

Una evaluación comparativa de modelos interpretables para el diagnóstico de Alzheimer a partir de datos de EEG

Clarivette Teresa Gopar-Nazariega, Verónica Rodríguez-López

Universidad Tecnológica de la Mixteca,
Heroica Ciudad de Huajuapán de León, Oaxaca,
México

clarit.gnaza@gs.utm.mx, veromix@gs.utm.mx

Resumen. La enfermedad de Alzheimer (A) es un trastorno neurológico progresivo y degenerativo que provoca pérdida de memoria y un deterioro cognitivo. Su detección temprana es importante, dado que permite ralentizar la progresión de los síntomas con un tratamiento oportuno. En este contexto, el presente trabajo de investigación desarrolla un análisis comparativo de modelos de aprendizaje automático (bosque aleatorio, árboles de decisión, regresión logística y Bayes ingenuo) para automatizar su diagnóstico. Se analiza el desempeño de los modelos de diagnóstico al utilizar diferentes grupos de características espectrales obtenidas de estudios de electroencefalograma y seleccionadas por el método Mínima Redundancia Máxima Relevancia. Los resultados experimentales muestran que los clasificadores Bayes Ingenuo y regresión logística obtienen la configuración más robusta y equilibrada para una mejor interpretabilidad.

Palabras clave: Enfermedad de Alzheimer (A), electroencefalograma (EEG), modelos interpretables, análisis espectral.

A Comparative Evaluation of Interpretable Models for the Diagnosis of Alzheimer's Disease based on EEG Data

Abstract. Alzheimer's disease (AD) is a progressive and degenerative neurological disorder that causes memory loss and cognitive decline. Early detection is important, as it allows for slowing the progression of symptoms with timely treatment. In this context, this research presents a comparative analysis of machine learning models (random forest, decision trees, logistic regression, and naive Bayes) to automate its diagnosis. The performance of the diagnostic models is analyzed using different sets of spectral features obtained from electroencephalogram studies and selected using the Minimum Redundancy Maximum Relevance (MRR) method. The experimental results show that the naive Bayes and logistic

regression classifiers achieve the most robust and balanced configuration for better interpretability.

Keywords: Alzheimer's disease (A), electroencephalogram (EEG), interpretable models, spectral analysis.

1. Introducción

La enfermedad de Alzheimer (A) es un padecimiento neurodegenerativo progresivo que deteriora gravemente la memoria, las funciones cognitivas y el comportamiento. A nivel mundial, es considerada como una de las causas principales de demencia [8], y muerte, impactando en mayor proporción a la población femenina [12]. Debido a que los cambios iniciales en el cerebro son sutiles, los métodos clínicos tradicionales suelen ser insuficientes, lo que ha llevado al uso de la electroencefalografía (EEG), combinada con inteligencia artificial, como una alternativa objetiva y de bajo costo para diagnosticar la enfermedad de Alzheimer.

Esta investigación analiza el desempeño de diversos modelos para el diagnóstico del Alzheimer. En particular, se analiza el desempeño de clasificadores Bayes ingenuo, árboles de decisión, regresión logística y bosques aleatorios al utilizar diversos grupos de características espectrales obtenidas de estudios de electroencefalogramas. Para reducir la redundancia de las características y potenciar la interpretabilidad del sistema; es decir, asegurar que un experto pueda descifrar el camino lógico seguido por el algoritmo para transformar las entradas en una conclusión determinada [6], se explora la aplicación del método de Mínima Redundancia Máxima Relevancia en combinación con el Coeficiente de Fisher y de Información Mutua. Los resultados experimentales indican que los clasificadores Bayes ingenuo y regresión logística obtienen el mejor desempeño.

El resto del artículo está estructurado de la siguiente forma. En la Sección 2 se presenta la metodología aplicada en esta investigación. Los experimentos y resultados se describen en la Sección 3. Finalmente, se presentan las conclusiones en la Sección 4.

2. Metodología

La metodología aplicada en esta investigación se resume en la Figura 1 e incluye el conjunto de datos, la extracción y selección de características, finalizando con el aprendizaje y evaluación de modelos.

2.1. Datos de EEG

Para el desarrollo de este trabajo se utilizó el conjunto *A dataset of EEG recordings from: Alzheimer's disease, Frontotemporal dementia and Healthy subjects* del repositorio OpenNEURO obtenidos por el segundo departamento de Neurología del Hospital General AHEPA de Tesalónica [7].



Fig. 1. Flujo de trabajo para la evaluación comparativa de modelos interpretables en el diagnóstico de Alzheimer.

El conjunto de datos incluye estudios de electroencefalografía (EEG) de 88 sujetos, registrados en estado de reposo y con los ojos cerrados, y distribuidos en tres categorías: 36 pacientes con enfermedad de Alzheimer (A), 23 con demencia frontotemporal (FTD) y 29 sujetos sanos como grupo de control (C). El etiquetado y la caracterización clínica de los participantes se realizaron mediante el examen Mini-Mental (MMSE), donde se registraron puntuaciones promedio de 17.75 para el grupo A, 22.17 para el FTD y 30 para el grupo C.

La adquisición de las señales se realizó mediante un equipo clínico Nihon Kohden EEG 2100, con una frecuencia de muestreo de 500 Hz y resolución de 10 $\mu\text{V}/\text{mm}$. La recolección siguió el Sistema Internacional 10-20, utilizando 19 electrodos de superficie y dos de referencia colocados en las mastoides.

Esta investigación trabajó con el conjunto denominado *derivates*, el cual se obtuvo al preprocesar las señales de EEG, primero con un filtro de Butterworth de paso de banda (0.5–45 Hz), seguido de la aplicación de la Reconstrucción del Subespacio de Artefactos (ASR, por sus siglas en inglés: Artifact Subspace Reconstruction) para la eliminación de segmentos ruidosos y finalizando con la supresión de interferencias oculares y mandibulares mediante el Análisis de Componentes Independientes (ICA, por sus siglas en inglés: Independent Component Analysis) basado en el algoritmo RunICA y la herramienta ICLabel.

Para este trabajo, únicamente se utilizaron los estudios que se describen en la Tabla 1 y que corresponden a los sujetos de control y a los sujetos con Alzheimer.

Tabla 1. Detalles del subconjunto de datos utilizado.

Clase	Sujetos	Duración (min)	Edad Promedio
A: Alzheimer	36	485.5	66.4
C: Control	29	402	67.9
Total	65	887.5	67.15

Tabla 2. Descripción de características del dominio espectral utilizadas en esta investigación.

Tipo	Característica
Potencia de bandas	Delta (0.5 - 4 Hz)
	Theta (4 - 8 Hz)
	Alpha (8 - 13 Hz)
	Beta (13 - 25 Hz)
	Gamma (25 - 45 Hz)
Ratio	$TAR \left(\frac{Theta}{Alpha} \right)$
	$DTABR \left(\frac{Delta+Theta}{Alpha+Beta} \right)$

2.2. Extracción de características

En esta sección se emplearon las 133 características espectrales descritas en la Tabla 2 y que de acuerdo a la literatura son adecuadas para el diagnóstico del Alzheimer [9,13].

La extracción de estas características se realizó para cada canal de los estudios de EEG de manera segmentada, utilizando ventanas de 4 segundos con un solapamiento del 50 % [15], siguiendo el proceso descrito en la Figura 2. En primer lugar, se aplicó la transformada de Welch para convertir la señal del dominio temporal al dominio espectral.

Posteriormente, se calcularon las potencias absolutas en las distintas bandas de frecuencia, así como los ratios derivados del cociente entre dichas potencias.

Finalmente, a los valores obtenidos se les aplicó una transformación logarítmica ($10 \times \log_{10}$), con el objetivo de ampliar su rango dinámico y conformar el vector de características correspondiente a cada época analizada.

2.3. Selección de características

En esta etapa, se combinaron técnicas de selección de características para generar vectores de características optimizados que permitieran mejorar el desempeño del sistema de diagnóstico. Para construir dichos vectores, se identificó el subconjunto de variables más apropiado a partir de la tabla de características extraídas, buscando eliminar información irrelevante o redundante que pudiera dificultar la convergencia y aumentar el riesgo de sobreajuste [9]. Estos vectores fueron procesados mediante métodos de filtro con el fin de reducir la complejidad del sistema propuesto, optimizar la eficiencia algorítmica y simplificar el proceso de entrenamiento de los clasificadores [4].

Se utilizaron métodos de filtro como Coeficiente de Fisher, Información Mutua (IM) y Mínima Redundancia Máxima Relevancia (mRMR), los cuales fueron seleccionados para evaluar individualmente la capacidad discriminativa de cada

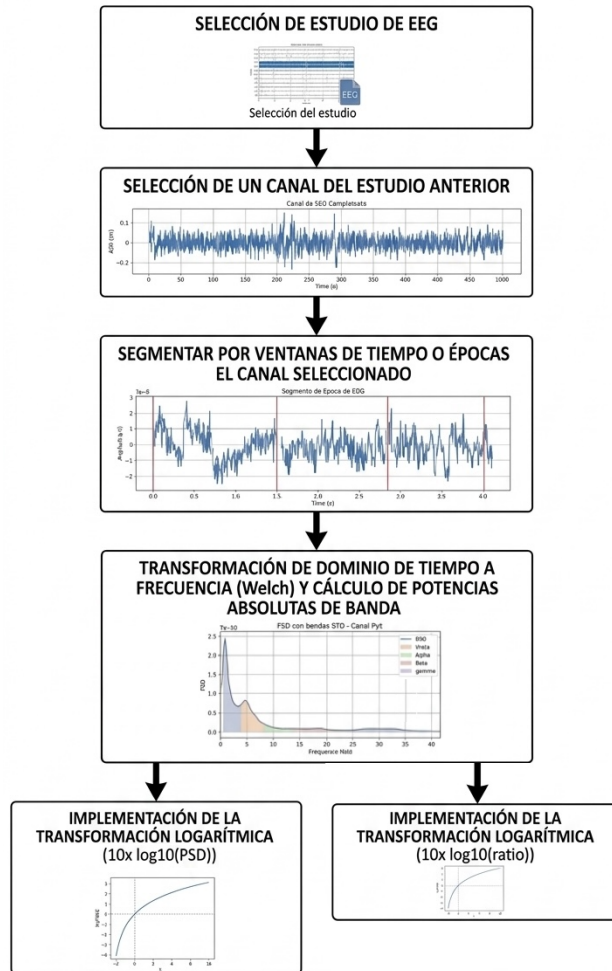


Fig. 2. Procedimiento aplicado para la extracción de las características espectrales de los estudios de EEG.

característica y determinar el mejor subconjunto de forma independiente al modelo de aprendizaje; de estos, el Coeficiente de Fisher se fundamenta en criterios estadísticos de media y varianza [3], mientras que la IM y el mRMR aplican principios de la teoría de la información.

2.4. Aprendizaje de modelos interpretables

En este estudio se evalúan cuatro tipos de clasificadores: árbol de decisión, bosques aleatorios, regresión logística y Bayes ingenuo. La elección de estos

modelos responde a su capacidad para equilibrar precisión, solidez estadística e interpretabilidad, factores esenciales en el ámbito del diagnóstico clínico.

En particular, el modelo de árbol de decisión destaca por su alta interpretabilidad, permitiendo representar de forma explícita la lógica de clasificación mediante estructuras jerárquicas basadas en reglas, lo que facilita la comprensión del proceso de decisión y resulta especialmente valioso en aplicaciones clínicas [5].

Por otro lado, bosques aleatorios proporciona una mayor robustez al disminuir el sobreajuste mediante su arquitectura de ensamble de múltiples árboles de decisión, permitiendo gestionar de manera eficiente la varianza y el ruido de las señales bioeléctricas [9].

Complementariamente, Bayes ingenuo sobresale por su eficiencia computacional y su capacidad para generar predicciones probabilísticas estables, incluso ante conjuntos de datos reducidos, lo que resulta útil para entornos de diagnóstico [11]. Por su parte, regresión logística destaca por el manejo de problemas de la alta dimensionalidad, donde el número de rasgos extraídos suele superar significativamente al tamaño de la muestra [14].

Para construir los modelos de clasificación se utilizaron los datos descritos en la Tabla 3. Estos datos corresponden a los vectores de características obtenidos de los estudios de EEG de los sujetos con Alzheimer y de control. Es importante mencionar que previo a la construcción de los clasificadores, este conjunto de datos fue sometido a un proceso de normalización, aplicando la transformada Z-score con respecto al grupo de control, la cual no alteró la dimensionalidad del conjunto de datos.

Tabla 3. Descripción del conjunto de datos utilizado para construir los clasificadores.

Clase	Dimensión
Alzheimer (A)	14,515 × 133
Control (C)	12,012 × 133
Total	26,527 × 133

2.5. Evaluación de modelos

Para evaluar el desempeño de los modelos de clasificación, el conjunto de datos se dividió a nivel de sujetos en una proporción de 80 % para entrenamiento y 20 % para validación, garantizando que los registros pertenecientes a un mismo individuo no estuvieran presentes simultáneamente en ambos subconjuntos. Con el propósito de obtener una estimación más robusta y confiable del rendimiento, este procedimiento se repitió durante 10 iteraciones independientes, variando aleatoriamente la distribución de los sujetos.

Sobre cada conjunto de entrenamiento generado se implementó un esquema de validación cruzada dejando un sujeto fuera (LOSO, por sus siglas en inglés: Leave-One-Subject-Out).

De esta manera, se redujo el riesgo de sobreajuste asociado a características particulares de cada sujeto y se favoreció la identificación de patrones relacionados con biomarcadores de la enfermedad de Alzheimer [2].

Bajo este esquema experimental, se definió como clase positiva a los sujetos diagnosticados con enfermedad de Alzheimer y como clase negativa al grupo de control. El desempeño de cada modelo fue evaluado tanto en entrenamiento y validación, utilizando métricas derivadas de la matriz de confusión. Esto permitió realizar un análisis multidimensional del comportamiento de los clasificadores.

En particular, la Exactitud (1) permite medir la proporción global de predicciones correctas sobre el total de casos evaluados [10]. Por otro lado, la Sensibilidad (2) destaca por su importancia en el diagnóstico clínico al cuantificar la habilidad del modelo para detectar correctamente a los pacientes afectados, siendo útil para reducir los falsos negativos [15]. Asimismo, la Especificidad (3) sobresale por su capacidad para identificar con precisión a los sujetos sanos, garantizando la fiabilidad del modelo al clasificar adecuadamente la clase de control y reducir la incidencia de falsos positivos [1]:

$$Exactitud = \frac{TP + TN}{TP + TN + FP + FN}, \quad (1)$$

$$Sensibilidad = \frac{TP}{TP + FN}, \quad (2)$$

$$Especificidad = \frac{TN}{TN + FP}. \quad (3)$$

3. Resultados

En esta sección se muestran los resultados de los dos experimentos realizados para evaluar el desempeño de los modelos:

- Experimento 1: Modelos de clasificación contruidos con las mejores características de acuerdo a los métodos de selección.
- Experimento 2: Modelos especializados contruidos con solo las 19 características correspondientes a esa banda o cociente.
- Ambos experimentos se implementaron en Python utilizando las librerías MNE, Pandas, NumPy y Scikit-learn; y se ejecutaron en una laptop con procesador Apple M2, memoria RAM de 8Gb y 8 núcleos.

3.1. Modelos de clasificación contruidos con las mejores características de acuerdo a los métodos de selección

El objetivo de este experimento es analizar el desempeño de los clasificadores al utilizar los grupos de características obtenidos por Mínima Redundancia Máxima Relevancia utilizando Coeficiente de Fisher (mRMRF) e Información

Mutua (mRMRIM). En particular, se analiza el desempeño de los clasificadores al construirse con las 10, 20, 30, 40 y 50 mejores características rankeadas por ambos métodos. Para evaluar la capacidad predictiva de estos subconjuntos y determinar su rendimiento, se implementaron los clasificadores árbol de decisión, Bayes ingenuo, regresión logística y bosques aleatorios (este último configurado con un ensamble de 10 árboles).

El desempeño de los clasificadores utilizando las mejores 10, 20, 30, 40 y 50 características se presentan en la Figura 3. Los resultados de esta figura indican que el método mRMRIM, en comparación con mRMRF, ayuda a identificar características más adecuadas para el diagnóstico de Alzheimer. Las 50 mejores características obtenidas por mRMRIM se muestran en la Figura 4.

Adicionalmente, en la Figura 3 se puede observar que los modelos con mayor potencial de clasificación son Bayes ingenuo y regresión logística, alcanzando su mejor desempeño bajo distintas configuraciones. Analizando los resultados y priorizando la interpretabilidad mediante el uso de conjuntos pequeños (entre 10 y 20 características), el clasificador Bayes ingenuo se establece como el modelo más eficiente. Este se destaca por alcanzar su máxima exactitud de 72.6 % con solo 10 características (obtenido por el método mRMRIM), manteniendo una sensibilidad robusta cercana al 80 %, lo que asegura una detección efectiva de la patología con una carga computacional y de variables mínima.

En contraste, el modelo de Regresión Logística, aunque alcanza la exactitud más alta (74.0 %), requiere para ello un conjunto de 50 características, lo que compromete la interpretabilidad clínica al aumentar la complejidad del sistema.

Aunque la Regresión Logística muestra una excelente capacidad para identificar casos positivos, su rendimiento en conjuntos pequeños es ligeramente inferior al de Bayes ingenuo.

El desempeño en entrenamiento y pruebas alcanzado por los mejores clasificadores se presentan en las Tablas 4 y 5. En general, los resultados de estas tablas indican que ambos tipos de modelos tienen una buena capacidad de generalización.

Tabla 4. Desempeño alcanzado (promedio $\pm$ desviación estándar) en entrenamiento y pruebas para los modelos de Bayes ingenuo usando las mejores 10, 20, 30, 40 y 50 características obtenidas por mRMRIM.

Método	Métrica		10	20	30	40	50
mRMRIM	Exactitud	Entrenamiento	72.6 $\pm$ 2.3	71.7 $\pm$ 2.3	71.1 $\pm$ 2.3	70.5 $\pm$ 2.4	70.7 $\pm$ 2.2
		Prueba	74.5 $\pm$ 9.2	73.2 $\pm$ 9.3	72.3 $\pm$ 9.4	71.7 $\pm$ 9.3	71.8 $\pm$ 9.0
	Especificidad	Entrenamiento	63.6 $\pm$ 3.8	62.6 $\pm$ 3.8	61.7 $\pm$ 3.8	61.3 $\pm$ 3.7	61.2 $\pm$ 3.6
		Prueba	67.1 $\pm$ 16.3	65.8 $\pm$ 16.7	64.9 $\pm$ 17.2	64.6 $\pm$ 17.3	64.0 $\pm$ 17.0
	Sensibilidad	Entrenamiento	79.8 $\pm$ 1.4	79.0 $\pm$ 1.5	78.6 $\pm$ 1.7	77.9 $\pm$ 1.8	78.3 $\pm$ 1.6
		Prueba	80.9 $\pm$ 5.5	79.6 $\pm$ 5.7	78.8 $\pm$ 5.2	77.8 $\pm$ 5.2	78.7 $\pm$ 4.2

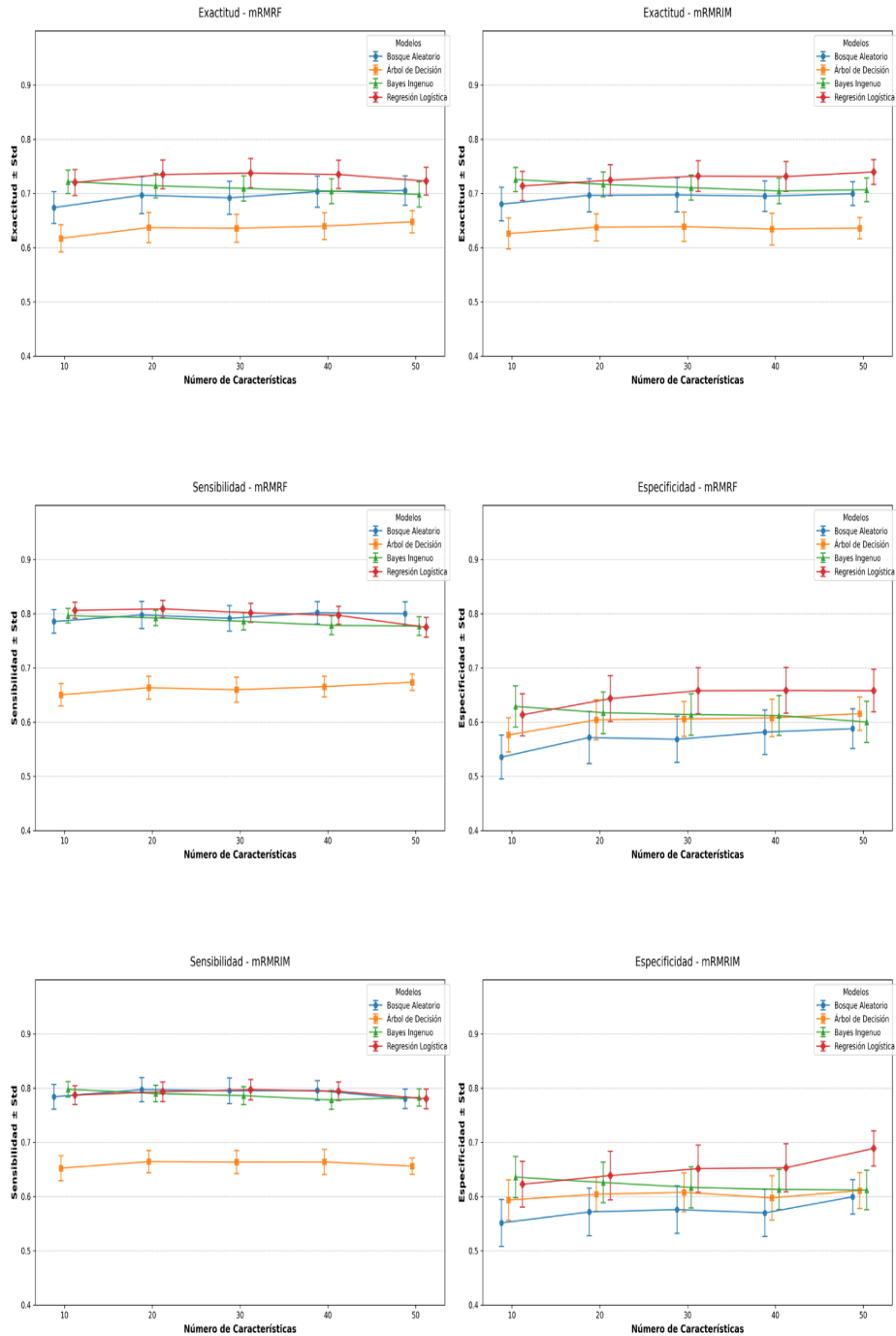


Fig. 3. Exactitud, Sensibilidad y Especificidad promedio, alcanzadas en entrenamiento por los modelos de árbol de decisión, bosque aleatorio, regresión logística y Bayes ingenuo al utilizar las mejores características seleccionadas con los métodos mRMRF y mRMRIM.

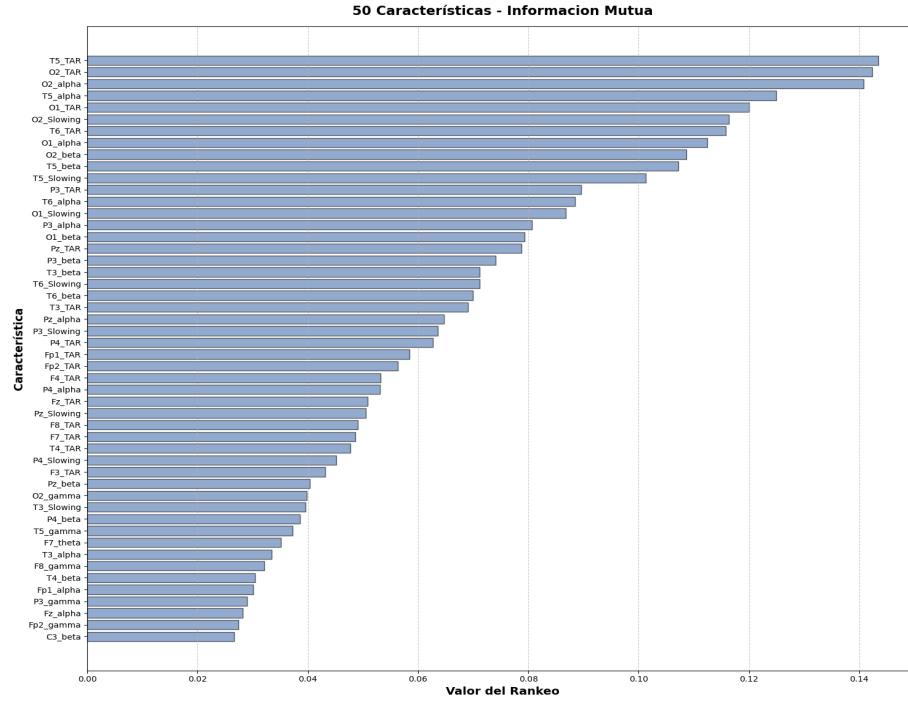


Fig. 4. Las 50 mejores características (Top 50 obtenidas por mRMRIM).

Tabla 5. Desempeño alcanzado (promedio $\pm$ desviación estándar) en entrenamiento y pruebas para los modelos de regresión logística usando las mejores 10, 20, 30, 40 y 50 características obtenidas por mRMRIM.

Método	Métrica		10	20	30	40	50
mRMRIM	Exactitud	Entrenamiento	71.4 $\pm$ 2.7	72.5 $\pm$ 2.9	73.2 $\pm$ 2.8	73.2 $\pm$ 2.8	74.0 $\pm$ 2.3
		Prueba	75.8 $\pm$ 9.1	75.9 $\pm$ 9.3	76.4 $\pm$ 8.8	75.4 $\pm$ 7.8	75.3 $\pm$ 6.0
	Especificidad	Entrenamiento	62.3 $\pm$ 4.2	63.9 $\pm$ 4.5	65.2 $\pm$ 4.4	65.3 $\pm$ 4.4	68.9 $\pm$ 3.2
		Prueba	68.3 $\pm$ 15.5	67.4 $\pm$ 16.3	67.7 $\pm$ 15.1	66.0 $\pm$ 12.3	67.6 $\pm$ 10.5
	Sensibilidad	Entrenamiento	78.7 $\pm$ 1.7	79.4 $\pm$ 1.8	79.7 $\pm$ 1.9	79.4 $\pm$ 1.7	78.1 $\pm$ 1.8
		Prueba	82.5 $\pm$ 7.2	83.7 $\pm$ 7.2	84.4 $\pm$ 7.2	84.0 $\pm$ 7.0	82.5 $\pm$ 8.3

3.2. Modelos especializados contruidos con solo las 19 características correspondientes a esa banda o cociente

El objetivo de este experimento es analizar el desempeño de los clasificadores al utilizar los grupos de características correspondientes a las potencias de banda individuales (Alpha, Beta, Delta y Theta) y a los cocientes fisiológicos (TAR y

DTABR). Esto debido a que la literatura señala que tanto las potencias de banda como los ratios de frecuencia son descriptores robustos y eficaces para diferenciar estados de neurodegeneración en señales de EEG [5,9]. Para evaluar la capacidad predictiva de estos conjuntos, se implementaron los clasificadores árbol de decisión, Bayes ingenuo, regresión logística y bosques aleatorios (este último configurado con un ensamble de 10 árboles), construidos únicamente con sus 19 características espectrales correspondientes.

El desempeño alcanzado en entrenamiento por los clasificadores especializados usando tanto bandas de frecuencia y cocientes fisiológicos se presenta en la Figura 5. Los resultados de esta figura revelan que el uso de características basadas en los cocientes de frecuencia permiten construir modelos de diagnóstico de Alzheimer con un desempeño superior al obtenido mediante el uso de bandas aisladas. Los modelos basados en bandas individuales de baja frecuencia, específicamente para Delta y Theta, muestran un rendimiento inferior y una limitada sensibilidad diagnóstica, evidenciando la dificultad para generar modelos competitivos.

En contraste, en la Figura 5 también se puede encontrar que los modelos basados en el cociente TAR (Theta/Alpha) alcanzan el mejor desempeño, siendo el modelo TAR de regresión logística el mejor modelo con una exactitud del 72.3 % y una sensibilidad del 80.0 %, lo que asegura una detección efectiva de la patología. Por su parte, el modelo TAR de Bayes ingenuo, aunque con un desempeño menor al de regresión logística, se mantiene como una alternativa sólida con una exactitud del 69.3 %.

El desempeño en entrenamiento y pruebas alcanzado por los mejores clasificadores especializados se muestran en las Tablas 6 y 7. Como se puede observar, en estas tablas estos modelos especializados alcanzan un rendimiento comparable en entrenamiento y pruebas, destacando los modelos TAR y DTABR. En particular, el modelo TAR alcanza un buen desempeño tanto en entrenamiento como en pruebas, lo que indica que tiene una buena capacidad de generalización.

Tabla 6. Desempeño alcanzado (promedio $\pm$ desviación estándar) en entrenamiento y pruebas para los modelos de Bayes ingenuo especializados por banda de potencia y cocientes fisiológicos.

Métrica		Delta	Theta	Alpha	Beta	TAR	DTABR
Exactitud	Entrenamiento	47.0 $\pm$ 0.6	54.8 $\pm$ 1.0	66.6 $\pm$ 2.4	65.4 $\pm$ 1.6	69.3 $\pm$ 2.0	64.6 $\pm$ 1.8
	Prueba	49.2 $\pm$ 1.3	56.6 $\pm$ 3.7	66.6 $\pm$ 9.6	66.2 $\pm$ 6.1	70.7 $\pm$ 8.1	65.5 $\pm$ 8.2
Especificidad	Entrenamiento	63.5 $\pm$ 2.3	70.5 $\pm$ 5.5	52.9 $\pm$ 4.5	58.8 $\pm$ 7.9	60.8 $\pm$ 3.2	54.2 $\pm$ 3.1
	Prueba	64.1 $\pm$ 2.1	70.9 $\pm$ 7.6	55.3 $\pm$ 19.0	59.0 $\pm$ 16.3	63.3 $\pm$ 14.6	56.9 $\pm$ 15.4
Sensibilidad	Entrenamiento	33.7 $\pm$ 1.8	42.1 $\pm$ 5.5	77.6 $\pm$ 1.5	70.7 $\pm$ 5.0	76.2 $\pm$ 1.9	73.0 $\pm$ 1.7
	Prueba	35.5 $\pm$ 3.5	43.5 $\pm$ 6.1	76.6 $\pm$ 6.1	72.5 $\pm$ 11.6	77.4 $\pm$ 3.9	73.1 $\pm$ 5.2

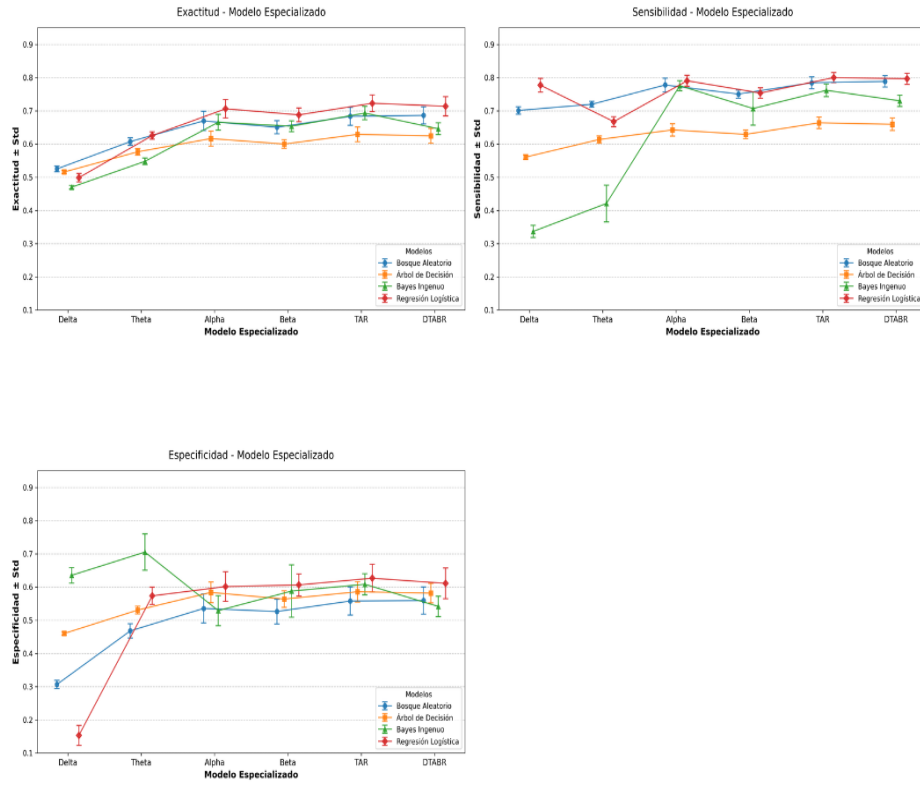


Fig. 5. Exactitud, Sensibilidad y Especificidad promedio alcanzadas en entrenamiento por los modelos especializados por banda de potencia y cocientes fisiológicos.

Tabla 7. Desempeño alcanzado (promedio $\pm$ desviación estándar) en entrenamiento y pruebas para los modelos de regresión logística especializados por banda de potencia y cocientes fisiológicos.

Métrica		Delta	Theta	Alpha	Beta	TAR	DTABR
Exactitud	Entrenamiento	49.9 $\pm$ 1.3	62.6 $\pm$ 1.1	70.6 $\pm$ 2.8	68.8 $\pm$ 2.1	72.3 $\pm$ 2.5	71.4 $\pm$ 2.9
	Prueba	52.9 $\pm$ 3.2	64.1 $\pm$ 3.0	72.3 $\pm$ 8.2	69.2 $\pm$ 6.3	74.9 $\pm$ 8.1	72.0 $\pm$ 8.4
Especificidad	Entrenamiento	15.4 $\pm$ 3.0	57.4 $\pm$ 2.6	60.2 $\pm$ 4.5	60.6 $\pm$ 3.4	62.7 $\pm$ 4.2	61.1 $\pm$ 4.6
	Prueba	21.9 $\pm$ 5.2	57.6 $\pm$ 7.1	61.0 $\pm$ 14.5	59.6 $\pm$ 11.7	64.3 $\pm$ 14.6	61.2 $\pm$ 14.3
Sensibilidad	Entrenamiento	77.7 $\pm$ 2.1	66.8 $\pm$ 1.5	79.1 $\pm$ 1.7	75.4 $\pm$ 1.6	80.0 $\pm$ 1.5	79.7 $\pm$ 1.7
	Prueba	81.7 $\pm$ 3.8	70.3 $\pm$ 7.5	82.7 $\pm$ 10.1	78.0 $\pm$ 8.6	84.4 $\pm$ 7.4	81.7 $\pm$ 8.7

4. Conclusiones

Esta investigación presentó una evaluación comparativa orientada al diagnóstico de la enfermedad de Alzheimer mediante el uso de distintos modelos de aprendizaje automático, incluyendo Bayes ingenuo, árboles de decisión, regresión logística y bosques aleatorios. A partir del análisis de desempeño bajo diferentes configuraciones de características, se observó que el clasificador Bayes ingenuo destacó por ofrecer un equilibrio entre rendimiento e interpretabilidad, manteniendo una sensibilidad cercana al 80 % y una exactitud estable con conjuntos reducidos de 10 características. Esto lo convierten en una alternativa adecuada para entornos clínicos, donde la transparencia y facilidad de interpretación en la toma de decisiones representan factores esenciales. Por otra parte, la regresión logística alcanzó un mayor potencial predictivo en el estudio, obteniendo una exactitud máxima de 73.97 %. Sin embargo, dicho rendimiento estuvo asociado al empleo de un conjunto más amplio de 50 características. Asimismo, los resultados permitieron identificar al cociente TAR (Theta/Alpha) como un biomarcador relevante para la cuantificación de la ralentización cerebral asociada al Alzheimer. En particular, el uso de este cociente fisiológico, en modelos como regresión logística y Bayes ingenuo, proporcionó una detección más robusta y consistente de la enfermedad en comparación con el uso aislado de bandas de potencia individuales.

Finalmente, aunque el presente trabajo se centró en identificar las características espectrales y los modelos más eficientes para optimizar la detección temprana del Alzheimer, se reconoce que el diagnóstico diferencial continúa siendo un desafío importante. Por ello, como línea futura de investigación se plantea el desarrollo de modelos multiclase capaces de integrar patologías como la demencia frontotemporal dentro del esquema de clasificación automática.

Agradecimientos. Los autores desean agradecer a la Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI) por el apoyo financiero brindado para la realización de los estudios de maestría. Asimismo, los autores agradecen a la Universidad Tecnológica de la Mixteca por el respaldo académico e institucional proporcionado durante el desarrollo de esta investigación.

Referencias

1. Alessandrini, M., Biagetti, G., Crippa, P., Falaschetti, L., Luzzi, S., Turchetti, C.: EEG-based Alzheimer's disease recognition using robust-PCA and LSTM recurrent neural network. *Sensors* **22**(10), 3696 (2022). <https://doi.org/10.3390/s22103696>
2. Aljalal, M., Aldosari, S.A., Molinas, M., Alturki, F.A.: Selecting EEG channels and features using multi-objective optimization for accurate MCI detection: validation using leave-one-subject-out strategy. *Sci. Rep.* **14**, 12185 (2024). <https://doi.org/10.1038/s41598-024-63180-y>
3. Gu, Q., Li, Z., Han, J.: Generalized Fisher score for feature selection. *arXiv preprint arXiv:1202.3725* (2012)
4. Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. *J. Mach. Learn. Res.* **3**, 1157–1182 (2003)

5. Miltiadous, A., Tzimourta, K.D., Giannakeas, N., Tsipouras, M.G., Afrantou, T., Ioannidis, P., Tzallas, A.T.: Alzheimer's disease and frontotemporal dementia: a robust classification method of EEG signals and a comparison of validation methods. *Diagnostics* **11**(8), 1437 (2021). <https://doi.org/10.3390/diagnostics11081437>
6. Miller, T. *Explanation in artificial intelligence: Insights from the social sciences*. Artificial Intelligence, Vol. 267, pp. 1–38, 2019. <https://doi.org/10.1016/j.artint.2018.07.007>
7. Miltiadous, A., Tzimourta, K.D., Afrantou, T., Ioannidis, P., Grigoriadis, N., Tsalikakis, D.G., Angelidis, P., Tsipouras, M.G., Glavas, E., Giannakeas, N., Tzallas, A.T.: A dataset of EEG recordings from: Alzheimer's disease, frontotemporal dementia and healthy subjects. OpenNeuro (2024). <https://doi.org/10.18112/openneuro.ds004504.v1.0.8>
8. Secretaría de Salud: Enfermedad de Alzheimer, demencia más común que afecta a personas adultas mayores. Gobierno de México, <https://www.gob.mx/salud/es/articulos/enfermedad-de-alzheimer-demencia-mas-comun-que-afecta-a-personas-adultas-mayores>, last accessed 2026/01/28
9. Senkaya, Y., Kurnaz, C., Özbilgin, F.: Enhancing Alzheimer's diagnosis with machine learning on EEG: a spectral feature-based comparative analysis. *Diagnostics* **15**(17), 2190 (2025). <https://doi.org/10.3390/diagnostics15172190>
10. Stefanou, K., Tzimourta, K.D., Bellos, C., Stergios, G., Markoglou, K., Gionanidis, E., Tsipouras, M.G., Giannakeas, N., Tzallas, A.T., Miltiadous, A.: A novel CNN-based framework for Alzheimer's disease detection using EEG spectrogram representations. *J. Pers. Med.* **15**(1), 27 (2025). <https://doi.org/10.3390/jpm15010027>
11. Wang, J., Wang, M.: Review of the emotional feature extraction and classification using EEG signals. *Cogn. Robot.* **1**, 29–40 (2021). <https://doi.org/10.1016/j.cogr.2021.04.001>
12. World Health Organization: The top 10 causes of death, <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>, last accessed 2026/02/05
13. Yuan, Y., Zhao, Y.: The role of quantitative EEG biomarkers in Alzheimer's disease and mild cognitive impairment: applications and insights. *Front. Aging Neurosci.* **17**, 1522552 (2025). <https://doi.org/10.3389/fnagi.2025.1522552>
14. Zhan, L., Liu, Y., Wang, Y., Zhou, J., Jahanshad, N., Ye, J., Thompson, P. M. (2015). Boosting brain connectome classification accuracy in Alzheimer's disease using higher-order singular value decomposition. *Frontiers in Neuroscience*, *9*, 257. <https://doi.org/10.3389/fnins.2015.00257>
15. Zheng, X., Wang, B., Liu, H., Wu, W., Sun, J., Fang, W., Jiang, R., Hu, Y., Jin, C., Wei, X., Chen, S.S.C.: Diagnosis of Alzheimer's disease via resting-state EEG: integration of spectrum, complexity, and synchronization signal features. *Front. Aging Neurosci.* **15**, 1288295 (2023). <https://doi.org/10.3389/fnagi.2023.1288295>