ISSN 1870-4069

Clasificación de la enfermedad de Alzheimer utilizando Redes Neuronales Profundas Multimodales

Ayrton Santos, Claudia I. González, Mario García

Instituto Tecnológico de Tijuana/TECNM,
División de Estudios de Posgrado e Investigación,
México

ayrton.santos@tectijuana.edu.mx, cgonzalez@tectijuana.mx,
mario@tectijuana.edu.mx

Resumen Las enfermedades neurodegenerativas, como la enfermedad de Alzheimer, representan un desafío creciente para la salud pública global. Con el envejecimiento de la población y los avances en tecnologías de reconocimiento de patrones y aprendizaje automático, la detección temprana de estas patologías ha mejorado considerablemente. En este trabajo se propone el uso de redes neuronales multimodales para la predicción de la enfermedad de Alzheimer integrando imágenes de resonancia magnética por sus siglas en inglés MRI (Magnetic Resonance Imaging) y las Calificaciones Clínicas de Demencia CDR (Clinical Dementia Rating). En la propuesta se evalúan dos enfoques arquitectónicos: fusión temprana y fusión tardía. A través de un análisis estadístico, los resultados obtenidos demuestran que la fusión temprana proporciona una ventaja significativa en el análisis de datos médicos, optimizando la precisión en la clasificación. El modelo se aplica específicamente a la clasificación de la enfermedad de Alzheimer utilizando la base de datos OASIS-3.

Palabras clave: Aprendizaje profundo multimodal, Redes Neuronales Multimodales, clasificación de la enfermedad de Alzheimer, IA en áreas médicas.

Classification of Alzheimer's Disease Using Multimodal Deep Neural Networks

Abstract Neurodegenerative diseases, such as Alzheimer's disease, represent a growing challenge for global public health. With population aging and advances in pattern recognition and machine learning technologies, early detection of these pathologies has considerably improved. This work proposes the use of multimodal neural networks for the prediction of Alzheimer's disease by integrating Magnetic Resonance Imaging (MRI) and Clinical Dementia Ratings (CDR). Two architectural approaches are evaluated in the proposal: early fusion and late fusion.

Through statistical analysis, the results obtained demonstrate that early fusion provides a significant advantage in the analysis of medical data, optimizing classification accuracy. The model is specifically applied to the classification of Alzheimer's disease using the OASIS-3 database.

Keywords: Multimodal deep learning, Multimodal Neural Networks, Alzheimer's disease classification, AI in medical fields.

1. Introducción

Las enfermedades neurodegenerativas son un grupo de trastornos relacionados con la edad que provocan la muerte de tipos específicos de células neuronales. Son más comunes en la población de la tercera edad, lo que las convierte en un grave problema de salud global. Los factores que contribuyen a las enfermedades neurodegenerativas incluyen mutaciones genéticas, apoptosis neuronal, pérdida de proteínas y reacciones de las células gliales [3]. Existen diferentes tipos de enfermedades neurodegenerativas, que comparten síntomas. Estas se caracterizan por el deterioro progresivo de funciones fisiológicas y cognitivas. Las enfermedades que afectan el cerebro son variadas, incluyendo el Alzheimer y la enfermedad de Parkinson [15].

La integración de tecnologías potentes en el área de la salud, como el aprendizaje automático y el aprendizaje profundo, se encuentra en pleno apogeo y se desarrolla a velocidades cada vez más altas, aprovechando el conocimiento acumulado durante años. Esto representa un gran beneficio, ya que, gracias a su precisión, la experiencia tanto para médicos como para pacientes ha mejorado notablemente. Además, los pronósticos pueden realizarse antes de que las enfermedades se vuelvan potencialmente mortales. A esto se suma que las técnicas de aprendizaje automático permiten evaluar la eficacia de nuevos medicamentos de forma más rápida y precisa [8].

Las enfermedades neurodegenerativas abarcan una amplia gama de problemas neurológicos progresivos que son dependientes de la edad y afectan aproximadamente a 50 millones de personas en todo el mundo, especialmente a adultos mayores. En las últimas décadas, el aumento en la población anciana proyecta una tasa de mortalidad del 42 %, representando el 11.8 % de todas las muertes. Es importante mencionar que algunas enfermedades no se comprenden completamente, y para otras no existe cura o tratamiento, lo que finalmente lleva a la muerte del paciente [16].

En un estudio publicado en 2020, aproximadamente 800,000 personas estaban afectadas por demencia. La prevalencia actual varía entre el 7.9 % y el 9 %, colocando a México en el 5° lugar con una alta incidencia. Según este estudio, se estima que para el año 2050 hasta 3 millones de personas estarán afectadas por demencia [14]. Como dato interesante, la relación de riesgo de mortalidad en áreas urbanas es mayor, con un 2.7. Esto se debe a muchos factores, como la calidad de la atención médica, la contaminación, entre otros, mientras que en áreas rurales es de 1.6 [17].

La enfermedad de Alzheimer, una de las principales causas de demencia en el mundo, continúa representando un desafío significativo para la salud pública debido a su impacto en la calidad de vida de los pacientes y en los sistemas de atención sanitaria. A medida que la población global envejece, la necesidad de métodos eficaces para la detección temprana y el diagnóstico preciso de esta enfermedad se vuelve aún más urgente. La multimodalidad ha cobrado gran relevancia en los últimos años debido a la necesidad de integrar diferentes tipos de datos, como texto, imagen y audio, en un mismo modelo de aprendizaje. En este contexto, las tecnologías de aprendizaje automático y el análisis de datos multimodales han abierto nuevas posibilidades para mejorar la precisión de la clasificación de enfermedades neurodegenerativas. Este trabajo presenta una propuesta para el diseño y desarrollo de arquitecturas multimodales basadas en redes profundas para la predicción de la enfermedad de Alzheimer. Combinando imágenes de resonancia magnética (MRI) y las Calificaciónes Clínicas de Demencia (CDR), se analizan dos enfoques arquitectónicos: fusión temprana y fusión tardía, mostrando que la fusión temprana mejora de manera significativa la precisión en el diagnóstico. Los resultados obtenidos, aplicados sobre la base de datos OASIS-3, demuestran el potencial de estas técnicas para avanzar en la detección temprana de la enfermedad, mejorando tanto la eficiencia en el diagnóstico. Este enfoque permite mejorar el desempeño en tareas complejas al aprovechar la complementariedad de las distintas fuentes de información.

En la Sección 2 se definen algunos de los conceptos básicos sobre la multimodalidad, redes profundas multimodales y principales arquitecturas multimodales. En la Sección 3 se presentan algunos antecedentes de investigaciones relacionadas a arquitecturas multimodales. En la Sección 4 se describe la metodología propuesta para el diseño de arquitecturas multimodales con fusión temprana y fusión tardía. En la Sección 5 se detallan los resultados obtenidos y finalmente, en la Sección 7, se abordan las conclusiones y el trabajo futuro.

2. Multimodalidad

2.1. Redes Neuronales Profundas Multimodales

Las redes neuronales son una subdivisión del aprendizaje profundo que usa imágenes para clasificar imágenes, aunque también se ha visto que se utilizan también para áreas como el procesamiento de sonido. En comparación con las redes neuronales tradicionales, en las que cada neurona está conectada a la siguiente (lo cual se conoce como capa densa o completamente conectada), las redes neuronales convolucionales utilizan la capa densa solo en la última parte de la red neuronal. Las redes neuronales logran el reconocimiento de formas mediante la combinación de agrupación (pooling) y capas densas. Las capas convolucionales procesan los datos de entrada usando múltiples filtros, se aplica una función de activación, y posteriormente las capas de agrupación extraen las características más significativas mediante operaciones como agrupamiento máximo (max pooling) o agrupamiento promedio (average pooling). Después

del aprendizaje, las capas completamente conectadas se convierten en un vector unidimensional, que es introducido en una función Softmax para la construcción del modelo [7].

Estas redes combinan tanto entradas visuales como lingüísticas para aprender correspondencias entre imágenes (referentes) y palabras. Algunos ejemplos de sus aplicaciones son la generación automática de subtítulos para imágenes y respuestas a preguntas visuales. Las redes multimodales emplean codificadores de imágenes y palabras que transforman las entradas en un espacio de representación multimodal compartido. Se utiliza una función de pérdida contrastiva para alinear los pares palabra-referente mientras se separan los que no están relacionados. Estas redes muestran potencial como modelos cognitivos al aprender de estímulos visuales y lingüísticos sin procesar, lo cual imita cómo los niños podrían aprender palabras en entornos ambiguos. Este proceso es conocido como aprendizaje cross-situacional. Se realizan experimentos bajo condiciones como ambigüedad referencial, mapeo rápido y exclusividad mutua para comprobar si las redes multimodales replican comportamientos de aprendizaje humano. Estos experimentos demuestran cómo los modelos multimodales pueden manejar tanto ambigüedad lingüística como tareas de generalización visual [19].

En [2], se presenta la idea de que el mundo es multimodal, ya que podemos ver objetos, sentir texturas, escuchar sonidos y experimentar sabores. Para que los algoritmos de inteligencia artificial puedan lograr mejores resultados, es necesario, en primer lugar, enfocarse en la forma de capturar y resumir los datos multimodales. Además, se debe considerar la traducción, que implica mapear los datos de una modalidad a otra, tomando en cuenta la heterogeneidad; la alineación, que consiste en encontrar relaciones entre los elementos de las modalidades; y finalmente, la fusión, que se centra en combinar la información de varias modalidades para hacer predicciones integrando la información multimodal. Hay mucha discusión sobre la diferencia entre las redes neuronales, multimodales y las modulares. La tecnología multimodal se refiere a la capacidad de un sistema para procesar y combinar múltiples tipos de datos, o modalidades, como texto, imágenes, audio, etc., con el fin de obtener mejores resultados.

En [20] se analizan los modelos multimodales actuales y se clasifican distintos tipos de arquitecturas de modelos multimodales de última generación; se describe de manera detallada de los tipos de fusión temprana y fusión tardía para entender las diversas arquitecturas, las cuales ser dividen en A, B, C y D como se muestra en la Figura 1.

La arquitectura tipo A abarca los modelos multimodales tempranos, los datos multimodales como imagen, audio y video, se procesan a través de codificadores, específicos para cada modalidad, un resampleador genera un número de tokens fijo, que se alinean con la capa decodificadora. La arquitectura tipo B está hecha usando un modelo de lenguaje pre-entrenado una capa lineal, que puede aprender de un módulo Q Former, capas de atención cruzada personalizada o capas personalizadas, y codificadores de modalidad. La arquitectura tipo C es la más usada, caracterizándose por incluir módulos y ser simple en

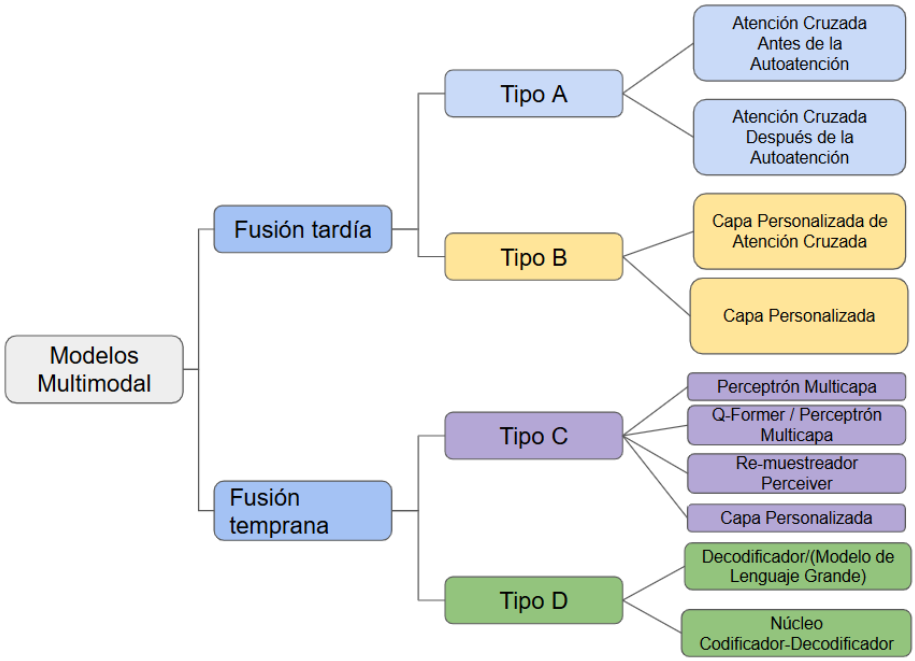

Fig. 1. Arquitecturas multimodales principales [20].

desarrollo y entrenamiento. La salida del codificador de modalidad se dirige y fusiona en la entrada del modelo sin interacciones con las capas internas, lo que permite una fusión temprana. El entrenamiento consta de tres etapas: pre-entrenamiento, ajuste con instrucciones y ajuste para alineación. Es la que menos recursos demanda, y este tipo de arquitectura permite facilitar la adopción y el entrenamiento en tareas multimodales. La clase D tokeniza las entradas multimodales usando un tokenizador común o específicos para cada modalidad. Las entradas tokenizadas se alimentan a un modelo de lenguaje grande o a un transformer tipo codificador-decodificador, generando salidas multimodales. La ventaja de tokenizar las entradas es que permite generar tokens de diversas modalidades (imagen, audio y texto) de manera autoregresiva aunque la tipo D tiene un enfoque más generativo que de clasificación como lo requerimos en este caso [20].

3. Antecedentes

En el estudio de Chattopadhyay [4], se utilizó una red neuronal convolucional 3D para predecir la acumulación de la proteína beta amiloide, un factor para desarrollar la enfermedad de alzheimer basándose en resonancias magnéticas, utilizando biomarcadores. El estudio consistió en 762 pacientes ancianos: 459 sanos y 67 con deterioro cognitivo leve. Los resultados mostraron que 236 pacientes tenían demencia, y la precisión obtenida fue del 76 %.

En [11] se presentó una selección de características multimodal de 3 tipos para la detección de Alzheimer: imágenes de resonancia magnética, tomografía por emisión de positrones y análisis de líquido cefalorraquídeo. El aprendizaje multimodal implica el uso de diferentes tipos de datos y el aprendizaje de lo mejor de esa información combinada, ya sea texto, audio o video. El problema es que los algoritmos comunes trabajan con solo un tipo de datos a la vez, lo que significa que no siempre se puede utilizar toda la información disponible.

En [5] el trabajo se centra en la enfermedad de Alzheimer; se realiza una predicción temprana de Alzheimer y deterioro cognitivo leve utilizando imágenes de resonancia magnética cerebral y aprendizaje automático. Se utilizaron dos conjuntos de datos: la Serie de Estudios de Imágenes de Acceso Abierto (OASIS) y la Iniciativa de Neuroimagen de la Enfermedad de Alzheimer (ADNI). Algunos de los algoritmos utilizados incluyeron máquinas de vectores de soporte, árboles de decisión, bosques aleatorios, árboles extremadamente aleatorios, análisis discriminante lineal, regresión logística y regresión logística con descenso de gradiente estocástico. Los mejores resultados se obtuvieron con los algoritmos de bosque aleatorio y árboles extremadamente aleatorios.

En [18], se exploraron múltiples herramientas estadísticas y algoritmos de aprendizaje automático para el diagnóstico de la enfermedad de Alzheimer en personas mayores de 75 años. A medida que la tecnología mejoró, se aplicó a la clasificación de imágenes médicas. Las muestras se dividieron en 75 % y 25 %, utilizadas para entrenar los algoritmos, que se ejecutaron en potentes GPUs. Utilizando redes neuronales convolucionales, se empleó la arquitectura optimizada llamada OViTAD. Estas tuberías lograron resultados promedio de 94.32 % y 97.88 % para las tuberías de fMRI y MRI, respectivamente.

En [6] se propone un estudio que compara el rendimiento de tres algoritmos de aprendizaje automático: Random Forest, Gradient Boosting y eXtreme Gradient Boosting, utilizando biomarcadores multimodales de sujetos con deterioro cognitivo leve (MCI) obtenidos de la base de datos ADNI. La tasa de predicción está relacionada con la naturaleza de los datos, como pruebas neuropsicológicas, proteínas relacionadas con el Alzheimer, líquido cefalorraquídeo, MRI, entre otros. Los datos sMRI (MRI estructurales), por sí solos tuvieron menor precisión (79 %), pero al tratarse de datos multimodales, combinando medidas clínicas y biológicas, se logró una mayor precisión (90 %).

En [13], se presentó un estudio que discute cómo las personas con la enfermedad de Parkinson pueden desarrollar demencia, pero no todos los pacientes la desarrollan de manera gradual o al mismo tiempo. Se recopilaron datos de 48 pacientes con Parkinson, en los que se evaluaron 38 características de riesgo, como habilidades motoras, capacidades cognitivas, moléculas en sangre, entre otros factores. Se utilizó el modelo de Random Forest, que es un algoritmo de aprendizaje automático, y una técnica llamada Tree SHAP, que explica por qué ciertos factores son importantes para las predicciones del modelo. Random Forest fue muy preciso al clasificar qué pacientes desarrollaron demencia, con un área bajo la curva (AUC) de 0.84.

Tabla 1. Resumen de técnicas y estudios sobre Alzheimer

Técnicas	**Base de Datos**	**Autores y Año**
CNN 3D	MRI + biomarcadores	Chattopadhyay (2023) [4]
Selección multimodal + Graph Anchors	MRI, PET, CSF	Li (2022) [11]
SVM, Árboles, RF, Reg. Logística	OASIS, ADNI, MRI, PET, CSF	Diogo (2022) [5]
CNN (OViTAD)	fMRI, MRI	Sarraf (2021) [18]
RF, Gradient Boosting, XGBoost	ADNI, MCI	Franciotti (2023) [6]
RF, Tree SHAP	Datos clínicos (Parkinson)	McFall (2023) [13]
Fusión multimodal tiempo-frecuencia	Neuroimagen y genética	Anand (2024) [1]
MC-RVAE, KNN, RF, GFA, ANOVA	MRI + cognitivas	Martí-Juan (2023) [12]
Contrastive Learning + ResNet + Tabular	MRI + datos tabulares	Huang (2023) [9]

En [1], se propone un modelo que es multimodal para lograr mejorar la detección de enfermedades neurodegenerativas. Usa distintos métodos avanzados para analizar la variedad de datos, análisis de tiempo-frecuencia, resonancias electromagnéticas, y datos genéticos. Estas características de datos permiten un diagnóstico preciso, con un aumento del 10 % en precisión y una reducción del tiempo del 2,9 %. Este estudio se enfocó en una amplia gama de enfermedades neurodegenerativas.

El modelo MC-RVAE [12], diseñado para manejar la enfermedad de Alzheimer de manera multimodal, trabaja con MRI y puntuaciones cognitivas. Este modelo fue entrenado con datos sintéticos y de ADNU con aproximadamente 3000 épocas. Se usaron los vecinos más cercanos, bosques aleatorios y análisis de factor de grupo. Los resultados se compararon usando un modelo ANOVA para cada tarea; este es flexible y escalable.

El modelo [9] utiliza aprendizaje contrastivo, que alinea las imágenes de resonancia y los datos tabulares, creando un espacio embebido conjunto para diferentes modalidades. Esta propuesta incluye capacidades de entrenamiento multimodal. El encoder, basado en ResNet, se encarga de procesar imágenes e integra un módulo de atención tabular que resalta los datos más relevantes y mejora la interpretación de los mismos. El modelo asigna puntuaciones para capturar las relaciones entre los datos. Es entrenado con 64 épocas, usando un optimizador Adam. El tamaño del batch es de 4 para imágenes 3D y 32 para imágenes 2D, alcanzando una precisión del 95.5 %. La adición de datos tabulares mejora significativamente el rendimiento del modelo.

4. Metodología

4.1. Datos multimodales

OASIS-3 es una recopilación longitudinal multimodal especializada en la enfermedad de Alzheimer. Contiene información de 1378 pacientes, con edades entre 42 y 95 años, recopilada a lo largo de 30 años a través de diversos estudios del Knight Alzheimer's Disease Research Center de la Universidad de Washington en St. Louis. El conjunto incluye datos crudos de MRI y de tomografías PET (Positron Emission Tomography), así como evaluaciones clínicas y cognitivas, estructuradas en más de 2800 sesiones de resonancia magnética (T1w, FLAIR, ASL, DTI, entre otras), y más de 2100 sesiones PET con distintos trazadores (PIB, AV45, FDG y Tau). Además, cuenta con una sección preprocesada mediante FreeSurfer, ideal para investigadores que no deseen limpiar manualmente los datos crudos de MRI. Gracias a esta herramienta, científicos e ingenieros pueden desarrollar modelos de inteligencia artificial con alta precisión [10].

4.2. Preprocesamiento de datos

Primero, se solicitó acceso al dataset. Una vez concedidos los permisos, se procedió a descargar los datos: 800 GB de información cruda que incluía evaluaciones médicas, historiales clínicos y familiares de los pacientes, así como tomografías computarizadas e imágenes de resonancia magnética (MRI). La etapa inicial del procesamiento se centró en la limpieza de los MRI. Afortunadamente, el equipo que proporciona el dataset incluye una sección específica con imágenes en un formato compatible con la herramienta FreeSurfer, lo que facilita su manipulación. Un aspecto destacable es que estas imágenes ya vienen preprocesadas: el cerebro está segmentado, es decir, aislado de otras estructuras como órganos o huesos, lo cual optimiza el reconocimiento de patrones cerebrales. Con esta base, se seleccionaron 26 cortes axiales del cerebro, ya que en esas secciones se encuentra el hipocampo, una región clave para la detección temprana del Alzheimer. A continuación, se utilizó un script para convertir las imágenes a escala de grises y redimensionarlas a 128×128 píxeles. Continuando con el uso del CDR, se consideró el diagnóstico más reciente de cada paciente para determinar la presencia o ausencia de Alzheimer. Con esta información, se ejecutó un script que emparejaba las imágenes de resonancia magnética (MRI) con sus respectivos valores de CDR, generando así el conjunto de datos necesario para entrenar la red neuronal. El pre-procesamiento previamente mencionado, derivó una base de datos con 1,248 registros de MRI y de CDR, de los cuales se dividieron en un 70 % para entrenamiento, 15 % para validación y 15 % para pruebas.

4.3. Flujo de trabajo

En las propuestas de arquitectura existen dos líneas principales: la rama de CDR, donde se toman datos volumétricos del cerebro, además del indicativo de

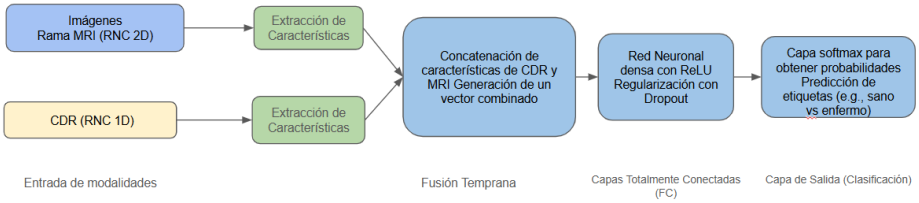

Fig. 2. Arquitectura de la fusión temprana.

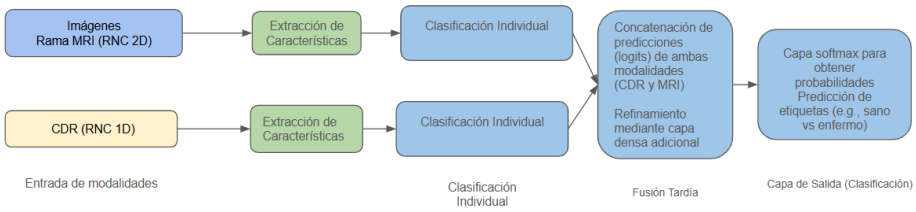

Fig. 3. Arquitectura de la fusión tardía.

paciente enfermo o sano; y, por otro lado, la rama de imágenes de resonancias magnéticas (MRI). La fusión se lleva a cabo en una etapa diferente dependiendo del tipo de fusión. Para la fusión temprana (Early Fusion), como se muestra en la Figura 2, se cuenta con una red neuronal 1D para procesar el CDR y una red neuronal convolucional 2D para las MRI. La fusión ocurre al concatenar las dos salidas y pasarlas por capas totalmente conectadas (fully connected). Cabe destacar que en este enfoque multimodal se utilizaron técnicas para mitigar el sobreajuste con dropout en las capas totalmente conectadas, se normaliza el batch tras la primera capa lineal que trabaja con los datos de CDR, y se emplea data augmentation, que genera rotaciones aleatorias en las imágenes para forzar al modelo a generalizar en vez de aprender de memoria.

La fusión tardía, como se muestra en la Figura 3, posee una arquitectura con una red neuronal convolucional de una capa que procesa el CDR, y otra red neuronal convolucional de dos capas que trabaja con MRI. A diferencia de la fusión temprana, que utiliza un vector de características para la concatenación, en la fusión tardía se obtienen los logits de cada red neuronal; luego, dichos logits se concatenan y se pasan por una capa final que genera la predicción.

Se crean las funciones que entrenan el modelo, permitiendo que procese la información del dataset, calcule posibles errores y ajuste los parámetros. Una segunda función se encarga de evaluar el modelo, registrando las predicciones en un log y calculando diferentes métricas como la exactitud, precisión, sensibilidad, especificidad y el área bajo la curva (AUC), para analizar su rendimiento.

5. Resultados

Se llevaron a cabo 30 experimentos independientes, cada uno con 100 épocas y un tamaño de lote (batch size) de 32, con el objetivo de evaluar el desempeño de

Tabla 2. Resultados del modelo con fusión temprana.

Pérdida entrenam.	Pérdida validación	Exactitud	Precisión	Sensibilid.	Especificidad	AUC
0.02	0.04	0.99	0.99	0.99	0.98	1
0.02	0.02	0.99	1	0.99	0.99	1
0.02	0.02	0.99	1	0.99	0.99	1
0.02	0.05	0.99	1	0.98	0.99	1
0.02	0.04	0.99	0.99	0.99	0.98	1
0.02	0.04	0.99	0.99	1	0.97	1
0.02	0.03	0.99	0.99	0.99	0.98	1
0.02	0.05	0.99	0.99	0.99	0.96	1
0.02	0.06	0.99	0.98	0.99	0.96	1
0.02	0.05	0.99	0.99	0.99	0.97	1
0.02	0.05	0.99	0.99	1	0.98	1
0.02	0.04	0.99	0.99	0.98	0.98	1
0.02	0.03	0.99	0.99	0.99	0.98	1
0.02	0.02	0.99	0.99	1	0.98	1
0.02	0.02	0.99	1	0.99	0.99	1
0.02	0.04	0.99	0.99	1	0.97	1
0.02	0.06	0.99	0.99	0.99	0.97	1
0.02	0.03	0.99	0.99	0.99	0.99	1
0.02	0.02	0.99	0.99	0.99	0.98	1
0.02	0.03	0.99	1	0.99	0.99	1
0.02	0.03	0.99	0.99	1	0.98	1
0.02	0.02	0.99	1	0.99	0.99	1
0.02	0.05	0.99	0.99	0.99	0.98	0.99
0.02	0.09	0.99	0.99	0.97	0.99	1
0.02	0.06	0.99	0.99	0.99	0.97	1
0.02	0.02	0.99	0.99	1	0.98	1
0.02	0.05	0.99	0.99	1	0.97	1
0.02	0.05	0.99	0.99	1	0.96	1
0.02	0.03	0.99	0.99	0.98	0.99	1
0.02	0.02	0.99	0.99	0.99	0.98	1

dos arquitecturas: fusión temprana y fusión tardía. El entrenamiento completo de la arquitectura de fusión temprana tomó aproximadamente 44 horas, mientras que la fusión tardía requirió alrededor de 39 horas El entrenamiento se llevo en un equipo con un procesador Intel Core i9-13900KF, con dos módulos de RAM combinados que suman 128 GB de memoria RAM, y una tarjeta de video NVIDIA GeForce RTX 4080 SUPER con 16 GB de memoria VRAM GDDR6X, con el framework PyTorch, el módulo Python OS para trabajar con los archivos del sistema, NumPy para operaciones matemáticas, Pandas para gestionar datos tabulares y PIL para trabajar con imágenes. Antes de seleccionar estas arquitecturas finales, se realizaron pruebas preliminares con distintas

Tabla 3. Resultados del modelo con fusión tardía.

Pérdida entrenam.	**Pérdida validación**	**Exactitud**	**Precisión**	**Sensibilid.**	**Especificidad**	**AUC**
0.05	0.03	0.99	0.99	0.99	0.97	1
0.05	0.05	0.98	0.98	0.99	0.95	0.99
0.04	0.03	0.99	0.99	0.99	0.97	1
0.05	0.03	0.99	0.99	0.99	0.98	1
0.05	0.02	0.99	0.99	0.99	0.98	1
0.05	0.02	0.99	0.99	0.99	0.98	1
0.04	0.07	0.97	0.97	0.99	0.93	0.99
0.04	0.05	0.98	0.98	0.99	0.96	1
0.04	0.03	0.99	0.99	1	0.97	1
0.05	0.03	0.99	0.99	1	0.98	1
0.05	0.03	0.99	0.99	1	0.97	1
0.04	0.04	0.98	0.98	0.99	0.96	1
0.05	0.02	0.99	1	0.99	0.99	1
0.04	0.03	0.99	0.99	1	0.98	1
0.04	0.03	0.99	0.99	1	0.97	1
0.04	0.02	0.99	0.99	1	0.98	1
0.04	0.03	0.99	0.99	0.99	0.97	1
0.04	0.02	0.99	0.99	0.99	0.98	1
0.04	0.03	0.99	0.99	1	0.97	1
0.05	0.02	0.99	0.99	0.99	0.98	1
0.06	0.03	0.99	0.99	1	0.96	1
0.04	0.03	0.99	0.99	1	0.96	1
0.04	0.05	0.98	0.99	0.99	0.96	1
0.04	0.05	0.98	0.99	0.99	0.97	1
0.05	0.02	0.99	0.99	0.99	0.98	1
0.04	0.03	0.99	0.99	0.99	0.98	1
0.05	0.04	0.98	0.98	1	0.96	1
0.05	0.03	0.99	0.99	0.99	0.97	1
0.04	0.04	0.98	0.98	0.99	0.95	1
0.05	0.03	0.99	0.99	1	0.97	1

configuraciones para identificar cuál ofrecía la mayor precisión, resultando seleccionadas las mencionadas por su superior desempeño. Para la evaluación de los experimentos, se consideraron las siguientes métricas: exactitud, precisión, sensibilidad, especificidad y área bajo la curva (AUC). Asimismo, se incluyeron la pérdida promedio durante el entrenamiento y la pérdida promedio en la fase de validación. Los resultados obtenidos con la arquitectura de fusión temprana se presentan en la Tabla 2, mientras que los correspondientes a la fusión tardía se muestran en la Tabla 3.

5.1. Análisis estadístico

En esta Sección se argumenta que la arquitectura de fusión temprana presenta una ventaja significativa en términos de rendimiento frente a la fusión tardía. Para evaluar si la diferencia en el rendimiento promedio es

Tabla 4. Resultados de precisión en 30 experimentos para fusión temprana y fusión tardía.

No. de experimento	Fusión Temprana (%)	Fusión Tardía (%)
1	0.99	0.99
2	0.99	0.98
3	0.99	0.99
4	0.99	0.99
5	0.99	0.99
6	0.99	0.99
7	0.99	0.97
8	0.99	0.98
9	0.99	0.99
10	0.99	0.99
11	0.99	0.99
12	0.99	0.98
13	0.99	0.99
14	0.99	0.99
15	0.99	0.99
16	0.99	0.99
17	0.99	0.99
18	0.99	0.99
19	0.99	0.99
20	0.99	0.99
21	0.99	0.99
22	0.99	0.99
23	0.99	0.98
24	0.99	0.98
25	0.99	0.99
26	0.99	0.99
27	0.99	0.98
28	0.99	0.99
29	0.99	0.98
30	0.99	0.99
Media (μ)	**0.990**	**0.987**
Desviación estándar (σ)	**0.000**	**0.0053**

estadísticamente significativa, se aplicó una prueba Z. La métrica utilizada en este análisis fue la precisión, cuyos resultados se detallan en la Tabla 4.

A partir de los resultados anteriores, se realizó una prueba estadística para evaluar si la diferencia entre ambas técnicas es significativa.

Hipótesis nula (H_0): No hay diferencia significativa entre los desempeños de la red de fusión temprana y la red de fusión tardía:

$$H_0 : \mu_1 = \mu_2.$$

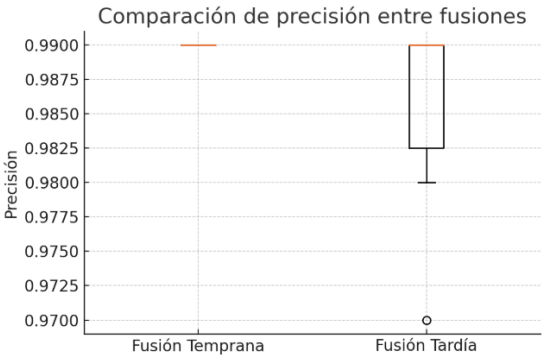

Fig. 4. Comparación de precisión entre fusión temprana y fusión tardía (Boxplot).

Hipótesis alternativa (H_a): Existe una diferencia significativa en el desempeño entre ambas redes:

$$H_a : \mu_1 \neq \mu_2.$$

Se utilizó una prueba Z para dos muestras independientes con los siguientes resultados:

- Estadístico Z: **3.071**,
- Valor p: **0.0021**.

Dado que el valor $p < 0{,}05$, se **rechaza la hipótesis nula**, indicando que la diferencia entre los modelos de fusión temprana y tardía es estadísticamente significativa.

En la Figura 4 se presenta una gráfica comparativa que ilustra la variabilidad y tendencia central de ambas técnicas.

Las redes neuronales multimodales logran obtener precisiones altas cuando se optimizan de manera correcta y se tienen datasets afinados y correctamente distribuidos. La fusión temprana nos da resultados consistentes; por otra parte, la fusión tardía permite mayor flexibilidad, aunque puede tener mayor variabilidad. A pesar de tener ventajas en rendimiento, también la complejidad técnica puede jugar en contra, ya que es necesario diseñar y distribuir los datos adecuadamente, puesto que no siempre estarán balanceados.

6. Conclusiones y trabajo futuro

Este estudio aborda el diagnóstico temprano de la enfermedad de Alzheimer con aprendizaje profundo multimodal, integrando dos tipos de datos, calificaciones clínicas de demencia (CDR) y resonancias magnéticas (MRI). Se evalúan dos tipos de arquitectura, la arquitectura de fusión temprana y la arquitectura de fusión tardía. La arquitectura de fusión temprana es más

lenta a la hora de entrenar, pero logró mejores resultados en la clasificación de la enfermedad de Alzheimer, ayudado por potente hardware y herramientas como PyTorch. En cuanto a la fusión tardía, se llevó menos tiempo de entrenamiento, pero su clasificación se vio reducida, lo cual nos deja con la fusión temprana como una opción viable para seguir trabajando y mejorando. La fusión de distintos tipos de datos cuidadosamente preprocesados logró resultados visiblemente buenos.

Esto es especialmente importante, ya que se estima que un gran número de personas mayores se verá afectado por enfermedades neurodegenerativas, y es crucial generar conciencia, ya que esto podría tener un impacto en muchas áreas de la sociedad. Las investigaciones futuras deberían integrar diversos tipos de datos, como imágenes médicas, información genética, registros clínicos e información de sensores, para mejorar la precisión de las predicciones y las capacidades diagnósticas en el ámbito de las enfermedades neurodegenerativas. Uno de los grandes desafíos del aprendizaje profundo es su naturaleza de caja negra.

El trabajo futuro debe centrarse en crear modelos más explicables que permitan a los profesionales médicos comprender y confiar mejor en los resultados. Además, fomentar una mayor colaboración entre científicos de datos, neurólogos y profesionales de la salud conducirá a modelos mejor alineados con las necesidades y prácticas clínicas. Finalmente, a medida que el aprendizaje profundo se vuelve más integral en la atención médica, abordar las preocupaciones éticas y garantizar la privacidad y seguridad de los datos será fundamental, especialmente al tratar con información médica sensible.

Agradecimientos. Agradecemos al TECNM/Instituto Tecnológico de Tijuana y a la SECIHTI por el apoyo financiero otorgado mediante el proyecto CF-2023-I-555.

Referencias

1. Anand, V. R., Priyan, S. T., Brahmam, M. G., Balusamy, B., Benedetto, F.: IMNMAGN: Integrative multimodal approach for enhanced detection of neurodegenerative diseases using fusion of multidomain analysis with graph networks. IEEE Access, (2024) doi: 10.1109/ACCESS.2024.3403860
2. Baltrušaitis, T., Ahuja, C., Morency, L.-P.: Multimodal machine learning: A survey and taxonomy. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 2, pp. 423–443 (2019) doi: 10.1109/TPAMI.2018.2798607
3. Candelise, N., Baiardi, S., Franceschini, A., Rossi, M., Parchi, P.: Towards an improved early diagnosis of neurodegenerative diseases: the emerging role of in vitro conversion assays for protein amyloids. Acta Neuropathologica Communications, vol. 8, no. 1, pp. 1–16 (2020) doi: 10.1186/s40478-020-00940-w
4. Chattopadhyay, T., Ozarkar, S. S., Buwa, K., Thomopoulos, S. I., Thompson, P. M.: Predicting brain amyloid positivity from T1 weighted brain MRI and MRI-derived gray matter, white matter and CSF maps using transfer learning on 3D CNNs. bioRxiv, (2023) doi: 10.1101/2023.02.15.528705
5. Diogo, V. S., Ferreira, H. A., Prata, D.: Early diagnosis of alzheimer's disease using machine learning: a multi-diagnostic, generalizable approach. Alzheimer's Research & Therapy, vol. 14, no. 107, pp. 1–15 (2022) doi: 10.1186/s13195-022-01047-y

6. Franciotti, R., Nardini, D., Russo, M., Onofrj, M., Sensi, S. L., Alzheimer's Disease Neuroimaging Initiative, Alzheimer's Disease Metabolomics Consortium: Comparison of machine learning-based approaches to predict the conversion to Alzheimer's disease from mild cognitive impairment. Neuroscience, vol. 514, pp. 143–152 (2023)
7. González Berrelleza, C. I.: Método de detección de bordes por medio de lógica difusa tipo-2 generalizada. Ph.D. thesis, Universidad Autónoma de Baja California (2016), tesis de doctorado, Facultad de Ciencias Químicas e Ingeniería, Tijuana, Baja California
8. Hajdu Macelaru, M., Chiuzbaian, R., Pop, P.: Machine learning approaches in the detection of amyotrophic lateral sclerosis disease using orofacial gestures (2024), manuscript
9. Huang, W.: Multimodal contrastive learning and tabular attention for automated Alzheimer's disease prediction. http://arxiv.org/abs/2308.15469v1 (2023)
10. LaMontagne, P. J., Benzinger, T. L. S., Morris, J. C., Keefe, S., Hornbeck, R., Xiong, C., Grant, E., Hassenstab, J., Moulder, K., Vlassenko, A., Raichle, M. E., Cruchaga, C., Marcus, D.: OASIS-3: Longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer disease. medRxiv, (2019) doi: 10.1101/2019.12.13.19014902
11. Li, J., Xu, H., Yu, H., Jiang, Z., Zhu, L.: Multi-modal feature selection with anchor graph for Alzheimer's disease. Frontiers in Neuroscience, vol. 16, pp. 1036244 (2022) doi: 10.3389/fnins.2022.1036244
12. Martí-Juan, G., Lorenzi, M., Piella, G.: Mc-rvae: Multi-channel recurrent variational autoencoder for multimodal Alzheimer's disease progression modelling. NeuroImage, vol. 268, pp. 119892 (2023) doi: 10.1016/j.neuroimage.2023.119892
13. McFall, G. P., Bohn, L., Gee, M., Drouin, S. M., Fah, H., Han, W., Dixon, R. A.: Identifying key multi-modal predictors of incipient dementia in Parkinson's disease: a machine learning analysis and tree SHAP interpretation. Frontiers in Aging Neuroscience, vol. 15, pp. 1124232 (2023) doi: 10.3389/fnagi.2023.1124232
14. Pan American Health Organization: Dementia in Latin America and the Caribbean: prevalence, incidence, impact, and trends over time. PAHO, Washington, DC (2023), https://doi.org/10.37774/9789275326657
15. Pao, P., Patnaik, D., Watson, L., Gao, F., Pan, L., Wang, J., Tsai, L.: HDAC1 modulates OGG1-initiated oxidative DNA damage repair in the aging brain and Alzheimer's disease. Nature Communications, vol. 11, no. 1, pp. 2484 (2020) doi: 10.1038/s41467-020-16298-6
16. Qiao, J., Wang, T., Shao, Z., Zhu, Y., Zhang, M., Huang, S., Zeng, P.: Genetic correlation and gene-based pleiotropy analysis for four major neurodegenerative diseases with summary statistics. Neurobiology of Aging, vol. 124, pp. 117–128 (2023) doi: 10.1016/j.neurobiolaging.2023.03.017
17. Romo-Galindo, D. A., Padilla-Moya, E.: Usefulness of brief cognitive tests for detecting dementia in the Mexican population. Archives of Neurosciences, vol. 23, no. 4, pp. 26–34 (2018)
18. Sarraf, S., DeSouza, D. D., Anderson, J., Tofighi, G.: Alzheimer's disease neuroimaging initiative. bioRxiv, (2017) doi: 10.1101/070441
19. Vong, W. K., Lake, B. M.: Cross-situational word learning with multimodal neural networks. Cognitive Science, vol. 46, no. 4, pp. e13122 (2021) doi: 10.1111/cogs.13122
20. Wadekar, S. N., Chaurasia, A., Chadha, A., Culurciello, E.: The evolution of multimodal model architectures. https://arxiv.org/abs/2405.17927 (2024)