

Proyecto final - Brain Tumor MRI Challenge

Introducción a Los Sistemas Inteligentes – 2025-1

Fecha límite de entrega: Martes 22 de Julio 2025

Grupos de máximo 3 personas

Objetivo

Desarrollar modelos de clasificación que permitan clasificar automáticamente si una imagen médica por resonancia magnética (MRI) muestra la presencia de un tumor cerebral y, en caso afirmativo, identificar el tipo de tumor. Adicionalmente, se busca comparar el desempeño del mejor modelo de clasificación, respecto a qué tan bien lo hace un LLM.

Descripción del problema

El análisis de imágenes cerebrales mediante resonancia magnética es un proceso rutinario en el diagnóstico de afecciones cerebrales, pero por la naturaleza en que estas se toman, representan también un desafío en entornos clínicos. Por esta razón, la clasificación automática de tumores cerebrales busca ser una manera de reducir la carga del personal médico y mejorar los tiempos de diagnóstico, especialmente en zonas con escaso acceso a especialistas en neurología.

Para este proyecto, se les plantea la construcción de un modelo capaz de clasificar imágenes MRI cerebrales en cuatro categorías:

- Glioma
- Meningioma
- Pituitario
- No tumor

¿Qué deben tener en cuenta?

- La categoría "no tumor" representa imágenes normales o benignas sin evidencia de masas cerebrales.
- El modelo no sólo debe ser capaz de detectar tumores, sino también de reconocer correctamente la ausencia de estos.
- **IMPORTANTE DADO EL CONTEXTO:** El diagnóstico erróneo en cualquiera de los sentidos (falsos positivos o falsos negativos) puede tener consecuencias graves. Por esto, se requiere un modelo sensible y preciso.

¿Qué son las imágenes MRI que contiene el Dataset utilizado?

Las imágenes hechas por resonancia magnética (MRI) son evidencia de una técnica avanzada de imagen médica la cual utiliza campos magnéticos y ondas de radio para generar imágenes detalladas del cuerpo humano, en particular del cerebro.

Clases en el dataset

- **Glioma:** Tumores originados en las células gliales. Suelen ser invasivos y clínicamente agresivos.

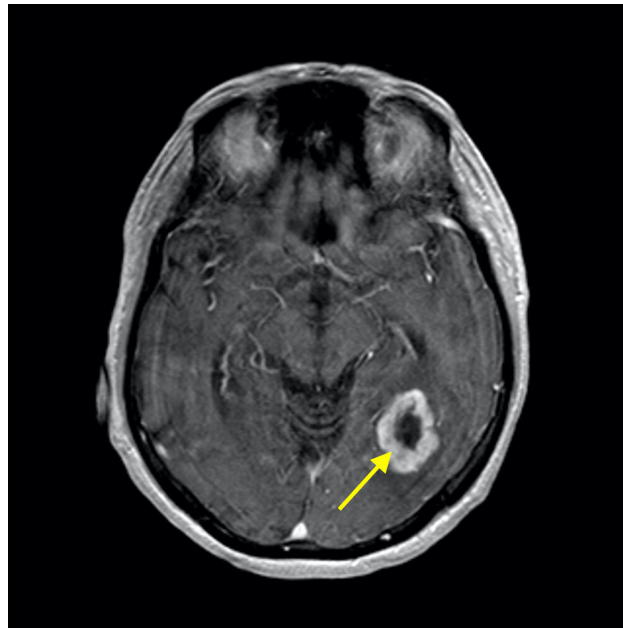


Figure 1: Tumor glioma

- **Meningioma:** Tumores en las meninges, habitualmente benignos, aunque su crecimiento puede afectar estructuras cerebrales adyacentes.



Figure 2: Tumor Meningioma

- **Pituitario:** Tumores en la glándula pituitaria. Pueden afectar funciones endocrinas y causar síntomas variados.

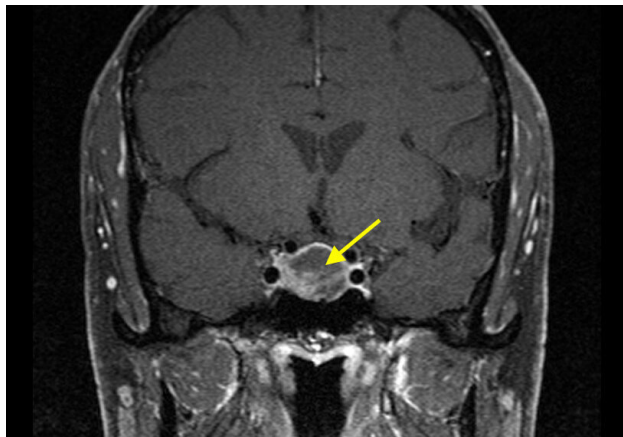


Figure 3: Tumor Pituitario

- **No tumor:** Imágenes cerebrales sin evidencia patológica detectable. Es esencial distinguirlas con alta certeza para evitar sobretratamientos.

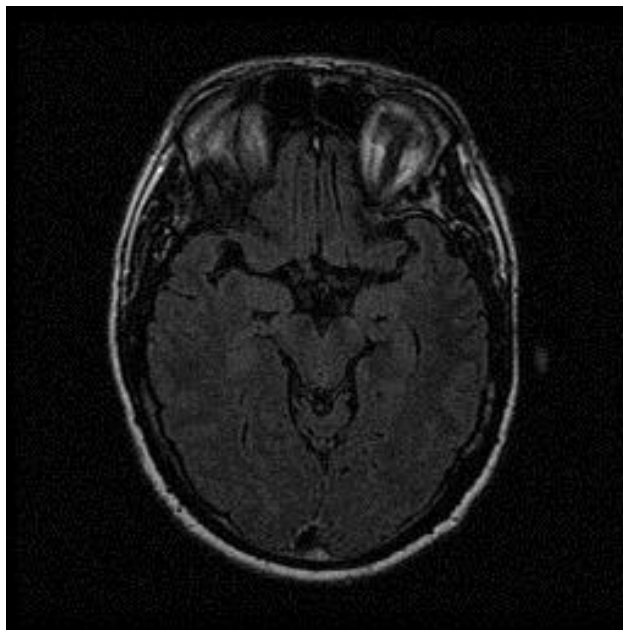


Figure 4: Cerebro sin tumor

Importancia clínica y métrica de evaluación

En problemas médicos, minimizar los errores de diagnóstico es prioritario. Una mala clasificación puede derivar en intervenciones quirúrgicas innecesarias o en la omisión de tratamientos vitales. Para capturar adecuadamente este riesgo, se utilizará una métrica compuesta basada en:

- **Recall Macro:** promedio de la sensibilidad (recall) para cada clase. Penaliza el mal desempeño en clases menos representadas.
- **Accuracy:** proporción general de predicciones correctas.

La métrica final será:

$$\text{Score Final} = \frac{\text{Recall Macro} + \text{Accuracy}}{2}$$

Esto obliga a los modelos a balancear correctamente las clases, evitando sesgos hacia clases dominantes.

Investigación de técnicas de preprocesamiento

Cada grupo deberá investigar y documentar al menos dos técnicas modernas de preprocesamiento para imágenes MRI, justificando su uso con referencias bibliográficas. Se espera que incluyan una sección en el notebook explicando:

- Qué técnicas consideraron
- Por qué las aplicaron (o decidieron no hacerlo)
- Qué impacto esperaban tener sobre el desempeño del modelo

Acceso a la competencia

Se habilitó una competencia en la plataforma Kaggle para medir qué tan bien clasifica su modelo de acuerdo a la métrica compuesta previamente hablada. En Kaggle:

- Deben registrarse en la plataforma con su correo institucional.
- Tienen acceso el dataset de entrenamiento y prueba.
- Deberán subir sus predicciones en formato `.csv`
- Obtienen un score basado en la métrica descrita.
- Con base a este score se organizarán por cada equipo este puntaje en un leaderboard de la competencia, una métrica más alta lo pondrá en un puesto más alto aquí.

Enlace para unirse a la competencia de kaggle:

<https://www.kaggle.com/t/2c8c60c70ec72653cd44d53217feaadd>

Formato del archivo de envío:

Pueden ver el formato del archivo de envío en la descripción de la competencia en Kaggle

Estructura de los archivos contenidos en la competencia

- **train/**: imágenes MRI de entrenamiento
- **train_labels.csv**: etiquetas correspondientes
- **test/**: imágenes MRI sin etiquetas

Todas las imágenes están en formato `.jpg`.

Etapas del proyecto

Deben entregar un notebook para resolver el problema el cual debe tener las siguientes secciones:

- **A. Descripción del problema**
Deben hacer una descripción clara del problema que se quiere resolver. Cuál es la relación del problema clínico con el uso de modelos de aprendizaje profundo, definir los términos técnicos clave y establecer cómo se aplican técnicas de clasificación de imágenes al contexto médico.

- **B. Importación de datos**

En esta sección se deben importar todas las bibliotecas necesarias y cargar los datos desde las fuentes correspondientes, esto incluye archivos `.csv` con las etiquetas y las imágenes MRI de entrenamiento y prueba.

- **C. Exploración de datos**

Se espera un análisis exploratorio completo que permita entender la estructura del dataset, usted con su grupo puede realizar:

- Visualización de imágenes por clase.
- Conteo de instancias por clase para analizar el balance.
- Revisión de tamaños y formatos de imágenes.
- Verificación del canal de color y rangos de intensidad.
- Detección de imágenes corruptas u outliers.

- **D. Preprocesamiento de imágenes**

Aquí deben reflejar el resultado de la investigación previa realizada por el grupo donde luego de realizar una búsqueda de literatura o documentación de técnicas apropiadas de preprocesamiento para imágenes de resonancia magnética, puede encontrar técnicas como:

- Normalización de intensidades.
- Redimensionamiento uniforme.
- Conversión de escala de grises o estándar RGB.
- Eliminación de ruido.
- Otras técnicas.

La implementación debe estar acompañada de una justificación clínica, computacional o de porqué creen que mejoraría el desempeño del modelo y debe incorporarse directamente en el pipeline de este.

- **E. Modelos y ajuste de hiperparámetros**

Aquí se deben entrenar los modelos de clasificación resultantes. Las opciones incluyen redes convolucionales desde cero, arquitecturas preentrenadas mediante Transfer Learning, enfoques híbridos, etc. Deben mostrarse también:

- Curvas de aprendizaje (loss y accuracy)
- Ajuste de hiperparámetros
- Técnicas de regularización (si utilizan)
- Validación cruzada, si creen que les serviría.

- **F. Evaluación del desempeño del modelo final**

Se debe evaluar el modelo con un enfoque centrado en la sensibilidad y exactitud, considerando la naturaleza crítica del contexto médico. Las métricas de evaluación obligatorias son:

- **Recall por clase (macro)**
- **Accuracy global**
- **Métrica compuesta: $(\text{Recall_macro} + \text{Accuracy}) / 2$**
- Matriz de confusión con análisis detallado por clase
- Curva ROC por clase y cálculo del área bajo la curva (AUC)
- Visualización de errores y análisis de casos mal clasificados

Cada métrica debe acompañarse de una interpretación clara.

- **G. Prueba y evaluación con un LLM**

En esta sección deberán implementar un segundo enfoque de clasificación de imágenes utilizando un **Large Language Model (LLM)**, además de su mejor modelo entrenado. La idea es comparar el desempeño de un modelo generativo (como Gemini o ChatGPT) frente a su modelo de clasificación. Ahora, no importa que no puedan clasificar todas las imágenes de prueba con el LLM por límites intrínsecos al modelo, lo importante es que hagan el ejercicio de clasificar con estos modelos.

Pasos sugeridos:

- Utilicen una plataforma que permita la carga de imágenes en LLMs, como **Google AI Studio**. Se recomienda crear un proyecto y usar uno de los modelos allí dispuestos, para utilizarlos pueden generar un API key.
- Clasifiquen las imágenes del conjunto de prueba solicitando al modelo una predicción de clase por medio de **prompts** diseñados por ustedes.
- Registren las respuestas del modelo para cada imagen.

Evaluación requerida:

- Así como lo realizaron con el modelo de clasificación deben calcular las mismas métricas:
- **Respondan:**
 - * ¿Cómo se desempeñó el LLM frente a cada clase?
 - * ¿Qué técnicas mejoraron el rendimiento? (por ejemplo, reformulación del prompt, forma de ingreso de las imágenes, contexto adicional, etc.)
 - * ¿Qué limitaciones identificaron en el enfoque con LLM?
 - * ¿Qué ventajas ofrece este tipo de modelo respecto al modelo tradicional?

Documentación de referencia:

Para explorar cómo interactuar con un LLM a partir de imágenes, pueden consultar la guía oficial de Gemini:

<https://ai.google.dev/gemini-api/docs/image-understanding?hl=es-419>

Nota: Pueden usar cualquier modelo LLM con capacidad multimodal. Gemini (via Google AI Studio) es recomendado por su facilidad de uso y soporte para carga directa de imágenes. Recuerden también registrarse con su cuenta personal, ya que google AI studio no está disponible con el correo institucional.

- **H. Generación de los archivos de envío para la competencia**

Se deben generar un archivo **submissionModel.csv** que contenga las predicciones de

los modelo sobre el conjunto de prueba, siguiendo el formato requerido por la competencia en Kaggle.

- **I. Conclusiones y resultados**

En esta sección se deben discutir los hallazgos generales del proyecto, tanto de su modelo de clasificación como la comparación de este frente al LLM de elección

Entregable

Además de realizar los envíos al sitio de la competencia, se debe entregar:

- **Jupyter Notebook:** El Notebook del proyecto junto con todas sus etapas debe estar estructurado, explicado, comentado y funcional, con las secciones descritas previamente. Asegúrense de que el archivo se ejecute correctamente, que los resultados de cada celda se visualicen de manera apropiada y que las celdas de texto expliquen con claridad cada paso del trabajo.
- **Estructura de nombre del archivo IPYNB:**
`isi-proj-userunal1_userunal2_userunal3.ipynb`
- **Enlace de entrega:**
<https://www.dropbox.com/request/rzNu5hzkl9Kudg0Dv4Xu>
- **Formato:**
El archivo debe tener ese nombre y entregarse antes de la medianoche en la fecha límite. No se aceptan entregas fuera de tiempo o con nombres incorrectos.

Evaluación

Todos los grupos deben participar en la competencia de clasificación en Kaggle y alcanzar al menos un **40% de score**. Si esto no se cumple, no se evaluará el notebook. Adicionalmente la evaluación estará estructurada de la siguiente manera:

- **80%** Calidad técnica del notebook: investigación, visualización, estructura, claridad y análisis. Se revisará especialmente que completen todas las etapas que debe contener el modelo.
- **20%** Desempeño en la competencia por el submission del modelo de clasificación.

Nota: Lo calificación del desempeño sera proporcional a la métrica compuesta alcanzada en el leaderboard oficial.