

La investigación reportada en esta tesis es parte de los programas de investigación del CICESE (Centro de Investigación Científica y de Educación Superior de Ensenada, Baja California).

La investigación fue financiada por el SECIHTI (Secretaría de Ciencia, Humanidades, Tecnología e Innovación).

Todo el material contenido en esta tesis está protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México). El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo o titular de los Derechos de Autor.

**Centro de Investigación Científica y de Educación  
Superior de Ensenada, Baja California**



---

**Maestría en Ciencias  
en Ciencias de la Computación**

---

**Aprendizaje federado para la clasificación de enfermedades  
mediante imágenes de resonancia magnética**

Tesis  
para cubrir parcialmente los requisitos necesarios para obtener el grado de  
Maestra en Ciencias

Presenta:

**Ixchel Dafne Avendaño Pacheco**

Ensenada, Baja California, México  
2025

Tesis defendida por  
**Ixchel Dafne Avendaño Pacheco**

y aprobada por el siguiente Comité

**Dr. Andrey Chernykh**  
Codirector de tesis

**Dr. Luis Bernardo Pulido Gaytan**  
Codirector de tesis

**Dr. Jesús Favela Vara**

**Dr. Raúl Rivera Rodríguez**

**Dr. Gleb Radchenko**



**Dr. Pedro Gilberto López Mariscal**  
Coordinador del Posgrado en Ciencias de la  
Computación

**Dra. Ana Denise Re Araujo**  
Directora de Estudios de Posgrado

Resumen de la tesis que presenta **Ixchel Dafne Avendaño Pacheco** como requisito parcial para la obtención del grado de Maestra en Ciencias en Ciencias de la Computación.

**Aprendizaje federado para la clasificación de enfermedades mediante imágenes de resonancia magnética**

Resumen aprobado por:

**Dr. Andrey Chernykh**  
Codirector de tesis

**Dr. Luis Bernardo Pulido Gaytan**  
Codirector de tesis

El uso de redes neuronales profundas ha demostrado un alto rendimiento en tareas de clasificación de imágenes médicas, alcanzando niveles de precisión comparables a los de expertos clínicos. Sin embargo, su aplicación en entornos reales enfrenta importantes limitaciones, como la escasez de datos representativos, el desbalance entre clases y las restricciones legales y éticas para compartir información sensible. En este contexto, el aprendizaje federado se presenta como una alternativa viable, al permitir el entrenamiento colaborativo de modelos sin necesidad de centralizar los datos. Esta tesis propone el diseño, desarrollo y validación de un sistema federado optimizado para la clasificación multiclase de imágenes de resonancia magnética cerebral en la detección de los estadios de la enfermedad de Alzheimer. Se evaluaron siete arquitecturas convolucionales del estado del arte y se desarrolló una versión mejorada del modelo AHANet, denominada AHANet-C, que incorpora un bloque de super-convergencia para reducir los tiempos de entrenamiento sin pérdida de precisión. Asimismo, se integraron estrategias de aumento de datos y generación de imágenes sintéticas para enfrentar el desbalance de clases y mejorar la capacidad de generalización. El sistema se evaluó en dieciséis escenarios experimentales que consideraron variaciones en el número de clientes, la cantidad de rondas de comunicación y el uso de datos sintéticos. Los resultados superaron el 98% de precisión promedio por estadio, con un desempeño comparable al reportado en la literatura. Este trabajo demuestra la viabilidad de aplicar aprendizaje federado en el diagnóstico automatizado de Alzheimer, preservando la privacidad de los datos clínicos y manteniendo un rendimiento robusto y eficiente.

**Palabras clave:** aprendizaje federado, clasificación de imágenes médicas, redes neuronales convolucionales, imágenes de resonancia magnética.

Abstract of the thesis presented by **Ixchel Dafne Avendaño Pacheco** as a partial requirement to obtain the Master of Science degree in Computer Science.

### **Federated Learning for Disease Classification Using Magnetic Resonance Imaging**

Abstract approved by:

**Dr. Andrey Chernykh**  
Codirector of thesis

**Dr. Luis Bernardo Pulido Gaytan**  
Codirector of thesis

Deep neural networks have shown high performance in medical image classification tasks, reaching accuracy levels comparable to those of clinical experts. However, their application in real-world settings faces significant limitations, such as the scarcity of representative data, class imbalance, and legal and ethical constraints related to the sharing of sensitive information. In this context, federated learning emerges as a viable alternative, enabling the collaborative training of models without centralizing data. This thesis presents the design, development, and validation of an optimized federated system for multiclass classification of brain magnetic resonance images in the detection of Alzheimer's disease stages. Seven state-of-the-art convolutional architectures were evaluated, and an enhanced version of the AHANet model, named AHANet-C, was developed by incorporating a superconvergence block to reduce training time without compromising accuracy. In addition, data augmentation and synthetic image generation strategies were integrated to address class imbalance and improve the model's generalization capacity. The system was evaluated in sixteen experimental scenarios that varied the number of clients, communication rounds, and use of synthetic data. The best results exceeded 98% average accuracy per stage, achieving performance comparable to that reported in the literature. This work demonstrates the feasibility of applying federated learning to automated Alzheimer's diagnosis, preserving clinical data privacy while ensuring robust and efficient performance.

**Keywords:** federated learning, medical image classification, convolutional neural networks, magnetic resonance imaging.

## Dedicatoria

A mi compañero de vida Andres Calzada

## Agradecimientos

En primer lugar, quiero expresar mi sincera gratitud a mis directores de tesis, el Dr. Andrey Chernykh y el Dr. Luis Bernardo Pulido Gaytan por su invaluable orientación y apoyo como directores de tesis.

A los miembros de mi comité de tesis, Dr. Jesús Favela Vara, Dr. Raúl Rivera Rodríguez, Dr. Gleb Radchenko, gracias por el tiempo y dedicación que esto representa.

Agradezco al Centro de Investigación Científica y de Educación Superior de Ensenada, Baja California por concederme el privilegio de formar parte de tan distinguida institución.

Agradezco a la Secretaría de Ciencia, Humanidades, Tecnología e Innovación por el financiamiento otorgado para la realización de esta maestría bajo la modalidad de beca.

Agradezco a los Pelícanos viajeros por permitirme recorrer senderos a su lado y por las invaluable vistas que me otorgaron de Baja California.

Finalmente, quiero expresar mi más sincero y profundo agradecimiento a Paty, Javier, Jazmin, Denise, Paola y Valentina por ser mi isla de paz y amor. Los quiero infinitamente.

# Tabla de contenido

	Página
Resumen en español.....	ii
Resumen en inglés.....	iii
Dedicatoria .....	iv
Agradecimientos.....	v
Lista de figuras.....	ix
Lista de tablas .....	x
<b>Capítulo 1. Introducción .....</b>	<b>1</b>
1.1 Objetivos .....	2
1.1.1 Objetivo general .....	2
1.1.2 Objetivos específicos.....	2
<b>Capítulo 2. Antecedentes .....</b>	<b>4</b>
2.1 Redes neuronales convolucionales .....	4
2.1.1 Optimización de redes neuronales convolucionales.....	9
2.2 Super-convergencia y esquemas cíclicos de tasa de aprendizaje .....	10
2.3 Aprendizaje federado.....	11
2.3.1 Tipos de aprendizaje federado.....	13
2.3.1.1 Aprendizaje federado horizontal .....	14
2.3.1.2 Aprendizaje federado vertical .....	15
2.3.1.3 Aprendizaje federado transversal .....	16
2.3.2 Algoritmos de agregación .....	17
2.3.2.1 FedAvg .....	17
2.3.2.2 FedAdam .....	18
2.3.2.3 FedYogi .....	19
2.3.2.4 FedProx.....	19

2.4	Aumento de datos y generación de imágenes sintéticas.....	20
2.4.1	Aumento de datos.....	20
2.4.2	Generación de imágenes sintéticas.....	21
2.5	Enfermedad de Alzheimer.....	23
2.5.1	Patología y síntomas .....	23
2.5.2	Factores de riesgo .....	24
2.5.3	Importancia de la detección temprana.....	24
2.5.4	Biomarcadores .....	24
2.6	Imágenes de resonancia magnética para diagnóstico .....	25
2.7	Bases de datos.....	27
<b>Capítulo 3.</b>	<b>CNN y aprendizaje federado para la clasificación de MRI en la EA.....</b>	<b>31</b>
3.1	Redes neuronales convolucionales para clasificación de MRI en la EA .....	31
3.2	Aprendizaje federado en enfermedades neurodegenerativas .....	33
3.3	Uso de datos sintéticos para balanceo de clases .....	35
<b>Capítulo 4.</b>	<b>Metodología .....</b>	<b>37</b>
4.1	Conjuntos de datos .....	38
4.1.1	Preprocesamiento y normalización.....	38
4.1.2	Aumento de datos y generación sintética.....	39
4.2	Descripción de modelos de CNN .....	42
4.2.1	Hiperparámetros .....	42
4.2.2	Modelos evaluados .....	43
4.2.3	Optimización de AHA Net con bloque de super-convergencia.....	48
4.3	Configuración del AF .....	49
4.3.1	Distribución de los datos entre clientes.....	49
4.3.2	Configuración del entrenamiento federado .....	51
4.3.3	Uso de datos sintéticos .....	53
4.4	Métricas de evaluación .....	54

4.5	Configuración Experimental.....	56
<b>Capítulo 5.</b>	<b>Resultados.....</b>	<b>58</b>
5.1	Comparación de arquitecturas CNN en entorno centralizado.....	58
5.1.1	Evaluación del modelo AHANet-C en entorno centralizado .....	62
5.1.2	Experimentos con aumento de datos y generación de imágenes sintéticas en entorno centralizado.....	63
5.2	Comparación de los diferentes escenarios de AF .....	66
5.2.1	Resultados globales por escenario.....	66
5.2.1.1	Análisis multiobjetivo: error y tiempo de ejecución .....	67
5.2.2	Análisis comparativo por factor experimental.....	69
5.2.3	Comparación por estadio .....	71
5.3	Análisis estadístico del efecto de los factores experimentales.....	72
<b>Capítulo 6.</b>	<b>Discusión .....</b>	<b>74</b>
6.1	Análisis comparativo con modelos de CNNs del estado del arte.....	74
6.2	Comparación con estudios en la literatura sobre escenarios de AF .....	76
6.3	Limitaciones generales del estudio .....	78
<b>Capítulo 7.</b>	<b>Conclusiones y trabajo futuro.....</b>	<b>80</b>
	<b>Literatura citada .....</b>	<b>82</b>
	<b>Acrónimos.....</b>	<b>89</b>
	<b>Anexos.....</b>	<b>91</b>

## Lista de figuras

Figura	Página
1. Esquema general de aprendizaje federado horizontal. ....	14
2. Esquema general de aprendizaje federado vertical.....	16
3. Esquema general de aprendizaje federado transversal.....	17
4. Transformaciones de imágenes para aumento de datos.....	21
5. Etapas de biomarcadores de la EA. Imagen recuperada del ADNI.....	25
6. Representación general de la metodología propuesta.....	37
7. Arquitectura general del modelo Pix2Pix.....	41
8. Arquitectura AlexNet.....	44
9. Arquitectura VGGNet.....	45
10. Arquitectura DenseNet.....	45
11. Arquitectura empleada en AbdulAzeem et al. (2021).....	46
12. Arquitectura empleada en Arafa et al. (2024).....	46
13. Arquitectura empleada en AHANet. (Illakiya et al., 2023).....	47
14. Arquitectura empleada en DEMNet. (Murugan et al., 2021).....	47
15. Arquitectura AHANet-C propuesta con esquema de super-convergencia.....	49
16. Escenarios A y B de análisis experimental.....	50
17. Organización del conjunto de datos.....	50
18. Curvas de aprendizaje de los modelos AHANet y AHANet-C.....	62
19. Frente aproximado de Pareto para los 16 experimentos en el entorno federado.....	68

## Lista de tablas

Tabla	Página
1. Comparativa de bases de datos de MRI para la enfermedad de Alzheimer.....	29
2. Características de las imágenes del conjunto de datos.....	38
3. División de datos para los conjuntos de entrenamiento, validación y prueba para cada uno de los estadios.....	38
4. Aumento de imágenes mediante transformaciones geométricas en el conjunto de entrenamiento para cada uno de los estadios.....	40
5. Aumento de imágenes mediante imágenes sintéticas generadas mediante GAN en el conjunto de entrenamiento para cada uno de los estadios.....	40
6. Cantidad de imágenes sintéticas generadas e incorporadas para cada estadio.....	42
7. Hiperparámetros empleados en los modelos de redes neuronales convolucionales.....	43
8. Modelos de redes neuronales convolucionales comparados.....	43
9. Relación entre el número de rondas de comunicación y el número de épocas.....	52
10. Configuración de los experimentos ejecutados en el entorno federado.....	54
11. Resultados globales de las siete arquitecturas de CNN evaluadas en el entorno centralizado.....	58
12. Resultados por estadio de las siete arquitecturas de redes neuronales convolucionales evaluadas en el entorno centralizado.....	61
13. Comparación de resultados globales entre AHANet y AHANet-C.....	62
14. Comparación por estadio entre AHANet y AHANet-C.....	63
15. Comparación global de AHANet-C bajo distintas estrategias de balanceo.....	64
16. Desempeño por estadio de AHANet-C bajo distintas estrategias de balanceo.....	65
17. Precisión global y tiempo de entrenamiento en los 16 escenarios federados.....	67
18. Promedio por número de rondas.....	69
19. Promedios por número de clientes.....	70
20. Promedios por uso de imágenes sintéticas.....	70
21. Comparación por estadio de los 16 escenarios de AF.....	71

22. Resultados del análisis de varianza multifactorial sobre la variable Precisión. ....	73
23. Resultados reportados en la literatura para los modelos de CNN evaluadas.....	75
24. Comparación con estudios de la literatura sobre aprendizaje federado para diagnóstico de Alzheimer.....	77
25. Acrónimos.....	89
26. Funciones de activación. ....	91

## Capítulo 1. Introducción

---

El análisis y procesamiento automatizado de imágenes médicas mediante Aprendizaje Profundo (DL, por sus siglas en inglés) ha transformado el desarrollo de herramientas de apoyo al diagnóstico en diversas especialidades, al permitir la identificación de patrones complejos que pueden pasar inadvertidos al ojo humano. Las redes neuronales profundas, y en particular las Redes Neuronales Convolucionales (CNN, por sus siglas en inglés), han alcanzado un rendimiento superior al 90% de precisión en tareas de clasificación de imágenes médicas, alcanzando niveles de precisión cercanos a los de los expertos clínicos en ciertas aplicaciones (Chen et al., 2025).

Sin embargo, la implementación de estos modelos en entornos clínicos reales enfrenta importantes desafíos técnicos y éticos. Uno de los principales obstáculos es la necesidad de grandes volúmenes de datos etiquetados y representativos para entrenar modelos robustos. En la práctica, los conjuntos de datos médicos suelen ser desbalanceados, ya que las patologías de interés, especialmente en estadios tardíos, están poco representadas. Este desbalance limita la capacidad de generalización de los modelos y puede llevar a resultados sesgados.

Otro desafío crucial es el manejo de los datos sensibles. La transferencia y centralización de imágenes médicas para entrenamiento de CNNs plantea riesgos relacionados con la confidencialidad y la privacidad de la información, además de contravenir en muchos casos las normativas legales vigentes. A esto se suman las restricciones prácticas para compartir datos entre instituciones debido a políticas internas y consideraciones éticas.

De esta forma, el Aprendizaje Federado (AF) se ha posicionado como una posible solución que permite entrenar modelos colaborativos sin necesidad de centralizar los datos. Este enfoque preserva la privacidad de la información y facilita la cooperación entre diferentes instituciones o dispositivos. No obstante, el diseño de sistemas federados robustos plantea retos técnicos adicionales, como la sincronización eficiente entre clientes, la gestión del desbalance de datos entre nodos, la heterogeneidad de los dispositivos y redes participantes, y el incremento en los costos computacionales y de comunicación asociados a la agregación de modelos (Antunes et al., 2022).

Esta tesis tiene como objetivo desarrollar y validar un modelo de AF para la clasificación de imágenes médicas, optimizado para garantizar la preservación de la privacidad, la reducción de los tiempos de

entrenamiento y el equilibrio entre precisión y costo computacional. En particular, se aborda la clasificación multiclase de Imágenes de Resonancia Magnética (MRI, por sus siglas en inglés) para la detección de los distintos estadios de la Enfermedad de Alzheimer (EA) dada su relevancia clínica. EA representa la demencia con mayor prevalencia a nivel mundial (World Health Organization, 2025). En el caso particular de México, se estima que entre un 60 y 70 por ciento de los diagnósticos de demencia corresponden a EA (Secretaría de Salud, 2021). Es por ello por lo que se plantean los siguientes objetivos:

## **1.1 Objetivos**

### **1.1.1 Objetivo general**

Desarrollar y validar un modelo de aprendizaje federado para la clasificación de imágenes de resonancia magnética en la detección de los diferentes estadios de la enfermedad de Alzheimer, garantizando la preservación de la privacidad de los datos y reduciendo el tiempo de entrenamiento.

### **1.1.2 Objetivos específicos**

- Identificar, proponer y evaluar diferentes arquitecturas para la clasificación de la enfermedad de Alzheimer empleando imágenes de resonancia magnética para modelos de redes neuronales profundas y aprendizaje federado.
- Comparar el tiempo de entrenamiento y precisión del modelo de aprendizaje federado desarrollado, comparándolo con modelos de redes neuronales profundas no federados para determinar su rendimiento en la clasificación de imágenes de resonancia magnética en diferentes estadios de enfermedad de Alzheimer.

Se evaluaron y compararon siete arquitecturas CNN reconocidas en el estado del arte, con el fin de identificar el modelo base más adecuado para el problema. Posteriormente, se desarrolló una versión optimizada del modelo seleccionado mediante la incorporación de un bloque de super-convergencia, destinado a acelerar el entrenamiento sin pérdida de precisión. Además, se exploraron estrategias de aumento de datos y generación de imágenes sintéticas para abordar el desbalance de clases y fortalecer

la capacidad de generalización del sistema. Además, se diseñó un entorno de AF simulado que permitió evaluar el desempeño del modelo en diferentes escenarios, variando el número de clientes, las rondas de comunicación y el uso de imágenes sintéticas. Finalmente, se compararon los resultados con los reportados en la literatura.

La tesis se estructura de la siguiente manera: el Capítulo 2. presenta los fundamentos teóricos relacionados con el aprendizaje profundo, redes convolucionales, super-convergencia, aprendizaje federado y generación de datos sintéticos. El Capítulo 3. ofrece una revisión del estado del arte en clasificación de imágenes médicas, aprendizaje federado en la EA. El Capítulo 4. describe la metodología implementada, los conjuntos de datos, las arquitecturas evaluadas y la configuración de los experimentos. El Capítulo 5. presenta los resultados obtenidos. El Capítulo 6. discute y analiza los resultados obtenidos en el contexto de la literatura. Finalmente, el Capítulo 7. expone las conclusiones y las principales líneas de trabajo futuro.

## Capítulo 2. Antecedentes

---

En este capítulo se presenta los fundamentos y enfoques utilizados en esta tesis. En primer lugar, se describen las CNNs. Posteriormente, se aborda el concepto de super-convergencia y los esquemas cíclicos de tasa de aprendizaje, como estrategias para optimizar el entrenamiento de modelos profundos. A continuación, se introduce el aprendizaje federado, detallando sus principales tipos; el horizontal, vertical y transversal, así como los principales algoritmos de agregación. Posteriormente, se revisan las técnicas de aumento de datos y generación de imágenes sintéticas, que contribuyen a mejorar el rendimiento de los modelos cuando el volumen de datos es limitado. Asimismo, se presentan aspectos clave de la enfermedad de Alzheimer, incluyendo su patología, síntomas, factores de riesgo, importancia de la detección temprana y biomarcadores asociados. Finalmente, se describen las bases de datos empleadas para el diagnóstico de la EA.

### 2.1 Redes neuronales convolucionales

Las Redes Neuronales Artificiales (ANN, por sus siglas en inglés) se inspiran en la función y estructura de las redes neuronales biológicas, además, su unidad fundamental es la neurona artificial, un nodo que establece conexiones ponderadas con otras neuronas dentro de la red. En (2016), Warren McCulloch y Walter Pitts, demostraron que una red de neuronas artificiales podía reproducir cualquier proposición lógica. Con el tiempo, el concepto se refinó hasta derivar en arquitecturas profundas capaces de procesar millones de parámetros. Sin embargo, su aplicación directa a imágenes planteaba un obstáculo práctico: el número de pesos crecía con el producto del tamaño de la entrada y el de las neuronas de la primera capa. La solución llegó con las CNNs, que reemplazan la conexión densa por pequeños filtros compartidos y, con ello, preservan la estructura espacial de la imagen al tiempo que contenían la complejidad del modelo. La eficacia de ese enfoque quedó demostrada en 2012, cuando AlexNet redujo a casi a la mitad el error de clasificación en el concurso ImageNet y marcó el renacimiento del DL alimentado por Unidad de Procesamiento Gráfico (GPU, por sus siglas en inglés) (Krizhevsky et al., 2012).

Una ANN hace referencia a un grafo dirigido cuyos nodos, es decir las neuronas artificiales, realizan transformaciones sobre los datos y transmiten el resultado por las artistas ponderadas. En 1958, Fran Resenblatt propone el *Perceptron*, el primer modelo entrenable capaz de ajustar sus pesos para clasificar dos clases, siempre que estas presentaran una frontera lineal, lo que presentó un reto para funciones no

lineales como la función XOR. Fue así como se desarrolló el Perceptrón Multicapa (MLP, por sus siglas en inglés) el cual intercala capas ocultas; dicho concepto fue respaldado con la llegada del teorema de aproximación universal propuesto por Hornik (1989) donde demostraron que con una sola capa oculta y activaciones adecuadas se puede aproximar cualquier función continua.

Cada neurona artificial realiza una combinación lineal de sus entradas,  $z = \mathbf{w}^T \mathbf{x} + b$ , después aplica una función de activación para introducir no linealidad, es decir,  $y = f(z)$ . Las activaciones clásicas — sigmoide, que comprime a  $(0, 1)$ , y la tangente hiperbólica, que oscila en  $(-1, 1)$ — dominaron los primeros sistemas, pero tienden a saturar el gradiente cuando  $|z|$  es grande. El problema se mitigó con la Unidad Lineal Rectificada (ReLU, por sus siglas en inglés), la cual mantiene un gradiente constante para valores positivos, es trivial de calcular y, por ello, favorece el entrenamiento de redes profundas. Variantes como Leaky ReLU, ELU o Softplus modifican la región negativa del dominio para prevenir neuronas inactivas o suavizar la transición, pero siempre con el mismo objetivo: romper la linealidad sin encarecer el costo computacional. En el Anexo se muestran las funciones de activación más comunes empleadas en ANN.

El proceso de entrenamiento de una ANN implica tres fases aplicadas iterativamente. Primero, la propagación directa, donde la red transforma la entrada en una predicción  $\hat{y}$ . Después, una función de pérdida cuantifica el error entre la predicción y la etiqueta verdadera; para clasificación suele emplearse la entropía cruzada, mientras que para regresión es común el error cuadrático medio. Por último, la retro-propagación calcula el gradiente de la pérdida respecto a cada peso y un optimizador actualiza los parámetros en la dirección que reduce ese error. Con suficientes datos y regularización adecuada, este ciclo permite que la red generalice más allá de los ejemplos vistos.

Un aspecto importante en el entrenamiento de ANN es la inicialización de pesos. Este puede ser desde un enfoque arbitrario como asignar valores aleatorios en todas las capas, hasta el uso de heurísticas. Xavier Glorot y Yoshua Bengio (2010) proponen una heurística que iguala la varianza de las activaciones a la de los gradientes, manteniendo un flujo de información estable en redes moderadamente profundas. Sin embargo, aun con una buena inicialización las activaciones cambian de escala durante el entrenamiento, fenómeno conocido como *desvío covariante interno*. Para combatirlo, Sergey Ioffe y Christian Szegedy (2015) introdujeron la Normalización por Lotes (BN, por sus siglas en inglés): cada mini-lote se redimensiona para tener media cero y varianza unitaria antes de pasar por la no linealidad, lo que permite usar tasas de aprendizaje mayores y reducir la sensibilidad a la inicialización. La normalización se ha generalizado: cuando el tamaño del lote es pequeño, se prefiere *normalización por capas*, y cuando el

objetivo es estilizar imágenes o trabajar con lotes de tamaño uno, *normalización por instancia* ha demostrado producir representaciones más robustas al contraste local.

Otro aspecto importante son los métodos de regularización. Nitish Srivastava et al. (2014) proponen *Dropout*, una técnica que consiste en apagar aleatoriamente un porcentaje de neuronas durante el entrenamiento. El efecto es análogo al de entrenar y promediar un gran número de redes más pequeñas, lo que reduce de manera drástica el sobreajuste sin multiplicar el coste de cálculo.

La convolución es el operador lineal que distingue a las CNN de los MLP. Al deslizar un pequeño filtro sobre la imagen, la red detecta patrones locales con pesos compartidos, lo que asegura equivarianza traslacional: un borde vertical se reconoce, aunque aparezca en posiciones distintas. Después de cada convolución se aplica la función de activación y posteriormente, una operación de agrupamiento que reduce la resolución para ampliar el campo receptivo y conferir cierta invariancia a traslaciones pequeñas.

Las CNN constituyen una aportación importante dentro de la visión por computadora moderna. Nacieron en 1980 del *Neocognitron* de Kunihiko Fukushima (1980), un modelo jerárquico que emplea filtros locales y comparte pesos para reconocer caracteres independientemente de su posición. Aquel concepto, perfeccionado por Yann LeCun (1998) en LeNet-5, el modelo que hizo viable la lectura automática de dígitos manuscritos, demostró que una estrategia basada en convoluciones podía aprender características útiles con muchos menos parámetros que un MLP denso. En lugar de conectar cada píxel con cada neurona de la siguiente capa, un filtro pequeño se desliza sobre la imagen. En cada posición se calcula un producto punto entre los valores del filtro y los de la región cubierta, produciendo un número que forma parte de un mapa de activaciones. Como el mismo filtro se aplica en toda la imagen, los pesos se comparten y la red aprende a detectar patrones locales; bordes, texturas, esquinas, sin importar dónde aparezcan. Esta propiedad, denominada equivarianza traslacional, reduce drásticamente el número de parámetros y refuerza la capacidad de generalizar.

Una CNN puede entenderse como una función diferenciable  $f(x; \theta)$  que transforma un tensor de entrada  $x \in \mathbb{R}^{H \times W \times C}$  en un vector de salidas: probabilidades, mapas de segmentación o características latentes mediante la composición de capas convolucionales, capas de agrupamiento y capas completamente conectadas. Cada capa aplica una convolución:

$$y_k(u, v) = b_k + \sum_{c=1}^{c_{in}} \sum_{i=0}^{k_H-1} \sum_{j=0}^{k_W-1} W_c^{(k)}(i, j) x_c(u + i, v + j) \quad (1)$$

donde  $y_k$  describe cómo se obtiene cada valor  $(u, v)$  del mapa de características  $k$  que genera una capa convolucional. Para cada posición de salida, el filtro  $k$  recorre todos los canales de la entrada  $y$ , dentro de cada canal, superpone su filtro de tamaño  $k_H \times k_W$ . En esa superposición multiplica cada peso  $W_c^{(k)}(i, j)$  por el valor del píxel correspondiente  $x_c(u + i, v + j)$  luego suma todos esos productos  $y$ , finalmente, añade el sesgo  $b_k$ . Así, la capa calcula una combinación lineal ponderada de los píxeles que quedan dentro del filtro en cada desplazamiento, produciendo un nuevo mapa que resalta patrones específicos, como bordes o texturas, aprendidos durante el entrenamiento.

En ANN, existen ciertos valores que deben establecerse previamente al proceso de entrenamiento, ya que no se aprenden automáticamente a partir de los datos. A estos valores se les denomina hiperparámetros. Su elección influye directamente en el comportamiento, capacidad de generalización y eficiencia del modelo. Dos hiperparámetros que determinan el comportamiento espacial de las capas convolucionales son el desplazamiento y el relleno, los cuales influyen directamente en el tamaño de la salida y en la cantidad de información conservada de la entrada.

El desplazamiento controla el número de píxeles que el filtro se desplaza en cada paso durante la convolución. Un desplazamiento de 1 significa que el filtro se desliza de forma adyacente, píxel por píxel, lo que produce un mapa de activaciones de tamaño similar al de la entrada, dependiendo del relleno. Aumentar el desplazamiento provoca una reducción más agresiva del tamaño de la salida, ya que el filtro se mueve en pasos más amplios, lo cual implica una forma de submuestreo implícito. Esto permite controlar la resolución de las representaciones intermedias y reducir el costo computacional, a costa de perder detalle espacial.

Por otro lado, el relleno consiste en añadir valores (generalmente ceros) alrededor de los bordes de la imagen antes de aplicar la convolución. Sin relleno, los filtros no pueden cubrir completamente los píxeles cercanos a los bordes, lo que hace que el tamaño de la salida disminuya con cada capa convolucional. Esta situación puede ser indeseable cuando se desea preservar el tamaño espacial o cuando los bordes contienen información relevante. Existen diferentes estrategias, entre las cuales destacan:

1. Relleno válido: no se añade ningún relleno, por lo que el mapa de salida es más pequeño que la entrada.
2. Relleno mismo: se añade la cantidad justa de ceros para que la salida conserve el mismo tamaño espacial que la entrada, cuando el desplazamiento es 1.

Estos hiperparámetros, junto con el tamaño del filtro, determinan el tamaño de salida de cada capa convolucional. La dimensión espacial de salida puede calcularse como:

$$\text{Salida} = \left\lfloor \frac{N + 2P - F}{S} \right\rfloor + 1 \quad (2)$$

donde  $N$  es el tamaño de la entrada,  $F$  el tamaño del filtro,  $P$  el relleno aplicado y  $S$  el desplazamiento.

La elección adecuada de desplazamiento y relleno es crucial para diseñar arquitecturas eficientes y mantener un equilibrio entre resolución, campo receptivo y costo computacional. La capa de agrupación es una operación de submuestreo que suele aplicarse después de las capas convolucionales y contribuye a dotar al modelo de invarianza espacial, es decir, la capacidad de reconocer patrones independientemente de pequeñas traslaciones o distorsiones. Las dos variantes más utilizadas son el agrupamiento máximo, que selecciona el valor máximo dentro de una ventana deslizante, y el agrupamiento promedio, que calcula el promedio de los valores dentro de esa ventana.

El uso del agrupamiento máximo tiende a conservar las características más relevantes, aquellas con mayor activación, mientras que el promedio suaviza la salida y puede resultar útil para tareas donde se desea una representación más global. Entre sus ventajas, destacan la reducción del número de parámetros y del costo computacional, así como la mitigación del sobreajuste al disminuir la sensibilidad a pequeñas variaciones en la entrada. Como desventaja, el agrupamiento implica una pérdida de información, ya que descarta detalles finos que podrían ser relevantes para ciertas tareas.

Por otro lado, las capas completamente conectadas reciben como entrada un vector aplanado y conectan cada neurona con todas las neuronas de la capa siguiente. Su función principal es combinar las características extraídas por las capas convolucionales y realizar la clasificación o regresión final. Si bien son efectivas para la toma de decisiones, su principal desventaja es el elevado número de parámetros que

introducen, lo que puede incrementar el riesgo de sobreajuste y el costo computacional en redes profundas.

### 2.1.1 Optimización de redes neuronales convolucionales

En el contexto del DL, los optimizadores desempeñan un papel esencial durante el entrenamiento de las CNN, ya que son los encargados de ajustar los hiperparámetros y pesos sinápticos, a fin de minimizar el error en las predicciones. A lo largo del tiempo, se han desarrollado diversos algoritmos de optimización que buscan mejorar la eficiencia, la estabilidad y la velocidad de convergencia durante el aprendizaje, especialmente cuando se trabaja con grandes volúmenes de datos.

Uno de los métodos ampliamente utilizados es el Descenso por Gradiente Estocástico, (SGD, por sus siglas en inglés). Este algoritmo actualiza los pesos del modelo utilizando gradientes calculados a partir de pequeños subconjuntos del conjunto de entrenamiento (mini-lotes), lo que permite que el proceso de entrenamiento sea más rápido y menos costoso en términos de memoria. Sin embargo, una de sus principales limitaciones es que puede presentar oscilaciones o una convergencia lenta, especialmente en funciones de pérdida con múltiples mínimos locales o valles muy estrechos (Rumelhart et al., 1986).

Para superar algunas de estas limitaciones, se han propuesto variantes que incorporan información adicional sobre la trayectoria del gradiente. Una de las más destacadas es SGD con *momentum*, la cual añade un término que considera el historial de actualizaciones anteriores, permitiendo que el optimizador avance con mayor firmeza en direcciones consistentes y se desacelere en aquellas con cambios abruptos. Esta técnica ayuda a reducir las fluctuaciones y mejora significativamente la velocidad de convergencia en superficies de error complejas.

Otros optimizadores como Adagrad introducen mecanismos adaptativos que modifican la LR individualmente para cada peso, lo cual resulta útil cuando se trabaja con datos dispersos o cuando algunas características del conjunto de datos son más frecuentes que otras. No obstante, Adagrad tiende a reducir en exceso la LR conforme avanza el entrenamiento, lo que puede dificultar el ajuste fino del modelo (Duchi et al., 2011).

Para abordar esta limitación, surgieron métodos como RMSprop cuyo nombre proviene de *Root Mean Square Propagation*, debido a su estrategia de actualizar los pesos según la media móvil de los gradientes

al cuadrado. Este algoritmo mantiene una media móvil de los cuadrados de los gradientes para normalizar las actualizaciones. Este enfoque permite controlar mejor la magnitud de los pasos y ha demostrado ser particularmente efectivo en tareas con datos no estacionarios o en redes neuronales recurrentes.

Uno de los optimizadores más populares y robustos en la actualidad es *Adaptive Moment Estimation (Adam)*. Este método combina las ideas de *momentum* y RMSProp, estimando tanto la media como la varianza de los gradientes para ajustar de manera eficiente la LR de cada parámetro. Gracias a su capacidad de adaptación y su buen desempeño en una amplia variedad de tareas, Adam se ha convertido en el optimizador predeterminado en muchas aplicaciones prácticas (Kingma & Welling, 2013).

Existen también variantes como Adadelta, que ajusta dinámicamente el tamaño de los pasos de actualización sin necesidad de especificar una LR inicial (Zeiler, 2012), y AdamW, que mejora a Adam separando explícitamente la regularización del proceso de actualización de los pesos, lo que resulta especialmente beneficioso en redes profundas (Loshchilov & Hutter, 2017).

En conjunto, estos métodos de optimización no solo permiten que los modelos aprendan con mayor precisión, sino que también influyen en la estabilidad, el tiempo de entrenamiento y la capacidad general de la red para generalizar a datos no vistos. La elección del optimizador adecuado depende de múltiples factores, como la arquitectura de la red, el tipo de datos, el tamaño del conjunto de entrenamiento y los recursos computacionales disponibles. Por tanto, comprender las características y ventajas de cada optimizador es fundamental para diseñar estrategias de entrenamiento efectivas en el desarrollo de modelos basados en CNN.

## 2.2 Super-convergencia y esquemas cíclicos de tasa de aprendizaje

La Tasa de Aprendizaje (LR, por sus siglas en inglés) es el hiperparámetro que regula la magnitud de cada actualización de pesos durante el Descenso de Gradiente (GD, por sus siglas en inglés). Un valor menor ralentiza el entrenamiento, mientras que un valor alto provoca inestabilidad y potencialmente divergencia. Tradicionalmente se recurre a un valor inicial pequeño que se reduce de forma gradual. No obstante, de acuerdo con Smith et al. (2017), si la LR se incrementa inicialmente hasta alcanzar valores elevados y posteriormente se reduce, la red no solo logra converger, sino que lo hace entre cinco y diez veces más rápido, fenómeno conocido como super-convergencia.

La clave reside en la relación entre LR y regularización. Una LR máxima inusualmente alta actúa como regularizador fuerte, empujando la solución lejos de mínimos estrechos y reduciendo la necesidad de otras técnicas como la penalización de pesos. Cuando la LR desciende, el optimizador aterriza en una cuenca ancha del paisaje de la pérdida, favoreciendo la generalización. Smith et al. (2018) demostró que, mediante una prueba preliminar, es posible explorar la tasa de aprendizaje desde valores muy bajos hasta muy altos en unas pocas iteraciones para estimar el rango en el que la pérdida disminuye; este rango define posteriormente los límites del ciclo. El autor formalizó este procedimiento dentro del enfoque de Tasas de Aprendizaje Cíclicas (CLR, por sus siglas en inglés), proponiendo perfiles triangulares como estrategia de ajuste dinámico. Este método permite acelerar la convergencia y facilita la búsqueda de hiperparámetros al evitar un ajuste manual extenso.

La variante más utilizada en la práctica es la política de un ciclo, que combina un único ciclo LR ascendente-descendente con un ciclo inverso de *momentum*, alto-bajo-alto. El ascenso rápido en LR favorece la exploración del espacio de parámetros; el descenso lento afina la solución, y el pico final de *momentum* estabiliza la trayectoria. Esta política proporciona un protocolo sencillo: elegir una LR mínima diez veces menor y una LR máxima de tres a diez veces mayor que la LR tradicional, entrenar durante un solo ciclo que abarque todas las épocas.

Super-convergencia no es un atajo universal. Si la LR máxima se elige demasiado alta, la pérdida puede explotar antes de que el ciclo descendente actúe; si se impone una penalización de pesos excesivo, el doble efecto regularizador puede infra ajustar la red. Además, en nodos federados de muy pequeño lote la oscilación de LR debe sincronizarse con el esquema de agregación para evitar que varios clientes aterricen en regiones dispares del espacio de parámetros.

## 2.3 Aprendizaje federado

El AF es una estrategia de entrenamiento distribuido que permite construir un modelo global sin centralizar los datos sensibles. En lugar de enviar los datos a un servidor, cada institución o dispositivo entrena localmente una copia del modelo y comparte únicamente los parámetros actualizados. Un servidor agrega esas actualizaciones y envía de vuelta un modelo mejorado. Al repetir el ciclo se obtiene una red que ha aprendido de todo el colectivo sin que ninguna institución haya expuesto su información. El concepto fue introducido por McMahan et al. (2016), favoreciendo la preservación de la privacidad y

disminuyendo la necesidad de transferir grandes volúmenes de datos (Yang et al., 2019). AF se define como;

Sea  $D_i$  el conjunto de datos local del  $i$ -ésimo participante en la federación, y  $f_i$  representa el objetivo de aprendizaje local para ese participante. El problema AF se puede formular como la optimización de una función objetivo global  $F$ , es decir,

$$\min_{\theta} F(\theta) = \min_{\theta} \sum_i^N w_i f_i(\theta; D_i) \quad (3)$$

donde  $\theta$  representa los parámetros del modelo global,  $N$  es el número de participantes y  $w_i$  son las ponderaciones que reflejan la importancia relativa o el tamaño de cada conjunto de datos local  $D_i$ .

A continuación, se presenta una descripción de los aspectos algorítmicos del AF. Esta descripción abarca pasos como la partición de datos, el entrenamiento de modelos locales y la agregación global de los mismos. Asimismo, se proporciona una visión general de los pasos habituales en la implementación de AF en la atención médica (Mothukuri et al., 2021).

**Entrada:** Conjuntos de datos de entrenamiento  $D_1, D_2, \dots, D_m$  de  $m$  diferentes instituciones sanitarias, modelo global  $M$ , número de rondas de comunicación  $R$ , épocas de formación local  $E$ , tasa de aprendizaje  $\eta$ .

**Salida:** Modelo global entrenado  $M$ .

Partición de datos:

Dividir los datos de entrenamiento generales en  $m$  subconjuntos,  $D_1, D_2, \dots, D_m$  donde  $D_i$  representa el conjunto de datos local de la  $i$ -ésima institución de atención médica.

Entrenamiento de modelos locales:

Para cada ronda de comunicación  $r$  de 1 a  $R$ :

Enviar el modelo global  $M^{(r-1)}$  a todas las  $m$  instituciones.

Cada institución  $i$  inicializa su modelo local  $M_i^r$  con  $M^{(r-1)}$ .

Cada institución  $i$  entrena  $M_i^r$  en su conjunto de datos local  $D_i$  para  $E$  épocas utilizando un optimizador local LR  $\eta$ , dando como resultado un modelo actualizado  $M_i^{(r,E)}$ .

Agregación global de modelos:

Recopile los modelos actualizados  $M_1^{(r,E)}$ ,  $M_2^{(r,E)}$ ,  $M_m^{(r,E)}$  de todas las  $m$  instituciones.

Agregue los modelos locales para actualizar el modelo global:

$$M^{(r)} = \frac{1}{m} \sum_{i=1}^m M_i^{(r,E)} \quad (4)$$

Predicción:

Dado un nuevo ejemplo  $x$  utilice el modelo global entrenado  $M^{(R)}$  para predecir la etiqueta  $\hat{y}$ .

Devuelve la etiqueta predicha  $\hat{y}$ .

### 2.3.1 Tipos de aprendizaje federado

El AF permite proteger la privacidad de los datos en el entrenamiento de modelos de CNN, permitiendo la colaboración entre múltiples instituciones sin necesidad de compartir datos individuales. Esta sección explora tres modalidades de AF: horizontal, donde los datos distribuidos son similares en diferentes instituciones; vertical, donde las entidades tienen conjuntos de datos diferentes que pueden vincularse mediante identificadores comunes; y transversal, que combina ambos enfoques para maximizar la utilidad de los datos disponibles.

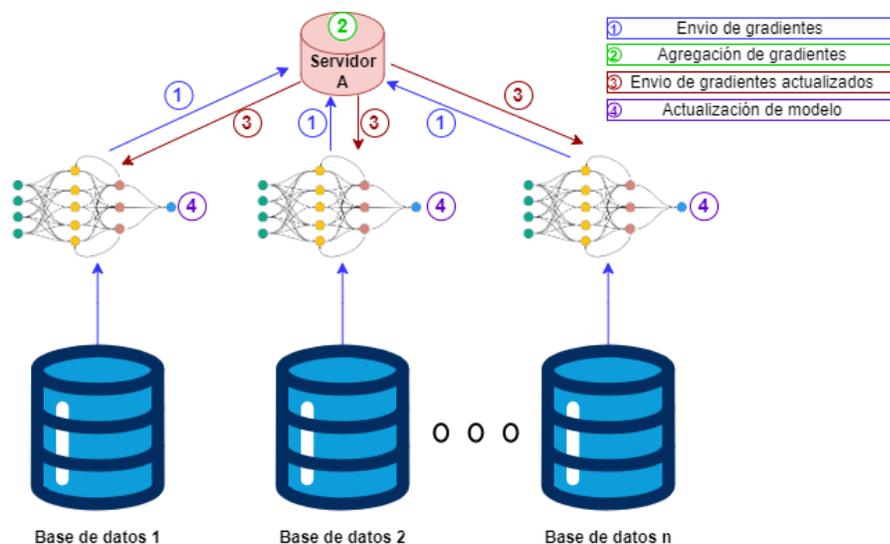
### 2.3.1.1 Aprendizaje federado horizontal

El Aprendizaje Federado Horizontal (HFL, por sus siglas en inglés) es un paradigma de aprendizaje automático distribuido que se aplica cuando las características de los datos de usuario en dos conjuntos de datos se superponen significativamente. En este enfoque, los conjuntos de datos se dividen horizontalmente a lo largo de la dimensión de los usuarios, permitiendo que cada nodo procese su porción de datos localmente. HFL se define como

$$X_i = X_j, Y_i = Y_j, I_i \neq I_j, \forall D_i, D_j, i \neq j \quad (5)$$

donde  $D_i, D_j$  representan dos conjuntos de datos locales pertenecientes a distintos clientes  $i, j$ .  $X_i, X_j$  son los espacios de características disponibles en los conjuntos  $D_i, D_j$ , respectivamente.  $Y_i = Y_j$  representan las etiquetas o variables objetivo.  $I_i \neq I_j$  son los índices o identificadores de las instancias (ejemplos concretos) contenidas en cada conjunto de datos (Hu et al., 2024).

Esta técnica, frecuentemente aplicada en escenarios que involucran dispositivos de Internet de las Cosas (IoT, por sus siglas en inglés), facilita el aprendizaje colaborativo preservando al mismo tiempo la privacidad y la seguridad de los datos (Lim et al., 2019). Por ejemplo, consideremos dos hospitales en diferentes ciudades: aunque ambos atienden a pacientes con condiciones similares, están ubicados en diferentes lugares.



**Figura 1.** Esquema general de aprendizaje federado horizontal.

En 2017, Google desarrolló un esquema de actualización de modelos utilizando un modelo HFL. En este modelo, todos los usuarios cargan los parámetros a la nube y los actualizan de manera centralizada después de completar la actualización local (Hu et al., 2024). La Figura 1 muestra este proceso, respecto a cómo los datos se dividen horizontalmente y cómo los parámetros del modelo se agregan y actualizan de manera centralizada.

### 2.3.1.2 Aprendizaje federado vertical

El Aprendizaje Federado Vertical (VFL, por sus siglas en inglés) es un paradigma distribuido utilizado cuando los usuarios de dos conjuntos de datos se superponen, pero las características de los datos de usuario en cada conjunto difieren (Kairouz et al., 2021). VFL se define como

$$X_i \neq X_j, Y_i \neq Y_j, I_i = I_j, \forall D_i, D_j, i \neq j \quad (6)$$

En el aprendizaje federado vertical, cada cliente posee información distinta sobre un mismo conjunto de individuos, es decir,  $I_i = I_j$  para todo par  $i \neq j$ . Sin embargo, los atributos disponibles en cada cliente son diferentes  $X_i \neq X_j$  y lo mismo puede ocurrir con las etiquetas  $Y_i \neq Y_j$ . Esto refleja un escenario típico donde cada institución tiene una vista parcial sobre los mismos sujetos, como ocurre entre hospitales y laboratorios (Hu et al., 2024).

Por ejemplo, consideremos dos instituciones diferentes, como una compañía de seguros y un hospital ubicados en la misma ciudad. Si bien es probable que sus grupos de usuarios sean los mismos, los hospitales registran la historia clínica y las condiciones de salud de los individuos, mientras que las compañías de seguros registran el estado de sus pólizas. La intersección de las características de usuario en estos dos conjuntos de datos es pequeña. VFL a menudo agrega estas diferentes características en un estado cifrado para mejorar la capacidad del modelo (Hu et al., 2024)

Este enfoque permite que instituciones con datos complementarios colaboren y utilicen sus datos para entrenar modelos más robustos y precisos, sin necesidad de compartir información sensible directamente. La Figura 2 muestra este proceso, mostrando cómo se combinan las diferentes características de usuario de varias fuentes en un entorno seguro y cifrado para el entrenamiento colaborativo del modelo.

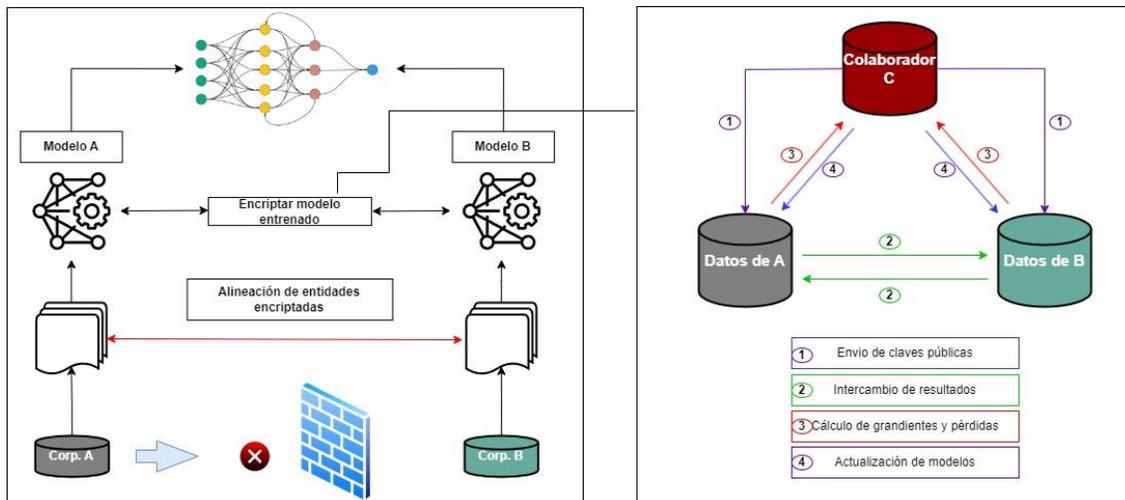


Figura 2. Esquema general de aprendizaje federado vertical.

### 2.3.1.3 Aprendizaje federado transversal

El Aprendizaje Federado Transversal (FTL, por sus siglas en inglés) es un paradigma distribuido que se utiliza cuando los usuarios y las características de los datos en dos conjuntos de datos tienen poca superposición. En lugar de dividir los datos, se emplea FTL para superar la falta de datos o etiquetas. Este enfoque se define como

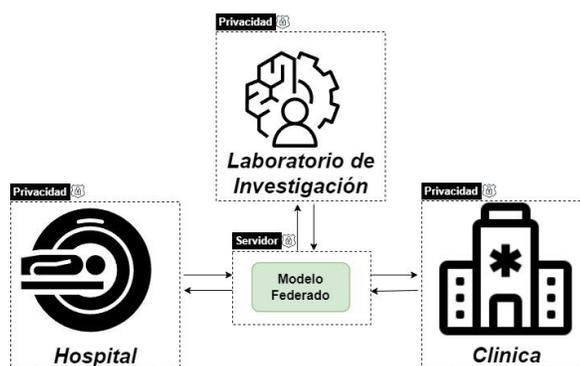
$$X_i \neq X_j, Y_i \neq Y_j, I_i \neq I_j, \forall D_i, D_j, i \neq j \quad (7)$$

En FTL, no existe coincidencia entre los atributos, las etiquetas ni las instancias disponibles en los diferentes nodos. Es decir, cada cliente trabaja con un dominio de datos completamente distinto:  $X_i \neq X_j$ ,  $Y_i \neq Y_j$  y  $I_i \neq I_j$  para todo  $i \neq j$ . Este escenario es característico de tareas de transferencia del conocimiento, donde se busca construir modelos colaborativos a partir de dominios heterogéneos y disjuntos (Hu et al., 2024).

Por ejemplo, en una colaboración entre bancos y supermercados en diferentes regiones, las características de diferentes espacios de características pueden ser transferidas a la misma representación y luego entrenadas utilizando etiquetas de los datos etiquetados recolectados por las diversas partes. Este método permite que instituciones con datos heterogéneos y no coincidentes colaboren eficazmente.

FTL facilita la integración de características diversas en una representación común, permitiendo el entrenamiento de modelos. Este enfoque es especialmente útil en situaciones donde la recopilación de grandes cantidades de datos etiquetados es costosa o impracticable. Al transferir características entre diferentes dominios, FTL permite aprovechar al máximo los datos disponibles, mejorando el rendimiento del modelo sin comprometer la privacidad de los datos (Hu et al., 2024)

La Figura 3 muestra este proceso, mostrando cómo se transfieren y combinan las características de diferentes fuentes en una representación latente común para el entrenamiento colaborativo del modelo.



**Figura 3.** Esquema general de aprendizaje federado transversal.

### 2.3.2 Algoritmos de agregación

Una vez que cada cliente ha efectuado su entrenamiento local, el servidor debe agregar las actualizaciones en un único modelo global. El mecanismo concreto de agregación determina la velocidad de convergencia, la robustez frente a datos heterogéneos y la estabilidad numérica de AF. A continuación, se describen los cuatro algoritmos más comunes en la literatura: FedAvg, FedAdam, FedYogi y FedProx.

#### 2.3.2.1 FedAvg

FedAvg es el algoritmo base y más ampliamente utilizado en AF. Dicho algoritmo fue propuesto por McMahan et al. (2016) y consiste en un proceso de entrenamiento iterativo donde cada cliente entrena un modelo local con su propio subconjunto de datos y, al final de cada ronda, estos modelos son promediados ponderadamente para actualizar el modelo global. La actualización se define como:

Cada cliente  $k$  parte del modelo global  $w_t$  y realiza  $E$  épocas locales de GD sobre su propia función de pérdida  $F_k$ . Tras finalizar, envía el modelo actualizado  $w_k^{t+1}$  al servidor, que calcula un promedio ponderado por el número de ejemplos  $n_k$ :

$$w_{t+1} = \sum_{k=1}^K \frac{n_k}{n} w_k^{t+1}, \quad n = \sum_{k=1}^K n_k \quad (8)$$

FedAvg es simple, eficiente y funciona bien en escenarios con datos Independientes e Idénticamente Distribuidos (IID). Sin embargo, su desempeño puede deteriorarse en contextos con alta heterogeneidad entre clientes, ya que las actualizaciones locales pueden ser divergentes, lo que ralentiza la convergencia y reduce la precisión del modelo global (Reddi et al., 2020).

### 2.3.2.2 FedAdam

FedAdam adapta el optimizador Adam, ampliamente utilizado en DL, al contexto federado. A diferencia de FedAvg, que utiliza únicamente el promedio de los pesos, FedAdam acumula momentos del gradiente global (como Adam) para realizar actualizaciones más informadas y estables. La actualización del modelo global se realiza con las siguientes ecuaciones:

$$\begin{aligned} m_t &= \beta_1 m_{t-1} + (1 - \beta_1) \Delta_t, \\ v_t &= \beta_2 v_{t-1} + (1 - \beta_2) \Delta_t^2, \\ w_{t+1} &= w_t - \eta \frac{m_t}{\sqrt{v_t} + \varepsilon} \end{aligned} \quad (9)$$

donde  $\beta_1, \beta_2 \in (0,1)$  son factores de decaimiento y  $\eta$  es la LR del servidor. El primer momento  $m_t$  corresponde a la media móvil de los gradientes, mientras que el segundo momento  $v_t$  representa la media móvil de los cuadrados de los gradientes. Este último se utiliza para ajustar adaptativamente la magnitud de las actualizaciones, reduciendo su tamaño en presencia de gradientes grandes y suavizando las primeras rondas de comunicación, lo que acelera la convergencia en presencia de clientes heterogéneos (Reddi et al., 2020).

FedAdam permite realizar pasos de optimización adaptativos que aceleran la convergencia, especialmente en entornos no IID, donde FedAvg puede resultar inestable. FedAdam acumula el segundo momento sin

restricciones, lo cual puede llevar a una LR efectiva más conservadora conforme avanza el entrenamiento

### 2.3.2.3 FedYogi.

FedYogi es un agregador basado en el optimizador Yogi, una variante adaptativa propuesta por Zaheer et al. (2018). Este optimizador busca superar una de las principales limitaciones de los métodos adaptativos clásicos, relacionada con el comportamiento del segundo momento  $v_t$ , el cual puede crecer indefinidamente cuando los gradientes locales son grandes o altamente ruidosos. Este crecimiento excesivo puede llevar a que la LR efectiva se reduzca drásticamente, afectando negativamente la velocidad de convergencia. Para mitigar este problema, Yogi modifica la actualización del segundo momento mediante el siguiente esquema:

$$v_t = v_{t-1} - (1 - \beta_2)\text{sign}(\Delta_t^2 - v_{t-1})\Delta_t^2 \quad (10)$$

donde  $\Delta_t$  representa el gradiente estocástico,  $\beta_2$  es el parámetro de decaimiento del segundo momento, y  $\text{sign}(\cdot)$  indica la dirección de ajuste. A diferencia de Adam, que siempre incrementa el valor de  $v_t$ , Yogi lo ajusta direccionalmente dependiendo de la diferencia entre el nuevo gradiente cuadrado y el valor anterior, evitando así una acumulación excesiva de la varianza.

### 2.3.2.4 FedProx

FedProx es una extensión directa de FedAvg propuesta para abordar uno de los principales desafíos en el AF: la heterogeneidad entre los datos locales de los clientes. En escenarios donde los datos no están distribuidos de manera independiente e idéntica (no IID), las actualizaciones de los modelos locales pueden divergir significativamente, afectando la estabilidad y la convergencia del modelo global.

Para mitigar este problema, FedProx introduce un término de regularización proximal que restringe la magnitud del cambio entre el modelo local y el modelo global. Esta modificación se refleja en la función de pérdida local optimizada por cada cliente:

$$\min_w F_k(w) + \frac{\mu}{2} \|w - w_t\|^2 \quad (11)$$

de modo que el GD local se modifica a

$$w_k^{t+1} = w_k^t - \eta[\nabla F_k(w_k^t) + \mu(w_k^t - w_t)] \quad (12)$$

donde:  $F_k(w)$  es la función de pérdida del cliente  $k$ ,  $w_t$  es el modelo global en la ronda  $t$ ,  $\mu$  es el hiperparámetro que controla la intensidad de penalización. FedProx no modifica el algoritmo de agregación central, sino el proceso de optimización local en cada cliente. Esta estrategia ofrece mejoras notables en entornos federados con alta heterogeneidad, donde FedAvg tiende a converger lentamente o a soluciones subóptimas (Reddi et al., 2020).

## 2.4 Aumento de datos y generación de imágenes sintéticas

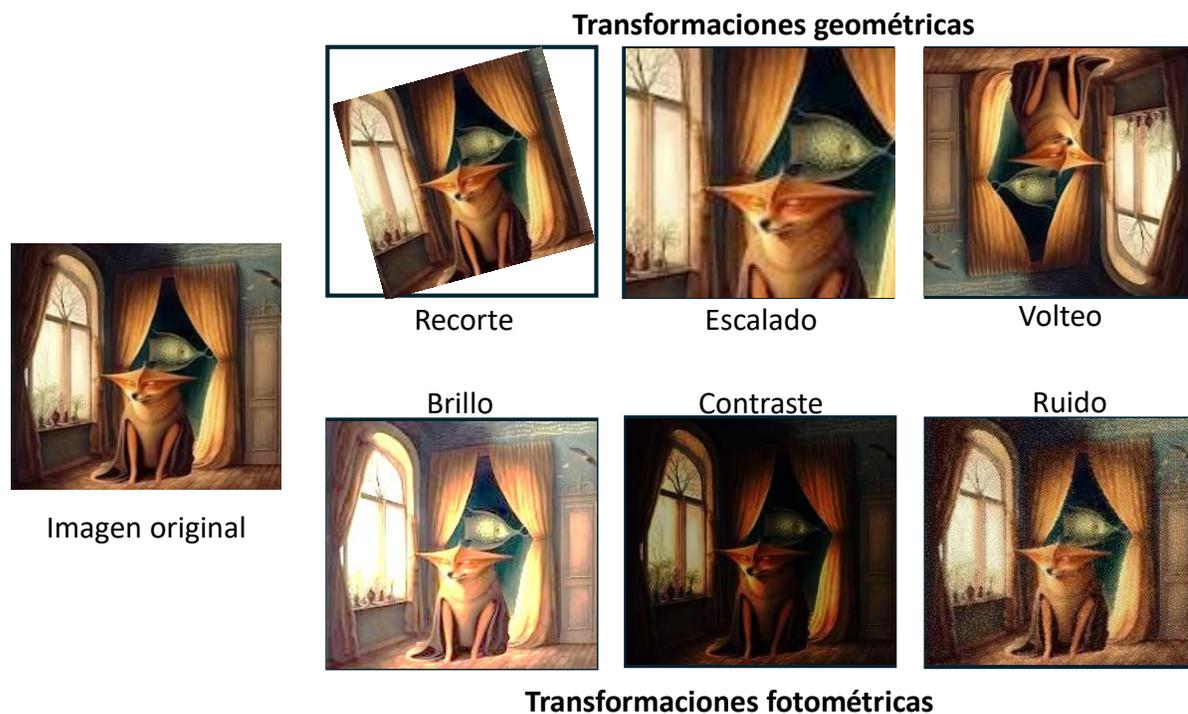
El desempeño final de un modelo de visión por computador está estrechamente ligada al tamaño y la diversidad del conjunto de entrenamiento. Sin embargo, en la práctica suele haber escasez de datos etiquetados, o bien las clases presentan desbalances. Dos familias de técnicas que mitigan ese problema sin necesidad de recolectar nuevos datos son: el aumento de datos y la síntesis de imágenes. A continuación, se muestran métodos más habituales y sus limitaciones principales.

### 2.4.1 Aumento de datos

El aumento de datos consiste en aplicar transformaciones que preservan la semántica de la imagen, pero alteran su representación píxel a píxel. Con ello se presentan al modelo variaciones plausibles de la fuente original, de modo que aprenda una función más robusta y menos propensa a sobreajuste. Además, si una clase minoritaria se amplía artificialmente, se reduce la varianza estadística del estimador y mejora la capacidad de generalización.

Las técnicas convencionales se clasifican en geométricas y fotométricas. Las primeras incluyen rotación,

escalado, recorte, volteo horizontal o pequeñas deformaciones elásticas; las segundas modifican brillo, contraste, saturación o añaden ruido (véase Figura 4). A pesar de su simplicidad, estas operaciones siguen siendo esenciales dado que introducen invariancia a transformaciones comunes: punto de vista, iluminación, desplazamiento, entre otros.



**Figura 4.** Transformaciones de imágenes para aumento de datos.

#### 2.4.2 Generación de imágenes sintéticas

Cuando el conjunto real es muy pequeño o se necesita simular condiciones extremas, los datos sintéticos puede aportar ejemplos completamente nuevos. La idea es entrenar un modelo generativo capaz de producir imágenes que sigan la distribución original.

Mientras las transformaciones clásicas amplían la varianza sin alterar la semántica, la generación sintética pretende acrecentar la varianza intrínseca: introduce texturas, estructuras y combinaciones de rasgos que el conjunto original quizá no contenga, pero que siguen siendo plausibles para el dominio. La literatura ha englobado en cuatro familias a modelos generativos, cada una con su propio mecanismo probabilístico y sus implicaciones prácticas.

El punto de partida lo marcan los Autocodificadores Variacionales (VAE, por sus siglas en inglés). Su premisa es condicionar la generación a un espacio latente de distribución bien comportada, normal isotrópica y entrenar, de manera conjunta, un codificador que proyecta las imágenes a esa latente y un decodificador que realiza el proceso inverso. El objetivo maximiza la similitud aproximada y un término de regularización obliga al codificador a producir latentes cercanos a la distribución gaussiana. La reconstrucción tiende a ser algo borrosa, resultado del compromiso entre fidelidad y suavidad, pero la solidez teórica del enfoque ha permitido extenderlo a dominios semánticos, condicionales y secuenciales con relativa facilidad (Kingma & Welling, 2013).

Las Redes Generativas Adversarias (GAN, por sus siglas en inglés) supusieron un salto cualitativo de fidelidad visual. El generador aprende a transformar ruido en imágenes convincentes, mientras el discriminador intenta distinguir muestras reales de sintéticas; ambos se entrenan en un juego de suma cero. Al final, cuando el discriminador ya no consigue decidir con seguridad, el generador produce imágenes que capturan con detalle los matices de la distribución original.

La formulación básica, sin embargo, adolece de inestabilidad: el equilibrio de Nash puede ser difícil de alcanzar y algunas regiones del espacio de datos quedan desatendidas, fenómeno conocido como colapso de modos. Se refinó la idea con penalizaciones de gradiente, discriminadores auxiliares y arquitecturas de estilo jerárquico que desacoplan global y detalle fino. En la práctica, los GAN se han convertido en un método para síntesis fotográfica y para esquemas de traducción imagen-a-imagen donde la tupla «dominio origen / dominio destino» se aprende sin emparejamientos explícitos (Goodfellow et al., 2014).

Frente al carácter adversario de los GAN, los modelos de flujo normalizado proponen una estrategia exacta: construyen funciones invertibles cuya jacobiana es eficiente de calcular. Gracias a esa invertibilidad, la densidad de probabilidad de cada muestra puede evaluarse sin aproximaciones, lo que convierte al modelo no solo en generador sino también en estimador de densidades. Su principal limitación es la demanda de memoria y cómputo, pues la transformación invertible suele requerir redes amplias (Dinh et al., 2016).

El avance más reciente lo conforman los modelos de difusión. En su fase directa añaden ruido gaussiano de forma secuencial hasta convertir la imagen en ruido puro; la fase inversa, entrenada con probabilidad variacional, aprende a revertir cada paso eliminando ruido. Al finalizar el recorrido inverso se obtiene una imagen nítida, a menudo con un nivel de detalle que supera al de los GAN y, sobre todo, con mayor estabilidad de entrenamiento: el objetivo se reduce a minimizar una pérdida de predicción de ruido, un

problema de regresión ordinario sobre distribuciones muy regulares. La contrapartida es el tiempo de muestreo: generar una sola imagen puede requerir centenares de pasos inversos, aunque recientes técnicas de integración numérica han reducido drásticamente ese coste (Ho et al., 2020).

La generación sintética no está exenta de riesgos. Si el modelo absorbe sesgos inadvertidos del conjunto real, los amplificará en las muestras nuevas. Además, la superposición de datos reales y sintéticos obliga a diseñar protocolos de validación estratificada para asegurarse de que el clasificador final no se confunde al ver su propio reflejo en el conjunto de prueba. No obstante, cuando se monitoriza la síntesis puede transformar conjuntos modestos de entrenamiento lo bastante ricos como para competir con conjuntos de datos mucho mayores, democratizando así el acceso a modelos de alto rendimiento.

## 2.5 Enfermedad de Alzheimer

La EA es una afección neurodegenerativa progresiva que constituye la causa más común de demencia entre las personas mayores. Afecta aproximadamente a 50 millones de personas en todo el mundo, con millones de nuevos casos diagnosticados cada año (World Health Organization, 2025). La enfermedad se caracteriza por la pérdida de neuronas y las sinapsis en la corteza cerebral y ciertas regiones subcorticales del cerebro, lo que provoca atrofas severas en las áreas afectadas.

### 2.5.1 Patología y síntomas

La patología de la EA se caracteriza por la acumulación de dos tipos de estructuras: las placas amiloides y los ovillos neurofibrilares. Las placas amiloides son depósitos extracelulares densamente agrupados de proteína beta-amiloide ( $A\beta$ ), que resultan de un procesamiento anómalo de la proteína precursora amiloide (Takahashi et al., 2017). Los ovillos neurofibrilares, por su parte, son agregados intracelulares de proteína tau hiperfosforilada, que se ensamblan en filamentos helicoidales emparejados y se acumulan dentro de las neuronas (Y. Gao et al., 2018). Además, la EA afecta principalmente a los tejidos grises del cerebro. Al monitorear los cambios cerebrales, se observa que las transformaciones pueden ocurrir antes de que una persona desarrolle completamente la EA. Las áreas de daño inicial incluyen las regiones que ayudan a formar recuerdos (Thompson et al., 2003), como el hipocampo, la corteza entorrinal, y otras áreas corticales y subcorticales. Estas alteraciones neuropatológicas interfieren con la función sináptica y la integridad neuronal, desencadenando disfunción y muerte celular. Clínicamente, estas alteraciones se

manifiestan como un deterioro cognitivo progresivo, que incluye pérdida de memoria, desorientación, y cambios en el comportamiento y la personalidad.

### 2.5.2 Factores de riesgo

Varios factores de riesgo contribuyen al desarrollo de la EA, incluyendo la edad avanzada, la genética, y ciertos factores de estilo de vida. La presencia de la apolipoproteína E (ApoE) alelo  $\epsilon 4$  es el predictor genético más fuerte de la enfermedad, su presencia puede afectar diversos mecanismos biológicos, incluyendo la acumulación de placas amiloides y la formación de ovillos neurofibrilares, procesos clave en la patogénesis de la enfermedad (Serrano-Pozo et al., 2021). Otros factores, como la hipertensión, la obesidad, el tabaquismo, y la diabetes (Zhang et al., 2021), también aumentan el riesgo de desarrollar esta condición, sugiriendo que las intervenciones en el estilo de vida podrían tener un impacto en la prevención o el retraso del inicio de la enfermedad.

### 2.5.3 Importancia de la detección temprana

La detección temprana de EA es fundamental para mitigar el avance de la demencia y mejorar la calidad de vida de los pacientes. Diagnosticar la enfermedad en sus etapas iniciales permite la implementación oportuna de intervenciones terapéuticas que pueden ralentizar la progresión de los síntomas neurológicos y preservar la funcionalidad cognitiva por un periodo más prolongado (Tan et al., 2014). Además, el diagnóstico precoz facilita la planificación adecuada para el manejo de la enfermedad, tanto para los pacientes como para sus familiares, quienes a menudo asumen el rol de cuidadores principales. Desde el punto de vista de los sistemas de salud, la detección temprana puede significar una reducción en los costos asociados con el cuidado a largo plazo y las intervenciones de emergencia, ya que permite un enfoque más gestionado y menos intensivo en recursos. Por lo tanto, mejorar la capacidad para identificar la enfermedad antes de que se manifiesten signos clínicos evidentes es crucial y requiere el desarrollo continuo de técnicas diagnósticas avanzadas y accesibles.

### 2.5.4 Biomarcadores

Inicialmente, se observa la formación de placas amiloides, seguida de la aparición de ovillos

neurofibrilares. Posteriormente, se detectan cambios morfológicos significativos en el cerebro, tales como la atrofia en regiones específicas como el hipocampo, la corteza entorrinal, y la corteza temporal medial. A medida que la enfermedad progresa, se manifiesta la pérdida de memoria (Mantzavinos & Alexiou, 2017), y finalmente, se observan alteraciones en las funciones clínicas que comprometen significativamente la calidad de vida del paciente como se muestra en la Figura 5. El reconocimiento temprano de estos biomarcadores es crucial para iniciar tratamientos que puedan alterar el curso de la enfermedad y ofrecer una intervención más efectiva.

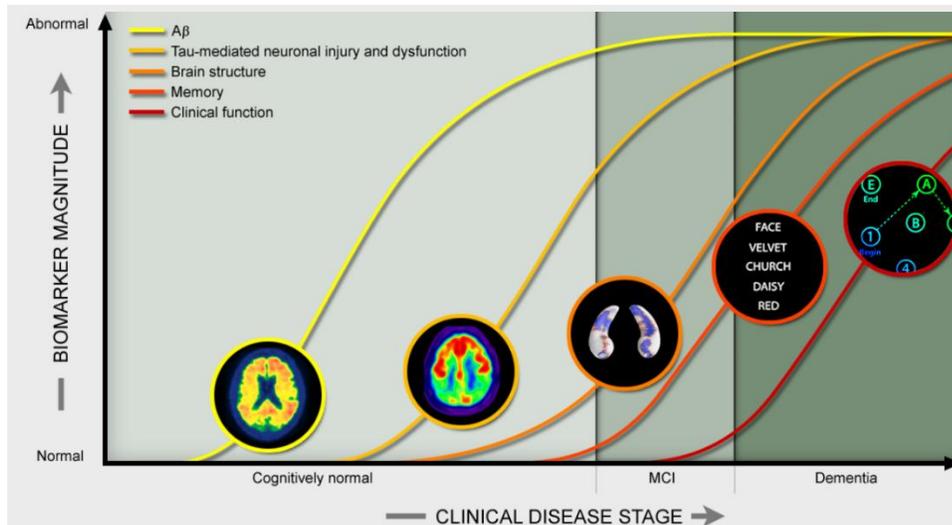


Figura 5. Etapas de biomarcadores de la EA. Imagen recuperada del ADNI.

## 2.6 Imágenes de resonancia magnética para diagnóstico

La Resonancia Magnética (MR, por sus siglas en inglés) es la aplicación clínica del fenómeno de Resonancia Magnética Nuclear (NMR, por sus siglas en inglés), descubierto de manera independiente en 1946 por Felix Bloch (1946) y Edward M. Purcell (1946). Ambos demostraron que determinados núcleos atómicos, en particular el protón ( $^1\text{H}$ ), absorben y reemiten energía de radiofrecuencia cuando se encuentran en un campo magnético estático.

Dos décadas más tarde, Paul C. Lauterbur (1973) publicó las primeras imágenes bidimensionales mediante la introducción de gradientes de campo magnético, y Peter Mansfield (1977) optimizó la adquisición rápida con gradientes de eco. Estos avances convirtieron la NMR de un experimento de laboratorio en una herramienta de imagen médica tridimensional, cuyo fin primordial es visualizar tejidos blandos con alto contraste sin emplear radiación ionizante.

Para contextualizar el funcionamiento de la MR, conviene identificar cuatro pilares físicos que actúan de forma secuencial y complementaria: (1) la alineación y precesión de los espines nucleares en un campo magnético estático, (2) los procesos de excitación y relajación que regulan la recuperación de la magnetización, (3) la codificación espacial y la posterior reconstrucción de la señal en imágenes clínicas, y (4) el contraste intrínseco generado por las diferencias de  $T_1$ ,  $T_2$  y densidad protónica entre tejidos. Estos componentes, descritos en detalle a continuación, constituyen la base sobre la cual se construyen las secuencias de adquisición y, en última instancia, la capacidad diagnóstica de la MR (Mansfield, 1977).

#### 1. Alineación y precesión.

En un campo estático  $B_0$ , generalmente 1.5 T o 3 T en clínica, los momentos magnéticos nucleares se alinean en dos estados energéticos; la ligera diferencia de poblaciones produce una magnetización neta  $M_0$  que precesa a la frecuencia de Larmor  $\omega_0 = \gamma B_0$  (42.58 MHz·T<sup>-1</sup> para <sup>1</sup>H).

#### 2. Excitación y relajación.

Cuando se aplica un pulso de radiofrecuencia sintonizado en  $\omega_0$ , la magnetización se aparta de su equilibrio. Tras el pulso, las interacciones moleculares provocan su retorno:

$T_1$  (relajación longitudinal) mide el tiempo que tarda  $M$  en recuperar el 63 % de  $M_0$

$T_2$  (relajación transversal) cuantifica la pérdida de coherencia en el plano perpendicular a  $B_0$

#### 3. Codificación espacial y reconstrucción.

Gradientes lineales ( $G_x, G_y, G_z$ ) hacen que la frecuencia o fase de las señales varíe con la posición, codificando cada punto del cuerpo en el espacio- $k$ . Una transformada de Fourier 2-D o 3-D traduce esos datos a una matriz de intensidades  $I(x, y, z)$ .

#### 4. Contraste intrínseco.

Las variaciones naturales de  $T_1$ ,  $T_2$  y densidad protónica ( $\rho$ ) entre tejidos generan diferencias de señal. Ajustar los tiempos de repetición (TR) y eco (TE) permite ponderar la imagen hacia uno u otro parámetro, multiplicando las posibilidades diagnósticas.

La importancia de las MRI radica en su capacidad para resaltar selectivamente distintos tejidos simplemente ajustando los parámetros de pulso (Brown et al., 2014; Dale et al., 2015; Mugler & Brookeman, 1990). La grasa contiene protones que intercambian energía con su entorno rápidamente, lo que se traduce en tiempos  $T_1$  cortos. Por ello, en una imagen  $T_1$ -ponderada ( $TR \approx 500$  ms,  $TE < 10$  ms) la grasa aparece brillante. Si el objetivo es suprimirla, por ejemplo, para estudiar médula ósea o lesiones vasculares, se recurre a secuencias con inversión o a técnicas de separación de agua-grasa.

El líquido cefalorraquídeo, por el contrario, tiene mucho espacio libre para que los protones mantengan su coherencia, de modo que presenta tiempos  $T_2$  prolongados ( $\approx 2000$  ms a 1.5 T). En una imagen  $T_2$  ponderada donde tanto el TR y el TE son largos, brilla intensamente, destacando quistes, edema o inflamación. Cuando se desea atenuar su señal para visualizar la sustancia blanca adyacente se utiliza Recuperación de Inversión Atenuada por Fluidos (FLAIR, por sus siglas en inglés), que combina un pulso de inversión con  $T_2$  largo para apagar selectivamente el líquido. La sustancia blanca cerebral, rica en mielina lipídica, tiene un  $T_1$  más corto y  $T_2$  más breve que la sustancia gris; por ello, en  $T_1$  se observa más clara que la corteza, mientras que en  $T_2$  ocurre lo inverso. Esta inversión de contraste ayuda a los radiólogos a delinear la corteza, cuantificar atrofia o detectar lesiones desmielinizantes.

En neurología, la MR revela la morfología cerebral con una claridad imposible en otras modalidades. Por un lado, las secuencias  $T_1$  ponderadas de alta resolución permiten cuantificar la atrofia del hipocampo, un marcador temprano de AD. Por otro lado, las imágenes FLAIR  $T_2$  ponderadas realzan las lesiones intracerebrales y permiten detectar las placas desmielinizantes características de la esclerosis múltiple, incluso antes de la aparición de síntomas clínicos (Jack et al., 2008).

Desde su concepción como curiosidad de laboratorio hasta su consolidación como pilar de la imagen médica, la MR ha mostrado una capacidad única para revelar la anatomía y la fisiología humanas sin infligir daño al paciente. Su fundamento físico, la interacción de los espines nucleares con campos magnéticos y radiofrecuencia, ofrece contrastes múltiples que permiten detectar y cuantificar enfermedades con precisión.

## 2.7 Bases de datos

Las bases de datos de MRI son esenciales para la investigación y el desarrollo de métodos de diagnóstico y clasificación de la EA. Estas bases de datos proporcionan un vasto conjunto de imágenes cerebrales junto

con información clínica y demográfica, lo que permite desarrollar y validar algoritmos de DL. A continuación, se presentan algunas de las bases de datos más reconocidas y utilizadas en el campo de la investigación de EA:

La Iniciativa de Neuroimagen de la Enfermedad de Alzheimer (ADNI, por sus siglas en inglés) es una de las fuentes más extensas y completas de datos de neuroimágenes (Alzheimer's Disease Neuroimaging Initiative, 2024). Esta base de datos de acceso libre incluye MRI, Tomografía por Emisión de Positrones (PET, por sus siglas en inglés), biomarcadores de líquido cefalorraquídeo y datos clínicos de individuos con EA, Deterioro Cognitivo Leve (MCI, por sus siglas en inglés) y Controles Sanos (HC, por sus siglas en inglés). La ADNI tiene como objetivo principal investigar el uso de estas herramientas para la detección temprana y el seguimiento de la progresión de la EA. La cantidad de imágenes MRI en esta base de datos asciende a aproximadamente 8000. Las imágenes están clasificadas en tres categorías: HC, MCI y EA.

La serie OASIS proporciona imágenes MRI estructurales de adultos sanos y pacientes con EA. Esta base de datos es ampliamente utilizada para estudios de análisis de imágenes y desarrollo de métodos de segmentación y clasificación de estructuras cerebrales afectadas por la EA. OASIS cuenta con aproximadamente 1100 sujetos con MRI y de igual manera es de acceso libre con previa autorización. Las imágenes están clasificadas en tres categorías: HC, MCI y EA (LaMontagne et al., 2019).

El Estudio Australiano sobre Envejecimiento, Imagen Médica, Biomarcadores y Estilo de Vida (AIBL, por sus siglas en inglés) recopila datos de neuroimágenes, biomarcadores y factores de estilo de vida para estudiar el envejecimiento y la progresión de la EA. Esta base de datos incluye imágenes MRI, PET y datos clínicos, proporcionando una perspectiva integral sobre los factores que influyen en la EA. AIBL tiene alrededor de 3000 sujetos con MRI disponibles y es de acceso libre con previa autorización (AIBL, 2024). Las imágenes están clasificadas en cinco categorías de acuerdo con el marco de la Clasificación Clínica de la Demencia (CDR, por sus siglas en inglés): Sanos (CDR 0), Demencia cuestionable (CDR 0,5), Demencia leve (CDR 1), Demencia moderada (CDR 2) y Demencia grave (CDR 3) (Morris, 1997).

El Estudio de Resonancia Magnética con Intervalos Mínimos en pacientes con Enfermedad de Alzheimer (MIRIAD, por sus siglas en inglés) incluye imágenes MRI de pacientes con EA obtenidas en intervalos cortos de tiempo. Esta colección de datos es valiosa para estudiar los cambios cerebrales a lo largo del tiempo y evaluar la progresión rápida de la atrofia cerebral en la EA. MIRIAD dispone de aproximadamente 70 sujetos con MRI y es de acceso libre con previa autorización. (UCL, 2018) Las imágenes están clasificadas

en cinco categorías: Sanos (CDR 0), Demencia cuestionable (CDR 0,5), Demencia leve (CDR 1), Demencia moderada (CDR 2) y Demencia grave (CDR 3).

Kaggle (Alzheimer's Dataset - 4 class of Images), un repositorio de datos que alberga varios conjuntos de datos de imágenes MRI para la EA, incluyendo conjuntos de datos derivados de la ADNI y otras fuentes. Estos conjuntos de datos se utilizan ampliamente en competencias de ciencia de datos y son accesibles libremente (Pinamonti, 2022). El conjunto de datos específico de Kaggle contiene 6400 imágenes, las imágenes están clasificadas en cuatro categorías: Sano, Demencia Muy Leve (VMD, por sus siglas en inglés), Demencia Leve (MiD, por sus siglas en inglés) y Demencia Moderada (MD, por sus siglas en inglés).

La base de datos del Estudio Harvard sobre el Envejecimiento Cerebral (HABS, por sus siglas en inglés) se distingue entre otras bases de datos, ya que se enfoca específicamente en la caracterización de adultos mayores cognitivamente normales durante las etapas preclínicas del deterioro neurodegenerativo. A diferencia de otras bases como ADNI u OASIS, que incluyen sujetos con diversos grados de deterioro cognitivo e incluso con diagnóstico confirmado de EA, HABS recopila datos longitudinales de participantes sin síntomas clínicos, lo que permite el análisis de biomarcadores sutiles y progresiones subclínicas mediante MRI, PET y evaluaciones neuropsicológicas repetidas en el tiempo (HABS, 2024).

**Tabla 1.** Comparativa de bases de datos de MRI para la enfermedad de Alzheimer.

Conjunto de datos	No. de pacientes	Formato	Estadios	Nomenclatura	Requiere pre procesamiento	Se requiere autorización
ADNI	2804	DICOM	3	HC, MCI, EA	✓	✓
Alzheimer 4 classes	200	JPG	4	ND, VMD, MiD, MD		
AIBL	3045	NIFTI	5	CDR (0-3)	✓	✓
HABS	290	DICOM	2	HC, EA	✓	✓
MIRIAD	69	NIFTI	5	CDR (0-3)	✓	✓
OASIS-3	1098	NIFTI	3	HC, MCI, EA	✓	✓

La Tabla 1 presenta una comparación detallada de las principales bases de datos utilizadas en la investigación sobre la EA. En ella se destacan: la cantidad de pacientes disponibles en cada base; el formato en que se encuentran las imágenes; el número de estadios clínicos en los que se clasifican las imágenes y

la nomenclatura correspondiente a cada estadio —es decir, las clases o categorías asignadas a las imágenes—; la indicación de si las imágenes requieren preprocesamiento, entendido como la necesidad de aplicar procedimientos adicionales como la conversión de formato, la normalización de intensidad o la remoción de estructuras óseas antes de su uso; y finalmente, si el acceso a los datos requiere autorización previa, como el registro del proyecto o la aprobación de un comité de datos.

Estas bases de datos permiten la implementación de estudios a gran escala y la validación de modelos de DL para la clasificación y diagnóstico de la EA. La disponibilidad de datos longitudinales y de múltiples modalidades de imagen facilita una comprensión más profunda de la progresión de la enfermedad y el desarrollo de estrategias de intervención temprana.

## Capítulo 3. CNN y aprendizaje federado para la clasificación de MRI en la EA

---

El auge reciente del aprendizaje automático en medicina se explica, en buena medida, por la eficacia demostrada de las CNN cuando se entrenan con grandes volúmenes de datos clínicos. Esta combinación de modelos profundos y bases de imágenes en expansión ha disparado la precisión diagnóstica en múltiples especialidades. No obstante, la velocidad con que surgen nuevas arquitecturas y técnicas complica distinguir qué enfoques siguen vigentes, dónde se sitúan sus límites y qué vacíos de conocimiento persisten. El presente capítulo ofrece una visión del estado actual de las CNN en la detección de EA.

### 3.1 Redes neuronales convolucionales para clasificación de MRI en la EA

Como primer ejemplo de los enfoques basados en CNN aplicados a la clasificación de MRI, AbdulAzeem et al. (2021), propusieron una arquitectura sencilla compuesta por tres capas convolucionales, agrupamiento máximo, dos capas densas y un esquema de aumento de datos que expandió el conjunto ADNI a más de 211,000 cortes. Su modelo alcanzó un 99.6 % de precisión en clasificación binaria (EA vs CN) y un 97.5 % en la tarea multiclase (EA, MCI, CN). Este estudio demostró que, con un preprocesamiento adecuado — incluyendo conversión a escala de grises y umbralización adaptativa— y un conjunto diverso de transformaciones, una CNN de arquitectura reducida puede obtener un rendimiento comparable al de modelos considerablemente más complejos.

En la misma línea de eficiencia sin sacrificar rendimiento, El-Assy et al. (2024) propusieron una CNN de 17 capas con bloques compresión-expansión que permite reducir la cantidad de parámetros sin perder capacidad expresiva. Entrenada desde cero en 2,750 estudios  $T1$  de ADNI: consiguió 94,8% de exactitud multiclase y redujo la inferencia a 11 ms por volumen, lo que confirma la relevancia de las arquitecturas compactas en escenarios donde la latencia resulta un factor crítico.

El preentrenamiento en ImageNet sigue siendo valioso cuando la muestra clínica es escasa. Hussain et al. (2025) mostraron que basta reentrenar los dos últimos bloques de un ResNet-50 para alcanzar un AUC de 0,942 en 1800 estudios de tres centros distintos, manteniendo un número reducido de pesos actualizables y, por ende, menor riesgo de sobreajuste.

Conjuntos multiclase desequilibrados inspiran estrategias en cascada. Ali et al. (2024) diseñaron una red de 26 capas que primero separa controles de casos y luego discrimina MCI de EA; con validación cruzada 3-fold sobre 4,100 volúmenes obtuvieron 93,1% de F1-score global y redujeron los falsos negativos en etapas iniciales.

Insertar atención bien calibrada mejora la saliencia sin inflar el tamaño del modelo. Gao et al. (2023) intercalaron un módulo espacial-canal en EfficientNet-B4 e incrementaron la precisión de 88,7% a 92,4% con solo un 3% extra de parámetros. De forma parecida, Illakiya et al. (2023) proponen el modelo AHANet en donde incorporaron información de contexto larga y Atención Coordinada a DenseNet-169, alcanzando 98,5% de precisión sobre 158,100 cortes ADNI, lo que refuerza la idea de que la atención potencia las regiones discriminativas clave.

La realidad hospitalaria exige modelos capaces de ejecutarse en entornos con Memoria de Video (VRAM, por sus siglas en inglés) limitada, es decir, la capacidad de almacenamiento temporal disponible en la tarjeta gráfica para procesar datos y realizar cálculos. Jabason et al. (2025) diseñaron Slim-3DNet, una red volumétrica de 13 capas con convoluciones por grupo ( $g = 8$ ) y poda estructurada al 55 %. El modelo resultante, de solo 7 MB, alcanzó un 90.2 % de exactitud y una latencia de 45 ms al procesar 240 sujetos de un mini-ADNI en una GPU GTX-1650 Ti (4 GB). Su consumo pico de VRAM fue de 0.9 GB, lo que permite su implementación en estaciones radiológicas sin necesidad de tarjetas gráficas de gama alta.

Fusionar representaciones profundas con clasificadores tradicionales puede mejorar la calibración. Chen et al. (2022) extrajeron vectores latentes mediante una CNN y los alimentaron a un algoritmo de bosque aleatorio iterado, alcanzando AUC-ROC de 90,6% (EA vs CN) y 85,1% (EA vs MCI). Cuando la resonancia se procesa como una secuencia axial ordenada, las redes de memoria aportan ventajas: una red de memoria a corto-largo plazo de Salehi et al. (2023) logró 98,6 % de precisión y AUC 0,97 con apenas 6,400 imágenes del conjunto Kaggle OASIS.

Para conjuntos desbalanceados, Murugan et al. (2021) proponen el modelo DEMNet donde demostraron que el uso de SMOTE; un método de sobre muestreo que genera nuevas instancias sintéticas de la clase minoritaria, aumenta la precisión comparado con simplemente duplicar ejemplos existentes. Lo hace interpolando entre ejemplos reales cercanos de esa clase, combinado con una CNN tipo VGG y RMSprop, eleva la exactitud multiclase a 95,2 %. Hazarika et al. (2022) compararon 20 arquitecturas y, tras optimizar DenseNet-121, subieron del 88,8 % al 90,2 %. Arafa et al. (2024) subrayaron que un diseño ad-hoc puede superar a VGG-16 pre-entrenado: su CNN propia alcanzó 99,99 % frente al 97,4 % de la red transferida.

Más allá de las tareas de clasificación discreta, Lee et al. (2022) entrenaron un modelo DenseNet-3D para estimar la edad cerebral a partir de 4,127 estudios combinados de MRI y FDG-PET. El modelo alcanzó un error medio absoluto de 3.1 años para las imágenes PET y de 3.5 años para las imágenes MRI, lo que significa que, si el modelo predice que un sujeto tiene 70 años, su estimación podría desviarse en promedio  $\pm 3.1$  años en el caso de PET y  $\pm 3.5$  para el caso de MRI. Además, la discrepancia entre la edad cerebral estimada y la edad cronológica mostró una correlación significativa con el deterioro cognitivo, lo que sugiere una vía potencial para el diagnóstico temprano sin necesidad de un etiquetado exhaustivo.

En conjunto, la literatura revisada muestra que las CNNs aplicadas a la clasificación de MRI en la EA han alcanzado desempeños con precisiones superiores al 95 % en escenarios multiclase y reducciones significativas en la complejidad computacional. Sin embargo, persisten desafíos que limitan su adopción clínica generalizada, como la dependencia de conjuntos de datos específicos, lo que restringe la generalización a distintas poblaciones y dispositivos, el desbalance de clases que puede sesgar la detección en etapas tempranas, y la escasa validación en entornos hospitalarios reales con recursos de hardware limitados. Asimismo, la mayoría de los modelos se entrenan y evalúan en condiciones centralizadas, sin considerar las restricciones de privacidad y disponibilidad de datos que caracterizan a los sistemas de salud. Estas limitaciones abren oportunidades para el desarrollo de arquitecturas más robustas, eficientes y adaptables, así como para la integración de enfoques de aprendizaje federado y técnicas de aumento de datos que mitiguen la escasez y heterogeneidad de las imágenes médicas.

### **3.2 Aprendizaje federado en enfermedades neurodegenerativas**

El AF ha pasado de ser un concepto teórico para convertirse en una solución práctica para entrenar modelos sin que los datos abandonen los hospitales, los laboratorios o los dispositivos de origen. En los últimos tres años, los grupos de investigación han ido más allá de las primeras pruebas de concepto y han empezado a explorar problemas de neurología, en particular EA y Parkinson, con estrategias tan diversas como optimización evolutiva, hardware acelerado o cifrado homomórfico. A continuación, se integran los trabajos más recientes, destacando el tipo de señal, la estrategia federada y los resultados clínicos que consiguieron.

Altable et al., (2023) entrenaron redes profundas con grabaciones de voz de pacientes con Parkinson repartidas entre dos clínicas. Para equilibrar los estadios de la enfermedad aplicaron SMOTE y, aun con solo 82,000 registros crudos, el sistema federado alcanzó una AUC próxima al 99 %. El trabajo puso de

relieve que la ganancia no depende exclusivamente de imágenes: otros biomarcadores, como la prosodia del habla, también pueden beneficiarse de la colaboración sin compartir datos.

En el ámbito radiológico, Basnin et al., (2025) propusieron un marco evolutivo que combina CNN con optimización por enjambre de partículas para afinar los hiperparámetros de cada cliente. Tras un preprocesado de MRI con *remoción de cráneo*; técnica que consiste en eliminar el cráneo y otros tejidos no cerebrales, y corrección de distorsiones de baja frecuencia, el modelo global llegó al 99,9 % de precisión, superando en más de diez puntos a los clasificadores tradicionales como bosques aleatorios y máquinas de soporte vectorial, entrenados de forma centralizada sobre los mismos datos.

Castro et al. (2024) introdujeron un paso adicional de autenticación biométrica de las imágenes antes de enviarlas al servidor. Con cinco rondas de comunicación el sistema federado superó a su equivalente centralizado 92% de precisión en OASIS y 88,5% en ADNI. Además, resistió inyecciones de ruido que simulaban ataques de intermediario.

Kapoor et al., (2024) combinaron AF con aprendizaje continuo y consolidación elástica de pesos; es una técnica utilizada en redes neuronales para mitigar el problema del olvido catastrófico, que ocurre cuando un modelo de DL olvida tareas anteriores al aprender nuevas, y transferencia desde un InceptionV3 pre-entrenado, construyendo así un ensamble: FedCL-Ensemble, que retuvo memoria de dominios anteriores y alcanzó 97,8% de exactitud. Por su parte, Mandawkar et al., (2022) integraron un algoritmo bio-inspirado: Optimización Tawny Flamingo, dentro de cada cliente para buscar la mejor topología de la CNN antes de participar en la ronda federal; el resultado fue un 98,25% de precisión, validado con k-fold estratificado.

Otro frente emergente es el hardware especializado. Khalil et al., (2023) desplegaron una red de MLP sobre dispositivos de borde basados en FPGA para clasificar biomarcadores sanguíneos de ADNI. Aun con un consumo inferior a 2 W, el sistema mantuvo 89 % de precisión y 87 % de sensibilidad, ilustrando que la inferencia federada puede llegar a consultorios o laboratorios sin grandes servidores.

La seguridad ha sido examinada con el mismo rigor. Mitrovskaja et al., (2024) compararon el esquema FedAvg puro con su versión protegida por agregación Segura (SecAgg) en tres bases públicas: ADNI, OASIS-3, AIBL. La defensa criptográfica apenas redujo la precisión un punto porcentual, pero eliminó los riesgos de ataque de inferencia de pertenencia. En la misma línea, Stripelis et al., (2023) presentaron MetisFL, un

sistema que introduce Cifrado Homomórfico (HE, por sus siglas en inglés) para todo el flujo y demostró que el sobre coste de tiempo de entrenamiento es manejable ( $\approx 25\%$ ) en mallas de cuatro a seis centros.

Por otro lado, Qian et al., (2023) ofrecieron FeDeFo, un bosque aleatorio federado entrenado exclusivamente con volúmenes hipocampales. A diferencia de las redes, los bosques profundos conservan interpretabilidad y, aun así, FeDeFo registró 91,6 % de precisión binaria. Recientemente, Wei et al. (2023) exploraron la señal de voz para la detección de EA mediante un enfoque denominado FedCPC que combina el aprendizaje contrastivo predictivo —un método de preentrenamiento auto supervisado basado en la predicción de representaciones latentes— con el AF. Esta técnica permite entrenar modelos distribuidos sin centralizar los datos, aprendiendo representaciones útiles a partir de la señal de voz. Con este enfoque, lograron elevar la macro F1 a 78.8 %, un valor competitivo para tareas de clasificación acústica.

En conjunto, los estudios revisados evidencian que el AF en salud ha alcanzado avances sustanciales, demostrando su viabilidad en dominios diversos como neuroimagen, biomarcadores acústicos y sanguíneos, e incorporando estrategias de optimización, seguridad y despliegue en hardware. Los resultados reportados muestran precisiones que, en muchos casos, superan el 90 %, así como la capacidad de preservar la privacidad y adaptarse a diferentes modalidades de datos. No obstante, persisten retos que limitan su adopción generalizada, entre ellos la dependencia de conjuntos de datos limitados o no representativos, la heterogeneidad de las fuentes y protocolos de adquisición, la necesidad de reducir latencias y consumo de recursos en entornos clínicos reales, y la falta de validación multicéntrica a gran escala. Estas limitaciones abren oportunidades para investigar arquitecturas más robustas y eficientes, diseñar esquemas de agregación resilientes a la variabilidad de los clientes, integrar técnicas avanzadas de aumento y preentrenamiento, y establecer marcos de evaluación estandarizados que faciliten su transición hacia aplicaciones clínicas rutinarias.

### **3.3 Uso de datos sintéticos para balanceo de clases**

En el entrenamiento de modelos de DL para el análisis de imágenes médicas, es frecuente que las clases presenten una distribución desigual debido a factores como la baja prevalencia de ciertas enfermedades, las restricciones éticas, y las limitaciones en la adquisición de datos. Este desbalance provoca que los modelos tiendan a favorecer las clases con mayor número de muestras, reduciendo su capacidad para identificar correctamente las clases minoritarias y afectando su generalización. En este contexto, la generación de datos sintéticos mediante redes generativas se presenta como una estrategia para

incrementar el número de instancias en las clases subrepresentadas. Estas imágenes sintéticas permiten ampliar el conjunto de datos manteniendo características visuales y estructurales coherentes con las imágenes reales, mitigando el sesgo de entrenamiento y mejorando el rendimiento del modelo en escenarios donde los datos reales son escasos (Skandarani et al., 2023).

Malihi et al. (2023) parte de un conjunto limitado y desbalanceado de 987 fotografías clínicas: 480 úlceras diabéticas y 507 venosas. Tras segmentar cada lesión y normalizar la resolución a 256×256 píxeles, los autores entrenaron un StyleGAN3 sobre una GPU RTX 8000 para generar imágenes verosímiles que sirvieran de refuerzo a un clasificador Xception. El estudio tuvo como objetivos principales: (i) evaluar si la inclusión de imágenes sintéticas mejora el rendimiento del clasificador basado en CNN, y (ii) determinar si las muestras generadas resultan suficientemente realistas para los especialistas clínicos.

En términos de métrica, la respuesta fue matizada. Añadir una cantidad de imágenes sintéticas equivalente al 50 % del conjunto original elevó la exactitud de 0.851 a 0.870 y el F1-score de 0.844 a 0.867; con un 75% de datos artificiales los valores subieron modestamente hasta 0.878 y 0.878, respectivamente. Sin embargo, la mejora no fue uniforme: precisión y recall fluctuaron y, en ocasiones, quedaron por debajo del entrenamiento con imágenes reales puras, lo que sugiere que cantidad no siempre equivale a calidad cuando la síntesis introduce artefactos.

El juicio humano confirmó esa cautela. En una encuesta con 217 clínicos solo el 31% de las imágenes sintéticas fue considerada “real”, mientras que también hubo ejemplos auténticos clasificados como artificiales. Los autores concluyen que filtrar las muestras que los expertos perciben como creíbles podría aportar más beneficio que volcar todo el lote generado, proponiendo una curación automática basada en calidad visual antes de alimentar al modelo.

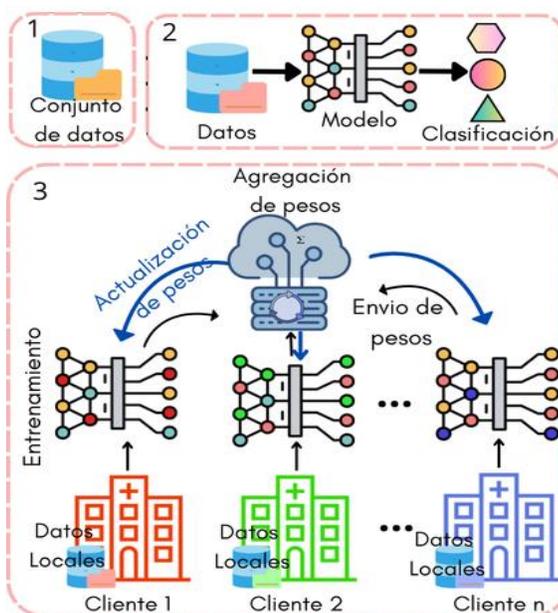
El estudio de Malihi et al. (2023) demuestra que las imágenes sintéticas pueden mejorar los clasificadores, pero sólo cuando se controla su fidelidad. Para esta tesis utilizaremos la síntesis de manera limitada y curada: equilibrar clases minoritarias de MRI tras validar el realismo mediante un clasificador, evitando recrear artefactos que perjudiquen la generalización.

## Capítulo 4. Metodología

La metodología empleada en este trabajo se estructura en tres bloques fundamentales, tal como se ilustra en la Figura 6. El primer bloque corresponde a la selección del conjunto de datos.

El segundo bloque abarca la selección del modelo de CNN, en este bloque se evaluaron diferentes arquitecturas para determinar aquella que ofrecía el desempeño más elevado en la clasificación de la EA, para ello se realizaron experimentos con el conjunto de datos seleccionado en el primer bloque. Aunado a ello se propone la optimización de uno de los modelos evaluados (AHANet) agregando un bloque de super-convergencia dando origen a la propuesta de AHANet-C. Por último, se realizaron experimentos en torno al aumento de datos en un entorno centralizado mediante dos técnicas; transformaciones geométricas y generación de imágenes sintéticas mediante GAN.

Finalmente, el tercer bloque comprende la implementación del entorno de AF y la ejecución de 16 experimentos con el modelo AHANet-C, empleando diferentes configuraciones de entrenamiento. Estos experimentos incluyen variaciones en el número de rondas de comunicación, el número de clientes simulados y la utilización de imágenes generadas sintéticamente mediante GAN en el esquema federado. Los resultados obtenidos en esta fase fueron analizados y comparados para evaluar el impacto de cada una de estas configuraciones.



**Figura 6.** Representación general de la metodología propuesta.

## 4.1 Conjuntos de datos

Todos los experimentos se basan en el Alzheimer MRI 4-Classes Dataset alojado en Kaggle (Pinamonti, 2022). Este repositorio de libre acceso consiste en 6400 MRI T1 axiales (176x208 píxeles) distribuidas en cuatro clases que denotan el estadio clínico. Se decidió re-etiquetar los estadios de la enfermedad de acuerdo con el marco proporcionado por la Asociación de Alzheimer (Asociación de Alzheimer, 2025). La Tabla 2 detalla el contenido de la base de datos, así como dicho proceso de re-etiquetado.

**Tabla 2.** Características de las imágenes del conjunto de datos.

Clase	Nueva etiqueta (Estadio)	No. de sujetos	No. de imágenes
<b>Non Dementia</b>	Normal	100	3200
<b>Very Mild Dementia</b>	Temprano	70	2240
<b>Mild Dementia</b>	Medio	28	896
<b>Moderate Dementia</b>	Tardío	2	64

Se adoptó una división estratificada a nivel de imagen —70 % entrenamiento, 20 % validación y 10 % prueba— de modo que las cuatro clases mantuvieran su proporción en cada subconjunto (véase Tabla 3). El repositorio no facilita identificadores de paciente y el estadio Tardío procede de únicamente dos sujetos; esta limitación queda registrada y se considera al interpretar los resultados, tal como se discute en el capítulo 5. Las siguientes secciones describen el preprocesado uniforme aplicado a todas las imágenes, las arquitecturas evaluadas y la configuración federada, respetando siempre la misma partición para asegurar comparabilidad interna.

**Tabla 3.** División de datos para los conjuntos de entrenamiento, validación y prueba para cada uno de los estadios.

	Validación	Prueba	Entrenamiento
<b>Normal</b>	640	321	2239
<b>Temprano</b>	448	225	1567
<b>Medio</b>	179	90	627
<b>Tardío</b>	13	7	44

### 4.1.1 Preprocesamiento y normalización

Aunque el conjunto de Kaggle ofrece los cortes axiales ya delimitados a 176x208 píxeles en formato JPEG, se aplicó un preprocesado homogéneo antes de alimentar cualquier modelo a fin de minimizar la varianza

de intensidad y geometría entre imágenes procedentes de distintos tipos de escáner. El flujo se implementó en PyTorch/MONAI y encapsula cada paso en transformaciones reproducibles con semilla fija.

Carga y tipado: Las imágenes se abren con Pillow, se convierten a escala de grises (modo L) y se convierten a Float32 en el intervalo [0, 255].

Normalización de intensidad: Para atenuar píxeles saturados se recortan los valores por fuera del rango P1–P99 (percentiles intra-imagen). El resto se escala linealmente a [0, 1]. Esta operación conserva el contraste relativo, pero evita que unos pocos píxeles extremos dominen la propagación del gradiente.

Estandarización  $z$ -score por imagen: Se sustrae la media  $\mu$  y se divide por la desviación estándar  $\sigma$  calculadas tras la delimitación anterior, de modo que cada corte tenga distribución  $N(0, 1)$ . Este paso favorece la estabilidad numérica en lotes pequeños—condición habitual en AF.

Conversión a tensor y almacenamiento en caché: El resultado se normaliza nuevamente a Float32, se convierte en torch.Tensor y se guarda en un cache loader (LMDB) para acelerar las rondas federadas, donde los nodos cargan y liberan lotes de forma asíncrona.

Los parámetros anteriores—percentiles, semilla, tamaño de trío—previamente definidos, se aplicaron de manera idéntica a las tres divisiones (entrenamiento, validación, prueba) con el fin de garantizar comparabilidad y reproducibilidad en todas las arquitecturas y escenarios federados.

#### 4.1.2 Aumento de datos y generación sintética

Durante la etapa inicial de evaluación de arquitecturas CNN (sección 4.2) no fueron aplicadas técnicas de aumento de datos, esto a fin de mantener condiciones experimentales homogéneas y comparables entre modelos. No obstante, una vez seleccionado un modelo, se realizaron experimentos adicionales en un entorno centralizado para estudiar el impacto de estrategias para abordar el desbalance de clases, estas son:

1. Aumento de datos mediante transformaciones geométricas, y
2. generación de imágenes sintéticas mediante GAN.

Se analizaron dos escenarios experimentos considerando la arquitectura del modelo seleccionado en un entorno centralizado, a fin de evaluar su rendimiento bajo diferentes condiciones de entrenamiento:

Condición A – Aumento de datos: se aplicaron transformaciones aleatorias a las imágenes de los estadios: Temprano, Medio y Tardío, incluyendo rotaciones de hasta 15°, volteo horizontal y vertical, y escalado. Estas imágenes fueron añadidas exclusivamente al conjunto de entrenamiento como se observa en la Tabla 4, cabe mencionar que tanto el conjunto de validación como de prueba se mantuvieron inalterables durante los experimentos.

**Tabla 4.** Aumento de imágenes mediante transformaciones geométricas en el conjunto de entrenamiento para cada uno de los estadios.

Estadio	Validación	Prueba	Entrenamiento		
			Reales	Img por aumento de datos	Totales
<b>Normal</b>	640	321	2239	0	2239
<b>Temprano</b>	448	225	1567	433	2000
<b>Medio</b>	179	90	627	1373	2000
<b>Tardío</b>	13	7	44	1956	2000

Condición B – Imágenes sintéticas: se generaron MRI para los estadios: Temprano, Medio y Tardío, utilizando una red Pix2Pix GAN entrenada con las imágenes reales correspondientes (Isola et al., 2016). Estas imágenes fueron añadidas exclusivamente al conjunto de entrenamiento como se observa en la Tabla 5. Al igual que en el caso de aumento de datos mediante transformaciones geométricas, las imágenes se añadieron exclusivamente al conjunto de entrenamiento, manteniendo los conjuntos de validación y prueba sin cambios.

**Tabla 5.** Aumento de imágenes mediante imágenes sintéticas generadas mediante GAN en el conjunto de entrenamiento para cada uno de los estadios.

	Validación	Prueba	Entrenamiento		
			Reales	Imágenes sintéticas	Totales
<b>Normal</b>	640	321	2239	0.00	2239
<b>Temprano</b>	448	225	1567	475	2042
<b>Medio</b>	179	90	627	1511	2138
<b>Tardío</b>	13	7	44	2151	2195

A continuación, se describe el modelo seleccionado para realizar la generación de imágenes sintéticas.

Pix2Pix está conformada por dos redes principales: un generador (G) y un discriminador (D), cuya interacción permite aprender una función de mapeo entre una imagen de entrada y su correspondiente salida sintética. El generador adopta una estructura compuesta por bloques codificadores y decodificadores enlazados mediante conexiones de salto, lo que facilita la preservación de la información espacial durante la reconstrucción de la imagen. Por su parte, el discriminador sigue un diseño PatchGAN, que evalúa regiones locales de la imagen generada en lugar de analizarla de forma global. Esto permite al modelo capturar mejor los detalles finos, lo que redundará en una mayor fidelidad de las imágenes sintéticas respecto a las reales. La arquitectura general del modelo se ilustra en la Figura 7.

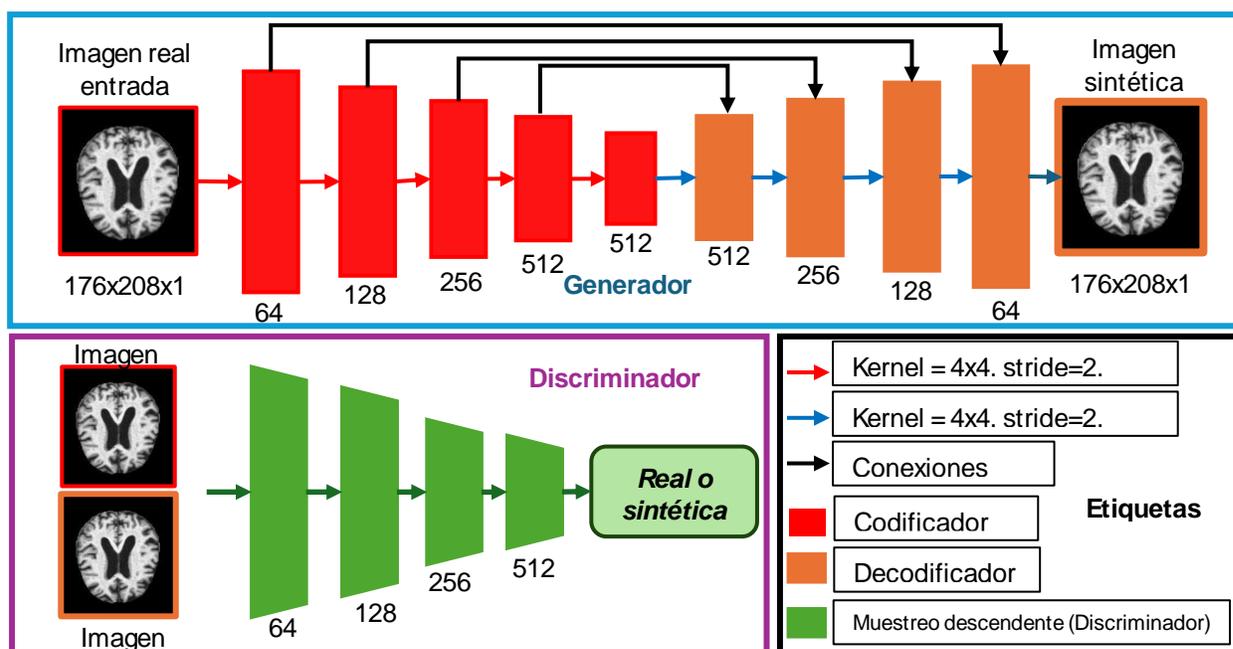


Figura 7. Arquitectura general del modelo Pix2Pix.

Con el objetivo de mejorar la generalización del modelo en entornos con conjuntos de datos reducidos o desbalanceados, se introdujeron dos técnicas de regularización: dropout y detención temprana. El dropout se aplicó al generador, desactivando aleatoriamente el 50 % de sus neuronas durante el entrenamiento para promover la diversidad en las muestras generadas y reducir la dependencia de patrones específicos en los datos de entrada (Mordido et al., 2018). La detención temprana se implementó como un criterio automático que interrumpe el entrenamiento si no se observa mejora en la pérdida del generador durante un número determinado de épocas consecutivas, evitando así el sobreajuste y reduciendo el costo computacional asociado a ciclos innecesarios (Saad et al., 2024).

Estas modificaciones fortalecen el proceso de entrenamiento del modelo GAN, favoreciendo la generación de imágenes sintéticas realistas y variadas, especialmente en escenarios donde el acceso a datos reales es limitado. La Tabla 6 presenta el número total de imágenes sintéticas generadas y utilizadas para cada estadio de la enfermedad.

**Tabla 6.** Cantidad de imágenes sintéticas generadas e incorporadas para cada estadio.

Estadio	No. de sujetos	No. de imágenes	No. de imágenes sintéticas	No. final de imágenes
Normal	100	3200	0	3200
Temprano	70	2240	476	2716
Medio	28	896	1511	2407
Tardío	2	64	2151	2215

## 4.2 Descripción de modelos de CNN

A fin de identificar el modelo de CNN más adecuado para la clasificación multiclase MRI cerebral, se seleccionaron siete modelos propuestos en la literatura, cada uno con distintos niveles de profundidad y estrategias de aprendizaje. Las arquitecturas comparadas fueron: AlexNet (Krizhevsky et al., 2012), DenseNet (Huang et al., 2017), VGGNet (Simonyan & Zisserman, 2014), AHA Net (Illakiya et al., 2023), DEMNet (Murugan et al., 2021), así como los modelos propuestos por Arafa et al. (2024) y AbdulAzeem et al. (2021).

Cada una de estas redes fue implementada, entrenada y evaluada en un entorno centralizado utilizando exactamente el conjunto de datos descrito en la sección 4.1. De esta forma, se garantizó una comparación justa y homogénea entre modelos, libre de sesgos inducidos por diferencias en bases de datos o condiciones experimentales.

### 4.2.1 Hiperparámetros

Cada modelo fue entrenado utilizando una configuración específica de hiperparámetros, adaptada a sus características arquitectónicas y siguiendo, en lo posible, las recomendaciones originales de sus autores. La Tabla 7 resume los hiperparámetros utilizados para cada modelo, incluyendo número de épocas, LR, tamaño del lote, tasa de dropout, optimizador y proporciones de división de los datos.

Cada modelo fue entrenado de forma supervisada utilizando la función de pérdida *función de pérdida de entropía cruzada*, adecuada para clasificación multiclase. Las métricas empleadas para la evaluación de desempeño fueron: precisión global (accuracy), sensibilidad (recall), precisión (precision) y F1-score, calculadas a partir de las predicciones sobre el conjunto de prueba independiente.

**Tabla 7.** Hiperparámetros empleados en los modelos de redes neuronales convolucionales.

	AlexNet	DenseNet	VGGNet	AbdulAzeem	AHANet	Arafa	DEMNet	AHANet-C
<b>No. Épocas</b>	520	100	100	100	50	512	50	50
<b>Tasa de Aprendizaje</b>	0.001	0.001	0.001	0.001	0.001	0.0001	0.001	LCR
<b>Tamaño lote</b>	64	10	64	18	10	64	16	10
<b>Dropout</b>	0.5	0.2	0.5	0.2		0.2	0.2	
<b>Optimizador</b>	SGD	ADAM	ADAM	ADAM	SGD	ADAM	RMS prop	SGD

#### 4.2.2 Modelos evaluados

Con el fin de contrastar soluciones de uso general frente a diseños pensados específicamente para neuroimagen, se entrenaron siete CNN sobre los cortes axiales normalizados descritos en la sección 4.1. Las redes se agrupan en dos categorías como se ilustra en la Tabla 8.

**Tabla 8.** Modelos de redes neuronales convolucionales comparados.

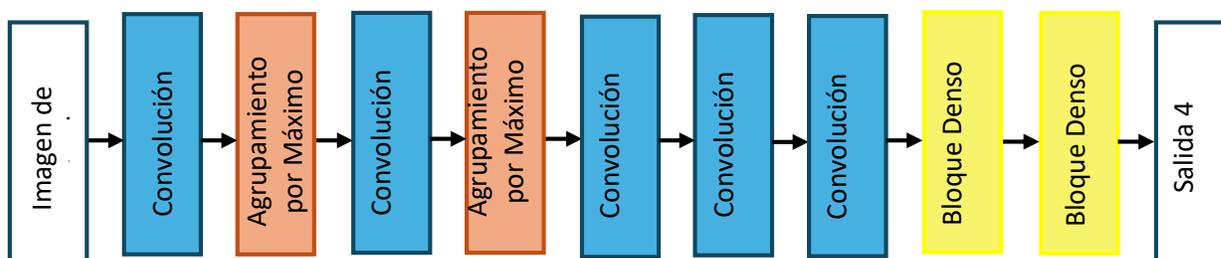
Modelos de CNN	
General	Específica
AlexNet	AbdulAzem
DenseNet	AHANet
VGGNet	Arafa
	DEMNet

Modelos de propósito general como AlexNet, VGG-16 y DenseNet-121 constituyen referencias habituales en estudios comparativos de la AD. Modelos especializados, tales como la CNN propuesta por Arafa et al. (2024), la CNN de AbdulAzeem et al. (2021), DEMNet (Murugan et al., 2021) y AHANet (Illakiya et al., 2023) fueron seleccionados porque incorporan mecanismos como atención, fusión multiescala o estrategias

contra el desbalance, que la literatura reciente señala como favorable para MRI de demencia.

La primera de ellas, AlexNet, introducida por Krizhevsky et al. (2012) fue la pionera en demostrar que una red profunda podía escalar a resoluciones altas sin perder generalización. Su estructura combina cinco capas convolucionales y dos capas densas: la primera convolución utiliza filtros de  $11 \times 11$  con un paso de cuatro píxeles para capturar rasgos de gran escala, mientras que las capas siguientes, con filtros progresivamente menores, abstraen características cada vez más complejas como puede observarse en la Figura 8.

Todas las funciones de activación son ReLU, lo que acelera la convergencia respecto a las funciones sigmoide clásicas, y las dos capas densas superiores emplean dropout al 50% para contener el sobreajuste. En esta tesis la red se entrenó desde cero con SGD (LR inicial  $10^{-3}$ ) por 520 épocas, lo que ofrece un punto de comparación con arquitecturas modernas (Fu'adah et al., 2021).



**Figura 8.** Arquitectura AlexNet.

VGG-16, propuesta por Simonyan y Zisserman (2014), se basa en un principio de simplicidad: apila bloques de dos o tres convoluciones  $3 \times 3$ , separadas por agrupamiento máximo  $2 \times 2$  como se observa en la Figura 9. Esta repetición permite profundizar sin aumentar el número de hiperparámetros y, sobre todo, facilita la reutilización de pesos por aprendizaje por transferencia.

En la implementación utilizada aquí se re-entrenaron únicamente las tres capas densas finales, mientras que los bloques convolucionales se inicializaron con los pesos de ImageNet y se dejaron fijos; la estrategia se acompañó de dropout 0,5 y del optimizador Adam con tasa de  $10^{-3}$  durante cien épocas. Gracias a ello la red conserva su capacidad de extraer rasgos genéricos, pero adapta rápidamente la cabeza clasificadora a las cuatro categorías de EA (Pradhan et al., 2021).

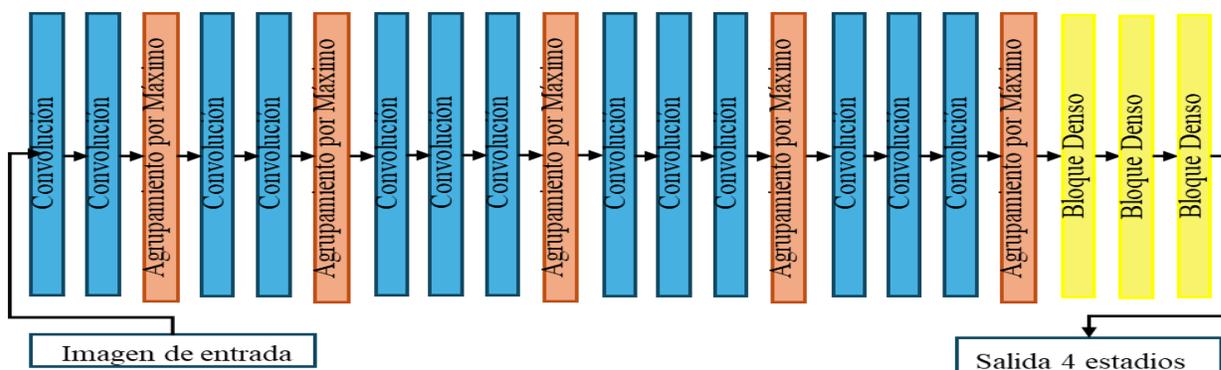


Figura 9. Arquitectura VGGNet.

DenseNet-121, propuesta por Huang et al., (2017), introduce conexiones densas dentro de cada bloque, de modo que cada capa recibe como entrada las salidas concatenadas de todas las capas anteriores como se observa en la Figura 10. Este diseño facilita un flujo sin pérdida de gradiente, reduce el número de parámetros —unos ocho millones en la variante 121 empleada— y, al mismo tiempo, fomenta la reutilización de rasgos finos y gruesos a lo largo de la red. En nuestros experimentos se empleó un crecimiento de canales de 32, transición  $1 \times 1$  seguida de agrupamiento promedio y dropout 0,2; se entrenó con Adam durante cien épocas, ajustando la LR de  $10^{-3}$  a  $10^{-4}$  mediante detención temprana cuando la pérdida de validación se estabilizaba (Pradhan et al., 2021).

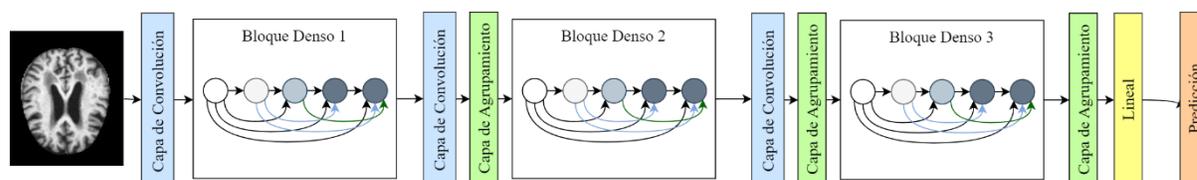
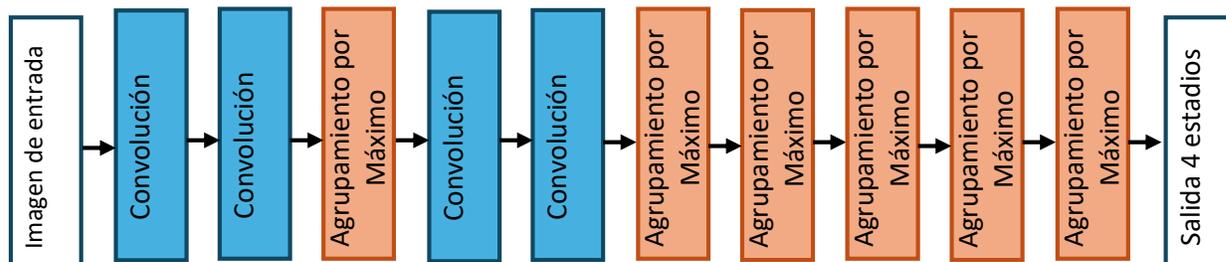


Figura 10. Arquitectura DenseNet.

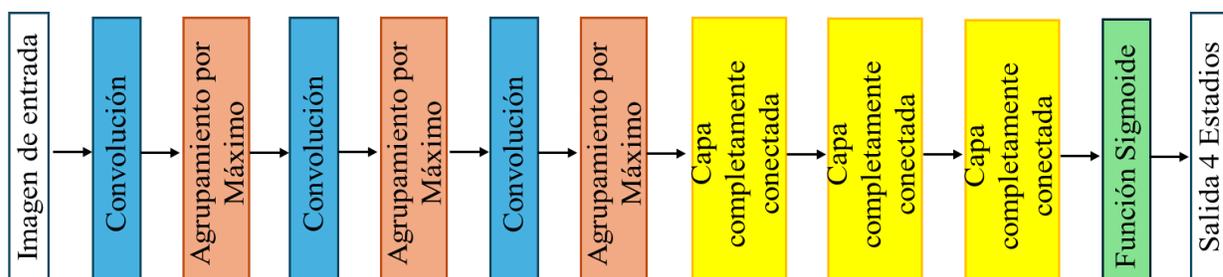
AbdulAzeem et al. (2021) propusieron un modelo sencillo diseñado específicamente para clasificar MRI de pacientes con la AD. Este modelo está compuesto por tres bloques principales, y cada uno realiza tres pasos: primero, una convolución, que ayuda al modelo a detectar patrones en pequeñas regiones de la imagen, como bordes o formas; segundo, una BN, que ajusta automáticamente los valores intermedios para hacer que el aprendizaje sea más rápido y estable; y tercero, una función, que permite pasar solo los valores positivos, ayudando al modelo a enfocarse en los patrones útiles. Después de cada uno de estos bloques, se aplica una técnica que reduce el tamaño de la imagen mientras conserva la información más importante. Al final del modelo, las características aprendidas se envían a dos capas totalmente conectadas, y finalmente a una capa de salida que asigna cada imagen a uno de los cuatro estadios posibles (véase Figura 11). Durante el entrenamiento, los valores iniciales del modelo se eligen mediante Glorot,

que ayuda a evitar errores comunes en las primeras etapas del aprendizaje (Glorot & Bengio, 2010). El modelo se entrenó usando Adam, con parámetros estándar para controlar la velocidad de aprendizaje. Para evaluar qué tan bien aprendía, se usó una función que compara las predicciones del modelo con las respuestas correctas, incorporando además una técnica llamada suavizado de etiquetas, que evita que el modelo se vuelva demasiado rígido en sus decisiones y este a su vez ayuda a generalizar mejor.



**Figura 11.** Arquitectura empleada en AbdulAzeem et al. (2021).

Arafa et al. (2024) propusieron un modelo diseñado como una solución integral y ligera, capaz de entrenarse incluso en estaciones de trabajo clínicas con recursos limitados. El modelo consiste en una red convolucional (CNN) desarrollada desde cero, cuya arquitectura se muestra en la Figura 12. Esta red procesa cortes axiales previamente reescalados a  $64 \times 64$  píxeles y está compuesta por tres bloques, cada uno con una capa convolucional, seguida de BN, una función de activación ReLU y una capa de agrupamiento máximo. Para el entrenamiento, las imágenes se normalizan y la optimización se realiza con el algoritmo Adam, utilizando una LR inicial de  $10^{-3}$ , lotes de tamaño 64 y durante un total de 512 épocas.



**Figura 12.** Arquitectura empleada en Arafa et al. (2024).

El modelo Adaptive Hybrid Attention Network (AHANet) propuesto por Illakiya et al. (Illakiya et al., 2023) parte de una DenseNet-169 como esqueleto extractor y lo refuerza con tres bloques que actúan en cascada como se observa en la Figura 13. Esta red cuenta con: (i) un módulo de Atención Mejorada No Local (ENLA, por sus siglas en inglés) que capta dependencias globales mediante operaciones no locales y realimenta

los canales con una vía residual promediada y soft-max; (ii) un módulo de atención coordinada que preserva información posicional al separar y reagrupar promedios globales sobre los ejes horizontal y vertical antes de recalibrar cada canal; y (iii) un módulo de Agregación Adaptativa de Características (AFA, por sus siglas en inglés) basado en Compresión y Excitación Adaptativa que fusiona de forma guiada las salidas globales y locales, suprimiendo ruido de fondo y potenciando la saliencia espacial (Illakiya et al., 2023).

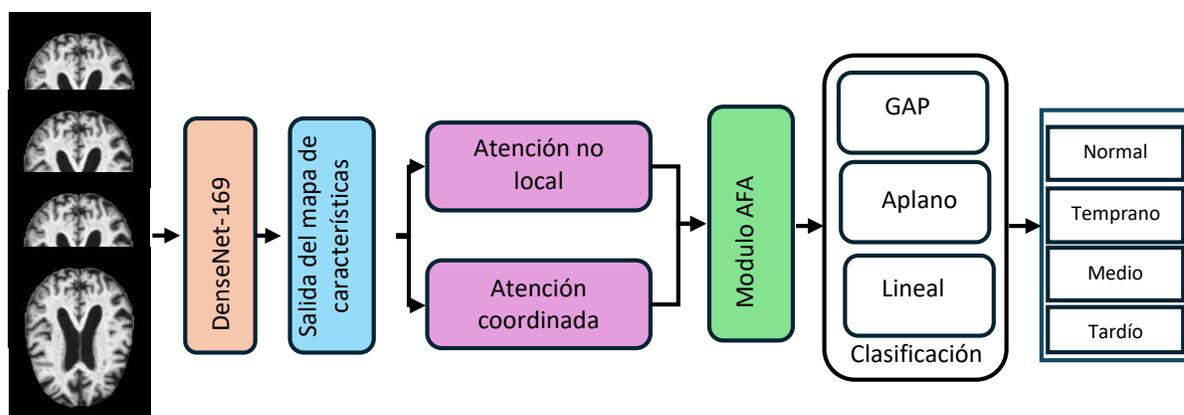


Figura 13. Arquitectura empleada en AHANet. (Illakiya et al., 2023).

DEMNet propuesto por Murugan et al. (Murugan et al., 2021), consiste en una red ligera capaz de ser entrenada sin hardware de alto nivel. DEMNet parte de un flujo 2-D sobre cortes axiales de  $176 \times 176$  píxeles. La arquitectura se organiza en cuatro bloques DEMNET, cada uno integrado por dos convoluciones seguido de BN, ReLU y submuestreo máximo, con profundidades progresivas de 32, 64, 128 y 256 canales; a la salida se aplanan los mapas y se encadenan tres capas densas intercaladas con dropout (0.7-0.5-0.2), para culminar en un soft-max de cuatro salidas (véase Figura 14). El diseño completo suma apenas 4,53 M de parámetros, manteniendo bajo el coste de memoria y cómputo. DEMNet se entrenó durante 50 épocas con RMSprop; LR =  $10^{-3}$ , lote = 16

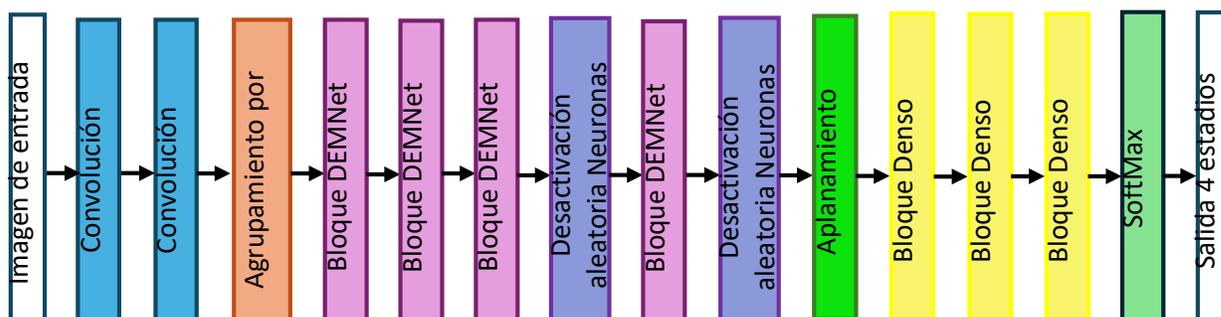


Figura 14. Arquitectura empleada en DEMNet. (Murugan et al., 2021).

### 4.2.3 Optimización de AHANet con bloque de super-convergencia

El modelo AHANet-C es una propuesta desarrollada en el marco de esta tesis, derivada de la arquitectura AHANet, con el objetivo de mejorar la eficiencia del proceso de entrenamiento y reducir la posibilidad de converger a mínimos locales. AHANet-C integra los tres bloques de atención (véase Sección 4.2.2.): ENLA, Atención Coordinada, y AFA, sobre una DenseNet-169. Esta configuración ha demostrado una notable capacidad para extraer y combinar características globales y locales relevantes a partir de MRI en tareas de clasificación multiclase de EA (Illakiya et al., 2023).

AHANet-C incorpora la técnica de super-convergencia, mediante el uso del calendarizador CyclicLR de PyTorch. Esta estrategia consiste en variar de forma cíclica la LR durante el entrenamiento, oscilando entre dos valores extremos definidos por el usuario. Esta oscilación controlada favorece una exploración más amplia del espacio de soluciones durante las primeras etapas del entrenamiento, y una posterior estabilización en las fases finales (Smith, 2017).

La configuración específica utilizada para CyclicLR consiste en:

- Mínimo LR: 0.0001
- Máximo LR: 0.001
- Modo: triangular
- Número de pasos de entrenamiento: ajustado dinámicamente en función del número de lotes por época
- Momentum cíclico: desactivado debido al uso del optimizador SGD sin momentum acumulado

Esta técnica permitió mejorar la tasa de convergencia sin necesidad de aumentar el número de épocas, logrando una mayor estabilidad en el aprendizaje. A diferencia del entrenamiento clásico con LR constante o decreciente, la super-convergencia evita estancamientos prematuros y mejora la generalización del modelo.

El entrenamiento de AHANet-C se llevó a cabo durante 50 épocas, utilizando los siguientes parámetros: Optimizador SGD, tamaño de lote 10, dropout: 0.2 en capas densas. La arquitectura interna del modelo se mantuvo idéntica a la versión original de AHANet la única modificación consistió en incorporar un planificador cíclico CLR que ajusta dinámicamente el valor de la LR en función del número de pasos por época, propiciando una convergencia más rápida y estable. Este cambio en el esquema de optimización se ilustra en la Figura 15, resultando en una mejora significativa en la precisión global del modelo.

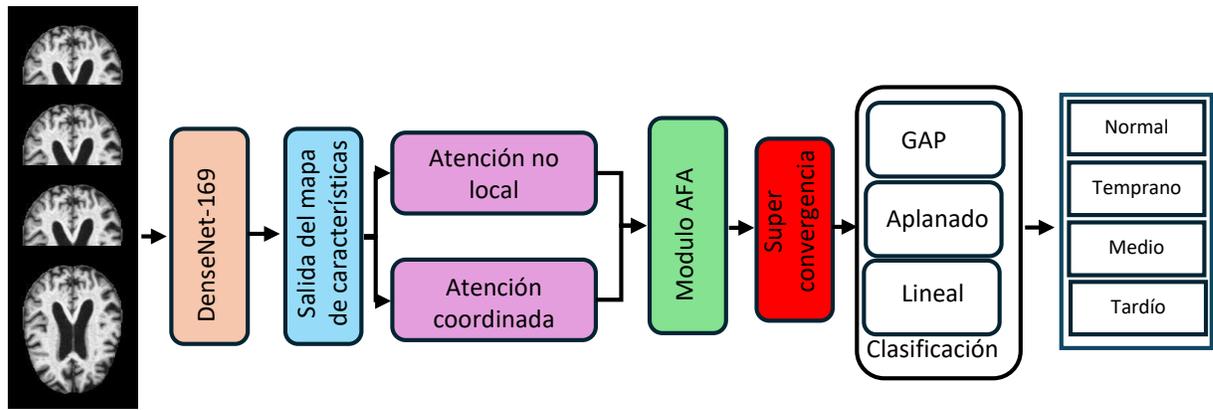


Figura 15. Arquitectura AHANet-C propuesta con esquema de super-convergencia.

### 4.3 Configuración del AF

Con base en los resultados obtenidos en el entorno centralizado, se seleccionó la arquitectura AHANet-C como modelo base para implementar un sistema de clasificación de MRI en un entorno de AF. Se busca evaluar la viabilidad del aprendizaje distribuido entre múltiples clientes, simulando escenarios reales donde los datos médicos se encuentran fragmentados entre distintas instituciones y no pueden ser centralizados por restricciones legales o éticas.

#### 4.3.1 Distribución de los datos entre clientes

Para simular un entorno de AF realista, se diseñó un esquema de distribución de datos donde el conjunto de entrenamiento fue particionado entre distintos clientes. Esta partición simula la situación en la que múltiples instituciones de salud poseen subconjuntos distintos de datos y no pueden compartirlos entre sí por restricciones legales, éticas o de privacidad.

Se trabajaron dos escenarios experimentales principales, diferenciados por el número de clientes: Escenario A: dos clientes y Escenario B: tres clientes, como se muestra en la Figura 16.

En ambos casos, la distribución se realizó de manera equilibrada y estratificada, es decir, cada cliente recibió aproximadamente el mismo número de imágenes y se preservó la proporción original de estadio en cada subconjunto. Esta estrategia busca garantizar una condición de datos IID como punto de partida, permitiendo el análisis del rendimiento sin introducir inicialmente heterogeneidad entre clientes.

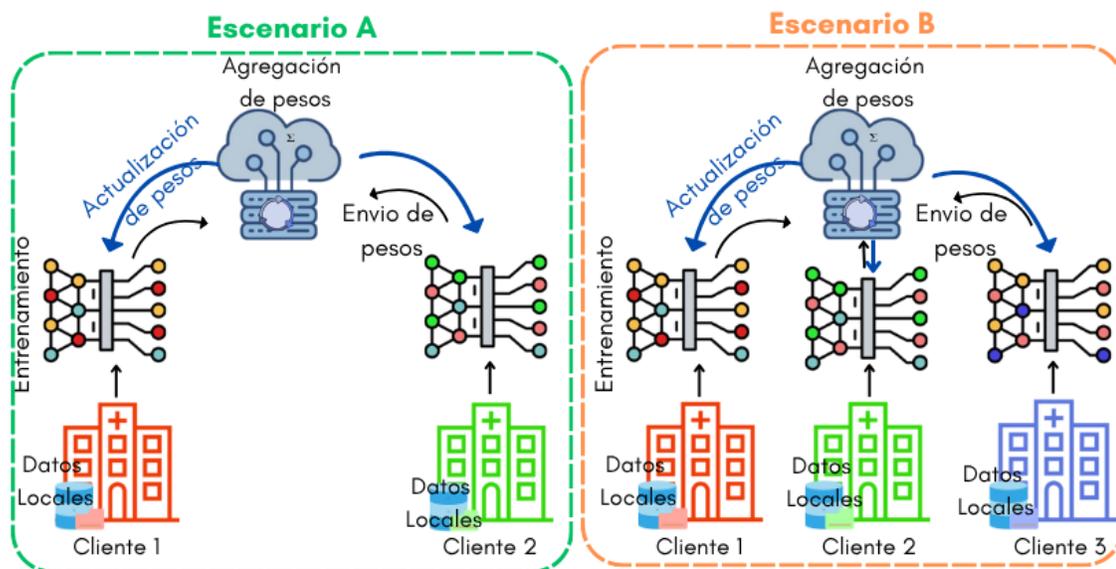


Figura 16. Escenarios A y B de análisis experimental.

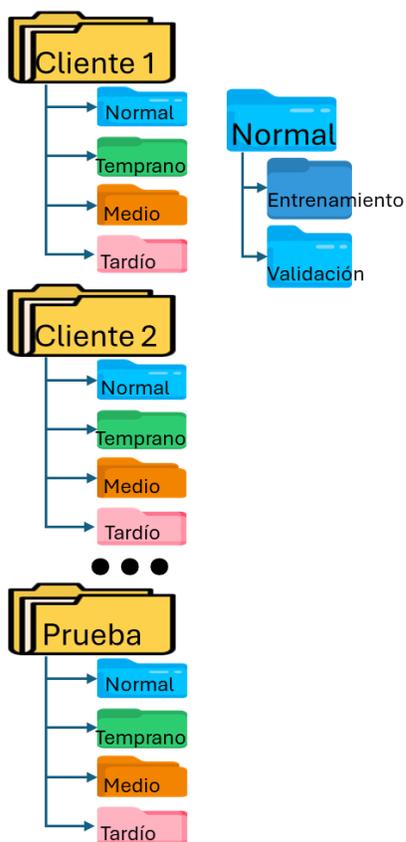


Figura 17. Organización del conjunto de datos.

La organización del conjunto de datos siguió la estructura de carpetas ilustrada en la Figura 17. Cada cliente

contiene su propio conjunto de entrenamiento y validación, con imágenes organizadas por clase. El conjunto de validación se construyó tomando el 20% de las imágenes asignadas a cada cliente, mientras que el 80% restante se utilizó para el entrenamiento local. Esta proporción fue mantenida en todos los experimentos, tanto con como sin imágenes sintéticas.

El conjunto de prueba se mantuvo fuera del entorno federado y no fue accesible por ninguno de los clientes durante el entrenamiento. Este conjunto fue utilizado exclusivamente por el servidor central al final de cada experimento para evaluar el modelo global agregado, asegurando una evaluación objetiva y no sesgada del rendimiento final.

En los experimentos con generación de imágenes sintéticas, las muestras artificiales fueron añadidas únicamente a los subconjuntos de entrenamiento de los clientes. Ni el conjunto de validación ni el conjunto de prueba fueron modificados por estas muestras generadas.

#### 4.3.2 Configuración del entrenamiento federado

Una vez definidos los subconjuntos locales para cada cliente, se procedió a configurar el proceso de entrenamiento federado empleando el modelo AHANet-C como arquitectura base. El proceso de entrenamiento se estructuró en rondas de comunicación. En cada ronda, se llevaron a cabo los siguientes pasos:

1. El servidor central envía el modelo global a todos los clientes participantes.
2. Cada cliente entrena localmente el modelo recibido durante un número definido de épocas utilizando su subconjunto de datos.
3. Los pesos actualizados son enviados nuevamente al servidor.
4. El servidor agrega los modelos locales utilizando FedAvg para actualizar el modelo global.

Este proceso se repitió durante un número determinado de rondas, definido en cada experimento.

Cada cliente entrenó el modelo AHANet-C utilizando la siguiente configuración:

- Tamaño del lote: 10

- Optimización: SGD
- LR: inicial de 0.001, ajustada dinámicamente
- Calendarizador: CyclicLR para implementar la estrategia de super-convergencia
- Dropout: 0.2 en capas densas
- Función de pérdida: pérdida de entropía cruzada
- Evaluación local: al finalizar cada ronda, con conjunto de validación interno

Para actualizar el modelo global después de cada ronda, se utilizó la estrategia FedAvg, que consiste en calcular el promedio ponderado de los parámetros del modelo entrenado en cada cliente. La ponderación se basó en la cantidad de datos locales de cada cliente. No se aplicaron técnicas adicionales de compresión, regularización o protección diferencial de datos, con el fin de mantener una configuración base y observar directamente el impacto del número de rondas, cantidad de clientes y balanceo de datos.

Por otro lado, se diseñaron cuatro configuraciones diferentes en cuanto al número de rondas de comunicación para los escenarios A y B, con el objetivo de analizar cómo la frecuencia de sincronización entre clientes y servidor afecta el rendimiento del modelo global. En todos los casos, el número total de épocas locales por cliente se definió en 50. Por lo tanto, el número de épocas por ronda varió según la configuración específica (véase Tabla 9).

**Tabla 9.** Relación entre el número de rondas de comunicación y el número de épocas.

No. rondas	No. épocas
50	1
25	2
10	5
2	25

Con 50 rondas, se realizó una comunicación al final de cada época, mientras que con 25 rondas los clientes entrenaron 2 épocas consecutivas antes de comunicarse con el servidor. Con 10 y 2 rondas, la comunicación ocurrió cada 5 y 25 épocas antes de sincronizarse con el servidor central, respectivamente. Dicho diseño experimental permite analizar en qué medida afecta la frecuencia de comunicación al

rendimiento del modelo global, siendo un factor crucial en entornos federados donde el costo de sincronización entre nodos es significativo. Esta variación también permite analizar la robustez del modelo AHANet-C bajo condiciones de comunicación limitada, comunes en entornos federados reales.

### 4.3.3 Uso de datos sintéticos

Como parte del análisis experimental, se diseñaron dos variantes principales para cada configuración del entorno federado: 2 o 3 clientes, y diferentes cantidades de rondas de comunicación. Estas variantes permitieron estudiar el impacto del desbalance de clases y del uso de técnicas para corregir el desbalance en los datos en el desempeño del modelo global:

**Escenario 1 – Sin generación de datos sintéticos:** Cada cliente entrenó el modelo AHANet-C utilizando exclusivamente imágenes reales. Este escenario representa la situación original del conjunto de datos, con una distribución naturalmente desbalanceada entre estadios. De igual manera, no se aplicaron técnicas de aumento de datos.

**Escenario 2 – Con generación de datos sintéticos:** En este caso, los conjuntos de entrenamiento locales fueron complementados con imágenes sintéticas generadas previamente mediante Pix2Pix GAN, con el objetivo de balancear los estadios. Las imágenes artificiales fueron distribuidas equitativamente entre los clientes, en función del número de muestras reales presentes en cada cliente.

En ambos escenarios, los conjuntos de validación y prueba se mantuvieron inalterados y compuestos únicamente por imágenes reales, con el fin de asegurar una evaluación objetiva y comparable entre experimentos. La integración de imágenes generadas se realizó exclusivamente en el entrenamiento local de cada cliente, donde las muestras sintéticas fueron incorporadas en su subconjunto de entrenamiento. No se modificaron las proporciones del conjunto de validación ni se incluyeron imágenes sintéticas en el conjunto de prueba. Aunque el servidor central no tuvo acceso directo a dichas imágenes generadas, su efecto se reflejó indirectamente en el modelo global a través de la actualización de los pesos locales. Esto permitió evaluar si el balanceo local por cliente influye positivamente en el rendimiento global del modelo, aun sin compartir explícitamente los datos sintéticos con el servidor.

La elección de evaluar ambos escenarios surge de los resultados obtenidos en la etapa centralizada

(Sección 4.2.5), donde se observó que el uso de MRI sintéticas generadas por GAN proporcionaba mejores resultados que el aumento de datos tradicional. Por esta razón, solo la estrategia de generación sintética fue considerada en el entorno federado, como una forma de simular balanceo local en contextos descentralizados. La Tabla 10 resume los dieciséis escenarios experimentales definidos en el análisis experimental. Cada combinación representa una configuración específica del entorno AF, variando en tres aspectos clave: el número de clientes, el número de rondas de comunicación y la inclusión de imágenes sintéticas generadas mediante GAN para el balanceo de estadios.

**Tabla 10.** Configuración de los experimentos ejecutados en el entorno federado.

Escenario	No. de clientes	No. de rondas	Imágenes Sintéticas
E1	2	50	-
E2	2	50	✓
E3	3	50	-
E4	3	50	✓
E5	2	25	-
E6	2	25	✓
E7	3	25	-
E8	3	25	✓
E9	2	10	-
E10	2	10	✓
E11	3	10	-
E12	3	10	✓
E13	2	2	-
E14	2	2	✓
E15	3	2	-
E16	3	2	✓

#### 4.4 Métricas de evaluación

Con el objetivo de evaluar de forma integral el desempeño del modelo AHANet-C en cada uno de los escenarios experimentales definidos en la Tabla 10, se utilizaron métricas estándar para problemas de

clasificación multi clase, es decir, la capacidad de las CNN para distinguir entre los diferentes estadios de AD. Para cada experimento, se calcularon las siguientes métricas de desempeño:

Exactitud global: proporción de predicciones correctas respecto al total de muestras evaluadas.

$$Exactitud = \frac{\sum_{i=1}^C TP_i}{\sum_{i=1}^C (TP_i + FP_i + FN_i + TN_i)} \quad (13)$$

donde  $TP$  son los verdaderos positivos: predicciones correctas de la clase positiva.  $FP$  son los falsos positivos: casos incorrectamente clasificados como positivos.  $FN$  son los falsos negativos: casos reales no detectados por el modelo.  $TN$  son los verdaderos negativos: casos correctamente identificados como negativos. Y  $C$  es el número total de clases o estadios.

Precisión: proporción de verdaderos positivos sobre el total de predicciones positivas para cada estadio.

$$Precisión_i = \frac{TP_i}{TP_i + FP_i} \quad (14)$$

Sensibilidad o Recall: proporción de verdaderos positivos sobre el total de verdaderos casos en cada estadio.

$$Recall_i = \frac{TP_i}{TP_i + FN_i} \quad (15)$$

F1-score: media armónica entre precisión y sensibilidad, útil para evaluar el rendimiento en escenarios con desbalance.

$$F1 = \frac{2 \cdot Precisión_i \cdot Recall_i}{Precisión_i + Recall_i} \quad (16)$$

Pérdida: valor de la función de pérdida de entropía cruzada al final de cada ronda de comunicación.

$$\mathcal{L} = - \sum_{i=1}^c y_i \log(\hat{y}_i) \quad (17)$$

donde  $y_i$  es la etiqueta verdadera y  $\hat{y}_i$  es la probabilidad predicha por el modelo para la clase  $i$ .

Los resultados fueron registrados en dos niveles: evaluación global centralizada y evaluación local por cliente. La evaluación global centralizada se realizó exclusivamente sobre el conjunto de prueba externo, al finalizar todas las rondas de comunicación. Este conjunto fue utilizado por el servidor para evaluar el modelo global agregado y calcular las métricas finales por escenario. Por otro lado, la evaluación local por cliente evalúa el modelo en cada ronda de comunicación, considerando las métricas calculadas por cada uno de los clientes sobre su conjunto de validación local. Estos valores permiten analizar el comportamiento individual de los clientes a lo largo del entrenamiento federado.

## 4.5 Configuración Experimental

Todos los modelos de CNN fueron implementados en Python 3.12.14, utilizando el marco de trabajo PyTorch 2.4.0, en un servidor de alto rendimiento con las siguientes características:

- Sistema operativo: Debian 6
- Procesadores: 2 × Intel Xeon Gold 5420+
- Núcleos totales: 56
- Memoria RAM: 100 GB
- GPU: NVIDIA A100-PCIE-40GB (40 GB de VRAM)

La implementación del entorno de AF se llevó a cabo utilizando el marco de trabajo Flower (FLWR), una herramienta flexible y extensible desarrollada en Python para el diseño y simulación de algoritmos de AF. Flower permite ejecutar múltiples clientes en una misma máquina, lo que resulta especialmente útil para pruebas controladas (Beutel et al., 2020). Además, el entrenamiento local en cada cliente fue acelerado mediante el uso de GPU, y la simulación de nodos se gestionó mediante la ejecución de procesos

independientes en paralelo, cada uno representando un cliente federado. La coordinación del servidor central, responsable de la agregación de modelos, también se ejecutó en el mismo entorno.

El entrenamiento de las CNNs fue acelerado mediante cómputo en GPU, aprovechando extensiones CUDA para reducir significativamente los tiempos de ejecución. Las principales librerías utilizadas fueron *torchvision* para la gestión de modelos y transformaciones de imágenes, *numpy* para operaciones vectorizadas, *matplotlib* y *seaborn* para visualización de resultados, y *scikit-learn* para el cálculo de métricas de evaluación y construcción de matrices de confusión.

## Capítulo 5. Resultados

---

Este capítulo presenta los resultados obtenidos durante el desarrollo experimental del estudio, así como el análisis de estos. Se abordan tanto los experimentos realizados en un entorno centralizado como aquellos llevados a cabo bajo un esquema de AF, con el objetivo de evaluar el desempeño del modelo AHANet-C en distintos contextos y configuraciones.

La Sección 5.1 describe la comparación entre diferentes arquitecturas de CNNs empleadas en un entorno centralizado, incluyendo una discusión comparativa con estudios previos, la evaluación detallada del modelo AHANet-C y los experimentos de aumento de datos. Posteriormente, la Sección 5.2 se enfoca en los resultados obtenidos en el entorno federado, donde se analizan métricas globales por escenario, la influencia de diversos factores experimentales, el comportamiento por estadio de la EA y una comparación con trabajos relacionados en la literatura.

### 5.1 Comparación de arquitecturas CNN en entorno centralizado

Como fase inicial del diseño experimental, se evaluaron siete arquitecturas de CNN ampliamente utilizadas y propuestas en la literatura para tareas de clasificación de imágenes médicas: AlexNet, DenseNet, VGGNet, ArafaNet, AbdulAzeemNet, AHANet y DEMNet. Estos experimentos se realizaron en un entorno centralizado, sin técnicas de aumento de datos ni generación de imágenes sintéticas, garantizando condiciones homogéneas para todos los modelos.

**Tabla 11.** Resultados globales de las siete arquitecturas de CNN evaluadas en el entorno centralizado.

Modelo	Precisión (%)	Tiempo (s)
AlexNet	95.12	2,843.76
DenseNet	96.39	3,416.87
VGGNet	96.38	736.92
AbdulAzeem	98.20	<b>263.88</b>
AHANet	<b>98.40</b>	1,214.87
Arafa	91.01	1,598.29
DEMNet	96.39	335.52

El objetivo fue identificar la arquitectura con mejor desempeño global que sirviera como base para los

experimentos posteriores en entornos centralizado optimizado y federado. La Tabla 11 presenta los resultados de precisión y tiempo de entrenamiento de cada arquitectura CNN.

De acuerdo con la Tabla 11, AHANet y el modelo propuesto por AbdulAzeem et al. (2021) presentan los mejores resultados en precisión global con 98.40% y 98.20% , respectivamente. Estos resultados se alinean con lo reportado en la literatura, donde ambos modelos destacan por incorporar mecanismos que enriquecen la extracción de características: AHANet mediante bloques de atención (ENLA, Atención Coordinada, AFA), y AbdulAzeemNet mediante el preprocesamiento con umbralización adaptativa que refuerza el contraste de las regiones cerebrales clave.

Por otro lado, VGGNet y DenseNet ofrecieron un rendimiento aceptable (96.38% y 96.39%, respectivamente), con un costo computacional heterogéneo VGGNet se posiciona como un modelo eficiente en cuanto a tiempo de entrenamiento (736.92 s), mientras que DenseNet, a pesar de su precisión, presentó el mayor tiempo (3416.87 s), dado el uso de conexiones densas entre capas.

DEMNet obtuvo la misma precisión que DenseNet (96.39%) con un tiempo de entrenamiento 90% menor. Sin embargo, como se detalla en la Tabla 13, presenta un rendimiento no uniforme en todos los estadios, especialmente en el segundo estadio. En contraste, AlexNet, pese a ser ampliamente utilizada en la literatura, mostró un desempeño limitado (95.12%) y un tiempo de entrenamiento de 2843.76 s, lo que refleja las limitaciones de su diseño frente a arquitecturas específicamente diseñadas para la clasificación de AD. Por último, AfaNet presentó una precisión de 91.01% y un tiempo de entrenamiento de 1598.29 s, lo que sugiere que su arquitectura no se adapta bien al conjunto de datos y tarea específica abordada en esta investigación, a pesar de haber sido propuesta para clasificación de EA en otros contextos.

La relación entre precisión y costo computacional es clave. Modelos como AbdulAzeemNet y DEMNet podrían ser opciones atractivas en entornos con recursos limitados; sin embargo, el análisis por estadio evidencia que estos modelos no siempre logran un rendimiento homogéneo, lo que puede ser un factor crítico en aplicaciones clínicas. Por estas razones, AHANet fue seleccionado como modelo base para el desarrollo de la versión optimizada (AHANet-C) y su posterior aplicación en el entorno AF, priorizando la precisión y la consistencia sobre la eficiencia computacional.

#### La

Tabla 12 muestra la precisión, recall y F1-score por estadio, lo que permite analizar la capacidad de cada arquitectura para diferenciar entre los distintos estadios de AD. Los resultados por estadio refuerzan la

elección de AHANet, que no solo alcanzó una alta precisión global, sino que demostró un rendimiento aceptable en todos los estadios.

AbdulAzeemNet mostró un desempeño aceptable en los estadios más críticos, aunque con ligeras caídas en consistencia comparado con AHANet. Por otro lado, DenseNet y VGGNet destacaron especialmente en el estadio Tardío, logrando un 100% de recall y F1-score; sin embargo, presenta un rendimiento irregular en el estadio Temprano. AlexNet y DEMNet evidenciaron dificultades para detectar el estadio Tardío (F1-score de 80.00% y 66.67%, respectivamente), lo que plantea preocupaciones sobre su aplicabilidad en contextos donde la precisión diagnóstica de estadios avanzados es crucial. Arafanet, pese a su diseño orientado a AD, tuvo el rendimiento más bajo de todos los modelos, tanto en precisión global como por estadio, lo que sugiere una falta de adaptabilidad a los datos y configuraciones de esta investigación.

**Tabla 12.** Resultados por estadio de las siete arquitecturas de redes neuronales convolucionales evaluadas en el entorno centralizado.

Estadio	Normal			Temprano			Medio			Tardío		
	Precisión (%)	Recall (%)	F1 score (%)	Precisión (%)	Recall (%)	F1 score (%)	Precisión (%)	Recall (%)	F1 score (%)	Precisión (%)	Recall (%)	F1 score (%)
AlexNet	96.74	95.19	95.96	87.65	95.24	91.29	96.10	83.15	89.16	100.00	66.67	80.00
VGGNet	96.48	<b>100.00</b>	98.21	89.05	90.00	86.42	<b>100.00</b>	94.44	97.14	100.00	100.00	100.00
DenseNet	<b>99.63</b>	87.83	93.36	85.92	<b>99.58</b>	92.25	<b>100.00</b>	<b>97.78</b>	<b>98.88</b>	100.00	100.00	100.00
Abdul	98.82	97.67	97.35	96.49	98.21	97.35	97.50	97.50	97.50	100.00	100.00	100.00
AHANet	99.07	97.85	<b>98.46</b>	<b>96.95</b>	99.11	<b>98.02</b>	97.59	96.42	97.00	100.00	100.00	100.00
Arafa	91.79	96.90	94.28	89.50	88.69	89.09	93.83	86.36	89.94	88.93	87.85	85.73
DEMNet	95.76	97.53	96.64	93.27	93.27	93.27	96.51	92.22	94.32	100.00	50.00	66.67

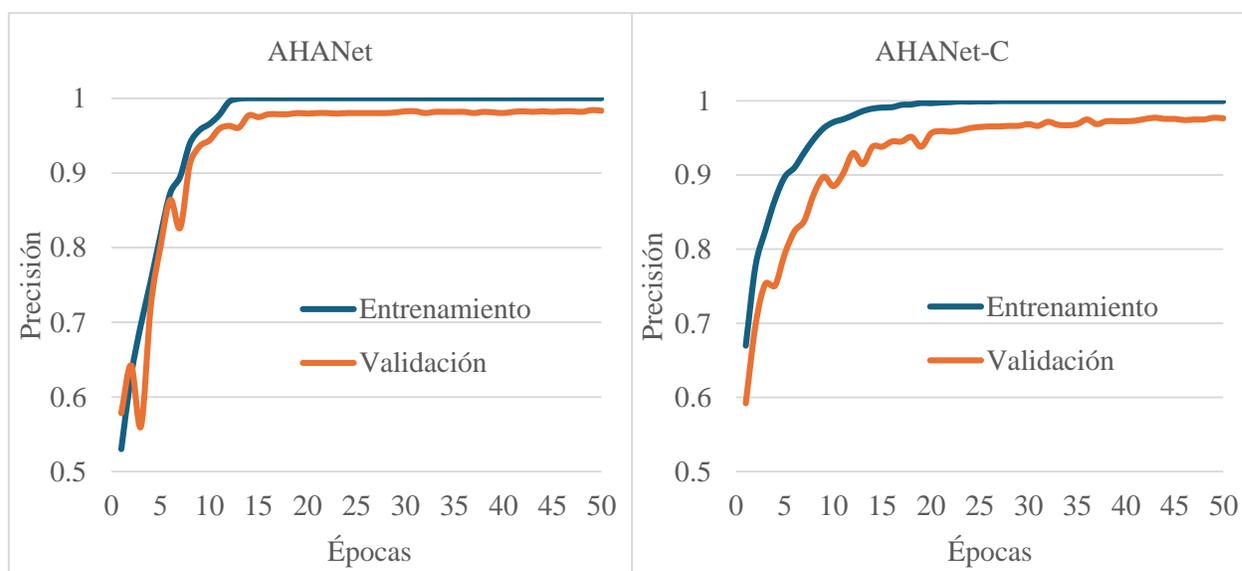
### 5.1.1 Evaluación del modelo AHANet-C en entorno centralizado

Tras identificar a AHANet como la arquitectura con mejor desempeño global y por estadio, se procedió a desarrollar una versión optimizada denominada AHANet-C. Esta variante incorpora un bloque de super-convergencia, implementado mediante el calendarizador CyclicLR en PyTorch, que ajusta dinámicamente la LR durante el entrenamiento. El objetivo de esta optimización fue acelerar la convergencia, mejorar la estabilidad del proceso de aprendizaje y potenciar la capacidad del modelo para generalizar, sin necesidad de incrementar el número de épocas.

La Tabla 13 presenta los resultados globales obtenidos por AHANet y AHANet-C. AHANet alcanzó una precisión promedio del 98.40% con un tiempo de entrenamiento de 1214.87 segundos. Por su parte, AHANet-C logró una precisión superior del 98.89% y redujo el tiempo de entrenamiento a 508.61 segundos, evidenciando la efectividad de la super-convergencia tanto en rendimiento como en la eficiencia computacional.

**Tabla 13.** Comparación de resultados globales entre AHANet y AHANet-C.

Modelo	Precisión (%)	Tiempo (s)
AHANet	98.40	1214.87
AHANet-C	98.89	508.61



**Figura 18.** Curvas de aprendizaje de los modelos AHANet y AHANet-C.

La Figura 18 presenta las curvas de aprendizaje para ambos modelos. Se observa que AHANet-C converge de forma más rápida y estable, alcanzando altos niveles de precisión en menos épocas efectivas gracias al ajuste dinámico de LR. Este comportamiento es coherente con lo esperado al implementar super-convergencia y refuerza la decisión de adoptar AHANet-C como modelo base para los experimentos en entorno federado.

La Tabla 14 desglosa la precisión alcanzada por ambos modelos en cada estadio. Los resultados muestran que AHANet-C no solo mantuvo el alto rendimiento de AHANet, sino que lo superó en los estadios Temprano y Medio, los cuales son especialmente relevantes para el diagnóstico temprano y el seguimiento de la progresión de la AD.

**Tabla 14.** Comparación por estadio entre AHANet y AHANet-C.

Estadio	AHANet Precisión (%)	AHANet-C Precisión (%)
Normal	99.07	98.45
Temprano	96.95	99.02
Medio	97.59	98.10
Tardío	100.00	100.00

Este desempeño reforzó la selección de AHANet-C como modelo para el AF, al ofrecer no solo un excelente rendimiento global, sino también un equilibrio adecuado entre precisión en estadios críticos y eficiencia computacional. La reducción del tiempo de entrenamiento en un 58% en comparación con AHANet, es relevante para aplicaciones de AF, donde el entrenamiento debe realizarse en dispositivos locales con recursos limitados y en tiempos razonables para facilitar la sincronización y agregación de los modelos.

### 5.1.2 Experimentos con aumento de datos y generación de imágenes sintéticas en entorno centralizado

Tras optimizar el modelo AHANet mediante la incorporación del bloque de super-convergencia generando AHANet-C, se realizaron dos experimentos adicionales en entorno centralizado con el propósito de abordar el desbalance de estadios, en particular la baja representación del estadio Tardío. Este desbalance constituye un desafío común en aplicaciones clínicas reales, donde los casos de etapas avanzadas o muy tempranas de la enfermedad son menos frecuentes que los casos normales o de etapas intermedias.

Los experimentos evaluaron el impacto de dos estrategias distintas para balancear el conjunto de entrenamiento:

- Aumento de datos tradicional: mediante transformaciones geométricas simples aplicadas a las imágenes reales (rotaciones, escalados, desplazamientos y volteos).
- Generación de imágenes sintéticas: utilizando una GAN para producir nuevas muestras del estadio subrepresentado.

El objetivo fue determinar cuál de estas estrategias permitía mejorar el rendimiento del modelo sin comprometer su capacidad de generalización ni inducir un sobreajuste, y cuál sería más viable para implementarse en AF.

La Tabla 15 resume los resultados globales de AHANet-C bajo las tres condiciones experimentales: conjunto original (sin balanceo), conjunto balanceado mediante aumento de datos, y conjunto balanceado mediante imágenes sintéticas.

**Tabla 15.** Comparación global de AHANet-C bajo distintas estrategias de balanceo.

	<b>Precisión (%)</b>	<b>Tiempo (s)</b>
<b>Conjunto de datos original</b>	98.89	508.61
<b>Aumento de datos</b>	93.90	4,031.52
<b>Imágenes Sintéticas</b>	98.76	747.25

Los resultados muestran que el uso de imágenes sintéticas permitió mantener una precisión global prácticamente equivalente al del conjunto sin balanceo (98.76% vs. 98.89%), logrando además un equilibrio aceptable entre rendimiento y tiempo de entrenamiento. En cambio, el aumento de datos tradicional no solo incrementó de forma considerable el tiempo de entrenamiento —87.39% superior al escenario normal—, sino que produjo un descenso notable en la precisión global (93.90%).

Esta diferencia se explica por las limitaciones inherentes a las técnicas de aumento de datos tradicionales: aunque generan mayor cantidad de muestras, estas no introducen variaciones sustanciales en el patrón estructural de las imágenes, y en ocasiones inducen un sobreajuste a las transformaciones aplicadas. Por el contrario, las imágenes sintéticas generadas mediante GAN enriquecieron el conjunto con patrones

realistas y variados, capturando de forma más efectiva las características del estadio Tardío y evitando un impacto negativo en los estadios restantes.

El tiempo de entrenamiento asociado al uso de imágenes sintéticas (747.25 s) fue mayor al del escenario sin balanceo, lo que resulta natural al incrementarse el volumen de datos. Sin embargo, este tiempo es razonable en comparación con el costo computacional del aumento de datos tradicional y es compatible con aplicaciones en entornos distribuidos. La Tabla 16 detalla la precisión alcanzada por AHANet-C en cada estadio bajo las tres condiciones experimentales.

**Tabla 16.** Desempeño por estadio de AHANet-C bajo distintas estrategias de balanceo.

Estadio	Conjunto de datos original Precisión (%)	Aumento de datos Precisión (%)	Imágenes sintéticas Precisión (%)
Normal	98.45	95.52	98.76
Temprano	99.02	97.18	99.55
Medio	98.10	95.40	96.74
Tardío	100.00	87.50	100.00

El escenario con el conjunto de datos original mostró un buen desempeño global y por estadio, con un 100% de precisión en el estadio Tardío y una precisión global de 98.89%. Sin embargo, el uso de imágenes sintéticas permitió no solo mantener estos resultados en el estadio Tardío, sino también mejorar la precisión a 99.55% en el estadio Temprano. Esto sugiere que las imágenes generadas por GAN aportaron una variabilidad útil al modelo para refinar la clasificación en estadios con diferencias morfológicas sutiles.

En contraste, el aumento de datos tradicional obtuvo una reducción en rendimiento en los estadios subrepresentados. Este resultado refuerza la idea de que las transformaciones geométricas no aportan información estructural relevante al modelo y que incluso pueden inducir sobreajuste a patrones artificiales. De esta forma, la inclusión de imágenes sintéticas se presenta como una estrategia preventiva ante escenarios clínicos o federados donde el desbalance puede variar entre clientes y donde es crucial mantener un rendimiento estable aun con datos limitados

La evaluación conjunta de las tres estrategias permitió confirmar que el diseño original de AHANet-C ofrece un rendimiento robusto y que la generación de imágenes sintéticas mediante GAN agrega un valor adicional al incrementar la capacidad del modelo para generalizar en estadios subrepresentados, sin penalizar el rendimiento en el resto ni incrementar el costo computacional. Por estas razones, se optó por emplear imágenes sintéticas como estrategia de balanceo en los experimentos del entorno federado.

## 5.2 Comparación de los diferentes escenarios de AF

En esta sección se presentan y discuten los resultados obtenidos al implementar el modelo AHANet-C en un entorno de AF bajo los 16 escenarios experimentales definidos (ver Tabla 10). Los resultados se organizaron de manera jerárquica para facilitar su interpretación y análisis, priorizando la claridad y la comparación entre escenarios.

### 5.2.1 Resultados globales por escenario

La Tabla 17 presenta los resultados globales obtenidos en los 16 escenarios experimentales definidos para el entorno de AF. Cada escenario corresponde a una combinación específica de número de clientes, número de rondas de comunicación y la incorporación de imágenes sintéticas para balancear el conjunto de datos.

Las métricas reportadas incluyen la precisión global alcanzado al evaluar el modelo agregado en el servidor central sobre el conjunto de prueba externo, así como el tiempo total de entrenamiento de cada experimento.

Los resultados muestran que el número de rondas de comunicación tuvo un impacto notable en el rendimiento global. Los escenarios con 50 y 25 rondas alcanzaron los mayores valores en precisión, destacando particularmente los que utilizaron imágenes sintéticas:

- E2 (2 clientes, 50 rondas, sintéticas): 99.14%
- E8 (3 clientes, 25 rondas, sintéticas): 99.03%

Esto indica que un mayor número de rondas permite una mejor sincronización y refinamiento del modelo global, y que el uso de imágenes sintéticas contribuye a mejorar el aprendizaje en los estadios subrepresentados.

Por el contrario, los escenarios E13 a E16 con solo 2 rondas mostraron un claro deterioro en rendimiento con valores por debajo del 65%. Esto podría atribuirse a un desbalance entre la autonomía de los clientes

y la falta de suficiente comunicación con el servidor para lograr una adecuada agregación de los modelos locales.

**Tabla 17.** Precisión global y tiempo de entrenamiento en los 16 escenarios federados.

Escenario	No. de clientes	No. de rondas	Imágenes Sintéticas	Precisión (%)	Tiempo (s)
E1	2	50	-	97.49	7,370.90
E2	2	50	✓	<b>99.14</b>	14,641.80
E3	3	50	-	96.59	3,448.56
E4	3	50	✓	98.57	5,395.74
E5	2	25	-	97.94	7,365.31
E6	2	25	✓	98.61	10,990.64
E7	3	25	-	96.99	2,761.79
E8	3	25	✓	99.03	4,822.72
E9	2	10	-	97.91	4,877.57
E10	2	10	✓	98.96	14,266.25
E11	3	10	-	97.09	2,417.23
E12	3	10	✓	98.53	4,430.76
E13	2	2	-	93.83	6,964.79
E14	2	2	✓	62.19	14,311.45
E15	3	2	-	62.81	2,113.88
E16	3	2	✓	53.77	4,099.66

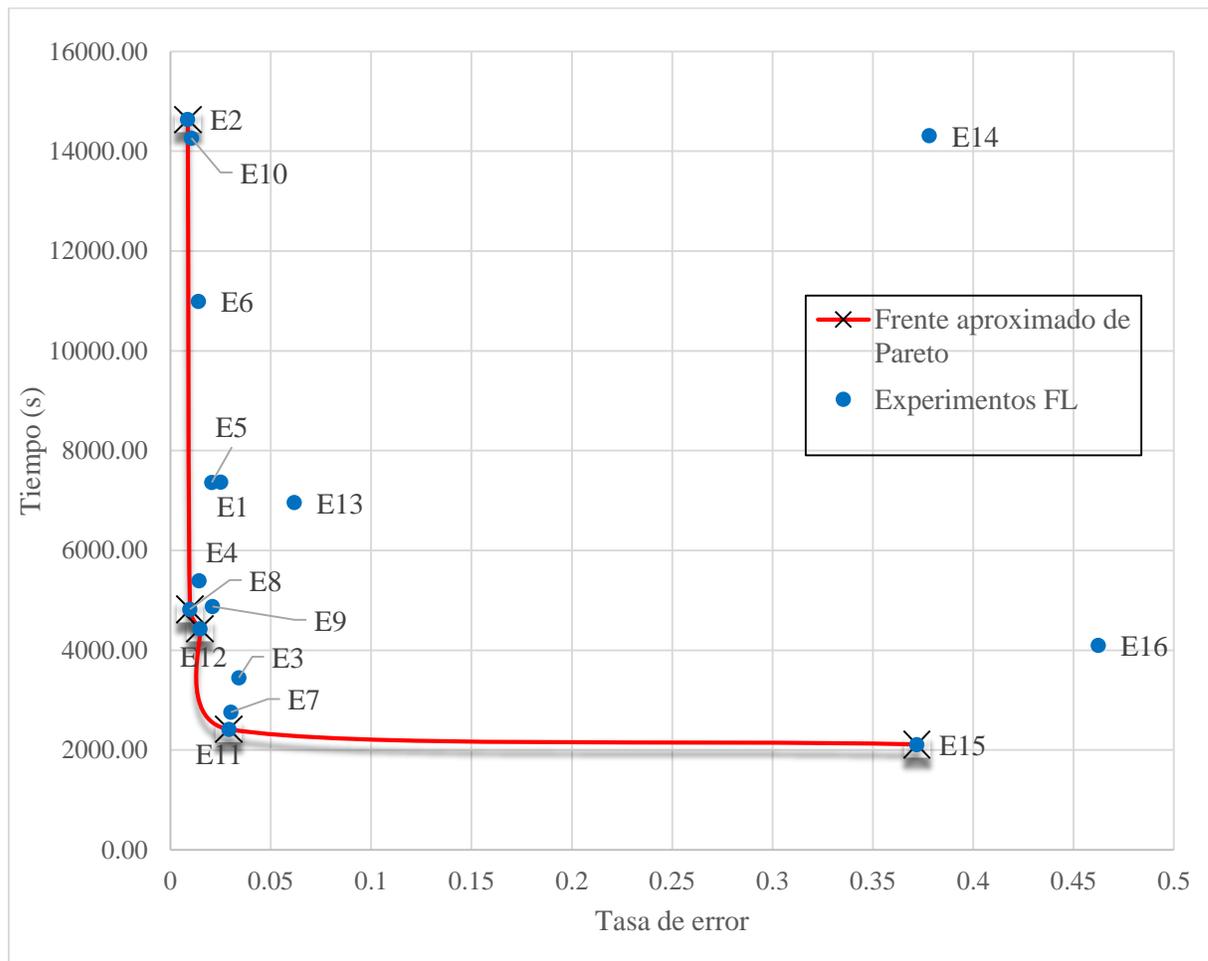
En cuanto al tiempo de entrenamiento, los escenarios con imágenes sintéticas implicaron un mayor costo computacional debido al mayor volumen de datos locales en cada cliente. Sin embargo, este incremento fue proporcional al beneficio en precisión en la mayoría de los casos, salvo en los escenarios con pocas rondas, donde el tiempo adicional no se tradujo en un mejor rendimiento.

### 5.2.1.1 Análisis multiobjetivo: error y tiempo de ejecución

Como complemento al análisis de los resultados globales por escenario, se construyó un frente aproximado de Pareto con el objetivo de identificar los escenarios que ofrecen un compromiso óptimo entre precisión y eficiencia computacional. En este análisis, se consideraron dos objetivos:

- Tasa de error
- Tiempo total de ejecución (s), correspondiente a la suma del tiempo de entrenamiento de todos los clientes y de las rondas de comunicación.

La Figura 19 muestra el frente aproximado de Pareto obtenido. Cada punto representa un escenario, y el frente está conformado por aquellos escenarios no dominados: es decir, aquellos para los que no existe otro escenario que sea mejor simultáneamente en ambos objetivos, menor tasa de error y menor tiempo.



**Figura 19.** Frente aproximado de Pareto para los 16 experimentos en el entorno federado.

El análisis multiobjetivo revela que los escenarios que logran un compromiso óptimo entre baja tasa de error y tiempo de ejecución se encuentran, en su mayoría, entre aquellos que:

- Usan imágenes sintéticas (E2, E8, E12), y

- Tienen un número intermedio o alto de rondas de comunicación (10, 25, 50 rondas).

El escenario E2 (2 clientes, 50 rondas, imágenes sintéticas) presenta la tasa de error más baja (0.86%) del frente, aunque con el mayor tiempo de ejecución, reflejando el costo computacional asociado a un alto número de rondas y uso de imágenes generadas. Por el contrario, E11 y E15 muestran soluciones con tiempos más bajos, pero a costa de una mayor tasa de error, especialmente en el caso de E15 (37.19%), lo que lo convierte en una opción menos viable para aplicaciones donde la precisión es prioritaria.

El frente aproximado de Pareto, por lo tanto, ofrece una herramienta para seleccionar configuraciones que balanceen de forma adecuada precisión y eficiencia computacional, de acuerdo con las necesidades específicas del entorno de aplicación.

### 5.2.2 Análisis comparativo por factor experimental

Para analizar el impacto de las variables experimentales principales (número de rondas de comunicación, número de clientes y uso de imágenes sintéticas), se calcularon los promedios de precisión y tiempo de ejecución agrupando los 16 escenarios de acuerdo con cada uno de estos factores. Esta comparación permite identificar tendencias generales y evaluar qué variable tuvo mayor influencia sobre el rendimiento y el costo computacional del modelo global. La Tabla 18 presenta los resultados promedio obtenidos al finalizar los 16 escenarios experimentales definidos para el entorno de AF enfocándose específicamente en el número de rondas de comunicación.

**Tabla 18.** Promedio por número de rondas.

No. de rondas	Precisión promedio (%)	Tiempo promedio (s)
50	97.95	6940.00
25	98.14	5725.12
10	98.12	6422.20
2	68.15	6862.20

El análisis muestra que los escenarios con 25 y 50 rondas alcanzaron los mejores promedios de precisión, evidenciando que un mayor número de rondas facilita una mejor sincronización y agregación del modelo global. Los escenarios con 2 rondas presentan un descenso notable en la precisión (68.15%), confirmando que la baja frecuencia de comunicación limita la capacidad del modelo global para generalizar correctamente. En términos de tiempo, los escenarios con 2 rondas no implicaron necesariamente un

menor costo computacional, ya que el tiempo total se vio afectado por el uso de imágenes sintéticas y la duración de las rondas de entrenamiento local.

La Tabla 19 resume la precisión y el tiempo promedio alcanzados por los escenarios con 2 y 3 clientes.

**Tabla 19.** Promedios por número de clientes.

No. de clientes	Precisión promedio (%)	Tiempo promedio (s)
2	89.64	9029.28
3	88.02	3686.3

De acuerdo con la Tabla 19, los escenarios con 2 clientes alcanzaron una precisión promedio superior (89.64%) al de los escenarios con 3 clientes (88.02%). Esto podría explicarse por la mayor cantidad de datos locales disponible por cliente en los escenarios de 2 clientes, lo que facilita un aprendizaje local más robusto. Sin embargo, el costo computacional promedio fue mayor, reflejando el mayor volumen de datos procesados por cada cliente.

La Tabla 20 presenta el promedio de precisión y tiempo de ejecución de los escenarios con y sin imágenes sintéticas.

**Tabla 20.** Promedios por uso de imágenes sintéticas.

Imágenes sintéticas	Precisión promedio (%)	Tiempo promedio (s)
Sí	88.5	8957.93
No	89.16	4915.1

La Tabla 20 muestra que el uso de imágenes sintéticas tuvo un impacto mixto: en escenarios con un número suficiente de rondas (25 y 50), contribuyó a mejorar la precisión, mientras que en escenarios con pocas rondas (especialmente con 2 rondas) no ofrece beneficios, y en algunos casos incluso genera una penalización en el rendimiento. En todos los casos, el uso de imágenes sintéticas incrementó el tiempo de ejecución, lo que era esperable dado el mayor volumen de datos.

El análisis agrupado por factor experimental, respaldado por las Tablas 16-18, confirma que el número de rondas de comunicación fue el factor de mayor impacto sobre el rendimiento global del modelo, seguido por el uso de imágenes sintéticas y el número de clientes. Estos hallazgos destacan la necesidad de un

diseño equilibrado que considere tanto la precisión como el costo computacional en sistemas de AF aplicados a la clasificación de imágenes médicas.

### 5.2.3 Comparación por estadio

La Tabla 21 presenta la precisión promedio obtenido en cada estadio; Normal, Temprano, Medio y Tardío para los 16 escenarios experimentales. Estos valores corresponden al promedio de los nodos o clientes en cada escenario. El análisis por estadio es crucial en aplicaciones clínicas, ya que permite evaluar la capacidad del modelo global para distinguir los distintos estadios de la enfermedad, con especial énfasis en aquellos de difícil clasificación, como el Temprano y el Medio.

**Tabla 21.** Comparación por estadio de los 16 escenarios de AF

Escenario	Normal	Temprano	Medio	Tardío
<b>E1</b>	97.52	96.52	97.87	100
<b>E2</b>	99.07	<b>99.11</b>	<b>98.92</b>	100
<b>E3</b>	96.28	96.31	96.63	100
<b>E4</b>	98.74	96.9	98.85	100
<b>E5</b>	98.43	95.63	97.7	100
<b>E6</b>	99.07	98.65	96.76	100
<b>E7</b>	96.02	96.28	95.48	100
<b>E8</b>	98.16	99.1	98.85	100
<b>E9</b>	97.24	97.77	96.73	100
<b>E10</b>	<b>99.39</b>	98.66	97.78	100
<b>E11</b>	92.7	97.22	98.67	100
<b>E12</b>	98.73	96.45	98.85	100
<b>E13</b>	94.06	84.5	97.15	100
<b>E14</b>	96.86	95.69	52.62	3.4
<b>E15</b>	76.37	78.19	97.44	0
<b>E16</b>	99.15	85.51	27.43	2.09

Los resultados muestran que el estadio Tardío alcanzó una precisión del 100% en la mayoría de los escenarios, lo que se explica por las características morfológicas distintivas de este estadio en las MRI. Sin embargo, este alto rendimiento se redujo significativamente en los escenarios con pocas rondas de comunicación o con configuraciones desfavorables. Por ejemplo, los escenarios E14, E15 y E16, con solo 2

rondas de comunicación, muestran una precisión promedio en el estadio Tardío de 3.39%, 0% y 2.09% respectivamente, evidenciando la incapacidad del modelo global para aprender características consistentes en estas condiciones.

En el estadio Temprano, se observa una gran variabilidad en la precisión según el escenario. Los mejores resultados se lograron en los escenarios con mayor número de rondas de comunicación y el uso de imágenes sintéticas, como E2 (99.11%) y E8 (99.10%). En cambio, los escenarios con configuraciones reducidas, como E13 (84.50%) y E16 (85.51%), presentan un descenso notable, reflejando la dificultad de la clasificación temprana en condiciones de AF limitadas.

El estadio Medio presenta un comportamiento intermedio. Escenarios como E2, E4, E8 y E12 alcanzaron precisiones superiores al 98%, confirmando que el número de rondas y el uso de imágenes sintéticas favorecieron su adecuada clasificación. Por el contrario, en E14, E15 y E16, los valores se redujeron drásticamente (52.62%, 97.44%, 27.43%, respectivamente), reforzando el impacto negativo de las restricciones en comunicación y la ausencia de un diseño balanceado en el entorno federado.

En el estadio Normal, la precisión se mantuvo alto en la mayoría de los escenarios, especialmente en los configurados con mayor número de rondas y datos sintéticos, por ejemplo, E2 con 99.07% y E10 con 99.39%. Sin embargo, los escenarios más limitados, como E15 (76.37%) y E14 (96.86%), demostraron que incluso en este estadio bien representado en el conjunto de datos, el rendimiento puede verse comprometido por un diseño inadecuado del entorno federado. En conjunto, estos resultados confirman que el número de rondas de comunicación y el uso de imágenes sintéticas fueron factores determinantes para un desempeño homogéneo en los distintos estadios. La correcta clasificación de estadios críticos como el Temprano y el Medio requiere un entorno federado bien configurado que permita al modelo global capturar patrones morfológicos sutiles

### **5.3 Análisis estadístico del efecto de los factores experimentales**

Con el objetivo de evaluar el efecto de diferentes configuraciones experimentales sobre el rendimiento del modelo, se realizó un análisis de varianza multifactorial considerando tres factores: número de hospitales 2 o 3, número de rondas de comunicación 50, 25, 10 y 2 y uso de imágenes sintéticas sí o no. La variable dependiente analizada fue la precisión obtenida en cada experimento. La Tabla 22 muestra los resultados, en el que se incluyen los efectos principales de cada factor y sus interacciones de dos vías. El

análisis revela que el número de rondas de comunicación presentó un efecto estadísticamente significativo sobre la precisión ( $p = 0.0098$ ). En contraste, ni el número de hospitales ni el uso de imágenes sintéticas mostraron efectos significativos. Del mismo modo, ninguna de las interacciones entre factores alcanzó significancia estadística ( $p > 0.05$ ).

**Tabla 22.** Resultados del análisis de varianza multifactorial sobre la variable Precisión.

Factor	Suma de cuadrados	gl	F	p-valor
Número de hospitales	113.9	1	3.79	0.1467
Rondas de comunicación	2685.86	3	29.8	<b>0.0098</b>
Uso de imágenes sintéticas	63.4	1	2.11	0.2422
Hospitales × Rondas	275.98	3	3.06	0.1913
Hospitales × Sintéticas	38.1	1	1.27	0.3421
Rondas × Sintéticas	356.99	3	3.96	0.144
Residual	90.12	3	—	—

Los resultados del análisis muestran que el número de rondas de comunicación es un factor determinante en el rendimiento del modelo, lo que coincide con reportes previos donde se ha evidenciado que un mayor número de rondas favorece la convergencia y estabilidad del entrenamiento en entornos de AF. La ausencia de significancia en el factor número de hospitales sugiere que, bajo las condiciones experimentales de este trabajo, variar entre dos y tres nodos no ejerce un impacto sustancial en la precisión, posiblemente debido a que el volumen total de datos y su distribución se mantuvieron constantes. Por otra parte, el uso de imágenes sintéticas no mostró un efecto principal significativo, lo que podría indicar que, si bien su inclusión puede contribuir al balance del conjunto de datos, su impacto sobre la precisión no es suficiente para generar diferencias detectables estadísticamente. La falta de interacciones significativas entre factores sugiere que estos operan de forma relativamente independiente en su influencia sobre la métrica de precisión.

## Capítulo 6.      Discusión

---

En este capítulo se analizan los resultados obtenidos a lo largo de la investigación, contrastándolos con los reportados en la literatura tanto en entornos centralizados como en escenarios de aprendizaje federado. El objetivo es examinar las diferencias y similitudes con estudios previos, considerando factores como el número de clases, las características de los conjuntos de datos, las técnicas de aumento de datos y las metodologías empleadas. Asimismo, se identifican las limitaciones del estudio, señalando aspectos técnicos y experimentales que influyen en la interpretación de los hallazgos y que pueden orientar trabajos futuros.

### 6.1 Análisis comparativo con modelos de CNNs del estado del arte

Además de evaluar el desempeño de las arquitecturas CNN seleccionadas bajo condiciones homogéneas en un entorno centralizado, se realizó una comparación de los resultados obtenidos en esta investigación con los valores reportados en los artículos originales donde se proponen estas arquitecturas. La Tabla 23 resume las métricas reportadas por los autores de cada modelo, junto con las características principales de sus configuraciones experimentales.

Al contrastar estos valores con los obtenidos en esta investigación, se observan diferencias significativas en algunos casos. Por ejemplo, mientras el modelo de AbdulAzeemNet reporta una precisión del 99.98% en un escenario binario (Normal vs. AD) con un conjunto de datos extenso (211,665 imágenes) y aumento de datos intensivo, en nuestro estudio alcanzó un 98.20% de precisión en una tarea de clasificación de cuatro estadios y con un conjunto de datos de menor tamaño y sin emplear técnicas de aumento de datos.

De manera similar, ArafaNet reporta un 99.98% en un escenario binario, mientras que en el análisis obtuvo 91.01% al enfrentarse a una tarea multiclase. En contraste, AHANet, que reportó un 98.53% en clasificación de tres clases, obtuvo un 98.40% clasificando cuatro estadios, evidenciando su capacidad de generalización y robustez.

Dichas diferencias pueden explicarse por tres principales factores. En primer lugar, el conjunto de datos empleado: en esta investigación se utilizó un conjunto de datos distinto al de los estudios originales, por

ejemplo, un conjunto con imágenes MRI de cuatro estadios, mientras que varios artículos trabajaron con ADNI o con escenarios binarios.

En segundo lugar, el número de estadios: las precisiones reportadas en los artículos suelen corresponder a tareas con menor número de clases (2 o 3 estadios), mientras que en este trabajo se evaluaron todas las arquitecturas bajo una clasificación de cuatro estadios. Por último, el tamaño y balance del conjunto de datos: muchos de los estudios originales emplean conjuntos de datos con mayor número de imágenes o número de pacientes, o aplicaron técnicas de aumento de datos, facilitando la obtención de mayores precisiones. En contraste, en este estudio a fin de garantizar una comparación justa, se optó por evaluar todas las arquitecturas bajo condiciones homogéneas y sin aumento de datos.

**Tabla 23.** Resultados reportados en la literatura para los modelos de CNN evaluadas.

Modelo	Conjunto de datos	Aumento de datos	No. de Img	Tamaño img	Estadios	Precisión (%)	Referencia
<b>AlexNet</b>	Alzheimer MRI 4 classes		664	64x64	4	95.50	(Fu'adah et al., 2021)
<b>DenseNet</b>	Alzheimer MRI 4 classes	✓	6,400	176x208	4	82.00	(Pradhan et al., 2021)
<b>VGGNet</b>	Alzheimer MRI 4 classes	✓	6,400	176x208	4	86.70	(Pradhan et al., 2021)
<b>Abdul</b>	ADNI	✓	211,665	64x64	2	99.98	(AbdulAzeem et al., 2021)
<b>AHANet</b>	ADNI	✓	158,100	-	3	98.53	(Illakiya et al., 2023)
<b>Arafa</b>	Alzheimer MRI 4 classes	✓	6,400	64x64	2	99.98	(Arafa et al., 2024)
<b>DEMNet</b>	Alzheimer MRI 4 classes	✓	12,800	176x176	4	95.75	(Murugan et al., 2021)E

Por lo tanto, aunque los valores absolutos de precisión obtenidos en este trabajo puedan diferir de los reportados por los autores originales, el análisis realizado tiene un valor añadido: permite comparar de manera directa el desempeño de las distintas arquitecturas en un mismo entorno experimental y sobre un conjunto de datos común, proporcionando un marco más robusto para seleccionar el modelo óptimo para su aplicación en entornos federados.

## 6.2 Comparación con estudios en la literatura sobre escenarios de AF

Diversos trabajos recientes han explorado el uso del AF para el diagnóstico de la EA y otras enfermedades neurodegenerativas. La mayoría de estas investigaciones se enfocan en configuraciones binarias o con un número reducido de estadios, empleando diferentes metodologías, conjuntos de datos y modalidades de entrada, la Tabla 24 resume los resultados.

Basnin et al., (2025) proponen un enfoque basado en AF evolutivo aplicado al conjunto ADNI, obteniendo una precisión del 99.9% para la clasificación en tres estadios (HC, MCI y EA). Si bien el valor reportado es ligeramente superior a los alcanzados en este trabajo, es importante considerar que su problema de clasificación es menos complejo, al incluir una clase menos que nuestro estudio, el cual aborda cuatro estadios.

Castro et al., (2024) emplean AF combinado con reconocimiento biométrico sobre imágenes de los conjuntos OASIS y ADNI, obteniendo una precisión del 92% en una tarea binaria. De manera similar, Mitrovska et al., (2024) aplicaron promedio federado con agregación segura a un problema binario, alcanzando un 83.3% de precisión. Ambos resultados son notablemente inferiores a los obtenidos en este trabajo, donde se enfrentó un reto mayor al tratar cuatro estadios distintos.

Por otro lado, Kapoor et al., (2024) emplearon una combinación de AF, aprendizaje continuo y aprendizaje por ensamblado (FedCL-Ensemble Learning) sobre un conjunto de datos de Kaggle, logrando un 97.8% de precisión en cinco estadios (ND, VMD, MiD y MD). Este resultado es comparable a los de nuestros mejores escenarios (E2, E8, E10 y E12), que superan el 98% de precisión global en la clasificación de cuatro estadios.

Otros estudios como el de Mandawkar y Diwan (2022), que integraron MRI, PET y CSF mediante un CNN profundo (Tawny Flamingo) con AF, reportaron una precisión del 98.25% en una tarea binaria. De forma similar, Qian et al., (2023) utilizaron bosque profundo federado con ADNI logrando una precisión de 91.6% en tres clases. En ambos casos, pese al uso de datos multimodales o enfoques alternativos, los resultados son comparables o inferiores a los obtenidos en este trabajo con imágenes MRI y clasificación multiclase.

Estudios con modalidades no convencionales, como biomarcadores sanguíneos (Khalil et al., 2023) o voz (Wei et al., 2023), reportan precisiones entre 79% y 89%, destacando el valor añadido de las imágenes MRI como fuente de datos para la clasificación en un entorno federado.

**Tabla 24.** Comparación con estudios de la literatura sobre aprendizaje federado para diagnóstico de Alzheimer.

<b>Año</b>	<b>Modalidad</b>	<b>Conjunto de datos</b>	<b>Metodología</b>	<b>Estadios</b>	<b>Cientes</b>	<b>Precisión</b>	<b>Ref.</b>
<b>2025</b>	Imagen	ADNI	AF Evolutivo	3	3	99.90	(Basnin et al., 2025)
<b>2024</b>	Imagen	OASIS, ADNI	AF + Reconocimiento Biométrico	2	2	92.00	(Castro et al., 2024)
<b>2024</b>	Imagen	ADNI, OASIS-3, AIBL, UK Base de datos	FedAvg + SecAgg	2	3	83.30	(Mitrovska et al., 2024)
<b>2024</b>	Imagen	Kaggle	AF + Aprendizaje Continuo + Aprendizaje por Transferencia Mediante ensamblado	4	5	97.80	(Kapoor et al., 2024)
<b>2023</b>	Imagen	ADNI (MRI, PET, CSF)	Tawny Flamingo + AF	2	-	98.25	(Mandawkar & Diwan, 2022)
<b>2023</b>	Imagen	ADNI	Bosque Profundo Federado (FeDeFo)	3		91.60	(Qian et al., 2023)
<b>2023)</b>	Imagen	ADNI, OASIS-3, AIBL, UK Base de datos	AF + Encriptación Homomórfica (MetisFL)	2	3	82.82	(Stripelis et al., 2023)
<b>2023</b>	Biomarcadores de sangre	Biomarcadores de sangre (ADNI)	AF + FPGA	2	-	89.00	(Khalil et al., 2023)
<b>2023</b>	Voz	NCMMSC EA Reto de Reconocimiento	Aprendizaje Federado Contrastivo (FedCPC)	3	3	79.50	(Wei et al., 2023)
<b>2023</b>	Voz	Conjunto de datos de voz con trastornos de Kay Elemetrics y conjunto de datos de voz para telemonitoreo de Parkinson (UCI)	AF	2	2 y 3	3 clientes 80.66 2 clientes 95.00	(Altalbe & Rehman Javed, 2023)

### 6.3 Limitaciones generales del estudio

A lo largo de esta tesis se llevaron a cabo diversos experimentos orientados a la clasificación de la AD mediante DL y AF, incluyendo la comparación de siete arquitecturas CNN, la propuesta y evaluación de AHANet-C, el análisis del impacto del aumento de datos y de las imágenes sintéticas, y finalmente, el diseño de un entorno federado con múltiples configuraciones. Si bien los resultados obtenidos son competitivos y aportan evidencia del potencial de los enfoques propuestos, el estudio presenta una serie de limitaciones que deben ser consideradas al interpretar los hallazgos. En primer lugar, el conjunto de datos utilizado, aunque representativo, presenta un desbalance importante, particularmente en el estadio Tardío, el cual está constituido únicamente por imágenes correspondientes a dos sujetos.

Esta limitación compromete la generalización de los resultados para dicho estadio, contribuyendo tanto a los altos valores de precisión observados en ciertos escenarios como a la caída significativa en escenarios con restricciones severas de comunicación o sin imágenes sintéticas. En segundo lugar, todos los experimentos se realizaron con imágenes MRI en formato 2D, lo que, si bien permitió estandarizar el preprocesamiento y facilitar la implementación, limita el potencial diagnóstico respecto al uso de volúmenes 3D completos, los cuales podrían capturar más información estructural del cerebro y ofrecer un diagnóstico más robusto.

Finalmente, al no contar con un identificador por paciente no se puede garantizar que imágenes pertenecientes al mismo sujeto permanezcan separadas y de esta forma evitar que imágenes con las que se realizó el entrenamiento se encuentren dentro del conjunto de prueba. Esta problemática se acentúa en el caso del estadio tardío, donde solo se cuenta con dos sujetos, por ende, imágenes de ambos sujetos se encuentran tanto en los conjuntos de entrenamiento validación y prueba generando así valores de precisión elevados 100% de precisión.

En cuanto a los experimentos centralizados, aunque se compararon múltiples arquitecturas y se propuso una optimización mediante super-convergencia, el trabajo se limitó a un solo conjunto de datos y no se exploraron otros tipos de datos clínicos o multimodales, como PET, CSF o biomarcadores, que podrían enriquecer el modelo.

En lo referente al entorno federado, la simulación se llevó a cabo en un entorno controlado, con nodos virtuales dentro de un mismo sistema físico, lo que impide reproducir completamente las condiciones de

un entorno federado real, donde podrían presentarse problemas adicionales relacionados con la conectividad, la heterogeneidad de los dispositivos y las restricciones de comunicación. Además, aunque el estudio preservó la privacidad mediante la estructura federada, no se incorporaron técnicas avanzadas de seguridad como HE, agregación segura o privacidad diferencial, que podrían reforzar la protección de los datos en aplicaciones reales.

Por último, el coste computacional asociado a los escenarios más complejos, con mayor número de rondas de comunicación y uso de imágenes sintéticas, fue elevado en términos de tiempo de entrenamiento, lo que plantea un desafío para su implementación en entornos clínicos con recursos limitados. Estas limitaciones ofrecen oportunidades claras para futuras investigaciones, orientadas a validar los modelos en entornos federados reales, incorporar datos multimodales, utilizar imágenes en 3D, equilibrar las clases mediante nuevos métodos y reforzar la seguridad de la información con técnicas criptográficas avanzadas.

## Capítulo 7. Conclusiones y trabajo futuro

---

El entrenamiento de modelos de DL para la clasificación de MRI requiere grandes volúmenes de datos centralizados, lo cual representa un desafío en entornos clínicos donde la privacidad de los datos es indispensable y la disponibilidad de información, especialmente en estadios avanzados de enfermedad, es limitada. El AF surge como una solución para entrenar modelos robustos sin necesidad de compartir los datos sensibles, preservando su confidencialidad. No obstante, este enfoque introduce nuevos retos en términos de eficiencia computacional, balance de clases, y estabilidad del entrenamiento.

Esta tesis presentó el diseño y validación de un sistema de clasificación multi estadio de la EA mediante MRI, empleando una arquitectura basada en AF. Inicialmente, se evaluaron siete arquitecturas de CNN reconocidas en la literatura. El análisis comparativo demostró que AHANet ofrece un equilibrio favorable entre precisión y eficiencia. A partir de este modelo, se propuso AHANet-C, una versión optimizada mediante bloques de super-convergencia, logrando acelerar el entrenamiento sin pérdida de precisión.

Posteriormente, se desarrolló una simulación de AF basada en AHANet-C, evaluada en 16 escenarios experimentales, con distintas configuraciones de número de clientes, rondas de comunicación y presencia de imágenes sintéticas. Se demostró que el modelo de AF alcanzó niveles de precisión superiores al 98%, con tiempos de entrenamiento significativamente menores, validando así su aplicabilidad en tareas de clasificación clínica.

La incorporación de imágenes sintéticas generadas por técnicas de aumento de datos y GAN permitió mitigar el desbalance de clases, especialmente en el estadio tardío de la EA. Entre las principales contribuciones de este trabajo se encuentran: (i) la propuesta de un modelo CNN optimizado con super-convergencia para entornos de AF; (ii) la validación de una estrategia AF multi estadio preservando la privacidad de los datos; (iii) la integración de imágenes sintéticas para el balanceo del conjunto de datos; y (iv) el análisis del impacto de variables de diseño; clientes, rondas de comunicación y balance de clases, en el rendimiento final.

Como trabajo a futuro, se propone validar el modelo en entornos reales con bases de datos heterogéneas, dispositivos físicamente distribuidos, heterogéneos en capacidad de cómputo y conectividad, para evaluar su robustez frente a desafíos clínicos reales como la latencia y la asincronía. También se plantea ampliar la evaluación a nuevas bases de datos con mayor diversidad estructural y demográfica, con el fin de mejorar la generalización y reducir el sesgo asociado a clases subrepresentadas. En materia de privacidad,

se propone incorporar técnicas como HE y agregación segura, fortaleciendo la protección de datos sensibles durante el entrenamiento. Además, se sugiere analizar la resiliencia del sistema frente a ataques adversarios, explorando mecanismos AF de defensa, detección de clientes maliciosos y estrategias de agregación robusta, avanzando hacia modelos federados seguros, confiables y escalables para su aplicación en el ámbito clínico.

## Literatura citada

- AbdulAzeem, Y., Bahgat, W. M., & Badawy, M. (2021). A CNN based framework for classification of Alzheimer's disease. *Neural Computing and Applications*, 33(16), 10415-10428. <https://doi.org/10.1007/s00521-021-05799-w>
- Ali, M. U., Kim, K. S., Khalid, M., Farrash, M., Zafar, A., & Lee, S. W. (2024). Enhancing Alzheimer's disease diagnosis and staging: A multistage CNN framework using MRI. *Frontiers in Psychiatry*, 15, 1-15, 1395563. <https://doi.org/10.3389/fpsy.2024.1395563>
- Altalbe, A., & Rehman Javed, A. (2023). Privacy Preserved Brain Disorder Diagnosis Using Federated Learning. *Computer Systems Science and Engineering*, 47(2), 1-14, 2187-2200. <https://doi.org/10.32604/csse.2023.040624>
- Alzheimer's Disease Neuroimaging Initiative (ADNI). (2024). Sharing Alzheimer's research data with the world. Consultado en noviembre 2023. <https://adni-lde.loni.usc.edu/>
- Antunes, R. S., André Da Costa, C., Küderle, A., Yari, I. A., & Eskofier, B. (2022). Federated Learning for Healthcare: Systematic Review and Architecture Proposal. *ACM Transactions on Intelligent Systems and Technology*, 13(4), 1-23. <https://doi.org/10.1145/3501813>
- Arafa, D. A., Moustafa, H. E.-D., Ali, H. A., Ali-Eldin, A. M. T., & Saraya, S. F. (2024). A deep learning framework for early diagnosis of Alzheimer's disease on MRI images. *Multimedia Tools and Applications*, 83(2), 3767-3799. <https://doi.org/10.1007/s11042-023-15738-7>
- Asociación de Alzheimer. (2025). *Etapas del Alzheimer*. Consultado en mayo 2025. <https://www.alz.org/es-mx/alzheimer-demencia/etapas>
- Australian Imaging, Biomarker and Lifestyle (AIBL). (2024). Research. Consultado en noviembre 2023. <https://aibl.org.au/research/>
- Basnin, N., Mahmud, T., Islam, R. U., & Andersson, K. (2025). An Evolutionary Federated Learning Approach to Diagnose Alzheimer's Disease Under Uncertainty. *Diagnostics*, 15(1), 1-37, 80. <https://doi.org/10.3390/diagnostics15010080>
- Beutel, D. J., Topal, T., Mathur, A., Qiu, X., Fernandez-Marques, J., Gao, Y., Sani, L., Li, K. H., Parcollet, T., de Gusmão, P. P. B., & Lane, N. D. (2020). *Flower: A Friendly Federated Learning Research Framework* (Version 5). arXiv. <https://doi.org/10.48550/ARXIV.2007.14390>
- Bloch, F., Hansen, W. W., & Packard, M. (1946). Nuclear Induction. *Physical Review*, 69(3-4), 127-127. <https://doi.org/10.1103/PhysRev.69.127>
- Brown, R. W., Cheng, Y. N., Haacke, E. M., Thompson, M. R., & Venkatesan, R. (2014). *Magnetic Resonance Imaging: Physical Principles and Sequence Design*. Wiley. <https://doi.org/10.1002/9781118633953>
- Castro, F., Impedovo, D., & Pirlo, G. (2024). A Federated Learning System with Biometric Medical Image Authentication for Alzheimer's Diagnosis: *Proceedings of the 13th International Conference on Pattern Recognition Applications and Methods*, 951-960. <https://doi.org/10.5220/0012550200003654>

- Chen, C., Mat Isa, N. A., & Liu, X. (2025). A review of convolutional neural network based methods for medical image classification. *Computers in Biology and Medicine*, 185, 1-12, 109507. <https://doi.org/10.1016/j.compbiomed.2024.109507>
- Chen, X., Tang, M., Liu, A., & Wei, X. (2022). Diagnostic accuracy study of automated stratification of Alzheimer's disease and mild cognitive impairment via deep learning based on MRI. *Annals of Translational Medicine*, 10(14), 765-765. <https://doi.org/10.21037/atm-22-2961>
- Dale, B. M., Brown, M. A., & Semelka, R. C. (2015). *MRI Basic Principles and Applications*. Wiley. <https://doi.org/10.1002/9781119013068>
- Dinh, L., Sohl-Dickstein, J., & Bengio, S. (2016). *Density estimation using Real NVP* (Version 3). arXiv. <https://doi.org/10.48550/ARXIV.1605.08803>
- Duchi, J., Hazan, E., & Singer, Y. (2011). Adaptive Subgradient Methods for Online Learning and Stochastic Optimization. *Journal of Machine Learning Research*, 12(61), 2121-2159. <https://www.jmlr.org/papers/volume12/duchi11a/duchi11a.pdf>
- El-Assy, A. M., Amer, H. M., Ibrahim, H. M., & Mohamed, M. A. (2024). A novel CNN architecture for accurate early detection and classification of Alzheimer's disease using MRI data. *Scientific Reports*, 14(1), 1-19, 3463. <https://doi.org/10.1038/s41598-024-53733-6>
- Fu'adah, Y. N., Wijayanto, I., Pratiwi, N. K. C., Taliningsih, F. F., Rizal, S., & Pramudito, M. A. (2021). Automated Classification of Alzheimer's Disease Based on MRI Image Processing using Convolutional Neural Network (CNN) with AlexNet Architecture. *Journal of Physics: Conference Series*, 1844(1), 1-9, 012020. <https://doi.org/10.1088/1742-6596/1844/1/012020>
- Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36(4), 193-202. <https://doi.org/10.1007/bf00344251>
- Gao, X., Cai, H., & Liu, M. (2023). A Hybrid Multi-Scale Attention Convolution and Aging Transformer Network for Alzheimer's Disease Diagnosis. *IEEE Journal of Biomedical and Health Informatics*, 27(7), 3292-3301. <https://doi.org/10.1109/JBHI.2023.3270937>
- Gao, Y., Tan, L., Yu, J.-T., & Tan, L. (2018). Tau in Alzheimer's Disease: Mechanisms and Therapeutic Strategies. *Current Alzheimer Research*, 15(3), 283-300. <https://doi.org/10.2174/1567205014666170417111859>
- Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, 249-256. <https://proceedings.mlr.press/v9/glorot10a.html>
- Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). *Generative Adversarial Networks*. arXiv. <https://doi.org/10.48550/ARXIV.1406.2661>
- Harvard Aging Brain Study (HABS)*. (2024). *Recursos*. Consultado en noviembre 2024 <https://habs.mgh.harvard.edu/es/recursos/>
- Hazarika, R. A., Kandar, D., & Maji, A. K. (2022). An experimental analysis of different Deep Learning based Models for Alzheimer's Disease classification using Brain Magnetic Resonance Images. *Journal of*

- King Saud University Computer and Information Sciences, 34(10), 1-23, 8576-8598. <https://doi.org/10.1016/j.jksuci.2021.09.003>
- Ho, J., Jain, A., & Abbeel, P. (2020). *Denoising Diffusion Probabilistic Models* (Version 2). arXiv. <https://doi.org/10.48550/ARXIV.2006.11239>
- Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5), 359-366. [https://doi.org/10.1016/0893-6080\(89\)90020-8](https://doi.org/10.1016/0893-6080(89)90020-8)
- Hu, K., Gong, S., Zhang, Q., Seng, C., Xia, M., & Jiang, S. (2024). An overview of implementing security and privacy in federated learning. *Artificial Intelligence Review*, 57(8), 1-66, 204. <https://doi.org/10.1007/s10462-024-10846-8>
- Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely Connected Convolutional Networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2261-2269. <https://doi.org/10.1109/CVPR.2017.243>
- Hussain, M. Z., Shahzad, T., Mehmood, S., Akram, K., Khan, M. A., Tariq, M. U., & Ahmed, A. (2025). A fine-tuned convolutional neural network model for accurate Alzheimer's disease classification. *Scientific Reports*, 15(1), 1-24, 11616. <https://doi.org/10.1038/s41598-025-86635-2>
- Illakiya, T., Ramamurthy, K., Siddharth, M. V., Mishra, R., & Udainiya, A. (2023). AHA-Net: Adaptive Hybrid Attention Network for Alzheimer's Disease Classification Using Brain Magnetic Resonance Imaging. *Bioengineering*, 10(6), 1-20, 714. <https://doi.org/10.3390/bioengineering10060714>
- Ioffe, S., & Szegedy, C. (2015). *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift* (Version 3). arXiv. <https://doi.org/10.48550/ARXIV.1502.03167>
- Isola, P., Zhu, J.-Y., Zhou, T., & Efros, A. A. (2016). *Image-to-Image Translation with Conditional Adversarial Networks* (Version 3). arXiv. <https://doi.org/10.48550/ARXIV.1611.07004>
- Jabason, E., Ahmad, M. O., & Swamy, M. N. S. (2025). A Lightweight Deep Convolutional Neural Network Extracting Local and Global Contextual Features for the Classification of Alzheimer's Disease Using Structural MRI. *IEEE Journal of Biomedical and Health Informatics*, 29(3), 2061-2073. <https://doi.org/10.1109/JBHI.2024.3512417>
- Jack, C. R., Bernstein, M. A., Fox, N. C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P. J., L. Whitwell, J., Ward, C., Dale, A. M., Felmlee, J. P., Gunter, J. L., Hill, D. L. G., Killiany, R., Schuff, N., Fox-Bosetti, S., Lin, C., Studholme, C., Weiner, M. W. (2008). The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging*, 27(4), 685-691. <https://doi.org/10.1002/jmri.21049>
- Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R., D'Oliveira, R. G. L., Eichner, H., Rouayheb, S. E., Evans, D., Gardner, J., Garrett, Z., Gascón, A., Ghazi, B., Gibbons, P. B., ... Zhao, S. (2021). *Advances and Open Problems in Federated Learning* (No. arXiv:1912.04977). arXiv. <http://arxiv.org/abs/1912.04977>
- Kapoor, R., Joshua, J., & Vijayarangan, M. (2024). *FedCL-Ensemble Learning: A Framework of Federated Continual Learning with Ensemble Transfer Learning Enhanced for Alzheimer's MRI Classifications while Preserving Privacy*. <https://doi.org/10.48550/arXiv.2411.12756>

- Khalil, K., Khan Mamun, M. M. R., Sherif, A., Elserly, M. S., Imam, A. A.-A., Mahmoud, M., & Alsabaan, M. (2023). A Federated Learning Model Based on Hardware Acceleration for the Early Detection of Alzheimer's Disease. *Sensors*, 23(19), 1-17, 8272. <https://doi.org/10.3390/s23198272>
- Kingma, D. P., & Welling, M. (2013). *Auto-Encoding Variational Bayes* (Version 11). arXiv. <https://doi.org/10.48550/ARXIV.1312.6114>
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84-90. <https://doi.org/10.1145/3065386>
- LaMontagne, P. J., Benzinger, T. L., Morris, J. C., Keefe, S., Hornbeck, R., Xiong, C., Grant, E., Hassenstab, J., Moulder, K., Vlassenko, A. G., Raichle, M. E., Cruchaga, C., & Marcus, D. (2019). *OASIS-3: Longitudinal Neuroimaging, Clinical, and Cognitive Dataset for Normal Aging and Alzheimer Disease* (p. 2019.12.13.19014902). medRxiv. <https://doi.org/10.1101/2019.12.13.19014902>
- Lauterbur, P. C. (1973). Image Formation by Induced Local Interactions: Examples Employing Nuclear Magnetic Resonance. *Nature*, 242(5394), 190-191. <https://doi.org/10.1038/242190a0>
- Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278-2324. <https://doi.org/10.1109/5.726791>
- Lee, J., Burkett, B. J., Min, H.-K., Senjem, M. L., Lundt, E. S., Botha, H., Graff-Radford, J., Barnard, L. R., Gunter, J. L., Schwarz, C. G., Kantarci, K., Knopman, D. S., Boeve, B. F., Lowe, V. J., Petersen, R. C., Jack, C. R., & Jones, D. T. (2022). Deep learning-based brain age prediction in normal aging and dementia. *Nature Aging*, 2(5), 412-424. <https://doi.org/10.1038/s43587-022-00219-7>
- Lim, W. Y. B., Luong, N. C., Hoang, D. T., Jiao, Y., Liang, Y.-C., Yang, Q., Niyato, D., & Miao, C. (2019). *Federated Learning in Mobile Edge Networks: A Comprehensive Survey* (Version 2). arXiv. <https://doi.org/10.48550/ARXIV.1909.11875>
- Loshchilov, I., & Hutter, F. (2017). *Decoupled Weight Decay Regularization* (Version 3). arXiv. <https://doi.org/10.48550/ARXIV.1711.05101>
- Malihi, L., Hübner, U., Richter, M. L., Moelleken, M., Przysocka, M., Busch, D., Heggemann, J., Hafer, G., Wiemeyer, S., Heidemann, G., Dissemond, J., Erfurt-Berge, C., Barkhau, C., Hendriks, A., & Hüsers, J. (2023). Can Synthetic Images Improve CNN Performance in Wound Image Classification? M. Hägglund, M. Blusi, S. Bonacina, L. Nilsson, I. Cort Madsen, S. Pelayo, A. Moen, A. Benis, L. Lindsköld, & P. Gallos (Eds.), *Studies in Health Technology and Informatics*. IOS Press. <https://doi.org/10.3233/SHTI230311>
- Mandawkar, U., & Diwan, T. (2022). Alzheimer disease classification using tawny flamingo based deep convolutional neural networks via federated learning. *The Imaging Science Journal*, 70(7), 459-472. <https://doi.org/10.1080/13682199.2023.2172524>
- Mansfield, P. (1977). Multi-planar image formation using NMR spin echoes. *Journal of Physics C: Solid State Physics*, 10(3), L55-L58. <https://doi.org/10.1088/0022-3719/10/3/004>
- Mantzavinos, V., & Alexiou, A. (2017). Biomarkers for Alzheimer's Disease Diagnosis. *Current Alzheimer Research*, 14(11), 1-6. <https://doi.org/10.2174/1567205014666170203125942>

- McMahan, H. B., Moore, E., Ramage, D., Hampson, S., & Arcas, B. A. y. (2016). *Communication-Efficient Learning of Deep Networks from Decentralized Data*. <https://doi.org/10.48550/ARXIV.1602.05629>
- Mitrovska, A., Safari, P., Ritter, K., Shariati, B., & Fischer, J. K. (2024). Secure federated learning for Alzheimer's disease detection. *Frontiers in Aging Neuroscience*, 16, 1-11, 1324032. <https://doi.org/10.3389/fnagi.2024.1324032>
- Mordido, G., Yang, H., & Meinel, C. (2018). *Dropout-GAN: Learning from a Dynamic Ensemble of Discriminators (Version 2)*. arXiv. <https://doi.org/10.48550/ARXIV.1807.11346>
- Morris, J. C. (1997). Clinical Dementia Rating: A Reliable and Valid Diagnostic and Staging Measure for Dementia of the Alzheimer Type. *International Psychogeriatrics*, 9, 173-176. <https://doi.org/10.1017/S1041610297004870>
- Mothukuri, V., Parizi, R. M., Pouriye, S., Huang, Y., Dehghantanha, A., & Srivastava, G. (2021). A survey on security and privacy of federated learning. *Future Generation Computer Systems*, 115, 619-640. <https://doi.org/10.1016/j.future.2020.10.007>
- Mugler, J. P., & Brookeman, J. R. (1990). Three-dimensional magnetization-prepared rapid gradient-echo imaging (3D MP RAGE). *Magnetic Resonance in Medicine*, 15(1), 152-157. <https://doi.org/10.1002/mrm.1910150117>
- Murugan, S., Venkatesan, C., Sumithra, M. G., Gao, X.-Z., Elakkiya, B., Akila, M., & Manoharan, S. (2021). DEMNET: A Deep Learning Model for Early Diagnosis of Alzheimer Diseases and Dementia From MR Images. *IEEE Access*, 9, 1-11, 90319-90329. <https://doi.org/10.1109/ACCESS.2021.3090474>
- Pinamonti, M. (2022). *Alzheimer MRI 4 classes dataset*. Alzheimer MRI 4 Classes Dataset. <https://www.kaggle.com/datasets/marcopinamonti/alzheimer-mri-4-classes-dataset>
- Pradhan, A., Gige, J., & Eliazer, M. (2021). Detection of Alzheimer's Disease (AD) in MRI Images using Deep Learning. *International Journal of Engineering Research*, 10(03), 1-6. <https://www.ijert.org/research/detection-of-alzheimers-disease-ad-in-mri-images-using-deep-learning-IJERTV10IS030310.pdf>
- Purcell, E. M., Torrey, H. C., & Pound, R. V. (1946). Resonance Absorption by Nuclear Magnetic Moments in a Solid. *Physical Review*, 69(1-2), 37-38. <https://doi.org/10.1103/PhysRev.69.37>
- Qian, C., Xiong, H., & Li, J. (2023). *FeDeFo: A Personalized Federated Deep Forest Framework for Alzheimer's Disease Diagnosis*. 1-6, 572-577. <https://doi.org/10.18293/SEKE2023-013>
- Reddi, S., Charles, Z., Zaheer, M., Garrett, Z., Rush, K., Konečný, J., Kumar, S., & McMahan, H. B. (2020). *Adaptive Federated Optimization (Version 5)*. arXiv. <https://doi.org/10.48550/ARXIV.2003.00295>
- Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6), 386-408. <https://doi.org/10.1037/h0042519>
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533-536. <https://doi.org/10.1038/323533a0>
- Saad, M. M., Rehmani, M. H., & O'Reilly, R. (2024). *Early Stopping Criteria for Training Generative Adversarial Networks in Biomedical Imaging*. arXiv. <https://doi.org/10.48550/ARXIV.2405.20987>

- Salehi, W., Baglat, P., Gupta, G., Khan, S. B., Almusharraf, A., Alqahtani, A., & Kumar, A. (2023). An Approach to Binary Classification of Alzheimer's Disease Using LSTM. *Bioengineering*, 10(8), 1-22, 950. <https://doi.org/10.3390/bioengineering10080950>
- Secretaría de Salud. (2021). *Enfermedad de Alzheimer, demencia más común que afecta a personas adultas mayores*. gob.mx. Consultado en noviembre 2024. <http://www.gob.mx/salud/es/articulos/enfermedad-de-alzheimer-demencia-mas-comun-que-afecta-a-personas-adultas-mayores>
- Serrano-Pozo, A., Das, S., & Hyman, B. T. (2021). APOE and Alzheimer's disease: Advances in genetics, pathophysiology, and therapeutic approaches. *The Lancet Neurology*, 20(1), 68-80. [https://doi.org/10.1016/S1474-4422\(20\)30412-9](https://doi.org/10.1016/S1474-4422(20)30412-9)
- Simonyan, K., & Zisserman, A. (2014). *Very Deep Convolutional Networks for Large-Scale Image Recognition* (Version 6). arXiv. <https://doi.org/10.48550/ARXIV.1409.1556>
- Skandarani, Y., Jodoin, P.-M., & Lalande, A. (2023). GANs for Medical Image Synthesis: An Empirical Study. *Journal of Imaging*, 9(3), 1-16, 69. <https://doi.org/10.3390/jimaging9030069>
- Smith, L. N. (2017). *Cyclical Learning Rates for Training Neural Networks*. arXiv. <http://arxiv.org/abs/1506.01186>
- Smith, L. N. (2018). *A disciplined approach to neural network hyper-parameters: Part 1 -- learning rate, batch size, momentum, and weight decay* (Version 2). arXiv. <https://doi.org/10.48550/ARXIV.1803.09820>
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). *Dropout: A Simple Way to Prevent Neural Networks from Overfitting*. <https://jmlr.org/papers/v15/srivastava14a.html>
- Stripelis, D., Gupta, U., Saleem, H., Dhinagar, N., Ghai, T., Anastasiou, R. C., Asghar, A., Steeg, G. V., Ravi, S., Naveed, M., Thompson, P. M., & Ambite, J. L. (2023). *Secure & Private Federated Neuroimaging*. arXiv. <https://doi.org/10.48550/arXiv.2205.05249>
- Takahashi, R. H., Nagao, T., & Gouras, G. K. (2017). Plaque formation and the intraneuronal accumulation of  $\beta$ -amyloid in Alzheimer's disease. *Pathology International*, 67(4), 185-193. <https://doi.org/10.1111/pin.12520>
- Tan, C.-C., Yu, J.-T., & Tan, L. (2014). Biomarkers for Preclinical Alzheimer's Disease. *Journal of Alzheimer's Disease*, 42(4), 1051-1069. <https://doi.org/10.3233/JAD-140843>
- Thompson, P. M., Hayashi, K. M., De Zubicaray, G., Janke, A. L., Rose, S. E., Semple, J., Herman, D., Hong, M. S., Dittmer, S. S., Doddrell, D. M., & Toga, A. W. (2003). Dynamics of Gray Matter Loss in Alzheimer's Disease. *The Journal of Neuroscience*, 23(3), 994-1005. <https://doi.org/10.1523/JNEUROSCI.23-03-00994.2003>
- University College London (UCL). *Minimal Interval Resonance Imaging in Alzheimer's Disease (MIRIAD)*. Dementia Research Centre. Consultado en marzo de 2025. <https://www.ucl.ac.uk/drc/research-clinical-trials/minimal-interval-resonance-imaging-alzheimers-disease-miriad>

- Wei, W., Yang, Z., Gao, Y., Li, J., Chu, C., Okada, S., & Li, S. (2023). *FedCPC: An Effective Federated Contrastive Learning Method for Privacy Preserving Early-Stage Alzheimer's Speech Detection* (No. arXiv:2311.13043). arXiv. <https://doi.org/10.48550/arXiv.2311.13043>
- World Health Organization. (2025). *Demencia*. Consultado en octubre 2023. <https://www.who.int/es/news-room/fact-sheets/detail/dementia>
- Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated Machine Learning: Concept and Applications. *ACM Transactions on Intelligent Systems and Technology*, 10(2), 1-19. <https://doi.org/10.1145/3298981>
- Zaheer, M., Reddi, S., Sachan, D., Kale, S., & Kumar, S. (2018). *Adaptive Methods for Nonconvex Optimization*. <https://dl.acm.org/doi/pdf/10.5555/3327546.3327647>
- Zeiler, M. D. (2012). *ADADELTA: An Adaptive Learning Rate Method* arXiv. <https://doi.org/10.48550/ARXIV.1212.5701>
- Zhang, X.-X., Tian, Y., Wang, Z.-T., Ma, Y.-H., Tan, L., & Yu, J.-T. (2021). The Epidemiology of Alzheimer's Disease Modifiable Risk Factors and Prevention. *The Journal of Prevention of Alzheimer's Disease*, 1-9. <https://doi.org/10.14283/jpad.2021.15>

## Acrónimos

**Tabla 25. Acrónimos.**

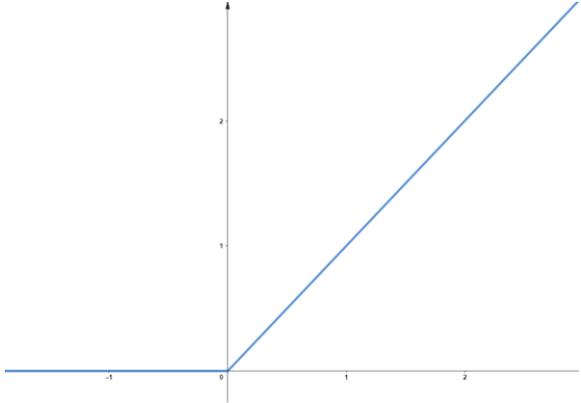
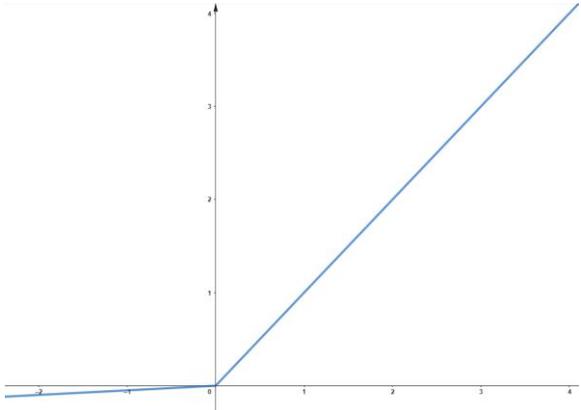
<b>Acrónimo</b>	<b>Descripción</b>
AF	Aprendizaje Federado
HFL	Aprendizaje Federado Horizontal
FTL	Aprendizaje Federado Transversal
AFA	Agregación Adaptativa de Características
VFL	Aprendizaje Federado Vertical
DL	Aprendizaje Profundo
ENLA	Atención Mejorada No Local
VAE	Autocodificadoras variacionales
HE	Cifrado Homomórfico
CDR	Clasificación Clínica de la Demencia
HC	Controles Sanos
MiD	Demencia Leve
MD	Demencia Moderada
VMD	Demencia Muy Leve
GD	Descenso de Gradiente
SDG	Descenso por Gradiente Estocástico
MCI	Deterioro Cognitivo Leve
EA	Enfermedad de Alzheimer
AIBL	Estudio Australiano sobre Envejecimiento, Imagen Médica, Biomarcadores y Estilo de Vida
HABS	Estudio Harvard sobre el Envejecimiento Cerebral
MRI	Imágenes de Resonancia Magnética
IID	Independientes e Idénticamente Distribuidos
ADNI	Iniciativa de Neuroimagen de la Enfermedad de Alzheimer
IoT	Internet de las Cosas
VRAM	Memoria de Video
MIRIAD	Estudio de Resonancia Magnética con Intervalos Mínimos en pacientes con Enfermedad de Alzheimer
BN	Normalización por Lotes
MLP	Perceptrón Multicapa
FLAIR	Recuperación de Inversión Atenuada por Fluidos
GAN	Redes Generativas Antagónicas
ANN	Redes Neuronales Artificiales

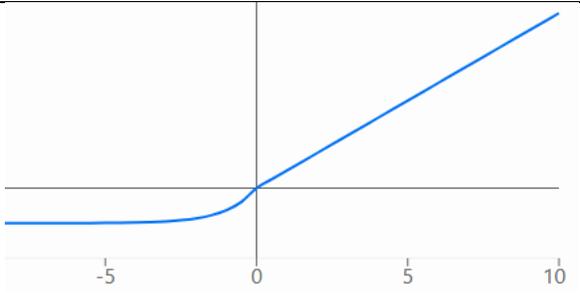
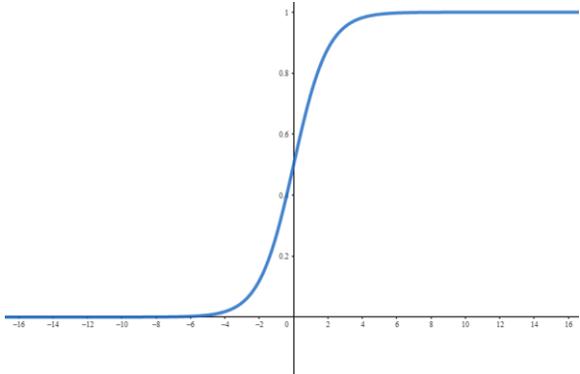
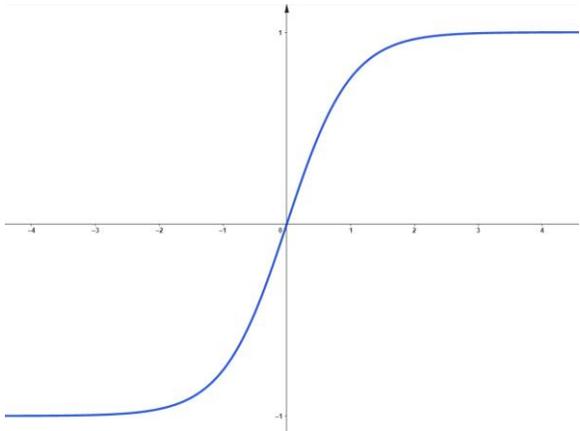
CNN	Redes Neuronales Convolucionales
MRI	Resonancia Magnética
NMR	Resonancia Magnética Nuclear
LR	Tasa de Aprendizaje
CLR	Tasas de Aprendizaje Cíclicas
PET	Tomografía por Emisión de Positrones
GPU	Unidad de Procesamiento Gráfico
ReLU	Unidad Lineal Rectificada

## Anexos

La Tabla 26 muestra las funciones de activación con mayor uso dentro del estado-del-arte.

**Tabla 26.** Funciones de activación.

Nombre de la función	Expresión	Representación gráfica
<b>ReLU</b>	$\phi(z) = \max(0, z)$	
<b>Leaky ReLU</b>	$\phi(z) = \max(\alpha z, z)$	<p>Con <math>\alpha = 0.1</math></p> 
<b>ELU</b>	$\phi(z) = \begin{cases} z & z \geq 0 \\ \alpha (e^z - 1) & z < 0 \end{cases}$	Con $\alpha = 2$

		
<b>Sigmoide</b>	$\phi(z) = \frac{1}{1 + e^{-z}}$	
<b>Tangente hiperbólica</b>	$\phi(z) = \tanh(z)$	
<b>Función SoftMax</b>	$\phi(z_i) = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}}$	