

Clasificación de tumores cerebrales con modelos de redes neuronales

Fernanda Quesada Chavarría¹, Yarin Vargas Valverde¹, Marie Sofia Villalobos Martínez¹

fernanda.quesada@ucr.ac.cr, yarin.vargas@ucr.ac.cr, marie.villalobos@ucr.ac.cr

RESUMEN

Este artículo presenta un análisis comparativo del rendimiento de diferentes modelos de redes neuronales aplicados a la clasificación de imágenes de resonancia magnética cerebral para la detección de tumores. Se evalúan cuatro modelos: un modelo básico, uno con regularización L2, otro con una tasa de aprendizaje ajustada y un modelo de redes neuronales convolucionales (CNN). Los resultados muestran que el modelo CNN alcanza el mejor desempeño, con una precisión del 99.90% en el conjunto de prueba y una precisión del 93.13% en validación. Sin embargo, los resultados sugieren la necesidad de optimizar el modelo para mejorar la clasificación de tumores específicos, como el meningioma.

PALABRAS CLAVE: Clasificación de imágenes, detección de tumores, modelos CNN, tumores cerebrales.

ABSTRACT

This article presents a comparative analysis of the performance of different neural network models applied to the classification of brain MRI images for tumor detection. Four models were evaluated: a basic model, one with L2 regularization, another with an adjusted learning rate, and a convolutional neural network (CNN) model. The results show that the CNN model achieved the best performance, with an accuracy of 99.90% on the test set and 93.13% accuracy in validation. However, the findings suggest the need to optimize the model to improve the classification of specific tumors, such as meningiomas.

KEYWORDS: Image classification, Brain tumors, Tumor detection, CNN models.

INTRODUCCIÓN

Los tumores representan una de las patologías más retadoras en el campo médico, teniendo un impacto en la salud de los pacientes, especialmente, tienen mayor complejidad en el área de la neurología y la oncología. Cada año, miles de personas son diagnosticadas con tumores cerebrales, los cuales pueden variar en gravedad y pronóstico. La detección temprana de estos tumores aporta a conocer el tipo de tratamiento que se debe aplicar y también, a conocer las probabilidades de recuperación del paciente.

Gómez et al. (2019) mencionan que los tumores cerebrales primarios en adultos, como los meningiomas, gliomas y adenomas pituitarios, presentan una considerable diversidad histológica y clínica. Los meningiomas, son los más frecuentes, suelen ser benignos y tratables mediante cirugía. Por otro lado, los gliomas, especialmente los glioblastomas, son altamente agresivos y tienen un mal pronóstico, con una baja tasa de supervivencia a cinco años. Los adenomas pituitarios, aunque son menos comunes, pueden causar síntomas neurológicos y

¹ Estudiantes de Estadística de la Universidad de Costa Rica

endocrinos significativos, la caracterización precisa de estos tumores es esencial para mejorar los resultados clínicos

Según Roy et al. (2013), existen diferentes métodos de detección y segmentación de tumores cerebrales, las imágenes de MRI (Imágenes por Resonancia Magnética) son de gran importancia para esto, ya que proporcionan información sobre la anatomía humana de los tejidos blandos y lo convierte en una herramienta eficaz y fiable para detectar el cáncer.

Estudios recientes han demostrado que el uso de redes neuronales y técnicas de aprendizaje puede mejorar la precisión de la detección de tumores, según Kouli et al. (2022) realizaron una revisión sistemática y un metanálisis de detección utilizando un modelo jerárquico unificado y realizando un análisis de sensibilidad para validar el estudio. El generar diagnósticos derivados de las resonancias demanda tiempo y es un desafío en un sistema de salud, por lo que en base a esto la solución de un sistema automatizado podría agilizar el proceso. Esta automatización del diagnóstico mediante técnicas de aprendizaje puede ayudar a los radiólogos u otros profesionales a identificar tumores cerebrales de manera rápida y precisa, aumentando la eficiencia del diagnóstico y la accesibilidad a la alta calidad de estos.

METODOLOGÍA

La base de datos utilizada fue obtenida de *Kaggle* llamada *Brain Tumor MRI Dataset* consta de 7023 imágenes de resonancia magnética cerebral organizadas en cuatro clases: glioma, meningioma, pituitario y no tumor. Cada imagen está en formato JPG con una resolución estándar de 512x512 píxeles. Las imágenes están agrupadas en carpetas según la categoría correspondiente. Este conjunto de datos es adecuado para proyectos de clasificación de imágenes, facilitando el entrenamiento de modelos de aprendizaje profundo para la detección automática de tumores cerebrales.

La preparación de los datos implica varios pasos fundamentales. Se cargan las imágenes desde las carpetas correspondientes y se verifica que cada imagen sea etiquetada correctamente. Luego, las imágenes se normalizan para que los valores de sus píxeles estén en un rango entre 0 a 1, lo cual es crucial para el correcto entrenamiento de las redes neuronales. Después de este preprocesamiento, los datos se distribuyen en tres conjuntos: entrenamiento, validación y prueba. El conjunto de entrenamiento se utiliza para ajustar el algoritmo, permitiendo que el modelo aprenda los patrones subyacentes en las imágenes. El conjunto de validación se emplea para evaluar el desempeño del modelo después de cada iteración, ayudando a ajustar la función de pérdida y actualizar los pesos mediante el optimizador seleccionado. Al finalizar el proceso de entrenamiento, el modelo se evalúa con el conjunto de pruebas, con el fin de medir su capacidad para generalizar y su efectividad en datos nuevos.

El presente trabajo implementa tres arquitecturas de redes neuronales distintas. El primer modelo es un Perceptrón Multicapa (MLP), que utiliza capas densas para procesar las imágenes aplanadas. El segundo modelo es una Red Neuronal Convolucional (CNN), diseñada para capturar patrones espaciales a través de capas convolucionales y de agrupamiento máximo (max pooling), siendo más adecuada para el procesamiento de imágenes. El tercer modelo también es un CNN, pero incorpora una aumentación de datos, aplicando

transformaciones como rotaciones y volteos durante el entrenamiento. En cada uno de estos modelos se optimiza al menos un hiper parámetro, esto podría ser la tasa de aprendizaje o el tipo de optimizador. Se prueban varias configuraciones, incluyendo optimizadores como Adam, con el objetivo de maximizar el rendimiento. La evaluación del desempeño se lleva a cabo utilizando el conjunto de pruebas, y se miden métricas como la precisión, la exactitud y el F1-Score. También se utiliza una matriz de confusión para identificar los errores de clasificación y analizar el comportamiento del modelo. Por otro lado, se utilizan técnicas de interpretabilidad, como Grad-CAM, para visualizar las áreas de las imágenes que el modelo considera relevantes al tomar decisiones. Esto facilita la comprensión de los patrones utilizados para identificar tumores, haciendo el sistema más transparente y confiable.

Se crearon cuatro modelos con el objetivo de comparar su rendimiento en la clasificación de imágenes de resonancia magnética cerebral. El modelo básico sirve como punto de referencia inicial sin ningún tipo de ajuste adicional. Posteriormente, se construyó un modelo con regularización L2, el cual incorpora una penalización en los pesos para evitar el sobreajuste y mejorar la generalización. El tercer modelo fue ajustado con una tasa de aprendizaje de 0.001, permitiendo un entrenamiento más controlado y progresivo, con la intención de optimizar la convergencia. Finalmente, se implementó un modelo de CNN, especializado en el procesamiento de imágenes, para capturar características espaciales de las imágenes de manera más eficiente. Estos modelos fueron evaluados y comparados para determinar cuál ofrecía el mejor rendimiento en términos de precisión y generalización. Cada modelo se estimó por medio de 15 epochs, a excepción del modelo CNN que fue estimado con 20 iteraciones.

El análisis de este proyecto se lleva a cabo utilizando el lenguaje de programación *Python* en la plataforma *Google Colab*, que permite trabajar directamente en la nube sin necesidad de instalar programas adicionales en las computadoras. Esta herramienta facilita el procesamiento de imágenes y el desarrollo de redes neuronales, ya que aprovecha los recursos de *Google*, optimizando el rendimiento y acelerando el entrenamiento de los modelos. Para este proyecto, se utilizaron librerías que facilitan el procesamiento de datos y la creación de modelos de aprendizaje profundo. Entre ellas se incluyen TensorFlow (TensorFlow Developers, 2024), matplotlib.pyplot (Hunter, 2007), pandas (McKinney, 2010), numpy (NumPy Developers, 2020), seaborn (Waskom et al., 2020) y sklearn.model_selection.train_test_split (Pedregosa et al., 2011).

RESULTADOS

Se evalúan distintos modelos de redes neuronales con el objetivo de identificar el de mejor rendimiento, cuyos resultados se resumen en la Tabla 1. El modelo CNN destaca como el más eficiente, logrando el error más bajo en el conjunto de prueba (0.0077) y la mayor precisión (99.90%), superando notablemente a los demás. El modelo con una tasa de aprendizaje diferente también muestra un buen desempeño, con un error de 0.1321 y una precisión del 95.91%. En contraste, tanto el modelo básico como el modelo con L2 Regular presentaron mayores errores y menor precisión, siendo este último el de peor desempeño general.

Al evaluar los modelos en el conjunto de validación, el CNN también resalta como el de mejor rendimiento, con una pérdida de 0.4603 y una precisión del 93.13%. Sin embargo, su desempeño en la clasificación de imágenes en el conjunto de validación, que incluye 1,311 imágenes distribuidas en cuatro clases (*glioma, meningioma, no tumor y pituitaria*), fue limitado (Tabla 2), con una precisión global de 0.26. El mejor desempeño en **recall** se da en la clase “*No tumor*” (0.33), mientras que las categorías tumorales, como meningioma, mostraron peores resultados, con un recall de solo 0.19.

Cuadro 1.

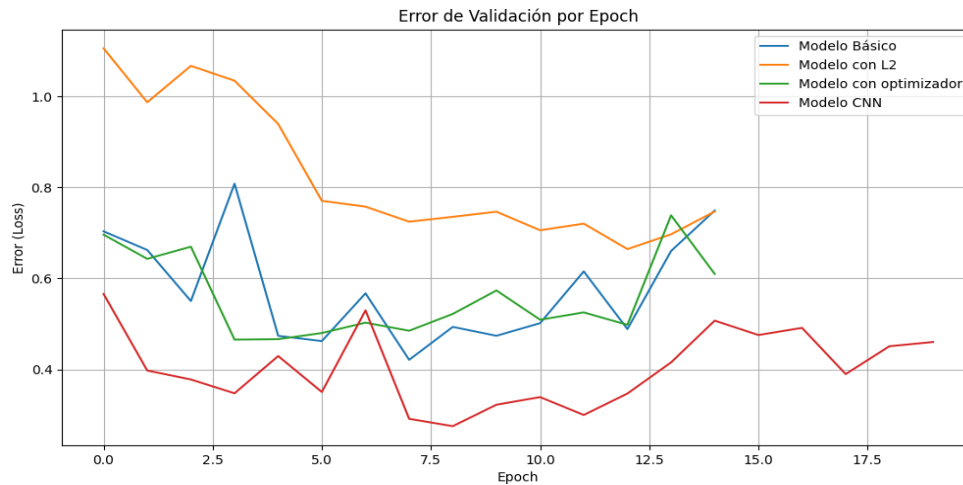
Métricas de rendimiento de modelos.

Modelo	Error (Conjunto Test)	Precisión (Conjunto Test)	Pérdida (Conjunto Validación)	Precisión (Conjunto Validación)
Básico	0.1453	94.77 %	0.7491	82.75 %
L2 regular	0.6256	82.99%	0.7469	77.98 %
Tasa de aprendizaje 0.001	0.1321	95.91 %	0.7385	82.84%
Redes neuronales convolucionales (CNN)	0.0077	99.90%	0.4603	93.13%

En la Figura 2 se muestra la evolución del error de validación por epoch para los diferentes modelos entrenados. Se observa que el Modelo CNN (*curva roja*) presenta consistentemente los errores de validación más bajos, con una tendencia estable después de aproximadamente 5 epochs. En contraste, el Modelo con L2 Regularización (*curva naranja*) tiene los errores más altos a lo largo de todo el entrenamiento, indicando un bajo rendimiento en comparación con los otros modelos. El Modelo Básico (*curva azul*) y el Modelo con tasa de aprendizaje 0.001 (*curva verde*) muestra una tendencia más variable, con errores que tienden a disminuir de forma menos consistente que el modelo CNN. Al igual que lo visto en Tabla 1 estos resultados sugieren que el Modelo CNN es el más eficiente para minimizar el error de validación, mientras que la regularización L2 parece haber perjudicado el rendimiento del modelo.

Figura 1.

Errores de validación por epoch para cada modelo



Los resultados del rendimiento del modelo de redes neuronales convolucionales escogido que se muestran en la Tabla 2 indican que la precisión global del modelo es de 0.26, indicando que el 26% de las predicciones fueron correctas en general. El recall más alto se obtuvo para la clase *No tumor* con un valor de 0.33, mientras que la clase *Meningioma* muestra el peor desempeño con un recall de 0.19. Las puntuaciones F1, que son el promedio armónico entre la precisión y el recall, también reflejan un bajo rendimiento, siendo la más alta de 0.32 para la clase *No tumor*. El promedio macro y el promedio ponderado para precisión, recall y F1 fueron de 0.26, lo que sugiere un rendimiento uniforme pero bajo en todas las clases. Estos resultados indican que el modelo CNN necesita mejoras, particularmente en la clasificación de las categorías tumorales como glioma y meningioma.

Cuadro 2.

Informe de clasificación de modelo CNN

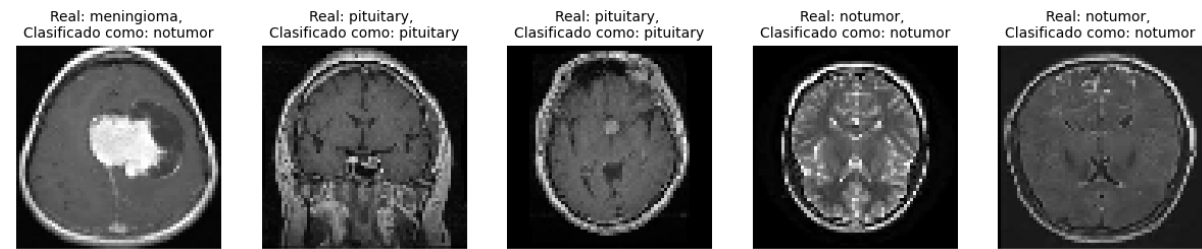
Categoría	Precisión	Recall (Sensibilidad)	Puntuación-F1	Support (Apoyo)
Glioma	0.26	0.27	0.26	300
Meningioma	0.22	0.19	0.2	306
No tumor	0.32	0.33	0.32	405
Pituitaria	0.24	0.25	0.24	300
Exactitud (Accuracy)			0.26	1311
Promedio macro	0.26	0.26	0.26	1311
Promedio ponderado	0.26	0.26	0.26	1311

En la Figura 2, se muestran imágenes de resonancia magnética utilizadas para la clasificación mediante un modelo de redes neuronales convolucionales. En el primer caso, la imagen real corresponde a un meningioma, sin embargo, el modelo lo clasifica incorrectamente como *no tumor*, lo que evidencia una limitación en la precisión del modelo para esta clase específica. Las siguientes dos imágenes corresponden a tumores pituitarios que fueron correctamente clasificados por el modelo, demostrando su efectividad en la detección de este tipo de lesiones. Finalmente, se presentan dos casos de imágenes sin presencia de

tumor *no tumor* que fueron clasificadas correctamente por el modelo, indicando un buen desempeño en la identificación de imágenes normales o sin patologías tumorales. Estos resultados reflejan un rendimiento variable del modelo dependiendo del tipo de tumor, con aciertos más consistentes en la detección de tumores pituitarios y ausencia de tumor, pero con dificultades para identificar meningiomas en algunos casos.

Figura 2.

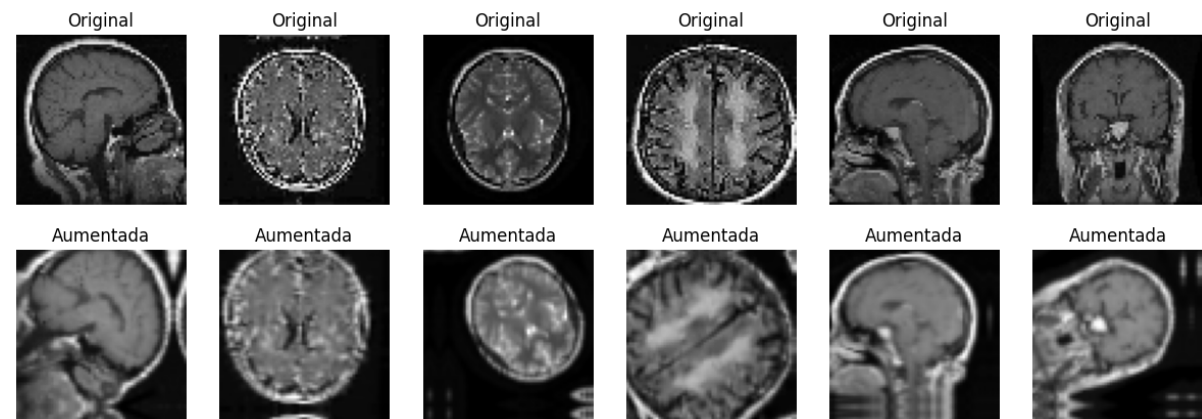
Imágenes para modelo CNN



La Figura 3 muestra las imágenes para la aumentación de datos en el proceso de entrenamiento de los modelos de clasificación. En la primera fila, se muestran las imágenes originales, que corresponden a diferentes vistas y cortes cerebrales. La segunda fila exhibe las versiones aumentadas de estas imágenes, las cuales fueron modificadas mediante técnicas como rotación, escalado o traslación, con el fin de incrementar la diversidad del conjunto de datos y mejorar la generalización del modelo. La comparación entre las imágenes originales y las aumentadas resalta cómo la aumentación permite generar nuevas variaciones que preservan las características clave de las estructuras cerebrales, mientras introducen ligeras distorsiones que ayudan al modelo a aprender de manera más robusta frente a cambios en la orientación o tamaño de las imágenes, lo que es esencial para evitar el sobreajuste y mejorar la precisión del modelo en datos nuevos no vistos durante el entrenamiento.

Figura 3.

Imágenes para aumentación de datos



Se presentan las métricas de rendimiento del modelo entrenado con aumentación de imágenes en la Tabla 3. El modelo alcanza un error de 0.4738 en el conjunto de prueba, con una precisión del 81.93 %, lo que indica un rendimiento moderado en la clasificación de nuevas imágenes. En cuanto al conjunto de validación, se observa una pérdida de 0.9435 y una precisión del 68.26 %, lo que sugiere que el modelo presenta dificultades para generalizar

correctamente, esto puede ser causa de un sobreajuste o a la alta complejidad de las transformaciones aplicadas durante la aumentación.

Cuadro 3.

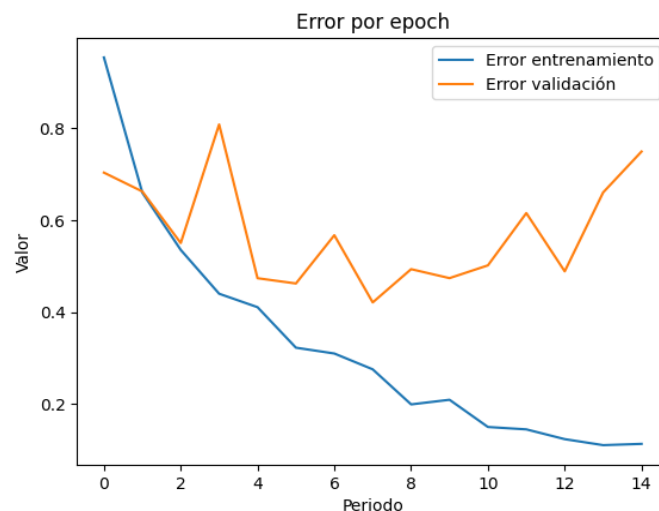
Métricas de rendimiento de modelo con aumentación de imágenes.

Error (Conjunto Test)	Precisión (Conjunto Test)	Pérdida (Conjunto Validación)	Precisión (Conjunto Validación)
0.4738	81.93 %	0.9435	68.26 %

La Figura 4 muestra la evolución del error de entrenamiento y validación por epoch para el modelo con aumentación de datos. El error de entrenamiento (*curva azul*) disminuye de manera constante a lo largo de las épocas, alcanzando un valor bajo y estable alrededor del periodo 10. Sin embargo, el error de validación (*curva naranja*) muestra una tendencia más volátil, con fluctuaciones a lo largo de los epoch, aunque inicialmente disminuye, a partir del periodo 6 comienza a aumentar nuevamente, lo que sugiere un posible sobreajuste del modelo. Estos resultados indican que, aunque la aumentación de datos mejora el rendimiento en el conjunto de entrenamiento, no garantiza una mejora consistente en la generalización del modelo sobre los datos de validación.

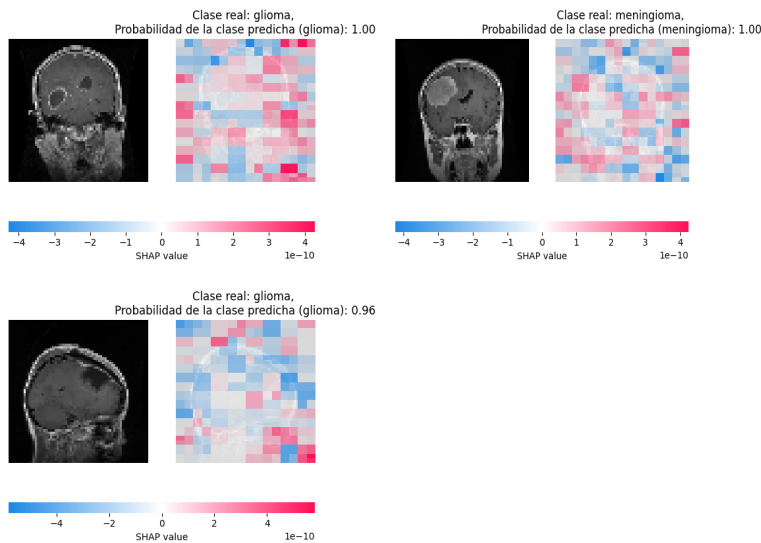
Figura 4.

Error con modelo con aumentación



El análisis SHAP se aplica para interpretar las predicciones de un modelo CNN en imágenes de resonancia magnética de pacientes con glioma y meningioma. En un caso de glioma, el modelo predijo esta clase con una probabilidad del 100%, mostrando alta confianza en la predicción. El mapa de calor SHAP revela que las áreas alrededor de la masa tumoral fueron las más relevantes para la clasificación, destacando las características clave identificadas por el modelo. Para un caso de meningioma, el modelo también muestra una predicción del 100%, y el mapa SHAP refleja una combinación de áreas positivas y negativas, sugiere que el modelo utiliza una variedad de características para realizar una predicción precisa. En otro caso de glioma, el análisis SHAP muestra una probabilidad de predicción del 96%, con áreas rojas que aumentaban la probabilidad de glioma y áreas azules que la reducían, indicando una alta confianza del modelo en su predicción.

Figura 5.
SHAP Glioma, Meningioma



CONCLUSIONES

Tras el análisis de los resultados se llega a la conclusión que el modelo de redes neuronales convolucionales (CNN) fue significativamente superior en términos de precisión, con un 99.90% en el conjunto de prueba y un 93.13% en validación.

Quiles (2023), resalta que el uso de redes neuronales convolucionales (CNN) ha demostrado ser particularmente efectivo en la segmentación y clasificación de tumores cerebrales en imágenes médicas.. Estas redes pueden detectar patrones y características clave en las imágenes de resonancia magnética, lo que facilita una segmentación precisa y una mejor identificación de áreas afectadas, respaldando así la importancia de su aplicación en entornos médicos.

Lo anterior respalda la efectividad de los modelos CNN en tareas de clasificación de imágenes complejas, este hallazgo apoya la creciente adopción de redes neuronales convolucionales en el análisis médico por su capacidad para identificar características clave en imágenes cerebrales.

Sin embargo, a pesar de los altos niveles de precisión obtenidos, se identificaron áreas donde el modelo podría mejorarse. La capacidad del modelo CNN para detectar tumores específicos, como el meningioma, fue limitada, con valores de recall más bajos en comparación con otras categorías. Este resultado refuerza la necesidad de ajustar el modelo, para mejorar la detección de tumores menos representados o con características más sutiles en las imágenes.

En conclusión, aunque el modelo de Red Neuronal Convolucional muestra un rendimiento general superior, los resultados destacan que existen desafíos persistentes en la clasificación precisa de ciertos tipos de tumores. Esto sugiere que futuras investigaciones

deberían centrarse en la optimización de las arquitecturas de red y en la incorporación de técnicas adicionales de regularización y aumentación de datos para mejorar el equilibrio en la clasificación de todas las categorías, estos avances serían esenciales para la aplicación clínica efectiva de este tipo de modelos en el diagnóstico automatizado.

BIBLIOGRAFÍA

- Gómez-Vega, J. C., Ocampo Navia, M. I., & Feo Lee, O. (2019). Epidemiología y caracterización general de los tumores cerebrales primarios en el adulto. *Universitas medica*, 60(1), 47-60.
- Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3), 90-95. <https://doi.org/10.1109/MCSE.2007.55>
- Kouli, O., Hassane, A., Badran, D., Kouli, T., Hossain-Ibrahim, K., & Steele, J. D. (2022). Automated brain tumor identification using magnetic resonance imaging: A systematic review and meta-analysis. *Neuro-Oncology Advances*, 4(1). <https://doi.org/10.1093/noajnl/vdac081>
- McKinney, W. (2010). Data analysis in Python with pandas. *Python for Data Analysis*. O'Reilly Media.
- NumPy Developers. (2020). NumPy. <https://numpy.org/>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830. <http://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf>
- Quiles Ferrández, Y. (2023). Segmentación multimodal de tumores cerebrales mediante redes neuronales convolucionales. *Universidad de Alicante*. https://rua.ua.es/dspace/bitstream/10045/145800/1/Segmentacion_Multimodal_de_Tumores_Cerebrales_mediant_Quiles_Ferrandez_Yeray.pdf
- Roy, S., Nag, S., Maitra, I. K., & Bandyopadhyay, S. K. (2013). A Review on Automated Brain Tumor Detection and Segmentation from MRI of Brain. *arXiv.org*. <https://arxiv.org/abs/1312.6150>
- TensorFlow. (2024). Guía de Keras: Deep learning con Python. <https://www.tensorflow.org/guide/keras?hl=es-419>
- Waskom, M. L., Botvinnik, O., Hobson, P., & Cole, J. B. (2020). Seaborn: statistical data visualization. *Journal of Open Source Software*, 5(49), 2372. <https://doi.org/10.21105/joss.02372>