

ESTRATEGIA COMPUTACIONAL PARA LA PREDICCIÓN DEL TRATAMIENTO DE
PACIENTES CON ACCIDENTE CEREBROVASCULAR ISQUÉMICO USANDO
HALLAZGOS RADIOLÓGICOS Y VARIABLES CLÍNICAS

LAURA RODRIGUEZ CALA

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FISICOMECÁNICAS
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
BUCARAMANGA

2025

ESTRATEGIA COMPUTACIONAL PARA LA PREDICCIÓN DEL TRATAMIENTO DE
PACIENTES CON ACCIDENTE CEREBROVASCULAR ISQUÉMICO USANDO
HALLAZGOS RADIOLÓGICOS Y VARIABLES CLÍNICAS

LAURA RODRIGUEZ CALA

Tesis presentada en cumplimiento de los requisitos para optar por el título de:
Ingeniera de Sistemas

Director:

Fabio Martínez Carrillo

Doctor en Ingeniería de Sistemas y Computación

Codirectora:

Alejandra Moreno Tarazona

Magíster en Ingeniería de Sistemas e Informática

Codirector:

Santiago Gómez Hernández

Magíster en Ingeniería de Sistemas e Informática

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FISICOMECAÑICAS
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
BUCARAMANGA

2025

AGRADECIMIENTOS

En primer lugar doy gracias a Dios por permitirme finalizar esta bonita etapa de mi vida. A mi familia, mi mamá y mi papá, que han confiado en mí y me han brindado su apoyo y cariño durante todos estos años. A mis hermanas, con quienes he compartido bellos momentos y superado diversas dificultades, cada una me ha enseñado cosas valiosas, gracias por escucharme, ayudarme y acompañarme en todo este proceso.

Agradezco a la Universidad Industrial de Santander por brindarme una formación integral durante la carrera, el estudio, el deporte y los idiomas han hecho que sea una experiencia enriquecedora. A cada uno de los profesores y profesoras de la Escuela de Ingeniería de Sistemas e Informática, cada clase, cada curso, aportó para alcanzar satisfactoriamente esta meta.

A mi director de tesis, el profesor Fabio, por su constante acompañamiento, apoyo y dedicación durante el desarrollo de este trabajo. Su aporte y compromiso como docente e investigador, motivan a continuar trabajando y contribuyendo desde la ciencia. Gracias por permitirme ser parte del grupo de investigación, una oportunidad que me permitió crecer personal y profesionalmente.

A mi codirectora Alejandra, por su guía, enseñanza, su amabilidad y paciencia, 고마워. A mi codirector Santiago, por su tiempo, su disposición para explicar y compartir de la mejor manera el conocimiento, merci. Excelentes personas y profesionales con quienes tuve la oportunidad de coincidir y aprender. Para ambos, mi respeto y admiración.

Agradezco también a cada uno de los integrantes del grupo de investigación *BIVL²ab* quienes hicieron del grupo un espacio más agradable para aprender, discutir y proponer ideas, gracias por los momentos compartidos. Los recordaré con mucho cariño.

A Juan David, Daniel, Oscar y Santiago, a quienes conocí durante la carrera y que con el paso del tiempo, los proyectos, las tristezas y las risas se han convertido en grandes amigos.

CONTENIDO

	pág.
INTRODUCCIÓN	11
1. FUNDAMENTOS Y TRABAJOS PREVIOS	14
1.1. ACCIDENTE CEREBROVASCULAR (ACV) ISQUÉMICO	14
1.1.1 Análisis desde la resonancia magnética ponderada por difusión	14
1.1.2 Análisis de variables Clínicas	15
1.1.3 Tratamiento del Accidente Cerebrovascular Isquémico	18
1.2. PROCESAMIENTO DE VARIABLES VISUALES Y TEXTUALES	20
1.2.1 Representaciones visuales basadas en atención	20
1.2.2 Procesamiento de datos tabulares	21
1.2.3 Metodologías multimodales	23
1.3. MÉTODOS COMPUTACIONALES PARA LA PREDICCIÓN DEL TRATAMIENTO DEL ACV ISQUÉMICO	25
2. PROBLEMA DE INVESTIGACIÓN	30
2.1. PREGUNTA DE INVESTIGACIÓN	30
3. OBJETIVOS	31
3.1. OBJETIVO GENERAL	31
3.2. OBJETIVOS ESPECÍFICOS	31
4. MÉTODO PROPUESTO	32
4.1. MODELO PARA LA EXTRACCIÓN DE CARACTERÍSTICAS IMAGENOLÓGI- CAS DE LA LESIÓN	33
4.2. REPRESENTACIÓN DE VARIABLES CLÍNICAS	35

4.3. PREDICCIÓN DEL TRATAMIENTO A PARTIR DE UNA FUSIÓN INTERMEDIA DE DESCRIPTORES MULTIMODALES	36
4.4. DISEÑO EXPERIMENTAL	36
4.4.1 Datos	36
4.4.2 Configuración de la arquitectura	38
5. EVALUACIÓN Y RESULTADOS	41
5.1. EVALUACIÓN DE MODELOS UNIMODALES PARA LA PREDICCIÓN DEL TRATA- MIENTO	41
5.2. EVALUACIÓN MULTIMODAL PARA LA PREDICCIÓN DEL TRATAMIENTO . .	43
6. CONCLUSIONES Y TRABAJO FUTURO	47
BIBLIOGRAFÍA	49

LISTA DE FIGURAS

	pág.
Figura 1. Imágenes de Resonancia Magnética de dos pacientes con ACV isquémico. Las imágenes de DWI muestran la lesión como una zona hiperintensa y en las imágenes de ADC la lesión se presenta como una zona hipointensa.	16
Figura 2. Mecanismo de auto atención para la extracción de características no locales. En este caso X representa una imagen de MRI a partir de la cual se puede obtener una caracterización de la lesión isquémica.	21
Figura 3. Estrategias de fusión	24
Figura 4. Método multimodal. a) Modelo de imagen que extrae las características de la lesión y adiciona información del territorio arterial afectado. b) Modelo clínico que obtiene una codificación de las variables clínicas. c) Fusión intermedia de los descriptores para predecir el tratamiento.	32
Figura 5. Matrices de confusión para la predicción multiclase del tratamiento en pacientes con ACV isquémico. Se compara el desempeño del modelo por ensamblaje secuencial de clasificadores binarios (izquierda) con la estrategia de clasificación multiclase directa (derecha).	46

LISTA DE TABLAS

	pág.
Tabla 1. Intervalos escala NIHSS junto al grado de severidad	17
Tabla 2. Escala mRS y su relación respecto a la condición física del paciente	17
Tabla 3. Descripción de las variables clínicas del conjunto de datos. TIS: Tiempo desde el inicio de los síntomas en horas, TAA: territorio arterial afectado.	38
Tabla 4. Comparación del desempeño de modelos imagenológicos bajo configuraciones de predicción binaria. Se contrastan los resultados del codificador imagenológico propuesto frente a una ResNet152 entrenada sobre mapas ADC. Las métricas reportadas corresponden a las configuraciones NoT vs. T (sin tratamiento vs. tratamiento) y TM vs. TF (trombectomía mecánica vs. trombólisis farmacológica). Los mejores resultados se encuentran resaltados en negrilla.	42
Tabla 5. Comparación del desempeño de modelos clínicos bajo configuraciones de predicción binaria. Se contrastan los resultados del codificador clínico propuesto frente a un Random Forest. Las métricas reportadas corresponden a las configuraciones NoT vs. T (sin tratamiento vs. tratamiento) y TM vs. TF (trombectomía mecánica vs. trombólisis farmacológica). Los mejores resultados se encuentran resaltados en negrilla.	43

Tabla 6. Comparación del desempeño de modelos multimodales bajo configuraciones de predicción binaria. Se contrastan los resultados del modelo propuesto frente a estrategias convolucionales y de atención. Las métricas reportadas corresponden a las configuraciones NoT vs. T (sin tratamiento vs. tratamiento) y TM vs. TF (trombectomía mecánica vs. trombólisis farmacológica). Los mejores resultados se encuentran resaltados en negrilla. 44

Tabla 7. Comparación del desempeño entre una estrategia de ensamblaje secuencial de modelos binarios y un modelo de clasificación multiclase en la predicción del tratamiento en pacientes con ACV isquémico. Los mejores resultados para cada métrica están resaltados en negrilla. 45

RESUMEN

TÍTULO: Estrategia computacional para la predicción del tratamiento de pacientes con accidente cerebrovascular isquémico usando hallazgos radiológicos y variables clínicas *

AUTOR: Laura Rodriguez Cala **

PALABRAS CLAVE: Accidente cerebrovascular, predicción del tratamiento, secuencias radiológicas, variables clínicas

DESCRIPCIÓN: El accidente cerebrovascular es la segunda causa de mortalidad y la tercera causa de discapacidad en todo el mundo. La atención inmediata y la planificación del tratamiento son claves para un pronóstico favorable del paciente. Según las guías clínicas, el tratamiento se define a partir de variables como la localización del territorio arterial afectado, el volumen de la lesión, las comorbilidades y la afectación neurológica. Sin embargo, tal definición y planificación dependen de la revisión completa y de la correlación compleja de variables por parte de los expertos, lo cual es subjetivo y repercute en la planificación efectiva del tratamiento. En este trabajo de investigación se desarrolló una estrategia multimodal que tiene la capacidad de codificar variables clínicas, así como también patrones imagenológicos de la lesión. En cuanto a las variables clínicas, se aprendió un diccionario para dar una codificación embebida de estas variables. En otra rama de procesamiento, la región de interés fue codificada a través de una arquitectura *transformer*, para obtener un embebido de la representación visual. Seguidamente, tanto el embebido de texto como el embebido visual son concatenados y ajustados para determinar la clasificación del tratamiento. En un estudio con 156 pacientes y tres posibles tratamientos (trombólisis farmacológica, trombectomía mecánica o tratamiento conservador), el enfoque multiclase propuesto alcanzó un F1 promedio de 54.1% y una exactitud de 55.8%, brindando una herramienta prometedora para la definición del tratamiento.

* Trabajo de investigación

** Facultad de Ingenierías Fisicomecánicas. Escuela de Ingeniería de Sistemas e Informática. Director: Fabio Martínez, PhD. Codirectores: Alejandra Moreno Tarazona, Msc., Santiago Gómez Hernández, Msc.

ABSTRACT

TITLE: Computational strategy for treatment prediction of ischemic stroke patients using radiological findings and clinical variables. *

AUTHOR: Laura Rodriguez Cala **

KEYWORDS: Stroke, treatment prediction, radiological sequences, clinical variables

DESCRIPTION:

Stroke is the second leading cause of death and the third leading cause of disability worldwide. Immediate attention and treatment planning are key for favorable output patient prognosis. According to clinical guides, the treatment is defined from variables such as the affected territorial artery localization, the lesion volume, the comorbidities, and the neurological affections. Nonetheless, such definition and planning rely on the comprehensive review and complex correlation of variables by experts, which is subjective and impacts the effective planning of treatment. In this research work, a multimodal strategy was developed with the capability of encoding clinical variables as well as imaging patterns of the lesion. Regarding clinical variables, a dictionary was learned to provide an embedded encoding of these variables. In another processing branch, the region of interest was encoded through a transformer architecture to obtain an embedding of the visual representation. Then, both the text embedding and the visual embedding are concatenated and fine-tuned to determine the treatment classification. In a study with 156 patients and three possible treatments (pharmacological thrombolysis, mechanical thrombectomy, or conservative treatment), the proposed multiclass approach achieved an average F1 of 54.1% and an accuracy of 55.8% providing a promising tool for treatment decision-making.

* Research work

** Faculty of Physics-Mechanics Engineering. School of Systems Engineering and Informatics. Advisor: Fabio Martínez Carrillo, PhD. Co-advisors: Alejandra Moreno Tarazona, Msc., Santiago Gómez Hernández, Msc.

INTRODUCCIÓN

El accidente cerebrovascular (ACV) es la segunda causa de mortalidad en el mundo y la tercera causa de discapacidad, donde un tercio de los supervivientes presentaron deficiencias motoras, sensoriales y del lenguaje ^{1,2}. La obstrucción isquémica es el tipo de ACV más común, representando aproximadamente el 87% de los casos. Para el análisis y caracterización del ACV, los estudios imagenológicos constituyen una base fundamental para la caracterización de las lesiones. Entre estos, estudios como la resonancia magnética (MRI, por sus siglas en inglés) permiten visualizar el tejido cerebral y establecer la ubicación y el volumen de la lesión ³. Por otro lado, este análisis es complementado con variables demográficas, comorbilidades y escalas que cuantifican el grado de discapacidad o daño neurológico del paciente. A partir del análisis complementario, entre hallazgos clínicos e imagenológicos, los expertos establecen el tratamiento, lo que puede conducir a un análisis subjetivo y dependiente de la experticia de los neurointervencionistas.

Recientemente se han propuesto estrategias computacionales en el estado del arte que permiten realizar la predicción del pronóstico de pacientes que sufrieron un ACV isquémico. Sin embargo, la mayoría de esfuerzos se han enfocado en resolver la tarea de predicción del pronóstico en términos de escalas de afectación como la escala de Rankin modificada (mRS, por sus siglas en inglés). Estos incluyen enfoques unimodales a partir de imágenes

-
- ¹ Valery L FEIGIN et al. “Global, regional, and national burden of stroke and its risk factors, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019”. In: *The Lancet Neurology* 20.10 (2021), pp. 795–820.
 - ² Valery L FEIGIN et al. “Pragmatic solutions to reduce the global burden of stroke: a World Stroke Organization–Lancet Neurology Commission”. In: *The Lancet Neurology* 22.12 (2023), pp. 1160–1206.
 - ³ Richard E LATCHAW et al. “Recommendations for imaging of acute ischemic stroke: a scientific statement from the American Heart Association”. In: *Stroke* 40.11 (2009), pp. 3646–3678.

diagnósticas, como secuencias MRI ponderadas por difusión (DWI), que extraen patrones visuales relacionados a la lesión o de información clínica como la edad, comorbilidades y escalas de afectación capturadas durante el ingreso del paciente ^{4,5,6}. A pesar de esto, al utilizar una única modalidad se pierde información complementaria que se tiene en cuenta durante la rutina clínica. También se han abordado enfoques multimodales que integran secuencias radiológicas e información clínica para extraer las características relevantes de cada modalidad y realizar la predicción del mRS90 ^{7,8,9,10}. No obstante, en algunos casos se entrenan redes separadas para cada modalidad para luego integrar la salida de cada modelo, lo cual no permite establecer relaciones entre cada modalidad durante el entrenamiento. Además, se ha abordado un trabajo que emplea imágenes de tomografía computarizada (CT, por sus siglas en inglés) e información clínica como la edad, el género, comorbilidades (fribilación auricular, diabetes) para predecir el tratamiento, sin embargo, este trabajo ha sido enfocado para el ACV hemorrágico, el cual tiene un protocolo de atención distinto, sus hallazgos radiológicos son marcados y se tienen en cuenta

-
- ⁴ Qaiser MAHMOOD et al. “Prediction of ischemic stroke lesion and clinical outcome in multi-modal MRI images using random forests”. In: *Workshop BrainLes MICCAI 2016*. Springer. 2016, pp. 244–255.
 - ⁵ Oskar MAIER et al. “Predicting stroke lesion and clinical outcome with random forests”. In: *Workshop BrainLes MICCAI 2016*. Springer. 2016, pp. 219–230.
 - ⁶ Nima HATAMI et al. “A Novel Autoencoders-LSTM Model for Stroke Outcome Prediction using Multi-modal MRI Data”. In: *arXiv preprint arXiv:2303.09484* (2023).
 - ⁷ Yongkai LIU et al. “Functional Outcome Prediction in Acute Ischemic Stroke Using a Fused Imaging and Clinical Deep Learning Model”. In: *Stroke* 54.9 (2023), pp. 2316–2327.
 - ⁸ Zeynel A SAMAK; Philip CLATWORTHY, and Majid MIRMEHDI. “Prediction of thrombectomy functional outcomes using multimodal data”. In: *Annual Conference on Medical Image Understanding and Analysis*. Springer. 2020, pp. 267–279.
 - ⁹ Stephen BACCHI et al. “Deep learning in the prediction of ischaemic stroke thrombolysis functional outcomes: a pilot study”. In: *Academic radiology* 27.2 (2020), e19–e23.
 - ¹⁰ Hongju JO et al. “Combining clinical and imaging data for predicting functional outcomes after acute ischemic stroke: an automated machine learning approach”. In: *Scientific Reports* 13.1 (2023), p. 16926.

diferentes variables ¹¹.

Este trabajo implementó una arquitectura multimodal que integra información clínica e imagenológica para soportar la predicción del tratamiento de pacientes con ACV isquémico. En cuanto a la rama de procesamiento textual, en este trabajo se aprendió una representación embebida de un diccionario, para las principales variables clínicas: edad, escala NIHSS y tiempo de inicio de los síntomas. En cuanto a la rama de procesamiento visual, desde imágenes MRI-ADC se procesó la región asociada a la lesión con el territorio afectado. Esta región es representada por un par de convoluciones, para obtener una representación embebida visual. La representación embebida es mapeada a una arquitectura Transformer (ViT) para obtener embebidos visuales, que a su vez contengan información sobre la codificación de la posición, relacionando la posición relativa de la región. Los embebidos resultantes se integran en una fusión intermedia, y se procesan en conjunto (*end-to-end*) con respecto a la tarea de discriminación entre tres posibles tratamientos: trombólisis farmacológica, trombectomía mecánica o tratamiento conservador. La arquitectura multimodal fue entrenada y validada utilizando un conjunto de datos privado que cuenta con 156 pacientes con mapas de coeficiente de difusión aparente y variables clínicas.

¹¹ Wenao MA et al. "Treatment Outcome Prediction for Intracerebral Hemorrhage via Generative Prognostic Model with Imaging and Tabular Data". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2023, pp. 715–725.

1. FUNDAMENTOS Y TRABAJOS PREVIOS

1.1. ACCIDENTE CEREBROVASCULAR (ACV) ISQUÉMICO

El ACV isquémico ocurre cuando el flujo sanguíneo en el cerebro se ve interrumpido por la obstrucción de un vaso sanguíneo. Estos ACV representan alrededor del 87% de los casos y cada año se producen más de 7.6 millones de nuevos casos ^{2,1}. Esta interrupción del flujo sanguíneo conlleva a una disminución en los niveles de oxígeno y glucosa en el cerebro provocando la muerte de las células del tejido cerebral, también llamado tejido infartado ¹². Dado que el tiempo es crucial en esta enfermedad, es necesario llevar a cabo un diagnóstico y tratamiento inmediato permitiendo restablecer el flujo sanguíneo. Para esto, los expertos realizan un análisis multimodal que abarca la caracterización de la lesión a partir de la adquisición de estudios radiológicos y el estudio de variables clínicas, cuando el paciente ingresa a urgencias. A continuación se detallan cada uno de estos procedimientos.

1.1.1. Análisis desde la resonancia magnética ponderada por difusión

La adquisición y análisis de imágenes diagnósticas es uno de los aspectos fundamentales para la determinación del tratamiento de pacientes con ACV isquémico ³. Este análisis permite estimar la localización del territorio arterial afectado, caracterizar morfológicamente la lesión y determinar el núcleo y la penumbra de la lesión ¹³. Particularmente, los estudios de MRI ponderados por difusión (DWI, por sus siglas en inglés) son la referencia para la caracterización de las lesiones isquémicas, ofreciendo una gran sensibilidad

¹² Mayank GOYAL et al. "Challenging the ischemic core concept in acute ischemic stroke imaging". In: *Stroke* 51.10 (2020), pp. 3147–3155.

¹³ Yang TANG et al. *Atlas of emergency neurovascular imaging*. Tech. rep. Springer, 2020.

y especificidad durante las primeras horas desde el inicio de los síntomas³. Estos estudios proporcionan un contraste de imagen a partir de las diferencias en la magnitud de la difusión de las moléculas de agua dentro del cerebro (b)¹⁴. Típicamente, en el protocolo de ACV se adquieren secuencias DWI con un valor b de 0 y 1000 s/mm², y mapas de coeficiente de difusión aparente (ADC, por sus siglas en inglés). Estas imágenes ADC son calculadas a partir de dos secuencias DWI y permiten calcular la magnitud de la difusión del agua en los tejidos. Para llevar a cabo una caracterización del tejido infartado, es necesario analizar conjuntamente el DWI B-1000 y ADC. En el DWI típicamente se visualizan regiones más grandes de la lesión y también lesiones que pueden corresponder a patologías pasadas. Contrario, los mapas de ADC son restrictivos en cuanto al tiempo de las lesiones, siendo su magnitud, una variable importante para evitar falsos positivos, observados en DWI. También, en el DWI se pueden presentar artefactos, propios de la resonancia, por ejemplo como el fenómeno *T2 shine-through* que ocurren en las imágenes B-1000. En la figura 1 se pueden observar ejemplos de imágenes tomadas desde secuencias DWI y mapas ADC.

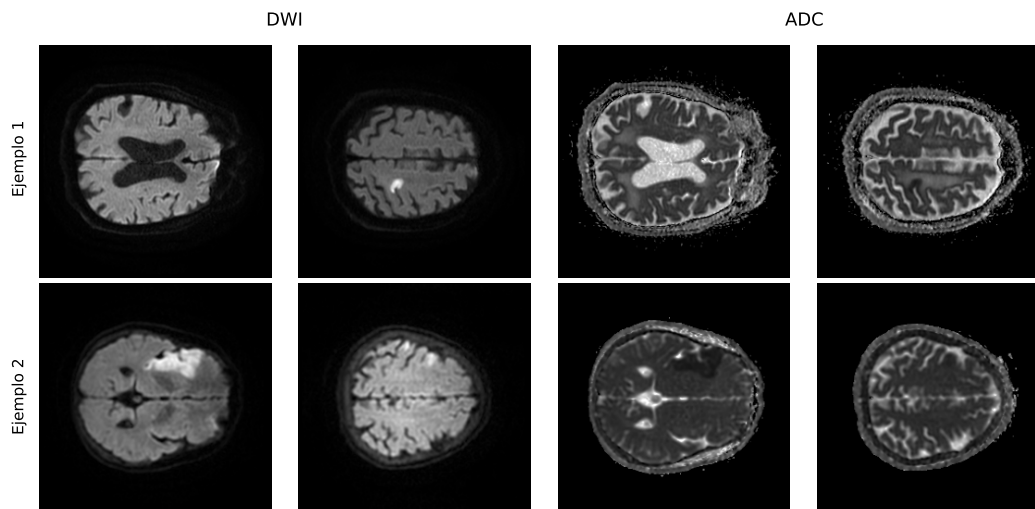
1.1.2. Análisis de variables Clínicas

Tras activar la alarma de un código ACV, es fundamental realizar simultáneamente estudios imagenológicos y registrar variables de interés clínico. En este protocolo, las variables clínicas más relevantes son el tiempo transcurrido desde el inicio de los síntomas del ACV, la cuantificación de la afectación neurológica y la identificación de los factores de riesgo del paciente. A continuación, se explica la utilidad de cada una de estas variables en el diagnóstico y determinación del tratamiento.

Escalas de afectación: Las escalas de afectación son herramientas utilizadas en el

¹⁴ Eloy MARTINEZ-HERAS et al. "Diffusion-weighted imaging: recent advances and applications". In: *Seminars in Ultrasound, CT and MRI*. Vol. 42. 5. Elsevier. 2021, pp. 490–506.

Figura 1. Imágenes de Resonancia Magnética de dos pacientes con ACV isquémico. Las imágenes de DWI muestran la lesión como una zona hiperintensa y en las imágenes de ADC la lesión se presenta como una zona hipointensa.



campo médico para evaluar el grado de discapacidad o dependencia en la vida diaria de las personas que han sufrido un ACV. Las principales escalas son:

- **NIHSS:** Esta escala permite cuantificar el daño neurológico del paciente teniendo en cuenta 11 aspectos, entre los que se encuentra el nivel de conciencia, el lenguaje y el movimiento de las extremidades. La puntuación se encuentra en un rango de [0 – 42] y se establece a partir de un cuestionario relacionado a la capacidad para realizar tareas de movimiento y actividades básicas como lectura, escritura, entre otras. Estas observaciones permiten asignar puntuaciones para cada aspecto y de esta forma establecer el déficit neurológico global del paciente, donde una puntuación más alta indica una mayor afectación neurológica ¹⁵. En la Tabla 1 se detallan las clasificaciones de severidad para distintos intervalos de NIHSS.

¹⁵ Eva A MISTRY et al. “National Institutes of Health Stroke Scale as an outcome in stroke research: value of ANCOVA over analyzing change from baseline”. In: *Stroke* 53.4 (2022), e150–e155.

Tabla 1. Intervalos escala NIHSS junto al grado de severidad

Puntuación NIHSS	Severidad
0	Sin síntomas
1-4	ACV leve
5-15	ACV moderado
16-20	ACV moderado a grave
21-42	ACV grave

- **mRS:** La escala mRS (*modified Rankin Scale*) es la medida más utilizada en la atención clínica posterior a un tratamiento, permitiendo cuantificar el grado de discapacidad a nivel global al incorporar factores médicos y sociales. La evaluación consiste en una serie de preguntas y análisis de las actividades diarias del paciente. También en esta escala se mide su capacidad para realizar estas tareas, donde los diferentes déficits físicos y neurológicos se representan en una única puntuación que se encuentra en un rango de [0–6]. En este caso, una mayor puntuación indica una mayor limitación física del paciente ¹⁶. Generalmente, un mRS de 2 o menos se considera como una evolución favorable, mientras que un valor superior a 2 indica una evolución desfavorable. En la Tabla 2 se pueden observar las distintas puntuaciones de mRS y su descripción.

Tabla 2. Escala mRS y su relación respecto a la condición física del paciente

Puntuación mRS	Descripción
0	Sin síntomas
1	Sin discapacidad significativa
2	Discapacidad leve
3	Discapacidad moderada
4	Discapacidad moderadamente grave
5	Discapacidad grave
6	Muerte

¹⁶ Jeffrey L SAVER et al. "Standardized nomenclature for modified rankin scale global disability outcomes: consensus recommendations from stroke therapy academic industry roundtable XI". In: *Stroke* 52.9 (2021), pp. 3054–3062.

1.1.3. Tratamiento del Accidente Cerebrovascular Isquémico

La administración efectiva de tratamientos adecuados, dependiendo de la naturaleza del ACV, puede prevenir complicaciones adicionales y facilitar una mejor recuperación ². Sin embargo, la determinación del tratamiento requiere de un análisis multifactorial del paciente con ACV isquémico. Actualmente, diversas guías clínicas ofrecen recomendaciones específicas para el manejo de pacientes con ACV, basadas en las características identificadas en los estudios radiológicos y en las variables clínicas pertinentes ¹⁷. A continuación, se describen los posibles tratamientos para un paciente que sufrió un ACV isquémico, junto con algunas directrices y variables según la Asociación Americana del Corazón.

- **Trombectomía mecánica**

La trombectomía mecánica (TM) es el tratamiento estándar para el ACV isquémico agudo causado por la oclusión de grandes vasos de la circulación anterior, dentro de las primeras seis horas posteriores al inicio de los síntomas. En casos seleccionados, este procedimiento puede realizarse en una ventana temporal extendida de hasta 24 horas ¹⁸. La TM se basa en la extracción del coágulo mediante el uso de *recuperadores de stent*, que se insertan y guían hasta la arteria afectada en el cerebro. Una vez posicionados correctamente, estos dispositivos capturan y extraen el coágulo, restableciendo el flujo sanguíneo y minimizando el daño cerebral ¹⁹. Los criterios de elegibilidad para este tratamiento incluyen que el paciente sea mayor

¹⁷ William J POWERS et al. "Guidelines for the early management of patients with acute ischemic stroke: 2019 update to the 2018 guidelines for the early management of acute ischemic stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association". In: *Stroke* (2019).

¹⁸ Wade S SMITH et al. "Mechanical thrombectomy for acute ischemic stroke: final results of the Multi MERCI trial". In: *Stroke* 39.4 (2008), pp. 1205–1212.

¹⁹ Jeffrey L SAVER et al. "Thrombectomy for distal, medium vessel occlusions: a consensus statement on present knowledge and promising directions". In: *Stroke* 51.9 (2020), pp. 2872–2884.

de 18 años, tenga una puntuación en la escala modificada de Rankin (mRS) entre 0 y 1, y una puntuación en la escala NIHSS de 6 o más ¹⁷. Asimismo, es crucial que el volumen del infarto cerebral sea menor a 70 mL, el volumen de la penumbra isquémica exceda los 15 mL, y la relación entre el volumen de la penumbra y el núcleo del infarto sea superior a 1.8.

- **Trombólisis farmacológica**

La trombólisis farmacológica (TF) emplea fármacos derivados de muestras biológicas, como la *uroquinasa* y la *desmoteplasa*, o producidos de manera sintética, como el activador tisular del plasminógeno recombinante (rt-PA) o la pro-uroquinasa recombinante, con el objetivo de disolver el coágulo y restaurar el flujo sanguíneo ^{20,21}. Los beneficios de este tratamiento están asociados a su aplicación temprana, así como de factores como la edad del paciente. El fármaco más comúnmente utilizado en TF es la alteplasa intravenosa (IV), cuya aplicación se recomienda en un rango de 3 a 4.5 horas desde la aparición de los síntomas. Este tratamiento está indicado para pacientes menores de 80 años, con una puntuación en la escala NIHSS menor o igual a 25, y cuya lesión no abarque más de un tercio del territorio de la arteria cerebral media.

- **Tratamiento Conservador**

El tratamiento conservador (TCon) se recomienda principalmente para pacientes con ACV leve o aquellos que presentan contraindicaciones para el uso de terapias trombolíticas. Esta es una estrategia no invasiva que se basa en modificaciones en

²⁰ Joanna M WARDLAW et al. "Thrombolysis for acute ischaemic stroke". In: *Cochrane database of systematic reviews* 7 (2014).

²¹ Martin HLAVICA et al. "Pharmacological recanalization therapy in acute ischemic stroke—evolution, current state and perspectives of intravenous and intra-arterial thrombolysis". In: *Journal of Neuroradiology* 42.1 (2015), pp. 30–46.

el estilo de vida, la implementación de fisioterapia y otras intervenciones complementarias para gestionar los síntomas del ACV y prevenir complicaciones futuras.

1.2. PROCESAMIENTO DE VARIABLES VISUALES Y TEXTUALES

1.2.1. Representaciones visuales basadas en atención

Generalmente, las arquitecturas convolucionales (CNNs) permiten capturar relaciones locales entre características logrando una robustez en la translación y rotación de las imágenes ²². No obstante, estas representaciones no logran capturar dependencias no-locales, que pueden robustecer la caracterización de la lesión isquémica. Los mecanismos de atención, podrían ponderar las características más importantes de la secuencia. Este concepto se empleó en primer lugar en el procesamiento de lenguaje natural (NLP, por sus siglas en inglés) ^{23,24} y más adelante se fue extendiendo a tareas de visión por computador ²⁵.

El mecanismo de auto atención consta de tres ramas codificadas desde la misma entrada, siendo el *key* (\mathbf{K}), *query* (\mathbf{Q}) y *value* (\mathbf{V}). Estas matrices están proyectadas por convoluciones 2D o 1×1 por medio de pesos \mathbf{W}^K , \mathbf{W}^Q , \mathbf{W}^V los cuales codifican la información espacial relevante, por ejemplo, de las lesiones isquémicas. A partir del *key* ($\mathbf{K} = \mathbf{XW}^K$) y el *query* ($\mathbf{Q} = \mathbf{XW}^Q$), se calcula la matriz de pesos de atención entre múltiples píxeles

²² Wangmeng ZUO et al. "Learning iteration-wise generalized shrinkage–thresholding operators for blind deconvolution". In: *IEEE Transactions on Image Processing* 25.4 (2016), pp. 1751–1764.

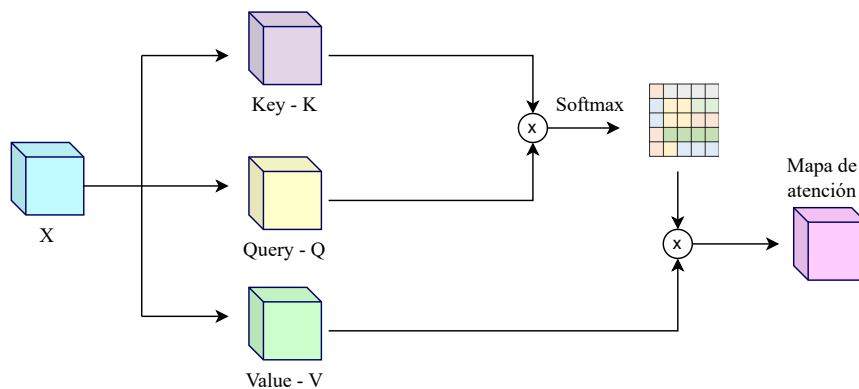
²³ Dzmitry BAHDANAU; Kyunghyun CHO, and Yoshua BENGIO. "Neural machine translation by jointly learning to align and translate". In: *arXiv preprint arXiv:1409.0473* (2014).

²⁴ Minh-Thang LUONG; Hieu PHAM, and Christopher D MANNING. "Effective approaches to attention-based neural machine translation". In: *arXiv preprint arXiv:1508.04025* (2015).

²⁵ Long CHEN et al. "Sca-cnn: Spatial and channel-wise attention in convolutional networks for image captioning". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 5659–5667.

$S_{i,j} = \mathbf{KQ}^T$ que permite realizar una ponderación entre las características, logrando capturar dependencias no-locales entre las filas de cada rama. Para ponderar la contribución de la matriz de pesos de atención en términos de probabilidades, se calcula la función *softmax*. Por último, esta matriz S se multiplica punto a punto con el *value* ($\mathbf{V} = \mathbf{X} \odot \mathbf{W}^V$) para así obtener el mapa de atención ²⁶.

Figura 2. Mecanismo de auto atención para la extracción de características no locales. En este caso X representa una imagen de MRI a partir de la cual se puede obtener una caracterización de la lesión isquémica.



1.2.2. Procesamiento de datos tabulares

Para problemas que incluyen el procesamiento de datos tabulares se han abordado principalmente métodos de aprendizaje de máquina ²⁷. Estos algoritmos son particularmente adecuados para analizar datos estructurados como variables clínicas, debido a su capacidad para manejar múltiples tipos de características y su robustez frente a la variabilidad en los datos. A continuación se resume los modos de representación de estos datos.

²⁶ Zhaoyang NIU; Guoqiang ZHONG, and Hui YU. "A review on the attention mechanism of deep learning". In: *Neurocomputing* 452 (2021), pp. 48–62.

²⁷ Vadim BORISOV et al. "Deep neural networks and tabular data: A survey". In: *IEEE transactions on neural networks and learning systems* (2022).

- **Información tabular** En muchos de los contextos, el mapeo se hace en crudo, proyectando la información tabular como las características que representan cada entrada. Así, luego se entrenan y ajustan clasificadores. Este método es utilizado en casos con un número de columnas limitado.
- **One-hot-encoding** Es una técnica para representar variables categóricas, en la que cada categoría se codifica como un vector binario de dimensión k , siendo k el número total de categorías posibles. Si una variable categórica $x \in \{c_1, c_2, \dots, c_k\}$, entonces su codificación one-hot es un vector $\mathbf{v} \in \{0, 1\}^k$, donde la posición i correspondiente a la categoría c_i toma el valor 1, y el resto son 0.
- **Word-embeddings** En contextos donde se incorporan variables textuales, los word-embeddings permiten mapear palabras a vectores, preservando relaciones semánticas entre ellas. Estos vectores pueden ser aprendidos durante el entrenamiento o preentrenados (Word2Vec o GloVe), y son especialmente útiles para representar descripciones en lenguaje natural.

A continuación se presentan algunos de los algoritmos más utilizados para la clasificación de estos datos tabulares.

Random Forest El método de Random Forest se utiliza en tareas de clasificación o regresión y consiste en un ensamble de múltiples árboles de decisión. Para la clasificación, se establecen k árboles, cada uno generado a partir de un vector aleatorio Θ_k , independiente de los anteriores ($\Theta_1, \dots, \Theta_{k-1}$), pero siguiendo la misma distribución. Cada árbol se construye mediante la selección aleatoria uniforme de características, aprendiendo un umbral por característica que forma un nodo y particiona el espacio de características. Este proceso se optimiza utilizando métricas como la entropía para maximizar la ganancia de información. Así, cada árbol actúa como un clasificador $h(x, \Theta_k)$, donde x es el vector de entrada, emitiendo un voto independiente y unitario. Con el ensamble de un

número significativo de árboles $\{h(x, \Theta_k), k = 1, \dots\}$, se determina la clasificación final por mayoría de votos. Según la ley de los grandes números, la precisión del método mejora con el aumento del número de árboles ²⁸.

XGBoost El método XGBoost (Extreme Gradient Boosting) implementa un ensamble de árboles de decisión. Inicialmente, se construye un solo árbol de decisión que se ajusta a los datos de entrenamiento mediante optimización de gradiente. Posteriormente, se añaden más árboles de manera iterativa y secuencial, cada uno corrigiendo los errores cometidos por los árboles previos. Mediante este refuerzo por gradiente, cada nuevo árbol agregado busca explicar mejor las variaciones que los árboles anteriores no pudieron, sumando sus predicciones para la estimación final ^{29,30}.

1.2.3. Metodologías multimodales

El ACV isquémico es dependiente de un análisis multifactorial para su correcto diagnóstico y tratamiento involucrando tanto secuencias radiológicas, como variables clínicas relacionadas al déficit neurológico del paciente. Por tanto, se considera necesario explorar la contribución entre los dos dominios, y proyectarse en un único espacio embebido. De hecho, el análisis multimodal permite realizar esta correspondencia entre diferentes fuentes de entrada, ya sean visuales o de texto, brindando una integración, separabilidad y discriminación entre los diferentes tratamientos realizados.

A continuación exploramos los mecanismos básicos para hacer análisis multimodal. Dado

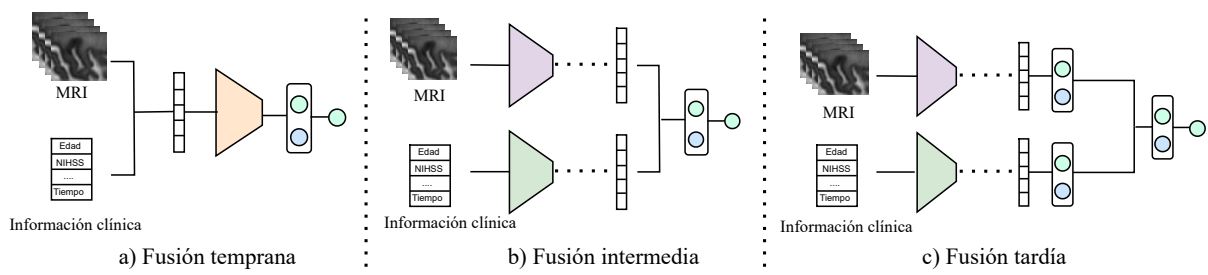
²⁸ Leo BREIMAN. "Random forests". In: *Machine learning* 45 (2001), pp. 5–32.

²⁹ Tianqi CHEN and Carlos GUESTRIN. "Xgboost: A scalable tree boosting system". In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 2016, pp. 785–794.

³⁰ Joos KORSTANJE. *Advanced forecasting with Python*. Springer, 2021.

un conjunto de datos N provenientes de un mismo paciente, con secuencias provenientes de dos dominios diferentes $X_i = \{x_i^1, \dots, x_i^N\}$ y $X_t = \{x_t^1, \dots, x_t^N\}$, se tiene un vector de características n -dimensionales para las modalidades i y t respectivamente. Además de esto, se tiene un conjunto de etiquetas $Y_n = \{y_1, \dots, y_N\}$ en donde el modelo multimodal entrena los conjuntos X_i, X_t al mismo conjunto categórico Y . Para realizar esta integración, se realizan diferentes estrategias de fusión tanto de concatenación, como adición de características. La concatenación significa que las características de ambas modalidades se unen y se introducen directamente en la siguiente capa de la red mientras que en la adición se realiza una suma elemento a elemento de cada modalidad ³¹. Por otro lado, dependiendo de cuándo sucede la fusión se encuentran estrategias como la fusión temprana, la fusión intermedia y la fusión tardía ^{32,33}.

Figura 3. Estrategias de fusión



- **Fusión temprana:** Consiste en realizar la fusión inmediatamente después de la ex-

³¹ Duoyi ZHANG; Richi NAYAK, and Md Abul BASHAR. “Exploring fusion strategies in deep learning models for multi-modal classification”. In: *Australasian Conference on Data Mining*. Springer. 2021, pp. 102–117.

³² Fan YANG; Bo NING, and Huaiqing LI. “An overview of multimodal fusion learning”. In: *International Conference on Mobile Computing, Applications, and Services*. Springer. 2022, pp. 259–268.

³³ Said Yacine BOULAHIA et al. “Early, intermediate and late fusion strategies for robust deep learning-based multimodal action recognition”. In: *Machine Vision and Applications* 32.6 (2021), p. 121.

tracción de características de cada modalidad X_i , X_t o incluso a nivel de los datos, normalmente mediante una operación de unión de las características, como se observa en la Figura 3a. Esta fusión permite aprovechar la correlación entre los elementos de bajo nivel de cada modalidad, entrenando las características de manera conjunta, lo cual permite encontrar relaciones inherentes entre los hallazgos radiológicos y las variables clínicas.

- **Fusión intermedia:** En el caso de la fusión intermedia, la integración de características X_i , X_t se realiza antes de la clasificación (Figura 3b), de esta forma, durante el entrenamiento se logran extraer las características más relevantes de cada modalidad que luego son integradas. Esto permite, por un lado, obtener de forma independiente las características más representativas de la lesión y por otro lado, las características textuales.
- **Fusión tardía:** La fusión tardía implica entrenar un modelo para cada modalidad X_i , X_t y fusionar los resultados de cada una de las diferentes fuentes de información (Figura 3c). Esto combina las decisiones de cada clasificador y produce una predicción final, permitiendo fusionar de manera directa las contribuciones de cada modalidad.

1.3. MÉTODOS COMPUTACIONALES PARA LA PREDICCIÓN DEL TRATAMIENTO DEL ACV ISQUÉMICO

En el estado del arte, se han propuesto diversas metodologías que tienen en cuenta variables clínicas e imagenológicas para tareas de predicción relacionadas al ACV isquémico. Esto se realiza, tanto con análisis unimodales y multimodales, dependiendo de la predicción de diferentes escalas de afectación dado el post-tratamiento (mRS, NIHSS) o incluso del tratamiento. Inicialmente, se han llevado a cabo análisis unimodales a partir del análisis de imágenes radiológicas o variables clínicas como edad, diabetes, comorbilidades y

tratamientos. Por ejemplo, Fernandez-Lozano *et al.* realizaron la predicción del pronóstico para pacientes con ACV isquémico y hemorragia intracerebral empleando una estrategia de bosque aleatorio. Este modelo utiliza datos clínicos como NIHSS, mRS previo, tiempo desde el inicio del ACV, deterioro neurológico, entre otros ³⁴. Sin embargo, estos enfoques no emplean características respecto a hallazgos radiológicos que brinden información respecto a la lesión. Por otro lado, en el trabajo de Hatami *et al.* propusieron a partir de diferentes modalidades de MRI realizar la predicción del mRS . En este caso, se emplean cinco auto-encoders para aprender características unimodales de cinco modalidades de MRI (ADC, CBF, CBV, DWI Y Tmax) logrando caracterizar la distribución de cada neuroimagen. Luego de ello, se usa un auto-encoder que permite integrar las características unimodales en una única representación embebida ⁶. No obstante, no se considera información clínica que pueda contribuir en la predicción. Por lo cual, en general las estrategias unimodales están sesgadas por la falta de información complementaria.

Alternativamente, se han propuesto modelos multimodales que aprovechan la información complementaria de las imágenes médicas y la información clínica donde dependiendo del post-tratamiento, se calcula la predicción de diversas escalas de afectación tales como el mRS90 y el NIHSS. Por ejemplo, Jo *et al.* combinaron información derivada de imágenes ADC y DWI, e información de historias clínicas tales como la edad, el NIHSS al ingreso y el deterioro neurológico temprano. El modelo de imagen obtiene la segmentación de la lesión a partir una red UNet y la concatena junto a las respectivas imágenes de ADC y DWI. Mediante una 3D DenseNet, obtiene la probabilidad de un pronóstico favorable o desfavorable. Esta probabilidad se fusiona con las variables clínicas y se realiza la predicción final del mRS mediante clasificadores clásicos de aprendizaje automático ¹⁰. No obstante, este enfoque puede limitar la contribución de la caracterización de la lesión de-

³⁴ Carlos FERNANDEZ-LOZANO *et al.* "Random forest-based prediction of stroke outcome". In: *Scientific reports* 11.1 (2021), p. 10071.

bido a la forma en la que se integra la información, dejando únicamente la contribución del modelo de imagen como una probabilidad junto a todas las variables clínicas. Por ejemplo, Bacchi *et al.* sugieren una predicción de los resultados de NIHSS o mRS90 basada en imágenes NCCT por medio de una CNN 3D e información clínica de pacientes las cuales son caracterizadas por medio de una DNN para pacientes con ACV tratados con trombólisis farmacológica intra-arterial. En este caso, realizan la integración de las dos modalidades por medio de una capa densamente conectada ⁹. Adicionalmente, Brugnara *et al.* emplearon un método de *Gradient Boosting* para predecir un pronóstico, en términos del mRS90 tras el tratamiento tromboectomía endovascular (EVT) en pacientes con ACV. La predicción se realizó a partir de variables clínicas como edad, mRS de ingreso, NIHSS e información establecida a partir del análisis imagenológico como el volumen de la lesión, el lugar de la oclusión y la escala ASPECT. No obstante, se emplearon datos tabulares como representación de las características de la lesión, por lo que no se tiene en cuenta información visual que permita brindar una mejor caracterización ³⁵. Asimismo, Samak *et al.* introdujeron una metodología para predecir los resultados del mRS tras aplicar una TM o TF, el modelo emplea imágenes NCCT de las cuales se extraen características a partir de una arquitectura *transformer* y combina características clínicas (género, edad, hipertensión, nivel de glucosa, entre otras) utilizando una red densa. La integración de cada modalidad se consigue mediante un esquema de fusión tardía ³⁶.

De forma alternativa, existen metodologías que agregando el tratamiento en el entrenamiento, predicen la probabilidad de presentar una determinada puntuación NIHSS o mRS de salida. Por ejemplo, Samak *et al.* realizó una estimación del pronóstico de pa-

³⁵ Gianluca BRUGNARA et al. "Multimodal predictive modeling of endovascular treatment outcome for acute ischemic stroke using machine-learning". In: *Stroke* 51.12 (2020), pp. 3541–3551.

³⁶ Zeynel A SAMAK; Philip CLATWORTHY, and Majid MIRMEHDI. "Transop: Transformer-based multimodal classification for stroke treatment outcome prediction". In: *arXiv preprint arXiv:2301.10829* (2023).

cientes de ACV después de una EVT a partir de la predicción de la escala mRS. En este escenario, las imágenes NCCT pasan a través de convoluciones 3D y esta información se integra mediante la concatenación con un vector codificado que contiene, la representación de la presencia o ausencia de EVT o información clínica como edad, sexo, NIHSS, ASPECTS. Sin embargo, este modelo puede verse limitado debido a la gran cantidad y el tipo de variables clínicas que emplea para ajustar el modelo, las cuales pueden no estar disponibles en todos los escenarios clínicos ⁸. Por otro lado, Liu *et al.* introdujeron un ensamblaje de modelos. El modelo de texto, incorpora datos tabulares que abarcan el historial médico del paciente, incluidos mRS y NIHSS de ingreso, incidencias de ACV previas y tratamientos administrados (tPA intravenoso y EVT) el cual genera una predicción al emplear una máquina de soporte vectorial (SVM), mientras que el modelo de visión emplea imágenes MRI (B0 y DWI) que pasan a través de una ResNet3D. Cada modelo genera una predicción y a partir de una técnica de ensamble se obtiene una predicción final ⁷. Pese a que se emplea información multimodal la técnica de ensamble puede no ser la mejor para realizar la predicción debido a que no se integran las características de cada modalidad sino que se realizan predicciones independientes.

Por otro lado, han emergido estrategias en el estado del arte que caracterizan la lesión afectada y emplean las variables clínicas para así realizar la predicción del tratamiento. En este aspecto, Ma *et al.* propusieron un modelo para la predicción del tratamiento que diferencia entre el tratamiento conservador y la trombectomía mecánica ¹¹. Esta metodología integra imágenes de CT e información clínica que incluye variables como el tabaquismo, la hipertensión, hiperlipidemia, entre otras. Para obtener características latentes representativas de cada modalidad se emplea la ResNet34 (en términos imagenológicos) y una DNN clásica (variables clínicas). Estas representaciones se integran mediante un módulo de distribución variacional para estimar probabilidades de tratamiento. Aunque este enfoque ha mostrado resultados prometedores, su aplicación se limita a ACV hemorrágicos

y no aborda múltiples opciones de tratamiento, lo que incrementa la complejidad del modelo al necesitar estimar distribuciones separadas para cada uno. Por otro lado, Chourib *et al.* implementaron un modelo para predecir siete clases de tratamiento basadas en el dispositivo empleado para la extracción del coágulo (*espiral, espiral y stent, espiral y direccionador de flujo, solo direccionador de flujo, stent, disruptor de flujo, disruptor de flujo y espiral*)³⁷. Esto se lleva a cabo mediante métodos de selección de características, como el método Chi-Cuadrado, Eliminación Recursiva de Características (RFE) o Random Forest. Las características calculadas son clasificadas por medio de algoritmos de aprendizaje de máquina tales como *K-Nearest Neighbor, Naive Bayesian classifier* o árboles de decisión. Sin embargo, este enfoque se limita al uso exclusivo de variables clínicas, lo cual restringe la extracción de características directamente relacionadas con la lesión. Además, el estudio se centra únicamente en la terapia endovascular y presenta un desbalance considerable entre las clases, lo que puede introducir sesgos y afectar la robustez del modelo.

³⁷ I CHOURIB *et al.* “Stroke treatment prediction using features selection methods and machine learning classifiers”. In: *IRBM* 43.6 (2022), pp. 678–686.

2. PROBLEMA DE INVESTIGACIÓN

El accidente cerebrovascular isquémico es una enfermedad cuyo diagnóstico y tratamiento oportuno es indispensable para reducir los daños neurológicos y el grado de discapacidad del paciente ². La determinación del diagnóstico y tratamiento se realiza a partir del análisis conjunto de secuencias radiológicas y variables clínicas que permiten realizar la caracterización de la lesión y conocer el estado físico y neurológico del paciente ¹⁷. De hecho, este proceso es dependiente de la experticia del radiólogo, lo cual puede inducir a una mayor afectación cerebral.

Actualmente, diversas propuestas computacionales se han enfocado en la estimación de la recuperación post-tratamiento de pacientes supervivientes de un ACV por medio de la predicción de escalas de afectación como el mRS. No obstante, en algunos casos se consideró únicamente las secuencias diagnósticas o las variables clínicas como datos de entrada, excluyendo información complementaria. Por otra parte, los enfoques multimodales presentan limitaciones en la integración de características específicas de cada modalidad. Además, otras aproximaciones reportadas en el estado del arte solo se enfocan en el ACV hemorrágico, el cual tiene un protocolo de atención diferente y una complejidad computacional mayor dependiendo del número de tratamientos. En la literatura se carecen de enfoques multimodales para la predicción del tratamiento en el ACV isquémico.

2.1. PREGUNTA DE INVESTIGACIÓN

¿Cómo implementar una arquitectura de aprendizaje profundo para el análisis de secuencias radiológicas y variables clínicas para determinar el tratamiento en pacientes con ACV isquémico?

3. OBJETIVOS

3.1. OBJETIVO GENERAL

Implementar una arquitectura multimodal que integre secuencias radiológicas y variables clínicas para realizar la predicción del tratamiento en pacientes con accidente cerebrovascular isquémico.

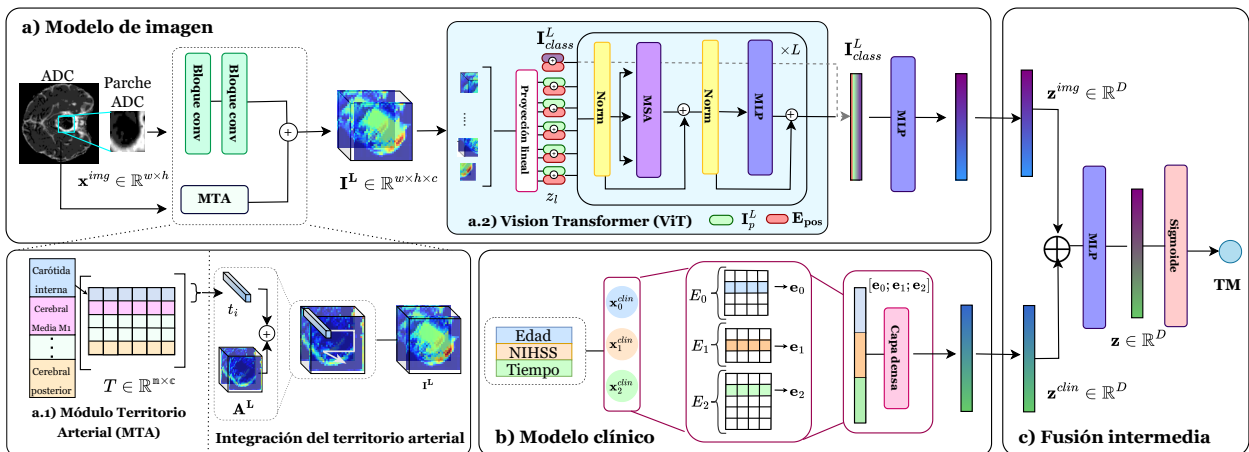
3.2. OBJETIVOS ESPECÍFICOS

- Seleccionar un conjunto de datos que cuente con secuencias radiológicas y variables clínicas de pacientes que sufrieron un ACV isquémico
- Implementar una representación profunda para la predicción del tratamiento desde hallazgos radiológicos.
- Implementar una estrategia para la caracterización de variables clínicas como indicadores del tratamiento del paciente.
- Integrar las representaciones radiológicas y variables clínicas para la predicción del tratamiento de pacientes que sufrieron un ACV isquémico.
- Evaluar el desempeño de los modelos propuestos en la tarea de predicción del tratamiento empleando métricas de clasificación.

4. MÉTODO PROPUESTO

En este trabajo se implementó una arquitectura multimodal para la predicción del tratamiento en pacientes con ACV isquémico (ver Figura 4). La arquitectura incluye un bloque de procesamiento y codificación de variables clínicas y un bloque de procesamiento visual, encargado de codificar hallazgos radiológicos sobre parches de mapas ADC. La representación imagenológica se complementó con embebidos que proporcionan información espacial sobre el territorio arterial afectado. Posteriormente, se empleó una estrategia de fusión intermedia para integrar ambos descriptores, permitiendo realizar un análisis conjunto de las variables y predecir el tratamiento más adecuado. En este trabajo se consideraron los siguientes tres posibles tratamientos: trombectomía mecánica, trombólisis farmacológica o tratamiento conservador.

Figura 4. Método multimodal. a) Modelo de imagen que extrae las características de la lesión y adiciona información del territorio arterial afectado. b) Modelo clínico que obtiene una codificación de las variables clínicas. c) Fusión intermedia de los descriptores para predecir el tratamiento.



4.1. MODELO PARA LA EXTRACCIÓN DE CARACTERÍSTICAS IMAGENOLÓGICAS DE LA LESIÓN

Con el objetivo de caracterizar los hallazgos radiológicos de ACV, observados en los mapas ADC, en este trabajo se desarrolló un módulo inicial, encargado de extraer estas características desde la región centrada en la observación de la lesión. En este módulo, se utiliza como entrada una región (parche de la imagen) demarcada alrededor de la lesión $\mathbf{x}^{img} \in \mathbb{R}^{w \times h}$, donde w y h son las dimensiones espaciales de ancho y alto, respectivamente. Entonces, este parche \mathbf{x}^{img} es proyectado sobre una representación convolucional, obteniendo un banco de características \mathbf{A}^L .

Por otra parte, la localización donde se ubica la lesión (territorio arterial afectado), juega un papel importante en la caracterización de la lesión y en la definición del tratamiento del accidente cerebrovascular. Por esta razón, es necesario considerar dicha información, integrándola con las características visuales obtenidas \mathbf{A}^L , permitiendo así, brindar contexto en la representación de la lesión ³⁸. Esta integración se hace en el módulo de territorio arterial (MTA), cómo se ilustra en la Figura 2a.1. Particularmente, a cada representación visual se le asocia un vector embebido de dimensión c mediante una capa que toma como entrada los n territorios, cada uno asociado a un índice $i \in \{0, 1, \dots, n - 1\}$, y lo representa como un vector contenido en una matriz de pesos $T \in \mathbb{R}^{n \times c}$. En esta matriz, cada fila $t_i = T[i, :] \in \mathbb{R}^c$ corresponde a la representación embebida del territorio i . Durante el entrenamiento, los valores de esta matriz se ajustan, obteniendo un embebido t_i para el territorio afectado, el cual se fusiona con la representación imagenológica, sumándolo a través de los c canales de \mathbf{A}^L . Esta integración $\mathbf{I}^L \in \mathbb{R}^{w \times h \times c}$ permite enriquecer la representación visual de la lesión con información contextual sobre su localización.

La representación integrada de la lesión \mathbf{I}^L es entonces proyectada a un *Vision Trans-*

³⁸ Chin-Fu LIU et al. "Digital 3D brain MRI arterial territories atlas". In: *Scientific Data* 10.1 (2023), p. 74.

former (ViT, por sus siglas en inglés) ³⁹, para aprovechar el contexto no-local que pueda estar contenido en el hallazgo acotado en la región. Inicialmente, este modelo transforma a \mathbf{I}^L en sub-parches no superpuestos de tamaño $P \times P$, los cuales se aplanan y representan como $\mathbf{I}_p^L \in \mathbb{R}^{N \times (P^2c)}$, donde $N = \frac{h}{P} \times \frac{w}{P}$ corresponde al número total de subparches generados. Cada sub-parche se proyecta a una capa lineal ($\mathbf{z} = \mathbf{I}_p^L E + b$), resultando en un nuevo vector de dimensión D . Dado que la imagen \mathbf{I}^L se divide en parches, para preservar la información espacial se añaden embebidos posicionales ($\mathbf{E}_{\text{pos}} \in \mathbb{R}^{(N+1) \times D}$) a cada parche embebido. Así, la representación de entrada al codificador del ViT es $\mathbf{z}_0 = [\mathbf{I}_{\text{class}}^L; \mathbf{I}_p^L E; \mathbf{I}_p^L E; \dots; \mathbf{I}_p^L E] + \mathbf{E}_{\text{pos}}$. Adicionalmente, se incorpora un token de clasificación ($\mathbf{I}_{\text{class}}^L$), que actúa como la representación visual global, que resume la información de todos los parches, y resulta en una representación esencial para la tarea de discriminación entre los tratamientos. Este vector se ingresa con valores aleatorios en la entrada del modelo y en cada fase de procesamiento de los sub-parches se irá procesando y actualizando con la información global de la imagen.

El codificador consiste en L capas, cada una compuesta por un mecanismo de autoatención multi-cabeza (MSA, por sus siglas en inglés) y un perceptrón multicapa (MLP). Para la autoatención, a partir de la representación \mathbf{z}_l en la capa l , se calculan las matrices *key* ($\mathbf{K}_l = \mathbf{z}_l \mathbf{W}^k$), *query* ($\mathbf{Q}_l = \mathbf{z}_l \mathbf{W}^q$) y *value* ($\mathbf{V}_l = \mathbf{z}_l \mathbf{W}^v$). La atención se calcula como $\text{SA}(\mathbf{Q}_l, \mathbf{K}_l, \mathbf{V}_l) = \text{softmax}\left(\frac{\mathbf{Q}_l \mathbf{K}_l^T}{\sqrt{d_k}}\right) \mathbf{V}_l$, donde $d_k = D/H$. Si hay H cabezas de atención, cada una produce una salida $\text{SA}_{h,l}$ y la salida final de la capa de autoatención se obtiene concatenando y proyectando todas las cabezas: $\text{MSA}_l = \text{Concat}(\text{SA}_{1,l}, \text{SA}_{2,l}, \dots, \text{SA}_{H,l}) \mathbf{W}^o$, donde $\mathbf{W}^o \in \mathbb{R}^{Hd_k \times D}$. Esta salida se suma a la entrada original mediante una conexión residual y se normaliza. Posteriormente, la salida de la autoatención pasa por un perceptrón multicapa (MLP), compuesto por dos capas completamente conectadas con activación no lineal intermedia, $\mathbf{z}'_l = \text{MLP}(\text{MSA}_l) = \sigma(\text{MSA}_l \mathbf{W}_1 + b_1) \mathbf{W}_2 + b_2$. Al igual que en

³⁹ Alexey DOSOVITSKIY et al. "An image is worth 16x16 words: Transformers for image recognition at scale". In: *arXiv preprint arXiv:2010.11929* (2020).

la etapa anterior, se utiliza una conexión residual seguida de normalización. Con la actualización de \mathbf{z}'_i se ponderan todos sus embebidos, incluyendo el $\mathbf{I}_{\text{class}}^L$, el cual engloba la descripción general de la imagen. Después de procesar todas las capas del codificador, se obtiene la representación final del token de clasificación $\mathbf{I}_{\text{class}}^L$, el cual contiene la caracterización de la imagen. Este vector embebido $\mathbf{I}_{\text{class}}^L$ pasa a través de una capa MLP para obtener la representación final $\mathbf{z}^{img} \in \mathbb{R}^D$.

4.2. REPRESENTACIÓN DE VARIABLES CLÍNICAS

Como complemento en el análisis para la predicción del tratamiento en pacientes con ACV, se desarrolló un modelo que representa la información clínica mediante el aprendizaje de vectores embebidos, apilados en una matriz que resume todo el vocabulario definido en la información clínica. Para cada variable clínica x_i^{clin} , se define una matriz de embebidos $\mathbf{E}_i \in \mathbb{R}^{n_i \times d}$, inicializada de forma aleatoria, donde n_i es el número de categorías posibles de dicha variable y d la dimensión del embebido. Esta matriz actúa como un diccionario entrenable donde el valor de la variable $x_i^{clin} \in \{0, 1, 2, \dots, n_i\}$ se utiliza como índice para recuperar su correspondiente representación vectorial, es decir, $\mathbf{e}_i = \mathbf{E}_i[x_i^{clin}]$. Durante el entrenamiento, el vector correspondiente al índice observado se actualiza, permitiendo al modelo ajustar individualmente las representaciones embebidas según su relación para la tarea de predicción. Los embebidos \mathbf{e}_i de todas las variables clínicas se concatenan para formar una representación conjunta $[\mathbf{e}_0; \mathbf{e}_1; \dots; \mathbf{e}_{i-1}] \in \mathbb{R}^D$, donde i es el número de variables consideradas y $D = i \cdot d$. Esta representación se transforma mediante una capa densamente conectada, que produce un descriptor clínico compacto $\mathbf{z}^{clin} \in \mathbb{R}^D$, útil para predecir el tratamiento adecuado en pacientes con ACV.

4.3. PREDICCIÓN DEL TRATAMIENTO A PARTIR DE UNA FUSIÓN INTERMEDIA DE DESCRIPTORES MULTIMODALES

Para determinar el tratamiento más adecuado para un paciente con ACV, se realizó una fusión intermedia de los descriptores que resumen la información imagenológica \mathbf{z}^{img} y clínica \mathbf{z}^{clin} . Específicamente, esta fusión se realiza concatenando ambos descriptores para formar un único embebido: $\mathbf{z} = [\mathbf{z}^{img}; \mathbf{z}^{clin}]$ donde $\mathbf{z} \in \mathbb{R}^{2D}$. Este vector de características fusionadas se procesa mediante un perceptrón multicapa (MLP), compuesto por una serie de capas densamente conectadas con activaciones ReLU intermedias. La salida final se obtiene aplicando una función sigmoide sobre la proyección del embebido en la última capa del MLP: $\mathbf{h} = \sigma(\mathbf{W}_2 \text{ReLU}(\mathbf{W}_1 \mathbf{z} + \mathbf{b}_1) + \mathbf{b}_2)$ donde σ representa la función sigmoide y $\mathbf{W}_1, \mathbf{W}_2$ son matrices de pesos ajustables durante el entrenamiento. Esta transformación permite obtener la probabilidad asociada al tratamiento. Para el aprendizaje de la predicción del tratamiento, la arquitectura propuesta se entrena en un esquema *end-to-end*. En este caso, teniendo en cuenta el desbalance de clases, como función de costo se utilizó una función Focal Loss (FL) definida como $\text{FL}(p) = -\alpha \cdot (1 - p)^\gamma \cdot \log(p)$, donde $\alpha \in [0, 1]$ es un factor de ponderación que permite balancear la importancia entre clases, $(1 - p)$ se conoce como factor modulador y $\gamma \geq 0$ es un hiperparámetro que controla cuánto se reduce la pérdida para las muestras bien clasificadas. Esto se realiza utilizando las etiquetas correspondientes al tratamiento, definido en el conjunto de datos del análisis. Cabe destacar que estas etiquetas son definidas por un experto, luego de hacer una valoración al paciente.

4.4. DISEÑO EXPERIMENTAL

4.4.1. Datos Para entrenar y validar el método propuesto se empleó un conjunto de datos privado. Para cumplir los criterios de inclusión, los pacientes debían ser mayores de 18 años y no mostrar signos de hemorragia cerebral. Los pacientes que mostraban

indicios de reperfusión parcial no fueron excluidos del estudio. El protocolo experimental del conjunto de datos fue aprobado por los comités de ética *CEINCI-UIS* de la Universidad Industrial de Santander (número de aprobación 4110, otorgado el 10 de febrero de 2023) y *CEI-FOSCAL* de la Fundación Oftalmológica de Santander - FOSCAL (número de aprobación 06075/2022, otorgado el 29 de abril de 2022).

En total, se consideraron 156 pacientes que presentaron síntomas de ACV y fueron tratados entre Octubre de 2021 y Junio de 2023. Para cada paciente se dispone de mapas ADC junto con la respectiva segmentación de la lesión isquémica y la identificación del territorio arterial afectado. La segmentación de lesiones isquémicas fue llevada a cabo por tres neurorradiólogos, cada uno con más de cinco años de experiencia, utilizando el software MRICroGL⁴⁰ tras revisar las historias clínicas al ingreso. En la Tabla 3 se detallan las variables clínicas incluidas en este estudio, junto con sus medidas estadísticas. Además, se contó con información sobre el tratamiento aplicado a cada paciente, clasificado en cuatro categorías: trombólisis farmacológica intravenosa (56 pacientes), trombólisis farmacológica intraarterial (13 pacientes), trombectomía mecánica (32 pacientes) y tratamiento conservador (55 pacientes). No obstante, debido al reducido número de casos tratados con trombólisis farmacológica intraarterial, se decidió agrupar ambos tratamientos farmacológicos (intravenoso e intraarterial) en una única clase.

En cuanto al preprocesamiento y con el fin de aislar el tejido cerebral, se removió el cráneo de cada uno de los mapas ADC, utilizando el software SynthStrip⁴¹. Posteriormente, los mapas ADC sin cráneo fueron normalizados restando la media de sus intensidades y dividiendo por la desviación estándar de las mismas. Además, teniendo como referencia las segmentaciones de las lesiones isquémicas, se ubicó el corte donde se encontraba

⁴⁰ Xiangrui et al. LI. "The first step for neuroimaging data analysis: DICOM to NIfTI conversion". In: *Journal of neuroscience methods* 264 (2016), pp. 47–56.

⁴¹ Andrew HOOPES et al. "SynthStrip: skull-stripping for any brain image". In: *NeuroImage* 260 (2022), p. 119474. DOI: <https://doi.org/10.1016/j.neuroimage.2022.119474>.

Tabla 3. Descripción de las variables clínicas del conjunto de datos. TIS: Tiempo desde el inicio de los síntomas en horas, TAA: territorio arterial afectado.

Variable	Valores
Edad [min-max] (media \pm desv)	[22-95] (71.78 \pm 14.18)
NIHSS [Sin síntomas / ACV leve / ACV moderado / ACV moderado a grave / ACV grave]	[8 / 59 / 70 / 9 / 10]
TIS [min-max] (media \pm desv)	[0.33-192] (15.15 \pm 26.79)
TAA [ICA / MCA / ACA / PCA / Vertebral / Basilar / Ganglio basal / Vaso pequeño]	[17 / 35 / 19 / 6 / 10 / 2 / 13 / 66]

la mayor cantidad de píxeles de lesión y se extrajo un parche (x^{img}) de dimensiones $64 \times 64 \times 1$. Estas dimensiones garantizan que cualquier lesión estuviera contenida en la región.

4.4.2. Configuración de la arquitectura Para el entrenamiento del modelo multi-modal, cada rama presenta unas características y detalles de implementación propias, que garantizan un rendimiento efectivo de la representación. En cuanto al modelo de imagen para la extracción de características de la lesión, el modelo incorpora dos convoluciones espaciales con un kernel 3×3 que se aplican al parche de mapas ADC. A partir de esto, se adiciona la información relacionada con el territorio afectado y se realiza el refinamiento de la lesión, cuyas configuraciones se explican a continuación:

- **Codificación del territorio arterial afectado.** Para cada uno de los territorios se obtuvo un vector embebido con una dimensión de 1×32 y se incorporó una estrategia de fusión temprana con los bloques convolucionales.
- **Módulo de refinamiento basado en atención (ViT).** El ViT recibe imágenes transformadas en parches de 16×16 , los cuales son proyectados a un espacio de dimensión $D = 90$ mediante una capa embedding, junto con codificaciones posicionales. El Módulo de Auto-atención Multi-Cabezas emplea 9 cabezas de atención, seguido de un MLP con dos capas ocultas, activadas mediante la función GELU.

En cuanto a la rama de procesamiento de las variables clínicas, se realizó en primera medida un análisis con XGBoost, donde se identificaron las variables que tienen una mayor contribución en la predicción del tratamiento. Estas variables fueron: edad, NIHSS y el tiempo de inicio de los síntomas. Teniendo en cuenta estas variables como entrada, el modelo clínico genera tres vectores embebidos cada uno de dimensión 1×30 , los cuales se concatenan y pasan a través de una capa densamente conectada para obtener como salida un vector embebido de tamaño $D = 90$ que contiene la representación de las características clínicas. Además, para valores no vistos en el entrenamiento de cada vector embebido, se emplea una interpolación basada en categorías similares para evitar asignaciones aleatorias. Las dimensiones de las matrices de los vectores embebidos que representan cada variable seleccionada se presentan a continuación:

- **Edad.** Esta variable esta definida en un rango entre 22 y 95 años. De acuerdo a los valores definidos en el conjunto de datos se tomaron un total de 100 valores. Entonces la matriz de vectores embebidos tiene una dimensión de 100×30 , donde 100 es el número de edades y 30 el valor del tamaño del embebido.
- **NIHSS.** Esta variable, de acuerdo al conjunto de datos tiene valores entre 0 y 29. Entonces se definió una matriz de 42×30 , donde 42 es el valor máximo que puede tener la escala NIHSS y 30 son las dimensiones del vector embebido. Los valores NIHSS aquí considerados representan una valoración que comprende la usencia de síntomas a un episodio de ACV grave.
- **Tiempo de inicio síntomas.** Esta variable representa el tiempo transcurrido desde el inicio de los síntomas. En este caso se consideraron los tiempos entre 0.3 horas y 192. Entonces, la matriz de representación tiene dimensiones 200×30 .

Para el entrenamiento, se empleó el optimizador Adam con una tasa de aprendizaje de 0.00001 y 200 épocas. Teniendo en cuenta el desbalanceo de clases de tratamiento, se

utilizó una función de pérdida Focal con $\gamma = 2.0$ para reducir la contribución a la pérdida de los casos más sencillos de clasificar y aumentar la penalización para los casos más difíciles de clasificar. En total, se siguieron dos representaciones binarias para la predicción de los tratamientos. La primera distingue entre No Tratamiento vs Tratamiento (NoT vs T), donde la clase NoT corresponde al Tratamiento Conservador, y T agrupa las clases Trombectomía Mecánica (TM) y Trombólisis Farmacológica (TF). La segunda configuración diferencia entre TM vs TF. Para cada configuración se realizó una validación cruzada de 6 folds, utilizando métricas como AUROC, F1 y Exactitud. Además, se ajustó el umbral de clasificación a partir de los datos de validación de cada fold.

5. EVALUACIÓN Y RESULTADOS

Esta sección presenta la evaluación cuantitativa y los resultados obtenidos por el método propuesto para la predicción del tratamiento en pacientes con ACV isquémico. El objetivo es determinar su capacidad para codificar patrones relevantes a partir de datos clínicos e imagenológicos, para así compararla con estrategias de referencia en la literatura. La exposición de los resultados sigue una estructura progresiva. Primero, se analiza el desempeño de modelos unimodales para identificar los componentes más eficaces en la codificación de información clínica e imagenológica. Posteriormente, se evalúan configuraciones multimodales que integran ambos tipos de información, contrastando el enfoque propuesto con variantes basadas en mecanismos convolucionales y de atención. También, se explora un enfoque de ensamblaje secuencial de modelos binarios que permite abordar la predicción del tratamiento como una tarea de clasificación multiclase, demostrando su aplicabilidad y su capacidad para soportar la toma de decisiones clínicas. A continuación se ordenan estos resultados en secciones y se detallan cada uno de los experimentos realizados.

5.1. EVALUACIÓN DE MODELOS UNIMODALES PARA LA PREDICCIÓN DEL TRATAMIENTO

Como primer paso en la validación del método propuesto, se evaluó el desempeño de sus codificadores clínico e imagenológico por separado, con el objetivo de determinar la capacidad individual para predecir el tratamiento en pacientes con ACV isquémico. El codificador clínico fue entrenado utilizando todas las variables disponibles en el conjunto de datos, mientras que el codificador imagenológico fue entrenado sobre mapas de ADC, incorporando además la codificación del territorio arterial afectado. Ambos modelos fueron comparados con enfoques representativos de la literatura: un clasificador Random Forest

para el caso clínico y una red ResNet152 para el caso imagenológico. La comparación se llevó a cabo para las dos configuraciones binarias (NoT vs. T) y (TM vs. TF), permitiendo analizar la capacidad discriminativa de cada modelo frente a decisiones clínicas específicas.

En resumen, los resultados obtenidos en cuanto a la predicción del tratamiento a partir de mapas de ADC se reportan en la Tabla 4. Como se ilustra en la tabla, el codificador imagenológico del método propuesto supera consistentemente a la arquitectura ResNet152 en la mayoría de las métricas evaluadas, a excepción de la especificidad. En la configuración de predicción binaria NoT vs. T, el modelo propuesto alcanza mejoras del 4.8% en AUROC, 6.4% en exactitud, 7.9% en F1 y 13.9% en sensibilidad, pero con una disminución del 7.4% en especificidad. Una tendencia similar se observa en la configuración TM vs. TF, donde las mejoras son aún más pronunciadas: 7% en AUROC, 7% en exactitud, 16.2% en F1 y 25.1% en sensibilidad, aunque acompañadas por una reducción del 31.1% en especificidad. Las diferencias en especificidad para ambas configuraciones sugieren un sesgo del modelo hacia la clasificación de la clase positiva. No obstante, el desempeño del codificador imagenológico propuesto en métricas que resaltan una clasificación más equilibrada como AUROC, exactitud y F1, refuerza su capacidad discriminativa general frente a la ResNet152.

Tabla 4. Comparación del desempeño de modelos imagenológicos bajo configuraciones de predicción binaria. Se contrastan los resultados del codificador imagenológico propuesto frente a una ResNet152 entrenada sobre mapas ADC. Las métricas reportadas corresponden a las configuraciones NoT vs. T (sin tratamiento vs. tratamiento) y TM vs. TF (trombectomía mecánica vs. trombólisis farmacológica). Los mejores resultados se encuentran resaltados en negrilla.

Conf.	Modelo	AUROC (%)	Exactitud (%)	F1 (%)	Sens. (%)	Espec. (%)
NoT vs. T	Propuesto	69.9 ± 9.4	66.0 ± 8.7	73.1 ± 6.6	73.2 ± 13.3	52.6 ± 14.2
	ResNet152	65.1 ± 11.7	59.6 ± 9.1	65.2 ± 9.4	59.3 ± 10.8	60.0 ± 10.3
TM vs. TF	Propuesto	69.9 ± 15.2	63.3 ± 14.4	72.4 ± 11.3	71.2 ± 13.3	46.1 ± 20.4
	ResNet152	62.9 ± 16.1	56.3 ± 16.2	56.2 ± 22.7	46.1 ± 22.0	77.2 ± 21.5

En cuanto a la predicción del tratamiento a partir de variables clínicas, los resultados se

resumen en la Tabla 5. Como se muestra, el Random Forest obtiene un mejor desempeño para la discriminación entre NoT y T, con una diferencia del 13.9% en AUC y una diferencia considerable entre la sensibilidad (42.6%). Por otro lado, el método propuesto muestra una ventaja en la distinción entre TM y TF con un F1 de 77.1%, sin embargo, para los dos enfoques se obtienen valores bajos para la especificidad, lo cual indica que existe un sesgo, esto podría estar relacionado con el desbalanceo de clases. Además se evidencia que las variables por sí solas no logran representaciones compactas que permitan la determinación del tratamiento, mostrando un comportamiento inconsistente entre los enfoques implementados. Dado que en este escenario los modelos evaluados no logran una discriminación robusta, se considera que el enfoque propuesto puede ser ventajoso al integrarse en esquemas multimodales. Su diseño puede dar mayor flexibilidad en la optimización end-to-end de la estrategia multimodal, teniendo en cuenta el ajuste de vectores embebidos para cada variable, lo que puede facilitar el aprendizaje de mejores representaciones para la tarea de predicción del tratamiento.

Tabla 5. Comparación del desempeño de modelos clínicos bajo configuraciones de predicción binaria. Se contrastan los resultados del codificador clínico propuesto frente a un Random Forest. Las métricas reportadas corresponden a las configuraciones NoT vs. T (sin tratamiento vs. tratamiento) y TM vs. TF (trombectomía mecánica vs. trombólisis farmacológica). Los mejores resultados se encuentran resaltados en negrilla.

Conf.	Modelo	AUROC (%)	Exactitud (%)	F1 (%)	Sens. (%)	Espec. (%)
NoT vs. T	Propuesto	55.6 ± 10.3	47.4 ± 12.7	44.1 ± 21.5	36.5 ± 21.2	66.9 ± 16.0
	Random Forest	69.5 ± 8.8	62.2 ± 7.6	66.1 ± 7.2	79.11 ± 9.5	71.11 ± 1.42
TM vs. TF	Propuesto	69.1 ± 7.9	67.3 ± 4.3	77.1 ± 6.5	84.3 ± 16.9	31.1 ± 28.8
	Random Forest	59.1 ± 9.3	58.6 ± 16.4	62.9 ± 29.3	69.2 ± 34.2	37.8 ± 28.5

5.2. EVALUACIÓN MULTIMODAL PARA LA PREDICCIÓN DEL TRATAMIENTO

Tras validar la eficacia individual de los codificadores clínico e imagenológico del método propuesto, se procedió a evaluar su desempeño en configuraciones multimodales orientadas a la predicción del tratamiento en las configuraciones binarias. Se busca con

estas representaciones multimodales, aprovechar el carácter conjunto de los hallazgos, para obtener una estimación apropiada en la predicción del tratamiento. Para ello, se integraron las representaciones clínicas seleccionadas previamente con la codificación imagenológica basada en mapas ADC. La arquitectura resultante fue comparada con estrategias multimodales de referencia que emplean mecanismos convolucionales y de atención. La primera aproximación consiste en una arquitectura Transformer donde los mapas de ADC son procesados en el codificador y las representaciones clínicas en el decodificador, integrándose mediante mecanismos de atención cruzada. Por otro lado, se evaluó un modelo que emplea mecanismos de atención y bloques de campo receptivo (RFB, por sus siglas en inglés) para obtener una representación de la lesión la cual se integra con la representación clínica obtenida de embebidos ajustados para cada variable clínica.

Tabla 6. Comparación del desempeño de modelos multimodales bajo configuraciones de predicción binaria. Se contrastan los resultados del modelo propuesto frente a estrategias convolucionales y de atención. Las métricas reportadas corresponden a las configuraciones NoT vs. T (sin tratamiento vs. tratamiento) y TM vs. TF (trombectomía mecánica vs. trombólisis farmacológica). Los mejores resultados se encuentran resaltados en negrilla.

Conf.	Modelo	AUROC (%)	Exactitud (%)	F1 (%)	Sens. (%)	Espec. (%)
NoT vs. T	Propuesto	71.2 ± 7.8	70.5 ± 8.5	77.1 ± 7.5	78.3 ± 12.5	56.7 ± 16.9
	Transformer	66.3 ± 10.3	57.7 ± 8.6	62.7 ± 8.1	55.5 ± 9.4	61.7 ± 14.3
	Atención y RFB	68.1 ± 11.0	64.7 ± 11.0	73.3 ± 9.9	76.3 ± 13.5	43.9 ± 20.6
TM vs. TF	Propuesto	74.6 ± 13.8	74.1 ± 9.2	81.1 ± 6.9	81.2 ± 9.6	58.9 ± 21.0
	Transformer	66.8 ± 15.4	72.2 ± 8.2	79.8 ± 6.2	81.3 ± 11.4	52.2 ± 17.0
	Atención y RFB	79.3 ± 6.1	68.4 ± 12.8	69.6 ± 24.0	66.4 ± 27.5	72.2 ± 23.9

Los resultados presentados en la Tabla 6 muestran un mejor desempeño del enfoque propuesto en ambas configuraciones de clasificación. Para la distinción entre NoT vs. T, el modelo propuesto alcanzó la mayor exactitud (70.5%), F1 (77.1%) y sensibilidad (78.3%), indicando una adecuada capacidad para identificar correctamente a los pacientes candidatos a tratamiento. Aunque la especificidad fue inferior en comparación con el modelo Transformer (56.7% vs. 61.7%), el F1 sugiere un mayor equilibrio entre sensibilidad y

precisión, favoreciendo al modelo propuesto. En la configuración TM vs. TF, el enfoque propuesto también obtuvo la mejor exactitud (74.14%) y F1 (81.1%). Si bien el método basado en mecanismos de atención y RFB alcanzó un mejor AUROC (79.3%) y especificidad (72.2%), su desempeño fue más inconsistente entre los folds, con una desviación estándar elevada en el F1 y la sensibilidad, lo que sugiere menor estabilidad del modelo. Con el fin de extender la capacidad del modelo propuesto hacia la predicción multiclase del tratamiento (TM, TF o TCon), se implementó una estrategia de ensamblaje secuencial basada en modelos de clasificación binaria previamente entrenados. En esta configuración, se predice inicialmente si el paciente requiere tratamiento o no (NoT vs. T). En los casos clasificados como tratamiento, la muestra se redirige a un segundo modelo encargado de discriminar entre TF y TM. Esta aproximación permite reutilizar modelos binarios ajustados y validados en etapas previas para abordar una tarea multiclase más compleja. El desempeño del ensamblaje fue comparado con una versión del modelo propuesto entrenado específicamente bajo una configuración multiclase, con el objetivo de evaluar los beneficios de la descomposición secuencial.

Tabla 7. Comparación del desempeño entre una estrategia de ensamblaje secuencial de modelos binarios y un modelo de clasificación multiclase en la predicción del tratamiento en pacientes con ACV isquémico. Los mejores resultados para cada métrica están resaltados en negrilla.

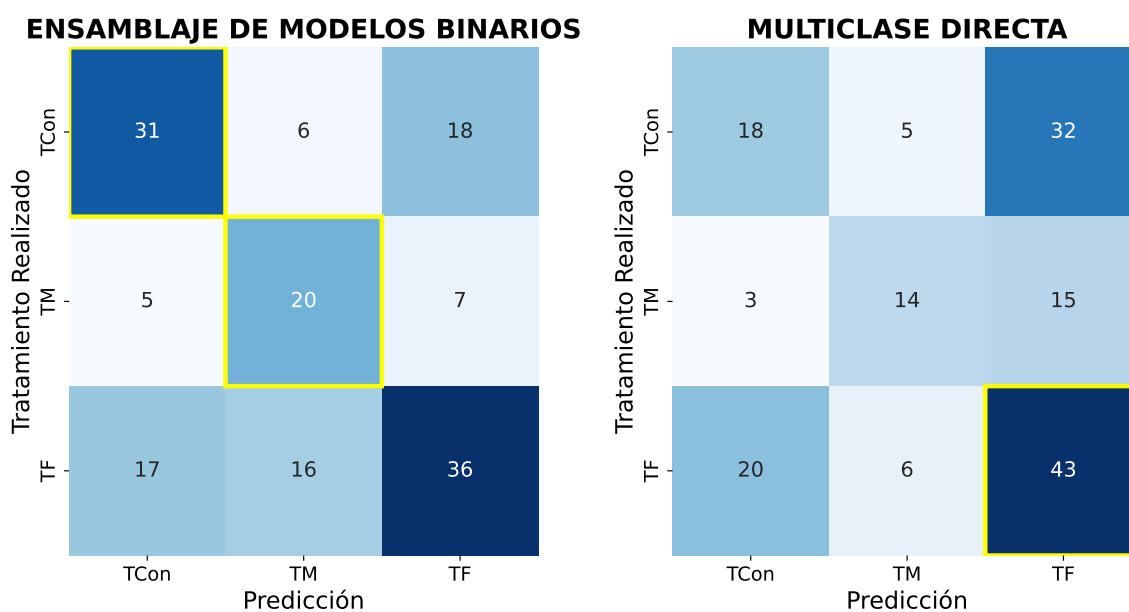
Esquema	Exactitud (%)	F1 (%)	Sensibilidad (%)	Especificidad (%)
Ensamblaje	55.8 ± 6.6	54.9 ± 6.2	57.0 ± 5.9	77.4 ± 3.3
Directo	48.1 ± 7.6	47.1 ± 9.6	45.5 ± 8.9	71.7 ± 4.4

Los resultados observados en la Tabla 7 muestran que el enfoque basado en ensamble supera al modelo multiclase directo en todas las métricas evaluadas. Esta diferencia puede atribuirse a la especialización de los modelos binarios en tareas específicas, lo que permite capturar de manera más eficiente los patrones asociados a cada clase. En contraste, el modelo multiclase directo enfrenta una mayor complejidad al intentar aprender simultáneamente las características de las tres clases.

Un análisis detallado de las predicciones muestra que el modelo basado en ensamble pre-

presenta una mayor capacidad para discriminar entre las tres clases evaluadas (ver Figura 5). En la clase TCon, se obtuvieron 31 predicciones correctas, aunque persistieron 18 errores principalmente hacia la clase TF, posiblemente debido a similitudes clínicas como el territorio arterial afectado. La distinción entre TCon y TM fue más clara, con solo 6 errores. La clase TF alcanzó 36 aciertos, lo que indica una buena capacidad de identificación por parte del modelo. En comparación, el enfoque de clasificación directa presentó mayores dificultades, especialmente entre las clases TCon y TF, con un total de 52 errores entre ambas. Aunque logró 43 aciertos en TF, se evidenció un sesgo hacia esta clase. En el caso de TM, el modelo por ensamble también mostró mejor desempeño, con 20 aciertos sobre 32 posibles. En conjunto, estos resultados indican que la estrategia de ensamble ofrece una clasificación más equilibrada y precisa que el modelo multiclase directo.

Figura 5. Matrices de confusión para la predicción multiclase del tratamiento en pacientes con ACV isquémico. Se compara el desempeño del modelo por ensamble secuencial de clasificadores binarios (izquierda) con la estrategia de clasificación multiclase directa (derecha).



6. CONCLUSIONES Y TRABAJO FUTURO

En este trabajo se propuso una estrategia computacional multimodal que integra información imagenológica con variables clínicas del paciente con el fin de predecir el tratamiento más adecuado para pacientes con ACV isquémico. Para entrenar y validar el enfoque propuesto, se utilizó un conjunto de datos privado compuesto por 156 pacientes que fueron tratados después de presentar síntomas de ACV. El modelo propuesto demostró un desempeño para configuraciones binarias con un F1 de 77.1% para NoT vs T, mientras que para TM vs TF se logró un F1 de 81.1%. Además, teniendo en cuenta el impacto y la utilidad clínica, se abordó una estrategia de ensamblaje de modelos binarios para permitir la clasificación entre los tres tipos de tratamiento (TCon, TM y TF). Esta estrategia demostró una mejor capacidad de discriminación en comparación con emplear el modelo propuesto para la discriminación multiclase directa (con un F1 de 54.9% vs 47.1%), con un acierto más equilibrado entre clases, donde se evidenció de forma general cierta dificultad para la discriminación entre TCon y TM.

Considerando el carácter multimodal y la información analizada en este trabajo, se consideraron versiones unimodales para explorar la capacidad de cada fuente de información en la tarea de predicción. Teniendo en cuenta la información imagenológica, se implementó un modelo que extrae características no-locales de la lesión utilizando un Vision Transformer, el cual fue comparado con una red preentrenada ResNet152. El modelo propuesto mostró un mejor desempeño alcanzando un F1 de 73.1% para NoT vs T y 72.4% para TM vs TF. En cuanto a la información clínica, se diseñó un modelo que genera representaciones embebidas para cada variable, el cual se contrastó con un clasificador Random Forest, comúnmente utilizado en la literatura para el procesamiento de datos tabulares. Sin embargo, los resultados demostraron que el uso exclusivo de variables clínicas no es suficiente para lograr representaciones predictivas robustas al obtener métricas considerablemente bajas para la especificidad en ambos enfoques (31.1% y 37.8% respectiva-

mente).

Según nuestros conocimientos a la fecha, el estudio propuesto en el marco de este trabajo de investigación representa la primera aproximación que integra hallazgos radiológicos con información clínica del episodio del ACV isquémico para predecir el tipo de tratamiento. En la literatura sí existen diversos trabajos que predicen escalas de afectación o que soportan tareas de segmentación basados en ambas modalidades (imágenes e información clínica). Sin embargo, la predicción del tratamiento aún no está ampliamente estudiada, lo cual resulta clave en la guía clínica para evitar riesgos adicionales en el paciente, así como también para garantizar un pronóstico favorable, según sus condiciones de ingreso.

Finalmente, como trabajo futuro se podría incorporar información imagenológica proveniente de otras modalidades diagnósticas, como las imágenes DWI, que podrían aportar detalles adicionales sobre la lesión y complementar las características extraídas de los mapas de ADC. Así mismo, sería pertinente explorar la inclusión de nuevas variables clínicas relevantes para la tarea de clasificación, como el volumen de la lesión. Por otro lado, se plantea la posibilidad de extender el enfoque propuesto hacia la predicción de escalas clínicas asociadas al déficit neurológico y al nivel de discapacidad en pacientes con ACV isquémico dependiendo de un tratamiento aplicado para medir la progresión de la enfermedad. Se espera que este trabajo pueda además ser extendido con herramientas computacionales que permitan aprovechar de forma más efectiva la información recuperada durante el protocolo de atención. Si bien, la arquitectura propuesta logra modelar datos tabulares y regiones de la lesión, su carácter puede ser dependiente de una correcta localización del ACV, así como también está asociada a la subjetividad de los clínicos a la hora de dar ingreso y caracterización inicial al paciente.

BIBLIOGRAFÍA

BACCHI, Stephen et al. “Deep learning in the prediction of ischaemic stroke thrombolysis functional outcomes: a pilot study”. In: *Academic radiology* 27.2 (2020), e19–e23 (cit. on pp. 12, 27).

BAHDANAU, Dzmitry; CHO, Kyunghyun, and BENGIO, Yoshua. “Neural machine translation by jointly learning to align and translate”. In: *arXiv preprint arXiv:1409.0473* (2014) (cit. on p. 20).

BORISOV, Vadim et al. “Deep neural networks and tabular data: A survey”. In: *IEEE transactions on neural networks and learning systems* (2022) (cit. on p. 21).

BOULAHIA, Said Yacine et al. “Early, intermediate and late fusion strategies for robust deep learning-based multimodal action recognition”. In: *Machine Vision and Applications* 32.6 (2021), p. 121 (cit. on p. 24).

BREIMAN, Leo. “Random forests”. In: *Machine learning* 45 (2001), pp. 5–32 (cit. on p. 23).

BRUGNARA, Gianluca et al. “Multimodal predictive modeling of endovascular treatment outcome for acute ischemic stroke using machine-learning”. In: *Stroke* 51.12 (2020), pp. 3541–3551 (cit. on p. 27).

CHEN, Long et al. “Sca-cnn: Spatial and channel-wise attention in convolutional networks for image captioning”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 5659–5667 (cit. on p. 20).

CHEN, Tianqi and GUESTRIN, Carlos. “Xgboost: A scalable tree boosting system”. In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 2016, pp. 785–794 (cit. on p. 23).

CHOURIB, I et al. “Stroke treatment prediction using features selection methods and machine learning classifiers”. In: *IRBM* 43.6 (2022), pp. 678–686 (cit. on p. 29).

DOSOVITSKIY, Alexey et al. “An image is worth 16x16 words: Transformers for image recognition at scale”. In: *arXiv preprint arXiv:2010.11929* (2020) (cit. on p. 34).

FEIGIN, Valery L et al. “Global, regional, and national burden of stroke and its risk factors, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019”. In: *The Lancet Neurology* 20.10 (2021), pp. 795–820 (cit. on pp. 11, 14).

FEIGIN, Valery L et al. “Pragmatic solutions to reduce the global burden of stroke: a World Stroke Organization–Lancet Neurology Commission”. In: *The Lancet Neurology* 22.12 (2023), pp. 1160–1206 (cit. on pp. 11, 14, 18, 30).

FERNANDEZ-LOZANO, Carlos et al. “Random forest-based prediction of stroke outcome”. In: *Scientific reports* 11.1 (2021), p. 10071 (cit. on p. 26).

GOYAL, Mayank et al. “Challenging the ischemic core concept in acute ischemic stroke imaging”. In: *Stroke* 51.10 (2020), pp. 3147–3155 (cit. on p. 14).

HATAMI, Nima et al. “A Novel Autoencoders-LSTM Model for Stroke Outcome Prediction using Multimodal MRI Data”. In: *arXiv preprint arXiv:2303.09484* (2023) (cit. on pp. 12, 26).

HLAVICA, Martin et al. “Pharmacological recanalization therapy in acute ischemic stroke—evolution, current state and perspectives of intravenous and intra-arterial thrombolysis”. In: *Journal of Neuroradiology* 42.1 (2015), pp. 30–46 (cit. on p. 19).

HOOPEES, Andrew et al. “SynthStrip: skull-stripping for any brain image”. In: *NeuroImage* 260 (2022), p. 119474. DOI: <https://doi.org/10.1016/j.neuroimage.2022.119474> (cit. on p. 37).

JO, Hongju et al. “Combining clinical and imaging data for predicting functional outcomes after acute ischemic stroke: an automated machine learning approach”. In: *Scientific Reports* 13.1 (2023), p. 16926 (cit. on pp. 12, 26).

KORSTANJE, Joos. *Advanced forecasting with Python*. Springer, 2021 (cit. on p. 23).

LATCHAW, Richard E et al. “Recommendations for imaging of acute ischemic stroke: a scientific statement from the American Heart Association”. In: *Stroke* 40.11 (2009), pp. 3646–3678 (cit. on pp. 11, 14, 15).

LI, Xiangrui et al. “The first step for neuroimaging data analysis: DICOM to NIfTI conversion”. In: *Journal of neuroscience methods* 264 (2016), pp. 47–56 (cit. on p. 37).

LIU, Chin-Fu et al. “Digital 3D brain MRI arterial territories atlas”. In: *Scientific Data* 10.1 (2023), p. 74 (cit. on p. 33).

LIU, Yongkai et al. “Functional Outcome Prediction in Acute Ischemic Stroke Using a Fused Imaging and Clinical Deep Learning Model”. In: *Stroke* 54.9 (2023), pp. 2316–2327 (cit. on pp. 12, 28).

LUONG, Minh-Thang; PHAM, Hieu, and MANNING, Christopher D. “Effective approaches to attention-based neural machine translation”. In: *arXiv preprint arXiv:1508.04025* (2015) (cit. on p. 20).

MA, Wenao et al. “Treatment Outcome Prediction for Intracerebral Hemorrhage via Generative Prognostic Model with Imaging and Tabular Data”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2023, pp. 715–725 (cit. on pp. 13, 28).

MAHMOOD, Qaiser et al. “Prediction of ischemic stroke lesion and clinical outcome in multi-modal MRI images using random forests”. In: *Workshop BrainLes MICCAI 2016*. Springer. 2016, pp. 244–255 (cit. on p. 12).

MAIER, Oskar et al. “Predicting stroke lesion and clinical outcome with random forests”. In: *Workshop BrainLes MICCAI 2016*. Springer. 2016, pp. 219–230 (cit. on p. 12).

MARTINEZ-HERAS, Eloy et al. “Diffusion-weighted imaging: recent advances and applications”. In: *Seminars in Ultrasound, CT and MRI*. Vol. 42. 5. Elsevier. 2021, pp. 490–506 (cit. on p. 15).

MISTRY, Eva A et al. “National Institutes of Health Stroke Scale as an outcome in stroke research: value of ANCOVA over analyzing change from baseline”. In: *Stroke* 53.4 (2022), e150–e155 (cit. on p. 16).

NIU, Zhaoyang; ZHONG, Guoqiang, and YU, Hui. “A review on the attention mechanism of deep learning”. In: *Neurocomputing* 452 (2021), pp. 48–62 (cit. on p. 21).

POWERS, William J et al. “Guidelines for the early management of patients with acute ischemic stroke: 2019 update to the 2018 guidelines for the early management of acute

ischemic stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association”. In: *Stroke* (2019) (cit. on pp. 18, 19, 30).

SAMAK, Zeynel A; CLATWORTHY, Philip, and MIRMEHDI, Majid. “Prediction of thrombectomy functional outcomes using multimodal data”. In: *Annual Conference on Medical Image Understanding and Analysis*. Springer. 2020, pp. 267–279 (cit. on pp. 12, 28).

— “Transop: Transformer-based multimodal classification for stroke treatment outcome prediction”. In: *arXiv preprint arXiv:2301.10829* (2023) (cit. on p. 27).

SAVER, Jeffrey L et al. “Standardized nomenclature for modified rankin scale global disability outcomes: consensus recommendations from stroke therapy academic industry roundtable XI”. In: *Stroke* 52.9 (2021), pp. 3054–3062 (cit. on p. 17).

SAVER, Jeffrey L et al. “Thrombectomy for distal, medium vessel occlusions: a consensus statement on present knowledge and promising directions”. In: *Stroke* 51.9 (2020), pp. 2872–2884 (cit. on p. 18).

SMITH, Wade S et al. “Mechanical thrombectomy for acute ischemic stroke: final results of the Multi MERCI trial”. In: *Stroke* 39.4 (2008), pp. 1205–1212 (cit. on p. 18).

TANG, Yang et al. *Atlas of emergency neurovascular imaging*. Tech. rep. Springer, 2020 (cit. on p. 14).

WARDLAW, Joanna M et al. “Thrombolysis for acute ischaemic stroke”. In: *Cochrane database of systematic reviews* 7 (2014) (cit. on p. 19).

YANG, Fan; NING, Bo, and LI, Huaqing. “An overview of multimodal fusion learning”. In: *International Conference on Mobile Computing, Applications, and Services*. Springer. 2022, pp. 259–268 (cit. on p. 24).

ZHANG, Duoyi; NAYAK, Richi, and BASHAR, Md Abul. “Exploring fusion strategies in deep learning models for multi-modal classification”. In: *Australasian Conference on Data Mining*. Springer. 2021, pp. 102–117 (cit. on p. 24).

ZUO, Wangmeng et al. “Learning iteration-wise generalized shrinkage–thresholding operators for blind deconvolution”. In: *IEEE Transactions on Image Processing* 25.4 (2016), pp. 1751–1764 (cit. on p. 20).