

UNIVERSIDAD RICARDO PALMA
ESCUELA DE POSGRADO

MAESTRÍA EN INGENIERÍA INFORMÁTICA
CON MENCIÓN EN INGENIERÍA DE SOFTWARE



Tesis para optar el Grado Académico de Maestro en Ingeniería
Informática con Mención en Ingeniería de Software

Las Redes Neuronales Convolucionales y la mejora en el diagnóstico de
Neumonía – área de Radiología

Autor: Bach. Cabrejos Yalán, Víctor Manuel

Asesor: Mg. Saito Silva, Carlos Agustín

LIMA – PERÚ
2022

Miembros del Jurado Examinador para la evaluación de la sustentación de la tesis,
integrado por:

1. Presidente : Dra. Lenis Rossi Wong Portillo
2. Miembro : Mg. Carlos Alberto Janjachi Toribio
3. Miembro : Mg. Octavio V. Jara Villavicencio
4. Asesor : Mg. Carlos Agustín Saito Silva
5. Representante de la EPG : Mg. Olga Suarez Avelino

Dedicatoria

Este proyecto de investigación está dedicado a mis padres por el esfuerzo que realizaron en mi formación personal y profesional, así también es dedicado a Dios que ha sido la fuerza motivadora para afrontar todos los retos que se me ha presentado.

Por: Víctor Manuel Cabrejos Yalán

Agradecimientos

Quiero expresar mi agradecimiento al Mg. Carlos Agustín Saito Silva, quien ha sido mi asesor y con su apoyo he podido culminar el trabajo de investigación.

Además, agradecer a los miembros del jurado, que han sido excelentes profesores y mentores que me han ayudado a mejorar mucho durante mis estudios de maestría y siempre llevaré conmigo todas sus enseñanzas.

Por: Víctor Manuel Cabrejos Yalán

ÍNDICE

RESUMEN.....	xi
ABSTRACT	xii
INTRODUCCIÓN	1
CAPÍTULO I: PLANEAMIENTO DEL ESTUDIO	3
1.1. Descripción del problema.....	3
1.2. Formulación del problema	9
1.2.1 Problema General.....	9
1.2.2 Problemas Específicos	9
1.3. Importancia y Justificación del estudio.....	10
1.4. Delimitación del estudio.....	16
1.5. Objetivos generales y específicos	17
1.5.1 Objetivo general.....	17
1.5.2 Objetivos específicos	17
CAPÍTULO II: MARCO TEÓRICO	18
2.1. Marco histórico	18
2.2. Investigaciones relacionadas con el tema.....	25
2.3. Estructura teórica y científica que sustenta el estudio.....	32
2.4. Definición de términos básicos	60
2.5. Fundamentos teóricos que sustentan las hipótesis	61
2.6. Hipótesis.....	62
2.6.1 Hipótesis general.....	62
2.6.1 Hipótesis específicas	62
2.7. Variables.....	63
CAPÍTULO III: MARCO METODOLÓGICO	64
3.1. Tipo, método y diseño de la investigación	64
3.2. Población y muestra	69
3.3. Técnicas e instrumentos de recolección de datos.....	72
3.4. Descripción de procedimientos de análisis	75
Capítulo IV: RESULTADOS Y ANÁLISIS DE RESULTADOS	76
4.1. Resultados	76
4.2. Análisis de resultados.....	122
CONCLUSIONES Y RECOMENDACIONES	152
REFERENCIA	157
ANEXOS.....	160
Anexo 1: Declaración de Autenticidad.....	160
Anexo 2: Autorización de consentimiento para realizar la investigación	161
Anexo 3: Matriz de consistencia.....	162
Anexo 4: Matriz de Operacionalización	163

ÍNDICE DE TABLAS

Tabla 01: <i>Costos totales del paciente con diagnóstico de neumonía (EsSalud)</i>	4
Tabla 02: <i>Número de diagnósicos incorrectos hechos por doctores</i>	5
Tabla 03: <i>Incidencia y letalidad por infecciones nosocomiales</i>	6
Tabla 04: <i>Variabilidad entre Primera y Segunda interpretación de Radiografías torácicas</i>	7
Tabla 05: <i>Arquitectura de la Red Neuronal Convolutiva AlexNet</i>	44
Tabla 06: <i>Arquitectura de la Red Neuronal Convolutiva VGG</i>	47
Tabla 07: <i>Diseño de Investigación Cuasi-Experimental</i>	67
Tabla 08: <i>Diseño Cuasi-Experimental</i>	67
Tabla 09: <i>Población y Muestra PRE y POST por cada una de las variables</i>	71
Tabla 10: <i>Técnicas e instrumentos</i>	74
Tabla 11: <i>Matriz de Análisis de datos</i>	75
Tabla 12: <i>Muestra Pre Test Porcentaje de Diagnósicos Incorrectos</i>	79
Tabla 13: <i>Muestra Post Test Porcentaje de Diagnósicos Incorrectos (usando ResNet-50)</i>	89
Tabla 14: <i>Muestra Pre Test Porcentaje de Diagnósicos Incorrectos</i>	91
Tabla 15: <i>Muestra Post Test Porcentaje de Diagnósicos Incorrectos (usando Alexnet)</i> . 101	
Tabla 16: <i>Muestra Pre Test Porcentaje de Diagnósicos Incorrectos</i>	104
Tabla 17: <i>Muestra Post Test Porcentaje de Diagnósicos Incorrectos (usando VGG)</i>	112
Tabla 18: <i>Muestra Pre Test Horas de Tiempo de Espera</i>	114
Tabla 19: <i>Muestra Post Test Horas de Tiempo de Espera</i>	121
Tabla 20: <i>Muestra Pre Test y Post Test de porcentaje de diagnósicos incorrectos</i>	125
Tabla 21: <i>Resumen de procesamiento de datos –porcentaje de diagnósicos incorrectos muestras Pre Test y Post Test</i>	126
Tabla 22: <i>Estadísticas de grupo – Muestras pre y post test</i>	126
Tabla 23: <i>Prueba de Normalidad para porcentaje de diagnósicos incorrectos de las muestras Pre Test y Post Test</i>	127
Tabla 24: <i>Estadísticas de muestras emparejadas para porcentaje de diagnósicos incorrectos en promedio</i>	129

Tabla 25: <i>Prueba de hipótesis de T de Student de muestras emparejadas para porcentaje de diagnósis incorrectas en promedio</i>	130
Tabla 26: <i>Muestra Pre Test y Post Test de porcentaje de diagnósis incorrectas</i>	131
Tabla 27: <i>Resumen de procesamiento de datos –porcentaje de diagnósis incorrectas muestras Pre Test y Post Test</i>	132
Tabla 28: <i>Estadísticas de grupo – Muestras pre y post test</i>	132
Tabla 29: <i>Prueba de Normalidad para porcentaje de diagnósis incorrectas de las muestras Pre Test y Post Test</i>	134
Tabla 30: <i>Estadísticas de muestras emparejadas para porcentaje de diagnósis incorrectas en promedio</i>	135
Tabla 31: <i>Prueba de hipótesis de T de Student de muestras emparejadas para porcentaje de diagnósis incorrectas en promedio</i>	136
Tabla 32: <i>Muestra Pre Test y Post Test de porcentaje de diagnósis incorrectas</i>	137
Tabla 33: <i>Resumen de procesamiento de datos –porcentaje de diagnósis incorrectas muestras Pre Test y Post Test</i>	138
Tabla 34: <i>Estadísticas de grupo – Muestras pre y post test</i>	138
Tabla 35: <i>Prueba de Normalidad para porcentaje de diagnósis incorrectas de las muestras Pre Test y Post Test</i>	140
Tabla 36: <i>Estadísticas de muestras emparejadas para porcentaje de diagnósis incorrectas en promedio</i>	141
Tabla 37: <i>Prueba de hipótesis de T de Student de muestras emparejadas para porcentaje de diagnósis incorrectas en promedio</i>	141
Tabla 38: <i>Muestra Pre Test y Post Test de horas de tiempo de espera</i>	143
Tabla 39: <i>Resumen de procesamiento de datos – horas de tiempo de espera muestras Pre Test y Post Test</i>	144
Tabla 40: <i>Estadísticas de grupo – Muestras pre y post test</i>	144
Tabla 41: <i>Prueba de Normalidad para horas de tiempo de espera de las muestras Pre Test y Post Test</i>	146
Tabla 42: <i>Estadísticas de muestras emparejadas para horas de tiempo de espera en promedio</i>	147
Tabla 43: <i>Prueba de hipótesis de T de Student de muestras emparejadas para horas de tiempo de espera en promedio</i>	148
Tabla 44: <i>Resumen de resultados</i>	149
Tabla 45: <i>Tabla del valor t de cada algoritmo</i>	150

Tabla 46: <i>Matriz de Consistencia</i>	162
Tabla 47: <i>Matriz de Operacionalización</i>	163

ÍNDICE DE FIGURAS

<i>Figura 01:</i> Número de muertes de niños menores de 5 años. (UNICEF, 2019)	4
<i>Figura 02:</i> Porcentaje de Errores en diagnosis. Asociación de Aseguradores de Médicos de América Fuente: (Graber, 2013)	6
<i>Figura 03:</i> Comparación de una neurona biológica y una red neuronal artificial. (Vázquez Enríquez, 2019, p. 11)	36
<i>Figura 04:</i> Estructura de una red neuronal simple con sus diferentes capas. (Vázquez Enríquez, 2019, p. 12).....	37
<i>Figura 05:</i> Estructura de una red neuronal simple con sus diferentes capas. (Sewak, Karim, & Pujari, 2018, p. 101)	40
<i>Figura 06:</i> La entera arquitectura de la red neuronal convolucional AlexNet. (Sewak, Karim, & Pujari, 2018, p. 103)	42
<i>Figura 07:</i> Arquitectura completa del modelo VGG16. (Sewak, Karim, & Pujari, 2018, p. 104)	45
<i>Figura 08:</i> La entera arquitectura de la red neuronal convolucional VGG-19. Fuente: (Venkatesan & Li, 2018)	47
<i>Figura 09:</i> Bloque Residual de una Red Residual. (Sewak, Karim, & Pujari, 2018, p. 108)	50
<i>Figura 10:</i> Matriz de confusión extendida para incluir el valor predictivo positivo (PPV). (Ranschaert, Morozov, & Algra, 2019, p. 94)	58
<i>Figura 11:</i> Mapa conceptual que sustentan las hipótesis. Elaboración: propia.....	61
<i>Figura 12:</i> Diseño Cuasi experimental. Elaboración Propia.....	68
<i>Figura 13:</i> Magic Quadrant for Operational Database Management Services. Deloitte Cyber Strategy (2020).....	73
<i>Figura 14:</i> Diagrama de Ishikawa para el problema general. Elaboración propia.	77
<i>Figura 15:</i> Porcentaje de diagnosis incorrectas en diagnóstico de neumonía automatizado. Elaboración propia.	79
<i>Figura 16:</i> Bloque residual. Conexión de acceso directo. (Shanmugamani, 2018)	81
<i>Figura 17:</i> Transformaciones en PyTorch requeridas en preprocesamiento de imágenes. Elaboración propia.	82

<i>Figura 18:</i> Muestra de imágenes redimensionadas en preprocesamiento. Elaboración propia.	82
<i>Figura 19:</i> Red neuronal convolucional con Transfer Learning de ResNet-50. Elaboración propia.	83
<i>Figura 20:</i> Diagrama del sistema usando ResNet-50. Elaboración propia.....	84
<i>Figura 21:</i> Ejemplo de patrones encontrados en una imagen usando ResNet-50 durante el entrenamiento de una detección de Neumonía. Elaboración propia.	85
<i>Figura 22:</i> Lista de métricas y Curva de Aprendizaje del modelo basado en ResNet-50. Elaboración propia.	86
<i>Figura 23:</i> Matriz de confusión del sistema basado en ResNet-50. Elaboración propia.	87
<i>Figura 24:</i> Reducción de Porcentaje de diagnosis incorrectas después de implementar el sistema basado en ResNet-50. Elaboración propia.	88
<i>Figura 25:</i> Porcentaje de diagnosis incorrectas en diagnóstico de neumonía automatizado. Elaboración propia.	91
<i>Figura 26:</i> AlexNet contiene ocho capas: cinco convolucionales y tres completamente conectadas (Elgandy, 2020).	92
<i>Figura 27:</i> Red neuronal convolucional con Transfer Learning de Alexnet. Elaboración propia.	95
<i>Figura 28:</i> Código de Entrenamiento del modelo Alexnet. Preparación propia.	96
<i>Figura 29:</i> Diagrama del sistema usando Alexnet. Elaboración propia.	97
<i>Figura 30:</i> Lista de métricas y Curva de Aprendizaje del modelo basado en Alexnet. Elaboración propia.	98
<i>Figura 31:</i> Matriz de confusión del sistema basado en Alexnet. Elaboración propia. .	100
<i>Figura 32:</i> Reducción de Porcentaje de diagnosis incorrectas después de implementar el sistema basado en Alexnet. Elaboración propia.	101
<i>Figura 33:</i> Porcentaje de diagnosis incorrectas en diagnóstico de neumonía automatizado. Elaboración propia.	103
<i>Figura 34:</i> VGG-16 contiene 16 con una arquitectura uniforme y fácil de entender (Elgandy, 2020).	105
<i>Figura 35:</i> Red neuronal convolucional con Transfer Learning de Alexnet. Elaboración propia.	107
<i>Figura 36:</i> Código de Entrenamiento del modelo VGG. Preparación propia.	107

<i>Figura 37:</i> Código de FastAI creando un modelo para el despliegue usando VGG-16. Elaboración propia.	108
<i>Figura 38:</i> Diagrama del sistema usando VGG-16. Elaboración propia.....	109
<i>Figura 39:</i> Lista de métricas y Curva de Aprendizaje del modelo basado en VGG. Elaboración propia.	109
<i>Figura 40:</i> Matriz de confusión del sistema basado en VGG. Elaboración propia.	110
<i>Figura 41:</i> Reducción de Porcentaje de diagnosis incorrectas después de implementar el sistema basado en VGG. Elaboración propia.	111
<i>Figura 42:</i> Promedio mensual de número de horas en tiempo más alto de espera de pacientes con neumonía. Elaboración propia.	114
<i>Figura 43:</i> Código de modificación de red neuronal para poder después integrar Softmax. Elaboración propia.	117
<i>Figura 44:</i> Código de creación de la función de activación softmax usando torch.exp. Elaboración propia.	118
<i>Figura 45:</i> Diagrama del sistema con integración de activación softmax. Elaboración propia	118
<i>Figura 46:</i> Predicciones del sistema con visibilidad de probabilidades. Elaboración propia.	119
<i>Figura 47:</i> Reducción de Horas de Espera en pacientes después de implementar el sistema con la función de activación SoftMax. Elaboración propia.	121

RESUMEN

La neumonía es la infección que produce un mayor número de ingresos hospitalarios, siendo responsable del 15% de todas las defunciones de menores de 5 años a nivel mundial (UNICEF, 2019). Es una enfermedad compleja de diagnosticar debido a que puede ser causado por diferentes agentes infecciosos. De contraer neumonía, el tipo de tratamiento seleccionado dependerá de la diagnosis, y por lo tanto la precisión del diagnóstico es de alta criticidad. Además, actualmente la muerte en niños con neumonía puede ocurrir en tan solo 2 días después del inicio de la enfermedad, eso hace que la velocidad de la detección sea también altamente importante para poder salvar vidas.

Un diagnóstico de neumonía típicamente requiere de la revisión de placas radiográficas torácicas. El radiólogo o médico, basado en su experiencia, visualmente trata de localizar anomalías en las imágenes. Algunos centros de salud han implementado un sistema que automatiza este proceso, pero obteniendo variables resultados de precisión.

En esta investigación después de hacer un análisis exhaustivo se ha automatizado el proceso de diagnóstico mediante la implementación de un sistema basado en Inteligencia Artificial. Este sistema tiene tres motores cognitivos basados tres algoritmos del estado del arte de redes neuronales convolucionales para obtener una mayor precisión y a su vez reducir el tiempo de diagnóstico de neumonía.

Este sistema además fue desplegado a través de una aplicación web que permite visualizar (en el caso de múltiples pacientes) los porcentajes de probabilidad de padecer neumonía por cada imagen, esta información adicional permite al médico en tomar la más apropiada decisión en el tratamiento de los pacientes. Como resultado se logró disminuir el número de diagnosis incorrectas en un 80% utilizando el mejor algoritmo basado en la arquitectura de redes neuronales Alexnet, y se logró reducir los tiempos de espera de los pacientes en un 32%.

Palabras clave: Redes Neuronales Convolucionales, Deep Learning, Inteligencia Artificial, Diagnóstico de Neumonía, Radiología automática.

ABSTRACT

Pneumonia is the infection that produces the highest number of hospital admissions, being responsible for 15% of all deaths of children under 5 years of age worldwide (UNICEF, 2019). It is a complex disease to diagnose because it can be caused by different infectious agents. If you contract pneumonia, the type of treatment selected will depend on the diagnosis, and therefore the precision of the diagnosis is highly critical. In addition, currently death in children with pneumonia can occur in as little as 2 days after the onset of the disease, which makes the speed of detection also highly important in order to save lives.

A diagnosis of pneumonia typically requires a review of chest x-ray films. The radiologist or physician, based on her experience, visually tries to locate abnormalities in the images. Some health centers have implemented a system that automates this process, but obtaining variable precision results.

In this research, after doing an exhaustive analysis, the diagnostic process has been automated through the implementation of a system based on Artificial Intelligence. This system has three cognitive engines based on three algorithms of the state of the art of convolutional neural networks to obtain greater precision and in turn reduce the time of diagnosis of pneumonia.

This system was also deployed through a web application that allows viewing (in the case of multiple patients) the probability percentages of suffering from pneumonia for each image, this additional information allows the doctor to make the most appropriate decision in the treatment of the patients. As a result, it was possible to reduce the number of incorrect diagnoses by 80% using the best algorithm based on the Alexnet neural network architecture, and it was possible to reduce patient waiting times by 32%.

Keywords: Convolutional Neural Networks, Deep Learning, Artificial Intelligence, Pneumonia Diagnosis, Automatic Radiology.

INTRODUCCIÓN

La presente investigación ha sido motivada por el deseo de buscar mejorar el proceso de diagnóstico de neumonía con la finalidad de ayudar a salvar vidas implementando un sistema basado en inteligencia artificial.

Para la implementación de esta solución ha sido necesario analizar cuáles son las causas raíz de los problemas en el proceso de diagnóstico en un área de radiología, identificándolas mediante el diagrama de Ishikawa. Actuales sistemas de automatización basados en algoritmos tradicionales toman demasiado tiempo, no son adaptables a nuevas imágenes y la precisión es demasiado baja.

El desarrollo de la tesis se inicia con el primer capítulo, en este punto se describe la situación problemática, así como sus evidencias en forma de estadísticas actuales, se detallan las variables dependientes que son afectadas y las variables independientes, así como las propuestas para mejorar la situación actual en el área de salud. Además, en este capítulo también se plantean las hipótesis de la investigación.

Luego en el segundo capítulo, se realiza la investigación sobre el aspecto teórico de las variables. Todas las variables son parte del campo de Deep Learning y pertenecen al estado del arte en arquitecturas de redes neuronales convolucionales. En este capítulo se detalla cada arquitectura, su historia y evolución en el campo de lo que es Visión Artificial. Se detallan también cada uno de sus componentes, explicando como estas redes aprenden y se adaptan a cambios en los datos con énfasis en la aplicabilidad de estos modelos en el sector salud, específicamente en clasificación de imágenes médicas. Este capítulo también incluye las investigaciones relacionadas con el tema de investigación, las cuales incluyen los resultados de tesis internacionales validando los modelos de inteligencia artificial en el diagnóstico de neumonía en diversos países.

En el tercer capítulo, se describe la metodología de la investigación, donde se plantea en detalle el diseño de la investigación, la determinación de la población y muestra, la técnica e instrumento de recolección de datos del sistema del área de radiología. Finalmente se muestra la descripción de procedimientos de análisis de cada variable.

En el cuarto capítulo es donde se presenta la implementación del sistema de inteligencia artificial mostrando en detalle como cada arquitectura de redes neuronales ha sido codificada, que frameworks de Deep Learning han sido utilizados incluyendo las técnicas usadas para transferir el aprendizaje de las redes pre-entrenadas para luego ser entrenadas con las imágenes radiográficas. También en este capítulo se muestran los diagramas del sistema para el despliegue de la solución propuesta. Se comparan los datos originales de la situación pre con los resultados obtenidos después de la implementación del sistema de inteligencia artificial, mostrando el aporte de la solución implementada.

También en el cuarto capítulo se muestra el análisis estadístico de los resultados por cada objetivo planteado en la línea de investigación. Esto incluye las pruebas de normalidades, reglas de decisiones y pruebas de hipótesis. En las pruebas de hipótesis se demuestra en detalle como cada variable tuvo un efecto significativo positivo en lograr alcanzar cada uno de los objetivos trazados, y consecuentemente, ayudar a mejorar el proceso de diagnóstico de neumonía.

Finalmente, en las conclusiones se muestran cada uno de los aportes logrados por el sistema implementado. Debido a que la presente investigación compara diversos esquemas de redes neuronales convolucionales se han resaltado las ventajas de cada una de ellas. Este estudio también sirve para crear futuros sistemas de clasificación de alta precisión, eso significa que este sistema podría aprender de otros tipos de imágenes médicas, y poder diagnosticar otras enfermedades.

Se han probado las hipótesis del estudio corroborando que el uso de un sistema basado en redes neuronales convolucionales mejora el proceso de diagnóstico de neumonía reduciendo el número de diagnósticos incorrectos, y reduciendo el tiempo de espera de pacientes que padecen de neumonía significativamente.

CAPÍTULO I: PLANEAMIENTO DEL ESTUDIO

1.1. Descripción del problema

Hoy en día el sector salud enfrenta un reto grande debido a que la neumonía es la principal causa de mortalidad infantil en todo el mundo. La neumonía no solo afecta a niños sino también a sus familias. La UNICEF ha declarado que todos los días aproximadamente 2 200 niños mueren de neumonía (UNICEF, 2019).

En la Figura 01, de acuerdo a (UNICEF, 2019), se muestra que, en niños menores de 5 años, el número de muertes anuales por causa de la neumonía ha aumentado de 802 000 a 1 755 000 entre los años 2000 y 2018, convirtiéndose en la enfermedad infecciosa con mayor mortalidad infantil a nivel mundial.

En términos de costos económicos, se calcula que el costo en antibióticos necesarios para tratar la neumonía es de aproximadamente \$109 millones de dólares al año (World Health Organization, 2021).

En la Tabla 01 se muestra el desglose de los estimados costos económicos por paciente de neumonía en el Perú, con un aproximado total de 11 470,00 nuevos soles. Estos datos fueron obtenidos de un estudio hecho en el hospital II en Huánuco

EsSalud (Alcalde Rabanal, Lazo Gonzales, Macias, Contreras Ochoa, & Espinosa Henao, 2019).

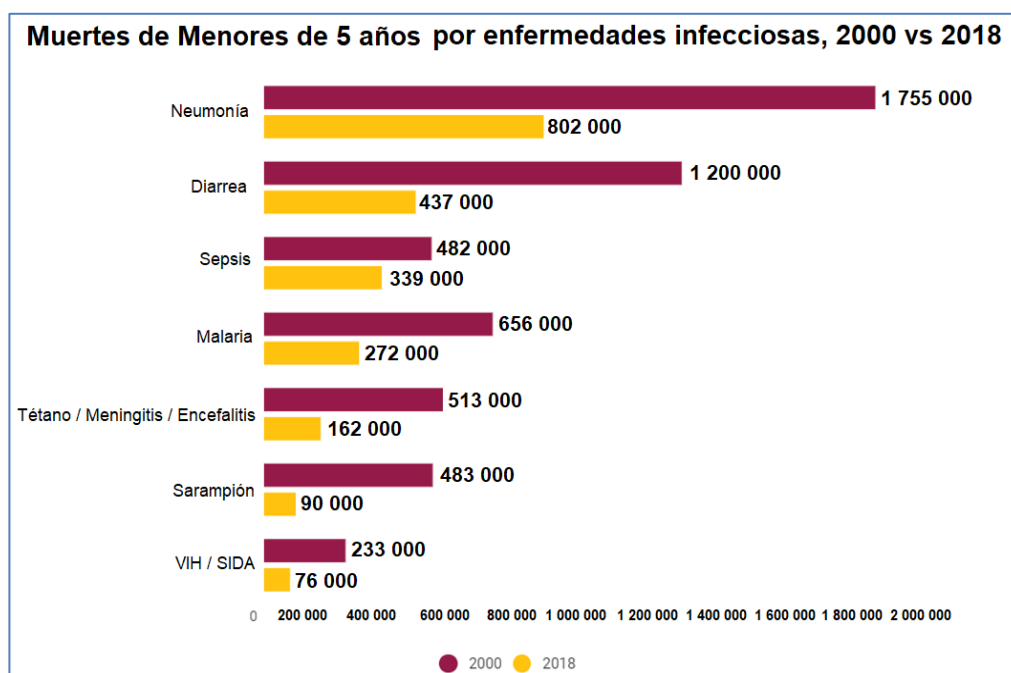


Figura 01: Número de muertes de niños menores de 5 años. (UNICEF, 2019)

Tabla 01:

Costos totales del paciente con diagnóstico de neumonía (EsSalud)

Costo por Actividad (nuevos soles)	Neumonía
Costos directos sanitarios	
▪ Hospitalización	1050,00
▪ Antibióticos	2730,00
▪ Exámenes Auxiliares	900,00
▪ Evaluaciones especializadas y otras medicaciones	2647,00
▪ Atenciones de hemodiálisis	1760,00
Costos directos no sanitarios	
▪ Gastos extra institucionales	414,00
▪ Contratación de rehabilitador	327,00
▪ Atención domiciliaria por personal de salud	700,00
▪ Cuidadores durante hospitalización	942,00
	11470,00

Fuente: (Alcalde Rabanal, Lazo Gonzales, Macias, Contreras Ochoa, & Espinosa Henao, 2019)

Además, el reto de no poder lograr una diagnosis rápida y precisa colabora a mantener esta problemática, esto es debido también a que la neumonía es una condición muy difícil de detectar y fácil de confundir con los síntomas de otras enfermedades respiratorias.

Esta falta de precisión en lograr llegar a la diagnosis correcta de neumonía afecta especialmente en los lugares donde los casos de neumonía son muy comunes, y en las regiones donde no siempre se tienen los recursos médicos necesarios para una segunda opinión (Álvarez Gutiérrez, Díaz Baquero, Medina Gallardo, & Romero Falcón, 2010).

En la Tabla 02 se muestran los resultados de un estudio científico (Ranschaert, Morozov, & Algra, 2019) donde se tomaron 1000 diagnosis de neumonía evidenciadas en las radiografías de los pacientes, y luego preguntando a doctores que diagnostiquen usando su juicio clínico donde se encontró que 71% de las diagnosis de neumonía diagnosticadas usando las radiografías no pudieron ser diagnosticadas por los doctores usando su juicio clínico.

Tabla 02:
Número de diagnosis incorrectas hechas por doctores

Juicio clínico	Neumonía
▪ Valor Predictivo Negativo	96%
▪ Sensibilidad	29%
▪ Número de Diagnosis de Neumonía diferentes a las detectadas en radiografías	71%

Fuente: (Ranschaert, Morozov, & Algra, 2019)

La conclusión de aquel estudio evidenció la importancia de usar radiografías y la imprecisión en diagnosis que se encuentra en el sector salud (Ranschaert, Morozov, & Algra, 2019).

La Figura 02 muestra los resultados obtenidos por la Asociación de Aseguradores de Médicos de América donde se revisó sistemáticamente a más de 8000 pacientes y la cantidad de errores en las diagnósis fue más del 25% (Graber, 2013).

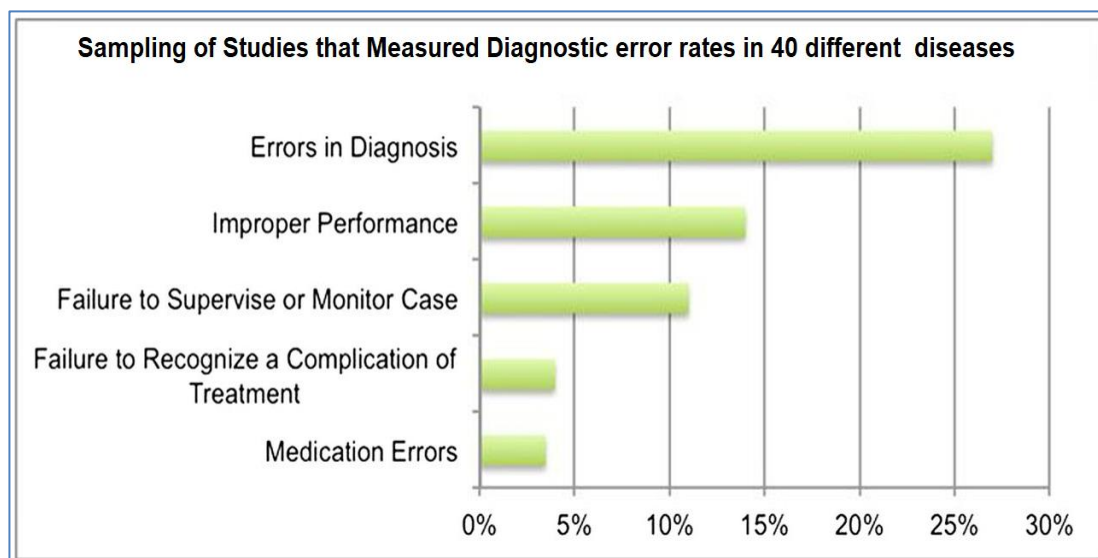


Figura 02: Porcentaje de Errores en diagnóstico. Asociación de Aseguradores de Médicos de América Fuente: (Graber, 2013)

De acuerdo a un estudio de perfil epidemiológico de infecciones nosocomiales se encontró que la neumonía ocupa el segundo lugar en orden de frecuencia dentro de las infecciones nosocomiales, y además constituye la principal causa de muerte por infecciones adquiridas en el hospital, con variaciones institucionales. Neumonía nosocomial es aquella que se presenta en las 48 a 72 horas tras el ingreso hospitalario (Elemraid, et al., 2014).

En la Tabla 03 se muestra que en el periodo estudiado la neumonía tuvo mayor letalidad que la flebitis y la infección de herida quirúrgica profunda (Ponce de León, Rangel-Frausto, Elías-López, & Romero Oliveros, 1999). Las infecciones más frecuentes fueron flebitis, bacteriemia relacionada a tratamiento intravenoso y neumonía (0.08 por 100 egresos y la letalidad de 20.00%).

Tabla 03:
Incidencia y letalidad por infecciones nosocomiales

Sitio de Infección	Incidencia	Letalidad
▪ Flebitis	4.09	1.87

Sitio de Infección	Incidencia	Letalidad
▪ Peritonitis	0.10	0.00
▪ Neumonía	0.08	20.00
▪ Infección de herida quirúrgica profunda	0.08	20.00
▪ Bacteriemia secundaria	0.05	0.00
▪ Tejidos blandos y piel	0.05	0.00
▪ Infección de vías urinarias	0.03	0.00

Fuente: (Ponce de León, Rangel-Frausto, Elías-López, & Romero Oliveros, 1999)

Globalmente es la primera causa de infección en las Unidades de Cuidados Intensivos (UCI). La incidencia es variable dependiendo del grupo de edad, siendo 5 casos / 1000 hospitalizados con edad inferior a 35 años y elevándose a más de 15 casos / 1000 en mayores de 65 años (Le, Wang, Carneiro, & Yang, 2019).

Esta incidencia se multiplica por 20 en pacientes conectados a Ventilación Mecánica Invasiva (VMI), en los que la mortalidad puede llegar hasta el 50%. El retraso en la instauración de un tratamiento antibiótico adecuado para la Neumonía Nosocomial Grave se acompaña de un peor pronóstico (Álvarez Gutiérrez, Díaz Baquero, Medina Gallardo, & Romero Falcón, 2010).

De acuerdo a otro estudio acerca de la exactitud en la interpretación de radiografías para la diagnosis de neumonía pediátrica se encontró una alta variabilidad en la interpretación de radiólogos especialmente en la interpretación de los cambios irregulares y perihilares en las radiografías (Elemraid, et al., 2014).

Tabla 04:

Variabilidad entre Primera y Segunda interpretación de Radiografías torácicas

First reading		Second reading (gold standard)				Disagreement ¹
Radiographic changes	<i>n</i> (%)	Lobar	Patchy	Perihilar	Normal	<i>n</i> (%)
Lobar	48 (28.4)	47	1	0	0	1 (2.1)
Patchy	43 (25.4)	7	22	5	9	21 (48.8)
Perihilar	32 (19.0)	4	0	23	5	9 (28.1)
Normal	46 (27.2)	4	2	0	40	6 (13.0)
<i>Total</i>	169	62	25	28	54	37 (22.0)

Fuente: (Elemraid, et al., 2014)

En la Tabla 04 se puede ver que los mayores cambios en la interpretación es en las características perihilares de 32 a 23 (28.1% de diferencia) y cavidades irregulares (Patchy) de 43 a 22 (48.8% de diferencia) debido a error humano y/o entrenamiento inadecuado en radiología (Elemraid, et al., 2014).

1.2. Formulación del problema

1.2.1 Problema General

¿De qué manera las Redes Neuronales Convolucionales podría mejorar la rapidez en el diagnóstico de Neumonía en un área de Radiología?

1.2.2 Problemas Específicos

- a) ¿De qué manera el Algoritmo ResNet-50 podría reducir el número de diagnósicos incorrectas?
- b) ¿De qué manera el Algoritmo Alexnet podría reducir el número de diagnósicos incorrectas?
- c) ¿De qué manera el Algoritmo VGG podría reducir el número de diagnósicos incorrectas?
- d) ¿De qué manera la Función de Activación SoftMax podría reducir el tiempo de espera de pacientes con neumonía?

1.3. Importancia y Justificación del estudio

✓ Importancia del estudio

Conociendo el problema por el que está pasando el sector salud en el mundo es que resulta de vital importancia el presente trabajo para lograr la implementación de un software de redes neuronales que permita mejorar el proceso de diagnóstico de la neumonía y poder reducir la tasa de mortalidad a nivel mundial. Este software permite detectar rápidamente los patrones en las placas radiográficas que indiquen si el paciente tiene neumonía o no con alto nivel de precisión.

El simple hecho de potencialmente automatizar el trabajo hecho por un especialista en radiología ya está acelerando el proceso de detección y por lo tanto entre más rápido sea detecte, más rápido el paciente puede ser tratado y aumentar las posibilidades de sobrevivir.

Además, no solo se reduce la mortalidad causada por la neumonía debido una detección rápida, sino también un software de redes neuronales descarta potenciales pacientes que adquieren la neumonía mientras esperan atención médica en el centro de salud.

Debido a que parte del aprendizaje de las redes neuronales incluye también aprender de miles de radiografías de pacientes con diagnóstico normal, este software puede detectar en la placa radiográfica si el paciente no ha contraído neumonía.

Por consiguiente, los doctores o incluso enfermeras tendrían acceso a esta información y de esa manera priorizar con facilidad a que paciente tratar y a cuál dar de alta y prevenir una neumonía intrahospitalaria.

Otro factor que hace que este estudio sea no solo importante sino también necesario es acerca de corregir la problemática de los diagnósticos de neumonía equivocados.

Esta es una problemática basada en el error humano y además porque la neumonía es frecuentemente confundida con otras enfermedades como Influenza, Bronquitis aguda, infección respiratoria alta, y Enfermedad pulmonar obstructiva crónica (EPOC), lo cual conlleva a que una radiografía sea mal interpretada terminando en una diagnosis incorrecta.

Con el uso de un software de redes neuronales, basado en múltiples algoritmos del estado del arte, se podrá tener acceso a la mayor precisión posible. Esta investigación no solo tendrá como resultado el acceso automatizado al diagnóstico en sí, sino también al porcentaje de nivel precisión del resultado gracias a que los diversos algoritmos son codificados para usar funciones que muestren también las probabilidades del diagnóstico mostrado.

Con los porcentajes disponibles el doctor o especialista médico ahora no solo tendrá una segunda opinión inmediata, sino que además será una segunda opinión no realizada por una interpretación visual humana.

Otro factor específico en el diagnóstico de la neumonía es la diferenciación entre neumonía bacteriana y neumonía viral. Es importante no solo diagnosticar si el paciente padece de neumonía, sino que además se necesita identificar si es una infección causada por la bacteria *Streptococcus Pneumoniae* o si es una neumonía causada por un virus.

Debido a que el tratamiento es completamente diferente para cada tipo de neumonía, la rápida identificación del tipo de neumonía es de vital importancia para poder escoger el medicamento correcto para el tratamiento.

La neumonía bacteriana es tratada con antibióticos, este tratamiento no ayudaría a una neumonía viral debido a que los virus no son receptivos a los antibióticos. En los casos de neumonía viral es recetado antivirales que ayudan a que el virus

no siga expandiéndose. El presente estudio consiste en el uso de un software de redes neuronales convolucionales que, al ser entrenadas con placas radiográficas de pacientes con ambos tipos de neumonía, adquiere también la capacidad de clasificar si el diagnóstico de neumonía es viral o bacteriana.

Eso lo convierte en una herramienta que acelera no solo el tiempo del diagnóstico, sino que además también reduce la complejidad y costos necesarios para identificar la etiología de la enfermedad.

Actualmente la etiología de la enfermedad se establece con cultivo bacteriológico del esputo y de la sangre (hemocultivo). La prueba de cultivo de esputo que se hace tarda al menos 48 horas en incubar e identificar esas bacterias del esputo, mientras que estos pacientes pueden deteriorarse en unas horas.

Este estudio es importante porque al acelerar el proceso de detección del tipo de neumonía ayuda a la atención más rápida y más apropiada para cada paciente. En una enfermedad de alta criticidad como la neumonía, este podría ayudar salvar la vida del paciente.

Este estudio ahora toma una mayor importancia y necesidad con la actual catástrofe que está viviendo el mundo entero donde la falta de recursos humanos y médicos que puedan rápidamente tratar pacientes contagiados por el virus COVID-19.

Y especialmente en países de bajos recursos como el Perú, este software apoya a reducir la carga en centros hospitalarios porque facilita el proceso de diagnóstico en pacientes con neumonía.

Otro factor que hace que este estudio sea importante es el objetivo de mayor eficacia y eficiencia en el flujo de trabajo para reducir la carga de trabajo de los radiólogos. En la actualidad la tendencia conocida como Big Data en el sector salud muestra que los datos de imágenes radiológicas continúan creciendo a un ritmo desproporcionado en comparación con el número de radiólogos capacitados disponibles, de acuerdo a estudios publicados por la Organización

Mundial de la Salud (OMS) en la actualidad en algunos casos, un radiólogo promedio debe interpretar una imagen cada 3 a 4 segundos en un día laboral de 8 horas para satisfacer las demandas de carga de trabajo.

El presente estudio crea un sistema de apoyo a la demanda que va en crecimiento al ser un software que reduce la complejidad de definir el límite entre una imagen normal y anormal.

Al usar redes neuronales que aprenden de datos específicos, estas redes pueden aprender una representación jerárquica normal de un tipo específico de imagen a partir de una gran cantidad de exámenes normales. Usando detección automática, un software de redes neuronales puede rápidamente separar lo que es considerado imágenes de diagnósticos normales (sin neumonía).

Y de la misma manera, la red neuronal puede extraer imágenes similares de una base de datos para su revisión cuando se encuentra con casos inusuales. Con esta automatización los radiólogos pueden escoger la prioridad de lectura, lo cual acelera la presentación de informes y ayuda a la mejora en la calidad de los resultados del paciente.

Así mismo esta investigación contribuye al campo del uso de Inteligencia Artificial aplicada al sector salud porque es una aplicación que une directamente el campo de redes neuronales y la detección de patrones en la región torácica y pulmonar del ser humano.

En la actualidad el campo de la inteligencia artificial aún está en desarrollo, los algoritmos de redes convolucionales usados en este estudio son basados en teorías que con cada contribución van creando nuevo conocimiento de cómo realmente funcionan.

En la actualidad los médicos tienen que emplear una combinación de historias clínicas personales, exámenes físicos, bio-marcadores individuales de cada paciente y sistemas de evaluación como el CURB-65 (puntaje de gravedad que

estima la mortalidad de la neumonía adquirida en la comunidad) para diagnosticar pacientes.

El presente estudio toma una importancia adicional debido a que colabora a la investigación del uso de Inteligencia Artificial para ayudar a los médicos al diagnosticar enfermedades usando algoritmos complejos que combinen todos esos datos con fuentes de información médica externas. El presente estudio analiza cada uno de estos algoritmos complejos a profundidad para identificar las ventajas y desventajas de cada algoritmo.

Debido a este análisis este estudio contribuye a un futuro análisis de cómo combinar estos algoritmos creando una herramienta que no solo aprenda de cientos de bio-marcadores y resultados de imágenes de placas radiográficas de millones de pacientes, sino que también aprenda de investigaciones clínicas publicadas en miles de bases de datos para hacer un diagnóstico ligado a lo último en medicina de precisión para toda enfermedad.

✓ **Justificación del estudio**

Justificación Teórica

La investigación cuenta con un valor teórico y utilidad a nivel metodológico ya que busca hacer un buen uso de las tecnologías emergentes incluyendo un enfoque de inteligencia artificial en el campo de redes convolucionales. Adicionalmente, este estudio podría ser empleado como contribución a la teoría de Transferencia de Aprendizaje y algoritmos de precisión aplicada al sistema respiratorio humano.

Justificación Metodológica

Se ha podido evidenciar que aún es escaso encontrar temas de Deep Learning e inteligencia artificial relacionados con diagnósticos médicos, por lo que se espera, que este proyecto sea un punto de partida para futuras investigaciones.

Justificación Práctica

Esta investigación tiene como objetivo aplicar tecnología de estado de arte de Inteligencia Artificial.

Justificación Económica

El software de redes convolucionales al automatizar el diagnóstico de neumonía permitirá disminuir la necesidad de contratación de adicionales especialistas en interpretación de placas radiográficas. Al apoyar en mejorar reducir el número de diagnósticos incorrectos, permitirá reducir los gastos innecesarios de pacientes en medicinas y tratamientos que no lo necesitan.

Justificación Social

La automatización de procesos en el sector salud ayuda a liberar el tiempo del talento humano médico, para que puedan desarrollar nuevas competencias y puedan contribuir a la investigación médica.

Justificación Legal

Esta investigación se sustenta en el Perú en la Ley N° 30024, ley de registro de nacional de historias clínicas electrónicas.

Justificación Ecológica

Esta investigación debido a la automatización del proceso de diagnóstico y con acceso a la información en forma virtual a través de una aplicación web, contribuye a disminuir el uso de papel y por consecuencia a disminuir la tala de árboles a nivel mundial.

1.4. Delimitación del estudio

✓ **Delimitación espacial**

La presente investigación se realizó en una entidad del sector salud de los Estados Unidos.

✓ **Delimitación temporal**

Se contó con tres (03) periodos.

- El primero fue de 1 año para la recolección de información que corresponde al año 2017.
- El segundo periodo fue de 12 meses para la investigación e implementación del sistema propuesto, en el año 2019.
- El tercer periodo fue de 1 año para la recolección y validación del sistema implementado y los estudios post-test que fueron coordinados en el área de radiología en el año 2020.

✓ **Delimitación Teórica**

Se contó como base teórica la teoría de redes neuronales convolucionales correspondiente al campo de Aprendizaje Profundo e Inteligencia Artificial.

1.5. Objetivos generales y específicos

1.5.1 Objetivo general

Implementar las Redes Neuronales Convolucionales, para mejorar la rapidez en el diagnóstico de Neumonía en un área de Radiología.

1.5.2 Objetivos específicos

- a) Implementar el Algoritmo ResNet-50 para reducir el número de diagnosing incorrectas.
- b) Implementar el Algoritmo Alexnet para reducir el número de diagnosing incorrectas.
- c) Implementar el Algoritmo VGG para reducir el número de diagnosing incorrectas.
- d) Implementar la Función de Activación SoftMax, para reducir el tiempo de espera de pacientes con neumonía.

CAPÍTULO II: MARCO TEÓRICO

2.1. Marco histórico

Historia de las Redes Neuronales Convolucionales:

- 1959: David Hubel y Torsten Wiesel describieron células simples y células complejas en la corteza visual humana. Propusieron que ambos tipos de células se utilizan en el reconocimiento de patrones. Una celda simple responde a aristas y barras de orientaciones particulares (Goodfellow, Bengio, & Courville, 2016).
- 1962: Hubel y Wiesel propusieron que las células complejas logran la invariancia espacial al sumar la salida de varias células simples que prefieren la misma orientación (por ejemplo, barras horizontales) pero diferentes campos receptivos (por ejemplo, inferior, medio o superior de una imagen) (Goodfellow, Bengio, & Courville, 2016).

Al recopilar información de un grupo de células simples, las células complejas pueden responder a las barras horizontales que se producen en cualquier lugar.

Este concepto, que los detectores simples se pueden sumar para crear detectores más complejos, se encuentra en todo el sistema visual humano, y también es la base fundamental de los modelos de redes neuronales convolucionales.

- 1979: El doctor Kunihiro Fukushima se inspiró en el trabajo de Hubel y Wiesel sobre células simples y complejas, y propuso el modelo llamado Neocognitrón (documento original: Neocognitrón: un modelo de red neuronal autoorganizadora para un mecanismo de reconocimiento de patrones no afectado por el cambio de posición), (Venkatesan & Li, 2018)

El modelo neocognitrón incluye componentes denominados células S y células C. Estas no son células biológicas, sino operaciones matemáticas. Las células S se sientan en la primera capa del modelo y están conectadas a las células C que se encuentran en la segunda capa del modelo.

La idea general es capturar el concepto de simple a complejo y convertirlo en un modelo computacional para el reconocimiento de patrones visuales.

- 1982: David Marr, un neurocientífico británico, publicó otro artículo influyente llamado Visión: una investigación computacional sobre la representación humana y el procesamiento de la información visual. Basándose en las ideas de Hubel y Wiesel (quienes descubrieron que el procesamiento de la visión no comienza con objetos holísticos),

David nos dio la siguiente idea importante: estableció que la visión es jerárquica. Argumentó que la función principal del sistema de visión es crear representaciones 3D del entorno para que podamos interactuar con él. Introdujo un marco para la visión donde los algoritmos de bajo nivel que

detectan bordes, curvas, esquinas, etc., se usan como peldaños hacia una comprensión de alto nivel de los datos visuales.

- 1989: El primer trabajo sobre redes neuronales convolucionales modernas (CNN) se produjo inspirado en el neocognitrón. Yann LeCun et al., En su artículo

Aprendizaje basado en gradientes aplicado al reconocimiento de documentos (ahora citado 17.588 veces) demostró que un modelo CNN que agrega características más simples en características progresivamente más complicadas puede usarse con éxito para el reconocimiento de caracteres escritos a mano. Específicamente, LeCun et al. Entrenó a una CNN utilizando la base de datos MNIST de dígitos escritos a mano. MNIST es un conjunto de datos ahora famoso que incluye imágenes de dígitos escritos a mano emparejados con su verdadera etiqueta de 0, 1, 2, 3, 4, 5, 6, 7, 8 o 9 (Badgeley, 2019).

Un modelo de CNN se entrena en MNIST dando es una imagen de ejemplo, pidiéndole que prediga qué dígito se muestra en la imagen y luego actualizando la configuración del modelo en función de si predijo la identidad del dígito correctamente o no. Los modelos CNN de última generación pueden alcanzar hoy una precisión casi perfecta en la clasificación de dígitos MNIST.

- 1997: Jitendra Malik, profesor de la universidad de Berkeley llamado (junto con su alumno Jianbo Shi) publicó un documento en el que describía sus intentos de abordar la agrupación perceptiva. Los investigadores trataron de hacer que las máquinas esculpieran imágenes en partes sensibles (para determinar automáticamente qué píxeles de una imagen pertenecen y distinguir los objetos de su entorno) utilizando un algoritmo de teoría de gráficos (Aja-Fernandez, Garcia, Tao, & Li, 2009).
- 2001: Paul Viola y Michael Jones introdujeron el primer marco de detección de rostros que funcionó en tiempo real. Aunque no se basa en el aprendizaje

profundo, el algoritmo todavía tenía un sabor de aprendizaje profundo, ya que, al procesar imágenes, aprendió qué características (características muy simples, similares a Haar) podrían ayudar a localizar caras (Goodfellow, Bengio, & Courville, 2016).

- 2006: Se lanzó el proyecto Pascal VOC. Proporcionó un conjunto de datos estandarizado para la clasificación de objetos, así como un conjunto de herramientas para acceder a dicho conjunto de datos y anotaciones. Los fundadores también realizaron una competencia anual, de 2006 a 2012, que permitió evaluar el desempeño de diferentes métodos para el reconocimiento de la clase de los objetos (Tursyngaliyeva, 2019).

Historia del sector salud en el Perú:

- 1972: La III Reunión Especial de ministros de Salud de las Américas aprobó el Plan Decenal de Salud para las Américas 1971-1980, en el que se fijó como meta: La cobertura total de la población con sistemas de salud en todos los países de la región (Alcalde Rabanal, Lazo Gonzales, Macias, Contreras Ochoa, & Espinosa Henao, 2019) .
- 1975 - 1979: Llegaba al Perú en calidad de representante de la Oficina Sanitaria Panamericana (OSP) Martín Vázquez Vigo, quien organizó un valioso apoyo técnico sobre sistemas nacionales de salud, logrando la difusión de los nuevos conocimientos y su complementación con la visita de expertos latinoamericanos, entre ellos durante dos años del reconocido planificador Antonio José González, exministro de salud de Argentina.

Se conformó así un grupo de profesionales peruanos conocedores de sistemas de salud, y se generó un proceso que llevó al país a ser el primero en América del Sur en crear y organizar un moderno sistema nacional de salud.

- 1978: Coincidiendo con el año de la Declaración de Alma-Ata, se dio un avance pionero con la creación del Sistema Nacional de Servicios de Salud (SNSS) por Decreto Ley 22365. El sistema tuvo año y medio de organizada

y fructífera coordinación de todas las instituciones de salud a través del Consejo Nacional de Salud, y numerosos consejos regionales, concretó logros trascendentes para construir equidad, como el primer plan nacional de atención primaria de salud, un importante programa nacional de medicamentos básicos y la aprobación y desarrollo del Régimen de Prestaciones de Salud del Seguro Social para la extensión del seguro social a la familia (Decreto Ley 22482).

- 1980: El innovador proceso del Régimen de Prestaciones de Salud del Seguro Social fue singularmente destacado por el director general de la Organización Mundial de la Salud Halfdan T. Mahler con la visita específica que hizo al Perú.
- 1985: El Sistema Nacional de Servicios de Salud fue desacelerado y luego desactivado sin evaluación alguna.
- 1990: El Sistema Nacional de Servicios de Salud fue derogado. Por Decreto Legislativo 584 fue creado el Sistema Nacional de Salud (SNS) en 1990 y nunca fue reglamentado.
- 1992: Se crearon los sistemas locales de salud (Zonadis).
- 1994: Los sistemas locales de salud (Zonadis) fueron desactivados.
- 2001: Por gestiones previas con el recientemente elegido presidente constitucional de la República, fue creada la Comisión de Alto Nivel del Ministerio de Salud encargada de reiniciar el proceso de organización del sistema nacional de salud, la cual realizó durante cinco meses una labor concertada con 127 profesionales de todas las instituciones de salud y el apoyo de importantes organismos internacionales
- 2002: Los sistemas locales de salud (Zonadis) fueron eliminados. La Comisión de Alto Nivel identificó en su informe final de enero de 2002, los diferentes problemas de salud existentes y planteó 51 propuestas de solución

con acciones concertadas. Sin embargo, la única propuesta que tramitó el Ministerio de Salud fue la creación del Sistema Nacional Descentralizado de Salud (SNDS) y, con ella, se gestionó y aprobó la Ley 27813 del Sistema Nacional Coordinado y Descentralizado de Salud (SNCDS).

- 2004: Anteproyecto de ley del Servicio de Medicina y Salud Comunitaria – SERMEC con la finalidad de desarrollar y fortalecer el primer nivel de atención del sistema nacional de salud con la estrategia de atención primaria de salud y recursos humanos calificados, fue elaborado por la Comisión de Apoyo a la Reforma en Salud con participación y concertación de los diferentes comités nacionales del Consejo Nacional de Salud.
- 2004: Anteproyecto de Ley de Aseguramiento Universal en Salud que tenía como finalidad el aseguramiento universal en salud como medio para lograr el acceso al cuidado integral de la salud de toda la población. Este anteproyecto, con base social y solidaria, fue preparado, también, por la Comisión de Apoyo a la Reforma en Salud. Pero fue archivado sin explicación alguna.
- 2009 – 2014: Inició la implementación del Aseguramiento Universal en Salud (AUS), pese a no contar con el acuerdo de algunos de los más importantes actores socio-sanitarios. Se puede afirmar que fue implementado con abierta oposición de los más importantes colegios profesionales, organizaciones de la sociedad civil y gremios laborales.

Esta situación se expresó en conflictos sindicales y múltiples cuestionamientos, que obligaron a las autoridades sanitarias a generar consensos mínimos para su operación a través del Foro por el Acuerdo Nacional.

El AUS se puso en marcha a partir de la promulgación de la Ley marco N° 29344 y su reglamento, D.S. N° 008-2010-SA. Este dispositivo legal creó la Superintendencia Nacional de Aseguramiento en Salud (Sunasa) como entidad supervisora de la operación del sistema.

Posteriormente, mediante el Decreto Legislativo N° 1158, la Sunasa vio precisadas y fortalecidas sus competencias y cambió de denominación a Superintendencia Nacional de Salud (Susalud). En suma, se transfirió a Susalud tanto la regulación de la administración financiera como la supervisión de los aparatos prestacionales público y privado, ambos instrumentos de política de primer orden.

- 2014 – presente: El Gobierno ofrece servicios de salud a la población no asegurada a cambio del pago de una cuota de recuperación de montos variables, a través del Seguro Integral de Salud (SIS) que subsidia la provisión de servicios a la población en situación de pobreza.

La prestación de servicios, tanto para el régimen subsidiado de población abierta como para la población afiliada al SIS, se realiza mediante la red de establecimientos de los Gobiernos regionales y del Ministerio de Salud (MINSA). Los otros cuatro subsistemas que brindan servicios de salud son (Alcalde Rabanal, Lazo Gonzales, Macias, Contreras Ochoa, & Espinosa Henao, 2019):

- i) el Seguro Social de Salud - EsSalud adscripto al Ministerio de Trabajo y Promoción del Empleo, que opera con su propia red de hospitales y centros de salud;
- ii) las Sanidades de las Fuerzas Armadas (Marina, Aviación y Ejército), adscritas al Ministerio de Defensa, que cuenta con sus propias instalaciones;
- iii) la Sanidad de la Policía Nacional del Perú (PNP), adscrita al Ministerio del Interior, que también cuenta con sus propias instalaciones; y
- iv) las instituciones del sector privado: entidades prestadoras de salud (EPS), aseguradoras privadas, clínicas y organizaciones de la sociedad civil (OSC).

2.2. Investigaciones relacionadas con el tema

Luego de haber realizado el levantamiento de información respectivo con relación al tema objeto de estudio se encontraron los siguientes trabajos de investigación:

- ✓ Título: Machine Learning en el Sector Salud: Identificando Neumonía con Inteligencia Artificial

Tesis para optar el grado de Bachiller Ingeniería de Tecnología de la Información que presenta

Autor: María Syed

Centro de Estudio: Universidad de Ciencias Aplicadas de Helsinki

Ciudad / País: Helsinki / Finlandia – 2018

<https://www.theseus.fi/handle/10024/153415>

Fecha de captura: 24 de noviembre 2019

Análisis, Síntesis y Discusión:

La meta de esta investigación es descubrir los fundamentos de Machine Learning (o también llamado Aprendizaje Automático), redes neuronales, visión artificial y su aplicación en diagnóstico de enfermedades en general con un hincapié en la Neumonía.

La metodología usada fue usar la técnica llamada Transfer Learning (o Transferencia de Aprendizaje) para poder tomar ventaja de redes neuronales convolucionales pre-entrenadas como por ejemplo Inception v3 y MobileNet v1. Usando esta técnica se modificaron las últimas capas de estas redes neuronales para que aprendan a detectar diagnóstico de neumonía en radiografías torácicas.

Usando las herramientas especializadas de Deep Learning: Tensorflow y Keras, se construyeron los modelos y se comparó cuál de los dos obtuvo la mayor precisión en las diagnósticos. Se determinó que el modelo basado en MobileNet v1 obtuvo la mejor precisión y

en menor tiempo. Se concluyó que el re-entrenamiento de arquitecturas de redes neuronales convolucionales son útiles en el uso de predicción de diagnóstico de neumonía, ya que produce una alta precisión en el diagnóstico y lo hace rápidamente.

En este estudio se destaca la importancia de los diferentes algoritmos de aprendizaje de redes neuronales convolucionales, y sus respectivas arquitecturas pre-entrenadas, en la rápida identificación de patrones después de re-entrenar con varias imágenes de radiografías.

- ✓ Título: Deep Learning, Lectura de Radiografías Ayudada por Computadora para la Tuberculosis: Un Estudio sobre la Exactitud de Diagnóstico en un Hospital Terciario en la India

Tesis para optar el grado de Magister en Epidemiología y Bioestadística que presenta

Autor: Madlen Nash

Centro de Estudio: Universidad McGill

Ciudad / País: Montreal, Canadá – 2019

<https://escholarship.mcgill.ca/concern/theses/hd76s4540>

Fecha de captura: 24 de noviembre 2019

Análisis, Síntesis y Discusión:

La meta de esta investigación es evaluar la precisión del diagnóstico de tuberculosis ya microbiológicamente confirmada, usando Deep Learning (o también llamado Aprendizaje profundo), redes neuronales y radiografías torácicas.

En el estudio se utilizó un software basado en redes neuronales llamado qXR Deep Learning System desarrollado por la empresa Qure.ai comparando los resultados con la interpretación de las radiografías por un radiólogo. Se comparó la especificidad y sensibilidad obtenida usando el software qXR.

Como resultado de la investigación se encontró que un software que usa Aprendizaje Profundo (redes neuronales) da resultados suficientemente óptimos para ayudar a centros de salud donde el acceso a radiólogos es limitado. Sin embargo, se recomienda más investigación al respecto para confirmar estos resultados.

En este estudio se destaca como un factor que podría influir en la disminución de la precisión del software de redes neuronales en el diagnóstico de Tuberculosis es que el paciente tenga otras enfermedades respiratorias como neumonía. Y esta variabilidad puede causar una confusión en la clasificación de solo Tuberculosis.

- ✓ Título: Visualización e Interpretación de Predicciones de Redes Neuronales Convolucionales en la Detección de Neumonía en Radiografías de Tórax Pediátricas

Artículo científico

Autor: Sivaramakrishnan Rajaraman, Sema Candemir, Incheol Kim, George Thoma y Sameer Antani

Centro de Estudio: Centro Nacional de Comunicaciones Biomédicas Lister Hill

Ciudad / País: Bethesda, Maryland / Estados Unidos – 2018

<https://www.mdpi.com/2076-3417/8/10/1715>

Fecha de captura: 24 de noviembre 2019

Análisis, Síntesis y Discusión:

La meta de esta investigación es usar métodos estadísticos para validar la precisión de modelos basados en redes neuronales convolucionales para detectar neumonía usando una versión modificada del algoritmo VGG16.

Usando radiografías pediátricas como datos de entrenamiento, se construyeron tres modelos que puedan detectar si la radiografía pertenecía a un paciente con neumonía bacteriana o viral. Una red

convolucional secuencial, una red convolucional con conexiones residuales y una basada en el algoritmo Inception.

Como resultado de la investigación se encontró que la red VGG16 tuvo la mejor precisión en la predicción del diagnóstico, pero la diferencia con los otros tres modelos no es significativa. De los tres modelos, la de mejor rendimiento fue la red convolucional residual.

- ✓ Título: Un Enfoque de Deep Learning para la Detección de Neumonía en Radiografías de Tórax

Tesis para optar el grado de Magister Ingeniería de Telecomunicaciones que presenta

Autor: Manuel Vázquez Enríquez

Centro de Estudio: Universidad de Vigo

Ciudad / País: Pontevedra, España – 2019

<http://castor.det.uvigo.es:8080/xmlui/handle/123456789/400>

Fecha de captura: 24 de noviembre 2019

Análisis, Síntesis y Discusión:

La meta de esta investigación es el uso de Deep Learning (o también llamado Aprendizaje profundo) y diferentes métodos de redes neuronales convolucionales para la clasificación correcta de radiografías, resonancias magnéticas y tomografías.

Para la detección de neumonía se entrenaron 3 diferentes modelos de redes neuronales convolucionales: Densenet, Inception v4 e Inception-resnet v2. Estos modelos fueron modificados y unidos en un solo módulo para poder que no solamente puedan individualmente hacer predicciones, sino además también se pueda combinar sus predicciones individuales en una sola predicción total.

Como resultado de la investigación se encontró que las predicciones hechas por esta arquitectura de múltiples modelos fueron altamente

precisas en comparación a otra arquitectura llamada CheXNet, también basada en redes neuronales convolucionales. Sin embargo, el uso de múltiples modelos usa demasiados recursos en términos de hardware y además el tiempo de entrenamiento y predicción no son óptimos.

En este estudio se destacan las limitaciones de usar varios modelos al mismo tiempo. Deja la puerta abierta a investigaciones que se enfoquen en usar un solo modelo que dé resultados similares en menor tiempo y sin los altos requerimientos computacionales.

- ✓ Título: Deep Learning para la Segmentación de Imágenes Médicas: Detección de Neumonía

Tesis para optar el grado de Magister Ingeniería de Ciencias de la Computación que presenta

Autor: Assem Tursyngaliyeva

Centro de Estudio: Universidad de Nazarbayev

Ciudad / País: Astana, Kazajistán – 2019

https://www.researchgate.net/publication/335378335_Deep_Learning_for_Medical_Image_Segmentation_Pneumonia_Detection

Fecha de captura: 24 de noviembre 2019

Análisis, Síntesis y Discusión:

La meta de esta investigación es la implementación de una arquitectura de Deep Learning (o también llamado Aprendizaje profundo) para diagnosticar casos de neumonía basada en datos obtenidos de la Sociedad Radiológica de Norteamérica.

Se crearon las siguientes arquitecturas de redes neuronales convolucionales: UNet (con aumento de datos), WNet (construida con dos UNets). Además de redes pre-entrenadas como ResNet152, Dense121 y CheXNet (121-capas). Todas siendo entrenadas con radiografías torácicas.

Como resultado de la investigación se encontró que el modelo WNet a pesar de ser una arquitectura de dos UNets, no solo demoró más entrenar, sino que también tuvo menor precisión. Pero ambas tuvieron mejores resultados que las redes pre-entrenadas. Se destaca del estudio que es posible que UNet haya obtenido la mejor precisión a pesar de tener solo 23 capas debido a su habilidad de aumentar datos.

- ✓ Título: Deep Learning Multimodal para Mejorar la Práctica de la Radiología
Tesis para optar el grado de Doctor de Ciencias Biomédicas que presenta
Autor: Marcus Alexander Badgeley
Centro de Estudio: Escuela de Medicina Icahn
Ciudad / País: New York, Estados Unidos – 2018
<https://pqdtopen.proquest.com/pubnum/10974774.html>
Fecha de captura: 24 de noviembre 2019

Análisis, Síntesis y Discusión:

La meta de esta investigación es la implementación de redes neuronales convolucionales para escanear radiografías de Tomografía computarizada para buscar evidencia de patología crítica con énfasis en la velocidad de resultados más que la precisión.

Obteniendo las radiografías de 3 diferentes centros de salud, se creó una aplicación web llamada CANDI que usa la arquitectura de redes convolucionales Inceptionv3 pre-entrenada con pesos de ImageNet. Para poder visualizar se modificaron los modelos. Y se usó Regresión logística para obtener las predicciones.

Luego este mismo modelo es usado en radiografías para diagnosticar neumonía. Como resultado de la investigación se encontró que el modelo propuesto fue suficientemente óptimo, sin embargo, el modelo ampliamente depende de la cantidad de datos de pacientes

que use para entrenar. Se destaca la importancia de la necesidad de una alta cantidad de datos cuando se entrena la red para obtener una mejor predicción.

2.3. Estructura teórica y científica que sustenta el estudio

La estructura teórica que soporta la presente investigación está en función a los siguientes términos que se han considerados como primordiales para el estudio:

Redes Convolucionales:

“El aprendizaje profundo (Deep Learning), a veces denominado redes neuronales profundas, es un subconjunto del aprendizaje automático que involucra múltiples capas de datos en lugar de solo uno. A diferencia del aprendizaje automático, las redes de aprendizaje profundo no necesitan ser programadas con criterios específicos que definan un elemento. Las redes de aprendizaje profundo pueden definir elementos simplemente al estar expuestas a grandes cantidades de datos.” (Nash, 2019, p. 14)

En (Rajaraman, Candemir, Kim, Thoma, & Antani, 2018) los autores señalan acerca de las redes neuronales convolucionales:

“En los últimos años, se ha demostrado que los métodos de aprendizaje profundo (DL) basados en datos evitan los problemas con las características artesanales mediante la extracción y clasificación de características de extremo a extremo. Las redes neuronales convolucionales (CNN) pertenecen a una clase de modelos DL que se utilizan de manera prominente en la visión por computadora.

Estos modelos tienen múltiples capas de procesamiento para aprender representaciones jerárquicas de características a partir de los datos de píxeles de entrada. Las características en las primeras capas de la red se abstraen a través de los mecanismos de los campos receptivos locales, el reparto de peso y la agrupación para formar representaciones de características ricas hacia el aprendizaje y la clasificación de las entradas a sus respectivas clases.

Debido a la falta de datos de imágenes médicas suficientemente extensos, las CNN capacitadas en colecciones de datos a gran escala, como ImageNet, se utilizan para transferir el conocimiento de las representaciones aprendidas en forma de características de imágenes genéricas a la tarea actual.

Las CNN también muestran resultados prometedores en tareas de detección y localización de objetos. El sorprendente éxito de las CNN profundas junto con la falta de una toma de decisiones explicable ha dado lugar a una percepción de duda. Este modelo de comportamiento poco conocido ha limitado su uso en la práctica clínica habitual.” (Rajaraman, Candemir, Kim, Thoma, & Antani, 2018, pág. 2)

En (Ranschaert, Morozov, & Algra, 2019) el autor define a las redes neuronales convolucionales:

“Es un tipo específico de red neuronal artificial que utiliza perceptrones, un algoritmo de unidad de aprendizaje automático, para aprendizaje supervisado, para analizar datos. Las redes neuronales convolucionales se aplican al procesamiento de imágenes, procesamiento del lenguaje natural y otros tipos de tareas cognitivas.

Una red neuronal convolucional también se conoce como ConvNet. Una red neuronal convolucional consiste en una capa de entrada y salida, así como múltiples capas ocultas que se forman como operaciones matemáticas. Las capas ocultas incluyen capa convolucional, capa de agrupación, normalización y capas completamente conectadas.” (Ranschaert, Morozov, & Algra, 2019, p. 352)

En (Badgeley, 2019) el autor define a las redes neuronales convolucionales en su aplicación a reconocimiento de imágenes:

“Los algoritmos clásicos de reconocimiento de imágenes requerían que los médicos e ingenieros desarrollaran descripciones matemáticas de cómo se ve la enfermedad, y luego los modelos de aprendizaje automático fueron entrenados

para aprender qué combinaciones de características indican la enfermedad. Los modelos de aprendizaje profundo utilizados para el reconocimiento de imágenes son redes neuronales con capas especiales para la codificación de patrones espaciales. Estos modelos se denominan redes neuronales convolucionales (CNN) porque la operación de convolución matemática se utiliza para codificar características espaciales.

Las CNN supuestamente automáticamente aprenden estas características, en contraste con las características de ingeniería de reconocimiento de imágenes clásicas. Una CNN aprende una tarea ajustando tanto las características convolucionales (aprendizaje de características de imagen) como las asociaciones de objetivos de características (identificación / caracterización). La entrada estándar para una CNN es una matriz de píxeles de imagen. Las CNN se pueden usar para numerosas tareas con distintos formatos de salida.

Para problemas de clasificación, una CNN transforma una radiografía completa en una probabilidad de clase escalar. Modifico modelos de varias maneras para adaptarlos a aplicaciones médicas. Debido a que una radiografía tiene más de una enfermedad u objetivo de entrenamiento, utilizamos modelos de tareas múltiples. Los modelos de tareas múltiples utilizan una representación de imagen subyacente compartida y un nodo de salida para cada variable de predicción. Para la entrada, combino radiografías con rasgos escalares del paciente y variables de adquisición de imágenes al concatenarlas en el vector de representación de imagen incrustado.

La mayoría de los parámetros del modelo están contenidos dentro de las capas convolucionales de codificación de imágenes del modelo. Los conjuntos de datos de atención médica son órdenes de magnitud más pequeños que los conjuntos de datos de imágenes cotidianos. Entonces, en lugar de capacitar a las CNN de nuevo, utilizamos conjuntos de datos de imágenes masivos para entrenar incrustaciones de imágenes de uso general y luego realizar una segunda fase de capacitación con los datos de atención médica.

Esta estrategia de entrenamiento paso a paso se llama Transfer Learning (transferencia de aprendizaje), y en la práctica las imágenes se caracterizan por características robustas de propósito general, y luego los patrones de características para una tarea de atención médica se aprenden en la fase de ajuste secundaria del entrenamiento.” (Badgeley, 2019, p. 6)

En (Goodfellow, Bengio, & Courville, 2016) el autor define a las redes neuronales convolucionales:

“Son un tipo especializado de red neuronal para procesar datos que tiene una topología conocida similar a una cuadrícula. Los ejemplos incluyen datos de series temporales, que pueden considerarse como una cuadrícula 1D que toma muestras a intervalos de tiempo regulares, y datos de imagen, que pueden considerarse como una cuadrícula de píxeles en 2D.

Las redes convolucionales han sido tremendamente exitosas en aplicaciones prácticas. El nombre "red neuronal convolucional" indica que la red emplea una operación matemática llamada convolución. La convolución es un tipo especializado de operación lineal. Las redes son simplemente redes neuronales que utilizan convolución en lugar de una multiplicación de matriz en al menos una de sus capas.” (Goodfellow, Bengio, & Courville, 2016, p. 330)

El autor Manuel Vázquez Enríquez señala acerca de redes neuronales artificiales y Deep Learning:

“Deep Learning es una rama del aprendizaje automático. La característica principal y común es su capacidad de aprender automáticamente características y funciones representativas de la información de entrada, que se llama aprendizaje de representación o aprendizaje de características.

Deep Learning lleva a cabo la capacitación y con ella el aprendizaje a través de múltiples nodos o neuronas distribuidas en múltiples capas, simulando el funcionamiento del sistema nervioso humano sistema, esta similitud se muestra.

En el caso de redes neuronales artificiales, comenzamos desde una unidad de procesamiento básica llamada nodo o neurona. Cada neurona es en sí misma

Un modelo clasificador simple que genera una salida en función de las evidencias de las capas anteriores.” (Vázquez Enríquez, 2019, pág. 11)

En la Figura 03 el autor Manuel Vázquez Enríquez muestra la similitud entre una neurona biológica y una red neuronal artificial:

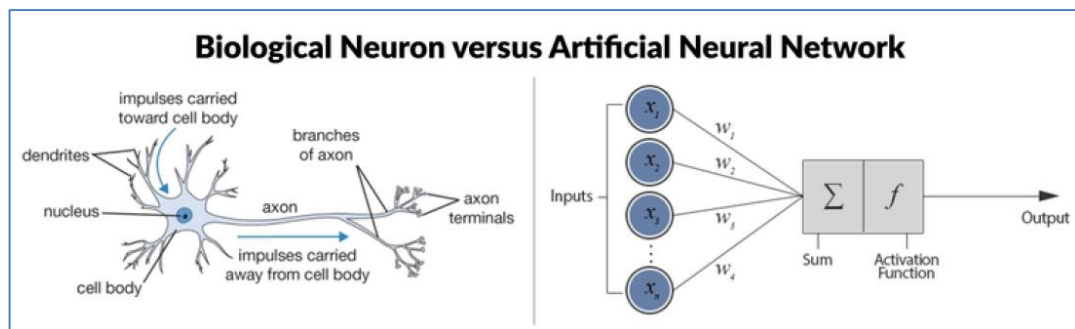


Figura 03: Comparación de una neurona biológica y una red neuronal artificial. (Vázquez Enríquez, 2019, p. 11)

Manuel Vázquez Enríquez agrega acerca de redes neuronales artificiales:

“Cada neurona realiza tres acciones: Operación lineal, que es una función de propagación que agrega las señales provenientes de las anteriores capas con sus respectivos pesos y un valor de sesgo. Operación no lineal, que es la función de propagación donde se aplica una función de activación que determina si la salida de la neurona está activada o no.

Existen varias funciones de activación y dependiendo del espacio ocupado por el nodo dentro la red neuronal es conveniente usar uno u otro. Las más comunes funciones de activación son: Sigmoid, Tanh, ReLU y Softmax. Operación de Salida. El resultado de la función de activación se encuentra a la salida del nodo y se propagará a todos los nodos ubicados en la siguiente capa.

Estos nodos o neuronas se organizan en capas y a través de la interacción con otras neuronas distribuidas en diferentes capas permiten obtener más compleja

y jerárquicas representaciones. Cada capa actúa esencialmente sobre las construcciones características de las capas anteriores generando una construcción de características de nivel superior.” (Vázquez Enríquez, 2019, pág. 11)

Manuel Vázquez Enríquez agrega acerca de la estructura de una red neuronal:

“Las tres partes principales de una red neuronal se diferencian: capa de entrada, capas ocultas y la capa de salida. El número de capas de entrada y salida son fijas y su forma depende del conjunto de datos de entrada y la formulación del problema o función objetivo buscados respectivamente. Pero el tamaño de las capas ocultas puede ser variable. Si el modelo usa una sola capa oculta, se llamará red neuronal superficial e introduciendo más capas tendremos un modelo más profundo.” (Vázquez Enríquez, 2019, pág. 12)

En la Figura 04 el autor Manuel Vázquez Enríquez muestra la estructura de una red neuronal artificial con sus diferentes capas:

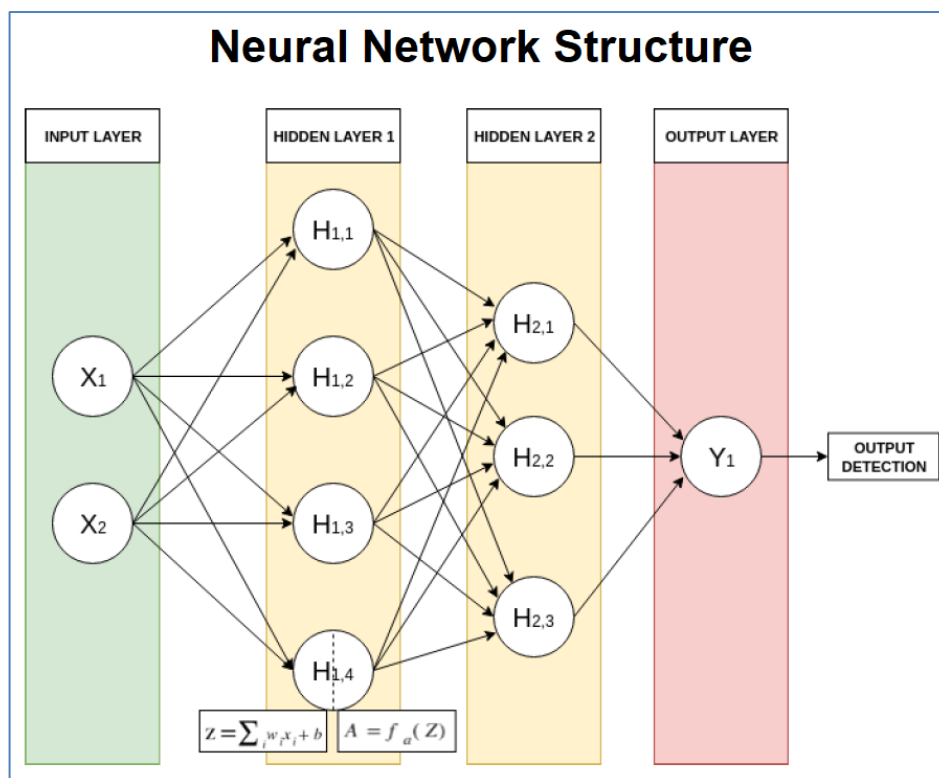


Figura 04: Estructura de una red neuronal simple con sus diferentes capas. (Vázquez Enríquez, 2019, p. 12)

El autor Manuel Vázquez Enríquez señala acerca de las redes convolucionales:

“El esquema básico de una red neuronal que muestra estructura de red neuronal de avance donde cada nodo está conectado a todos los anteriores nodos de capa. Pero esta estructura no es viable para trabajar con imágenes. La imagen es una entrada multidimensional donde los píxeles que lo componen mantienen una relación de local correlación con los píxeles vecinos, si utilizamos la red neuronal de avance el número de los parámetros necesarios haría que su entrenamiento sea inviable.

Por lo tanto, para trabajar con imágenes surgen redes neuronales convolucionales. Las redes neuronales convolucionales son un tipo de red neuronal especializada para trabajar con imágenes. Estas redes se inspiran en el procesamiento de la Corteza visual cuando el cerebro humano procesa una imagen. Cada nodo cubre una región de la imagen y junto con el resto de los nodos toda la imagen es cubierta.” (Vázquez Enríquez, 2019, pág. 17)

El autor Manuel Vázquez Enríquez agrega acerca de la estructura de las redes convolucionales:

“En una red neuronal convolucional se combinan 3 tipos de capas principales:

- La Capa convolucional, que actúa como un extractor de características para la imagen de entrada. Aplica una serie de filtros o núcleos, que contienen los pesos entrenables y que funcionan a través de toda la imagen ejecutando operaciones de convolución y generando un mapa de características. Cada uno de los filtros termina por especializarse en detectar ciertas características. En las primeras capas detectan patrones simples como diferentes líneas o curvas y en las capas posteriores detectan características o patrones más complejos, como los objetos o formas. Después de una capa convolucional se aplica una función de activación, más comúnmente usando la unidad lineal rectificadora (ReLU).

- La Capa de agrupamiento. Esta capa tiene la función de reducir la dimensionalidad de los mapas de características. Las operaciones más comunes son, por un lado, detectar el máximo valor de la región de submuestra descartando el resto para esta operación o para calcular el promedio de los elementos de cada región.
- La Capa totalmente conectada. Esta es una o más capas completamente conectadas y está ubicada cerca de la salida. Estas capas aprenden las relaciones entre mapas de características y obtiene predicciones de salida que minimizan la función de costo.” (Vázquez Enríquez, 2019, pág. 17)

Algoritmo Alexnet

Alexnet es una arquitectura de redes convolucionales basada en una anterior arquitectura llamada LeNet. En (Sewak, Karim, & Pujari, 2018) se explica primero el modelo LeNet:

“En 2010, salió un desafío de ImageNet (conocido como ILSVRC 2010) con una arquitectura de red neuronal convolucional, LeNet 5, construida por Yann Lecun. Esta red toma una imagen de 32 x 32 como entrada, que va a las capas de convolución (C1) y luego a la capa de submuestreo (S2). Hoy, la capa de submuestreo se reemplaza por una capa de agrupación.

Luego, hay otra secuencia de capas de convolución (C3) seguida de una capa de agrupación (es decir, submuestreo) (S4). Finalmente, hay tres capas completamente conectadas, incluida la capa de SALIDA al final. Esta red se utilizó para el reconocimiento de código postal en las oficinas de correos. Desde entonces, cada año se introdujeron varias arquitecturas de redes neuronales convolucionales con la ayuda de esta competencia.” (Sewak, Karim, & Pujari, 2018, p. 101)

En la Figura 05 el autor muestra la estructura de una red neuronal convolucional LeNet 5 con sus diferentes capas:

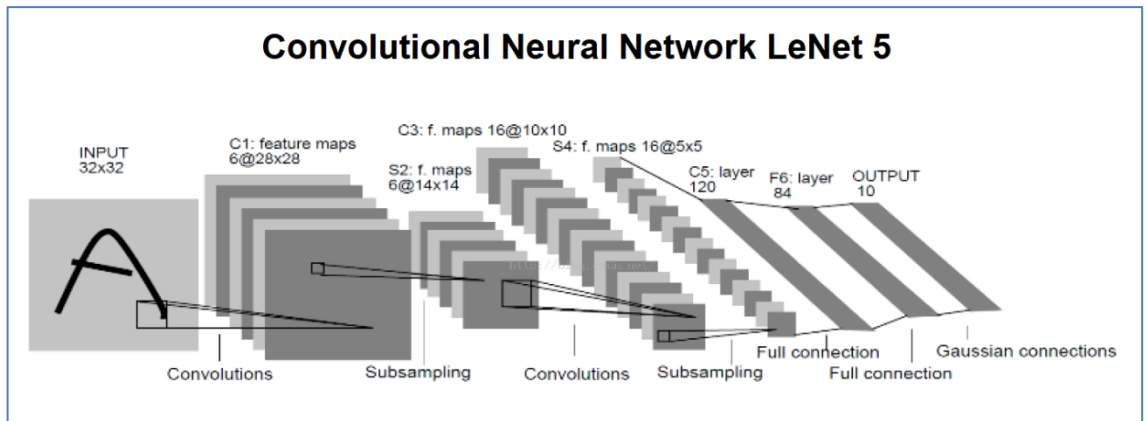


Figura 05: Estructura de una red neuronal simple con sus diferentes capas. (Sewak, Karim, & Pujari, 2018, p. 101)

En (Venkatesan & Li, 2018) se explica acerca de AlexNet:

“Al igual que el conjunto de datos MNIST fué influyente para hacer que LeNet5 fuese ampliamente conocido y utilizado, ImageNet es un conjunto de datos a gran escala en el que muchos investigadores confiaron mucho en la tercera ola de redes neuronales. En el momento de AlexNet, ImageNet era un conjunto de datos que tenía más de 15 millones de imágenes en más de 22,000 categorías.

La competencia ImageNet, el desafío de reconocimiento visual a gran escala ImageNet (ILSVRC), utilizó alrededor de 1000 de estas categorías. Las categorías de ImageNet son mucho más complicadas que otros conjuntos de datos y, a menudo, son difíciles de clasificar incluso para los humanos. El rendimiento promedio a nivel humano es de aproximadamente el 96% en este conjunto de datos.

Los sistemas de reconocimiento que utilizan técnicas convencionales pueden lograr, en el mejor de los casos, un rendimiento con solo alrededor del 75% de precisión, y en 2012 AlexNet fue el primer sistema en romper la marca del 80%.” (Venkatesan & Li, 2018, pág. 103)

En (Venkatesan & Li, 2018) se señala la relación entre la arquitectura LeNet y AlexNet, y también lo innovador del modelo AlexNet:

“AlexNet en su núcleo es simplemente una LeNet más profunda. Alexnet tiene cinco capas de agrupación convolucional seguidas de dos capas completamente conectadas y una capa softmax de 1000 nodos, una para cada categoría. Aunque a primera vista esta red parece una extensión directa y una profundización de LeNet en términos de su filosofía arquitectónica, la gran cantidad de parámetros y, por lo tanto, la complejidad del modelo conduce a cantidades drásticas de sobreajuste.

Esta red por sí sola nunca podría entrenarse de manera estable con el conjunto de datos ImageNet a pesar de que hay más de 15 millones de imágenes. Computacionalmente, la capacitación de una red tan grande también es una preocupación importante. En el momento de su implementación, AlexNet era una de las redes neuronales más grandes y profundas.

Además del algoritmo de entrenamiento básico para esta CNN que utiliza el descenso de gradiente estocástico, Krizhevsky et al. Tuvo que usar varias técnicas nuevas en combinación para lograr que esta red entrenara sin sobreajustar y de una manera firme. Estudiaremos estas técnicas en detalle a continuación. Una novedad importante de AlexNet fue la elección de usar la activación ReLU en lugar de las activaciones sigmoideas o tanh más tradicionales (Nair y Hinton, 2010).

Las activaciones sigmoideas y de Tanh, como discutimos en los capítulos anteriores, son activaciones saturantes. Las ReLU no se saturan cerca de 1 como otras funciones de activación y, por lo tanto, aún tienen un gradiente mayor a medida que nos acercamos a la unidad. Esto ayuda a aumentar la velocidad de aprendizaje. Otra contribución importante de AlexNet es el uso de múltiples GPU para el aprendizaje.

La implementación original todavía está disponible como una red convolucional cuda, aunque desde entonces se han implementado versiones avanzadas en forma de muchas cajas de herramientas.” (Venkatesan & Li, 2018, pág. 104)

En la Figura 06 el autor muestra la estructura de una red neuronal convolucional AlexNet con sus diferentes capas:

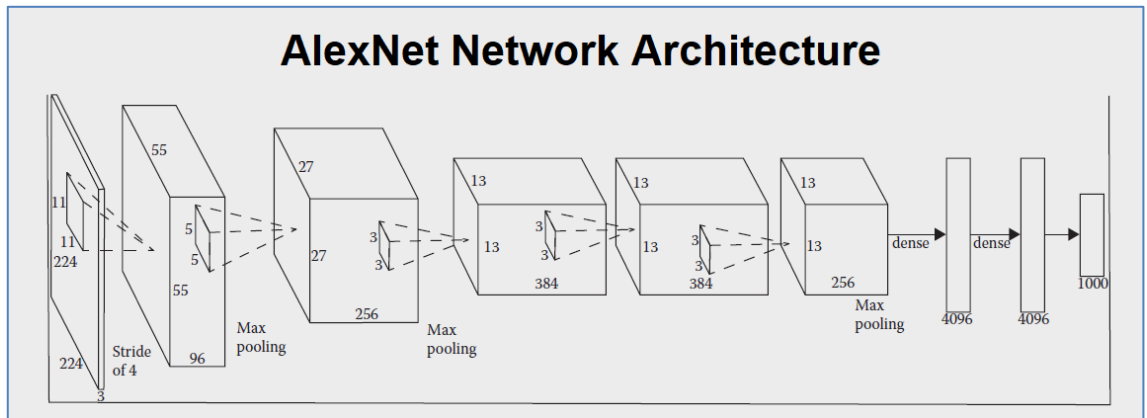


Figura 06: La entera arquitectura de la red neuronal convolucional AlexNet. (Sewak, Karim, & Pujari, 2018, p. 103)

En (Venkatesan & Li, 2018), el autor agrega acerca de los avances creados por el modelo AlexNet:

“Otro avance significativo en AlexNet fue el uso de Dropouts. AlexNet es una red con una gran cantidad de parámetros y el entrenamiento de una red de este tamaño sin la regularización adecuada conducirá a un sobreajuste severo. El Dropout es un método para evitar la coadaptación de mapas de características (Srivastava et al., 2014).

Idealmente, para mejorar el rendimiento, nos gustaría entrenar a muchos modelos y combinar sus decisiones tomando una encuesta promedio de todos los modelos que hemos entrenado. Esta técnica se llama aprendizaje conjunto y se usa típicamente en bosques aleatorios y árboles de decisión. Cuando se trata de redes neuronales, el aprendizaje conjunto plantea un problema computacional.

Entrenar a un AlexNet ya era difícil en dos GPU; entrenar a decenas o incluso cientos de ellos sería una imposibilidad logística. La deserción es un método que aborda estos problemas. La deserción es una técnica en la que la salida de

una neurona se fuerza a cero con una probabilidad aleatoria. Esta probabilidad aleatoria es típicamente 0.5 Bernoulli.

Esto significa que, para cada neurona, asociamos una variable aleatoria de la cual extraemos un estado de una distribución de Bernoulli con una probabilidad de 0.5. Si la muestra extraída era 1, permitimos que la señal atravesara la neurona; si era 0, reducimos la salida de la neurona a 0, independientemente de si la neurona estaba activa o no. Esto implica que aproximadamente la mitad de las neuronas en la red normalmente se apagan al azar mientras aprenden.

Esto crea muchos efectos ventajosos. Dado que la red ahora se ejecuta con solo la mitad de las representaciones, cada neurona está ahora bajo el doble de estrés para aprender características significativas. Esto permite un aprendizaje más rápido sin sobreajustar.

Dado que las neuronas se apagan y encienden al azar, las neuronas ya no pueden adaptarse y deben aprender características que son independientes de otras neuronas.

Más importante aún, durante cada retropropagación efectivamente, se crea una nueva arquitectura mediante el uso de abandonos. Cada propagación directa es un modelo de red completamente nuevo que estamos aprendiendo en nuestro conjunto. La única diferencia es que estos modelos ahora comparten parámetros. Los Dropouts generalmente no se usan en una capa convolucional en el contexto tradicional, ya que eliminarían un mapa de características completo.

Los Dropouts se aplican a respuestas de entidades individuales que abandonan solo unas pocas ubicaciones en los mapas de entidades al azar, lo que en sí mismo no es ideal.” (Venkatesan & Li, 2018, págs. 105-106)

En la Tabla 05 se muestra en detalle la estructura de una red neuronal convolucional AlexNet escribiendo el número entradas y salidas en cada capa, además del tipo de capa:

Tabla 05:
Arquitectura de la Red Neuronal Convolutiva AlexNet

LAYER NUMBER	INPUT SHAPE	RECEPTIVE FIELD	NUMBER OF KERNELS	TYPE OF NEURONS
1	224 × 224 × 3	11 × 11, stride 4	96	Convolutional
2	-	3 × 3, stride 2	-	Pooling
3	55 × 55 × 96	5 × 5	256	Convolutional
4	-	3 × 3, stride 2	-	Pooling
5	27 × 27 × 256	3 × 3, padded	384	Convolutional
6	13 × 13 × 384	3 × 3, padded	384	Convolutional
7	13 × 13 × 384	3 × 3	256	Convolutional
8	13 × 13 × 256	-	4096	Fully connected
9	4096	-	4096	Fully connected
10	4096	-	1000	Softmax

Fuente: (Venkatesan & Li, 2018)

Algoritmo VGG

En (Sewak, Karim, & Pujari, 2018) el autor define la arquitectura de redes convolucionales VGG:

“Esta red neuronal convolutiva es una arquitectura simple y elegante con una tasa de error de 7.3%. Tiene dos versiones: VGG16 y VGG19. VGG16 es una red neuronal de 16 capas, sin contar la capa de agrupación máxima y la capa softmax. Por lo tanto, se conoce como VGG16. VGG19 consta de 19 capas.

La consideración clave del diseño aquí es la profundidad. Se lograron aumentos en la profundidad de la red agregando más capas de convolución, y se hizo debido a los pequeños filtros de convolución 3 x 3 en todas las capas. El tamaño de entrada predeterminado de una imagen para este modelo es 224 x 224 x 3. La imagen se pasa a través de una pila de capas de convolución con una zancada de 1 píxel y un relleno de 1.

Utiliza convolución de 3 x 3 en toda la red. La agrupación máxima se realiza sobre una ventana de 2 x 2 píxeles con una zancada de 2, luego otra pila de capas de convolución seguidas por tres capas completamente conectadas. Las primeras dos capas completamente conectadas tienen 4,096 neuronas cada una,

y las terceras capas completamente conectadas son responsables de la clasificación con 1,000 neuronas.

La capa final es una capa softmax. VGG16 utiliza una ventana de convolución 3 x 3 mucho más pequeña, en comparación con la ventana de convolución 11 x 11 mucho más grande de AlexNet. Todas las capas ocultas se crean con la función de activación ReLU. Debido al pequeño filtro de convolución 3 x 3, se aumenta la profundidad de VGGNet.

El número de parámetros en esta red es de aproximadamente 140 millones, principalmente de la primera capa totalmente conectada. En las arquitecturas de los últimos días, las capas completamente conectadas de VGGNet se reemplazan con capas de agrupación de promedio global (GAP) para minimizar el número de parámetros.” (Sewak, Karim, & Pujari, 2018, pp. 103-104)

En la Figura 07 el autor muestra la estructura de una red neuronal convolucional VGG16 con sus diferentes capas:

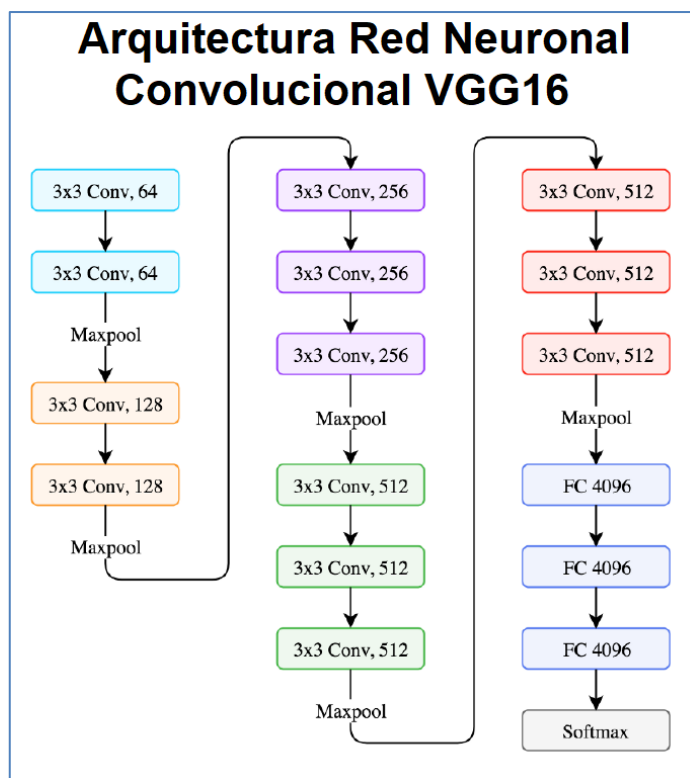


Figura 07: Arquitectura completa del modelo VGG16. (Sewak, Karim, & Pujari, 2018, p. 104)

En (Venkatesan & Li, 2018), el autor explica acerca de la red neuronal VGG:

“La red VGG funciona con una filosofía de tamaño de filtro diferente de las redes anteriores. En todas las redes anteriores, consideramos filtros que son más grandes en las capas anteriores y más pequeños en las capas más profundas. VGG sigue la misma idea, pero comienza con un campo receptivo que es tan pequeño como 3×3 y crece a un ritmo constante de 3×3 .

La red VGG considera solo filtros convolucionales 3×3 con una capa de agrupación máxima de 2×2 y los apila para crear una red convolucional. Esta red está muy cerca de LeNet en todos sus paradigmas arquitectónicos, sin módulos de inicio o similares. El documento original de VGG incluye cinco arquitecturas diferentes, pero solo nos interesa la quinta (E), ya que es la más utilizada.

La red E es una red convolucional simple que tiene 24 capas de profundidad y se conoce comúnmente como VGGNet o VGG-19. La estrategia para construir una red grande en capas usando filtros pequeños más pequeños fue originalmente analizar el efecto de la profundidad en el rendimiento y, como se esperaba, el estudio encuentra que cuanto más profunda sea la red, mejor será el rendimiento.

La afirmación de ser más profundo ha sido cuestionada por algunos resultados recientes y, por lo tanto, no es una conclusión establecida, aunque los resultados reportados por VGG funcionan para capas de los tamaños estudiados. Al ser un filtro 3×3 en la primera capa, es interesante ver el tipo de plantillas que la red está aprendiendo a detectar.

Tenga en cuenta que los filtros son predominantemente detectores de esquinas, detectores de bordes y manchas de color. Estos son los bloques de construcción fundamentales para construir características de orden superior como discutimos en capítulos anteriores. VGG19 se ha convertido en una red prototípica y sus características de primera capa parecen ser una prueba más de que las CNN se

pueden utilizar para aprender detectores de características.” (Venkatesan & Li, 2018, págs. 110-111)

En la Figura 08 el autor muestra la estructura de una red neuronal convolucional VGG-19 con sus diferentes capas:

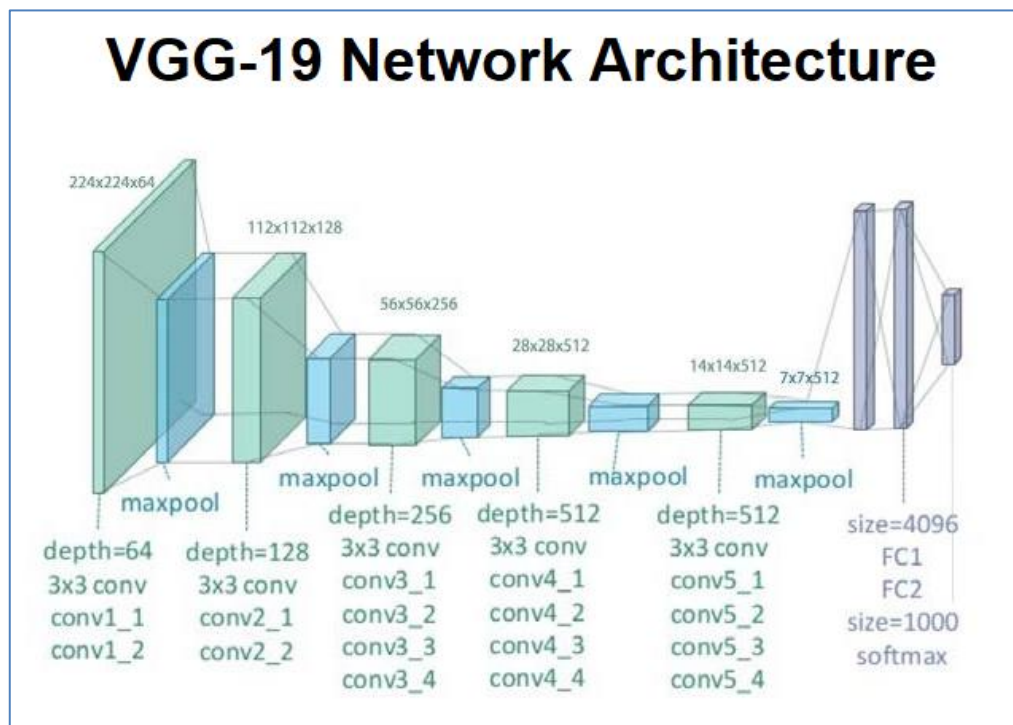


Figura 08: La entera arquitectura de la red neuronal convolucional VGG-19. Fuente: (Venkatesan & Li, 2018)

En la Tabla 06 se muestra en detalle la estructura de una red neuronal convolucional VGG escribiendo el número entradas y salidas en cada capa, además del tipo de capa (Elgendy, 2020):

Tabla 06:

Arquitectura de la Red Neuronal Convolucional VGG

LAYER NUMBER	RECEPTIVE FIELD	NUMBER OF KERNELS	TYPE OF NEURONS
1	3 × 3, stride 1	64	Convolutional
2	3 × 3, stride 1	64	Convolutional
3	2 × 2, stride 1	-	Pooling
4	3 × 3, stride 1	128	Convolutional
5	3 × 3, stride 1	128	Convolutional
6	2 × 2, stride 1	-	Pooling
7	3 × 3, stride 1	256	Convolutional
8	3 × 3, stride 1	256	Convolutional

LAYER NUMBER	RECEPTIVE FIELD	NUMBER OF KERNELS	TYPE OF NEURONS
9	3 × 3, stride 1	256	Convolutional
10	3 × 3, stride 1	256	Convolutional
11	2 × 2 stride 1	-	Pooling
12	3 × 3, stride 1	512	Convolutional
13	3 × 3, stride 1	512	Convolutional
14	3 × 3, stride 1	512	Convolutional
15	3 × 3, stride 1	512	Convolutional
16	2 × 2, stride 1	-	Pooling
17	3 × 3, stride 1	512	Convolutional
18	3 × 3, stride 1	512	Convolutional
19	3 × 3, stride 1	512	Convolutional
20	3 × 3, stride 1	512	Convolutional
21	2 × 2, stride 1	-	Pooling
22	-	4096	Fully connected
23	-	4096	Fully connected
24	-	1000	Softmax

Fuente: (Elgendy, 2020)

Algoritmo ResNet-50

“Red Residual, es una red muy profunda que consta de capas con conexiones omitidas. Fue entrenado en el conjunto de datos ImageNet. ResNet es el ganador de ILSVRC 2015 en clasificación, detección y localización de imágenes. La red aprende las funciones de representación residual en lugar de reconocer la representación de la señal inmediatamente.

ResNet practica la conexión de omisión, es decir, la conexión de acceso directo, para ajustar la entrada de la capa anterior a la capa siguiente sin cambiar la entrada. Entonces la salida $H(x) = F(x) + x$. Dicha conexión de omisión permite tener una red más profunda.” (Tursyngaliyeva, 2019, p. 18)

En (Ranschaert, Morozov, & Algra, 2019) se señala acerca de Redes Residuales:

“Una Red neuronal residual (RNN) es una red que omite las conexiones a través de capas de red, haciendo accesos directos o saltos. Una red residual se salta sobre una sola capa.” (Ranschaert, Morozov, & Algra, 2019, p. 361)

Además, en (Ranschaert, Morozov, & Algra, 2019) se agrega la importancia de una capa residual:

“Hay algunos tipos de capas adicionales que se están desarrollando y aplicando con gran ventaja. Un ejemplo reciente es la capa residual que recibe su nombre porque usa una capa de derivación que es esencialmente la función de identidad, y luego la salida de una capa o grupo de capas se compara con esa función de identidad.

Esto obliga efectivamente a las capas sin derivación a funcionar mejor que una función de identidad y, por lo tanto, puede aprender de manera más efectiva con menos capas. Esto es importante porque la reducción en las capas reduce la cantidad de parámetros potenciales para ajustar cuando se aprende y también reduce la posibilidad de sobreajuste a los datos de entrenamiento. (Ranschaert, Morozov, & Algra, 2019, p. 44)

En (Sewak, Karim, & Pujari, 2018) se señala acerca de la arquitectura ResNet en conjunto a las redes convolucionales:

“Después de una cierta profundidad, la adición de capas adicionales para las redes convolucionales de avance genera un error de entrenamiento más alto y un error de validación más alto. Al agregar capas, el rendimiento aumenta solo hasta una cierta profundidad y luego disminuye rápidamente.

En el artículo de ResNet (Red Residual), los autores argumentaron que esta adaptación insuficiente es improbable debido al Problema del desvanecimiento del gradiente, porque esto ocurre incluso cuando se usa la técnica de normalización por lotes. Por lo tanto, han agregado un nuevo concepto llamado bloque residual. El equipo de ResNet agregó conexiones que pueden omitir capas.

Redes Residuales utiliza redes neuronales estándares y agregan conexiones que omiten algunas capas de convolución a la vez. Cada desvío crea un bloque residual.” (Sewak, Karim, & Pujari, 2018, p. 108)

En la Figura 09 el autor muestra la estructura de un bloque residual de una red residual:

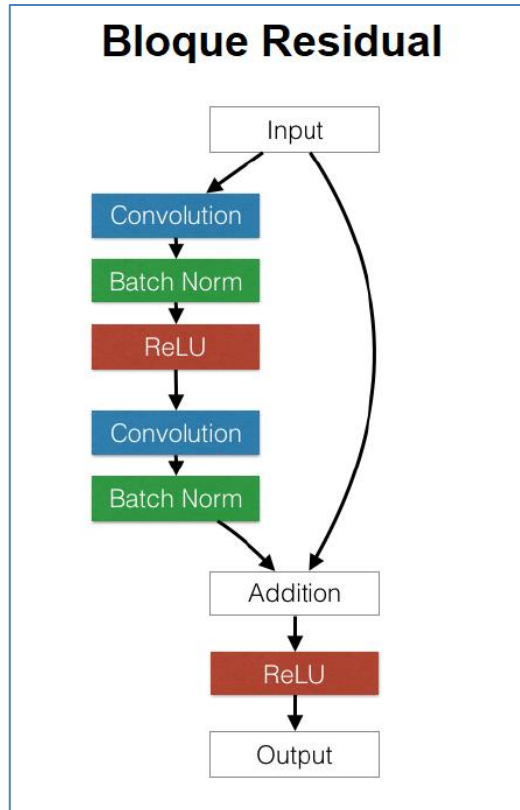


Figura 09: Bloque Residual de una Red Residual. (Sewak, Karim, & Pujari, 2018, p. 108)

En (Venkatesan & Li, 2018) el autor señala acerca de la arquitectura ResNet:

“Las redes residuales o ResNets son una arquitectura novedosa, que utiliza la idea de dividir una capa en dos ramas, donde una rama no hace nada a la señal y la otra la procesa como lo haría una capa típica (He et al., 2015). Los datos no procesados o el residual se agregan a la señal original que pasa por la red sin alteraciones.

Esto crea una división en la red, donde una rama simplemente propaga el gradiente sin alterarlo. Esto permite que una red más profunda aprenda con fuertes gradientes que la atraviesan. Mientras tanto en GoogLeNet como en

ResNet los datos (y, por lo tanto, el gradiente) pasan, el módulo de inicio todavía tenía una capa de agrupación, que el módulo ResNet evita por completo.

Tenga en cuenta que hay algunos artículos que demuestran que una ResNet muy profunda es equivalente a una red neuronal recurrente poco profunda (donde una capa se alimenta de su yo temporalmente pasado, con o sin memoria de almacenamiento) (Liao y Poggio, 2016) y similares. También se propuso una arquitectura llamada redes de carreteras, inspirada en las redes neuronales recurrentes (Srivastava et al., 2015).

Se han propuesto varias modificaciones a esta arquitectura original de ResNet. Una idea es crear una arquitectura en la que algunas capas puedan desactivarse aleatoriamente durante el tiempo de entrenamiento y compensarse durante el tiempo de prueba bastante similar a los abandonos (Huang et al., 2016). Esta arquitectura de red reemplazará el módulo descartado utilizando una identidad (lo que implica que la parte procesada del residuo simplemente produce un cero).

Si bien es contrario a la intuición colocar capas en el momento del entrenamiento, se debe tener en cuenta que la red en el momento de la prueba sigue siendo una red profunda y en el momento del entrenamiento tiene una profundidad estocástica (la profundidad es aleatoria).

Un resultado de la arquitectura ResNet es que, si bien las redes tradicionales arbitrariamente profundas tienden a sobreajustarse y a veces incluso producen un rendimiento más pobre que una red con menos capas, ResNets parecen ser fieles a la idea de que una mayor profundidad de la red implica un mayor rendimiento.

ResNets superó los récords de rendimiento establecidos por todas las redes discutidas anteriormente y una de sus variantes tiene el récord en el desafío de ImageNet al momento de escribir este libro. La ResNet que ganó la competencia ImageNet tiene hasta 152 capas y también tiene una implementación en el

momento de 1,000 capas para el conjunto de datos CIFAR-10.” (Venkatesan & Li, 2018, págs. 113-114)

Neumonía

“Una enfermedad que infecta a 450 millones de personas en todo el mundo, afecta los pulmones de los infectados, inflamando los alvéolos (sacos de aire) y descargando pus en ellos, causando tos espasmódica con flema además de escalofríos e inhalación y exhalación incómodas.

Los alvéolos afectados con neumonía están ocupados con líquido y pocas células sanguíneas, mientras que los alvéolos sanos están vacíos y pueden contener suficiente aire para llevar a cabo sus funciones normales. Si bien la neumonía puede ser causada por varios tipos de gérmenes, generalmente es causada por bacterias o virus presentes en el aire inhalado.

En la mayoría de los casos, el cuerpo puede combatir estos gérmenes evitando cualquier infección en los pulmones, pero en algunos casos el sistema inmunológico del cuerpo no es lo suficientemente fuerte. La neumonía se puede adquirir de varias maneras, por ejemplo, en hospitales o instituciones de atención médica donde el paciente entra en contacto con la enfermedad cuando visita un hospital por otro motivo, como cuando los pacientes usan ventiladores.

La neumonía adquirida de esta manera es bastante peligrosa ya que las bacterias que la causan pueden ser resistentes a los antibióticos y, además, los efectos de la infección pueden ser más graves ya que la condición del paciente es considerablemente débil. Por otro lado, la neumonía también se puede adquirir fuera de los hospitales / centros de salud. Se dice que esta adquisición es adquirida por la comunidad y es la forma más común de adquirir neumonía.

Si bien la fuente puede ser una bacteria que se puede propagar sin esfuerzo a una persona infectada con una gripe común, ciertos hongos que se encuentran en el suelo o en las heces de las aves también pueden ser la causa de este tipo de neumonía.

Otras fuentes incluyen organismos similares a bacterias y ciertos virus que causan el resfriado común, pero la infección causada por ellos es bastante leve. Además de la comunidad y la atención hospitalaria / de salud, la neumonía también se puede adquirir cuando se inhalan accidentalmente alimentos, líquidos, saliva o vómito en los pulmones.

Aquellos con alguna interrupción en su reflejo nauseoso natural corren el mayor riesgo de este tipo de adquisición. Dado que la neumonía se puede adquirir de varias maneras, puede ser difícil evaluar la fuente de la infección. Existen varias técnicas de diagnóstico administradas para inferir esta información, así como para identificar correctamente la presencia de la infección en sí.” (Syed, 2018, págs. 3-4)

En (Rajaraman, Candemir, Kim, Thoma, & Antani, 2018) los autores señalan acerca de la neumonía y el rol de la radiografía torácica:

“La neumonía es una causa importante de mortalidad en niños de todo el mundo. Según la Organización Mundial de la Salud (OMS), se informan alrededor de 2 millones de muertes relacionadas con la neumonía cada año en niños menores de 5 años, lo que la convierte en la causa más importante de muerte pediátrica. La neumonía proveniente de patógenos bacterianos y virales son las dos causas principales y requieren diferentes formas de manejo. La neumonía bacteriana se trata inmediatamente con antibióticos, mientras que la neumonía viral requiere atención de apoyo, lo que hace que el diagnóstico oportuno y preciso sea importante. El análisis de rayos X del tórax es el examen radiográfico más comúnmente realizado para diagnosticar y diferenciar los tipos de neumonía. Sin embargo, los diagnósticos y tratamientos radiográficos rápidos se ven afectados negativamente por la falta de radiólogos expertos en regiones con recursos limitados donde la neumonía pediátrica es altamente endémica con tasas de mortalidad alarmantes.” (Rajaraman, Candemir, Kim, Thoma, & Antani, 2018, pág. 1)

En (Syed, 2018) la autora señala acerca del proceso de diagnóstico de neumonía y el rol de la radiografía torácica:

“Tradicionalmente, para diagnosticar la neumonía, se realiza un examen físico que incluye escuchar los pulmones del paciente, con un estetoscopio, para detectar sonidos atípicos que indiquen la presencia de neumonía. En caso de sospecha, se administran una serie de pruebas que incluyen análisis de sangre para confirmar la presencia de neumonía e identificar el organismo exacto que causa la neumonía si está presente. También se puede realizar una oximetría de pulso para determinar el nivel de oxígeno presente en la sangre, ya que la neumonía puede evitar que los pulmones muevan gran parte del oxígeno hacia la sangre. Sin embargo, el método principal de diagnóstico es una radiografía de tórax. Con esta radiografía, se puede determinar el alcance y la ubicación de la neumonía. Interpretarlo implica buscar ciertas manchas blancas llamadas infiltrados en la radiografía de los pulmones que muestran la presencia de una infección.” (Syed, 2018, pág. 5)

Función de activación Softmax

“La regresión de Softmax es una función utilizada en la arquitectura del modelo de trabajo del proyecto que inicialmente calcula un puntaje por clase y luego aplica la función de softmax a los puntajes para calcular una probabilidad aproximada para cada clase. La función softmax es la siguiente:

$$\hat{P}_k = \sigma(s(x))_k = \frac{\exp(s_k(x))}{\sum_{j=1}^k \exp(s_j(x))}$$

- k es el número de clases,
- $s(x)$ es el vector que consiste en las puntuaciones correspondientes a cada clase para x
- $\delta(s(x))_k$ es la certeza aproximada de que la instancia x cae en la clase k

El objetivo de estos modelos de aprendizaje supervisado es calcular los parámetros que reducen el valor de la función de costo (o función de pérdida)

con cada iteración. Una función de costo es esencialmente una función que calcula qué tan lejos está la predicción del modelo de los valores verdaderos. La idea es que se calcule una métrica particular sobre los errores de predicción de la red.

Estos errores se promedian en todo el conjunto de datos produciendo un valor que representa la proximidad del modelo al óptimo. En el caso de la regresión logística y softmax, un modelo ideal predice una gran certeza para la clase objetivo y una pequeña certeza para el resto de las clases.” (Syed, 2018, p. 9)

En (Syed, 2018) la autora además señala de la relación entre la clasificación de neumonía, función de pérdida y la función:

“Aunque el diagnóstico de si una persona tiene neumonía o no es bastante distinto, es decir, si una persona tiene o no neumonía, la certeza del modelo de Machine Learning también es bastante interesante, es decir, una persona tiene un 60% de posibilidades de tener neumonía. Por esta razón, se selecciona una función de pérdida de entropía cruzada binaria.

En esta función de pérdida, la salida sería un valor entre 0 y 1 para cada clase que representa la probabilidad de que la imagen pertenezca a esa clase. Por esta razón, lógicamente, la capa final en la red neuronal debe ser una función softmax que, como se explicó, genera el valor tal como se requiere. El valor de esta función de pérdida aumenta a medida que la probabilidad de salida del modelo se aleja de la probabilidad real.” (Syed, 2018, p. 10)

En (Syed, 2018) la autora además señala el propósito de la función de activación Softmax en la última capa de una red neuronal:

“En la capa de salida final en la red neuronal, existe un nodo por clase que el modelo puede generar junto con una función de activación softmax para calcular un número entre cero y uno para cada clase. Este valor generado puede interpretarse como la probabilidad de que la imagen pertenezca a la clase correspondiente.” (Syed, 2018, p. 25)

En (Vázquez Enríquez, 2019) el autor también señala el uso de la función de activación Softmax en la última capa de una red neuronal cuando se debe hacer una múltiple clasificación:

“Capa de salida. Hemos formulado el problema como una tarea de clasificación multiclase, por lo que después de la última capa del modelo convolucional debemos agregar una capa softmax. La capa softmax asignará una probabilidad decimal a cada clase según su relevancia y la suma de las probabilidades de todas las clases sumará 1.

Estas características proporcionadas por la función softmax permiten una convergencia más rápida en este tipo de problemas y proporcionan una distribución de probabilidad válida con valores entre 0 y 1 para cada clase.” (Vázquez Enríquez, 2019, pág. 42)

En (Ranschaert, Morozov, & Algra, 2019) también se describe el uso de la función de activación Softmax en la última capa de una red neuronal para clasificación:

“La capa de salida final es un caso especial de una función de activación, y para eso, a menudo se usan tipos de capa más sofisticados. Si la tarea es la regresión (por ejemplo, estimar la edad del paciente en base a una radiografía de la mano), es apropiado un resultado lineal (por ejemplo, un valor de punto flotante en lugar de un "sí" binario versus "no").

Si la tarea es la clasificación (por ejemplo, la lesión es un cáncer en lugar de benigna), la función softmax a menudo funciona bien. La función softmax tomará un vector de valores (de la capa anterior) y los convertirá en un vector de salida de tamaño arbitrario (el número de clases posibles), y la suma de todos los valores de salida es uno.” (Ranschaert, Morozov, & Algra, 2019, pág. 44)

En (Le, Wang, Carneiro, & Yang, 2019) se menciona como una capa Softmax en una red neuronal convolucional es la que produce una puntuación probabilística de cada clase que se busca predecir:

“La última capa de convolución a menudo es alimentada a capas conectadas localmente que actúan de manera similar a una capa convolucional, pero sin compartir pesos. Finalmente, la clasificación se logra mediante capas de redes neuronales completamente conectadas que a menudo están terminan en una capa softmax que proporciona una puntuación probabilística para cada clase. Intuitivamente, la red convolucional codifica la imagen de entrada con una abstracción creciente a medida que las características se propagan a través de las capas de la red, concluyendo con una elección de clasificación abstracta final.” (Le, Wang, Carneiro, & Yang, 2019, p. 38)

En (Raschka & Mirjalili, 2019) el autor señala el uso de la función Softmax:

“En la sección Construcción de un perceptrón multicapa para clasificar flores en el conjunto de datos de Iris, determinamos la activación = 'softmax' en la última capa del modelo MLP. La función softmax es una forma de la función argmax; en lugar de dar un índice de una sola clase, proporciona la probabilidad de cada clase. Por lo tanto, nos permite calcular probabilidades de clase significativas en entornos multiclase (regresión logística multinomial).” (Raschka & Mirjalili, 2019, pág. 465)

Métricas de Precisión de un Modelo

“Las métricas, si bien dependen de la tarea, debe intentar aclarar qué tipo de errores comete su modelo. Por ejemplo, en las tareas de clasificación, utilice métricas que expliquen los falsos positivos, los falsos negativos y el papel de la prevalencia. Un buen conjunto inicial de métricas a considerar es la sensibilidad (recuerdo), la especificidad y el valor predictivo positivo (precisión). AUC (Área bajo la curva) es una gran métrica general para usar en tareas de clasificación / diagnóstico.” (Ranschaert, Morozov, & Algra, 2019, p. 103)

En (Ranschaert, Morozov, & Algra, 2019) el autor además agrega la importancia de medir la prevalencia de la condición médica:

“Aquí hay un ejemplo simple, donde un modelo de Inteligencia Artificial para detectar cáncer tiene una sensibilidad del 90% y una especificidad del 90%. Si asumimos una prevalencia del 1%, entonces vemos que el modelo generará un falso negativo por 1000 casos y 99 falsos positivos. Esto es muy diferente de la impresión dada por la sensibilidad y la especificidad que, a su valor nominal, sugieren que las tasas de falsos positivos y falsos negativos están equilibradas.

Para reflejar mejor el alto número de falsos positivos que realmente veremos en la práctica, debemos usar una métrica dependiente de la prevalencia, como el valor predictivo positivo (también conocido como precisión).

Mientras que la especificidad es del 90% en el ejemplo anterior, el PPV (valor predictivo positivo) o precisión es solo del 8,3%. Este es un reflejo mucho mejor de la gran cantidad de falsos positivos que el modelo producirá en la práctica. Finalmente, un enfoque común al evaluar el rendimiento de un sistema de Inteligencia Artificial médica en las tareas de diagnóstico es comparar el modelo con el rendimiento de los humanos en la misma tarea." (Ranschaert, Morozov, & Algra, 2019, p. 94)

En la Figura 10 el autor una matriz de confusión más detallada incluyendo la fórmula del Valor Predictivo Positivo:

Matriz de Confusión Valor Predictivo Positivo (Precisión)			
	Case positive	Case negative	
Predicted positive	True positive	False positive	$PPV = \frac{\text{True positives}}{\text{Predicted positives}}$
Predicted negative	False negative	True negative	
	$Sensitivity = \frac{\text{True positives}}{\text{Case positives}}$	$Specificity = \frac{\text{True negatives}}{\text{Case negatives}}$	

Figura 10: Matriz de confusión extendida para incluir el valor predictivo positivo (PPV). (Ranschaert, Morozov, & Algra, 2019, p. 94)

En (Syed, 2018) la autora muestra la fórmula para calcular la precisión, llamada técnicamente Exactitud (accuracy):

“La exactitud (accuracy) de estos modelos se puede calcular como:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

- *TP* es el número de verdaderos positivos,
- *FP* es el número de falsos positivos” (Syed, 2018, p. 51)

2.4. Definición de términos básicos

Machine Learning

Es un marco algorítmico unificado diseñado para identificar modelos computacionales que describen con precisión los datos empíricos y los fenómenos subyacentes, con poca o ninguna participación humana que pueda usarse para enseñar a las computadoras a realizar una amplia gama de tareas útiles, incluida la detección automática de objetos en imágenes.” (Watt, Borhani, & Katsaggelos, 2020, p. 1)

Algoritmo

“Una fórmula o conjunto de reglas (o procedimiento, procesos o instrucciones) para resolver un problema o realizar una tarea. En Inteligencia Artificial, el algoritmo le dice a la máquina cómo encontrar respuestas a una pregunta o soluciones a un problema. En el aprendizaje automático, los sistemas utilizan muchos tipos diferentes de algoritmos. Ejemplos comunes incluyen algoritmos de clasificación.” (Ranschaert, Morozov, & Algra, 2019, p. 348)

Algoritmo de Aprendizaje

“Un algoritmo de aprendizaje es un algoritmo utilizado en Machine Learning para ayudar a la tecnología a imitar el proceso de aprendizaje humano. En combinación con tecnologías como las redes neuronales, los algoritmos de aprendizaje crean programas de aprendizaje sofisticados e involucrados.” (Ranschaert, Morozov, & Algra, 2019, pág. 356)

2.5. Fundamentos teóricos que sustentan las hipótesis

En la Figura 11, se muestra el mapa conceptual que sustentan las hipótesis de la investigación. En nuestro marco teórico tenemos como teoría raíz las redes neuronales convolucionales. De esa teoría se desglosan las tres diferentes redes ResNet-50, Alexnet y VGG, cada una siendo una arquitectura con un algoritmo diferente. Como parte del mismo fundamento teórico se encuentra la función SoftMax que se es incluida en una de las capas de las redes convolucionales. Por otro lado, tenemos nuestra problemática central en cómo mejorar la rapidez en el diagnóstico de neumonía, basada en disminuir el número de diagnósticos incorrectos y disminuir el tiempo de espera de pacientes. Cada componente de nuestro fundamento teórico fue utilizado para resolver la problemática central.

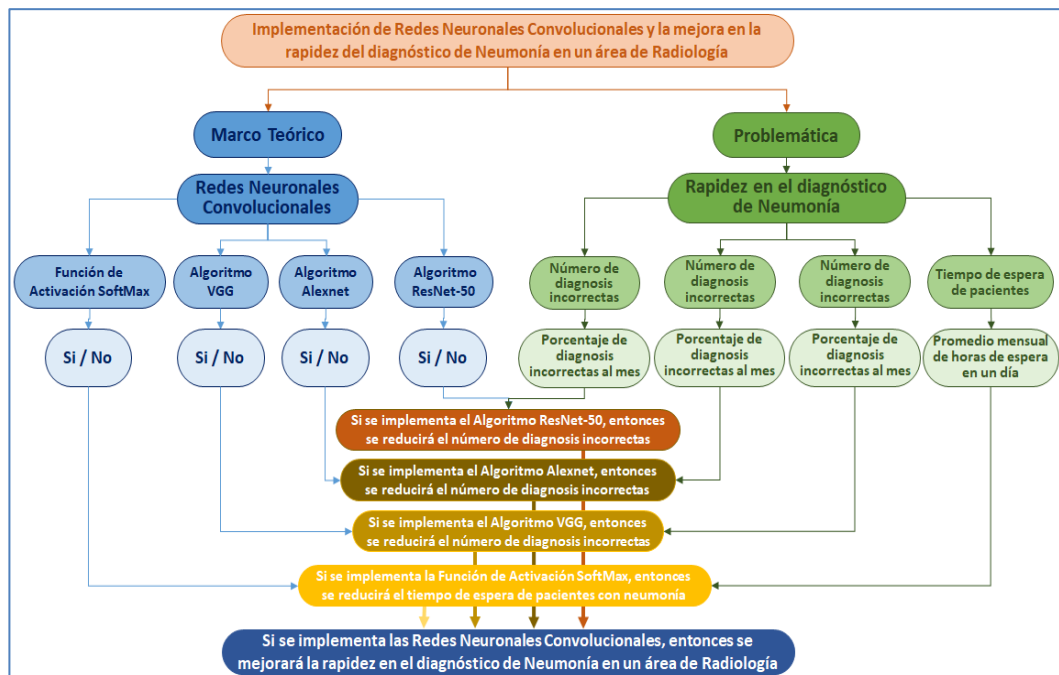


Figura 11: Mapa conceptual que sustentan las hipótesis. Elaboración: propia

2.6. Hipótesis

2.6.1 Hipótesis general

Si se implementa las Redes Neuronales Convolucionales, entonces se mejorará la rapidez en el diagnóstico de Neumonía en un área de Radiología.

2.6.1 Hipótesis específicas

- a) Si se implementa el Algoritmo ResNet-50, entonces se reducirá el número de diagnosis incorrectas.
- b) Si se implementa el Algoritmo Alexnet, entonces se reducirá el número de diagnosis incorrectas.
- c) Si se implementa el Algoritmo VGG, entonces se reducirá el número de diagnosis incorrectas.
- d) Si se implementa la Función de Activación SoftMax, entonces se reducirá el tiempo de espera de pacientes con neumonía.

2.7. Variables

✓ **Independiente**

- ✓ *Redes Neuronales Convolucionales*
- ✓ Algoritmo ResNet-50
- ✓ Algoritmo Alexnet
- ✓ Algoritmo VGG
- ✓ Función de Activación SoftMax

✓ **Dependiente**

- ✓ *Rapidez en el diagnóstico de Neumonía*
- ✓ Número de diagnosis incorrectas
- ✓ Tiempo de espera de pacientes

✓ **Indicadores**

- ✓ Porcentaje de diagnosis incorrectas al mes
- ✓ Promedio mensual de horas de espera en un día

✓ **Matriz de Operacionalización**

Las variables independientes como las variables dependientes i sus indicadores, presentadas anteriormente permitieron trasladar el marco metodológico en un plan de acción, donde se pudo determinar en detalle el método a través del cual cada una de las variables fueron medidas y analizadas.

En el Anexo 4 se muestra la matriz de operacionalización utilizada para el estudio de la investigación.

CAPÍTULO III: MARCO METODOLÓGICO

3.1. Tipo, método y diseño de la investigación

Enfoque

La presente investigación corresponde al enfoque cuantitativo porque se realizó la recolección de datos para probar la hipótesis basándonos en la medición numérica y el análisis estadístico.

El enfoque cuantitativo se define como un conjunto de procesos “secuencial y probatorio, donde cada etapa precede a la siguiente, impidiendo así saltar u omitir pasos.” (Hernández Sampieri, Fernández Collado, & Baptista Lucio, 2014, p. 4)

✓ Tipo de la investigación

La presente investigación es del tipo aplicada porque se hizo uso de diversas herramientas y técnicas para la recolección de datos tales como base de datos, para poder así resolver un problema.

La investigación aplicada es definida como “la utilización de los conocimientos en la práctica, para aplicarlos en provecho de los grupos que participan en esos procesos y en la sociedad en general, además del bagaje de nuevos conocimientos que enriquecen la disciplina.” (Martínez, 2004)

Debido a que se trata de un enfoque cuantitativo y aplicado, la implementación fue de manera inmediata, no estando orientada solo al desarrollo de teorías. “La investigación explicativa pretende establecer las causas de los eventos, sucesos, o fenómenos que se estudian” (Hernández Sampieri, Fernández Collado, & Baptista Lucio, 2014, p. 95)

✓ **Método de la investigación**

El tipo de método elegido para esta investigación es científico, ya que se realizó una experimentación para la comprobación de las hipótesis planteadas.

La investigación explicativa es definida de la siguiente manera: “la investigación es explicativa o de comprobación de hipótesis causales dado que se profundizará en la causa que está originando el evento, en otras palabras, no solo se buscará dar solución al problema, sino también a dar explicación a sus causas.” (Sanchez Carlessi & Reyes Meza, 2015)

El presente estudio busca descartar las posibles asociaciones aleatorias y determinar la relación de causa y efecto entre las variables de tipo dependiente e independiente.

✓ **Diseño de la investigación**

Para dar una respuesta a las preguntas que se plantea en la investigación, lograr los objetivos de estudio y someter a las hipótesis formuladas a prueba, se empleó el diseño de investigación cuasi experimental.

“Este diseño consiste en aplicar a un grupo una prueba previa al estímulo o tratamiento experimental, para luego administrar el estímulo y luego realizar la medición.” (Hernández Sampieri, Fernández Collado, & Baptista Lucio, 2014)

“Los diseños cuasi-experimentales tienen el mismo propósito que los estudios experimentales: probar la existencia de una relación causal entre dos o más variables. Cuando la asignación aleatoria es imposible, los cuasi-experimentales (semejantes a los experimentos) permiten estimar los impactos del tratamiento o programa, dependiendo de si llega a establecer una base de comparación apropiada” (Hedrick, Bickman, & Rog, 1993)

Para el diseño de la muestra será según No Probabilístico, debido a que la elección de los elementos no depende de la probabilidad, sino de causas relacionadas con las características de la investigación o de quien hace la muestra.

Un diseño experimental en su modalidad cuasi-experimental tiene los siguientes argumentos:

- Las variables dependientes (problemas) de la investigación son cuantitativas
- Se aplicó el estímulo intencionalmente la variable independiente (teórica)
- El nivel o método es explicativo, o lo que es lo mismo de causalidad
- No se usó grupo de control
- El diseño muestral es no aleatorio

Para responder a las preguntas de investigación y lograr los objetivos del estudio se sometieron las cuatro (04) hipótesis formuladas a pruebas, usando un diseño cuasi experimental, se trabajó con las variables independientes (algoritmo ResNet-50, algoritmo Alexnet, Algoritmo VGG y Función de Activación SoftMax) para observar el efecto en los resultados de las variables dependientes (Número de diagnósticos incorrectos y Tiempo de espera de pacientes).

Este diseño permitió la toma de mediciones antes y después de la implementación del nuevo modelo considerando un lapso de tiempo.

“Este diseño consiste en aplicar a un grupo una prueba previa al estímulo o tratamiento experimental, para luego administrar el estímulo y luego realizar la medición.” (Hernández Sampieri, Fernández Collado, & Baptista Lucio, 2014)

A continuación, detallaremos el diseño que se utilizara para cada una de las hipótesis, ver Tabla 07, Tabla 08, y también la Figura 12.

Tabla 07:

Diseño de Investigación Cuasi-Experimental

Nombre	Esquema
Diseño de un grupo de medición antes y después	G O ₁₁ O ₁₂ O ₁₃ ... X O ₂₂ O ₂₃ O ₂₄ ...

Fuente: (Hernández Sampieri, Fernández Collado, & Baptista Lucio, 2014)

Donde: G: Grupo de análisis o variable experimental.

O_{1n}; Mediciones Pre test de la variable dependiente.

O_{2m}: Mediciones Post test de la variable dependiente.

X: Tratamiento o estímulo.

Tabla 08:

Diseño Cuasi-Experimental

Objetivo Especifico	Hipótesis Especifica	Variable Dependiente	Indicador
Implementar el Algoritmo ResNet-50 para reducir el número de diagnósis incorrectas.	Si se implementa el Algoritmo ResNet-50, entonces se reducirá el número de diagnósis incorrectas.	Número de diagnósis incorrectas	Porcentaje de diagnósis incorrectas al mes
Implementar el Algoritmo Alexnet para reducir el número de diagnósis incorrectas.	Si se implementa el Algoritmo Alexnet, entonces se reducirá el número de diagnósis incorrectas.	Número de diagnósis incorrectas	Porcentaje de diagnósis incorrectas al mes
Implementar el Algoritmo VGG para reducir el número de diagnósis incorrectas.	Si se implementa el Algoritmo VGG, entonces se reducirá el número de diagnósis incorrectas.	Número de diagnósis incorrectas	Porcentaje de diagnósis incorrectas al mes
Implementar la Función de Activación SoftMax, para reducir el tiempo de espera de pacientes con neumonía.	Si se implementa la Función de Activación SoftMax, entonces se reducirá el tiempo de espera de pacientes con neumonía.	Tiempo de espera de pacientes	Promedio mensual de horas de espera en un día

Fuente: Elaboración Propia.

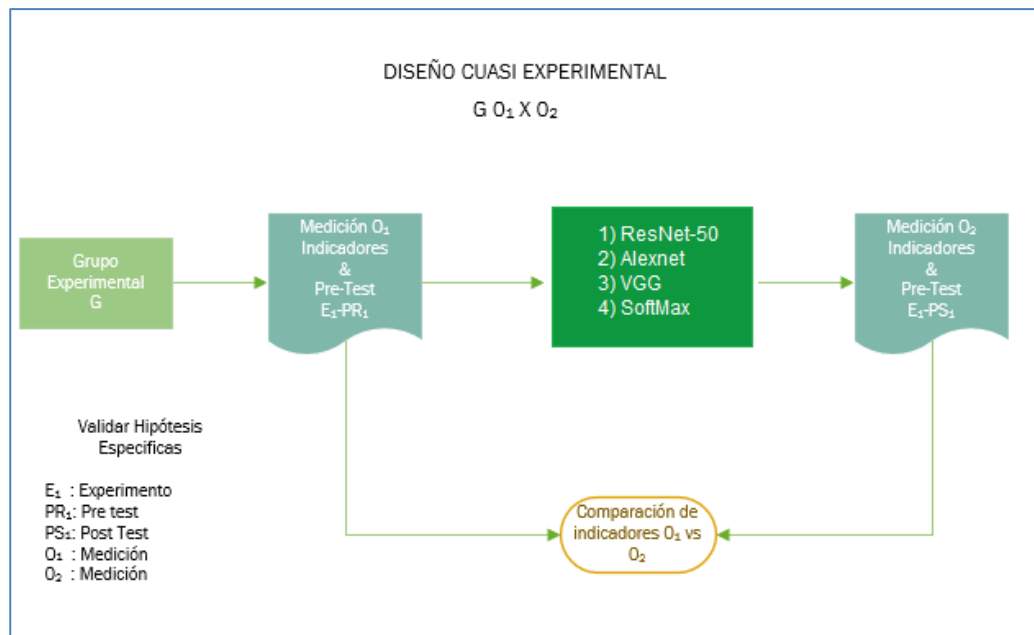


Figura 12: Diseño Cuasi experimental. Elaboración Propia.

3.2. Población y muestra

- **Población General**

De acuerdo a (Hernández Sampieri, Fernández Collado, & Baptista Lucio, 2014) la población general es definida de la siguiente manera:

“Una población es el conjunto de todos los casos que concuerdan con una serie de especificaciones. Es la totalidad del fenómeno a estudiar, donde las entidades de la población poseen una característica común la cual se estudia y da origen a los datos de la investigación.” (Hernández Sampieri, Fernández Collado, & Baptista Lucio, 2014)

Para la presente investigación se tuvo una población de estudio finita y abarca 5232 placas radiográficas torácicas.

La población es la misma para todas las variables que se trabaje durante la ejecución de la investigación. Más adelante se detallará la población y muestra por cada una de las variables.

- **Muestra General**

De acuerdo a (Hernández Sampieri, Fernández Collado, & Baptista Lucio, 2014) la muestra general es definida de la siguiente manera:

“La muestra es un subgrupo de la población, un subconjunto de elementos que pertenecen a ese conjunto definido en sus características al que llamamos población.” (Hernández Sampieri, Fernández Collado, & Baptista Lucio, 2014)

Para la presente investigación se tuvo una muestra del tipo probabilístico. Esta cantidad de placas radiográficas a revisar de la base de datos se ha obtenido a partir de la fórmula de acuerdo a (Ñaupas, Mejía, Novoa, & Villagómez, 2014):

La fórmula es:

$$n = \frac{N * z^2 * p * q}{(N - 1) * e^2 + z^2 * p * q}$$

Donde: n: muestra

z: nivel de confianza: 95%

p: probabilidad de éxito: 50%

q: probabilidad de fracaso: 50%

e: nivel de error, 5.0%

N: población “ (Ñaupas, Mejía, Novoa, & Villagómez, 2014)

A continuación, se presenta la población y la muestra que se emplearon por cada una de las Variables Dependientes planteadas en esta investigación.

Para las cuatro (04) variables de la investigación tenemos:

- Número de diagnósis incorrectas – Porcentaje de diagnósis incorrectas al mes usando Algoritmo ResNet-50
- Número de diagnósis incorrectas – Porcentaje de diagnósis incorrectas al mes usando el Algoritmo Alexnet
- Número de diagnósis incorrectas – Porcentaje de diagnósis incorrectas al mes usando el Algoritmo VGG
- Tiempo de espera de pacientes – Promedio mensual de horas de espera en un día

✓ Población

La población para la variable dependiente tiempo en la transacción tendremos una población Pre y Post respectivamente para esta investigación.

- **Pre:** Para el presente estudio se tomaron 12 meses del año 2017 que está representado por 5232 placas radiográficas (Estudio pre-test).
- **Post:** Para el presente estudio se tomaron 12 meses del año 2020 que está representado por 5232 placas radiográficas (Estudio post-test).

✓ **Muestra**

Tanto las muestras Pre y Post de la investigación serían igual a la población.

En la Tabla 09 se muestran las poblaciones y las muestras en una situación PRE Test y POST Test.

Tabla 09:
Población y Muestra PRE y POST por cada una de las variables

Variable Dependiente	Indicador	Población PRE	Muestra PRE	Población POST	Muestra POST
Número de diagnosing incorrectas	Porcentaje de diagnosing incorrectas al mes	5,232 placas radiográficas 2017	5,232 placas radiográficas 2017	5,232 placas radiográficas 2020	5,232 placas radiográficas 2020
Número de diagnosing incorrectas	Porcentaje de diagnosing incorrectas al mes	5,232 placas radiográficas 2017	5,232 placas radiográficas 2017	5,232 placas radiográficas 2020	5,232 placas radiográficas 2020
Número de diagnosing incorrectas	Porcentaje de diagnosing incorrectas al mes	5,232 placas radiográficas 2017	5,232 placas radiográficas 2017	5,232 placas radiográficas 2020	5,232 placas radiográficas 2020
Tiempo de espera de pacientes	Promedio mensual de horas de espera en un día	5,232 placas radiográficas 2017	5,232 placas radiográficas 2017	5,232 placas radiográficas 2020	5,232 placas radiográficas 2020

Fuente: Elaboración propia

3.3. Técnicas e instrumentos de recolección de datos

La técnica utilizada es el conjunto de reglas y de procedimientos que permitió al investigador a establecer la relación con el objeto o sujeto de la investigación. El instrumento es el mecanismo que se utilizó para recolectar y registrar la información. (Pino Gotuzzo, 2013)

Es la etapa donde se buscó elegir el instrumento que sea confiable y que nos garantizara que los datos que se fueron recabando han permitido interpretar la realidad. Esto suponía que la muestra de estudio comprendía las preguntas que se les formulo. (Pino Gotuzzo, 2013)

Para las cuatro (04) variables de la investigación tenemos:

- Número de diagnósticos incorrectos –Porcentaje de diagnósticos incorrectos al mes usando Algoritmo ResNet-50
- Número de diagnósticos incorrectos –Porcentaje de diagnósticos incorrectos al mes usando el Algoritmo Alexnet
- Número de diagnósticos incorrectos –Porcentaje de diagnósticos incorrectos al mes usando el Algoritmo VGG
- Tiempo de espera de pacientes –Promedio mensual de horas de espera en un día

a. Técnicas e instrumentos

✓ Técnicas

Base de datos: Esta se realizó con la ayuda de elementos técnicos, SAP Hana DB y el Sistema de información denominado: Sistema de Repositorio Mendeley Data.

✓ Instrumentos

Se utilizó el software gestor de base de datos SAP Hana DB en su versión SAP HANA 1.0 SP11.

b. Criterio de validez del instrumento

La validez del instrumento DBMS SAP Hana DB en su versión SAP HANA 1.0 SP11, está dado por ser el líder en los softwares de gestión de base de datos, como se puede apreciar en la Figura 13 extraída del reporte de evaluación de Gartner de Julio del 2020.



Figura 13: Magic Quadrant for Operational Database Management Services. Deloitte Cyber Strategy (2020)

c. Criterio de confiabilidad de instrumento

SAP HANA DB se aprovecha del bajo coste de la memoria principal (RAM), la capacidad del procesamiento de datos de los procesadores multi-núcleo y el acceso rápido a datos de unidades de estado sólido con respecto a los discos duros

tradicionales para ofrecer un mejor rendimiento de las aplicaciones analíticas y transaccionales.

Ofrece un entorno de consulta multi-motor de procesamiento que le permite soportar tanto datos relacionales (con tanto en fila y columna orientado a representaciones físicas en un motor híbrido) así como el tratamiento gráfico y de texto para la gestión de datos no estructurados y dentro del mismo sistema. HANA DB es 100% compatible con ACID.

En la Tabla 10 se muestran las técnicas empleadas en el presente estudio; así como, los instrumentos utilizados para cada una de ellas.

Tabla 10:
Técnicas e instrumentos

Variable Dependiente	Indicador	Técnica	Instrumento
Número de diagnosing incorrectas	Porcentaje de diagnosing incorrectas al mes	Base de Datos	Sistema de Información y Base de datos
Número de diagnosing incorrectas	Porcentaje de diagnosing incorrectas al mes	Base de Datos	Sistema de Información y Base de datos
Número de diagnosing incorrectas	Porcentaje de diagnosing incorrectas al mes	Base de Datos	Sistema de Información y Base de datos
Tiempo de espera de pacientes	Promedio mensual de horas de espera en un día	Base de Datos	Sistema de Información y Base de datos

Fuente: Elaboración propia

3.4. Descripción de procedimientos de análisis

Con las variables y sus indicadores ya establecidos anteriormente, se permite medir, analizar y verificar los datos, y así obtener la información suficiente y necesaria para el análisis de los resultados de la investigación. Para ello se desarrolló la matriz de análisis de datos que se muestra a continuación (Ver Tabla 11).

Tabla 11:
Matriz de Análisis de datos

Variable Dependiente	Indicador	Escala de medición	Estadísticos descriptivos	Análisis inferencial
Número de diagnosis incorrectas	Porcentaje de diagnosis incorrectas al mes	Razón	Media Mediana Varianza	t de Student de muestras emparejadas
Número de diagnosis incorrectas	Porcentaje de diagnosis incorrectas al mes	Razón	Media Mediana Varianza	t de Student de muestras emparejadas
Número de diagnosis incorrectas	Porcentaje de diagnosis incorrectas al mes	Razón	Media Mediana Varianza	t de Student de muestras emparejadas
Tiempo de espera de pacientes	Promedio mensual de horas de espera en un día	Razón	Media Mediana Varianza	t de Student de muestras emparejadas

Fuente: Elaboración propia

Capítulo IV: RESULTADOS Y ANÁLISIS DE RESULTADOS

4.1. Resultados

En el presente capítulo se presentan los resultados obtenidos en el PRE Test, la aplicación de la teoría, los resultados obtenidos en el POST Test para cada objetivo planteado en la presente investigación.

✓ Generalidades

La presente investigación está enfocada en un área de radiología cuya misión es la rápida detección de neumonía usando especialistas que visualmente examinan placas radiográficas torácicas para detectar patrones que podrían revelar si el paciente tiene neumonía.

Anteriormente se había implementado un sistema programado para la automatización del proceso y encontrar patrones en las imágenes, sin embargo, los diagnósticos fueron de baja precisión y alto tiempo del proceso.

Para poder mejorar la problemática se ha propuesto un modelo de redes neuronales convolucionales con la finalidad de incrementar la precisión del diagnóstico y reducir el tiempo de diagnóstico y por lo tanto permitir reducir la tasa de mortalidad de pacientes al permitir un diagnóstico temprano y un tratamiento más preciso.

En la Figura 14 podemos observar las causas para el problema general de la investigación. Se puede

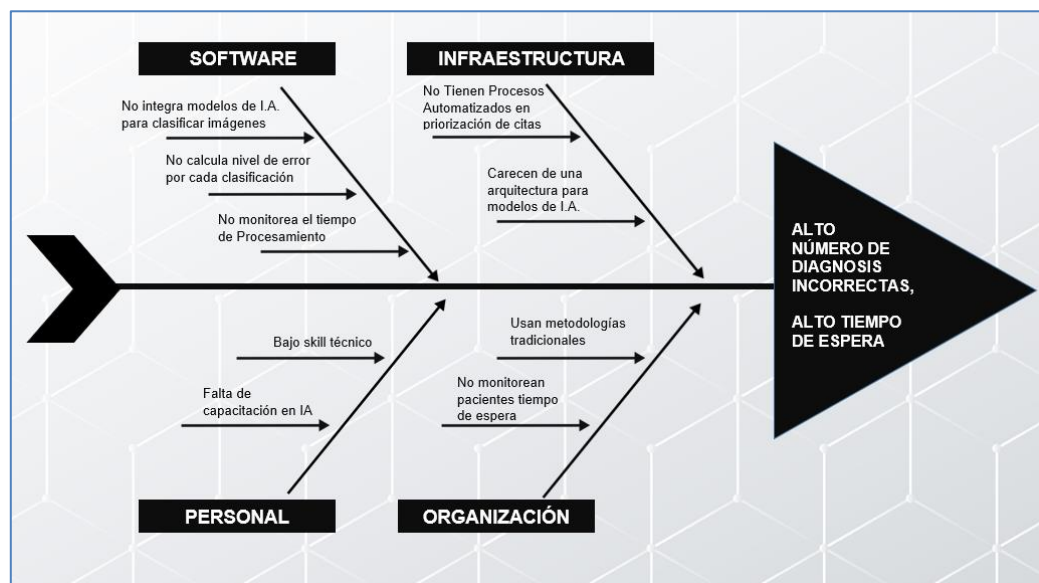


Figura 14: Diagrama de Ishikawa para el problema general. Elaboración propia.

Se puede ver en la Figura 14 que se identificaron las causas primordiales de las variables dependientes (alto número de diagnosis incorrectas y alto tiempo de espera de pacientes de neumonía) mediante el diagrama de Ishikawa y se priorizaron las causas raíces mediante la ponderación de hallazgos que fueron recolectados de la base de datos del área de radiología de enero a diciembre del 2017. Se puede ver que el sistema previamente utilizado carece de un robusto motor cognitivo (modelos de I.A.) que pueda predecir la diagnosis con suficiente alta exactitud. Y eso impide a que la organización misma no pueda comenzar a implementar un sistema de I.A. debido a que carecen del skill técnico para diseñar un sistema de ese tipo.

- ✓ **Algoritmo ResNet-50:** Aplicar el algoritmo ResNet-50 para reducir el número de diagnósicos incorrectas en el área de Radiología.

- **Situación PRE Test**

El proceso de diagnóstico de neumonía en el área de radiología se ha venido realizando primero de manera tradicional, es decir, solamente con la revisión visual de placas radiográficas del área torácica de los pacientes por parte de un especialista radiólogo, que, basado en el nivel de experiencia y conocimientos, puede inferir si la placa analizada contiene patrones asociados con la enfermedad de neumonía.

En el esfuerzo de automatizar este proceso, el área de radiología adoptó un sistema programado para reconocimiento de objetos en imágenes. Sin embargo, para evaluar los resultados se compararon los diagnósicos hechos por los especialistas versus los diagnósicos hechos por el sistema.

Utilizando la métrica de exactitud, se calculó el promedio de precisión del sistema en un periodo de 1 año.

En la Figura 15 se detalla el porcentaje de diagnósicos incorrectas hechas por este sistema automatizado del área de radiología por mes durante el 2017.

Aunque las cifras aproximan en total anual de un 25% de diagnósicos incorrectas, esa es una cifra demasiado alta debido a la criticidad de que un diagnóstico de neumonía sea altamente preciso.

A continuación, en la Tabla 12 se puede apreciar el detalle de los 12 registros (correspondientes a los 12 meses del año 2017) de la muestra PRE Test antes de la implementación de un sistema de inteligencia artificial basado en redes neuronales convolucionales basado en redes neuronales convolucionales basadas en el algoritmo ResNet-50.

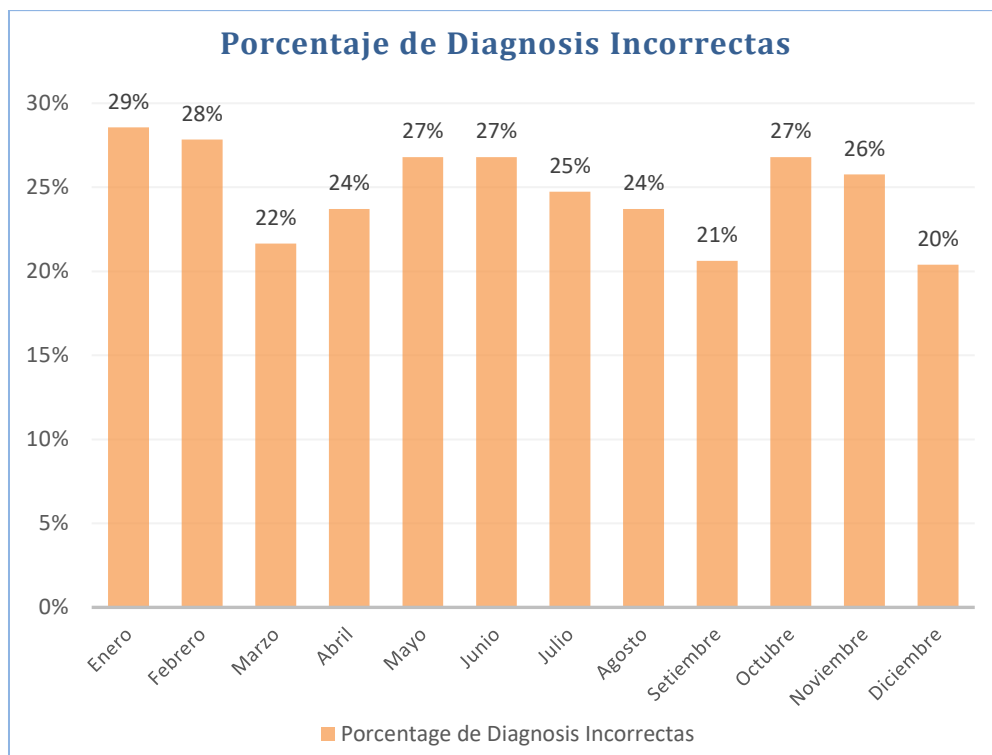


Figura 15: Porcentaje de diagnóstico incorrectas en diagnóstico de neumonía automatizado. Elaboración propia.

Tabla 12:

Muestra Pre Test Porcentaje de Diagnóstico Incorrectas

Mes	Muestra Pre Test
Enero	0.2857
Febrero	0.2784
Marzo	0.2165
Abril	0.2371
Mayo	0.2680
Junio	0.2680
Julio	0.2474
Agosto	0.2371
Setiembre	0.2062
Octubre	0.2680
Noviembre	0.2577
Diciembre	0.2039
TOTAL	0.2478

Fuente: SAP Hana DB. Área de Radiología. Elaboración propia.

▪ **Aplicación de la teoría**

Para el presente estudio, frente a la problemática que se observaba debido a los bajos niveles de precisión en los diagnósticos automáticos producidos, se ha optado por implementar un sistema basado en redes neuronales convolucionales cuya arquitectura es un algoritmo de redes residuales llamado ResNet-50.

ResNet-50 es, en el campo de la inteligencia artificial y visión artificial, una arquitectura de redes neuronales de 50 capas basada en redes residuales, y además es la arquitectura ganadora del concurso ImageNet en el 2015.

En el campo de clasificación de imágenes, para poder encontrar patrones complejos y mejorar la precisión de las predicciones la arquitectura debe ser muy profunda (Chakraborty, Shaw, & Aich, 2018).

Las redes más profundas pueden crear representaciones profundas porque en cada capa la red aprende una representación nueva y más abstracta. Sin embargo, ante mayor profundidad se ocasionan otros problemas, de acuerdo a (Shanmugamani, 2018)

ResNet-50 es un algoritmo que demostró que se pueden entrenar redes más profundas. Cuanto más profunda es la red, más saturada se vuelve la precisión. Ni siquiera se debe a un sobreajuste ni a la presencia de un número elevado de parámetros, sino a una reducción del error de entrenamiento.

Esto se debe a la incapacidad de propagar hacia atrás los gradientes. Esto se puede superar enviando los gradientes directamente a las capas más profundas con un bloque residual. Cada dos capas se conectan formando un bloque residual.

Puede ver que el entrenamiento se pasa entre las capas. Mediante esta técnica, la propagación hacia atrás puede llevar el error a capas anteriores. Este bloque residual se puede ver en Figura 16.

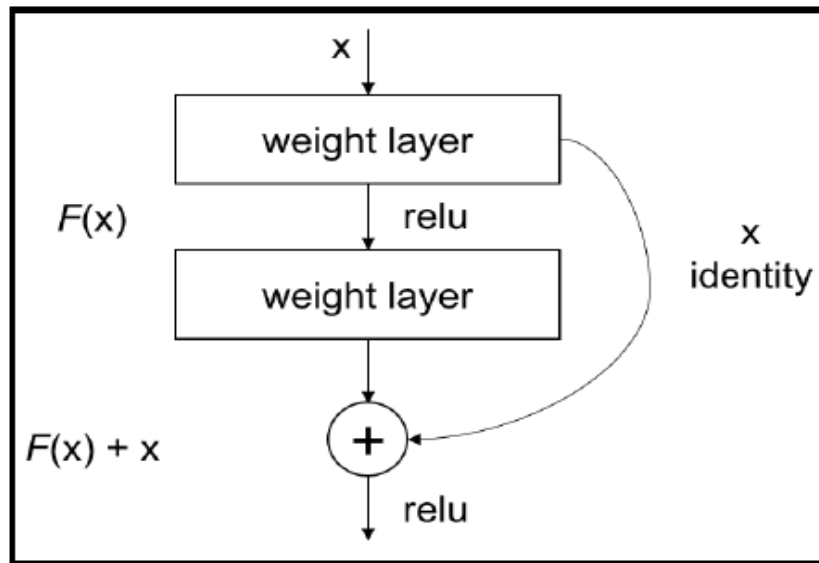


Figura 16: Bloque residual. Conexión de acceso directo. (Shanmugamani, 2018)

Existen muchas maneras de implementar un modelo basado en redes residuales ResNet-50, en el presente estudio se utilizó un framework de Deep Learning llamado PyTorch (Chollet, 2016) que permite la creación y programación de la red neuronal y además permite configurarla en más detalle.

La primera fase fue la programación del preprocesamiento de las imágenes para que puedan ser ingresadas en el modelo de redes neuronales. En PyTorch el preprocesamiento consiste en transformaciones de los datos.

La Figura 17 muestra las funciones de transformación realizadas. Las transformaciones tienen el propósito de estandarizar el tamaño de las imágenes, centrar las imágenes, y luego redimensionarlas con rotaciones aleatorias (ver Figura 18), esto permite que el modelo pueda ser generalizable, es decir, que el modelo simplemente no memorice las imágenes.

Las imágenes luego son transformadas a matrices de 3 dimensiones con los valores de cada pixel, estas matrices son llamadas Tensores (Aja-Fernandez, Garcia, Tao, & Li, 2009). Estos tensores son normalizados para mejorar el desempeño del modelo de redes neuronales.

```
transformers = {'train_transforms' : transforms.Compose([
    transforms.Resize(224),
    transforms.CenterCrop(224),
    transforms.RandomRotation(20),
    transforms.RandomHorizontalFlip(),
    transforms.ToTensor(),
    transforms.Normalize([0.5,0.5,0.5],[0.5,0.5,0.5])
]),
'test_transforms' : transforms.Compose([
    transforms.Resize(224),
    transforms.CenterCrop(224),
    transforms.ToTensor(),
    transforms.Normalize([0.5,0.5,0.5],[0.5,0.5,0.5])
]),
'valid_transforms' : transforms.Compose([
    transforms.Resize(224),
    transforms.CenterCrop(224),
    transforms.ToTensor(),
    transforms.Normalize([0.5,0.5,0.5],[0.5,0.5,0.5])
])}
```

Figura 17: Transformaciones en PyTorch requeridas en preprocesamiento de imágenes. Elaboración propia.

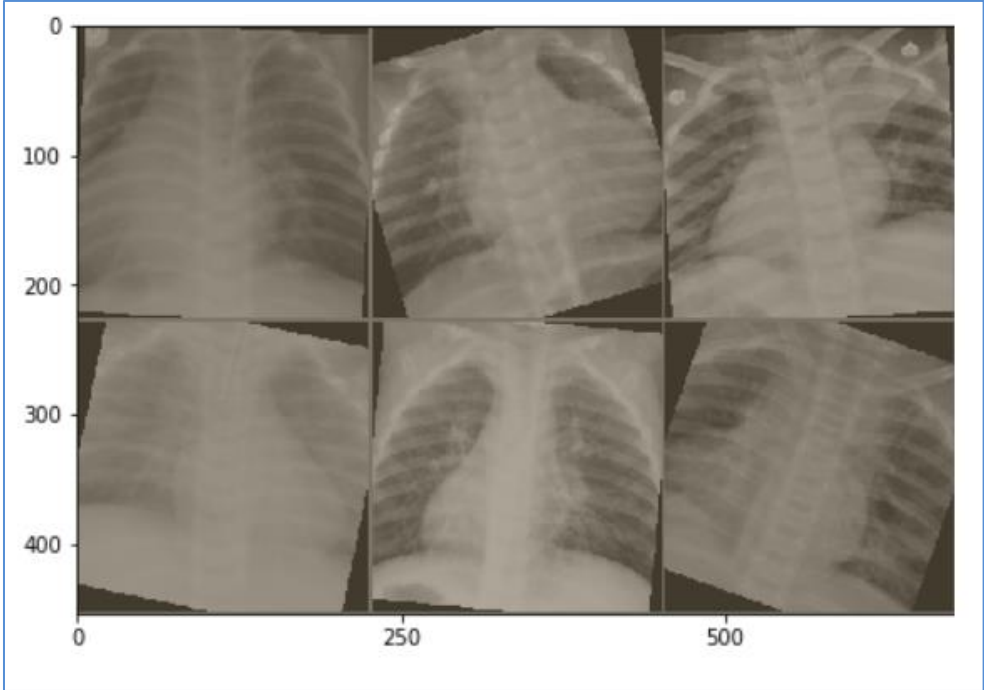


Figura 18: Muestra de imágenes redimensionadas en preprocesamiento. Elaboración propia.

La segunda fase fue la creación de la red neuronal usando la biblioteca Torchvision lo cual permitirá tener acceso al modelo ResNet-50 (ver Figura 19), de esta manera estamos usando la técnica llamada Transfer Learning o Transferencia de Aprendizaje.

```
class Model(nn.Module):
    def __init__(self):
        super(Model, self).__init__()

        # Transfer Learning
        # leveraging the feature extractor of ResNet-152
        # #obtain the ResNet model from torchvision.model library
        self.model = torchvision.models.resnet50(pretrained=True)

        # Building classifier and since we are classifying the images
        # into NORMAL and PNEMONIA, we output a two-dimensional tensor.
        self.classifier = nn.Sequential(
            nn.Linear(self.model.fc.in_features,2),

        )

        # Pytorch provides us with the ability to take and freeze
        # these powerful feature extractors, attach our own classifiers
        # depending on our problem domain and train the resulting model
        # Requires_grad = False
        # denies the ResNet model the ability
        # to update its parameters hence make it unable to train.
        for params in self.model.parameters():
            params.requires_grad = False
        self.model.fc = self.classifier

        # Every model built from the nn.Module requires
        # that we override the forward function
        def forward(self, x):
            return self.model(x)
```

Figura 19: Red neuronal convolucional con Transfer Learning de ResNet-50. Elaboración propia.

Mediante esta técnica, la red convolucional pre-entrenada transfiere su capacidad de clasificación de imágenes a la red neuronal que hemos creado. Este método es especialmente útil de detección de anomalías en imágenes médicas debido a que a las limitaciones de acceso de datos médicos.

De acuerdo a (Yang, Zhang, Dai, & Pan, 2020) las imágenes médicas tienen las siguientes características específicas:

- Datos pequeños y etiquetado caro: Los datos de imágenes médicas se recopilan a través de equipos especiales en contextos muy privados, por lo

que los datos de imágenes médicas a menudo solo tienen una muestra pequeña, medida en el orden de cientos de muestras solamente. El etiquetado de imágenes médicas a menudo se basa en expertos humanos experimentados y bien capacitados, como médicos y radiólogos, lo que hace que el etiquetado de imágenes médicas sea mucho más costoso.

- Datos complejos: Las tareas de visión por computadora generalmente se enfocan en imágenes o videos bidimensionales (2D). Sin embargo, las imágenes médicas tienen una formación de datos mucho más compleja. Las imágenes de rayos X a menudo consisten en varias vistas de un área del cuerpo de un paciente.

Para la fase de despliegue se utilizó la plataforma como servicio (PaaS) llamada Render (Goel, 2019) que es una plataforma unificada para desplegar automáticamente todas las aplicaciones web desde Git.

Como se puede visualizar en el diagrama del sistema en la Figura 20 el flujo del proceso de clasificación integra el modelo ResNet-50 creado en PyTorch con FastAI, guardando las predicciones en la base de datos SQL de Render.

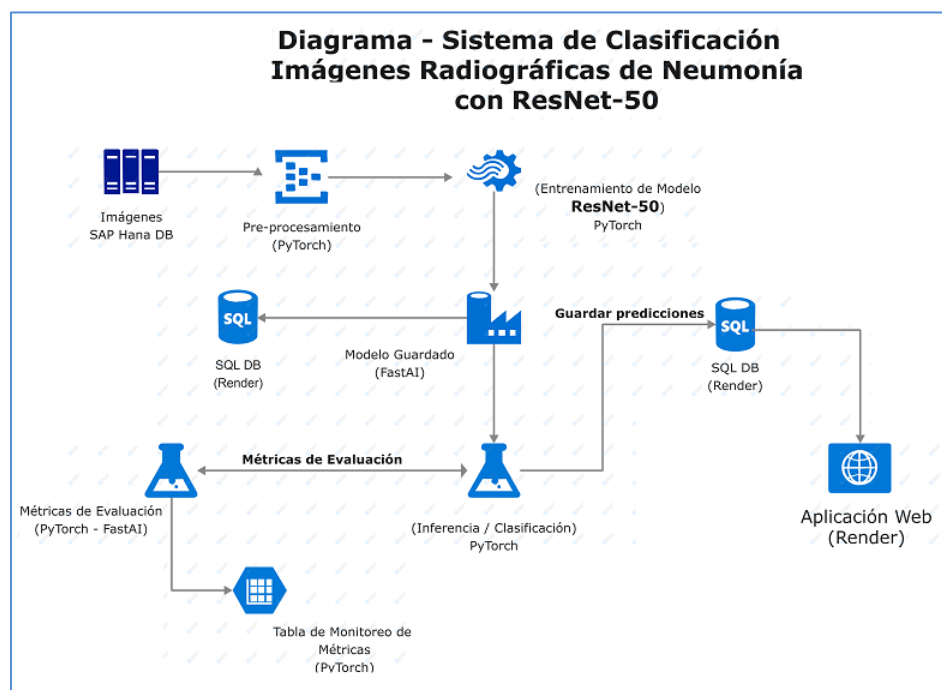


Figura 20: Diagrama del sistema usando ResNet-50. Elaboración propia.

Durante el proceso de entrenamiento podemos observar como el algoritmo ResNet-50 encuentra los patrones (segmentación) en la imagen, estos son grupos de pixeles encontrados que ayudaran a clasificar si la imagen contiene patrones consistentes con neumonía o no, ver Figura 21.

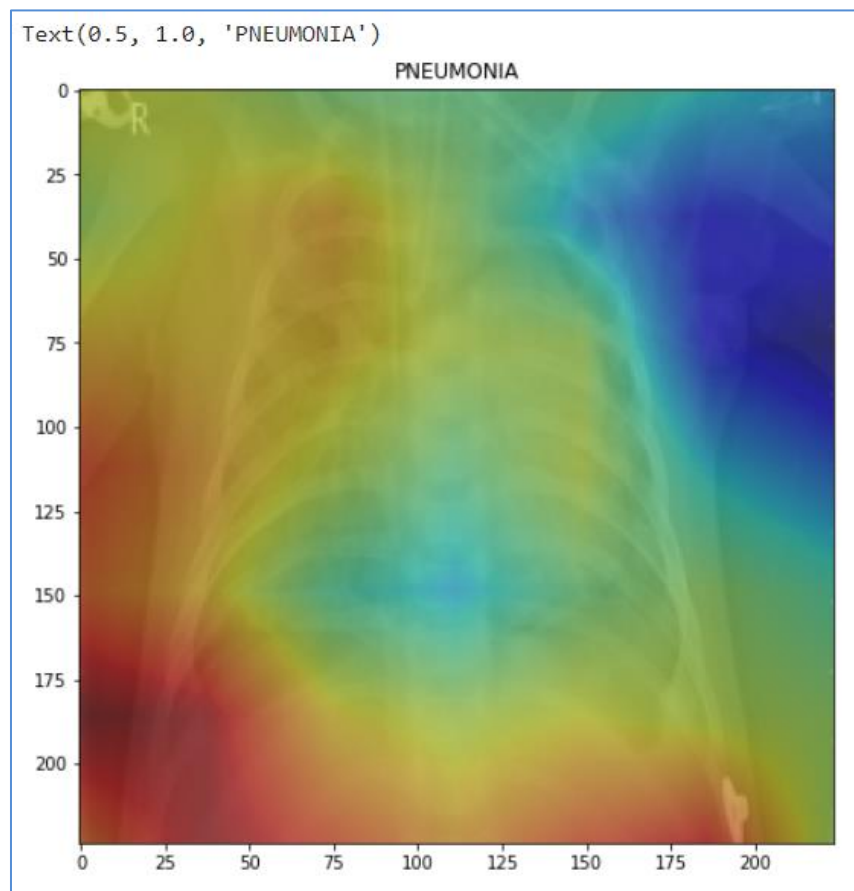


Figura 21: Ejemplo de patrones encontrados en una imagen usando ResNet-50 durante el entrenamiento de una detección de Neumonía. Elaboración propia.

Durante la fase de evaluación de métricas y monitoreo, una vez implementado el sistema de clasificación de imágenes basado en ResNet-50 (ver Figura 22) se pudo observar que el modelo obtuvo una alta exactitud (accuracy) de 95% en tan solo 7 minutos después de haber iniciado el entrenamiento.

En la misma figura se puede apreciar las métricas de evaluación y curva de aprendizaje del modelo ResNet-50, después de tres épocas de entrenamiento.

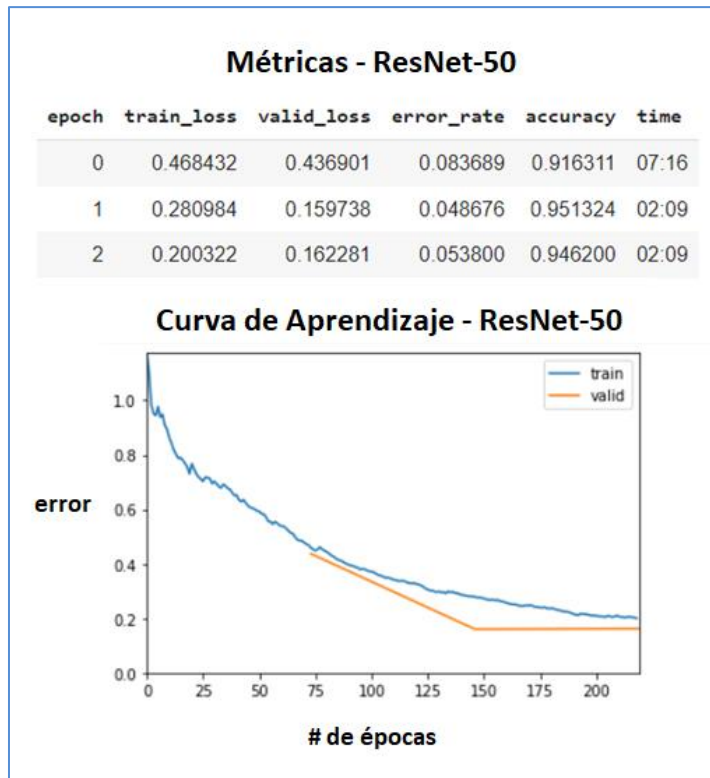


Figura 22: Lista de métricas y Curva de Aprendizaje del modelo basado en ResNet-50. Elaboración propia.

Otras métricas que nuestro sistema puede mostrar son:

- la cantidad de error de entrenamiento (train_loss), el cual se refiere al valor de error final alcanzado en cada época (iteración) de entrenamiento
- la cantidad de error de validación (valid_loss), el cual refiere al valor de error final alcanzado en cada época (iteración) de evaluación.

Una manera de medir el desempeño de un modelo de inteligencia artificial es observando ambas métricas train_loss y valid_loss cuyos valores deben ser similares, de no ser así modelaría estaría incurriendo en lo que es llamado Sobreajust (overfitting) donde el modelo no aprende sino solo memoriza de los datos. En este caso los valores son similares, con train_loss de 0.2 y valid_loss de 0.16, demostrando que el modelo generaliza (aprende) correctamente.

Como parte de las métricas de monitoreo se encuentra curva de aprendizaje, un modelo de inteligencia artificial con buen desempeño muestra una curva

descendente similar a la de Figura 22 con ambas proporciones (train_loss y valid_loss) descendiendo de manera similar por cada época (iteración).

- **Situación POST Test**

Después de la implementación del nuevo sistema de clasificación de imágenes, el área de radiología registró una alta cantidad de diagnósticos automatizados correctamente clasificados, lo cual fue corroborado por 10 médicos especialistas.

Podemos ver en la Figura 23 la matriz de confusión mostrando específicamente cuantas radiografías fueron diagnosticadas por el sistema correctamente para cada clase: Neumonía y Normal.

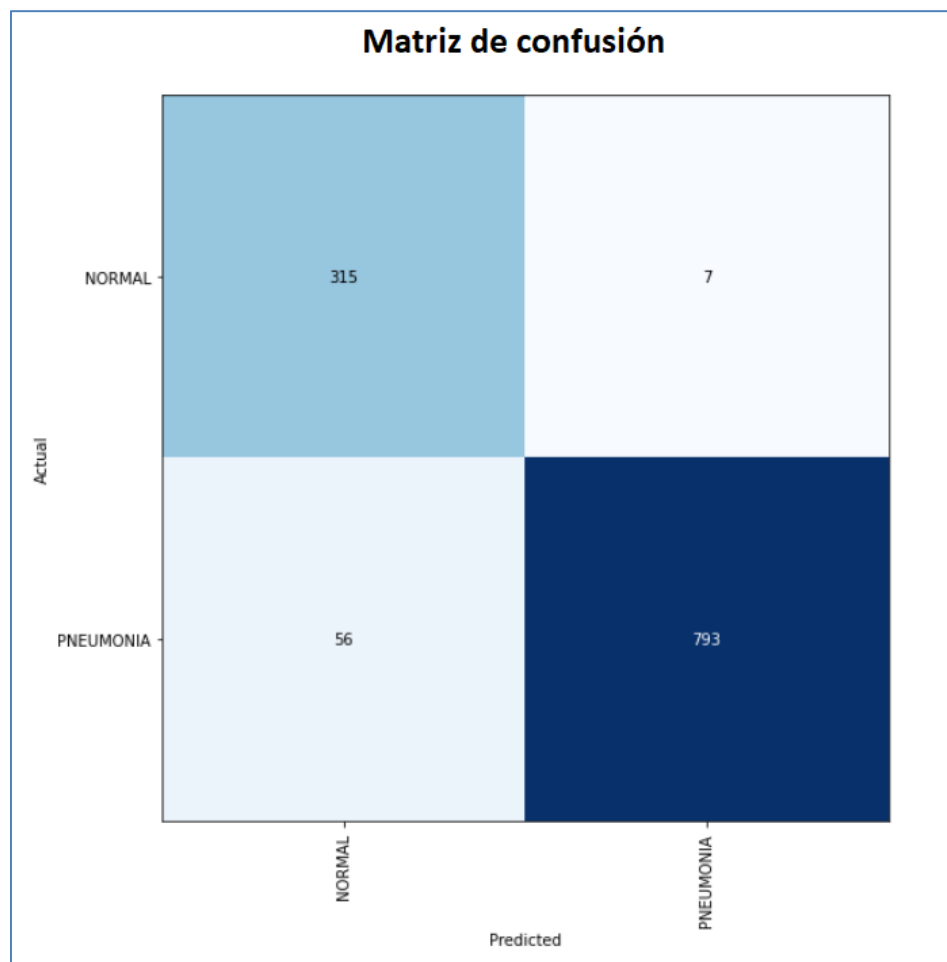


Figura 23: Matriz de confusión del sistema basado en ResNet-50. Elaboración propia.

En esta etapa de evaluación estamos incluyendo también las 1171 imágenes que fueron separadas para evaluar el modelo (1171 de un total de 5232 radiografías), estas son imágenes no antes vistas por el modelo.

Y de estas 1171 imágenes el sistema clasificó dos particiones (clases): 793 pacientes con neumonía correctamente clasificados como neumonía, y acerca de pacientes que no padecen de neumonía (normal), el sistema clasificó correctamente 315 de 322, para una exactitud total por arriba del 95%.

Además, al aumentar la precisión del diagnóstico al usar el sistema implementado, consecuentemente el porcentaje de diagnósticos incorrectos fue disminuido. En la Figura 24 se puede visualizar esta reducción drástica de porcentajes en promedio mensual de 25% a solo 5%.

A continuación, en la Tabla 13 se puede apreciar el detalle de los 12 registros (correspondientes a los 12 meses del año 2020) de la muestra POST Test después de la implementación de un sistema de inteligencia artificial basado en redes neuronales convolucionales ResNet-50.

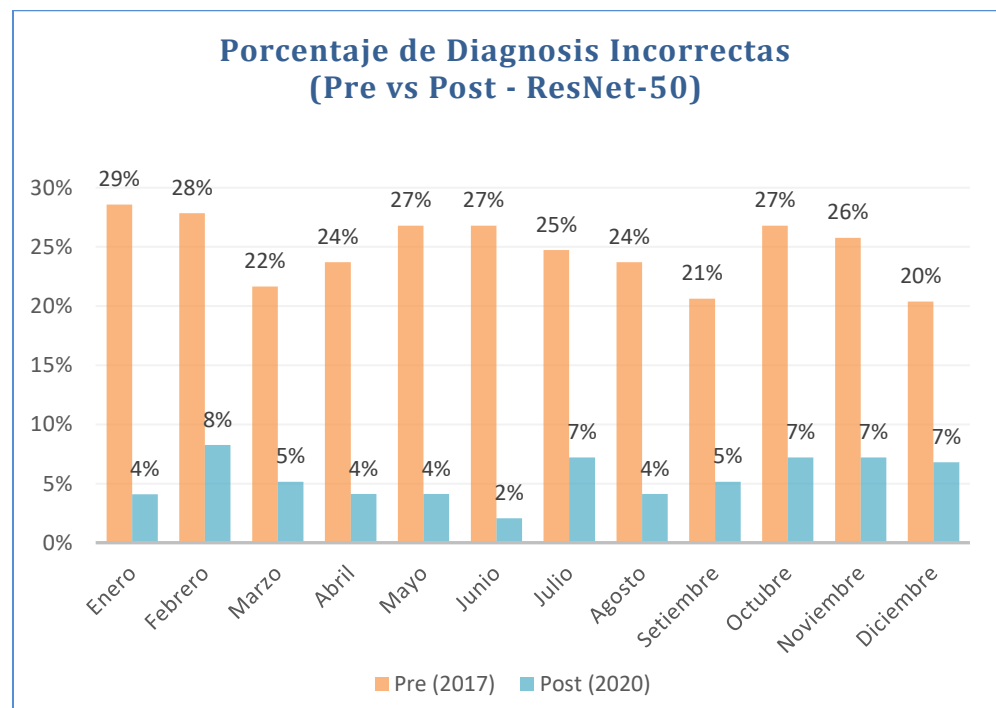


Figura 24: Reducción de Porcentaje de diagnóstico incorrectas después de implementar el sistema basado en ResNet-50. Elaboración propia.

Tabla 13:
Muestra Post Test Porcentaje de Diagnoses Incorrectas (usando ResNet-50)

Mes	Muestra Post Test
Enero	0.0408
Febrero	0.0825
Marzo	0.0515
Abril	0.0412
Mayo	0.0412
Junio	0.0206
Julio	0.0722
Agosto	0.0412
Setiembre	0.0515
Octubre	0.0722
Noviembre	0.0722
Diciembre	0.0680
TOTAL	0.0546

Fuente: SAP Hana DB. Área de Radiología. Elaboración propia.

Aplicando un sistema basado en redes neuronales convolucionales basados en el algoritmo ResNet-50 se ha logrado disminuir en un 20% (de 25% a solo 5%) en promedio el porcentaje mensual de diagnoses incorrectas con un modelo predictivo de alta precisión que además no sufre de problemas de Sobreajuste lo cual es importante para generalizar el modelo a todo tipo de imágenes.

- ✓ **Algoritmo Alexnet:** Aplicar el algoritmo Alexnet para reducir el número de diagnósicos incorrectos en el área de Radiología.

- **Situación PRE Test**

Como anteriormente mencionado, el proceso de diagnóstico de neumonía en el área de radiología se ha venido realizando primero de manera tradicional, hasta la implementación de un sistema de diagnóstico automático que utiliza un algoritmo de reconocimiento de imágenes, pero este sistema original produce diagnósicos de precisión baja (incorrectos).

Este es un gran problema debido a la severidad de esta enfermedad donde el número de muertes anuales por causa de la neumonía ha aumentado de 802 000 a 1 755 000 en tan solo entre los años 2000 y 2018.

Se necesita por lo tanto de un nuevo sistema informático que ayude a acelerar el proceso de diagnóstico para poder salvar vidas, pero esto conlleva a que el nuevo sistema reduzca el número de diagnósicos incorrectos al predecir diagnósicos con alta precisión.

En la Figura 25 se pueden ver el porcentaje de diagnósicos incorrectos durante el año 2017, con un promedio mensual de alrededor de 25% de diagnósicos incorrectos, bastante alto.

A continuación, en la Tabla 14 se puede apreciar el detalle de los 12 registros (correspondientes a los 12 meses del año 2017) de la muestra PRE Test antes de la implementación de un sistema de inteligencia artificial basado en redes neuronales convolucionales basadas en el algoritmo Alexnet.

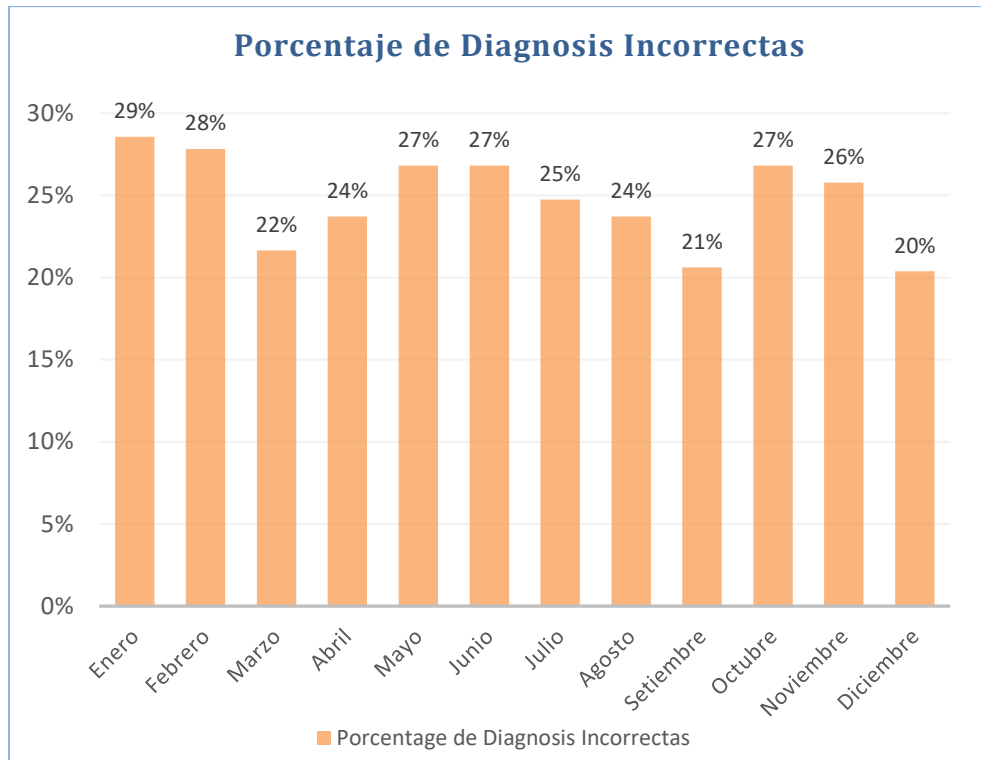


Figura 25: Porcentaje de diagnóstico incorrectas en diagnóstico de neumonía automatizado. Elaboración propia.

Tabla 14:

Muestra Pre Test Porcentaje de Diagnóstico Incorrectas

Mes	Muestra Pre Test
Enero	0.2857
Febrero	0.2784
Marzo	0.2165
Abril	0.2371
Mayo	0.2680
Junio	0.2680
Julio	0.2474
Agosto	0.2371
Setiembre	0.2062
Octubre	0.2680
Noviembre	0.2577
Diciembre	0.2039
TOTAL	0.2478

Fuente: SAP Hana DB. Área de Radiología. Elaboración propia.

- **Aplicación de la teoría**

Para el presente estudio, frente a la problemática que se observaba debido a los bajos niveles de precisión en los diagnósticos automáticos producidos, se ha optado por implementar un sistema basado en redes neuronales convolucionales cuya arquitectura es un algoritmo de redes residuales llamado Alexnet.

Alexnet es el modelo de inteligencia artificial que ganó la competencia ImageNet 2013. Reconocida por ser una arquitectura relativamente simple con cinco capas de convolución (ver Figura 26).

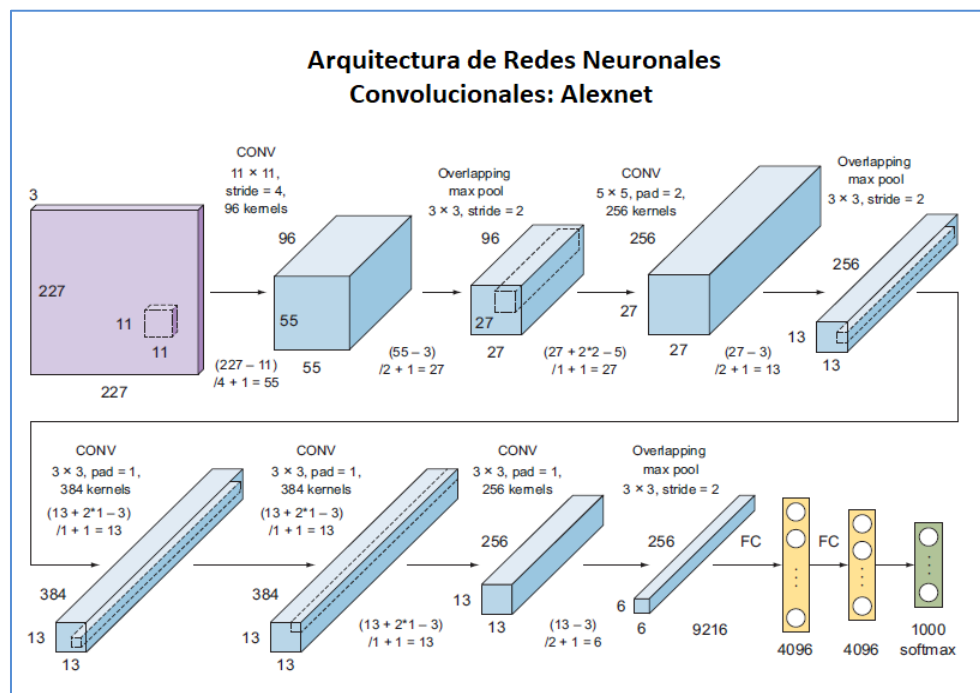


Figura 26: AlexNet contiene ocho capas: cinco convolucionales y tres completamente conectadas (Elgendy, 2020).

A diferencia de la arquitectura ResNet-50, el sistema basado en AlexNet usó ReLU (Rectified Linear Unit) como función de activación y descubrió que entrenaba varias veces más rápido que otras funciones de activación (Shanmugamani, 2018).

Los modelos de Deep Learning usados en clasificación de imágenes obtienen alta precisión en sus predicciones debido a su profundidad, es decir, la alta

cantidad de capas. Sin embargo, al mayor número de capas, se produce el fenómeno llamado desvanecimiento de gradiente (Elgendy, 2020).

Alexnet soluciona este problema al usar la función de activación ReLU en reemplazo a la función de activación sigmoidea que es utilizada en otras arquitecturas, esto se debe a que, cuando la salida de la función de activación sigmoidea está muy cerca de 0 o 1, el gradiente de estas regiones es casi 0, por lo que la propagación hacia atrás no puede continuar actualizando algunos de los parámetros del modelo.

Por el contrario, el gradiente de la función de activación de ReLU en el intervalo positivo es siempre 1 (Zhang, Lipton, Li, & Smola, 2021).

Además de ReLU, para Alexnet también se adoptaron nuevas técnicas de regularización y normalización que permitieron que el modelo entrene más rápido, mejorando el rendimiento del modelo en general.

La primera técnica utilizada fue la inclusión de capas de abandono (Dropout layers). Según (Elgendy, 2020) las capas de abandono se utilizan para evitar que la red neuronal se sobreajuste (overfitting), donde se abandonan neuronas específicas que no contribuyen en el entrenamiento durante la etapa de propagación y que no participan en la retro-propagación.

Esto significa que cada vez que se presenta una nueva entrada al modelo, la red neuronal muestra una arquitectura diferente, pero todas estas arquitecturas comparten los mismos pesos. Esta técnica reduce las co-adaptaciones complejas de las neuronas, ya que una neurona no puede depender de la presencia de otras neuronas en particular.

Por lo tanto, la neurona se ve obligada a aprender características más sólidas que son útiles junto con muchos subconjuntos aleatorios diferentes de las otras neuronas. Alexnet utiliza abandono con una probabilidad de 0.5 en las dos capas completamente conectadas de su arquitectura.

La otra técnica que Alexnet utiliza es Normalización de respuesta local que ayuda a acelerar la convergencia de los pesos del modelo (Elgendy, 2020). Según (NG, 2018) esta técnica es útil especialmente cuando se trata de neuronas ReLU.

Las neuronas ReLU tienen activaciones ilimitadas y necesitamos la normalización de la respuesta local (LRN) para normalizarlas. Para hacer esto, necesitamos identificar características de alta frecuencia. Al aplicar LRN, las neuronas se vuelven más sensibles que las neuronas vecinas.

La combinación de ambas técnicas hace que Alexnet sea una arquitectura de alta precisión sin ser tan profunda lo cual la hace ideal para el despliegue en una aplicación web. Al respecto (Bressem, et al., 2020) describe que el uso de redes neuronales más pequeñas tiene la ventaja de que los requisitos de hardware y el tiempo de entrenamiento son menores en comparación con las redes más profundas.

Los tiempos de entrenamiento más cortos además permiten probar más hiperparámetros, simplificando el proceso de entrenamiento en general. Los requisitos de hardware más bajos también permiten el uso de mayores resoluciones de imagen.

Esto podría ser de relevancia para la evaluación de radiografías torácicas con resoluciones altas generalmente de 2048 x 2048 píxeles, donde hallazgos específicos requieren mayores resoluciones de imágenes de entrada, porque de lo contrario la información crucial con respecto a su presencia podría perderse debido a una reducción de escala.

Además, los tiempos de entrenamiento más cortos podrían simplificar la integración de métodos de mejora en los datos de entrenamiento, como la implementación de supervisada por un experto humano, que puede intervenir y corregir la red en pasos críticos.

Por ejemplo, el experto humano puede comprobar las clasificaciones erróneas con la mayor pérdida en busca de etiquetas incorrectas, reduciendo así eficazmente el ruido de las etiquetas. Con tiempos de entrenamiento más cortos, estos ciclos de retroalimentación se pueden ejecutar más rápido.

En el presente estudio Alexnet fue implementado utilizando el framework de Deep Learning llamado PyTorch que permite la creación y programación de la red neuronal y además permite configurarla en más detalle.

Procediendo con la creación de la red neuronal (ver Figura 27) al crear la clase Model del módulo de PyTorch nn.Module. Se obtuvo el modelo pre-entrenado Alexnet del módulo Torchvision.

```
class Model(nn.Module):
    def __init__(self):
        super(Model, self).__init__()

        # Transfer Learning
        # leveraging the feature extractor of ResNet-152
        # #obtain the ResNet model from torchvision.model library
        self.model = torchvision.models.alexnet(pretrained=True)

        # Building classifier and since we are classifying the images
        # into NORMAL and PNEMONIA, we output a two-dimensional tensor.
        self.classifier = nn.Sequential(
            nn.Linear(self.model.fc.in_features,2),

        )

        for params in self.model.parameters():
            params.requires_grad = False
        self.model.fc = self.classifier

        # Every model built from the nn.Module requires
        # that we override the forward function
        def forward(self, x):
            return self.model(x)
```

Figura 27: Red neuronal convolucional con Transfer Learning de Alexnet. Elaboración propia.

Luego se codificó en PyTorch la función de aprendizaje (fit) (ver Figura 28), y específicamente para el uso del modelo basado en Alexnet se usó el optimizador de estimación de momento adaptativo (Adam).

```

def fit(self, dataloaders, num_epochs):
    train_on_gpu = torch.cuda.is_available()
    optimizer = optim.Adam(self.model.fc.parameters())
    scheduler = optim.lr_scheduler.StepLR(optimizer, 4)
    criterion = nn.NLLLoss()
    since = time.time()

    best_model_wts = copy.deepcopy(self.model.state_dict())
    best_acc = 0.0
    if train_on_gpu:
        self.model = self.model.cuda()
    for epoch in range(1, num_epochs+1):
        print("epoch {}/{}".format(epoch, num_epochs))
        print("-" * 10)

        for phase in ['train', 'test']:
            if phase == 'train':
                scheduler.step()
                self.model.train()
            else:
                self.model.eval()

            running_loss = 0.0
            running_corrects = 0.0

            for inputs, labels in dataloaders[phase]:
                if train_on_gpu:
                    inputs = inputs.cuda()
                    labels = labels.cuda()
                optimizer.zero_grad()

                with torch.set_grad_enabled(phase == 'train'):
                    outputs = self.model(inputs)
                    _, preds = torch.max(outputs, 1)

```

Figura 28: Código de Entrenamiento del modelo Alexnet. Preparación propia.

Según (Kandel, Castelli, & Popovic, 2020) el modelo de redes neuronales aprende (entrena) en un conjunto de datos comparando la etiqueta real de la entrada (disponible en el conjunto de entrenamiento) con la etiqueta predicha, minimizando así la función de costo.

Hipóticamente, si la función de costo es cero, el modelo ha aprendido el conjunto de datos correctamente. Sin embargo, se necesita un algoritmo de optimización para lograr el mínimo de una función de costo, minimizar el error con mayor velocidad.

Para la fase de despliegue se utilizó la plataforma como servicio (PaaS) llamada Render (Goel, 2019) que es una plataforma unificada para desplegar automáticamente todas las aplicaciones web desde Git. Como se puede visualizar en el diagrama del sistema en Figura 29 el flujo del proceso de clasificación integra el modelo Alexnet creado en PyTorch con FastAI, guardando las predicciones en la base de datos SQL de Render.

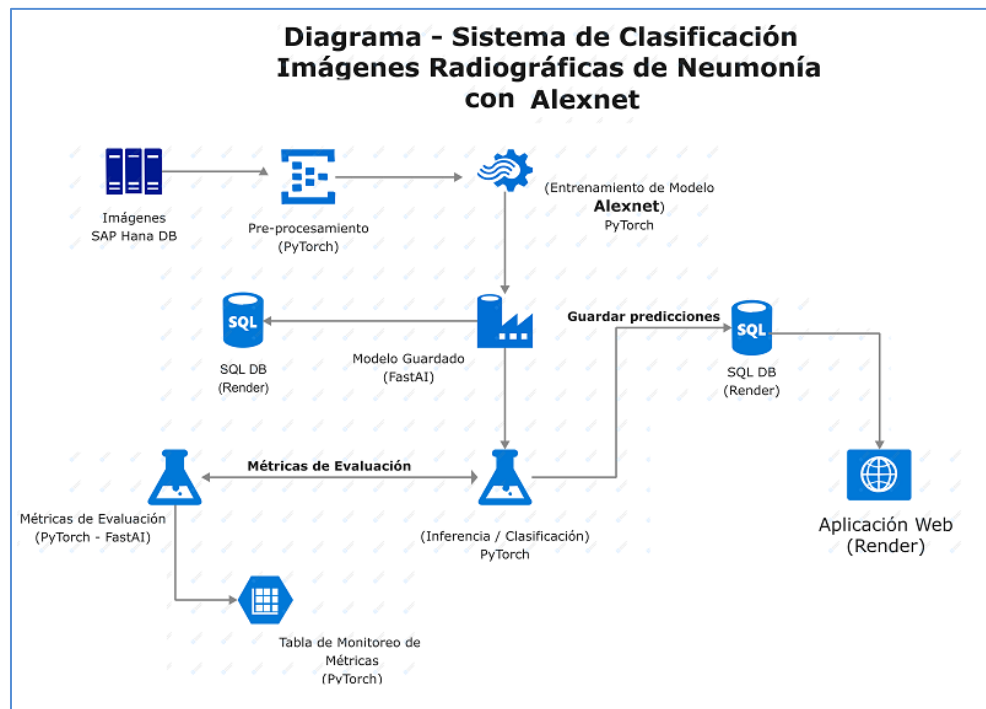


Figura 29: Diagrama del sistema usando Alexnet. Elaboración propia.

La implementación se hizo usando la técnica de Transfer Learning para transferir el modelo pre-entrenado Alexnet y modificándolo solo para que se adecue a las nuevas imágenes radiográficas de las que está aprendiendo.

Al igual que en la implementación de ResNet-50 también fue necesario un preprocesamiento (transformación) de las imágenes, después de extraerlas de la base de datos, con el propósito de estandarizar el tamaño de las imágenes, centrar las imágenes, y luego redimensionarlas con rotaciones aleatorias.

Durante la fase de evaluación de métricas y monitoreo, una vez implementado el sistema de clasificación de imágenes basado en Alexnet (ver Figura 30) se

pudo observar que el modelo obtuvo una alta exactitud (accuracy) de 94% en tan solo 2 minutos después de haber iniciado el entrenamiento y con incluso menor proporción de error (0.058924).

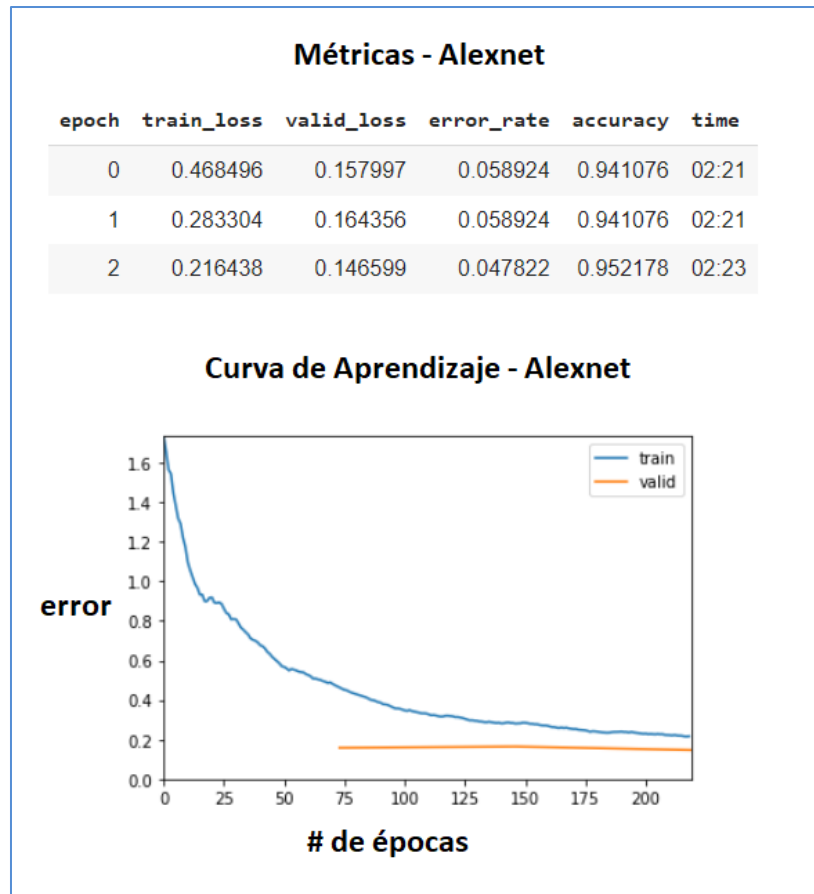


Figura 30: Lista de métricas y Curva de Aprendizaje del modelo basado en Alexnet. Elaboración propia.

Es decir, alta precisión con mayor velocidad de aprendizaje debido a que es una arquitectura de menor profundidad. En la misma figura se puede apreciar las métricas de evaluación y curva de aprendizaje del modelo Alexnet, después de tres épocas de entrenamiento.

Otras métricas que nuestro sistema puede mostrar son:

- la cantidad de error de entrenamiento (train_loss), el cual se refiere al valor de error final alcanzado en cada época (iteración) de entrenamiento
- la cantidad de error de validación (valid_loss), el cual refiere al valor de error final alcanzado en cada época (iteración) de evaluación.

Una manera de medir el desempeño de un modelo de inteligencia artificial es observando ambas métricas `train_loss` y `valid_loss` cuyos valores deben ser similares, de no ser así modelaría estaría incurriendo en lo que es llamado Sobreajust (overfitting) donde el modelo no aprende sino solo memoriza de los datos.

En este caso los valores son similares, con `train_loss` de 0.2 y `valid_loss` de 0.15, demostrando que el modelo generaliza (aprende) correctamente.

Como parte de las métricas de monitoreo se encuentra curva de aprendizaje, podemos observar que la tasa de error correspondiente al aprendizaje basado en los datos de entrenamiento (`train_loss`) no sigue la misma trayectoria que la tasa de error correspondiente al aprendizaje basado en los datos de validación (`valid_loss`).

Esto demuestra que el modelo basado en Alexnet aprende o converge con más rapidez y por lo tanto alcanza la una alta precisión desde la primera época. En general esto demuestra que aparte de alcanzar una exactitud alta, también el modelo es adaptado para la implementación de una aplicación web en plataformas con limitaciones de capacidad (dispositivos móviles).

▪ **Situación POST Test**

Después de la implementación del nuevo sistema de clasificación de imágenes, el área de radiología registró una alta cantidad de diagnósticos automatizados correctamente clasificados, lo cual fue corroborado por 10 médicos especialistas.

Podemos ver en la Figura 31 la matriz de confusión mostrando específicamente cuantas radiografías fueron diagnosticadas por el sistema correctamente para cada clase: Neumonía y Normal.

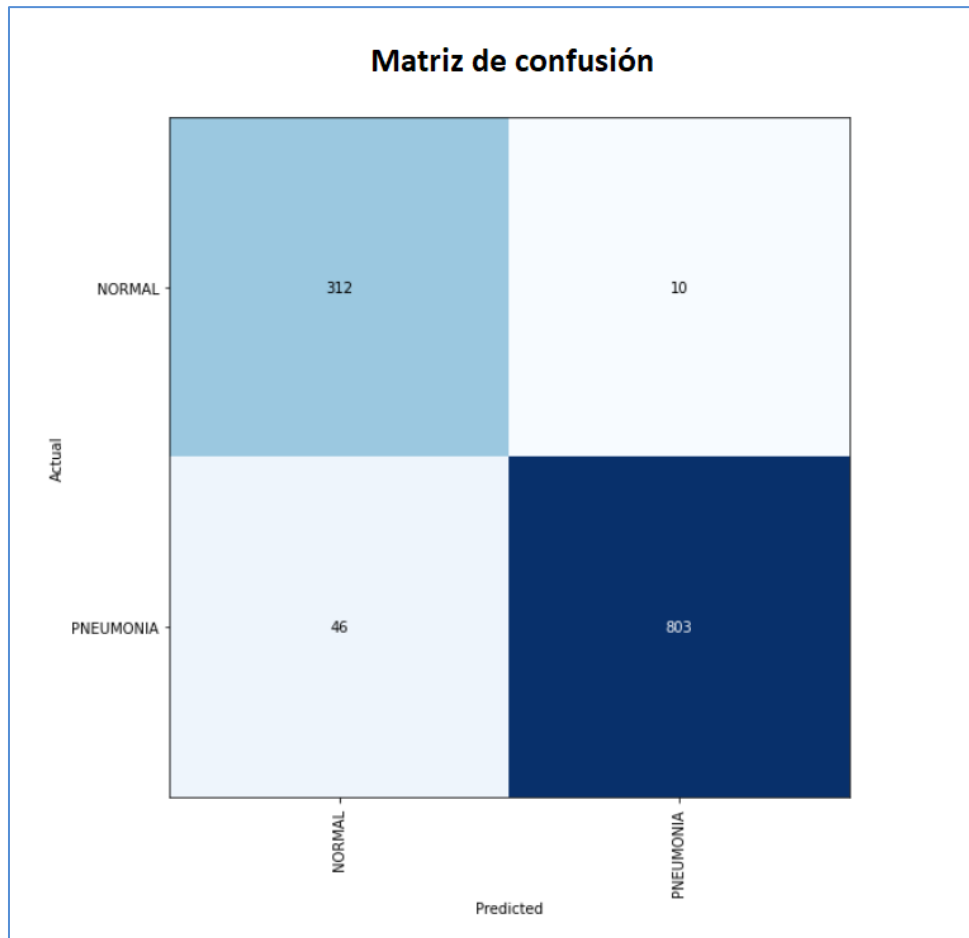


Figura 31: Matriz de confusión del sistema basado en Alexnet. Elaboración propia.

En esta etapa de evaluación estamos incluyendo también las 1171 imágenes que fueron separadas para evaluar el modelo (1171 de un total de 5232 radiografías), estas son imágenes no antes vistas por el modelo.

Y de estas 1171 imágenes el sistema clasificó dos particiones (clases): 803 pacientes con neumonía correctamente clasificados como neumonía, y acerca de pacientes que no padecen de neumonía (normal), el sistema clasificó correctamente 312 de 322, para una exactitud total por arriba del 95%.

Además, al aumentar la precisión del diagnóstico al usar el sistema implementado, consecuentemente el porcentaje de diagnosis incorrectas fue disminuido. En la Figura 32 se puede visualizar esta reducción drástica de porcentajes en promedio mensual de 25% a solo 5%.

A continuación, en la Tabla 15 se puede apreciar el detalle de los 12 registros (correspondientes a los 12 meses del año 2020) de la muestra POST Test después de la implementación de un sistema de inteligencia artificial basado en redes neuronales convolucionales Alexnet.

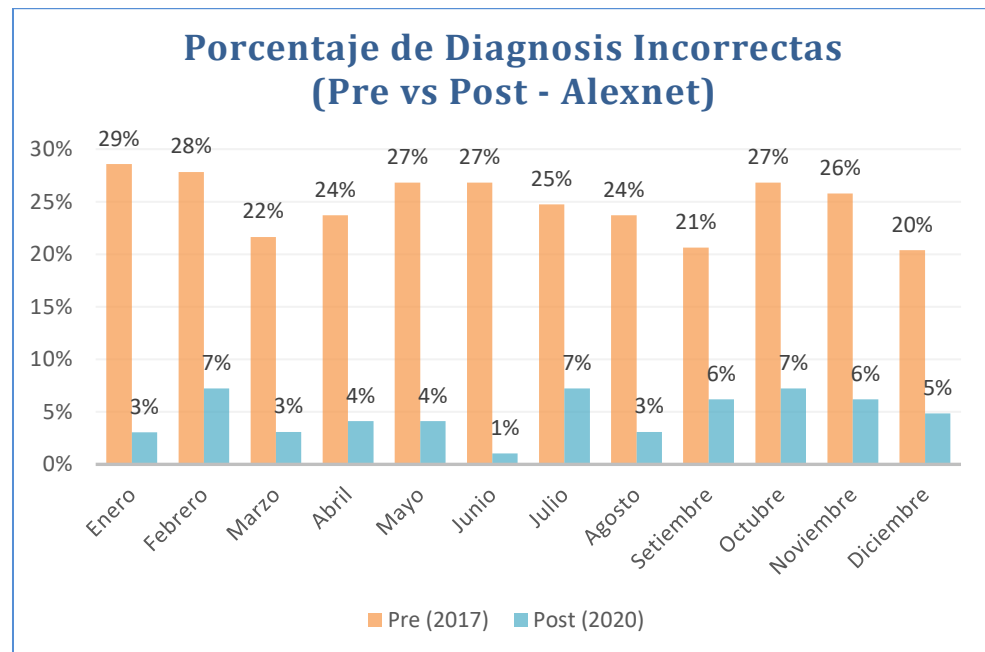


Figura 32: Reducción de Porcentaje de diagnóstico incorrectas después de implementar el sistema basado en Alexnet. Elaboración propia.

Tabla 15:

Muestra Post Test Porcentaje de Diagnóstico Incorrectas (usando Alexnet)

Mes	Muestra-Post-Test
Enero	0.0306
Febrero	0.0722
Marzo	0.0309
Abril	0.0412
Mayo	0.0412
Junio	0.0103
Julio	0.0722
Agosto	0.0309
Setiembre	0.0619
Octubre	0.0722
Noviembre	0.0619
Diciembre	0.0485
TOTAL	0.0478

Fuente: SAP Hana DB. Área de Radiología. Elaboración propia.

Después de implementar el nuevo sistema esta vez basado en Alexnet, el área de radiología además de ver la reducción en diagnósticos incorrectos, a diferencia de ResNet-50 esta versión del sistema también permitió que al ser desplegado pueda disminuir la latencia en el tiempo de entrenamiento (ayudando al objetivo de acelerar el proceso en general).

- ✓ **Algoritmo VGG:** Aplicar el algoritmo VGG para reducir el número de diagnósicos incorrectos en el área de Radiología.

- **Situación PRE Test**

Como anteriormente mencionado, el proceso de diagnóstico de neumonía en el área de radiología se ha venido realizando primero de manera tradicional, hasta la implementación de un sistema de diagnóstico automático que utiliza un algoritmo de reconocimiento de imágenes con resultados de precisión baja, en la Figura 33 se pueden ver el porcentaje de diagnósicos incorrectos durante el año 2017.

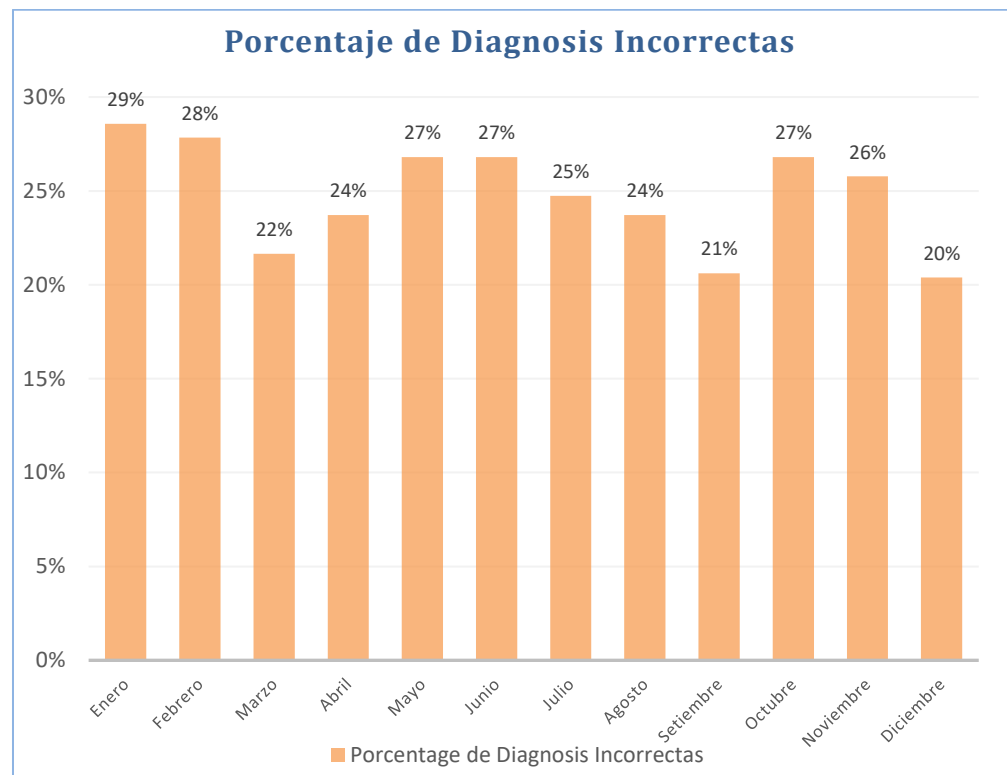


Figura 33: Porcentaje de diagnósicos incorrectos en diagnóstico de neumonía automatizado. Elaboración propia.

A continuación, en la Tabla 16 se puede apreciar el detalle de los 12 registros (correspondientes a los 12 meses del año 2017) de la muestra PRE Test antes de la implementación de un sistema de inteligencia artificial basado en redes neuronales convolucionales basadas en el algoritmo VGG.

Tabla 16:
Muestra Pre Test Porcentaje de Diagnósis Incorrectas

Mes	Muestra Pre Test
Enero	0.2857
Febrero	0.2784
Marzo	0.2165
Abril	0.2371
Mayo	0.2680
Junio	0.2680
Julio	0.2474
Agosto	0.2371
Setiembre	0.2062
Octubre	0.2680
Noviembre	0.2577
Diciembre	0.2039
TOTAL	0.25

Fuente: SAP Hana DB. Área de Radiología. Elaboración propia.

▪ **Aplicación de la teoría**

Para el presente estudio, frente a la problemática que se observaba debido a los bajos niveles de precisión en los diagnósticos automáticos producidos, se ha optado por implementar un sistema basado en redes neuronales convolucionales cuya arquitectura es un algoritmo de redes residuales llamado VGG-16.

VGG (Visual Geometry Group), es un algoritmo también conocido como VGG16, es un modelo de inteligencia artificial cuyos componentes de la arquitectura son exactamente los mismos que AlexNet, excepto que VGG es una red aún más profunda con más capas convolucionales, capas agrupadas y capas densas. Según (Elgandy, 2020) VGG, consta de 16 capas: 13 capas convolucionales y 3 capas completamente conectadas. Su arquitectura uniforme es la que hace popular en la comunidad de Deep Learning porque es muy fácil de entender. Mejora a la arquitectura AlexNet al reemplazar filtros de gran tamaño de kernel (11 y 5 en la primera y segunda capas

convolucionales) con múltiples filtros de tamaño de grupo de 3×3 uno tras otro (ver Figura 34).

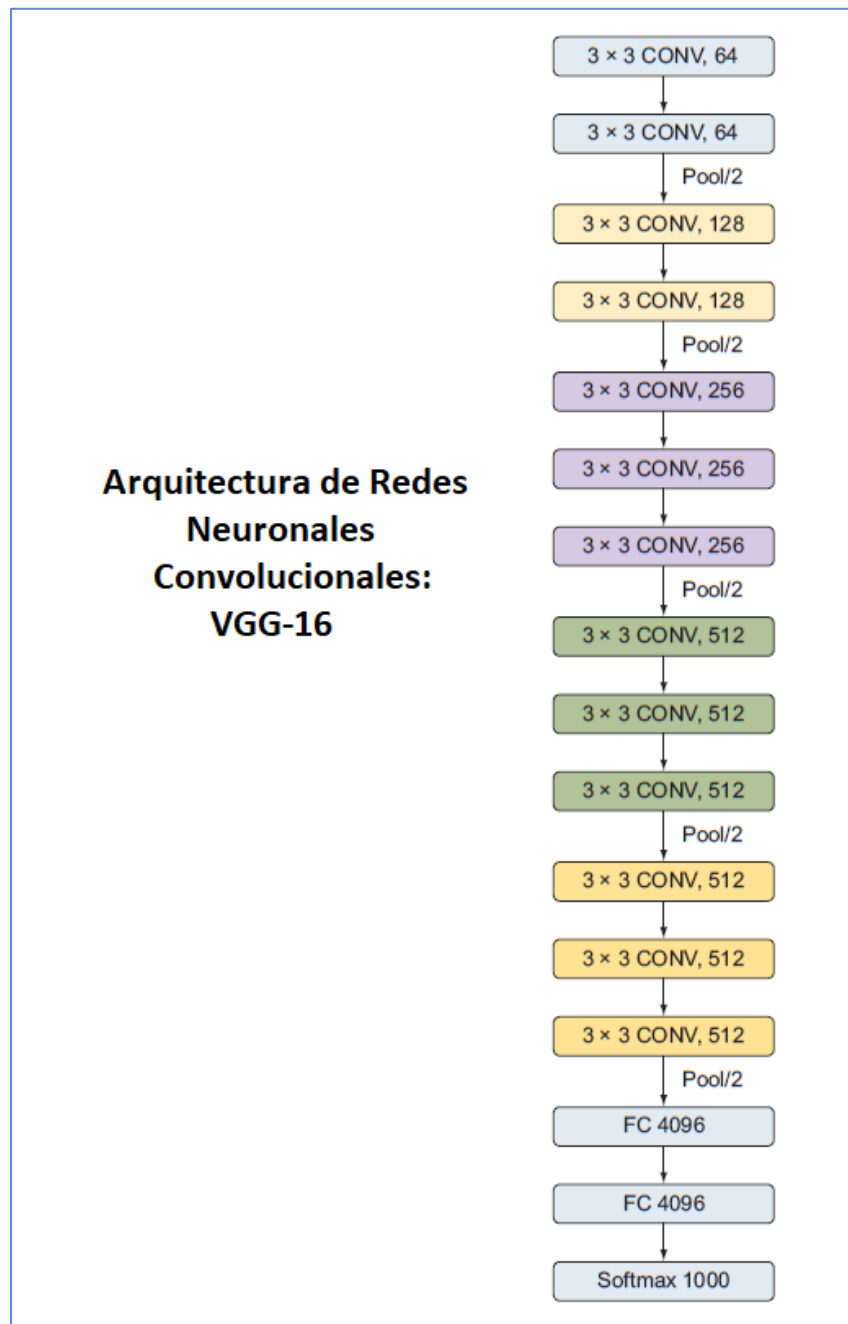


Figura 34: VGG-16 contiene 16 con una arquitectura uniforme y fácil de entender (Elgandy, 2020).

Este tamaño de los filtros permite que la red extraiga características de nivel más fino de la imagen en comparación con los núcleos grandes de AlexNet (11×11 y 5×5).

La idea es que con un campo receptivo convolucional dado, múltiples núcleos más pequeños apilados es mejor que un núcleo más grande porque tener múltiples capas no lineales aumenta la profundidad de la red; esto le permite aprender características más complejas a un costo menor porque tiene menos parámetros de aprendizaje.

Por lo tanto, al usar convoluciones de 3×3 con mayor profundidad, obtiene los beneficios de usar más capas de rectificación no lineal (ReLU), lo que hace que la función de decisión sea más discriminativa.

En segundo lugar, esto disminuye el número de parámetros de entrenamiento porque cuando se usa un convolucional 3×3 de tres capas con canales C , la pila se parametriza con pesos $3^2C^2 = 9C^2$ en comparación con una sola capa convolucional 7×7 que requiere pesos $7^2C^2 = 49C^2$, que es un 82% más de parámetros.

Esta configuración unificada de los componentes convolucionales y agrupados simplifica la arquitectura de la red neuronal, lo que la hace muy fácil de entender e implementar (Elgandy, 2020).

Al igual que en la implementación de ResNet-50 y Alexnet también fue necesario un preprocesamiento (transformación) de las imágenes con el propósito de estandarizar el tamaño de las imágenes, centrar las imágenes, y luego redimensionarlas con rotaciones aleatorias.

En todo modelo de redes neuronales es necesario este proceso para optimizar el rendimiento.

Procediendo se procedió con la creación de la red neuronal (ver Figura 35) al crear la clase Model del módulo de PyTorch nn.Module. Se obtuvo el modelo pre-entrenado VGG16 del módulo Torchvision.

```

class Model(nn.Module):
    def __init__(self):
        super(Model, self).__init__()

        # Transfer Learning
        # leveraging the feature extractor of ResNet-152
        # #obtain the ResNet model from torchvision.model library
        self.model = torchvision.models.vgg16(pretrained=True)

        # Building classifier and since we are classifying the images
        # into NORMAL and PNEUMONIA, we output a two-dimensional tensor.
        self.classifier = nn.Sequential(
            nn.Linear(self.model.fc.in_features,2),

        )

        for params in self.model.parameters():
            params.requires_grad = False
        self.model.fc = self.classifier

        # Every model built from the nn.Module requires
        # that we override the forward function
        def forward(self, x):
            return self.model(x)

```

Figura 35: Red neuronal convolucional con Transfer Learning de Alexnet. Elaboración propia.

Luego se codificó en PyTorch la función de aprendizaje (fit) (ver Figura 36), específicamente para el uso del modelo VGG-16 implementado tomó como base una versión que resultó con alta precisión en diagnóstico de cáncer pulmonar en la investigación de (Pang, et al., 2020).

```

def fit(self, dataloaders, num_epochs):
    train_on_gpu = torch.cuda.is_available()
    optimizer = optim.Adam(self.model.fc.parameters())
    scheduler = optim.lr_scheduler.StepLR(optimizer, 4)
    criterion = nn.NLLLoss()
    since = time.time()

    best_model_wts = copy.deepcopy(self.model.state_dict())
    best_acc = 0.0
    if train_on_gpu:
        self.model = self.model.cuda()
    for epoch in range(1, num_epochs+1):
        print("epoch {}/{}".format(epoch, num_epochs))
        print("-" * 10)

        for phase in ['train', 'test']:
            if phase == 'train':
                scheduler.step()
                self.model.train()
            else:
                self.model.eval()

            running_loss = 0.0
            running_corrects = 0.0

            for inputs, labels in dataloaders[phase]:
                if train_on_gpu:
                    inputs = inputs.cuda()
                    labels = labels.cuda()
                optimizer.zero_grad()

```

Figura 36: Código de Entrenamiento del modelo VGG. Preparación propia.

Donde los autores describen que la red VGG puede procesar bien las imágenes al aproximarse rápidamente a la función objetivo y abstraer representaciones de características significativas. VGG-16 funciona bien en el diagnóstico de imágenes radiográficas, por lo que lo eligieron como la estructura principal de su modelo.

La implementación se hizo usando FastAI (Howard, 2018) que es una biblioteca de Deep Learning basada en PyTorch dedicada al rápido despliegue de un modelo de redes neuronales usando la función `cnn_learner` en conjunto con `models.vgg16_bn` (ver Figura 37).

```
from fastai.vision import *
from fastai.metrics import error_rate

# VGG

metrics = [error_rate, accuracy]

learn_vgg = cnn_learner(
    data,
    models.vgg16_bn,
    metrics = metrics,
    model_dir = "/tmp/model/",
    callback_fns = ShowGraph
)

defaults.device = torch.device('cuda') # makes sure the gpu is used

learn_vgg.fit_one_cycle(3)
```

Figura 37: Código de FastAI creando un modelo para el despliegue usando VGG-16. Elaboración propia.

Para la fase de despliegue se utilizó la plataforma como servicio (PaaS) llamada Render (Goel, 2019) que es una plataforma unificada para desplegar automáticamente todas las aplicaciones web desde Git.

Como se puede visualizar en el diagrama del sistema en Figura 38 el flujo del proceso de clasificación integra el modelo VGG-16 creado en PyTorch con FastAI, guardando las predicciones en la base de datos SQL de Render.

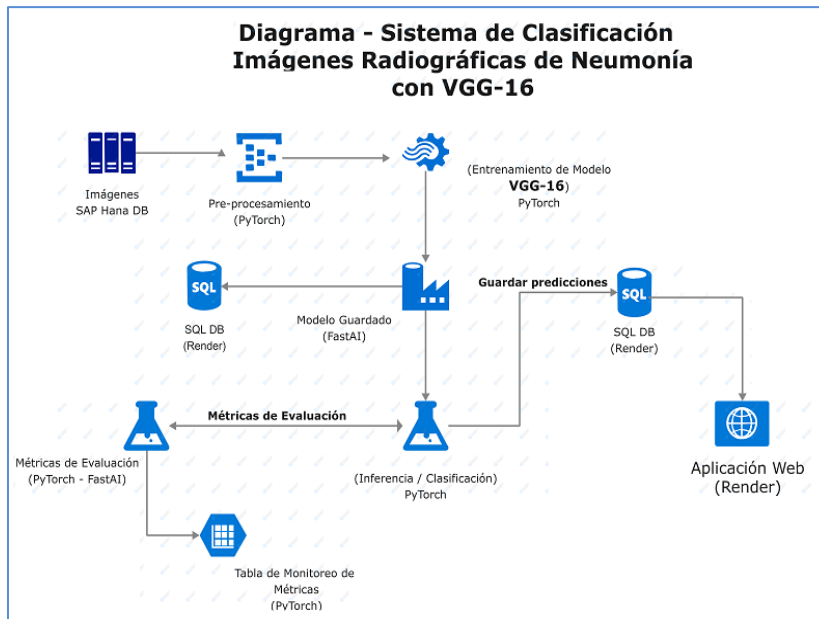


Figura 38: Diagrama del sistema usando VGG-16. Elaboración propia.

Durante la fase de evaluación de métricas y monitoreo, una vez implementado el sistema de clasificación de imágenes basado en VGG se pudo observar que el modelo obtuvo una alta exactitud (accuracy) de 95%. En la Figura 39 se puede apreciar las métricas de evaluación y curva de aprendizaje del modelo VGG, después de tres épocas de entrenamiento.

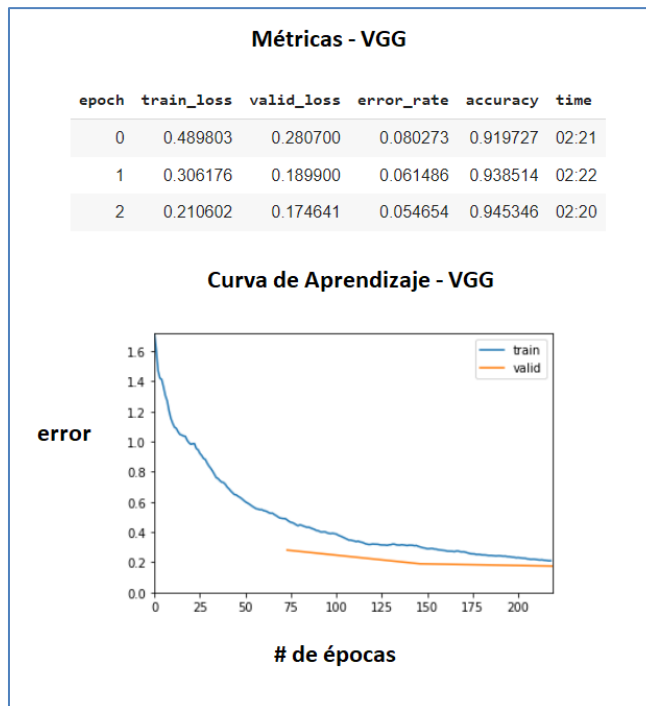


Figura 39: Lista de métricas y Curva de Aprendizaje del modelo basado en VGG. Elaboración propia.

- **Situación POST Test**

Después de la implementación del sistema se identificó una mayor precisión en todas las radiografías evaluadas, alcanzando una precisión en promedio de entre 99% y 100% para los pacientes que padecían de neumonía.

Podemos ver en la Figura 40 la matriz de confusión mostrando específicamente cuantas radiografías fueron diagnosticadas por el sistema correctamente para cada clase: Neumonía y Normal.

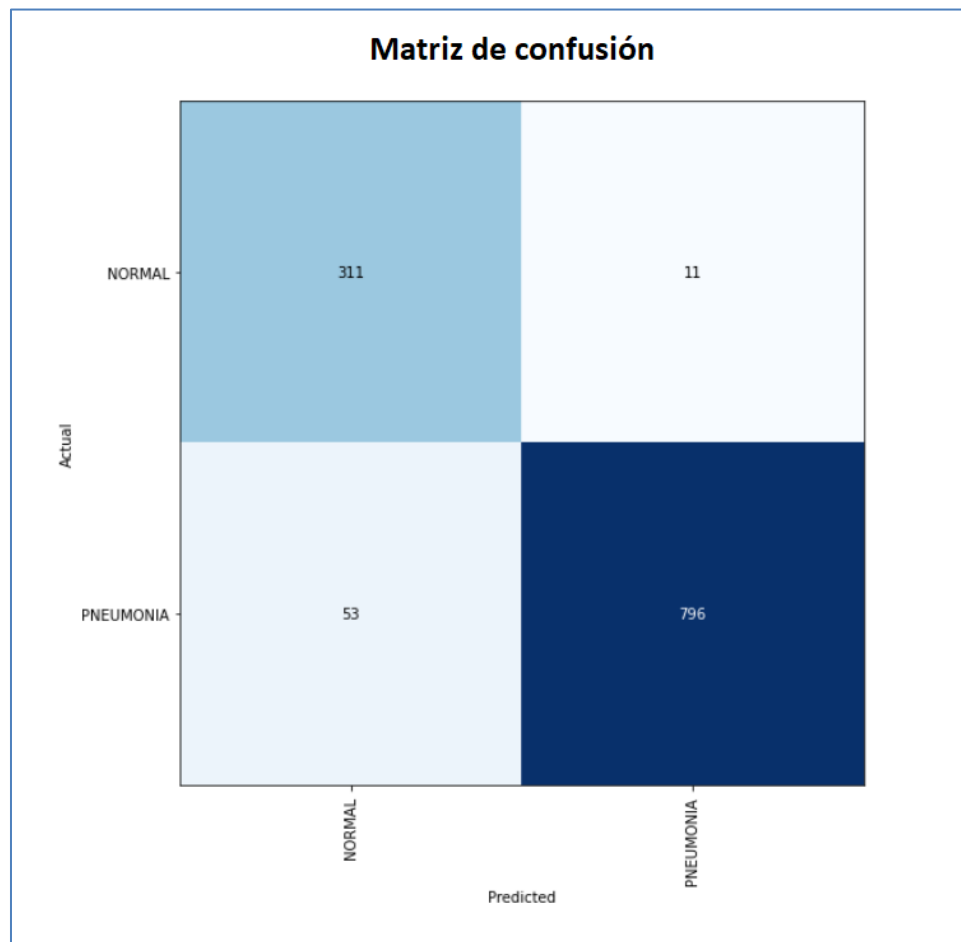


Figura 40: Matriz de confusión del sistema basado en VGG. Elaboración propia.

En la etapa de evaluación usamos las 1171 imágenes que fueron separadas para evaluar el modelo (1171 de un total de 5232 radiografías), estas son imágenes no antes vistas por el modelo.

El sistema clasificó correctamente 796 radiografías de pacientes con neumonía, y acerca de pacientes que no padecen de neumonía (normal), el sistema clasificó correctamente 311 de 322, para una exactitud total por arriba del 95%.

Además, al aumentar la precisión del diagnóstico al usar el sistema implementado, consecuentemente el porcentaje de diagnósticos incorrectos fue disminuido. En la Figura 41 se puede visualizar esta reducción drástica de porcentajes en promedio mensual de 25% a solo 5%.

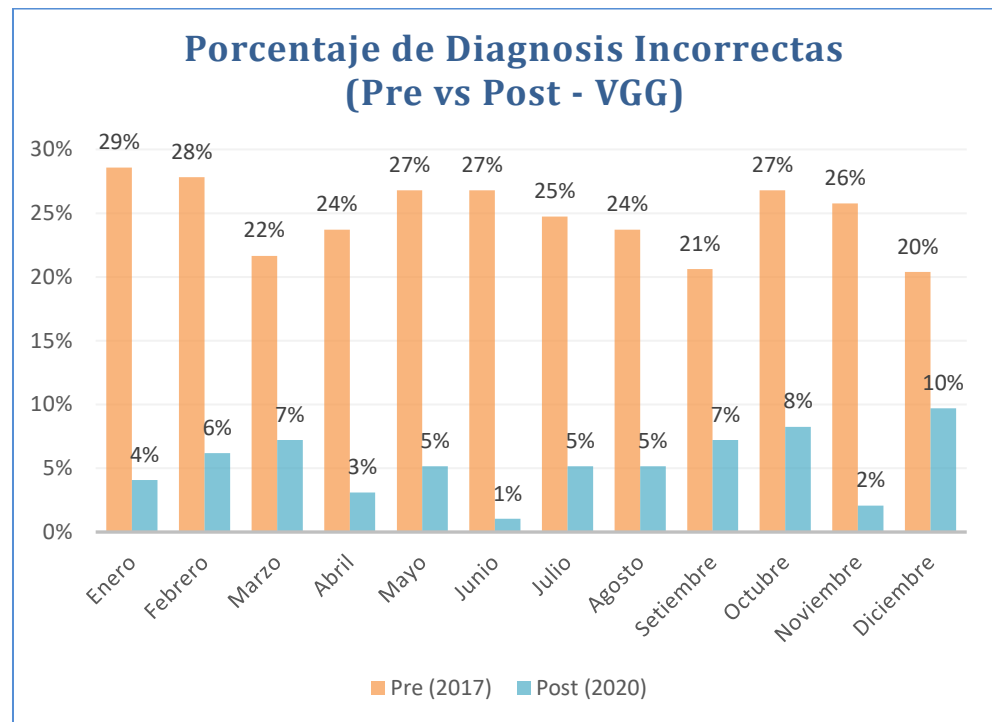


Figura 41: Reducción de Porcentaje de diagnóstico incorrectas después de implementar el sistema basado en VGG. Elaboración propia.

A continuación, en la Tabla 17 se puede apreciar el detalle de los 12 registros (correspondientes a los 12 meses del año 2020) de la muestra POST Test después de la implementación de un sistema de inteligencia artificial basado en redes neuronales convolucionales VGG.

Tabla 17:
Muestra Post Test Porcentaje de Diagnósis Incorrectas (usando VGG)

Mes	Muestra Post Test
Enero	0.0408
Febrero	0.0619
Marzo	0.0722
Abril	0.0309
Mayo	0.0515
Junio	0.0103
Julio	0.0515
Agosto	0.0515
Setiembre	0.0722
Octubre	0.0825
Noviembre	0.0206
Diciembre	0.0971
TOTAL	0.0536

Fuente: SAP Hana DB. Área de Radiología. Elaboración propia.

Después de implementar el nuevo sistema esta vez basado en VGG, el área de radiología además de ver la reducción en diagnósticos incorrectos también pudo notar los beneficios de este nuevo sistema intermedio en términos de profundidad de la arquitectura.

Es un punto equilibrado, capaz de lidiar con imágenes de alta resolución al tener más capas que Alexnet, y a la vez manteniendo una simplicidad para un entrenamiento más rápido, al tener menos capas que ResNet-50.

- ✓ **Función de Activación Softmax:** Implementar la Función de Activación SoftMax para reducir el tiempo de espera de pacientes con neumonía.

- **Situación PRE Test**

En la actualidad uno de los retos más grandes en el objetivo de acelerar el proceso de diagnóstico y tratamiento de neumonía es la falta de visibilidad de los pacientes que llegan a un área de radiología para descartar si padece o no de neumonía.

Esta falta de visibilidad ocasiona que los pacientes sean admitidos para el análisis en orden de llegada, y por lo tanto casos urgentes de pacientes con complicaciones (por ejemplo: absceso pulmonar) terminan esperando un número de horas valiosas que aumenta el riesgo de severidad de la enfermedad en también la mortalidad al ser admitidos horas después que pacientes que no padecen de la enfermedad.

Accediendo la base de datos del sistema del área de radiología (2017) encontramos el número de horas más altas de espera en pacientes que necesitaron una priorización debido a padecer de la enfermedad, estos datos fueron una intersección de hora de llegada, hora de atención y diagnóstico final (después de ser atendido).

Estos datos han sido promediados por mes, y se pueden visualizar en la Figura 42 donde se observa que en promedio hay tiempos de espera desde 4 a casi 6 horas de espera que podrían disminuirse si (excluyendo demoras por causas administrativas) se implementara un sistema que integre una visibilidad al estado de múltiples pacientes basado simplemente en las imágenes.

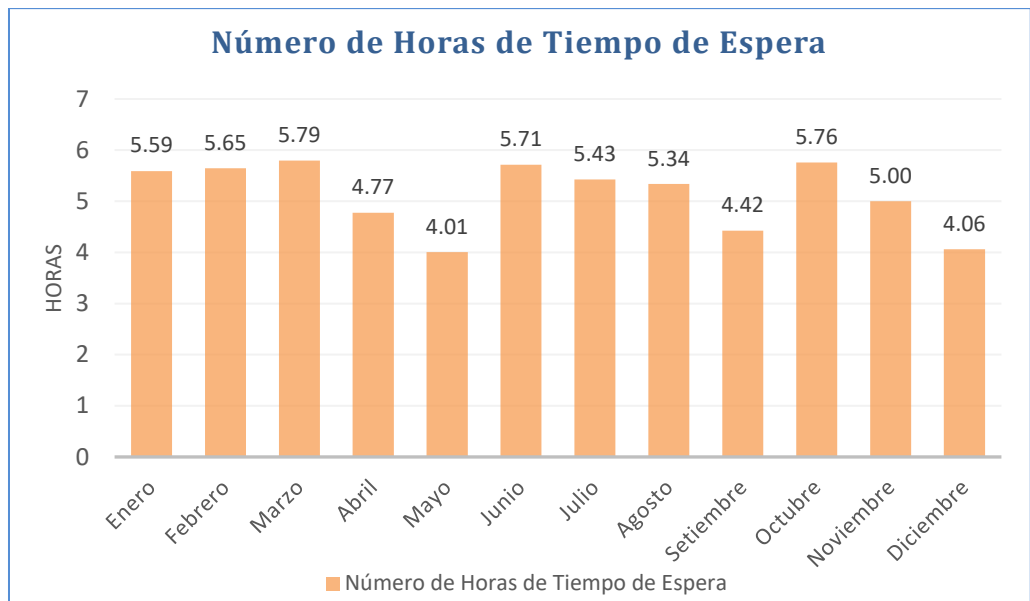


Figura 42: Promedio mensual de número de horas en tiempo más alto de espera de pacientes con neumonía. Elaboración propia.

A continuación, en la Tabla 18 se puede apreciar el detalle de los 12 registros (correspondientes a los 12 meses del año 2017) de la muestra PRE Test antes de la implementación de un nuevo sistema que permita visibilidad al estado de los pacientes en referencia a si padecen de neumonía o no.

Tabla 18:
Muestra Pre Test Horas de Tiempo de Espera

Mes	Muestra Pre Test
Enero	5.59
Febrero	5.65
Marzo	5.79
Abril	4.77
Mayo	4.01
Junio	5.71
Julio	5.43
Agosto	5.34
Setiembre	4.42
Octubre	5.76
Noviembre	5.00
Diciembre	4.06
TOTAL	5.1275

Fuente: SAP Hana DB. Área de Radiología. Elaboración propia.

▪ **Aplicación de la teoría**

Para el presente estudio, frente a la problemática que se observaba debido a la falta de visibilidad basado en solo las imágenes radiográficas de múltiples pacientes, se ha optado por agregar a nuestro sistema de diagnóstico automatizado basado en redes neuronales convolucionales, una capa extra que permita procesar múltiples imágenes y mostrar las probabilidades de que cada imagen sea de un paciente que padezca de neumonía o no, esta es una capa con la función de activación SoftMax.

Según (Shanmugamani, 2018) Softmax es una forma de obligar a las redes neuronales a generar la suma de 1. Por lo tanto, los valores de salida de la función softmax se pueden considerar como parte de una distribución de probabilidad.

Esto es útil en problemas de clasificación de clases múltiples. Softmax es un tipo de función de activación con la especialidad de sumar la salida a 1.

Convierte las salidas en probabilidades dividiendo la salida por la suma de todos los demás valores.

Y acerca de la implementación de la función de la función softmax (Zhang, Lipton, Li, & Smola, 2021) menciona que cuando el enfoque es interpretar las salidas de nuestro modelo como probabilidades uno podría tener la tentación de sugerir que interpretemos los logits directamente como nuestras salidas de interés.

Sin embargo, existen algunos problemas al interpretar directamente la salida de la capa lineal como una probabilidad. Por un lado, nada limita estos números a sumar 1. Por otro lado, dependiendo de las entradas, pueden tomar valores negativos.

Para interpretar nuestras salidas como probabilidades, debemos garantizar que (incluso con datos nuevos), serán no negativos y suman 1.

Además, necesitamos un objetivo de entrenamiento que anime al modelo a estimar probabilidades fielmente.

De todas las instancias en las que un clasificador genera 0.5, esperamos que la mitad de esos ejemplos pertenezcan realmente a la clase predicha.

Acerca de la implementación de softmax en clasificación de imágenes médicas (Syed, 2018) menciona que la salida sería un valor entre 0 y 1 para cada clase que representa la probabilidad de que la imagen pertenezca a esa clase.

Por esta razón, lógicamente, la capa final en la red neuronal debe ser una función softmax que, como se explicó, genera el valor tal como se requiere.

Y (Vázquez Enríquez, 2019) explica que la capa de salida para tareas de clasificación multiclase, por lo que después de la última capa del modelo convolucional debemos agregar una capa softmax.

La capa softmax asignará una probabilidad decimal a cada clase de acuerdo con su relevancia y la suma de las probabilidades de todas las clases sumará 1.

Es decir, debido a que ya tenemos un sistema basado en redes neuronales convolucionales, es posible agregar una capa con activación softmax.

Para lograr esto en usando PyTorch se guardó el modelo pre-entrenado (ResNet-50, Alexnet o VGG-16) y en el código del modelo usamos el método `requires_grad` y lo configuramos en `False` (ver Figura 43). Este código es esencial para modificar la última capa.

```

self.classifier = nn.Sequential(
    nn.Linear(self.model.fc.in_features,2),
)

for params in self.model.parameters():
    params.requires_grad = False
self.model.fc = self.classifier

def fit(self, dataloaders, num_epochs):
    train_on_gpu = torch.cuda.is_available()
    optimizer = optim.Adam(self.model.fc.parameters())
    scheduler = optim.lr_scheduler.StepLR(optimizer, 4)
    criterion = nn.NLLLoss()
    since = time.time()

```

Figura 43: Código de modificación de red neuronal para poder después integrar Softmax. Elaboración propia.

Lo que el código hizo que al final del primer pase a través de las capas pre-entrenadas se detenga al final y no se retro-propague, luego re-definimos esa última capa con `self.classifier` usando la función de Pytorch de red neuronal `nn.Sequential`.

Nuestro modelo ha sido programado para que el cálculo del error sea basado en el criterio llamado `nn.NLLLoss()` (pérdida de probabilidad logarítmica negativa) que produce como salida el logaritmo de las probabilidades.

En Pytorch creamos la función softmax usando `torch.exp(logps)` (ver Figura 44) y teniendo como resultado final las probabilidades de que cada imagen pertenezca a la clase neumonía o a la clase normal.

Para la fase de despliegue se utilizó la plataforma como servicio (PaaS) llamada Render (Goel, 2019) que es una plataforma unificada para desplegar automáticamente todas las aplicaciones web desde Git.

Como se puede visualizar en el diagrama del sistema en Figura 45 el flujo del proceso de clasificación integra el modelo ResNet-50 creado en PyTorch con FastAI, guardando las predicciones en la base de datos SQL de Render.

```

model_save_name = 'classifier.pt'
# model.model.load_state_dict(state_dict)
model_ft = model.model
#load image and perform a forward pass through our model.
img = image_loader(image_path)

logps = model_ft(img.cuda() if torch.cuda.is_available() else img)

#since we have performed a forward pass through our model, we can obtain activations
# from layer(model.layer4) defined in the LayerActivation class from the features
# list and take out the batch dimension.
out_features = acts.features[0].squeeze(0)

# Changes shape from 2048 x 7 x7 to 7 x 7 x 2048. Just performs
# a matrix transpose on the output features tensor.
out_features = np.transpose(out_features.cpu(),(1,2,0))

# Our final model layer is a log-softmax activation.
# We perform torch.exp to take out the log and obtain the softmax values.
ps = torch.exp(logps)

# Obtain the axis of the predicted class.
pred = np.argmax(ps.cpu().detach())

W = model_ft.fc[0].weight
top_probs, top_classes = torch.topk(ps, k=2)
pred = np.argmax(ps.detach())
w = W[pred,:]

```

Figura 44: Código de creación de la función de activación softmax usando torch.exp. Elaboración propia.

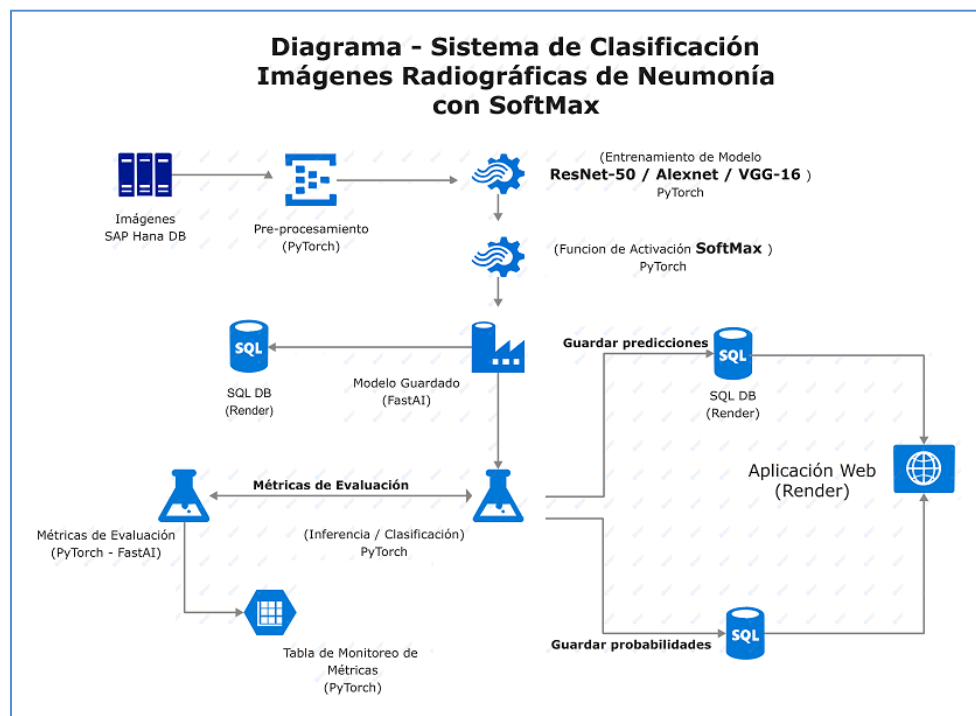


Figura 45: Diagrama del sistema con integración de activación softmax. Elaboración propia

En el sistema implementado hemos integrado la solución completa, donde nuestro sistema trabaja con los tres diferentes modelos pre-entrenados como un motor cognitivo de clasificación binaria (neumonía vs normal). Después del entrenamiento se ha transformado la salida para integrar la activación softmax y obtener las probabilidades de múltiples imágenes.

Ambos resultados son guardados en la base de datos SQL de Render y son mostrados en la aplicación web para el uso del radiólogo o médico y tenga completa visibilidad a las probabilidades y tenga la posibilidad de tomar una mejor decisión en el orden de atención de los pacientes.

df_resnet50.head()					
	actual	diagnosis	error	predicted	pred_diagnosis
0	1	pneumonia	0.044644	0.96	pneumonia
1	1	pneumonia	0.007706	0.99	pneumonia
2	0	normal	0.002187	0.00	normal
3	1	pneumonia	0.000532	1.00	pneumonia
4	1	pneumonia	0.018100	0.98	pneumonia

df_alexnet.head()					
	actual	diagnosis	error	predicted	pred_diagnosis
0	1	pneumonia	0.014450	0.99	pneumonia
1	1	pneumonia	0.001731	1.00	pneumonia
2	0	normal	0.001413	0.00	normal
3	1	pneumonia	0.001197	1.00	pneumonia
4	1	pneumonia	0.009361	0.99	pneumonia

df_vgg16.head()					
	actual	diagnosis	error	predicted	pred_diagnosis
0	1	pneumonia	0.012884	0.99	pneumonia
1	1	pneumonia	0.008451	0.99	pneumonia
2	0	normal	0.046550	0.05	normal
3	1	pneumonia	0.001771	1.00	pneumonia
4	1	pneumonia	0.102325	0.90	pneumonia

Figura 46: Predicciones del sistema con visibilidad de probabilidades. Elaboración propia.

Debido a que todos los datos guardados pueden ser convertido a tablas usando la biblioteca Pandas basada en Python, podemos fácilmente acceder a las probabilidades de cada paciente como se visualiza en Figura 46 donde la columna llamada predicted contiene valores de 0.00 al 1.00, más cerca al 0 son las probabilidades de que la imagen correspondiente a ese paciente tiene una diagnosis de normal basada en el algoritmo del sistema (ResNet-50, Alexnet o VGG-16), y entre más cercano a 1, la diagnosis es neumonía.

▪ **Situación POST Test**

Una vez implementado el sistema, el área de radiología notó una gran diferencia al poder tener acceso a las probabilidades de los pacientes esperando a su correspondiente análisis.

El sistema al mostrar las probabilidades de que cada uno de esos pacientes citados padezcan de neumonía o no, permitió que se estructure mejor la priorización de la atención a cada paciente y de esa manera contribuir a la disminución de los tiempos de espera más altos.

En la Figura 47 se muestran la reducción aproximada del número de horas máximas de tiempo de espera en promedio, por mes. Podemos ver que el sistema ayudo en disminuir este número de horas, de lo que originalmente se encontraban entre 4 a 6 horas, se disminuyó a entre 2 a casi 5 horas. Esta disminución es un resultado muy importante para alcanzar el objetivo principal del estudio de acelerar el proceso de diagnóstico de neumonía en el área de radiología y ayudar a salvar vidas.

A continuación, en la Tabla 19 se puede apreciar el detalle de los 12 registros (correspondientes a los 12 meses del año 2020) de la muestra POST Test después de la implementación de la función de activación SoftMax que permitió la visibilidad al estado de los pacientes en referencia a si padecen de neumonía o no y de esa manera disminuir los tiempos de espera.

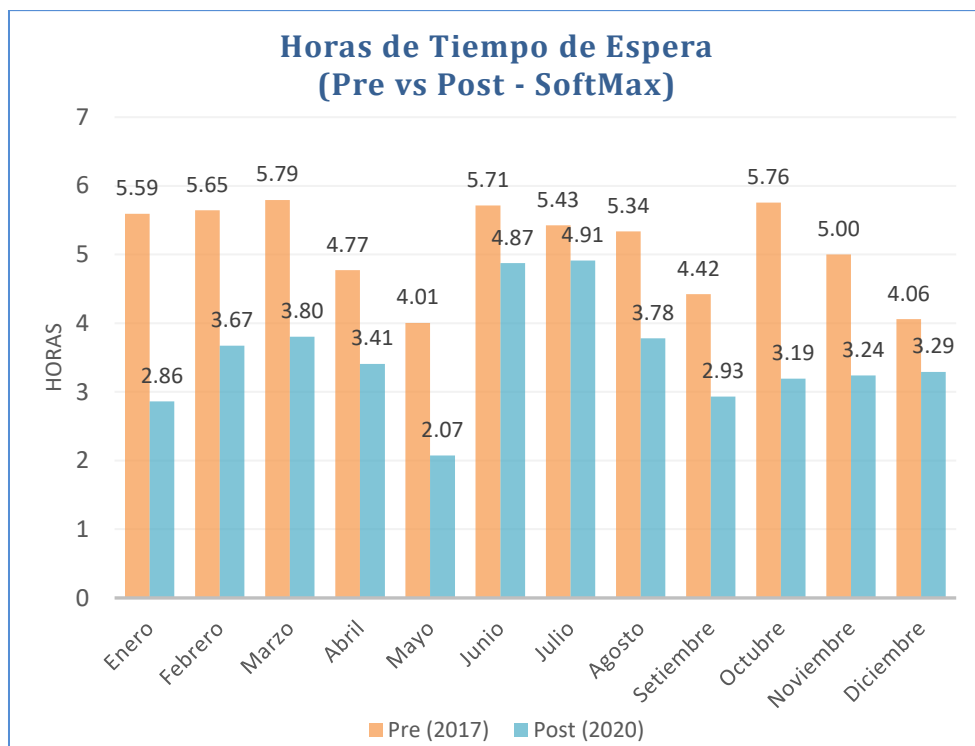


Figura 47: Reducción de Horas de Espera en pacientes después de implementar el sistema con la función de activación SoftMax. Elaboración propia.

Tabla 19:

Muestra Post Test Horas de Tiempo de Espera

Mes	Muestra Post Test
Enero	2.86
Febrero	3.67
Marzo	3.80
Abril	3.41
Mayo	2.07
Junio	4.87
Julio	4.91
Agosto	3.78
Setiembre	2.93
Octubre	3.19
Noviembre	3.24
Diciembre	3.29
TOTAL	3.5017

Fuente: SAP Hana DB. Área de Radiología. Elaboración propia.

4.2. Análisis de resultados

Generalidades

En esta sección se presentan los planteamientos y los resultados de las pruebas de normalidad y de las pruebas de hipótesis de esta investigación, donde se expone el detalle de la información levantada de las muestras en situación pre test y en situación post test, de manera que se pueda comprobar y verificar el contraste de las muestras, a través del análisis de la estadística inferencial planteadas en la investigación para cada una de las hipótesis específicas.

Para todos los resultados de las pruebas se ha utilizado el software estadístico IBM SPSS Statistics en su versión 26. Este software ofrece análisis estadístico avanzado, una amplia biblioteca de algoritmos de machine learning, análisis de texto, extensibilidad de código abierto, integración con big data y un fácil despliegue en las aplicaciones.

Su facilidad de uso, flexibilidad y escalabilidad hacen que SPSS sea accesible para usuarios con cualquier nivel de conocimiento. Además, es adecuado para proyectos de todos los tamaños y niveles de complejidad, y puede ayudar a los investigadores a encontrar nuevas oportunidades, mejorar la eficiencia y minimizar el riesgo.

✓ Prueba de Normalidad

Para las pruebas de normalidad se plantean las siguientes hipótesis:

H₀: Hipótesis Nula – Los datos de la muestra, SI son paramétricos, siguen una distribución normal

H₁: Hipótesis Alterna – Los datos de la muestra, NO son paramétricos, no siguen una distribución normal

Nivel de significancia: Sig. = 0.05

Regla de decisión:

- Si el nivel de significancia Sig. resulta ser un valor mayor o igual al 5,00% (Sig. $\geq 0,05$), entonces, se acepta la hipótesis nula (H_0)

Por lo tanto, los datos de la muestra, SI son paramétricos, siguen una distribución normal.

- Si el nivel de significancia Sig. resulta ser un valor menor al 5,00% (Sig. $< 0,05$), entonces, se acepta la hipótesis alterna (H_1)

Por lo tanto, los datos de la muestra, NO son paramétricos, no siguen una distribución normal.

✓ **Prueba de Hipótesis**

Para la contrastación de hipótesis se plantea la siguiente validez de la hipótesis:

H₀: Hipótesis Nula – NO existe diferencia estadística significativa entre la muestra Pre-Test y la muestra Post Test

H₁: Hipótesis Alterna – SI existe diferencia estadística significativa entre la muestra Pre-Test y la muestra Post Test (Hipótesis del investigador)

Nivel de significancia: Sig. = 0.05

Regla de decisión:

- Si el nivel de significancia Sig. resulta ser un valor mayor o igual al 5,00% (Sig. $\geq 0,05$), entonces, se acepta la hipótesis nula (H_0), o lo que es lo mismo, se rechaza la hipótesis del investigador.

Por lo tanto: NO se aplica la Variable Independiente (Variable Teórica) del investigador

- Si el nivel de significancia Sig. resulta ser un valor menor al 5,00% (Sig. < 0,05), entonces, se acepta la hipótesis alterna (H_1), o lo que es lo mismo, se acepta la hipótesis del investigador.

Por lo tanto: SI se aplica la Variable Independiente (Variable Teórica) del investigador

- **Hipótesis específica 01:** Si se implementa el Algoritmo ResNet-50, entonces se reducirá el número de diagnósicos incorrectos en el área de radiología.

✓ **Pruebas de Normalidad**

- **Muestra Pre Test y Post Test:**

De acuerdo a lo descrito en el punto 3.2 las muestras para la variable Algoritmo ResNet-50 son relacionadas por ser los mismos pacientes que han sido analizados.

Consta de un total de 12 datos de porcentaje de diagnósicos incorrectos en promedio, en la muestra antes (Pre Test) y en la muestra después (Post Test), de aplicar la variable independiente en la investigación para esta primera hipótesis específica. Ver Tabla 20.

Tabla 20:
Muestra Pre Test y Post Test de porcentaje de diagnósicos incorrectos

Porcentaje de diagnósicos incorrectos en promedio		
MES	Muestra Pre Test 2017	Muestra Post Test 2020
Enero	0.2857	0.0408
Febrero	0.2784	0.0825
Marzo	0.2165	0.0515
Abril	0.2371	0.0412
Mayo	0.2680	0.0412
Junio	0.2680	0.0206
Julio	0.2474	0.0722
Agosto	0.2371	0.0412
Setiembre	0.2062	0.0515
Octubre	0.2680	0.0722
Noviembre	0.2577	0.0722
Diciembre	0.2039	0.0680
TOTAL	0.2478	0.0546

Fuente: SAP HANA. Empresa de la investigación
Elaboración propia

▪ **Prueba Pre Test y Post Test**

En el cuadro de resumen de procesamiento de casos, obtenido mediante el software IBM SPSS Versión 26, se verifica que, del total de 12 muestras procesadas, el 100% han sido validadas, es decir, no hubo ningún dato perdido. Ver Tabla 21.

Tabla 21:

Resumen de procesamiento de datos –porcentaje de diagnosis incorrectas muestras Pre Test y Post Test

Resumen de procesamiento de casos						
	Casos				Total	
	Válido		Perdidos		N	Porcentaje
	N	Porcentaje	N	Porcentaje		
% Diagnosis Incorrectas - PRE	12	100,0%	0	0,0%	12	100,0%
% Diagnosis Incorrectas - POST	12	100,0%	0	0,0%	12	100,0%

Fuente: IBM SPSS Versión 26

Estadísticos descriptivos

Con los estadísticos descriptivos podemos contar con un resumen conciso de los datos para poder analizarlos por tendencia central o por dispersión. Ver Tabla 22.

Tabla 22:

Estadísticas de grupo – Muestras pre y post test

			Descriptivos	
			Estadístico	Error estándar
% Diagnosis Incorrectas - PRE	Media		,247844	,0080564
	95% de intervalo de confianza para la media	Límite inferior	,230112	
		Límite superior	,265577	
	Media recortada al 5%		,248183	
	Mediana		,252577	
	Varianza		,001	
	Desviación estándar		,0279082	
	Mínimo		,2039	
	Máximo		,2857	
	Rango		,0818	
% Diagnosis Incorrectas - POST	Media		,054597	,0053514
	95% de intervalo de confianza para la media	Límite inferior	,042819	
		Límite superior	,066376	

Descriptivos			
		Estadístico	Error estándar
	Media recortada al 5%	,054936	
	Mediana	,051546	
	Varianza	,000	
	Desviación estándar	,0185378	
	Mínimo	,0206	
	Máximo	,0825	
	Rango	,0619	

Fuente: IBM SPSS Versión 26

De la Tabla 22, podemos ver que se ha obtenido las medidas de tendencia central, así como, como medidas de dispersión, para las muestras Pre Test y Post Test.

- Muestra Pre Test:
 - o Media: 0,247844
 - o Mediana: 0,252577
 - o Varianza: 0,001
 - o Desviación estándar: 0,0279082

- Muestra Post Test
 - o Media: 0,054597
 - o Mediana: 0,051546
 - o Varianza: 0,000
 - o Desviación estándar: 0,0185378

Prueba de normalidad

Por la cantidad de datos que tenemos (12 datos) en Pre Test y Post Test respectivamente, las muestras son sometidas a la prueba de normalidad de Shapiro - Wilk a través del programa software IBM SPSS Versión 26, a fin de verificar si la distribución es normal, es decir, si es paramétrica. Ver Tabla 23.

Tabla 23:

Prueba de Normalidad para porcentaje de diagnosis incorrectas de las muestras Pre Test y Post Test

Pruebas de normalidad	
-----------------------	--

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
% Diagnóstico Incorrectas - PRE	,182	12	,200*	,924	12	,320
% Diagnóstico Incorrectas - POST	,181	12	,200*	,919	12	,274

*. Esto es un límite inferior de la significación verdadera.
a. Corrección de significación de Lilliefors
Fuente: IBM SPSS Versión 26

De acuerdo a los resultados obtenidos en la prueba de normalidad de Shapiro - Wilk podemos determinar que:

- Para las muestras Pre Test y Post Test del porcentaje de diagnóstico incorrectas en el presente estudio, los valores de la Sig. son: 0,320 y 0,274 respectivamente
- Estos valores son mayores que el valor de la significancia 0,05, de modo que, se acepta la Hipótesis Nula, con lo cual se concluye que los datos de la muestra Pre Test y Post Test provienen de una distribución normal.

✓ **Prueba de Hipótesis**

H₀: Si se implementa el Algoritmo ResNet-50, entonces NO se reducirá el número de diagnóstico incorrectas.

H₁: Si se implementa el Algoritmo ResNet-50, entonces SI se reducirá el número de diagnóstico incorrectas.

▪ **Prueba de significancia**

Dado que los datos son de naturaleza numérica; de muestras relacionadas, debido a que es el mismo grupo de análisis para la muestra Pre Test y Post Test; y que, además, ambas muestras provienen de una distribución normal, se determinó utilizar la Prueba de T de Student de muestras emparejadas, la cual es una prueba de hipótesis que permite evaluar si en los resultados hay diferencia estadística de manera significativa respecto a sus medias.

T de Student de Muestras emparejadas

Para la prueba de T de Student de muestras emparejadas tenemos:

- Estadísticas de muestras emparejadas
- Correlaciones de muestras emparejadas
- Prueba de hipótesis de T de Student de muestras emparejadas

En las estadísticas de muestras emparejadas, se puede observar que la desviación estándar, error promedio, de la muestra Pre Test es 0,0080564 y de la muestra Post Test es 0,0053514. Ver Tabla 24.

Tabla 24:

Estadísticas de muestras emparejadas para porcentaje de diagnosis incorrectas en promedio

Estadísticas de muestras emparejadas					
		Media	N	Desv. Desviación	Desv. Error promedio
Par 1	% Diagnosi Incorrectas - PRE	,247844	12	,0279082	,0080564
	% Diagnosi Incorrectas - POST	,054597	12	,0185378	,0053514

Fuente: IBM SPSS Versión 26

En la prueba de hipótesis de T de Student de muestras emparejadas (ver Tabla 25).

Aunque se puede observar que la significancia Sig es de 0,000, debido a que el valor es tan bajo que no se puede visualizar en la tabla mostrada en SPSS, el valor completo es $1,5558 \times 10^{-8}$, el cual es menor que 0,05, por lo tanto, podemos concluir que se rechaza la hipótesis nula (H_0) y se acepta la hipótesis alterna (H_1)

Tabla 25:
Prueba de hipótesis de T de Student de muestras emparejadas para porcentaje de diagnosis incorrectas en promedio

Prueba de muestras emparejadas								
Diferencias emparejadas								
	Media	Desv. Desviación	Desv. Error promedio	95% de intervalo de confianza de la diferencia		t	gl	Sig. (bilateral)
				Inferior	Superior			
% Diagnosis Incorrectas PRE	,1932471	,0340009	,0098152	,1716439	,2148502	19,689	11	,000
% Diagnosis Incorrectas POST								

Fuente: IBM SPSS Versión 26

Dado que la significancia es igual a $1,5558 \times 10^{-8}$, menor que 0,05 y respetando el criterio de evaluación, se rechazó la hipótesis nula H_0 y se aceptó la hipótesis alterna H_1 , afirmando que existe una diferencia estadística significativa entre el porcentaje de diagnosis incorrectas en promedio pre test y post test respectivamente.

Por lo tanto, se llegó a concluir que: Si se implementa el Algoritmo ResNet-50, entonces se reducirá el número de diagnosis incorrectas.

Con lo cual, además, de todo lo antes expuesto se evidencia claramente que la implementación del algoritmo ResNet-50 (variable independiente), tuvo un efecto positivo y significativo en la reducción del número de diagnosis incorrectas en el área de radiología (variable dependiente).

- **Hipótesis específica 02:** Si se implementa el Algoritmo Alexnet, entonces se reducirá el número de diagnósicos incorrectos en el área de radiología.

✓ **Pruebas de Normalidad**

- **Muestra Pre Test y Post Test:**

De acuerdo a lo descrito en el punto 3.2 las muestras para la variable Algoritmo Alexnet son relacionadas por ser los mismos pacientes que han sido analizados.

Consta de un total de 12 datos de porcentaje de diagnósicos incorrectos en promedio, en la muestra antes (Pre Test) y en la muestra después (Post Test), de aplicar la variable independiente en la investigación para esta primera hipótesis específica. Ver Tabla 26.

Tabla 26:
Muestra Pre Test y Post Test de porcentaje de diagnósicos incorrectos

Porcentaje de diagnósicos incorrectos en promedio		
MES	Muestra Pre Test 2017	Muestra Post Test 2020
Enero	0.2857	0.0306
Febrero	0.2784	0.0722
Marzo	0.2165	0.0309
Abril	0.2371	0.0412
Mayo	0.2680	0.0412
Junio	0.2680	0.0103
Julio	0.2474	0.0722
Agosto	0.2371	0.0309
Setiembre	0.2062	0.0619
Octubre	0.2680	0.0722
Noviembre	0.2577	0.0619
Diciembre	0.2039	0.0485
TOTAL	0.2478	0.0478

Fuente: SAP HANA. Empresa de la investigación
Elaboración propia

▪ **Prueba Pre Test y Post Test**

En el cuadro de resumen de procesamiento de casos, obtenido mediante el software IBM SPSS Versión 26, se verifica que, del total de 12 muestras procesadas, el 100% han sido validadas, es decir, no hubo ningún dato perdido. Ver Tabla 27.

Tabla 27:

Resumen de procesamiento de datos –porcentaje de diagnosis incorrectas muestras Pre Test y Post Test

Resumen de procesamiento de casos						
	Casos				Total	
	Válido		Perdidos		N	Porcentaje
	N	Porcentaje	N	Porcentaje		
% Diagnosis Incorrectas - PRE	12	100,0%	0	0,0%	12	100,0%
% Diagnosis Incorrectas - POST	12	100,0%	0	0,0%	12	100,0%

Fuente: IBM SPSS Versión 26

Estadísticos descriptivos

Con los estadísticos descriptivos podemos contar con un resumen conciso de los datos para poder analizarlos por tendencia central o por dispersión. Ver Tabla 28.

Tabla 28:

Estadísticas de grupo – Muestras pre y post test

			Descriptivos	
			Estadístico	Error estándar
% Diagnosis Incorrectas - PRE	Media		,247844	,0080564
	95% de intervalo de confianza para la media	Límite inferior	,230112	
		Límite superior	,265577	
	Media recortada al 5%		,248183	
	Mediana		,252577	
	Varianza		,001	
	Desviación estándar		,0279082	
	Mínimo		,2039	
	Máximo		,2857	
	Rango		,0818	
% Diagnosis Incorrectas - POST	Media		,047833	,0058602
	95% de intervalo de confianza para la media	Límite inferior	,034935	
		Límite superior	,060732	
	Media recortada al 5%		,048566	

Descriptivos		
	Estadístico	Error estándar
Mediana	,044890	
Varianza	,000	
Desviación estándar	,0203002	
Mínimo	,0103	
Máximo	,0722	
Rango	,0619	

Fuente: IBM SPSS Versión 26

De la Tabla 28, podemos ver que se ha obtenido las medidas de tendencia central, así como, como medidas de dispersión, para las muestras Pre Test y Post Test.

- Muestra Pre Test:
 - o Media: 0,247844
 - o Mediana: 0,252577
 - o Varianza: 0,001
 - o Desviación estándar: 0,0279082

- Muestra Post Test
 - o Media: 0,047833
 - o Mediana: 0,044890
 - o Varianza: 0,000
 - o Desviación estándar: 0,0203002

Prueba de normalidad

Por la cantidad de datos que tenemos (12 datos) en Pre Test y Post Test respectivamente, las muestras son sometidas a la prueba de normalidad de Shapiro - Wilk a través del programa software IBM SPSS Versión 26, a fin de verificar si la distribución es normal, es decir, si es paramétrica. Ver Tabla 29.

Tabla 29:
Prueba de Normalidad para porcentaje de diagnosis incorrectas de las muestras Pre Test y Post Test

Pruebas de normalidad						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
% Diagnosis Incorrectas - PRE	,182	12	,200*	,924	12	,320
% Diagnosis Incorrectas - POST	,172	12	,200*	,915	12	,245

*. Esto es un límite inferior de la significación verdadera.

a. Corrección de significación de Lilliefors

Fuente: IBM SPSS Versión 26

De acuerdo a los resultados obtenidos en la prueba de normalidad de Shapiro - Wilk podemos determinar que:

- Para las muestras Pre Test y Post Test del porcentaje de diagnosis incorrectas en el presente estudio, los valores de la Sig. son: 0,320 y 0,245 respectivamente
- Estos valores son mayores que el valor de la significancia 0,05, de modo que, se acepta la Hipótesis Nula, con lo cual se concluye que los datos de la muestra Pre Test y Post Test provienen de una distribución normal.

✓ Prueba de Hipótesis

H₀: Si se implementa el Algoritmo Alexnet, entonces NO se reducirá el número de diagnosis incorrectas.

H₁: Si se implementa el Algoritmo Alexnet, entonces SI se reducirá el número de diagnosis incorrectas.

▪ Prueba de significancia

Dado que los datos son de naturaleza numérica; de muestras relacionadas, debido a que es el mismo grupo de análisis para la muestra Pre Test y Post Test; y que, además, ambas muestras provienen de una distribución normal.

Se determinó utilizar la Prueba de T de Student de muestra emparejadas, la cual es una prueba de hipótesis que permite evaluar si en los resultados hay diferencia estadística de manera significativa respecto a sus medias.

T de Student de Muestras emparejadas

Para la prueba de T de Student de muestras emparejadas tenemos:

- Estadísticas de muestras emparejadas
- Correlaciones de muestras emparejadas
- Prueba de hipótesis de T de Student de muestras emparejadas

En las estadísticas de muestras emparejadas, se puede observar que la desviación estándar, error promedio, de la muestra Pre Test es 0,0080564 y de la muestra Post Test es 0,0058602. Ver Tabla 30.

Tabla 30:

Estadísticas de muestras emparejadas para porcentaje de diagnósis incorrectas en promedio

Estadísticas de muestras emparejadas					
		Media	N	Desv. Desviación	Desv. Error promedio
Par 2	% Diagnósis Incorrectas - PRE	,247844	12	,0279082	,0080564
	% Diagnósis Incorrectas - POST	,047833	12	,0203002	,0058602

Fuente: IBM SPSS Versión 26

En la prueba de hipótesis de T de Student de muestras emparejadas (ver Tabla 31). Aunque se puede observar que la significancia Sig es de 0,000, debido a que el valor es tan bajo que no se puede visualizar en la tabla mostrada en SPSS, el valor completo es $5,1867 \times 10^{-10}$, el cual es menor que 0,05.

Por lo tanto, podemos concluir que se rechaza la hipótesis nula (H_0) y se acepta la hipótesis alterna (H_1)

Tabla 31:
Prueba de hipótesis de T de Student de muestras emparejadas para porcentaje de diagnosis incorrectas en promedio

		Prueba de muestras emparejadas							
		Diferencias emparejadas					t	gl	Sig. (bilateral)
		Media	Desv. Desviación	Desv. Error promedio	95% de intervalo de confianza de la diferencia				
					Inferior	Superior			
	% Diagnosis Incorrectas PRE								
Par 2	% Diagnosis Incorrectas POST	,2000110	,0345482	,0099732	,1780601	,2219620	20,055	11	,000

Fuente: IBM SPSS Versión 26

Debido a que el valor de la significancia es $5,1867 \times 10^{-10}$, el cual es menor que 0,05, y respetando el criterio de evaluación, se rechazó la hipótesis nula H_0 y se aceptó la hipótesis alterna H_1 , afirmando que existe una diferencia estadística significativa entre el porcentaje de diagnosis incorrectas en promedio pre test y post test respectivamente.

Por lo tanto, se llegó a concluir que: Si se implementa el Algoritmo Alexnet, entonces se reducirá el número de diagnosis incorrectas.

Con lo cual, además, de todo lo antes expuesto se evidencia claramente que la implementación del algoritmo Alexnet (variable independiente), tuvo un efecto positivo y significativo en la reducción del número de diagnosis incorrectas en el área de radiología (variable dependiente).

- **Hipótesis específica 03:** Si se implementa el Algoritmo VGG, entonces se reducirá el número de diagnósicos incorrectos en el área de radiología.

✓ **Pruebas de Normalidad**

- **Muestra Pre Test y Post Test:**

De acuerdo a lo descrito en el punto 3.2 las muestras para la variable Algoritmo VGG son relacionadas por ser los mismos pacientes que han sido analizados.

Consta de un total de 12 datos de porcentaje de diagnósicos incorrectos en promedio, en la muestra antes (Pre Test) y en la muestra después (Post Test), de aplicar la variable independiente en la investigación para esta primera hipótesis específica. Ver Tabla 32.

Tabla 32:
Muestra Pre Test y Post Test de porcentaje de diagnósicos incorrectos

Porcentaje de diagnósicos incorrectos en promedio		
MES	Muestra Pre Test 2017	Muestra Post Test 2020
Enero	0.2857	0.0408
Febrero	0.2784	0.0619
Marzo	0.2165	0.0722
Abril	0.2371	0.0309
Mayo	0.2680	0.0515
Junio	0.2680	0.0103
Julio	0.2474	0.0515
Agosto	0.2371	0.0515
Setiembre	0.2062	0.0722
Octubre	0.2680	0.0825
Noviembre	0.2577	0.0206
Diciembre	0.2039	0.0971
TOTAL	0.2478	0.0536

Fuente: SAP HANA. Empresa de la investigación
Elaboración propia

▪ **Prueba Pre Test y Post Test**

En el cuadro de resumen de procesamiento de casos, obtenido mediante el software IBM SPSS Versión 26, se verifica que, del total de 12 muestras procesadas, el 100% han sido validadas, es decir, no hubo ningún dato perdido. Ver Tabla 33.

Tabla 33:

Resumen de procesamiento de datos –porcentaje de diagnosis incorrectas muestras Pre Test y Post Test

Resumen de procesamiento de casos						
	Casos					
	Válido		Perdidos		Total	
	N	Porcentaje	N	Porcentaje	N	Porcentaje
% Diagnosis Incorrectas - PRE	12	100,0%	0	0,0%	12	100,0%
% Diagnosis Incorrectas - POST	12	100,0%	0	0,0%	12	100,0%

Fuente: IBM SPSS Versión 26

Estadísticos descriptivos

Con los estadísticos descriptivos podemos contar con un resumen conciso de los datos para poder analizarlos por tendencia central o por dispersión. Ver Tabla 34.

Tabla 34:

Estadísticas de grupo – Muestras pre y post test

			Descriptivos	
			Estadístico	Error estándar
% Diagnosis Incorrectas - PRE	Media		,247844	,0080564
	95% de intervalo de confianza para la media	Límite inferior	,230112	
		Límite superior	,265577	
	Media recortada al 5%		,248183	
	Mediana		,252577	
	Varianza		,001	
	Desviación estándar		,0279082	
	Mínimo		,2039	
	Máximo		,2857	
	Rango		,0818	
% Diagnosis Incorrectas - POST	Media		,053588	,0073517
	95% de intervalo de confianza para la media	Límite inferior	,037407	
		Límite superior	,069769	
	Media recortada al 5%		,053576	

Descriptivos		
	Estadístico	Error estándar
Mediana	,051546	
Varianza	,001	
Desviación estándar	,0254671	
Mínimo	,0103	
Máximo	,0971	
Rango	,0868	

Fuente: IBM SPSS Versión 26

De la Tabla 34, podemos ver que se ha obtenido las medidas de tendencia central, así como, como medidas de dispersión, para las muestras Pre Test y Post Test.

– Muestra Pre Test:

- Media: 0,247844
- Mediana: 0,252577
- Varianza: 0,001
- Desviación estándar: 0,0279082

– Muestra Post Test

- Media: 0,053588
- Mediana: 0,051546
- Varianza: 0,001
- Desviación estándar: 0,0254671

Prueba de normalidad

Por la cantidad de datos que tenemos (12 datos) en Pre Test y Post Test respectivamente.

Las muestras son sometidas a la prueba de normalidad de Shapiro - Wilk a través del programa software IBM SPSS Versión 26, a fin de verificar si la distribución es normal, es decir, si es paramétrica. Ver Tabla 35.

Tabla 35:
Prueba de Normalidad para porcentaje de diagnosis incorrectas de las muestras Pre Test y Post Test

Pruebas de normalidad						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
% Diagnosis Incorrectas - PRE	,182	12	,200*	,924	12	,320
% Diagnosis Incorrectas - POST	,135	12	,200*	,982	12	,989

*. Esto es un límite inferior de la significación verdadera.

a. Corrección de significación de Lilliefors

Fuente: IBM SPSS Versión 26

De acuerdo a los resultados obtenidos en la prueba de normalidad de Shapiro - Wilk podemos determinar que:

- Para las muestras Pre Test y Post Test del porcentaje de diagnosis incorrectas en el presente estudio, los valores de la Sig. son: 0,320 y 0,989 respectivamente
- Estos valores son mayores que el valor de la significancia 0,05, de modo que, se acepta la Hipótesis Nula, con lo cual se concluye que los datos de la muestra Pre Test y Post Test provienen de una distribución normal.

✓ Prueba de Hipótesis

H₀: Si se implementa el Algoritmo VGG, entonces NO se reducirá el número de diagnosis incorrectas.

H₁: Si se implementa el Algoritmo VGG, entonces SI se reducirá el número de diagnosis incorrectas.

▪ Prueba de significancia

Dado que los datos son de naturaleza numérica; de muestras relacionadas, debido a que es el mismo grupo de análisis para la muestra Pre Test y Post Test; y que, además, ambas muestras provienen de una distribución normal, se determinó utilizar la Prueba de T de Student de

muestra emparejadas, la cual es una prueba de hipótesis que permite evaluar si en los resultados hay diferencia estadística de manera significativa respecto a sus medias.

T de Student de Muestras emparejadas

Para la prueba de T de Student de muestras emparejadas tenemos:

- Estadísticas de muestras emparejadas
- Correlaciones de muestras emparejadas
- Prueba de hipótesis de T de Student de muestras emparejadas

En las estadísticas de muestras emparejadas, se puede observar que la desviación estándar, error promedio, de la muestra Pre Test es 0,0080564 y de la muestra Post Test es 0,0073517. Ver Tabla 36.

Tabla 36:

Estadísticas de muestras emparejadas para porcentaje de diagnosis incorrectas en promedio

Estadísticas de muestras emparejadas					
		Media	N	Desv. Desviación	Desv. Error promedio
Par 3	% Diagnosis Incorrectas - PRE	,247844	12	,0279082	,0080564
	% Diagnosis Incorrectas - POST	,053588	12	,0254671	,0073517

Fuente: IBM SPSS Versión 26

En la prueba de hipótesis de T de Student de muestras emparejadas (ver Tabla 37), aunque se puede observar que la significancia Sig es de 0,000, debido a que el valor es tan bajo que no se puede visualizar en la tabla mostrada en SPSS, el valor completo es $6,314 \times 10^{-10}$, el cual es menor que 0,05, por lo tanto, podemos concluir que se rechaza la hipótesis nula (H_0) y se acepta la hipótesis alterna (H_1)

Tabla 37:

Prueba de hipótesis de T de Student de muestras emparejadas para porcentaje de diagnosis incorrectas en promedio

Prueba de muestras emparejadas	
--------------------------------	--

	Diferencias emparejadas						t	gl	Sig. (bilateral)
	Media	Desv. Desviación	Desv. Error promedio	95% de intervalo de confianza de la diferencia					
				Inferior	Superior				
% Diagnóstico Incorrectas PRE	,1942563	,0462068	,0133388	,1648979	,2236147	14,563	11	,000	
% Diagnóstico Incorrectas POST									

Fuente: IBM SPSS Versión 26

Dado que la significancia es igual a $6,314 \times 10^{-10}$, menor que 0,05 y respetando el criterio de evaluación, se rechazó la hipótesis nula H_0 y se aceptó la hipótesis alterna H_1 , afirmando que existe una diferencia estadística significativa entre el porcentaje de diagnósticos incorrectos en promedio pre test y post test respectivamente.

Por lo tanto, se llegó a concluir que: Si se implementa el Algoritmo VGG, entonces se reducirá el número de diagnósticos incorrectos.

Con lo cual, además, de todo lo antes expuesto se evidencia claramente que la implementación del algoritmo VGG (variable independiente), tuvo un efecto positivo y significativo en la reducción del número de diagnósticos incorrectos en el área de radiología (variable dependiente).

- **Hipótesis específica 04:** Si se implementa la Función de Activación SoftMax, entonces se reducirá el tiempo de espera de pacientes con neumonía.

✓ **Pruebas de Normalidad**

- **Muestra Pre Test y Post Test:**

De acuerdo a lo descrito en el punto 3.2 las muestras para la variable Función de Activación SoftMax son relacionadas por ser los mismos pacientes que han sido analizados.

Consta de un total de 12 datos de porcentaje de diagnosis incorrectas en promedio, en la muestra antes (Pre Test) y en la muestra después (Post Test), de aplicar la variable independiente en la investigación para esta primera hipótesis específica. Ver Tabla 38.

Tabla 38:

Muestra Pre Test y Post Test de horas de tiempo de espera

Horas al día de Tiempo de Espera en promedio		
MES	Muestra Pre Test 2017	Muestra Post Test 2020
Enero	5.59	2.86
Febrero	5.65	3.67
Marzo	5.79	3.80
Abril	4.77	3.41
Mayo	4.01	2.07
Junio	5.71	4.87
Julio	5.43	4.91
Agosto	5.34	3.78
Setiembre	4.42	2.93
Octubre	5.76	3.19
Noviembre	5.00	3.24
Diciembre	4.06	3.29
TOTAL	5.1275	3.5017

Fuente: SAP HANA. Empresa de la investigación
Elaboración propia

▪ **Prueba Pre Test y Post Test**

En el cuadro de resumen de procesamiento de casos, obtenido mediante el software IBM SPSS Versión 26, se verifica que, del total de 12 muestras procesadas, el 100% han sido validadas, es decir, no hubo ningún dato perdido. Ver Tabla 39.

Tabla 39:

Resumen de procesamiento de datos – horas de tiempo de espera muestras Pre Test y Post Test

Resumen de procesamiento de casos						
	Casos					
	Válido		Perdidos		Total	
	N	Porcentaje	N	Porcentaje	N	Porcentaje
horas tiempo de espera - PRE	12	100,0%	0	0,0%	12	100,0%
horas tiempo de espera - POST	12	100,0%	0	0,0%	12	100,0%

Fuente: IBM SPSS Versión 26

Estadísticos descriptivos

Con los estadísticos descriptivos podemos contar con un resumen conciso de los datos para poder analizarlos por tendencia central o por dispersión. Ver Tabla 40.

Tabla 40:

Estadísticas de grupo – Muestras pre y post test

			Descriptivos	
			Estadístico	Error estándar
horas tiempo de espera - PRE	Media		5,1273	,19143
	95% de intervalo de confianza para la media	Límite inferior	4,7060	
		Límite superior	5,5487	
	Media recortada al 5%		5,1526	
	Mediana		5,3810	
	Varianza		,440	
	Desviación estándar		,66312	
	Mínimo		4,01	
	Máximo		5,79	
	Rango		1,79	
	95% de intervalo de confianza para la media	Media		3,5039
Límite inferior			2,9948	
Límite superior			4,0130	

Descriptivos			
		Estadístico	Error estándar
horas tiempo de espera - POST	Media recortada al 5%	3,5050	
	Mediana	3,3497	
	Varianza	,642	
	Desviación estándar	,80125	
	Mínimo	2,07	
	Máximo	4,91	
	Rango	2,84	

Fuente: IBM SPSS Versión 26

De la Tabla 40, podemos ver que se ha obtenido las medidas de tendencia central, así como, como medidas de dispersión, para las muestras Pre Test y Post Test.

- Muestra Pre Test:
 - o Media: 5,1273
 - o Mediana: 5,3810
 - o Varianza: 0,440
 - o Desviación estándar: 0,66312

- Muestra Post Test
 - o Media: 3,5039
 - o Mediana: 3,3497
 - o Varianza: 0,642
 - o Desviación estándar: 0,80125

Prueba de normalidad

Por la cantidad de datos que tenemos (12 datos) en Pre Test y Post Test respectivamente, las muestras son sometidas a la prueba de normalidad de Shapiro - Wilk a través del programa software IBM SPSS Versión 26, a fin de verificar si la distribución es normal, es decir, si es paramétrica. Ver Tabla 41.

Tabla 41:
Prueba de Normalidad para horas de tiempo de espera de las muestras Pre Test y Post Test

Pruebas de normalidad						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
horas tiempo de espera - PRE	,207	12	,166	,862	12	,052
horas tiempo de espera - POST	,188	12	,200*	,931	12	,394

*. Esto es un límite inferior de la significación verdadera.

a. Corrección de significación de Lilliefors

Fuente: IBM SPSS Versión 26

De acuerdo a los resultados obtenidos en la prueba de normalidad de Shapiro - Wilk podemos determinar que:

- Para las muestras Pre Test y Post Test del porcentaje de diagnosis incorrectas en el presente estudio, los valores de la Sig. son: 0,052 y 0,394 respectivamente
- Estos valores son mayores que el valor de la significancia 0,05, de modo que, se acepta la Hipótesis Nula, con lo cual se concluye que los datos de la muestra Pre Test y Post Test provienen de una distribución normal.

✓ Prueba de Hipótesis

H₀: Si se implementa la Función de Activación SoftMax, entonces NO se reducirá el tiempo de espera de pacientes con neumonía.

H₁: Si se implementa la Función de Activación SoftMax, entonces SI se reducirá el tiempo de espera de pacientes con neumonía.

▪ Prueba de significancia

Dado que los datos son de naturaleza numérica; de muestras relacionadas, debido a que es el mismo grupo de análisis para la muestra Pre Test y Post Test; y que, además, ambas muestras provienen de una distribución normal, se determinó utilizar la Prueba de T de Student de

muestra emparejadas, la cual es una prueba de hipótesis que permite evaluar si en los resultados hay diferencia estadística de manera significativa respecto a sus medias.

T de Student de Muestras emparejadas

Para la prueba de T de Student de muestras emparejadas tenemos:

- Estadísticas de muestras emparejadas
- Correlaciones de muestras emparejadas
- Prueba de hipótesis de T de Student de muestras emparejadas

En las estadísticas de muestras emparejadas, se puede observar que la desviación estándar, error promedio, de la muestra Pre Test es 0,19143 y de la muestra Post Test es 0,23130. Ver Tabla 42.

Tabla 42:

Estadísticas de muestras emparejadas para horas de tiempo de espera en promedio

Estadísticas de muestras emparejadas					
		Media	N	Desv. Desviación	Desv. Error promedio
Par 4	horas tiempo de espera - PRE	5,1273	12	,66312	,19143
	horas tiempo de espera - POST	3,5039	12	,80125	,23130

Fuente: IBM SPSS Versión 26

En la prueba de hipótesis de T de Student de muestras emparejadas (ver Tabla 43), aunque se puede observar que la significancia Sig es de 0,000, debido a que el valor es tan bajo que no se puede visualizar en la tabla mostrada en SPSS, el valor completo es $5,00 \times 10^{-6}$, el cual es menor que 0,05, por lo tanto, podemos concluir que se rechaza la hipótesis nula (H_0) y se acepta la hipótesis alterna (H_1)

Tabla 43:
Prueba de hipótesis de T de Student de muestras emparejadas para horas de tiempo de espera en promedio

Prueba de muestras emparejadas								
	Diferencias emparejadas					t	gl	Sig. (bilateral)
	Media	Desv. Desviación	Desv. Error promedio	95% de intervalo de confianza de la diferencia				
				Inferior	Superior			
horas tiempo de espera PRE	1,62346	,68430	,19754	1,18867	2,05824	8,218	11	,000
horas tiempo de espera POST								

Fuente: IBM SPSS Versión 26

Dado que la significancia es igual a $5,00 \times 10^{-6}$, menor que 0,05 y respetando el criterio de evaluación, se rechazó la hipótesis nula H_0 y se aceptó la hipótesis alterna H_1 , afirmando que existe una diferencia estadística significativa entre el porcentaje de diagnósticos incorrectos en promedio pre test y post test respectivamente.

Por lo tanto, se llegó a concluir que: Si se implementa la función de activación SoftMax, entonces se reducirá el tiempo de espera de pacientes con neumonía.

Con lo cual, además, de todo lo antes expuesto se evidencia claramente que la implementación de la función de activación SoftMax (variable independiente), tuvo un efecto positivo y significativo en la reducción del tiempo de espera de pacientes con neumonía (variable dependiente).

✓ **Resumen de resultados**

Líneas abajo observamos el resumen de los resultados mostrados en esta investigación. Ver Tabla 44.

Tabla 44:
Resumen de resultados

Hipótesis Específica	VARIABLES Independiente	VARIABLES Dependiente	Indicador	Pre-Test	Post-Test	Diferencia
1	Algoritmo ResNet-50	Número de diagnosing incorrectas	Porcentaje de diagnosing incorrectas al mes	0,2478	0,0546	Disminuyó 0,1932 77,97 %
2	Algoritmo Alexnet	Número de diagnosing incorrectas	Porcentaje de diagnosing incorrectas al mes	0,2478	0,0478	Disminuyó 0,2000 80,70%
3	Algoritmo VGG	Número de diagnosing incorrectas	Porcentaje de diagnosing incorrectas al mes	0,2478	0,0536	Disminuyó 0,1943 78,38%
4	Función de Activación SoftMax	Tiempo de espera de pacientes	Promedio mensual de horas de espera en un día	5,1275	3,5017	Disminuyó 1,6258 31.,71%

Elaboración: Propia

- ✓ Aunque el programa del SPSS muestra los valores de las significancias como solo 0,000, los valores completos si muestran que hay un orden de significancia en los tres algoritmos comparados. Y por lo tanto fue el Algoritmo Alexnet (segunda hipótesis) el cual obtuvo la mayor precisión al obtener una significancia de $5,1867 \times 10^{-10}$, seguido por el Algoritmo VGG (tercera hipótesis) con una significación de $6,314 \times 10^{-10}$ y seguido por el Algoritmo ResNet-50 (primera hipótesis) con una significancia de $1,5558 \times 10^{-8}$.
- ✓ En la primera hipótesis se puede ver la disminución del 77,97 % del porcentaje de diagnosing incorrectas al mes al implementar el algoritmo

ResNet-50, es decir, el sistema basado en ResNet-50 mejora la precisión del diagnóstico de neumonía.

- ✓ En la segunda hipótesis se puede ver la disminución del 80,70% del porcentaje de diagnosis incorrectas al mes al implementar el algoritmo Alexnet, es decir, el sistema basado en Alexnet mejora la precisión del diagnóstico.
- ✓ En la tercera hipótesis se puede ver la disminución del 78,38% del porcentaje de diagnosis incorrectas al mes al implementar el algoritmo VGG, es decir, el sistema basado en VGG mejora la precisión del diagnóstico.
- ✓ En la cuarta hipótesis se puede ver la disminución del 31,71% de horas de tiempo espera al implementar la función de activación SoftMax, es decir, el sistema diagnóstico de neumonía que además integra un nodo de activación SoftMax permite acceder a las probabilidades de que cada paciente padezca de neumonía, ayudando a la priorización de pacientes en un día.

Tabla 45:
Tabla del valor t de cada algoritmo

	Media	Desv. Desviación	Desv. Error promedio	95% de intervalo de confianza de la diferencia		t
				Inferior	Superior	
Algoritmo ResNet-50	0,1932471	0,0340009	0,0098152	0,1716439	0,2148502	19,689
Algoritmo Alexnet	0,2000110	0,0345482	0,0099732	0,1780601	0,2219620	20,055
Algoritmo VGG	0,1942563	0,0462068	0,0133388	0,1648979	0,2236147	14,563

Elaboración: Propia

- ✓ Reforzando la idea de cuál es el algoritmo con mayor exactitud, podemos también ver Tabla 45 mostrando los valores t de cada uno de los tres algoritmos. El valor T es un valor estadístico que mide el tamaño de la diferencia en la relación con la variación en los datos. Cuanto mayor sea la magnitud de T, mayor es la evidencia en contra de la hipótesis nula. Es el

Algoritmo Alexnet (segunda hipótesis) el que obtuvo un valor t de 20,055, a comparación del Algoritmo VGG (tercera hipótesis) que obtuvo un valor t de 14,563 y del Algoritmo ResNet-50 (primera hipótesis) que obtuvo un valor t de 19,689.

CONCLUSIONES Y RECOMENDACIONES

✓ Conclusiones

1. La aplicación de un sistema de inteligencia artificial basado en redes neuronales convolucionales ha brindado un gran aporte al área de radiología frente a la creciente cantidad de pacientes que padecen de neumonía, no solo para confirmar la enfermedad de manera automática con alta precisión, sino también para poder descartar si el paciente no padece de neumonía con mayor velocidad.
2. Se encontró que esta investigación también ha contribuido a crear un sistema no solo aplicable a detectar la existencia de neumonía en pacientes, sino que también poder descartar la enfermedad, de esa manera el área de radiología rápidamente pudo referir a los pacientes que no padecían de neumonía a otras especialidades para comenzar a evaluar otras potenciales enfermedades respiratorias, resultando en el aumento de las probabilidades de encontrar el tratamiento apropiado para aquellos pacientes y con mayor rapidez.
3. La implementación de diversas arquitecturas de inteligencia artificial ha contribuido a la reducción de diagnósticos incorrectos en más de 80%, y cada algoritmo contribuyo en la precisión de los diagnósticos en niveles diferentes.
4. Se encontró que el Algoritmo Alexnet, es la arquitectura de redes neuronales convolucionales que logró reducir la mayor cantidad de diagnósticos incorrectos con una reducción total de 80,70%. Esto demuestra que es la simpleza de su arquitectura con la inclusión de capas ReLU que tuvo mayor facilidad para aprender de las imágenes, generalizar y encontrar los patrones que ayuden a clasificar si la imagen representa si el paciente padece de neumonía o no (con una exactitud de encima del 95%).
5. Además, continuando con Alexnet, debido a que es una red neuronal convolucional simple de ocho capas, se encontró que especialmente útil para el

despliegue de este modelo en dispositivos con limitaciones de hardware, como dispositivos móviles y dispositivos Edge. La arquitectura de Alexnet simplifica el tiempo de entrenamiento y esto permite que el sistema en conjunto obtenga una menor latencia.

6. En el caso de ResNet-50, aunque en este estudio no logró la exactitud más alta, si obtuvo una precisión muy alta. Usar ResNet-50 como motor cognitivo del sistema crea una solución más robusta debido a la profundidad de su arquitectura, haciéndola apropiada para el manejo de imágenes más complejas. Esto quiere decir que esta arquitectura es más útil para cuando se integren imágenes radiográficas de otros centros médicos cuyo incremento en complejidad se deba a la calidad de las nuevas imágenes, es decir imágenes que puedan incluir artefactos que obstruyan la visualización de patrones de neumonía.
7. En el caso de VGG, aunque no logró la más alta reducción de diagnosis incorrectas, sigue siendo una buena opción como arquitectura alterna en el motor cognitivo de nuestro sistema. Es decir, VGG es una arquitectura similar a Alexnet pero con mayor número de capas, esto convierte a VGG en el punto intermedio entre ResNet-50 y Alexnet, demostrando ser útil en casos en se necesite un modelo más robusto pero con una mínima cantidad de parámetros reduciendo el tiempo de entrenamiento.
8. El sistema propuesto es un sistema que integra todos los modelos pre-entrenados y a la vez incluye la implementación de una capa SoftMax que ha ayudado a brindar visibilidad a los especialistas radiólogos y médicos para poder priorizar a los casos más urgentes y reducir el tiempo de espera en pacientes que buscan ser evaluados rápidamente. Se redujo el tiempo máximo de espera en un 32%, o en promedio, 1,6 horas. Este tiempo ha sido valioso en detectar la enfermedad con más velocidad y brindar el tratamiento correspondiente.
9. Finalmente se concluye que, aunque el sistema implementado ha logrado una mejora en el proceso de diagnostico de neumonía, se encontró que sirve más como una herramienta de ayuda. Sin embargo, al integrar I.A., el sistema ha

obtenido la habilidad de mejorar por sí solo, es decir, el sistema mejora conforme encuentre nuevas imágenes de nuevos pacientes debido a que es justamente los datos de entrada los que adaptan el comportamiento (cálculo) de estas redes neuronales convolucionales.

Recomendaciones

1. En el estudio realizado se demostró que la integración del estado del arte en Inteligencia Artificial orientado a imágenes médicas contribuye a un mejor proceso de diagnóstico de neumonía. Sin embargo, una vez que se ha introducido elementos de I.A. en un sistema informático, se enfatiza la necesidad de una capacitación del personal del área de radiología. En términos de skill técnico se recomienda una capacitación en el área de T.I. para conocer los retos presentados en la operación de sistemas inteligentes, también conocido como MLOps. Y en el caso del personal médico también se recomienda una capacitación en como funciona el sistema en términos generales para reforzar la noción que este sistema es una herramienta de ayuda y no un reemplazo de especialistas. Esto es importante debido a la criticidad del diagnóstico de una enfermedad humana.
2. Como parte de la introducción de I.A. es la ventaja principal de que estos algoritmos estudiados aprenden por sí solos a encontrar patrones en las imágenes, en este caso el estudio se enfocó en el uso de imágenes radiográficas torácicas para la clasificación de la enfermedad respiratoria neumonía debido a la criticidad actual de esta enfermedad. Pero, se recomienda usar este mismo sistema para la investigación en otros procesos de diagnóstico similares, es decir, procesos donde el diagnóstico pueda ser inferido basado en imágenes médicas, por ejemplo, detección de cáncer piel usando imágenes dermatológicas, detección de tumores cerebrales usando tomografías y placas de resonancia magnéticas.
3. De implementarse lo recomendado en el punto 2, en caso los modelos utilizados en este estudio no logren una alta exactitud, se recomienda crear un ensamble de estos modelos, combinarlos, y compararlos con el desempeño de cada uno individualmente. En caso se proceda con un ensamble de modelos de redes neuronales convolucionales, esto introduciría a su vez una mayor complejidad del sistema en total y se recomienda que la infraestructura sea la adecuada antes del despliegue.

4. Se recomienda expandir este estudio al campo de compresión / aceleración de redes neuronales convolucionales. Esta solución implementada siempre necesita una etapa de entrenamiento que por lo tanto tomara un lapso de tiempo proporcional a la cantidad (y tipo) de imágenes que consuma.
5. Hoy en día con el uso de dispositivos móviles, Edge / IoT, se necesitan modelos de menor latencia, y para eso se deben implementar técnicas de compresión de los modelos, antes, durante y después del entrenamiento.
6. Las técnicas recomendadas a implementar del punto anterior, son conocidas como Pruning, que incluye:
 - ✓ Weight Pruning (podado de los pesos de la red neuronal) donde parámetros que no contribuyen al aprendizaje del modelo son desechados, y
 - ✓ Channel Pruning donde se eliminan grupos enteros de neuronas, capas y filtros, resultando en una reducción considerable de cálculos necesarios.

Ambas técnicas de pruning no solo reducen el tamaño de la red neuronal convolucional, sino que además la reducción de los cálculos permite a estos modelos ser usados en dispositivos / servidores con componentes de procesamiento limitado.

REFERENCIA

Bibliografía

- Aja-Fernandez, S., Garcia, R. d., Tao, D., & Li, X. (2009). *Tensors in Image Processing and Computer Vision*. London: Springer.
- Alcalde Rabanal, J., Lazo Gonzales, O., Macias, N., Contreras Ochoa, C., & Espinosa Henao, O. (2019). Sistema De Salud En El Perú: Situación Actual, Desafíos Y Perspectivas. *Revista Internacional de Salud Materno Fetal*, 4(3), 8-18.
- Álvarez Gutiérrez, F., Díaz Baquero, A., Medina Gallardo, J., & Romero Falcón, A. (2010). Neumonías adquiridas en la comunidad. *Medicine - Programa de Formación Médica Continuada Acreditado*, 10(67), 4573-4581. doi:10.1016/s0304-5412(10)70252-3
- Badgeley, M. A. (2019). *Deep Learning Multimodal para Mejorar la Práctica de la Radiología*. New York, United States: Icahn School of Medicine at Mount Sinai.
- Bressem, K. K., Adams, L., Erxleben, C., Hamm, B., Niehues, S., & Vahldiek, J. (2020). Comparing Different Deep Learning Architectures for Classification of Chest Radiographs. *Scientific Reports*, 10(1), 13590. doi:https://doi.org/10.1038/s41598-020-70479-z
- Chakraborty, B., Shaw, B., & Aich, J. (2018). Does Deeper Network Lead to Better Accuracy: A Case Study on Handwritten Devanagari Characters. *2018 13th IAPR International Workshop on Document Analysis Systems (DAS)* (págs. 411-416). Vienna: IEEE. doi:10.1109/DAS.2018.72
- Chollet, F. (21 de September de 2016). *Keras: the Python deep learning API*. Obtenido de <https://keras.io/>
- Elemraid, M. A., Muller, M., Spencer, D. A., Rushton, S. P., Gorton, R., Thomas, M. F., . . . Clark, J. E. (2014). Accuracy of the Interpretation of Chest Radiographs for the Diagnosis of Paediatric Pneumonia. *PLoS ONE*, 9(8), e106051. doi:doi.org/10.1371/journal.pone.0106051
- Elgendy, M. (2020). *Deep Learning for Vision Systems*. Shelter Island, New York: Manning Publications.
- Goel, A. (2019). *Render: Run your Deep Learning code in production*. Obtenido de <https://github.com/renderinc>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. Cambridge, MA, United States: MIT Press.



- Graber, M. L. (2013). The incidence of diagnostic error in medicine. *BMJ Quality & Safety*, 22(Suppl 2), ii21-ii27. doi:10.1136/bmjqs-2012-001615
- Hedrick, T. E., Bickman, L., & Rog, D. J. (1993). *Applied Research Design: A Practical Guide*. California: Sage Publications Ltd.
- Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. (12 de 09 de 2014). *Metodología de la investigación* (Quinta ed.). (M. G. S.A., Ed.) Mexico, Mexico: Mcgraw Hill.
- Howard, J. (2018). *FastAI: the deep learning library with with high-level components*. Obtenido de <https://github.com/fastai/fastai>
- Kandel, I., Castelli, M., & Popovic, A. (2020). Comparative Study of First Order Optimizers for Image Classification Using Convolutional Neural Networks on Histopathology Images. *Journal of Imaging*, 6(9), 92. doi:<https://doi.org/10.3390/jimaging6090092>
- Le, L., Wang, X., Carneiro, G., & Yang, L. (2019). *Deep Learning and Convolutional Neural Networks for Medical Imaging and Clinical Informatics*. Cham, Switzerland : Springer International Publishing.
- Martínez, M. (2004). *Los Grupos Focales de Discusión como Método de Investigación*. Caracas, Venezuela: Universidad Simón Bolívar.
- Nash, M. (2019). *Deep learning, computer-aided radiography reading for tuberculosis: a diagnostic accuracy study from a tertiary hospital in India*. Montreal, Canada: McGill University.
- Ñaupas, H., Mejía, E., Novoa, E., & Villagómez, A. (2014). *Metodología de la Investigación*. Colombia: Ediciones de la U.
- NG, K. (2018). *Machine Learning Projects for Mobile Applications*. Birmingham: Packt Publishing Ltd.
- Pang, S., Meng, F., Wang, X., Wang, J., Song, T., Wang, X., & Cheng, X. (2020). VGG16-T: A Novel Deep Convolutional Neural Network with Boosting to Identify Pathological Type of Lung Cancer in Early Stage by CT Images. *International Journal of Computational Intelligence Systems*, 13(1), 771. doi:10.2991/ijcis.d.200608.001
- Pino Gotuzzo, R. (2013). *Metodología de la Investigación(tercera edicion)*. Lima: San Marcos.
- Ponce de León, S., Rangel-Frausto, M. S., Elías-López, J., & Romero Oliveros, C. (1999). Infecciones nosocomiales: tendencias seculares de un programa de control en México. *Salud Publica de Mexico*, 5-11.
- Rajaraman, S., Candemir, S., Kim, I., Thoma, G., & Antani, S. (09 de 2018). Visualization and Interpretation of Convolutional Neural Network Predictions in Detecting Pneumonia in Pediatric Chest Radiographs. *Applied Sciences*, 8, 1715.

- Ranschaert, E. R., Morozov, S., & Algra, P. R. (2019). *Artificial Intelligence in Medical Imaging: Opportunities, Applications and Risks*. Cham, Switzerland: Springer International Publishing.
- Raschka, S., & Mirjalili, V. (2019). *Python Machine Learning*. Birmingham, United Kingdom: Packt.
- Sanchez Carlessi, H., & Reyes Meza, C. (2015). *Metodología y Diseños en la Investigación Científica*. Lima, Peru: Universidad Ricardo Palma.
- Sewak, M., Karim, R., & Pujari, P. (2018). *Practical Convolutional Neural Networks*. Birmingham, United Kingdom: Packt.
- Shanmugamani, R. (2018). *Deep Learning for Computer Vision*. Birmingham, UK: Packt Publishing.
- Syed, M. (2018). *Machine Learning in Healthcare: Identifying Pneumonia with Artificial Intelligence*. Helsinki, Finlandia: Helsinki Metropolia University of Applied Sciences. Obtenido de <https://www.theseus.fi/handle/10024/153415>
- Tursyngaliyeva, A. (2019). *Deep Learning for Medical Image Segmentation: Pneumonia Detection*. Astana, Kazakhstan: Nazarbayev University.
- UNICEF. (November de 2019). *Fighting for Breath Call to Action: End Childhood Pneumonia Deaths*. Obtenido de UNICEF: https://stopppneumonia.org/wp-content/uploads/2019/11/Fighting-for-Breath-briefing-8th-pp-low-res_rev-20-Nov.pdf
- Vázquez Enríquez, M. (2019). *A deep learning approach for pneumonia detection in chest X-ray*. Pontevedra, España: Universidad de Vigo.
- Venkatesan, R., & Li, B. (2018). *Convolutional Neural Networks in Visual Computing*. Boca Raton, United States: CRC Press.
- Watt, J., Borhani, R., & Katsaggelos, A. K. (2020). *Machine Learning Refined*. Cambridge, United Kingdom: Cambridge University Press.
- World Health Organization. (20 de 10 de 2021). *Pneumonia*. Obtenido de WHO Pneumonia Fact Sheets: <https://www.who.int/news-room/fact-sheets/detail/pneumonia>
- Yang, Q., Zhang, Y., Dai, W., & Pan, S. J. (2020). *Transfer Learning*. New York, USA: Cambridge University Press.
- Zhang, A., Lipton, Z. C., Li, M., & Smola, A. J. (2021). *Dive into Deep Learning*. Hong Kong: Amazon Science.

ANEXOS


Anexo 1: Declaración de Autenticidad

A continuación, se muestra el formato de autenticidad y no plagio.

	UNIVERSIDAD RICARDO PALMA	Escuela de Posgrado
DECLARACIÓN DE AUTENTICIDAD Y NO PLAGIO		
DECLARACIÓN DEL GRADUANDO		
Por el presente, el graduando: (Apellidos y nombres)		
<input type="text" value="Cabrejos Yalán, Víctor Manuel"/>		
en condición de egresado del Programa de Posgrado:		
<input type="text" value="Ingeniería Informática con mención en Ingeniería de Software"/>		
deja constancia que ha elaborado la tesis intitulada:		
<input type="text" value="Las Redes Neuronales Convolucionales y la mejora en el diagnóstico de Neumonía – área de Radiología"/>		
<p>Declara que el presente trabajo de tesis ha sido elaborado por el mismo y no existe plagio/copia de ninguna naturaleza, en especial de otro documento de investigación (tesis, revista, texto, congreso, o similar) presentado por cualquier persona natural o jurídica ante cualquier institución académica, de investigación, profesional o similar.</p> <p>Deja constancia que las citas de otros autores han sido debidamente identificadas en el trabajo de investigación, por lo que no ha asumido como suyas las opiniones vertidas por terceros, ya sea de fuentes encontradas en medios escritos, digitales o de la Internet.</p> <p>Asimismo, ratifica que es plenamente consciente de todo el contenido de la tesis y asume la responsabilidad de cualquier error u omisión en el documento y es consciente de las connotaciones éticas y legales involucradas.</p> <p>En caso de incumplimiento de esta declaración, el graduando se somete a lo dispuesto en las normas de la Universidad Ricardo Palma y los dispositivos legales vigentes.</p>		
 _____ Firma del graduando		30/10/2020 _____ Fecha

Anexo 2: Autorización de consentimiento para realizar la investigación

A continuación, se muestra el formato de autorización para realizar la investigación.

	UNIVERSIDAD RICARDO PALMA	Escuela de Posgrado
AUTORIZACIÓN PARA REALIZAR LA INVESTIGACIÓN		
DECLARACIÓN DEL RESPONSABLE DEL ÁREA O DEPENDENCIA DONDE SE REALIZARÁ LA INVESTIGACIÓN		
Dejo constancia que el área o dependencia que dirijo, ha tomado conocimiento del proyecto de tesis titulado:		
Las Redes Neuronales Convolucionales y la mejora en el diagnóstico de Neumonía – área de Radiología		
el mismo que es realizado por el Sr. / Srta. Estudiante (Apellidos y nombres):		
Cabrejos Yalán, Víctor Manuel		
en condición de estudiante – investigador del Programa de:		
Ingeniería Informática con mención en Ingeniería de Software		
Así mismo señalamos, que según nuestra normativa interna procederemos con el apoyo al desarrollo del proyecto de investigación, dando las facilidades del caso para aplicación de los instrumentos de recolección de datos.		
En razón de lo expresado doy mi consentimiento para el uso de la información y/o la aplicación de los instrumentos de recolección de datos:		
Nombre de la empresa: Repositorio de Datos: Mendeley Data Y Guangzhou Women and Children's Medical Center (respaldado por Data Seal of Approval / OAIS - ISO 14721)	Autorización para el uso del nombre de la Empresa en el Informe Final	SI NO
Apellidos y Nombres del Jefe/Responsable del área: Huimin Xia	Cargo del Jefe/Responsable del área: Jefe del Departamento de Cirugía Pediátrica	
Teléfono fijo (incluyendo anexo) y/o celular: +86-20 38076001	Correo electrónico de la empresa: data@mendeley.com huiminxia@hotmail.com	
<i>Huimin Xia</i>	30/10/2020	
Firma	Fecha	

Anexo 3: Matriz de consistencia

A continuación, se presenta la Matriz de Consistencia utilizada en la investigación del estudio. (Ver Tabla 46).

Tabla 46:
Matriz de Consistencia

Problemas Principal	Objetivos General	Hipótesis General	Variables Independiente	Indicador V..I	Variables Dependiente	Indicador V.D.
¿De qué manera las Redes Neuronales Convolucionales podría mejorar la rapidez en el diagnóstico de Neumonía en un área de Radiología?	Implementar las Redes Neuronales Convolucionales, para mejorar la rapidez en el diagnóstico de Neumonía en un área de Radiología.	Si se implementa las Redes Neuronales Convolucionales, entonces se mejorará la rapidez en el diagnóstico de Neumonía en un área de Radiología.	<i>Redes Neuronales Convolucionales</i>	--,--	<i>Rapidez en el diagnóstico de Neumonía</i>	--,--
Problemas Especifico	Objetivos Especificos	Hipótesis Especificas				
¿De qué manera el Algoritmo ResNet-50 podría reducir el número de diagnósicos incorrectas?	Implementar el Algoritmo ResNet-50 para reducir el número de diagnósicos incorrectas.	Si se implementa el Algoritmo ResNet-50, entonces se reducirá el número de diagnósicos incorrectas.	Algoritmo ResNet-50	Si / No	Número de diagnósicos incorrectas	Porcentaje de diagnósicos incorrectas al mes
¿De qué manera el Algoritmo Alexnet podría reducir el número de diagnósicos incorrectas?	Implementar el Algoritmo Alexnet para reducir el número de diagnósicos incorrectas.	Si se implementa el Algoritmo Alexnet, entonces se reducirá el número de diagnósicos incorrectas.	Algoritmo Alexnet	Si / No	Número de diagnósicos incorrectas	Porcentaje de diagnósicos incorrectas al mes
¿De qué manera el Algoritmo VGG podría reducir el número de diagnósicos incorrectas?	Implementar el Algoritmo VGG para reducir el número de diagnósicos incorrectas.	Si se implementa el Algoritmo VGG, entonces se reducirá el número de diagnósicos incorrectas.	Algoritmo VGG	Si / No	Número de diagnósicos incorrectas	Porcentaje de diagnósicos incorrectas al mes
¿De qué manera la Función de Activación SoftMax podría reducir el tiempo de espera de pacientes con neumonía?	Implementar la Función de Activación SoftMax para reducir el tiempo de espera de pacientes con neumonía.	Si se implementa la Función de Activación SoftMax, entonces se reducirá el tiempo de espera de pacientes con neumonía.	Función de Activación SoftMax	Si/No	Tiempo de espera de pacientes	Promedio mensual de horas de espera en un día

Fuente: Elaboración Propia

Anexo 4: Matriz de Operacionalización

A continuación, se presenta la Matriz de Operacionalización utilizada en la investigación del estudio. (Ver Tabla 47).

Tabla 47:
Matriz de Operacionalización

Variable Independiente	Indicador	Definición Conceptual	Definición Operacional
Algoritmo ResNet-50	Si / No	Es una arquitectura de redes neuronales de 50 capas de profundidad que utiliza la técnica de Aprendizaje Residual para clasificar imágenes complejas. (Fuente: Definición del investigador)	Diseñar e implementar una arquitectura de redes neuronales como parte de un software que clasifique placas radiográficas de pacientes con Neumonía.
Algoritmo Alexnet	Si / No	Es una arquitectura de redes neuronales pre-entrenada con 5 capas convolucionales, que reduce el problema de sobreajuste, y es utilizada para clasificar imágenes. (Fuente: Definición del investigador)	Diseñar e implementar una arquitectura de redes neuronales como parte de un software que clasifique placas radiográficas de pacientes con Neumonía.
Algoritmo VGG	Si / No	Es una arquitectura de redes neuronales que incluye de 16 a 19 capas con filtros reducidos, para mejor precisión, y es utilizada para clasificar imágenes. (Fuente: Definición del investigador)	Diseñar e implementar una arquitectura de redes neuronales como parte de un software que clasifique placas radiográficas de pacientes con Neumonía.
Función de Activación SoftMax	Si/No	Es un nodo ubicado al final de la última capa de una red neuronal, utilizada para clasificación múltiple al normalizar el resultado y convertirlo en una distribución de probabilidades. (Fuente: Definición del investigador)	Diseñar e implementar una arquitectura de redes neuronales que al recibir múltiples placas radiográficas pueda mostrar el resultado en probabilidades de tener Neumonía para poder priorizar con facilidad.
Variable Dependiente	Indicador	Definición Conceptual	Definición Operacional
Número de diagnóstico incorrectas	Porcentaje de diagnóstico incorrectas al mes	Es la métrica que evalúa que tan cerca es la predicción hecha una red neuronal al valor verdadero. (Fuente: Definición del investigador)	Aplicar una métrica de precisión a la red neuronal que sigue el algoritmo ResNet-50 cuando predice si una placa radiográfica para saber qué tan correcta es la predicción.
Número de diagnóstico incorrectas	Porcentaje de diagnóstico incorrectas al mes	Es la métrica que evalúa que tan cerca es la predicción hecha una red neuronal al valor verdadero. (Fuente: Definición del investigador)	Aplicar una métrica de precisión a la red neuronal que sigue el algoritmo Alexnet cuando predice si una placa radiográfica para saber qué tan correcta es la predicción.
Número de diagnóstico incorrectas	Porcentaje de diagnóstico incorrectas al mes	Es la métrica que evalúa que tan cerca es la predicción hecha una red neuronal al valor verdadero. (Fuente: Definición del investigador)	Aplicar una métrica de precisión a la red neuronal que sigue el algoritmo VGG cuando predice si una placa radiográfica para saber qué tan correcta es la predicción.
Tiempo de espera de pacientes	Promedio mensual de horas de espera en un día	Es el resultado de ordenar los resultados de las predicciones mostradas en porcentajes de probabilidad de que el paciente tenga neumonía de mayor probabilidad a menor probabilidad. (Fuente: Definición del investigador)	Facilitar la identificación y priorización de casos más urgentes de neumonía usando el resultado de una lista de porcentajes de probabilidad de tener Neumonía.

Fuente: Elaboración propia