Universidad del Valle de Guatemala
Facultad de Ingeniería
Data Science
Departamento de Ciencias de la Computación
Ciclo II, 2024
Fecha de entrega: 28 oct 2024

Proyecto II

# Resultados Parciales y Visualizaciones Estáticas

| | |
|---|---|
| Adrián Ricardo Flores Trujillo | Carné 21500 |
| Daniel Armando Valdez Reyes | Carné 21240 |
| Emilio José Solano Orozco | Carné 21212 |
| Andrea Ximena Ramírez Recinos | Carné 21874 |

Guatemala, 28 de octubre de 2024

# 1. Introducción

Los accidentes cerebrovasculares isquémicos representan una de las principales causas de mortalidad a nivel mundial. Para reducir el riesgo de recurrencia de estos eventos y mejorar el diagnóstico de los pacientes, es esencial identificar con precisión la etiología de los coágulos sanguíneos. Este proyecto se enfoca en la diferenciación entre dos subtipos de coágulos: la embolia cardioembólica (CE) y la aterosclerosis de grandes arterias (LAA), empleando algoritmos de Deep Learning y Machine Learning para mitigar el riesgo de un pronóstico equivocado.

El conjunto de datos utilizado contiene más de 700 observaciones obtenidas del Registro de Trombosis por Embolismo de Imágenes y Patología (STRIP) de la Clínica Mayo, disponible en la página oficial de Kaggle. En esta segunda fase, se ha realizado una revisión exhaustiva de los algoritmos más efectivos para abordar esta problemática. La investigación, basada en trabajos similares, ha identificado enfoques prometedores, especialmente en el uso de Redes Neuronales Convolucionales (CNN) y sus variantes.

Un estudio reciente de Azatyan (2023) resalta que el uso de CNN, en particular con la arquitectura Efficient Net, es uno de los métodos más efectivos para la clasificación de imágenes de coágulos. Este modelo ha demostrado un rendimiento superior en tareas de clasificación médica. Otro trabajo relevante de Krishnan et al. (2023) también confirma la eficacia de Efficient Net para clasificar coágulos, con resultados favorables que consolidan su viabilidad para este tipo de aplicación. Finalmente, la arquitectura SWIN y los Transformers Visuales (ViT) han mostrado gran potencial en visión por computadora, destacándose como alternativas valiosas para este proyecto.

En base a todo lo anteriormente expuesto, el proyecto se orienta a implementar y evaluar estas arquitecturas para lograr una diferenciación precisa entre CE y LAA, optimizando así la atención médica y el tratamiento oportuno para pacientes con riesgo de accidentes cerebrovasculares.

# 2. Problemática a abordar

## 2.1. Situación Problemática

El accidente cerebrovascular isquémico, como se mencionó previamente, es una de las principales causas de muerte a nivel mundial. Según la Organización Mundial de Accidentes Cerebrovasculares (WSO por sus siglas en inglés), más de 7,6 millones de personas sufren estos episodios anualmente, y un porcentaje significativo experimenta un segundo evento, lo que agrava las probabilidades de supervivencia. Determinar el origen del coágulo que causa estos episodios es crucial para el manejo terapéutico apropiado, pero los métodos actuales son limitados. Con el avance de la trombectomía mecánica, ahora es posible analizar los coágulos recuperados, pero los formatos de datos y el tamaño de las imágenes patológicas hacen que el análisis sea un desafío.

(World Stroke Organization, 2022)

## 2.2 Problema Científico

Con base a lo anteriormente descrito, el problema científico que se plantea es la correcta y precisa identificación de la etiología de los accidentes cerebrovasculares para reducir el riesgo de episodios recurrentes. Se pretende abordar esto por medio de un modelo de aprendizaje profundo, como por ejemplo una red convolucional, aunque esto supone

enfrentar dificultades técnicas relacionadas con el análisis de grandes conjuntos de datos y la naturaleza compleja de las imágenes. Todo esto, con el propósito de generar una herramienta que mejore las decisiones clínicas y reduzca con ello, la probabilidad de más accidentes de esta índole.

## 3. Objetivos

### 3.1 Objetivo General

- Desarrollar un modelo predictivo capaz de clasificar entre dos tipos de orígen de coágulos sanguíneos en accidentes cerebrovasculares isquémicos, coágulos cardioembólicos (CE) y ateroscleróticos de grandes arterias (LAA).

### 3.2 Objetivos específicos

- Diseñar un preprocesado ideal para las imágenes, con dinámicas de reescalado y aplicación de filtros.

- Evaluar la efectividad entre variantes del modelo desarrollado mediante el ajuste de sus parámetros.

- Comparar el rendimiento entre modelos.

- Seleccionar el modelo más óptimo para la implementación de la problemática planteada.

## 4. Investigación de Enfoques en Contextos Similares

### 4.1. Advancing Ischemic Stroke Diagnosis: A Novel Two-Stage Approach for Blood Clot Origin Identification.

Este trabajo al igual que el presente, se centró en el análisis y clasificación de coágulos sanguíneos responsables de accidentes cerebrovasculares isquémicos. Se aplicaron redes neuronales convolucionales (CNN) y modelos avanzados como Visual Transformers para clasificar imágenes digitales de coágulos. Estos modelos permitieron preprocesar las imágenes patológicas, aumentar los datos mediante transformaciones y utilizar técnicas de aprendizaje profundo para mejorar la precisión en la identificación de la etiología de los coágulos. Los mejores resultados se obtuvieron con arquitecturas de vanguardia, como SwinTransformerV2 y PoolFormer, alcanzando valores de precisión superiores al 90%.

### 4.2. Image Classification of Stroke Blood Clot Origin using Deep Convolutional Neural Networks and Visual Transformers

En este trabajo se realizó un estudio sobre el mismo dataset que se utilizará en este proyecto. Utilizando cuatro arquitecturas avanzadas de redes neuronales convolucionales (CNN) y un método de ensamble simple, se analizaron imágenes patológicas de coágulos. Los resultados mostraron que la combinación de arquitecturas avanzadas, incluyendo ResNet y Visual Transformers, permitió una clasificación precisa entre los coágulos CE y LAA, lo que representa un avance importante en el diagnóstico de accidentes cerebrovasculares. Los modelos alcanzaron altos niveles de exactitud, lo cual sugiere que estas tecnologías podrían ser aplicadas en entornos clínicos para mejorar la identificación de la etiología de los coágulos y optimizar las decisiones terapéuticas. Además, el trabajo destaca la capacidad de estos

enfoques para adaptarse a grandes conjuntos de datos y resolver desafíos relacionados con la variabilidad de las imágenes.

## 5. Investigación de Algoritmos a Utilizar

### 5.1. Redes Neuronales Convolucionales (CNNs):

En este trabajo, se propone el uso de redes neuronales profundas, específicamente redes neuronales convolucionales (CNN), para abordar la clasificación de imágenes médicas, una técnica que ha demostrado ser efectiva en estudios previos. Azatyan (2023) explora cómo las CNN pueden emplearse para diferenciar entre dos subtipos principales de coágulos en pacientes con accidentes cerebrovasculares isquémicos. Este enfoque se basa en arquitecturas ampliamente utilizadas, como ResNet y EfficientNet, las cuales implementan el aprendizaje transferido para acelerar el entrenamiento de los modelos en este contexto y mejorar su precisión.

### 5.2. Transformadores Visuales (ViTs):

Se propone la utilización de redes neuronales convolucionales (CNN) en combinación con transformadores visuales (ViTs) para la clasificación de imágenes de coágulos sanguíneos. Este enfoque, estudiado por Azatyan (2023), muestra que la integración de ambas arquitecturas permite capturar tanto características locales como dependencias globales en las imágenes, logrando así una mayor precisión en tareas complejas de clasificación. Además, existen variantes como el Swin Transformer, que se destaca por su capacidad de segmentar las imágenes en ventanas de atención desplazables. Esto permite captar patrones con gran precisión en diversas escalas, mejorando el análisis de imágenes y patrones complejos.

## 6. Selección de algoritmos a probar

Para el presente proyecto, los modelos basados en CNN resultan viables por su comprobada capacidad en la clasificación de imágenes médicas, especialmente en escenarios donde los detalles sutiles en la estructura de los tejidos son críticos. Arquitecturas como EfficientNet ofrecen un equilibrio óptimo entre precisión y eficiencia computacional, escalando de manera controlada el tamaño de la red y el número de parámetros, lo cual resulta beneficioso al trabajar con grandes volúmenes de datos de imágenes patológicas. Asimismo, los transformadores visuales (Visual Transformers) aportan una perspectiva novedosa en tareas de clasificación de imágenes, al modelar dependencias a largo plazo dentro de las imágenes; esto es especialmente útil cuando las características relevantes están distribuidas a lo largo de la imagen, como suele suceder en imágenes médicas.

Para conjuntos de datos similares a los de este proyecto, EfficientNet destaca por su eficiencia y capacidad para extraer características discriminativas con poca pérdida de información relevante, demostrando aplicaciones exitosas en contextos similares. Por su parte, el uso de transformadores visuales (ViTs) es prometedor, ya que pueden modelar dependencias de largo alcance dentro de las imágenes y mejorar la clasificación en situaciones donde las características están distribuidas de manera heterogénea. La combinación de estas técnicas permitirá abordar de forma robusta y eficiente la complejidad de las imágenes médicas en nuestro proyecto.

Estas arquitecturas nos permitirán explorar diferentes enfoques para la clasificación, seleccionando el más adecuado según el tipo de imágenes y la naturaleza del problema, optimizando tanto la precisión como el tiempo de procesamiento. Dada la escalabilidad de los modelos mencionados, es posible ajustarlos en función de los recursos disponibles y las

características particulares de nuestro conjunto de datos, maximizando su aplicabilidad clínica.

## 7. Preprocesamiento de Imágenes

### 7.1. Primer Enfoque de Procesamiento

El primer paso en el preprocesamiento del conjunto de datos consistió en seleccionar una ventana de tamaño 5000x5000 en todas las imágenes, comenzando desde el punto (1000,1000), con el objetivo de reducir la dimensionalidad de cada entrada. Esta decisión se basó en un análisis exploratorio previo, donde se identificó que muchas imágenes contenían amplios espacios vacíos. Posteriormente, se redimensiona las imágenes a un tamaño de 512x512 para el modelo EfficientNet y a 384x384 para el transformador Swin. Finalmente, se estandarizó el formato de color a RGB y se normalizaron los valores de píxeles en cada imagen.

Además de trabajar con las imágenes existentes, para abordar el desbalance en el conjunto de datos original se aplicó aumentación de datos únicamente a las imágenes correspondientes a coágulos provenientes de grandes arterias. El proceso de aumentación incluyó rotaciones aleatorias, así como modificaciones aleatorias de brillo, contraste y saturación, con el objetivo de duplicar la cantidad de muestras en las clases con menor cantidad de datos.

### 7.2. Segundo Enfoque de Procesamiento

Inicialmente, debido al gran tamaño y alta resolución de las imágenes de portaobjetos completos (Whole-Slide Images), se extrajeron parches más pequeños de 512x512 píxeles. Esta fragmentación permite un manejo más eficiente de los datos y enfoca el análisis en regiones más manejables y relevantes de las imágenes. Luego, para asegurar que los parches extraídos contenían realmente información relevante y no fueran solo áreas de fondo. Se evaluó la desviación estándar en escala de grises para determinar la variabilidad dentro del parche; aquellos con desviaciones estándar bajas fueron considerados como fondo y se descartaron.

Además, se analizaron las diferencias de media entre los canales de color (rojo, verde, azul) para identificar y eliminar parches sin suficiente contraste o diversidad cromática, de modo que se garantizara que solo se incluyeran en el conjunto de datos los parches con información relevante para el modelo.

Adicionalmente, para abordar el desequilibrio en la distribución de las clases(siendo una clase significativamente menos representada que la otra), se aplicaron técnicas de aumento de datos específicas para cada clase. La clase minoritaria recibió mayor cantidad de aumentaciones agresivas, como volteos horizontales y verticales, rotaciones aleatorias, ajustes de brillo, contraste y saturación, distorsiones de perspectiva y la introducción de dropout en regiones de la imagen. Estas transformaciones aumentaron tanto la cantidad como la diversidad de los datos de la clase minoritaria. Por otro lado, la clase mayoritaria también fue sometida a aumentaciones similares pero de forma más moderada, para evitar el sobreajuste del modelo. De esta forma se obtuvo un conjunto de datos más homogéneo y diverso.

# 8. Análisis de Desempeño de Modelos (Visualizaciones Estáticas)

## 8.1. Pérdida de entrenamiento y validación (Folds)

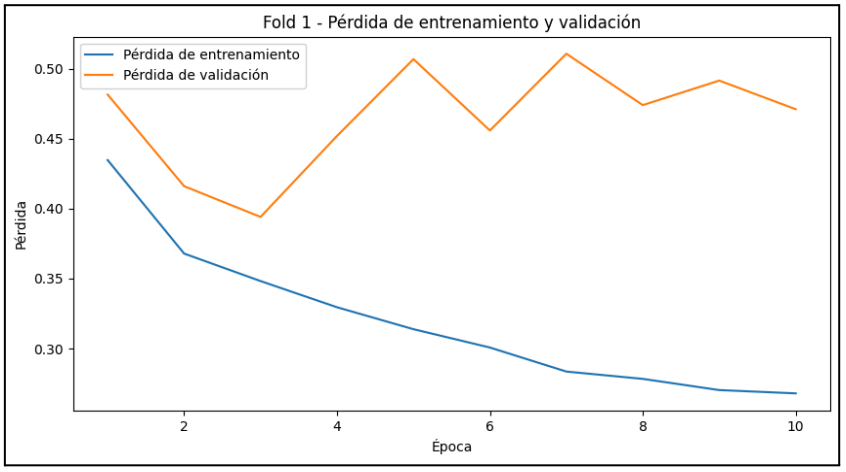


*Figura 1.* Pérdida de entrenamiento y validación (Fold 1, versión final).

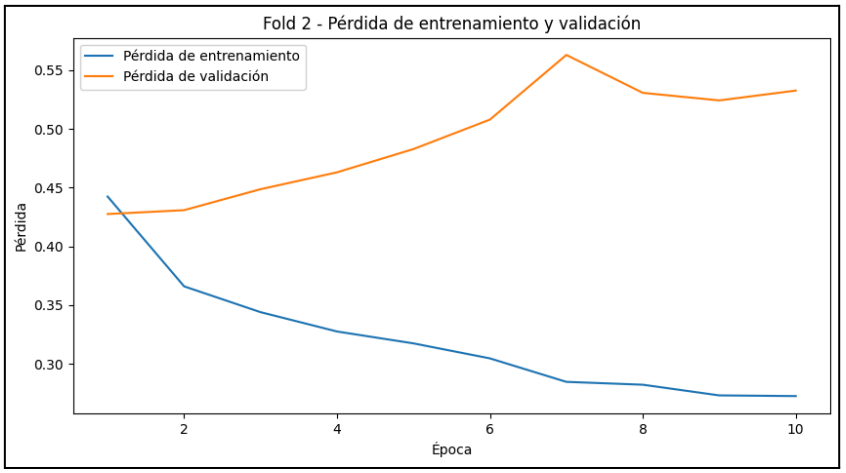


*Figura 2.* Pérdida de entrenamiento y validación (Fold 2, versión final).

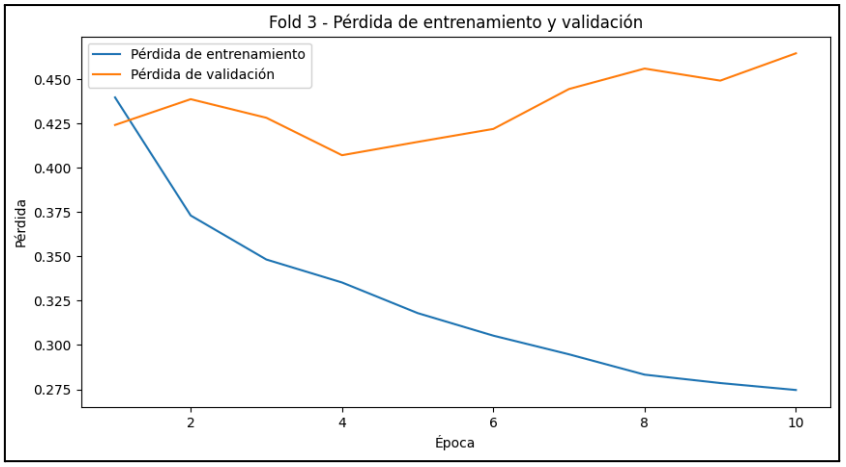


*Figura 3.* Pérdida de entrenamiento y validación (Fold 3, versión final).

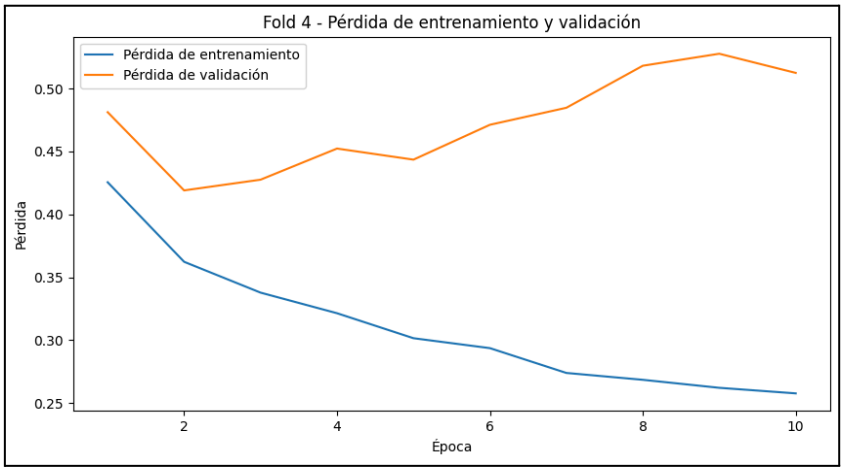


*Figura 4.* Pérdida de entrenamiento y validación (Fold 4, versión final).

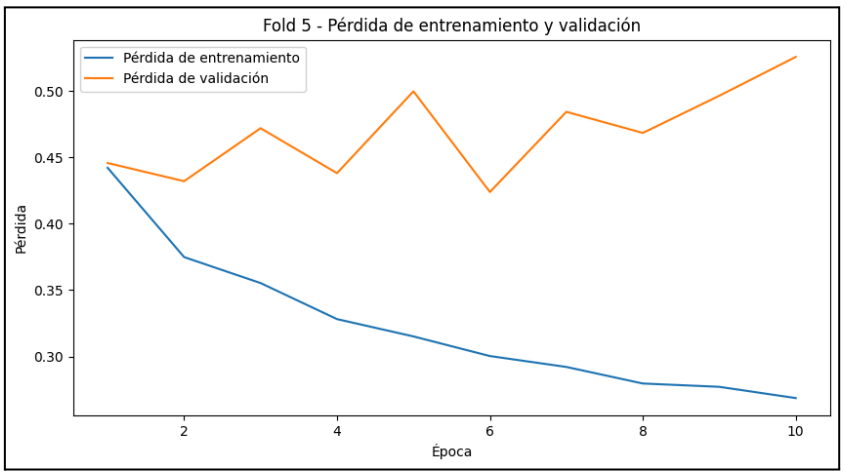


*Figura 5.* Pérdida de entrenamiento y validación (Fold 5, versión final).

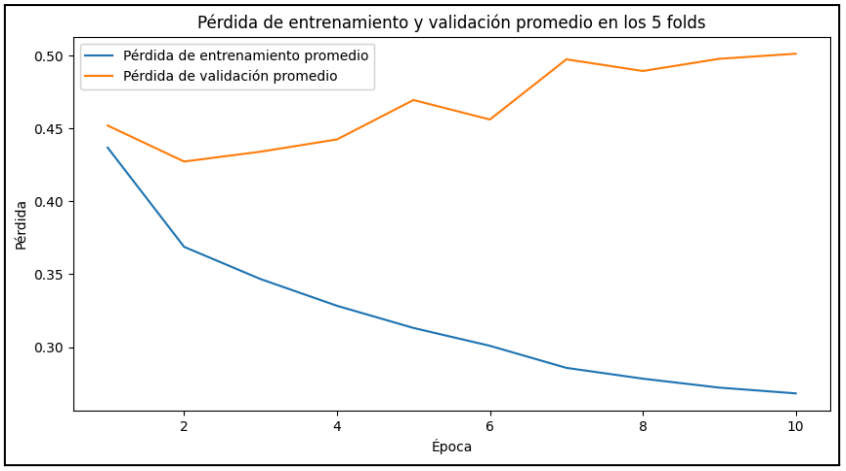


*Figura 6.* Pérdida de entrenamiento y validación (Todos los folds, versión final)

## 8.2. EfficientNetB1 Sin Aumentación de Datos

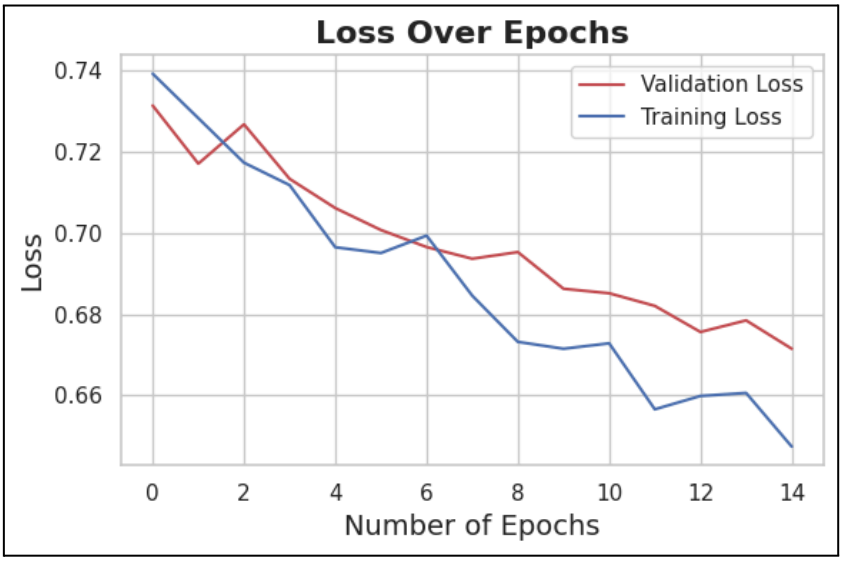


*Figura 7.* Pérdida en entrenamiento del modelo EfficientNetB1 sin aumento de datos.

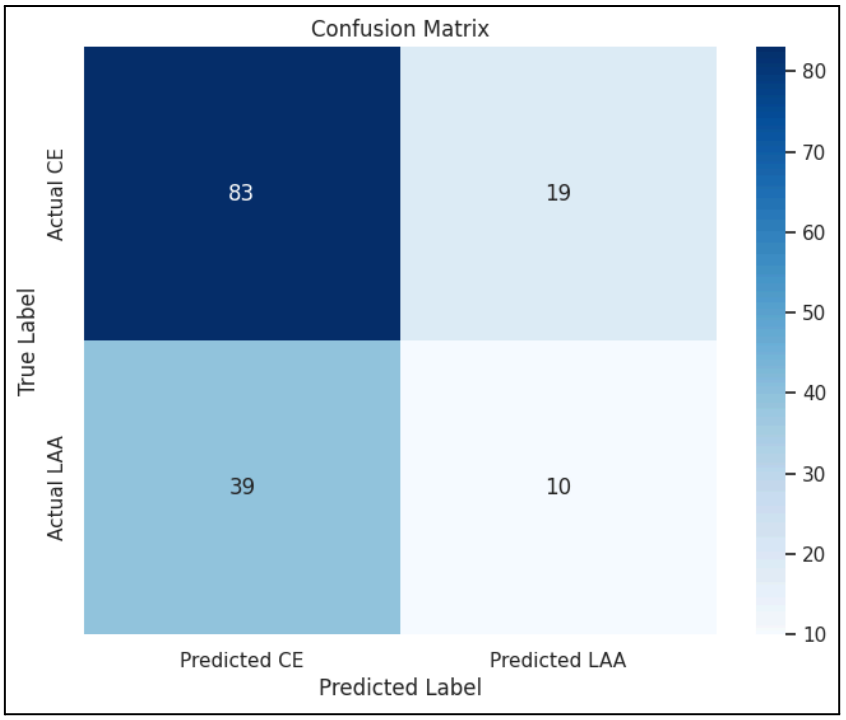


*Figura 8.* Matriz de confusión del modelo EfficientNetB1 sin aumento de datos.

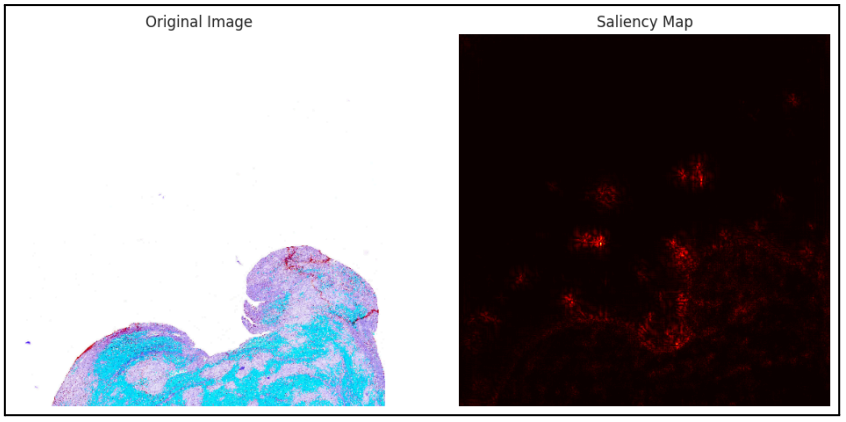


*Figura 9.* Mapa de saliencia de una muestra con EfficientNetB1 sin aumento de datos.

## 8.3. EfficientNetB1 con Aumentación de Datos

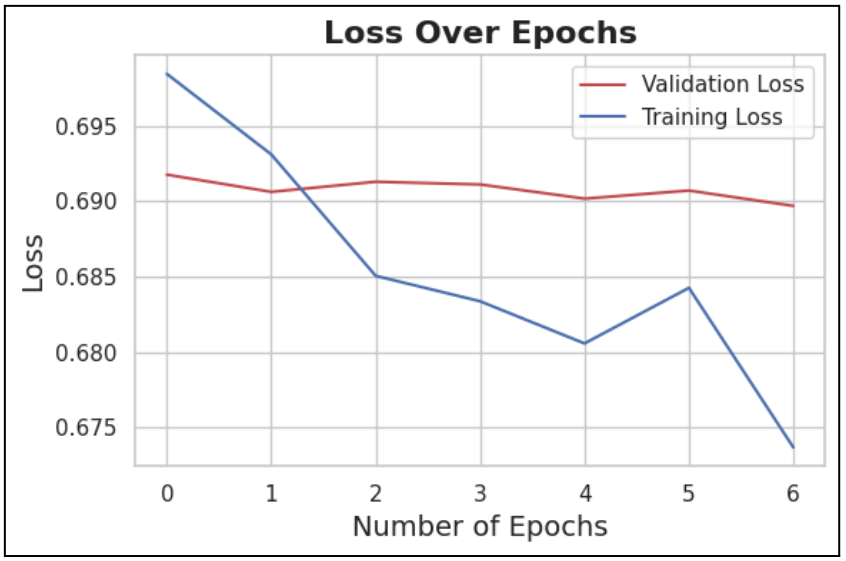


*Figura 10.* Pérdida en entrenamiento del modelo EfficientNetB1 con aumentación de datos.

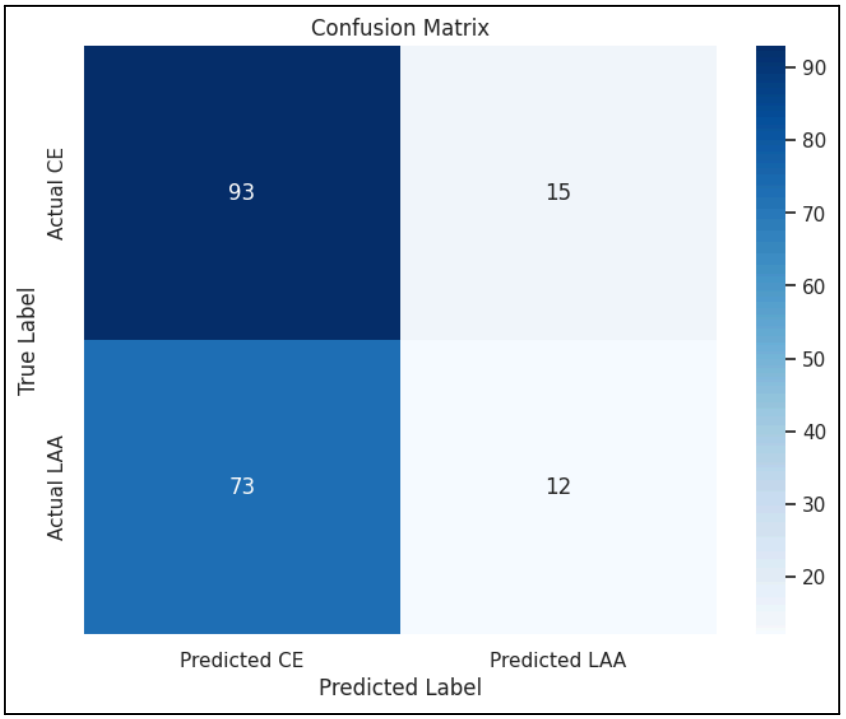


*Figura 11.* Matriz de confusión del modelo EfficientNetB1 con aumentación de datos.

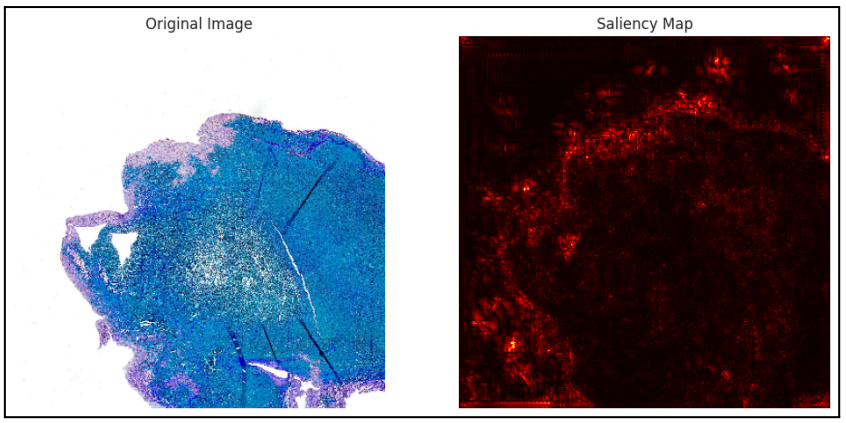


*Figura 12.* Mapa de saliencia de una muestra con EfficientNetB1 con aumento de datos.

## 8.4. Transformador Swin (328 x 328) con Aumentación de Datos

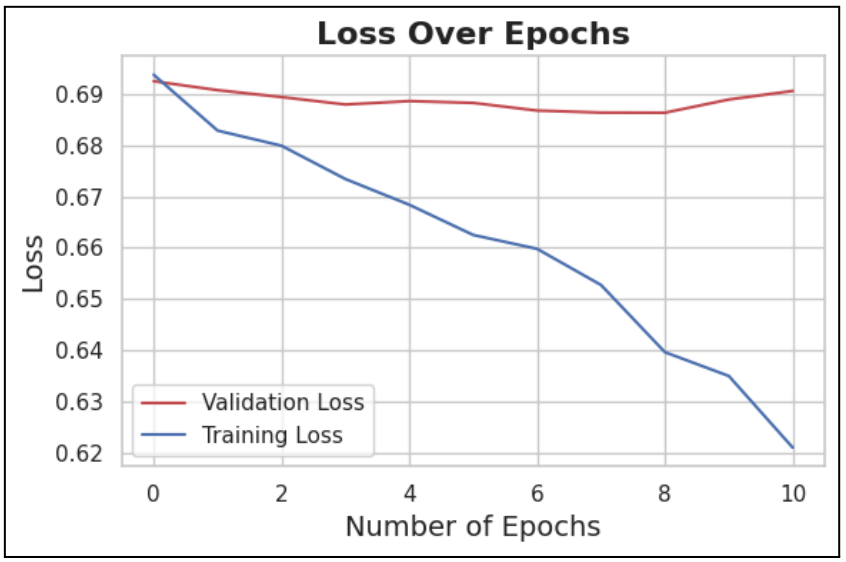


*Figura 13.* Pérdida en entrenamiento del transformador Swin con aumentación de datos.

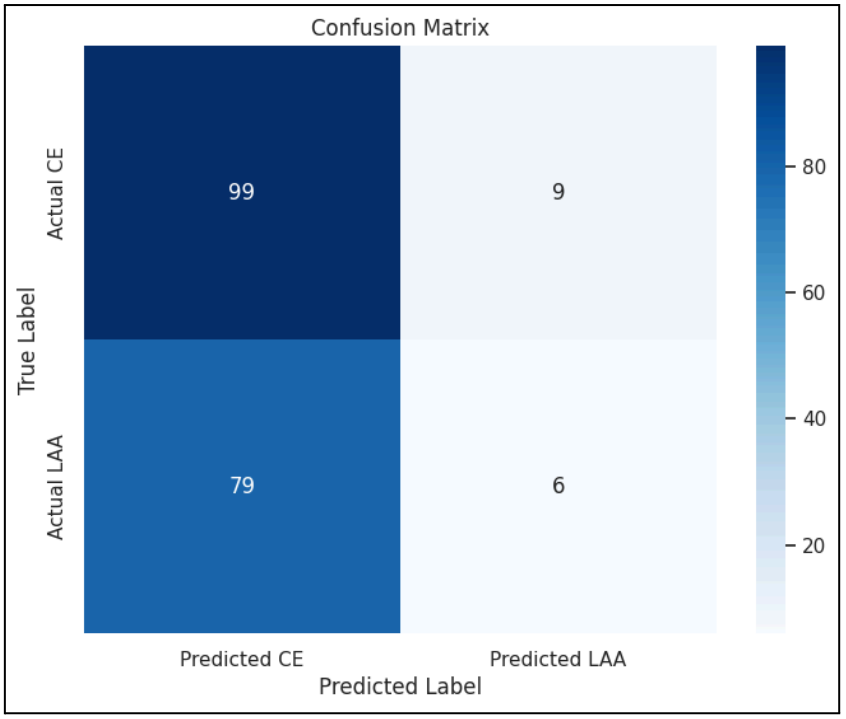


*Figura 14.* Matriz de confusión del transformador Swin con aumentación de datos.

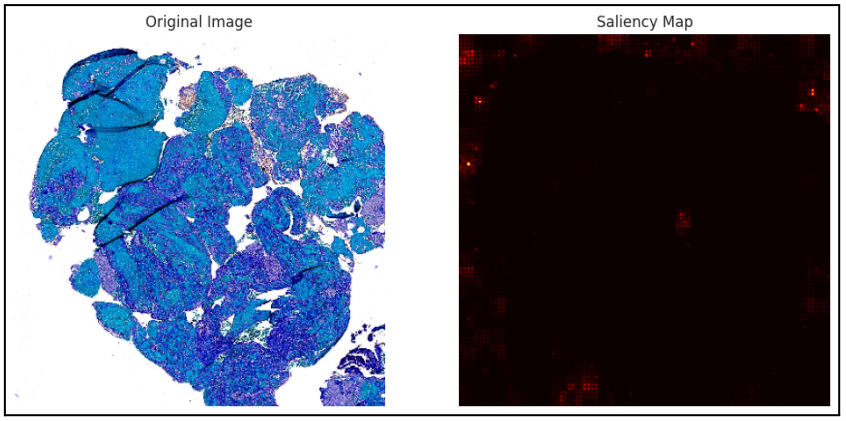


*Figura 15.* Mapa de saliencia de una muestra con transformador Swin.

## 8.5. Mejor modelo Segundo Enfoque (EfficientNet Versión 1)

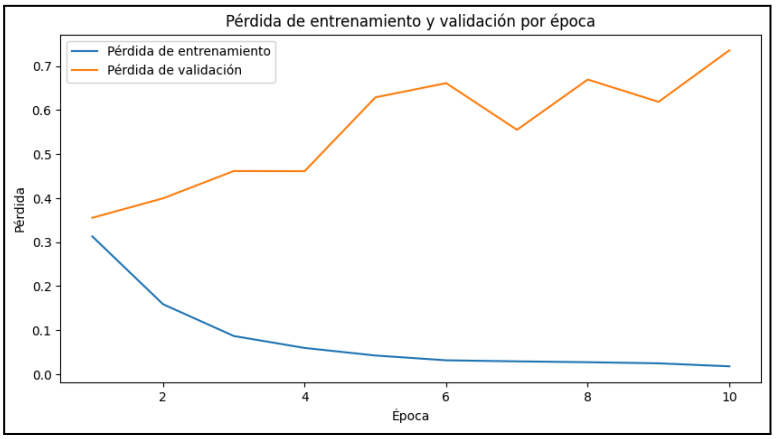


*Figura 16.* Pérdida en entrenamiento del modelo EfficientNet Versión 1

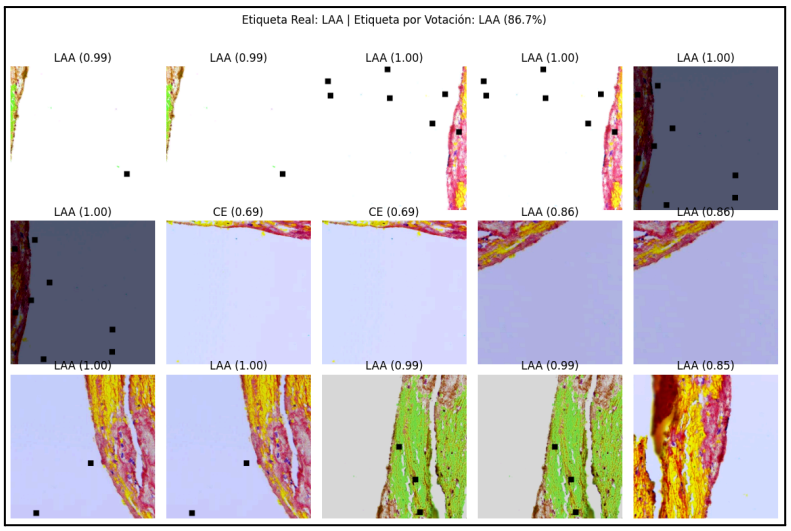


*Figura 17.* Clasificación por votación hecha por el modelo EfficientNet Versión 1

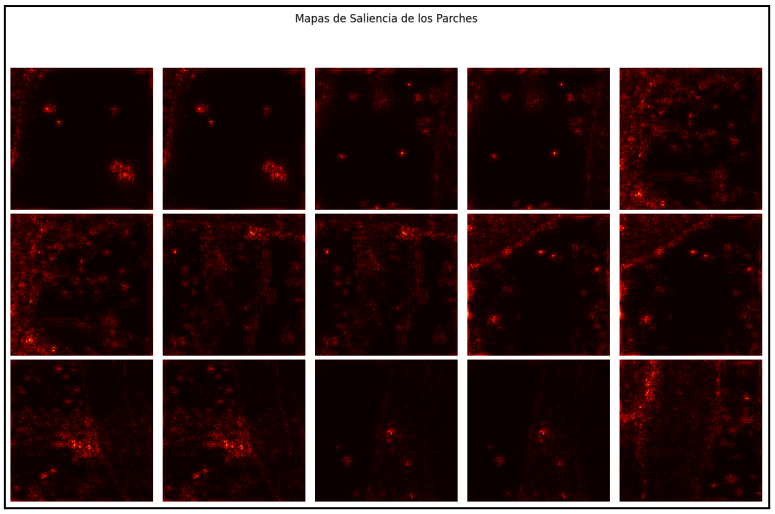


*Figura 18.* Mapas de saliencia para los parches de la figura 17

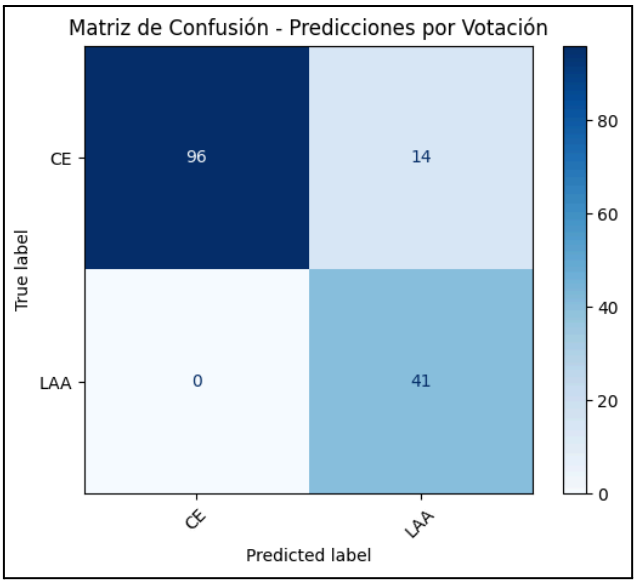


*Figura 19.* Matriz de confusión del modelo EfficientNet Versión 1

## 9. Discusión de Resultados

Durante la primera prueba, para el desarrollo del modelo de clasificación se diseñaron los modelos planteados se decidió hacer un preprocesamiento de la data algo más básico, para analizar el rendimiento de los modelos de manera preliminar. Los modelos seleccionados para esta fase incluyen una arquitectura EfficientNet B1, adaptada para clasificación binaria, y un modelo basado en Vision Transformers (ViTs) para comparar el desempeño. La implementación inicial de EfficientNet B1 incluye ajustes en la tasa de aprendizaje, épocas, tamaño de lote y regularización mediante weight_decay, buscando optimizar el aprendizaje y evitar sobreajuste. En general en base a resultados le fue bastante bien para ser un modelo inicial y nos permitió comprender el enfoque que debíamos seguir. Sin embargo, denota un problema importante que es el sesgo hacia una de las dos clases de coágulos que queríamos evaluar.

Luego se realizó un modelo basado en ViTs, hay que aclarar que éstos han demostrado una mayor capacidad en el análisis de dependencias globales dentro de las imágenes. Este enfoque, tenía potencial para la identificación de los coágulos, ya que es esencial el considerar tanto detalles finos como la relación espacial en áreas amplias de la imagen. Sin embargo no le fue tan bien como al primer modelo y presentó mucho mayor sesgo que el anterior.

Luego se realizó un modelo basado en ViTs, hay que aclarar que éstos han demostrado una mayor capacidad en el análisis de dependencias globales dentro de las imágenes. Este enfoque, tenía potencial para la identificación de los coágulos, ya que es esencial el considerar tanto detalles finos como la relación espacial en áreas amplias de la imagen. Sin embargo no le fue tan bien como al primer modelo y presentó mucho mayor sesgo que el anterior.

En base a esto, se denota un claro problema, y es que los modelos anteriores no son capaces de generalizar bien los rasgos de los coágulos y presentan un claro sesgo que disminuye su precisión. Por ende en este punto se buscó realizar un nuevo procesamiento de para las imágenes que permitiese a los modelos generalizar mejor las características importantes. Sin embargo, uno de los problemas más grandes de estos tipos de modelos es que no son muy buenos para realizar esta tarea al utilizar imágenes tan grandes y muchos de los rasgos se podrían perder al solo reducir o escalar las imágenes. Por ende se optó por un enfoque distinto, dividir cada imagen en distintos parches que contienen la mayor cantidad de información posible y de esta forma pasarsela al modelo.

Por lo anterior, ya que el procesamiento y entrenamiento de los modelos se volvió más complejo para el segundo intentó, solo se realizó para la arquitectura que presentó mejores resultados en este caso, EfficientNet. Este modelo fue entrenado usando de forma individual cada parche, de manera que no hacía distinción por imagen sino por un parche en específico de la imagen. Este enfoque resultó mucho más eficiente y permitió que el modelo se enfocara más en diversas características de los coágulos, presentando un sesgo mucho menor que el que se evaluó inicialmente. Sin embargo debido a lo mismo y la diversidad de la data preprocesada con múltiples transformaciones. El modelo tendió ligeramente a sobre-ajustarse debido a una pérdida en aumento de la pérdida en la data de entrenamiento. Sin embargo, los resultados demostraron una enorme y muy potente capacidad de predicción.

Además este enfoque permitió un manejo de error adicional. Dado que el modelo fue entrenado con cada parche de forma individual, solo puede predecir por cada parche. Por lo tanto cada imagen debe ser sometida a una división por parches antes de que se dé la predicción, lo cual involucra que la predicción final sea decidida en base a votación. Un sistema que además es favorecido a nivel de imagen debido a la muy buena capacidad predictiva del modelo, lo que hace que un falso positivo o un falso negativo sea aún más raro de darse.

Dicho esto, se puede establecer que los resultados obtenidos remarcan la importancia de poder seleccionar un modelo adecuado según el contexto, las circunstancias y las necesidades de un diagnóstico. Por un lado, las CNN sobresalen en entornos con datos bien definidos y una complejidad moderada en cuanto a aspectos visuales, a pesar de carecer de flexibilidad en la detección de relaciones contextuales complejas. Por otro lado, los ViTs demostraron superioridad con los patrones distribuidos, aunque su aplicabilidad se ve mermada a su necesidad de altos volúmenes de datos y su alto costo computacional involucrado. Cabe destacar que las CNN presentan una ligera mejora en cuanto a evitar falsos positivos en la clasificación, lo cual podrá ser de utilidad para diagnósticos que tengan como objetivo evitar tratamiento innecesarios. EfficientNet puede ser una solución balanceada entre precisión y eficiencia y que puede ayudar a centros de salud con capacidades computacionales limitadas. Tanto ésta opción como el uso de transformadores visuales ayudarán en la estandarización de diagnósticos y ayudarán a reducir la variabilidad entre médicos ante casos complejos.

## 10. Hallazgos y conclusiones

Los modelos realizados permitieron evaluar y comparar el rendimiento de modelos basados en redes neuronales convolucionales, EfficientNet, y modelos basados en Vision Transformers en la identificación de la etiología de coágulos isquémicos. Un hallazgo clave fue que, aunque ambos enfoques tienen el potencial de detectar patrones visuales complejos en imágenes médicas, su rendimiento varía significativamente dependiendo del preprocesamiento de datos y del enfoque de segmentación de imágenes que sea utilizado.

Inicialmente, el modelo EfficientNet B1 mostró un rendimiento superior en la clasificación binaria de los coágulos, aunque también exhibió un sesgo significativo hacia una de las clases. Al implementar ViTs, se observó que aunque los transformadores tienen una capacidad notable para captar dependencias globales dentro de las imágenes, presentaron un sesgo aún mayor y un menor rendimiento en comparación con EfficientNet. Con ello, se sugiere que en este caso específico, los modelos convolucionales fueron más apropiados para el proceso de clasificación en cuanto a precisión y balance de clases.

El modelo EfficientNet logró una mejora considerable al implementar un enfoque de segmentación de las imágenes en parches. Este método de división de imágenes permitió que el modelo se enfocara en detalles más específicos de los coágulos sin perder información clave por reducción o escalamiento. Aunque se incrementó la complejidad en el procesamiento y el riesgo de sobreajuste, los resultados demostraron que la capacidad predictiva del modelo tuvo una mejora notable. Además, el sistema de votación a nivel de imagen, apoyó en la reducción de errores de clasificación, mejorando la precisión general del modelo.

En conclusión, el uso de EfficientNet puede ser beneficioso para centros médicos con capacidades computacionales limitadas, ofreciendo un rendimiento confiable en la identificación de coágulos sin necesidad de infraestructura avanzada. Los ViTs, por su parte, siguen siendo una alternativa viable al contar con grandes volúmenes de datos y recursos computacionales elevados, aunque presentan limitaciones prácticas en escenarios clínicos que requieren un diagnóstico rápido y eficiente. Estos hallazgos resaltan la importancia de adaptar el modelo según las necesidades y el contexto clínico, y sugieren enfoques híbridos, los cuales combinan la capacidad de detección local de las CNN y la atención global de los ViTs para mejorar aún más la precisión y aplicabilidad de estos modelos en la práctica médica y en el futuro.

## 11. Referencias


Azatyan, D. (2023). *Image Classification of Stroke Blood Clot Origin using Deep Convolutional Neural Networks and Visual Transformers*. arXiv preprint. https://arxiv.org/abs/2305.16492

Krishnan, K. S., John, J. N. P., Gnanasekar, S., & Krishnan, K. S. (2023). *Advancing Ischemic Stroke Diagnosis: A Novel Two-Stage Approach for Blood Clot Origin Identification*. https://ar5iv.labs.arxiv.org/html/2304.13775

World Stroke Organization. (2022). *Global stroke fact sheet 2022*. https://www.world-stroke.org


## 12. Anexos

Enlace a presentación