

FACULTAD DE INGENIERÍA

Carrera de Ingeniería de Sistemas Computacionales

“CLASIFICACIÓN DE IMÁGENES MÉDICAS PARA LA DETECCIÓN DEL CÁNCER DE MAMA MEDIANTE REDES NEURONALES”

Tesis para optar al título profesional de:

Ingeniero de Sistemas Computacionales

Autores:

Marcos Pavel Riera Iziga

Waldir Sotelo Hernandez

Asesor:

Ing. Neicer Campos Vásquez

<https://orcid.org/0000-0003-1508-6575>

Lima - Perú

JURADO EVALUADOR

Jurado 1 Presidente(a)	Denis Christian Ovalle Paulino	40234321
	Nombre y Apellidos	N° DNI

Jurado 2	Jorge Rosvin Narvaez Villacorta	41455569
	Nombre y Apellidos	N° DNI

Jurado 3	Maritza Cabana Cáceres	40501601
	Nombre y Apellidos	N° DNI

DEDICATORIA

El presente trabajo de investigación lo dedico a mis padres Marcos y Krupskaia, por su apoyo incondicional, constante consejo, entrega y amor. A mis abuelos Roger y Paquita quienes me motivaron a seguir adelante con el amor que siempre los caracterizó y guiándome para ser una buena persona. A mis tíos Roger, Carlos y Pepe quienes a pesar de la distancia nunca dejan de preocuparse por mí y siempre buscan la forma de apoyarme. A mi hermano Rodrigo quien fue uno de mis principales soportes durante esta etapa tan dura que me tocó vivir. Finalmente, a mi amigo Waldir, con quien logramos concluir de manera satisfactoria este trabajo, a pesar de las dificultades que se presentaron a lo largo de su elaboración. El apoyo recibido siempre será una de las razones por las que les estaré eternamente agradecido.

Marcos Riera Iziga

El presente trabajo de investigación, lo dedico en primer lugar a Dios por habernos guiado y dado fortaleza suficiente para continuar con el logro de nuestro objetivo. A mis amados padres Pelayo y Enid, por su incondicional amor, consejos, trabajo y sacrificio, por todo ello que me hace estar tan agradecido y orgulloso de ser su hijo. A mi querida hermana Susana, por su apoyo moral y constante presencia en cada etapa de este largo camino. A mi estimado amigo Marcos que es parte importante de esta gran aventura de conocimiento, que con su esfuerzo y dedicación logramos grandes cosas. Y a quienes nos alentaron, enriquecieron en conocimientos y nos abrieron las puertas permitiendo que, por más difícil que sea la situación, podamos continuar y llegar a finalizar este que es uno de nuestros sueños más preciados.

Waldir Sotelo Hernández

AGRADECIMIENTO

Agradecer a Dios por guiarnos en cada momento y bendecirnos con vida durante este proceso. Gracias a nuestros padres por confiar y creer en nosotros, por ser los primeros impulsores para no dejar de luchar por lo que queremos, por los principios y muchos valores inculcados durante nuestros años de vida y por su eterna paciencia ante los errores cometidos. Queremos agradecer a nuestro asesor en este trabajo de investigación, Neicer Campos Vásquez, por sus conocimientos, amistad y experiencias brindadas, por su tiempo invertido en nosotros para que lleguemos con éxito al fin de este arduo camino. A nuestra alma mater la Universidad Privada del Norte sede Lima Centro y a nuestros docentes, en especial a todos los de nuestra carrera Ingeniería de Sistemas Computacionales por su dedicación y enseñanzas, que nos ayudaron a crecer como profesionales en cada clase impartida. Por último, gracias a nuestras familias y amistades, por estar presentes y apoyarnos de manera desinteresada. A todos y cada uno, nuestro agradecimiento eterno.

**Marcos Riera Iziga y
Waldir Sotelo Hernández**

Clasificación de Imágenes Médicas para la Detección del Cáncer de mama mediante Redes Neuronales

Riera-Iziga Marcos, Sotelo-Hernández Waldir, Campos-Vásquez Neicer, Maestro en Ciencias Económicas, Universidad Privada del Norte, Perú, N00132734@upn.pe, N00147775@upn.pe, neicer.campos@upn.edu.pe

Resumen — En la Actualidad no es fácil la interpretación de imágenes de mamografía y brindar un diagnóstico médico por un profesional de salud, Es que a pesar de la experiencia del profesional no se detecta al 100 % las anomalías ó tumores de manera rápida. Puesto a que no es fácil de interpretar. objetivo principal fue realizar un clasificador de imágenes y aplicar una neurona convolucional. posteriormente de un entrenamiento riguroso de la neurona brindar un diagnóstico eficiente, se muestra que mediante el aumento de entrenamiento el resultado es más eficiente y se disminuye el factor de error. Solo en la segunda prueba de llego 64.3% la efectividad. En la investigación no se presentaron imágenes incompatibles con la red neuronal por lo que no se perdió atributos si se realizaron ajustes de parámetros. Por lo que es demostrable que el sistema tiene gran alcance practico debido a que los recursos usados son de fácil alcance.

Palabras Clave— *deep learning, mamografía, clasificación de imágenes, diagnostico.*

Abstract — At present, it is not easy to interpret mammography images and provide a medical diagnosis by a health professional. Despite the professional's experience, anomalies or tumors are not detected 100% quickly. Since it is not easy to interpret. main objective was to perform an image classifier and apply a convolutional neuron. After a rigorous training of the neuron to provide an efficient diagnosis, it is shown that by increasing training the result is more efficient and the error factor is decreased. Only in the second test was 64.3% effective. In the research, no incompatible images with the neural network were presented, so no attributes were lost if parameter adjustments were made. So it is demonstrable that the system has great practical scope because the resources used are easy to reach.

keywords— *deep learning, mammography, image classification, diagnosis.*

I. INTRODUCCIÓN

En la actualidad, los avances para mejorar la calidad de vida humana y optimizar recursos para su preservación vienen logrando objetivos insospechados en corto plazo desde el inicio de investigaciones tecnológicas hasta la puesta en marcha de posteriores herramientas que logran concretamente un cambio sustancial en nuestras vidas o salud. Un claro ejemplo de esto es el Deep Learning que surge como una propuesta de aprendizaje automático artificial pues trata de imitar el mecanismo del cerebro humano imitando las capas o métodos similares de redes neuronales apiladas para el proceso de reconocimiento o ingreso de información significativa [1].

Es por esto en la búsqueda constante de tecnologías emergentes que realizan labores de soporte en importantes campos de estudio como lo es la medicina en el análisis y diagnóstico clínico, se alcanzó a optimizar la que hoy en día es la más emergente de todas las nuevas tecnologías en este campo como es el Deep Learning. Según Tsochatzidis, Costaridou & Pratikakis [2], los métodos de Deep Learning en cuanto a aprendizaje automático han superado la eficiencia de los métodos convencionales de detección y clasificación de objetos, puesto que se requería la extracción de características realizadas a mano por un especialista concentrado en tal labor, en suma los métodos de Deep Learning con cualquiera de sus varios enfoques aprenden de forma adaptativa mediante los datos de entrada y así reduciría el tiempo de discriminación en beneficio de la reproductibilidad de nuevas metodologías.

Entonces teniendo presente el estudio y optimización con el tiempo del Deep Learning en beneficio de la ciencia médica se puede precisar la gran utilidad, para lo que hoy representa una de las enfermedades más peligrosas del mundo, ya que un estudio reciente del año 2018 muestra que 3.1 millones de mujeres en todo el mundo ya habían sido diagnosticadas con cáncer de mama hasta ese año [3] y representa aproximadamente un tercio del total de cánceres diagnosticados en los Estados Unidos, y esto representa el 17% de las muertes por cáncer en general [2]. La atención médica de una paciente con cáncer de seno es costosa y, dado el costo y el valor de la preservación de la salud de las personas, la prevención del cáncer de seno se ha convertido en una prioridad en la salud pública. En los últimos 20 años se han propuesto varias técnicas para este propósito, como la mamografía, que se usa con frecuencia para el diagnóstico del cáncer de seno [4].

De tal manera que la principal causa en el incremento de la tasa de mortalidad por este tipo de cáncer en mujeres es el diagnóstico tardío, pues un paciente con un tumor maligno puede ser debidamente guiado por un especialista para su pronta recuperación. Y esto sumado a la dificultad de acceso, en algunas regiones como Europa del Este, Asia, América latina y África, a mamografías de calidad empeora la tendencia de mortalidad [3].

Los países de Latinoamérica tienen diferencias cruciales en las tasas de incidencia (estandarizadas por edad) de cáncer de mama con la mayor incidencia en Argentina (73.0% en 100,000 cantidad de personas), Uruguay (65.2% en 100,000 personas muestreadas) y Brasil (62.9% en 100,000 personas muestreadas), mientras que la menor incidencia se registra en Guatemala (26.2% en 100,000 personas muestreadas) y finalmente Bolivia (26.5% en 100,000 personas muestreadas) [5]. Mientras que en el Perú se identifica que es el cáncer más frecuente presentado en las mujeres, posicionándolo como la segunda causa mayor de muertes en mujeres adultas con un

resultado de 17039 y un riesgo de morir antes de los 75 años con un porcentaje de 9.3%.

II. PROCESO

A. Materiales

Es de suma importancia la inclusión de materiales en la presente investigación ya que nos brindarán el apoyo necesario para la realización del software, es por ello que estos deben tener concordancia con los objetivos planteados con anterioridad. Esto con el fin de poder realizar un proceso investigativo con resultados óptimos y válidos, razón por la cual se tomará en consideración la siguiente lista de materiales:

- Laptop o PC con recursos necesarios para la fluidez de este sistema.
- Programación en lenguaje Python el cual es compatible con la implementación de la red neuronal que se encargará de clasificar las imágenes médicas.
- Librerías de Python.
- Editor de texto Sublime.
- Base de datos pública con imágenes médicas mamarias.

Como población se cuenta con un total de 322 imágenes médicas dentro de la base de datos mini-MIAS, sin embargo, para la etapa de pruebas se tomarán en cuenta 104 mamografías las cuales se encuentran segmentadas de acuerdo con el siguiente cuadro:

Tabla 1. Resumen de segmentación de imágenes médicas

	ENTRENAMIENTO			
	Training inicial	Primer testeo	Training final	Segundo testeo
Con rastro maligno (cáncer)	22	5	16	9
Sin rastro maligno (cáncer)	22	5	16	9
Total imágenes	104			

Fuente: Elaboración propia

B. Red Neuronal Convolutiva

La implementación de una red neuronal convolutiva de fácil acceso nos garantiza poder realizar las pruebas correspondientes al momento de ingresar las mamografías a ella con el fin de que pueda ser entrenada por un periodo determinado de tiempo. Dicho entrenamiento podrá garantizar una precisión mucho mayor proporcionalmente al tiempo en el que la red haya sido entrenada, es decir, a mayor tiempo de entrenamiento se obtendrán resultados más precisos.

C. Regla de Tres Simple

El uso de esta fórmula matemática será de gran apoyo en la presente investigación y desarrollo del software clasificador de imágenes ya que brindará datos estadísticos confiables y precisos, los cuales facilitarán la interpretación de los resultados obtenidos.

$$(c * b) / a = X \quad (1)$$

En la presente fórmula (1), la letra X representa el resultado de eficiencia, mientras que la letra “b” el total porcentual (100%), la letra “c” cantidad de resultados

obtenidos en el testeo y la letra “a” simboliza el total de imágenes utilizadas.

D. Unidad de Estudio.

La unidad de estudio que se utilizará en la presente investigación consta de 104 imágenes mamográficas recolectadas de la base de datos mini-MIAS, las cuales se encuentran clasificadas de acuerdo con el estado que poseen, es decir, si presentan cáncer de mama o no.

E. Proceso de recolección de datos

En la variable Deep Learning, tendremos que someter la red neuronal convolutiva a una fase previa de entrenamiento con el fin de que pueda distinguir correctamente el estado de las mamografías con las cuales se pondrá a prueba la efectividad del software, en primera instancia se realizarán 2 fases de entrenamiento para que puedan realizarse comparaciones entre un antes y un después.

F. Proceso de análisis de datos

En primera instancia se ingresarán las mamografías dentro del clasificador de imágenes con el fin de que estas sean distinguidas correctamente. El factor decisivo que nos dará la constancia de si el resultado es correcto o no serán los datos de cada mamografía que ya se encuentra dentro de la base de datos mini-MIAS. Si la comparación entre el resultado obtenido entre el software clasificador y la información mostrada en la base de datos es la misma, se puede entonces concluir que el resultado es correcto, en caso de que no se cumpla esta condición el resultado obtenido fue incorrecto. Una vez que se haya realizado la comparación de todas las mamografías analizadas, se elaborará una tabla que mostrará cada uno de los resultados obtenidos y el estado que se muestra en mini-MIAS con el fin de que puedan visualizarse en tiempo real cuales debieron haber sido los resultados obtenidos.

III. DESARROLLO

Se empleó un algoritmo de clasificación de imágenes dentro del lenguaje de programación Python. Asimismo, se empleó el software Anaconda Navigator con el fin de poder instalar las librerías necesarias con las cuales la red neuronal pudiese funcionar de manera óptima. Finalmente habiendo instalado y configurado todas las librerías correspondientes se procedió con la programación del código dentro del editor de texto Sublime. Posteriormente a ello, el progreso del desarrollo fue continuado con el ingreso de 104 imágenes médicas (mamografías), extraídas de la base de datos pública mini-MIAS. Como fue mencionado con anterioridad, se realizó una segmentación en cuanto a las imágenes seleccionadas, con el fin de entrenar la red neuronal convolutiva, de tal manera que esta sea capaz de lograr distinguir entre mamografías que contengan cáncer de mama o no. A continuación, se mostrará el flujo de trabajo del software y los prototipos de las interfaces en las que se podrá visualizar el estado de cada imagen médica.



Figura 1. Flujo de funcionalidad



Figura 2. Prototipo de interfaz

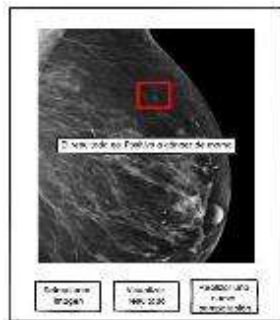


Figura 3. Prototipo de interfaz de resultado

La siguiente imagen muestra líneas de código, en las cuales se inicia el desarrollo del algoritmo principal, siendo este una red neuronal convolucional. Dicha red neuronal es de gran ayuda para el siguiente proyecto ya que es una de las que se enfoca en la clasificación de imágenes mediante un entrenamiento reforzado el cual se realiza mediante el ingreso de mamografías en este caso.

```

def main():
    extract_img('dataset')
    extract_img('dataset/normal', extract_img(0, 0), activation='relu', last_node=(0, 0))
    extract_img('dataset/cancer', extract_img(0, 0), activation='relu')
    extract_img('dataset/normal', kernel_size=(3, 3), activation='relu')
    extract_img('dataset/cancer', kernel_size=(3, 3), activation='relu')
    extract_img('dataset/normal', kernel_size=(3, 3), activation='relu')
    extract_img('dataset/cancer', kernel_size=(3, 3), activation='relu')
    extract_img('dataset/normal', activation='relu')
    extract_img('dataset/cancer', activation='relu')
    extract_img('dataset/normal', activation='sigmoid')
    extract_img('dataset/cancer', activation='sigmoid')
    
```

Figura 4. Algoritmo principal de la red neuronal convolucional

Para implementar el entrenamiento reforzado de la red neuronal convolucional, se tuvieron que crear directorios los cuales almacenan 104 imágenes mamográficas que luego fueron convertidas de formato pgm de origen a jpg para que puedan ser utilizadas y leídas correctamente por las librerías empleadas. Luego de haber realizado ello, se declaran las etapas de entrenamiento y la cantidad de veces que este será repetido con el fin de que la red pueda realizar una distinción mucho mejor.

```

def main():
    # Crear directorios para almacenar las imágenes
    os.makedirs('dataset/normal', exist_ok=True)
    os.makedirs('dataset/cancer', exist_ok=True)
    # Leer imágenes de la base de datos
    images = load_images('dataset/normal')
    images = load_images('dataset/cancer')
    # Entrenar la red neuronal
    train_model(images)
    
```

Figura 5. Algoritmo para el entrenamiento de la red neuronal convolucional

En la imagen presentada a continuación se pueden apreciar los códigos de algunas de las mamografías empleadas en el software, siendo estos: mdb124, mdb125, mdb126 y mdb127. Con el fin de poder realizar las distinciones en la clasificación de imágenes, se tuvieron en cuenta los estados de cada una de las utilizadas, teniendo como estado B las que no presentan cáncer de mama y M las que si lo poseen. Debido a que el clasificador de imágenes solamente está enfocado a la detección del cáncer de mama, se obviaron los datos mostrados en la segunda y tercera columna, los cuales indican el estado del tejido y el tipo de la anomalía.

mdb124	G	ARCH	M
mdb125	D	ARCH	M
mdb126	D	ARCH	B
mdb127	G	ARCH	B

Figura 6. Detalle para la clasificación de imágenes

IV. RESULTADOS

Con el fin de lograr resultados con mayor precisión, se realizaron 2 fases de entrenamiento; en la primera se incluyeron 54 imágenes médicas y contó con 5 épocas de training correspondientes a la red neuronal. Dichas mamografías fueron agrupadas de acuerdo con el siguiente cuadro:

	TRAINING INICIAL	PRIMER TESTEO
Con rastros malignos (con cáncer)	22	5
Sin rastros malignos (sin cáncer)	22	5
Total imágenes	54	

Tabla 2. Segmentación de imágenes médicas
Fuente: Elaboración propia

Según lo descrito con anterioridad, las imágenes mamográficas fueron divididas entre las que fueron empleadas para el entrenamiento de la red neuronal y las que se utilizaron para las pruebas realizadas (testeo). Dichas imágenes se introdujeron manualmente dentro del código de programación mediante el editor de texto Sublime guardado como códigos con extensión Python, por lo que los resultados se fueron obteniendo uno por uno. En la siguiente tabla se pueden apreciar los códigos que manejan las imágenes dentro de la base de datos mini-MIAS, el estado en el que se encuentran y el resultado que fue obtenido por cada una tras haber sido utilizada por el software.

Tabla 3. Relación de resultados obtenidos

MAMOGRAFÍAS		
ID mini-MIAS	ESTADO	RESULTADO
57	SIN CANCER	CORRECTO
60	SIN CANCER	CORRECTO

64	SIN CANCER	CORRECTO
68	SIN CANCER	CORRECTO
70	SIN CANCER	INCORRECTO
238	CON CANCER	CORRECTO
256	CON CANCER	INCORRECTO
270	CON CANCER	INCORRECTO
271	CON CANCER	CORRECTO
274	CON CANCER	INCORRECTO
Total	10	

En esta tabla se muestran los resultados obtenidos tras haber ingresado las mamografías dentro del software (Fuente: Elaboración propia).

Como puede apreciarse en la tabla mostrada con anterioridad, de un total de 10 imágenes mamográficas empleadas, se obtuvieron 6 imágenes con resultado esperado, mientras que otras 4 imágenes mamográficas no fueron reconocidas correctamente, todo esto luego de las 5 épocas de entrenamiento de la red neuronal, y tras haber probado su eficiencia con las mamografías extraídas de la base de datos mini-MIAS (la cual sirve como guía para poder identificar la validez de los resultados obtenidos). Una vez realizado el proceso de clasificación de imágenes, se procedió con la utilización de la fórmula (2) para una regla de tres simple con la cual se pueden apreciar los valores empleados para obtener el porcentaje de eficiencia obtenido, resultando esto en un 60% tras haber realizado un entrenamiento previo a la red neuronal por un periodo de una hora y con 44 imágenes:

$$(6 * 100) / 10 = X \quad (2)$$

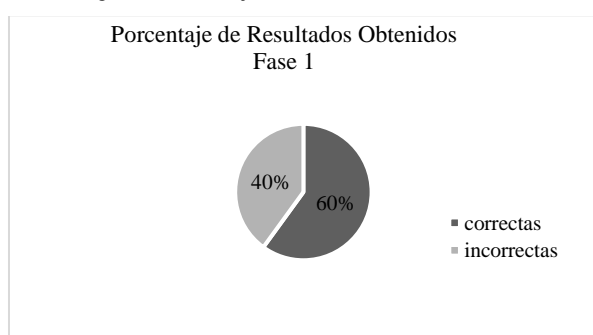
$$X = 60$$

Finalmente, el otro 40% restante perteneciente al porcentaje de error obtenido tras haber realizado las pruebas con el software, se obtuvo empleando la misma fórmula (3) reemplazando las variables con los siguientes valores:

$$(4 * 100) / 10 = X \quad (3)$$

$$X = 40$$

Figura 7. Porcentaje de Resultados Obtenidos Fase 1



Fuente: Elaboración propia.

Para la siguiente fase de entrenamiento se incrementó a 30 épocas del software se ingresa 10 imágenes con rastro oncológico y 16 imágenes mamarias libres de padecimiento como se observa en la figura 3, siendo usadas de lo que resta del total del banco de imágenes para el testeo o prueba de esta segunda fase de entrenamiento, es decir 9 imágenes sin rastro

de cáncer y 9 imágenes con sospecha de padecimiento oncológico.

Tabla 4. Segmentación de imágenes para el segundo entrenamiento

	ENTRENAMIENTO (30 Épocas)	
	Training final	Segundo testeo
Con rastro maligno (con cáncer)	16	9
Sin rastro maligno (sin cáncer)	16	9
Total de imágenes	50	

Fuente: Elaboración propia.

Tras haber realizado el entrenamiento de la red neuronal por un periodo de 6 horas (30 épocas), se obtuvieron los siguientes resultados plasmados en la tabla a continuación:

Tabla 5. Resultados obtenidos etapa 2

MAMOGRAFÍAS		
ID mini-MIAS	ESTADO	RESULTADO
54	SIN CANCER	CORRECTO
55	SIN CANCER	CORRECTO
56	SIN CANCER	INCORRECTO
57	SIN CANCER	CORRECTO
60	SIN CANCER	CORRECTO
61	SIN CANCER	CORRECTO
62	SIN CANCER	INCORRECTO
64	SIN CANCER	INCORRECTO
65	SIN CANCER	CORRECTO
66	SIN CANCER	INCORRECTO
67	SIN CANCER	INCORRECTO
68	SIN CANCER	CORRECTO
70	SIN CANCER	CORRECTO
71	SIN CANCER	CORRECTO
181	CON CANCER	CORRECTO
238	CON CANCER	CORRECTO
239	CON CANCER	CORRECTO
241	CON CANCER	CORRECTO
245	CON CANCER	CORRECTO
249	CON CANCER	CORRECTO
253	CON CANCER	CORRECTO
256	CON CANCER	CORRECTO
264	CON CANCER	CORRECTO
265	CON CANCER	INCORRECTO
267	CON CANCER	INCORRECTO
270	CON CANCER	INCORRECTO
271	CON CANCER	INCORRECTO
274	CON CANCER	INCORRECTO
Total	28	

En esta tabla se muestran los resultados obtenidos en la segunda fase (30 épocas) tras haber ingresado las mamografías dentro del software (Fuente: Elaboración propia).

Según los resultados obtenidos tras haber entrenado por un periodo de tiempo mucho más largo a la red neuronal, se puede identificar que de las 28 imágenes empleadas para las pruebas 18 fueron clasificadas correctamente, mientras que 10 fueron identificadas erróneamente. Por lo tanto, se aplicará la fórmula para una regla de tres simple nuevamente para poder identificar tanto el porcentaje de eficiencia como el de

error que se lograron obtener en esta segunda fase de pruebas (4).

Tras haber entrenado la red neuronal con 76 imágenes y siendo puesta a prueba por 28, se obtuvo un 64.3%, aumentando en un 4.3% a comparación de la primera prueba realizada.

$$(18 * 100) / 28 = X \quad (4)$$

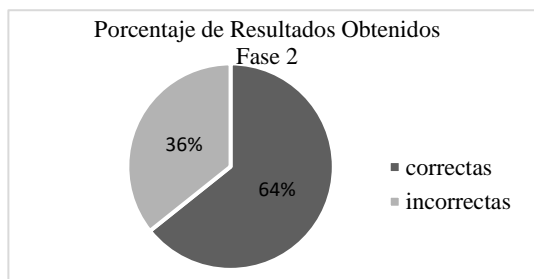
$$X = 64.3$$

Finalmente, se pudo reducir el porcentaje de error obteniendo un 35.7% en esta ocasión (5).

$$(10 * 100) / 28 = X \quad (5)$$

$$X = 35.7$$

Figura 8. Porcentaje de Resultados Obtenidos Fase 2.



Fuente: Elaboración propia.

Para finalizar la evidencia de resultados, se presenta el detalle de la ejecución del software en tiempo real vía el programa compilador Anaconda Navigator, mostrando mediante la interfaz diseñada, un resultado positivo a cáncer y uno negativo al padecimiento oncológico.



Figura 9. Interfaz con clasificación de mamografía libre de cáncer.



Figura 10. Interfaz con clasificación de mamografía con cáncer.

V. DISCUSIÓN

Se logró identificar que la capacidad de recursos computacionales con los cuales se dispuso para la presente investigación, no fueron lo suficientemente óptimos para poder realizar un entrenamiento largo ya que este necesita una gran cantidad de procesamiento de datos con el fin de fortalecer y optimizar el entrenamiento de la red neuronal, por lo que fácilmente podrán lograrse obtener mejores resultados si el entrenamiento se realiza con equipos mucho más “potentes y sofisticados”.

Tomando investigaciones previamente referenciadas que enriquecen nuestra investigación, podemos mencionar que, en base a los resultados obtenidos en sus respectivos estudios, estos podrán ser fácilmente contrastados. Como Tsochatzidis, Costaridou & Pratikakis [2] indican en su artículo “Deep Learning for Breast Cancer Diagnosis from Mammograms – A Comparative Study” en Grecia, emplearon un total de 10239 imágenes mamarias entre malignas y benignas extraídas de la base de datos privada DDSM con variedad de redes neuronales como AlexNet y GoogleNet para comparar el desempeño de cada una de ellas, sin embargo, esto conllevó a largos procesos de aprendizaje importantes y llegaron a la conclusión de que las imágenes empleadas no son de utilidad para todas las redes neuronales con la misma eficiencia y que aún tenían un largo camino por recorrer si se querían eliminar las complicaciones en cuanto a los contrastes y sobreajuste de los mamogramas, además de contar con un presupuesto de investigación a gran escala.

A diferencia de ello podemos destacar que dentro de la presente investigación no se tuvo como factor desfavorable imágenes médicas incompatibles con la red neuronal convolucional, es decir, no existieron pérdidas de atributos ni se tuvieron que reajustar parámetros con el fin de obtener resultados más precisos, adicionalmente a ello, tampoco se tuvo que contar con exorbitantes presupuestos para la realización, entrenamiento y distinción de mamografías con la red neuronal.

En base a lo explicado con anterioridad podemos destacar que el software para clasificación de imágenes mamográficas tiene un gran alcance práctico, ya que los recursos sobre los cuales fue programado son de fácil acceso por lo que puede ser replicado minimizando costos dentro de las entidades médicas que contemplan hacer uso de él. Sin embargo, se

debe contar con una cantidad de procesamiento mucho mayor para que la red neuronal pueda entrenarse de manera mucho más prolongada con el fin de que los resultados sean más precisos y la distinción de mamografías entre las que poseen cáncer de mama o no, se realice de manera correcta; se cuenta también con una flexibilidad sustancial dentro de la fase de entrenamiento, ya que la red neuronal puede ser fácilmente entrenada debido a su programación, en la cual pueden insertarse diversos conjuntos de datos que contengan mamografías (en los formatos de imagen más apropiados para una óptima visualización) o incluso puede realizarse el entrenamiento con imágenes médicas que posean los propios hospitales.

VI. CONCLUSIONES

Para concluir con toda la investigación realizada, se logró aplicar el conocimiento obtenido con el cual fue factible realizar el software para la clasificación de imágenes, todo ello gracias a la recolección de información e investigaciones previas referentes al campo de las redes neuronales y su aplicación en beneficio de la calidad de vida humana.

De manera complementaria, se procedió con la identificación de los recursos necesarios para la correcta realización del entrenamiento al cual fue finalmente sometida la red neuronal. Dicho esto, fueron escogidas ciertas imágenes médicas mamarias extraídas de la base de datos pública mini-MIAS, con las cuales se ejecutaron diversas fases de entrenamiento en las que se logró detectar un incremento de la eficiencia y precisión de los resultados de manera progresiva.

Para lograr poner en tela de juicio la efectividad y precisión de la red neuronal, esta fue sometida a una etapa de pruebas en las que fueron ingresadas mamografías como objetos de comparación con el fin de que se verifique que su estado sea clasificado correctamente, con el fin de lograr un avance progresivo en su eficiencia y cuando los resultados de la clasificación no fueran óptimos, esto sirve como indicador para saber cuándo fortalecer la red neuronal, ya que se prolongaba el tiempo de entrenamiento en busca de superar los porcentajes anteriormente obtenidos.

Finalmente se comprobó que la realización de un software de bajo costo que pueda servir de apoyo al área oncológica del sector salud de la ciudad de Lima es factible y replicable, debido a que la implementación y posterior etapa de entrenamiento pudieron ser realizadas con recursos de fácil acceso e insumos primarios recolectados a través de un banco de imágenes público, en el cual se encontraban alojadas las mamografías necesarias para la clasificación. Desde este punto, se sugiere como oportunidad de mejora para futuras repeticiones o etapas de desarrollo, maximizar el enfoque de la investigación hacia otros ámbitos para que puedan clasificarse distintos tipos de imágenes médicas (tomografías, rayos x).

REFERENCIAS

[1] Xiaofan, L., Dong, F., Zhang, S., & Weibin, G. (2019). "A Survey on Deep Learning Techniques in Wireless Signal Recognition". *Wireless Communications and Mobile Computing* 2019, 5629572. Recuperado de: <https://www.hindawi.com/journals/wcmc/2019/5629572/>

[2] Tsochatzidis, L.; Costaridou, L.; Pratikakis, I. (2019) "Deep Learning for Breast Cancer Diagnosis from Mammograms—A Comparative Study". *J. Imaging*, 5, 37. Recuperado de: <https://www.mdpi.com/2313-433X/5/3/37/html>

[3] Jiménez Garay, G. A. (2019). "Deep Learning for Semantic Segmentation versus Classification in Computational Pathology: Application to mitosis analysis in Breast Cancer grading". Recuperado de: <http://tesis.pucp.edu.pe/repositorio/handle/20.500.12404/13969>

[4] Mambou, S.J.; Maresova, P.; Krejcar, O.; Selamat, A.; Kuca, K. (2018). "Breast Cancer Detection Using Infrared Thermal Imaging and a Deep Learning Model". *Sensors*, 18, 2799. Recuperado de: <https://www.mdpi.com/1424-8220/18/9/2799>

[5] Ferlay J, Soerjomataram I, Dikshit R, et al. (2018). "Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods". *Int J Cancer* 2015;136: E359–86. Recuperado de: <https://onlinelibrary.wiley.com/doi/full/10.1002/ijc.31937>

[6] Abdullah, K. A., & Reed, W. (2018). "3D printing in medical imaging and healthcare services". *Journal of Medical Radiation Sciences*, 65(3), 237-239. Recuperado de: <https://search.proquest.com/docview/2290547549/C6CD91BBE30D4285PQ/4?accountid=36937>

[7] Bugada, J (1956). "Los instrumentos de la investigación en las ciencias sociales". *Revista de estudios políticos*, pp. 137–162, 1956. Recuperado de: <https://dialnet.unirioja.es/servlet/articulo?codigo=2128996>

[8] Cauas, D. (2015). Definición de las variables, enfoque y tipo de investigación. Bogotá: biblioteca electrónica de la universidad Nacional de Colombia, 2, 1-11.

[9] Cheikh, B. (2017). "Morphologie mathématique sur les graphes pour la caractérisation de l'organisation spatiale des structures histologiques dans les images haut-contenu : application aumicroenvironnement tumoral dans le cancer du sein", Ph.D. thesis, Université Pierre et Marie Curie, 2017.

[10] Couture, H. D., Williams, L. A., Gerads, J., Nyante, et al. (2018). "Image analysis with deep learning to predict breast cancer grade, ER status, histologic subtype, and intrinsic subtype". *NPJ Breast Cancer*, 4, 1-8. Recuperado de: <https://search.proquest.com/docview/2099032482/483F4A4C23E142CEPQ/2?accountid=36937>

[11] Duggento, A.; Aiello, M; Cavaliere, C; Cascella, G. et al. (2019). "An Ad Hoc Random Initialization Deep Neural Network Architecture for Discriminating Malignant Breast Cancer Lesions in Mammographic Images". *Contrast Media & Molecular Imaging*, 2019, 5982834. Recuperado de: <https://www.hindawi.com/journals/cmim/2019/5982834/#abstract>

[12] Fitch, H., & Friedrich, K. (2018). "Digital matters: Processes of normalization in medical imaging". *Catalyst: Feminism, Theory, Technoscience*, 4(2). Recuperado de: <https://search.proquest.com/docview/2133459265/C6CD91BBE30D4285PQ/10?accountid=36937>

[13] Freudenberg, M; Nölke, N; Agostini, A; Urban, K; Wörgötter, F & Kleinn, C. (2019). "Large Scale Palm Tree Detection in High Resolution Satellite Images Using U-Net". *Remote Sens*. 2019, 11(3), 312. Recuperado de: <https://www.mdpi.com/2072-4292/11/3/312/html>

[14] Han, Z., Wei, B., Zheng, Y., Yin, Y., Li, K., & Li, S. (2017). "Breast cancer multi-classification from histopathological images with structured deep learning model". *Scientific Reports (Nature Publisher Group)*, 7, 1-10. Recuperado de: <https://search.proquest.com/docview/1955970724/E80D174FEB5D47D0PQ/1?accountid=36937>

[15] Hopwood, P., Fletcher, I., Lee, A. y Al Ghazal, S. (2001). A body image scale for use with cancer patients. *European Journal of Cancer*, 37, 189-197.

[16] Huang, B.; Lu, K.; Audebert, N.; Khalel, A.; Tarabalka, Y.; Malof, J.; Boulch, A.; Le Saux, B.; Collins, L.; Bradbury, K.; et al. (2018). "Large-scale semantic classification: Outcome of the first year of Inria aerial image labeling benchmark". In *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium—IGARSS 2018*, Valencia, Spain, 22–27 July 2018. Recuperado de: <https://ieeexplore.ieee.org/abstract/document/8518525/>

[17] Irshad, H; Jalali, S; Roux, L; Racoceanu, D; Naour, G.; Hwee, L & Capron, F., (2013). "Automated mitosis detection using texture, SIFT features and HMAX biologically inspired approach," *Journal of Pathology Informatics*, vol. 4, no. 2, pp. 12, 2013.

[18] J Suckling et al (1994) "The Mammographic Image Analysis Society Digital Mammogram Database" *Exerpta Medica. International Congress Series* 1069 pp375-378. Recuperado de: <http://peipa.essex.ac.uk/pix/mias/Info.txt>

[19] Kaur, P.; Singh, G.; Kaur, P., (2019). "Intellectual detection and validation of automated mammogram breast cancer images by multi-

- class SVM using deep learning classification". *Informatics in Medicine Unlocked* Volume 16, 2019, 100239. Recuperado de: <https://www.sciencedirect.com/science/article/pii/S2352914819302448>
- [20] LeCun Y, Bengio Y, Hinton G. (2015). "Deep learning". *Nature*. 521: 436–444. PMID:26017442. Recuperado de: <https://www.nature.com/articles/nature14539>
- [21] Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; Van Der Laak, J.A.; Van Ginneken, B.; Sánchez, C.I. (2017). "A survey on deep learning in medical image analysis". *Med. Image Anal.* 2017, 42, 60-88. Recuperado de: <https://www.sciencedirect.com/science/article/abs/pii/S1361841517301135>
- [22] Muhammad, N. B., Muta, K., Malik, M. I., Shoaib, A. S., et al. (2020). "Computer-aided diagnosis of skin diseases using deep neural networks". *Applied Sciences*, 10(7), 2488. Recuperado de: <https://search.proquest.com/docview/2387868778/583313C50B4B4325PQ/12?accountid=36937>
- [23] Rosenberg, M. (1965). *Society and adolescent self-image*. Princeton, NJ: Princeton University Press.
- [24] Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. (2015). "Imagenet large scale visual recognition challenge". *Int. J. Comput. Vis.* 2015, 115, 211-252. Recuperado de: <https://link.springer.com/article/10.1007%2Fs11263-015-0816-y#citeas>
- [25] Saha, M.; Chakraborty, C & Racoceanu D., (2018) "Efficient deep learning model for mitosis detection using breast histopathology images", *Computerized Medical Imaging and Graphics*, vol. 64, pp. 29–40, 2018.
- [26] Shen, L., Margolies, L. R., Rothstein, J. H., Fluder, E., McBride, R., & Sieh, W. (2019). "Deep learning to improve breast cancer detection on screening mammography". *Scientific Reports (Nature Publisher Group)*, 9, 1-12. Recuperado de: <https://search.proquest.com/docview/2282431384/35A7E2AD3B944D01PQ/1?accountid=36937>
- [27] Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. (2015). "Going deeper with convolutions". In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA, 7–12 June 2015; pp. 1–9. Recuperado de: <https://ieeexplore.ieee.org/document/7298594>
- [28] Teare, P.; Fishman, M.; Benzaquen, O.; Toledano, E.; Elnekave, E. (2017). "Malignancy detection on mammography using dual deep convolutional neural networks and genetically discovered false color input enhancement". *J. Dig. Imaging* 2017, 30, 499-505. Recuperado de: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5537100/>
- [29] Wang, J., Yang, X., Cai, H., Tan, W., Jin, C., & Li, L. (2016). "Discrimination of breast cancer with microcalcifications on mammography by deep Learning". *Scientific Reports (Nature Publisher Group)*, 6, 27327. Recuperado de: <https://search.proquest.com/docview/1794561958/2FA7FD6B2D854C4APQ/2?accountid=36937>
- [30] Wen, L.T.; Tanaka, A.; Nonoyama, M. (1988). "Identification of Marek's disease virus nuclear antigen in latently infected lymphoblastoid cells". *J. Virol.*, 62, 3764-3771. Recuperado de: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC253520/>
- [31] Zheng, X., Yao, Z., Huang, Y. et al. (2020). "Deep learning radiomics can predict axillary lymph node status in early-stage breast cancer". *Nat Commun* 11, 1236. Recuperado de: <https://www.nature.com/articles/s41467-020-15027-z>
- [32] Zigmond, A. y Snaith, R. (1983). *The Hospital Depression and Anxiety Scale*. *Acta Psychiatrica Scandinavica*, 67, 361-370