

Clasificación de Tumores Cerebrales

Estudiante:	Arbey de Jesús Villegas Carvajal
Documento:	CC 98695639
Asignatura:	Introducción a Deep Learning
Profesor:	Raúl Ramos
Carrera:	Ingeniería de Sistema (Virtual)
Institución:	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 Github para este propósito y utilizaremos los conocimientos aprendidos en este curso. En el siguiente apartado se describe el dataset.

Dataset (número de datos y tamaño en disco), distribución de las clases

General

Número de imágenes:	3264
Tamaño en disco:	95.2 MB
Ubicación dataset:	brain-tumor-classification-dataset

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.

Recall: el recall calcula cómo se comporta el modelo para identificar las muestras que deben ser positivas. Esta métrica es muy importante cuando es muy alto el costo con los Falsos Positivos.

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

F1 score: esta métrica es necesaria cuando se busca tener un balance entre Precisión y Recall. Primero que todo debemos definir qué es la precisión. La precisión determina la cantidad de predicciones que son de hecho 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.

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