

PROYECTO FINAL - Predicción de Accidentes Cerebrovasculares Usando Perceptrón Multicapa

Diego Antonio Rosario Palomino
Ciencias de la computación (100%)

Katherine Coralie Figueroa Avalos
Ingeniería civil (100%)

Xiomara Mayela Siche Eusebio
Bioingeniería (100%)

I. INTRODUCCIÓN

El Accidente Cerebrovascular (ACV), o "stroke", es una de las principales causas de mortalidad y discapacidad a nivel global, ocurriendo cuando el suministro de sangre al cerebro se interrumpe. La detección temprana de individuos en riesgo es crucial para la prevención. En este proyecto, abordaremos el desafío de predecir la probabilidad de un ACV utilizando Machine Learning y el "Stroke Prediction Dataset" de Kaggle. Nuestro objetivo es desarrollar una solución de clasificación binaria que estime el riesgo de ACV basándose en características clínicas y demográficas de los pacientes, transformando un modelo predictivo fundamental en un sistema inteligente y aplicable.

Más allá de la predicción, el proyecto busca construir un sistema robusto para un contexto real. Integraremos un análisis de interpretabilidad del modelo con técnicas como integrated gradient, lo que nos permitirá entender el "porqué" de cada predicción, algo esencial en el ámbito médico.

Esto nos podría permitir determinar si el modelo está aprendiendo una fórmula para clasificar los datos de los pacientes. O si el modelo usado está memorizando los datos. Algo muy útil para este dataset, ya que los MLP tienden a presentar underfitting u overfitting en este.

II. OBJETIVOS

Objetivo General:

- Desarrollar un sistema de Machine Learning para la predicción interpretable de accidentes cerebrovasculares (ACV).

Objetivos Específicos:

- Preprocesar y analizar el dataset de ACV, abordando el desbalance de clases.
- Entrenar y evaluar modelos de clasificación robustos para la predicción de ACV.
- Implementar un análisis de interpretabilidad del modelo para explicar las predicciones.

III. CONJUNTO DE DATOS

A. Origen

El conjunto de datos utilizado (Stroke Prediction Dataset de Kaggle) compila información de 5.110 pacientes adultos, cada uno con 12 atributos o columnas relevantes. Estos incluyen variables demográficas (por ejemplo, identificación de paciente, edad, género, estado civil, tipo de residencia),

clínicas (hipertensión, enfermedad cardíaca, nivel promedio de glucosa en sangre, índice de masa corporal BMI) y de estilo de vida (tipo de trabajo, hábito de fumar), además de la etiqueta de clase binaria "stroke" que indica si el paciente ha sufrido un ACV (1) o no (0). En la Tabla 1 a continuación se resume la composición del dataset, destacando el marcado desbalance de clases:

- **Total de registros:** 5110
- **Casos positivos (ACV = 1):** 249 (aproximadamente 4.87% del total)
- **Casos negativos (ACV = 0):** 4861 (aproximadamente 95.1%)

Este desbalance de clases es evidente, con la condición de ACV siendo relativamente rara (aproximadamente 5% de los casos). Un conjunto de datos tan sesgado puede llevar a que modelos predictivos alcancen una alta exactitud aparente simplemente por aprender a predecir siempre la clase mayoritaria, ignorando la minoritaria. Por ello, desde el inicio se anticipa la necesidad de emplear métricas de evaluación más informativas (ver sección de Evaluación) y considerar técnicas de ajuste como el re-muestreo o la asignación de pesos para mitigar el impacto del desbalance.

B. Características

Variable	Tipo	Descripción / Observaciones
id	Numérica	Identificador único del paciente. No aporta valor predictivo.
gender	Categorica	Género del paciente (Male, Female, Other). Un valor atípico ('Other').
age	Numérica	Edad del paciente. Varía entre 0.08 y 82 años.
hypertension	Binaria	Indica presencia de hipertensión. Desbalanceada.
heart_disease	Binaria	Indica enfermedad cardíaca. También desbalanceada.
ever_married	Categorica	Si el paciente ha estado casado. Valores: Yes / No.
work_type	Categorica	Tipo de ocupación: Private, Govt_job, Self-employed, etc.
Residence_type	Categorica	Zona de residencia: Urbano o Rural. Casi balanceado.
avg_glucose_level	Numérica	Nivel promedio de glucosa. Con presencia de outliers.
bmi	Numérica	Índice de masa corporal. Contiene valores faltantes (N/A).
stroke	Binaria	Variable objetivo. Muy desbalanceada (5% positivos).

TABLE I: Variables presentes en el dataset y principales observaciones

C. Análisis exploratorio

Se realizó una exploración de las relaciones entre las características y la incidencia de ACV. Algunos hallazgos destacables incluyeron:

- **Edad:** La incidencia de ACV aumenta notablemente con la edad. La tasa de *strokes* es baja en adultos jóvenes, pero crece con cada grupo etario, observándose el pico de casos alrededor de los 80 años.

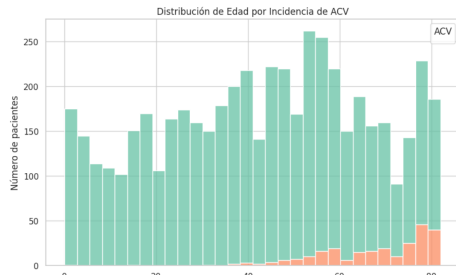


Fig. 1: Distribución de edad por incidencia de ACV

- **Género:** Se aprecia una ligera mayor frecuencia de ACV en mujeres que en hombres, si bien la diferencia no es muy pronunciada.

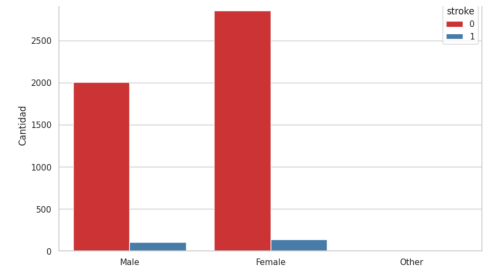


Fig. 2: Distribución de género por incidencia de ACV

- **Hipertensión y enfermedad cardíaca:** Como era esperable, los pacientes con antecedente de hipertensión arterial presentan incidencias de ACV más altas. En el dataset se ve una clara asociación de mayor riesgo de *stroke* en individuos hipertensos frente a no hipertensos, y de igual forma, aquellos con enfermedad cardíaca. Estas comorbilidades son factores de riesgo conocidos, lo que refuerza la validez del dataset.

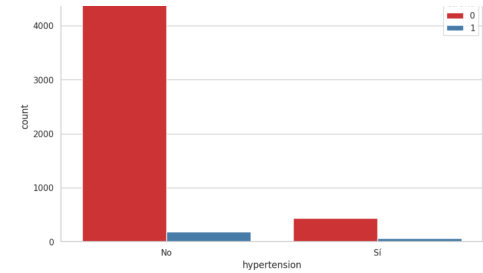


Fig. 3: Distribución de hipertensión por incidencia de ACV

- **Estado civil:** Los pacientes categorizados como “casados” muestran una tasa ligeramente mayor de ACV comparado con los no casados. La diferencia no es muy grande, pero podría reflejar correlaciones con la edad u otros factores socioeconómicos.

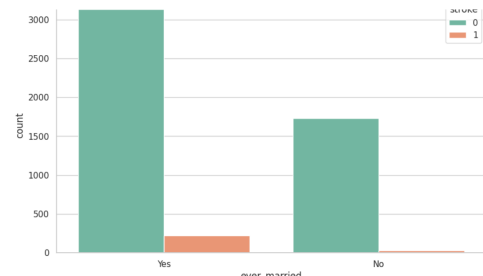


Fig. 4: Distribución de estado civil por incidencia de ACV

- **Tipo de residencia (urbano/rural):** No se halló una diferencia significativa en la proporción de ACV entre habitantes de zonas urbanas versus rurales. Esto sugiere que, en este dataset específico, el entorno residencial no impacta fuertemente el riesgo de *stroke*.

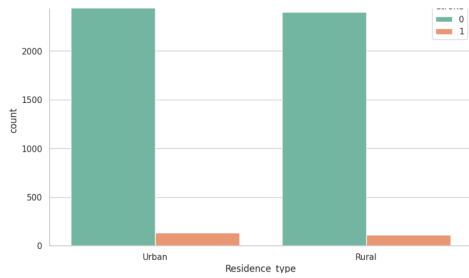


Fig. 5: Distribución de residencia por incidencia de ACV

- **Nivel de glucosa promedio:** La relación entre la glucemia promedio y el ACV parece no lineal. Se observaron mayores incidencias de ACV en pacientes con niveles de glucosa en ciertos rangos elevados (aproximadamente 200 mg/dL), pero también en algunos pacientes con niveles moderadamente bajos (60–120 mg/dL). Esto podría indicar que tanto la hiperglucemia crónica como ciertos episodios hipoglucémicos podrían ser indicadores de riesgo.

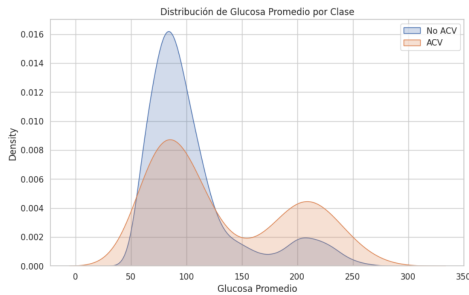


Fig. 6: Distribución de glucosa por incidencia de ACV

- **Índice de masa corporal (BMI):** La mayoría de casos de ACV se concentran en pacientes con BMI en el rango de 20 a 40. Es decir, los pacientes con peso insuficiente (<20) o con obesidad mórbida (>40) no predominan en los casos de ACV del dataset, aunque esto podría deberse a que son minoría en la población general también.

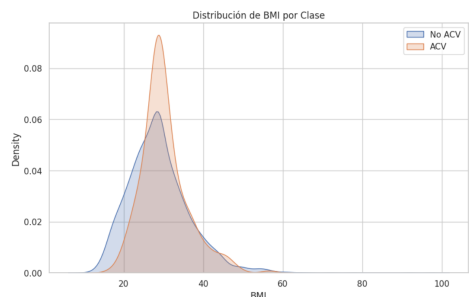


Fig. 7: Distribución de BMI por incidencia de ACV

- **Hábito de fumar:** La exploración sugiere que los fumadores actuales no presentan la mayor tasa de ACV en comparación con exfumadores o personas que nunca fumaron. Los datos indican mayor incidencia de ACV entre exfumadores, seguidos de quienes nunca han fumado,

mientras que los fumadores actuales muestran un número menor de casos. Esta aparente anomalía podría explicarse por sesgos en la muestra.

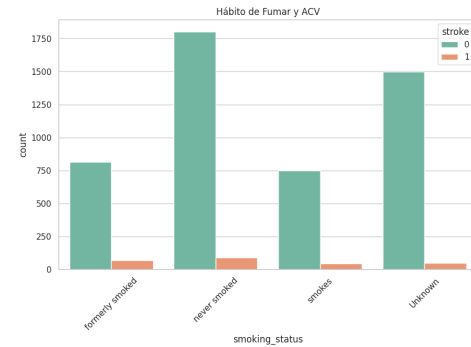


Fig. 8: Distribución de hábito por incidencia de ACV

- **Ocupación:** Al comparar la incidencia de ACV según el tipo de trabajo, se encontró un ligero aumento de casos entre quienes trabajan en el sector privado o son autónomos, en comparación con empleados gubernamentales, amas de casa u otros. Este patrón podría asociarse a niveles de estrés, largas jornadas o hábitos de vida menos saludables en ciertos sectores laborales.

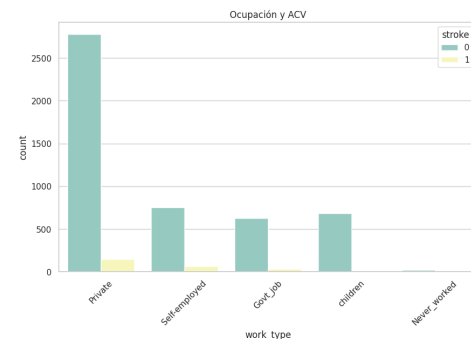


Fig. 9: Distribución de ocupación por incidencia de ACV

D. Preprocesamiento aplicado

- 1) **Carga del dataset:** Se utilizó el archivo `healthcare-data.csv`, el cual contiene 5110 registros y 12 atributos.
- 2) **Eliminación de columnas irrelevantes:** Se descartó la columna `id`, al no aportar información predictiva.
- 3) **Tratamiento de valores atípicos:** Se eliminó el único registro con género etiquetado como `Other`.
- 4) **Imputación de valores faltantes:** La variable `bmi` contenía valores nulos, los cuales fueron completados utilizando la media. Asimismo, los valores faltantes en `smoking_status` fueron reemplazados con la categoría `unknown`.
- 5) **Codificación de variables categóricas:** Se aplicó codificación *one-hot encoding* a las variables categóricas, excluyendo la primera categoría para evitar multicolinealidad.

- 6) **Normalización de variables numéricas:** Se escalaron las variables `age`, `avg_glucose_level` y `bmi` mediante *StandardScaler* para mejorar el rendimiento del modelo.
- 7) **Balanceo de clases:** Dado que la variable objetivo `stroke` estaba altamente desbalanceada, se utilizó la técnica *SMOTE* (Synthetic Minority Over-sampling Technique) para generar ejemplos sintéticos de la clase minoritaria.

IV. METODOLOGÍA

A. Modelos seleccionados

- **Red Neuronal Multicapa (Multilayer Perceptron)**

Para este proyecto se seleccionó como único modelo de predicción el Perceptrón Multicapa (MLP), una red neuronal de tipo feed-forward ampliamente utilizada en tareas de clasificación supervisada. La elección se fundamenta en su capacidad de modelar relaciones no lineales complejas entre variables, lo cual resulta especialmente útil en contextos médicos donde múltiples factores pueden interactuar de manera no trivial [4].

A diferencia de modelos como los árboles de decisión, que generan reglas explícitas mediante particiones jerárquicas del espacio de características, el MLP permite aprender representaciones distribuidas a través de sus capas ocultas, lo que lo vuelve más expresivo en términos de patrones subyacentes en los datos. Esta capacidad es particularmente ventajosa al tratar con indicadores biomédicos (edad, presión, glucosa, etc.) que, en conjunto, pueden contribuir al riesgo de ACV sin una relación lineal evidente [5].

Se optó por una arquitectura compacta pero suficientemente expresiva, con dos capas ocultas de 64 y 32 neuronas respectivamente, activación ReLU y una capa de salida con activación sigmoide para clasificación binaria. Esta estructura se considera adecuada para conjuntos de datos tabulares de tamaño moderado, y ha mostrado un buen equilibrio entre capacidad de representación y riesgo de sobreajuste en literatura previa.

Además, el MLP permite incorporar con facilidad técnicas de regularización avanzadas (como *activation* y *excitation dropout*), que son más difíciles de aplicar en modelos no neuronales. Por tanto, su selección responde tanto a su flexibilidad como a su potencial de alcanzar una mayor sensibilidad predictiva frente a eventos de baja frecuencia como el ACV.

- **Interpretabilidad del modelo.** A pesar de su capacidad predictiva, el Perceptrón Multicapa (MLP) carece de interpretabilidad directa, lo cual representa un desafío en aplicaciones clínicas donde la transparencia es fundamental. Para mitigar esta limitación, se contempla el uso de técnicas post-hoc de explicabilidad, específicamente *Integrated Gradients* [6]. Esta técnica permite atribuir de manera coherente la predicción de un modelo neuronal a sus características de entrada, al integrar los gradientes de la función de salida a lo largo de un trayecto desde una entrada base (por ejemplo, ceros) hasta la entrada actual.

Con ello, se obtiene una estimación del aporte de cada variable (edad, glucosa, hipertensión, etc.) a la decisión del modelo, ofreciendo una visualización interpretable y cuantificable del razonamiento del MLP. Esto permite balancear el uso de modelos complejos con la necesidad de explicar sus decisiones, aspecto crítico en contextos sensibles como el de la salud.

- 1) **Activation Dropout :** Es una técnica de regularización propuesta para reducir el sobreajuste en redes neuronales profundas. Fue introducida en el mismo papel académico que la siguiente técnica ya que son similares, sin embargo *activation dropout* es una aproximación computacionalmente barata aunque un poco menos efectiva en el performance del modelo dado, pero sigue siendo mejor que *dropout* normal.
- 2) **Excitation Dropout :** Es una variación avanzada de *dropout* introducida para fomentar la plasticidad en redes neuronales profundas. A diferencia del *dropout* convencional (que desactiva neuronas al azar con probabilidad uniforme), este método guía la desactivación de neuronas basándose en la importancia (saliencia) de cada neurona en la predicción de la red [8]. En concreto, se evalúa la evidencia que cada neurona aporta a la decisión final del modelo, y se asigna mayor probabilidad de desactivación a las neuronas que más contribuyen a la salida (es decir, aquellas con activaciones más relevantes) [8]. De este modo, las neuronas de alta saliencia son penalizadas durante el entrenamiento, forzando al MLP a encontrar rutas alternativas a través de neuronas menos utilizadas para seguir minimizando la pérdida. Este comportamiento induce una forma de reorganización plástica en la red, similar a la adaptación que ocurre en cerebros biológicos, evitando que el modelo dependa excesivamente de un pequeño subconjunto de neuronas destacadas. Gracias a este enfoque dirigido, la red distribuye mejor la carga de aprendizaje entre sus unidades. La literatura reporta que *Excitation Dropout* logra mejorar la generalización del modelo al reducir la dependencia de características específicas, a la vez que aumenta el aprovechamiento del conjunto completo de neuronas y otorga mayor resiliencia ante la compresión del modelo (reducción o poda de parámetros) sin una caída notable en el rendimiento [8].

B. Funciones de pérdida

Cabe destacar que la incorporación de estas técnicas de regularización no requirió modificar la función de pérdida empleada en el entrenamiento. Se utilizó la pérdida estándar de entropía cruzada (propia de tareas de clasificación) para optimizar el MLP, ya que tanto *Activation drop-out* como *Excitation Dropout* actúan sobre la activación de neuronas durante el *forward pass* sin alterar directamente el cálculo del error objetivo [9]. Esto significa que el proceso de aprendizaje subyacente permanece centrado en minimizar la misma métrica de

pérdida, mientras las técnicas introducidas ayudan a mejorar la generalización controlando la actividad de las neuronas.

V. EXPERIMENTACIÓN

1) Diseño experimental

El modelo utilizado fue un Perceptrón Multicapa (MLP) con dos capas ocultas de 64 y 32 neuronas respectivamente. La red se implementó en *PyTorch*, permitiendo aplicar técnicas avanzadas de regularización y mayor control sobre el proceso de entrenamiento. Los datos se dividieron en conjuntos de entrenamiento (80%) y prueba (20%), asegurando una distribución balanceada y representativa. Para garantizar la reproducibilidad de los resultados, se fijaron semillas de aleatoriedad utilizando `torch.manual_seed(42)`, `numpy.random.seed(42)` y `random.seed(42)`. Esto asegura que tanto la inicialización de pesos como la partición de los datos sean consistentes en ejecuciones sucesivas.

2) Semilla y reproducibilidad

Con el fin de garantizar la reproducibilidad de los experimentos, se fijaron semillas de aleatoriedad en todas las librerías involucradas. Esto asegura que la partición de los datos, la inicialización de los pesos y cualquier operación estocástica sea consistente entre ejecuciones. Las semillas utilizadas fueron:

- `torch.manual_seed(42)`: fija la semilla para *PyTorch*.
- `numpy.random.seed(42)`: fija la semilla para operaciones aleatorias de *NumPy*.
- `random.seed(42)`: asegura reproducibilidad en funciones aleatorias de la librería estándar de Python.

3) Matriz de Confusión

La matriz de confusión obtenida para el conjunto de prueba se muestra en la Tabla II. Esta representa el desempeño del modelo en la clasificación binaria del evento cerebrovascular (ACV), donde la clase positiva corresponde a la presencia del evento ($ACV = 1$) y la clase negativa a su ausencia ($ACV = 0$).

TABLE II: Matriz de confusión del modelo MLP en test

Clase real \ Predicción	Negativo (0)	Positivo (1)
Negativo (0)	486	0
Positivo (1)	25	0

Como se observa, el modelo obtuvo una **exactitud de prueba del 95.11%**, pero esta cifra resulta engañosa debido al severo desbalance de clases. El modelo predijo correctamente todos los casos negativos (486 verdaderos negativos) pero **falló en identificar los 25 casos positivos**, clasificándolos incorrectamente como negativos

(falsos negativos). Esto refleja un sesgo del modelo hacia la clase mayoritaria, comprometiendo su utilidad clínica. Este comportamiento pone en evidencia la necesidad de utilizar métricas adicionales a la exactitud (como *recall*, *precision* o *F1-score*) y aplicar técnicas de balanceo de clases o regularización más efectivas para mejorar la detección de la clase minoritaria.

4) Interpretabilidad

A. Visualización de la atribución mediante Integrated Gradients

- La siguiente imagen presenta la importancia media de cada variable en la predicción del modelo. Se observa que *avg_glucose_level* y *bmi* tienen una mayor contribución, lo cual es coherente con su relevancia clínica. Variables como *residence_type* o *heart_disease* presentan atribuciones cercanas a cero, indicando un impacto limitado en la salida del modelo.

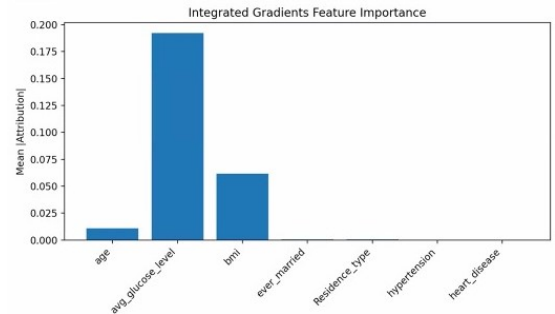


Fig. 10: Importancia media por variable según Integrated Gradients

- Se presentan dos gráficos de dispersión que muestran las atribuciones individuales para las variables x_1 (izquierda) y x_2 (derecha). En el caso de x_1 , la atribución varía entre valores positivos y negativos, lo que indica una influencia bidireccional sobre la salida del modelo. En cambio, la atribución de x_2 es mayormente negativa, lo que sugiere que esta variable tiende a reducir la salida del modelo. Estos gráficos permiten observar la variabilidad y dirección de la influencia de cada característica, a diferencia de los gráficos de barras que solo muestran promedios absolutos.

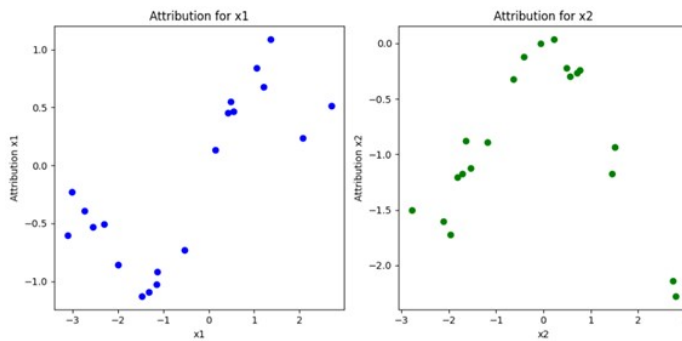


Fig. 11: Visualización de la atribución individual de las variables x_1 y x_2 en función de sus valores de entrada.

- La figura presenta dos gráficos de dispersión que comparan las derivadas verdaderas del modelo con respecto a las variables de entrada ($\partial f/\partial x_1$ y $\partial f/\partial x_2$) frente a las atribuciones estimadas mediante Integrated Gradients. En ambos casos, las curvas reales siguen patrones sinusoidales, lo que refleja la naturaleza de las funciones de entrada. Las atribuciones de IG muestran una tendencia similar, aunque con mayor dispersión, lo que indica que IG logra capturar de forma aproximada la sensibilidad del modelo ante variaciones en las entradas, validando así su utilidad como herramienta interpretativa para relaciones no lineales.

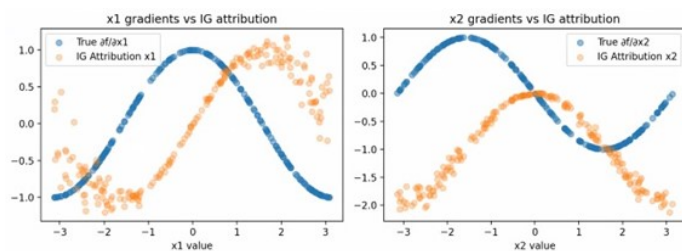


Fig. 12: Comparación entre gradientes reales y atribuciones de IG

- La visualización corresponde a un experimento sintético donde se grafican las atribuciones generadas por IG frente a variables con comportamientos no lineales (\sin y \cos). Se aprecia que las atribuciones siguen una forma sinusoidal, lo cual evidencia que IG puede capturar correctamente relaciones complejas entre las entradas y la predicción.

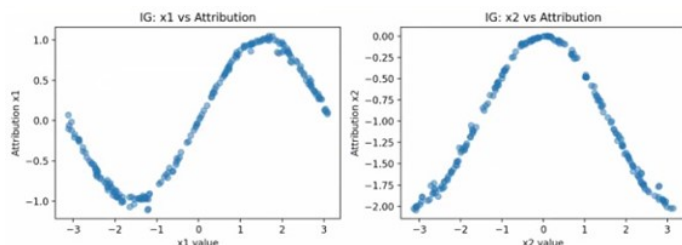


Fig. 13: Relación entre el valor de entrada y su atribución con IG en funciones no lineales

- La figura muestra la comparación de las atribuciones medias absolutas para dos variables sintéticas: $x_1 = \sin(\cdot)$ y $x_2 = \cos(\cdot)$. Se observa que x_2 tiene una mayor relevancia en las predicciones del modelo. Este análisis permite visualizar qué entradas tuvieron mayor influencia en la salida de la red, contribuyendo a una interpretación más transparente del modelo.

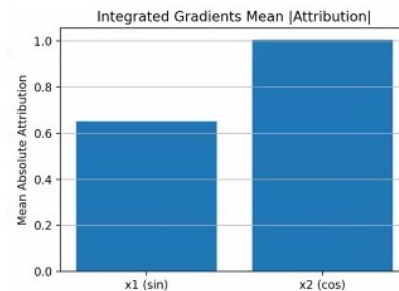


Fig. 14: Atribución media absoluta por variable (x_1 y x_2) usando Integrated Gradients

- Se presenta la saliencia media absoluta de las variables sintéticas $x_1 = \sin(\cdot)$ y $x_2 = \cos(\cdot)$, reflejando qué tan influyente fue cada característica en las predicciones del modelo. Se observa que ambas variables tienen una relevancia similar, aunque x_2 muestra una atribución ligeramente mayor. Esto sugiere que, en promedio, el modelo considera al coseno como una entrada ligeramente más determinante que el seno.

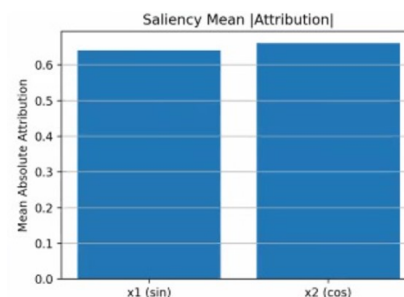


Fig. 15: Importancia de características según saliencia media absoluta

- A continuación se muestra la atribución media absoluta de las variables $x_1 = \sin(\cdot)$ y $x_2 = \cos(\cdot)$ utilizando el método GradientSHAP. Se observa que x_2 alcanza una atribución cercana a 1.0, mientras que x_1 se aproxima a 0.70, lo que indica que, bajo este enfoque, la variable coseno fue considerablemente más influyente en las decisiones del modelo. Esta diferencia resalta cómo distintos métodos de explicabilidad pueden revelar patrones diversos sobre la relevancia de las características.

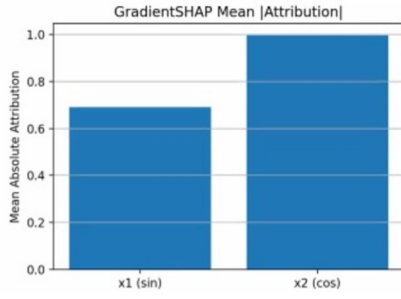


Fig. 16: Importancia de características según GradientSHAP

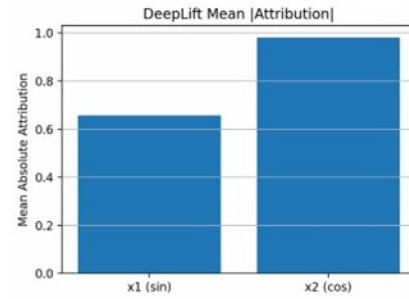


Fig. 18: Comparación de atribuciones medias absolutas mediante *Saliency* y *DeepLIFT*.

- La Figura muestra una comparación de dos técnicas de interpretabilidad: $\text{Input} \times \text{Gradient}$ y *Saliency*, aplicadas a las variables de entrada x_1 (sin) y x_2 (cos). En el gráfico superior, ambas variables presentan atribuciones absolutas medias prácticamente iguales (aprox. 1.10), lo que sugiere que ambas influyen fuertemente y de forma equilibrada en las predicciones. En cambio, en el gráfico inferior (*Saliency*), se observa una ligera diferencia a favor de x_2 (cos), aunque ambas variables siguen siendo relevantes. Esta comparación resalta cómo diferentes métodos de atribución pueden generar resultados similares en magnitud, pero con matices distintos en la percepción de la importancia relativa.

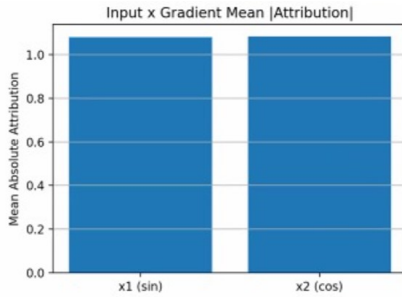


Fig. 17: Comparación entre $\text{Input} \times \text{Gradient}$ y *Saliency* Mean —Attribution—

- Se muestra una comparación entre los métodos de interpretabilidad *Saliency* y *DeepLIFT* para las variables x_1 (sin) y x_2 (cos). Mientras que *Saliency* presenta valores similares para ambas entradas (0.64 y 0.66), *DeepLIFT* resalta una clara dominancia de x_2 , cuya atribución alcanza casi el valor máximo (1.00). Esta diferencia evidencia cómo cada técnica puede ofrecer perspectivas distintas sobre la importancia relativa de las características del modelo.

VI. DISCUSIÓN

Los resultados obtenidos mediante la técnica de *Integrated Gradients* indican que la incorporación de *Activation Dropout* en el entrenamiento del modelo no produjo una alteración significativa en la lógica interna de toma de decisiones del Perceptrón Multicapa. En concreto, las variables que recibieron mayor atribución —como edad, nivel de glucosa y presión arterial— se mantuvieron prácticamente constantes en términos de su relevancia relativa antes y después de aplicar esta técnica de regularización.

Aunque *Activation Dropout* ha sido propuesta como una estrategia para fomentar un aprendizaje más robusto y distribuido dentro de las redes neuronales, los resultados de este estudio no muestran evidencia empírica clara de un cambio sustancial en el comportamiento del modelo. Las curvas de atribución generadas se mantuvieron similares, y no se observó una redistribución notoria del peso asignado a características menos predominantes.

Este resultado puede explicarse por diversas razones de índole metodológica. En primer lugar, la codificación de los datos podría haber limitado la capacidad del modelo para beneficiarse plenamente de la regularización. Se empleó una combinación de encoding para variables categóricas y *StandardScaler* para variables numéricas, lo cual, si bien es una práctica común, puede introducir ciertas rigideces que dificultan la adaptación del modelo a dinámicas más complejas.

En segundo lugar, los hiperparámetros seleccionados —como la tasa de *dropout*, la arquitectura del modelo y la tasa de aprendizaje— podrían no haber sido los más adecuados para que *Activation Dropout* despliegue su efecto completo. Finalmente, es posible que la naturaleza del conjunto de datos en sí, con ciertas distribuciones de clases y correlaciones internas, haya limitado el margen de mejora observable al aplicar esta técnica.

En conjunto, estos hallazgos subrayan la importancia de considerar cuidadosamente tanto la codificación de los datos como la sintonización de hiperparámetros al evaluar técnicas de regularización avanzadas. La interpretación debe realizarse con cautela y dentro del contexto metodológico adoptado.

VII. CONCLUSIONES

En este experimento, **no se logró demostrar una ventaja clara del uso de Activation Drop-out frente al entrenamiento estándar del Perceptrón Multicapa (MLP)**. Si bien esta técnica ha mostrado mejoras en la capacidad de generalización en contextos distintos, **los resultados obtenidos en este trabajo fueron comparables o incluso inferiores** a los del modelo sin esta regularización. Las atribuciones generadas por *Integrated Gradients* revelaron que la lógica interna del modelo —en cuanto a las variables a las que se asigna mayor relevancia— **permanece prácticamente inalterada al aplicar Activation Drop-out**. Esto sugiere que:

- El modelo **no modificó sustancialmente sus rutas internas de decisión**.
- No se observó una redistribución significativa de la importancia de las variables ni una disminución de la dependencia de un subconjunto reducido de características.

Este resultado podría atribuirse a **limitaciones en el diseño experimental**, particularmente en la codificación de las variables de entrada categóricas y en ciertos aspectos de la arquitectura y los hiperparámetros seleccionados.

Futuras mejoras.

Para una evaluación más precisa del impacto de Activation Drop-out, se proponen las siguientes mejoras metodológicas:

- Explorar distintos valores de probabilidad para *drop* ($p = 0.1, 0.3, 0.5$) y combinarlos con diferentes esquemas de optimización (por ejemplo, Adam, SGD y aproximación de Natural Gradient Descent).
- Probar codificaciones alternativas, como *embeddings* para variables categóricas, que podrían revelar mejor las diferencias de comportamiento en las atribuciones.
- Realizar estudios de sensibilidad variando la semilla aleatoria y la partición entrenamiento/validación para descartar que los resultados sean producto de una división específica.
- Comparar con otras técnicas de explicación como SHAP o LIME para verificar si resaltan mejor las diferencias entre modelos con y sin regularización avanzada.

VIII. DECLARACIÓN DE CONTRIBUCIONES

A continuación, se detalla la contribución individual de cada integrante del equipo en el desarrollo del presente proyecto:

- **Diego Rosario Palomino:** Diseño del modelo Perceptrón Multicapa (MLP), implementación del modelo en PyTorch y elaboración de la sección de discusión e interpretación de resultados.
- **Xiomara Siche Eusebio:** Análisis exploratorio de datos, desarrollo de la metodología y diseño de la interfaz de usuario para la visualización de resultados.

- **Katherine Coralie Figueroa Avalos:** Ejecución de los experimentos, redacción del informe en LaTeX, investigación bibliográfica y preprocesamiento del conjunto de datos.

Todos los integrantes participaron en la planificación general del proyecto, validación de resultados y revisión final del documento.

IX. REFERENCIAS BIBLIOGRÁFICAS

- [1] L. Xia, Y. Yang, Z. Chen, Z. Yang, and S. Zhu, “Movie Recommendation with Poster Attention via Multi-modal Transformer Feature Fusion,” arXiv preprint arXiv:2407.09157, 2024. [Online]. Available: <https://arxiv.org/abs/2407.09157>
- [2] U. K. Nareti, C. Adak, S. Chattopadhyay, and P. Wang, “Unraveling Movie Genres through Cross-Attention Fusion of Bi-Modal Synergy of Poster,” arXiv preprint arXiv:2410.19764, 2024. [Online]. Available: <https://arxiv.org/abs/2410.19764>
- [3] Y. Mu and Y. Wu, “Multimodal Movie Recommendation System Using Deep Learning,” *Mathematics*, vol. 11, no. 4, p. 895, 2023. doi: 10.3390/math11040895
- [4] Montañez, M. C., Montesinos, J., Martínez, C. M. (2021). Deep learning techniques for early stroke prediction in health datasets. *International Journal of Advanced Computer Science and Applications*, 12(9), 568–575. <https://doi.org/10.14569/IJACSA.2021.0120966>
- [5] Dong, X., Wang, G., Han, L., Xu, L., Hou, L. (2020). Application of machine learning in predicting stroke outcomes. *Frontiers in Neurology*, 11, 575295. <https://doi.org/10.3389/fneur.2020.575295>
- [6] Sundararajan, M., Taly, A., Yan, Q. (2017). Axiomatic attribution for deep networks. *Proceedings of the 34th International Conference on Machine Learning*, 70, 3319–3328. <https://arxiv.org/abs/1703.01365>
- [7] Zhang, Q., Yang, X., Yao, X. (2018). Improving generalization of deep neural networks via regularization and architecture design. arXiv preprint arXiv:1811.05850. Retrieved from <https://arxiv.org/abs/1811.05850>
- [8] Zunino, A., Bargal, S. A., Morerio, P., Zhang, J., Sclaroff, S., Murino, V. (2018). Excitation dropout: Encouraging plasticity in deep neural networks. arXiv preprint arXiv:1805.09092. <https://arxiv.org/abs/1805.09092>
- [9] Liang, B., Xing, E. P., Caruana, R. (2020). Drop-Activation: reducing overfitting and improving generalization in deep neural networks. arXiv preprint arXiv:1811.05850.h

X. LINKS IMPORTANTES

Repositorio de Descarga de Datos: <https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset>

Repositorio de Entrenamiento: