentó cuando se combinaron con los biomarcadores morfométricos, especialmente para diferenciar entre NC y MCI. La

validación cruzada estratificada (5-fold) confirmó la robustez y estabilidad del modelo, con una desviación estándar de la precisión de ±0.04 entre las particiones, lo que indica un bajo riesgo de sobreajuste y una buena capacidad de generalización a nuevos datos dentro de la cohorte ADNI (Figura 5). El pipeline integrado demostró consistencia en diferentes subconjuntos de la base ADNI, sugiriendo buena generalización a nuevos datos.

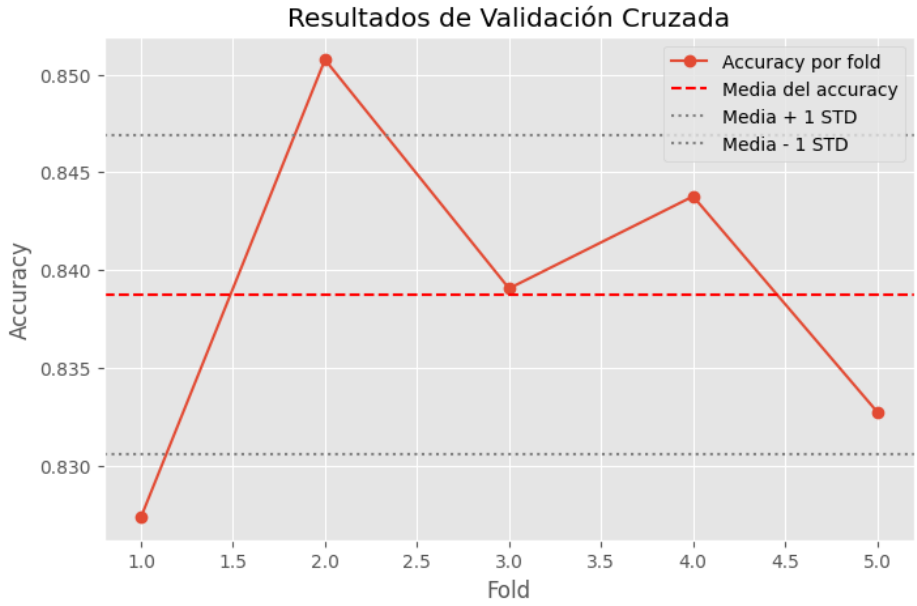


Figura 5. Rendimiento del modelo XGBoost en la validación cruzada estratificada (5-fold), donde se observa la estabilidad de la precisión entre particiones y la baja variabilidad del modelo.

Los resultados de este estudio muestran que la integración multimodal de biomarcadores morfométricos derivados de resonancia magnética con evaluaciones cognitivas permite mejorar de forma significativa la clasificación de las etapas de progresión de la enfermedad de Alzheimer, en comparación con modelos basados exclusivamente en variables cognitivas. En particular, el modelo XGBoost optimizado alcanzó una precisión equilibrada superior al 83%, con un rendimiento especialmente destacado en la identificación de sujetos con deterioro cognitivo leve, etapa crítica para la intervención temprana.

Estos hallazgos son coherentes con la evidencia previa que señala la utilidad de la neuroimagen estructural y de los volúmenes hipocampales y de corteza entorrinal como marcadores sensibles de neurodegeneración en la enfermedad de Alzheimer. Sin embargo, el presente trabajo aporta evidencia adicional al demostrar que la combinación explícita de estas medidas morfométricas con pruebas cognitivas estandarizadas, en una cohorte amplia de la base ADNI, incrementa la sensibilidad y mantiene una elevada especificidad respecto a enfoques unimodales. De este modo, el estudio refuerza la importancia de abordar el diagnóstico desde una perspectiva integradora que considere simultáneamente los cambios estructurales cerebrales y el rendimiento cognitivo.

Una contribución relevante de este trabajo es la implementación de un pipeline completo de aprendizaje automático que incluye imputación de valores faltantes, escalado de características, optimización bayesiana de hiperparámetros y validación cruzada estratificada, aplicado a múltiples algoritmos de clasificación. La comparación sistemática mostró que los modelos basados en árboles de decisión potenciados, como XGBoost y LightGBM, superan a otros métodos clásicos como regresión logística, SVM o KNN, lo que sugiere que las relaciones no lineales y las

interacciones entre variables juegan un papel importante en la discriminación entre NC, MCI y EA. Además, el análisis de importancia de características confirmó el papel central de los volúmenes hipocampales, la corteza entorrinal y la sustancia gris total, junto con pruebas como el Mini-Mental State Examination y la Functional Activities Questionnaire, como predictores claves en el contexto multimodal.

Desde la perspectiva del marco NIA-AA, que promueve una definición biológica de la enfermedad basada en biomarcadores, los resultados de este estudio respaldan el valor de integrar marcadores de neurodegeneración con medidas cognitivas para caracterizar con mayor precisión el continuo clínico de la enfermedad de Alzheimer. En particular, la mejora en la identificación de la etapa MCI tiene implicaciones clínicas directas, ya que puede facilitar la selección de pacientes candidatos a intervenciones tempranas y a ensayos clínicos dirigidos a fases prodrómicas o iniciales de la enfermedad. Al mismo tiempo, la capacidad del modelo para mantener buenas tasas de clasificación en los grupos de controles y de demencia establecida contribuye a reducir errores diagnósticos potencialmente relevantes en la práctica asistencial.

A pesar de sus fortalezas, este estudio presenta limitaciones que deben ser consideradas al interpretar los resultados. En primer lugar, el diseño observacional y retrospectivo, basado en datos de la base ADNI, puede introducir sesgos de selección y limita la posibilidad de extrapolar directamente los hallazgos a poblaciones más diversas. En segundo lugar, la validación del modelo se realizó dentro de la propia cohorte ADNI, por lo que será necesario evaluar su desempeño en bases de datos independientes y en entornos clínicos reales para confirmar su generalización. Finalmente, aunque se integraron múltiples biomarcadores estructurales y pruebas cognitivas, no se incluyeron otros tipos de marcadores, como datos genéticos, de líquido cefalorraquídeo o de PET, que podrían enriquecer aún más la capacidad predictiva de los modelos y abrir nuevas líneas de investigación futura.

CONCLUSIONES

Este estudio demuestra que la integración multimodal de biomarcadores morfométricos obtenidos por resonancia magnética con evaluaciones cognitivas mejora de forma significativa la precisión en la clasificación de controles cognitivos, deterioro cognitivo leve y enfermedad de Alzheimer, en comparación con modelos basados únicamente en variables cognitivas. El modelo XGBoost optimizado alcanzó una precisión equilibrada superior al 83% y una alta capacidad discriminativa global, con especial eficacia para la identificación de sujetos con deterioro cognitivo leve, etapa clave para la intervención temprana.

Los hallazgos refuerzan la pertinencia de enfoques diagnósticos alineados con el marco NIA-AA, que combinan información estructural cerebral y desempeño cognitivo para definir de manera más precisa la enfermedad de Alzheimer a lo largo de su continuo clínico. La metodología propuesta, basada en un pipeline reproducible de aprendizaje automático y en una cohorte amplia de la base ADNI, sienta las bases para el desarrollo de herramientas de apoyo a la decisión clínica que podrían facilitar una selección más adecuada de pacientes para ensayos clínicos y estrategias terapéuticas personalizadas.

De cara al futuro, será necesario validar estos modelos en otras poblaciones y contextos asistenciales, así como explorar la incorporación de biomarcadores adicionales (genéticos, de líquido cefalorraquídeo o PET) que puedan potenciar aún más el rendimiento predictivo y la aplicabilidad clínica de los enfoques multimodales.

AGRADECIMIENTOS

Agradecemos al Centro de Neurociencias de Cuba y a la comunidad científica de la Iniciativa ADNI, cuyo compromiso con la investigación colaborativa hizo posible el acceso a datos esenciales para este estudio.

REFERENCIAS

Alzheimer's Association. (2023). 2023 Alzheimer's disease facts and figures. Alzheimer's & Dementia, 19(4), 1598-1695. https://doi.org/10.1002/alz.13016

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

Barnes, D. E., Yaffe, K., Byers, A. L., McCormick, M., Schaefer, C., & Whitmer, R. A. (2009). Midlife vs late-life depressive symptoms and risk of dementia: Differential effects for Alzheimer disease and vascular dementia. Archives of General Psychiatry, 66(5), 527-533. https://doi.org/10.1001/archgenpsychiatry.2009.39

Braak, H., & Braak, E. (1991). Neuropathological stageing of Alzheimer-related changes. Acta Neuropathologica, 82(4), 239-259. https://doi.org/10.1007/BF00308809

Dubois, B., Feldman, H. H., Jacova, C., Hampel, H., Molinuevo, J. L., Blennow, K., ... & Cummings, J. L. (2021). Revised criteria for diagnosis and staging of Alzheimer's disease. The Lancet Neurology, 20(6), 484-496. https://doi.org/10.1016/S1474-4422(21)00066-1

Dubois, B., Feldman, H. H., Jacova, C., Cummings, J. L., DeKosky, S. T., Barberger-Gateau, P., ... & Scheltens, P. (2016). Advancing research diagnostic criteria for Alzheimer's disease: The IWG-2 criteria. The Lancet Neurology, 13(6), 614-629. https://doi.org/10.1016/S1474-4422(14)70090-0

Falahati, F., Westman, E., & Simmons, A. (2014). Multimodal prediction of conversion to Alzheimer's disease in mild cognitive impairment. Journal of Alzheimer's Disease, 41(4), 1205-1219. https://doi.org/10.3233/JAD-140201

Fischl, B. (2012). FreeSurfer. NeuroImage, 62(2), 774-781. https://doi.org/10.1016/j.neuroimage.2012.01.021

Frisoni, G. B., Fox, N. C., Jack, C. R., Scheltens, P., & Thompson, P. M. (2010). The clinical use of structural MRI in Alzheimer disease. Nature Reviews Neurology, 6(2), 67-77. https://doi.org/10.1038/nrneurol.2009.215

Jack, C. R., Bennett, D. A., Blennow, K., Carrillo, M. C., Dunn, B., Haeberlein, S. B. ... & Sperling, R. (2018). NIA-AA research framework: Toward a biological definition of Alzheimer's disease. Alzheimer's & Dementia, 14(4), 535-562. https://doi.org/10.1016/j.jalz.2018.02.018

Jack, C. R., Petersen, R. C., O'Brien, P. C., & Tangalos, E. G. (1992). MR-based hippocampal volumetry in the diagnosis of Alzheimer's disease. Neurology, 42(1), 183-188. https://doi.org/10.1212/wnl.42.1.183

McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Kawas, C. H., ... & Phelps, C. H. (2011). The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's & Dementia, 7(3), 263–269. https://doi.org/10.1016/j.jalz.2011.03.005

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O. ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830. http://www.jmlr.org/papers/v12/pedregosa11a.html

Risacher, S. L., Saykin, A. J., West, J. D., Shen, L., Firpi, H. A., & McDonald, B. C. (2019). MRI biomarkers in Alzheimer's disease. NeuroImage: Clinical, 22, 101771. https://doi.org/10.1016/j.nicl.2019.101771

Thompson, P. M., et al. (2017). The Alzheimer's Disease Neuroimaging Initiative: Progress and innovation. Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, 8, 1-8. https://doi.org/10.1016/j.dadm.2017.02.001

Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Green, R. C., ... & Trojanowski, J. Q. (2013). The Alzheimer's Disease Neuroimaging Initiative: A review of papers published since its inception. Alzheimer's & Dementia, 9(5), e111-e194. https://doi.org/10.1016/j.jalz.2013.05.1769

World Health Organization. (2022). Global action plan on the public health response to dementia 2017-2025. https://www.who.int/publications/i/item/global-action-plan-on-the-public-health-response-to-dementia-2017---2025

Copyright © 2026, Autores: León Pino, Sheyla, Garea Llano, Eduardo

Esta obra está bajo una licencia de Creative Commons Atribución-No Comercial 4.0 Internacional