

Interpretando Modelos de IA en Cáncer de Mama con SHAP y LIME

Interpreting AI Models in Breast Cancer with SHAP and LIME

Guzmán-Ponce, A.^{1*}, López-Bautista, J.², Fernandez-Beltrán, R.³ (*aguzman@uji.es)

Recibido: 31 de mayo de 2024 Aceptado: 17 de junio de 2024

RESUMEN

La Inteligencia Artificial (IA) ha transformado múltiples disciplinas, en concreto en el área médica se ha distinguido por su capacidad de mejorar diagnósticos y tratamientos. No obstante, la falta de transparencia en los sistemas ha generado desafíos prácticos. La Inteligencia Artificial Explicable (XAI) ha surgido como una solución para abordar estos problemas, permitiendo a los profesionales de la salud comprender y verificar las decisiones de los algoritmos. En este trabajo se aplicarán dos técnicas comúnmente usadas de XAI, específicamente SHapley Additive

exPlanations (SHAP) y Local Interpretable Model-Agnostic Explanations (LIME), para interpretar un modelo de red neuronal convolucional usado en el diagnóstico de cáncer de mama a partir de imágenes histológicas del Carcinoma Ductal Invasivo (IDC). Los resultados muestran que ambas técnicas son efectivas para identificar las contribuciones individuales de las características en las predicciones del modelo, resaltando que LIME ofrece una granularidad detallada a nivel de píxel.

Palabras clave: *Inteligencia Artificial Explicable, SHAP, LIME, Carcinoma Ductal Invasivo*

ABSTRACT

Artificial Intelligence (AI) has transformed multiple disciplines, particularly medicine, where it has been distinguished by its ability to improve

diagnostics and treatments. However, the lack of transparency in the systems has raised practical challenges. Explainable Artificial Intelligence

¹Departamento de Lenguajes y Sistemas Informáticos, Instituto de Nuevas Tecnologías de Imagen, Universitat Jaume I, 12071 Castelló de la Plana, España.

²Hospital Regional ISSEMyM Toluca, Médico Adscrito al Departamento de Anatomía Patológica, Calle Dr. Nicolás San Juan S/N, Delegación San Lorenzo Tepaltitlán I, Delegación San Lorenzo Tepaltitlán, 50010 Santa Cruz Atzacapotzaltongo, Méx.

³Departamento de Informática y Sistemas, Universidad de Murcia, Murcia, España.



(XAI) has emerged as a solution to address these issues, allowing healthcare professionals to understand and verify the decisions made by algorithms. In this work, two commonly used XAI techniques, Shapley Additive explanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME), are applied to interpret a convolutional neural network model used in the diagnosis of breast cancer from histological images of Invasive Ductal Carcinoma (IDC). The results show that both techniques effectively identify the individual contributions of features in the predictions of the model, highlighting that LIME offers detailed granularity at the pixel level.

Keywords: *Explainable Artificial Intelligence, SHAP, LIME, Invasive Ductal Carcinoma*

INTRODUCCIÓN

En las últimas décadas, la Inteligencia Artificial (IA) ha revolucionado disciplinas desde la agricultura hasta la biotecnología, destacando su impacto en medicina por mejorar diagnósticos y tratamientos. La IA ha sido clave en el reconocimiento de patrones y la predicción de comportamientos, como se evidencia en su aplicación durante la pandemia de COVID-19, donde facilitó la toma de decisiones informadas al prever dinámicas virales [1]. Además, en radiología, las técnicas de IA han permitido distinguir eficazmente entre imágenes de rayos X de pacientes con COVID-19 y otras enfermedades respiratorias, demostrando su eficacia [2] [3].

La implementación de la IA en medicina ha facilitado avances notables en análisis y diagnóstico, mejorando la precisión y eficacia de los tratamientos [4]. No obstante, la falta de transparencia en los procesos de decisión de muchos sistemas de IA presenta retos tanto éticos como prácticos. En el área médica es importante contar con un entendimiento claro de cómo los algoritmos toman decisiones, dada la necesidad de justificar cada diagnóstico o recomendación de tratamiento de manera detallada y comprensible [5].

Ante esta necesidad, ha surgido la Inteligencia Artificial Explicable (eXplainable Artificial Intelligence, XAI, por sus siglas en inglés), como una respuesta a la “caja negra” que se genera en los modelos de IA [6]. XAI permite a los profesionales de la salud comprender y verificar los procesos mediante los cuales se llegan a tomar decisiones automáticas, fomentando así la confianza y facilitando la integración de estos sistemas en las prácticas médicas regulares [7].

A pesar del esfuerzo que la comunidad científica ha realizado, existe una falta de estudios en la aplicación de técnicas de XAI específicamente en imágenes de histología del cáncer de mama IDC. Derivado de este hecho, este trabajo explora cómo diferentes metodologías, como SHAP y LIME, pueden optimizar la interpretación y el análisis de imágenes de histología. La pregunta de investigación central de nuestro estudio se centra en determinar cuál de estas técnicas de XAI proporciona mayor claridad en

el diagnóstico del cáncer de mama.

INTELIGENCIA ARTIFICIAL EXPLICABLE

XAI se ha utilizado para hacer que los procesos de toma de decisiones de la IA sean transparentes y comprensibles, especialmente en campos críticos como la salud, donde las metodologías de IA a menudo resultan complejas para muchos profesionales médicos. XAI busca responder a la pregunta: ¿Qué hace que esta predicción sea confiable? [8]. Por lo tanto, se considera una técnica analítica posterior al proceso de extracción de conocimiento en las metodologías convencionales de IA. Su objetivo es destacar las características que influyen en el rendimiento del modelo, aumentando así la confianza en los sistemas de IA y facilitando su integración en áreas sensibles como la medicina [4] .

Existen dos técnicas de XAI agnósticas al modelo de IA, lo que significa que pueden aplicarse a cualquier tipo de modelo de extracción de conocimiento, sin importar su estructura interna. Las técnicas de Local Interpretable Model Agnostic Explanations (LIME) y SHapley Additive exPlanations (SHAP) son métodos ampliamente utilizados [6]. Ambos enfoques se centran en proporcionar interpretaciones locales de las predicciones, ayudando a entender cómo cada característica contribuye a una decisión específica del modelo.

SHapley Additive exPlanations

SHAP se basa en los valores de Shapley, que cuantifica la contribución de cada característica a la predicción de un modelo. Si se considera un conjunto de datos DS con n instancias, cada una representada por un par (x, y) , donde x es un arreglo de m características $[x_1, x_2, \dots, x_m]$, y es la clase. Dado M el conjunto de todas las características. El valor de Shapley para la i -ésima característica se define como la Ec. (1):

$$\phi_i = \sum_{S \subseteq M \setminus \{i\}} \frac{|S|! (|M| - |S| - 1)!}{|M|!} [f(S \cup \{i\}) - f(S)] \quad (1)$$

Donde S es un subconjunto de M sin la característica i -ésima; $f(S \cup \{i\})$ es la predicción del modelo con la característica i y $f(S)$ es la predicción sin ella [8].

Para aplicar la fórmula, primero se calculan todas las posibles combinaciones de características excluyendo la i -ésima característica. Luego, se evalúan las predicciones del modelo con y sin la característica i para cada combinación. La diferencia en las predicciones se multiplica por un peso que depende del tamaño del subconjunto y del total de características. Finalmente, se suman todas estas contribuciones ponderadas para obtener el valor SHAP, que representa la contribución promedio de la característica i a lo largo de todas las posibles combinaciones. Las características que aumentan las predicciones generan

valores SHAP positivos, mientras que las que disminuyen las predicciones tienen valores negativos.

Local Interpretable Model Agnostic Explanations

LIME se basa en la aproximación local del modelo en la vecindad de una instancia específica. La explicación de LIME para la instancia x se define en la Ec. (2):

$$\xi(x) = \arg \arg \min_{g \in G} L(f, g, \pi_x) + \Omega(g) \quad (2)$$

Donde G son las posibles explicaciones; f es el modelo original que se quiere explicar; g es el modelo local explicable; L es la pérdida que mide la cercanía de la explicación a la predicción del modelo original; π_x es una medida de proximidad que pondera las instancias de acuerdo con su similitud con x . $\Omega(g)$ es una medida de la complejidad del modelo de explicación $g \in G$, por ejemplo, el número de características de entrada.

Para aplicar LIME, primero se generan muestras locales perturbando de manera controlada la instancia x original. Luego, se evalúa el modelo en estas muestras perturbadas para obtener las predicciones correspondientes. A continuación, se asignan pesos a estas muestras usando una medida de proximidad. Con las muestras ponderadas, se entrena un modelo explicable g . Finalmente, se optimiza este modelo para minimizar la pérdida ponderada y la penalización de complejidad, obteniendo así una aproximación localmente interpretable del modelo completo. La interpretación de los valores de LIME es similar a la de SHAP. Las características que aumentan las predicciones generan valores LIME positivos, mientras que las que disminuyen las predicciones tienen valores LIME negativos.

XAI EN IMÁGENES MÉDICAS

En la investigación médica, especialmente en el diagnóstico por imágenes Borys et al. [9], resaltan cómo la disponibilidad de datos y los avances en el alto rendimiento computacional han aumentado el potencial de la IA, pero también señalan la necesidad de entender sus procesos internos en contextos críticos. El estudio se enfoca en métodos de XAI basados en la saliencia, es decir, la cualidad por la cual ciertas partes de una imagen atraen más la atención que otras, son cruciales en la imagenología médica, ya que permiten proporcionar explicaciones directamente sobre los datos de entrada (imágenes), mejorando la confianza y la comprensión de los resultados de IA por parte de médicos y pacientes.

Por otro lado, Shovon et al. [10] abordan la clasificación de imágenes histopatológicas de cáncer de mama HER2, un subtipo de cáncer que se caracteriza por la sobreexpresión del receptor del factor de crecimiento epidérmico humano 2, usando un modelo que integra una red DenseNet201 que optimiza el entrenamiento al conectar densamente cada capa con todas las demás, reduciendo el sobreajuste; y

Xception que utiliza convoluciones separables en profundidad para procesar eficientemente las características. Para interpretar los resultados implementaron Grad-CAM el cual usa gradientes de la última capa convolucional para crear mapas de calor que destacan regiones críticas en imágenes.

Mientras que, Kaplun et al. [11] desarrollaron un método basado en los momentos de Zernike (son polinomios ortogonales en el círculo unitario, útiles para análisis de imágenes por su invarianza rotacional) para caracterizar las imágenes de células cancerosas de mama y posteriormente usar una red neuronal simple para la clasificación. Integraron un modelo local de interpretación, lo que permitió que los resultados obtenidos por el modelo de clasificación de imágenes de células cancerosas fueran más sencillos de interpretar.

METODOLOGÍA

Las imágenes histológicas son imágenes microscópicas de tejidos biológicos que han sido especialmente preparados y teñidos para revelar detalles estructurales, siendo fundamentales en el campo de la patología, ya que permiten examinar la arquitectura microscópica de los tejidos para diagnosticar enfermedades, evaluar la severidad de las condiciones patológicas y entender la patogénesis de diversas enfermedades. En la actualidad, este tipo de imágenes han sido usadas en arquitecturas de aprendizaje profundo donde se pueden aplicar tareas de detección, segmentación y clasificación de tejidos con el objetivo de producir resultados comparables o superiores a los enfoques basados en características diseñadas manualmente, tradicionales en la patología [12].

Conjunto de datos

Para el análisis del objetivo de este artículo, se emplea un conjunto de datos compuesto por imágenes histológicas, centrado específicamente en el estudio del Carcinoma Ductal Invasivo (IDC) sin patrón específico NOS (Not Otherwise Specified), la variante más común de cáncer de mama, que representa aproximadamente el 70-80 % de todos los casos de esta enfermedad [12]. Este tipo de cáncer se distingue por la invasión de las células neoplásicas hacia el estroma mamario causando una reacción llamada desmoplasia.

El conjunto tomado de [13] está compuesto de imágenes en fragmentos de 50x50 píxeles teñidos con Hematoxilina y Eosina (H&E), que es una técnica de tinción común utilizada en histopatología para destacar las características de las células y tejidos. Cada imagen se encuentra etiquetada como clase 1 si contienen células IDC (IDC +), en caso contrario se etiqueta como clase 0 (IDC -), esta etiqueta fue asignada manualmente [12]. Teniendo un total de 5547 imágenes, que se distribuyen en 2759 imágenes con ausencia de IDC y 2788 con presencia de IDC. La Figura 1 muestra un ejemplo representativo para cada categoría.

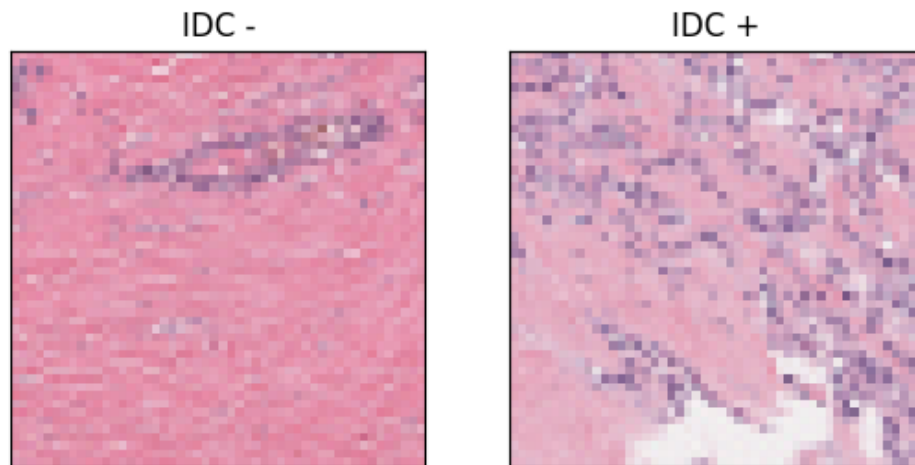


FIGURA 1. *Imágenes de tejido mamario H&E, a la izquierda con ausencia de cáncer, a la derecha con presencia de cáncer.*

Modelo de Aprendizaje

Para la extracción de conocimiento, se implementó un modelo de red neuronal convolucional (CNN) usado en [12], [13], como se observa en la Figura 2. El modelo se compone de diversas capas que ayudan a capturar características esenciales de las imágenes de entrada para realizar la clasificación de IDC. Inicialmente, se utilizan dos capas de convolucionales con 32 y 64 filtros respectivamente, seguidas cada una por una función de activación ReLU que introduce no linealidades al modelo. Luego, una capa de agrupación reduce la dimensionalidad de los datos, lo que ayuda a hacer el modelo más eficiente y a reducir el sobreajuste.

Para combatir aún más el sobreajuste, se incluyen capas de Dropout que, aleatoriamente, ignoran algunos nodos durante el entrenamiento, lo que ayuda a mejorar la generalización del modelo en datos no vistos. Las últimas capas convierten los mapas de características multidimensionales en un vector unidimensional, permitiendo que las capas densas procesen estas características para realizar la clasificación final, el uso de la función de activación softmax distribuye las probabilidades a través de las clases definidas.

El modelo emplea el optimizador Adadelta, adecuado para datos con características dinámicas por su adaptabilidad en el tamaño de pasos de aprendizaje. Se configuró un lote de 128 y se entrenó durante 200 épocas, las capas de Dropout tienen un porcentaje de 0.25 después de la segunda capa convolucional y 0.5 después de la capa densa.

El conjunto de datos fue dividido en 80 % para entrenamiento y el 20 % restante para la fase de prueba.

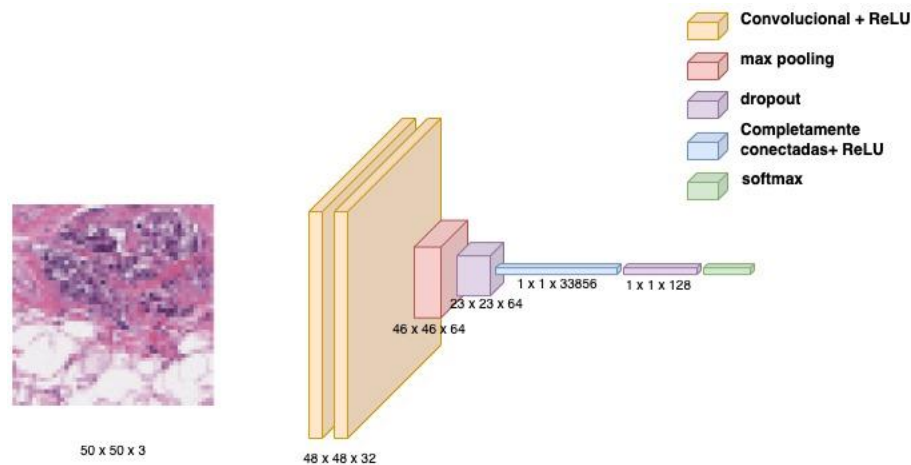


FIGURA 2. Diagrama de la red neuronal convolucional para la clasificación de imágenes histológicas.

La evaluación del modelo se realizó utilizando como métrica la exactitud (accuracy) definida como la proporción de predicciones correctas (tanto verdaderos positivos como verdaderos negativos) respecto al total de casos evaluados.

XAI

Una vez que el modelo CNN ha sido entrenado, es posible aplicar métodos que permitan identificar las porciones que han sido más influyentes en la decisión de clasificación de la red. Esta tarea se enmarca en el campo de XAI mediante SHAP y LIME, estas herramientas facilitan la interpretación de las decisiones del modelo, ayudando a identificar posibles sesgos en los datos o en el aprendizaje del modelo.

RESULTADOS

El modelo CNN empleado para la clasificación de imágenes de IDC alcanzó una precisión del 75 %, condicionada por una resolución de 50x50 píxeles y variaciones en la tinción H&E. Aunque la tinción H&E es estándar en histopatología, sus variaciones limitan significativamente la captura de características celulares en el modelo utilizado.

XAI en IDC

Una vez evaluado el modelo, se deduce que existen oportunidades para mejorar su desempeño mediante la optimización de los hiperparámetros, tales como el número y tamaño de los filtros en las capas convolucionales. En línea con el objetivo de este trabajo, los resultados del análisis de las percepciones sobre la contribución individual de cada característica en las predicciones del modelo es especialmente útil para:

1. Casos en los que el modelo ha clasificado incorrectamente una instancia, identificando qué características influyen negativamente en la decisión del modelo y cómo podrían ajustarse.
2. Casos en los que el modelo ha clasificado correctamente una instancia, entendiendo las características que contribuyen positivamente a las decisiones acertadas.

La metodología SHAP utiliza una escala cromática en la que el color azul simboliza una disminución en la probabilidad de la clase asociada, orientada específicamente hacia la ausencia de cáncer, con una intensidad que es directamente proporcional al impacto de la característica evaluada. En contraste, la metodología LIME emplea el color verde para indicar características que mejoran positivamente la precisión de las predicciones, mientras que el color rojo señala aquellos atributos que tienen un efecto negativo.

La Figura 3 muestra que hay áreas para SHAP con píxeles rojos que deberían incrementar la predicción hacia la presencia de cáncer, pero el modelo ha predicho incorrectamente la clase. Esto sugiere que, a pesar de la presencia de factores que indican cáncer, las zonas no coinciden con las infiltraciones del carcinoma, por lo que otros factores, como los píxeles azules, han llevado al modelo a una conclusión errónea, dado que coinciden en zonas con ausencia de infiltración. Esto puede deberse al sesgo del modelo, dada la representación insuficiente en los datos de entrenamiento de características similares.

La Figura 4 muestra la explicación generada por LIME para una imagen histológica incorrectamente clasificada como ausente de cáncer, cuando en realidad tiene presencia de IDC. Esta representación visual evidencia que el modelo ha asignado de manera uniforme el color verde a todos los píxeles de la imagen, lo que indica que ha interpretado estas características como señales de ausencia de cáncer, contribuyendo

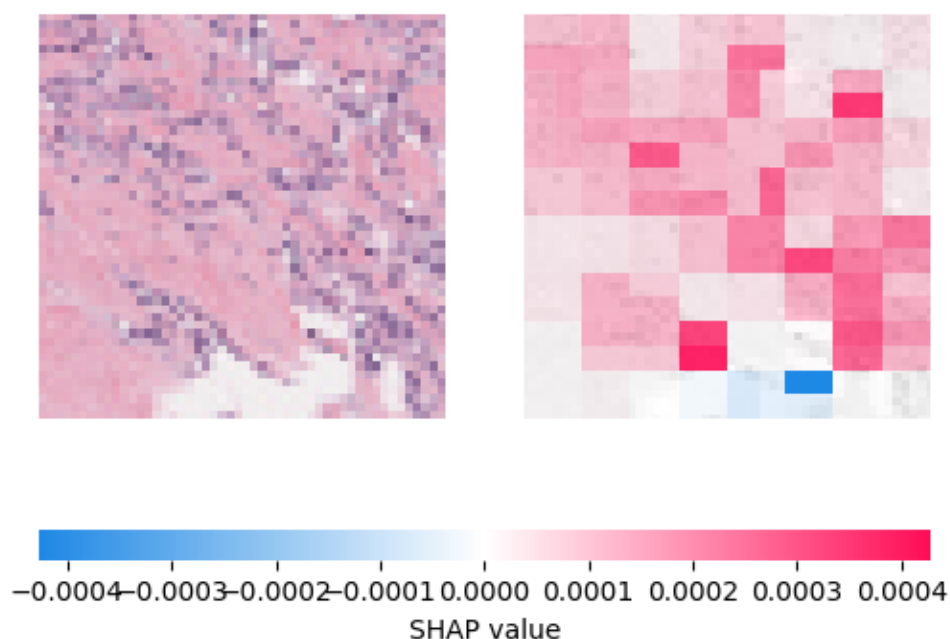


FIGURA 3. Valores SHAP para una imagen clasificada erróneamente como sin cáncer, cuando en realidad la etiqueta correcta indica la presencia de cáncer.

así erróneamente a la predicción de la clase. Esta observación sugiere la necesidad de reentrenar el modelo incorporando un volumen ampliado de imágenes, con el objetivo de mejorar su capacidad de discriminación y precisión en la clasificación de imágenes histológicas.

En la Figura 5 se observan los valores de SHAP correspondientes a una imagen correctamente clasificada con ausencia de cáncer. Los valores marcados en azul se localizan en áreas donde no hay reacción a la tincura, indicando la ausencia de cáncer.

De la Figura 6 se observa que los píxeles marcados en verde, correspondientes a LIME, indican zonas donde no se detecta la infiltración del carcinoma.

Un caso particular se ilustra en la Figura 7, donde el modelo predijo la presencia de cáncer, a pesar de que el conjunto de datos clasifica esta muestra como carente de cáncer. Sin embargo, es notable que existen áreas dentro de la imagen que el análisis SHAP identifica como características relevantes, coincidiendo con el análisis visual realizado por una experta en la materia. Esta observación sugiere la posible presencia de instancias ruidosas en el conjunto de datos debido a un etiquetado inadecuado. Esta situación resalta la importancia de verificar la precisión de las etiquetas en los datos utilizados para entrenar modelos de aprendizaje automático, especialmente en aplicaciones críticas como el diagnóstico médico.

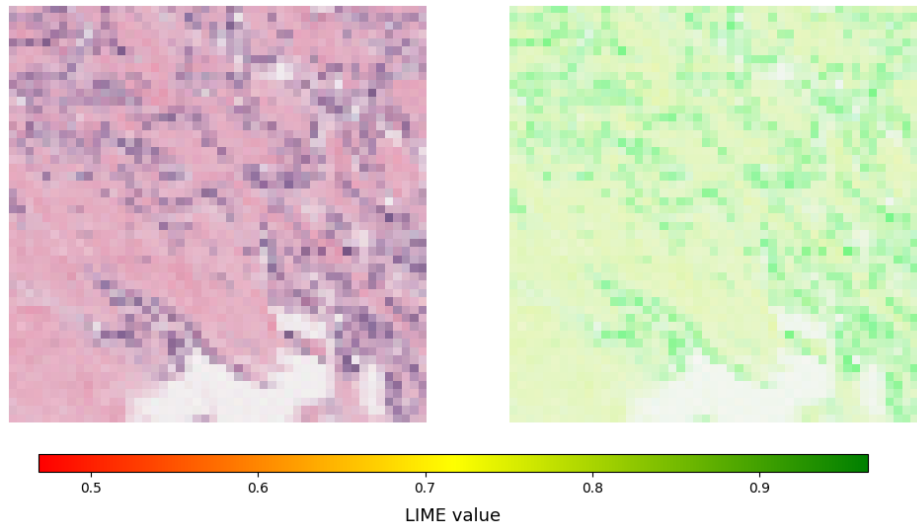


FIGURA 4. *Explicación con LIME para una imagen clasificada erróneamente como sin cáncer, cuando en realidad la etiqueta correcta indica la presencia de cáncer.*

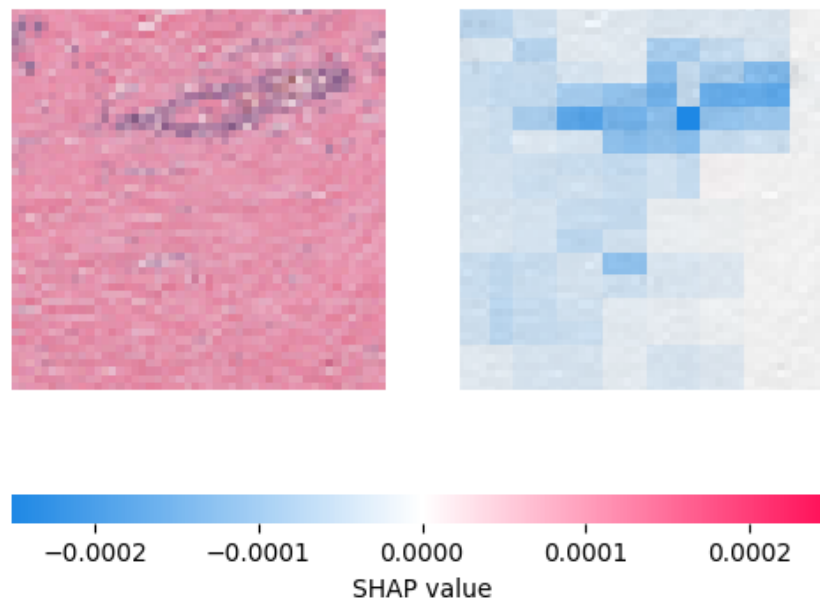


FIGURA 5. *Valores SHAP para una imagen correctamente clasificada como sin cáncer.*

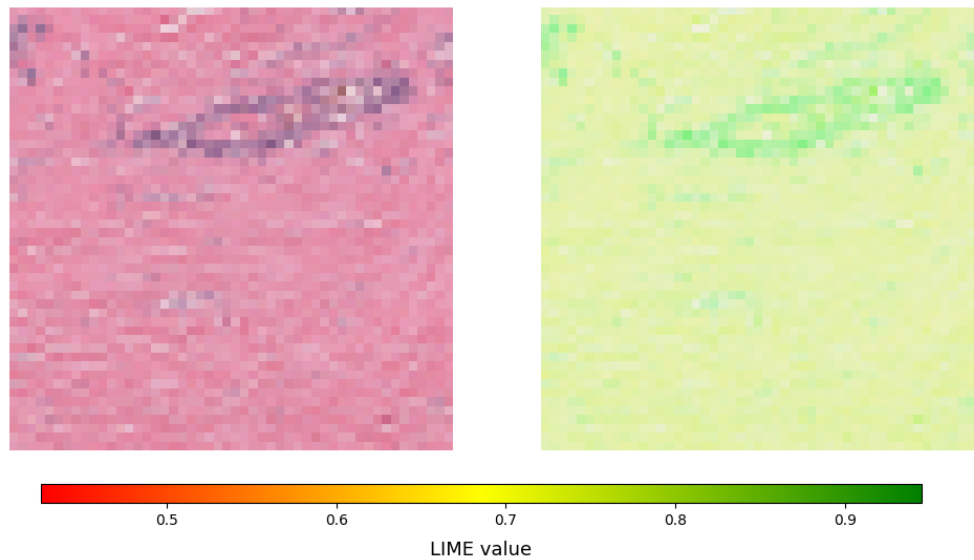


FIGURA 6. Explicación de *LIME* para una imagen correctamente clasificada como sin cáncer.

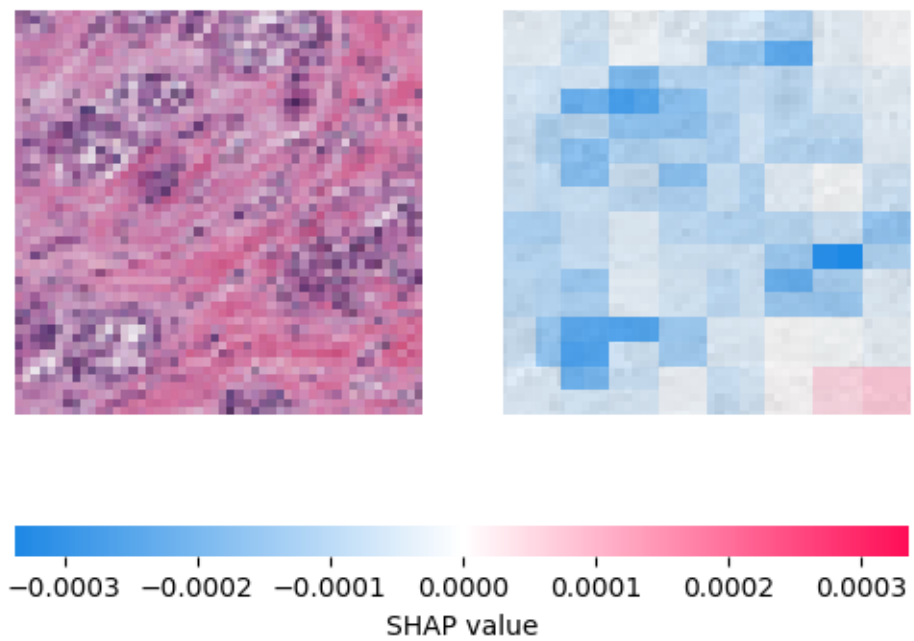


FIGURA 7. Valores *SHAP* para una imagen clasificada erróneamente con presencia de cáncer, cuando en realidad la etiqueta a priori indica ausencia de cáncer.

En la Figura 8, relativa a la aplicación de LIME, se observa un comportamiento similar al descrito por SHAP, donde las zonas marcadas en verde también coinciden con el análisis visual de la experta. Esto refuerza la relevancia de utilizar herramientas de interpretación local, como LIME y SHAP, para validar y entender mejor las predicciones del modelo frente a las etiquetas originales del conjunto de datos.

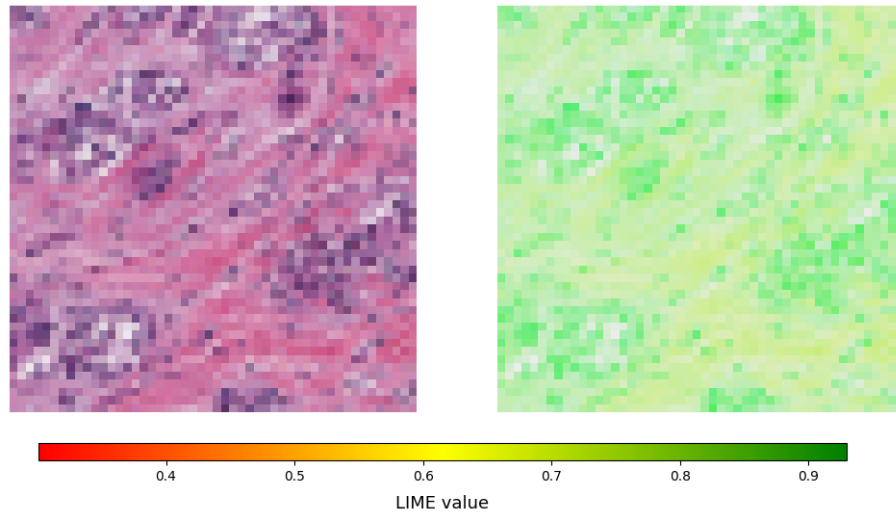


FIGURA 8. *Explicación de LIME para una imagen clasificada erróneamente con presencia de cáncer, cuando en realidad la etiqueta a priori indica ausencia de cáncer.*

Al realizar una comparación entre ambas técnicas de interpretación de modelos, LIME y SHAP, se observa que LIME proporciona una representación detallada, píxel a píxel, de la importancia de las características en la decisión del modelo. Esta granularidad es particularmente útil en el análisis de imágenes de histopatología coloreadas mediante la técnica Hematoxilina y Eosina, ya que permite una elucidación más precisa del aprendizaje del modelo en relación con la base de conocimiento proporcionada.

LIME facilita una mejor comprensión de los patrones específicos y las características morfológicas que el modelo está aprendiendo de las imágenes de histopatología. Esto contribuye a la validación y refinamiento del modelo, asegurando que las inferencias hechas se alineen estrechamente con los criterios diagnósticos establecidos por especialistas médicos.

El conjunto de datos para análisis de cáncer de mama, al centrarse en un subtipo específico, no refleja toda la variedad clínica, lo que lleva a una sobre-optimización del modelo hacia características dominantes no extrapolables a otros contextos. Las técnicas de dropout y regularizaciones implementadas no garantizan eliminar los sesgos hacia características visuales no representativas.

CONCLUSIONES

Este trabajo ha explorado el uso de técnicas de XAI en la interpretación de modelos de IA utilizados para el diagnóstico de cáncer de mama a partir de imágenes histológicas del Carcinoma Ductal Invasivo. A través del uso de SHAP y LIME, se buscó proporcionar transparencia y claridad en las decisiones tomadas por un modelo de red neuronal convolucional entrenado para esta tarea.

Los resultados obtenidos demuestran que ambas técnicas SHAP y LIME, son efectivas para revelar las contribuciones individuales de las características en las predicciones del modelo. El uso de LIME en el análisis de imágenes histopatológicas no solo proporciona una visión integral de las dinámicas de aprendizaje del modelo, sino que también ofrece oportunidades para ajustes dirigidos que pueden mejorar la sensibilidad y especificidad del modelo en tareas de diagnóstico médico.

El uso de XAI en la interpretación de modelos de IA para el diagnóstico de IDC es fundamental para aumentar la confianza en estos sistemas y para facilitar su integración en la práctica médica. La capacidad de explicar y justificar las decisiones del modelo no solo mejora la transparencia, sino que también promueve una mayor aceptación de la IA por parte de los profesionales de la salud.

Como trabajo a futuro se centrará en la mejora del modelo en cuestión de parámetros y el incremento de imágenes en la base de conocimiento del modelo. Adicionalmente, se explorará el impacto que los ajustes en los hiperparámetros pueden tener sobre la precisión y la interpretabilidad del modelo.

Agradecimientos

Angélica Guzmán-Ponce contó con el apoyo del contrato postdoctoral Margarita Salas MGS/2021/23 (UP2021-021) financiado por la Unión Europea-NextGenerationEU

REFERENCIAS

- [1] A. Guzmán-Ponce, J. L. González-Ruíz, and I. Francisco-Valencia. Review on the application of big data algorithms to understand a pandemic virus. In *Handbook of Research on Applied Artificial Intelligence and Robotics for Government Processes*. IGI Global, 2023, pp. 211–233.
- [2] H. Li, N. Zeng, P. Wu, and K. Clawson. Cov-net: A computer-aided diagnosis method for recognizing covid-19 from chest x-ray images via machine vision. *Expert Systems with Applications*, vol. 207, p. 118029, 2022. DOI: 10.1016/j.eswa.2022.118029
- [3] F. Kamalov, K. Rajab, A. K. Cherukuri, A. Elnagar, and M. Safaraliev. Deep learning for covid-19 forecasting: State-of-the-art review. *Neurocomputing*, vol. 511, pp. 142–154, 2022. DOI:

10.1016/j.neucom.2022.09.005

- [4] L. Longo, M. Brcic, F. Cabitza, J. Choi, R. Confalonieri, J. Del Ser, R. Guidotti, Y. Hayashi, F. Herrera, A. Holzinger, R. Jiang, H. Khosravi, F. Lecue, G. Malgieri, A. Páez, W. Samek, J. Schneider, T. Speith, and S. Stumpf. Explainable artificial intelligence (XAI) 2.0: A manifesto of open challenges and interdisciplinary research directions. *Information Fusion*, vol. 106, p. 102301, 2024. DOI: 10.1016/j.inffus.2024.102301
- [5] L. Malinverno, V. Barros, F. Ghisoni, G. Visonà, R. Kern, P. J. Nickel, B. E. Ventura, I. imi, S. Stryeck, F. Manni, C. Ferri, C. Jean-Quartier, L. Genga, G. Schweikert, M. Lovri, and M. Rosen-Zvi. A historical perspective of biomedical explainable ai research. *Patterns*, vol. 4, no. 9, p. 100830, 2023. DOI: 10.1016/j.patter.2023.100830
- [6] R. O. Alabi, M. Elmusrati, I. Leivo, A. Almangush, and A. A. Mäkitie. Machine learning explainability in nasopharyngeal cancer survival using lime and shap. *Scientific Reports*, vol. 13, no. 1, p. 8984, 2023. DOI: 10.1038/s41598-023-35795-0
- [7] M. M. Abdelsamea, U. Zidan, Z. Senousy, M. M. Gaber, E. A. Rakha, and M. Ilyas. A survey on artificial intelligence in histopathology image analysis. *WIREs Data Mining and Knowledge Discovery*, vol. 12, 2022. DOI: 10.1002/widm.1474
- [8] M. N. S. Choudary, V. B. Bommineni, G. Tarun, G. P. Reddy, and G. Gopakumar. Predicting Covid-19 Positive Cases and Analysis on the Relevance of Features using SHAP (SHapley Additive exPlanation). In *2021 Second ICESC*, Coimbatore, India, 2021, pp. 1892–1896. DOI: 10.1109/ICESC51422.2021.9532829
- [9] K. Borys, Y. A. Schmitt, M. Nauta, C. Seifert, N. Krämer, C. M. Friedrich, and F. Nensa. Explainable ai in medical imaging: An overview for clinical practitioners – saliency-based xai approaches. *European Journal of Radiology*, vol. 162, p. 110787, 2023. DOI: 10.1016/j.ejrad.2023.110787
- [10] M. S. H. Shovon, M. F. Mridha, K. M. Hasib, S. Alfarhood, M. Safran, and D. Che. Addressing uncertainty in imbalanced histopathology image classification of her2 breast cancer: An interpretable ensemble approach with threshold filtered single instance evaluation (sie). *IEEE Access*, vol. 11, pp. 122 238–122 251, 2023. DOI: 10.1109/ACCESS.2023.3327898
- [11] D. Kaplun, A. Krasichkov, P. Chetyrbok, N. Oleinikov, A. Garg, and H. S. Pannu. Cancer cell profiling using image moments and neural networks with model agnostic explainability: A case

study of breast cancer histopathological (breckhis) database. *Mathematics*, vol. 9, no. 20, p. 2616, 2021. DOI: 10.3390/math9202616

- [12] A. Janowczyk and A. Madabhushi. Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. *Journal of Pathology Informatics*, vol. 7, no. 1, p. 29, 2016. DOI: 10.4103/2153-3539.186902
- [13] Predicting IDC in Breast Cancer Histology Images. *Kaggle*, 2018, consultado el 29 de mayo del 2024. [en línea]. Disponible: <https://www.kaggle.com/code/paultimothymooney/predicting-idc-in-breast-cancer-histology-images>