



**UNIVERSIDAD NACIONAL AUTÓNOMA DE
MÉXICO**

Programa de Maestría y Doctorado en Ingeniería
Ingeniería Eléctrica - Procesamiento Digital de Señales

**ANÁLISIS DE IMÁGENES DE
ULTRASONIDO PARA DETECCIÓN
DE ENFERMEDADES PULMONARES**

TESIS

Para optar por el grado de:
MAESTRA EN INGENIERÍA

Presenta

Melisa Mateu Villa

Tutora

Dra. Jimena Olveres Montiel
Facultad de Ingeniería

Ciudad de México, 2022

JURADO ASIGNADO:

Presidente: DRA. LUCÍA MEDINA GÓMEZ

Secretario: DR. FERNANDO ARÁMBULA COSÍO

1er. Vocal: DRA. JIMENA OLVERES MONTIEL

2do. Vocal: DR. BORIS ESCALANTE RAMÍREZ

3er. Vocal: DR. FABIÁN TORRES ROBLES

Lugar donde se realizó la tesis: FACULTAD DE INGENIERÍA, CU

TUTOR DE TESIS:
DRA. JIMENA OLVERES MONTIEL

FIRMA

*Para todas las personas con enfermedades mentales, como recordatorio de
que se puede aunque nuestra mente diga lo contrario.*

AGRADECIMIENTOS

Este trabajo de tesis recibió apoyo de los proyectos UNAM PAPIIT IV100420 y TA101121 y de una beca de nivel maestría del Consejo Nacional de Ciencia y Tecnología.

Quiero comenzar por agradecer a mi asesora, la Dra. Jimena por todo el apoyo que me brindó durante mis dos años de posgrado, por confiar en mi y motivarme. De igual manera al Dr. Boris por su ayuda y apoyo. A todos los integrantes de LaPI, en especial a Toño por ser mi compañero estos años.

A mi hermana, Chawi, por todo el amor que siempre me otorgas y a mis padres por su apoyo.

A mi familia de CDMX: Armando, Fernanda Macal, Julio y Fernanda por ser un apoyo incondicional en todo este proceso, por su cariño y sus palabras.

To Eric, for your love and support. For being a loving raccoon at any time.

A Lulú, Rafael, Army, Josué, Nao y Saúl porque a la distancia no dejaron de apoyarme y quererme. A Gaby, por todas tus palabras y acompañamiento.

A mi familia, a mis compañeros y cualquier persona involucrada en este proyecto, gracias...

RESUMEN

La enfermedad del Coronavirus 2019 (COVID-19) afecta principalmente a los pulmones, es por ello que monitorearlos de forma constante es una manera efectiva de observar la evolución de esta enfermedad. Siendo la tomografía computarizada (TC) la técnica de imagenología más comúnmente empleada para observar y seguir los cambios en los pulmones de los pacientes con COVID-19.

Debido a los costos y efectos sobre la salud que tiene la TC, se han comenzado a explorar otras; como el ultrasonido pulmonar o LUS por sus siglas en inglés (*Lung Ultrasound*). Siendo el LUS una buena alternativa, ya que no expone al paciente a radiación ionizante y minimiza el riesgo de contaminación.

Una de las principales estructuras que se puede observar con el LUS es la pleura, que al ser una membrana que rodea a los pulmones sufre de afectaciones con las distintas enfermedades que estos órganos puedan presentar. Se ha observado que en el LUS, estas anomalías de la pleura se presentan como artefactos en la imagen.

Sin embargo, aunque la técnica de LUS pueda detectar estas enfermedades regularmente se requiere de un experto clínico para su interpretación; lo cual tiende a incrementar tiempos de diagnóstico o bien disminuir el desempeño de este.

Debido a ello en este proyecto se describen y comparan, varios métodos aprendizaje máquina para clasificación de distintas imágenes de LUS con COVID-19, neumonía y pacientes sanos.

Los clasificadores usados son: Naive Bayes (NB), Máquina de Soporte Vectorial (SVM), K-Vecinos Más Cercanos (KNN) y Bosques Aleatorios (RF). Estos usan para clasificar características de las imágenes previamente ex-

traídas mediante el uso de la matriz de co-ocurrencia de nivel de gris (GLCM) y de los estadísticos de su histograma correspondiente. Además, este trabajo compara los métodos clásicos antes descritos con dos diferentes arquitecturas de redes neuronales convolucionales (CNN): DenseNet 201 y VGG-19; las cuales clasifican las mismas imágenes de ultrasonido de las enfermedades pulmonares antes mencionadas.

ABSTRACT

Since lungs are the most affected organs with Coronavirus Disease 2019 (COVID-19), monitoring them constantly is an effective way to observe sickness evolution.

The most common technique for lung-imaging and evaluation is Computed Tomography (CT). However, its costs and effects over human health has made Lung Ultrasound (LUS) a good alternative. LUS does not expose the patient to ionizing radiation and minimizes the risk of contamination.

Also, there is evidence of a relation between different artifacts on LUS and lung's diseases coming from the pleura, whose abnormalities are related with most acute respiratory disorders. However, LUS often requires an expert clinical interpretation that may increase diagnosis time or decrease diagnosis performance.

This work describes and compares machine learning classification methods namely Naive Bayes (NB), Support Vector Machine (SVM), K-Nearest Neighbor (K-NN) and Random Forest (RF) over several LUS images. They obtain a classification between lung images with COVID-19, pneumonia, and healthy patients, using image's features previously extracted from Gray Level Co-Occurrence Matrix (GLCM) and histogram's statistics.

Furthermore, this work compares the above classic methods with two different Convolutional Neural Networks (CNN) architectures namely DenseNet 201 and VGG-19. These classify the ultrasound images in order to identify these lung's diseases.

ÍNDICE GENERAL

Agradecimientos	III
Resumen	I
Abstract	III
1. Introducción	1
1.1. Motivación	2
1.2. Objetivo	2
1.3. Antecedentes en la clasificación de imágenes médicas con COVID-19	2
1.4. Estructura de la tesis	5
2. Conceptos generales	6
2.1. Ultrasonido	6
2.2. Ultrasonido pulmonar	7
2.2.1. Artefactos	8
2.3. Extracción de características	11
2.3.1. Histograma	11
2.3.2. Matriz de co-ocurrencia de niveles de gris (GLCM)	11
2.3.3. Selección de características	13
2.4. Métodos de Clasificación	15
2.4.1. Aprendizaje máquina supervisado	15

2.4.2. Aprendizaje profundo	17
2.5. Métricas	19
3. Metodología para la clasificación de imágenes de LUS	20
3.1. Conjunto de datos	20
3.2. Comparativa de métodos de aprendizaje máquina	21
3.2.1. Extracción y selección de características	22
3.2.2. Entrenamiento de los métodos de aprendizaje máquina	22
3.3. Comparativa de arquitecturas de aprendizaje profundo	23
3.4. Comparativa entre los métodos de clasificación seleccionados .	24
4. Resultados	25
4.1. Evaluación y comparativa de los métodos de aprendizaje máqui- na	25
4.2. Comparativa entre el desempeño de las redes neuronales con- volucionales	31
4.3. Evaluación de los métodos de aprendizaje profundo seleccionados	32
4.4. Comparativa entre los métodos de clasificación de aprendizaje máquina y aprendizaje profundo seleccionados	35
5. Conclusiones	36
A. Artículo	41
B. Curvas de aprendizaje profundo	51

CAPÍTULO

1

INTRODUCCIÓN

El síndrome respiratorio agudo severo coronavirus 2 (SARS-Cov-2), también conocido como enfermedad del Coronavirus 2019 (COVID-19) ha causado una pandemia mundial en los últimos años. Debido a que los pulmones son los órganos más afectados por esta enfermedad, monitorearlos de forma constante es una manera efectiva de observar la evolución del COVID-19 en ellos.

La técnica radiológica principalmente empleada para el diagnóstico de los pacientes infectados con COVID-19 son las imágenes pulmonares de tomografía computarizada (TC) y rayos X, pero debido a que estas técnicas significan una exposición constante a radiación ionizante se han comenzado a explorar técnicas menos invasivas para el paciente, como lo es el ultrasonido pulmonar o LUS por sus siglas en inglés (*Lung Ultrasound*), el cual además de ser una buena herramienta en el monitoreo de los pacientes, ha mostrado una buena correlación con la TC.

Dentro de los estudios con la técnica LUS una de las principales estructuras que se puede observar es la pleura, que al ser una membrana que rodea a los pulmones sufre de afectaciones con las distintas enfermedades que estos órganos presentan. Se ha observado que estas anomalías en la pleura se

muestran como artefactos en la imagen del ultrasonido pulmonar.

Conociendo que existe una relación entre las enfermedades pulmonares y la pleura, y la relación de esta membrana con los artefactos en el LUS se pueden proporcionar indicios de diversas enfermedades pulmonares, observando los artefactos en esta técnica de imagenología. Pudiendo entonces, estas imágenes de LUS emplearse para identificar o clasificar algunos de los padecimientos presentados en los pulmones.

1.1. Motivación

El área de procesamiento digital de señales puede aportar un apoyo para el diagnóstico médico, facilitando tareas que ahorren tiempo al personal de salud, o bien, generando algoritmos que permitan procesar con mayor eficiencia una gran cantidad de información. En el caso específico de este proyecto, ayudando a la clasificación de imágenes de ultrasonido pulmonar con COVID-19, neumonía y pacientes sin afecciones.

1.2. Objetivo

Siendo el objetivo de este trabajo la comparación de distintos métodos de aprendizaje máquina y aprendizaje profundo en la tarea de clasificación de tres clases de imágenes de ultrasonido pulmonar, siendo estas clases: imágenes de pacientes con COVID-19, imágenes de pacientes con neumonía e imágenes de pacientes sin afecciones.

1.3. Antecedentes en la clasificación de imágenes médicas con COVID-19

El problema de la clasificación de imágenes médicas entre diversas enfermedades es recurrente debido a la cantidad de información que se tiene y la utilidad que la automatización aporta en diagnóstico.

Como se mencionó anteriormente, las técnicas de imagenología principalmente empleadas en el diagnóstico y evolución del COVID-19 han sido la tomografía computarizada y los rayos X. Debido a esto, existen muchas más bases de datos con estas imágenes en comparación a las del ultrasonido pulmonar; siendo tanto las imágenes de rayos X y TC las principales empleadas

para trabajar en un problema de clasificación con esta enfermedad.

Dentro de los primeros trabajos que existen al respecto se encuentra el de Ozturk(2020), [1] que empleando imágenes de rayos X realiza tanto una clasificación binaria entre pacientes con y sin COVID-19, obteniendo una exactitud de 98,08 %; además de realizar una clasificación de tres clases entre pacientes con COVID-19, neumonía y sanos obteniendo una exactitud de 87,02 %; ambas clasificaciones entrenadas con la red neuronal convolucional (CNN), DarkNet.

En el mismo año Xu(2020) [2] realizó una clasificación de las mismas tres clases que Ozturk, con la diferencia de que las imágenes estudiadas eran de TC. Xu realizó un preprocesamiento en el que obtuvo la región de interés (ROI) pulmonar, extrajo características de esta área mediante el uso de una ResNet 18 y después realizó la clasificación de las tres clases, obteniendo una exactitud del 86,7 %.

Apostolopoulos(2020) [3] realiza los mismos tipos de clasificaciones que Ozturk, binaria y tres clases, usando también imágenes de rayos X pulmonar. Apostolopoulos compara distintas arquitecturas de CNN, en específico: VGG-19, MobileNet V2, Inception, Xception, ResNet V2. La comparativa de estas redes tenía la finalidad de encontrar las de mejor desempeño para realizar las clasificaciones, siendo estas la VGG-19 y MobileNet V2.

Para el siguiente año se contaba con más imágenes para poder realizar entrenamientos de clasificación, Halder(2021) [4] trabaja con imágenes de TC y CNNs en una tarea de clasificación binaria para COVID-19. Halder realiza una comparativa entre las imágenes con y sin técnicas de aumentado y a su vez una comparación entre las arquitecturas: DenseNet 201, VGG-16, ResNet 50 y MobileNet; teniendo como resultado que la red DenseNet 201 es la de mejor desempeño, con exactitudes de 1.00 y 0.99 para las imágenes originales y aquellas con técnicas de aumentado, respectivamente.

Kumari(2021) [5] de igual manera realiza, una comparativa entre arquitecturas; en específico entre ResNet 50, VGG-16, Inception V3 y Xception para lograr una clasificación binaria en una base de imágenes de rayos X y TC. Obteniendo exactitudes de 94 %, 98 %, 97 % y 97 % respectivamente para las arquitecturas mencionadas.

Narin(2021) [6] hace otra comparativa entre cinco arquitecturas: ResNet 50, ResNet 101, ResNet 152, Inception V3 e Inception-ResNet V2 para tres problemas binarios de cuatro clases: COVID-19, paciente sano, neumonía viral y neumonía bacteriana. Siendo la ResNet 50 la de mejor desempeño con exactitudes de 96,10 %, 99,5 % y 99,70 % para las clases COVID-19/Paciente sano, COVID-19/Neumonía viral y COVID-19/neumonía bacteriana.

Bohmrah(2021) [7] de igual forma realiza una comparativa de CNN con la diferencia de que evalúa la misma arquitectura: una DenseNet, pero con distinto número de capas, 121, 169 y 201. Siendo la DenseNet 201 la que obtuvo la exactitud más alta, 95,2 %, en la clasificación de imágenes de rayos X pulmonar para las tres clases: COVID-19, neumonía y pacientes sin afecciones.

Además de las comparativas de CNN, ensamblar varias arquitecturas o fusionar sus resultados también se presentan como soluciones para las tareas de clasificación de imágenes con COVID-19; Kundu(2021) [8] emplea los resultados fusionados de cuatro arquitecturas pre-entrenadas; VGG-11, GoogleNet, SqueezeNet y ResNet 50, para realizar la clasificación binaria con imágenes de TC, siendo 98,93 % su exactitud resultante.

Mamalakis(2021) [9] usa un pipeline denominado DenResCov-19 compuesto de una DenseNet 121 y una ResNet 50, para una evaluación binaria entre COVID-19 y neumonía, así como una clasificación de tres clases; COVID-19, neumonía y pacientes sanos, además de una tercera clasificación de cuatro clases que incluye las tres anteriores además de imágenes de rayos X con tuberculosis. Las exactitudes que obtuvo fueron 99,60 %, 96,51 % y 91,77 % respectivamente para cada clasificación.

Durante el último año, Kong(2022) [10] y Soni(2022) [11] realizaron también entrenamientos para tareas de clasificación. Kong sobre imágenes de rayos X pulmonar utilizando tres arquitecturas: ResNet 34, para segmentar la ROI, DenseNet 201 y VGG-16 para extraer características de las imágenes originales y de las segmentadas, respectivamente. Estas características obtenidas fueron fusionadas para realizar una clasificación binaria y tres clases; COVID-19, neumonía y pacientes sanos, obteniendo una exactitud del 98 % y 97,3 % respectivamente.

Soni empleó imágenes de TC pulmonar con preprocesamiento en los píxeles más brillantes, normalizó los valores y empleó filtros de mediana. Usando una red neuronal generativa (GAN) generó más imágenes de TC, después las segmentó y extrajo sus características con ayuda de una U-Net; y por último con un perceptrón multicapa (MLP) realizó la tarea de clasificación binaria entre casos de pacientes con y sin COVID-19, obteniendo una exactitud de 97,1 %.

Otra forma en la que se comenzó a abordar el problema de clasificación de imágenes médicas de pacientes con COVID-19, fue mediante la extracción de características empleando la matriz de coocurrencia de nivel de gris (GLCM).

Pi(2021) [12] utiliza estas matrices para realizar una clasificación bina-

ria con aprendizaje máquina externo en imágenes de TC preprocesadas mediante su histograma para mejorar su contraste, Pi obtuvo una exactitud del 76 %. Bakheet(2021) [13] también emplea esta técnica de extracción de características sobre imágenes de rayos X segmentadas en la ROI, preprocesadas con filtro de mediana y ecualizadas con su histograma para lograr una clasificación binaria, obteniendo una exactitud del 95,88 %.

Además del uso de las imágenes de rayos X y TC, se han empleado imágenes de LUS para realizar la tarea de clasificación, La Salvia(2021) [14] las utiliza realizando una separación de cuatro clases, asignando un puntaje de cero a tres dependiendo del daño pulmonar que el LUS presenta. La Salvia emplea una ResNet 18 y ResNet 50 para realizar esta clasificación de cuatro clases.

Tsai(2021) [15] utilizando también imágenes de LUS y una *Spatial Transformer Network*, realiza una clasificación entre pacientes con y sin COVID-19 para ello emplea una identificación de líneas de derrame pleural como criterio de clasificación, si las imágenes poseen esta línea los pacientes con considerados con COVID-19.

Otro trabajo que aborda una clasificación de tres clases de pacientes con COVID-19, neumonía y sin afecciones es el de Erfanian Ebadi(2021) [16]; empleando vídeos de LUS, así como dos 3D-ConvNet realiza una clasificación de un 90 % de exactitud.

1.4. Estructura de la tesis

El siguiente trabajo se compone de cinco capítulos, en el capítulo 2 se comienza con una descripción más amplia del ultrasonido pulmonar, así como los artefactos que se observan en este y cómo se relacionan con las enfermedades empleadas para la clasificación realizada. También se especifican los métodos usados para la extracción de características de las imágenes de ultrasonido, así como los métodos de clasificación tanto de aprendizaje máquina supervisado como de aprendizaje profundo con las que se llevó a cabo el trabajo y las métricas empleadas en su evaluación.

Posteriormente en el capítulo 3 se explica la metodología seguida para la clasificación de las imágenes de LUS, para en el capítulo 4 mostrar los resultados obtenidos y la discusión de los mismos. Finalmente en el capítulo 5 se exponen las conclusiones de este trabajo.

CAPÍTULO

2

CONCEPTOS GENERALES

2.1. Ultrasonido

El ultrasonido (US) es una onda mecánica cuya frecuencia en uso clínico es de 1 - 15MHz, estas ondas de US son producidas por un transductor.

La adquisición de una imagen de US clínico, conocido también como modalidad B; puede entenderse en una manera simple como se ilustra en la Figura 2.1. En esta imagen se observa como el transductor emite un haz de ultrasonido de manera secuencial, y conforme el haz de US pasa a través del tejido, una pequeña fracción de energía es reflejada por sus bordes y detectada nuevamente por el transductor. Los bordes del tejido tienen propiedades físicas y acústicas ligeramente distintas al resto del tejido.

La energía remanente del haz, es decir, aquella que no fue reflejada por el borde; es transmitida a través del tejido. Nuevamente las ondas reflejadas son detectadas por el transductor y la distancia a cada borde del tejido es calculada por el tiempo entre la transmisión y la detección de la señal.

Usando la velocidad del sonido a través del tejido, la profundidad del borde del tejido puede ser determinada. Además de que la intensidad de cada pixel en la imagen es proporcional a la intensidad de la señal detectada de ese punto [17].

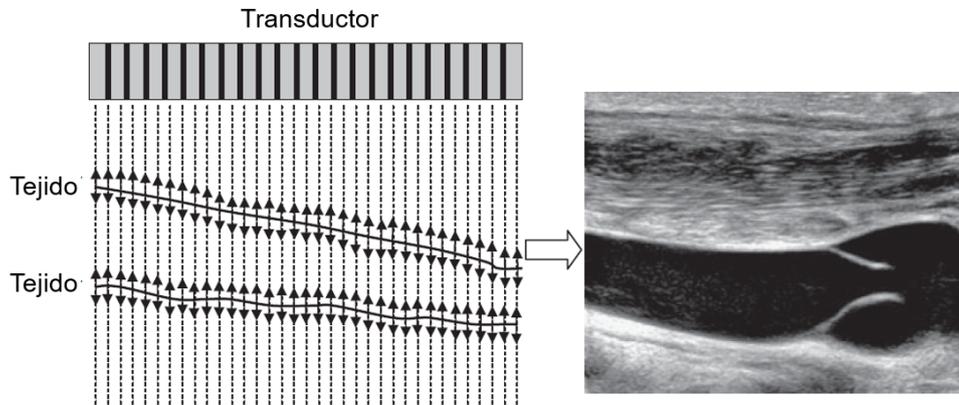


Figura 2.1: Principio básico de la adquisición de una imagen de ultrasonido, se observa como la onda mecánica es emitida, reflejada por el tejido y detectada de nuevo por el transductor [17]

2.2. Ultrasonido pulmonar

La enfermedad del coronavirus 2019 (COVID-19) es una enfermedad altamente contagiosa causada por el síndrome respiratorio agudo severo coronavirus 2 (SARS-CoV-2). Los órganos más afectados por esta enfermedad son los pulmones, esto se debe a que el virus entra directamente por sus células alveolares.

El COVID-19 es caracterizado por el síndrome de dificultad respiratoria aguda (ARDS) en el 20% de los pacientes infectados con alta mortalidad y morbilidad [18–20].

La tomografía computarizada (TC) es capaz de jugar un papel crucial en la evaluación de la evolución del COVID-19, así como en seleccionar apropiadamente el manejo de los pacientes infectados. Sin embargo, las desventajas de la TC incluyen la necesidad de transportar al paciente hasta la sala del tomógrafo, la exposición a la radiación ionizante y el riesgo de esparcimiento del virus [13, 21].

El ultrasonido pulmonar (LUS) tiene un buen desempeño como herramienta de diagnóstico cuando es realizada apropiadamente, es libre de los daños biológicos que aplicar radiación ionizante causa. Entre otras ventajas que el LUS posee, se encuentran la portabilidad del equipo que permite examinar al paciente sin tenerlo que transportar y la reducción de costos [22].

El LUS parece apropiado para observar los pulmones de los pacientes que involucran sospechas o averiguamientos de infección por SARS-CoV-2, debido a que es de fácil aplicabilidad, además que puede ser un método relevante

para monitorear a los pacientes [23].

2.2.1. Artefactos

La pleura es una membrana que rodea a los pulmones, esta sufre de afectaciones con distintas enfermedades respiratorias que los pulmones puedan presentar, debido a que la mayoría de los desórdenes agudos respiratorios involucran a la pleura [24, 25].

Con el ultrasonido pulmonar es posible detectar esta membrana, e inclusive se han observado que los artefactos, que se presentan en esta técnica de imagenología provienen de la pleura; por ello debe ser detectada para evitar errores [22].

Los tres artefactos principales de LUS en los que se centra este trabajo son mencionados a continuación:

Líneas A

Son líneas hiperecoicas paralelas horizontales, sin movimiento, representadas en intervalos regulares bajo la pleura, como se muestran en la Figura 2.2. Sus múltiples repeticiones, o reverberaciones, son equidistantes a la línea pleural y corresponden a artefactos de la misma; usualmente estos son interpretados como hallazgos normales en pacientes sanos [20, 22].

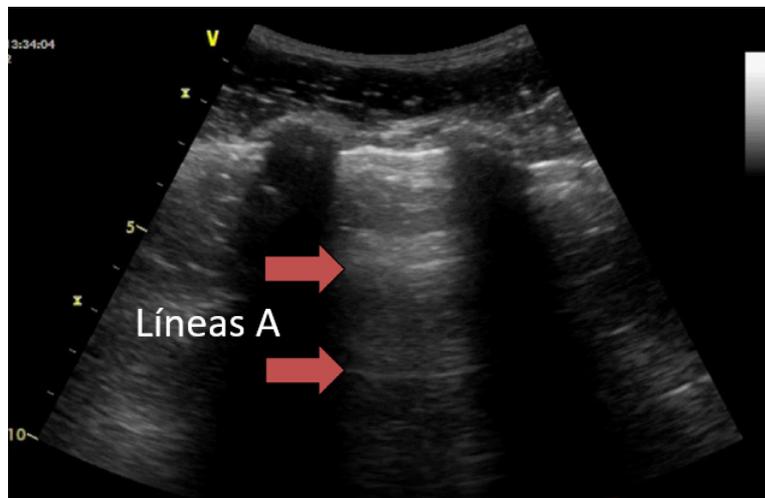


Figura 2.2: Líneas A mostradas en ultrasonido pulmonar [26]

Líneas B

Son reverberaciones longitudinales verticales, definidas como líneas hiperecoicas que comienzan desde la línea pleural y se extienden directamente hasta el fondo del monitor en el equipo de US, o bien, de la imagen obtenida. Las líneas B borran a las líneas A y se mueven conforme el transductor avanza, como se muestran en la Figura 2.3, siendo las líneas B relacionadas a pacientes con COVID-19 [21–23,27].

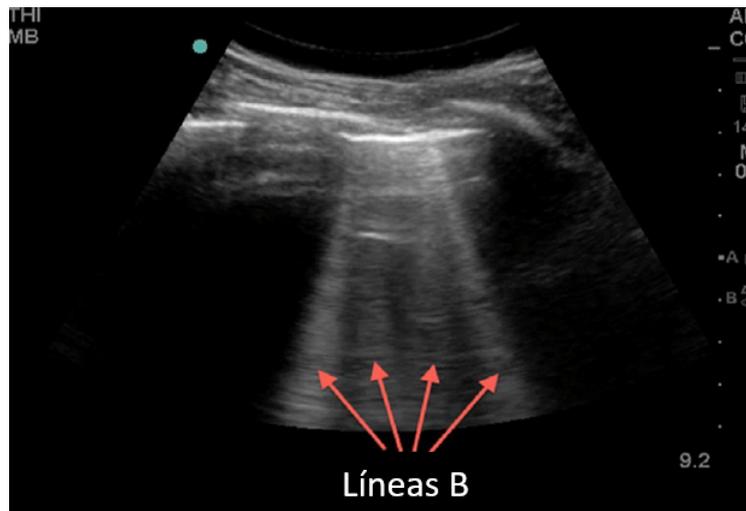


Figura 2.3: Líneas B mostradas en ultrasonido pulmonar [26]

Las líneas B son causadas por las reverberaciones del haz de US entre el aire alveolar ligeramente disminuido y el incremento de los fluidos intersticiales, estas son generadas por el engrosamiento interlobular [20, 28].

El número de líneas B está correlacionado con la severidad de la enfermedad, múltiples líneas B son consideradas signos ecográficos de síndrome intersticial pulmonar, y su número incrementa junto con el decremento de contenido de aire y el incremento de la densidad pulmonar [20, 23]

Consolidaciones

Cuando existe una consolidación del pulmón su estructura se puede observar en el LUS, representando el aire dentro del alvéolo o bronquio rodeado por inflamación o pus y la línea pleural completamente oscurecida [26], como se observa en la Figura 2.4. Las consolidaciones pulmonares son visualizadas en LUS como regiones hipoeoicas, que reflejan alta reducción en flujo de

aire y un incremento en la cantidad de exudado inflamatorio celular, estas son relacionadas a neumonía [22, 28].

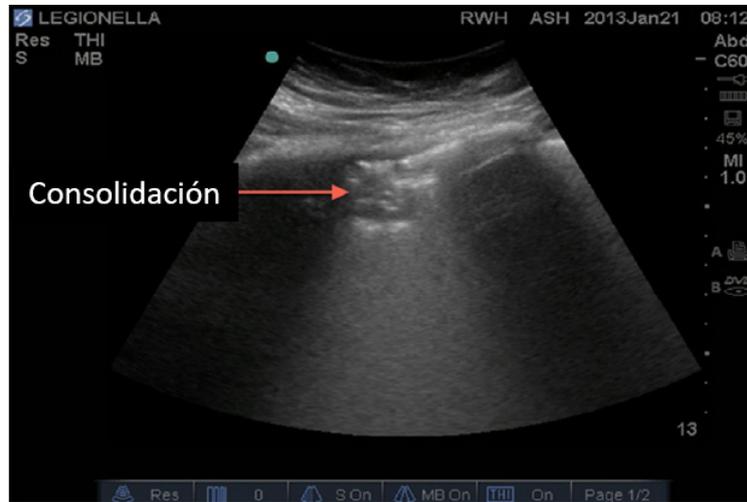


Figura 2.4: Consolidaciones mostradas en ultrasonido pulmonar [26]

2.3. Extracción de características

La extracción de características juega un rol importante en el reconocimiento de patrones de las imágenes, al ser uno de los métodos para facilitar el proceso de analizar una imagen.

La extracción de características es la técnica de determinar aquellas características más importantes de todos los rasgos que pueden ser medidos de la imagen, estas características extraídas proveen información acerca de la composición y las propiedades fundamentales de los elementos de la imagen. Además estas características representan a la imagen y pueden ser usadas para determinar la clase a la que pertenece la imagen en una tarea de clasificación [29, 30].

2.3.1. Histograma

El histograma de una imagen digital, es una gráfica con los L posibles niveles de intensidad en el rango de $[0, G]$, definido como una función discreta $h(r_k) = n_k$ donde r_k es el k -ésimo nivel de intensidad en un intervalo $[0, G]$ y n_k es el número de píxeles en la imagen cuyo nivel de intensidad de r_k .

El análisis estadístico derivado del histograma puede servir para comparar contrastes e intensidades entre imágenes [31].

2.3.2. Matriz de co-ocurrencia de niveles de gris (GLCM)

La matriz de co-ocurrencia de niveles de gris (GLCM) describe la interrelación espacial de los tonos de grises de una imagen. La GLCM contiene elementos que son recuento de un número de píxeles separados por cierta distancia y en alguna dirección angular.

Por lo general la GLCM es calculada en pequeñas ventanas, las cuales exploran la imagen completa, asociando una textura con cada pixel. Estas características tienen cambio rápido de valores en el caso de imágenes basadas en texturas finas y un cambio lento en los valores en imágenes basadas en texturas gruesas [32, 33].

- **Correlación:** Indica la similaridad de la textura de la imagen en dos direcciones perpendiculares, nombradas la dirección horizontal y vertical

$$Cor = \sum_{i,j=0}^{N-1} P_{i,j} \left[\frac{(i - \mu_i)(j - \mu_j)}{\sqrt{(\sigma_i^2)(\sigma_j^2)}} \right] \quad (2.1)$$

- **Contraste:** Indica la variación de la profundidad y la suavidad de las regiones de la imagen

$$Con = \sum_{i,j=0}^{N-1} P_{i,j}(i-j)^2 \quad (2.2)$$

- **Homogeniedad:** También llamada momento diferente inverso (IDM), indica la similitud estructural de una imagen

$$Hom = \sum_{i,j=0}^{N-1} \frac{P_{i,j}}{1+(i-j)^2} \quad (2.3)$$

- **ASM:** Segundo Momento Angular

$$ASM = \sum_{i,j=0}^{N-1} P_{i,j}^2 \quad (2.4)$$

- **Energía:** Es calculada como la raíz cuadrada del ASM

$$Ene = \sqrt{ASM} \quad (2.5)$$

- **Disimilaridad:**

$$Dis = \sum_{i,j=0}^{N-1} P_{i,j}|i-j| \quad (2.6)$$

Donde $P_{i,j}$ es el valor de probabilidad obtenido en celda (i, j) de la ventana de la imagen, y N es el número de fila o columna [34].

2.3.3. Selección de características

Es el proceso donde se seleccionan automática o manualmente aquellas características en los datos que contribuyen más a la variable de predicción o salida en la que se está interesado. Teniendo como ventajas la reducción del sobreajuste, el mejoramiento de la exactitud y la reducción del tiempo de entrenamiento [35].

Dentro de este proyecto se probaron tres técnicas de selección de características, las cuales fueron:

Importancia de características

Esta técnica computa rangos de importancia de características, otorga un puntaje para cada característica de los datos. Entre más alto sea este puntaje más importancia o relevancia tiene la característica [36,37].

Este puntaje es calculado por el método de importancia de Gini, el cual se basa en modelos de árboles, y calcula la importancia de cada característica como la suma del número de particiones en todos los árboles que incluyen esta característica, entre el número de muestras en las que se particiona [38–40].

Los puntajes pueden ser útiles en diferentes situaciones como lo son: entender mejor los datos o el modelo, así como reducir el número de características de entrada.

La importancia de características en un conjunto de datos indica cuáles características son más relevantes para el objetivo, y a la inversa cuáles son las menos relevantes [41].

Análisis de Componentes Principales (PCA)

Este método consiste en una transformación ortogonal para convertir muestras pertenecientes a variables correlacionadas en muestras de características linealmente no correlacionadas, utilizando álgebra lineal para transformar el conjunto de datos en una forma más comprimida. Este método extrae las características más relevantes de información de un conjunto de datos redundantes o ruidosos.

Las nuevas características son llamadas componentes principales, y son el mismo número al de las variables iniciales. Sin embargo, se pueden seleccionar el número de dimensiones o componentes principales esperadas en el resultado, con la finalidad de reducir dimensiones.

Los componentes principales son nuevas variables con dos propiedades: cada componente principal es una combinación lineal de las variables y los

componentes principales no están correlacionados entre ellos además de que la información redundante es eliminada.

Este método reduce el número original de variables eliminando el último componente principal que no contribuye significativamente a la variabilidad observada. Es una transformación lineal de los datos que minimiza la redundancia (medida a través de covarianza) y maximiza la información (medida por varianza).

El PCA es una técnica no supervisada que como tal no incluye la información de las etiquetas de los datos, si los datos están distribuidos normalmente entonces los componentes principales serán independientes [35, 42].

Selección univariante

Esta es una técnica de test estadístico que puede ser usado para seleccionar las características que tienen una mayor relación con la variable de salida. Devuelve una lista ranqueada de características basada en diferentes funciones de puntuación estadística.

El método de selección univariante no considera las dependencias entre las características del conjunto de datos, son independiente entre ellas [36, 43].

2.4. Métodos de Clasificación

2.4.1. Aprendizaje máquina supervisado

Los métodos de aprendizaje máquina, o ML por sus siglas en inglés (*Machine Learning*), permiten a las computadoras aprender sofisticados patrones y perspectivas de los datos sin tener que estar explícitamente programados.

Usando técnicas estadísticas y datos, los métodos de aprendizaje máquina aprenden para poder realizar predicciones o clasificaciones. Esto permite que la computadora tome los datos de entrada, aprenda sobre ellos y prediga una salida, además de que mejore su propio aprendizaje para cuando se le presenten datos similares en tareas futuras [44, 45].

Los métodos de aprendizaje máquina son denominados supervisados cuando los datos de entrada de los algoritmos fueron etiquetados, es decir, asociados con la respuesta verdadera o la salida de interés. Este método permite a la computadora clasificar o predecir objetos, problemas o situaciones con base en los datos etiquetados que la alimentaron. El objetivo del aprendizaje máquina supervisado es tener una computadora que reciba nuevas variables de entrada y prediga o clasifique correctamente las variables de salida [45, 46].

Naive Bayes

Es un método de clasificación supervisada, basado en un enfoque probabilístico empleando el teorema de Bayes, requiere todas las suposiciones para ser incorporados explícitamente en modelos que luego se utilizan para derivar reglas de decisión o clasificaciones óptimas. Es llamado ingenuo porque hace la suposición de que las características de ocurrencia son independientes una de la otra [47].

Es construido usando los datos de entrenamiento para estimar la probabilidad de cada clase dados los vectores de características de una nueva instancia. Dado un ejemplo representado por el vector de características X , el teorema de Bayes proporciona un método para calcular el probabilidad de que X pertenezca a la clase C_i , denotado como $p(C_i|X)$:

$$P(C_i|X) = \frac{C_i \prod_{j=1}^N P(x_j|C_i)}{X} \quad (2.7)$$

esto significa que el clasificador ingenuo de Bayes aprende la probabilidad condicional para cada atributo $x_j (j = 1, 2, \dots, N)$ de X dada la etiqueta de clase C_i . Por lo tanto, el problema de clasificación puede declararse como dado

un conjunto de características observadas x_j , de un objeto X , clasifique X en una de sus clases, en la que presenta mayor probabilidad de pertenencia [48].

Máquina de Soporte Vectorial (SVM)

Es un algoritmo de aprendizaje máquina supervisado que tiene como objetivo hacer la mejor línea o límite de decisión llamado hiperplano, el cual puede separar un espacio de n -dimensiones en clases, y con ello posteriormente poder insertar fácilmente nuevos datos en la categoría correcta.

Para crear el límite de decisión SVM elige los puntos extremos, los cuales se denominan vectores de soporte. Este algoritmo considera el valor de cada propiedad como el valor de una coordenada particular, cada elemento de los datos es dividido como un punto en un espacio n -dimensional y después de esto se ejecuta una clasificación para encontrar el hiperplano que diferencie bien las clases.

Una buena separación es lograda por un hiperplano que tiene la mejor distancia al punto más cercano de los datos de entrenamiento de cualquier clase. SVM tiene como fin llegar al hiperplano óptimo que divide los datos en diferentes clases [49].

K-Vecinos más cercanos (K-NN)

El método de K-NN supervisado puede servir para clasificar datos con etiquetas discretas o realizar regresión para datos con etiquetas continuas. El principio detrás de este método es encontrar un número predefinido de muestras en el entrenamiento cercanas en distancia a un nuevo punto y predecir la etiqueta a partir de esto. El número de puntos nuevos puede ser definido por el usuario como una constante K .

A pesar de su simplicidad, K-NN es exitoso en varios problemas de clasificación y regresión, incluso en situaciones donde el límite de decisión es bastante irregular. Este método de clasificación es calculado por una simple mayoría de votos al vecino más cercano de cada punto, a este se le asigna la clase de datos que tiene la mayor cantidad de representantes dentro de los vecinos más cercanos del punto [50].

Bosques aleatorios (RF)

Es un algoritmo de aprendizaje máquina supervisado que puede ser usado tanto para tareas de clasificación como de regresión. Está formado por un conjunto de árboles de decisión individuales, cada uno entrenado con una muestra aleatoria extraída de los datos de entrenamiento originales, esto implica que cada árbol se entrena con unos datos ligeramente distintos.

En cada árbol individual, las observaciones se van distribuyendo por bifurcaciones (nodos) generando la estructura del árbol hasta alcanzar un nodo terminal y la predicción de una nueva observación se obtiene agregando las predicciones de todos los árboles individuales que forman el modelo. RF selecciona la mejor solución por promedios de votación, y también provee un muy buen indicador de la importancia de las características [51, 52].

2.4.2. Aprendizaje profundo

El aprendizaje profundo, DL por sus siglas en inglés (*Deep Learning*) es una sub-rama del aprendizaje máquina inspirado por la estructura del cerebro. Las técnicas de aprendizaje profundo se basan en configuraciones multicapas para procesar los datos, estas son llamadas redes neuronales artificiales. Por lo regular, las redes neuronales artificiales tienen cientos de capas de profundidad y pueden entrenarse a sí mismas con un largo conjunto de datos para poder realizar predicciones exactas en nuevas entradas de datos desconocidos [6, 45].

Una red neuronal convolucional, CNN por sus siglas en inglés (*Convolutional Neural Network*), es una clase de red de aprendizaje profundo usada para los problemas con imágenes. Las CNN consisten principalmente en tres diferentes tipos de capas: las convolucionales, las de pooling y las completamente conectadas. Los procesos de extracción de características de las imágenes suceden en los dos primeros tipos de capas, convolucionales y de pooling. El proceso de clasificación sucede en el tipo de capa completamente conectada [6].

Aprendizaje por transferencia y fine-tuning

El principal principio detrás del aprendizaje por transferencia es que las habilidades aprendidas por una red neuronal convolucional (CNN) cuando estudia un modelo pueden ser aplicadas a diferentes tareas de aprendizaje. Se pueden cargar versiones pre-entrenadas de CNN, previamente entrenadas con más de un millón de imágenes provenientes de distintas bases de datos, como lo es ImageNet [6, 7].

En comparación con los métodos de aprendizaje máquina tradicionales, los modelos basados en CNN por aprendizaje por transferencia poseen las siguientes ventajas: se requiere menor pre-procesamiento del conjunto de datos, el proceso de aprendizaje es más rápido, optimizar los numerosos parámetros puede ajustar la complejidad del tiempo y estos modelos trabajan increíblemente bien con bases de datos limitadas [4].

El Fine-tuning puede ser usado para adaptar una CNN pre-entrenada a un conjunto de datos diferente. Esto se realiza transfiriendo los pesos de una red pre-entrenada a la CNN a entrenar, con la diferencia que la última capa completamente conectada de la red pre-entrenada es reemplazada con una nueva capa completamente conectada que tiene el mismo número de neuronas que clases en la nueva aplicación [53, 54].

Dentro de este trabajo se emplearon varias arquitecturas ya pre-entrenadas, las cuales fueron seleccionadas por ser las que se repitieron con mayor frecuencia en la revisión del estado del arte presentado en la Sección 1.3.

Aplicando aprendizaje por transferencia y fine-tuning a todas las arquitecturas para realizar una tarea de clasificación de tres clases en imágenes de LUS, se compararon y terminaron por elegirse las siguientes arquitecturas:

Arquitecturas de las redes neuronales convolucionales

- **VGG-19**

El grupo geométrico visual (VGG) fue desarrollado por la universidad de Oxford en 2014, esta red es caracterizada por la simplicidad y uniformidad de su arquitectura. En específico la red VGG-19 consiste en 19 capas convolucionales apiladas una tras la otra con un tamaño de kernel de 3×3 , que incrementa el número de convoluciones de acuerdo al crecimiento de la profundidad en la red [5].

- **DenseNet 201**

La red convolucional densa 201 (DenseNet 201) es una CNN de 201 capas de profundidad, en esta red cada capa se conecta con todas las capas anteriores y estas funcionan como entrada de la siguiente capa.

La DenseNet 201 es fácil de entrenar debido a que la manera en la que está conectada permite un amplio flujo de información [4, 10].

2.5. Métricas

Exactitud

Es definida como la fracción de todas las decisiones que son correctas. Se puede denotar como se muestra en la ecuación 2.8; donde VN denota los verdaderos negativos, es decir, aquellos que son realmente negativos y fueron clasificados como negativos, VP como verdaderos positivos, aquellos positivos clasificados como positivos. Los FP y FN son aquellos clasificados como positivos y negativos, respectivamente, sin serlo [55].

$$A_c = \frac{VN + VP}{VN + VP + FP + FN} \quad (2.8)$$

Recall

Es el número de elementos identificados correctamente como positivos del total de positivos verdaderos. El recall es definido por la ecuación 2.9 donde los verdaderos positivos son divididos entre la suma de totales positivos verdaderos, es decir, los verdaderos positivos y falsos negativos [56].

$$Recall = \frac{VP}{VP + FN} \quad (2.9)$$

El recall en pruebas de diagnóstico médico recibe el nombre de sensibilidad, y corresponde a la probabilidad de que un enfermo sea identificado correctamente [57], es decir, cuan sensible es una prueba a la presencia de la enfermedad.

Precisión

Es el número de elementos identificados correctamente como positivos de un total de elementos identificados como positivos. La precisión es definida con la ecuación 2.10 donde los verdaderos positivos son divididos entre la suma de verdadero positivo con falso positivo [56].

$$Precision = \frac{VP}{VP + FP} \quad (2.10)$$

También es conocido como el valor predictivo positivo (VPP) e indica la probabilidad de que un individuo con prueba positiva tenga la enfermedad [57].

CAPÍTULO

3

METODOLOGÍA PARA LA CLASIFICACIÓN DE IMÁGENES DE LUS

3.1. Conjunto de datos

Para realizar la clasificación se obtuvo una base de imágenes de LUS dividida en tres clases: COVID-19, neumonía y regular. Este conjunto de datos está compuesto por dos bases distintas, la primera proveniente de un repositorio de Kaggle [58] con 988 imágenes de COVID-19, 731 de neumonía y 1,276 de pacientes sanos.

La segunda base de datos fue recolectada de "Butterfly's COVID-19 gallery" [59] una galería que consistía en videos de LUS; ocho de COVID-19, nueve de neumonía y dos de pulmones sin afecciones. Estos fueron adquiridos con anterioridad por médicos y radiólogos mediante el uso de un transductor llamado "The Butterfly iQ", el cual solo requiere conectarse a un teléfono celular y mediante una aplicación móvil hace posible el obtener estas imágenes.

De cada video fueron extraídos los cuadros con la finalidad de obtener solamente las imágenes, el conjunto final de los datos consistía en 1,562 imágenes.

nes de COVID-19, 1,705 de neumonía y 832 imágenes de pulmones sanos.

Combinando ambas bases de datos se obtuvo un total de 2,550 imágenes de COVID-19, 2,436 de neumonía y 2,108 de pacientes sin afecciones, con distintas resoluciones y en escala RGB. Se realizó una selección aleatoria de 700 imágenes por clase; este total de 2,100 imágenes fue usado para el entrenamiento y evaluación de los métodos de aprendizaje máquina y aprendizaje profundo.

3.2. Comparativa de métodos de aprendizaje máquina

En la Figura 3.1 se ilustra el proceso llevado a cabo con un diagrama de bloques. Primero las imágenes de LUS, descritas en la Sección 3.1, fueron convertidas a escala de grises y redimensionadas a una resolución de 256×256 píxeles. Después de esto se extrajeron sus características, y mediante el uso de distintos métodos de selección de características fueron entrenados cuatro algoritmos de aprendizaje máquina, y sus resultados se evaluaron con las métricas mencionadas en la Sección 2.5.

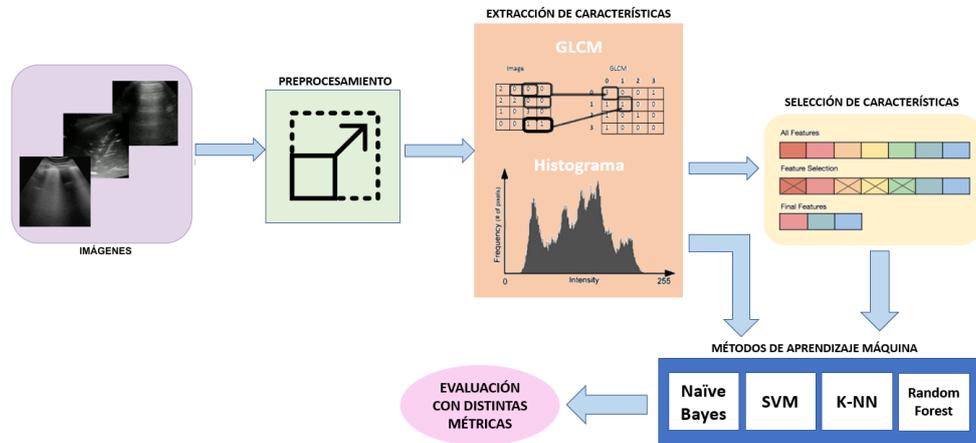


Figura 3.1: Diagrama de bloques del proceso de clasificación de imágenes de LUS empleando métodos de extracción y selección de características para entrenar métodos de aprendizaje máquina y evaluar sus resultados con distintas métricas

3.2.1. Extracción y selección de características

Las características extraídas de cada imagen consistían en los estadísticos de sus GLCM e histograma. Se extrajeron 36 parches por imagen en un solo ángulo de 0° , estos parches se tomaban cada 40×40 píxeles, y a cada parche se le calcularon los seis estadísticos mencionados en la Sección 2.3.2.

Por lo tanto una imagen proporcionaba 216 características de textura; esta información es de interés debido a que al estar trabajando con LUS en escala de grises es importante conocer los cambios en sus texturas dentro de la imagen y al hacerlo por ventanas se podrían obtener mejores valores dependiendo del cambio que tuviera esa parte de la imagen, siendo un análisis más específico.

Además de lo anterior, cinco estadísticos por histograma fueron calculados: la media, desviación estándar, entropía, skewness y kurtosis.

En total por cada imagen se obtenían 221 características, que podían ser usadas para el entrenamiento de la tarea de clasificación.

Debido a que se contaba con un gran número de características por imagen se emplearon los métodos de selección de características, mencionados en la Sección 2.3.3, para determinar los más importantes a usarse en la tarea de clasificación con aprendizaje máquina.

El método de selección de características univariante eligió tres características, siendo el mismo número seleccionado por PCA y el método de importancia de características escogió seis.

3.2.2. Entrenamiento de los métodos de aprendizaje máquina

Empleando los métodos de aprendizaje máquina descritos en la Sección 2.4.1, se realizaron cuatro entrenamientos: el primero de ellos fue usando las 221 características totales y otros tres entrenamientos fueron empleando los conjuntos reducidos de características obtenidos con los métodos de selección de características.

Los resultados de cada método de aprendizaje máquina fueron evaluados con las métricas descritas en la Sección 2.5, estos resultados se pueden observar en la Sección 4.1.

3.3. Comparativa de arquitecturas de aprendizaje profundo

Además de lo anterior descrito, se hicieron varias pruebas con distintas arquitecturas de redes neuronales empleando el mismo conjunto de imágenes usadas en las pruebas en la Sección 3.2. Se entrenaron ocho redes neuronales convolucionales (CNN) usando técnicas de aprendizaje por transferencia y fine-tuning, para lograr la tarea de clasificación de las imágenes de COVID-19, neumonía y pacientes regulares. Las ocho arquitecturas empleadas fueron: VGG-13, VGG-19, ResNet 18, ResNet 50, ResNet 152, DenseNet 201, MobileNet V2 y GoogleNet.

Los resultados de estos entrenamientos fueron evaluados con las métricas descritas en la Sección 2.5, estos valores fueron comparados entre todas las redes para al final elegir las que tuvieran el desempeño más adecuado, las cuales fueron DenseNet 201 y VGG-19. En la Figura 3.2 se muestra el diagrama de bloques del proceso anteriormente descrito.

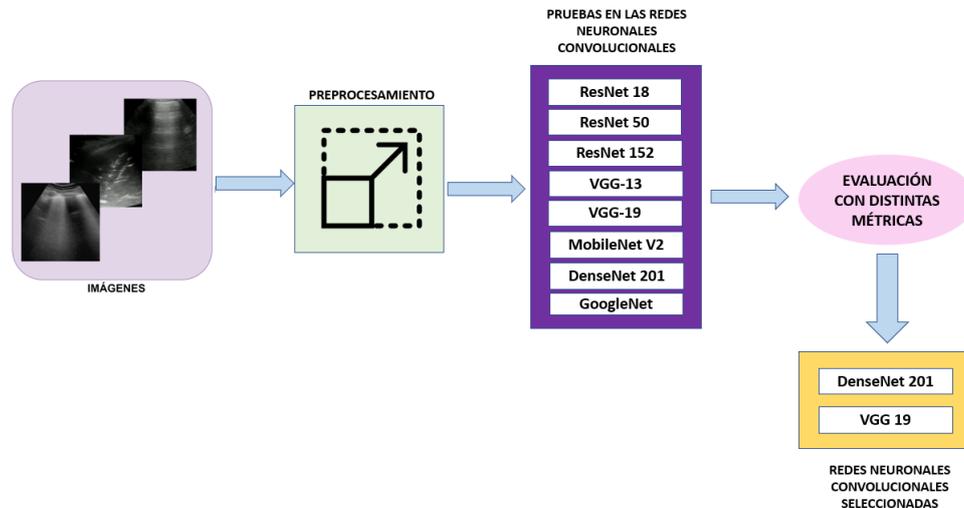


Figura 3.2: Diagrama de bloques del proceso seguido para aprendizaje profundo donde se probaron ocho arquitecturas y mediante la comparativa de las métricas obtenidas seleccionar tres de mejor desempeño

Para todas las redes se empleó el optimizador Adam, usando una tasa de aprendizaje de $1E - 3$ para cada CNN y entropía cruzada como función de pérdida. Los resultados obtenidos pueden observarse en la Sección 4.2.

3.4. Comparativa entre los métodos de clasificación seleccionados

Después de realizar las pruebas tanto de aprendizaje máquina como en aprendizaje profundo se obtuvo la metodología final empleada para la clasificación a tres clases de las imágenes de LUS con COVID-19, neumonía y pacientes sin afecciones; la cual está ilustrada en la Figura 3.3.

En esta metodología se emplea el mismo conjunto de imágenes, descrito en la Sección 3.1, para ambos entrenamientos. La metodología final puede dividirse en dos partes; en la primera se extraen las características de estas imágenes mediante el uso del GLCM y el histograma, después se emplean métodos de selección de características y con estos resultados se entrenan los métodos de aprendizaje máquina.

En la segunda parte solo se emplean las dos redes seleccionadas con el mismo conjunto de imágenes, sin hacer extracción o selección de características. El resultado de ambos entrenamientos es evaluado con las métricas descritas en la Sección 2.5 para observar el desempeño de estos algoritmos y compararlos en la tarea de clasificación de las tres clases.

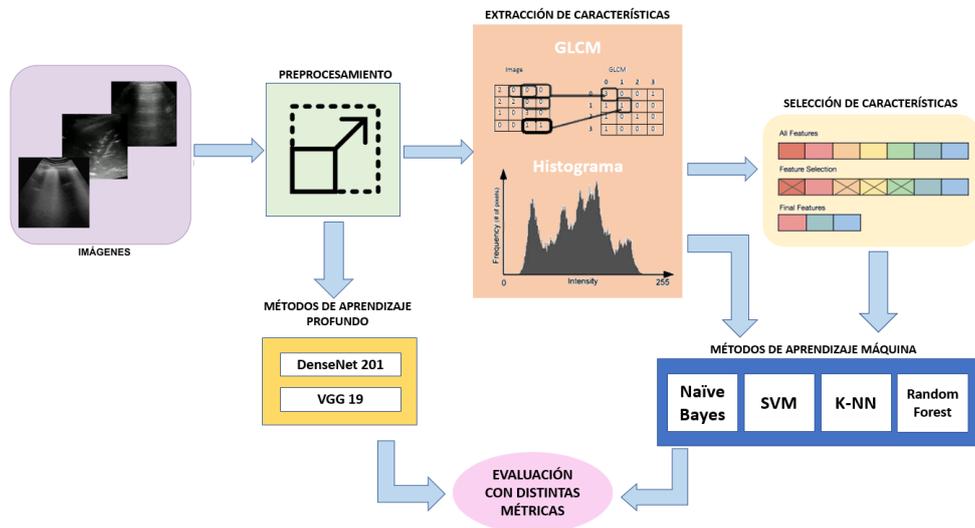


Figura 3.3: Diagrama de bloques del proceso completo

CAPÍTULO

4

RESULTADOS

Como parte preliminar de este trabajo se logró una publicación en la conferencia "SPIE Photonics Europe 2022", este artículo se incluye en el anexo A.

4.1. Evaluación y comparativa de los métodos de aprendizaje máquina

Los resultados obtenidos de los entrenamientos de los métodos de aprendizaje máquina fueron evaluados con las métricas mencionadas en la Sección 2.5, en la Tabla 4.1 podemos observar una comparativa de los valores obtenidos para la exactitud en cada uno de los métodos tanto de aprendizaje máquina como de selección de características. Además en la Figura 4.1 también se muestran los resultados de la evaluación de exactitud con una gráfica comparativa.

Tanto en la Tabla 4.1 y Figura 4.1 podemos observar que los valores de exactitud son los mismos para casi todos los métodos independientemente de la técnica de selección de características empleada, se puede ver gráficamente como las barras azules correspondientes al método de Naive Bayes tienen el mismo tamaño entre ellas al tener los mismos valores, siendo también las

Tabla 4.1: Comparativa de exactitud entre los métodos de ML y los métodos selección de características para la clasificación de las imágenes de LUS de pacientes con COVID-19, neumonía o sin afecciones

Métodos de Aprendizaje Máquina	Características - Exactitud			
	Todos los datos	Selección Univariante	PCA	Importancia de característica
<i>Naive Bayes</i>	77.14	77.14	77.14	77.14
<i>SVM</i>	93.38	93.38	93.38	93.38
<i>K-NN</i>	96.90	96.90	96.90	96.90
<i>Random Forest</i>	97.95	98.05	98.10	98.00

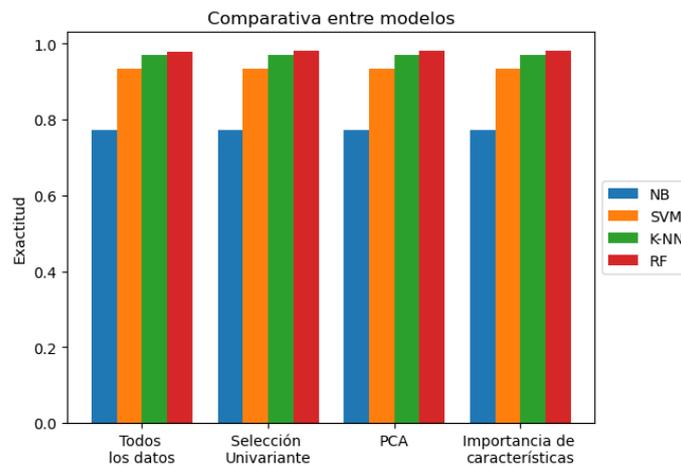


Figura 4.1: Gráfica comparativa de exactitud entre los métodos de ML y los métodos selección de características para la clasificación de las imágenes de LUS de pacientes con COVID-19, neumonía o sin afecciones

de menor exactitud en comparativa con los otros métodos de aprendizaje máquina empleados.

A su vez las barras naranjas de SVM tienen el mismo valor sin importar el subconjunto de características evaluadas, si bien su exactitud es mayor que la de NB siguen sin ser la más alta. Observando el K-NN, las barras verdes, nos encontramos con que de igual forma poseen el mismo tamaño en todos los conjuntos de datos pero esta exactitud es mayor que en los casos anteriores.

La única diferencia entre los valores de cada conjunto de características proviene del método de Random Forest, aunque esta diferencia no es muy significativa; no puede ser apreciada en las barras rojas de la gráfica compa-

rativa pero si en los valores de la Tabla 4.1, donde se tienen las cifras para cada conjunto de datos siendo el seleccionado por el método de PCA el que alcanzó la exactitud más alta, con un valor del 98.10.

Otra comparativa entre los conjuntos de características usados para el entrenamiento y los métodos de clasificación es mostrada en las gráficas de las Figuras 4.2 y 4.3, estas muestran respectivamente la comparativa entre precisión y recall, además los valores obtenidos de las métricas evaluadas se encuentran descritos en la Tabla 4.2.

Tabla 4.2: Comparativa entre Precisión (P) y Recall (R) de los métodos de aprendizaje máquina y selección de características para la clasificación de las imágenes de LUS de pacientes con COVID-19, neumonía o sin afecciones

Clase	Todos los datos				Clase	Selección Univariante			
	NB	SVM	K-NN	RF		NB	SVM	K-NN	RF
COVID-19	P: 0.75 R: 0.77	P: 0.95 R: 0.88	P: 0.96 R: 0.97	P: 0.97 R: 0.98	COVID-19	P: 0.75 R: 0.77	P: 0.95 R: 0.88	P: 0.96 R: 0.97	P: 0.98 R: 0.98
Neumonía	P: 0.81 R: 0.69	P: 0.97 R: 0.94	P: 0.98 R: 0.94	P: 0.98 R: 0.97	Neumonía	P: 0.81 R: 0.69	P: 0.97 R: 0.94	P: 0.98 R: 0.94	P: 0.98 R: 0.97
Regular	P: 0.75 R: 0.84	P: 0.88 R: 0.96	P: 0.96 R: 0.99	P: 0.98 R: 0.98	Regular	P: 0.75 R: 0.84	P: 0.88 R: 0.96	P: 0.96 R: 0.99	P: 0.97 R: 0.98
Clase	PCA				Clase	Importancia de Característica			
	NB	SVM	K-NN	RF		NB	SVM	K-NN	RF
COVID-19	P: 0.75 R: 0.77	P: 0.95 R: 0.88	P: 0.96 R: 0.97	P: 0.98 R: 0.97	COVID-19	P: 0.75 R: 0.77	P: 0.95 R: 0.88	P: 0.96 R: 0.97	P: 0.97 R: 0.98
Neumonía	P: 0.81 R: 0.69	P: 0.97 R: 0.94	P: 0.98 R: 0.94	P: 0.98 R: 0.97	Neumonía	P: 0.81 R: 0.69	P: 0.97 R: 0.94	P: 0.98 R: 0.94	P: 0.98 R: 0.97
Regular	P: 0.75 R: 0.84	P: 0.88 R: 0.96	P: 0.96 R: 0.99	P: 0.97 R: 0.98	Regular	P: 0.75 R: 0.84	P: 0.88 R: 0.96	P: 0.96 R: 0.99	P: 0.98 R: 0.98

En las gráficas de precisión, Figura 4.2, podemos observar que para las tres clases el método de Naives Bayes (barra azul) fue el que tuvo un desempeño más bajo comparado a los otros, sin importar que subconjunto de características se estuviera empleando en su entrenamiento, obteniendo precisiones por debajo de 0.82.

Por otro lado el mejor desempeño de los clasificadores en cuanto a la precisión para las clases de COVID-19 y regular, mostradas en las Figuras 4.2(a) y 4.2(c) respectivamente, fue el obtenido con el método de Random Forest (barra roja) y para la clase de neumonía, presentada en la Figura 4.2(b), K-NN (barra verde) fue el que mostró un mejor desempeño en precisión.

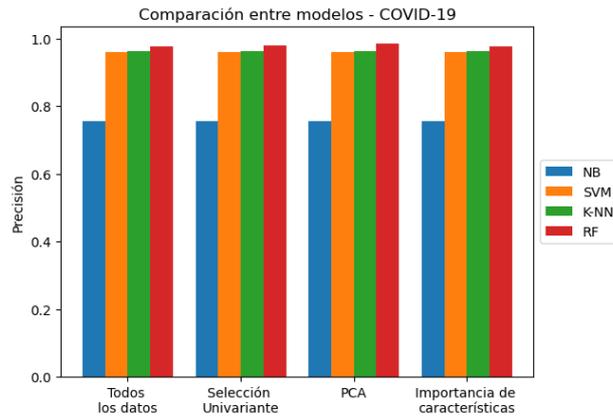
Ahora para las gráficas de recall de la Figura 4.3, ocurre como con la precisión, donde el Naive Bayes (barra azul) tuvo el desempeño más bajo para todas las clases y conjuntos de datos, con valores menores a 0.85.

El mejor desempeño fue de nuevo RF (barra roja) para la clase COVID-19 y neumonía, mostradas respectivamente en las Figuras 4.3(a) y 4.3(b); y para la clase regular de la Figura 4.3(c) el desempeño más alto de recall se obtuvo con K-NN (barra verde)

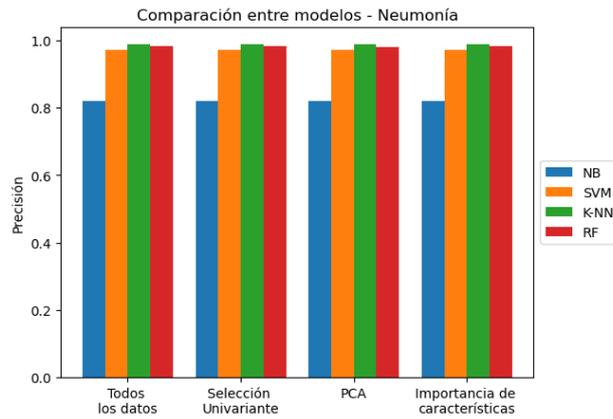
Tomando como referencia la Tabla de comparaciones 4.2 podemos observar los valores numéricos obtenidos para precisión y recall de cada clase evaluadas en los cuatro conjuntos de datos entrenados (todos los datos, selección univariante, PCA e importancia de características) con los cuatro métodos de aprendizaje máquina (NB, SVM, K-NN y RF).

En esta Tabla podemos observar como el método de Naive Bayes tanto en precisión como en recall es el de menor desempeño en todas las clases y conjuntos de características. También los demás métodos de aprendizaje máquina como en el caso de la exactitud, no varían los valores independientemente del método y el conjunto de datos. La variación proviene, al igual que en la exactitud, del método de RF aunque estas son muy pequeñas los valores alcanzados son muy altos encontrándose arriba del 0.97 para todas las clases.

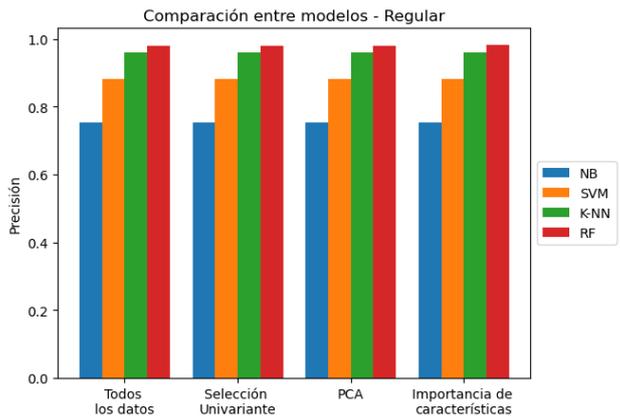
Considerando estos resultados (Tabla 4.2) y los de la Tabla 4.1, es decir, tomando la combinación de todas las métricas: exactitud, precisión y recall; podemos decir que el mejor desempeño es alcanzado por la combinación PCA – Random Forest, en esta tarea de clasificación de tres clases. Esto debido a que en exactitud fue la combinación que obtuvo el valor mayor, para precisión el mejor desempeño es también alcanzado por el método de RF en los subconjuntos de características seleccionados por PCA y por Selección Univariante. Con referencia al recall RF muestra los mejores valores para las clases de COVID-19 y neumonía, y KNN para la clase de pacientes sin afecciones. Por lo tanto RF fue tomado como el mejor método de clasificación de aprendizaje máquina, y PCA como el subconjunto de características al tener la exactitud más alta, ser de lo mayores en precisión y tener valores de recall altos.



(a) COVID-19

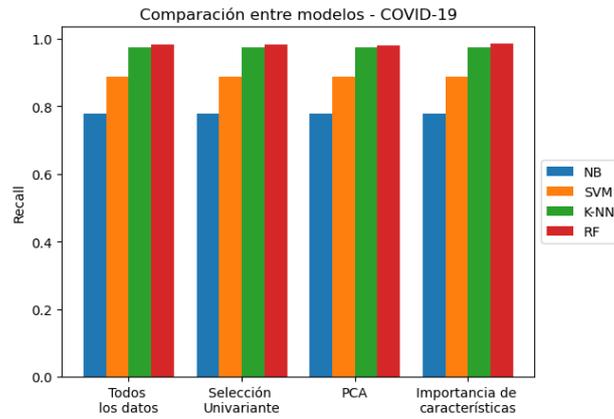


(b) Neumonía

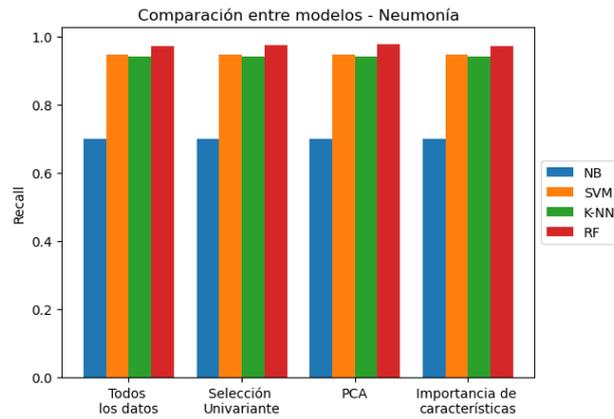


(c) Regular

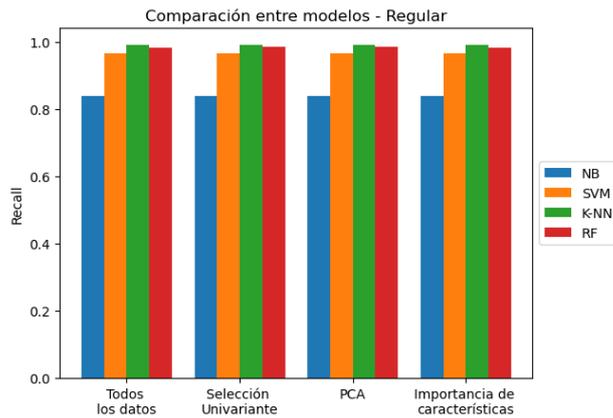
Figura 4.2: Comparación de precisión entre los distintos métodos de aprendizaje máquina y los métodos selección de características para la clasificación de las imágenes de LUS de pacientes con COVID-19, neumonía o sin afecciones



(a) COVID-19



(b) Neumonía



(c) Regular

Figura 4.3: Comparación de recall entre los distintos métodos de aprendizaje máquina y los métodos de selección de características para la clasificación de las imágenes de LUS de pacientes con COVID-19, neumonía o sin afecciones

4.2. Comparativa entre el desempeño de las redes neuronales convolucionales

Primeramente las ocho arquitecturas comparadas fueron: VGG-13, VGG-19, ResNet 18, ResNet 50, ResNet 152, DenseNet 201, MobileNet V2 y GoogleNet. En la Tabla 4.3, se muestran exactitudes, precisiones y recalls alcanzados por cada una de las redes entrenadas además en el anexo B se pueden encontrar las gráficas de pérdida y exactitud de cada una de estas redes.

Tabla 4.3: Comparativa de las métricas evaluadas entre todas las redes neuronales convolucionales para la clasificación de las imágenes de LUS de pacientes con COVID-19, neumonía o sin afecciones

Métrica	Arquitectura de la CNN			
	<i>VGG-13</i>	<i>VGG-19</i>	<i>ResNet 18</i>	<i>ResNet 50</i>
Exactitud	0.94	0.97	0.84	0.88
Precisión - Recall				
<i>COVID - 19</i>	0.83 - 1.00	0.91 - 1.00	0.80 - 0.50	0.82 - 0.90
<i>Neumonía</i>	1.00 - 0.94	1.00 - 1.00	0.87 - 0.93	0.90 - 0.75
<i>Regular</i>	1.00 - 0.83	1.00 - 0.83	0.83 - 1.00	0.91 - 1.00
Métrica	Arquitectura de la CNN			
	<i>ResNet 152</i>	<i>DenseNet 201</i>	<i>MobilNet V2</i>	<i>GoogleNet</i>
Exactitud	0.94	0.97	0.91	0.91
Precisión - Recall				
<i>COVID - 19</i>	0.88 - 1.00	1.00 - 1.00	0.89 - 1.00	0.89 - 1.00
<i>Neumonía</i>	1.00 - 0.93	1.00 - 0.93	1.00 - 0.79	1.00 - 0.79
<i>Regular</i>	0.90 - 0.90	0.91 - 1.00	0.83 - 1.00	0.83 - 1.00

Se decidió seleccionar las mejores arquitecturas, se tomaron aquellas que tuvieran la exactitud más alta así como buenos valores de precisión y recall. Al final con estas condiciones se seleccionaron solo dos arquitecturas: VGG-19 y DenseNet 201.

4.3. Evaluación de los métodos de aprendizaje profundo seleccionados

Los datos obtenidos con las métricas correspondientes para los métodos de aprendizaje profundo después de su entrenamiento se muestran en la Tabla 4.4, donde se observa que las dos CNN obtuvieron una exactitud de 0.97 siendo esto un comportamiento satisfactorio.

En cuanto a la precisión y recall la arquitectura VGG-19 mostró un menor desempeño comparado con la DenseNet 201.

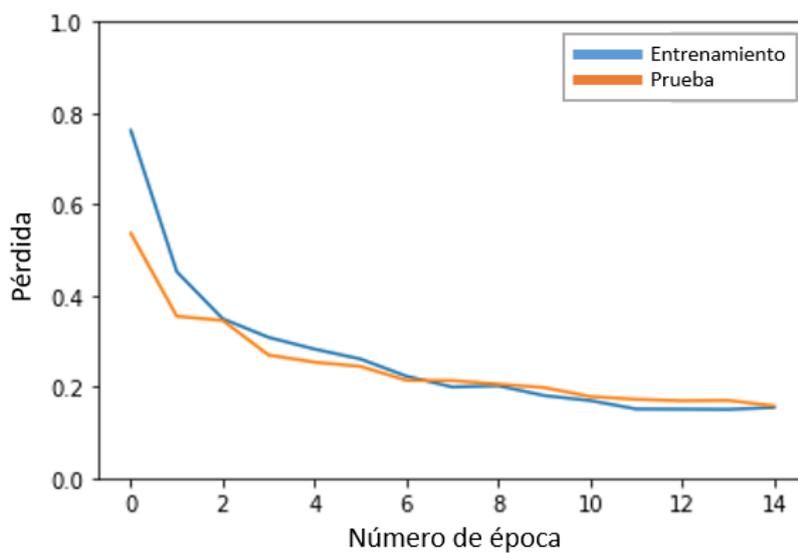
Aunque las arquitecturas VGG-19 y DenseNet 201 tienen el mismo valor de exactitud, su precisión y recall varían; con todos estos datos se puede observar que la arquitectura con el mejor desempeño en las tres métricas evaluadas es DenseNet 201.

Tabla 4.4: Comparativa entre los métodos de aprendizaje profundo para la clasificación de las imágenes de LUS de pacientes con COVID-19, neumonía o sin afecciones

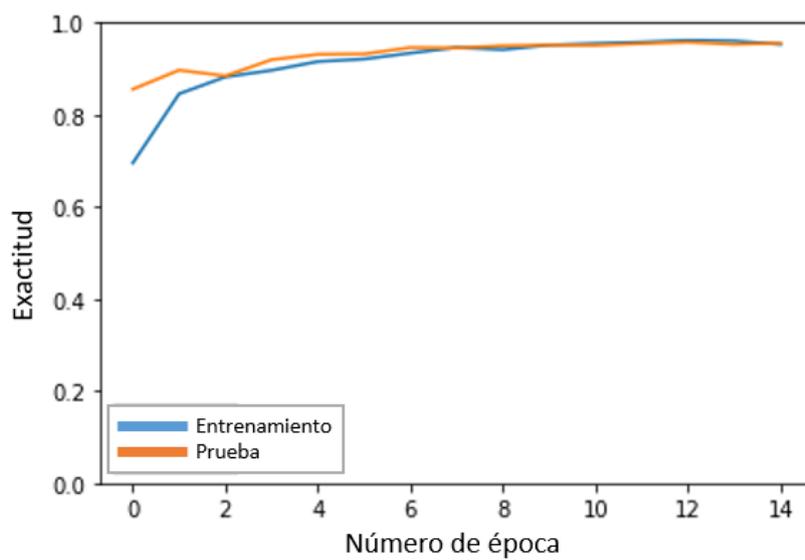
Métrica	Arquitectura de CNN	
	<i>DenseNet 201</i>	<i>VGG-19</i>
Exactitud	0.97	0.97
Precisión - Recall		
<i>COVID-19</i>	1.00 - 1.00	0.91 - 1.00
<i>Neumonía</i>	1.00 - 0.93	1.00 - 1.00
<i>Regular</i>	0.91 - 1.00	1.00 - 0.83

Las curvas de pérdida y exactitud de las dos CNN entrenadas se muestran en las Figuras 4.4 y 4.5. Comenzando por las curvas de pérdida de las Figuras 4.4(a) y 4.5(a), se puede observar que todas ellas coinciden con un decaimiento rápido que converge tanto en el conjunto de entrenamiento como en el de evaluación.

Además en las curvas de exactitud de las Figuras 4.4(b) y 4.5(b), podemos observar como alcanzan altos niveles de exactitud en pocas épocas. Los desempeños tanto en pérdida como en exactitud sugieren que la tarea de clasificación fue fácilmente lograda por los métodos de aprendizaje profundo.

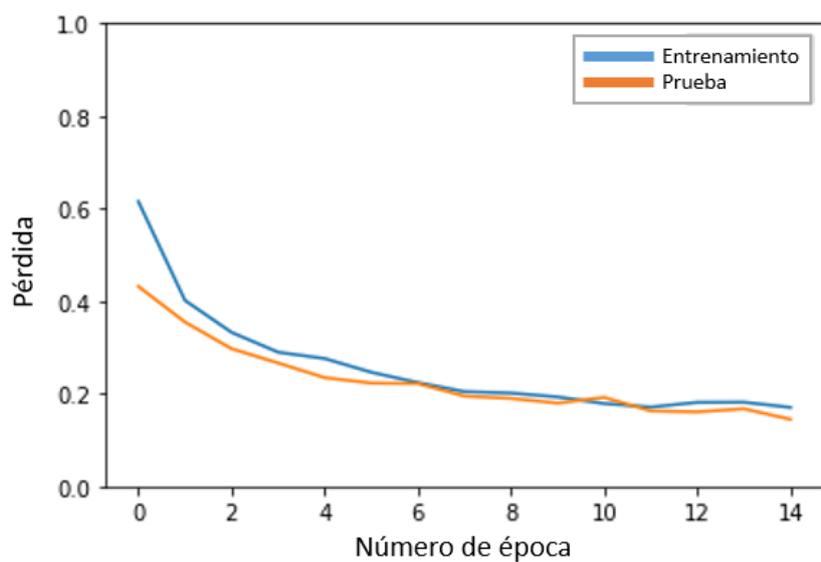


(a) Curva de pérdida

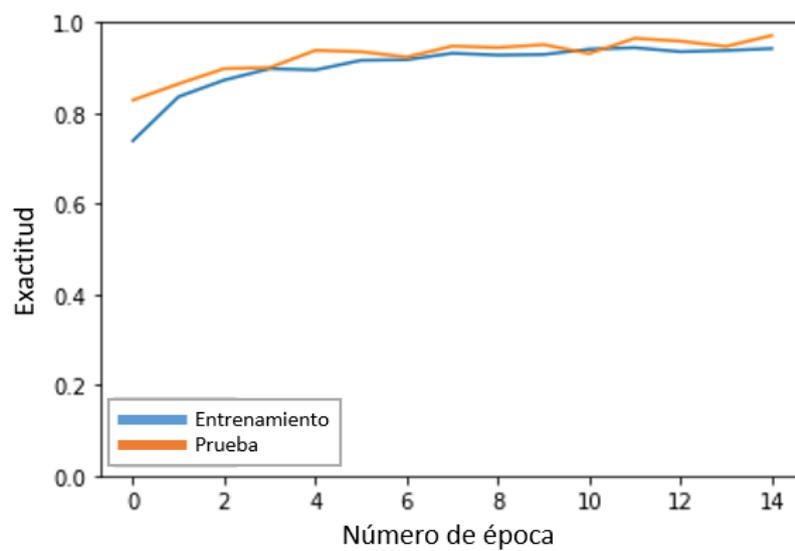


(b) Curva de exactitud

Figura 4.4: DenseNet 201



(a) Curva de pérdida



(b) Curva de exactitud

Figura 4.5: VGG-19

4.4. Comparativa entre los métodos de clasificación de aprendizaje máquina y aprendizaje profundo seleccionados

Los valores resultantes de la evaluación de las métricas, enumeradas en la Sección 2.5, sobre los métodos de aprendizaje máquina y aprendizaje profundo se presentan comparados en la Tabla 4.5. Para los valores mostrados en la parte de aprendizaje máquina de esta Tabla, se eligieron los correspondientes al método PCA; debido a que el subconjunto de características seleccionado por este método fue el que obtuvo uno de los mejores desempeños, como se puede ver en las Tablas 4.1 y 4.2.

En la Tabla de comparativa 4.5, se observan los cuatro métodos de aprendizaje máquina: Naive Bayes (NB), Máquina de Soporte Vectorial (SVM), K-Vecinos más Cercanos (K-NN) y Bosques Aleatorios (RF) comparados con las dos arquitecturas de los métodos de aprendizaje profundo: DenseNet 201 y VGG-19. La comparativa es realizada en las tres métricas evaluadas: exactitud, precisión y recall.

Tabla 4.5: Comparativa entre los métodos de aprendizaje máquina y aprendizaje profundo para la clasificación de las imágenes de LUS de pacientes con COVID-19, neumonía o sin afecciones

Métrica	Aprendizaje máquina Métodos				Aprendizaje profundo Métodos	
	<i>NB</i>	<i>SVM</i>	<i>K-NN</i>	RF	<i>DenseNet 201</i>	<i>VGG-19</i>
Exactitud	0.77	0.93	0.96	0.98	0.97	0.97
Precision - Recall						
<i>COVID-19</i>	0.75 - 0.77	0.95 - 0.88	0.96 - 0.97	0.98 - 0.97	1.00 - 1.00	0.91 - 1.00
<i>Neumonía</i>	0.81 - 0.69	0.97 - 0.94	0.98 - 0.94	0.98 - 0.97	1.00 - 0.93	1.00 - 1.00
<i>Regular</i>	0.75 - 0.84	0.88 - 0.96	0.96 - 0.99	0.97 - 0.98	0.91 - 1.00	1.00 - 0.83

La comparación de todos los métodos muestra que el método de Random Forest presenta el mejor desempeño en cuanto a exactitud y valores altos en precisión y recall, mostrando diferencias de ± 0.01 aproximadamente entre estos valores. En comparativa, el desempeño del resto de los métodos varía con mayores diferencias, mostrando que pueden clasificar de mejor manera algunas de estas clases pero son más susceptibles a fallar en otras.

CAPÍTULO

5

CONCLUSIONES

A lo largo de este trabajo se ha expuesto cómo los distintos métodos tanto de aprendizaje máquina como de aprendizaje profundo pueden resolver una tarea de clasificación de tres clases aplicada a imágenes médicas, en específico de ultrasonido pulmonar; demostrando la utilidad de estas técnicas para análisis y clasificación de una gran cantidad de información en beneficio del apoyo a los médicos, ofreciendo apoyo en tareas rutinarias, herramientas para mejorar el diagnóstico y con ello una disminución en los tiempos empleados.

Este trabajo recopila una base de datos de más de 2,000 imágenes por clase de ultrasonido pulmonar en pacientes con COVID-19, neumonía y sin afecciones; siendo bastante amplia para la cantidad de imágenes que existen con estas enfermedades en ultrasonido pulmonar en la actualidad. Esto está relacionado a, como se mencionó con anterioridad, el tipo de imagen principalmente empleado para COVID-19 es la TC y los rayos X.

En los resultados se muestra la capacidad de los métodos de aprendizaje máquina y de aprendizaje profundo de completar exitosamente la tarea de clasificación de tres clases, sin embargo los métodos de aprendizaje máquina también demuestran tener varias ventajas sobre los de aprendizaje profundo; como el requerir menos poder computacional para el entrenamiento de los algoritmos, haciéndolos más fáciles de entrenar y haciendo posible obtener buenos resultados con una base de datos mucho menor en comparación con

los métodos de aprendizaje profundo.

Siendo el caso de este trabajo, donde el mejor desempeño para la resolución de esta tarea fue logrado por el método de bosques aleatorios en el conjunto de características seleccionadas por análisis de componentes principales.

A su vez puede observarse que no es necesario contar con un gran número de parámetros para entrenar los métodos de aprendizaje máquina, sino que realizando una selección de aquellos con mayor importancia pueden obtenerse valores altos en las métricas evaluadas, lo que significaría menos tiempo de entrenamiento y poder computacional requerido.

El conocer las imágenes con las que se trabaja, entender cuáles son sus características principales y qué es lo que ese está buscando es de vital importancia ya que permite realizar una buena selección de los métodos de extracción requeridos en la tarea deseada. Para los métodos de aprendizaje máquina comparados en este trabajo se empleó el mismo conjunto de características previamente extraídas; siendo estas de textura, contraste e intensidad correspondientes al ultrasonido pulmonar.

La metodología empleada para la resolución de este problema sirve como una base para clasificaciones y detecciones sobre otro tipo de imágenes médicas, o bien, en otro tipo de enfermedades; ya que se han expuesto herramientas tanto de extracción de características, selección de las mismas y distintos tipos de clasificadores.

Como perspectiva a futuro de este trabajo se plantea el uso de estos algoritmos en imágenes de tomografía computarizada y rayos X, para realizar el mismo de tipo de clasificación entre COVID-19, neumonía y pacientes sin afecciones. Además de re-entrenar estos mismos métodos con imágenes de ultrasonido pulmonar segmentado para observar las diferencias que podría aportar el trabajar solamente sobre una región de interés, así como también identificar el área y características específicas que están teniendo un mejor y mayor aporte en las imágenes.

ÍNDICE DE FIGURAS

2.1.	Principio básico de la adquisición de una imagen de ultrasonido, se observa como la onda mecánica es emitida, reflejada por el tejido y detectada de nuevo por el transductor [17] . . .	7
2.2.	Líneas A mostradas en ultrasonido pulmonar [26]	8
2.3.	Líneas B mostradas en ultrasonido pulmonar [26]	9
2.4.	Consolidaciones mostradas en ultrasonido pulmonar [26] . . .	10
3.1.	Diagrama de bloques del proceso de clasificación de imágenes de LUS empleando métodos de extracción y selección de características para entrenar métodos de aprendizaje máquina y evaluar sus resultados con distintas métricas	21
3.2.	Diagrama de bloques del proceso seguido para aprendizaje profundo donde se probaron ocho arquitecturas y mediante la comparativa de las métricas obtenidas seleccionar tres de mejor desempeño	23
3.3.	Diagrama de bloques del proceso completo	24
4.1.	Gráfica comparativa de exactitud entre los métodos de ML y los métodos selección de características para la clasificación de las imágenes de LUS de pacientes con COVID-19, neumonía o sin afecciones	26

Índice de figuras

4.2. Comparación de precisión entre los distintos métodos de aprendizaje máquina y los métodos selección de características para la clasificación de las imágenes de LUS de pacientes con COVID-19, neumonía o sin afecciones	29
4.3. Comparación de recall entre los distintos métodos de aprendizaje máquina y los métodos de selección de características para la clasificación de las imágenes de LUS de pacientes con COVID-19, neumonía o sin afecciones	30
4.4. DenseNet 201	33
4.5. VGG-19	34
B.1. VGG-13	52
B.2. VGG-19	53
B.3. ResNet 18	54
B.4. ResNet 50	55
B.5. ResNet 152	56
B.6. DenseNet 201	57
B.7. MobileNet V2	58
B.8. GoogleNet	59

ÍNDICE DE TABLAS

4.1. Comparativa de exactitud entre los métodos de ML y los métodos selección de características para la clasificación de las imágenes de LUS de pacientes con COVID-19, neumonía o sin afecciones	26
4.2. Comparativa entre Precisión (P) y Recall (R) de los métodos de aprendizaje máquina y selección de características para la clasificación de las imágenes de LUS de pacientes con COVID-19, neumonía o sin afecciones	27
4.3. Comparativa de las métricas evaluadas entre todas las redes neuronales convolucionales para la clasificación de las imágenes de LUS de pacientes con COVID-19, neumonía o sin afecciones	31
4.4. Comparativa entre los métodos de aprendizaje profundo para la clasificación de las imágenes de LUS de pacientes con COVID-19, neumonía o sin afecciones	32
4.5. Comparativa entre los métodos de aprendizaje máquina y aprendizaje profundo para la clasificación de las imágenes de LUS de pacientes con COVID-19, neumonía o sin afecciones	35

APÉNDICE

A

ARTÍCULO

COVID-19 detection from lung ultrasound images

Melisa Mateu^a, Jimena Olveres^{b,c}, and Boris Escalante-Ramírez^{b,c}

^aPosgrado en Ingeniería, Universidad Nacional Autónoma de México, Mexico City, Mexico

^bFacultad de Ingeniería, Universidad Nacional Autónoma de México, Mexico City, Mexico

^cCentro de Estudios en Computación Avanzada, Universidad Nacional Autónoma de México, Mexico City, Mexico

ABSTRACT

Early-stage detection of Coronavirus Disease 2019 (COVID-19) is crucial for patient medical attention. Since lungs are the most affected organs, monitoring them constantly is an effective way to observe sickness evolution. The most common technique for lung-imaging and evaluation is Computed Tomography (CT). However, its costs and effects over human health has made Lung Ultrasound (LUS) a good alternative. LUS does not expose the patient to radiation and minimizes the risk of contamination. Also, there is evidence of a relation between different artifacts on LUS and lung's diseases coming from the pleura, whose abnormalities are related with most acute respiratory disorders. However, LUS often requires an expert clinical interpretation that may increase diagnosis time or decrease diagnosis performance. This paper describes and compares machine learning classification methods namely Naive Bayes (NB) Support Vector Machine (SVM), K-Nearest Neighbor (K-NN) and Random Forest (RF) over several LUS images. They obtain a classification between lung images with COVID-19, pneumonia, and healthy patients, using image's features previously extracted from Gray Level Co-Occurrence Matrix (GLCM) and histogram's statistics. Furthermore, this paper compares the above classic methods with different Convolutional Neural Networks (CNN) that classifies the images in order to identify these lung's diseases.

Keywords: Machine learning, deep learning, comparison, metrics.

1. INTRODUCTION

COVID-19 is an infectious disease caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-Cov-2). It is characterized by severe pneumonia or acute respiratory distress syndrome (ARDS) in about 20% of infected patients with high morbidity and mortality.^{1,2} Radiological imaging techniques such as thoracic computed tomography (CT) is used for early screening, diagnosis and treatment of patients with suspected or confirmed COVID-19 infections.³ Although CT has excellent ability to detect lung's changes, it has disadvantages of large radiation exposure, lack of portability for bedside imaging and risk of cross infections between patients.¹

On the other hand, lung ultrasound (LUS) is an emerging non-invasive bedside technique in the diagnosis of interstitial lung syndrome. LUS has provided the physical bases and patterns in COVID-19 patients, suggesting that it can be a useful tool to diagnose and monitor this sickness.^{2,4} The main LUS findings in COVID-19 are B lines, which are represented by vertical hyperechoic artifacts that depart from the pleura to the bottom of the screen,^{5,6} as shown in Fig. 1b. The presence of B lines suggests ARDS and can be related to COVID-19. Besides B lines, LUS has more artifacts that are related to the majority of acute respiratory disorders because they involve the pleura. For these reasons some artifacts can be used to detect potential lung's conditions. For instance A Lines can be related to healthy patients. These type of lines are motionless and regularly spaced lines horizontal to the pleura,^{2,6} as shown in Fig. 1a. In addition, consolidations are another artifact that can be related to pneumonia.⁶ The echo structure of the lung itself becomes visible with characteristic air inside the alveolus or surrounded by inflammation or pus,¹ as shown in Fig. 1c.

These different LUS artifacts allow to study medical images in order to identify possible patterns that may lead to the automatic diagnosis of the disease. Machine learning (ML) and deep learning (DL) have become

Further author information: (Send correspondence to M.M.)

M.M.: E-mail: mateu.melisa@gmail.com

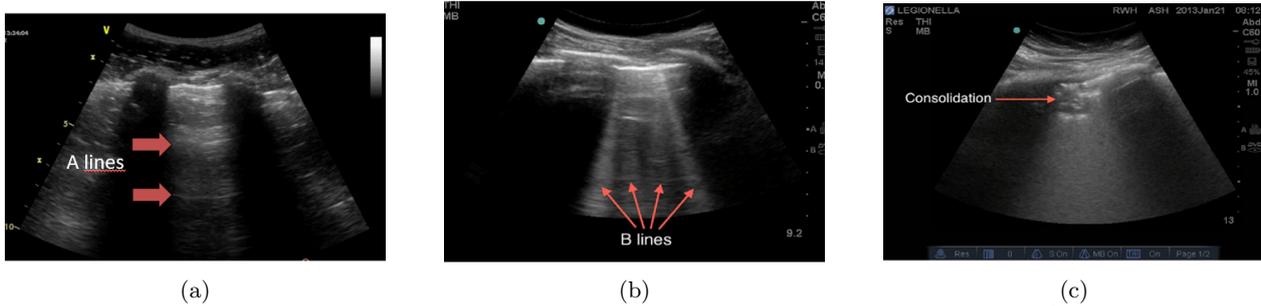


Figure 1: Artifacts of LUS: (a) A lines, (b) B lines and (c) consolidations¹

established disciplines in applying artificial intelligence to mine, analyze, and recognize patterns from data. The advances of those techniques can benefit clinical decision making. As new data emerges, the use of computer-aided systems is becoming increasingly applied in clinical settings.⁷

The purpose of this paper is to compare the supervised machine learning classification methods trained by several LUS images. By using the images features extracted from Gray Level Co-Occurrence Matrix (GLCM) and histogram's statistics these algorithms can classify the lung images of patients with COVID-19, pneumonia, and healthy ones. These ML methods were compared with different Convolutional Neural Networks (CNN) that also classified the same images in order to identify lung diseases. Lastly, this paper evaluates ML and DL methods using metrics like accuracy, precision and recall to identify the best method for the classification task.

2. METHODS

2.1 Dataset

The LUS image database was constructed from two different datasets. Firstly from a Kaggle's repository⁸ with 988 images of COVID-19, 731 images of pneumonia and 1,276 images of patients with no affections. The second dataset was collected from the Butterfly's COVID-19 gallery,⁹ it consisted of LUS videos; eight of COVID-19, nine of pneumonia and two of normal lungs. After we extracted the frames from each video, the dataset consisted of 1,562 COVID-19 images, 1,705 pneumonia's images and 832 images of normal lungs. A random selection of 700 images per class; COVID-19, pneumonia and regular patients, was generated from the combination of both datasets. These 2,100 images were used for the training and evaluation of machine and deep learning methods.

2.2 Features

The machine learning algorithms consisted of several features extracted from the images on the database. The images were first resized to a resolution of 254×254 pixels, then the features were extracted and finally different feature selection methods were tested when training the ML algorithms.

2.2.1 Extraction

The features extracted for each image consisted of GLCM and histogram's statistics. For GLCM 25 patches were extracted per image and to each patch six statistics were calculated; dissimilarity, correlation, energy, contrast, homogeneity and angular second moment (ASM). In total one image had 150 textural features. In addition, there were five statistics calculated per histogram: mean, standard deviation, entropy, skewness and kurtosis. Therefore, each image had a total of 155 features used to classify them.

2.2.2 Selection

Due to the large amount of features per image, feature selection methods were used to determine the most important features for the ML classification task. Univariate feature selection method picked ten features, the same number of features was selected by the principal component analysis (PCA) and six features were retained with the feature importance method.

2.3 Machine Learning

We trained four classic supervised methods: Naive Bayes (NB), Support Vector Machine (SVM), K-Nearest Neighbor (K-NN) and Random Forest (RF). All 155 features and also the sets of reduced ones, mentioned in the Sec. 2.2.2, were used separately to train these ML algorithms for a three-classes classification task. To evaluate these method's performance, the metrics from Sec. 2.5 were calculate for each one of them. The results are shown in Sec. 3.

2.4 Deep Learning

Using the same resolution as in Sec. 2.2 three CNN's were trained. Transfer learning techniques were used to learn the classification task of COVID-19, pneumonia and regular images. Adam optimizer was used with a learning rate of $1E - 3$ for each CNN. Cross Entropy was used as the loss function.

The trained CNN were ResNet18, MobilNet V2 and GoogleNet. They were evaluated with the metrics from Sec. 2.5 and the results are shown in Sec. 3.

2.5 Metrics

To evaluate categorization predictions in deep and machine learning methods, three types of metrics were computed. The accuracy determines how correct the diseases classifications were, precision shows how many of the predictions were correct and recall indicates how correctly the disease was recognized.

The values obtained and their comparison are shown in the tables 1 - 4 on the Sec. 3.

3. RESULTS

3.1 Machine Learning

For accuracy training results, different values can be observed depending on the machine learning method used, as shown in Fig. 2 and in table 1.

Also accuracy values are the same from almost all the methods independently of the feature selection technique that was used. The only difference between values comes with the Random Forest method but the difference is not significant. The RF method achieved the highest accuracy value compared to the other ML methods.

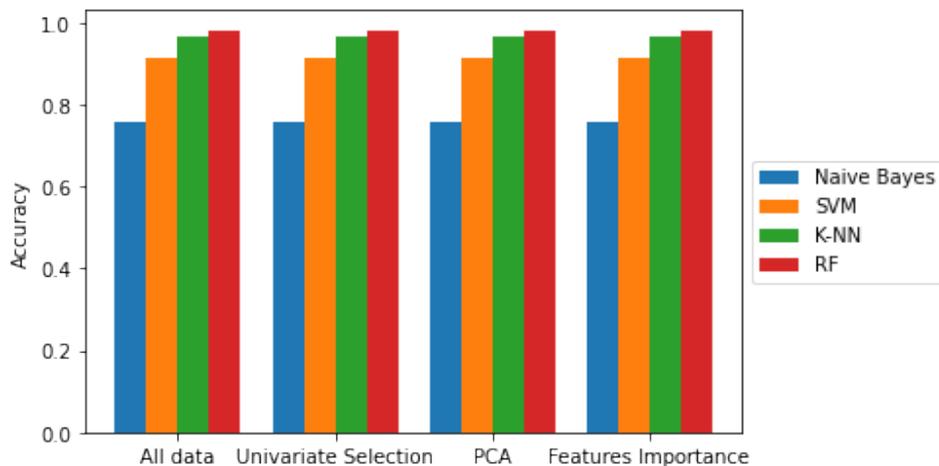


Figure 2: Accuracy comparison between the different methods of machine learning and feature selection

Figures 3 and 4 show a comparison between precision and recall values obtained for every class. These values are shown in table 2

Table 1: Accuracy comparison between ML and feature selection methods

Machine Learning Methods	Features - Accuracy			
	All data	Univariate Selection	PCA	Feature Importance
<i>Naive Bayes</i>	75.67	75.67	75.67	75.67
<i>SVM</i>	91.38	91.38	91.38	91.38
<i>K-NN</i>	96.62	96.62	96.62	96.62
<i>Random Forest</i>	98.10	98.24	98.24	98.19

Table 2: Precision (P) and Recall (R) comparison between ML and features selection methods

Class	All data				Class	Univariate Selection			
	<i>NB</i>	<i>SVM</i>	<i>K-NN</i>	<i>RF</i>		<i>NB</i>	<i>SVM</i>	<i>K-NN</i>	<i>RF</i>
<i>COVID-19</i>	P: 0.72 R: 0.76	P: 0.89 R: 0.89	P: 0.95 R: 0.97	P: 0.98 R: 0.98	<i>COVID-19</i>	P: 0.72 R: 0.76	P: 0.89 R: 0.89	P: 0.95 R: 0.97	P: 0.97 R: 0.99
<i>Pneumonia</i>	P: 0.83 R: 0.66	P: 0.96 R: 0.90	P: 0.99 R: 0.94	P: 0.98 R: 0.97	<i>Pneumonia</i>	P: 0.83 R: 0.66	P: 0.96 R: 0.90	P: 0.99 R: 0.94	P: 0.98 R: 0.97
<i>Regular</i>	P: 0.73 R: 0.84	P: 0.88 R: 0.94	P: 0.95 R: 0.98	P: 0.98 R: 0.97	<i>Regular</i>	P: 0.73 R: 0.84	P: 0.88 R: 0.94	P: 0.95 R: 0.98	P: 0.97 R: 0.98
Class	PCA				Class	Feature Importance			
	<i>NB</i>	<i>SVM</i>	<i>K-NN</i>	<i>RF</i>		<i>NB</i>	<i>SVM</i>	<i>K-NN</i>	<i>RF</i>
<i>COVID-19</i>	P: 0.79 R: 0.76	P: 0.89 R: 0.89	P: 0.95 R: 0.97	P: 0.98 R: 0.98	<i>COVID-19</i>	P: 0.72 R: 0.76	P: 0.89 R: 0.89	P: 0.95 R: 0.97	P: 0.98 R: 0.99
<i>Pneumonia</i>	P: 0.83 R: 0.66	P: 0.96 R: 0.90	P: 0.99 R: 0.94	P: 0.98 R: 0.97	<i>Pneumonia</i>	P: 0.83 R: 0.66	P: 0.96 R: 0.90	P: 0.99 R: 0.94	P: 0.98 R: 0.97
<i>Regular</i>	P: 0.73 R: 0.84	P: 0.88 R: 0.94	P: 0.95 R: 0.98	P: 0.98 R: 0.98	<i>Regular</i>	P: 0.73 R: 0.84	P: 0.88 R: 0.94	P: 0.95 R: 0.98	P: 0.97 R: 0.97

According to the comparisons the best ML method in all the feature selection methods was RF, it shows higher values than NB, SVM and K-NN in all the cases. Taking RF as the best ML method we can observe that the highest precision and recall was achieved with the PCA method. Thus, considering all metrics (accuracy, recall and precision) the best performance is PCA - Random Forest combination.

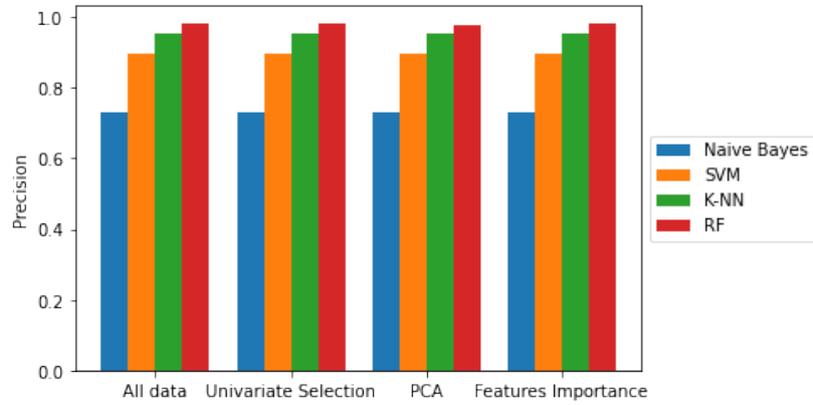
3.2 Deep Learning

Table 3 shows a comparison between the DL methods trained. All three displayed satisfactory accuracy above 0.90. In precision and recall the ResNet 18 architecture showed the worst performance compared to the other two. Although MobilNet V2 and GoogleNet have the same accuracy value, their precision and recall values vary. The best performance of the three metrics was achieved by Google Net architecture.

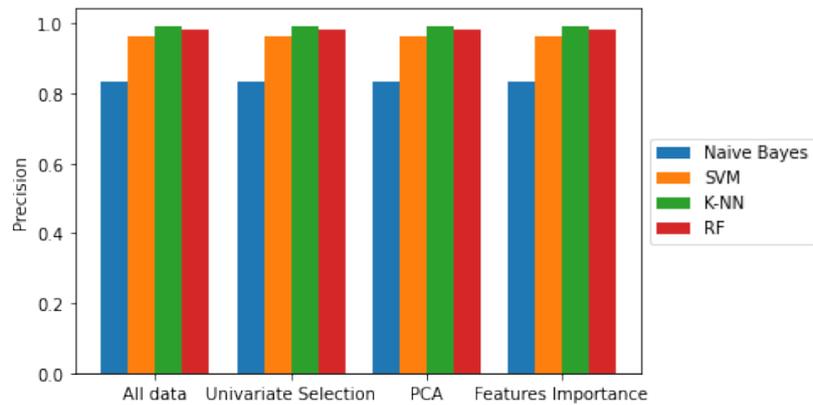
Table 3: Comparison between DL methods

Metric	CNN architecture		
	<i>ResNet 18</i>	<i>MobilNet V2</i>	<i>GoogleNet</i>
Accuracy	0.91	0.94	0.94
Precision - Recall			
<i>COVID-19</i>	0.89 - 1.00	1.00 - 0.92	1.00 - 1.00
<i>Pneumonia</i>	0.93 - 0.93	1.00 - 0.88	0.93 - 0.93
<i>Regular</i>	0.89 - 0.80	0.85 - 1.00	0.90 - 0.90

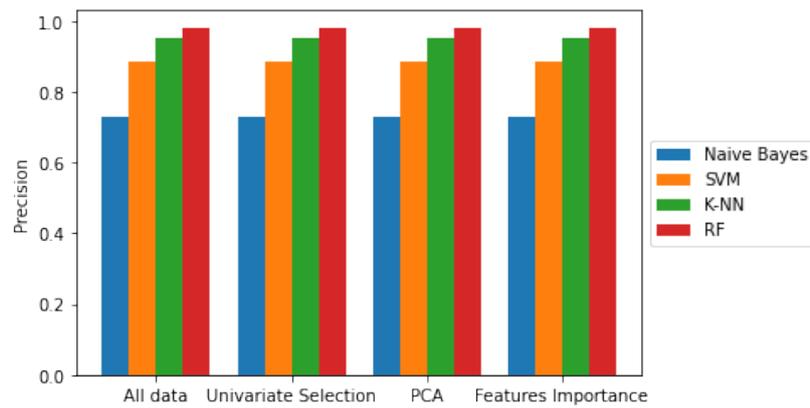
The loss and accuracy curves are shown in Figs. 5, 6 and 7. All loss curves coincide in a fast decay and



(a) COVID-19

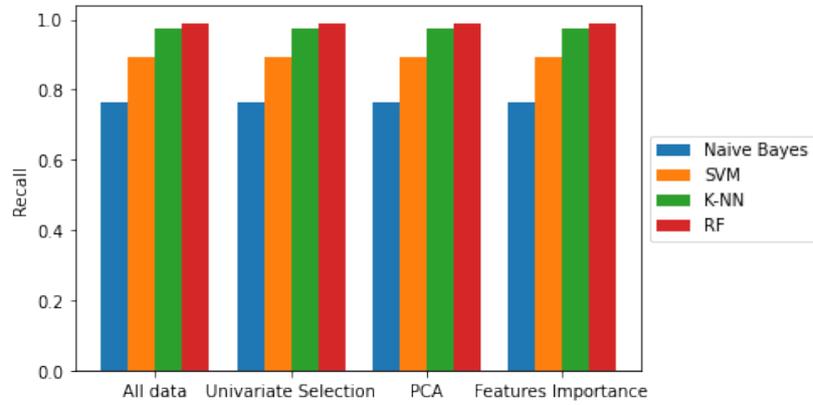


(b) Pneumonia

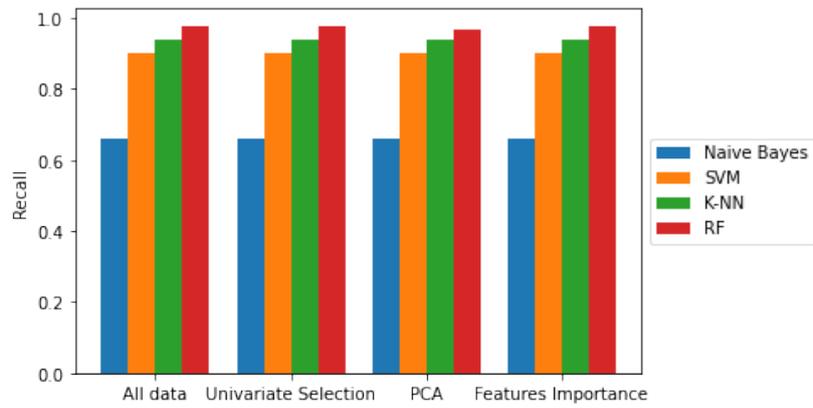


(c) Regular

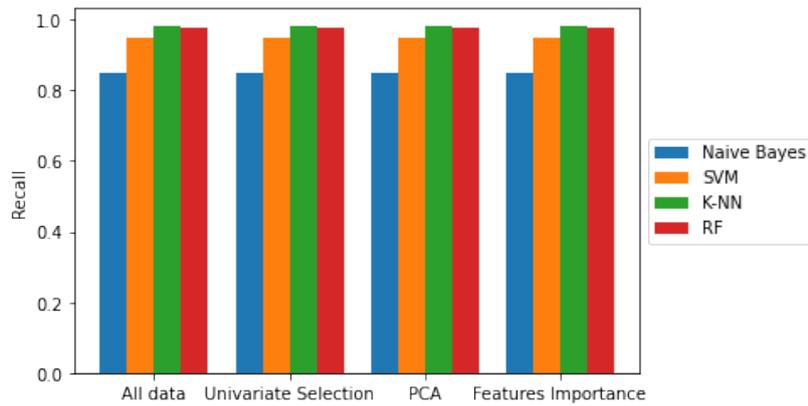
Figure 3: Precision comparison between the different methods of ML and feature selection



(a) COVID-19

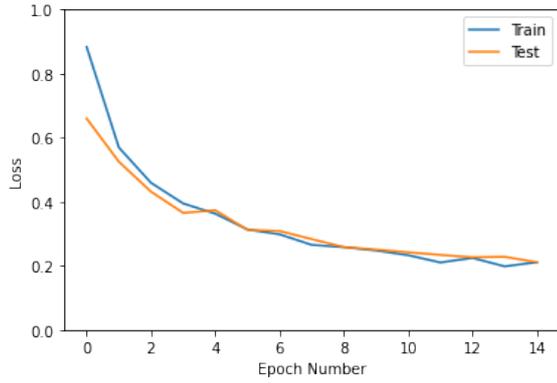


(b) Pneumonia

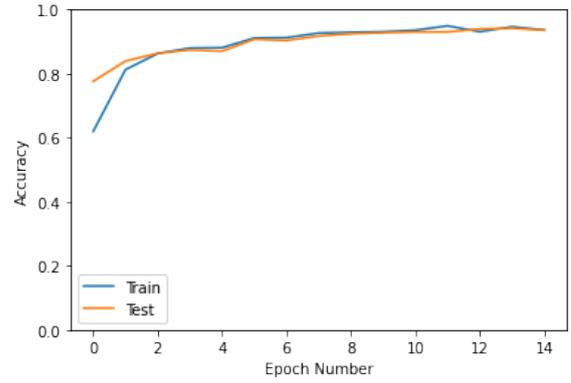


(c) Regular

Figure 4: Recall comparison between the different methods of ML and feature selection

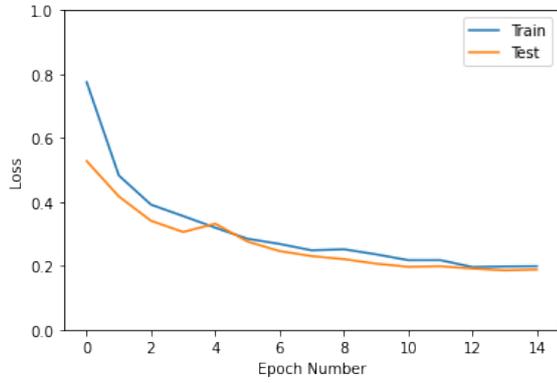


(a)

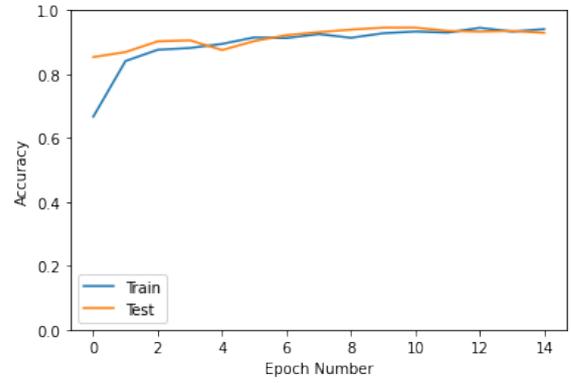


(b)

Figure 5: ResNet 18, (a) loss curve and (b) accuracy



(a)



(b)

Figure 6: MobilNet V2, (a) loss curve and (b) accuracy

converge on the training and testing sets. Also, the accuracy curves show that high levels of accuracy are achieved in very few epochs. Their performances, in loss and accuracy, suggest that this classification task is easily achieved for the DL methods.

3.3 Comparison

In table 4 all metrics for ML and DL methods obtained are compared. The ML method selected for the comparison corresponds to the PCA feature selection since it showed the highest performance in the metrics evaluated as seen in tables 1 and 2.

Table 4: Comparison between DL and ML methods

Metric	Machine Learning Methods				Deep Learning Methods		
	<i>NB</i>	<i>SVM</i>	<i>K-NN</i>	RF	<i>ResNet18</i>	<i>MobilNet V2</i>	<i>GoogleNet</i>
Accuracy	0.75	0.91	0.96	0.98	0.91	0.94	0.94
Precision - Recall							
<i>COVID-19</i>	0.72 - 0.76	0.89 - 0.89	0.95 - 0.97	0.98 - 0.98	0.89 - 1.00	1.00 - 0.92	1.00 - 1.00
<i>Pneumonia</i>	0.83 - 0.66	0.96 - 0.90	0.99 - 0.94	0.98 - 0.97	0.93 - 0.93	1.00 - 0.88	0.93 - 0.93
<i>Regular</i>	0.73 - 0.84	0.88 - 0.94	0.95 - 0.98	0.98 - 0.98	0.89 - 0.80	0.85 - 1.00	0.90 - 0.90

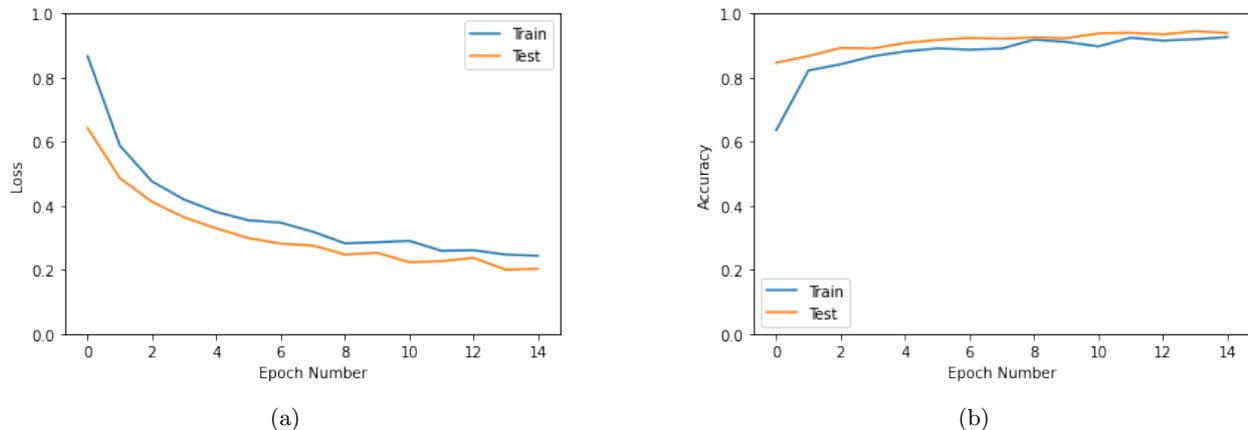


Figure 7: GoogleNet, (a) loss curve and (b) accuracy

The all methods comparison table (table 4) evidenced that Random Forest presents the best accuracy performance and very high precision and recall values, showing little differences of ± 0.01 between them. In comparative, the remaining method's performance values vary with bigger differences, showing they can classify better some classes but are more susceptible to fail in others.

4. CONCLUSION

As the results show, ML and DL were able to successfully complete the three-classes classification task. However, ML methods have several advantages over DL ones. First, less computational power is required to train ML algorithms. For DL methods a GPU is needed whereas only a CPU is required on ML methods. Additionally, less data is required to train ML methods. ML algorithms were able to achieve good metrics with less LUS images than the DL algorithms.

In brief, artificial intelligence methods can be used to analyze and classify large amounts of medical information which can be useful as it can help doctors with diagnosis or treatments. This study has demonstrated that ML and DL methods are useful tools in a clinical setting. The approach for classifying LUS images of patients with COVID-19, pneumonia and without affections can be performed with a variety of machine learning and deep learning methods. Despite both methods achieved the task successfully, ML methods are easier to train as they require less computational power and can obtain good results with a smaller database.

ACKNOWLEDGMENTS

This work has been sponsored by UNAM PAPIIT grants TA101121, IV100420 and Consejo Nacional de Ciencia y Tecnología (CONACYT)

REFERENCES

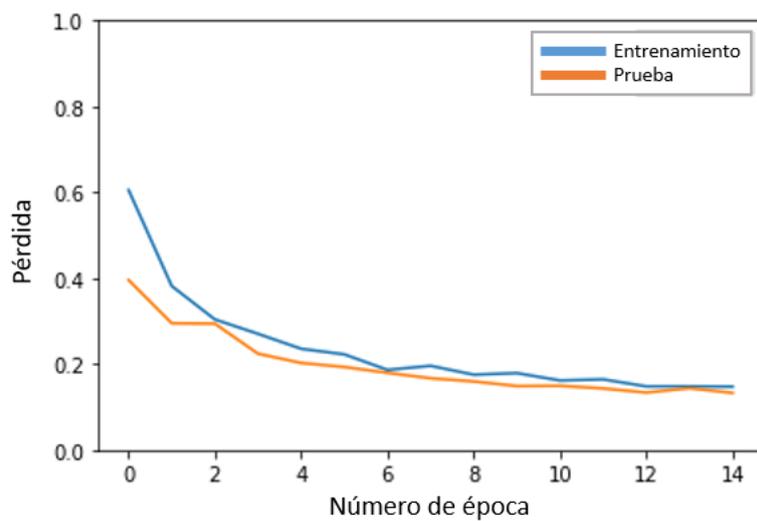
- [1] Sultan, L. R. and Sehgal, C. M., "A review of early experience in lung ultrasound in the diagnosis and management of covid-19," *Ultrasound in Medicine and Biology* **46**, 2530–2545 (9 2020).
- [2] Allinovi, M., Parise, A., Giacalone, M., Amerio, A., Delsante, M., Odone, A., Franci, A., Gigliotti, F., Amadasi, S., Delmonte, D., Parri, N., and Mangia, A., "Lung ultrasound may support diagnosis and monitoring of covid-19 pneumonia," *Ultrasound in Medicine and Biology* **46**, 2908–2917 (11 2020).
- [3] Bakheet, S. and Al-Hamadi, A., "Automatic detection of covid-19 using pruned glcm-based texture features and lcrf classification," *Computers in Biology and Medicine* **137**, 104781 (10 2021).
- [4] Musolino, A. M., Supino, M. C., Buonsenso, D., Ferro, V., Valentini, P., Magistrelli, A., Lombardi, M. H., Romani, L., D'Argenio, P., and Campana, A., "Lung ultrasound in children with covid-19: Preliminary findings," *Ultrasound in Medicine and Biology* **46**, 2094–2098 (8 2020).

- [5] Sofia, S., Boccatonda, A., Montanari, M., Spampinato, M., Cocco, G., Accogli, E., Cipollone, F., and Schiavone, C., “Thoracic ultrasound and sars-covid-19: a pictorial essay,” *Journal of Ultrasound* **23**, 217–221 (2020). ARDS - Síndrome de dificultad respiratoria aguda.
- [6] Serafino, M. D., Notaro, M., Rea, G., Iacobellis, F., Paoli, V. D., Acampora, C., Ianniello, S., Brunese, L., Romano, L., and Vallone, G., “The lung ultrasound: facts or artifacts? in the era of covid-19 outbreak,” *Radiologia Medica* **125**, 738–753 (8 2020).
- [7] Apostolopoulos, I. D. and Mpesiana, T. A., “Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks,” **43**, 635–640 (2020).
- [8] “covid19 ultrasound original dataset.” (Accessed: 8 june 2021).
- [9] “Covid-19 lung ultrasound imaging: Fighting together.” (Accessed: 16 february 2022).

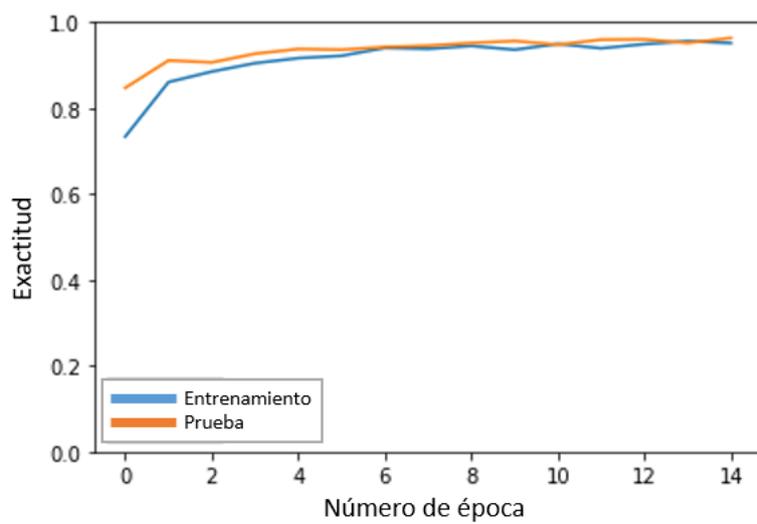
APÉNDICE

B

CURVAS DE APRENDIZAJE
PROFUNDO

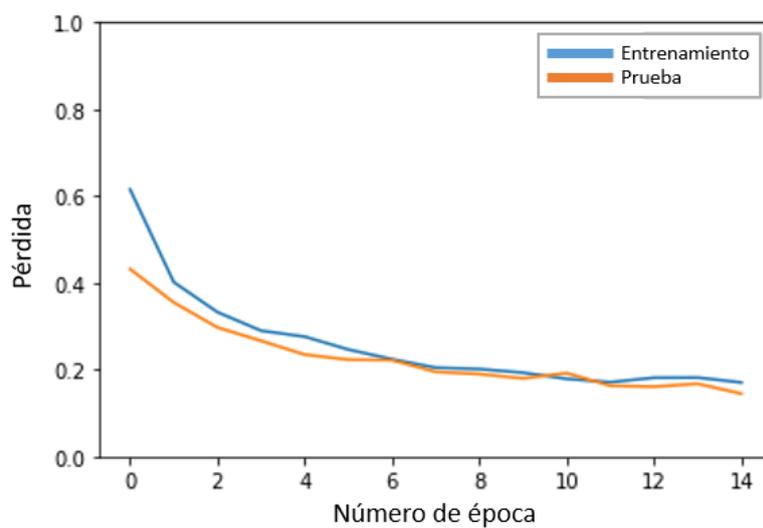


(a) Curva de pérdida

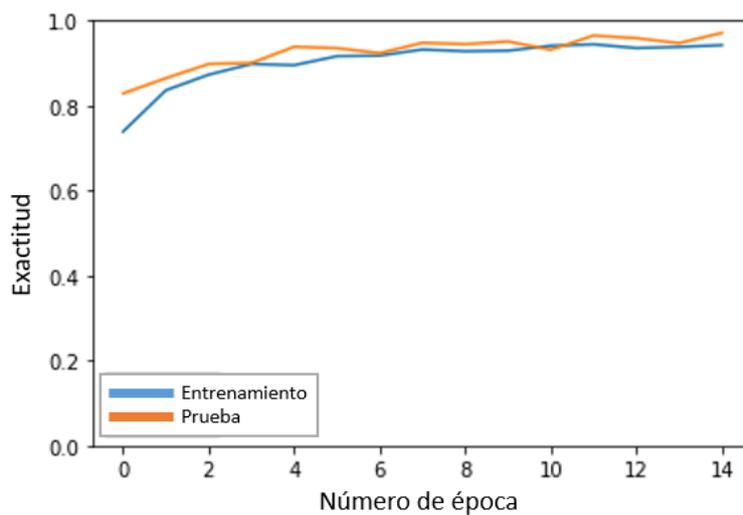


(b) Curva de exactitud

Figura B.1: VGG-13

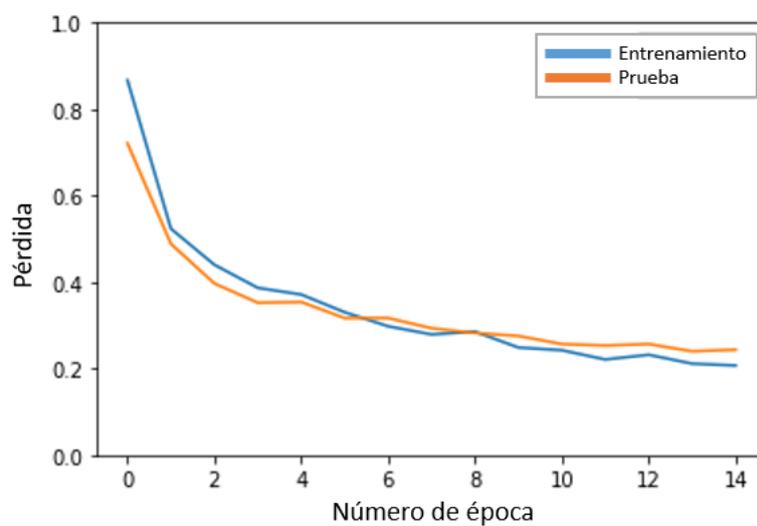


(a) Curva de pérdida

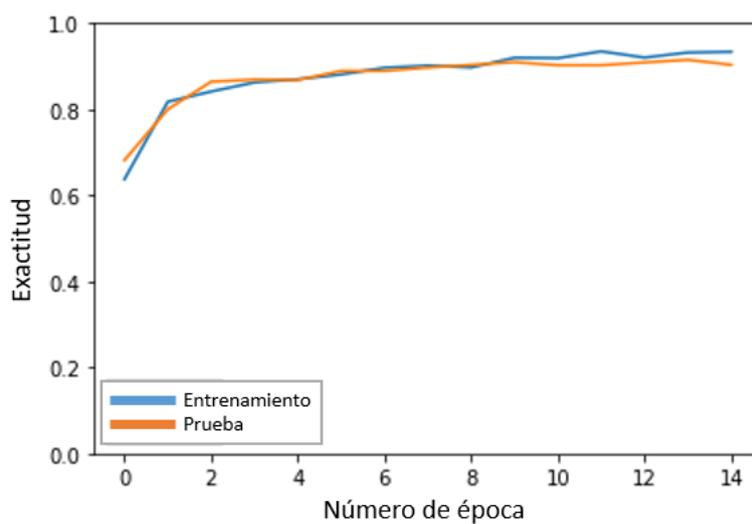


(b) Curva de exactitud

Figura B.2: VGG-19

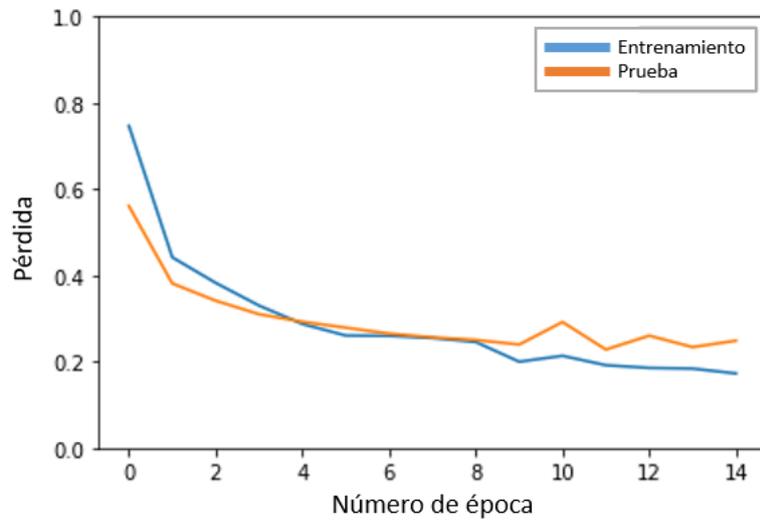


(a) Curva de pérdida

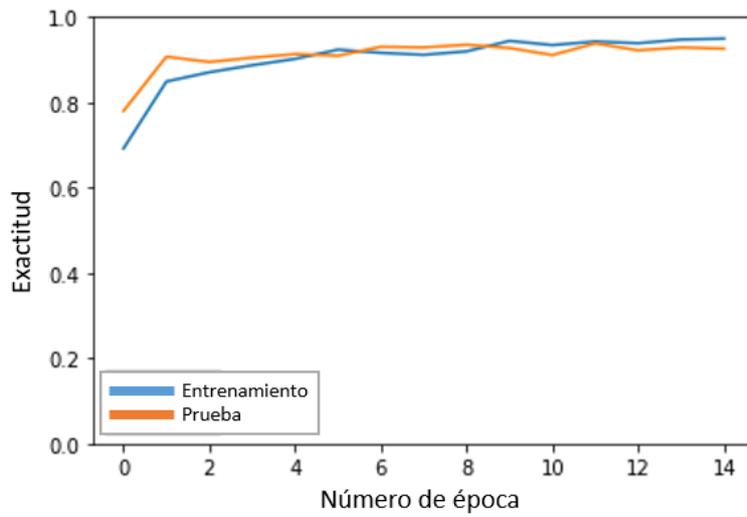


(b) Curva de exactitud

Figura B.3: ResNet 18

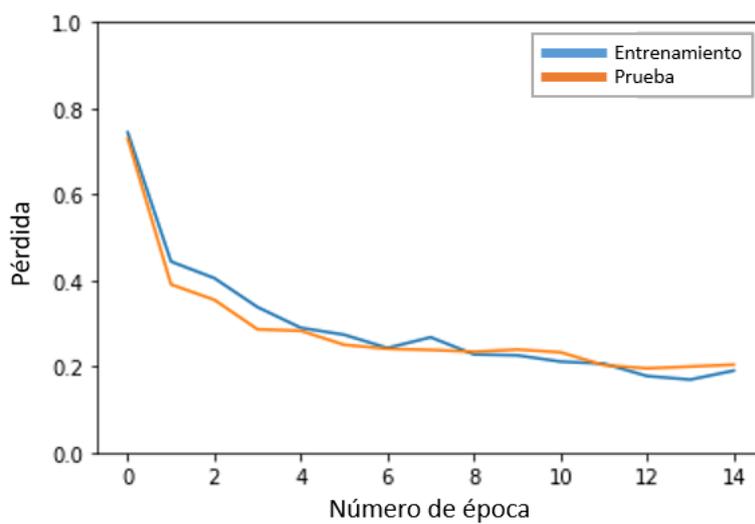


(a) Curva de pérdida

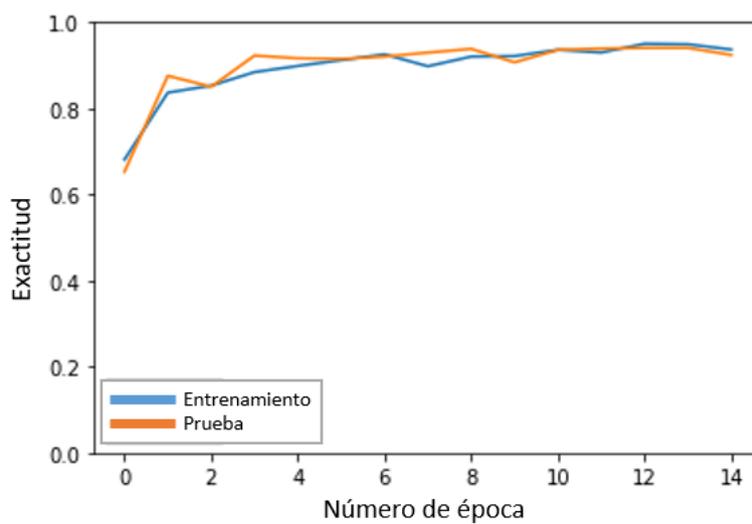


(b) Curva de exactitud

Figura B.4: ResNet 50

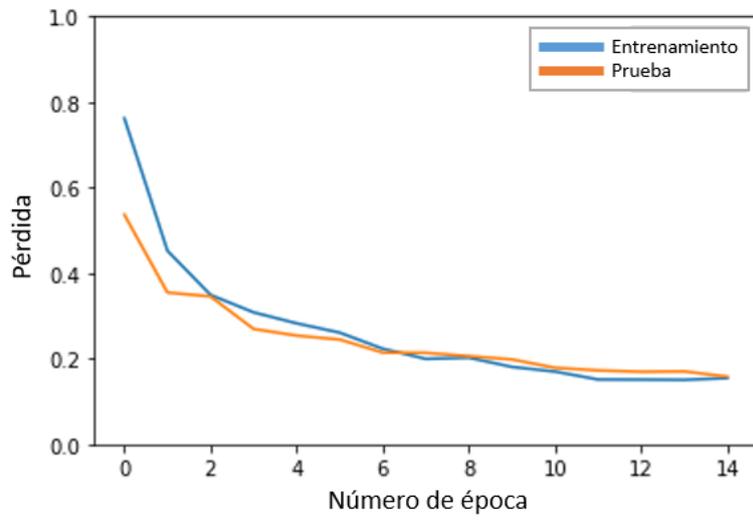


(a) Curva de pérdida

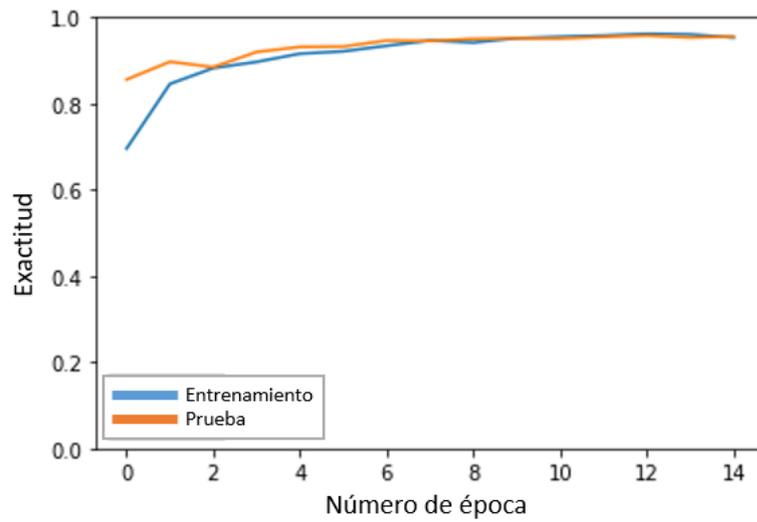


(b) Curva de exactitud

Figura B.5: ResNet 152

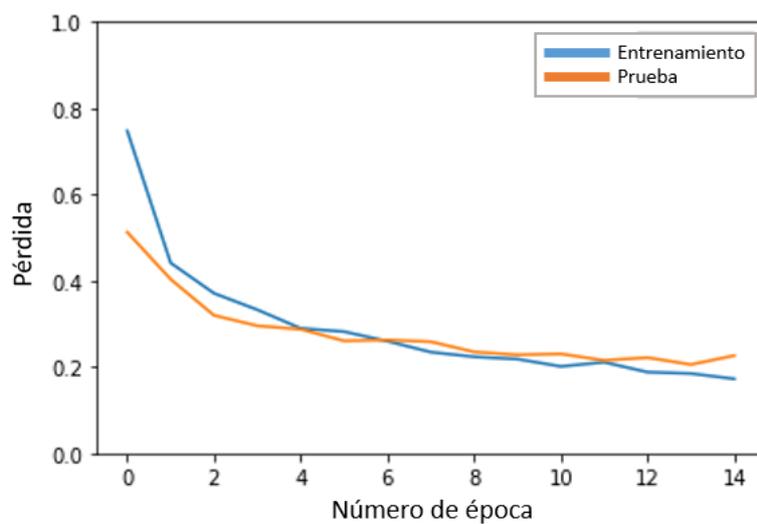


(a) Curva de pérdida

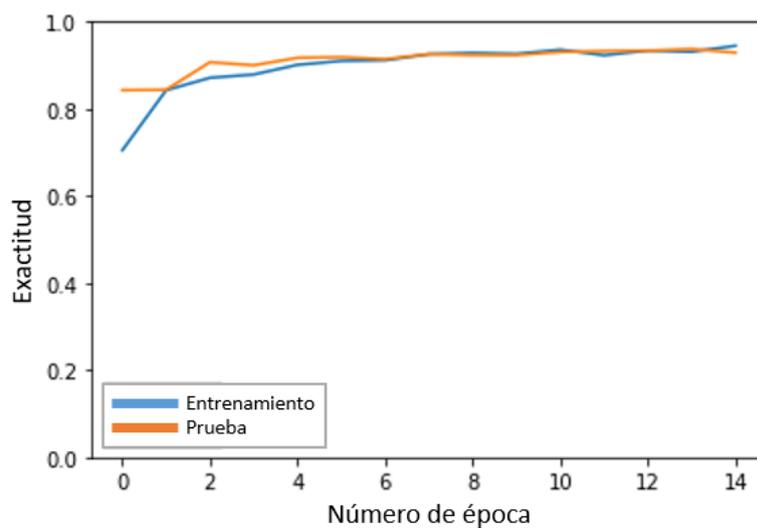


(b) Curva de exactitud

Figura B.6: DenseNet 201

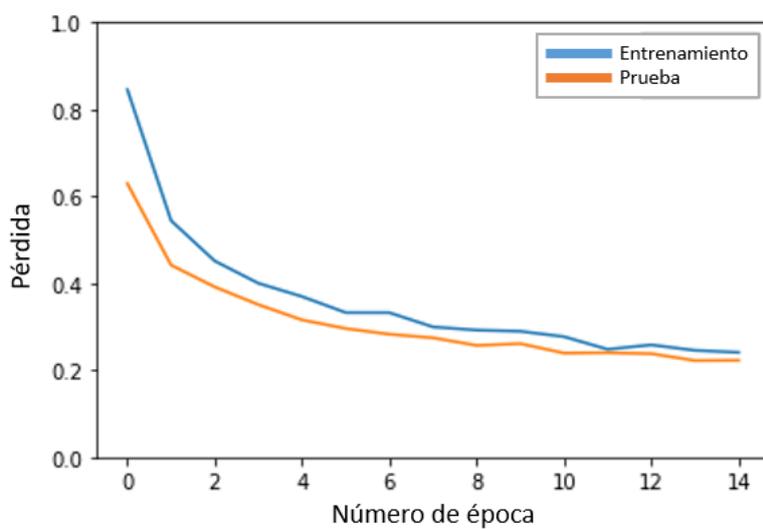


(a) Curva de pérdida

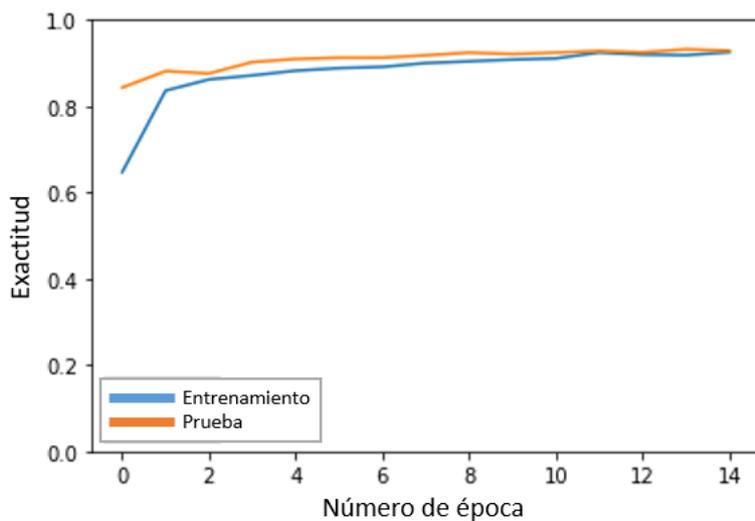


(b) Curva de exactitud

Figura B.7: MobileNet V2



(a) Curva de pérdida



(b) Curva de exactitud

Figura B.8: GoogleNet

BIBLIOGRAFÍA

- [1] T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. R. Acharya, “Automated detection of covid-19 cases using deep neural networks with x-ray images,” *Computers in Biology and Medicine*, vol. 121, p. 103792, 6 2020.
- [2] X. Xu, X. Jiang, C. Ma, P. Du, X. Li, S. Lv, L. Yu, Q. Ni, Y. Chen, J. Su, G. Lang, Y. Li, H. Zhao, J. Liu, K. Xu, L. Ruan, J. Sheng, Y. Qiu, W. Wu, T. Liang, and L. Li, “A deep learning system to screen novel coronavirus disease 2019 pneumonia,” *Engineering*, vol. 6, pp. 1122–1129, 10 2020.
- [3] I. D. Apostolopoulos and T. A. Mpesiana, “Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks,” vol. 43, pp. 635–640, 2020.
- [4] A. Halder and B. Datta, “COVID-19 detection from lung CT-scan images using transfer learning approach,” *Machine Learning: Science and Technology*, vol. 2, p. 045013, jul 2021.
- [5] S. Kumari, E. Ranjith, A. Gujjar, S. Narasimman, and H. S. A. S. Zeelani, “Comparative analysis of deep learning models for covid-19 detection,” *Global Transitions Proceedings*, vol. 2, pp. 559–565, 11 2021.

- [6] A. Narin, C. Kaya, and Z. Pamuk, “Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks,” vol. 24, pp. 1207–1220, 2021.
- [7] M. K. Bohmrah and H. Kaur, “Classification of Covid-19 patients using efficient fine-tuned deep learning DenseNet model,” *Global Transitions Proceedings*, vol. 2, pp. 476–483, nov 2021.
- [8] R. Kundu, P. K. Singh, S. Mirjalili, and R. Sarkar, “COVID-19 detection from lung CT-Scans using a fuzzy integral-based CNN ensemble,” *Computers in Biology and Medicine*, vol. 138, p. 104895, nov 2021.
- [9] M. Mamalakis, A. J. Swift, B. Vorselaars, S. Ray, S. Weeks, W. Ding, R. H. Clayton, L. S. Mackenzie, and A. Banerjee, “Denrescov-19: A deep transfer learning network for robust automatic classification of covid-19, pneumonia, and tuberculosis from x-rays,” *Computerized Medical Imaging and Graphics*, vol. 94, p. 102008, 12 2021.
- [10] L. Kong and J. Cheng, “Classification and detection of covid-19 x-ray images based on densenet and vgg16 feature fusion,” *Biomedical Signal Processing and Control*, vol. 77, p. 103772, 8 2022.
- [11] M. Soni, A. K. Singh, K. S. Babu, S. Kumar, A. kumar, and S. singh, “Convolutional neural network based ct scan classification method for covid-19 test validation,” *Smart Health*, vol. 25, p. 100296, 9 2022.
- [12] P. Pi and D. Lima, “Gray level co-occurrence matrix and extreme learning machine for covid-19 diagnosis,” *International Journal of Cognitive Computing in Engineering*, vol. 2, pp. 93–103, 6 2021.
- [13] S. Bakheet and A. Al-Hamadi, “Automatic detection of covid-19 using pruned glcm-based texture features and ldcrf classification,” *Computers in Biology and Medicine*, vol. 137, p. 104781, 10 2021.
- [14] M. L. Salvia, G. Secco, E. Torti, G. Florimbi, L. Guido, P. Lago, F. Salinaro, S. Perlini, and F. Leporati, “Deep learning and lung ultrasound for covid-19 pneumonia detection and severity classification,” *Computers in Biology and Medicine*, vol. 136, p. 104742, 9 2021.
- [15] C.-H. Tsai, J. van der Burgt, D. Vukovic, N. Kaur, L. Demi, D. Canty, A. Wang, A. Royse, C. Royse, K. Haji, J. Dowling, G. Chetty, and D. Fontanarosa, “Automatic deep learning-based pleural effusion classification in lung ultrasound images for respiratory pathology diagnosis,” *Physica Medica*, vol. 83, pp. 38–45, mar 2021.

- [16] S. Erfanian Ebadi, D. Krishnaswamy, S. E. S. Bolouri, D. Zonoobi, R. Greiner, N. Meuser-Herr, J. L. Jaremko, J. Kapur, M. Noga, and K. Punithakumar, “Automated detection of pneumonia in lung ultrasound using deep video classification for COVID-19,” *Informatcs in Medicine Unlocked*, vol. 25, p. 100687, jan 2021.
- [17] B. N. Smith and A. G. Webb, *Introduction to medical imaging: Physics, engineering and clinical applications*. Cambridge University Press, 2012.
- [18] Y. Tung-Chen, M. M. de Gracia, A. Díez-Tascón, R. Alonso-González, S. Agudo-Fernández, M. L. Parra-Gordo, S. Ossaba-Vélez, P. Rodríguez-Fuertes, and R. Llamas-Fuentes, “Correlation between chest computed tomography and lung ultrasonography in patients with coronavirus disease 2019 (covid-19),” *Ultrasound in Medicine and Biology*, vol. 46, pp. 2918–2926, 11 2020.
- [19] F. Guarracino, L. Vetrugno, F. Forfori, F. Corradi, D. Orso, P. Bertini, A. Ortalda, N. Federici, R. Copetti, and T. Bove, “Lung, heart, vascular, and diaphragm ultrasound examination of covid-19 patients: A comprehensive approach,” 6 2020.
- [20] M. Allinovi, A. Parise, M. Giacalone, A. Amerio, M. Delsante, A. Odone, A. Franci, F. Gigliotti, S. Amadasi, D. Delmonte, N. Parri, and A. Mangia, “Lung ultrasound may support diagnosis and monitoring of covid-19 pneumonia,” *Ultrasound in Medicine and Biology*, vol. 46, pp. 2908–2917, 11 2020.
- [21] L. Zhao, K. Yu, Q. Zhao, R. Tian, H. Xie, L. Xie, P. Deng, G. Xie, A. Bao, and J. Du, “Lung ultrasound score in evaluating the severity of coronavirus disease 2019 (covid-19) pneumonia,” *Ultrasound in Medicine and Biology*, vol. 46, pp. 2938–2944, 11 2020.
- [22] M. D. Serafino, M. Notaro, G. Rea, F. Iacobellis, V. D. Paoli, C. Acampora, S. Ianniello, L. Brunese, L. Romano, and G. Vallone, “The lung ultrasound: facts or artifacts? in the era of covid-19 outbreak,” *Radiologia Medica*, vol. 125, pp. 738–753, 8 2020.
- [23] S. Sofia, A. Boccatonda, M. Montanari, M. Spampinato, G. Cocco, E. Accogli, F. Cipollone, and C. Schiavone, “Thoracic ultrasound and sars-covid-19: a pictorial essay,” *Journal of Ultrasound*, vol. 23, pp. 217–221, 2020. ARDS - Síndrome de dificultad respiratoria aguda.
- [24] G. J. Tortora and B. Derrickson, *Principios de Anatomía y Fisiología*. Medica Panamericana, 11 ed., 2006.

- [25] A. C. Guyton and J. E. Hall, *Tratado de Fisiologia Medica*. Elsevier, 12 ed., 2011.
- [26] L. R. Sultan and C. M. Sehgal, “A review of early experience in lung ultrasound in the diagnosis and management of covid-19,” *Ultrasound in Medicine and Biology*, vol. 46, pp. 2530–2545, 9 2020.
- [27] C. Baloescu, G. Toporek, S. Kim, K. McNamara, R. Liu, M. M. Shaw, R. L. McNamara, B. I. Raju, and C. L. Moore, “Automated lung ultrasound b-line assessment using a deep learning algorithm,” *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 67, pp. 2312–2320, 11 2020.
- [28] Y. Zhang, H. Xue, M. Wang, N. He, Z. Lv, and L. Cui, “Lung ultrasound findings in patients with coronavirus disease (covid-19),” *American Journal of Roentgenology*, vol. 216, pp. 80–84, 1 2021.
- [29] N. R. Ratnasari, A. Susanto, I. Soesanti, and Maesadji, “Thoracic X-ray features extraction using thresholding-based ROI template and PCA-based features selection for lung TB classification purposes,” *Proc. of 2013 3rd Int. Conf. on Instrumentation, Communications, Information Technol., and Biomedical Engineering: Science and Technol. for Improvement of Health, Safety, and Environ., ICICI-BME 2013*, pp. 65–69, 2013.
- [30] H. A. Nugroho, M. Rahmawaty, Y. Triyani, I. Ardiyanto, L. Choridah, and R. Indrastuti, “Texture analysis and classification in ultrasound medical images for determining echo pattern characteristics,” *2017 7th IEEE International Conference on System Engineering and Technology, ICSET 2017 - Proceedings*, pp. 23–26, nov 2017.
- [31] W.-R. Gonzalez, R.C. and S. Eddins, *Digital Image Processing using MATLAB*. Gatesmark Publishing, 2 ed., 2009.
- [32] X. Wang and N. D. Georganas, “GLCM texture based fractal method for evaluating fabric surface roughness,” *Canadian Conference on Electrical and Computer Engineering*, pp. 104–107, 2009.
- [33] S. L. Jany Shabu and C. Jayakumar, “Brain tumor classification with mri brain images using 2-level glcm features and sparse representation based segmentation,” in *2020 3rd International Conference on Intelligent Sustainable Systems (ICISS)*, pp. 793–799, 2020.
- [34] M. Hall-Beyer, “Glm texture: A tutorial v. 3.0 march 2017,” 2017.

- [35] J. Brownlee, *Machine Learning Mastery with Python: Understand Your Data, Create Accurate Models and Work Projects End-to-end*. Jason Brownlee, 2016.
- [36] R. Shaikh, “Feature selection techniques in machine learning with python,” Oct 2018. (Accesado: 18 junio 2022).
- [37] G. K. Rajbahadur, S. Wang, G. Ansaldi, Y. Kamei, and A. E. Hassan, “The impact of feature importance methods on the interpretation of defect classifiers,” *IEEE Transactions on Software Engineering*, jul 2021.
- [38] A. Perrier, “Feature importance in random forests,” Aug 2015. (Accesed: 10 october 2022).
- [39] C. Lee, “Feature importance measures for tree models - part i,” Sep 2020. (Accessed: 10 october 2022).
- [40] T. Shin, “Understanding feature importance and how to implement it in python,” Feb 2021. (Accessed: 10 october 2022).
- [41] J. Brownlee, “How to calculate feature importance with python,” Aug 2020. (Accessed: 2 september 2022).
- [42] S. Khalid, T. Khalil, and S. Nasreen, “A survey of feature selection and feature extraction techniques in machine learning,” *Proceedings of 2014 Science and Information Conference, SAI 2014*, pp. 372–378, oct 2014.
- [43] M. R. H. Subho, M. R. Chowdhury, D. Chaki, S. Islam, and M. M. Rahman, “A Univariate Feature Selection Approach for Finding Key Factors of Restaurant Business,” *Proceedings of 2019 IEEE Region 10 Symposium, TENSYP 2019*, pp. 605–610, jun 2019.
- [44] N. Kagiya, M. Tokodi, and P. P. Sengupta, “Machine Learning in Cardiovascular Imaging,” *Heart Failure Clinics*, vol. 18, pp. 245–258, apr 2022.
- [45] A. Davis, K. Billick, K. Horton, M. Jankowski, P. Knoll, J. E. Marshall, A. Paloma, R. Palma, and D. B. Adams, “Artificial Intelligence and Echocardiography: A Primer for Cardiac Sonographers,” *Journal of the American Society of Echocardiography*, vol. 33, pp. 1061–1066, sep 2020.
- [46] C. R. Olsen, R. J. Mentz, K. J. Anstrom, D. Page, and P. A. Patel, “Clinical applications of machine learning in the diagnosis, classification, and prediction of heart failure,” *American Heart Journal*, vol. 229, pp. 1–17, nov 2020.

- [47] H. T. Zaw, N. Maneerat, and K. Y. Win, “Brain tumor detection based on naïve bayes classification,” *Proceeding - 5th International Conference on Engineering, Applied Sciences and Technology, ICEAST 2019*, 7 2019.
- [48] C.-F. Tsai, W.-Y. Lin, Z.-F. Hong, and C.-Y. Hsieh, “Distance-based features in pattern classification,” *EURASIP Journal on Advances in Signal Processing*, vol. 2011, pp. 1–11, 12 2011.
- [49] N. Absar, B. Mamur, A. Mahmud, T. B. Emran, M. U. Khandaker, M. Faruque, H. Osman, A. Elzaki, and B. A. Elkhader, “Development of a computer-aided tool for detection of covid-19 pneumonia from cxr images using machine learning algorithm,” *Journal of Radiation Research and Applied Sciences*, vol. 15, pp. 32–43, 3 2022.
- [50] “1.6. nearest neighbors.” (Accesado: 14 junio 2022).
- [51] “Sklearn random forest classifiers in python tutorial.” (Accesado: 18 junio 2022).
- [52] “Random forest con python.” (Accesado: 18 junio 2022).
- [53] A. Kumar, J. Kim, D. Lyndon, M. Fulham, and D. Feng, “An ensemble of fine-tuned convolutional neural networks for medical image classification,” *IEEE Journal of Biomedical and Health Informatics*, vol. 21, pp. 31–40, 1 2017.
- [54] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, “Convolutional neural networks for medical image analysis: Full training or fine tuning?,” *IEEE Transactions on Medical Imaging*, vol. 35, pp. 1299–1312, 5 2016.
- [55] D. Chakraborty, *Observer Performance Methods for Diagnostic Imaging: Foundations, Modeling, and Applications with R-Based Examples*. CRC Press, 1 ed., 2018.
- [56] H. Kaur, A. S. Sharma, and S. Kaur, “Liver tissue classification for ultrasound images,” *Proceedings - 2016 International Conference on Advances in Computing, Communication and Automation (Fall), ICACCA 2016*, 11 2016.
- [57] C. G. González, J. F. P. Castán, and J. L. Martín, “Curso de introducción a la investigación clínica. capítulo 8: Pruebas diagnósticas. concordancia,” *Medicina de Familia. SEMERGEN*, vol. 33, pp. 509–519, 12 2007.

Bibliografía

- [58] “covid19 ultrasound original dataset.” (Accessed: 8 june 2021).
- [59] “Covid-19 lung ultrasound imaging: Fighting together.” (Accessed: 16 february 2022).