

**APOYO AL DIAGNÓSTICO DE NEUMONÍA Y DETECCIÓN DE OPACIDADES
PULMONARES USANDO SEGMENTACIÓN SEMÁNTICA (MASK-RCNN)
UTILIZANDO IMÁGENES DE RX DE TÓRAX.**

VICTOR MANUEL ASTUDLLO DELGADO

**CORPORACION UNIVERSITARIA COMFACAUCA
FACULTAD DE INGENIERIAS
INGENIERIA MECATRONICA
POPAYAN
2021**

**APOYO AL DIAGNÓSTICO DE NEUMONÍA Y DETECCIÓN DE OPACIDADES
PULMONARES USANDO SEGMENTACIÓN SEMÁNTICA (MASK-RCNN)
UTILIZANDO IMÁGENES DE RX DE TÓRAX.**

VICTOR MANUEL ASTUDLLO DELGADO

Trabajo de grado para optar al título de Ingeniería Mecatrónica

**Director de tesis:
M.Sc. DAVID ARMANDO REVELO LUNA**

**CORPORACION UNIVERSITARIA COMFACAUCA
FACULTAD DE INGENIERIAS
INGENIERIA MECATRONICA
POPAYAN
2021**

Nota de aceptación

El Director y los Jurados han leído el presente documento, escucharon la sustentación del mismo por sus autores y lo encuentran satisfactorio.

Director

Presidente del Jurado

Jurado

Popayán, 2 de agosto de 2021

AGRADECIMIENTOS

Primeramente, le agradezco a Dios por tenerme con vida y llegar a esta etapa final de mi carrera, agradezco a mi familia por brindarme su apoyo durante todo este tiempo de mi carrera, por darme motivación de seguir con mis metas propuestas.

Agradezco de manera atenta al M. Sc. David Armando Revelo Luna, profesor e investigador del programa Ingeniería mecatrónica, quien con su experiencia, conocimiento y motivación me oriento durante todo el proceso de formación tanto académica como investigativa.

Así mismo agradezco a la Corporación Universitaria Comfacauca y al cuerpo profesoral del programa ingeniería mecatrónica, por el conocimiento brindando durante todo este proceso de la carrera, tanto académicamente como para formar profesionales íntegros que aporten al desarrollo del país.

CONTENIDO

	pág.
CAPITULO 1. GENERALIDADES	14
1.1 PLANTEAMIENTO DEL PROBLEMA	14
1.2 PREGUNTA DE INVESTIGACIÓN	17
1.3 JUSTIFICACION	17
1.4 OBJETIVOS	18
1.4.1 Objetivo general	18
1.4.2 Objetivo especifico	18
CAPITULO 2. MARCO TEORICO	19
2.1 NEUMONÍA	19
2.2 DIAGNÓSTICO DE NEUMONÍA	19
2.2.1 Conteo sanguíneo completo (CSC)	19
2.2.2 Hemocultivo	20
2.2.3 Radiografías de tórax	20
2.2.4 Prueba esputo	21
2.2.5 Tomografía computarizada	21

	pág.
2.3 DEEP LEARNING	21
2.3.1 Redes neuronales convolucionales (CNN)	22
2.3.2 Segmentación semántica	22
2.3.3 Instanciación semántica (MaskRCNN)	23
2.4 PRODUCCIÓN DE MODELOS DE DEEP LEARNING	24
2.4.1 Interfaz de programación de aplicaciones (API)	25
2.4.2 Streamlit	25
2.5 MÉTRICAS DE DESEMPEÑO	26
2.5.1 Intersección sobre la unión (IoU)	26
2.5.2 Métricas convencionales	27
CAPITULO 3. MARCO ANTECEDENTES	29
3.1 Enfoque general	29
3.2 Enfoque específico	29
CAPITULO 4. METODOLOGIA	36
4.1 DATASEET	36
4.1.1 Etiquetado de imágenes	36

	pág.
4.2 PROCESAMIENTO DE LOS DATOS	37
4.3 ARQUITECTURA CNN	38
4.4 ENTRENAMIENTO DE LOS MODELOS	40
4.5 MÉTRICAS DE DESEMPEÑO	41
4.6 UMBRALES DE SELECCIÓN	42
4.7 PUESTA EN PRODUCCIÓN DEL MODELO	43
CAPITULO 5. RESULTADOS Y DISCUSION	45
5.1 PRE-PROCESAMIENTO DE IMÁGENES RX	45
5.2 ENTRENAMIENTO DE LOS MODELOS	46
5.3 DETECCIÓN DE OPACIDADES MASK-RCNN	47
5.4 DESEMPEÑO DE LOS MODELOS EN LA DETECCIÓN DE OPACIDADES	51
5.5 DESEMPEÑO DE LOS MODELOS EN LA DETECCIÓN DE NEUMONÍA	52
5.6 PUESTA EN PRODUCCIÓN DEL PREDICTOR DE NEUMONÍA	54
6. CONCLUSIONES	57

	pág.
7. RECOMENDACIONES	58
8. BIBLIOGRAFÍA	59

LISTADO DE FIGURAS

	pág.
Figura 1. Radiografía de Tórax completa.	20
Figura 2. Arquitecta general de una CNN.	22
Figura 3. Segmentación semántica de detección de objetos: a. Imagen original, b. Imagen segmentada por clases.	23
Figura 4. Instanciación semántica de objetos.	23
Figura 5. El Framework de Mask R-CNN para instanciación semántica.	24
Figura 6. Flujo de trabajo de streamlit.	26
Figura 7. Etiquetado de imágenes utilizando el software VGG Image Annotator.	36
Figura 8. Histograma de una imagen del dataset inicial.	37
Figura 10. Modelo basado en Mask-RCNN para identificación de opacidad y segmentación de enfermedades por píxeles.	39
Figura 11. Plataforma de Google Colab.	41
Figura 13. Diagrama de selección de umbrales para las métricas de desempeño.	43
Figura 14. Diagrama de solicitudes de la APP en Streamlit.	44
Figura 15. Histograma de imagen con ecualización.	45
Figura 16. Aplicación de ecualización de histograma al dataset: a. Imagen original, b. Imagen con ecualización de histograma.	45
Figura 17. Ecualización de histograma en opacidad: a. Opacidad original, b. Opacidad con ecualización de histograma.	46

	pág.
Figura 18. Loss entrenamiento y validación de los experimentos.	47
Figura 19. Aplicación de la métrica IoU individual: a. Anotación mascara, b. Resultado mascara, c. Unión e intersección.	48
Figura 20. Opacidades pulmonares segmentadas en pacientes.	48
Figura 21. Diagramas de desempeño de la métrica IoU global de grupo de imágenes test.	49
Figura 22. Resultados de grupos individuales de la prueba Kruskal-Wallis.	50
Figura 23. Predicción de los modelos en la detección de opacidades.	51
Figura 24. Matriz de confusión de los experimentos: a. Experimento 1, b. Experimento 2, c. Experimento 3.	53
Figura 25. Main principal del Predictor de neumonía APP.	54
Figura 26. Cargar imagen diagnostico a la APP.	55
Figura 27. Predictor APP: a. Resultado del predictor, b. Resultado predictor: imagen segmentada y por cuadros delimitadores.	55

LISTADO DE ECUACIONES

	pág.
Ecuación 1. IoU (Intersection over union).	26
Ecuación 2. Métrica de desempeño: Accuracy.	27
Ecuación 3. Métrica de desempeño: Precisión.	27
Ecuación 4. Métrica de desempeño: Recall.	27
Ecuación 5. Métrica de desempeño: F1Score.	28
Ecuación 6. Métrica de desempeño: Especificidad.	28
Ecuación 7. Ecuación de ecualización de histograma: niveles de grises	38
Ecuación 8. Step epoch para entrenamiento.	39
Ecuación 9 . Media de la métrica Intersección sobre la Unión para el grupo de test.	41
Ecuación 10. Media de la métrica intersección sobre la unión para cada imagen.	41
Ecuación 11. Desviación estándar de la métrica Intersección y Unión.	41

LISTADO DE TABLAS

	pág.
Tabla 1. Características generales de los antecedentes	33
Tabla 2. Transformaciones geométricas utilizadas en data augmentation.	38
Tabla 3. Parámetros de entrenamiento de la red Mask-RCNN.	40
Tabla 4. Resultados de la prueba de Kolmogorov-Smirnov de una muestra	49
Tabla 5. Resultados de la prueba de Kruskal-Wallis para datos no paramétricos.	50
Tabla 6. Métricas de desempeño de los experimentos para identificar opacidades pulmonares.	51
Tabla 7. Métricas de desempeño de los experimentos para clasificar imágenes de neumonía.	53
Tabla 8. Tiempos de predicción de la APP.	56

RESUMEN

La neumonía es una enfermedad que ha causado gran cantidad de muertes a nivel mundial. Los métodos de detección de neumonía: pruebas de sangre, prueba de esputo, tomografía computarizada y las radiografías (RX) de tórax, en general requieren de un médico especialista para la lectura de los exámenes correspondientes y tiempo para su análisis. El objetivo de esta investigación fue generar un modelo basado en instancias semánticas y aprendizaje profundo (Mask-RCNN) que permita el apoyo en el diagnóstico de neumonía, identificando opacidades pulmonares asociadas a la neumonía y clasificando las imágenes diagnósticas de pacientes con neumonía sanos, asociando a pacientes enfermos como verdaderos positivos y pacientes sanos como verdaderos negativos. Usando el dataset publicado en el año 2018 por la RSNA (Sociedad Radiológica Norteamericana) y la Society of Thoracic Radiology, etiquetado en conjunto con MD.ai. Este conjunto de datos contiene 26.636 imágenes de RX de tórax en formato (.dcm) de las cuales 20.672 son de imágenes de pacientes sanos y 6.012 de pacientes con neumonía. Se utilizaron 12.024 imágenes de radiografías de RX de tórax en formato PNG. Se realizaron 3 experimentos, donde se variaron las condiciones de las imágenes del dataset (exp1: imágenes RX pacientes con neumonía, exp2: imágenes RX pacientes sanos y con neumonía, exp3: imágenes RX pacientes sanos y con neumonía, más *data augmentation*). Para los 3 experimentos se realizó un pre-procesamiento de ecualización de histograma. Para evaluar el desempeño de los modelos de identificación de opacidades pulmonares y clasificación de pacientes con neumonía, se utilizaron la métrica intersección sobre la unión (IoU), precisión, recall, F1Score y accuracy. Finalmente, los modelos obtenidos se subieron a un servidor local utilizando el framework de streamlit. Se encontró que la métrica de mejor desempeño fue Recall para el experimento 1 con (95% y 96%) para la detección de opacidades y clasificación de pacientes con neumonía respectivamente. Por medio del sistema de predicción que permite realizar un diagnóstico de neumonía desde el navegador web, como aplicativo de servidor, se pretende agilizar procesos en la detección de neumonía para generar tratamientos oportunos.

Palabras clave: aumento de datos, intersección sobre la unión, redes neuronales convolucionales, Resnet50, Sociedad Radiológica Norteamérica, Streamlit.

CAPITULO 1. GENERALIDADES

1.1 PLANTEAMIENTO DEL PROBLEMA

La neumonía es un tipo de infección respiratoria aguda (IRA) que afecta a los pulmones, los cuales están conformados por alvéolos que en las personas sanas se llenan de aire al respirar. En personas enfermas con neumonía los alveolos están llenos de pus y líquido, lo que genera dolor al respirar y limita el oxígeno entrante a los pulmones (Organización Mundial de la Salud, 2019). La neumonía puede ser causada por diversos agentes infecciosos como: virus, bacterias y hongos. Entre ellos, la más frecuente es "*Streptococcus pneumoniae*", bacteria que causa neumonía en los niños. Por otra parte, el "*virus sincitial*" respiratorio es la causa más frecuente de neumonía vírica que afecta igualmente a los niños, pero además a las personas adultas. El contagio se presenta por contacto con una persona diagnosticada, ya que puede ingresar el virus o la bacteria por la boca o nariz y así ingresar al sistema respiratorio (Biblioteca Nacional de Medicina de EE. UU, 2018).

Los síntomas más comunes que se le atribuyen a la neumonía son: tos, mucosidad amarillenta e incluso tos con sangre, fiebre, escalofríos, confusión, dolor de cabeza por causa de la mala absorción de oxígeno, dolor torácico o leuconiquia esto anterior dependiendo de la gravedad de la enfermedad (Mayo Clinic, 2020).

De este modo, una persona que posea los síntomas anteriormente expuestos se le realizarán pruebas para detectar si tiene neumonía. En este sentido, las pruebas más comunes son: el examen de sangre, consiste en un conteo sanguíneo completo que permitirá medir componentes y características de la sangre, como los glóbulos rojos, encargados de llevar oxígeno por todo el cuerpo, glóbulos blancos encargados de combatir infecciones, el porcentaje de plaquetas y el hematocrito (Biblioteca Nacional de Medicina de EE.UU, 2020). Este examen permitirá diagnosticar una enfermedad de la sangre, una infección y otras enfermedades del sistema inmunitario, por lo cual su precisión para la detección de neumonía es baja, ya que puede detectar cualquier infección presente en el cuerpo, pero no identificará la bacteria o virus que causó la infección; el hemocultivo es una prueba adicional que permite determinar si el paciente tiene una infección bacteriana que se ha diseminado al torrente sanguíneo. La prueba permite detectar qué tipo de bacteria es la que ha causado la infección, sin embargo, no se puede determinar el avance de la enfermedad en el organismo (Biblioteca Nacional de Medicina de EE.UU, 2020).

Asimismo, la prueba de esputo permitirá buscar bacterias en la saliva o flema (sustancia viscosa que proviene de los pulmones) que estén causando una infección en los pulmones, de este modo se puede identificar si existe una bacteria que pueda causar la neumonía o este causando una infección en los pulmones (Biblioteca Nacional de Medicina de EE.UU, 2019). La principal limitación de esta prueba es que permite identificar si el paciente tiene la bacteria que causa la neumonía o alguna infección en ellos, pero no brindará la información que pueda establecer si ha causado algún daño en los pulmones ya que es una prueba que analiza el esputo que proviene de los pulmones, pero no permite una visualización de los mismos.

El examen de prueba de esputo suele complementarse con la toma de una radiografía de tórax (RX) ya que permite observar la presencia de líquido en los pulmones o la acumulación de aire en el espacio alrededor de ellos, o si existen opacidades pulmonares asociadas a la neumonía. De esta forma, el médico especialista, puede determinar el avance que ha tenido la neumonía al afectar los pulmones, permitiendo proporcionar el tratamiento adecuado (Mayo Clinic, 2020). Las limitaciones de la radiografía de tórax están asociadas con incertidumbres en el diagnóstico, ya que otras enfermedades pueden causar los mismos daños en los pulmones como el cáncer o la tuberculosis (Neuman, et al., 2012). Por esto, es importante complementar los estudios con la prueba de esputo para así determinar de manera acertada en el diagnóstico de la neumonía.

Por otra parte, se utiliza la tomografía computarizada como un examen adicional que permite visualizar los pulmones y la afectación de los mismos. Este examen permite también determinar si tiene complicaciones como abscesos pulmonares o derrames pleurales de una forma muy detallada, favoreciendo un posible diagnóstico adecuado para la enfermedad. Las limitaciones de este método están relacionadas con los altos costos del examen, ya que requiere un equipo especializado (National Institute of Biomedical Imaging and Bioengineering, 2012). Además, el acceso a los equipos de Tomografía Computarizada es limitado, sobre todo en clínicas y hospitales de bajo nivel, lo cual restringe el acceso al examen.

Existen métodos tecnológicos alternativos utilizando redes neuronales convolucionales que podrían permitir apoyar a los especialistas en la toma de decisión para clasificar imágenes de RX de pacientes sanos y enfermos de neumonía. En (Jain, Nagrath, Kataria, Kaushik, & Hemanth, 2019) utilizaron 5.840 imágenes de radiografías de tórax para entrenamiento y test, en los modelos VGG16 y VGG19, con un accuracy mayor a 85%, el modelo ResNet50, obtuvo un accuracy del 77%, pero al contrario obtuvo un recall del 97% mientras que en los otros modelos fue menor a 95%. En (Yu, HuaWang, & DongZhang, 2021) se utilizó el mismo número de imágenes para el entrenamiento y validación, en los modelos

XceptionNet, AlexNet y ResNet101, se obtuvo un accuracy mayor a 81%. Por el contrario, en (Militante & Sibbaluca, 2020), se utilizó un mayor número de imágenes para el entrenamiento y validación (26.684 imágenes) con los modelos AlexNet, LeNet, GoogleNet, y VGGNet, obtuvieron una precisión del 95 % y 97%, el recall mayor fue del 98 % para los modelos GoogleNet y VGGNet.

Otros estudios se han enfocado en la detección de opacidades en imágenes RX. Es decir, más allá de clasificarlas en una categoría (paciente con neumonía o paciente sano), se ha evaluado el desempeño de las CNN para identificar si existen opacidades localizadas en las imágenes que sugieran que el paciente padece neumonía. En las investigaciones (Guan, et al., 2018; Yao, et al., 2017; Rajpurkar, Irvin, 2017), se utilizó el conjunto de datos “ChestX-ray14” que contiene 112.120 imágenes de RX con catorce etiquetas de enfermedades comunes pulmonares, para los modelos CheXNet - AG-CNN, utilizando como base la ResNet50 – DenseNet respectivamente, usando la métrica AUROC (Área bajo la curva). El modelo de mejor desempeño fue el presentado en (Guan, et al., 2018) con 77,6%. En (Sirazitdinov, et al., 2019) se utilizaron 29.684 imágenes para entrenamiento y test, en los modelos ResNet50 y Densenet121. Se obtuvo una precisión mayor a 85%. Además, en el mismo estudio se propuso unir en un solo modelo la técnica de instancias semánticas (Mask-RCNN) y el modelo de detección RetinaNet obteniendo una precisión inferior a los anteriores modelos del estudio, pero logrando un recall mayor a 79%. En las investigaciones (Chen, Han, Lin, Zhao, & Kong, 2020; Li, Zhu, Wu, & Wang, 2018; Li, Shen, & Luo, 2018), se utilizó el mismo dataset de imágenes y se obtuvo un recall mayor a 90%. En (Chen, Han, Lin, Zhao, & Kong, 2020) se presentó el valor más alto de recall con 97%.

Las radiografías digitales de tórax se han convertido en una herramienta valiosa para generar tratamientos que ayuden a diagnosticar la neumonía, ya que el equipamiento médico utilizado es más accesible al usado para las tomas de tomografías. De este modo, las radiografías de tórax permiten obtener imágenes de los pulmones e información detallada de la gravedad de la infección. Sin embargo, existe una cantidad limitada de profesionales especialistas enfocados al diagnóstico de la neumonía en Colombia, lo que representa una limitación en el sistema médico nacional, debido a que la tasa de contagio supera en muchos casos las capacidades instaladas de los hospitales (Gómez Hernandez, González Santacruz, & Ibarguen Mondragón, 2014).

Dentro de este contexto, esta investigación propone desarrollar una red neuronal convolucional que permita, por medio de imágenes RX torácicas, apoyar el diagnóstico de neumonía, generando un aplicativo (APP) que realice el diagnóstico desde el navegador web mejorando el tiempo de detección y disminuyendo la carga

hacia el especialista, favoreciendo la toma de decisiones en relación al tratamiento adecuado para cada paciente.

1.2 PREGUNTA DE INVESTIGACIÓN

Considerando las ventajas de los sistemas de aprendizaje automático y sus limitaciones, se presenta la siguiente pregunta de investigación: **¿Cómo apoyar el diagnóstico de la neumonía mediante un sistema de inteligencia artificial basado en redes neuronales convolucionales aplicado a imágenes torácicas de rayos X?**

1.3 JUSTIFICACION

La Our World in Data estimó que en el 2017 2.56 millones de personas murieron por neumonía a nivel mundial. De estas, casi un tercio de las víctimas fueron niños menores de cinco años, ubicando a esta enfermedad como la principal causante de muerte de esta población, lo que supone el 15% de todas las defunciones de niños menores de 5 años (Dadonaite & Roser, 2019). Asimismo, el 99% de muertes por neumonía que sufren los niños menores de 5 años en todo el mundo se producen en los países en desarrollo (Organización Panamericana de la Salud, 2014).

Según los datos de OMS publicados en 2018 los principales países en Latinoamérica con muertes atribuidas a la neumonía son, Brasil con 92,498, Perú con 22,743, México con 22.441, Colombia con 10.748, Venezuela con 6.740, Ecuador con 5.368, Paraguay con 1,468 (World Life Expectancy, 2018). En Colombia los departamentos con mayor número de muertes son Antioquia con 968, Bolívar 243, Tolima con 201, Nariño 163, Huila con 121 y Cauca 119 (Gómez Hernández, González Santacruz, & Iburguen Mondragón, 2014).

Colombia es un país dependiente de otros países para la adquisición de equipamientos médicos, en consecuencia, los costos de inversión en estos equipos aumentan debido a los procesos de importación. Además, existe escases de especialistas que puedan interpretar imágenes de rayos X para detectar la neumonía. Para el año 2017, el Ministerio de Salud estimó 920 radiólogos en el país (Ministerio de Salud, 2018). Existe una motivación en la comunidad científica en generar alternativas tecnológicas que puedan apoyar al diagnóstico de neumonía para reducir su tasa de mortalidad y el tiempo de diagnóstico.

Debido a la necesidad planteada, este proyecto pretende abordar métodos que apoyen a los especialistas en el diagnóstico de la neumonía, permitiendo generar tratamientos oportunos para los pacientes. El desarrollo de esta investigación propone el uso de redes neuronales convolucionales para el apoyo a la detección de neumonía a partir de imágenes de RX de tórax identificando anomalías en los pulmones que impliquen la presencia de la enfermedad, con el fin de obtener un método rápido de diagnóstico.

1.4 OBJETIVOS

1.4.1 Objetivo general

- Desarrollar un sistema basado en redes neuronales convolucionales como apoyo al análisis de imágenes diagnósticas de tórax para la detección de neumonía.

1.4.2 Objetivo específico

- Definir un dataset de imágenes torácicas RX con anotaciones para el entrenamiento de un sistema de aprendizaje profundo.
- Diseñar un sistema basado en redes neuronales convolucionales que permita encontrar características clínicas en imágenes torácicas RX.
- Evaluar el desempeño del sistema de detección basado en redes neuronales convolucionales en imágenes torácicas RX.

CAPITULO 2. MARCO TEORICO

2.1 NEUMONÍA

La neumonía es una infección en uno o ambos pulmones. Causa que los alvéolos pulmonares se llenen de líquido o pus. Puede variar de leve a grave, según el tipo de germen que causa la infección, su edad y su estado general de salud (Mayo Clinic, 2020), diversos agentes infecciosos como virus o bacterias causan neumonía, siendo las más comunes son: “*Streptococcus pneumoniae*” afecta principalmente a niños, “*Haemophilus influenzae*” de tipo b (Hib) la segunda causa más común de neumonía bacteriana, virus “*sincitia*” respiratorio es la causa más frecuente de neumonía vírica, “*Pneumocystis jiroveci*” es una causa importante de neumonía en niños menores de seis meses con VIH/SIDA (Organización Mundial de Salud, 2019).

2.2 DIAGNÓSTICO DE NEUMONÍA

La neumonía puede ser difícil de diagnosticar porque puede causar algunos de los mismos síntomas que un resfriado o gripe. Para la cual existen algunos métodos convencionales como son:

2.2.1 Conteo sanguíneo completo (CSC)

Es un análisis de sangre que mide componentes y características de la sangre del paciente, entre ellos mide: (Loyola Medicine, 2020),

- **Glóbulos rojos:** Llevan oxígeno de los pulmones al resto del cuerpo.
- **Glóbulos blancos:** Combaten las infecciones. Hay cinco tipos principales de glóbulos blancos. Un CSC mide el número total de glóbulos blancos en la sangre. Una prueba llamada CSC con fórmula leucocitaria mide además los porcentajes de cada tipo de glóbulo blanco.
- **Plaquetas:** Ayudan a que la sangre coagule y detienen el sangrado.
- **Hemoglobina:** Proteína de los glóbulos rojos que lleva oxígeno de los pulmones al resto del cuerpo.
- **Hematocrito:** Medida del volumen total de glóbulos rojos en la sangre.

2.2.2 Hemocultivo

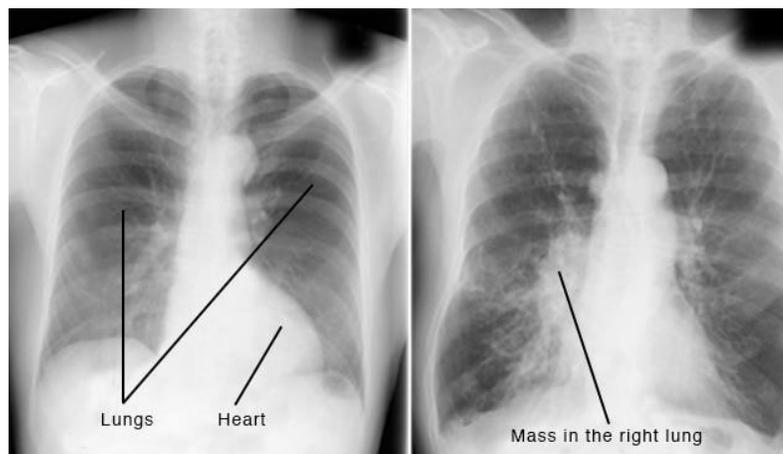
Las bacterias son un grupo amplio de organismos unicelulares. Una prueba de cultivo de bacterias puede detectar bacterias perjudiciales en el cuerpo. Durante una prueba de cultivo de bacterias, se toma una muestra de sangre, orina, piel u otra parte del cuerpo. El tipo de prueba depende del lugar en el que se sospecha la infección. Las células de la muestra se llevan a un laboratorio y se colocan en un medio especial para promover la reproducción celular de esta forma determinar si el paciente tiene algún tipo de bacteria ubicada en el sitio de la muestra (Biblioteca Nacional de Medicina de EE. UU, 2020).

2.2.3 Radiografías de tórax

Las radiografías torácicas producen imágenes del corazón, los pulmones, los vasos sanguíneos, las vías respiratorias, los huesos del tórax y de la columna vertebral, ver Figura 1. También pueden revelar la presencia de líquido en los pulmones o alrededor de ellos, o la presencia de aire que rodea a los pulmones (Mayo Clinic, 2020).

La imagen le proporciona al médico herramientas para determinar si el paciente presenta problema cardíaco, un pulmón colapsado, neumonía, una fractura de costillas, enfisema y cáncer (Mayo Clinic, 2020).

Figura 1. Radiografía de Tórax completa.



Fuente: (Mayo Clinic, 2020).

Las radiografías de tórax se utilizan para diagnosticar:

- **Infecciones o afecciones pulmonares:** Se puede detectar la presencia de neumonía, tuberculosis o cáncer de pulmón.
- **Cáncer mamario:** Una mamografía es un tipo de radiografía especial que se utiliza para examinar el tejido mamario.
- **Corazón dilatado:** Este signo de insuficiencia cardíaca congestiva se ve con claridad en las radiografías.
- **Vasos sanguíneos obstruidos:** La inyección de medios de contraste con yodo puede ayudar a resaltar algunas partes del aparato circulatorio para que sean visibles en las radiografías.

2.2.4 Prueba esputo

Un cultivo de esputo es una prueba que busca bacterias y otros gérmenes que pueden causar una infección en los pulmones o las vías respiratorias. El esputo, también conocido como flema, es un tipo de mucosidad espesa que se produce en los pulmones. Si tiene una infección o una enfermedad crónica que afecta los pulmones o las vías respiratorias (Biblioteca Nacional de Medicina de EE.UU, 2019; CanalSalud, 2016).

2.2.5 Tomografía computarizada

Es un procedimiento computarizado de imágenes por RX en el que se proyecta un haz angosto de RX a un paciente y se gira rápidamente alrededor del cuerpo. Para generar imágenes transversales del cuerpo y obtener unas imágenes tridimensionales con mayor detalle de los pulmones en comparación con las radiografías de tórax, permite fácilmente la identificación de presencia de tumores, embolias pulmonares (coágulos de sangre), exceso de fluido y otros padecimientos como enfisema o neumonía (National Institute of Biomedical Imaging and Bioengineering, 2012).

2.3 DEEP LEARNING

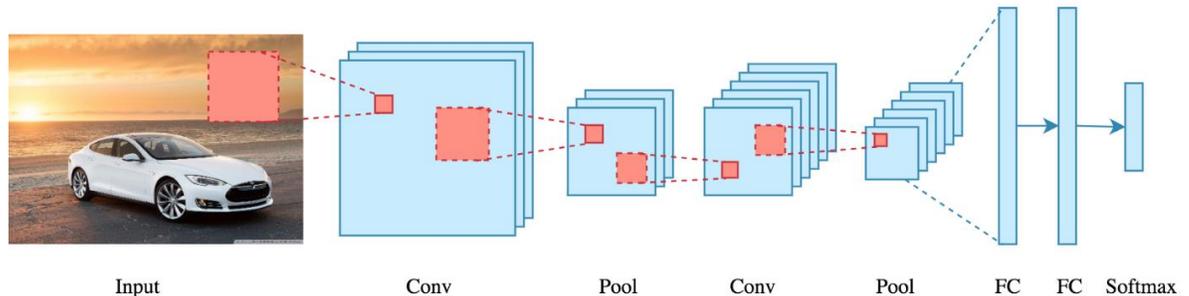
El aprendizaje profundo es una rama del aprendizaje automático, que se basa en un grupo de algoritmos que buscan obtener abstracciones de alto nivel de los datos

utilizando varias capas de procesamiento, Compuesto por varias alteraciones lineales y no lineales, para realizar tareas: como el reconocimiento de voz, la identificación de imágenes y la proyección. Este aprendizaje determina los patrones básicos de esta información y enseña a las computadoras a desarrollarse a través de la identificación de patrones en las capas de procesamiento (Oliveira, 2019).

2.3.1 Redes neuronales convolucionales (CNN)

Las CNN es un algoritmo de aprendizaje profundo que actualmente es utilizado en la resolución de problemas de visión computacional, dado su alto desempeño en el reconocimiento de objetos en imágenes y video. El objetivo de la CNN es extraer todas las características de una imagen y luego usar dichas características para detectar o clasificar los objetos en ciertas categorías ver Figura 2. Los parámetros de los filtros que se aplican en las operaciones de convolución se ajustarán y optimizarán junto con los componentes de clasificación para minimizar el error de clasificación total (Fukushima, 1988; LeCun, Bottou, Orr, & Müller, 1998).

Figura 2. Arquitecta general de una CNN.



Fuente: (Silva & Freire, 2019).

2.3.2 Segmentación semántica

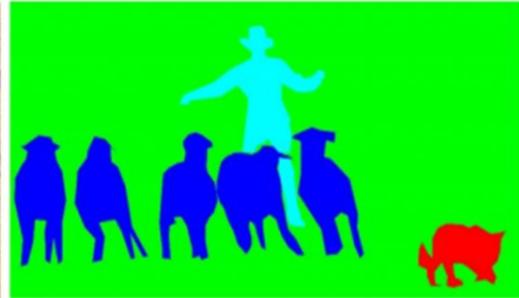
La segmentación semántica es un algoritmo de *Deep Learning*, que permite diferenciar tipos de objetos del fondo de la imagen. Detecta todos los objetos presentes en una imagen a nivel de píxel y produce regiones agrupando objetos a las clases que correspondan. Esta técnica detecta claramente objetos que tienen una forma irregular como se observa en la Figura 3. Al contrario que la detección de objetos, en donde los objetos deben encajar en un *bounding box* (Palomino & Concha, 2009).

Figura 3. Segmentación semántica de detección de objetos: a. Imagen original, b. Imagen segmentada por clases.

a. Imagen original



b. Imagen segmentada por clases



Fuente. (Universitat Oberta de Catalunya, 2016).

2.3.3 Instanciación semántica (MaskRCNN)

La segmentación de instancias utiliza el mismo principio de la segmentación semántica, asocia una etiqueta de clase a cada objeto encontrado dentro de la imagen, pero en cambio trata los múltiples objetos asociados a la misma clase como objetos individuales ver Figura 4, utilizando cuadros delimitadores y creando máscaras para cada instancia. Esto lo convierte en un híbrido de segmentación semántica y detección de objetos (Girshick, Donahue, Darrell , & Malik, 2014; Liu, et al., 2016). Mask R-CNN utiliza instanciación semántica para la identificación de objetos en imágenes. Mask R-CNN es conceptualmente similar a Faster R-CNN, tiene dos salidas para cada objeto candidato, una etiqueta de clase, un cuadro delimitador y una máscara de objeto (He, Gkioxari, Dollar, & Girshick, 2017).

Figura 4. Instanciación semántica de objetos.

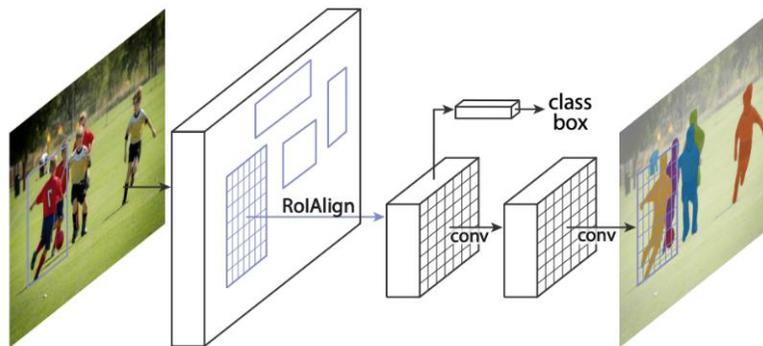


Fuente. (Universitat Oberta de Catalunya, 2016).

2.3.3.1 Arquitectura de Mask R-CNN

La primera etapa, denominada Region Proposal Network (RPN), utiliza una CNN para generar múltiples regiones de interés (RoI) mediante un clasificador binario. Posteriormente, la red RoI Align genera múltiples cuadros delimitadores en lugar de uno único definido y los escala a una dimensión fija. Las RoI escaladas se introducen en capas completamente conectadas para realizar la clasificación utilizando activación de tipo softmax y la predicción del cuadro delimitador utilizando un modelo de regresión (He, Gkioxari, Dollar, & Girshick, 2017). Las regiones también se introducen en el clasificador de máscara, que consta de dos CNN para generar una máscara binaria para cada RoI. Mask Classifier permite a la red generar máscaras para cada clase, ver Figura 5.

Figura 5. El Framework de Mask R-CNN para instanciación semántica.



Fuente: (He, Gkioxari, Dollar, & Girshick, 2017).

2.4 PRODUCCIÓN DE MODELOS DE DEEP LEARNING

Esta etapa se considera como la transición entre el desarrollo de un modelo en entornos experimentales a un modelo listo para ser consumidos por un cliente ya sea incorporado en un programa, microcontrolador o a través de servidores remotos como las plataformas Google Cloud, Amazon Web Services, Microsoft Azure, Docker, Flask y Streamlit utilizando servicios API.

2.4.1 Interfaz de programación de aplicaciones (API)

Las API son un conjunto de comandos, funciones y protocolos informáticos que permiten a los desarrolladores crear programas. Las API simplifican en gran medida el trabajo de un creador de programas, ya que no tiene que escribir códigos desde cero. Estas permiten al desarrollador usar funciones predefinidas para interactuar con el programa (ABC tecnología, 2015). Le otorgan flexibilidad, simplifican el diseño, la administración y el uso de las aplicaciones y proporcionan oportunidades de innovación, lo cual es ideal al momento de diseñar herramientas y productos nuevos (o de gestionar los actuales) (Red Hat, 2021).

2.4.2 Streamlit

Es un framework de Python de código abierto que facilita la creación y el intercambio de aplicaciones web personalizadas para el aprendizaje automático y la ciencia de datos. Streamlit presenta una primitiva de caché ver Figura 6, se comporta como un almacén de datos persistente e inmutable por defecto que permite que las aplicaciones Streamlit reutilicen la información de forma segura (Treuille, 2019).

En resumen, streamlit funciona así:

1. El script completo se ejecuta desde cero para cada interacción del usuario.
2. Streamlit asigna a cada variable un valor actualizado según los estados del widget.
3. El almacenamiento en caché permite a streamlit omitir la obtención y el cálculo de datos redundantes.

Figura 6. Flujo de trabajo de streamlit.



Fuente: (Treuille, 2019).

2.5 MÉTRICAS DE DESEMPEÑO

Las métricas de rendimiento juegan un papel importante en la prueba de los modelos para determinar si cumplen con el objetivo planteado. Para este caso, se utilizaron las siguientes métricas: Intersección sobre la unión (IoU), precisión, especificidad, recall, F1Score y accuracy.

2.5.1 Intersección sobre la unión (IoU)

Relaciona los píxeles de la región segmentada por el modelo y la máscara real. La intersección está compuesta por los píxeles que se encuentran tanto en la máscara de predicción ($B_{predicted}$) como en la máscara real ($B_{ground-truth}$), mientras que la unión se compone de todos los píxeles que se encuentran en la predicción y máscara real (Rezatofighi, et al., 2019).

Ecuación 1. IoU (Intersection over union).

$$IoU_{region}(B_{predicted}, B_{ground-truth}) = \frac{B_{predicted} \cap B_{ground-truth}}{B_{predicted} \cup B_{ground-truth}}$$

2.5.2 Métricas convencionales

Las métricas relevantes en modelos de Deep learning para la clasificación utilizando imágenes son las siguientes: (Borja Robalino, Monleón Getino, & José , 2020).

El accuracy es el porcentaje total de elementos clasificados correctamente, por lo que es directamente proporcional a la suma de los verdaderos positivos y negativos e inversamente proporcional a la sumatoria de todas las imágenes clasificadas, lo que indica que entre mayor sea el accuracy identificara en mayor proporción verdaderos positivos y negativos. Métrica importante para medir el desempeño de modelos que realicen clasificación.

Ecuación 2. Métrica de desempeño: Accuracy.

$$Accuracy = \frac{Ntp+Ntn}{Ntp+Nfp+Ntn++Nfn}$$

Precisión es el número de elementos identificados correctamente como positivos de un total de elementos identificados como positivos, por lo que es directamente proporcional al número de verdaderos positivos e inversamente proporcional a la suma de positivos verdaderos y falsos.

Ecuación 3. Métrica de desempeño: Precisión.

$$Precision = \frac{Ntp}{Ntp+Nfp}$$

Recall es el número de elementos identificados correctamente como positivos del total de positivos verdaderos, por lo que es directamente proporcional al número de verdaderos positivos e inversamente proporcional a la suma de verdaderos positivos y falsos negativos. Métrica importante en aplicaciones médicas, ya que tiene en cuenta los falsos negativos (clasificar un paciente enfermo como sano).

Ecuación 4. Métrica de desempeño: Recall.

$$Recall = \frac{Ntp}{Ntp+Nfn}$$

F1Score fusiona las métricas accuracy con recall y constituye el promedio armónico entre estas dos métricas.

Ecuación 5. Métrica de desempeño: F1Score.

$$F1Score = 2 \frac{precision \times recall}{precision + recall}$$

Especificidad es el número de elementos correctamente identificados como negativos fuera del total de negativos.

Ecuación 6. Métrica de desempeño: Especificidad.

$$Especificidad = \frac{Ntn}{Ntn + Nfp}$$

Donde, Ntp representan los verdaderos positivos, Nfp falsos positivos, Nfn falsos negativos y Ntn verdaderos negativos.

CAPITULO 3. MARCO ANTECEDENTES

3.1 Enfoque general

En un enfoque general, existen diversas investigaciones que se han venido trabajando con la finalidad de ayudar a la medicina por medio de la inteligencia artificial. (Xu, et al., 2015) realizaron un entrenamiento de redes neuronales convolucionales (CNN) para la clasificación y segmentación de imágenes de tumores cerebrales de grandes dimensiones, en donde se utilizaron técnicas de aumento de datos (*data augmentation*), selección de características y agrupación de características, obteniendo una precisión de 84% y del 97,5% en las tareas de segmentación y clasificación.

Las CNN se han usado exitosamente en el apoyo a diagnóstico de varias enfermedades. (Lakhani & Sundaram, 2017) usaron redes neuronales profundas para la clasificación automática de la tuberculosis pulmonar a partir de radiografías de tórax. El modelo se utilizó con fines de clasificación, en donde el conjunto de datos se procesó previamente antes de la evaluación, este modelo tenía un accuracy del 98% y un recall del 97,3%. Por otro lado, en la investigación realizada por Roth, et al., (2014) usaron un modelo CNN para la detección del ganglio linfático en imágenes de Tomografías computarizadas (TC), este modelo obtuvo un recall del 70% y 83%. Otras investigaciones han trabajado con enfermedades pulmonares entre ellas la propuesta por (Anthimopoulos, Christodoulidis, Ebner, Christe, & Mougiakakou, 2016), donde presentaron un modelo CNN para la clasificación de Patrones ILD (Enfermedad pulmonar intersticial). Para ello, propusieron un modelo con 5 capas convolucionales utilizando las activaciones ReLU. El conjunto de datos utilizado contiene siete clases, y el conjunto de datos tiene 14.696 imágenes derivadas de tomografías computarizadas, este modelo tenía una precisión del 85,5%.

3.2 Enfoque específico

En un enfoque específico, se describen investigaciones relacionadas con la detección de neumonía utilizando redes neuronales convolucionales. Existen diversas investigaciones donde se han empleado diferentes técnicas y hacen uso de plataformas como Kaggle, Deep Learning ChexNet, entre otros. Estas plataformas ponen a disposición los dataset que posteriormente serán utilizados para el entrenamiento de la red con el objetivo de detectar la neumonía. Diversos

estudios han usado estos dataset para desarrollar sistemas de apoyo al diagnóstico médico basado en redes neuronales convolucionales, ver Tabla 1.

Las investigaciones realizadas se han enfocado en la clasificación de imágenes de pacientes que padecen neumonía. En (Jain, Nagrath, Kataria, Kaushik, & Hemanth, 2019) se presentan seis redes neuronales convolucionales, el primer y segundo modelo constan de dos y tres capas convolucionales, los otros cuatro modelos son modelos previamente entrenados, que son VGG16, VGG19, ResNet50 y Inception-v3; utilizaron 5.840 imágenes de radiografías de tórax para entrenamiento y test. El primer y segundo modelo lograron una precisión del 85,26% y el 92,31% respectivamente, la precisión de VGG16, VGG19, ResNet50 e Inception-v3 fueron 87,28%, 88,46%, 77,56% y 70,99% respectivamente; los modelos con un mayor accuracy del 85% son VGG16 y VGG19, así mismo el modelo ResNet50 obtuvo un accuracy del 77% y un recall del 97%, mientras que en los otros modelos fue menor a 95%. En (Yu, HuaWang, & DongZhang, 2021) se utilizó el mismo número de imágenes para el entrenamiento y validación, se realizaron pruebas con distintos modelos, donde los más destacados fueron XceptionNet, AlexNet y ResNet101 obteniendo un accuracy mayor a 81%.

En (Militante & Sibbaluca, 2020) desarrollaron 5 modelos de CNN utilizando un mayor número de imágenes para el entrenamiento y validación (26.684 imágenes) con dimensiones de 1024x1024 para los modelos AlexNet, LeNet, GoogleNet, ResNet y VGGNet, además se utilizó *data augmentation* utilizando las técnicas *rotation* de 15 grados, *horizontal flipping*, *zoom* de 0,2, *shifting* de rango de 0,1 para el ancho y el alto de la imagen. Obtuvieron precisiones del 100%, 99%, 100%, 89%, 97% respectivamente y el modelo con mayor recall de 99% fue GoogleNet.

Algunas investigaciones, más allá de clasificar un paciente con o sin neumonía a partir de su imagen diagnóstica, se han enfocado en identificar opacidades en las imágenes de RX, haciendo uso de mapas de calor para la identificación de opacidades. En (Rajpurkar, et al., 2017) se utilizó el dataset "ChestX-ray14" que contiene 112.120 imágenes que consta de 6.643 con neumonía y 60.412 de pacientes sanos. Además, dicho dataset consta de otras 13 enfermedades torácicas. Las imágenes se redimensionaron 224x224, para la validación del modelo utilizaron la métrica AUROC (área bajo la curva) obteniendo 76,80%, este modelo fue comparado con 4 especialistas médicos radiólogos utilizando las mismas imágenes de test donde se destacó el modelo obteniendo en la métrica F1Score 43,5% superando a 3 radiólogos.

Las investigaciones a continuación utilizaron el mismo dataset “ChestX-ray14” y la métrica AUROC. En (Yao, et al., 2017) se utilizó el modelo DenseNet, el cual consta de *DenseBlocks* y *TransitionBlocks*. Dentro de cada *DenseBlock*, hay varios *ConvBlocks* que representan capas convolucionales. Las imágenes se redimensionaron a 512x512, además utilizaron la técnica *data augmentation* y lograron obtener 71,3%. En (Guan, et al., 2018) implementaron el modelo ResNet50 proponiendo una red neuronal convolucional guiada por *attention* (AG-CNN) utilizando mapas de calor generaron una máscara para recortar la región de interés, obtuvieron un desempeño de 77,6%. Esta última investigación obtuvo la métrica mayor en comparación a las anteriores investigaciones que utilizaron el mismo dataset.

En (Sirazitdinov, et al., 2019) se implementó cuadros delimitadores que enmarcan las opacidades encontradas en las imágenes diagnóstico. Utilizaron un dataset que contiene 29.684 imágenes redimensionadas a 512x512 para entrenamiento y test, además se usó *data augmentation* utilizando las transformaciones: *flipping*, *rotation*, *random brightness*, *gamma transforms*, *random Gaussian noise* y *blur*, en los modelos ResNet50 y DenseNet121, obteniendo una precisión del 85% y 88% respectivamente. Además, en el mismo estudio se propuso unir en un solo modelo, la técnica de instancias semánticas (Mask-RCNN) y el modelo de detección RetinaNet por medio del ajuste de pesos dándole prioridad a RetinaNet, obtuvieron una precisión inferior a los anteriores modelos del estudio, pero lograron un recall mayor que corresponde al 79%. Otros estudios utilizaron un dataset que contiene 247 imágenes de RX de 14 instituciones médicas con dos conjuntos de datos de 154 imágenes de nódulos pulmonares y 93 imágenes normales. En (Chen, Han, Lin, Zhao, & Kong, 2020) utilizaron este dataset, y la técnica *data augmentation* con las transformaciones geométricas: *flipping*, *rotation*, voltear imagen y *zoom*, obtuvieron un recall del 97%. En la investigación realizada por (Li, Zhu, Wu, & Wang, 2018) obtuvieron un recall menor que corresponde al 94%.

A lo largo de los años, países latinoamericanos han comenzado a realizar investigaciones con el fin de ayudar a la medicina y detectar la neumonía de forma rápida y precisa. En Colombia, un grupo de investigación de la Universidad del Rosario usó 9.000 imágenes radiológicas de tórax para el desarrollo de una CNN, el cual consiste en segmentar los pulmones del resto de la radiografía, luego se utilizó un algoritmo *k-means* para la clasificación de pacientes sanos y pacientes con diagnóstico de neumonía, logrando así una precisión de 83,3% y una especificidad de 76,6% con respecto al desempeño de la clasificación (Romero, Castellanos, Gutiérrez, & Cancino, 2020). Otro país que se unió a esta causa fue México, teniendo como representantes de la investigación a Roldán, et al., (2019), los autores de este estudio diseñaron una CNN en un ambiente pseudo-distribuido, para la clasificación automática de radiografías de paciente con neumonía.

Utilizaron 1.341 imágenes de RX de pacientes sanos y 3.883 de pacientes con neumonía, con la ayuda de la tecnología de Apache Spark para la configuración de la red y las librerías de dist-keras realizaron el análisis de datos en memoria basada en el procesamiento distribuido, finalmente obtuvieron una precisión del 86%.

Tabla 1. Características generales de los antecedentes

Titulo	Características Generales	Aportes al proyecto	Referencia
Deep convolutional activation features for large scale Brain Tumor histopathology image classification and segmentation.	# imágenes de entrenamiento: 45. Precisión:84 %. Modelo de CNN: Realizado por CognitiveVision team in ImageNet LSVRC 2013.	<ul style="list-style-type: none"> Beneficios de utilizar la segmentación a imágenes médicas. 	(Xu, et al., 2015)
Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks.	# imágenes de entrenamiento: 1007. Precisión:99 %. Modelo de CNN: AlexNet, GoogleNet.	<ul style="list-style-type: none"> Beneficios y desventajas de utilizar AlexNet, GoogleNet. Métodos de desempeño de rendimiento. utilizando el método DeLong. Imágenes para el Dataset. 	(Lakhani & Sundaram, 2017)
Representation for Lymph Node Detection using Random Sets of Deep Convolutional Neural Network Observations.	# imágenes de entrenamiento: 388. Sensibilidad :70% y 83%. Modelo de CNN: CNN propia.	<ul style="list-style-type: none"> Como manejar <i>Support Vector Machine (SVM)</i>. Utilizar la técnica "DropOut". 	(Roth, et al., 2014)
Lung Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural Network.	# imágenes de entrenamiento: 14.696. Precisión :85%. Modelo de CNN: CNN propia 5 capas convolucionales	<ul style="list-style-type: none"> Implementar activaciones LeakyReLU. Implementar Optimizador Adam. Métricas de desempeño convencionales (Accuracy). 	(Anthimopoulos, Christodoulidis, Ebner, Christe, & Mougiakakou, 2016)
Pneumonia detection in chest X-ray images using convolutional neural networks and transfer learning	# imágenes de entrenamiento: 5.840. Precisión :85%, 92%, 87%, 88%, 77% y 71%. Modelo de CNN: modelo 1, modelo 2, VGG16, VGG19, ResNet50 e Inception-v3.	<ul style="list-style-type: none"> Como utilizar los modelos VGG16, VGG19, ResNet-50 y Inception-v3. Utilizar la técnica "DropOut". 	(Jain, Nagrath, Kataria, Kaushik, & Hemanth, 2019)

Titulo	Características Generales	Aportes al proyecto	Referencia
CGNet: A graph-knowledge embedded convolutional neural network for detection of pneumonia.	# imágenes de entrenamiento: 5.840. Accuracy: 85%, 81% y 82%. Modelo de CNN: XceptionNet, AlexNet y ResNet101.	<ul style="list-style-type: none"> Métricas de desempeño convencionales (Sensitivity, Specificity, Accuracy, Precision y F1Score). 	(Yu, HuaWang, & DongZhang, 2021)
Pneumonia Detection Using Convolutional Neural Networks	# imágenes de entrenamiento: 26.684. Accuracy: 100%, 99%, 100%, 89%, 97%. Modelo de CNN: AlexNet, LeNet, GoogleNet, ResNet y VGGNet.	<ul style="list-style-type: none"> Como utilizar los modelos AlexNet, LeNet, GoogleNet, ResNet y VGGNet. Manejo del método ReLU. Utilizar la técnica Data Augmentation. 	(Militante & Sibbaluca, 2020)
Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning.	# imágenes de entrenamiento: 100.000. AUROC: 76 %. Modelo de CNN: ChexNet.	<ul style="list-style-type: none"> Beneficios de utilizar CheXNet. Proporción para las imágenes de entrenamiento y validación. Métodos de desempeño utilizando intersección sobre la unión IoU. 	(Rajpurkar, Irvin, CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays, 2017)
Learning to diagnose from scratch by exploiting dependencies among labels	# imágenes de entrenamiento: 112.000 AUROC: 71 %. Modelo de CNN: DenseNet	<ul style="list-style-type: none"> Utilizar el modelo DenseNet. Utilizar la métrica AUC. 	(Yao, et al., 2017)
Diagnose like a Radiologist: Attention Guided Convolutional Neural Network for Thorax Disease Classification	# imágenes de entrenamiento: 112.000. AUROC: 77 %. Modelo de CNN: ResNet50.	<ul style="list-style-type: none"> Utilizar mapas de calor para clasificación de imágenes. Implementación de la red AG-CNN Utilizar la métrica AUC. 	(Guan, et al., 2018)
Deep neural network ensemble for pneumonia localization from a large-scale chest x-ray database	# imágenes de entrenamiento: 29.684 Precision: 85% y 88%. Modelo de CNN: ResNet50 y DenseNet121.	<ul style="list-style-type: none"> Ensamble de modelos por ajuste de pesos. Utilizar la métrica IoU. Implementar Optimizador Adam. Implementar la técnica cross-entropy. 	(Sirazitdinov, et al., 2019)
Pulmonary nodule detection on chest radiographs using balanced convolutional neural network and classic candidate detection.	# imágenes de entrenamiento: 3.166. Precisión: 90 %. Modelo de CNN: GoogleNet.	<ul style="list-style-type: none"> Técnicas de pre-procesamiento de imágenes. Segmentación de pulmones y nódulos pulmonares. Beneficios de utilizar GoogleNet. 	(Chen, Han, Lin, Zhao, & Kong, 2020)

Titulo	Características Generales	Aportes al proyecto	Referencia
Red neuronal convolucional en un ambiente pseudo-distribuido para la clasificación de radiografías de pacientes con neumonía.	# imágenes de entrenamiento: 154. Precisión: 97,1 %. Modelo de CNN: GoogLeNet.	<ul style="list-style-type: none"> <li data-bbox="1176 293 1701 391">Métricas de desempeño, <i>Accuracy</i>, <i>Sensitivity</i>, <i>Specificity</i>. <li data-bbox="1176 391 1701 488">Utilización del FrameWork Apache Spark para la configuración de redes neuronales convolucionales. <li data-bbox="1176 488 1701 576">Utilizar un ambiente pseudo-distribuido para la clasificación de imágenes de rayos x con neumonía. 	(Roldán & Karina, 2019)

CAPITULO 4. METODOLOGIA

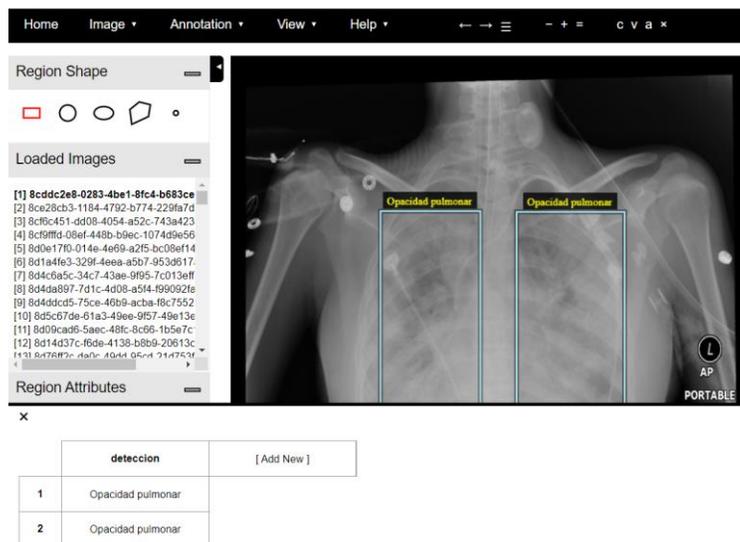
4.1 DATASET

El conjunto de datos utilizado se obtuvo desde la plataforma Kaggle con el nombre “RSNA Pneumonia Detection Challenge” publicado por la RSNA (Sociedad radiológica de norteamericana) en el año 2018, trabajaron en conjunto con la Sociedad de Radiología Torácica y MD.ai que apoyo en el etiquetado de las imágenes (Radiological Society of North America, 2018). Este conjunto de datos contiene 26.636 imágenes en formato (.dcm) de las cuales 20.672 son de imágenes de pacientes sanos y 6.012 de pacientes con neumonía. Se escogieron para esta investigación las 6.012 imágenes de pacientes con neumonía y la misma proporción de imágenes de pacientes sanos convertidas a formato PNG (en total 12.024 imágenes). El dataset se dividió 75% para entrenamiento, 20% para validación de la red y 5% de imágenes para test.

4.1.1 Etiquetado de imágenes

Para el etiquetado de imágenes del dataset, se realizaron anotaciones manualmente. Utilizando el software “VGG Image Annotator” versión 1 que es una plataforma de etiquetado en línea.

Figura 7. Etiquetado de imágenes utilizando el software VGG Image Annotator.



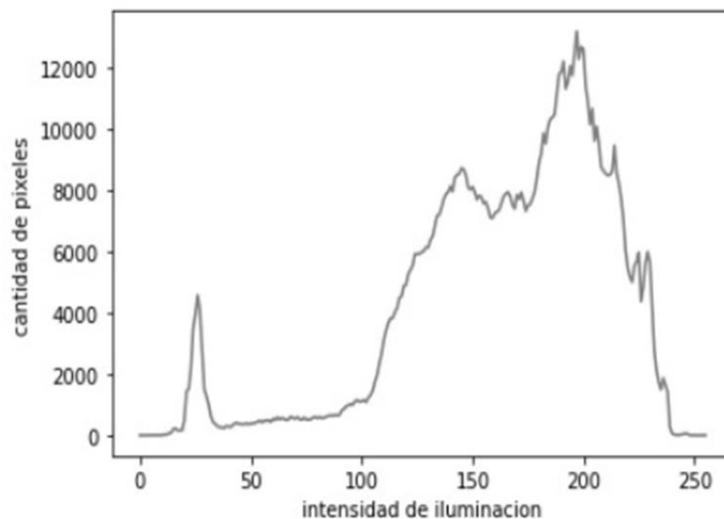
El software permite cargar el dataset y etiquetar imagen por imagen según la región de atributos que correspondan para cada imagen, ver Figura 7. Posteriormente permite exportar los datos del etiquetado en formato JSON, que será el tipo de formato que se le ingresará a la red para entrenamiento, validación y test para los modelos.

4.2 PROCESAMIENTO DE LOS DATOS

En este apartado se realizaron los tratamientos previos a las imágenes antes del entrenamiento de los modelos, como primera instancia se dimensionaron todas las imágenes a 1024x1024 píxeles (Militante & Sibbaluca, 2020).

Posteriormente, se realizó una ecualización de histograma como etapa de pre-procesamiento, en base al histograma de cantidad de píxeles por intensidad de iluminación ver Figura 8, donde se realiza una distribución uniforme de intensidad en todo el rango de píxeles de la imagen, haciendo uso de la librería de OpenCV (Luis, Daniel, & Roberto, 2005).

Figura 8. Histograma de una imagen del dataset inicial.



Ecuación 7. Ecuación de histograma: niveles de grises

$$T(r_k) = \sum_{j=1}^k p_r(r_j) = \sum_{j=1}^k \frac{n_j}{n}$$

Donde, k es el número de niveles de gris, p_r probabilidad de un cierto nivel de gris, r_j un nivel de gris "j" específico, n_j es un número de píxeles con nivel "j", n total de píxeles. Con la Ecuación 7 se busca una distribución uniforme en los valores de intensidad en el histograma mejorando el brillo y contraste de la imagen.

Además, se aplicó la técnica de *data augmentation* al dataset (Taylor & Nitschke, 2018), con el fin de aumentar el tamaño del conjunto de datos, usando las transformaciones: voltear la imagen (*flipping*), aumentar y disminuir el tamaño (*scale*), trasladar los píxeles (*translation*), rotar imagen (*rotate*) y cortar imagen (*shear*) utilizando la librería *imgaug augmenters*, ver Tabla 2.

Tabla 2. Transformaciones geométricas utilizadas en data augmentation.

Transformaciones Geométricas	Unidades de Deformación
Flipping	0,5
Scale	"x": 0,8, 1,2, "y": 0,8, 1,2
Translation	"x": (-0,2, 0,2), "y": -0,2, 0,2
Rotate	-25, 25
Shear	-8, 8

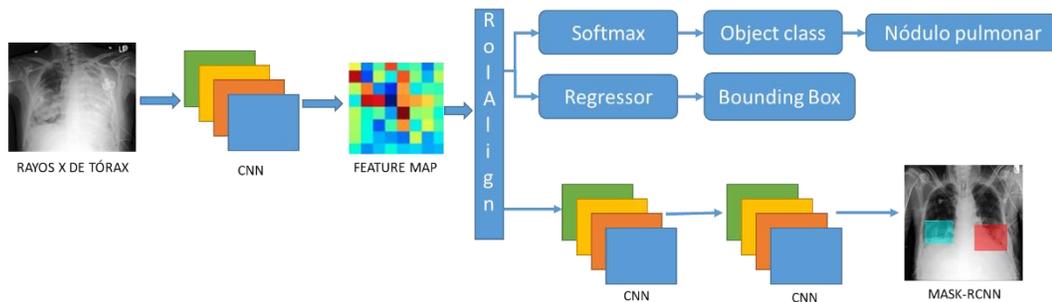
4.3 ARQUITECTURA CNN

Para esta investigación se utilizó la red neuronal convolucional basada en el modelo de instancias semánticas Mask-RCNN (He, Gkioxari, Dollar, & Girshick, 2017), con la finalidad de identificar opacidades pulmonares y clasificar imágenes de radiografías de tórax de pacientes con neumonía ver Figura 9, por lo que en la entrada de la red se le ingresa las imágenes de RX que pasan por las primeras CNN para generar los mapas de características. Posteriormente, la Region Proposal Network (RPN) utiliza una CNN para generar múltiples regiones de interés (RoI) mediante un clasificador binario ligero. Lo hace usando 9 cuadros de anclaje sobre la imagen. Después el clasificador devuelve puntuaciones de objeto o no objeto.

La red RoI Align genera múltiples cuadros delimitadores en lugar de uno único definido y los deforma en una dimensión fija. Luego, las características deformadas se introducen en capas completamente conectadas para hacer la clasificación usando softmax, y la predicción del cuadro de límite se refina aún más usando el modelo de regresión.

Las características deformadas también se introducen en el clasificador de máscara, que consta de dos CNN para generar una máscara binaria para cada RoI. Mask Classifier permite que la red genere máscaras para cada clase. Usando este modelo de red neuronal, se plantearon 3 experimentos con la arquitectura ResNet50 de base, en donde se variaron las características del dataset de entrada. Los detalles de los experimentos se presentan en la Tabla 3.

Figura 9. Modelo basado en Mask-RCNN para identificación de opacidad y segmentación de enfermedades por píxeles.



Para escoger los hiper-parametros para el entrenamiento. Primero, se tomó batch-size de 8 y 10 como máximo considerando las limitaciones en memoria del entorno de entrenamiento, para step epoch se utilizó la Ecuación 8.

Ecuación 8. Step epoch para entrenamiento.

$$step\ epoch = \frac{\# \text{ imagenes de entrenamiento}}{batch\ zise}$$

Para learning rate se tomó de 0,001 considerando los antecedentes de investigaciones anteriores (Rajpurkar, et al., 2017; Jain, Nagrath, Kataria, Kaushik, & Hemanth, 2019; Sirazitdinov, et al., 2019).

Para las epochs se tomó el valor que permita obtener un loss de entrenamiento lo más cercano a 0, además, para el experimento 1 por cada epoch se demoraba 18 minutos, para el experimento 2 y 3 se demora 30 minutos lo que corresponde al tiempo de entrenamiento total que son de 18 Horas para todos los experimentos.

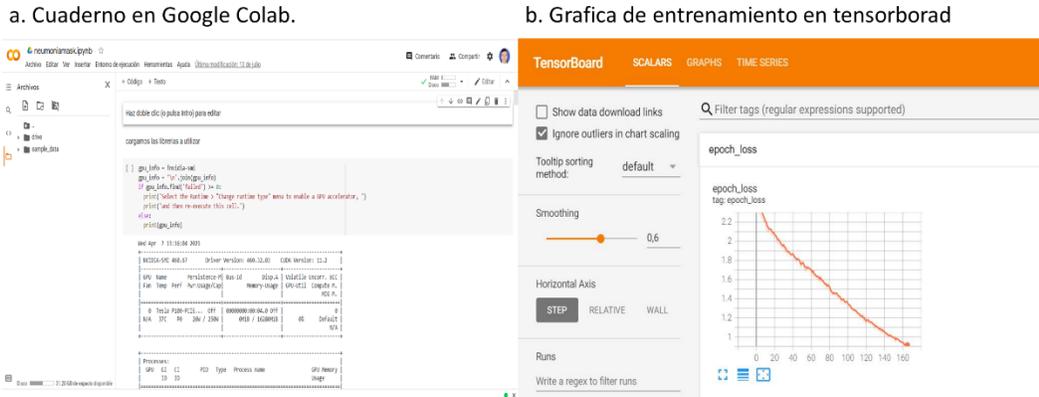
Tabla 3. Parámetros de entrenamiento de la red Mask-RCNN.

	Experimento 1	Experimento 2	Experimento 3
Tamaño Imagen	[1024 1024 3]	[1024 1024 3]	[1024 1024 3]
Arquitectura De La Red	ResNet50		
Batch Size	8	10	10
Datos entrenamiento	Solo imágenes con neumonía (6.010)	Imágenes pacientes con neumonía y sanos, dataset balanceado (12.020)	Imágenes pacientes con neumonía y sanos, dataset balanceado, más <i>data augmentation</i> .
Learning - Rate	0,001	0,001	0,001
Epochs	100	60	60
Steps Epoch	563	900	1000
Tiempo entrenamiento	18H	18H	18H

4.4 ENTRENAMIENTO DE LOS MODELOS

El entrenamiento para los modelos se realizó en Google Colab es un servicio Cloud de Google, basado en los Notebooks de Jupyter, que permite el uso de GPU's y TPU's. Además, permite conectarse a Google Drive para poder visualizar y editar los documentos alojados en el drive ver Figura 10.a, por motivos de desempeño se utilizó la versión PRO ya que permite el uso de la GPU Tesla P100-PCIE de 16 GB, RAM disponible 27,4 GB y tamaño en Disco de 170 GB. Además, se usó TensorBoard para graficar el comportamiento del loss de entrenamiento y validación con respecto a las epochs ver Figura 10.b.

Figura 10. Plataforma de Google Colab.



4.5 MÉTRICAS DE DESEMPEÑO

Se utilizó la media de la medida de intersección sobre la unión (IoU) Ecuación 1, para cuantificar el porcentaje de superposición entre la máscara real y la salida del modelo. Se calculó la métrica IoU para cada una de las imágenes en el conjunto de datos de test, y luego se tomó su promedio, ver Ecuación 9. Además, se calculó la desviación estándar que permite ver la dispersión de los datos respecto a la media, ver Ecuación 11.

Ecuación 9 . Media de la métrica Intersección sobre la Unión para el grupo de test.

$$\bar{x}(IoU) = \frac{\sum_i^{image} IoU_I}{image}$$

Ecuación 10. Media de la métrica intersección sobre la unión para cada imagen.

$$\bar{x}(IoU)_{image} = \frac{\sum_i^{\# opacidades} IoU_I}{\# opacidades}$$

Ecuación 11. Desviación estándar de la métrica Intersección y Unión.

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (X_i - \bar{x}(IoU))^2}$$

Se establecieron los **Verdaderos positivos**: tiene neumonía y el modelo lo clasifica con neumonía, **Verdaderos Negativos**: no tiene neumonía y el modelo lo clasifica sin neumonía, **Falsos Negativos**: tiene neumonía y el modelo lo clasifica sin neumonía, **Falsos Positivos**: no tiene neumonía y el modelo lo clasifico con neumonía.

4.6 UMBRALES DE SELECCIÓN

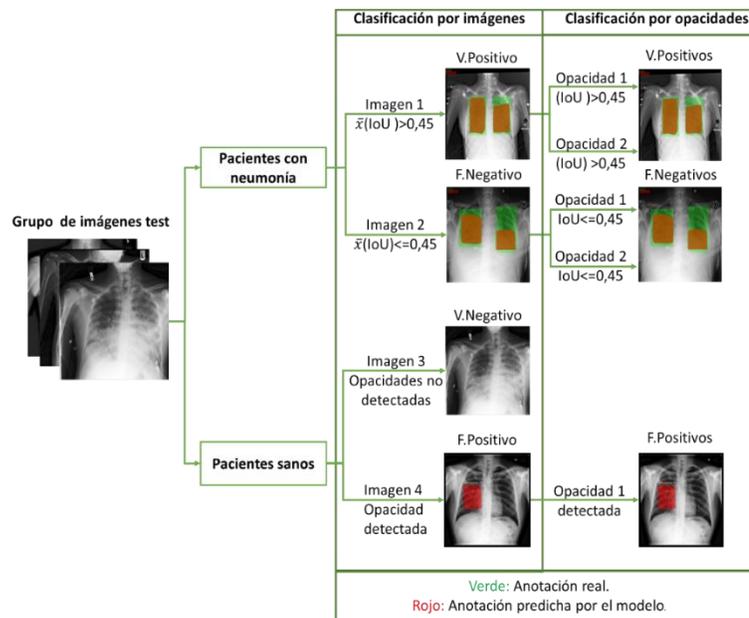
Para el umbral mínimo de selección, se usó el valor de $IoU > 0,45$ tomando como referencia diversas investigaciones (Sirazitdinov, et al., 2019; Jaiswal, et al., 2019), que establecieron valores aceptables cerca del 0,5 para la métrica IoU.

Para identificar el desempeño del sistema en la detección de opacidades, se consideraron como verdaderos positivos todas las opacidades detectadas con un valor de IoU mayor a 0,45. Se consideraron falsos negativos aquellas opacidades que existen en las imágenes de RX y que el sistema no detecta, o detecta con un IoU menor a 0,45. Se consideraron como falsos positivos todas aquellas detecciones del sistema que no existían en las imágenes de RX, ver Figura 11.

Para identificar el desempeño del sistema en la clasificación de pacientes con neumonía y pacientes sanos se utilizó la $\bar{x}(IoU)$ ver Ecuación 10, por cada imagen del dataset de test. El dataset se divide en dos partes: pacientes enfermos y pacientes sanos. En los pacientes enfermos todas las imágenes con $\bar{x}(IoU)$ mayores a 0,45 se consideran verdaderos positivos y las imágenes con $\bar{x}(IoU)$ menor a 0,45 se consideran falsos negativos. Aquellas detecciones del modelo en imágenes RX de pacientes sanos, se consideraron como falsos positivos. Las imágenes Rx de pacientes sanos, donde el sistema no detectó opacidades, se consideraron verdaderos negativos, ver Figura 11.

Finalmente, para los dos casos se utilizaron precisión (Ecuación 3), recall (Ecuación 4), F1Score (Ecuación 5). Accuracy (Ecuación 2) ésta última solo para la clasificación entre pacientes con neumonía y pacientes sanos.

Figura 11. Diagrama de selección de umbrales para las métricas de desempeño.

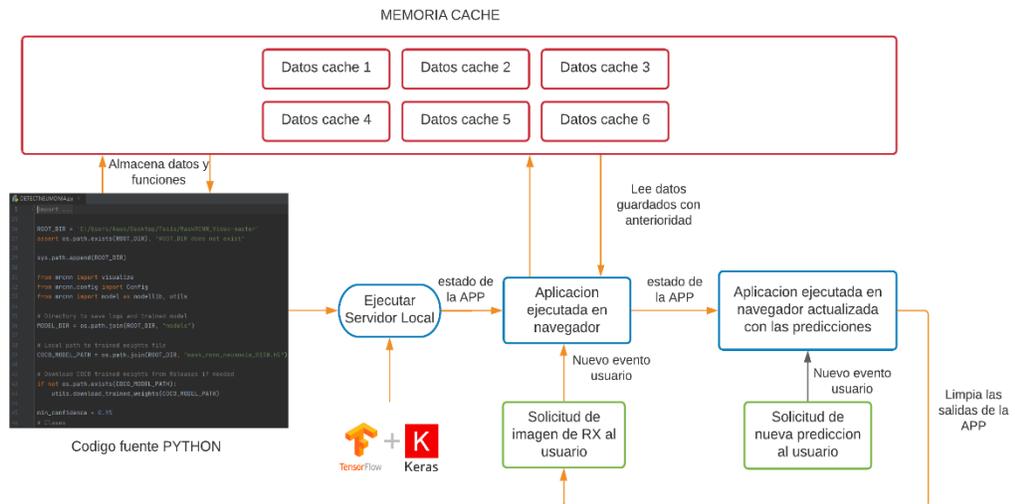


4.7 PUESTA EN PRODUCCIÓN DEL MODELO

Se creó un Dashboard usando el framework streamlit con el uso de Tensorflow-Keras. Streamlit ejecutará las funciones del script principal programado en Python el cual contiene los modelos entrenados y que estarán disponibles para el usuario. La plataforma activará un servidor streamlit local y la aplicación se abrirá en una nueva pestaña en el navegador web predeterminado para que el usuario tenga acceso al APP predictor de neumonía.

El funcionamiento del Dashboard consiste en una APP por solicitudes a usuarios, ver Figura 12. Primeramente, le solicitará una imagen de RX frontal, la APP le permitirá subir la imagen al servidor, de esta manera la APP le muestra la imagen subida y le solicitará al usuario si desea realizar la predicción para identificar opacidades pulmonares asociadas a la neumonía, finalmente la APP segmentará en color si la imagen cargada tiene opacidades y los cuadros delimitadores de las opacidades encontradas. Además, el usuario puede guardar la predicción generada o puede generar nuevas predicciones.

Figura 12. Diagrama de solicitudes de la APP en Streamlit.



CAPITULO 5. RESULTADOS Y DISCUSION

5.1 PRE-PROCESAMIENTO DE IMÁGENES RX

Como pre-procesamiento se realizó ecualización de histograma a las imágenes del dataset como se observa en la Figura 13, se presenta una mejor distribución de la intensidad de iluminación con respecto a los píxeles de la imagen. El efecto de la ecualización mostró mejoras en el brillo y contraste de las imágenes RX, resaltando los detalles como bordes o variaciones en la iluminación y zonas blancas que pueden corresponder a opacidades pulmonares, ver Figura 14.

Figura 13. Histograma de imagen con ecualización.

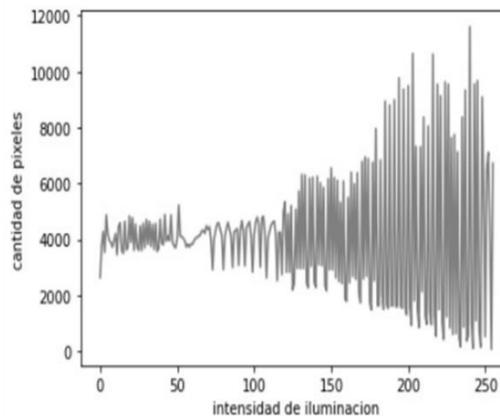
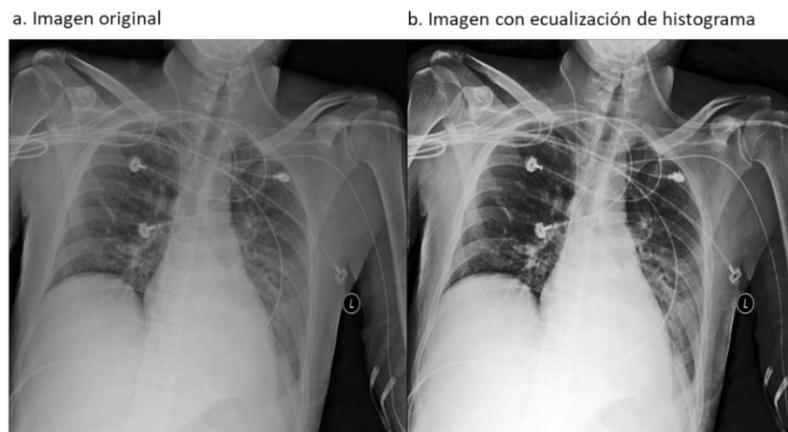
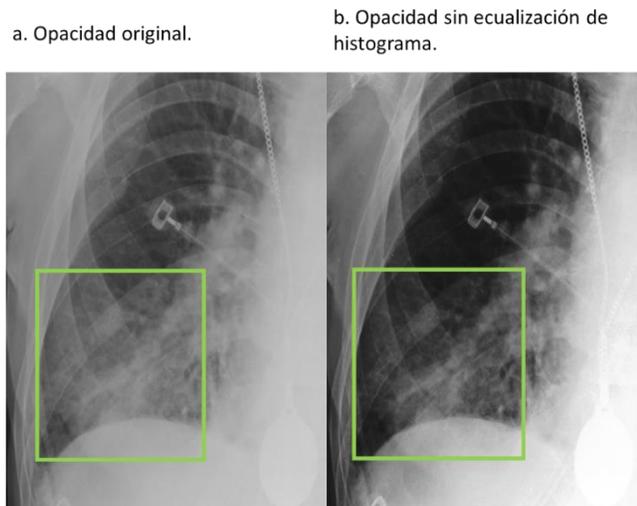


Figura 14. Aplicación de ecualización de histograma al dataset: a. Imagen original, b. Imagen con ecualización de histograma



En la Figura 15 se observa con mayor detalle el resultado de la ecualización de histograma sobre una opacidad de un paciente. En la imagen se puede ver que los bordes y la definición de la opacidad en comparación con la original es mayor, de esta forma podemos concluir que la ecualización del histograma como pre-procesamiento para imágenes con este tipo de aplicaciones es favorable.

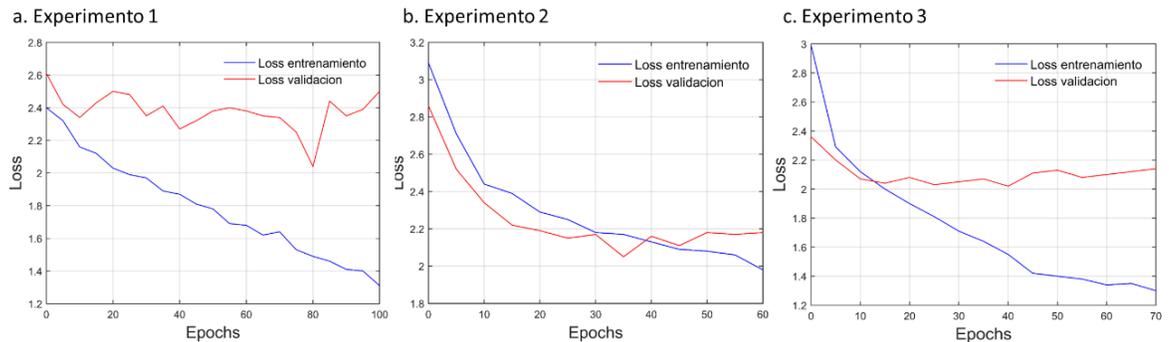
Figura 15. Ecualización de histograma en opacidad: a. Opacidad original, b. Opacidad con ecualización de histograma.



5.2 ENTRENAMIENTO DE LOS MODELOS

Se graficó el comportamiento del loss de entrenamiento y validación para los tres experimentos utilizando tensorboard, se utilizaron las epochs con el menor valor en el loss de validación para los tres experimentos. En la Figura 16.a que corresponde al experimento 1 se escogió la epoch 80, en la Figura 16.b que corresponde al experimento 2 se escogió la epoch 35 y en la Figura 16.c que corresponde al experimento 3 se escogió la epoch 40. En los tres experimentos se presentó overfitting después de las epochs escogidas por lo que se descartaron las posteriores a ellas ya que el sistema se está sobre-entrenando.

Figura 16. Loss entrenamiento y validación de los experimentos.



En (Jain, Nagrath, Kataria, Kaushik, & Hemanth, 2019) en las arquitecturas VGG16, VGG19, Inception-v3 y ResNet utilizaron las técnicas de data augmentation y dropout, que permite desconectar neuronas durante el proceso de entrenamiento. Durante la validación se presentó overfitting, en las arquitecturas de ResNet e Inception-v3. En (Jaiswal, et al., 2019) utilizando las arquitecturas de ResNet con Mask-RCNN y (Anthimopoulos, Christodoulidis, Ebner, Christe, & Mougiakakou, 2016) utilizando su arquitectura propia presentaron overfitting, para reducirlo implementaron data augmentation. En nuestro caso se presentó overfitting en todos los modelos después de las epochs escogidas, el modelo 3 utilizó la técnica data augmentation donde se redujo el sobre-ajuste en el proceso de entrenamiento del modelo, ya que obtuvo menor valor en el loss de validación y entrenamiento, en comparación con el modelo 2 que no utilizó data augmentation.

5.3 DETECCIÓN DE OPACIDADES MASK-RCNN

Se evaluaron los modelos obtenidos en el proceso de entrenamiento correspondiente a los 3 experimentos antes descritos, cada uno de ellos se evaluó usando las métricas presentadas en la sección anterior. Se estimó el IoU para cada imagen de test usando los tres modelos obtenidos. En la Figura 17.a se muestran las etiquetas que se asignaron manualmente. En la Figura 17.b se muestran las máscaras predichas por el modelo Mask-RCNN, y en la Figura 17.c se observa la unión (color verde) y la intersección (color rojo) de las etiquetas reales y las predichas por el modelo. En la Figura 18 se muestran las opacidades pulmonares segmentadas encontradas en imágenes diagnósticas.

Figura 17. Aplicación de la métrica IoU individual: a. Anotación máscara, b. Resultado máscara, c. Unión e intersección.

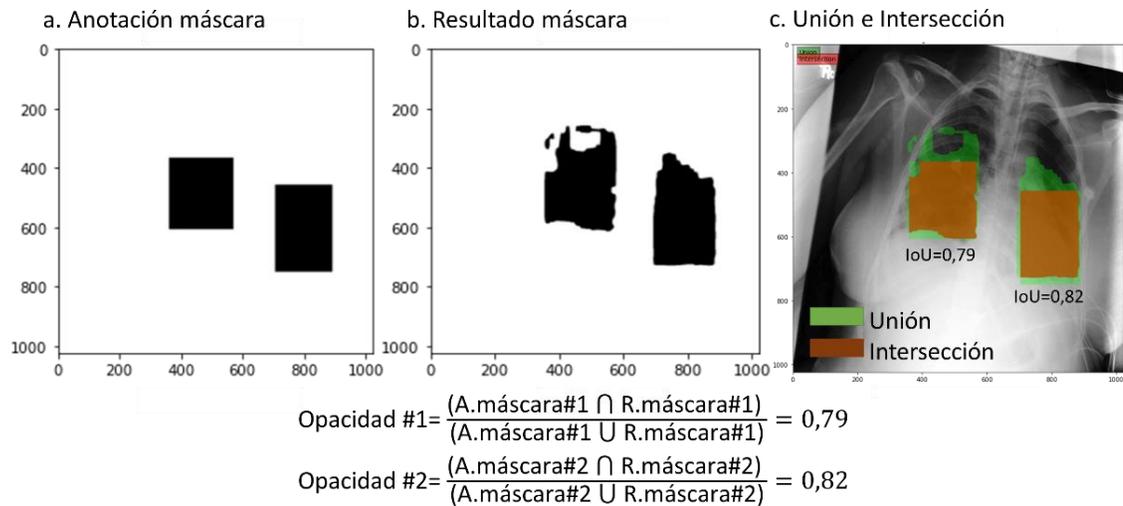
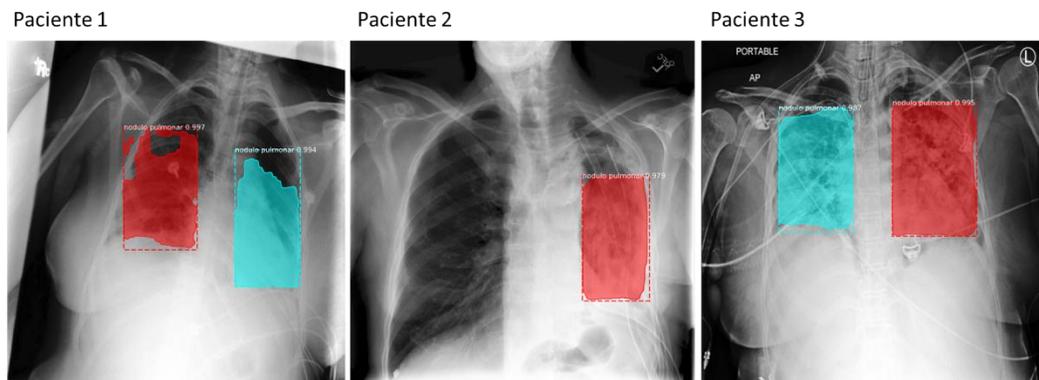
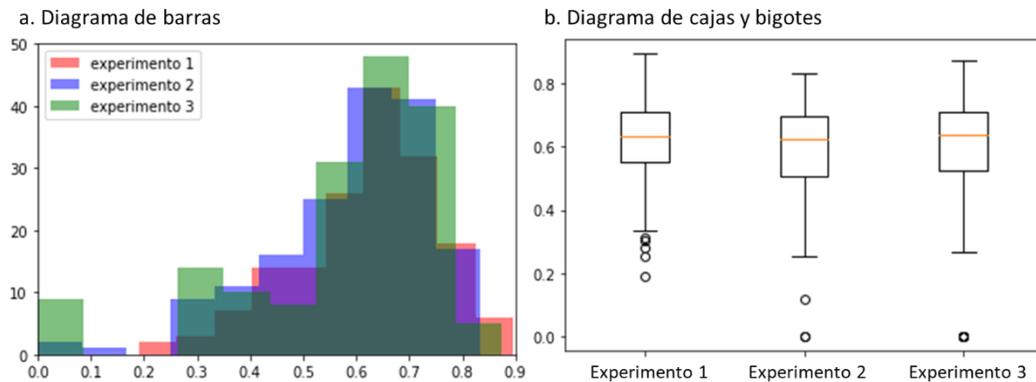


Figura 18. Opacidades pulmonares segmentadas en pacientes.



En la Figura 19.a se presenta la distribución de valores de IoU de los tres experimentos utilizando el mismo grupo de imágenes de test, en la Figura 19.b se observa la dispersión y la simetría de los datos. Se calculó además la media total de las IoU y la desviación estándar de los datos, para cada modelo. Para el experimento 1 $\bar{x}(IoU)$: 0,6207 y σ :0,1336, el experimento 2 $\bar{x}(IoU)$: 0,5882 y σ :0,1523 y el experimento 3 $\bar{x}(IoU)$: 0,5777 y σ :0,1953. Con base en los gráficos y la desviación estándar se infiere que las IoU obtenidas de los tres experimentos se centran en unos valores cercanos en el rango de 0,6 y 0,7, lo que significa que los 3 experimentos obtuvieron un comportamiento similar en la identificación de verdaderos positivos ya que el valor central es mayor al establecido (>0,45).

Figura 19. Diagramas de desempeño de la métrica IoU global de grupo de imágenes test.



Se aplicó la prueba Kolmogorov-Smirnov para los datos IoU asociados a las imágenes de prueba para cada experimento. Con este test estadístico se evaluó el tipo de distribución estadística de los datos (Romero Saldaña, 2016). Se utilizó para comprobar si los datos de las IoU obtenidos de los pacientes con neumonía de los tres experimentos se distribuyen normalmente o son no normal.

Tabla 4. Resultados de la prueba de Kolmogorov-Smirnov de una muestra

	Valor de H	Nivel de Significancia(α)	Probabilidad (P)
Experimento 1	1	5%	$2,3287 \times 10^{-53}$
Experimento 2	1		$3,6274 \times 10^{-50}$
Experimento 3	1		$1,7263 \times 10^{-44}$

En la Tabla 4, se muestran los valores para los tres experimentos donde se obtuvo $H=1$, lo que indica que los datos de las IoU de los experimentos son de distribución no normal, por lo que se utilizó una prueba no paramétrica para realizar la prueba de hipótesis a los tres experimentos. Se utilizó la prueba Kruskal-Wallis, que compara las medianas de los grupos de datos, para determinar si las muestras provienen de la misma población (o, de manera equivalente, de diferentes poblaciones con la misma distribución) para dicha prueba se utilizó un nivel de significancia de 1% (Amat Rodrigo, 2016).

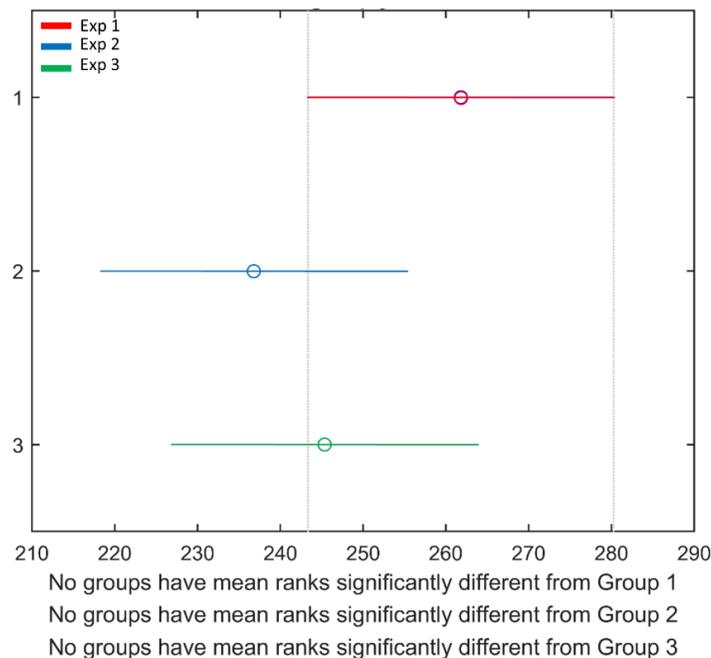
Tabla 5. Resultados de la prueba de Kruskal-Wallis para datos no paramétricos.

	Suma de Cuadrados (SS)	Grados de Libertad (df)	Media Cuadrática (Ms)	Chi Cuadrado (χ^2)	Probabilidad (χ^2)
Datos	53318,4	2	26659,2	2,61	0,2117
Error	10053811,6	492	20434		
Total	10107130	494			

La fila datos, representa la variabilidad debido a las diferencias entre las medias de los grupos (variabilidad entre los grupos), y la fila de error, la variabilidad debido a las diferencias entre los datos de cada grupo y la media del grupo (variabilidad dentro de los grupos).

En la Tabla 5, se observa el valor de probabilidad de 0,2117, donde se evidencia que es mayor al 0,01 del nivel de significancia, lo que indica que no se rechaza la hipótesis nula, lo que quiere decir que los grupos provienen de una misma distribución.

Figura 20. Resultados de grupos individuales de la prueba Kruskal-Wallis.



Como se observa en la Figura 20, no existe una diferencia significativa entre los experimentos, por lo que también se concluye que la prueba no rechaza la hipótesis nula, dando como resultado que los tres experimentos provienen de la misma distribución. Es decir, los resultados muestran que la capacidad de segmentar opacidades de los modelos obtenidos en los 3 experimentos no presenta diferencias significativas en un nivel general.

5.4 DESEMPEÑO DE LOS MODELOS EN LA DETECCIÓN DE OPACIDADES

En la Tabla 6, se muestran las métricas de desempeño que se obtuvieron de los experimentos para predecir opacidades pulmonares. El experimento 3 fue el de mayor precisión, además, fue la configuración que utilizó el grupo de entrenamiento de imágenes balanceado más *data augmentation*. La métrica recall, es una medida relevante en modelos que involucran aplicaciones médicas, ya que es importante que el modelo no clasifique pacientes enfermos como sanos. El experimento 1 fue el de mayor recall, dicho experimento solo se entrenó con imágenes de neumonía. Este modelo tuvo un mejor desempeño para identificar opacidades pulmonares en los pacientes enfermos, disminuyendo los falsos negativos, pero aumentando los posibles falsos positivos en las imágenes.

Tabla 6. Métricas de desempeño de los experimentos para identificar opacidades pulmonares.

MASK-RCNN	Precisión	Recall	F1 Score
Experimento 1	0,7627	0,9513	0,8466
Experimento 2	0,7164	0,8801	0,7899
Experimento 3	0,8075	0,8801	0,8422

Figura 21. Predicción de los modelos en la detección de opacidades.



A pesar del buen desempeño de los modelos en la métrica recall en la detección de opacidades, el sistema presentó algunas detecciones erróneas en algunas de las imágenes, asociadas a falsos negativos (paciente con opacidad y el modelo lo clasifico como sano) como se observa en la Figura 21. Para los modelos usados en aplicaciones médicas se pretende que los falsos negativos sean mínimos. La detección de opacidades en algunas ocasiones representa un reto de cierta complejidad, debido al color blanco particular emitido en las imágenes de RX que puede llegar a confundirse con los órganos o hasta con las costillas (Pezzotti, 2014). Se podría mejorar los resultados si se incluyen más imágenes con esas características expuestas anteriormente, aumentando el dataset de entrenamiento como también se plantea en (Yu, HuaWang, & DongZhang, 2021; Sirazitdinov, et al., 2019), modificando los parámetros de data augmentation para el experimento 3 (Jaiswal, et al., 2019), segmentando los pulmones en la etapa de pre-procesamiento (Chen, Han, Lin, Zhao, & Kong, 2020).

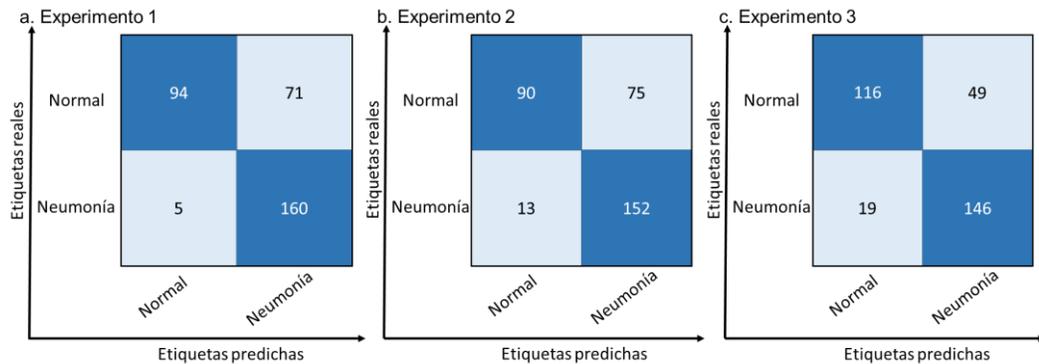
Se debe tener cuenta que el especialista medico es la persona que toma la última decisión para el diagnóstico de paciente, el sistema solo le brinda apoyo dándole información que obtuvo a partir de las imágenes de RX del paciente.

5.5 DESEMPEÑO DE LOS MODELOS EN LA DETECCIÓN DE NEUMONÍA

Se estimaron las métricas para la clasificación de imágenes con neumonía utilizando el mismo grupo de imágenes test. En la Figura 22, se observa la matriz de confusión para cada experimento. En las matrices de confusión se identificó que el experimento 1 y 2 lograron identificar un mayor número de aciertos de verdaderos positivos lo que indica que identificaran un mayor número de pacientes con neumonía. Por el contrario, el experimento 3 obtuvo un menor número de aciertos en verdaderos positivos, pero mejoró en disminuir los falsos positivos lo que indica que el modelo con neumonía y el paciente es sano. Así mismo, fue el que obtuvo el mayor número de aciertos en verdaderos negativos, lo que significa que logró identificar un mayor número de pacientes sanos en comparación con los otros experimentos.

Por último, el experimento 1 fue el que obtuvo el menor valor en falsos negativos, parámetro importante en aplicaciones médicas que buscan no clasificar pacientes enfermos como pacientes sanos.

Figura 22. Matriz de confusión de los experimentos: a. Experimento 1, b. Experimento 2, c. Experimento 3.



En la Tabla 7 se presentan las métricas de desempeño que tuvieron los experimentos para clasificar imágenes de pacientes con neumonía y pacientes sanos. El modelo del experimento 3 presentó el mayor accuracy en comparación con los modelos obtenidos de los experimentos 1 y 2, en consecuencia, este identificará en mayor proporción pacientes enfermos y sanos correctamente. Además, el experimento 1 fue el que obtuvo el recall más alto métrica importante en aplicaciones médicas, lo que significa que identificará en menor proporción falsos negativos.

Tabla 7. Métricas de desempeño de los experimentos para clasificar imágenes de neumonía.

MASK-RCNN	Precisión	Recall	F1 Score	Accuracy
Experimento 1	0,7026	0,9696	0,8080	0,7696
Experimento 2	0,6696	0,9212	0,7755	0,7333
Experimento 3	0,7487	0,8848	0,8111	0,7939

El tamaño y la configuración del dataset de entrenamiento han mostrado que influyen en el desempeño del modelo. Existen varios estudios donde usaron el dataset de RSNA (26.684 imágenes) para la detección de opacidades pulmonares y neumonía. En (Militante & Sibbaluca, 2020) se redimensionó el dataset a 1024x1024 y realizaron *data augmentation* utilizando las técnicas: *rotation*, *horizontal flipping* y *scale*; los resultados mostraron una precisión del 89% utilizando ResNet50. Así mismo en (Sirazitdinov, et al., 2019) redimensionaron las imágenes a 512x512, entrenaron Mask-RCNN solo con imágenes de neumonía y usaron RetinaNet con el dataset completo. Integrando los dos modelos obtuvieron una precisión del 76%. En (Wang, et al., 2017) utilizaron el dataset “ChestX-ray14” que

contiene 112.120 imágenes que consta de 6.643 con neumonía y 60.412 de pacientes sanos, obtuvieron una precisión del 89%. En comparación con los resultados obtenidos en el estudio aquí presentado, el modelo obtenido en el exp1 con la configuración del dataset de solo imágenes con neumonía fue superior en el recall a los referentes consultados, con un valor de 96%.

En (Jain, Nagrath, Kataria, Kaushik, & Hemanth, 2019) se utilizó un dataset de 5.840 imágenes con dimensiones de 64x64, además se empleó la técnica dropout para reducir el sobre-ajuste del modelo ResNet50 obteniendo una accuracy 77,56% y recall del 97%. En comparación con el modelo obtenido del exp3 aquí presentado, con la configuración del dataset balanceado más *data augmentation*, los referentes mostraron un accuracy menor, dado que el resultado obtenido en esta investigación fue del 79%.

5.6 PUESTA EN PRODUCCIÓN DEL PREDICTOR DE NEUMONÍA

Se utilizó el *framework* Streamlit para montar un servidor local, en la Figura 23 se observa el *main* principal del predictor APP, que le permite al usuario montar o subir una imagen de RX frontal al servidor para su posterior predicción ver Figura 24.

Figura 23. Main principal del Predictor de neumonía APP.

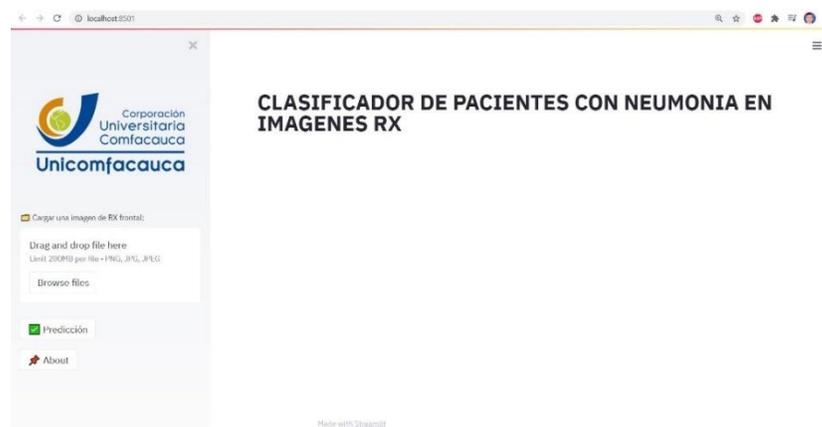
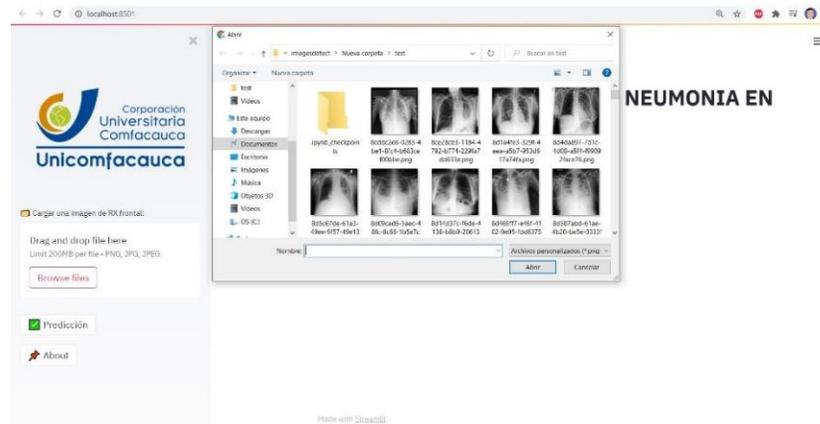


Figura 24. Cargar imagen diagnostico a la APP.



En la Figura 25 se muestra el resultado de la predicción de la APP. Si el paciente padece neumonía muestra la imagen diagnostico indicando las opacidades pulmonares asociadas a neumonía de manera segmentada (por mascarar) y por cuadros delimitadores con sus respectivos porcentajes de precisión del modelo. Por otro lado, si el paciente es sano mostrará nuevamente la imagen diagnostico indicando que la predicción es paciente sano. Por último, el predictor le permite al usuario guardar los resultados para un posterior análisis.

Figura 25. Predictor APP: a. Resultado del predictor, b. Resultado predictor: imagen segmentada y por cuadros delimitadores.



En la Tabla 8, se observan los tiempos de predicción de neumonía utilizando la APP. Se escogió una muestra de 10 imágenes diagnóstico diferentes para las cuales se calculó el tiempo de diagnóstico individual. Se obtuvo un promedio de 4,59s que sería el tiempo estimado en realizar el diagnóstico de neumonía de una imagen con la APP.

Tabla 8. Tiempos de predicción de la APP.

# Paciente	1	2	3	4	5	6	7	8	9	10	$\bar{x}(t)$
Tiempo(s)	5,11	4,75	4,41	4,32	4,45	4,46	4,77	4,75	4,42	4,52	4,59

Por medio de esta APP se pretende llegar a médicos del país, para que pueda ser un apoyo en el diagnóstico de neumonía para reducir tiempos de detección de la enfermedad. El sistema de predicción, partiendo de la imagen de RX y de manera inmediata puede obtener los resultados, identificando opacidades pulmonares y detección de neumonía, brindando apoyo a los especialistas médicos en la lectura de las imágenes diagnóstico. De esta manera, el especialista puede generar tratamientos oportunos al paciente y además diagnosticar a un mayor número de pacientes con ayuda del predictor APP.

6. CONCLUSIONES

- En esta investigación se presentó el modelo Mask-RCNN para la identificación de opacidades pulmonares y para la clasificación de pacientes con neumonía, usando imágenes RX. Se desarrollaron 3 experimentos con diferentes configuraciones en el dataset. El modelo que obtuvo mayor precisión (80% y 74%), para la detección de opacidades y neumonía, fue el modelo asociado al experimento 3, donde se usó un dataset balanceado y *data agumentation en comparación con los estudios presentados por* (Militante & Sibbaluca, 2020) que obtuvo una precisión del 89%, en (Sirazitdinov, et al., 2019) obtuvieron una precisión del 76% y (Wang, et al., 2017) obtuvieron una precisión del 89%.
- El tamaño del dataset, la simetría del mismo y además el uso de técnicas de aumento de imágenes de entrenamiento como *data augmentation* influenciaron en el desempeño de los modelos, para el experimento 3 donde se utilizó el dataset simétrico y *data augmentation*, se obtuvo el mayor accuracy (79%) para la detección de neumonía lo que permitirá clasificar de manera correcta pacientes con neumonía en comparación con los otros experimentos que obtuvieron unos accuracy del 76% y 73% respectivamente para los experimentos 1 y 2.
- Considerando la importancia de la minimización de falsos negativos en los sistemas de detección y clasificación en aplicaciones médicas, se prioriza el modelo con mayor recall sobre los demás. El mejor modelo bajo esa perspectiva con recall (95% y 96%), para la detección de opacidades y neumonía, fue el modelo asociado al experimento 1 donde se usó solamente imágenes RX de pacientes con neumonía.
- La aplicación de la prueba Kruskal-Wallis a los datos obtenidos de las IoU de las imágenes de pacientes con neumonía, dio como resultado una probabilidad de 0,2117, lo que indica que los datos de las IoU obtenidos de los tres experimentos provienen de una misma distribución.
- La implementación del detector de neumonía APP con streamlit demostró el buen desempeño que pueden llegar a tener un predictor web, por su interfaz amigable con el cliente y la rapidez en los resultados del predictor, teniendo un tiempo promedio de diagnóstico de neumonía de 4,59s por paciente.

- Estos resultados sugieren que las técnicas de instancias semánticas pueden ser útiles en el apoyo al equipo médico para el diagnóstico de la neumonía, agilizando los procesos clínicos de diagnóstico y tratamiento de neumonía haciendo uso de aplicativos por medio de servidores web que permitan a médicos del país acceder a este tipo de desarrollos tecnológicos.

7. RECOMENDACIONES

Considerando la importancia que tiene esta investigación en el ámbito de desarrollos tecnológicos que apoyen a diagnósticos médicos, además en función de los resultados obtenidos se proponen las siguientes recomendaciones:

- Referente al dataset se sugiere aumentar la cantidad de imágenes de entrenamiento, además considerar el uso de meta-información: como el sexo, edad y antecedentes del paciente, como características adicionales útiles que pueden ayudar al apoyo del diagnóstico de neumonía y por último Incorporar a la investigación médicos especialistas o centros médicos que brinden apoyo en el etiquetado de nódulos pulmonares asociadas a la neumonía.
- Se propone el uso de otras arquitecturas de redes neuronales convolucionales como: LeNet, GoogleNet, y VGGNet, VGG16 y VGG19 como los especifican algunos autores que implementaron estas arquitecturas y obtuvieron excelente desempeño: (Militante & Sibbaluca, 2020), (Wang, et al., 2017) y (Jain, Nagrath, Kataria, Kaushik, & Hemanth, 2019), o modelos de segmentación semántica como U-Net, que puedan proporcionar herramientas adicionales para el apoyo del diagnóstico de neumonía.
- Para el predictor de neumonía APP se sugiere implementar un servidor que no solo se ejecute de manera local, de esta forma cualquier médico por medio del enlace de conexión con el servidor pueda acceder al predictor APP.

8. BIBLIOGRAFÍA

- Borja Robalino, R., Monleón Getino, A., & José , R. (2020). Estandarización de métricas de rendimiento para clasificadores Machine y Deep Learning. *Iberian Journal of Information Systems and Technologies*, 172–184.
- Chen, S., Han, Y., Lin, J., Zhao, X., & Kong, P. (2020, 7). Pulmonary nodule detection on chest radiographs using balanced convolutional neural network and classic candidate detection. *Artificial Intelligence in Medicine*, 107(101881). doi://doi.org/10.1016/j.artmed.2020.101881
- Chen, S., Han, Y., Lin, J., Zhao, X., & Kong, P. (2020, 7). Pulmonary nodule detection on chest radiographs using balanced convolutional neural network and classic candidate detection. *Artificial Intelligence in Medicine*, 107(101881). doi:https://doi.org/10.1016/j.artmed.2020.101881
- Fukushima, K. (1988). Neocognitron: A hierarchical neural network capable of visual pattern recognition. *Neural Networks*, 119-130. doi:https://doi.org/10.1016/0893-6080(88)90014-7
- Girshick, R., Donahue, J., Darrell , T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, (pp. 580-587). doi:https://doi.org/10.1109/CVPR.2014.81
- Guan, Q., Huang, Y., Zhong, Z., Zheng, Z., Zheng, L., & Yang, Y. (2018). Diagnose like a Radiologist: Attention Guided Convolutional Neural Network for Thorax Disease Classification. *ArXiv*, 1801.09927. Retrieved 05 20, 2021, from <https://arxiv.org/abs/1801.09927>
- He, K., Gkioxari, G., Dollar, P., & Girshick, R. (2017). Mask R-CNN. *2017 IEEE International Conference on Computer Vision (ICCV)*, (pp. 2980-2988). doi:https://doi.org/10.1109/ICCV.2017.322
- Lakhani, P., & Sundaram, B. (2017, 4 24). Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks. *Radiology*, 284(2), 9. doi:https://doi.org/10.1148/radiol.2017162326
- LeCun, Y., Bottou, L., Orr, G., & Müller, K. R. (1998). Efficient BackProp. In O. G. Montavon G., *Neural Networks: Tricks of the Trade*. Berlin, Heidelberg: Lecture Notes in Computer Science. doi:https://doi.org/10.1007/978-3-642-35289-8_3
- Li, C., Zhu, G., Wu, X., & Wang, Y. (2018). False-Positive Reduction on Lung Nodules Detection in Chest Radiographs by Ensemble of Convolutional Neural Networks. *IEEE Access*, 6, 16060-16067. doi:10.1109/ACCESS.2018.2817023

- Li, X., Shen, L., & Luo, S. (2018). A Solitary Feature-Based Lung Nodule Detection Approach for Chest X-Ray Radiographs. *IEEE Journal of Biomedical and Health Informatics*, 22(2), 516-524. doi:10.1109/JBHI.2017.2661805
- Liu, W., Dragomir, A., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., & Berg, A. (2016). *SSD: Single Shot MultiBox Detector*. Retrieved 07 13, 2021, from <https://arxiv.org/pdf/1512.02325.pdf>
- Luis, F., Daniel, D., & Roberto, D. (2005). Optimización de la ecualización del histograma en el procesamiento de imágenes digitales. *VII Workshop de Investigadores en Ciencias de la Computación*. Florencio Varela. Retrieved 05 21, 2021
- Pezzotti, W. (2014). Interpretación de la radiografía de tórax: algo más que blanco y negro. *Nursing*, 31(5), 30-37. doi:DOI: 10.1016/j.nursi.2014.10.012
- Romero, A., Castellanos, C., Gutiérrez, A., & Cancino, S. (2020). *ALGORITMO PARA LA CARACTERIZACIÓN DE NEUMONÍA EN RADIOGRAFÍA DE TÓRAX*. Universidad del Rosario, Ingeniería. Bogotá: Universidad del Rosario. Retrieved 10 25, 2020
- Silva, S., & Freire, E. (23 de 11 de 2019). *Bootcamp AI*. Obtenido de <https://bootcampai.medium.com/redes-neuronales-convolucionales-5e0ce960caf8>
- Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., & Summers, R. (2017). ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (pp. 3462-3471). Retrieved 05 21, 2021, from <https://arxiv.org/abs/1705.02315>
- Yu, X., HuaWang, S., & DongZhang, Y. (2021). CGNet: A graph-knowledge embedded convolutional neural network for detection of pneumonia. *Information Processing & Management*, 58(1), 25. doi:<https://doi.org/10.1016/j.ipm.2020.102411>
- ABC tecnología. (2015, 02 16). *ABC*. Retrieved 07 13, 2021, from <https://www.abc.es/tecnologia/consultorio/20150216/abci--201502132105.html?ref=https%3A%2F%2Fwww.google.com%2F>
- Amat Rodrigo, J. (2016). *Under a Attribution 4.0 International (CC BY 4.0)*. Retrieved 08 31, 2021, from https://www.cienciadedatos.net/documentos/20_Kruskal-Wallis_test.html
- Anthimopoulos, M., Christodoulidis, S., Ebner, L., Christe, A., & Mougiakakou, S. (2016). Lung Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural Network. *IEEE TRANSACTIONS ON MEDICAL IMAGING*, 35(5), 1207-1216. doi: doi: 10.1109/TMI.2016.2535865

- Biblioteca Nacional de Medicina de EE. UU. (2018, 6 16). *MedlinePlus*. (Biblioteca Nacional de Medicina de EE. UU) Retrieved 9 7, 2020, from <https://medlineplus.gov/spanish/ency/article/000145.htm#:~:text=Las%20formas%20como%20se%20puede,microbios%20directamente%20hacia%20los%20pulmones.>
- Biblioteca Nacional de Medicina de EE. UU. (2020). *MedlinePlus*. Retrieved 05 25, 2021, from <https://medlineplus.gov/spanish/pruebas-de-laboratorio/prueba-de-cultivo-de-bacterias/>
- Biblioteca Nacional de Medicina de EE.UU. (2019, enero 26). *MedlinePlus*. Retrieved octubre 28, 2020, from <https://medlineplus.gov/spanish/ency/article/003723.htm>
- Biblioteca Nacional de Medicina de EE.UU. (2020, 3 24). *MedlinePlus*. (Biblioteca Nacional de Medicina de EE. UU.) Retrieved 10 28, 2020, from <https://medlineplus.gov/spanish/pruebas-de-laboratorio/conteo-sanguineo-completo/>
- Biblioteca Nacional de Medicina de EE.UU. (2020, 2 24). *MedlinePlus*. Retrieved octubre 28, 2020, from <https://medlineplus.gov/spanish/pruebas-de-laboratorio/prueba-de-cultivo-de-bacterias/>
- CanalSalud. (2016). *MAPFRE*. Retrieved 05 31, 2021, from <https://www.salud.mapfre.es/pruebas-diagnosticas/respiratorias-pruebas/analisis-del-esputo/>
- Dadonaite, B., & Roser, M. (2019, 11). *Our World in Data*. (OurWorldInData.org, Editor) Retrieved 12 1, 2020, from <https://ourworldindata.org/pneumonia>
- Gómez Hernandez, E., González Santacruz, E., & Ibarguen Mondragón, E. (2014, julio). Elementos para la modelación matemática de la epidemiología de la neumonía. *Universidad y Salud*, 16(2), 9. Retrieved 09 2020, 7, from http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0124-71072014000200009
- Jain, R., Nagrath, P., Kataria, G., Kaushik, S., & Hemanth, D. (2019, 5 21). Pneumonia detection in chest X-ray images using convolutional neural networks and transfer learning. *Measurement*, 165, 10. Retrieved 04 10, 2021, from <http://www.sciencedirect.com/science/article/pii/S0263224120305844>
- Jaiswal, A. K., Tiwari, P., Kumar, S., Gupta, D., Khanna, A., & Rodrigues, J. (2019). Identifying pneumonia in chest X-rays: A deep learning approach. *Measurement*, 145, 511-518. doi:<https://doi.org/10.1016/j.measurement.2019.05.076>
- Jaiswal, A. K., Tiwari, P., Kumar, S., Gupta, D., Khanna, A., & Rodrigues, J. (2019). Identifying pneumonia in chest X-rays: A deep learning approach.

- Measurement*, 145, 511-518.
doi:<https://doi.org/10.1016/j.measurement.2019.05.076>
- Loyola Medicine. (2020). *Loyola Medicine*. Retrieved 05 31, 2021, from <http://loyolamedicine.adam.com/content.aspx?productid=118&pid=5&gid=003642>
- Mayo Clinic. (2020, mayo 2). *Mayo clinic*. Retrieved octubre 28, 2020, from <https://www.mayoclinic.org/es-es/tests-procedures/chest-x-rays/about/pac-20393494>
- Mayo Clinic. (2020, Mayo 5). *Mayo Clinic*. Retrieved Octubre 12, 2020, from <https://www.mayoclinic.org/es-es/tests-procedures/x-ray/about/pac-20395303>
- Mayo Clinic. (2020, 06 13). *Mayo Clinic*. Retrieved 11 28, 2020, from <https://www.mayoclinic.org/es-es/diseases-conditions/pneumonia/symptoms-causes/syc-20354204>
- Mayo Clinic. (2020, 6 13). *Mayo Clinic*. Retrieved 10 28, 2020, from <https://www.mayoclinic.org/es-es/diseases-conditions/pneumonia/symptoms-causes/syc-20354204#:~:text=Neumon%C3%ADa%20adquirida%20en%20el%20hospital,la%20contraen%20ya%20est%C3%A1n%20enfermas>
- Militante, S., & Sibbaluca, B. (2020). Pneumonia Detection Using Convolutional Neural Networks. *Revista Internacional de Investigación Científica y Tecnológica*, 9(04), 1332-1337. doi:<https://doi.org/10.46501/ijmtst070117>
- Ministerio de Salud. (2018, Agosto 10). *MINSALUD*. Retrieved octubre 28, 2020, from <https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/VS/ED/GCFI/estimacion-cantidad-especialistas-medicas-septiembre-2018.pdf>
- National Institute of Biomedical Imaging and Bioengineering. (2012, enero 24). *National Institute of Biomedical Imaging and Bioengineering*. Retrieved octubre 28, 2020, from <https://www.nibib.nih.gov/espanol/temas-cientificos/tomograf%C3%ADa-computarizada-tc>
- Neuman, M., Lee, E., Bixby, S., Diperna, S., Hellinger, J., Markowitz, R., . . . Shah, S. (2012). Variability in the interpretation of chest radiographs for the diagnosis of pneumonia in children. *Journal of hospital medicine*, 7(4), 294-298. doi:<https://doi.org/10.1002/jhm.955>
- Oliveira, E. T. (2019). Deep Learning y sus aplicaciones hoy. *Científica Multidisciplinar Núcleo do Conhecimento*, 4(5), 5-26. doi:2448-0959
- Organizacion Mundial de la Salud. (2019, 8 2). *Organizacion Mundial de la Salud*. (Organizacion Mundial de la Salud) Retrieved septiembre 7, 2020, from <https://www.who.int/es/news-room/fact-sheets/detail/pneumonia>

- Organizacion Mundial de Salud. (2019). *Organizacion Mundial de Salud*. Retrieved 05 13, 2021, from <https://www.who.int/es/news-room/fact-sheets/detail/pneumonia>
- Organizacion Panamericana de la Salud. (2014). *OPS*. Retrieved from https://www3.paho.org/hq/index.php?option=com_content&view=article&id=1894:2009-about-pneumococcus-disease&Itemid=1630&lang=es#:~:text=M%C3%A1s%20del%2095%25%20de%20todos,5%20graves%20o%20muy%20graves.
- Palomino, N. L., & Concha, U. R. (2009). Técnicas de Segmentación en Procesamiento Digital de Imágenes. *Revista de ingeniería de sistemas e informática*, 6(4), 9. Retrieved 11 29, 2020, from <https://revistasinvestigacion.unmsm.edu.pe/index.php/sistem/article/view/3299/2749>
- Radiological Society of North America. (2018). *Kaggle*. Retrieved 05 21, 2021, from <https://www.kaggle.com/c/rsna-pneumonia-detection-challenge>
- Rajpurkar, Irvin. (2017, 12 25). CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays. *Computer Vision and Pattern Recognition*, 7. Retrieved 10 25, 2020, from <https://arxiv.org/abs/1711.05225>
- Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., . . . Ng, A. (2017). Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. *ArXiv*, 1711.05225.
- Red Hat. (2021). *Red Hat*. Retrieved 07 13, 2021, from <https://www.redhat.com/es/topics/api/what-are-application-programming-interfaces>
- Rezatofghi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., & Savarese, S. (2019). Generalized Intersection over Union. *The IEEE Conference on Computer Vision and Pattern Recognition*.
- Roldán, M., & Karina, A. (2019). *RED NEURONAL CONVOLUCIONAL EN UN AMBIENTE PSEUDO-DISTRIBUIDO PARA LA CLASIFICACIÓN DE RADIOGRAFÍAS DE PACIENTES CON NEUMONÍA*. UNIVERSIDAD AUTÓNOMA DE CIUDAD JUÁREZ, Ingeniería Eléctrica y Computación. CIUDAD JUÁREZ: Instituto de Ingeniería y Tecnología Departamento de Ingeniería Eléctrica y Computación. Retrieved 10 25, 2020
- Romero Saldaña, M. (2016). Pruebas de bondad de ajuste a una distribución normal. *Revista Enfermería del Trabajo*, 6(3), 105-114. Retrieved 08 31, 2021, from <file:///C:/Users/Asus/Desktop/sustentacion%20practica/Dialnet-PruebasDeBondadDeAjusteAUnaDistribucionNormal-5633043.pdf>
- Roth, H., Lu, L., Seff, A., Cherry, K., Hoffman, J., Wang, S., . . . Summers, R. (2014). Representation for Lymph Node Detection using Random Sets of Deep Convolutional Neural Network Observations. *Medical Image Computing and*

- Computer-Assisted Intervention*, 8673, 520-527. doi:10.1007/978-3-319-10404-1_65
- Sirazitdinov, I., Kholiavchenko, M., Mustafaev, T., Yixuan, Y., Kuleev, R., & Ibragimov, B. (2019). Deep neural network ensemble for pneumonia localization from a large-scale chest x-ray database. *Computers and Electrical Engineering*, 78, 388–399. doi:<https://doi.org/10.1016/j.compeleceng.2019.08.004>
- Sirazitdinov, I., Kholiavchenko, M., Mustafaev, T., Yixuan, Y., Kuleev, R., & Ibragimov, B. (2019). Deep neural network ensemble for pneumonia localization from a large-scale chest x-ray database. *Computers and Electrical Engineering*, 78, 388–399. doi:<https://doi.org/10.1016/j.compeleceng.2019.08.004>
- Taylor, L., & Nitschke, G. (2018). Improving Deep Learning using Generic Data Augmentation. *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, (pp. 1542-1547). Retrieved 05 21, 2021, from <https://arxiv.org/abs/1708.06020>
- Treuille, A. (2019, 10 1). *Towards data science*. Retrieved from <https://towardsdatascience.com/coding-ml-tools-like-you-code-ml-models-dba3357eace>
- Universitat Oberta de Catalunya. (2016, 05 26). *El blog de los Estudios de Informática, Multimedia y Telecomunicación de la UOC*. Retrieved 07 12, 2021, from <https://informatica.blogs.uoc.edu/la-segmentacion-semantica-y-sus-benchmarks/>
- World Life Expectancy. (2018). *World Life Expectancy*. (OMS) Recuperado el 2020 de 09 de 7, de <https://www.worldlifeexpectancy.com/cause-of-death/influenza-pneumonia/by-country/>
- Xu, Y., Jia, Z., Ai, Y., Zhang, F., Lai, M., & Chang, E.-C. (2015, 8 6). Deep convolutional activation features for large scale Brain Tumor histopathology image classification and segmentation,. *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. doi:10.1109 / ICASSP.2015.7178109
- Yao, L., Poblenz, E., Dagunts, D., Covington, B., Bernard, D., & Lyman, K. (2017). Learning to diagnose from scratch by exploiting dependencies among labels. *ArXiv*, 1710.10501. Retrieved 05 20, 2021

