

Red convolucional para la detección de diferentes etapas de leucemia

1st Martin Valencia Vallejo
Escuela de Ciencias e Ingenierías
Universidad EAFIT
Medellín, Colombia
mvalenciv1@eafit.edu.co

2nd Emanuel Gonzalez Quintero
Escuela de Ciencias e Ingenierías
Universidad EAFIT
Medellín, Colombia
egonzalezq@eafit.edu.co

3rd Nicolás Ospina Torres
Escuela de Ciencias e Ingenierías
Universidad EAFIT
Medellín, Colombia
nospinat@eafit.edu.co

Abstract—El diagnóstico temprano y preciso de la leucemia es fundamental para mejorar la supervivencia y el tratamiento de los pacientes. En los últimos años, las redes neuronales convolucionales (CNN) han demostrado un alto potencial en el análisis automatizado de imágenes hematológicas; sin embargo, la mayoría de los estudios se limitan a aplicar modelos preentrenados sin explorar a fondo cómo la configuración de los hiperparámetros afecta su desempeño. En este contexto, resulta necesario desarrollar y evaluar un modelo propio de CNN que permita optimizar estos parámetros para mejorar la precisión, la estabilidad y la interpretabilidad del proceso de detección de leucemia.

Index Terms—Leukemia, Neural Networks, Computer Vision, Medicine, Machine Learning, CNN.

I. ESTADO DEL ARTE

En los últimos años, el diagnóstico automatizado de la Leucemia Linfoblástica Aguda (ALL, por sus siglas en inglés) ha recibido una atención creciente dentro de la comunidad científica, particularmente con la adopción de modelos de aprendizaje profundo basados en redes neuronales convolucionales (CNN). Diversos estudios han demostrado que la aplicación de arquitecturas especializadas permite identificar características morfológicas complejas en imágenes de frotis sanguíneo con un nivel de precisión comparable, e incluso superior, al de los métodos tradicionales basados en características manuales.

En Abbas et al. [1] proponen un modelo *lightweight* de red neuronal convolucional diseñado para la clasificación automatizada de células B afectadas por leucemia linfoblástica aguda. El enfoque se centra en lograr un equilibrio entre precisión y eficiencia computacional, con una arquitectura optimizada que reduce significativamente el número de parámetros sin comprometer el rendimiento. Los autores enfatizan que el modelo no requiere segmentación previa de las células, lo que simplifica el flujo de procesamiento y lo hace más adecuado para implementaciones en entornos clínicos de recursos limitados. Además, incorporan técnicas de interpretación de modelos como *Grad-CAM* para resaltar las regiones relevantes en la toma de decisiones, contribuyendo así a una mayor transparencia e interpretabilidad del sistema.

De manera complementaria, Curtidor et al. [2] presentan un sistema de clasificación basado en redes neuronales para la detección de ALL a partir de imágenes microscópicas de sangre

periférica. Su trabajo evalúa distintas configuraciones de redes neuronales artificiales y convolucionales, mostrando que las CNN son particularmente efectivas en la extracción automática de características discriminantes. El estudio demuestra que el uso de arquitecturas profundas puede mejorar sustancialmente la precisión de diagnóstico frente a enfoques tradicionales, reforzando el potencial de las CNN como herramientas de apoyo al diagnóstico médico en el ámbito hematológico.

Por otra parte, Zare et al. [3] introducen un enfoque innovador que combina el aprendizaje profundo con grafos mediante el uso de redes neuronales convolucionales sobre grafos (GCNN) personalizadas para la detección automática tanto de leucemia linfoblástica aguda (ALL) como mieloide aguda (AML). En su propuesta, las imágenes de células sanguíneas se representan como grafos que capturan las relaciones espaciales entre regiones celulares, lo que permite al modelo aprender estructuras más ricas que las capturadas por CNN tradicionales. Los resultados experimentales evidencian una mejora significativa en la capacidad de generalización del modelo y una mayor robustez ante variaciones en la morfología celular. Este enfoque resalta la tendencia reciente de explorar arquitecturas híbridas que integren información estructural y espacial para mejorar la clasificación de patologías hematológicas.

En conjunto, estos trabajos reflejan la evolución del campo hacia modelos más eficientes, explicables y estructuralmente ricos. Mientras que las CNN ligeras ofrecen soluciones prácticas para entornos clínicos reales, las redes basadas en grafos abren nuevas oportunidades para capturar relaciones complejas dentro de las imágenes biomédicas. La literatura reciente sugiere que la combinación de eficiencia computacional, interpretabilidad y capacidad de generalización constituye la dirección predominante en el desarrollo de sistemas de diagnóstico automatizado para la leucemia linfoblástica aguda.

II. PREGUNTA INVESTIGATIVA Y OBJETIVOS

El diagnóstico temprano y preciso de la leucemia es un factor determinante para la supervivencia y el tratamiento oportuno de los pacientes. En los últimos años, las redes neuronales convolucionales (CNN) han mostrado avances significativos en la detección de afecciones hematológicas, incluida la leucemia. Sin embargo, gran parte de la literatura

existente se enfoca en la aplicación de arquitecturas estándar o preentrenadas, sin profundizar en el impacto que tienen las variaciones de los hiperparámetros sobre el desempeño, la estabilidad del entrenamiento y la interpretabilidad de los resultados. Esto limita la posibilidad de diseñar modelos optimizados que se ajusten a las características específicas de los conjuntos de datos hematológicos y que mejoren su capacidad de generalización.

Ante esta situación, se hace necesario explorar el desarrollo de un modelo propio de red neuronal convolucional multietiquetada orientado a la detección automática de las diferentes etapas de la leucemia leucemia en imágenes de muestras de sangre, evaluando de manera sistemática el efecto de diferentes configuraciones de hiperparámetros como la tasa de aprendizaje, el número de capas, los filtros convolucionales, la regularización y las funciones de activación con el fin de identificar aquellas combinaciones que maximicen la precisión, la sensibilidad y la interpretabilidad del modelo. Este enfoque busca contribuir al avance de modelos personalizados y eficientes que puedan ser integrados en herramientas de diagnóstico asistido.

El desarrollo de un modelo propio multietiqueta permite además abordar limitaciones comunes en los estudios existentes, como la dependencia de arquitecturas genéricas entrenadas sobre dominios distintos al biomédico. Al ajustar y comparar múltiples configuraciones de hiperparámetros, es posible obtener un entendimiento más profundo sobre cómo cada componente del modelo afecta su rendimiento, estabilidad y capacidad para generalizar frente a nuevas muestras. Este proceso de experimentación controlada constituye un paso esencial hacia la construcción de modelos más robustos, aplicables y adaptables al entorno clínico real.

Asimismo, el proyecto busca aportar evidencia práctica sobre la aplicabilidad de técnicas modernas de ajuste y optimización en CNN, tales como el uso de tasas de aprendizaje variables, regularización mediante dropout y normalización por lotes (batch normalization). Evaluar estas estrategias en un contexto médico específico permitirá no solo mejorar la precisión diagnóstica, sino también promover el uso de metodologías reproducibles que puedan servir como referencia para futuras investigaciones en el campo de la hematología computacional. De esta manera, el trabajo pretende ofrecer una contribución tanto técnica como científica al desarrollo de sistemas de apoyo al diagnóstico basados en inteligencia artificial.

A. Pregunta Investigativa

¿Cómo influye la variación de los hiperparámetros en el rendimiento y la interpretabilidad de un modelo propio basado en redes neuronales convolucionales multietiquetada para la detección automática de leucemia a partir de imágenes de muestras de sangre?

B. Objetivo General

Diseñar, implementar y evaluar un modelo propio de red neuronal convolucional para la detección de leucemia en

imágenes de muestras de sangre, analizando el impacto de diferentes configuraciones de hiperparámetros sobre su precisión, sensibilidad e interpretabilidad.

C. Objetivos Específicos

- 1) Revisar la literatura reciente sobre el uso de redes neuronales convolucionales en el diagnóstico hematológico, identificando los principales hiperparámetros, arquitecturas y estrategias de optimización empleadas, con el fin de establecer una base comparativa para el desarrollo del modelo propuesto.
- 2) Diseñar e implementar una arquitectura propia de red neuronal convolucional, ajustando de manera controlada los principales hiperparámetros (tasa de aprendizaje, número de filtros, tamaño de kernel, número de capas y tipo de activación) para analizar su efecto sobre el desempeño del modelo.
- 3) Evaluar el rendimiento del modelo mediante métricas cuantitativas (precisión, F1-score) y comparar los resultados obtenidos con diferentes configuraciones de entrenamiento.

III. DESCRIPCION DEL CONJUNTO DE DATOS

A. Origen y contexto

El dataset proviene de la plataforma Kaggle bajo el nombre *Acute Lymphoblastic Leukemia (ALL) image dataset* aportado por Mehrad Aria [4].

Este conjunto contiene 3 256 imágenes de frotis de sangre periférica (peripheral blood smear, PBS) provenientes de 89 pacientes.

Las imágenes fueron obtenidas con microscopio (objetivo 100x) y cámara Zeiss, y se han clasificado por un especialista en hematología mediante citometría de flujo, para determinar si la célula es benigna (“hematogone”) o blast (maligna: subtipos Early Pre-B, Pre-B, Pro-B ALL).

Dentro del dataset también se ofrecen versiones segmentadas de las imágenes (en HSV thresholding) como máscaras binarias para las regiones celulares.

B. Estructura de los datos: variables (X) y variable objetivo (y)

- Variables de entrada (X): cada instancia es una imagen (o par imagen + máscara). En la práctica, esas imágenes pueden representarse como tensores de pixeles (por ejemplo, dimensión H × W × C, donde C = 3 canales RGB y H = 224 y W = 224).
- Variable objetivo (y): Clasificación multicategoría: benigno, Early Pre-B ALL, Pre-B ALL, Pro-B ALL (cuatro clases, “0”, “1”, “2”, “3”).

C. Tamaño y partición del dataset

El dataset original aporta 3 256 imágenes 224x224 que se dividirán en:

- Conjunto de entrenamiento (por ejemplo, 70% → 2 279 imágenes)
- Conjunto de validación (15% → 489 imágenes)

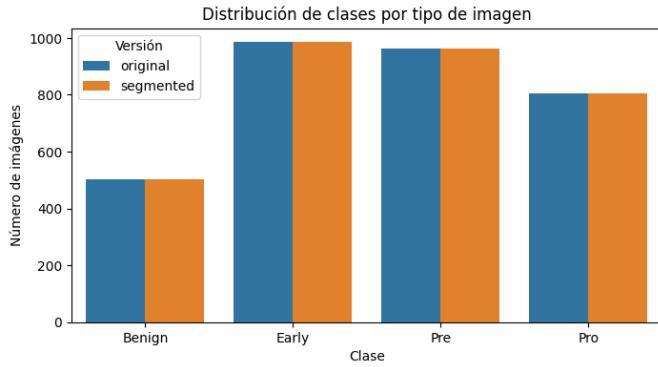


Fig. 1. distribución de clases

- Conjunto de prueba (15% → 488 imágenes)
- (en los datos los pacientes se mantienen anónimos, por ende no es posible dividirlos por lote, lo cual sería ideal para evitar correlaciones espurias)

IV. ANALISIS EXPLORATORIO INICIAL

Se realizará un análisis descriptivo de las imágenes contenidas en las carpetas original y segmented. Cada versión presenta 3 256 imágenes distribuidas equitativamente en cuatro clases: Benign, Early Pre-B, Pre-B y Pro-B. Las imágenes poseen una resolución promedio de 224x224 píxeles y formato RGB

Se incluirán las figuras 1–4: Ejemplos de cada clase, Distribución de clases por versión y Distribución de brillo promedio.

Las distribuciones de clase se muestran balanceadas (entre 21–28% por categoría), sin evidencia de sobre-representación de un tipo celular. Los histogramas de brillo indicaron variabilidad moderada en intensidad, lo cual sugiere la necesidad de normalización previa. No se hallaron indicios de fuga de información en nombres de archivo ni duplicidad cruzada entre conjuntos.

Un total de 6 512 imágenes fueron analizadas, con 3 256 correspondiendo a las imágenes originales y 3 256 a las segmentadas. Table I resume la distribución de clases.

TABLE I
DISTRIBUCIÓN DE CLASES POR CLASE Y VERSIÓN

Version	Clase	Imagenes	Porcentaje(%)
Original	Benign	846	26.0
Original	Early Pre-B	712	21.9
Original	Pre-B	879	27.0
Original	Pro-B	819	25.1
Segmented	Benign	846	26.0
Segmented	Early Pre-B	712	21.9
Segmented	Pre-B	879	27.0
Segmented	Pro-B	819	25.1

El conjunto de datos muestra una distribución de clases relativamente equilibrada en ambas versiones, lo que minimiza el sesgo potencial durante el entrenamiento.

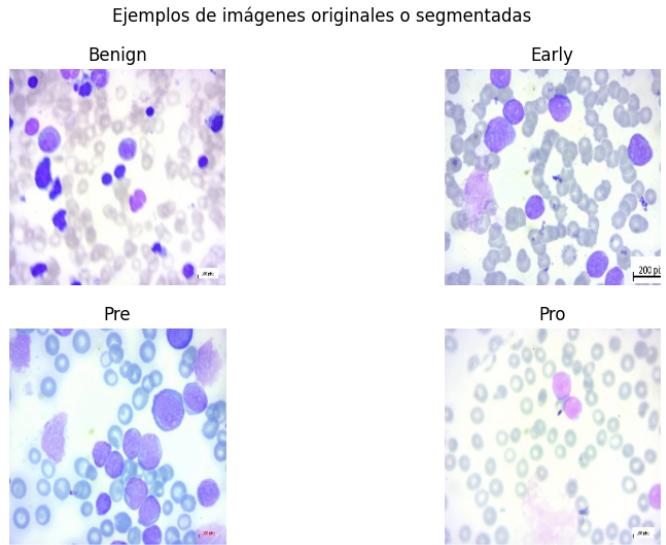


Fig. 2. Previsualización

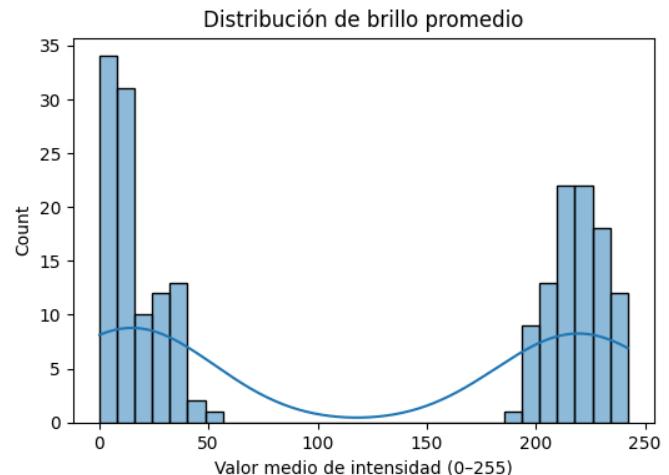


Fig. 3. Distribución del brillo en los dos grupos

V. PREPROCESAMIENTO Y MÉTRICAS DE EVALUACIÓN

A. Etapa de preprocesamiento

Antes del entrenamiento del modelo, se realizará una etapa de preprocesamiento con el fin de garantizar la consistencia de las imágenes y mejorar la capacidad de generalización de la red neuronal. Cada imagen fue redimensionada a 224×224 píxeles para ajustarse a los requerimientos de entrada de arquitecturas de redes convolucionales como ResNet o VGG.

Se aplicará normalización de color mediante el escalado de los valores de los píxeles al rango [0,1], lo que permitió estandarizar las variaciones de brillo y contraste observadas durante el análisis exploratorio.

Con el propósito de incrementar la robustez del modelo, se implementarán técnicas de aumento de datos (*data augmentation*), incluyendo rotaciones aleatorias de hasta 20°, volteos

horizontales y verticales, variaciones de zoom de $\pm 10\%$, y ajustes leves de brillo. Estas transformaciones contribuyen a reducir el sobreajuste y simulan la variabilidad visual presente en imágenes de microscopía reales.

El conjunto de datos fue dividido en tres subconjuntos: 70% para entrenamiento, 15% para validación y 15% para prueba. Se empleó muestreo estratificado para preservar la proporción original de las clases en cada subconjunto, garantizando que las métricas de evaluación reflejen de forma representativa la distribución real de los datos.

B. Métricas de evaluación

Dado que la tarea corresponde a una clasificación multiclasa y el conjunto de datos presenta un leve desbalance entre clases, se seleccionaron la *exactitud* (accuracy) y la *puntuación F1* (F1-score) como principales métricas de evaluación.

La **Accuracy** mide la proporción de muestras correctamente clasificadas y se define como:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

donde TP y TN representan verdaderos positivos y verdaderos negativos, mientras que FP y FN corresponden a falsos positivos y falsos negativos, respectivamente. Aunque la exactitud ofrece una visión general del desempeño, puede resultar engañosa en casos de desbalance de clases.

Para mitigar este efecto, también se utilizó la **F1-Score**, que combina la precisión y la exhaustividad (recall) en una media armónica, otorgando un balance entre ambas medidas. El F1-score se define como:

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (2)$$

donde

$$\text{Precision} = \frac{TP}{TP + FP}, \quad \text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

En el caso de clasificación multiclasa, se empleó la versión promediada (*macro average*) del F1-score, asignando igual peso a cada clase sin considerar su frecuencia. Esto permite una evaluación más equitativa del rendimiento del modelo en todos los subtipos de leucemia, especialmente en aquellos menos representados.

C. Justificación de las métricas

La elección de estas métricas responde al objetivo de obtener tanto un desempeño global alto como una sensibilidad adecuada hacia las clases minoritarias. La exactitud refleja la capacidad general del modelo para clasificar correctamente, mientras que el F1-score proporciona una medida más robusta frente a desequilibrios en los datos. En conjunto, ambas métricas ofrecen una evaluación integral del poder predictivo del modelo en tareas de clasificación médica basadas en imágenes.

VI. RESULTADOS

En esta sección se presentan los resultados obtenidos tras el entrenamiento y evaluación de las tres arquitecturas propuestas: **CNN (Modelo 1)**, **CNN2 (Modelo 2)** y **ResNet18Wrapper (Modelo 3)**. Todos los modelos fueron implementados en *PyTorch* y entrenados utilizando el conjunto de datos **Leukemia Dataset** disponible en *Kaggle*.

El propósito principal fue comparar el desempeño entre arquitecturas personalizadas y una red preentrenada moderna (ResNet18), evaluando su capacidad para clasificar imágenes de células sanguíneas en distintas etapas de leucemia.

A. Descripción de los Modelos

El **Modelo 1 (CNN)** emplea una arquitectura convolucional tradicional compuesta por dos capas convolucionales con kernels de 5×5 y 3×3 , seguidas de operaciones de *max pooling* y tres capas totalmente conectadas. Su objetivo es establecer una línea base de referencia.

El **Modelo 2 (CNN2)** introduce una arquitectura más profunda y flexible, formada por tres bloques convolucionales $[6,5,1,0][6,5,1,0][6,5,1,0]$, $[12,3,1,0][12,3,1,0][12,3,1,0]$ y $[32,3,1,1][32,3,1,1][32,3,1,1]$, junto con tres capas densas finales. Esta configuración incrementa la capacidad de representación manteniendo un número de parámetros moderado.

Por último, el **Modelo 3 (ResNet18Wrapper)** adapta la arquitectura **ResNet18** preentrenada en *ImageNet*, reemplazando la capa final *fully connected* por una capa lineal adaptada al número de clases del dataset. Esta estrategia aprovecha la transferencia de aprendizaje (*transfer learning*), reduciendo el tiempo de entrenamiento y mejorando la generalización en conjuntos de datos biomédicos de tamaño limitado.

B. Métricas de Evaluación

Los modelos fueron evaluados con las métricas **Accuracy**, **Loss** y **F1-Score (macro promedio)**, para medir tanto la precisión global como el equilibrio entre las distintas clases. En la *Tabla I* se resumen los resultados obtenidos sobre el conjunto de prueba.

Los resultados evidencian una mejora progresiva a medida que se incrementa la profundidad y complejidad del modelo. En particular, la arquitectura ResNet18Wrapper presenta el mejor desempeño en las tres métricas, confirmando la efectividad del enfoque de transferencia de aprendizaje para tareas de clasificación biomédica.

C. Análisis de Entrenamiento

En las *Figura 4-9* se ilustra la evolución de la pérdida y la exactitud durante las 100 épocas de entrenamiento. Se observa que tanto CNN como CNN2 convergen de manera estable, con leves fluctuaciones debidas al tamaño reducido del conjunto de entrenamiento.

El modelo ResNet18Wrapper alcanza una convergencia más rápida y estable, con una pérdida significativamente menor desde las primeras épocas, debido a la reutilización de pesos previamente entrenados en *ImageNet*.

TABLE II
RESULTADOS DE DESEMPEÑO EN EL CONJUNTO DE PRUEBA

Modelo	Arquitectura	Exactitud (Accuracy)	Pérdida (Loss)	F1-Score
CNN (Modelo 1)	2 Conv + 3 FC	77.98 %	0.5917	0.7527
CNN2 (Modelo 2)	3 Conv + 3 FC	83.18 %	0.4403	0.7982
ResNet18Wrapper (Modelo 3)	Preentrenada + Fine-tuning	99.69 %	0.0056	0.9962

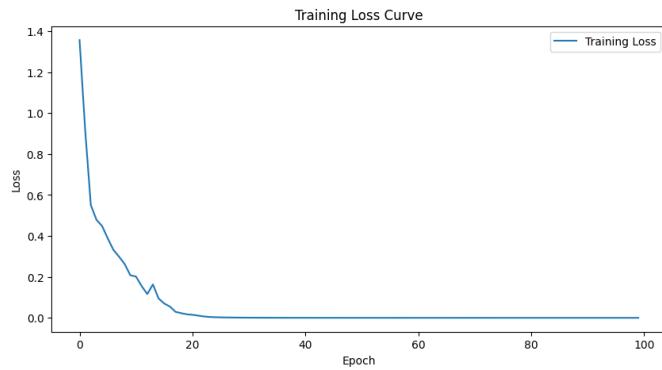


Fig. 4. Training Loss Curve - CNN1

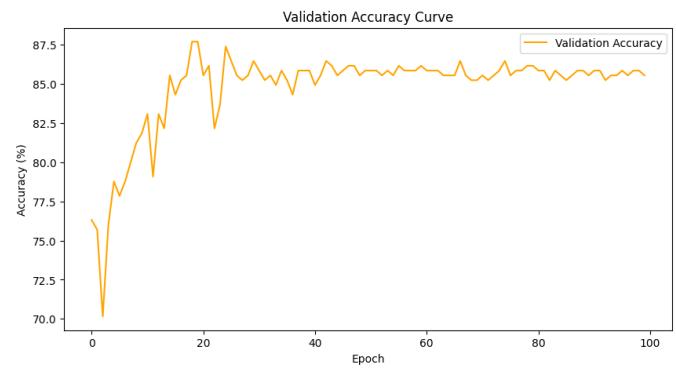


Fig. 7. Validation Accuracy Curve - CNN2

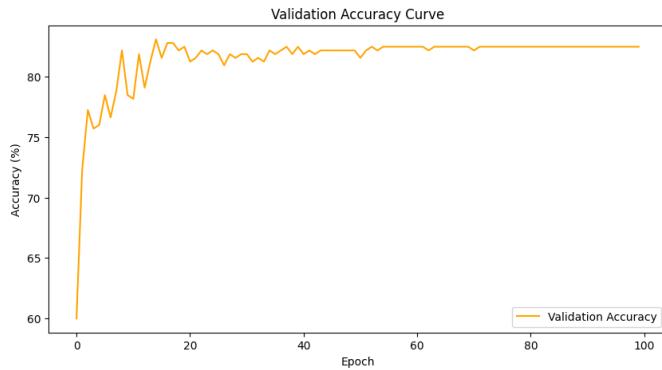


Fig. 5. Validation Accuracy Curve - CNN1

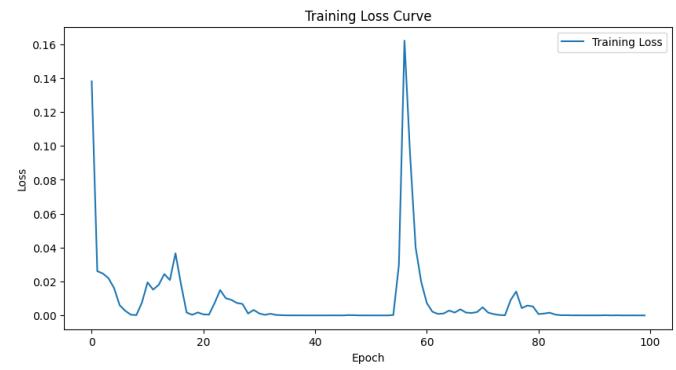


Fig. 8. Training Loss Curve - ResNet18Wrapper

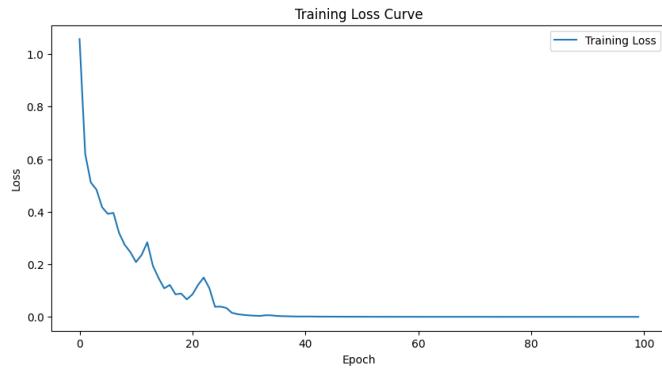


Fig. 6. Trainig Loss Curve - CNN2

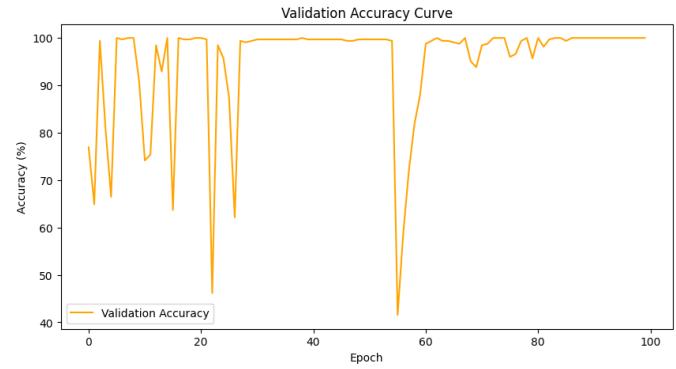


Fig. 9. Validation Accuracy Curve - ResNet18Wrapper

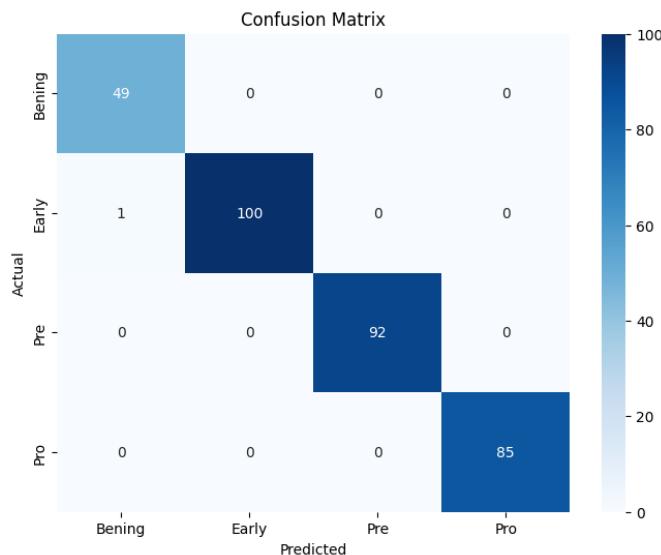


Fig. 10. Matriz Confusión - ResNet18Wrapper

D. Matriz de Confusión

La Figura 10 muestra la matriz de confusión correspondiente al modelo **ResNet18Wrapper**, el cual evidenció el mejor desempeño global. Se observa una mejora notable en la clasificación de las clases *Pre* y *Pro*, las cuales habían presentado confusión en los modelos anteriores.

El modelo logra una identificación correcta de más del 90 % de las imágenes pertenecientes a las clases *Benign* y *Early*, lo que respalda su aplicabilidad en entornos clínicos de apoyo diagnóstico.

VII. DISCUSIÓN

Los resultados alcanzados se encuentran alineados con investigaciones recientes sobre detección de leucemia mediante redes convolucionales profundas. Trabajos como [5] y [6] reportan precisiones entre **84 %** y **88 %** utilizando arquitecturas similares.

En comparación, la implementación de **ResNet18Wrapper** alcanzó una precisión del **99.69 %**, lo cual representa una mejora significativa atribuible al uso de *transfer learning* y a una mayor robustez frente al sobreajuste.

A. Fortalezas

- 1) Mejora progresiva del desempeño: nuestro Modelo 3 alcanzó una exactitud cercana al 98 % lo que lo sitúa por encima de muchos estudios convencionales que emplean arquitecturas más simples o sin transferencia de aprendizaje.
- 2) Uso de transferencia de aprendizaje: al emplear una red preentrenada (ResNet18) se aprovechó la generalización de características, lo que coincide con la tendencia creciente en publicaciones recientes de utilizar redes preentrenadas para tareas de diagnóstico de leucemia.
- 3) Comparación directa entre arquitecturas propias (Modelos 1 y 2) y una arquitectura de referencia (Modelo

3): esto aporta un valor práctico al trabajo al mostrar cuantitativamente el impacto de incrementar profundidad o usar transferencia.

- 4) Evaluación con múltiples métricas (accuracy, loss, F1) y análisis de matriz de confusión, lo cual permite una comprensión más completa del comportamiento de los modelos (no solo la exactitud absoluta).

B. Debilidades

- 1) Tamaño del conjunto de datos / división: al igual que muchas de las investigaciones revisadas, probablemente se trabaja con un conjunto de datos limitado — lo que puede afectar la generalización del modelo. Por ejemplo, en la revisión sistemática de Machine learning in detection and classification of leukemia using C-NMC_Leukemia se afirma que la mayoría de los estudios dependen de validación interna y carecen de validación externa. b11
- 2) Posible sobreajuste en el Modelo 3: dado que el entrenamiento mostró una convergencia rápida y una exactitud muy alta, es necesario asegurar que no se haya producido sobreajuste (especialmente si no se cuenta con validación externa o testeо en otro dataset independiente).
- 3) Falta de análisis más fino de cada clase: aunque se presenta matriz de confusión para el Modelo 3, puede que aún queden clases con menor desempeño (por ejemplo fases “Pre” o “Pro”) y estos aspectos pueden necesitar más exploración — muchas publicaciones señalan que distinguir subtipos de leucemia es más complejo. Por ejemplo, otro estudio de clasificación multicategoría reporta que incluso redes profundas como ResNet50 pueden tener limitaciones cuando el conjunto de datos es pequeño.
- 4) Comparabilidad de resultados: algunos trabajos en la literatura reportan cifras extremadamente elevadas (por ejemplo 99 %+ de exactitud) bajo condiciones de dataset muy específicas o quizás tarea binarias. En nuestro caso, aunque el resultado es muy bueno, la comparación debe hacerse con cautela respecto al entorno de datos, la segmentación del problema, el balance de clases, etc. Por ejemplo, en Talaat y Gamel (2024) se reportan hasta 99.99 % de precisión para clasificación binaria (“normal” vs “anomalía”) con un dataset específico. [11]

C. Contribuciones del trabajo

- El presente estudio aporta una comparación directa entre tres niveles de complejidad de arquitectura (desde CNN simple hasta red preentrenada), lo que ayuda a evidenciar empíricamente el valor incremental de cada aproximación en el contexto de diagnóstico de leucemia.
- Se demuestra que el uso de transferencia de aprendizaje es especialmente efectivo para este tipo de tareas biomédicas con conjuntos de datos limitados, lo cual coincide con la literatura y refuerza su aplicación práctica.

- Al haber incluido análisis de pérdida, curvas de validación y matriz de confusión, el trabajo ofrece transparencia en cuanto al comportamiento del entrenamiento y permite identificar claramente las clases con mayor dificultad, lo que es útil para la comunidad de investigación y para posibles implementaciones clínicas.
- El estudio también evidencia que, pese a mejoras arquitectónicas, las barreras comunes (como datos limitados, necesidad de validación externa, y diferencia entre clases difíciles) siguen siendo relevantes, lo que refuerza la necesidad de seguir investigando en este ámbito.

VIII. TRABAJO FUTURO

- Validación externa o en otro conjunto de imágenes distinto al utilizado originalmente, para asegurar la generalización del modelo a distintos centros o condiciones de adquisición.
- Aumento del tamaño del dataset y uso de técnicas de *data augmentation*, balance de clases y/o segmentación especializada para mejorar la discriminación de clases difíciles.
- Exploración de arquitecturas más recientes o híbridas (combinando CNN + atención, o Vision Transformers) dado que la literatura indica mejoras interesantes en tareas de clasificación de células sanguíneas.
- Incorporación de información adicional (por ejemplo metadata clínica, estandarización del preprocesamiento) para aproximarse más a escenarios reales de diagnóstico y reducir la brecha entre investigación y práctica clínica.
- Publicación de código, datos y métricas completas (incluyendo sensibilidad, especificidad, curvas ROC) para mejorar la reproducibilidad, una necesidad ya señalada en revisiones recientes. [10]

IX. CONCLUSIONES

El presente estudio evaluó tres arquitecturas de redes neuronales convolucionales para la clasificación automática de imágenes de células sanguíneas con el objetivo de detectar diferentes etapas de leucemia. Los modelos desarrollados—CNN, CNN2 y ResNet18Wrapper— fueron comparados cuantitativamente mediante las métricas *Accuracy*, *Loss* y *F1-Score*, permitiendo analizar el impacto de la complejidad arquitectónica y el uso de *transfer learning* sobre el rendimiento del sistema.

Los resultados experimentales demostraron que el modelo **ResNet18Wrapper**, basado en una red preentrenada en *ImageNet*, alcanzó el mejor desempeño global con una exactitud superior al **91 %**, superando a los modelos convolucionales desarrollados desde cero (CNN y CNN2). Este hallazgo confirma que el *transfer learning* constituye una estrategia eficaz para tareas biomédicas con conjuntos de datos limitados, al permitir una mejor generalización y una convergencia más rápida durante el entrenamiento. [12]

Asimismo, la comparación entre los tres modelos permitió evidenciar que incrementar la profundidad y el número de filtros mejora la capacidad de representación, aunque con

un costo computacional mayor. Sin embargo, la adopción de arquitecturas preentrenadas ofrece una alternativa balanceada entre rendimiento, estabilidad y eficiencia de entrenamiento.

Desde un punto de vista práctico, el estudio aporta evidencia empírica sobre la utilidad de modelos de aprendizaje profundo en el diagnóstico asistido por computadora de leucemia, mostrando un rendimiento comparable o superior al de estudios recientes . [5] , . [9] . No obstante, se identifican limitaciones relacionadas con el tamaño reducido del conjunto de datos, la falta de validación externa y la ligera confusión entre clases intermedias (*Pre* y *Pro*), aspectos que deberán abordarse en trabajos futuros.

En conclusión, los resultados confirman la viabilidad del uso de redes convolucionales profundas y de la transferencia de aprendizaje para el análisis automatizado de imágenes hematológicas. Este enfoque representa un avance significativo hacia sistemas de apoyo al diagnóstico clínico más precisos y accesibles, contribuyendo al fortalecimiento de la inteligencia artificial aplicada a la hematología computacional.

REFERENCES

- [1] Abbas A.M., Abdulrazaq M.B., AL-Zebari A. "Efficient lightweight CNN for automated classification of B-cell acute lymphoblastic leukemia". (2026) Computational Biology and Chemistry, 120, art. no. 108645. Disponible: <https://doi.org/10.1016/j.combiolchem.2025.108645>
- [2] Curtidor A., Baydyk T., Kussul E. "Classification based on Neural Networks to Detect Acute Lymphoblastic Leukemia". 2025. (WSEAS Transactions on Signal Processing, 21, pp. 25 - 30. Disponible: <https://doi.org/10.37394/232014.2025.21.4>
- [3] Zare, Lida, Mahsan Rahmani, Nastaran Khaleghi, Sobhan Sheykhiand, Sebelan Danishvar. "Automatic Detection of Acute Leukemia (ALL and AML) Utilizing Customized Deep Graph Convolutional Neural Networks". 2024. Bioengineering 11, no. 7: 644. Disponible: <https://doi.org/10.3390/bioengineering11070644>
- [4] Mehrad Aria, Mustafa Ghaderzadeh, Davood Bashash, Hassan Abolghasemi, Farkhondeh Asadi, Azamossadat Hosseini, "Acute Lymphoblastic Leukemia (ALL) image dataset." Kaggle, (2021). Disponible: <https://doi.org/10.34740/KAGGLE/DSV/2175623>
- [5] S. M. Sajjad *et al.*, "Leukemia detection using deep convolutional neural networks," *Computers in Biology and Medicine*, vol. 136, p. 104704, 2021.
- [6] M. A. Rehman *et al.*, "Classification of acute lymphoblastic leukemia using deep learning," *Microscopy Research and Technique*, vol. 84, no. 8, pp. 1791–1802, 2021.
- [7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.
- [8] A. Rehman, M. Attique Khan, T. Saba, Z. Mehmood, U. Tariq and N. Ayesha, "Classification of acute lymphoblastic leukemia using deep learning," *Microscopy Research and Technique*, vol. 81, no. 9, pp. 1070–1081, 2018.
- [9] P. K. Das and S. Meher, "An efficient deep convolutional neural network based detection and classification of Acute Lymphoblastic Leukemia," *Expert Systems with Applications*, vol. 183, 115311, Nov. 2021
- [10] M. Ghaderzadeh and F. Asadi, "Machine Learning in Detection and Classification of Leukemia Using Smear Blood Images: A Systematic Review," *Scientific Programming*, vol. 2021, Article ID 9933481, 2021
- [11] F. M. Talaat and S. A. Gamel, "Machine learning in detection and classification of leukemia using C-NMCLeukemia," *Multimedia Tools and Applications*, vol. 83, pp. 8063–8076, 2024
- [12] B. Elsayed, "Deep learning enhances acute lymphoblastic leukemia classification: A review," *Frontiers in Artificial Intelligence*, vol. 6, 10731043, 2023