

Evaluación, refinamiento y comparación de modelos

Diego Arturo Padilla Domínguez
A01552594

Introducción

En este reporte se llevará a cabo la comparación de distintos modelos de “Machine Learning” que tienen un mismo fin, la clasificación de 3 tipos de tumores cerebrales y también decir si no existe. Se utilizará como métrica de evaluación “accuracy” debido a que es la más usada en este tipo de situaciones [1].

Preprocesamiento de los datos

Se tiene un conjunto de imágenes categorizadas por el nombre de cada tipo de tumor, el conjunto de datos no se le puede hacer un aumento de datos muy grande, debido a las limitantes del problema, ya que cambiar la forma de la imagen podría empeorar el desempeño del modelo en casos reales.

Los tipos de modificación que se hicieron son los siguientes:

- horizontal_flip
- vertical_flip

Estos fueron debido a que solo cambian el punto de vista de la imagen pero no hacen cambios que modifiquen la forma de la imagen.

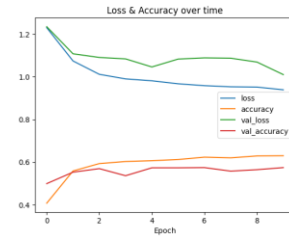
Modelo 1

En la creación de este modelo se utilizaron tan solo 5 capas básicas de keras, entre las que se encuentra:

- Conv2d (relu)
- GlobalAveragePooling2D
- Dropout
- Densa (relu)
- Densa (softmax)

El entrenamiento dio los siguientes resultados:

Precisión: 57%



Se puede observar que el modelo no está ‘overfited’ ya que tanto el dataset de entrenamiento como el de pruebas tienen una precisión bastante similar, siendo la diferencia cada vez mas pequeña con el paso de las épocas.

Modelo 2

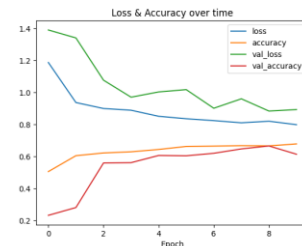
En el anterior modelo se vieron los resultados con un poco cantidad de capas y esto nos dio un resultado con una precisión muy baja, por ello se creó un nuevo modelo añadiendo nuevas capas:

- BatchNormalization
- Flatten
- 2 Dense (relu)
- Dropout

El objetivo de esto es mejorar la precisión del modelo

Tras el entrenamiento, estos fueron los resultados:

Precisión: 66.8%



Se puede ver que el modelo no se encuentra ‘overfited’, debido a que la precisión de ambas partes del dataset, entrenamiento y pruebas, son casi iguales.

Con respecto al modelo anterior se ha podido observar una mejora de

un poco más del 10%, siendo señal de que modelos más robustos pueden llegar a dar mejores resultados en este tipo de situación

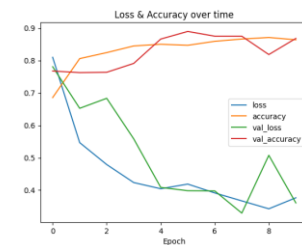
Modelo 3

Viendo el resultado de los modelos anteriores podemos concluir que solo agregando más capas al modelo puede mejorar.

Dadas las observaciones se procedió al uso de transfer learning, utilizando el modelo VGG16 debido a que es el modelo menos profundo de clasificación con mejores resultados [2], en donde este será la nueva capa de entrada, esto hará que el modelo sea mas robusto y de mejores resultados.

Estos fueron los resultados del entrenamiento

Precisión: 87.4%



Este modelo es más fluctuante entre cada una de sus épocas en comparación a los modelos anteriores, pero esto no provoca que el modelo este ‘overfited’.

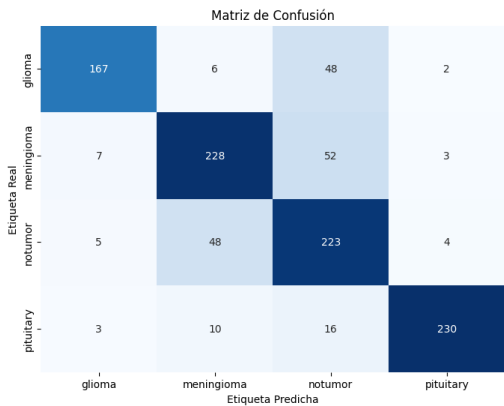
El modelo ha mejorado en gran medida con respecto a los dos modelos anteriores, siendo un resultado aceptable en el contexto académico como aprendizaje del entrenamiento de modelos de clasificación.

Nota: Todos los modelos fueron entrenados con los siguientes parámetros:

- 10 épocas
- 186 steps por época
- 32 steps de validación
- Guardada la mejor época

Comparación de modelos			
Parámetro / N modelo	1	2	3
Precisión	57%	66.8%	87.4%
Tiempo de entrenamiento	12 minutos	14 min	3.3 horas

Análisis resultados



Se puede observar que la diagonal esta bien marcada, simbolizando el buen comportamiento del modelo, aunque si esto fuese aplicarse a un caso real, no es un modelo correcto, debido a que al ser la detección de una condición médica, el tener falsos positivos en la parte de “nottumor” sería de alto riesgo, provocando poner en alto riesgo la salud de una persona, el modelo tendría que modificarse para forzar que esto no ocurra.

Conclusión

En conclusión, se compararon tres modelos de aprendizaje automático para la clasificación de tumores cerebrales en tres categorías diferentes, así como la detección de la ausencia de tumor.

En general, se observó que la adición de capas y el uso de transfer learning condujeron a mejoras significativas en la precisión del modelo. El modelo más complejo y robusto, basado en transfer learning con VGG16, demostró los mejores resultados con una precisión del 87.4%.

Es importante destacar que estos modelos fueron entrenados y evaluados en un contexto académico y su aplicabilidad en casos reales debe ser abordada con precaución. Por ejemplo, en el caso de la detección de tumores cerebrales, los falsos positivos podrían tener consecuencias graves para la salud de una persona. Por ello, en un escenario clínico real, se requerirían ajustes adicionales para minimizar los falsos positivos y garantizar la seguridad de la persona analizada.

Referencias

[1] Image Classification Metrics - Getting Started with Image Classification with PyTorch. Educative: Interactive Courses for Software Developers [en línea]. [consultado el 31 de mayo de 2023]. Disponible en: <https://www.educative.io/courses/getting-started-with-image-classification-with-pytorch/xVvy4y5WAQn>

[2] Keras documentation: Keras Applications. Keras: Deep Learning for humans [en línea]. [consultado el 31 de mayo de 2023]. Disponible en: <https://keras.io/api/applications/>