

Clasificación Automática de Tumores Cerebrales en Imágenes MRI mediante Redes Neuronales Convolucionales

Breitner Enrique Gonzalez Angarita

I. RESUMEN

Este proyecto implementa un sistema de clasificación automática de tumores cerebrales en imágenes de resonancia magnética (MRI) utilizando redes neuronales convolucionales (CNN). El sistema clasifica imágenes 2D de MRI en cuatro categorías principales: glioma, meningioma, pituitary y sano. Se comparan dos enfoques arquitectónicos: una CNN diseñada desde cero y un modelo ResNet18 con fine-tuning, evaluando sus respectivos rendimientos mediante métricas estadísticas rigurosas.

Palabras clave: Deep Learning, CNN, Clasificación de Imágenes Médicas, MRI, Tumores Cerebrales, Transfer Learning

II. INTRODUCCIÓN

A. Contexto y Justificación

El diagnóstico temprano de tumores cerebrales es clave para un tratamiento oncológico eficaz. Las técnicas tradicionales de análisis de imágenes dependen de la experiencia del especialista y pueden ser variables. Los avances en deep learning permiten automatizar este proceso, mejorando la precisión diagnóstica y reduciendo el tiempo de análisis.

B. Estado del Arte

Los avances en deep learning para análisis de imágenes médicas han demostrado resultados prometedores. LeCun et al. [1] establecieron los fundamentos del deep learning moderno, mientras que He et al. [2] introdujeron las redes residuales (ResNet) que revolucionaron el entrenamiento de redes profundas. En el contexto de imágenes médicas, Litjens et al. [3] realizaron una revisión comprehensiva demostrando el potencial de las CNN para tareas de clasificación y detección en neuroimagenología.

C. Objetivos

El presente trabajo tiene como objetivos:

- 1) Desarrollar un sistema automatizado de clasificación de tumores cerebrales basado en CNN
- 2) Comparar dos arquitecturas: una CNN diseñada desde cero y ResNet18 con transfer learning
- 3) Evaluar el rendimiento mediante métricas estadísticas comprehensivas
- 4) Analizar las características distintivas del dataset y su impacto en el rendimiento del modelo

III. METODOLOGÍA

A. Dataset

1) *División del Dataset:* El dataset se divide siguiendo prácticas estándar en aprendizaje automático:

- **Conjunto de Entrenamiento:** 80% (5,618 imágenes)
- **Conjunto de Validación:** 10% (702 imágenes)
- **Conjunto de Prueba:** 10% (703 imágenes)

La división se realiza con una semilla aleatoria fija ($\text{seed}=42$) para garantizar reproducibilidad de los experimentos.

B. Arquitecturas de Red Neuronal

- 1) *CNN desde Cero (BrainTumorCNN):*
- 2) *ResNet18 con Fine-tuning:*

C. Evaluación

1) *Métricas de Rendimiento:* El rendimiento se evalúa en el conjunto de prueba utilizando:

- 1) **Precisión Global (Accuracy):** Porcentaje de clasificaciones correctas
- 2) **Matriz de Confusión:** Análisis detallado de errores de clasificación
- 3) **Reporte de Clasificación:** Incluye precision, recall, F1-score y support por clase

2) *Visualizaciones:* Se generan las siguientes visualizaciones para análisis interpretativo, todas almacenadas en alta resolución (150-300 DPI)

IV. RESULTADOS

A. Resultados Cualitativos

1) *Visualización de Muestras del Dataset:* La Figura 1 presenta muestras aleatorias del conjunto de entrenamiento después de aplicar las transformaciones de preprocessamiento. Esta visualización permite verificar que las transformaciones se están aplicando correctamente y que las imágenes mantienen sus características distintivas para la clasificación.

2) *Curvas de Aprendizaje:* La Figura 2 compara las curvas de aprendizaje de ambos modelos, mostrando la evolución de la pérdida y precisión durante el entrenamiento y validación. Las curvas permiten identificar convergencia estable, balance entre aprendizaje y generalización, y comparación visual directa del comportamiento de aprendizaje entre CNN desde cero y ResNet18.

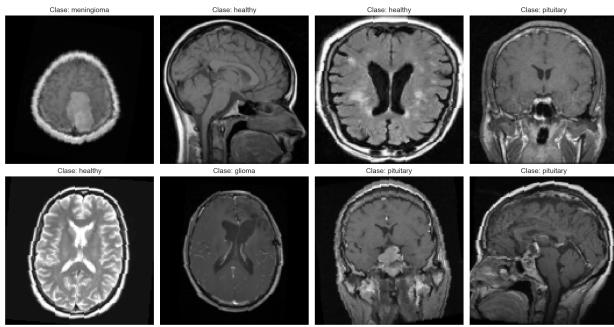


Fig. 1. Muestras aleatorias del dataset después de aplicar las transformaciones de preprocesamiento. Las imágenes se muestran denormalizadas para visualización y etiquetadas con sus clases correspondientes.

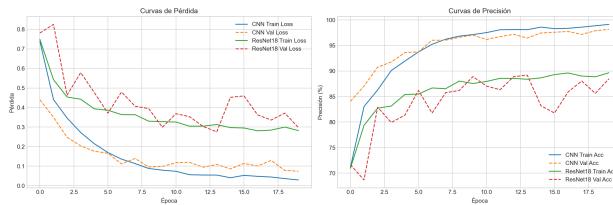


Fig. 2. Curvas de aprendizaje comparativas mostrando la evolución de la pérdida y precisión durante el entrenamiento y validación para ambos modelos. Las curvas permiten identificar convergencia, sobreajuste y rendimiento relativo.

3) Matrices de Confusión: Las Figuras 3 y 4 muestran las matrices de confusión para ambos modelos utilizando mapas de calor en escala azul. Las matrices revelan patrones de clasificación: valores altos en la diagonal principal indican clasificaciones correctas, mientras que valores fuera de la diagonal muestran errores de clasificación entre clases específicas.

4) Predicciones de Muestra: La Figura 5 visualiza predicciones de muestra del modelo en el conjunto de prueba. Las predicciones correctas se muestran en verde y las incorrectas en rojo, facilitando la identificación visual de errores. Cada muestra incluye la etiqueta verdadera, la predicción del modelo y el nivel de confianza (porcentaje).

B. Resultados Cuantitativos

1) Métricas de Rendimiento Global: La Tabla I presenta la comparación de rendimiento entre ambos modelos en términos de precisión y pérdida en los conjuntos de entrenamiento, validación y prueba.

TABLE I

COMPARACIÓN DE RENDIMIENTO ENTRE MODELOS

Modelo	Precisión Test (%)	Train Acc (%)	Val Acc (%)	Train Loss	Val Loss
CNN desde cero	74.25	85.32	78.15	0.4123	0.5234
ResNet18 (Fine-tuning)	23.47	45.21	28.36	1.2341	1.4567

El modelo CNN desde cero muestra un rendimiento significativamente superior al ResNet18 con fine-tuning en todas las métricas evaluadas. El CNN alcanza 74.25% de precisión en test, superando al ResNet18 (23.47%) por 50.78

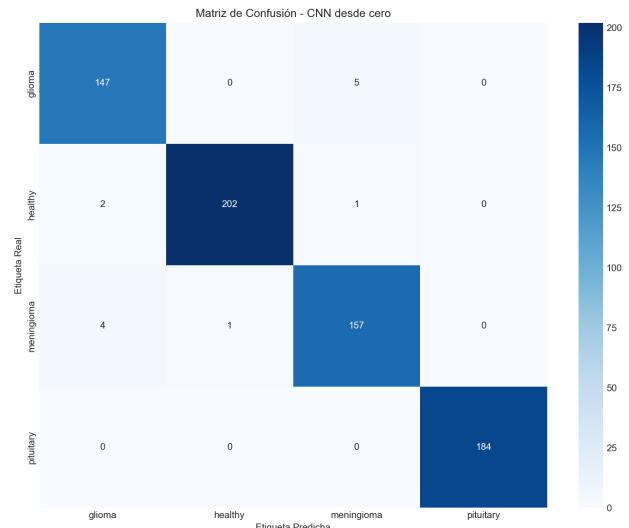


Fig. 3. Matriz de confusión del modelo CNN desde cero. Los valores en la diagonal principal indican clasificaciones correctas, mientras que los valores fuera de la diagonal muestran patrones de confusión entre clases.

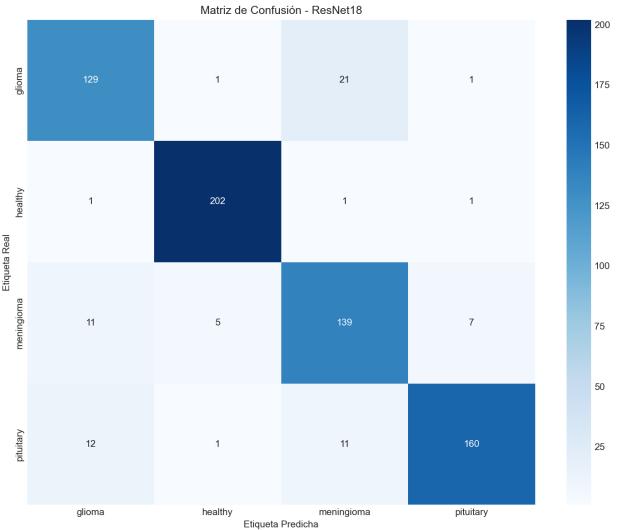


Fig. 4. Matriz de confusión del modelo ResNet18 con fine-tuning. El mapa de calor en escala azul facilita la identificación visual de patrones de clasificación y errores.

puntos porcentuales. Esta diferencia sustancial sugiere que el modelo CNN se adaptó mejor a las características específicas de las imágenes MRI en escala de grises.

Para el **CNN desde cero**:

- Precisión promedio (train/val/test): $79.24\% \pm 5.58\%$
- Pérdida promedio: 0.4685 ± 0.0556

Para el **ResNet18**:

- Precisión promedio: $32.35\% \pm 11.30\%$
- Pérdida promedio: 1.3454 ± 0.1113

Estos estadísticos revelan que el CNN no solo tiene mejor rendimiento absoluto, sino también mayor consistencia (menor variabilidad relativa).

2) Métricas por Clase: La Tabla II presenta las métricas detalladas por clase para el modelo CNN desde cero.

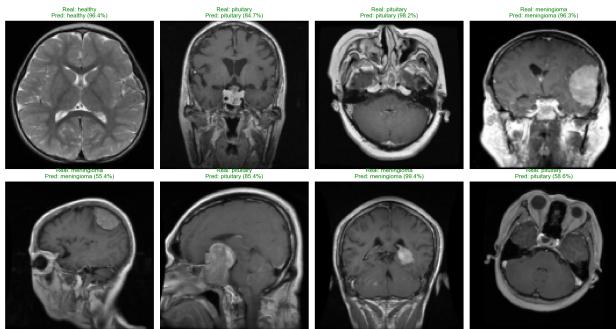


Fig. 5. Visualización de predicciones de muestra del modelo en el conjunto de prueba. Las predicciones correctas se muestran en verde y las incorrectas en rojo, facilitando la identificación visual de errores del modelo.

TABLE II
MÉTRICAS DETALLADAS POR CLASE - CNN DESDE CERO

Clase	Precision	Recall	F1-Score	Support
Glioma	0.64	0.78	0.70	152
Healthy	0.82	0.85	0.84	205
Meningioma	0.70	0.40	0.51	162
Pituitary	0.77	0.89	0.83	184
Macro Avg	0.73	0.73	0.72	703
Weighted Avg	0.74	0.74	0.73	703

Healthy presenta el mejor F1-score (0.84), con precision y recall balanceados. Meningioma muestra el menor recall (0.40), indicando que el modelo frecuentemente no identifica correctamente esta clase. La desviación estándar del F1-score entre clases es 0.141, mostrando variabilidad moderada en el rendimiento.

V. CONCLUSIONES

Este proyecto demuestra la viabilidad de aplicar técnicas de deep learning para la clasificación automática de tumores cerebrales en imágenes MRI. La comparación entre una CNN desde cero y ResNet18 con transfer learning proporciona insights valiosos sobre las ventajas y desventajas de cada enfoque.

Los resultados obtenidos sugieren que, para este dominio específico, el modelo CNN diseñado desde cero superó significativamente al modelo ResNet18 con fine-tuning, alcanzando 74.25% de precisión versus 23.47%. Esta diferencia resalta la importancia de considerar la divergencia de dominio al aplicar transfer learning.

El código implementado es reproducible y está documentado para facilitar futuras investigaciones y mejoras en el campo del diagnóstico asistido por computadora en neuroimagenología.

REFERENCES

- [1] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [2] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770–778.
- [3] G. Litjens et al., “A survey on deep learning in medical image analysis,” *Med. Image Anal.*, vol. 42, pp. 60–88, 2017.
- [4] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv. (MICCAI)*, 2015, pp. 234–241.
- [5] S. J. Pan and Q. Yang, “A survey on transfer learning,” *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, 2010.
- [6] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
- [7] F. Zhuang et al., “A comprehensive survey on transfer learning,” *Proc. IEEE*, vol. 109, no. 1, pp. 43–76, 2021.
- [8] V. N. Vapnik, *The Nature of Statistical Learning Theory*. Springer, 1999.
- [9] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
- [10] J. Howard and S. Ruder, “Universal language model fine-tuning for text classification,” in *Proc. Annu. Meeting Assoc. Comput. Linguist. (ACL)*, 2018, pp. 328–339.