



INAOE

**Explicaciones Visuales Post Hoc (XAI):
Un enfoque orientado hacia la
Leucemia Linfoblástica Aguda**

por

**M.C. José de Jesús Velázquez Arreola
INAOE**

Tesis sometida como requisito parcial
para obtener el grado de

**DOCTOR EN CIENCIAS EN LA
ESPECIALIDAD DE CIENCIAS
Y TECNOLOGÍAS BIOMÉDICAS**

en el

**Instituto Nacional de Astrofísica,
Óptica y Electrónica**

Agosto 2024
Tonantzintla, Puebla

Supervisada por:

Dra. Raquel Díaz Hernández
Investigadora Titular del INAOE

©INAOE 2024

Derechos Reservados

El autor otorga al INAOE el permiso de reproducir y
distribuir copias de esta tesis en su totalidad
o en partes mencionando la fuente.



Resumen

En este estudio se evaluaron las características morfológicas de componentes sanguíneos en imágenes mediante métodos de generación de mapas de relevancia, parte de la Inteligencia Artificial Explicativa (XAI). Se usaron seis métodos (Gradient, ABS Gradient, GradientInput, LRPz, Deep Taylor y Grad-CAM) sobre seis arquitecturas de redes neuronales convolucionales (VGG-16, VGG-19, MobileNet V1, ResNet18, ResNet50 y GoogleNet). La evaluación se realizó en dos fases, empleando las bases de datos “The Cancer Imaging Archive” y “ALL-IDB2”. En la primera fase, expertos hematólogos concluyeron que la combinación del método Gradient*Input con los modelos VGG-19 y ResNet50 fue la más efectiva en imágenes segmentadas. En la segunda fase, el modelo GoogleNet combinado con Grad-CAM mostró una mayor precisión en imágenes sin segmentar. La principal contribución de este estudio es la validación de diversas configuraciones de CNN y métodos de generación de mapas de relevancia, sugiriendo la integración de otros tipos de XAI para mejorar el diagnóstico de la ALL.

Palabras Clave: Explainable Artificial Intelligence (XAI), Heatmaps, Acute Lymphoblastic Leukemia (ALL), and Cell Morphology.

Abstract

This study evaluated the morphological characteristics of blood components in images using relevance map generation methods, part of Explainable Artificial Intelligence (XAI). Six methods (Gradient, ABS Gradient, Gradient*Input, LRPz, Deep Taylor, and Grad-CAM) were applied to six convolutional neural network architectures (VGG-16, VGG-19, MobileNet V1, ResNet18, ResNet50, and GoogleNet). The evaluation was conducted in two phases, using the databases "The Cancer Imaging Archive" and "ALL-IDB2". Expert hematologists concluded that combining the Gradient*Input method with the VGG-19 and ResNet50 models was the most effective in segmented images. In the second phase, the GoogleNet model combined with Grad-CAM showed a 43.61% accuracy in unsegmented images. The main contribution of this study is the validation of different configurations of CNNs and relevance map generation methods, suggesting the integration of other types of XAI to improve ALL diagnoses.

Keywords: Explainable Artificial Intelligence (XAI), Heatmaps, Acute Lymphoblastic Leukemia (ALL), and Cell Morphology.

Agradecimientos

Agradezco al Consejo Nacional de Humanidades, Ciencia y Tecnología (CONAHCYT) por la beca otorgada por un año, la cual sustentó parte de mis estudios permitiendo lograr esta meta personal.

Al Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE) por todo el apoyo recibido de todo tipo. A los doctores y compañeros que hoy en día son mis amigos les digo ¡gracias!, gracias por compartir su conocimiento, alentar y motivar mis deseos de seguir el camino de la ciencia.

Gracias a la Dra. Raquel Díaz Hernández, quien ha sido y es parte importante en el desarrollo de este trabajo de tesis y sobre todo por sus consejos que me ha brindado en todos los años de desarrollo de esta investigación.

Gracias al Dr. Leopoldo Altamirano Robles por la atención brindada, aportando sus conocimientos, ideas y las herramientas necesarias que fueron esenciales para esta tesis.

A mis compañeros a la M.C. Noemí Sánchez y al Ing. Oliver Zárraga por el apoyo y tiempo que me dieron en diferentes momentos del desarrollo de este trabajo. ¡Gracias!

Dedicatorias

Dedico este trabajo a mi familia, por ser mi soporte y mi principal inspiración para alcanzar este sueño y los que están por venir. Sin su apoyo, nada de esto sería posible. Gracias a mi padre por las grandes enseñanzas que me has dado a lo largo de la vida, en especial por enseñarme a no darme por vencido ante cualquier dificultad. Gracias por estar presente en cada momento y por guiar mis pasos. A mi madre le doy las gracias por estar siempre pendiente de mi salud, mis ánimos y, sobre todo, por no permitir que abandone mis sueños. Gracias por recordarme diariamente la presencia de Dios en cada paso que doy. Por estas razones y muchas más, también te dedico este trabajo. A mis hermanas, quiero dedicarles este trabajo y agradecerles todo el apoyo incondicional que me han dado a lo largo de la vida, así como por los hermosos sobrinos que me han dado y que también son parte de mi motivación. A todos ustedes les agradezco por hacerme sentir amado, aun en la distancia. A mi pareja Viri, gracias por animarme en los momentos de desánimo durante este último año y por recordarme que puedo lograr todo lo que me proponga. Por último, pero no menos importante, quiero dedicar este trabajo a Dios. Gracias por encontrarte en cada paso que doy y en cada persona con la que convivo día a día, por el gran amor que me tienes y por hacer todo con una sonrisa.

Tabla de contenido

.....	1
Resumen	2
Abstract.....	3
Agradecimientos.....	4
Dedicatorias	5
Tabla de contenido	6
1 Introducción	8
1.1 Motivación	10
1.2 Pregunta de Investigación.....	11
1.3 Hipótesis	11
1.4 Justificación	11
1.5 Objetivos	12
Objetivo General.....	12
Objetivos Específicos	12
1.6 Trabajo relacionado.....	13
1.7 Alcances y Limitaciones	14
Alcances	14
Limitaciones	15
1.8 Estructura del documento	15
2 Marco Teórico.....	17
2.1 Leucemia Linfoblástica Aguda (ALL).....	17
2.1.1 Examen y diagnóstico.....	19
2.1.2 Clasificación FAB y morfología.....	19
2.2 Base de datos.	23
2.2.1 Base de datos The Cancer Imaging Archive.....	23
2.2.2 Base de datos ALL-IDB2.....	24
2.3 Arquitecturas preentrenadas.....	25
2.3.1 Arquitecturas VGG-16 y VGG-19.....	26
2.3.2 Arquitectura MobileNet V1	27
2.3.3 Arquitectura ResNet50	28
2.3.4 Arquitectura GoogleNet	29

2.4 Métodos Explicativos	30
2.4.1. Método Gradient* Input	33
2.4.2. Método Layer-wise Relevance Propagation (LRPz).....	34
2.4.3. Método Deep Taylor.....	34
2.4.4. Método Grad-CAM.....	35
2.5 Segmentación Semántica	36
3 Metodología.....	38
3.1 Fase 1.....	38
3.1.1 Preparación de la base de datos segmentado	38
3.1.2 Reentrenamiento de modelos	39
3.1.3 Generación de mapas de calor para la primera fase.....	41
3.1.4 Desarrollo de la evaluación realizada por hematólogos expertos.	45
3.2 Fase 2.....	48
3.2.1 Preparación de la base datos ALL-IDB2.....	48
3.2.2 Segmentación semántica con morfología celular de la base de datos ALL-IDB2 para ground truth.....	49
3.2.3 Reentrenamiento de modelos de CNN con la base de datos ALL-IDB251	
3.2.4 Generación de mapas de calor	52
3.2.5 Método de evaluación de mapas de calor.....	55
4 Resultados y Discusión	57
4.1 Resultados Fase 1: Evaluación de Hematólogos	57
4.2 Discusión de los resultados de la fase 1	60
4.3 Resultados Fases 2: Evaluación morfológica	61
4.4 Discusión de los resultados de la fase 2.....	65
5 Conclusiones	66
6 Trabajo Futuro	68
7 Aportaciones Clave.....	68
8 Referencias.....	69

1 Introducción

La Inteligencia Artificial (IA) continúa demostrando mayor eficiencia, en comparación a los resultados obtenidos por humanos expertos en problemáticas de clasificación y/o identificación de patrones en bases de datos. Las Redes Neuronales Convolucionales (CNN, por sus siglas en inglés) son una rama del Aprendizaje Profundo (DL, por sus siglas en inglés) que sigue siendo parte de la IA.

Las arquitecturas de CNN tienen buenos resultados en la clasificación de imágenes. Por ejemplo, clasificar imágenes de perros según su raza, o identificación de objetos cotidianos en imágenes. Además, se ha comprobado que las arquitecturas entrenadas con una gran cantidad de imágenes se pueden reentrenar para la clasificación de nuevos objetos de interés, obteniendo mayor rendimiento en comparación a entrenar una red neuronal desde cero para dicha clasificación. Es cada vez más habitual emplear CNN para la clasificación de objetos en imágenes.

Sin embargo, una desventaja de estos algoritmos es que aún se consideran como cajas negras. Esto se debe a que los resultados obtenidos están basados en probabilidades de pertenecer a una u otra categoría, y el usuario no puede identificar las características específicas que utiliza la CNN para realizar la clasificación.

Derivado a esta problemática se ha desarrollado la Inteligencia Artificial Explicable (XAI), la cual se conforma de algoritmos que pueden dar explicación visual, textual o una combinación de ambos. Unos métodos de explicación visual pueden generar una nueva imagen que corresponde a un mapa de relevancias, esta imagen se conforma principalmente por píxeles con mayor importancia para la CNN en la clasificación de una imagen dada.

En la actualidad, las tecnologías de diagnóstico de enfermedades en el área médica con XAI pueden ayudar a los médicos a descifrar imágenes médicas

como rayos X, imágenes por resonancia magnética y tomografías computarizadas, lo que resulta en diagnósticos más rápidos y precisos. La implementación de este tipo de algoritmo proporciona información adicional al experto para corroborar el resultado obtenido por una CNN.

Dada estas ventajas es importante que, al desarrollar herramientas que empleen algún algoritmo XAI como auxiliar para el área médica, se deba incluir a expertos en el campo para validar los resultados que se van obteniendo en cada paso. En el presente trabajo, se involucra a cinco patólogos expertos quienes validan y aportan información para el desarrollo de esta investigación.

En la Figura 1. se ilustra el resumen gráfico del proceso realizado para este trabajo de tesis. Durante el proceso se experimentaron con diferentes arquitecturas de CNN y métodos explicativos visuales que generan mapas de relevancias, descritas en la sección 2. Así mismo, se realizaron dos evaluaciones, una realizada por los patólogos expertos y la otra realizada implementando segmentación semántica, estas evaluaciones se describen en el apartado 3.

De este trabajo de tesis se publicaron los siguientes artículos:

1. Evaluation of Heatmaps as an Explicative Method for Classifying Acute Lymphoblastic Leukemia Cells. Pattern Recognition. MCPR 2023. Lecture Notes in Computer Science, vol 13902. Springer, Cham. https://doi.org/10.1007/978-3-031-33783-3_24
2. Evaluation of Heat Map Methods Using Cell Morphology for Classifying Acute Lymphoblastic Leukemia Cells. Computación y Sistemas, Vol. 28, No. 1, 2024, pp. 221–231 <https://doi.org/10.13053/cys-28-1-4897>

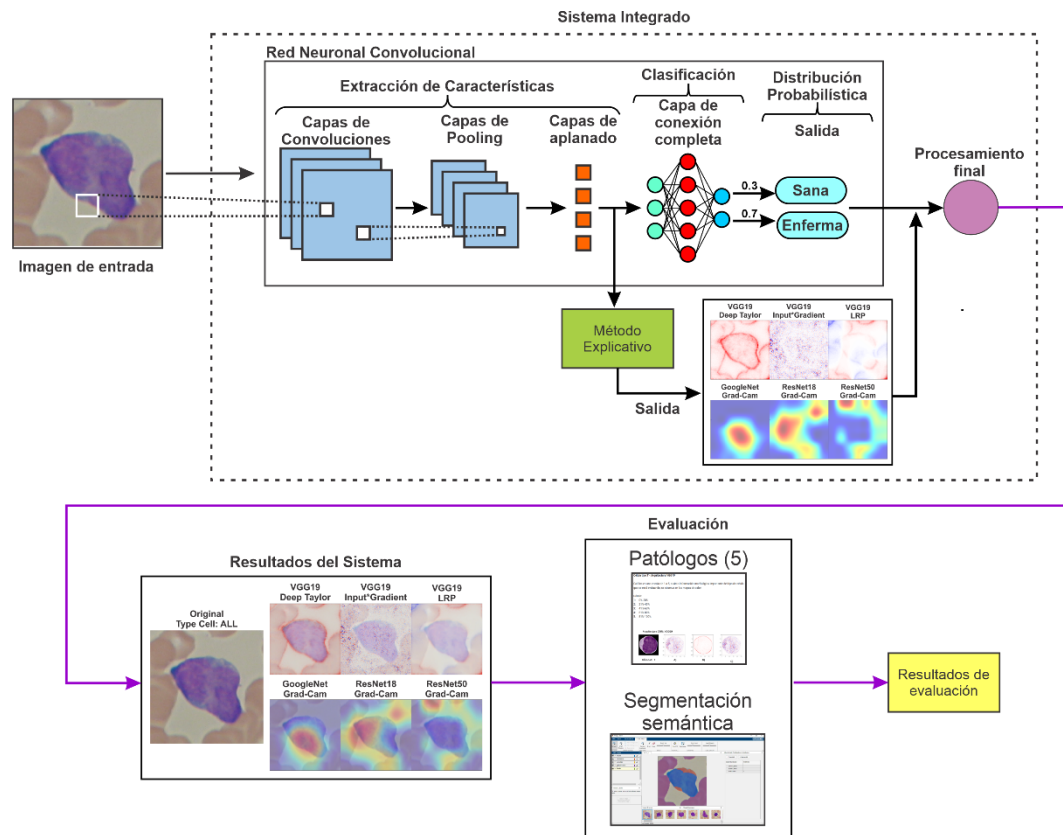


Figura 1. Resumen gráfico del proceso realizado en el presente trabajo.

1.1 Motivación

Según el INEGI [1], en México, 73 de cada 100 pacientes hospitalizados por cáncer son debido a problemas en tejidos linfoides, hematopoyéticos o similares. Parte relevante de estas cifras es la Leucemia Linfoblástica Aguda, que abarca el 61% del total de ingresos por cáncer. Esta enfermedad es predominante en infantes, representando el 80% de todas las leucemias agudas.

Para la atención de esta enfermedad de forma oportuna es crucial establecer estrategias para la detección precoz de la ALL, evitando los errores humanos al realizar el diagnóstico de esta enfermedad, garantizando así un tratamiento adecuado y oportuno para los afectados. Así mismo, reducir las muertes derivadas de un mal diagnóstico.

El presente trabajo está enfocado a la clasificación de ALL en imágenes y a la implementación de los métodos explicativos visuales, en particular a los métodos generadores de mapas de relevancias, como una herramienta auxiliar en el diagnóstico de dicha enfermedad.

1.2 Pregunta de Investigación

¿Cuál es el método generador de mapas de relevancia que puede asociar los píxeles importantes con las características morfológicas más sobresalientes de las células de leucemia linfoblástica aguda (ALL)?

1.3 Hipótesis

La combinación de un modelo de red neuronal convolucional (CNN) avanzado, como ResNet o GoogleNet, junto con algoritmos generadores de mapa de relevancia como Gradient-weighted Class Activation Mapping (Grad-CAM) o Layer-wise Relevance Propagation (LRP), permitirá identificar y relacionar de manera efectiva las características morfológicas distintivas de las células en imágenes de células sanas y enfermas de Leucemia Linfoblástica Aguda (ALL).

1.4 Justificación

La Leucemia Linfoblástica Aguda representa un problema grave en México debido a la alta incidencia de casos recurrentes en los hospitales. El diagnóstico oportuno es crucial para asignar un tratamiento adecuado lo más pronto posible. Una de las etapas en el proceso de diagnóstico consiste en analizar una muestra de sangre del paciente mediante un microscopio, con el objetivo de clasificar las células enfermas según sus características morfológicas.

Se ha demostrado que el uso de modelos de redes neuronales convolucionales (CNN) entrenados para la clasificación de enfermedades ofrece un rendimiento superior en comparación con el de un experto humano. No obstante, resulta imprescindible que los expertos validen los resultados obtenidos por estos modelos. Implementar algún método explicativo agilizaría este proceso de validación.

En este trabajo, se analiza y evalúa la relación entre los mapas de relevancia generados por la clasificación de una CNN y las características morfológicas presentes en la imagen de entrada de una célula sana o enferma de ALL. El objetivo es determinar la viabilidad de emplear estos métodos explicativos visuales como herramientas auxiliares en el diagnóstico de la LLA.

1.5 Objetivos

Objetivo General

Identificar la combinación óptima de un método generador de mapas de relevancia y un modelo de redes neuronales convolucionales (CNN) que establezca la mejor correlación entre los píxeles importantes y las características morfológicas de las células de leucemia linfoblástica aguda (ALL).

Los resultados serán evaluados y validados por hematólogos especializados en clasificación morfológica.

Objetivos Específicos

1. Analizar e identificar las combinaciones óptimas de modelos de redes neuronales convolucionales (CNN) y métodos generadores de mapas de relevancia a implementar.
2. Evaluar los mapas de relevancia obtenidos a través de la colaboración con hematólogos expertos.

3. Evaluar los mapas de relevancia utilizando segmentación semántica, correlacionando los píxeles relevantes con las características morfológicas presentes en las células de ALL.

1.6 Trabajo relacionado.

La comparación de los trabajos relacionados se muestra en la Tabla 1. Se puede observar que los trabajos realizados por los diferentes autores se centran en clasificar imágenes que contienen células de tipo ALL y emplean diferentes técnicas explicativas para relacionar las descripciones con la morfología de la célula estudiada. La mayoría de los autores utiliza modelos CNN, destacándose los modelos ResNet50, VGG e Inception. Por otro lado, tres autores utilizan un método de explicación visual, empleando los métodos explicativos LIME, DCN o Grad-CAM.

Tabla 1. Tabla comparativa de trabajos relacionados

Autor	AL L	Modelo CNN	XAI Método	Colaboración con expertos
N. Jiwani et al. in [2]	Si	No	No	No
Jiang et al. [3]	Si	*Wavelet	No	No
Abir et al. [4]	Si	Resnet50, DenseNet121, y VGG16	LIME	No
Nayeon Kim [5]	Si	InceptionV3, Res-Net101V2, InceptionResNetV2, y VGG19	LIME y DCN	No
Ochoa-Montiel et al. [6]	Si	*Random Forest, LeNet, y AlexNet	No	No
Maaliw R.R. et al. [7]	Si	Transfer learning InceptionV3, y Xception InceptionResNetV2	No	No
Díaz R. J. et al. [8]	Si	Modified ResNet-50	Grad-CAM	No
Velázquez A. J et al.	Si	VGG 16, VGG 19, MobileNet V1, ResNet18 ,ResNet-50, y GoogleNet	Deep Taylor, Gradient*Input, LRP, y GradCAM	Si

En [8] se propone un método de clasificación y explicación. El método propuesto contempla la segmentación de las características morfológicas de las células. Posteriormente, el método utiliza una red ResNet50 que realiza la clasificación, obtiene el respectivo mapa de relevancias empleando GradCAM y genera una explicación de características espaciales. Ambos pasos se consideran métodos explicativos visuales. Además, realizan experimentos de generación de mapas de relevancias, con la célula segmentada y no segmentada. Concluyen que se obtienen mejores resultados cuando el mapa de calor se genera con la célula segmentada.

La principal diferencia entre el trabajo presentado aquí y el trabajo discutido anteriormente radica en evaluar diferentes arquitecturas y métodos explicativos. Las evaluaciones se realizan por médicos expertos en el área, y mediante una evaluación por segmentación semántica. Se realizan experimentos combinando diferentes modelos CNN y métodos generativos de mapas de relevancias. Los experimentos tienen como objetivo identificar qué combinación es la más eficiente para relacionar píxeles relevantes con características morfológicas utilizando la imagen segmentada y no segmentada. En consecuencia, nuestro trabajo difiere sustancialmente del trabajo explicado anteriormente.

1.7 Alcances y Limitaciones

Alcances

En este estudio, se lleva a cabo una evaluación por cinco médicos especialistas en clasificación de ALL, así como una evaluación por segmentación semántica de la combinación de algoritmos generadores de mapas de relevancias y CNN. Los resultados permiten identificar una combinación que, según los expertos, podría ser implementada como una herramienta para el diagnóstico de la ALL.

Limitaciones

Las principales limitaciones del trabajo son las siguientes:

- Bases de datos: Las bases de datos de imágenes de células con ALL que son de acceso libres son escasas. Además, la cantidad de imágenes disponibles es limitada y muchas de ellas presentan un preprocesamiento o alteraciones previas. El equipo y condiciones con las que fueron tomadas las imágenes varían, lo cual implica que las imágenes no están homologadas ni estandarizadas.
- Acceso a médicos expertos. La evaluación por parte de los expertos se realizó con la colaboración de cinco médicos, quienes llevaron a cabo la evaluación de forma remota.

1.8 Estructura del documento

Capítulo 1: Introducción. En este capítulo se exponen las generalidades del trabajo de investigación. Se abordan los objetivos, la pregunta de investigación, la hipótesis, la justificación, los trabajos realizados, así como los alcances y limitaciones.

Capítulo 2: Marco Teórico. Este capítulo se centra en el marco teórico, comenzando con las generalidades de la enfermedad ALL y su clasificación. Se describen las bases de datos empleadas para este estudio, las redes neuronales convolucionales (CNN) reentrenadas, la información sobre los métodos explicativos visuales utilizadas en este trabajo y, finalmente, la segmentación semántica.

Capítulo 3: Metodología. El tercer capítulo se enfoca en la metodología desarrollada en el presente trabajo. Se explica el desarrollo de las dos fases llevadas a cabo para la evaluación de los mapas de relevancias obtenidos.

Capítulo 4: Resultados y Discusión. En este capítulo se presentan los resultados y la discusión del trabajo. El capítulo está dividido en dos secciones:

la primera sección corresponde a los resultados obtenidos por los médicos expertos, y la segunda sección expone los resultados de la evaluación utilizando la segmentación semántica.

Capítulo 5: Conclusiones y Trabajo Futuro. Finalmente, este capítulo recopila las conclusiones obtenidas de este estudio y describe el trabajo futuro que se espera realizar para continuar avanzando en este campo de investigación.

2 Marco Teórico

En este capítulo se abordan conceptos fundamentales para el desarrollo de este proyecto. Primero, se describe la Leucemia Linfoblástica Aguda (ALL) y se presentan sus conceptos generales. A continuación, se revisan las redes neuronales convolucionales (CNN) utilizadas para su reentrenamiento. Luego, se detallan los métodos explicativos visuales empleados en este trabajo, y finalmente, se describe la segmentación semántica utilizada para realizar una de las evaluaciones.

2.1 Leucemia Linfoblástica Aguda (ALL)

Las leucemias agudas son un tipo de cáncer de la sangre que se origina de la alteración en el crecimiento y desarrollo de células inmaduras, ya sean de linaje mieloide o linfoide. Estas células reemplazan a las células normales de la médula ósea que producen sangre. Las leucemias agudas de tipo linfoide son más comunes en niños, mientras que las mieloides son más frecuentes en adultos. Clínicamente, se identifican por la presencia de anemia, tendencia a sangrar debido a la baja de plaquetas, y susceptibilidad a infecciones, estas últimas a menudo debido a la neutropenia que suelen presentar estos pacientes. Los exámenes de laboratorio desempeñan un papel crucial en el diagnóstico de la enfermedad y en el seguimiento de su progreso.

De acuerdo con Iván Palomo et al. en [9], la leucemia linfoblástica aguda (ALL) es la neoplasia más frecuente de la infancia, constituyendo un 35 a 40 % de los cánceres en la edad pediátrica, con una incidencia anual de 3 a 4 casos por cada 100.000 niños menores de 15 años. Se presenta en todas las edades con un pico entre los 4 y 6 años, siendo más frecuente en varones. La causa sigue siendo desconocida, si bien en un número muy pequeño de casos se asocia a síndromes genéticos conocidos. Hasta mediados de los años 60, era una enfermedad fatal, pues 80% de los pacientes fallecían antes de 6

meses. Con la mejoría en el diagnóstico y en la calidad del tratamiento, se ha logrado la curación en más del 70% de los casos.

En la ALL, una célula progenitora linfoide sufre una alteración genética adquirida cuya consecuencia es una expansión clonal con detención de la diferenciación celular, proliferación, crecimiento descontrolado e invasión de la médula ósea; desde allí, se disemina a la sangre periférica, bazo, ganglios y el resto de los tejidos, haciéndose clínicamente detectable. Esta célula maligna o blasto leucémico comparte muchas de las características de los progenitores linfoides, de modo que la mayoría de ellos tiene un fenotipo que corresponde a la etapa de diferenciación en la que se detuvo el desarrollo.

Hay factores clásicamente asociados a la patogénesis de la leucemia como la exposición a radiación ionizante, algunos químicos (benceno, pesticidas) o agentes alquilantes, pero excepcionalmente se encuentran como antecedente en el paciente pediátrico. Del mismo modo, hay condiciones genéticas que tienen un mayor riesgo de presentar leucemia como trisomía 21 (1 cada 95), anemia de Fanconi (1 cada 12), síndrome de Bloom (1 cada 8), agammaglobulinemia congénita y enfermedad de Kostman. Estos cuadros son muy poco frecuentes, a excepción de la trisomía 21, de modo que la mayoría de los niños que presentan leucemia eran sanos previamente. En un lactante gemelo con leucemia, ésta puede presentarse también en el segundo gemelo, y el estudio de esta situación ha dado importante información para explicar la patogénesis de la ALL.

En [10], Emadi menciona que la leucemia linfoblástica aguda (ALL), es el cáncer pediátrico más frecuente, que también afecta a adultos de todas las edades. La transformación maligna y la proliferación no controlada de una célula progenitora hematopoyética con diferenciación anormal y de supervivencia prolongada determinan un alto número de blastocitos circulantes, reemplazo de la médula ósea normal por células malignas y posibilidad de infiltración leucémica del sistema nervioso central y los

testículos. Los síntomas son cansancio, palidez, infección, dolor óseo, síntomas del sistema nervioso central (p. ej., cefalea), hematomas de fácil aparición y hemorragias. El examen del frotis de sangre periférica y de la médula ósea suele ser un método de diagnóstico.

2.1.1 Examen y diagnóstico

La anemia es un proceso prácticamente constante en el paciente con leucemia. Por lo general, la cifra de leucocitos se halla aumentada (leucocitosis), aunque en algunos casos puede estar disminuida (leucopenia). La cifra de plaquetas es inferior a lo normal (trombopenia). Cuando se examina una muestra de sangre del paciente al microscopio, se observan glóbulos blancos inmaduros (blastos). Una biopsia de médula ósea servirá para confirmar el diagnóstico y determinar qué tipo de leucemia sufre el paciente.

Los trastornos bioquímicos que se registran con mayor frecuencia son hiperuricemia (ácido úrico elevado), hipocalcemia (calcio bajo), hiperfosfatemia (fósforo elevado), hiperpotasemia (potasio elevado) e incremento de la actividad sérica de la láctico-deshidrogenasa (LDH). Estas alteraciones se observan sobre todo en los casos con leucocitosis, grandes visceromegalias o adenopatías, y reflejan el elevado recambio celular. En algunos enfermos se detecta hipogammaglobulinemia (bajo nivel de gammaglobulinas).

Las pruebas médicas realizadas para detectar leucemias son: Biopsia de médula ósea, análisis de sangre, citometría de flujo, frotis de sangre periférica, punción lumbar, tomografía computarizada, resonancia magnética, ecografía, entre otras.

2.1.2 Clasificación FAB y morfología

La clasificación morfológica franco-estadounidense-británica (FAB) fue creada por un grupo de franceses, americanos y británicos con el objetivo de

clasificar la leucemia según el tipo de célula que origina la enfermedad y el grado de madurez que presentaba la célula en ese momento [11]. Se basa en las características de los blastos, incluido el tamaño celular, la relación núcleo-citoplasma, el número y tamaño de los nucléolos y el grado de basófila citoplasmática (Figura 2). En esta clasificación encontramos 3 tipos para las linfoblásticas L1, L2 y L3, y 8 subtipos diferentes para las mieloblásticas, que van del M0 al M7, [12].

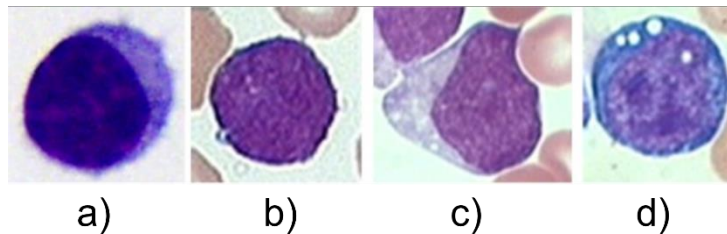


Figura 2. Variabilidad morfológica asociada a las células blásticas según la clasificación FAB: a) células linfocitos sanas de pacientes sin ALL, b-d) linfoblastos de pacientes con ALL donde b), c) y d) son L1, L2 y L3 respectivamente. Tomado de [13]

La definición de los subtipos inmunológicos de ALL depende de la presencia o ausencia de diversos antígenos citoplasmáticos y de superficie celular. Una clasificación comúnmente utilizada divide la ALL en subtipos tempranos de células pre-B, de células B y de células T. La ALL de células B maduras suele tener morfología de L3. La clasificación actual de la OMS divide la mayoría de los subtipos de ALL en leucemia/linfoma linfoblástico B o T bajo el título “neoplasias linfoides precursoras”. La ALL-B madura se incluye como linfoma de Burkitt.

En la siguiente lista se describen de forma general cada una de las clases según la clasificación FAB:

- **Leucemia Linfoblástica Aguda (ALL):**
 - L1: Células blásticas pequeñas y uniformes con escaso citoplasma.

- L2: Células blásticas grandes y heterogéneas con nucléolos y relación nuclear-citoplasmática baja.
- L3: Células blásticas vacuoladas basófilas. Blastos tipo célula de Burkitt

- **Leucemia Mieloblástica Aguda (AML):**

- M0 Leucemia mieloblástica aguda indiferenciada.
- M1 Leucemia mieloblástica aguda con maduración mínima.
- M2 Leucemia mieloblástica aguda con maduración.
- M3 Leucemia promielocítica aguda (APL).
- M4 Leucemia mielomonocítica aguda con eosinofilia.
- M5 Leucemia monocítica aguda.
- M6 Leucemia eritroide aguda.
- M7 Leucemia megacarioblástica aguda.

En la leucemia aguda, la cromatina nuclear puede ser reticular o no, los nucléolos pueden estar ausentes o presentes, y el color del nucléolo y su morfología de estos puede variar con la tinción. Todos estos cambios se pueden inducir simplemente cambiando la tinción empleada sobre la misma población blástica. Las características citoplasmáticas se usan menos frecuentemente para distinguir células blásticas. Sin embargo, los cuerpos de Auer o los gránulos citoplasmáticos son decisivos en un diagnóstico dado, aunque estos también sujetos a las variaciones técnicas de la tinción panóptica [12].

Los autores en [14], [15] y [16] sostienen que las células benignas y malignas pueden discriminarse por su estructura nuclear, relación núcleo-citoplasma, color y textura. La Figura 3 ilustra un ejemplo de la diferencia entre la estructura de células sanas y cancerosas. Muestra que las células cancerosas tienen una estructura irregular y la forma del núcleo está colapsada; así es como los hematólogos describen la ALL.

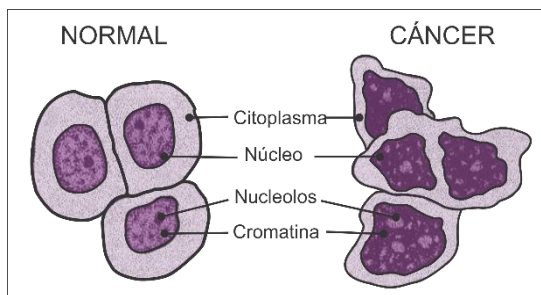


Figura 3. Ejemplo de las diferencias de estructuras entre células ALL y células sanas.

En la Tabla 2, se muestra la clasificación de las leucemias agudas de diferenciación linfóide. Se describe con mayor detalle las características morfológicas que se pueden presentar en las células de ALL según la clasificación FAB.

Tabla 2. Clasificación de las leucemias agudas de diferenciación linfóide.

Clasificación morfológica de la ALL según FAB			
Características	L1	L2	L3
Tamaño Celular	Pequeño, hasta 2 veces el linfocito	Grande, Heterogéneo en tamaño	Grande y homogéneo
Cromatina nuclear	Homogénea en cada caso	Heterogénea	Finamente punteada y homogénea
Forma del núcleo	Regular, indentaciones ocasionales	Irregular. Identación	Regular, oval o redondo
Nucleolos	No visible o pequeño	Una o más veces grandes	Prominente o más vesiculoso
Extensión del citoplasma	Escaso	Variable, a veces moderadamente abundante	Moderadamente abundante con múltiples vacuolas que se superponen incluso al núcleo
Basófila del citoplasma	Ligera o moderadamente intensa	Variable	Muy intensa
Vacuolización del citoplasma	Variables	Variables	A menudo prominentes
PAS (Periodic Acid-Schiff)	Positivas en porcentaje elevado	Escasamente positivas	Negativa

De acuerdo con [17], Las leucemias linfoblásticas agudas son las leucemias en las que las células no se parecen morfológicamente a mieloblastos, promielocitos o monocitos, ni muestran el patrón citoquímico correspondiente. Los atributos comunes son un núcleo celular generalmente ligeramente más pequeño y una estructura de cromatina más densa, cuya consistencia granulada sólo puede distinguirse con una técnica de frotis óptima (es decir, muy ligera). La clasificación como ALL se basa en las similitudes (a menudo remotas) de las células con los linfocitos o linfoblastos de los ganglios linfáticos y en su comportamiento como marcador celular inmunológico.

Un análisis morfológico insuficientemente detallado también puede dar lugar a una posible confusión con la leucemia linfocítica crónica (CLL), pero el análisis de marcadores de superficie celular corregirá este error. El diagnóstico avanzado comienza con pruebas de peroxidasa y esterasa en frotis frescos, realizadas en un laboratorio de hematología, junto con estudios de marcadores inmunológicos realizados en muestras de sangre fresca heparinizada en un laboratorio especializado.

2.2 Base de datos.

En esta sección se discutirán las dos bases de datos empleadas para llevar a cabo los reentrenamientos de las redes neuronales. La primera base de datos contiene imágenes segmentadas en la que la célula de interés ha sido segmentada del fondo. La segunda base de datos, por otro lado, incluye imágenes en las que las células de interés no han sido segmentadas del fondo.

2.2.1 Base de datos The Cancer Imaging Archive

La primera base de datos empleada fue publicada por The Cancer Imaging Archive [18]. Las imágenes utilizadas constan de 363 imágenes de

células segmentadas del fondo de tamaño 254 x 254 x 3 píxeles, en la Figura 4 se muestra un ejemplo de cada categoría que conforman la base de datos.

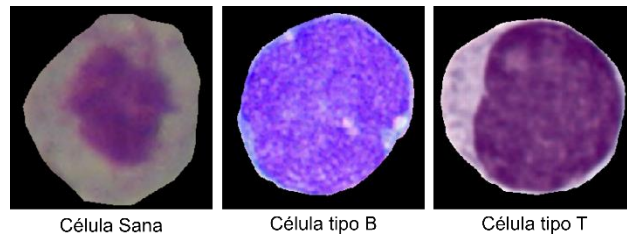


Figura 4. Ejemplo de imágenes de la base de datos The Cancer Imaging Archive.

El conjunto de imágenes se distribuyó en tres categorías: 135 imágenes para la categoría de células B (ALL-B), 135 para la clase de células T (ALL-T) y 93 para la categoría de células sanas.

2.2.2 Base de datos ALL-IDB2

La base de datos de imágenes de leucemia linfoblástica aguda para el procesamiento de imágenes (ALL-IDB) [13] es de acceso público y presenta categorías equilibradas. Contiene imágenes microscópicas de muestras de sangre y destinada a evaluar y comparar algoritmos de segmentación y clasificación de imágenes en Leucemia Linfoblástica Aguda (ALL). Cada imagen de la base de datos fue identificada y clasificada por un grupo de oncólogos con experiencia en detectar las células enfermas de ALL.

Las fotografías están divididas en dos categorías: enfermas (ALL) y sanas, con 130 imágenes para cada categoría. Las imágenes tienen una resolución de 257 x 257 píxeles en RGB. En este trabajo de tesis, se dividieron las imágenes en 100 por categoría. Con el subconjunto restante de imágenes, 30 por categoría, se creó una sección de imágenes nunca antes vistas, que se utilizará exclusivamente para evaluar los modelos tras todo el proceso de reentrenamiento y generar mapas de relevancia.

Como se describe en [6], los conjuntos de datos típicos utilizados para el reconocimiento de células leucémicas presentan inconvenientes relacionados con el desequilibrio de categorías. En otros casos, se construyeron a partir de diferentes fuentes o condiciones de adquisición. Las células leucémicas incluyen dos o más categorías de leucemia, así como sus subtipos (leucemia linfocítica y leucemia mieloide, crónica o aguda). Para nuestro propósito, este trabajo está enfocado al tipo ALL.

El conjunto de datos ALL-IDB2 [13] es pequeño; sin embargo, es uno de los conjuntos de datos más utilizados, como se menciona en [6], [7] y [16], debido a su acceso público y al equilibrio de sus categorías. Por este motivo, se seleccionó el conjunto de datos ALL-IDB2 para este trabajo. En la Figura 2 se muestra un ejemplo de las imágenes que componen esta base de datos.

2.3 Arquitecturas preentrenadas

En el área del Aprendizaje Profundo (Deep Learning), se utilizan técnicas de reentrenamiento de arquitecturas de redes neuronales convolucionales (CNN) preentrenadas. El objetivo es entrenar un nuevo modelo aprovechando las características extraídas por la arquitectura preexistente, acelerando así el proceso de aprendizaje para una nueva tarea de clasificación específica.

En esta investigación, se implementaron las siguientes arquitecturas: ResNet50, VGG16, VGG19 y MobileNet v1 para la primera fase. En la segunda fase, se añade la arquitectura GoogleNet. Se aplicó la transferencia de aprendizaje utilizando la técnica de fine-tuning. Esto se debe a que la base de datos empleada es pequeña, y aplicar esta técnica permite obtener altos valores de precisión sin necesidad de una gran capacidad de cómputo, como se menciona en la referencia [19].

Es importante mencionar que se utilizaron arquitecturas preentrenadas de tamaño pequeño, mediano y grande. Esto se hizo con el propósito de

evaluar el rendimiento de cada una de ellas según el tamaño y considerar las posibles aplicaciones en función de los resultados obtenidos.

2.3.1 Arquitecturas VGG-16 y VGG-19

Las arquitecturas VGG Net, especialmente los modelos VGG-16 y VGG-19, son redes neuronales convolucionales desarrolladas por el Visual Group de la Universidad de Oxford [20]. Estas redes se caracterizan por utilizar kernels de convolución de 3x3 en todas sus capas y un paso de convolución constante. Esto permite aumentar la profundidad de la red sin generar un exceso de parámetros.

La estructura general de una arquitectura VGG consiste en 5 conjuntos de capas convolucionales seguidas de una capa de Max-pooling. La diferencia entre los diferentes modelos radica en el número de capas convolucionales en cascada en cada conjunto. Por ejemplo, el modelo VGG-16 tiene 13 capas de convolución y 3 capas totalmente conectadas, mientras que el modelo VGG-19 cuenta con 16 capas de convolución y 3 capas totalmente conectadas.

Tanto el modelo VGG-16 como el VGG-19 utilizan la función de activación ReLU en todas sus capas, y la capa de salida aplica la función Softmax. El modelo VGG-16 cuenta con 133 millones de parámetros y tiene una tasa de error del 8.8%, mientras que el modelo VGG-19 tiene 144 millones de parámetros y una tasa de error del 9%.

Estas arquitecturas comparten las últimas tres capas totalmente conectadas, lo que permite que cada capa esté conectada a la capa anterior. Esto ayuda a preservar características útiles y significativas durante la propagación, eliminando características no relevantes.

Las arquitecturas VGG están diseñadas para procesar imágenes de tres canales RGB con dimensiones de 224 x 224 píxeles. Cada una de estas arquitecturas es capaz de clasificar imágenes en 1000 categorías diferentes

de objetos, gracias a su entrenamiento utilizando la base de datos de ImageNet [21].

En la Figura 5 y Figura 6 se muestra la estructura de las capas que conforman a cada una de estas arquitecturas.

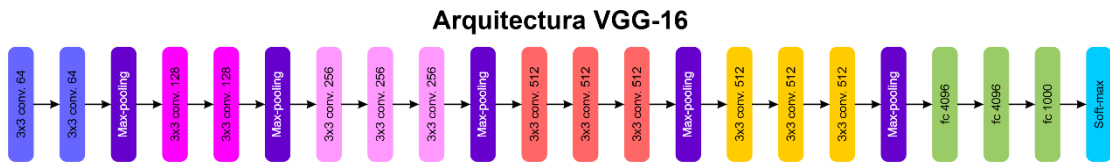


Figura 5. Representación de la arquitectura VGG-16

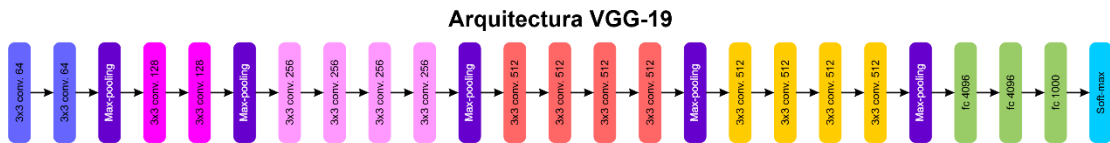


Figura 6. Representación de la arquitectura VGG-19

2.3.2 Arquitectura MobileNet V1

La arquitectura preentrenada de CNN MobileNet es ideal para aplicaciones de visión móvil debido a su uso de convoluciones separables en profundidad. En este diseño, se utiliza la convolución puntual para fusionar las salidas de la convolución profunda a través de una convolución 1x1.

La arquitectura general de MobileNet consta de 30 capas, como se muestra en la Figura 7. Comienza con una capa convolucional con un paso de 2, seguida de una capa en profundidad y una capa puntual que duplica el número de canales, entre otras características.

Debido a que MobileNet ha sido diseñada específicamente para aplicaciones móviles, el gasto y la velocidad de computación están directamente relacionados con la cantidad de multiplicaciones y acumulaciones utilizadas. Esta arquitectura se basa en otras arquitecturas como ResNet, incorporando convoluciones 1x1 y convoluciones 3x3 para

lograr una significativa reducción en el costo computacional y la recuperación de la dimensión del canal de salida.

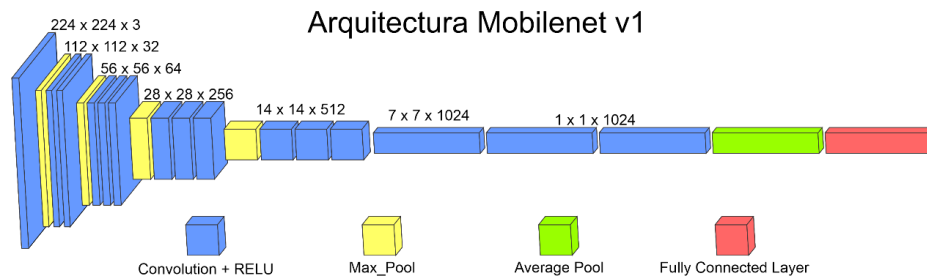


Figura 7. Representación de la arquitectura MobileNet V1.

2.3.3 Arquitectura ResNet50

Otra de las arquitecturas implementadas es la conocida como ResNet50. Esta estructura de red neuronal convolucional se basa en módulos de microarquitectura llamados "redes en red", lo que la distingue de las arquitecturas secuenciales como VGG o AlexNet. ResNet50 fue desarrollada por Kaiming He, Xiangyu Zhang, Shaoqing Ren y Jian Sun, y ganó el concurso de clasificación ILSVRC de 2015. Actualmente, es reconocida como una de las redes convolucionales más populares.

Esta arquitectura consta de un total de 152 capas que incluyen conexiones residuales, lo que permite la conexión de una capa con la capa anterior (ver Figura 8). Esto significa que estas capas pueden aprender características relevantes de las capas anteriores, un aspecto que la diferencia de arquitecturas como VGG.

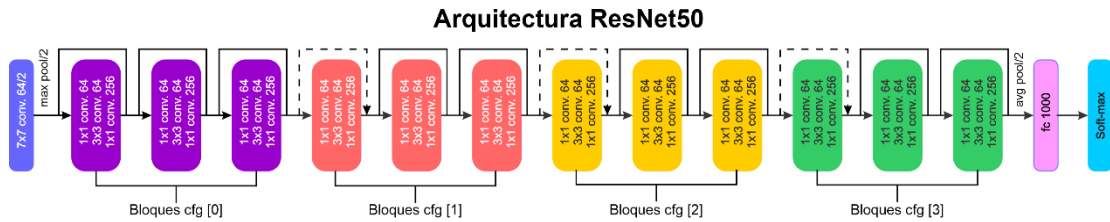


Figura 8. Representación de la arquitectura ResNet 50.

2.3.4 Arquitectura GoogleNet

GoogleNet, también conocida como Inception Network, es una arquitectura de red neuronal convolucional diseñada por Szegedy et al. de Google. Una de las características más destacadas de GoogleNet es su estructura profunda y eficiencia en términos de computación y memoria.

La estructura GoogleNet está compuesta por un total de 22 capas, que incluyen capas convolucionales y capas totalmente conectadas (ver Figura 9). Sin embargo, a diferencia de las arquitecturas convencionales, GoogleNet implementa módulos de Incepción, lo que permite una extracción de características más eficiente y diversa. Cada módulo de Inception toma la entrada y la procesa a través de diferentes tamaños de filtros de convolución (1x1, 3x3 y 5x5) y una operación de Max-Pooling (3x3). Las salidas de estos filtros se concatenan en la salida. Esta arquitectura tiene como objetivo permitir que la CNN aprenda características a diferentes escalas y abstracciones simultáneamente.

Otra característica novedosa de la arquitectura es el uso de "clasificadores auxiliares". Estos son redes convolucionales más pequeñas ubicadas en la mitad de la red principal. La idea detrás de estos clasificadores auxiliares es que la red principal "no sabe" si está en la capa de entrada o en una de las redes de clasificación, por lo que debe aprender características generalizadas.

Es importante destacar que GoogleNet no cuenta con una capa completamente conectada al final de la red, a diferencia de muchas otras arquitecturas. En cambio, utiliza una capa especial llamada Global Average Pooling, lo que resulta en un modelo más liviano y con menor riesgo de sobreajuste.

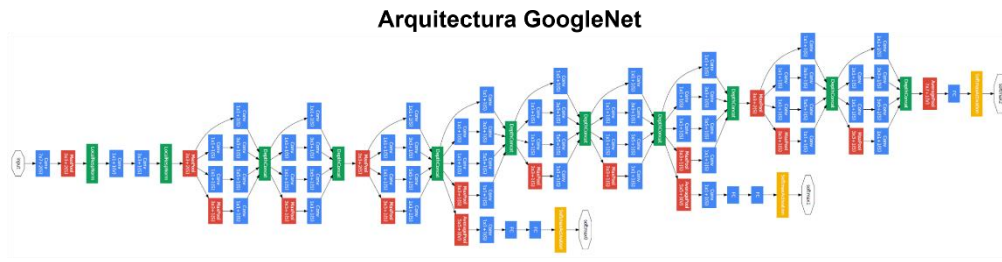


Figura 9. Representación gráfica de la arquitectura GoogleNet.

En la actualidad, se han publicado investigaciones que sugieren el uso de modelos de redes neuronales para la clasificación de células de leucemia linfoblástica aguda (ALL). Estos modelos han demostrado una precisión superior a la capacidad humana en la identificación y clasificación de esta patología particular. Dada la eficiencia notoria de estos modelos, es crucial adentrarse en el estudio de cómo estos sistemas de inteligencia artificial toman sus decisiones. Esto resalta la importancia de avanzar hacia una inteligencia artificial explicativa, donde se pueda comprender y explicar de manera clara el proceso de toma de decisiones de estos modelos.

2.4 Métodos Explicativos

La inteligencia artificial explicable (IAE) es un campo de investigación emergente que busca brindar transparencia a los modelos de redes neuronales convolucionales, algoritmos altamente complejos y opacos comúnmente conocidos como cajas negras debido a la falta de transparencia en el proceso de obtención de resultados. Aunque se han desarrollado múltiples métodos para explicar las decisiones de los clasificadores CNN, estas herramientas rara vez se utilizan más allá de la visualización.

Recientemente, los investigadores han comenzado a utilizar explicaciones visuales, como los mapas de relevancias, para abordar este problema.

La creciente digitalización de los flujos de trabajo en patología diagnóstica ha abierto la puerta a aplicaciones de inteligencia artificial (IA) que salvan vidas. Una rama de implementación de estos métodos explicativos se encuentra en el área médica, donde se utilizan como una herramienta auxiliar para corroborar las predicciones realizadas por las redes neuronales.

En este trabajo se emplean métodos generadores de mapas de calor, estos son considerados como una forma de explicar una red neuronal. Pero ¿Qué se entiende por Inteligencia Artificial Explicable (XAI)? Según Barredo et. al. [22], se definen tres conceptos que a menudo se utilizan incorrectamente e intercambian de manera incorrecta. Como resultado de su investigación llegan a las siguientes definiciones:

- Interpretabilidad (Interpretability, en inglés): se refiere a la capacidad de explicar o proporcionar un significado en términos comprensibles para los seres humanos.
- Explicabilidad (Explainability, en inglés): está asociada con la noción de explicación como una interfaz entre los humanos y el tomador de decisiones. Debe ser tanto un representante preciso del tomador de decisiones como comprensible para los humanos.
- Transparencia (Transparency, en inglés): Se considera que un modelo es transparente si es comprensible por sí mismo. Los modelos transparentes se dividen en tres categorías: modelos simulables, modelos que se puedan descomponer y modelos algorítmicamente transparentes.

La explicabilidad se identifica como un componente crítico para la seguridad, aprobación y aceptación de los sistemas de IA para uso clínico. En el trabajo de Weber et. al. [23], se presenta una descripción general de las técnicas que aplican XAI de manera práctica para mejorar diversas

propiedades de los modelos de aprendizaje automático (ML). Estas técnicas se clasifican sistemáticamente, comparando sus respectivas fortalezas y debilidades respectivas.

Dado que los modelos entrenados en este estudio están destinados al uso médico, es imperativo proporcionar una explicación detallada de los resultados obtenidos. Esto se aplica especialmente a la posible discriminación, perjuicio o daño que los resultados podrían causar en un contexto hospitalario. Por lo tanto, este sistema utiliza métodos de explicaciones visuales específicamente diseñados para las CNNs. Estos métodos explicativos permiten la generación de mapas de relevancias y de prominencia, que reflejan la relevancia de píxeles específicos en la imagen de entrada en relación con la clasificación realizada por el modelo. Además, proporcionan una representación visual de las áreas de la imagen que fueron cruciales para la clasificación, así como la identificación de áreas del modelo que no están funcionando de manera óptima.

Los métodos visuales empleados en este estudio se consideran métodos Post Hoc, ya que se agregan después del entrenamiento de la CNN. Para lograr esto, es necesario acceder a la información de la red neuronal antes de la capa de conexión completa (Fully Connected Layer, en inglés), donde generalmente se emplea la función Softmax.. En la Figura 10 se muestra una representación gráfica de este proceso.

En este estudio, se emplean los métodos explicativos proporcionados por el toolbox iNNvestigate [24], Los cuales permiten la generación de mapas de relevancias para identificar las áreas de interés relacionadas con la clasificación realizada por los modelos de redes neuronales utilizados.

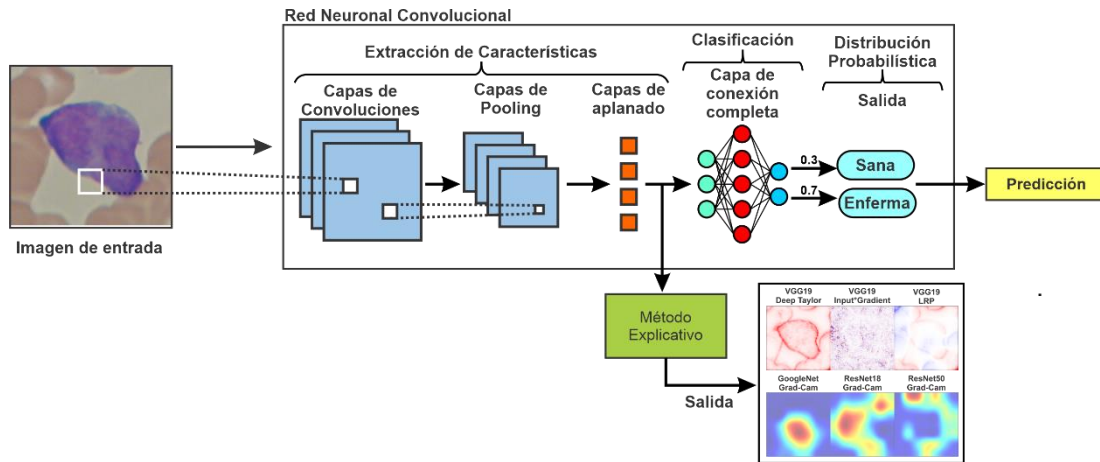


Figura 10. Representación gráfica del proceso de generación de mapas de calor.

2.4.1. Método Gradient* Input

El método de atribución Gradient*Input [24], descrito en la ecuación (1), es una técnica simple que emplea la información del gradiente de una función como su nombre lo indica. Aplicar el gradiente determina para cada dimensión de entrada si la función aumenta o disminuye si damos un pequeño paso en esa dirección. Este método multiplica el gradiente con la entrada para obtener la relevancia de cada píxel en la clasificación realizada por la red neuronal convolutiva.

$$R_{i,n} = x_{i,n} \frac{\partial \hat{F}}{\partial X_i} \Big|_{X_i = x_{i,n}} \quad (1)$$

En la ecuación (1) $x_{i,n}$ representa la característica de entrada de una parte específica de la muestra que se multiplica por el gradiente. Este método cuantifica la atribución de la salida con la entrada, siendo distinta de la sensibilidad de la salida con la de entrada. Esto significa que durante este proceso de retropropagación (backpropagation), se mantienen las relevancias hasta que se alcanza el espacio de entrada.

2.4.2. Método Layer-wise Relevance Propagation (LRPz)

El propósito del método LRPz [25], es utilizar los pesos de la red y las activaciones neuronales generadas durante el avance para propagar la salida a través de la red neuronal hasta la capa de entrada. De esta manera, es posible visualizar qué píxeles han contribuido realmente a la salida. A cada píxel o neurona intermedia que contribuye se le asigna una magnitud de relevancia, denotada como R de la capa de entrada. LRPz es una técnica conservadora, lo que significa que la magnitud de cualquier salida se conserva a lo largo del proceso de retropropagación y es igual a la suma del mapa de relevancia R de la capa de entrada. Esta propiedad se mantiene para cualquier capa consecutiva j y k , y por transitividad para la capa de entrada y salida.

Este método se basa en el gradiente multiplicado por la fórmula de entrada, como se describe en la ecuación (2). La relevancia $R_j^{(l+1)}$ se calcula a partir del nodo de salida j en la capa de superior $(l + 1)$ mientras que $R_i^{(l)}$ representa la relevancia de la neurona i en la capa inferior l . Mediante la retropropagación, se recorren todas las capas desde la salida hasta la entrada. En la regla de LRPz, la propagación se basa en las preactivaciones z_{ij} durante el tiempo de predicción, y la agregación $z_j = \sum_j z_{ij} + b_j$ en la neurona j . Existen diversas formas de calcular y propagar la relevancia, pero todas siguen esa misma idea general.

$$R_i^{(l)} = \sum_j \frac{z_{ij}}{z_j} R_j^{(l+1)} \quad (2)$$

2.4.3. Método Deep Taylor

El método Deep Taylor [26] permite el cálculo de gradientes con respecto a la entrada utilizando la ecuación (3):

$$R_i^{(l)} = \sum_j \frac{\partial R_j^{(l+1)}}{\partial x_i} \Big|_{x_i = \hat{x}_i^j} (x_i - \hat{x}_i^j), \quad (3)$$

Donde \hat{x}_i^j calcula el punto raíz cercano a la entrada x_i , en el cual la relevancia de la neurona es cero. Esto se logra mediante la resta $(x_i - \hat{x}_i^j)$, lo que nos permite calcular de manera recursiva la relevancia de las neuronas de la capa inferior. La redistribución de relevancias es calculada mediante $R_j^{(l+1)}$ que representa la relevancia de la neurona j .

Este método se basa en la expansión de la serie Taylor y es especialmente útil para calcular gradientes de manera eficiente en redes profundas.

2.4.4. Método Grad-CAM

La explicación visual Grad-CAM (Gradient-weighted Class Activation Mapping, [27]) es una técnica utilizada para visualizar y comprender qué partes de una imagen son importantes para la clasificación realizada por un modelo de red neuronal convolucional (CNN). A diferencia de otras técnicas de explicación que se enfocan en la importancia de las características de entrada individuales, Grad-CAM se centra en la importancia de las regiones específicas de la imagen.

Grad-CAM utiliza el gradiente de la salida de la capa convolucional final con respecto a la salida de la clase objetivo para calcular la importancia de cada píxel en la imagen. Luego, se realiza un promedio ponderado de los mapas de relevancia en cada canal convolucional para generar un mapa de activación que resalta las regiones relevantes de la imagen para la clasificación, como se define en la ecuación (4).

Esta técnica permite una visualización más intuitiva y explicativa, ya que nos permite identificar las regiones en la imagen que contribuyen de manera significativa a la clasificación realizada por el modelo. Grad-CAM ha demostrado ser una herramienta útil en el campo de la interpretación y explicación de modelos de aprendizaje profundo.

$$\alpha_k^c = \frac{1}{z} \sum_j \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \quad (4)$$

2.5 Segmentación Semántica

La segmentación semántica en el análisis de imágenes es una técnica que tiene como objetivo dividir una imagen en regiones semánticas significativas. A diferencia de la segmentación tradicional, que divide la imagen en regiones basadas en características como el color o la textura, la segmentación semántica asigna una etiqueta semántica a cada píxel de la imagen. Esto permite identificar y diferenciar objetos específicos y sus clases.

La segmentación semántica es una tarea fundamental en el campo de la visión por computadora, ya que proporciona un nivel más alto de comprensión de la estructura y contenido de una imagen. Al asignar etiquetas semánticas a los píxeles, se logra una comprensión más detallada de la escena. Esto permite reconocer objetos, segmentar objetos en primer plano y fondo, identificar áreas de interés y realizar análisis más avanzados, como seguimiento de objetos en movimiento o reconocimiento de acciones humanas.

Para lograr la segmentación semántica, se utilizan técnicas de aprendizaje automático y redes neuronales convolucionales, que han demostrado ser eficaces en la tarea al permitir la extracción de características relevantes de la imagen y su asociación con clases semánticas. Además, existen diversos enfoques que han mejorado significativamente la precisión y

el rendimiento de la segmentación semántica en aplicaciones prácticas, como reconocimiento de objetos, conducción autónoma, análisis médico y muchas otras áreas.

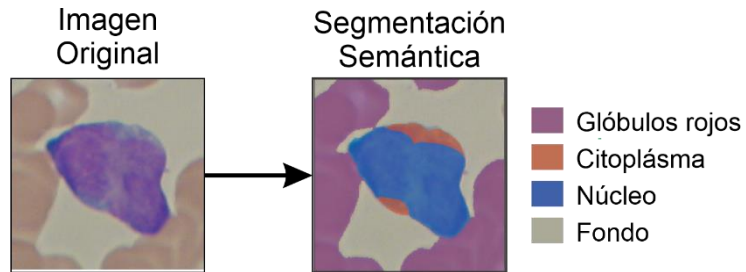


Figura 11. Ejemplo de Segmentación Semántica en imágenes de ALL.

En este trabajo se emplea la segmentación semántica manual para crear los datos que se considerarán como “Ground Truth”. Estos datos servirán para realizar la evaluación de los mapas de relevancia generados en la segunda fase. En la sección siguiente se describe con más detalles el proceso realizado.

3 Metodología

En la sección actual se detalla la ejecución de la metodología y sus etapas correspondientes. Se resaltan las acciones llevadas a cabo, como la organización de las bases de datos utilizadas, el fine-tuning de las redes neuronales convolucionales y la utilización de los métodos explicativos para la creación de los mapas de relevancias. Por último, se describen las evaluaciones de los mapas resultantes.

Como se muestra en el resumen ilustrativo (consultar Figura 1), se llevaron a cabo dos evaluaciones, una por cada fase en el desarrollo de esta tesis. La primera evaluación fue realizada por hematólogos expertos. La segunda evaluación se basó en la segmentación semántica, utilizando este criterio como medida de veracidad validado por los expertos.

3.1 Fase 1

A continuación, se describen los procesos llevados a cabo en esta primera fase. Primero, se realiza la preparación de la base de datos The Cancer Imaging Archive, descrito en la sección 2.2.1, Luego, se procede al reentrenamiento de los modelos VGG-16, VGG-19, MobileNet v1 y ResNet50.

Posteriormente, se describen en detalle la generación de los mapas de relevancias, los cuales se obtienen a partir de la combinación de los modelos mencionados anteriormente y los métodos Gradient*Input, LRP y Deep Taylor.

Por último, se especifica la evaluación de los mapas de relevancias realizados por los cinco hematólogos expertos.

3.1.1 Preparación de la base de datos segmentado

En primer lugar, para la ejecución de este proyecto, se particionó el conjunto de datos “The Cancer Imaging Archive”, como se mencionó

previamente en el Capítulo 2, dicha partición se comprende por 3 subcarpetas, las cuales corresponden a cada tipo de célula (B CELL, TCELL y NORMAL), cada carpeta contenía 135, 135 y 93, respectivamente, sumando un total de 363 imágenes.

Debido a que el conjunto de imágenes es pequeño, se ejecutó un algoritmo para generar un aumento de datos tradicional, aplicando transformaciones de rotaciones de 90° , 180° y 270° para todas las categorías. Sin embargo, como en la categoría Sanas la cantidad de imágenes era menor, a este conjunto de imágenes se le realizó una transformación de tipo espejo a la imagen original. Esto con la finalidad de que la base de datos quedara lo más balanceada posible. En la Figura 12 se muestra un ejemplo de las transformaciones realizadas.

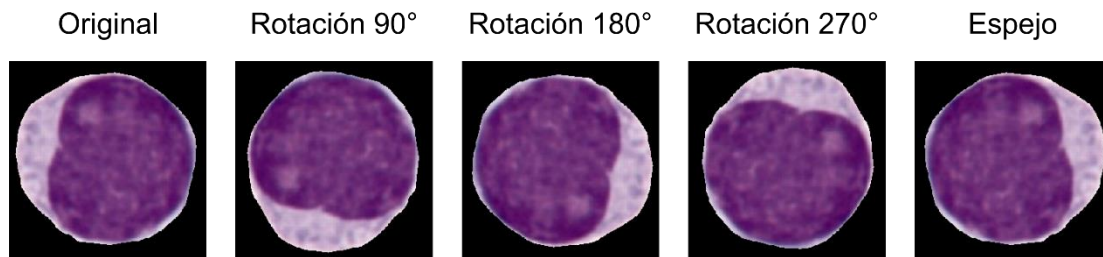


Figura 12. Ejemplo de las transformaciones aplicadas a la base de datos.

Después de finalizar el aumento de datos, se obtuvo un total de 1545 imágenes en la nueva base de datos. Esta base de datos se dividió en tres grupos: entrenamiento, prueba y validación. Se asignó un 80% para el grupo de entrenamiento, un 10% para el grupo de validación y un 10% para el grupo de prueba, lo que representa 1237, 154 y 154 imágenes respectivamente.

3.1.2 Reentrenamiento de modelos

En esta etapa del proyecto se realiza el reentrenamiento de modelos Visual Geometry Group 16 y 19 (VGG16 y VGG19), MobileNet V1 y la Residual

Neural Network 50 (ResNet50). El reentrenamiento se lleva a cabo aplicando la técnica de fine-tuning. En primer lugar, se entrena el modelo VGG16 utilizando 16 capas, de las cuales 13 son capas convolucionales para la normalización y el batch. Se incorporaron 3 capas dense y capas de aplanamiento, resultando en un total de 18,942,531 parámetros.

La arquitectura VGG19 fue la siguiente en ser reentrenada. Para ello, se emplearon 20 capas, de las cuales 16 son capas convolucionales para la normalización y el batch. Se incluyeron 4 capas dense y capas de aplanamiento, dando un total de 24,268,739 parámetros.

En un tercer momento, toca el turno al modelo MobileNet V1, utilizando 130 capas, de las cuales 142 son capas convolucionales para la normalización y el batch, se anexaron 4 capas dense y capas de aplanamiento, resultando en un total de 11,667,523 parámetros.

Concluyendo con el proceso de reentrenamientos, se aplicó la arquitectura de Resnet50 usando 141 capas, de las cuales 42 son capas de activación, 46 son capas convolucionales para la normalización y el batch. Se añadieron 4 capas dense y capas de aplanamiento, proporcionando un total de 40,414,979 parámetros.

3.1.2.1 Resultados de los reentrenamientos.

Al finalizar los reentrenamientos de los cuatro modelos, los resultados obtenidos fueron los siguientes:

- **VGG16** obtuvo una precisión del 99.35%, siendo este el modelo con mejor precisión de todos. Este modelo se entrenó con 100 épocas, un tamaño de batch size de 6 usando el optimizador Adamax.

- **VGG19** obtuvo una precisión del 84.42%. Sus hiperparámetros de entrenamiento utilizados fueron 100 épocas, un tamaño de batch size de 6 y el uso del optimizador Adam.
- **MobileNet V1** obtuvo una precisión del 92.86%. Este modelo se entrenó con los siguientes hiperparámetros: 110 épocas, un tamaño de batch size de 6 y se usó el optimizador RMSprop.
- **ResNet50** obtuvo una precisión del 70.13%, siendo este el valor más bajo de todos los modelos. Los hiperparámetros empleados fue de 110 épocas, un tamaño de batch size de 6 y se usó el optimizador Adadelta.

En la Tabla 3 se muestra la comparación de la exactitud obtenida por cada red neuronal convolucional y la precisión por clase.

Tabla 3. Tabla comparativa de exactitud y precisión de las CNN de la Fase 1.

Modelo	Exactitud	Clase	Precisión [%]
VGG-16	99.35%	BCELL	98.15
		NORMAL	100.00
		TCELL	100.00
VGG-19	84.42%	BCELL	62.96
		NORMAL	100.00
		TCELL	92.59
ResNet50	70.13%	BCELL	77.78
		NORMAL	84.78
		TCELL	50.00
MobileNetV1	92.86%	BCELL	85.19
		NORMAL	100.00
		TCELL	94.44

3.1.3 Generación de mapas de calor para la primera fase.

Al finalizar los entrenamientos de las CNN expuestos en la sección anterior, se aplican los algoritmos de los métodos explicativos con el toolbox iNNvestigate. Las pruebas se realizaron para las tres categorías de células que contiene el dataset, es decir, BCELL, TCELL y NORMAL. Asimismo, se

implementaron los métodos de Gradient, ABS Gradient, Deep Taylor y Gradient*Input.

En los resultados mostrados en las figuras Figura 13, Figura 14 y Figura 15 se pueden observar 19 mapas de relevancias generados, teniendo como entrada una imagen de célula de cada categoría. Estos mapas corresponden a la combinación de los cinco métodos explicativos implementados y los cuatro modelos de CNN reentrenados. Cabe destacar que para la combinación del modelo MobileNet V1 y el método LRPz no está definida la generación de ese mapa de relevancias.

Por otro lado, las arquitecturas VGG-16 y VGG19 y sus respectivos mapas de relevancias muestran resultados similares, a pesar de tener distintos valores de precisión. Sin embargo, los resultados de los pixeles más relevantes se visualizan de mejor manera en los mapas de VGG19. Es por esta razón que para la evaluación realizada por los hematólogos se emplearán estos mapas.

En lo que respecta a los mapas de relevancias obtenidos por el método Deep Taylor, se pudo observar de primera instancia que, para las arquitecturas VGG y ResNet50, los pixeles relevantes tienen a marcar principalmente el contorno de la célula. Por otro lado, el modelo MobileNet V1 tiende a marcar la relevancia en zonas determinadas que no siempre tienen relación con las características morfológicas. Finalmente, respecto a la combinación de la arquitectura ResNet50 y el método LRPz, los mapas de relevancias presentan escasos pixeles, lo que pudiera representar escasas características relacionadas con la célula clasificada.

En la siguiente sección se describen los resultados de la evaluación realizada por los hematólogos expertos. Una de las principales aportaciones de este trabajo se basa en el análisis de los resultados obtenidos de esta evaluación, con la finalidad identificar la mejor combinación entre arquitectura

de red neuronal y método generador de mapa de relevancias adecuado para la implementación como una herramienta auxiliar para la clasificación de ALL.

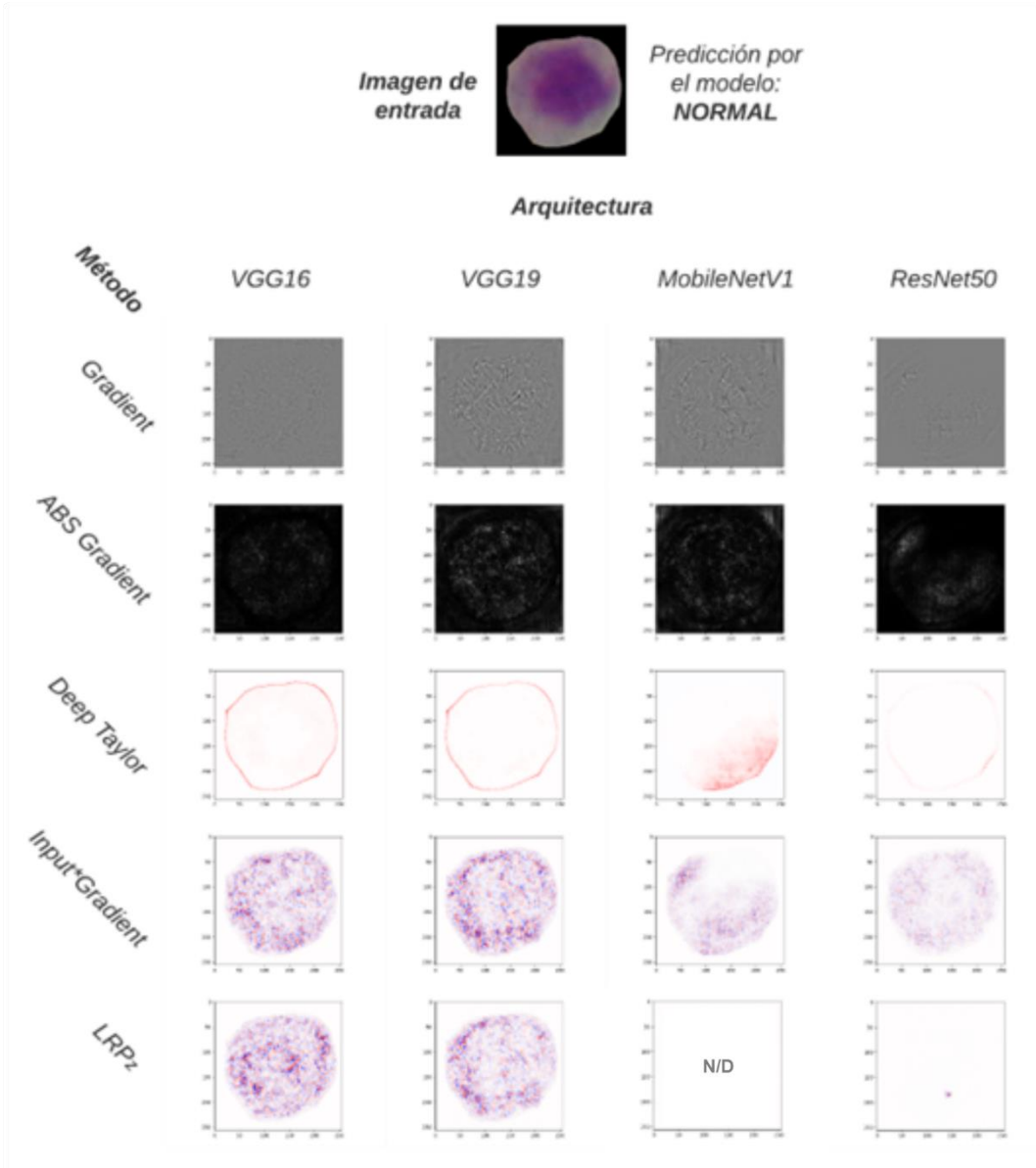


Figura 13. Mapas de calor generados para una imagen de entrada tipo NORMAL.

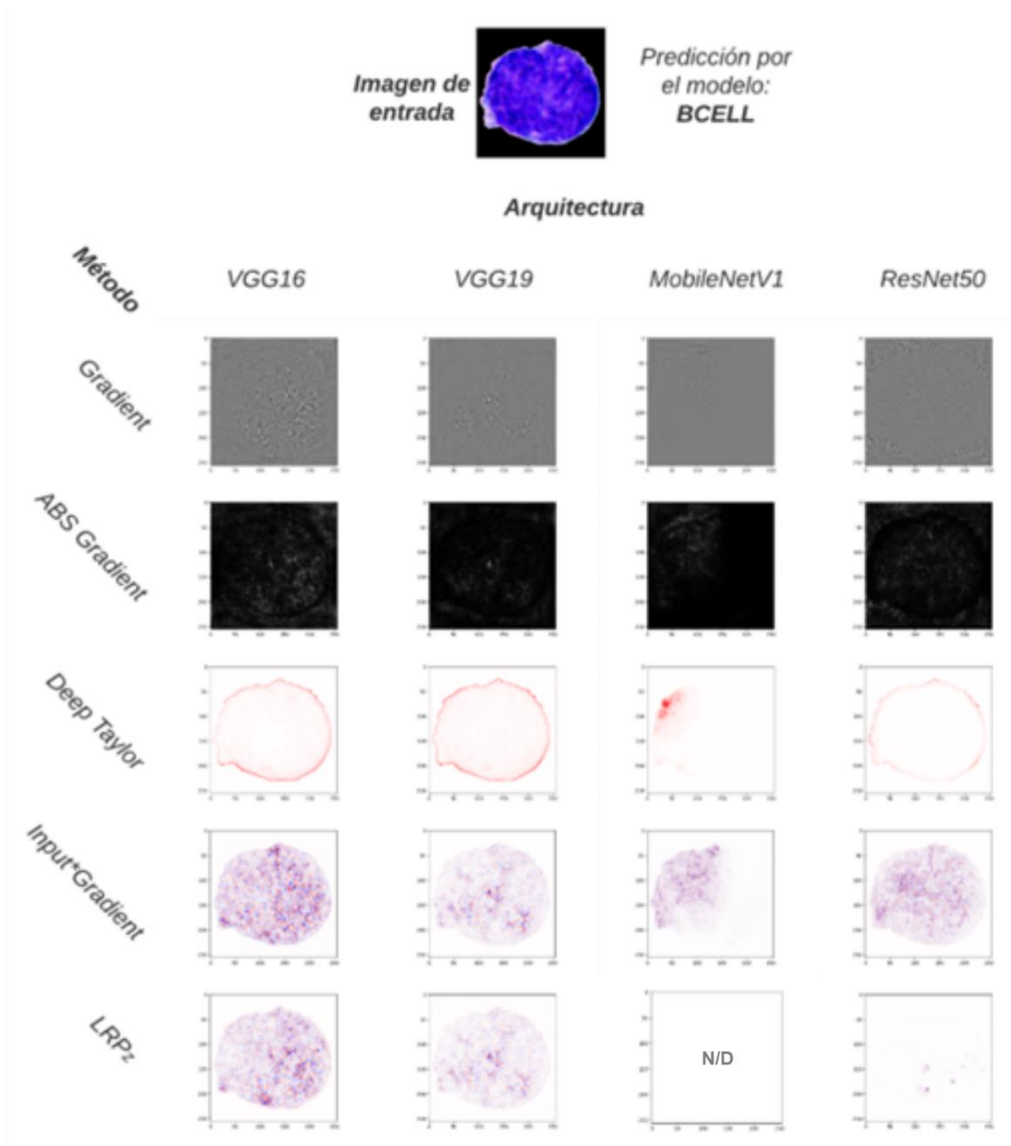


Figura 14. Mapas de calor generados para una imagen de entrada tipo BCELL.

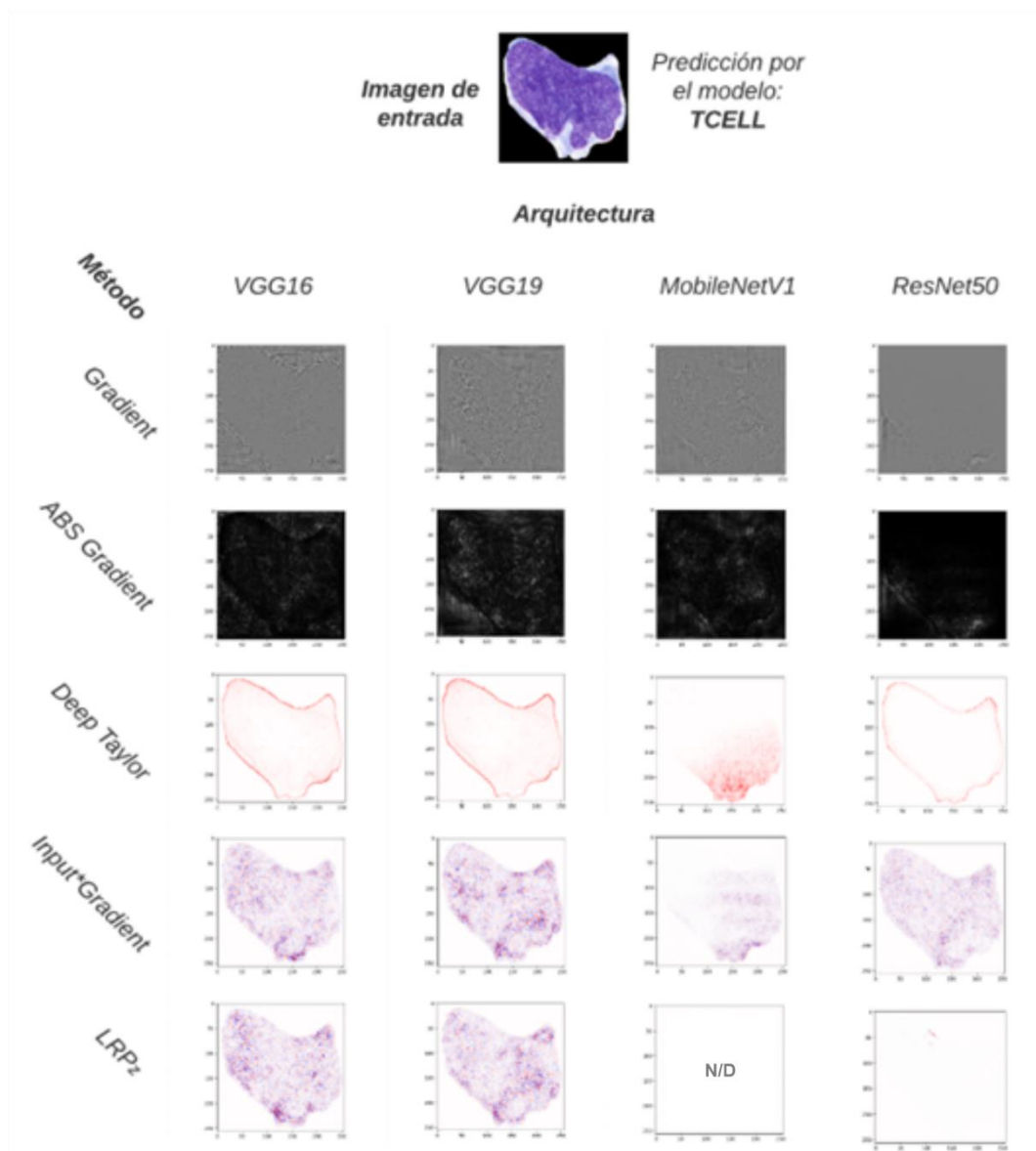


Figura 15. Mapas de calor generados para una imagen de entrada tipo TCELL.

3.1.4 Desarrollo de la evaluación realizada por hematólogos expertos.

Durante la etapa final de esta primera fase, se llevó a cabo una evaluación de los resultados obtenidos de los mapas de relevancias que fueron generados a partir de 15 imágenes, las cuales se muestran en la Figura 16. Los modelos de CNN evaluados fueron VGG19, MobileNet V1 y Resnet50,

descartando VGG16 por la similitud entre los mapas de relevancias con VGG19, ya antes mencionado.

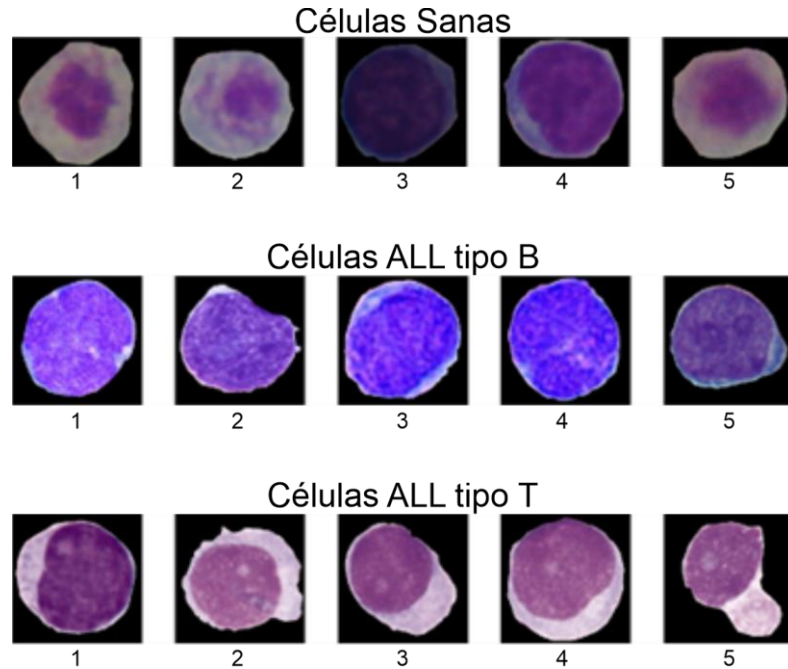


Figura 16. Células ALL tipo B, tipo T y Sanas empleadas para generar sus respectivos mapas de calor y ser evaluados por los hematólogos.

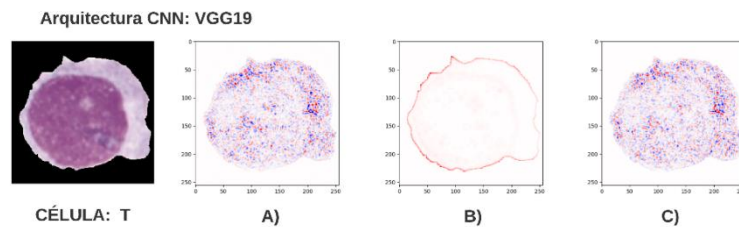
En lo que respecta a los métodos de mapas de relevancias implementados para la evaluación, se consideraron Gradient*Input, Deep Taylor y LRPz, descartando Gradient y ABS Gradient debido a la dificultada de visualizar la información de los pixeles en los mapas generados por estos métodos. El total de mapas de relevancias generados es de 120, cada uno de ellos serán calificados por los médicos expertos.

Para ello, se realizó un arreglo de las imágenes, es decir, cada imagen original de la célula y sus respectivos mapas de relevancias fueron colocados en disposiciones como se muestran en las figuras Figura 17, Figura 18 y Figura 19. La calificación asignada por los cinco expertos sería en una escala de 1 al 5, la cual representa una cantidad porcentual de acuerdo con la escala porcentual mostrada en la Tabla 4. El valor asignado correspondería a la cantidad de información que el mapa de relevancia les proporciona,

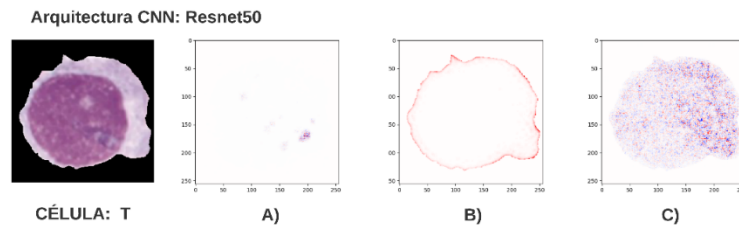
considerando que dicha información esté relacionada con las características morfológicas que ellos buscan cuando realizan una clasificación típica, y pensando si dichos mapas de relevancias pueden ser considerados como una herramienta auxiliar para facilitar y corroborar la clasificación obtenida por la CNN.

Tabla 4. Tabla escala porcentual para evaluación de mapas de calor.

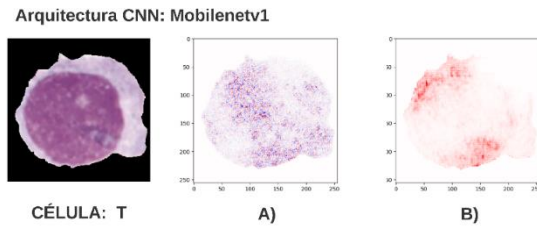
Calificación	Información proveída
1	0% - 20%
2	21% - 40%
3	41% - 60%
4	61% - 80%
5	81% - 100%



*Figura 17. Un ejemplo de la distribución para evaluación de una imagen con etiqueta TCELL y sus respectivos mapas de calor: A) LRP, B) Deep Taylor, C) Gradient*Input, obtenidos de la arquitectura VGG19.*



*Figura 18. Un ejemplo de la distribución para evaluación de una imagen con etiqueta TCELL y sus respectivos mapas de calor: A) LRP, B) Deep Taylor, C) Gradient*Input, obtenidos de la arquitectura ResNet50.*



*Figura 19. Un ejemplo de la distribución para evaluación de una imagen con etiqueta TCELL y sus respectivos mapas de calor: A) Gradient*Input, B) Deep Taylor, obtenidos de la arquitectura MobileNet V1.*

Al finalizar la evaluación, los cinco hematólogos evaluaron un total de 600 mapas de calor. Los resultados obtenidos se exponen en el capítulo 4.

3.2 Fase 2

En este apartado se describe el proceso realizado en esta segunda fase de la investigación. En primera instancia, se prepara la segunda base de datos implementada de imágenes de ALL sin segmentar los elementos ni el fondo (ALL-IDB2, descrita en la sección 2.2.2). Posteriormente, se genera la segmentación semántica de forma semimanual para crear el “ground truth”, la cual es validada por los hematólogos.

A continuación, se realiza el reentrenamiento de las redes CNN (VGG-19, ResNet18, ResNet50 y GoogleNet). En seguida, se implementan los métodos generadores de mapas de relevancias (Gradient*Input, Deep Taylor, LRPz y Grad-CAM). Finalmente, se realiza la evaluación de los mapas de relevancias empleando la segmentación semántica realizada previamente.

3.2.1 Preparación de la base datos ALL-IDB2

La base de datos de imágenes para el procesamiento de imágenes de Leucemia Linfoblástica Aguda (IDB-ALL2) es de acceso público y sus categorías están equilibradas. Esta base de datos está destinada a evaluar y comparar algoritmos de segmentación y clasificación de imágenes en ALL. Cada imagen en la base de datos fue identificada y clasificada por un grupo de oncólogos con experiencia en identificar células enfermas de ALL.

Las imágenes se dividen en dos categorías: enfermas y sanas, con 130 imágenes en cada categoría. Las imágenes tienen una resolución de 257 x 257 píxeles en RGB; para este trabajo, fueron redimensionadas a un tamaño de 254 x 254 píxeles en RGB. Además, se dividió las imágenes en 100 imágenes por cada categoría. Con las 30 imágenes restantes por categoría se creó una carpeta llamada “imágenes nunca antes vistas”. Estas imágenes se utilizarán exclusivamente para evaluar los modelos después de todo el proceso de reentrenamiento y para generar mapas de relevancias.

Utilizando las 200 imágenes para el reentrenamiento, aplicamos el aumento de datos mediante el método tradicional (rotación y reflexión) para tener al final 1000 imágenes para cada tipo de célula (Sana o ALL). Con estas imágenes, finalmente reentrenamos las redes neuronales GoogleNet, ResNet18, ResNet50 y VGG19

El conjunto de datos ALL-IDB2 es pequeño, pero es uno de los conjuntos de datos más utilizados. Por ejemplo, en [6], [7], [16], además de ser de acceso público, sus categorías están equilibradas. Por esta razón, se seleccionó un conjunto de datos ALL-IDB2 para esta segunda fase de este proyecto de investigación.

3.2.2 Segmentación semántica con morfología celular de la base de datos ALL-IDB2 para ground truth

Como se mencionó en la sección 2.1.2, la morfología celular se utiliza para identificar las células enfermas de ALL. Las características principales se centran en el núcleo y el citoplasma. Por esta razón, se realiza una segmentación semimanual, destacando cinco clases: núcleo, citoplasma, vacuolas, glóbulos rojos y fondo.

Los resultados de la segmentación serán considerados como la referencia base (ground truth), que luego se utilizará para evaluar los mapas

de relevancias. La evaluación compara los píxeles más relevantes en cada mapa de relevancias con el ground truth correspondiente a la imagen original de entrada.

La segmentación se realizó utilizando el etiquetador de imágenes de MATLAB [28]. Esta aplicación permite etiquetar imágenes de referencia a partir de una colección de imágenes, definiendo etiquetas como regiones de interés (ROI) rectangulares con ejes alineados o rotados, etiquetas de ROI de línea, etiquetas de ROI de píxel, etiquetas de ROI de polígono, etiquetas de ROI de punto, etiquetas de ROI de cuboide proyectado y etiquetas de escena.

La Figura 20 ilustra cómo se realizó la segmentación de imágenes de la base de datos IDB-ALL2 con la aplicación Image Labeler. La segmentación genera una matriz del mismo tamaño que la imagen segmentada. Estas matrices contienen el valor que corresponde a cada etiqueta de la segmentación, y serán la referencia de verdad absoluta utilizada para la evaluación. El proceso de segmentación fue llevado a cabo por dos estudiantes que trabajaron en esta investigación y supervisado por dos hematólogos con experiencia en morfología celular del sistema de salud pública mexicano. La imagen se segmentó en cinco categorías: núcleo, citoplasma, vacuolas, glóbulos rojos y fondo. Estas categorías fueron elegidas por sugerencia de los hematólogos.

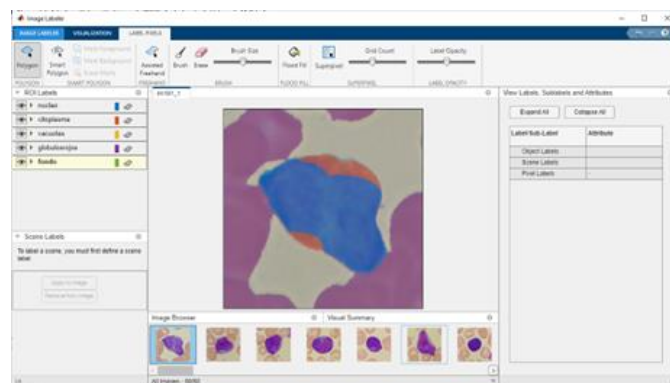


Figura 20. Segmentación de una imagen de ALL-IDB2 usando el software Image Labeler.

3.2.3 Reentrenamiento de modelos de CNN con la base de datos ALL-IDB2

Los resultados obtenidos de los reentrenamientos de los cuatro modelos de CNN se describen a continuación:

- **VGG19** obtuvo una precisión del 85%. Sus hiperparámetros de entrenamiento fueron 100 épocas con un tamaño de batch de 6 usando un optimizador Adam.
- **ResNet18** obtuvo una precisión del 93.33%. Los hiperparámetros empleados fue 110 épocas, un tamaño de batch de 6 y se usó el optimizador Adadelata.
- **ResNet50** obtuvo una precisión del 91.67%. Se emplearon los hiperparámetros de 110 épocas, un tamaño de batch de 6 y se usó el optimizador Adadelata.
- **GoogleNet** obtuvo una precisión del 83.33%, siendo este el valor más bajo de todos los modelos. Los hiperparámetros empleados fue de 110 épocas, un tamaño de batch de 6 y se usó el optimizador Adadelata.

En la Tabla 3 se muestra la comparación de la exactitud obtenida por cada CNN y la precisión por clase para cada una de las CNN reentrenadas.

Tabla 5. Tabla comparativa de exactitud y precisión de las CNN de la Fase 2

Modelo	Exactitud	Clase	Precisión [%]
VGG-19	85.00%	ALL	93.33%
		Sanas	76.67%
ResNet18	93.33%	ALL	93.33%
		Sanas	93.33%
ResNet50	91.67%	ALL	96.67%
		Sanas	86.67%
GoogleNet	83.33%	ALL	93.33%
		Sanas	73.33%

3.2.4 Generación de mapas de calor

En la sección 3.1.4 de este trabajo se explica las razones del por qué solo se utilizaron los métodos Gradient*Input, Deep Taylor y LRP. Los mismos procedimientos se realizan en esta segunda etapa, con la diferencia que ahora se emplean imágenes de entrada sin segmentar los componentes sanguíneos del fondo.

La Figura 21 muestra algunos mapas de calor obtenidos para una célula clasificada como sana y una célula clasificada como enferma. El mapa de calor generado con el método Deep Taylor muestra en toda la imagen diferentes grados de relevancia, utilizando un solo color para mostrar la importancia de los píxeles para la red neuronal. Sin embargo, la mayoría de las veces ofrece niveles más altos de relevancia en los bordes de las células de interés.

Los mapas de calor obtenidos por el método Gradient*Input generan píxeles de acuerdo con la escala de colores que va desde el azul al rojo, siendo el azul los píxeles con la menor relevancia y los rojos los píxeles con la mayor relevancia. Con este método, identificar las formas o regiones de mayor importancia para la red neuronal en la clasificación es un poco más complejo. La complejidad surge porque genera píxeles relevantes y no relevantes muy cerca uno del otro y con regiones mal definidas en comparación con otros métodos.

Finalmente, con el método LRPz, los mapas de calor obtenidos, al igual que el método anterior, manejan la escala de colores de azul a rojo. A diferencia del método Input*Gradient, estos mapas de calor muestran más claramente las regiones de mayor relevancia para la red neuronal que las que no son importantes.

La figura 22 muestra dos células donde el mapa de calor no está definido para el método de Deep Taylor. Estos casos, según [6], son inherentes a la técnica

ya que se calcula un valor que funciona como umbral. Si este umbral no se supera, no se genera el mapa de calor.

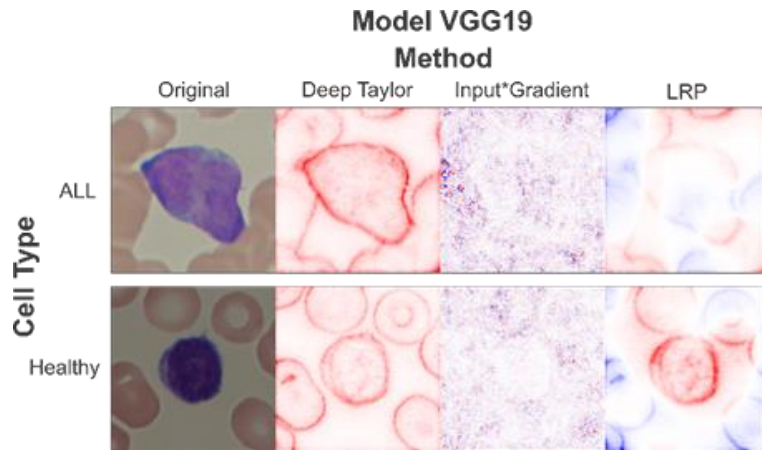


Figura 21. Ejemplos de mapas de calor de los métodos Deep Taylor, Gradient*Input y LRPz

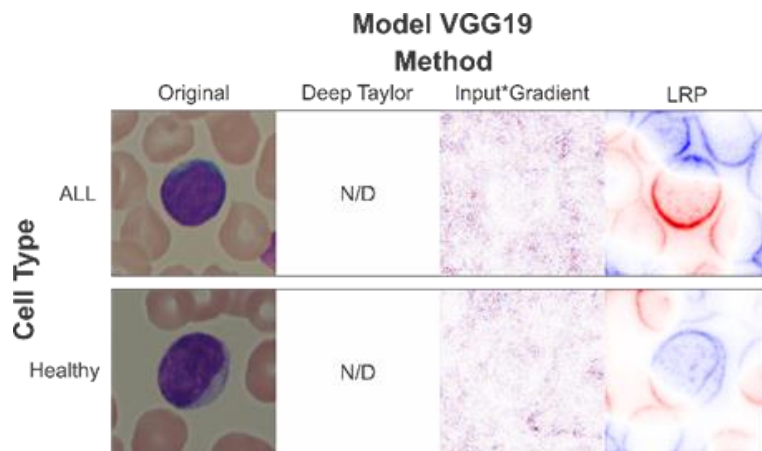


Figura 22. Ejemplo de mapas de calor donde para el método Deep Taylor no está definido.

En esta segunda fase, se anexó el método explicativo Grad-CAM, el cual no había sido contemplado en la primera fase. De acuerdo con Díaz R. en [8] obtiene buenos resultados con este método para células segmentadas.

La Figura 23 visualiza algunos mapas de calor obtenidos con el método Grad-Cam para los modelos de GoogleNet, ResNet18 y ResNet50, para una imagen de célula enferma y saludable. Esta figura muestra que los píxeles más

relevantes con el modelo GoogleNet están ubicados principalmente dentro de la misma ubicación que la célula de interés.

En los mapas de calor obtenidos con el modelo ResNet18, algunas regiones de interés están situadas dentro de la célula en cuestión. Sin embargo, la mayoría de estos píxeles se ubican fuera de la célula objetivo.

Finalmente, en los mapas de calor que corresponden al modelo ResNet50, las regiones de interés generalmente no están relacionadas con la célula representada en la imagen.

A primera vista, se podría decir que el método Grad-Cam, en combinación con el modelo GoogleNet, es la mejor combinación para generar mapas de calor que se relacionen los píxeles de relevancia con la morfología de la célula objetivo. Sin embargo, en la Figura 24 podemos ver los resultados obtenidos con células diferentes a las de la figura anterior. En esta imagen, se visualiza que el mapa de calor obtenido no siempre es el mejor.

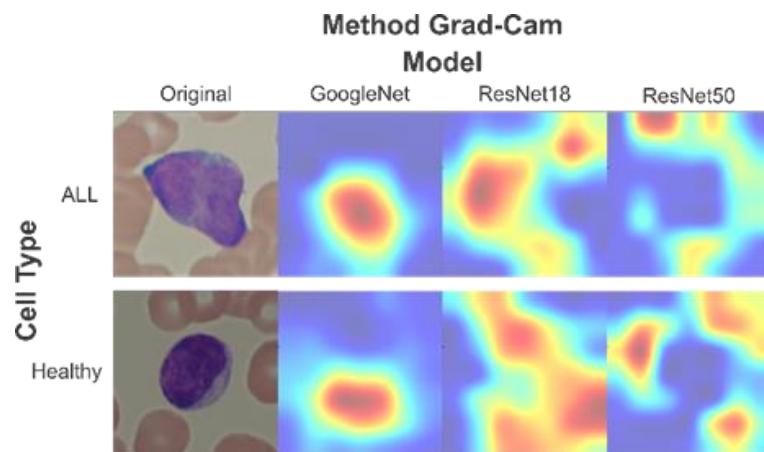


Figura 23. Mapas de calor generados con el método Grad-CAM

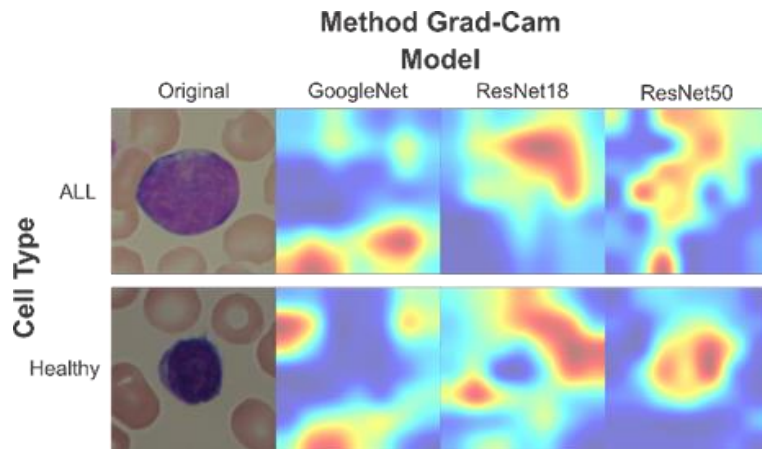


Figura 24. Mapas de calor obtenidos con el método Grad-CAM

En la siguiente sección se describe el método de evaluación realizado, con la finalidad de determinar cuál de los métodos XAI implementados, aporta mejores resultados que sea de utilidad para los expertos en la clasificación de células de ALL.

3.2.5 Método de evaluación de mapas de calor

Una vez obtenidos los mapas de calor con los modelos CNN reentrenados se procede a realizar la evaluación de los mapas de calor. Para este proceso, solo se consideran los píxeles más relevantes para la red neuronal durante la clasificación. Recordemos que algunos mapas de calor que emplean una escala de colores que van del azul al rojo.

Es evidente que los píxeles de mayor relevancia típicamente se encuentran en el canal rojo de las imágenes del mapa de calor. Por lo tanto, este canal se utiliza para evaluar los mapas de calor.

El proceso de evaluación se describe de la siguiente manera:

- Utilizamos los píxeles del canal rojo. Se calcula la profundidad de color promedio de todos los píxeles en este canal. El valor obtenido será el

valor de referencia o umbral para reducir los píxeles de mayor relevancia.

- Se identifica los píxeles por encima del valor de umbral en la imagen de mapas de calor, así mismo, el valor de la segmentación en la matriz de “ground truth”.
- Se realiza un conteo de todos los píxeles para cada clase según la segmentación.
- Guardamos el número de píxeles en cada categoría en una tabla para su posterior análisis y graficado de los resultados.

Los resultados y su análisis se describen en la siguiente sección.

4 Resultados y Discusión

En este capítulo se analizarán los resultados obtenidos en la fase 1, donde cinco hematólogos expertos evaluaron los resultados y de ellos obtuvimos retroalimentación de los resultados.

La segunda fase se empleó segmentación semántica, que fue validado por hematólogos. Dicha segmentación fue empleada como ground truth con la cual se realizar el análisis de los pixeles relevantes y dónde se ubican en regiones de interés para los expertos.

4.1 Resultados Fase 1: Evaluación de Hematólogos

El presente trabajo se enfoca en evaluar e identifica el mejor modelo de CNN y método de generación de mapas de calor, midiendo la información morfológica que proporcionan al experto médico.

En la Tabla 6 se muestra los resultados de los mapas de calor generados con los métodos LRP, Deep Taylor y Gradient*Input para el modelo VGG19, utilizando células de ALL tipo T como imagen de entrada. Según la evaluación, el proceso Gradient*Input obtuvo el mejor resultado con una puntuación promedio de 2.79 puntos, el 56% de la información de acuerdo con la morfología observada, obteniendo un total de 10.5 mapas con la puntuación más alta de un total de 15 imágenes evaluadas.

*Tabla 6. Resultados de la evaluación de los mapas de calor generados por el modelo VGG-19 y los métodos Gradient*Input, LRP y Deep Taylor para una célula de ALL tipo T.*

Evaluación Modelo VGG-19. Células tipo T			
Método	Promedio	% información morfológica	No. de mapas con calificación alta
LRP	2.56	51%	4
Deep Taylor	2.01	40%	0.5
Gradient*Input	2.79	56%	10.5

A continuación, se muestra en la Tabla 7, que el modelo Gradient*Input tiene la mejor puntuación. En esta ocasión, se evaluaron los mapas de calor generados con el modelo ResNet50 y los métodos LRP, Deep Taylor y Gradient*Input para clasificar imágenes de células B. Se obtuvo una puntuación de 2.45, equivalente al 49% de la información correspondiente a la morfología celular. Además, este método recibió las puntuaciones más altas en todos los casos.

*Tabla 7. Resultados de la evaluación de los mapas de calor generados por el modelo ResNet50 y los métodos Gradient*Input, LRP y Deep Taylor para una célula de ALL tipo B.*

Evaluación Modelo ResNet-50. Células tipo B			
Método	Promedio	% información morfológica	No. de mapas con calificación alta
LRP	1.00	20%	0
Deep Taylor	1.31	26%	0
Input*Gradient	2.45	49%	14.0

Por último, la Tabla 8 muestra los resultados de la evaluación del modelo MobileNet V1 y los métodos Deep Taylor y Gradient*Input, recordando que no está determinada la generación de los mapas de calor LRP para este modelo. En esta ocasión, el algoritmo mejor evaluado fue Deep Taylor, con una puntuación promedio de 1.76, equivalente al 35% de la información correspondiente a la morfología celular, solo un 2% por encima del modelo Gradient*Input. Sin embargo, el número de puntuaciones más altas fue el mismo para ambos métodos, es decir. De los 15 mapas evaluados ambos modelos obtuvieron la misma cantidad de mapas mejor evaluados. En uno de los mapas obtuvieron la misma calificación.

Tabla 8. Resultados de la evaluación de los mapas de calor generados por el modelo MobileNet V1 y los métodos Gradient*Input, LRP y Deep Taylor para una célula Normales.

Evaluación Modelo MobileNet-V1. Células tipo Normal			
Método	Promedio	% información morfológica	No. de mapas con calificación alta
LRP	0	0%	0
Deep Taylor	1.76	35%	7.5
Input*Gradient	1.67	33%	7.5

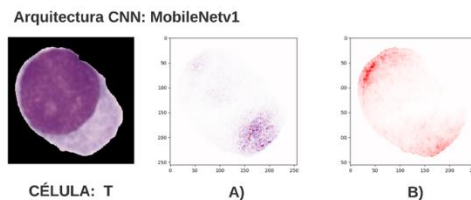
Según la Tabla 9, el método Gradient* Input es el mejor calificado considerando los tres modelos de redes neuronales, obteniendo una puntuación promedio de 2.30 puntos, equivalente al 46% de la información correspondiente a la morfología de las células. Del mismo modo, el método recibió el mayor número de mapas mejor calificados, sumando un total de 32 mapas generados con este método con la mayor calificación.

Tabla 9. Resultados de la evaluación general para los mapas de calor generados con los modelos VGG19, ResNet50 y MobileNet V1 y los métodos LRP, Deep Taylor e Input*Gradient correspondientes a las tres categorías de células.

Evaluación General de los tres modelos y tipos de células			
Método	Promedio	% información morfológica	No. de mapas con calificación alta
LRP	1.19	24%	4
Deep Taylor	1.70	34%	8
Input*Gradient	2.30	46%	32

Finalmente, se dio un paso más: replicar una de las evaluaciones, correspondiente al modelo MobileNet V1 con la imagen de entrada de la célula 3. Esta acción tuvo como objetivo medir la consistencia de la evaluación realizada por hematólogos expertos. A partir de los resultados obtenidos que se muestran en la Tabla 10, se corrobora que las puntuaciones obtenidas en ambas evaluaciones son consistentes para los cinco expertos que realizaron la actividad.

Tabla 10. Resultados de la medición de la consistencia de la evaluación realizada por hematólogos expertos.



	1ra Evaluación		2da Evaluación	
	Cell 3 tipo T Gradient*Input (A)	Cell 3tipo T DeepTaylor (B)	Cell 3 tipo T Gradient*Input (A)	Cell 3tipo T DeepTaylor (B)
Experto 1	1	3	2	3
Experto 2	1	1	1	1
Experto 3	2	4	2	3
Experto 4	1	1	1	2
Experto 5	1	1	1	1
Promedio	1.2	2	1.4	2
% información morfológica	24%	40%	28%	40%

4.2 Discusión de los resultados de la fase 1

Según los resultados, los métodos Deep Taylor y Gradient* Input produjeron los mejores mapas de calor que nos permitieron resaltar e identificar los píxeles de interés que tenían mayor importancia para el proceso de clasificación. Además, proporcionan más información que podría estar relacionada con características morfológicas presentes en las células en comparación con los otros métodos de generación de mapas de calor.

Según la evaluación realizada por los expertos, los mapas de calor con los métodos y modelos seleccionados no proporcionan la información ideal correspondiente a las características morfológicas de las células clasificadas. Algunos de los comentarios hechos por los médicos enfatizan que, para realizar una clasificación morfológica de las células de leucemia, a menudo es

necesario identificar más de una característica en la célula. Esta necesidad de más características morfológicas debería conducir a mejorar los métodos de generación de mapas de calor para aplicaciones como la clasificación de enfermedades celulares.

4.3 Resultados Fases 2: Evaluación morfológica

Derivado la retroalimentación realizada por los expertos, donde mencionan que los mapas de calor, evaluados en la primera fase, tienen poca información correlacionada con la morfología que es de su interés. Se determina realizar una evaluación computacional.

Esta evaluación se realiza empleando una segmentación semántica semimanual. Los expertos recomiendan segmentar cinco características morfológicas más significativas que buscan al clasificar células de ALL.

Además, se añade el método Grad-Cam para la generación del mapa de calor y modelo GoogleNet, que no habían sido considerados en la fase 1, como se menciona en la sección 3.2.

Los resultados obtenidos se muestran en la Figura 25, en ella se observan algunos de los mapas de calor generados. Por otro lado, en la Figura 26 visualiza una superposición de los mapas de calor y la célula de ALL con la que se generaron, esto con la finalidad de tener las imágenes para realizar una comparativa. Si conocemos el número de píxeles relevantes ubicados dentro de cada región segmentada, se puede calcular el porcentaje de píxeles relevantes para cada característica morfológica segmentada.

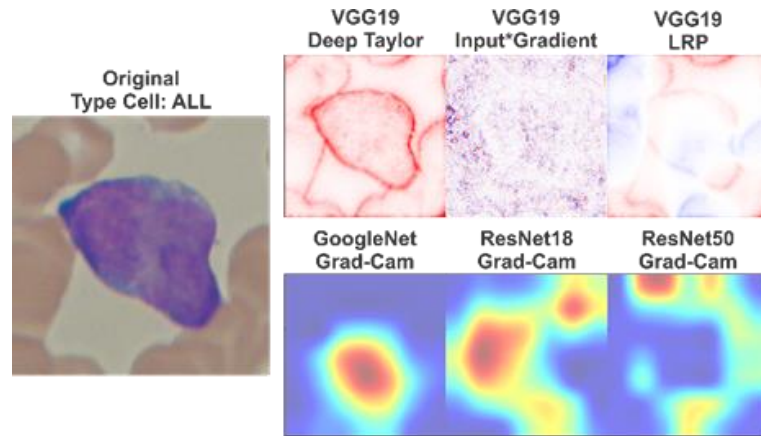


Figura 25. Imagen comparativa de los resultados obtenidos con los modelos y métodos de generación de mapas de calor evaluados.

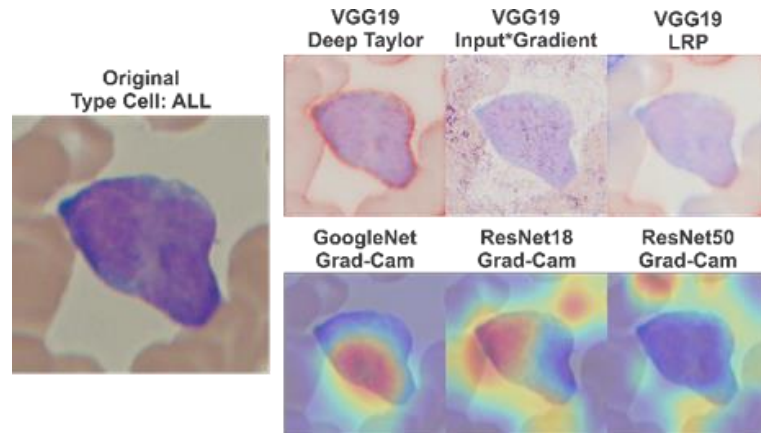


Figura 26. Imagen comparativa de los mapas de calor obtenidos superpuestos en la célula ALL con la que se generaron.

Los resultados de la evaluación realizada se muestran en el gráfico de la Figura 27. A partir de estos resultados, se observa que la combinación del modelo GoogleNet y el método Grad-Cam es la que mejor se relaciona con las características morfológicas de la célula de interés, con el 43.61% de los píxeles marcados como significativos ubicados en la célula, el 26.75% de los píxeles están posicionados en los glóbulos rojos y el restante 29.63% en el fondo.

En segundo lugar, se estableció el modelo ResNet18 y el método Grad-Cam con el 31.78% de los píxeles relevantes dentro de la célula de interés, 29.66% en los glóbulos rojos, y 38.56% en el fondo. La combinación que obtuvo el peor resultado fue el modelo VGG19 con el método Deep Taylor, con porcentajes de píxeles relevantes dentro de la célula de 20.81%, el 35.92% en los glóbulos rojos, y la tasa más alta en el fondo con el 43.28%.

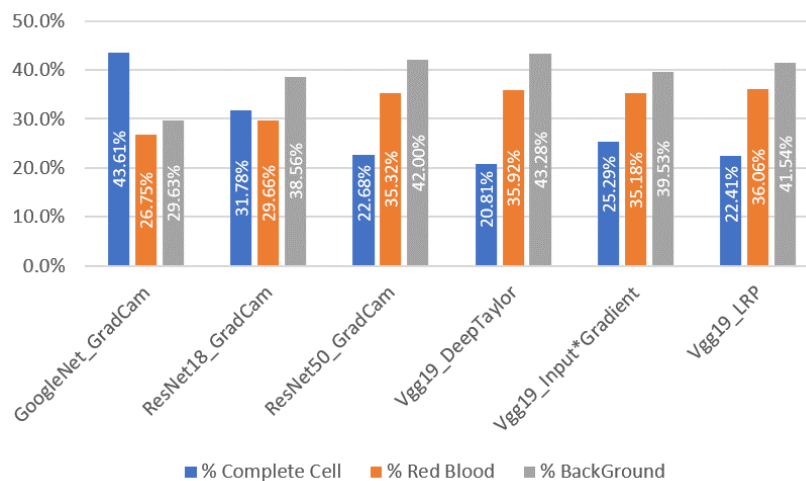


Figura 27. Comparación entre el modelo y el método de generación de mapa de calor, y su relación con la célula central, los glóbulos rojos y el fondo, en términos de porcentajes de píxeles relevantes.

Calculando la media de píxeles relevantes para cada categoría se obtiene los resultados de la gráfica ilustrada en la Figura 28. Se observa que la combinación del modelo GoogleNet y la arquitectura GradCam tiene valor medio de 8071 píxeles dentro de la célula. En contraste, las categorías de glóbulos rojos y fondo tienen los valores de 4951 y 5484, respectivamente. Siendo esta la única combinación donde la diferencia es considerablemente significativa entre la célula completa, y las categorías de glóbulos rojos y fondo.

Por el contrario, para el resto de las combinaciones, la mayor cantidad de píxeles relevantes se posicionan en el fondo o los glóbulos rojos, o bien, la diferencia del número de píxeles relevantes entre la célula principal, glóbulos

rojos y fondo son mínimos. Cuando lo deseable es que la mayoría de la relevancia esté en la célula central.

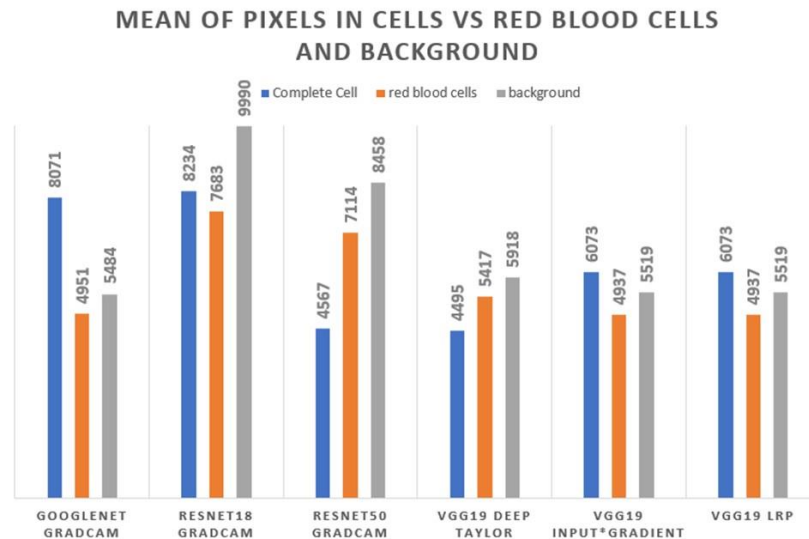


Figura 28. Media de píxeles relevantes para cada categoría utilizando una combinación de generación de mapas de calor y los distintos modelos reentrenados.

Posteriormente, se realiza el análisis en términos porcentuales la cantidad de relevancia contiene cada categoría morfológica de la célula de interés, es decir, la relevancia presente en el núcleo, citoplasma y vacuolas. Los resultados de la evaluación se muestran en la Figura 29.

Los resultados muestran que los píxeles más relevantes para las células saludables son 54.26% dentro del núcleo y 45.74% en el citoplasma. En el caso de las células ALL, el 91.90% están ubicadas dentro del núcleo, y el 8.10% se encuentran en el citoplasma. Este último está relacionado con las características morfológicas que describen la enfermedad porque, en las células ALL, el núcleo de la célula tiende a tener un área mayor, en comparación al núcleo de una célula saludable.

En la base de datos con la que se evaluaron los mapas de calor, hay tres imágenes donde se visualizaron vacuolas. Para esta combinación del modelo

CNN y el método de generación de mapas de calor, no hay píxeles de gran relevancia en sus ubicaciones. Por lo tanto, la categoría de vacuola tiene un 0% de relevancia.

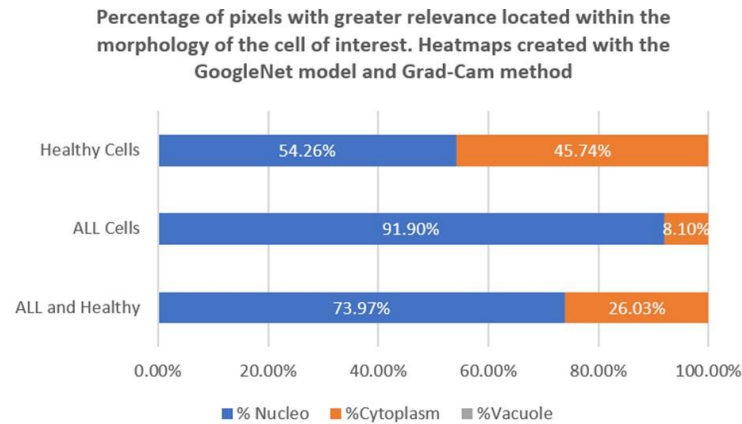


Figura 29. Porcentaje de píxeles con mayor relevancia ubicados dentro de la morfología celular de interés. Los mapas de calor se crearon con el modelo Google Net y el método Grad-Cam.

4.4 Discusión de los resultados de la fase 2

Realizar una evaluación computacional de los mapas de calor generados basándose en un mapa de características morfológicas (núcleo, citoplasma, vacuolas) para la clasificación de células ALL, así como de glóbulos rojos y fondo, permite evaluar si la CNN se enfoca en las características celulares de interés u otros elementos presentes en la imagen. Además, permite la comparación de métodos de generación de mapas de calor para definir cuál se correlaciona con dichas características morfológicas.

Según los resultados obtenidos en esta investigación, el modelo GoogleNet y el método Grad-Cam son los que mejor relacionan las características morfológicas naturales de la célula con los mapas de calor. De acuerdo con los resultados obtenidos en el presente trabajo, se puede corroborar la evaluación realizada por los hematólogos en la fase 1.

5 Conclusiones

La implementación de un algoritmo generador de mapas de relevancias en una red neuronal tiene como objetivo identificar las regiones o píxeles más críticos para el proceso de clasificación de la red neuronal.

En este trabajo, generamos mapas de relevancias con seis métodos diferentes Gradient, ABS Gradient, Gradient*Input, LRPz, Deep Taylor y Grad-CAM; aplicados a seis arquitecturas redes neuronales convolucionales distintas VGG-16, VGG-19, MobileNet V1, ResNet18, ResNet50 y GoogleNet. Se emplea la base de datos "The Cancer Imaging Archive", la que contiene imágenes de células ALL segmentadas del fondo y otros elementos sanguíneo, y la base de datos "ALL-IDB2", la que contiene imágenes de células son sementar los elementos alrededor.

En una primera fase, se evaluaron los modelos VGG19, MobileNet V1 y ResNet50, y los tres métodos de generación de mapas de relevancias que permitía visualizar la relevancia claramente, las cuales fueron LRPz, Gradient*Input y Deep Taylor. La evaluación fue realizada por especialistas en hematología quienes.

Se concluye de la evaluación que el método Input*Gradient puede estar relacionado con características morfológicas utilizadas para el diagnóstico celular. Del mismo modo, los comentarios de los expertos proporcionaron áreas de mejora en la generación de mapas de relevancias. Lo más importante es que la construcción de mapas de relevancias debería incluir características morfológicas para ser útiles para los especialistas médicos.

En la segunda fase, En este trabajo, generamos mapas de relevancias con cuatro métodos diferentes (Deep Taylor, Input*Gradient, LRP y Grad-Cam) implementados en cuatro arquitecturas diferentes (GoogleNet, ResNet18, ResNet50, VGG19). Se utilizó la base de datos ALL-IDB2 que contiene

imágenes no segmentadas de la célula de interés con fondo y otros elementos sanguíneos presentes.

Se generó una segmentación semántica, la cual fue validada por los hematólogos y se dividió en cinco clases: núcleo, citoplasma, vacuolas, glóbulos rojos y fondo. Utilizando esta segmentación semántica evaluamos los mapas de relevancias generados, contabilizando la cantidad de píxeles con mayor relevancia en las diferentes clases.

Esta evaluación concluye que el modelo GoogleNet se centra principalmente en las características presentes en la célula de interés. El método Grad-Cam es el método de generación de mapas de calor que mejor expresa la relevancia de las CNN. Combinado con el modelo GoogleNet, produce resultados que se centran exclusivamente en la célula objetivo. Por lo que este modelo de generación de mapas de relevancias es muy prometedor para aplicarse como un auxiliar para los médicos expertos.

6 Trabajo Futuro

La generación de mapas de calor como herramienta para explicar el resultado de una predicción en una imagen es prometedora. Sin embargo, aún se pueden mejorar los resultados implementando métodos explicativos textuales para complementar la información de las características presentes en la célula. Lo más importante es que la construcción de los mapas de relevancias que sean útiles para los especialistas médicos, por lo que continuaremos realizando trabajo colaborativo con los expertos para mejorar la implementación de estos métodos.

Por otro lado se explorará la implementación de modelos de procesamiento de lengua natural para la elaboración del reporte médico después de analizar la muestra para darle un uso a estas técnicas en el diagnóstico médico.

7 Aportaciones Clave

- Computacional:
Grad-CAM es el método en el que sus pixeles relevantes se relacionan de mejor forma con las características morfológicas de las células.

- Biomédico:
 - Evaluación y validación de resultados por hematólogos expertos en morfología celular.
 - Identificación de áreas de oportunidad para la implementación de métodos explicativos visuales.
 - Generar aceptación y seguridad del uso de este tipo de sistemas inteligentes en el área médica.

8 Referencias

- [1] INEGI, «Instituto nacional de estadística, geografía e informática.» INEGI, [En línea]. Available: <http://www.inegi.gob.mx>. [Último acceso: 27 04 2021].
- [2] N. Jiwani, K. Gupta, G. Pau and M. Alibakhshikenar, «"Pattern Recognition of Acute Lymphoblastic Leukemia (ALL) Using Computational Deep Learning",» *IEEE Access*, vol. 11, pp. 29541-29553, 2023.
- [3] Jiang, Z., Dong, Z., Wang, L., & Jiang, W., «Method for Diagnosis of Acute Lymphoblastic Leukemia Based on ViT-CNN Ensemble Model,» *Computational Intelligence and Neuroscience*, 2021.
- [4] Abir, W.H., Uddin, M.F., Khanam, F.R., Tazin, T., «Explainable AI in Diagnosing and Anticipating Leukemia Using Transfer Learning Method,» *Computational Intelligence and Neuroscience*, vol. 2022, p. 14, 2022.
- [5] Nayeon Kim, «Deep Learning Technology-based Model to Identify Benign and Pro-B Acute Lymphoblastic Leukemia (ALL): Xception + LIME,» *American Journal of Biomedical and Life Sciences*, vol. 9(5), pp. 279-285, 2021.
- [6] Rocio Ochoa-Montiel, Gustavo Olague, and Humberto , « "Expert knowledge for the recognition of leukemic cells," "Expert knowledge for the recognition of leukemic cells",» *Applied Optics*, vol. 59, nº 14, pp. 4448-4460, 2020.
- [7] R. R. Maaliw et al., «A Multistage Transfer Learning Approach for Acute Lymphoblastic Leukemia Classification,» *IEEE 13th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, pp. 0488-0495, 2022.
- [8] Diaz Resendiz, J.L., Ponomaryov, V., Reyes Reyes, R., & Sadovnychiy, S., «Explainable CAD System for Classification of Acute Lymphoblastic Leukemia Based on a Robust White Blood Cell Segmentation,» *Cancers*, vol. 15, nº 13, p. 3376, 2023.
- [9] Iván Palomo G., Jaime Pereira G., Julia Palma B., HEMATOLOGÍA Fisiopatología y Diagnóstico, Talca- Chile: UNIVERSIDAD DE TALCA, 2009.
- [10] Ashkan Emadi, Jennie York Law, «Manual MSD, Leucemia linfoblástica aguda (LLA),» octubre 2023. [En línea]. Available: <https://www.msmanuals.com/es-mx/professional/hematolog%C3%ADa-y-oncolog%C3%ADa/leucemias/leucemia-linfobl%C3%A1stica-aguda-lla>. [Último acceso: 2023].

- [11] Martin R Howard, Peter J Hamilton, Haematology an illustrated color text, 4th Edition, CHURCHILL LIVINGSTONE EL SERVIER, 2013.
- [12] Guido Arroyo, German F. Saenz, Eliecer Valenciano, «Interpretación morfológica y citológica de la leucemia aguda.» *Revista costarricense de ciencias médicas*, vol. 4 (supl 1), pp. 31-46, 1983.
- [13] R. Donida Labati, V. Piuri, F. Scotti, «ALL-IDB: THE ACUTE LYMPHOBLASTIC LEUKEMIA IMAGE DATABASE,» *2011 IEEE Int. Conf. on Image Processing (ICIP 2011)*, pp. 11-14, 2011.
- [14] Hariprasath, S. and Dharani, T. and Shaikh Mohammad, Bilal N., «Automated Detection of Acute Lymphocytic Leukemia Using Blast Cell,» *2nd International Conference on Advances in Science & Technology (ICAST) 2019*, 2019.
- [15] K. C. Jakkrich Laosai., «Classification of acute leukemia using medical-knowledge-based morphology and CD marker,» *Biomedical Signal Processing and Control*, vol. 44, pp. 127-137, 2018.
- [16] W. Lamberti., «Classification of White Blood Cell Leukemia with Low Number of Interpretable and Explainable Features,» *ArXiv*, 2022.
- [17] Harald Thiel, M.D., Heinz Diem, M.D., Torsten Haferlach, M.D., Color Atlas of Hematology, Practical Microscopic and Clinical Diagnosis, 2nd revised edition, Stuttgart, New York: Thieme, 2003.
- [18] Mourya, S., Kant, S., Kumar, P., Gupta, A., & Gupta, R., «ALL Challenge dataset of ISBI 2019 (C-NMC 2019),» *The Cancer Imaging Archive*, vol. 1, 2019.
- [19] M. K. Tusar, «Automated Detection of Acute Lymphoblastic Leukemia Subtypes,» *arXiv preprint arXiv:2208*, 2022.
- [20] Karen Simonyan, A. Z., «Very Deep Convolutional Networks for Large-Scale Image,» *arXiv:1409.1556,14*, 2014.
- [21] Olga Russakovsky, J. D.F., «ImageNet Large Scale Visual Recognition Challenge,» *arXiv*, 2015.
- [22] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador Garcia, Sergio Gil-Lopez, Daniel Molina, Richard Benjamins, Raja Chatila, Francisco Herrera, «Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI,» *Information Fusion*, vol. 58, pp. 82-115, 2020.

- [23] Leander Weber, Sebastian Lapuschkin, Alexander Binder, Wojciech Samek, «Beyond explaining: Opportunities and challenges of XAI-based model improvement,» *Information Fusion*, vol. 92, pp. 1566-2535, 2023.
- [24] Alber, M., Lapuschkin, S., Seegerer, P., Hägele, M., Schütt, K. T., Montavon, G., Samek, W., Müller, K. R., Dähne, S., & Kindermans, P. J., «iNNvestigate Neural Networks!,» *Journal of Machine Learning Research*, vol. 20, nº 93, pp. 1-8, 2019.
- [25] Bach S, Binder A, Montavon G, Klauschen F, Müller K-R, Samek W, «n Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation,» *PLoS ONE*, vol. 10, nº 7, 2015.
- [26] Grégoire Montavon, Sebastian Lapuschkin, Alexander Binder, Wojciech Samek, Klaus-Robert Müller, «Explaining nonlinear classification decisions with deep Taylor decomposition,» *Pattern Recognition, ELSEVIER*, vol. 65, pp. 211-222, 2017.
- [27] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh and D. Batra, «Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization,» *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 618-626, 2017.
- [28] Mathworks, «Mathworks. "Matlab Images Labeler application,» [En línea]. Available: <https://mathworks.com/help/vision/ref/imagelabeler-app.html..> [Último acceso: 2023].
- [29] Avanti Shrikumar, Peyton Greenside, Anna Y. Shcherbina, Anshul Kundaje, «ot Just A Black Box: Learning Important Features Through Propagating Activation Differences,» *ArXiv*, 2017.