

# Clasificación de Tumores Cerebrales

<b>Estudiante:</b>	Arbey de Jesús Villegas Carvajal
<b>Documento:</b>	CC 98695639
<b>Asignatura:</b>	Introducción a Deep Learning
<b>Profesor:</b>	Raúl Ramos
<b>Carrera:</b>	Ingeniería de Sistema (Virtual)
<b>Institución:</b>	Universidad de Antioquia

## Contexto

Un tumor cerebral es una enfermedad particularmente agresiva. En el año 2020, se estimó que 308.102 fueron diagnosticadas con tumor cerebral en todo el mundo. La mejor técnica para detectar tumores cerebrales son las Imágenes por Resonancia Magnética (IRM). Los tumores cerebrales, más que cualquier otro cáncer, pueden alterar fuertemente la vida de las personas a nivel físico, cognitivo y psicológico por lo que un diagnóstico temprano puede llevar a un tratamiento adecuado que pueda mejorar la esperanza de vida y el bienestar de los pacientes que padecen la enfermedad. Las redes neuronales han demostrado una precisión importante en los problemas de segmentación y clasificación de imágenes.

## Objetivo

El Aprendizaje Automático (ML) y la Inteligencia Artificial (AI) han demostrado consistentemente precisión al momento de usarse para tareas de clasificación. Por lo tanto, para la clasificación de las imágenes de tumores cerebrales, que son resultado de IRM, vamos a utilizar Redes Neuronales Convolucionales, un tipo de algoritmo del Deep Learning. Tomaremos un dataset de la plataforma Kaggle para este propósito y utilizaremos los conocimientos aprendidos en este curso. En el siguiente apartado se describe el dataset.

## Características del dataset

### General

Número de imágenes:	3264
Tamaño en disco:	95.2 MB
Ubicación dataset:	<a href="#">brain tumor classification dataset</a>

### Distribución de muestras para training

Clase	Número de imágenes
Tumor glioma	826
Tumor meningioma	822
Tumor en la pituitaria	827
Sin tumor	395
Total	2870

### Distribución de muestras para testing

Clase	Número de imágenes
Tumor glioma	100
Tumor meningioma	115
Tumor en la pituitaria	74
Sin tumor	105
Total	394

## Métricas de desempeño

Dado que es importante reducir la cantidad de falsos negativos, se tendrá en cuenta las métricas Accuracy, Recall y F1 score:

**Accuracy:** El accuracy es una métrica que nos permite identificar qué porcentaje de predicciones fueron correctas con respecto a las muestras totales.

**Recall:** el recall calcula cómo se comporta el modelo para identificar el porcentaje de muestras que fueron predichas correctamente como positivas entre el total de las predicciones que fueron y debieron ser pos. Esta métrica es muy importante cuando es muy alto el costo y son riesgosos los Falsos Negativos. En nuestro caso, dado que decirle a un paciente que no tiene un tumor cerebral y resulta que sí tenía, es un error que es muy delicado sobrellevar.

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

**F1 score:** esta métrica es necesaria cuando se busca tener un balance entre Precision y Recall. Primero que todo debemos definir qué es la precisión. La precisión determina la cantidad de predicciones que son de hecho positivas entre el total de predicciones positivas. La precisión es muy tenida en cuenta cuando el costo de los Falsos Positivos es muy alto.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

Ahora, el F1 score busca un balance entre Precision y Recall y la utilizaremos en el proyecto junto con Recall y Accuracy, como se había mencionado más arriba.

$$\text{F1} = 2 \times \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

## Referencias y resultados previos

El dataset que vamos a utilizar en este proyecto ha sido descargado en Kaggle 34862 veces y también tiene 281 referencias en código en el cual ha sido utilizado el dataset.

Existen 2 estudios importantes donde hacen uso de este dataset. Voy a hacer solo referencia a uno de ellos: A Review on Automated Brain Tumor Detection and Segmentation, donde los autores no se concentran en el dataset per se sino que con su uso buscan las mejores alternativas para analizar, transformar, procesar datos provenientes de IRM para así ser usados en modelos predictivos de clasificación que permitan con la mayor precisión posible identificar los tumores cerebrales automáticamente para el apoyo de diagnóstico temprano.