

Detección y clasificación automática de Hemorragia Intracerebral

A. Ochoa, E. Ortiz, M. Porras, S. Urrea

27 de mayo de 2024

1. Objetivo del proyecto

Desarrollar un algoritmo de inteligencia artificial basado en redes neuronales convolucionales (CNN) para la detección automática de hemorragias intracraneales (ICH) y la clasificación de sus subtipos en escáneres de tomografía computarizada (TC) de cabeza sin contraste. Este algoritmo será entrenado y validado utilizando un gran conjunto de datos públicos, específicamente 188,202 escáneres de TC de cabeza proporcionados por la Sociedad Americana de Radiología (RSNA) en el RSNA Intracranial Hemorrhage Detection Challenge, con el objetivo de mejorar la precisión y la velocidad del diagnóstico de ICH, apoyando así a los radiólogos en la toma de decisiones clínicas.

2. Justificación del proyecto

La hemorragia intracraneal (ICH) es una condición médica grave que puede causar discapacidad severa o incluso la muerte, representando más del 50 % de la mortalidad total por accidentes cerebrovasculares a pesar de constituir solo el 10-15 % de todos los accidentes cerebrovasculares. Un diagnóstico oportuno es crucial para determinar la necesidad de una intervención quirúrgica de emergencia, permitiendo tratamientos inmediatos que pueden prevenir la expansión de la hemorragia y complicaciones adicionales. Sin embargo, el diagnóstico de ICH y sus subtipos a partir de imágenes de TC es altamente desafiante debido a la variabilidad en la forma, tamaño y ubicación de la hemorragia, y a la posible confusión con calcificaciones y artefactos.

El método tradicional de revisión por pares, donde dos radiólogos revisan independientemente las mismas imágenes, ha demostrado ser efectivo para reducir errores diagnósticos. Sin embargo, este proceso es extremadamente laborioso y requiere una cantidad significativa de tiempo de radiólogos especializados. Los algoritmos de inteligencia artificial basados en el aprendizaje profundo han demostrado un rendimiento equivalente al de los especialistas en la interpretación de imágenes médicas. Estos algoritmos pueden actuar como un segundo lector para asegurar la detección adecuada de hallazgos sutiles y como herramientas de triage para diagnósticos oportunos. En este proyecto, utilizamos la arquitectura preentrenada DenseNet121, aplicando transfer learning para adaptarla a la clasificación de ICH y sus subtipos. El modelo se entrenará con una tasa de aprendizaje baja, sin congelar las capas existentes y añadiendo capas adicionales para mejorar la capacidad de clasificación. El resultado de este enfoque proporciona una herramienta valiosa para la práctica clínica.

3. Descripción del conjunto de datos

El conjunto de datos principal utilizado en este proyecto proviene del RSNA Intracranial Hemorrhage Detection Challenge. Este fue recopilado de tres instituciones (Stanford University, Universidade Federal de São Paulo y Thomas Jefferson University Hospital) y reanotado por la American Society of Neuroradiology (ASNR) con la contribución de más de 60 neurorradiólogos. El conjunto de datos consiste en más de un millón de imágenes provenientes de 25,272 estudios, proporcionando el mayor dataset multi-institucional y multinacional para el estudio de la detección y clasificación de subtipos de hemorragia intracraneal (ICH).

Durante la competencia RSNA, el conjunto de datos completo se publicó en dos etapas (stage 1 y stage 2). Los datos de la primera etapa no se encuentran públicos por lo que en este proyecto fueron usados los datos de la segunda etapa. Estos corresponden a 752,803 escáneres de TC de cabeza. Del total de estos datos, definimos

un sample del 25 % para realizar el entrenamiento del modelo, equivalente a 188,202 escáneres. Esto nos permite manejar un volumen de datos significativo y computable, dadas nuestras restricciones computacionales.

El conjunto de datos del RSNA tenía la siguiente estructura en su archivo CSV de etiquetas: los datos de entrenamiento se proporcionaban como un conjunto de IDs de imágenes y múltiples etiquetas, una para cada uno de los cinco subtipos de hemorragia, además de una etiqueta adicional para “any”, que siempre es verdadera si alguna de las etiquetas de subtipos es verdadera. Cada imagen tenía seis filas en el CSV, una por cada tipo de hemorragia más la columna “any”. Las etiquetas para una imagen en particular se estructuraban de la siguiente manera:

```
Id,Label
ID_1b8ac55df_epidural_hemorrhage,0
ID_1b8ac55df_intraparenchymal_hemorrhage,0
ID_1b8ac55df_intraventricular_hemorrhage,0
ID_1b8ac55df_subarachnoid_hemorrhage,1
ID_1b8ac55df_subdural_hemorrhage,0
ID_1b8ac55df_any,1
```

Para facilitar el entrenamiento, convertimos esta estructura a un formato tabular donde cada imagen tenía una única fila con columnas para cada subtipo de hemorragia y la columna “any”, como se muestra a continuación:

ID	epidural	intraparenchymal	intraventricular	subarachnoid	subdural	any
ID_1b8ac55df	0	0	0	1	0	1
ID_03579392d	0	0	0	0	0	0
ID_9989097e7	0	0	0	0	0	0
ID_8e27ea473	0	0	0	0	0	0

Todas las imágenes proporcionadas están en formato DICOM, que contiene metadatos asociados, incluyendo PatientID, StudyInstanceUID, SeriesInstanceUID, Window Center, Window Width, entre otras características.

3.1. Procesamiento de las Imágenes

Para preparar las imágenes para el entrenamiento del modelo, realizamos varios pasos de procesamiento. Primero, extraemos los metadatos relevantes de cada imagen DICOM, tales como el centro y ancho de ventana, intercepto y pendiente de rescalado. Utilizamos estos metadatos para ajustar la escala de las imágenes, aplicando una ventana de imagen que mejora la visibilidad de las características relevantes.

Luego, normalizamos las imágenes para que los valores de píxeles se encuentren en un rango uniforme, mejorando así la consistencia durante el entrenamiento del modelo. Las imágenes se redimensionan a una resolución uniforme adecuada para el procesamiento eficiente por la red neuronal convolucional. Finalmente, las imágenes procesadas se guardan en formato PNG para su uso en etapas posteriores del pipeline de entrenamiento.

4. Arquitectura del Modelo

Utilizamos un modelo DenseNet121 preentrenado, el cual adaptamos para la tarea de detección y clasificación de subtipos de hemorragia intracraneal (ICH). La red original fue modificada añadiendo capas adicionales para refinar la clasificación. En particular, añadimos una capa lineal seguida de una función de activación ReLU y una capa de Dropout para regularización, culminando en una capa lineal final que produce las predicciones para las seis clases de hemorragia.

La arquitectura del modelo se implementa de la siguiente manera:

```

class DenseNet121_Classifier(nn.Module):
    def __init__(self, num_classes=6):
        super(DenseNet121_Classifier, self).__init__()
        self.densenet = models.densenet121(weights='DEFAULT')
        num_ftrs = self.densenet.classifier.in_features
        self.densenet.classifier = nn.Sequential(
            nn.Linear(num_ftrs, 512),
            nn.ReLU(),
            nn.Dropout(0.2),
            nn.Linear(512, num_classes)
        )

    def forward(self, x):
        x = self.densenet(x)
        return x

```

El modelo fue entrenado usando la función de pérdida `BCEWithLogitsLoss()`. Utilizamos el optimizador Adam con una tasa de aprendizaje inicial de 0.0005 y parámetros específicos para betas, epsilon y weight decay para ajustar el modelo. Los parámetros del optimizador Adam fueron seleccionados de la siguiente manera: las betas (0.9 y 0.999) controlan los momentos de primer y segundo orden del gradiente, lo que ayuda a estabilizar y acelerar la convergencia. El valor de epsilon ($1e-08$) se utiliza para evitar divisiones por cero en la actualización de los parámetros. El weight decay (0.00002) ayuda a prevenir el sobreajuste al penalizar grandes valores de los pesos del modelo, fomentando así una mayor generalización. El entrenamiento se llevó a cabo en una GPU Nvidia A40, lo que permitió manejar el gran volumen de datos y la complejidad del modelo.

Adicionalmente, implementamos un mecanismo de *Early Stopping* para prevenir el sobreajuste durante el entrenamiento, monitoreando la pérdida de validación y deteniendo el entrenamiento si no se observaba mejora después de 8 épocas consecutivas.

Durante el entrenamiento, se registraron y evaluaron las siguientes métricas: precisión, recall y AUC (Área bajo la curva ROC), tanto para los conjuntos de entrenamiento como de validación.

5. Resultados

Se van a explicar los resultados obtenidos en esta investigación. Se tiene que definir lo que se hizo en entrenamiento, validación y en pruebas finales.

5.1. Entrenamiento y validación

En este apartado se obtuvo lo siguiente en las gráficas de la evolución de las métricas de precision, recall, loss y auc:

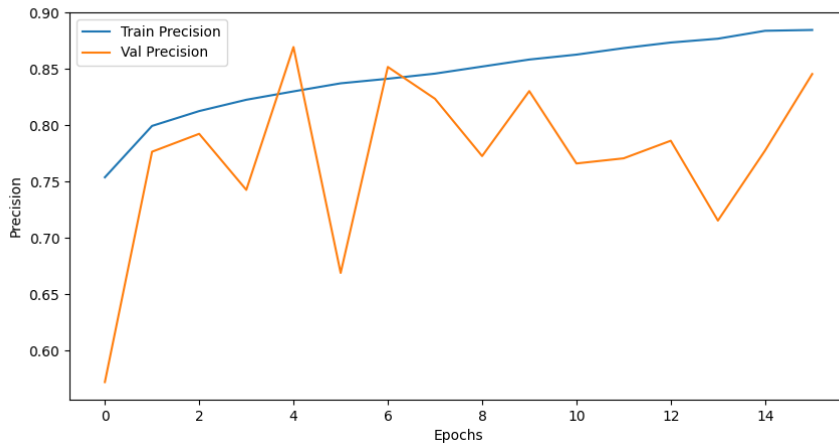


Figura 1: Gráfica de precisión por cada época.

Se puede observar en la Figura 1 que la precisión llega al punto en que tanto el train y el val se acercan bastante en tres puntos distintos de la gráfica , cualquiera de estos puntos de épocas son recomendables pero es mejor el último acercamiento.

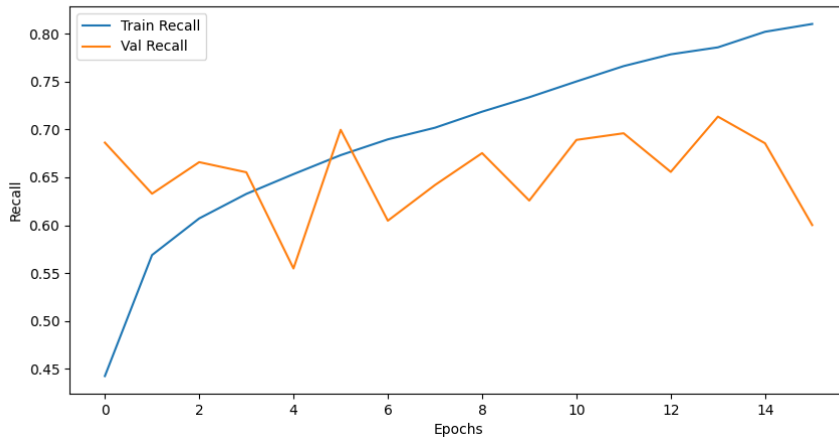


Figura 2: Gráfica de recall por cada época.

El recall si es un valor que se aleja un poco a medida de mayor número de épocas en el conjunto de validación.

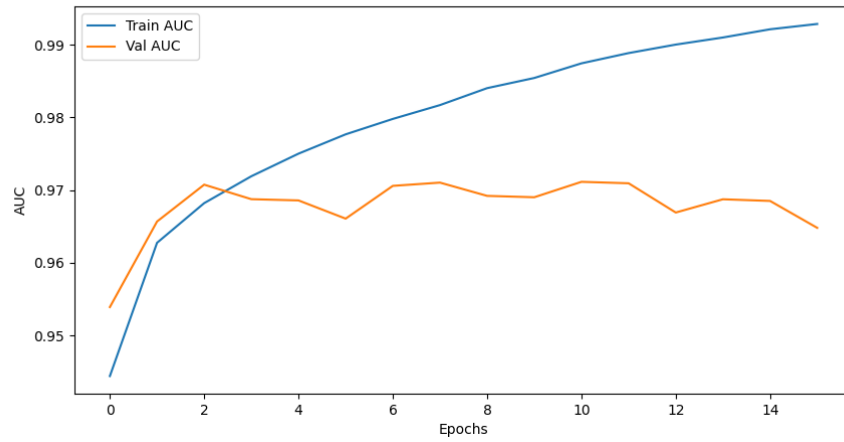


Figura 3: Gráfica de auc por cada época.

En esta gráfica de AUC muestra que un valor de AUC se estabiliza pero no mejora en el conjunto de validación.

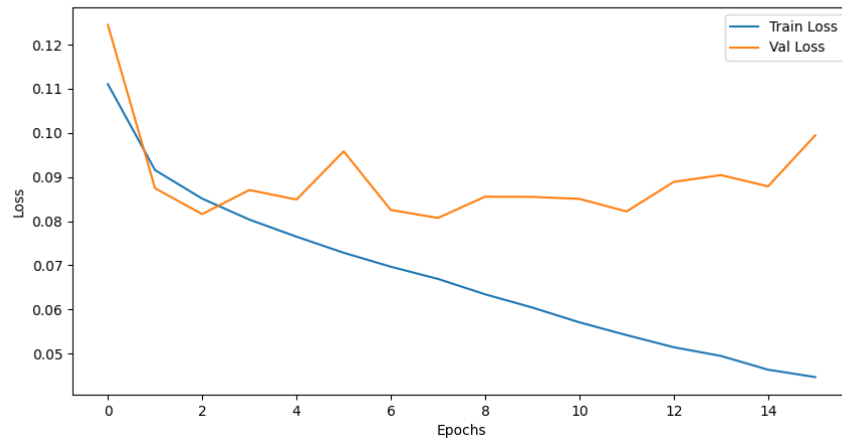


Figura 4: Gráfica de loss por cada época.

Finalmente , en la gráfica de loss , se puede observar que de nuevo al conjunto de validación le costó un poco más estabilizarse a la baja , pero se logra mantener en un valor aceptable.

5.2. Pruebas finales

Con el conjunto de datos de pruebas finales se obtuvo el siguiente informe de clasificación:

	precision	recall	f1-score	support
epidural	0.00	0.00	0.00	9
intraparenchymal	1.00	0.67	0.80	3
intraventricular	1.00	1.00	1.00	2
subarachnoid	1.00	0.12	0.22	8
subdural	0.57	0.36	0.44	11
any	0.92	0.42	0.58	26
micro avg	0.83	0.34	0.48	59
macro avg	0.75	0.43	0.51	59
weighted avg	0.73	0.34	0.44	59
samples avg	0.09	0.07	0.08	59

Figura 5: Reporte de clasificación.

El informe de clasificación proporciona las métricas basadas en un umbral específico (por ejemplo, 0.5). Si ninguno de los predicciones para la clase .`epidural` supera este umbral, entonces tanto la precisión, el recall y el f1-score serán 0. Esto puede suceder si el modelo está extremadamente conservador en predecir la clase .`epidural`.

La curva ROC (Receiver Operating Characteristic) y el AUC (Area Under the Curve) se calculan considerando todos los umbrales posibles. Un AUC de 0.98 indica que el modelo tiene una buena capacidad de distinguir entre las clases positivas y negativas en general, independientemente del umbral. El modelo puede estar asignando puntuaciones más altas a las instancias positivas de .`epidural` en comparación con las instancias negativas, incluso si ninguna de estas puntuaciones supera el umbral específico utilizado en el informe de clasificación.

Además , se obtuvieron los siguientes resultados con la gráfica ROC en cada una de las clases que predice el modelo:

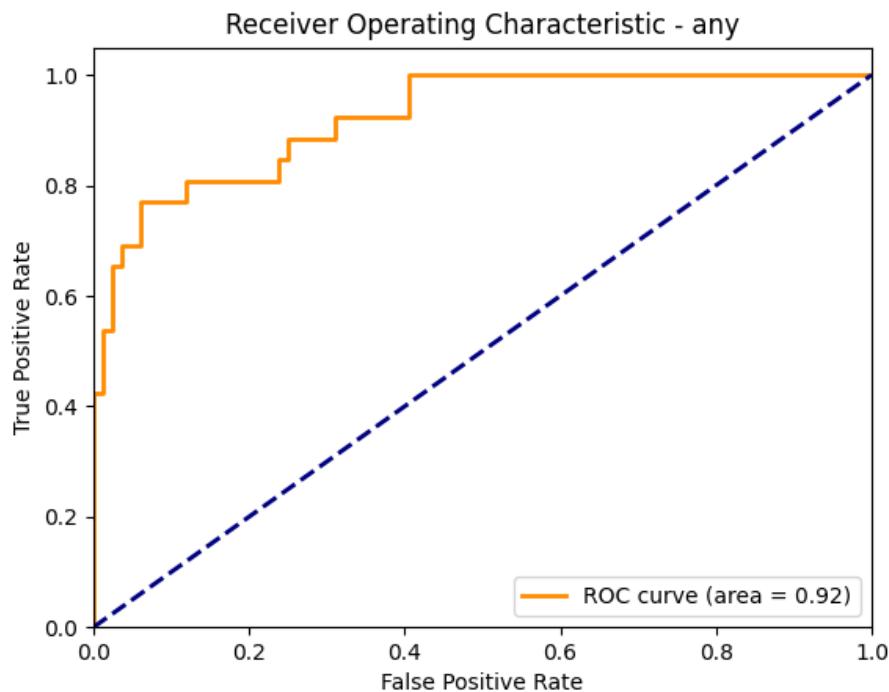


Figura 6: Curva roc sobre cualquier clase.

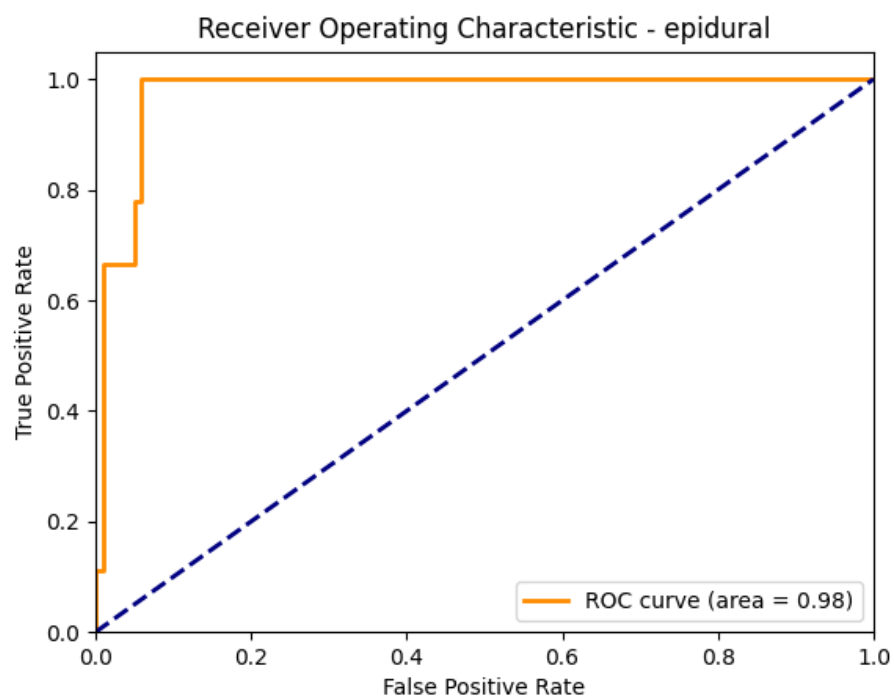


Figura 7: Curva roc sobre la clase epidural.

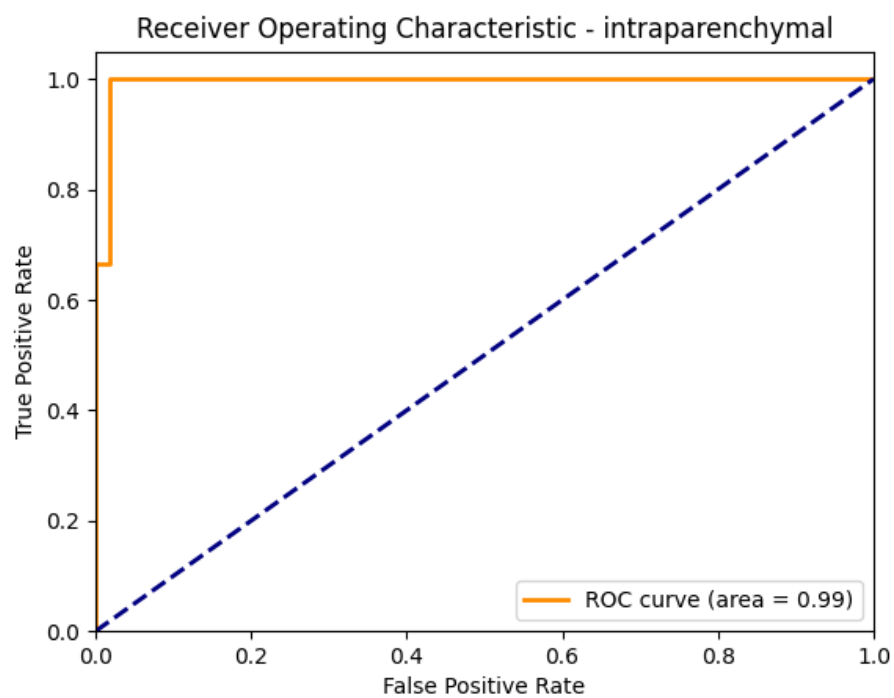


Figura 8: Curva roc sobre la clase intraparenchymal.

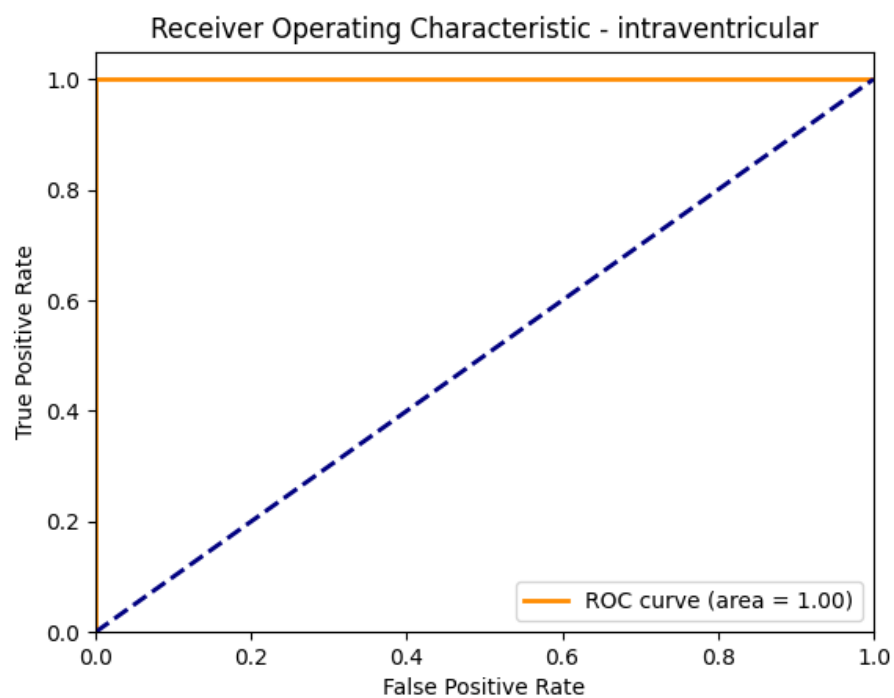


Figura 9: Curva roc sobre la clase intraventricular.

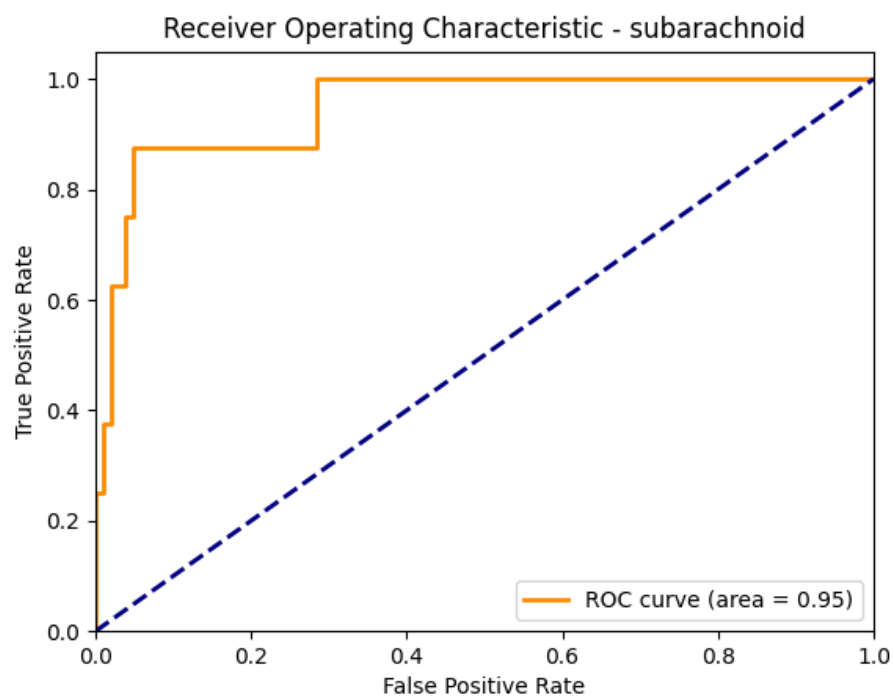


Figura 10: Curva roc sobre la clase subarachnoid.

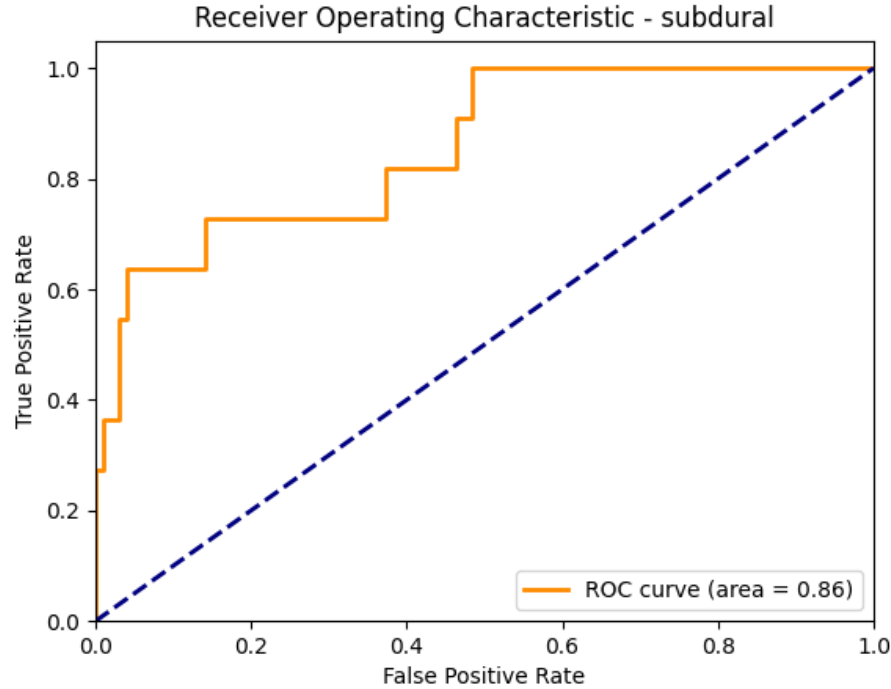


Figura 11: Curva roc sobre la clase subdural.

6. Trabajo futuro

Se propone el desarrollo de un pipeline de dos algoritmos de inteligencia artificial: uno dedicado a la clasificación y detección de hemorragias intracraneales, y otro enfocado en la segmentación de las mismas. El objetivo principal es asegurar la consistencia en los resultados obtenidos por ambos algoritmos, de modo que si en una imagen específica se detecta una hemorragia, se proceda con la segmentación correspondiente; de lo contrario, no se realizará ninguna segmentación. Adicionalmente, se plantea la implementación de una arquitectura unificada que permita la clasificación y segmentación de manera simultánea.

7. Artículos Relacionados

7.1. A deep learning algorithm for automatic detection and classification of acute intracranial hemorrhages in head CT scans

(Wang y cols., 2021) presenta un algoritmo deep learning para la detección y clasificación automática de hemorragias intracraneales agudas (ICH) en escáneres de tomografía computarizada (TC) de cabeza sin contraste. El algoritmo combina un modelo CNN 2D y dos modelos de secuencia para lograr una detección precisa de ICH y la clasificación de sus subtipos. Al igual que el nuestro fue desarrollado utilizando el conjunto de datos del RSNA Intracranial Hemorrhage Detection Challenge. El algoritmo alcanzó niveles de precisión comparables a los de radiólogos expertos. Los AUC obtenidos fueron 0.988 para ICH, y entre 0.983 y 0.996 para los subtipos de hemorragia. Las implementaciones en GitHub en PyTorch de este trabajo fueron de gran ayuda en la implementación de la arquitectura de nuestro proyecto.

7.2. The state-of-the-art 3D anisotropic intracranial hemorrhage segmentation on non-contrast head CT: The INSTANCE challenge

(Li y cols., 2023) aborda el problema de la segmentación automática de hemorragias intracraneales en escáneres de TC de cabeza sin contraste en 3D, destacando la naturaleza anisotrópica de los datos de TC. La

competencia INSTANCE 2022, organizada en conjunto con la Conferencia Internacional sobre Computación de Imágenes Médicas e Intervención Asistida por Computadora (MICCAI), proporcionó en volúmenes 3D un conjunto de entrenamiento con 100 casos y un conjunto de validación con 30 casos. Los métodos fueron evaluados utilizando métricas como el Coeficiente de Similitud de Dice (DSC) y la Distancia de Hausdorff (HD). El método ganador alcanzó un DSC promedio de 0.6925, superando significativamente el método base propuesto. Este desafío presenta el primer benchmark de segmentación de hemorragias intracraneales que aborda el problema anisotrópico en la segmentación de imágenes médicas 3D.

7.3. Improved performance and robustness of multi-task representation learning with consistency loss between pretexts for intracranial hemorrhage identification in head CT

(Kyung y cols., 2022) propone un marco de red de transferencia de representación multitarea supervisada (SMART-Net) para la identificación de ICH en TC de cabeza sin contraste, abordando los desafíos de la heterogeneidad de ICH y la necesidad de alto rendimiento en sensibilidad y especificidad. El marco consta de componentes de upstream y downstream. En el upstream, un codificador con pesos compartidos se entrena como extractor de características robustas mediante tareas multi-pretexto a nivel de corte (clasificación, segmentación y reconstrucción), añadiendo una pérdida de consistencia para regularizar las discrepancias entre las cabezas de clasificación y segmentación. En el downstream, se realiza la transferencia de aprendizaje con un codificador preentrenado y un operador 3D para tareas a nivel de volumen. Los resultados experimentales mostraron que SMART-Net tiene mejor robustez y rendimiento en clasificación y segmentación de ICH a nivel de volumen en comparación con métodos previos. El código disponible en GitHub permitió que analizáramos y consideráramos detalles importantes el proceso de extracción de características de la imagen.

Referencias

- Kyung, S., Shin, K., Jeong, H., Kim, K. D., Park, J., Cho, K., ... Kim, N. (2022, may). Improved performance and robustness of multi-task representation learning with consistency loss between pretexts for intracranial hemorrhage identification in head ct. *Med Image Anal*, 81, 102489.
- Li, X., Luo, G., Wang, K., Wang, H., Liu, J., Liang, X., ... Li, S. (2023). *The state-of-the-art 3d anisotropic intracranial hemorrhage segmentation on non-contrast head ct: The instance challenge*.
- Wang, X., Shen, T., Yang, S., Lan, J., Xu, Y., Wang, M., ... Han, X. (2021, aug). A deep learning algorithm for automatic detection and classification of acute intracranial hemorrhages in head ct scans. *Neuroimage Clin*, 32, 102785.