

Informe para hospitales sobre el proyecto de detección temprana de Alzheimer

Introducción

La enfermedad de Alzheimer (EA) es la causa más frecuente de demencia y provoca un impacto devastador en los pacientes y sus familias. Aunque no es un desenlace inevitable del envejecimiento, las estimaciones demográficas indican que el número de personas con EA o deterioro cognitivo leve (MCI) aumentará drásticamente en las próximas décadas **【600612056069495†L188-L200】**. Actualmente no existe una terapia capaz de revertir el daño, pero **la detección temprana y la intervención oportuna pueden retrasar la progresión de la demencia y mejorar la calidad de vida**

【600612056069495†L148-L156】. Además, identificar la EA en la fase de MCI permite a los pacientes y cuidadores planificar el futuro, acceder a tratamientos y servicios de apoyo y mantener la autonomía por más tiempo **【600612056069495†L360-L369】**. La detección precoz también genera ahorros sustanciales para los sistemas sanitarios **【600612056069495†L169-L171】**.

Las técnicas de neuroimagen (MRI) y el aprendizaje automático han demostrado ser herramientas prometedoras para detectar cambios cerebrales sutiles años antes de que aparezcan los síntomas clínicos **【600612056069495†L271-L276】**. **Este proyecto forma parte de mi trabajo de fin de posgrado en Machine Learning.** Pretende desarrollar un sistema de **clasificación multiclas** capaz de distinguir entre personas cognitivamente normales (CN), con MCI y con Alzheimer (AD) mediante imágenes de resonancia magnética (MRI) y algoritmos de aprendizaje profundo. El objetivo va más allá de cumplir con los requisitos académicos: se busca generar una herramienta operativa que, una vez entrenada, **será puesta a disposición de los hospitales de forma gratuita para apoyar el diagnóstico.** Para

entrenar modelos robustos se requiere una base de datos amplia y diversa; por ello solicitamos la colaboración de hospitales interesados en contribuir con datos anónimos.



Diagrama conceptual del pipeline

Objetivo del proyecto

El objetivo es **crear un modelo de aprendizaje automático que identifique de manera temprana y precisa la transición de normalidad cognitiva a MCI y AD**. El proyecto aspira a:

- 1. Construir un conjunto de datos balanceado** combinando recursos públicos (ADNI, OASIS) con neuroimágenes provenientes de hospitales colaboradores. El énfasis se coloca en adquirir imágenes de pacientes en fases tempranas de deterioro, que suelen estar infrarepresentadas.
- 2. Diseñar un pipeline de procesamiento de imágenes** que incluya eliminación de tejido extracerebral (skull-stripping), corrección de inhomogeneidades de intensidad (N4 bias correction), normalización espacial al atlas MNI152 y escalado de intensidades. Estas etapas mejoran la comparabilidad entre sujetos
【486845307368449†L69-L83】 【486845307368449†L93-L119】 .
- 3. Desarrollar y entrenar modelos de clasificación basados en redes convolucionales**, aprovechando transfer learning (ResNet50, MobileNetV2) para extraer patrones morfológicos del hipocampo y otras regiones asociadas con la EA. Se explorará el uso de volúmenes 3D completos y de cortes 2D para equilibrar riqueza espacial y eficiencia computacional.

4. **Evaluar el rendimiento con métricas robustas** (Precisión, Recall, F1 Score y AUC) y aplicar validación cruzada estratificada para asegurar la generalización a nuevas cohortes
【416413359508989†L80-L112】 【629241601495482†L130-L172】 .
5. **Proporcionar interpretabilidad** mediante técnicas de explicabilidad (Grad-CAM, valores SHAP) que generen mapas de calor y permitan verificar que el modelo focaliza su atención en biomarcadores conocidos (atrofia hipocampal, alargamiento de surcos, etc.).
6. **Desarrollar un modelo utilizable** que pueda integrarse en los sistemas hospitalarios y **cederse sin coste a los centros participantes** una vez completado el proyecto. Esta cesión busca facilitar la adopción de la IA en la práctica clínica sin requerir inversión adicional por parte de los hospitales.

Fuentes de datos y necesidad de colaboración

Datasets públicos

ADNI (Alzheimer's Disease Neuroimaging Initiative) es un estudio longitudinal, multicéntrico y observacional creado para validar biomarcadores de EA en ensayos clínicos 【338688597966393†L56-L60】 . Proporciona MRI, PET, datos clínicos y genéticos de miles de participantes seguidos durante años, lo que permite analizar la evolución de la enfermedad. El acceso requiere solicitar acreditación institucional y firmar un acuerdo de uso ético.

OASIS (Open Access Series of Imaging Studies) ofrece bases de datos abiertas para investigación en envejecimiento y demencia. OASIS-3 es un conjunto multimodal longitudinal con imágenes de más de un millar de sujetos (normales y con demencia) recogidas durante décadas 【407110464257885†L36-L54】 . Incluye cerca de 80 000 archivos de MRI y PET y facilita el estudio de la evolución temporal 【407110464257885†L65-L83】 .

Ambos recursos son valiosos, pero presentan limitaciones: requieren tiempo para el acceso, pueden carecer de determinados grupos de edad o etnias y, sobre todo, el **número de casos de MCI temprano es escaso**. Por ello solicitamos a los hospitales interesados que aporten MRI de pacientes con MCI y AD, garantizando la diversidad y la representatividad del conjunto.

Solicitud de colaboración hospitalaria

Los hospitales cuentan con archivos de neuroimágenes y diagnósticos clínicos que podrían enriquecer el proyecto. Proponemos firmar convenios de colaboración mediante los cuales los centros cedan imágenes MRI (preferentemente T1 ponderadas) y diagnósticos asociados en formato anónimo y desidentificado. Cumpliremos íntegramente las normativas europeas (GDPR) y los marcos éticos internacionales (HIPAA), aplicando estrategias de anonimización (eliminación de metadatos personales y re-etiquetado) y alojando los datos en servidores seguros. Se ofrecerán a los centros colaboradores:

- Acceso a los resultados y a las versiones finales del modelo, para que puedan evaluar su rendimiento en sus pacientes.
- Reconocimiento en publicaciones científicas y difusión en seminarios y congresos.
- Posibilidad de participar como coautores en artículos y estudios derivados.
- **Recepción del modelo entrenado sin coste alguno**, con soporte para su integración en las plataformas clínicas y guía para su uso.

Pipeline metodológico

1. Recopilación y organización de datos

- **Selección de sujetos**: elegir participantes CN, MCI y AD según criterios clínicos. Para cada sujeto se registrará edad, sexo, diagnóstico y fecha de adquisición (sin datos identificativos).
- **Etiquetado**: asignar clase (CN, MCI, AD) con base en las evaluaciones neurológicas y neuropsicológicas. El desequilibrio de clases se mitigará mediante sobremuestreo de la minoría o ponderación de la pérdida durante el entrenamiento.
- **Almacenamiento seguro**: alojar las imágenes en repositorios cifrados con control de acceso.

2. Preprocesamiento de imágenes

1. **Skull-stripping** – se aplicará un algoritmo de eliminación de tejido no cerebral para aislar el cerebro y reducir el ruido; se utilizan métodos basados en umbral, aprendizaje automático y atlas [486845307368449†L93-L119].
2. **Corrección N4 de sesgo** – el método N4 bias correction corrige inhomogeneidades de intensidad debidas a variaciones de campo,

mejorando la uniformidad y la calidad de las imágenes
【486845307368449†L69-L83】 .

3. **Normalización espacial** – registraremos cada MRI al atlas estándar MNI152 para alinear las estructuras anatómicas entre sujetos.
4. **Rescalado de intensidades y recorte** – se convertirán los valores de voxel a una escala uniforme y se recortará la región de interés para reducir memoria.
5. **Augmentación** – se generarán ejemplos sintéticos (rotaciones, desplazamientos, flip horizontal) para aumentar la variabilidad y mejorar la generalización.

3. Diseño de modelos

- **Modelos 2D:** se empleará transfer learning con arquitecturas preentrenadas (ResNet50, MobileNetV2) adaptadas para imágenes en escala de grises; se modificará la primera capa y se congelarán bloques iniciales para aprovechar características de bajo nivel. Estos modelos son eficientes y aptos para despliegues en dispositivos con recursos limitados.
- **Modelos 3D:** se evaluarán redes 3D CNN (por ejemplo, 3D-ResNet, VGG-like) que procesan volúmenes completos, capturando patrones volumétricos pero requiriendo más memoria. Se probará el compromiso entre complejidad y precisión.
- **Explicabilidad:** se aplicarán técnicas como Grad-CAM y SHAP para visualizar qué regiones cerebrales influyen en la decisión y validar que el modelo atiende a biomarcadores conocidos.

4. Entrenamiento y evaluación

- **División de datos:** 70 % para entrenamiento, 15 % para validación y 15 % para prueba, procurando no mezclar volúmenes de un mismo paciente entre conjuntos.
- **Validación cruzada:** se aplicará **k-fold cross-validation** para obtener estimaciones robustas, rotando los subconjuntos de entrenamiento y prueba 【629241601495482†L130-L172】 .
- **Métricas de rendimiento:** se calcularán Accuracy, Precision, Recall y **F1 Score**, este último relevante en datasets desbalanceados porque combina precisión y exhaustividad 【416413359508989†L80-L112】 . Para las tres clases se evaluará la curva ROC/AUC por clase y se utilizarán matrices de confusión para analizar errores.

5. Implementación y despliegue

El desarrollo se realizará en Google Colab utilizando cuadernos Jupyter (IPYNB). Los modelos se entrenarán en GPUs, y al finalizar se construirá una API que permita integrar la herramienta en sistemas hospitalarios de información. El código será abierto para que los socios puedan reproducir los resultados y proponer mejoras. Asimismo, se elaborará documentación detallada para facilitar la implementación del modelo en entornos clínicos.

Cronograma estimado

A continuación se propone un cronograma de 6 meses (24 semanas) dividido en fases. Cada fase puede solaparse en función de la disponibilidad de datos y recursos.

1. **Semanas 1-4 – Preparación y acuerdos:** firma de convenios con hospitales, solicitud de acceso a ADNI/OASIS, definición de criterios de inclusión y diseño del protocolo de anonimización.
2. **Semanas 5-8 – Recolección y preprocesamiento:** recopilación de MRI, aplicación de skull-stripping, corrección N4 y normalización; creación del set de datos etiquetado y balanceado.
3. **Semanas 9-12 – Modelos base:** entrenamiento de modelos 2D y 3D iniciales; evaluación con k-fold cross-validation y ajuste de hiperparámetros; selección de la arquitectura más prometedora.
4. **Semanas 13-16 – Optimización y explicabilidad:** implementación de técnicas de regularización (dropout, batch normalization), fine-tuning de capas y generación de mapas de calor con Grad-CAM/SHAP para interpretar resultados.
5. **Semanas 17-20 – Validación y análisis:** pruebas finales en el conjunto independiente; análisis estadístico, comparación con referencias de literatura; preparación de informes técnicos y ajustes para despliegue.
6. **Semanas 21-24 – Documentación y difusión:** redacción del informe final, preparación de presentación para los hospitales colaboradores, envío de resultados, elaboración de guía de implementación y discusión sobre posibles publicaciones científicas.

Ética y cumplimiento legal

El proyecto se regirá por los principios de confidencialidad, consentimiento informado y minimización de riesgos. Las imágenes y los datos clínicos serán tratados de manera anónima y se implementarán salvaguardas técnicas (cifrado, control de acceso, auditorías). Cualquier uso de datos estará amparado por acuerdos escritos con los centros y se revisará por comités de ética.

Respetaremos la normativa europea de protección de datos (GDPR) y la normativa española (LOPD-GDD), así como las directrices sobre datos médicos (HIPAA). La finalidad investigadora se comunicará claramente a los pacientes o sus representantes y se garantizará la posibilidad de retirada de la información.

Beneficios para los hospitales colaboradores

- **Contribución a la ciencia y a la sociedad:** su colaboración permitirá desarrollar un sistema que beneficie a miles de personas y ayude a anticipar el tratamiento del Alzheimer.
- **Valorización de la infraestructura:** los modelos podrán integrarse en flujos de trabajo hospitalarios, ofreciendo una herramienta de apoyo al diagnóstico que optimiza el tiempo de los neurorradiólogos.
- **Formación y difusión:** se organizarán talleres y seminarios para el personal clínico, donde se explicará el funcionamiento del modelo, sus limitaciones y su interpretación.
- **Reconocimiento académico:** los hospitales aparecerán como colaboradores en publicaciones científicas y memorias del proyecto.
- **Acceso libre a la tecnología:** al final del proyecto, el modelo entrenado se entregará sin coste, acompañado de soporte técnico y manuales, para que los hospitales puedan beneficiarse directamente de la herramienta.

Conclusión y llamado a la acción

La enfermedad de Alzheimer supone un reto sanitario global y, aunque actualmente no existan tratamientos curativos, **diagnosticarla en fases precoces** permite acceder a terapias que mejoran el pronóstico y a intervenciones que retrasan la aparición de la demencia

[600612056069495†L148-L156] [600612056069495†L271-L276].
Estudios demuestran que los pacientes tratados tras un diagnóstico

temprano presentan mejor supervivencia y menores tasas de institucionalización [【600612056069495†L372-L377】](#). Sin embargo, los datos públicos son insuficientes para entrenar modelos robustos; por ello necesitamos su colaboración.

Como alumno de posgrado, mi objetivo es culminar el proyecto académico y, al mismo tiempo, **poner a disposición de los hospitales una herramienta de detección temprana de Alzheimer sin coste alguno**. La colaboración con hospitales que aporten imágenes y datos clínicos hará posible entrenar modelos más precisos y generalizables. Invitamos a los centros sanitarios a unirse a este proyecto; recibirán reconocimiento académico y acceso a una tecnología avanzada que podrá ayudar a sus pacientes.