

# Algoritmo para la Optimización de Análisis en Ecografías Mamarias para la Detección Temprana del Cáncer de Mama en Entornos con Recursos Limitados

1<sup>st</sup> Alessandra Aldave Javier

*Facultad de Ciencias e Ingeniería*

*Universidad Peruana Cayetano Heredia*

Lima, Perú

alessandra.aldave@upch.pe

1<sup>st</sup> Valeria Zavaleta Jave

*Facultad de Ciencias e Ingeniería*

*Universidad Peruana Cayetano Heredia*

Lima, Perú

valeria.zavaleta@upch.pe

2<sup>nd</sup> Moisés Meza Rodríguez

*Facultad de Ciencias e Ingeniería*

*Universidad Peruana Cayetano Heredia*

Lima, Perú

moises.meza@upch.pe

**Abstract**—This study presents an algorithm to optimize breast ultrasound analysis aimed at improving early detection of breast cancer in resource-limited settings. A low computational cost deep learning model was developed to identify characteristic frames in breast ultrasound videos, classifying them as benign or malignant. The methodology included the use of two breast ultrasound image databases and the design of an architecture combining a U-Net for segmentation and a convolutional neural network (CNN) for classification. Although the U-Net model showed limitations in segmentation due to possible overfitting, the CNN model achieved a high accuracy in classifying significant frames in the test set. The results suggest that this approach could be promising for implementation in embedded systems for rapid analysis of breast ultrasounds, although further research is needed to improve model generalization and validate its effectiveness in real clinical settings.

**Index Terms**—ultrasonido mamario, redes neuronales convolucionales, cáncer de mama, segmentación de imágenes, frame significativo

## I. INTRODUCTION

Por muchos años y hasta la actualidad, el cáncer de mamario (CM) constituye uno de los más grandes desafíos de administración sanitaria a alrededor del mundo [1][2], en donde casi el 100% el cáncer de mama se caracteriza por ser una enfermedad en la cual las células que pertenecen al tejido mamario se encuentran en un proceso de crecimiento sin control [4]. Cuando la identificación y valoración del estado de esta enfermedad se efectúa de manera temprana, las probabilidades de éxito que se puede conseguir con el tratamiento son altas; no obstante, si la detección es tardía, el porcentaje de proliferación de células tumorales resulta elevado, lo cual sitúa la condición de la persona que adolece la enfermedad en el último estadio del cáncer de seno (estadio IV) [5] en el cual, el cáncer inicialmente concentrado en las mamas se disemina hacia otras áreas del cuerpo conduciendo a un cáncer metastásico que, pese a que puede ser tratado, permanece incurable [6]. A nivel global, las cifras son alarmantes: en el año 2020 se diagnosticaron aproximadamente 2.26 millones de nuevos casos de CM, lo que lo clasifica como la causa número cinco más

frecuente de defunción en la población femenina, con 685,000 fallecimientos registrados [7]. En el 2022, el CM ha sido la variedad de cáncer más extendida entre la población femenina en 157 países de 185, y en el mismo año, esta enfermedad produjo 670 000 muertes a nivel global [3]. Sin embargo, persisten desigualdades significativas en la distribución de la enfermedad. En naciones con un muy alto Índice de Desarrollo Humano, 1 de cada 12 mujeres son diagnosticadas y 1 en 71 mujeres mueren por cáncer de mama. En contraste, en países con bajo IDH, solamente a 1 de cada 27 mujeres se le detecta con CM y se estima que 1 de cada 48 mujeres morirán bajo esta causa [3]. En el contexto peruano, la situación no es menos preocupante. El CM es la forma de cáncer más prevalente en las mujeres peruanas [8]. Además, con una incidencia en constante aumento, anualmente, se reportan unos 28 casos nuevos por cada 100,000 residentes, reflejando una tendencia preocupante en las últimas dos décadas [9]. En este escenario, la detección temprana emerge como un factor crucial para elevar las tasas de supervivencia. El diagnóstico temprano de CM no solo aumenta las posibilidades de cura al detectar la enfermedad en etapas más tempranas, cuando el tumor es más tratable, sino que también abre la puerta a opciones de tratamiento menos invasivas, preservando así el estándar de vida de las pacientes [10]. Además, detectar la enfermedad a tiempo se refleja en una baja en las tasas de mortalidad asociada a la enfermedad, salvando vidas y mejorando el pronóstico general. Asimismo, cabe resaltar la desigualdad en los criterios de tamizaje en los sistemas de salud, así como a la falta de una cultura preventiva adecuada y a la concentración de mamógrafos en la ciudad de Lima. Según datos recopilados hasta el 2022 por La República, de la Dirección de Prevención y Control de Cáncer del Ministerio de Salud (Minsa), solo 31 de los 49 mamógrafos disponibles en los hospitales del país están operativos. Este panorama es preocupante considerando que, según el Instituto Nacional de Estadística e Informática (INEI), se tiene una media de 1 dispositivo de mamografía por cada 1 millón de residentes

que Perú presenta, pese a que actualmente supera los 33 millones [11]. Además, estos mamógrafos en funcionamiento no están distribuidos equitativamente en todas las regiones del país, lo que significa que solo algunas tienen acceso a estos equipos cruciales para el diagnóstico temprano del CM. La falta de diagnóstico se ve agravada por la escasez de profesionales capacitados para operar estas máquinas esenciales en el diagnóstico de la enfermedad. A diario se diagnostican 19 nuevos casos de CM en Perú, afectando aproximadamente a 6,860 personas al año. Es evidente que se requiere no solo más equipos, sino también una inversión significativa en la capacitación de personal médico para utilizar eficazmente estos recursos [12].

## II. MATERIALES Y MÉTODOS

El presente proyecto busca construir un modelo de Deep Learning de bajo costo computacional sin sacrificar el accuracy para poder identificar los frames característicos de videos de ultrasonido pertenecientes a cáncer de mama maligno o a cáncer de mama benigno. El objetivo de utilizar este alcance es poder realizar un posterior despliegue del modelo en un sistema embebido de NVIDIA que permita conseguir una rápida clasificación de videos de ultrasonido.

Para dicho propósito se utilizó dos modelos de Redes Neuronales Convolucionales (CNN) debido a que destacan y sobresalen por su desempeño en la clasificación de imágenes [13][14]. El modelo que utiliza una arquitectura pura de CNN se empleó para la tarea de clasificación de imágenes o frames característicos dentro de un video de ultrasonido mamario; mientras que para la tarea de segmentación de imágenes de ultrasonido y generación de máscaras se utilizó una arquitectura U-Net la cual usa únicamente redes neuronales totalmente convolucionales [15][16] y es considerada extensión y versión mejorada de las CNN para poder realizar tareas de segmentación al crear mapas de segmentación altamente detallados [17].

### A. Base de datos

Se utilizaron dos bases de datos distintas, la primera base de datos llamada Miccai 2022 BUV Dataset [18] corresponde al dataset principal del proyecto y el cual se busca emplear para la tarea de clasificación de frames característicos. Este data que contiene 188 videos de ecografías mamarias adquiridos con los ecógrafos LOGIQ-E9 y PHILIPS TIS L9-3, de los cuales 113 son de lesiones malignas y 75 de lesiones benignas. Los videos están etiquetados con rectángulos que delimitan las lesiones mamarias y fueron clasificados por dos patólogos con ocho años de experiencia.

El segundo conjunto de datos utilizado en el proyecto se trata de un dataset llamado Dataset BUSI with GT. Este dataset consta de 780 imágenes en formato PNG de ultrasonido mamario con un tamaño promedio de imagen de 500 x 500 píxeles. Todas las imágenes han sido clasificadas excluyentemente en las categorías normal, benigno o maligno. La data fue adquirida en 600 pacientes mujeres de entre 25 años a 75 años en el año 2018 en el Hospital Baheya en el Cairo, Egipto

[19]. Las imágenes del set de datos han sido adquiridas con los instrumentos: sistema de ultrasonido LOGIQ E9 y LOGIQ E9 Agile, los cuales se caracterizan por brindar una resolución de 1280x1024 [19].

### B. Preprocesamiento

En este estudio, el dataset auxiliar (Dataset BUSI with GT) atravesó un preprocesamiento de redimensionamiento de las imágenes a un tamaño de 256 píxeles por 256 píxeles con la finalidad de tener un tamaño fijo de las mismas además de una transformación posterior a tensores para que el modelo U-Net pueda manejar los datos ingresados.

La razón por la cual se decidió utilizar un procedimiento sucesivo de segmentación y clasificación es debido a que la predicción de máscaras sintéticas para el dataset principal a partir de un entrenamiento con el dataset auxiliar permite reducir la complejidad del modelo CNN, reducir el tiempo de procesamiento y simplifica la representación de la información relevante en las imágenes de ultrasonido mamario [20].

En cuanto al dataset principal (Miccai 2022 BUV Dataset), se comenzó trabajando con un conjunto de datos estructurado en formato JSON denominado 'imagenet\_vid\_train\_15frames.json', el cual contiene información detallada sobre videos ecográficos y sus frames significativos correspondientes.

Se iteró a través de los 120 primeros elementos presentes en la lista de datos, y se extrajo los frames significativos de cada video. Posteriormente, se procedió a cargar las imágenes desde la carpeta principal de archivos. Cada imagen se redimensionó a un tamaño de 224x224 píxeles, una dimensión estándar para la entrada de modelos de redes neuronales convolucionales (CNN).

Finalmente, cada imagen se convirtió en un array numpy para poder manipularla y utilizarla como entrada del modelo.

Las imágenes redimensionadas fueron etiquetadas según su relevancia como frames significativos o no significativos. Esta etiqueta se determinó comparando el nombre de cada imagen con la lista de frames significativos obtenida previamente. Si el nombre de la imagen coincidía con algún frame significativo, se asignaba la etiqueta 1; de lo contrario, se asignaba la etiqueta 0. Estas etiquetas se almacenaron en una lista junto con las imágenes correspondientes.

### C. Diseño del Algoritmo

Como primer paso, se adaptó un modelo de U-Net grande y preentrenado para la segmentación de tumores de mama a partir de imágenes de ultrasonido propuesto en la conferencia MICCAI 2023 [21]. Este modelo U-Net presenta las siguientes características:

- **Arquitectura de contracción y expansión:** El encoder presenta cuatro capas de convolución y agrupación máxima (max pooling), que redujeron progresivamente la dimensionalidad espacial mientras incrementaban la profundidad de características. El decoder utilizó capas de transposición de convolución (ConvTranspose2d) y

concatenación de características (UpConcat), seguidas por convoluciones para restaurar la resolución espacial.

- **Capas convolucionales:** Cada bloque convolucional aplicó dos capas de convolución 3x3, además de una normalización por lotes (BatchNorm), una función de activación ReLU, y un dropout del 30% para reducir el sobreajuste.

La arquitectura de este modelo se configuró para comenzar con 32 canales de características y duplicarse en cada etapa de la contracción, alcanzando hasta 512 canales en la capa más profunda.

El optimizador elegido fue AdamW debido a que supera la regularización que L2 podría lograr en variables con grandes gradientes, conduciendo así a una mejor pérdida de entrenamiento y error de generalización [22]. La función de pérdida para esta clasificación binaria o segmentación de imágenes fue Binary Cross Entropy (BCE).

Los hiperparámetros definidos consistieron en un tamaño de lote igual a 8 para el entrenamiento y validación del modelo, una tasa de aprendizaje de 0.001 y un entrenamiento y validación a lo largo de 3 épocas.

La elección de 3 épocas de entrenamiento fue debido a que en unas primeras iteraciones piloto realizadas se encontró que la tercera época ofrecía los mejores resultados ya que en la época sucesiva no se evidenciaba mejoras significativas en el aprendizaje del modelo y en su capacidad de clasificación.

El siguiente paso fue diseñar y entrenar un modelo de red neuronal convolucional (CNN) para clasificar los frames significativos en los videos ecográficos.

Una vez cargadas las imágenes ya etiquetadas (0 o 1 para saber que se trataba de un frame significativo), se procedió a normalizarlas dividiendo todos los valores de píxeles por 255, lo cual estandariza los datos y facilita el entrenamiento del modelo.

A continuación, se dividió el conjunto de datos en conjuntos de entrenamiento y prueba, reservando un 20% del total de datos para el conjunto de prueba para asegurar que el modelo fuera evaluado en datos no vistos durante el entrenamiento y proporcionando una medida robusta de su rendimiento.

Luego se procedió a la construcción del modelo CNN utilizando la biblioteca TensorFlow Keras teniendo en cuenta las siguientes consideraciones:

- **Regularización L2:** La regularización L2 se aplicó en las capas convolucionales y densas del modelo en el código proporcionado para mitigar el sobreajuste. En las capas convolucionales, se utilizó para penalizar los pesos de los filtros durante el entrenamiento, limitando su magnitud y ayudando así a controlar la complejidad del modelo. De manera similar, en las capas densas, la regularización L2 se utilizó para penalizar los pesos de las neuronas, asegurando que no crezcan demasiado y comprometan la capacidad del modelo para generalizar correctamente a datos nuevos.

- **Capas de convolución y pooling:** El modelo se configuró con tres capas convolucionales de 32, 64 y 64 filtros respectivamente, cada uno de tamaño (3, 3) con activación ReLU para introducir no linealidades en el modelo. Estas capas fueron seguidas por capas de MaxPooling con tamaño de (2, 2) para reducir la dimensionalidad y captar características importantes de las imágenes. Además, se añadieron capas de Dropout para regularizar el modelo y evitar el sobreajuste, y capas densas (Dense) para la clasificación binaria final utilizando una activación sigmoide.
- **Capas densas:** La capa de flattening transformó las salidas de las capas convolucionales en un vector unidimensional, que luego fue procesado por una capa densa de 64 unidades con activación ReLU. La capa de salida utilizó una activación sigmoide para la clasificación binaria (frame significativo o no).
- **Compilación y entrenamiento:** El modelo se compiló con el optimizador Adam, seleccionado por su capacidad de ajuste adaptativo del learning rate, y se usó la función de pérdida binary\_crossentropy adecuada para problemas de clasificación binaria. Se implementaron callbacks como Early Stopping y Model Checkpointing, para monitorear la pérdida de validación y guardar el modelo con los mejores pesos respectivamente. Estos callbacks ayudan a prevenir el sobreajuste y a asegurar que el modelo final capture las características generales de los datos de manera efectiva.

Finalmente, evaluamos el desempeño del modelo entrenado en el conjunto de prueba, calculando métricas de rendimiento como exactitud, recall y F1-score. Visualizamos estas métricas para demostrar la eficacia del modelo en identificar frames significativos en las ecografías de tumores mamarios, validando así su capacidad para asistir en la detección temprana de cáncer de mama.

Con esta metodología detallada, logramos no solo construir y entrenar una CNN robusta, sino también validar su capacidad de generalización y su rendimiento en un conjunto de datos independiente.

### III. RESULTADOS

El modelo U-Net para la segmentación de imágenes de ultrasonido mamario de clase normal, benigno y maligno se llevó a cabo utilizando como entradas para el entrenamiento un 75% (583) de las imágenes que componen el Dataset BUSI with GT y para la validación, un 25% (197) del mismo dataset.

Las métricas que se incluyeron para el análisis de resultados fueron la pérdida (loss), exactitud (accuracy) y el coeficiente de Dice (Dice coefficient). Esta última métrica es elegida dado que es usada ampliamente para evaluar la exactitud con la que se superpone una segmentación automatizada con una segmentación manual[23]. A continuación se reportan los resultados obtenidos para cada métrica analizada durante el entrenamiento y validación a través de 3 épocas.

- Época 1: Se obtiene una pérdida 0.4627 durante el entrenamiento y de 0.3638 durante la validación. Además, una exactitud de entrenamiento de 0.8568 y de validación de 0.9266.
- Época 2: La pérdida disminuye a 0.3669 y a 0.2973 para entrenamiento y validación respectivamente; mientras que la exactitud aumenta a 0.9210 para el set de entrenamiento; sin embargo, el valor para la validación se mantiene constante.
- Época 3: Se obtienen valores más bajos para la pérdida: 0.2548 y 0.2601 para entrenamiento y validación respectivamente. La exactitud se incrementa mínimamente para el entrenamiento y se mantiene constante para la validación.

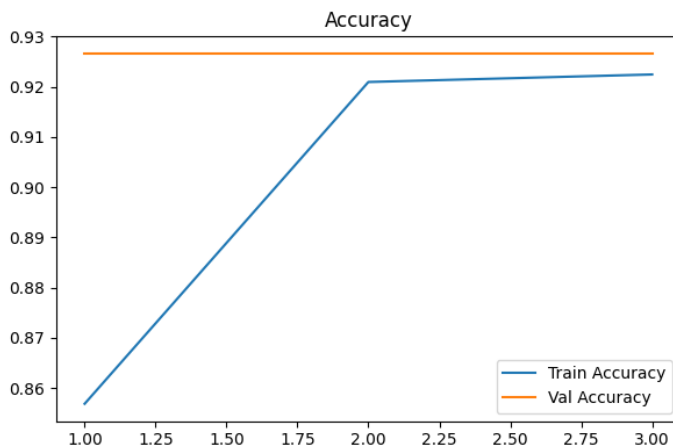


Fig. 1. Gráfica de exactitud (accuracy) de entrenamiento y validación a lo largo de las épocas.

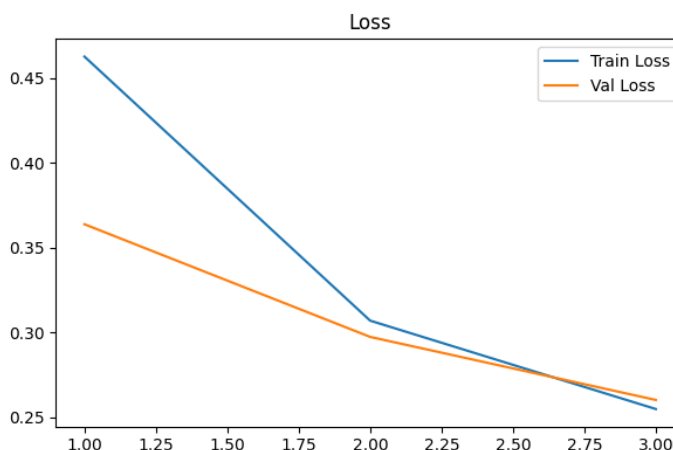


Fig. 2. Gráfica de pérdida (loss) de entrenamiento y validación a lo largo de las épocas.

En cuanto al modelo CNN para clasificación de frames significativos en videos ecográficos, el conjunto de entrenamiento

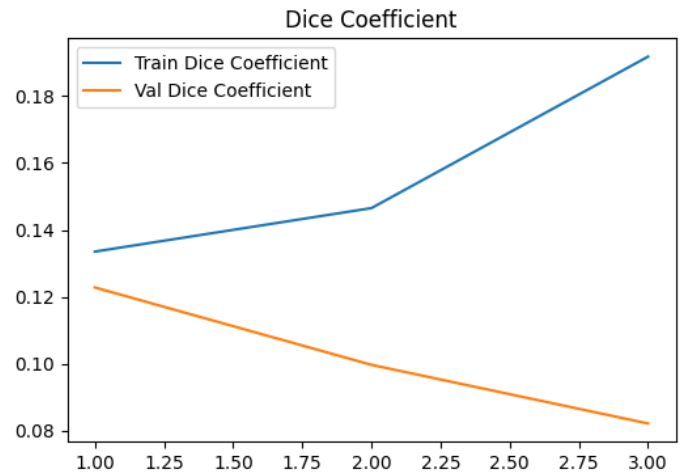


Fig. 3. Gráfica de Coeficiente de Dice (Dice Coefficient) de entrenamiento y validación a lo largo de las épocas.

contiene 13,054 imágenes de tamaño 224x224x3, mientras que el conjunto de prueba está compuesto por 3,264 imágenes del mismo tamaño.

El entrenamiento se limitó a 4 épocas debido a la rápida convergencia del modelo hacia el 100% de exactitud.

A continuación se detalla la progresión de los parámetros para cada época:

- Época 1: Se alcanzó una exactitud del 99.19% en el conjunto de entrenamiento, con una pérdida de 0.7207. En el conjunto de validación, se obtuvo una exactitud del 100% con una pérdida de 0.0079.
- Época 2: La exactitud en el conjunto de entrenamiento alcanzó el 100%, con una pérdida reducida a 0.0073. En validación, se mantuvo una exactitud del 100% con una pérdida aún menor de 0.000469.
- Época 3: Se mantuvo la exactitud del 100% en ambos conjuntos, con una pérdida de 0.0074 en entrenamiento y 0.0076 en validación.
- Época 4: Aunque la exactitud siguió siendo del 100%, se observó un aumento en la pérdida tanto en entrenamiento (0.0056) como en validación (0.0288).

Tras la conclusión de las 4 épocas, se evaluó el modelo en el conjunto de prueba, donde se confirmó una exactitud del 100% en la clasificación de los frames significativos. Esto se reflejó en el reporte de clasificación con métricas de exactitud, recall y f1-score del 100% para la clase positiva.

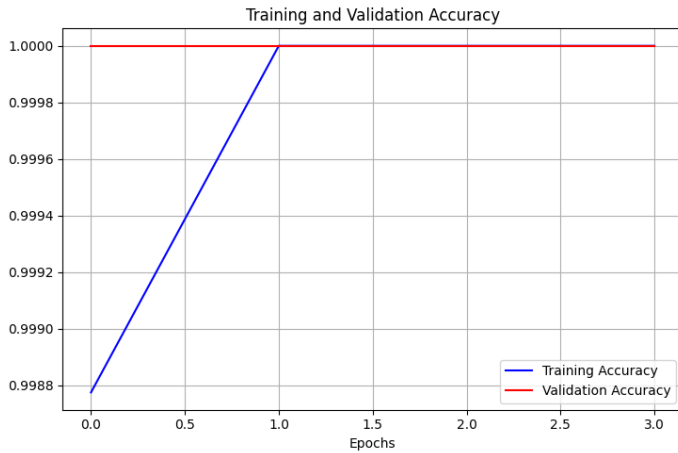


Fig. 4. Gráfica de exactitud (accuracy) de entrenamiento y validación a lo largo de las épocas.

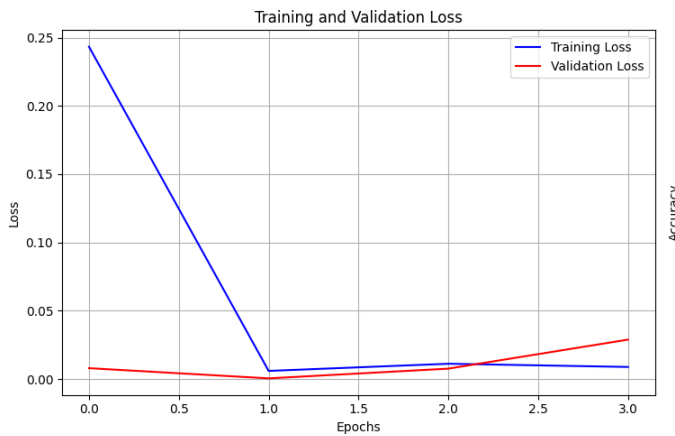


Fig. 5. Gráfica de pérdida (loss) de entrenamiento y validación a lo largo de las épocas.

#### IV. DISCUSIÓN

El presente proyecto busca construir un modelo de Deep Learning de bajo costo computacional sin sacrificar el accuracy al identificar los frames característicos de videos de ultrasonido pertenecientes a cáncer de mama maligno o a cáncer de mama benigno donde el objetivo de utilizar es lograr un posterior despliegue del modelo en un sistema embebido de NVIDIA para conseguir una rápida clasificación de vídeos de ultrasonido.

El objetivo inicial del proyecto presentado fue realizar una fusión y automatización del proceso de segmentación llevado a cabo por un modelo U-Net, seguido de una clasificación utilizando una Red Neuronal Convolutiva (CNN). Sin embargo, las métricas obtenidas durante la segmentación evidenciaron un sobreajuste pronunciado, marcando una diferencia notoria entre los resultados de entrenamiento y validación.

A pesar de que la pérdida disminuye y la exactitud (accuracy) incrementa gradualmente a medida que el proceso de aprendizaje evoluciona, el coeficiente de Dice es innegable-

mente pequeño. Esto se interpreta como una precaria superposición entre la región segmentada creada a partir de la predicción y la máscara binaria original del dataset para cada imagen específica. La visualización de las máscaras predichas respalda las métricas numéricas obtenidas, mostrando que las predicciones no coinciden adecuadamente con las máscaras reales.

El preprocesamiento inicialmente aplicado al Dataset BUSI with GT comprendía una etapa de data augmentation para incrementar la variedad y garantizar la generalización del modelo, utilizando técnicas como flip vertical, flip horizontal y rotación, además de resize y conversión a tensores. Sin embargo, este proceso de data augmentation no se incluyó finalmente debido a que los resultados de la segmentación y predicción eran de menor calidad.

Una de las razones que podrían haber influido en la calidad de la predicción y el rendimiento del modelo U-Net implementado en este trabajo es la elección de hiperparámetros como tasa de aprendizaje y tamaño del batch y la definición de las capas que componen la arquitectura U-Net. Este último factor influye determinadamente en el éxito del modelo [24]; sin embargo, debido a las limitaciones de poder computacional disponible no ha sido posible converger a un modelo robusto, preciso y exacto después de las pruebas iterativas realizadas.

Es debido a ello que en el resultado final no se implementó el modelo de segmentación debido a que las máscaras predichas no favorecerían la eficiencia del modelo CNN y podrían incluso afectar negativamente su rendimiento actual. El resultado del 100% de exactitud en el conjunto de prueba puede atribuirse principalmente a un posible sobreajuste del modelo a los datos de entrenamiento. Esto sugiere que el modelo pudo haber memorizado los detalles específicos de las imágenes del conjunto de entrenamiento en lugar de aprender patrones generales que se puedan aplicar a datos nuevos. Una de las razones para este sobreajuste podría ser la falta de suficiente variedad en las imágenes utilizadas para el entrenamiento, lo que limita la capacidad del modelo para generalizar a datos ecográficos nuevos y distintos de los que ya ha visto durante el entrenamiento. Por esta razón, inicialmente se planeó aplicar aumento de datos para mejorar la generalización del modelo, este paso no se implementó en el proceso actual debido a limitaciones en cuanto al poder computacional.

#### REFERENCES

- [1] Y. Xu et al., "Global trends and forecasts of breast cancer incidence and deaths," *Scientific Data*, vol. 10, no. 1, May 2023. doi:10.1038/s41597-023-02253-5
- [2] Y. Feng et al., "Breast cancer development and progression: Risk factors, cancer stem cells, signaling pathways, genomics, and Molecular Pathogenesis," *Genes diseases*, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6147049/> (accessed Apr. 9, 2024).
- [3] "Breast cancer," World Health Organization, <https://www.who.int/news-room/fact-sheets/detail/breast-cancer> (accessed Apr. 9, 2024).
- [4] "What is breast cancer?," Centers for Disease Control and Prevention, [https://www.cdc.gov/cancer/breast/basic\\_info/what\\_is\\_breast\\_cancer.htm](https://www.cdc.gov/cancer/breast/basic_info/what_is_breast_cancer.htm) (accessed Apr. 9, 2024). \Breastcancerstaging," *AmericanCollege* <http://www.facs.org/for-patients/home-skills-for->

*patients/breast – cancer – surgery/breast – cancer – types/breast – cancer – staging/ : : text = There*

- [5] Bcrf, “Metastatic breast cancer: Symptoms, treatment, research: BCRF,” Breast Cancer Research Foundation, <https://www.bcrf.org/blog/metastatic-breast-cancer-symptoms-treatment> (accessed Apr. 9, 2024).
- [6] “Cáncer de mama,” Cancer.net, Jun. 17, 2014. <https://www.cancer.net/es/tipos-de-c>
- [7] Z. D. Morante et al., “Diagnosis and treatment of HER2 + breast cancer: Clinical Practice Guide of the Peruvian Society of Cancerology,” SciELO, vol. 81, no. 4, pp. 458–465, Dec. 2020. [Online]. Available: <https://dx.doi.org/10.15381/anales.v81i4.18839>
- [8] “Vista de El cáncer de mama en el Perú y el mundo,” Intercienciamedica.com, 2024. <https://intercienciamedica.com/intercienciamedica/article/view/89/87> (accessed Apr. 09, 2024).
- [9] World, “El diagnóstico temprano del cáncer salva vidas y reduce los costos de tratamiento,” Who.int, Feb. 03, 2017. <https://www.who.int/es/news/item/03-02-2017-early-cancer-diagnosis-saves-lives-cuts-treatment-costs> (accessed Apr. 09, 2024).
- [10] A. Ortega, “En el Perú, el Minsa solo tiene un mamógrafo operativo por cada millón de habitantes,” Larepublica.pe, Oct. 20, 2022. <https://larepublica.pe/sociedad/2022/10/19/cancer-de-mama-en-el-peru-el-minsa-solo-tiene-1-mamografo-operativo-por-cada-millon-de-habitantes-dia-internacional-de-lucha-contra-el-cancer-de-mama> (accessed May 08, 2024).
- [11] A. RSNA, “Ultrasonido de Senos,” Radiologyinfo.org, 2022. <https://www.radiologyinfo.org/es/info/breastus> (accessed Apr. 09, 2024).
- [12] N. Jmour, S. Zayen, and A. Abdelkrim, “Convolutional neural networks for image classification,” 2018 International Conference on Advanced Systems and Electric Technologies (ICASET), Mar. 2018, doi : <https://doi.org/10.1109/aset.2018.8379889>. N. Siddique, S. Paheding, C. P. Elkin, and V. Devabhaktuni, “U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications,” *IEEE Access*, vol. 9, pp. 82031–82057, 2021, doi : <https://doi.org/10.1109/access.2021.3086020>.
- [13] M. H. Hesamian, W. Jia, X. He, and P. Kennedy, “Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges,” *Journal of Digital Imaging*, vol. 32, no. 4, pp. 582–596, May 2019, doi: <https://doi.org/10.1007/s10278-019-00227-x>.
- [14] Seshadri Sastry Kunapuli and Praveen Chakravarthy Bhallamudi, “A review of deep learning models for medical diagnosis,” Elsevier eBooks, pp. 389–404, Jan. 2021, doi: <https://doi.org/10.1016/b978-0-12-821777-1.00007-0>.
- [15] N. Siddique, S. Paheding, C. P. Elkin, and V. Devabhaktuni, “U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications,” *IEEE Access*, vol. 9, pp. 82031–82057, 2021, doi: <https://doi.org/10.1109/access.2021.3086020>.
- [16] Z. Lin, J. Lin, L. Zhu, H. Fu, J. Qin, and L. Wang, “A New Dataset and a Baseline Model for Breast Lesion Detection in Ultrasound Videos,” *Lecture notes in computer science*, pp. 614–623, Jan. 2022, doi: [https://doi.org/10.1007/978-3-031-16437-8\\_9](https://doi.org/10.1007/978-3-031-16437-8_9). W. Al – Dhabyani, M. Goma, H. Khaled, and A. Fahmy, “Dataset of breast ultrasound images,” *Data in Brief*, vol. 28, p. 104863, Feb. 2020, doi : [10.1016/j.dib.2019.104863](https://doi.org/10.1016/j.dib.2019.104863)
- [17] K. Y. Rao, M. J. Stephen, and D. S. Phanindra, “Classification based image segmentation approach,” *International Journal of Computer Science and Technology*, vol. 3, no. 1, pp. 112–115, 2012.
- [18] M. Li, K. Sun, Y. Gu, K. Zhang, Y. Sun, Z. Li, and D. Shen, “Developing Large Pre-trained Model for Breast Tumor Segmentation from Ultrasound Images,” in *\*MICCAI 2023\**, Jun. 2023. [Online]. Available: <https://conferences.miccai.org/2023/papers/191-Paper1896.html>. [Accessed: Jun. 9, 2024].
- [19] TensorFlow Team, “TensorFlow Addons Documentation,” TensorFlow, 2024. [Online]. Available: [https://www.tensorflow.org/addons/api\\_docs/python/tfa/optimizers/AdamW](https://www.tensorflow.org/addons/api_docs/python/tfa/optimizers/AdamW). [Accessed : Jun. 10, 2024]. A. A. Taha and A. Hanbury, “Metrics for evaluating 3D medical image segmentation : analysis, selection, and tool,” *BMC Medical Imaging*, vol. 15, no. 29, Aug. 2015, doi : <https://doi.org/10.1186/s12880-015-0068-x>.
- [20] H. Lu, Y. She, J. Tie, and S. Xu, “Half-UNet: A Simplified U-Net Architecture for Medical Image Segmentation,” *Frontiers in Neuroinformatics*, vol. 16, Jun. 2022, doi: <https://doi.org/10.3389/fninf.2022.911679>.