

**ANÁLISIS PREDICTIVO DE ALZHEIMER Y PARKINSON A PARTIR DE DATOS  
MÉDICOS USANDO MINERÍA DE DATOS  
PREDICTIVE ANALYSIS OF ALZHEIMER'S AND PARKINSON'S BASED ON MEDICAL  
DATA USING DATA MINING**

**Autores:** <sup>1</sup>Mercedes Carolina García Caiza y <sup>2</sup>José Renato Cumbal Simba.

<sup>1</sup>ORCID ID: <https://orcid.org/0000-0003-4512-6371>

<sup>2</sup>ORCID ID: <https://orcid.org/0000-0001-8182-5343>

<sup>1</sup>E-mail de contacto: [mcgarcia34@utpl.edu.ec](mailto:mcgarcia34@utpl.edu.ec)

<sup>2</sup>E-mail de contacto: [rcumbal@ups.edu.ec](mailto:rcumbal@ups.edu.ec)

Afiliación: <sup>1</sup>\*<sup>2</sup>Universidad Politécnica Salesiana, (Ecuador).

Artículo recibido: 15 de Enero del 2026

Artículo revisado: 28 de Enero del 2026

Artículo aprobado: 06 de Febrero del 2026

<sup>1</sup>Estudiante de la carrera de Ingeniería en Sistemas Computacionales egresada de la Universidad Politécnica Salesiana, (Ecuador).

<sup>2</sup>Ingeniero en Electrónica y Telecomunicaciones, Magíster en Gerencia de Sistemas de Información egresado de la Universidad Politécnica Salesiana, (Ecuador). Profesor de la Universidad Politécnica Salesiana, Quito, Ecuador, y miembro del Grupo de Investigación en Telecomunicaciones (GIETEC). Doctorante en Ingeniería, Universidad Pontificia Bolivariana, Medellín, (Colombia).

### Resumen

Las enfermedades neurodegenerativas como Alzheimer y Parkinson representan un desafío en la salud global debido a la complejidad del diagnóstico temprano. Este estudio aplica minería de datos y machine learning para crear modelos predictivos que permitan la diferenciación diagnóstica entre ambas patologías. Utilizando datasets públicos (ADNI para Alzheimer y PPMI para Parkinson), se integraron y analizaron variables clínicas, demográficas, cognitivas y funcionales mediante cinco algoritmos: Decision Trees (DT), Random Forest (RF), Gradient Boosting Machine (GBM), Support Vector Machines (SVM) y Artificial Neural Networks (ANN). El pipeline metodológico incluyó preprocesamiento con imputación y estandarización, selección de características y validación cruzada estratificada. Los resultados muestran que para Parkinson (PPMI), el modelo Gradient Boosting Machine alcanzó una precisión del 96.52%, F1-macro de 0.936 y AUC de 0.995. Para Alzheimer (ADNI), el mismo algoritmo obtuvo una precisión del 90.29%, F1-macro de 0.900 y AUC de 0.987. Se concluye que la integración de datos multimodales con técnicas de aprendizaje automático permite construir herramientas de apoyo clínico no invasivas, objetivas y escalables para mejorar el diagnóstico temprano y la estratificación de pacientes.

**Palabras clave:** Alzheimer, Parkinson, Aprendizaje automático, Minería de datos, Predicción, Enfermedades neurodegenerativas, Diagnóstico diferencial, Machine learning.

### Abstract

Neurodegenerative diseases like Alzheimer's and Parkinson's represent a global health challenge due to early diagnosis complexity. This study applies data mining and machine learning to create predictive models for diagnostic differentiation between both pathologies. Using public datasets (ADNI for Alzheimer's and PPMI for Parkinson's), clinical, demographic, cognitive and functional variables were integrated and analyzed through five algorithms: Decision Trees (DT), Random Forest (RF), Gradient Boosting Machine (GBM), Support Vector Machines (SVM) and Artificial Neural Networks (ANN). The methodological pipeline included preprocessing with imputation and standardization, feature selection and stratified cross-validation. Results show that for Parkinson's (PPMI), the Gradient Boosting Machine model achieved 96.52% accuracy, 0.936 F1-macro and 0.995 AUC. For Alzheimer's (ADNI), the same algorithm obtained 90.29% accuracy, 0.900 F1-macro and 0.987 AUC. The study concludes that multimodal data integration with machine learning techniques enables building non-

invasive, objective and scalable clinical support tools to improve early diagnosis and patient stratification.

**Keywords: Alzheimer, Parkinson, Machine learning, Data mining, Prediction, Neurodegenerative disorders, Differential diagnosis, Artificial intelligence.**

### **Sumário**

As doenças neurodegenerativas como Alzheimer e Parkinson representam um desafio na saúde global devido à complexidade do diagnóstico precoce. Este estudo aplica mineração de dados e machine learning para criar modelos preditivos que permitam a diferenciação diagnóstica entre ambas as patologias. Utilizando datasets públicos (ADNI para Alzheimer e PPMI para Parkinson), foram integradas e analisadas variáveis clínicas, demográficas, cognitivas e funcionais mediante cinco algoritmos: Decision Trees (DT), Random Forest (RF), Gradient Boosting Machine (GBM), Support Vector Machines (SVM) e Artificial Neural Networks (ANN). O pipeline metodológico incluiu pré-processamento com imputação e padronização, seleção de características e validação cruzada estratificada. Os resultados mostram que para Parkinson (PPMI), o modelo Gradient Boosting Machine alcançou uma precisão de 96.52%, F1-macro de 0.936 e AUC de 0.995. Para Alzheimer (ADNI), o mesmo algoritmo obteve uma precisão de 90.29%, F1-macro de 0.900 e AUC de 0.987. Conclui-se que a integração de dados multimodais com técnicas de aprendizagem automática permite construir ferramentas de apoio clínico não invasivas, objetivas e escaláveis para melhorar o diagnóstico precoce e a estratificação de pacientes.

**Palavras-chave: Alzheimer, Parkinson, Aprendizagem automática, Mineração de dados, Predição, doenças Neurodegenerativas, Diagnóstico diferencial, Inteligência artificial.**

### **Introducción**

Las enfermedades neurodegenerativas se han convertido en un reto prioritario para la salud

pública y la investigación biomédica, impulsado por su aumento sostenido, su curso clínico silencioso y el impacto irreversible que producen en la vida de los pacientes (World Health Organization, 2023; Kavitha et al., 2022). Entre estas patologías destacan las enfermedades de Alzheimer (EA) y Parkinson (EP) por su alta prevalencia, su asociación con el envejecimiento y el deterioro funcional progresivo (World Health Organization, 2023; Kavitha et al., 2022; Marek et al., 2018). Se estima que a nivel mundial, más de 55 millones de individuos viven con demencia, siendo la EA la principal causa, y se proyecta que esta cifra podría superar los 139 millones para el año 2050 si no se desarrollan intervenciones eficaces (World Health Organization, 2023). La EP afecta a aproximadamente 10 millones de individuos, constituyéndose como la segunda enfermedad neurodegenerativa más frecuente (Marek et al., 2018). La Organización Mundial de la Salud ha resaltado que los gastos asociados al cuidado de personas con demencia superan actualmente los 1.3 billones de dólares anuales (World Health Organization, 2023).

Los diagnósticos a menudo ocurren en etapas avanzadas cuando las intervenciones tienen efectividad limitada, ya que en las etapas iniciales hay pocas manifestaciones clínicas anormales claramente relacionadas con el inicio de la enfermedad (Kavitha et al., 2022; Marek et al., 2018). La heterogeneidad en la progresión y el número limitado de biomarcadores accesibles obstaculizan la detección temprana (Tanveer et al., 2020; Zhang et al., 2022; Sharma et al., 2023). Diversos estudios han evidenciado que las enfermedades neurodegenerativas presentan patrones fisiopatológicos complejos que dificultan su identificación en fases iniciales. En el caso del Alzheimer, investigaciones recientes muestran que procesos inflamatorios, alteraciones

inmunológicas y mecanismos neuropatológicos pueden influir en la progresión de la enfermedad (Sochocka et al., 2017). En Parkinson, estudios basados en análisis multi-ómicos han demostrado que la integración de biomarcadores sanguíneos permite identificar genes marcadores con alto potencial diagnóstico (Zhang et al., 2022).

La incorporación de técnicas computacionales ha abierto nuevas posibilidades para la predicción de progresión neurodegenerativa. Modelos basados en minería de datos y series temporales han permitido anticipar cambios clínicos relevantes mediante la integración de factores longitudinales y patrones complejos (Tanveer et al., 2020). Revisiones recientes señalan una tendencia sostenida hacia el uso de enfoques multimodales y aprendizaje profundo para mejorar el rendimiento predictivo (Kavitha et al., 2022; Marek et al., 2018; Myszczyńska et al., 2020). Se han propuesto modelos que integran múltiples fuentes de información e incorporan componentes de interpretabilidad (IA explicable) para facilitar su adopción en entornos clínicos (Sarica et al., 2017). En la EA, la literatura reporta arquitecturas profundas orientadas a clasificación y predicción, así como enfoques para pronóstico de deterioro cognitivo en horizontes longitudinales (Rathore et al., 2017; Liu et al., 2018; Grassi et al., 2018; Lin et al., 2018). En Parkinson, se han explorado estrategias basadas en características funcionales como la marcha y aproximaciones emergentes para el diagnóstico/predicción (Chen et al., 2016; Suarez-Revelo et al., 2021). Se han reportado esquemas basados en optimización y aprendizaje supervisado aplicados al análisis predictivo de trastornos neurodegenerativos (Sharma et al., 2023).

La literatura científica evidencia la necesidad de desarrollar modelos predictivos robustos que

integren datos clínicos, funcionales y biomarcadores para diferenciar de forma precisa el Alzheimer del Parkinson (Kavitha et al., 2022; Liu et al., 2018). Las comparaciones entre la EA y la EP se ven limitadas porque los estudios se centran en una sola enfermedad, dificultando encontrar discrepancias que podrían mejorar el diagnóstico temprano (Choi et al., 2017). Existe una brecha en el uso de datos integrados de múltiples niveles (clínicos, cognitivos, biomarcadores y funcionales) en el diagnóstico diferencial. Este estudio es necesario porque la identificación temprana y la diferenciación precisa entre EA y EP siguen siendo un desafío cuando los síntomas iniciales son sutiles y existe variabilidad clínica entre pacientes. La evaluación comparativa usando metodologías de minería de datos y aprendizaje automático permite avanzar hacia un enfoque de medicina de precisión (Kavitha et al., 2022; Marek et al., 2018).

El desarrollo de modelos predictivos no invasivos basados en datos clínicos, cognitivos, funcionales y biomarcadores puede complementar la evaluación médica tradicional, ofreciendo resultados más objetivos, reproducibles y sensibles a cambios tempranos (Tanveer et al., 2020; Sarica et al., 2017). El uso de datasets longitudinales consolidados como ADNI y PPMI posibilita entrenar y contrastar modelos de forma sistemática, fortaleciendo la reproducibilidad (Jack et al., 2008; Marek et al., 2011). El presente estudio propone un modelo predictivo basado en minería de datos para identificar y comparar patrones clínicos de EA y EP mediante la aplicación de algoritmos supervisados. Se busca evaluar la capacidad de estos modelos para discriminar entre ambas patologías utilizando datos clínicos, cognitivos, funcionales y demográficos, así como determinar qué variables aportan mayor relevancia predictiva.

### **Materiales y Métodos**

La metodología general se justificó dentro de un enfoque cuantitativo, con alcance comparativo y predictivo, y diseño observacional retrospectivo. No se justificó la recolección de nuevos datos puesto que toda la medición se obtuvo de repositorios y fue procesada mediante análisis *in silico*. Se utilizaron dos repositorios biomédicos longitudinales: la Iniciativa de Neuroimagen de la Enfermedad de Alzheimer (ADNI) (Jack et al., 2008) y la Iniciativa de Marcadores de Progresión de Parkinson (PPMI) (Marek et al., 2011). Ambos repositorios proporcionan datos anonimizados con protocolos estandarizados para evaluación clínica y pruebas neuropsicológicas. El acceso se realizó a través de los portales oficiales, en cumplimiento con sus políticas de uso académico (Jack et al., 2008; Marek et al., 2011). Como el análisis empleó datos secundarios anonimizados, evitó cualquier contacto directo con los participantes.

La conformación de la muestra siguió un muestreo no probabilístico de conveniencia, condicionado por los registros accesibles en las bases de datos. Para eliminar redundancias por evaluaciones longitudinales y preservar la equivalencia entre participantes, se limitó el trabajo a datos iniciales (baseline), manteniendo un solo registro por sujeto. En ADNI se tomaron instancias con `VISCODE = bl` y se agrupó por `RID`; en PPMI, se seleccionaron aquellas con `EVENT_ID = BL` y se agrupó por `PATNO`. Se requirió: (i) la existencia de un registro inicial; (ii) la presencia de la variable de interés (ADNI: `DX_bl`; PPMI: `COHORT`); y (iii) la disponibilidad de variables predictivas, admitiendo valores faltantes imputables. Se excluyeron registros con ausencia de etiqueta diagnóstica, inconsistencias de identificación, duplicidad de registros basales y variables incompatibles.

Para ADNI se utilizaron 12 variables: edad (`AGE`), género (`PTGENDER`), años de educación (`PTEDUCAT`), Mini-Mental State Examination (`MMSE`), Montreal Cognitive Assessment (`MOCA`), Clinical Dementia Rating Sum of Boxes (`CDRSB`), ADAS-Cog 13 (`ADAS13`), Functional Activities Questionnaire (`FAQ`), memoria diferida (`LDELTOTAL`),  $\beta$ -amiloide (`ABETA`), Tau total (`TAU`), y diagnóstico basal (`DX_bl`) como variable de salida con categorías: Normal Cognitivo (`CN`), Deterioro Cognitivo Leve (`MCI`) y Alzheimer (`AD`). Para PPMI se utilizaron 12 variables: edad (`AGE`), sexo (`SEX`), años de educación (`EDUC_YRS`), `MOCA`, Symbol Digit Modalities Test (`SDMTOTAL`), HVLT recuerdo inmediato (`HVLT_IMMEDIATE`), prueba verbal (`VLTANIM`), MDS-UPDRS Parte III (`UPDRS3_SCORE`), escala Hoehn y Yahr (`NHY`), Geriatric Depression Scale (`GDS`), alelos `APOE ε4` (`APOE_E4`), uricemia (`URATE`), y cohorte diagnóstica (`COHORT`) como variable de salida con categorías: Participante PD, Control Sano, SWEDD y Prodromal.

Se implementó un procedimiento automático para estandarizar nombres de columnas, eliminando espacios y aplicando normalización básica. Se realizó un filtrado inicial conservando únicamente variables predictoras y la variable objetivo, eliminando registros con etiqueta ausente. La etiqueta objetivo se transformó mediante Label Encoding para habilitar la clasificación multiclase. Para el manejo de valores faltantes y normalización, se construyó un preprocesador con `ColumnTransformer`: en variables numéricas se aplicó imputación por mediana seguida de estandarización con `StandardScaler`, mientras que en variables categóricas se utilizó imputación por moda y codificación one-hot.

Todo el preprocesamiento se integró en un pipeline para prevenir fugas de información (data leakage). Para cada repositorio se efectuó una partición train/test con proporción 80/20, de manera estratificada respecto a la variable objetivo para preservar la distribución de clases. Se fijó semilla (random\_state=42) para asegurar reproducibilidad.

Se realizó un análisis comparativo de cinco enfoques de clasificación supervisada: Decision Trees (DT), Random Forest (RF), Gradient Boosting Machine (GBM), Support Vector Machines (SVM) y Artificial Neural Networks (ANN)[19]. Todos los modelos fueron implementados mediante pipelines que integran preprocesamiento, manejo del desbalance de clases y ajuste de hiperparámetros. Formulación Matemática: Sea  $x_i \in \mathbb{R}^d$  el vector de variables predictoras de un individuo  $i$ , donde  $d$  corresponde al número de características ( $d = 12$  en ambos datasets), y sea  $y_i \in \{1, \dots, K\}$  la etiqueta diagnóstica. El objetivo consiste en aprender una función de clasificación:

$$f: \mathbb{R}^d \rightarrow \{1, \dots, K\}$$

Decision Trees (DT): Dividen el espacio de características mediante reglas de partición binarias. Cada nodo interno aplica una condición sobre una variable predictora utilizando la impureza de Gini como criterio:

$$G(t) = 1 - \sum_{k=1}^K p_k(t)^2$$

donde  $G(t)$  es la impureza del nodo  $t$  y  $p_k(t)$  es la proporción de muestras de clase  $k$  en el nodo  $t$ .

Random Forest (RF): Construye un ensemble de árboles de decisión entrenados sobre muestras bootstrap y subconjuntos aleatorios de características, agregando sus predicciones mediante votación por mayoría.

Gradient Boosting Machine (GBM): Construye un ensemble aditivo de modelos débiles de manera secuencial, donde cada nuevo modelo corrige los errores del ensemble previo:

$$F_M(x) = \sum_{m=1}^M v f_m(x)$$

donde  $v$  es la tasa de aprendizaje que controla la contribución de cada árbol. Support Vector Machines (SVM): Busca el hiperplano óptimo que maximiza el margen de separación entre clases en un espacio de características de alta dimensión. Artificial Neural Networks (ANN): Perceptrón multicapa con capas ocultas completamente conectadas que aplican transformaciones afines seguidas de función de activación ReLU:

$$h^{(l)} = \phi(W^{(l)} h^{(l-1)} + b^{(l)})$$

La capa de salida utiliza función softmax para estimar probabilidades de pertenencia a cada clase. Se ejecutó RandomizedSearchCV definiendo espacios de búsqueda específicos para cada modelo, con validación cruzada estratificada ( $k=5$ ) y métrica de optimización F1-macro. El modelo con mejor desempeño en validación cruzada se seleccionó para evaluación final en el conjunto de prueba independiente.

La eficacia de los modelos se evaluó mediante:

- Accuracy: Proporción de predicciones correctas
- Balanced Accuracy: Promedio de sensibilidades por clase
- F1-macro: Media armónica entre precisión y recall promediada por clase
- AUC-ROC OvR: Área bajo la curva ROC en esquema one-vs-rest para clasificación multiclase



Adicionalmente, se calculó la importancia de variables mediante `feature_importances_` para modelos basados en árboles, y mediante permutación para SVM y ANN.

## Resultados y Discusión

### Resultados para Parkinson (PPMI)

Los cinco modelos fueron entrenados y evaluados sobre el dataset PPMI. La Tabla 1 presenta los resultados de desempeño en el conjunto de prueba:

**Tabla 1.** Resultados de los modelos para el dataset PPMI (Parkinson)

Modelo	Accuracy	Balanced Accuracy	F1-macro	AUC-OvR
Decision Tree	91.30%	0.837	0.846	0.957
Random Forest	95.65%	0.899	0.927	0.993
Gradient Boosting	96.52%	0.906	0.936	0.995
SVM	93.91%	0.871	0.889	0.982
ANN	94.78%	0.885	0.908	0.988

Fuente: Elaboración propia

El modelo Gradient Boosting Machine demostró el mejor rendimiento global, alcanzando una precisión del 96.52%, un F1-macro de 0.936, una precisión equilibrada de 0.906 y un AUC de 0.995. Random Forest obtuvo el segundo mejor desempeño con 95.65% de precisión. Decision Tree mostró el rendimiento más bajo, lo cual es consistente con su tendencia al sobreajuste cuando no se limita su profundidad. El análisis de importancia de variables para GBM reveló que las características motoras (UPDRS3\_SCORE, NHY) y cognitivas (MOCA, SDMTOTAL) fueron los predictores más relevantes, seguidos de biomarcadores como URATE y factores demográficos como edad. A continuación se presentan los resultados para Alzheimer.

### Resultados para Alzheimer (ADNI)

Los cinco modelos fueron entrenados y evaluados sobre el dataset ADNI. La Tabla 2 presenta los resultados:

**Tabla 2.** Resultados de los modelos para el dataset ADNI (Alzheimer)

Modelo	Accuracy	Balanced Accuracy	F1-macro	AUC-OvR
Decision Tree	84.47%	0.823	0.831	0.941
Random Forest	88.35%	0.894	0.887	0.981
Gradient Boosting	90.29%	0.911	0.900	0.987
SVM	87.38%	0.878	0.875	0.972
ANN	86.89%	0.867	0.869	0.976

Fuente: Elaboración propia

Gradient Boosting Machine nuevamente demostró el mejor rendimiento, con una precisión del 90.29%, F1-macro de 0.900, precisión equilibrada de 0.911 y AUC de 0.987. Random Forest obtuvo el segundo lugar con 88.35% de precisión. El análisis de importancia reveló que las variables cognitivas (MMSE, MOCA, CDRSB, ADAS13) y de memoria (LDELTOTAL) fueron los predictores más relevantes, seguidos de biomarcadores (TAU, ABETA) y el cuestionario funcional (FAQ). Las variables demográficas mostraron menor importancia relativa.

Los resultados obtenidos demuestran que los modelos de ensamble, particularmente Gradient Boosting Machine, superan consistentemente a modelos individuales en ambos datasets. Este hallazgo es coherente con la literatura reciente que reporta la superioridad de métodos de ensamble en problemas de clasificación médica (Bhattacharya et al., 2021; Breiman, 2001; Chen & Guestrin, 2016). La mayor precisión observada en el dataset PPMI (96.52%) comparada con ADNI (90.29%) puede atribuirse a varios factores: (i) la EP presenta manifestaciones motoras más objetivamente

medibles (UPDRS3\_SCORE, NHY) que facilitan la discriminación diagnóstica; (ii) la EA muestra mayor heterogeneidad clínica, especialmente en la categoría MCI que puede incluir subtipos variados; (iii) el solapamiento sintomático entre CN y MCI temprano es más sutil que entre controles y PD.

La importancia de variables motoras en Parkinson coincide con estudios previos que identifican UPDRS como predictor robusto (Chen et al., 2016; Suarez-Revelo et al., 2021). Para Alzheimer, la relevancia de variables cognitivas (MMSE, MOCA) y de memoria valida el rol central del deterioro cognitivo en la progresión de la EA (Rathore et al., 2017; Liu et al., 2018; Grassi et al., 2018; Lin et al., 2018). Los biomarcadores (TAU, ABETA) mostraron importancia significativa pero no dominante, sugiriendo que la integración multimodal aporta mayor valor que biomarcadores aislados (Zhang et al., 2022; Sochocka et al., 2017). Los valores de AUC superiores a 0.98 en ambos datasets indican excelente capacidad discriminativa de los modelos, comparable con estudios recientes que utilizan aprendizaje profundo (Kavitha et al., 2022; Marek et al., 2018; Myszczyńska et al., 2020). Sin embargo, los modelos propuestos ofrecen ventajas de interpretabilidad mediante análisis de importancia de variables, facilitando la comprensión clínica de las decisiones predictivas (Sarica et al., 2017). Las métricas de balanced accuracy y F1-macro superiores a 0.90 demuestran robustez ante desbalance de clases, aspecto crítico en datos clínicos donde las categorías diagnósticas pueden tener prevalencias dispares.

### **Conclusiones**

Este estudio desarrolló y evaluó modelos predictivos basados en minería de datos para el diagnóstico diferencial de Alzheimer y Parkinson, demostrando que la integración de

datos clínicos, cognitivos, funcionales y biomarcadores mediante algoritmos de machine learning permite alcanzar alta precisión diagnóstica. Las principales conclusiones son:

- Gradient Boosting Machine demostró el mejor rendimiento en ambos datasets, alcanzando 96.52% de precisión para Parkinson (PPMI) y 90.29% para Alzheimer (ADNI), superando consistentemente a otros algoritmos evaluados.
- Los modelos de ensamble (Random Forest, Gradient Boosting) superaron a modelos individuales (Decision Tree, SVM, ANN), confirmando el valor de combinar múltiples predictores débiles para problemas complejos de clasificación médica.
- Las variables más relevantes difieren entre patologías: en Parkinson dominan características motoras (UPDRS3\_SCORE, NHY) y cognitivas (MOCA), mientras que en Alzheimer destacan pruebas cognitivas (MMSE, MOCA, CDRSB) y de memoria (LDELTOTAL).
- La integración multimodal de datos demográficos, clínicos, cognitivos, funcionales y biomarcadores aporta mayor valor predictivo que el uso de variables aisladas, validando el enfoque de medicina de precisión.
- Los modelos desarrollados pueden servir como herramientas de apoyo clínico no invasivas, objetivas y escalables para mejorar el diagnóstico temprano, la estratificación de pacientes y el seguimiento longitudinal en entornos asistenciales.
- La metodología propuesta es reproducible y aplicable a otros trastornos neurodegenerativos, permitiendo extender

el enfoque a nuevas patologías conforme se dispongan de datos estructurados.

### **Agradecimientos**

Los autores agradecen a las iniciativas ADNI y PPMI por proporcionar acceso a sus bases de datos longitudinales, así como a todas las instituciones participantes y pacientes que contribuyeron a estos repositorios. Este trabajo fue posible gracias al apoyo de [institución financiadora] bajo el proyecto [número].

### **Referencias Bibliográficas**

- Bhattacharya, S., Salil, P., Ranjan, P., y Goel, S. (2021). Deep learning and medical image processing for early detection of Alzheimer's disease. *Neural Computing and Applications*, 33(21), 14249–14262.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Chen, H., Wang, G., Ma, C., Cai, Z., Liu, W., y Wang, S. (2016). An efficient hybrid kernel extreme learning machine approach for early diagnosis of Parkinson's disease. *Neurocomputing*, 184, 131–144.
- Chen, T., y Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794).
- Choi, H., Ha, S., Im, H., Paek, S., y Lee, D. (2017). Refining diagnosis of Parkinson's disease with deep learning-based interpretation of dopamine transporter imaging. *NeuroImage: Clinical*, 16, 586–594.
- Grassi, M., Perna, G., Caldirola, D., Schruers, K., Duara, R., y Loewenstein, D. (2018). A clinically-translatable machine learning algorithm for the prediction of Alzheimer's disease conversion in individuals with mild and premild cognitive impairment. *Journal of Alzheimer's Disease*, 61(4), 1555–1573.
- Jack, C., Bernstein, M., Fox, N., Thompson, P., Alexander, G., y Harvey, D. (2008). The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging*, 27(4), 685–691.
- Kavitha, C., Mani, V., Srividhya, S., Khalaf, O., y Tavera, C. (2022). Early-stage Alzheimer's disease prediction using machine learning models. *Frontiers in Public Health*, 10, 853294.
- Lin, W., Tong, T., Gao, Q., Guo, D., Du, X., y Yang, Y. (2018). Convolutional neural networks-based MRI image analysis for the Alzheimer's disease prediction from mild cognitive impairment. *Frontiers in Neuroscience*, 12, 777.
- Liu, M., Zhang, J., Adeli, E., y Shen, D. (2018). Landmark-based deep multi-instance learning for brain disease diagnosis. *Medical Image Analysis*, 43, 157–168.
- Marek, K., Chowdhury, S., Siderowf, A., Lasch, S., y Coffey, C. (2018). The Parkinson's progression markers initiative (PPMI): Establishing a PD biomarker cohort. *Annals of Clinical and Translational Neurology*, 5(12), 1460–1477.
- Marek, K., Jennings, D., Lasch, S., Siderowf, A., Tanner, C., y Simuni, T. (2011). The Parkinson progression marker initiative (PPMI). *Progress in Neurobiology*, 95(4), 629–635.
- Myszczyńska, M., Ojamies, P., Lacoste, A., Neil, D., Saffari, A., y Mead, R. (2020). Applications of machine learning to diagnosis and treatment of neurodegenerative diseases. *Nature Reviews Neurology*, 16(8), 440–456.
- Rathore, S., Habes, M., Iftikhar, M., Shacklett, A., y Davatzikos, C. (2017). A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages. *NeuroImage*, 155, 530–548.
- Sarica, A., Cerasa, A., y Quattrone, A. (2017). Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: A systematic review. *Frontiers in Aging Neuroscience*, 9, 329.
- Sharma, N., Kolekar, M., y Jha, K. (2023). Alzheimer's disease prediction using machine learning and deep learning: A comprehensive review. *Electronics*, 12(7), 1723.



Sochocka, M., Zwolińska, K., y Leszek, J. (2017). The infectious etiology of Alzheimer's disease. *Current Neuropharmacology*, 15(7), 996–1009.

Suarez-Revelo, J., Ochoa-Gomez, J., y Tobón, C. (2021). Quantum machine learning for early Parkinson's disease detection. *Applied Sciences*, 11(19), 9345.

Tanveer, M., Richhariya, B., Khan, R., Rashid, A., Khanna, P., Prasad, M., y Lin, C. (2020). Machine learning techniques for the diagnosis of Alzheimer's disease: A review. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 16(1s), 1–35.

World Health Organization. (2023). Dementia: Key facts. WHO Global Report on Dementia. WHO Press.

Zhang, L., Wang, Y., Liu, Y., Wang, T., Zhang, Q., y Shi, H. (2022). Multi-omics integration analysis identifies novel genes for Parkinson's disease. *Frontiers in Genetics*, 13, 865208.



Esta obra está bajo una licencia de **Creative Commons Reconocimiento-No Comercial 4.0 Internacional**. Copyright © Mercedes Carolina García Caiza y José Renato Cumbal Simba.

