

UNIVERSIDAD AUTÓNOMA DE ZACATECAS

“Francisco García Salinas”



“Estudio comparativo entre arquitecturas de Redes Neuronales Convolucionales para la clasificación de lesiones gastrointestinales”

Tesis para obtener el grado de:
Maestro en Ciencias del Procesamiento de la Información

Presenta
Erik Orlando Cuevas Rodríguez

Director:
Dr. Eric Galván Tejada

Co-Directores:
Dr. Gamaliel Moreno Chávez
Dr. Jorge Isaac Galván Tejada

Asesores:
Dr. Huizilopoztli Luna García
Dr. José María Celaya Padilla

Zacatecas, Zac., septiembre del 2022



SOMOS
ARTE, CIENCIA Y
DESARROLLO
CULTURAL



Zacatecas, Zac., 18 de agosto de 2022.

C. Erik Orlando Cuevas Rodriguez
Estudiante de la MCPI
PRESENTE

Dr. Huizilopoztli Luna García
Responsable de la MCPI

Nos es grato comunicarle que después de haber sometido a revisión académica la propuesta de Tesis titulada “**Estudio comparativo entre arquitecturas de Redes Neuronales Convolucionales para la clasificación de lesiones gastrointestinales**”, presentada por el estudiante Ing. **Erik Orlando Cuevas Rodriguez** y habiendo efectuado todas las correcciones indicadas por este Comité Tutorial, se **AUTORIZA** el documento de tesis para su impresión. Sin más por el momento reciban un cordial saludo.

COMITÉ TUTORIAL
PROCESAMIENTO Y ANÁLISIS DE DATOS



Dr. Carlos Enrique Galván
Tejada



Dr. Huizilopoztli Luna
García



Dr. Gamaliel Moreno
Chávez



Dr. Jorge Isaac Galván Tejada



Dr. José María Celaya

c.c.p. Interesado.

c.c.p. Responsable de la Maestría en Ciencias del Procesamiento de la Información.



SOMOS
ARTE, CIENCIA Y
DESARROLLO
CULTURAL

**COORDINACIÓN DE
INVESTIGACIÓN Y POSGRADO**

Carta de similitud núm. 285/IyP
Zacatecas, Zacatecas 02/septiembre/2022

Dr. Huizilopoztli Luna García
Responsable de la MCPI – UAZ
Presente

Estimado Dr. Huizilopoztli,

Después de saludarlo, sirva el presente oficio para notificar que el documento

“Estudio comparativo entre arquitecturas de Redes Neuronales Convolucionales para la clasificación de lesiones gastrointestinales”
Por Erik Orlando Cuevas Rodríguez

Fue analizado con el software Copyleaks, con la intención de detectar similitudes; el resultado en cuestión fue

6 % de similitud

De acuerdo a lo anterior, el porcentaje se considera **ACEPTABLE** de acuerdo a los estándares internacionales.

Atentamente

"Somos Arte, Ciencia y Desarrollo Cultural"

Dr. Carlos Francisco Bautista Capetillo
Coordinador de Investigación y Posgrado
Universidad Autónoma de Zacatecas



SOMOS
ARTE, CIENCIA Y
DESARROLLO
CULTURAL



Zacatecas, Zac., 08 de agosto de 2022
Carta de cesión de derechos

A QUIEN CORRESPONDA

El que suscribe C. Erik Orlando Cuevas Rodríguez alumno del Programa **Maestría en Ciencias del Procesamiento de la Información** con numero de matrícula **25604477**, adscrito a la Unidad Académica de Ingeniería Eléctrica de la Universidad Autónoma de Zacatecas, manifiesta que es autor intelectual del presente trabajo de Tesis bajo la dirección del Dr. Carlos Eric Galván Tejada y cede los derechos del trabajo titulado: **“Estudio comparativo entre arquitecturas de Redes Neuronales Convolucionales para la clasificación de lesiones gastrointestinales”** a la Universidad Autónoma de Zacatecas para su difusión, con fines académicos y de investigación.

ATENTAMENTE

Erik Orlando Cuevas Rodríguez



SOMOS
ARTE, CIENCIA Y
DESARROLLO
CULTURAL



AGRADECIMIENTO ESPECIAL

Al Programa Nacional de Posgrados de Calidad (PNPC) del Consejo Nacional de Ciencia y Tecnología (CONACYT) por su apoyo económico a través de la convocatoria “Becas Nacional (Tradicional) 2020-22”.

Agradecimientos

Quiero agradecer a la vida por brindarme la oportunidad de seguir estudiando. Por permitirme hacer lo que a mí me gusta. A mi familia, por apoyarme incondicionalmente y tenerme paciencia. También quiero agradecer a mi director y asesores de tesis, quienes me apoyaron en todo el proceso del proyecto.

Dedicatoria

Quiero dedicar el presente trabajo a mi madre, quien es la persona que más admiro en la vida.

Resumen

El tracto gastrointestinal (GI) está sujeto a numerosos y diferentes hallazgos de la mucosa gástrica, que van desde molestias menores, hasta enfermedades altamente letales. La endoscopia es el procedimiento estándar utilizado para detectar diferentes patologías en el tracto GI, sin embargo, pequeñas lesiones del tracto GI suelen pasar desapercibidas durante el procedimiento debido a la dificultad para identificarlas, e incluso, algunas veces pueden ser precursoras de cáncer. En los últimos años, este problema ha sido abordado por diferentes enfoques de aprendizaje profundo y técnicas de aprendizaje automático, que detectan y clasifican de manera automática las diferentes lesiones y hallazgos patológicos del tracto GI. En el aprendizaje profundo, la red neuronal convolucional (CNN) es el algoritmo mayor utilizado debido a su gran precisión para modelar fenómenos complejos. Se han propuesto diversas CNN como herramienta de diagnóstico asistida por ordenador para la clasificación de lesiones del tracto gastrointestinal. Este trabajo compara cuatro arquitecturas de CNN diferentes: AlexNet, DenseNet-201, Inception-v3 y ResNet-101 para clasificar 10 tipos de lesiones del tracto GI utilizando HyperKvasir como conjunto de datos. Para evaluar el rendimiento de los modelos se utilizó un conjunto de métricas tales como: exactitud, precisión, sensibilidad, especificidad, F1-score y área bajo la curva (AUC). Todos los modelos lograron resultados estadísticamente significativos. El modelo DenseNet-201 logró una exactitud del 97.11%, 96.26% de sensibilidad, 99.67% de especificidad y 94.90% de AUC.

Palabras clave: Aprendizaje profundo, clasificación automática, comparación de arquitecturas, lesiones gastrointestinales, red neuronal convolucional.

Abstract

The gastrointestinal (GI) tract is subject to numerous different findings of the gastric mucosa, ranging from minor annoyances to highly lethal diseases. Endoscopy is the standard procedure used to detect different pathologies in the GI tract, however, small lesions of the GI tract often go undetected during the procedure due to the difficulty in identifying them, and sometimes they can even be precursors of cancer. In recent years, this problem has been addressed by different deep learning approaches and machine learning techniques, which automatically detect and classify different lesions and pathological findings of the GI tract. In deep learning, convolutional neural network (CNN) is the most widely used algorithm due to its high accuracy in modeling complex phenomena. Several CNNs have been proposed as a computer-aided diagnostic tool for GI tract lesion classification. This work compares four different CNN architectures: AlexNet, DenseNet-201, Inception-v3 and ResNet-101 to classify 10 types of GI tract lesions using HyperKvasir as a dataset. A set of metrics such as: accuracy, precision, sensitivity, specificity, F1-score, and area under the curve (AUC) were used to evaluate the performance of the models. All models achieved statistically significant results. The DenseNet-201 model achieved 97.11% accuracy, 96.26% sensitivity, 99.67% specificity and 94.90% AUC.

Keywords: Deep learning, automatic classification, architectures comparison, gastrointestinal lesions, convolutional neural network.

Contenido General

Agradecimiento	i
Dedicatoria.....	ii
Resumen	iii
Palabras clave:	iii
Abstract.....	iv
Keywords:	iv
Contenido General	1
Índice de Figuras	3
Índice de Tablas.....	4
Capítulo 1. Introducción.....	5
1.1 Antecedentes.....	5
1.2 Planteamiento del problema de investigación	6
1.3 Justificación del problema de investigación.....	7
1.4 Preguntas de investigación	8
1.5 Objetivo general	8
1.6 Objetivos específicos.....	8
1.7 Hipótesis	9
1.8 Estructura de la tesis	9
Capítulo 2. Marco Teórico	10
2.1 Descripción de Teorías base	10
2.1.1 Inteligencia Artificial	10
2.1.2 Aprendizaje Profundo	10
2.1.3 Redes Neuronales Convolucionales.....	11
2.1.4 Funciones de activación	15
2.1.5 Funciones de pérdida.....	16
2.1.6 Regularización.....	17
2.1.7 Optimización	18

2.1.8	Aprendizaje por transferencia	19
2.1.9	Arquitecturas de Redes Neuronales Convolucionales.....	20
2.1.10	Métricas de evaluación del desempeño.....	23
2.1.11	Endoscopia	28
2.1.12	Puntos de referencia anatómicos.....	30
2.1.13	Hallazgos Patológicos	32
2.2	Principales estudios relacionados	36
2.3	Modelo o esquema general de investigación.....	45
Capítulo 3. Método y propuesta de investigación		46
3.1	Modelo de investigación.....	46
3.2	Descripción de la propuesta.....	47
3.2.1	Adquisición de los datos	47
3.2.2	Entrenamiento de las CNNs	49
Capítulo 4. Resultados y Limitaciones		52
Capítulo 5. Conclusiones		58
Referencias.....		60

Índice de Figuras

Figura 1. Ejemplo de Red Neuronal Convolutiva.	12
Figura 2. Arquitectura de una CNN (Lee et al., 2017).	13
Figura 3. Tres tipos de operación de agrupación (Alzubaidi et al., 2021).	14
Figura 4. Estructura de AlexNet (J. Chen et al., 2021).	20
Figura 5. Bloque denso de 5 capas (Mocsari & Stone, 2017).	21
Figura 6. Arquitectura del modelo ResNet 18 (Damaševi, 2021).	23
Figura 7. Matriz de confusión.	25
Figura 8. Gastroscofia o Endoscopia Superior (Terese Wislow, 2005).	28
Figura 9. Colonoscopia o Coloscopia (Terese Wislow, 2005).	29
Figura 10. Proceso de desarrollo de las medidas de rendimiento (Kaminski et al., 2017).	30
Figura 11. Línea Z (Borgli et al., 2020).	31
Figura 12. Píloro (Borgli et al., 2020).	31
Figura 13. Ciego (Borgli et al., 2020).	32
Figura 14. Esofagitis (Borgli et al., 2020).	33
Figura 15. Esófago de Barret (Borgli et al., 2020).	34
Figura 16. Hemorroides (Borgli et al., 2020).	34
Figura 17. Pólipo (Borgli et al., 2020).	35
Figura 18. Colitis Ulcerosa (Borgli et al., 2020).	35
Figura 19. Diagrama de la metodología propuesta.	46
Figura 20. Diferentes clases del conjunto de datos.	48
Figura 21. Número de imágenes por clase del conjunto de datos.	49
Figura 22. Representación esquemática del proceso de aprendizaje por transferencia.	50
Figura 23. Resumen del rendimiento de las arquitecturas.	54
Figura 24. Matriz de confusión de AlexNet.	55
Figura 25. Matriz de confusión de DenseNet-201.	55
Figura 26. Matriz de confusión de Inception-v3.	56
Figura 27. Matriz de confusión de ResNet-101.	57
Figura 28. Curva ROC de la arquitectura DenseNet-201.	57

Índice de Tablas

Tabla 1. Comparación de los diferentes estudios de clasificación de lesiones gástricas, aquí, SGD = Stochastic Gradient Descent, ACC = Accuracy, MCC = Matthews Correlation Coefficient, SPEC = Specificity, PREC = Precision, SEN = Sensitivity	43
Tabla 2. Características generales de las arquitecturas.....	49
Tabla 3. Hiperparámetros implementados. *Stochastic Gradient Descent with Momentum (SGDM).....	50
Tabla 4. Características del área de trabajo.	51
Tabla 5. Rendimiento general de las arquitecturas.....	52

Capítulo 1. Introducción

El presente capítulo describe los antecedentes y da contexto al estudio, expone la problemática y relevancia del estudio, se plantean las preguntas de investigación, además, se definen los objetivos y se enuncia la hipótesis, por último, describe de manera breve la estructura del estudio.

1.1 Antecedentes

El tracto GI humano, está sujeto a numerosos y diferentes hallazgos anormales de la mucosa gastrointestinal que van desde molestias menores hasta enfermedades altamente letales (Borgli et al., 2020). Como el cáncer colorrectal, que ocupa el tercer lugar en términos de incidencia de cáncer y el segundo lugar en términos de mortalidad. El cáncer de esófago es el octavo más común y el sexto en causas de muerte (Bray et al., 2018; Sung et al., 2021; Zhao et al., 2019).

Por otro lado, la endoscopia es el procedimiento estándar para detectar y diagnosticar patologías del tracto GI (Glissen Brown et al., 2021; González-Huix Lladó et al., 2010; Yu et al., 2021). No obstante, es preocupante que existan deficiencias en la detección de cánceres y pólipos adenomatosos (tumores premalignos benignos o adenomas) debido a que no son fáciles de observar por los diferentes puntos ciegos que existen en el colon (Choi et al., 2014; Gómez-Zuleta et al., 2021; Kaminski et al., 2010, 2017; Komeda et al., 2013; Yu et al., 2021). Además, aproximadamente entre el 50% y 60% de las lesiones que se pasan por alto se convierten en cáncer de intervalo, es decir, una lesión no detectada que resulta en cáncer (Pohl & Robertson, 2010). Sin embargo, una detección temprana de estas lesiones precancerosas ofrece la oportunidad de prevenir el cáncer en su totalidad (Jha et al., 2021; Levin et al., 2008).

En los últimos años, la aplicación de la Inteligencia Artificial (IA) en el análisis y clasificación de imágenes médicas ha ido en aumento debido al potencial que tiene para mejorar la eficiencia y precisión en los métodos de diagnóstico (Borgli et al., 2020; Kochhar et al., 2021; Pannala et al., 2020; Petschamig et al., 2017; Rodríguez de Santiago et al., 2019; Smedsrud et al., 2021; Yang et al., 2021). En este sentido, el amplio uso de imágenes digitales como herramienta de diagnóstico de las afecciones gastrointestinales convierte a la IA en una herramienta de interés (Cepeda-Vásquez, 2021; Kochhar et al., 2021; Sánchez-Montes et al., 2020). El Aprendizaje Profundo (DL, por sus siglas en inglés), un campo de la IA tiene un futuro prometedor como asistente en la toma de decisiones, reduciendo teóricamente el número de lesiones no detectadas y con la consecuente reducción de biopsias y polipectomías (técnica para eliminar los pólipos presentes) innecesarias (Fernández-Esparrach & García-Rodríguez, 2020).

1.2 Planteamiento del problema de investigación

La Organización Mundial de la Salud (WHO, por sus siglas en inglés) estimó aproximadamente un total de 19.2 millones de nuevos casos de cáncer a nivel mundial, donde el 10% corresponde a cáncer colorrectal (CCR), ocupando el tercer lugar en incidencia de cáncer. Sin embargo, en cuanto a índice de mortalidad el CCR ocupa el segundo lugar (Sung et al., 2021).

En México, el CCR representa el 8.6% del total de las muertes causadas por cáncer, siendo el segundo lugar solo detrás del cáncer de mama. Y de acuerdo con la Agencia Internacional para la Investigación del Cáncer (IARC) se estima un aumento del 16.5% al año 2025 (Sung et al., 2021). Aunque no se sabe exactamente que provoca el CCR, existen diferentes factores de riesgo que aumentan las probabilidades de contraerlo, como cambios en el estilo de vida, y en la dieta, es decir, cambios hacia una mayor ingesta de alimentos de origen animal, el consumo excesivo de alcohol, tabaquismo y un estilo de vida más sedentario, lo que lleva a una disminución de la actividad física y aumento del peso corporal (Bray et al., 2018; Gómez-Zuleta et al., 2021).

Hoy en día, la endoscopia es la técnica diagnóstica de elección en CCR y otras lesiones del tracto GI, sin embargo, sus resultados pueden ser falsamente negativos, produciendo un retardo en el diagnóstico del CCR (Levin et al., 2008; Sastre Lozano, 2018). Diferentes artículos mencionan que hay una tasa de error de hasta el 26% en la detección de adenomas (Gómez-Zuleta et al., 2021; Zhao et al., 2019).

De acuerdo con lo anterior, y tomando en cuenta que el diagnóstico de afecciones gastrointestinales es mediante imágenes digitales, surge un escenario en donde la tecnología y la inteligencia artificial empiezan a mostrar resultados, el uso de redes neuronales convolucionales (CNNs, por sus siglas en inglés) se ha vuelto popular debido a la facilidad que presentan para la clasificación de imágenes (Agrawa et al., 2017; Chang et al., 2019; T. H. Hoang et al., 2018). Sin embargo, la clasificación de imágenes de lesiones del tracto GI sigue siendo un problema muy complejo de resolver.

1.3 Justificación del problema de investigación

El diagnóstico de las afecciones gastrointestinales se apoya en estudios endoscópicos utilizando imágenes digitales, esto permite la implementación de técnicas de análisis de imágenes y de IA, que en conjunto empiezan a mostrar resultados en la mejora de tasas de detección de las afecciones del tracto GI (Cepeda-Vásquez, 2021; Fernández-Esparrach & García-Rodríguez, 2020; Gómez-Zuleta et al., 2021; Yu et al., 2021), contribuyendo probablemente a reducir las desigualdades en la clasificación de lesiones gastrointestinales y mejorar la calidad de los diagnósticos médicos (Glissen Brown et al., 2021; Pogorelov, Ranheim Randel, et al., 2017).

Una limitante que se tenía en la implementación de técnicas de IA era la poca cantidad de información en las bases de datos, además no se contaba con un gran número de bases de datos (Cogan et al., 2019; Da-Chuan Cheng, Wen-Chien Ting, Yung-Fu Chen, 2008; J et al., 2019; Pogorelov, Ranheim Randel, et al., 2017), otro factor fue la calidad en las imágenes, lo que limitaba la implementación de modelos de IA (Yu et al., 2021). Hoy en día, están disponibles diferentes bases de datos con mayor cantidad de información (Borgli et al., 2020; Smedsrud et al., 2021) lo cual hace posible la utilización de algoritmos de DL,

específicamente, de arquitecturas de CNNs (Chang et al., 2019; Ding et al., 2019) por sus elevados porcentajes de precisión en la clasificación de lesiones gastrointestinales (Agrawa et al., 2017; Chang et al., 2019; T. Hoang et al., 2018). No obstante, estos clasificadores sufren una falta de interpretabilidad debido a que se consideran como “cajas negras” que dan buenos resultados, pero sin explicación alguna (Maeda Gutiérrez & Galván Tejada, 2020). Siendo necesario considerar la utilización de diferentes métricas para evaluar el rendimiento de los modelos y determinar su comportamiento, las métricas de rendimiento deben interpretarse siempre de forma conjunta en lugar de depender de una sola métrica (Thambawita et al., 2020).

1.4 Preguntas de investigación

1. ¿Cuál es la arquitectura de CNN que mejor modela el problema de clasificación de lesiones gastrointestinales?
2. ¿Qué efectos tiene la implementación de arquitecturas de CNNs con respecto a la reducción de la tasa de error en la detección de pólipos adenomatosos?

1.5 Objetivo general

Evaluar y comparar el desempeño de AlexNet, DenseNet-201, Inception-v3 y Resnet-101 para la clasificación de imágenes con lesiones gastrointestinales, utilizando HyperKvasir como conjunto de datos.

1.6 Objetivos específicos

1. Evaluar y comparar el rendimiento de AlexNet, DenseNet-201, Inception-v3 y Resnet-101 utilizando imágenes con lesiones gastrointestinales para su clasificación.
2. Analizar los resultados en la clasificación de pólipos y hacer una comparativa de la tasa de error en la detección de adenomas en procedimientos de endoscopia.

1.7 Hipótesis

La implementación y comparación de arquitecturas de CNNs para el análisis de imágenes de lesiones del tracto GI, a través de la evaluación y comparación en términos de rendimiento utilizando un conjunto de métricas como lo son: exactitud, precisión, sensibilidad, especificidad, F1-score y área bajo la curva permiten seleccionar la arquitectura que mejor modela el problema de clasificación de lesiones gastrointestinales.

1.8 Estructura de la tesis

El presente trabajo está estructurado de la siguiente manera: El capítulo 2 presenta una amplia descripción de los temas de referencia. Capítulo 3 aborda la metodología del trabajo de investigación, donde se detallan las técnicas, herramientas y recursos utilizados. Capítulo 4 contiene la discusión de los resultados obtenidos y por último en el capítulo 5 expone las conclusiones de la investigación.

Capítulo 2. Marco Teórico

El presente capítulo describe las teorías base. Contiene la descripción y definición de inteligencia artificial, aprendizaje profundo y redes neuronales convolucionales. También se mencionan las métricas de evaluación utilizadas para medir el rendimiento de los diferentes modelos que se implementaron en el presente trabajo. Así mismo, se describe el proceso de endoscopia, sus principales características, lugares anatómicos y lesiones del tracto GI. Por último, se presentan los trabajos relacionados de mayor relevancia para el estudio.

2.1 Descripción de Teorías base

2.1.1 Inteligencia Artificial

La IA es una de las disciplinas más recientes y se inició formalmente en 1956 (Russell, 2004). Es una rama de las ciencias de la computación que incluye conceptos muy transversales relacionados con la lógica y el aprendizaje, está formada por una serie de algoritmos lógicos suficientemente entrenados a partir de los cuales las máquinas son capaces de tomar decisiones para pasos concretos a partir de normas generales (Avila-Tomás et al., 2020). En definitiva, la inteligencia artificial tiene por objeto que los ordenadores hagan las mismas cosas que puede realizar la mente humana (Boden & Pérez Parra, 2017), con la ventaja de que puede articularse con sistemas automáticos que posibiliten la ejecución.

2.1.2 Aprendizaje Profundo

El DL es una potente forma de ML y a su vez un subcampo de la IA, que permite a los ordenadores resolver problemas de percepción, como el reconocimiento de imágenes y del habla (Mohanty et al., 2015; Rusk, 2015), debido a la semejanza con el sistema de cognición humana multicapa (Lee et al., 2017). El DL puede entenderse como un marco general para modelos construidos por múltiples capas de procesamiento que le permiten aprender presentaciones jerárquicas en diferentes niveles de abstracción (Lecun et al., 2015). Se pueden identificar tres factores importantes en el desarrollo del DL, uno de ellos

es el aumento de datos, que se puede ver como una escala que impulsa al DL, otro es el tamaño de los modelos, es decir, el número de capas que contienen las arquitecturas para procesar los datos de manera efectiva y, por último, la implementación de Unidades de Procesamiento Gráfico (GPU) para mejorar el rendimiento computacional (Maeda Gutiérrez & Galván Tejada, 2020).

2.1.3 Redes Neuronales Convolucionales

La base biológica de la inteligencia está en el cerebro. Aunque vamos conociendo cada vez más sobre su fisiología y las diferentes vías anatómicas y funcionales que participan en el proceso de aprendizaje humano, aún existen muchas incógnitas que deben despejarse. Para las máquinas existe una analogía similar a la estructura y funcionamiento del cerebro (Avila-Tomás et al., 2020).

Una CNN pertenece a una clase de algoritmos de DL que se utilizan más comúnmente para tareas de clasificación de imágenes (Abraham et al., 2021), su principal beneficio es que identifica automáticamente las características relevantes sin necesidad de intervención humana (Alzubaidi et al., 2021). Una CNN es una arquitectura de red neuronal inspirada en la corteza visual biológica de los animales. La arquitectura está formada por capas convolucionales con conjuntos compartidos de pesos bidimensionales. El algoritmo reconoce la información espacial y utiliza la agrupación de capas para filtrar los conocimientos comparativamente más importantes y transmitir solo las características concentradas (Hiriyannaiah et al., 2020; Kwak & Hui, 2019; Song & Cai, 2021; Subasi, 2020). La entrada \mathbf{x} de cada capa en un modelo CNN se organiza en tres dimensiones: altura, ancho, y profundidad o $\mathbf{m} \times \mathbf{m} \times \mathbf{r}$, donde la altura (\mathbf{m}) es igual a la anchura, mientras que la profundidad se refiere al número de canales. La Figura 1 ilustra un modelo de CNN con una imagen de entrada con \mathbf{r} igual a 3.

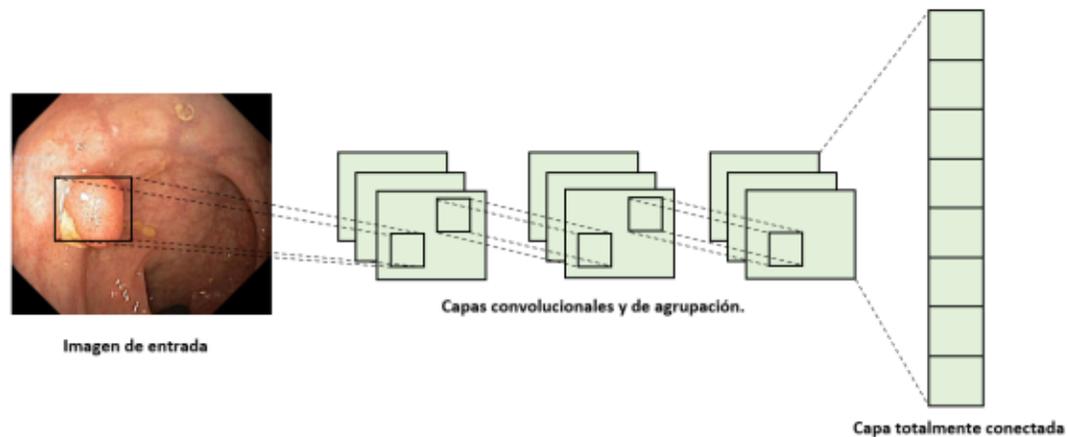


Figura 1. Ejemplo de Red Neuronal Convolutiva.

En una capa convolutiva, se calcula la similitud entre pequeños segmentos de la imagen y algunos núcleos aprendidos. A continuación, en una capa de agrupación, los valores de los píxeles que están cerca se agrupan y se combinan en un solo píxel. Esto reduce la dimensión de los datos, lo que conlleva menos cálculos y una selección de características que es robusta a los pequeños cambios. Las capas de agrupación y las capas convolucionales se alternan. El valor de un solo píxel después de la primera capa de agrupación representa un lote entero de píxeles de la imagen original, y después de dos capas de agrupación, cada valor representa un lote de píxeles de la imagen original. El primer par de capas de convolución y agrupación combina los píxeles de la imagen en características locales y los siguientes pares de capas de convolución y agrupación combinan estas características locales en características cada vez más globales (Reimers & Requena-Mesa, 2020).

La red neuronal utiliza una capa de salida para la clasificación multiclase, se utiliza un extractor de características en el proceso de entrenamiento y consiste en tipos especiales de redes neuronales que deciden los pesos a través del proceso de entrenamiento, lo que proporciona un mejor reconocimiento de imágenes cuando el extractor de características de la red neuronal se hace más profundo (contiene más capas), a costa de las complejidades del método de aprendizaje (Subasi, 2020).

2.1.3.1 Estructura de una Red Neuronal Convolutiva

La unidad principal de estas arquitecturas es la neurona, que proporciona una salida como función de las entradas a ella. Un arreglo de neuronas forma una capa o bloque y una red está compuesta por varios bloques elementales ordenados (Gómez-Zuleta et al., 2021). La arquitectura de una CNN (**Figura 2**) consiste en tres tipos de capas: (1) de convolución, (2) de agrupación y (3) completamente conectada (Shajun Nisha & Nagoor Meeral, 2021).

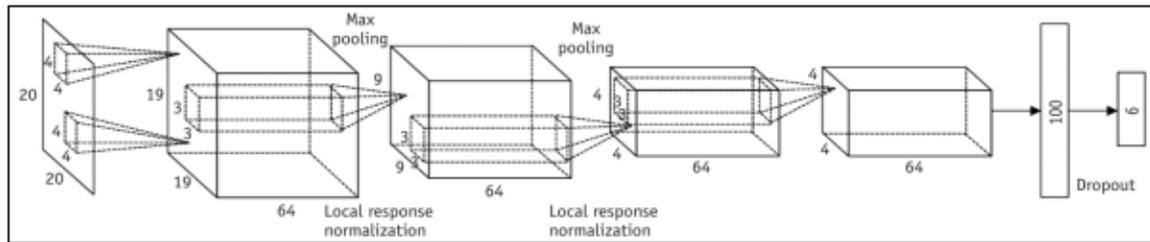


Figura 2. Arquitectura de una CNN (Lee et al., 2017).

Una capa de convolución está formada por múltiples mapas de neuronas, denominados mapas de características. Cada neurona de un mapa de características sólo está conectada a un parche local de neuronas de la capa anterior, el llamado campo receptivo, estas capas o kernels convolucionales denotados por k se distinguen por los pesos específicos de cada nodo al ser operado sobre una región específica de la imagen (Gómez-Zuleta et al., 2021; Maeda Gutiérrez & Galván Tejada, 2020), también tienen tres dimensiones ($n \times n \times q$) similares a la imagen de entrada, sin embargo, n debe de ser menor que m , mientras que q puede ser igual o menor que r , además, los kernels son la base de las conexiones locales, que comparten parámetros análogos como el sesgo b^k y el peso W^k para generar k mapas de características h^k con un tamaño de $(m - n - 1)$. La capa de convolución calcula el producto punto entre la entrada x y los pesos como en la **Ecuación 1**, a continuación, aplicando una función de activación a la salida de la capa de convolución se obtiene lo siguiente:

$$h^k = f(W^k * x + b^k) \quad (1)$$

Posteriormente, los datos de entrada son convalidados mediante el desplazamiento de los campos receptivos paso a paso y comparten los mismos parámetros en cada pequeña porción de la imagen, reduciendo en gran medida el número de hiperparámetros de la imagen (Zemouri et al., 2019). El objetivo principal de una capa de convolución es detectar bordes, líneas y otros elementos visuales.

Una capa de agrupación se utiliza para reducir eficazmente las dimensiones de los mapas de características, seguir siendo resistente a la forma y posición de las características semánticas detectadas dentro de la imagen e introducir la invariabilidad de la traslación. En otras palabras, este enfoque mantiene la mayoría de la información dominante (o características) en cada paso de la etapa de agrupación (Alzubaidi et al., 2021). Existen varios métodos de agrupación para utilizar en varias capas, las tres operaciones más comunes son la agrupación promedio, donde la unidad de salida será el promedio matemático de las unidades vecinas, la agrupación máxima, donde la unidad de salida será el valor máximo de la región y la agrupación máxima global donde se toma el promedio matemático de toda la capa (Indolia et al., 2018; Lee et al., 2017).

La **Figura 3** representa los tres tipos de operaciones de agrupación.

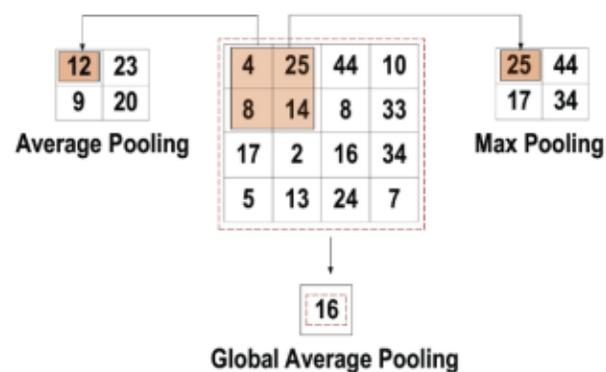


Figura 3. Tres tipos de operación de agrupación (Alzubaidi et al., 2021).

Una capa totalmente conectada está diseñada para transformar el mapa de características de las capas anteriores a vectores unidimensionales de características, funciona como una capa de clasificación de las características extraídas por las capas convolucionales y de agrupación anteriores y proporciona las probabilidades de la tarea de clasificación del

objetivo(Alzubaidi et al., 2021).

Un modelo de CNN utiliza varios módulos adicionales, como la activación, la normalización por lotes (Witten et al., 2017), la regularización y el abandono, para mejorar el proceso de aprendizaje (Niyas et al., 2021).

2.1.4 Funciones de activación

Una función de activación desempeña un papel fundamental en el entrenamiento de las CNNs (Maguolo et al., 2021), esta función agrega los factores no lineales para eliminar los datos redundantes mientras preserva las características, retiene la “característica de la neurona activa” y mapea estas características por medio de funciones no lineales (Lin & Shen, 2018), es decir, normaliza las probabilidades obtenidas de las capas completamente conectadas según una función específica, en las que se obtiene una probabilidad de 0 a 1 (Gómez-Zuleta et al., 2021). Es decir, una función de activación toma la decisión de activar o no una neurona con referencia a una entrada particular creando la salida correspondiente. En la actualidad, se han aplicado diversas funciones de activación para construir CNNs, como la Sigmoide denotada por la **Ecuación 2**.

$$f_{(x)sgm} = \frac{1}{1 + e^{-x}} \quad (2)$$

Esta función de activación recibe como entrada números reales, es decir, la variable x , mientras que la salida queda restringida a un valor entre uno y cero (Iliev et al., 2017). ReLu una de las funciones más utilizadas en el contexto de las CNNs, convierte todos los valores de la entrada x a números positivos, una ventaja de implementar esta función de activación es que su costo computacional es muy bajo (Brown et al., 2017), su representación matemática se muestra en la **Ecuación 3**.

$$f_{(x)ReLu} = \max(0, x) \quad (3)$$

Softplus puede ser considerada como una aproximación suave a ReLu que siempre es diferenciable y tiene un gradiente no nulo por debajo de cero. La principal razón para usar Softplus es que propaga el gradiente a través de su dominio. Esto puede hacer que las redes sean más fáciles de entrenar (Senior & Lei, 2014). La función se expresa matemáticamente en la **Ecuación 4**.

$$f_{(x)Soft} = \log(1 + \exp(x)) \quad (4)$$

2.1.5 Funciones de pérdida

Una función de pérdida se utiliza en la capa de salida para calcular el error predicho a lo largo de las muestras de entrenamiento. Este error revela la diferencia entre la salida actual y la esperada, posteriormente es optimizada a través del proceso de entrenamiento de la red. Diferentes funciones de pérdida se han utilizado en diversos tipos de problemas, como la entropía cruzada o función Softmax.

Esta función es muy utilizada para medir el rendimiento de la CNN, su salida es la probabilidad $p \in \{0,1\}$. Además, comúnmente se utiliza como remplazo de la función del error cuadrático medio en problemas de clasificación multiclase. En la capa de salida se emplean las activaciones Softmax para generar la salida con una distribución de probabilidad (Alzubaidi et al., 2021).

La representación matemática de la probabilidad de la clase de salida es la **Ecuación 5**.

$$p_i = \frac{e^{a_i}}{\sum_{k=1}^N e_k^a} \quad (5)$$

Donde e^{a_i} representa la salida no normalizada de la capa anterior, mientras que N representa el número de neuronas en la capa de salida. Finalmente, la representación matemática de la función de entropía cruzada se encuentra en la **Ecuación 6**. Donde y_i es la salida de la CNN.

$$H(p, y) = - \sum_i y_i \log p_i; \quad i \in [1, N] \quad (6)$$

2.1.6 Regularización

Es muy común que los modelos de CNNs sufran de sobreajuste, es decir, que el modelo consiga buen rendimiento con los datos de entrenamiento, pero con los datos de prueba su rendimiento disminuya, caso contrario sucede con el subajuste. Por otro lado, cuando un modelo responde bien para ambas situaciones se dice que el modelo se encuentra bien ajustado. Al proceso de controlar estos errores se le llama regularización y existen diferentes técnicas que se implementan para evitar estos problemas. El Dropout o técnica de abandono es una técnica muy utilizada en estos casos, consiste en desactivar neuronas de la red de manera aleatoria durante cada época de entrenamiento (Nandini et al., 2021).

Otro enfoque es el aumento de datos, es la manera más fácil de evitar el sobreajuste, para lograr esto se aplica un preprocesamiento a las imágenes originales, configuraciones como el cambio de tamaño, la rotación y la reflexión son algunas de las operaciones que se utilizan.

La normalización por lotes escala las salidas de las capas de tal manera que tengan una media cero y una varianza de uno, es decir, sigue una distribución Gaussiana unitaria, esto hace que se capacite a la red más rápido (Alzubaidi et al., 2021; Amin et al., 2020). Las principales ventajas son las siguientes:

- Evita que el gradiente se desvanezca.
- Puede controlar eficazmente la mala inicialización de los pesos.
- Reduce significativamente el tiempo de convergencia de la red.
- Procura disminuir la dependencia del entrenamiento a través de los hiperparámetros.
- Las posibilidades de sobreajuste se reducen, ya que tiene una influencia menor en la regularización.

2.1.7 Optimización

El rendimiento en el proceso de aprendizaje de una CNN depende directamente del optimizador que se seleccione debido a que los parámetros de la red deben actualizarse siempre, mientras que la red también debe buscar la respuesta optimizada localmente en todas las épocas de entrenamiento para minimizar el error. La tasa de aprendizaje es un hiperparámetro que afecta la velocidad a la que avanza el algoritmo de optimización, así como el número de épocas del entrenamiento. La época de entrenamiento representa una repetición completa de la actualización de los parámetros que incluye el conjunto de datos de entrenamiento completo. Por lo que es muy importante seleccionar ambos hiperparámetros de manera adecuada para no afectar el rendimiento del modelo.

El descenso por gradiente estocástico (SGD) es la técnica más utilizada para entrenar CNNs, este algoritmo actualiza constantemente los parámetros de la red a través de cada época de entrenamiento. En lugar de analizar todos los ejemplos de entrenamiento, el SGD realiza la actualización de los parámetros en una sola iteración. La representación matemática de esta operación es la **Ecuación 7**.

$$W_k = W_{k-1} - \alpha_{k-1} \nabla f(W_k - 1) \quad (7)$$

Donde el peso actual está denotado por W_k , mientras que el peso de la iteración anterior está indicado por W_{k-1} , la tasa de aprendizaje es α y el gradiente ∇f representa la colección de todas las derivadas parciales en forma de vector. Sin embargo, el SGD puede llegar a presentar algunas limitaciones. Hay ocasiones donde la inclinación de la curva varía en todas direcciones, esto genera que el SGD se vuelva inestable y decaiga el rendimiento debido a las oscilaciones a lo largo de la pendiente.

Para amortiguar las oscilaciones se agrega un término conocido como momento, este término mejora la exactitud y velocidad en el entrenamiento del modelo. Para actualizar los pesos con el descenso por gradiente estocástico con momento (SGDM) no solo considera el gradiente actual, sino que también lo reemplaza con un vector de velocidad (**Ecuación 8**),

donde α representa el factor de aprendizaje y es mayor a cero. $\mu \in (0,1)$ es el coeficiente del momento que controla la velocidad y al mismo tiempo realiza pasos descendientes más rápido. Una media móvil exponencial de los gradientes actual y anterior para la minimización de la función objetivo (G. Habib & Qureshi, 2020). Por lo tanto, el momento está dado por la **Ecuación 9**:

$$V_{k+1} = \mu_k - \alpha \nabla f \quad (8)$$

$$W_{k+1} = W_k + V_{k+1} \quad (9)$$

No obstante, existen otros métodos de optimización, como el gradiente acelerado de Nesterov (NAG), este método mejora la tasa de convergencia global utilizando un factor de funciones convexas demasiado suaves, además se comporta de forma agresiva hacia la optimización de funciones no convexas (Luo et al., 2021).

El algoritmo de estimación del momento adaptativo (ADAM) es otro método de optimización ampliamente utilizado. ADAM es una estrategia de aprendizaje creada específicamente para entrenar redes neuronales profundas. Sus principales ventajas son más memoria y menos costo computacional. Este método tiene la capacidad de ajustar la tasa de aprendizaje para cada parámetro durante el proceso de optimización utilizando estadísticas acumulativas de segundo orden (Anil et al., 2019; Yadav & Anubhav, 2020).

2.1.8 Aprendizaje por transferencia

Este enfoque consiste en entrenar la red utilizando un gran volumen de datos, donde aprende el sesgo y los pesos durante el proceso de entrenamiento. Estos pesos se transfieren a diferentes redes para volver a entrenar o probar un modelo, de este modo, el nuevo modelo puede preentrenar los pesos en lugar que tener que entrenar desde cero, el uso de estos pesos se realiza por medio de un proceso llamado ajuste fino (fine tuning), en el cual se toma toda

la red preentrenada y se retira la última capa completamente conectada. Esta capa es reemplazada por una nueva, en donde el número de neuronas es igual al número de clases en la tarea de clasificación (Gómez-Zuleta et al., 2021).

Existen muchas razones para utilizar un modelo preentrenado. En primer lugar, el entrenar modelos en conjuntos de datos de gran tamaño tiene un elevado costo computacional. En segundo lugar, el entrenamiento de modelos con muchas capas puede llevar mucho tiempo, incluso puede tardar semanas. Por último, un modelo preentrenado puede ayudar a la generalización de la red y acelerar la convergencia (Alzubaidi et al., 2021).

2.1.9 Arquitecturas de Redes Neuronales Convolucionales

Diferentes arquitecturas de CNN se han propuesto en la literatura para su desarrollo e implementación en la clasificación de imágenes médicas, a continuación, se describen brevemente las arquitecturas propuestas en el presente trabajo.

2.1.9.1 AlexNet

AlexNet fue propuesta por (Alex Krizhevsky, Ilya Sutskever, 2012), la arquitectura se presentó a la Comisión de Reconocimiento Visual a Gran Escala de ImageNet (ILSVRC) donde se obtuvieron buenos resultados (Kazemi, 2017). La red consta de 8 capas, las primeras 5 son capas convolucionales y las 3 capas restantes son totalmente conectadas. La estructura específica se observa en la **Figura 4**.

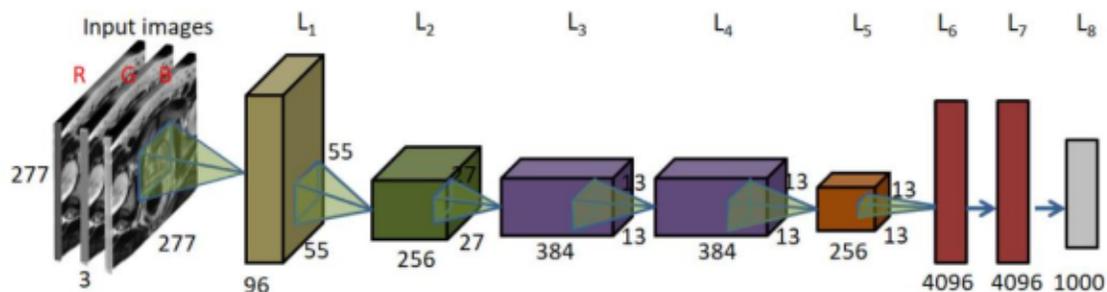


Figura 4. Estructura de AlexNet (J. Chen et al., 2021).

La primera capa convolucional filtra la imagen con 96 kernels, la segunda capa convolucional toma como entrada la salida de la primera capa y la filtra con 256 kernels.

La tercera y cuarta capa tienen 384 kernels conectados a las salidas normalizadas y agrupadas a la capa anterior, y la quinta capa tiene 256 kernels.

En las capas totalmente conectadas tiene 4096 neuronas y en total 650,000 neuronas, con 60 millones de parámetros (Adam et al., 2021; Maeda Gutiérrez & Galván Tejada, 2020).

2.1.9.2 DenseNet

El modelo DenseNet consta de un bloque denso que contiene la capa convolucional, la capa ReLU y la normalización por lotes. Por otro lado, dos bloques densos se conectan con la capa convolucional y de agrupación máxima y el último bloque denso se conecta con la agrupación media global y el clasificador Softmax (Bohmrah & Kaur, 2021).

Las DenseNets necesitan menos parámetros que las CNNs convencionales porque no necesitan mapas de características no esenciales, debido a que son estrechas e introducen nuevos mapas de características en cantidad insignificante (Chauhan et al., 2021).

La Red comprime las capas, cada una de las cuales implementa una transformación no lineal donde se indexa la capa. Para preservar la naturaleza de avance cada capa obtiene entradas adicionales de todas las capas precedentes y pasa sus propios mapas de características a todas las capas posteriores (Mocsari & Stone, 2017). La **Figura 5** ilustra esta disposición de forma esquemática.

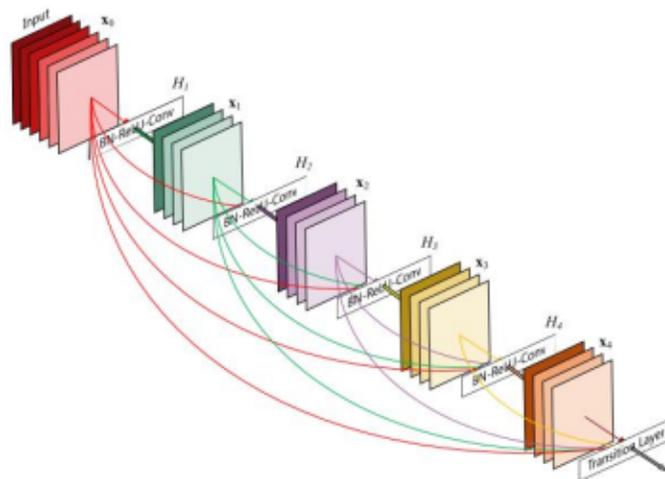


Figura 5. Bloque denso de 5 capas (Mocsari & Stone, 2017).

Una gran ventaja de la DenseNet es su flujo mejorado de información y gradientes a través de la red, lo que hace que sea fácil de entrenar. Cada capa tiene acceso directo a los gradientes de la función de pérdida y a la señal de la entrada original, lo que conduce a una supervisión profunda implícita. Además, las conexiones densas tienen un efecto regularizador que reduce el sobreajuste en tareas con conjuntos de entrenamiento pequeños.

2.1.9.3 Inception V3

Una red Inception es una red formada por módulos apilados unos sobre otros, con capas ocasionales de agrupación máxima para reducir a la mitad la resolución de la red (Szegedy, Reed, et al., 2014). La idea principal de la arquitectura se basa en averiguar cómo una estructura local dispersa óptima en una red de visión convolucional puede ser aproximada y cubierta por componentes densos fácilmente disponibles (Salama & Aly, 2021). Esta red contiene 48 capas y 24 millones de parámetros, por lo tanto, su costo computacional es muy alto. En gran parte estas capas están agrupadas en 11 módulos, en los cuales se extraen características a múltiples niveles. Cada módulo se compone de una configuración determinada de capas convolucionales y de agrupamiento rectificadas por la función ReLu (Gómez-Zuleta et al., 2021) y la regularización adicional con el suavizado de etiquetas de los clasificadores auxiliares normalizados por lotes permiten entrenar una red de alta calidad en conjuntos de entrenamiento de tamaño relativamente modesto (Maeda Gutiérrez & Galván Tejada, 2020).

2.1.9.4 Residual Networks (ResNet)

Los modelos ResNet fueron desarrollados por (Kaiming He, 2016), surgieron como una familia de arquitecturas profundas. Estos modelos obtuvieron el primer lugar en el ILSVRC y Common Objects in Context (COCO), se diferencian de otras arquitecturas en cuanto a la omisión de las conexiones y el uso excesivo de las capas ReLu (Kazemi, 2017). ResNet se construyó con varias unidades residuales aplicadas y se desarrolló con diferentes números de capas: 18, 34, 50, 101, 152 y 1202 (Maeda Gutiérrez & Galván Tejada, 2020).

El modelo ResNet18 (**Figura 6**) tiene un total de 17 capas convolucionales y una capa totalmente conectada, este modelo presenta una buena compensación entre la profundidad y rendimiento (Damaševi̇, 2021).

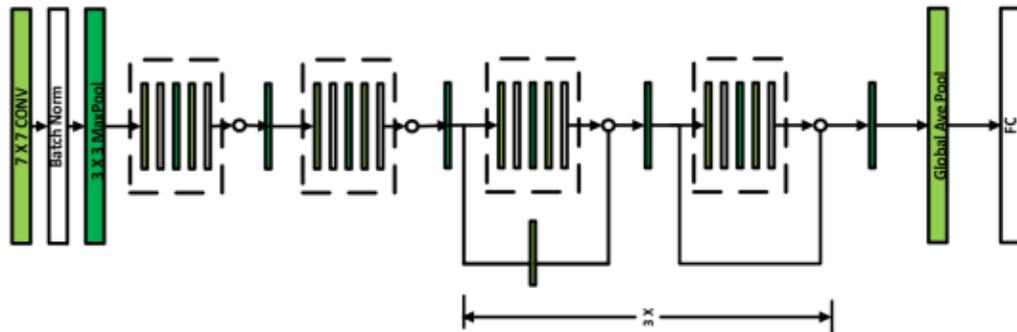


Figura 6. Arquitectura del modelo ResNet 18 (Damaševi̇, 2021).

La ResNet 50, también conocida como red residual profunda, es uno de los modelos pre entrenados que utiliza ImageNet (Base de datos con más de 15 millones de imágenes de alta resolución etiquetadas que pertenecen a unas 22,000 categorías), se compone de 50 capas y 26 millones de parámetros (Alex Krizhevsky, Ilya Sutskever, 2012). Es común que en arquitecturas muy profundas como la mencionada, el gradiente propagado se desvanezca en las últimas capas, para evitar esto, ciertas capas son entrenadas con el residuo del gradiente obtenido en la capa actual y el gradiente de una capa dos posiciones antes y por lo general, finaliza con una función de activación exponencial normalizada (Gómez-Zuleta et al., 2021; Salama & Aly, 2021).

2.1.10 Métricas de evaluación del desempeño

Se utilizan para medir el rendimiento de los modelos en la clasificación de datos y para optimizar el algoritmo de clasificación durante la etapa de entrenamiento. En otras palabras, las métricas juzgan la eficiencia del clasificador, por ejemplo, como evaluador, dentro de la etapa de prueba del modelo utilizando datos ocultos. Actualmente, existen una variedad de métricas, sin embargo, cada una de ellas proporciona información específica de una característica dentro del rendimiento del clasificador, por lo que es importante seleccionar de manera correcta el conjunto de métricas a utilizar.

Por otro lado, cuando se habla de los conjuntos de datos que se utilizan para entrenar a las CNNs las distribuciones de datos sesgadas surgen de forma natural en muchas aplicaciones, lo cual produce un desequilibrio intrínseco debido a las frecuencias naturales de los datos en donde la clase positiva se produce con una frecuencia reducida, incluidos los datos que se encuentran en el diagnóstico de enfermedades (Johnson & Khoshgoftaar, 2019). Así pues, es necesario seleccionar adecuadamente aquellas métricas que mejor representen el rendimiento de los modelos.

La matriz de confusión (**Figura 7**) es un mecanismo para visualizar el rendimiento de un clasificador, las filas representan la predicción del clasificador, mientras que las columnas representan el valor actual de cada clase. La matriz de confusión contiene cuatro parámetros que son:

- Verdaderos positivos (TP): El número de instancias positivas que han sido clasificadas con éxito.
- Verdaderos negativos (TN): El número de instancias negativas que han sido clasificadas con éxito.
- Falsos positivos (FP): El número de instancias positivas que han sido mal clasificadas.
- Falsos negativos (FN): El número de instancias negativas que han sido mal clasificadas.

En términos médicos, positivo significa que el paciente tiene lesiones anormales o que tiene el virus, mientras que el TP significa que los pacientes con lesiones anormales se someten a pruebas según un método de cribado de pruebas predefinido, y se etiquetan correctamente como lesiones anormales. Mientras que el FP se define como un diagnóstico médico erróneo de un paciente sin lesiones anormales. Negativo significa que el paciente está saludable o que no presenta una lesión anormal, y TN es el paciente sin lesiones y que es diagnosticado como normal. FN se define como la condición en donde el paciente con una lesión anormal es etiquetado como saludable, lo cual es una condición que causa un diagnóstico erróneo (Wang et al., 2019).

		Valor real	
		+	-
Valor predicho	⊕ Positivos	Verdadero Positivo (TP)	Falso Positivo (FP)
	⊖ Negativos	Falso Negativo (FN)	Verdadero Negativo (TN)

Figura 7. Matriz de confusión.

De acuerdo con lo anterior, se puede lograr un análisis más detallado basado en la combinación de los parámetros de la matriz de confusión para obtener diferentes métricas.

1. Exactitud: Calcula el ratio del número de instancias correctamente clasificadas y el número total de muestras evaluadas (**Ecuación 10**).

$$Exactitud = \frac{TP + TN}{TP + TN + FP + FN} \quad (10)$$

Sin embargo, cuando se trabaja con conjuntos de datos no balanceados la exactitud no es suficiente para describir el rendimiento del modelo ya que el valor resultante es dominado por la clase mayoritaria, es decir, la clase negativa (Johnson & Khoshgoftaar, 2019). Por lo que conviene añadir más métricas para evaluar el rendimiento.

2. Precisión: Mide el porcentaje de muestras etiquetadas positivamente que son realmente positivas. La precisión es sensible al desequilibrio de clases debido a que considera el número de instancias negativas incorrectamente etiquetadas como positivas (**Ecuación 11**).

$$Precisión = \frac{TP}{TP + FP} \quad (11)$$

Aun así, la precisión por sí sola es insuficiente para evaluar un modelo, ya que no proporciona información sobre el número de muestras del grupo positivo que fueron etiquetadas erróneamente como negativas.

3. Sensibilidad: También llamada “Recall” o “Tasa de verdaderos positivos” mide el porcentaje de instancias positivas que han sido correctamente clasificadas, es decir, es la probabilidad de clasificar las clases positivas correctamente (**Ecuación 12**).

$$\text{Sensibilidad} = \frac{TP}{TP + FN} \quad (12)$$

La sensibilidad no considera el número de muestras negativas que han sido clasificadas incorrectamente como positivas, lo cual puede ser problemático en situaciones donde se tenga un gran número de instancias negativas.

4. Especificidad: También conocida como “Tasa de verdaderos negativos” mide el porcentaje de instancias negativas correctamente clasificadas (**Ecuación 13**).

$$\text{Especificidad} = \frac{TN}{FP + TN} \quad (13)$$

5. F1-score: Combina la precisión y la sensibilidad utilizando la media armónica ponderada, donde el coeficiente β se utiliza para ajustar la importancia relativa entre la precisión y la sensibilidad (**Ecuación 14**).

$$F1 - score = \frac{(1 + \beta^2) \times recall \times precision}{\beta^2 \times recall + precision} \quad (14)$$

6. Área bajo la curva ROC: La curva ROC es una representación gráfica y bidimensional del rendimiento de un clasificador. Se utiliza para realizar comparaciones entre modelos de aprendizaje, así como para construir un modelo de aprendizaje que modele de mejor manera los datos. En contraste con la probabilidad y las métricas, el AUC expone el rendimiento general del clasificador. La **Ecuación 15** se utiliza para calcular el AUC en un problema de dos clases.

$$AUC = \frac{S_p - n_p(n_n + 1)/2}{n_p n_n} \quad (15)$$

Donde S_p representa la suma de todas las muestras clasificadas como positivas. El número de muestras positivas y negativas se denota con n_p y n_n respectivamente. Comparado con la métrica de exactitud, el valor del AUC ha sido verificado empírica y teóricamente. Por lo que es muy útil para identificar una solución adecuada y evaluar el rendimiento del modelo. Sin embargo, para problemas multiclase el cálculo del AUC es muy costoso computacionalmente cuando se discrimina un gran número de soluciones.

7. El AUC multiclase es una función probabilística que mide el índice de discriminabilidad de clases no ponderado (Hand & Till, 2001). El rendimiento global de la clasificación en la separación de las clases es entonces la media de las clases sobre todos los pares de estas (**Ecuación 16**).

$$M = \frac{2}{c(c-1)} \sum_{i < j} \hat{A}(i, j) \quad (16)$$

Donde las diferentes clases han sido etiquetadas como $0, 1, 2, \dots, c-1$ ($c > 2$). Para cualquier par de clases se puede computar la probabilidad $\hat{A}(i, j)$ de que un miembro de la clase j extraído al azar tenga una probabilidad estimada menor de pertenecer a la clase i que un miembro de la clase i extraído al azar.

2.1.11 Endoscopia

El sistema digestivo humano puede verse afectado por varias enfermedades o lesiones. Por ejemplo, tres de los ocho cánceres más comunes en general se localizan en el tracto GI. En conjunto el cáncer de esófago, estómago y colorrectal representan alrededor de 2.8 millones de nuevos casos y 1.8 millones de muertes al año (Hmoud Al-Adhaileh et al., 2021; J et al., 2019). Por otro lado, los exámenes endoscópicos son los procedimientos estándar para la investigación del tracto GI. La gastroscopia (**Figura 8**) es un examen del tracto GI superior que incluye el esófago, el estómago y la primera parte del intestino delgado (duodeno) mediante el uso de un gastroscopio que se pasa a través de la boca y el esófago, mientras que la colonoscopia (**Figura 9**) abarca el intestino grueso (colon) y el recto (Pogorelov, Ranheim Randel, et al., 2017).

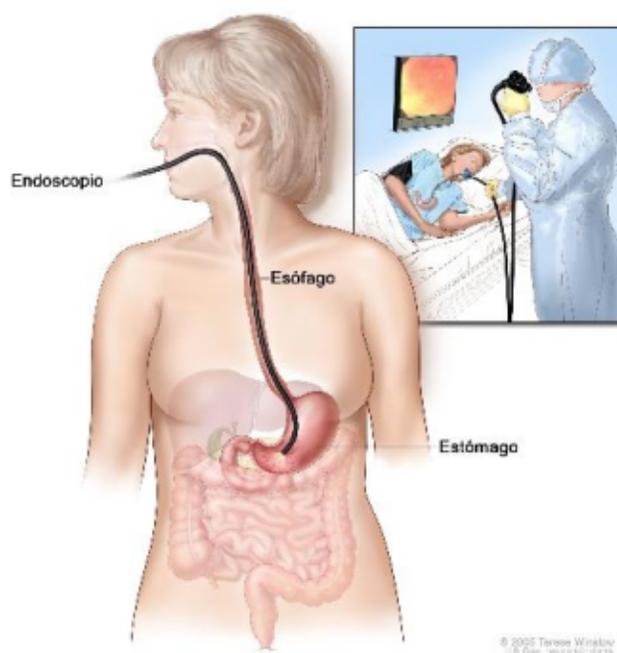


Figura 8. Gastroscopia o Endoscopia Superior (Terese Wislow, 2005).

La Sociedad Europea de Endoscopia Gastrointestinal y United European Gastroenterology (<https://ueg.eu/>) presentan una breve lista de medidas de rendimiento clave para la endoscopia gastrointestinal inferior (Kaminski et al., 2017), a continuación, se presentan las primeras cuatro.

1. El estándar mínimo es del 90% en la tasa de preparación intestinal adecuada.
2. El estándar mínimo en la tasa de intubación cecal es del 90%.
3. La norma mínima es del 25% en la tasa de detección de adenomas.
4. Técnica de polipectomía adecuada (estándar mínimo del 80%).

La intubación cecal se define como el paso de la punta del colonoscopio hasta un punto próximo a la válvula ileocecal y la visualización de todo el ciego. La tasa de intubación cecal es otro indicador de calidad y se define como la proporción de exámenes completos, ajustada por los exámenes incompletos debidos a una preparación muy deficiente o a una estenosis (engrosamiento de la apertura entre el estómago y el intestino delgado) causada por un tumor, el examen se considera completo cuando el endoscopista haya registrado la identificación de los puntos de referencia cecales o la intubación del fleon terrestre (Kaminski et al., 2010).



Figura 9. Colonoscopia o Coloscopia (Terese Wislow, 2005).

La tasa de detección de adenomas (TDA) es un indicador de calidad conocido durante la realización de la colonoscopia y es definido como la proporción de pacientes a los que se les detecta uno o más adenomas durante el procedimiento, y se considera aceptable una tasa global de detección mayor del 25% (Cepeda-Vásquez, 2021; Gómez-Zuleta et al., 2021).

Esta tasa de detección tiene una correlación inversa con la incidencia del CCR de intervalo y con la tasa de mortalidad asociada al CCR (Fernández-Esparrach & García-Rodríguez, 2020). En la **Figura 10**, se resume lo anterior expuesto y se indican los procesos de desarrollo de las medidas de rendimiento faltantes de descripción.

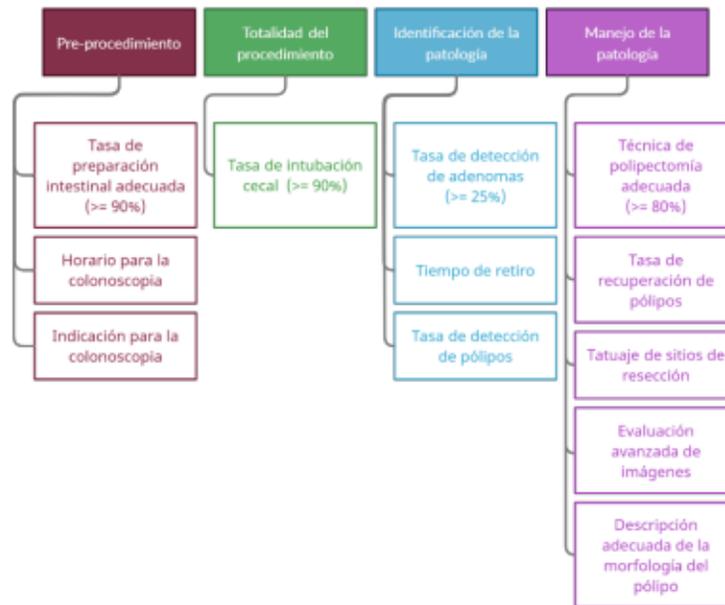


Figura 10. Proceso de desarrollo de las medidas de rendimiento (Kaminski et al., 2017).

2.1.12 Puntos de referencia anatómicos

Un punto de referencia anatómico es una característica reconocible dentro del tracto GI que es fácilmente visible a través del endoscopio. Son esenciales para la navegación y como punto de referencia para describir la ubicación de un determinado hallazgo. Los puntos de referencia también pueden ser lugares típicos de patologías como úlceras o inflamaciones (Pogorelov, Ranheim Randel, et al., 2017).

2.1.12.1 Línea Z

La línea Z (**Figura 11**) es la unión anatómica entre el epitelio escamoso del esófago y el epitelio columnar del estómago. El reconocimiento y evaluación de la línea Z es importante para determinar si una enfermedad está presente o no (Borgli et al., 2020).



Figura 11. Línea Z (Borgli et al., 2020).

2.1.12.2 El píloro

El píloro (**Figura 12**) es la unión anatómica entre el estómago y el intestino delgado y es un esfínter (músculo circular) que regula el vaciado del estómago en el duodeno. La identificación del píloro es necesaria para la instrumentación endoscópica al duodeno, una de las maniobras más difíciles de la gastroscopia (Smedsrud et al., 2021).



Figura 12. Píloro (Borgli et al., 2020).

2.1.12.3 El ciego

El ciego (**Figura 13**) es la parte más proximal del intestino grueso. Llegar al ciego es la prueba de una colonoscopia completa y la tasa de finalización ha demostrado ser un indicador de calidad valido para la colonoscopia.

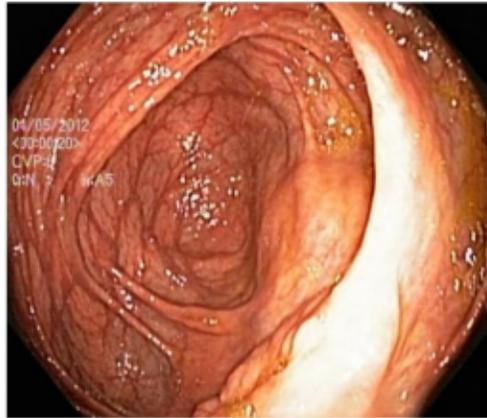


Figura 13. Ciego (Borgli et al., 2020).

2.1.13 Hallazgos Patológicos

Un hallazgo patológico es una característica anormal dentro del tracto GI. Endoscópicamente, es visible como un daño o cambio en la mucosa normal. A continuación, se describen algunos.

2.1.13.1 Esofagitis

La esofagitis (**Figura 14**) es una inflamación del esófago visible como una rotura de la mucosa esofágica en relación con la línea Z. La principal prueba diagnóstica es la endoscopia, que permite clasificar la esofagitis.



Figura 14. Esofagitis (Borgli et al., 2020).

El grado de inflamación se define por la longitud de las roturas de la mucosa y la proporción de la circunferencia afectada, de acuerdo con lo anterior se tiene la siguiente clasificación (Barcenilla Laguna et al., 2000).

1. Grado I: Placas blanquecinas de hasta 2 milímetros (mm) localizadas irregularmente con mucosa circundante normal.
2. Grado II: Mayor presencia de placas, o pseudomembranas, que tienden a confluir, con hiperemia y edema. No existen ulceraciones.
3. Grado III: Placas elevadas confluentes que abarcan toda la circunferencia, con hiperemia mucosa y ulceraciones.
4. Grado IV: Asocia a lo anterior necrosis y estenosis de la luz esofágica.

2.1.13.2 Esófago de Barret

El esófago de Barret (**Figura 15**) es una patología adquirida producto del reflujo gastroesofágico crónico que provoca la lesión de la mucosa esofágica normal y su remplazo por mucosa metaplásica. Representa una transformación metaplásica del epitelio escamoso del esófago en un epitelio columnar similar al gástrico y es considerada como una condición precancerosa (Borgli et al., 2020; R. & G., 2015).

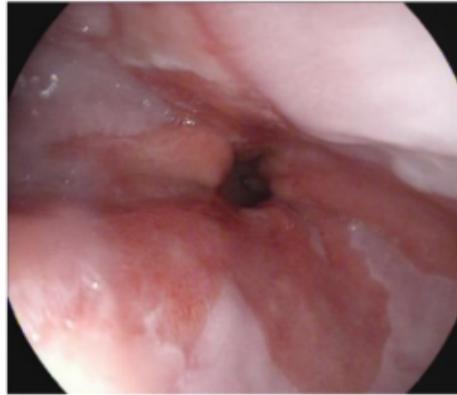


Figura 15. Esófago de Barret (Borgli et al., 2020).

2.1.13.3 Hemorroides

Las hemorroides (**Figura 16**) son dilataciones de las venas en el ano o en la parte inferior del recto. Cuando están presentes en el recto, se denominan hemorroides internas, y cuando se encuentran en el ano, se llaman hemorroides externas. La inflamación genera una obstrucción en el flujo de la sangre, que va acompañada de hipertrofia e infiltración de células redondas de tejido conjuntivo perivascular (Olga de Conte Vila, 2003).

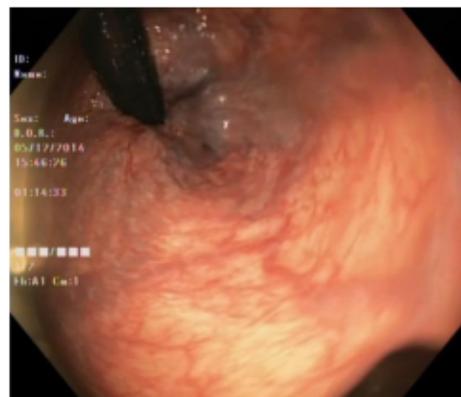


Figura 16. Hemorroides (Borgli et al., 2020).

2.1.13.4 Pólipos

Los pólipos (**Figura 17**) son tumores circunscritos que protruyen (sobresalen de sus límites) desde la pared hacia la luz intestinal. En relación con el número de lesiones pueden ser únicos o múltiples, cuando el número de pólipos es muy elevado se habla de poliposis intestinal. Se ha demostrado de manera inequívoca que el adenoma del colon constituye una lesión premaligna. La probabilidad de transformación carcinomatosa aumenta en relación con el tamaño de la lesión, la proporción del componente veloso y la edad (Moreira et al., 2021).



Figura 17. Pólipo (Borgli et al., 2020).

2.1.13.5 Colitis ulcerosa

La colitis ulcerosa (**Figura 18**) es una enfermedad inflamatoria crónica que afecta al colon, con periodos alternados de remisión y actividad. Las exacerbaciones pueden ser graves y asociadas a complicaciones y mortalidad. El grado de inflamación varía desde ninguno, leve, moderado y severo (Hernández-Rocha et al., 2017).



Figura 18. Colitis Ulcerosa (Borgli et al., 2020).

2.2 Principales estudios relacionados

En los últimos años, se han producido notables avances en el campo del reconocimiento computacional de imágenes, especialmente en el área médica, donde se han implementado diferentes técnicas de DL para la clasificación automática de lesiones gastrointestinales, un ejemplo de ello es el trabajo de (Pogorelov, Ranheim Randel, et al., 2017) donde presenta una clasificación multiclase utilizando Kvasir como conjunto de datos. El conjunto de datos contiene 4,000 imágenes y 8 clases diferentes anotadas y verificadas por médicos expertos, incluyendo lugares anatómicos, hallazgos patológicos y procedimientos endoscópicos. Se utilizan varias configuraciones de tres enfoques diferentes, el primer enfoque utiliza Bosques Aleatorios (Macaulay et al., 2021) y Árbol de Modelo Logístico (Landwehr et al., 2005). El segundo enfoque utiliza CNNs con una función de activación ReLu y agrupación máxima. El tercer enfoque se basa en aprendizaje por transferencia, es decir, la utilización de CNNs pre-entrenadas con la implementación del Descenso de Gradiente Estocástico (Hong et al., 2020) para conseguir el mejor rendimiento en términos de velocidad y exactitud. Cabe mencionar que no se utilizó ningún esquema de aumento de datos, sin embargo, se implementa una validación cruzada doble (técnica para la validación de datos mediante el entrenamiento de subconjuntos de datos) como estrategia para evaluar sus resultados. El enfoque con mejor rendimiento fue el Árbol de Modelo Logístico con una exactitud del 93.70%, el cual combinó todas las características extraídas, dando como resultado un vector de 1186 características.

El trabajo de (Petschamig et al., 2017), propone dos arquitecturas de CNNs con la particular característica de utilizar un módulo “inception” con el objetivo de disminuir el costo computacional. La idea básica del módulo inception es que la red puede seleccionar al momento del entrenamiento si la agrupación, la convolución pequeña o la convolución más amplia se adaptan mejor a los datos subyacentes. Se utiliza Kvasir como conjunto de datos, una arquitectura basada en GoogLeNet (CNN con 22 capas de profundidad) y un esquema de aumento de datos (Monshi et al., 2021) para incrementar el número de imágenes. En general, la arquitectura proporciona resultados aceptables incluso con pocos datos de entrenamiento, sin embargo, los autores concluyen que el modelo donde utilizan 2048 neuronas en las capas profundas sufre de sobreajuste y produce un menor rendimiento, caso

contrario sucede con el modelo donde utilizaron 1024 neuronas, obteniendo una exactitud general del 93.90%.

De igual manera, (Agrawa et al., 2017) utiliza Kvasir e implementa una combinación de CNNs pre-entrenadas con ImageNet (Russakovsky et al., 2015). Se emplea la configuración de 16 capas de VGGNet (Simonyan & Zisserman, 2015) como un extractor de características, utilizando las salidas de la primera capa totalmente conectada como características para la clasificación. Se utiliza una red Inception-v3 para extraer las características y finalmente, se implementa el algoritmo de Máquina de Soporte de Vectores (SVM) (Badr et al., 2021) para la clasificación multiclase empleando diferentes configuraciones. Los hiperparámetros (parámetros que afectan directamente el rendimiento del modelo, un ejemplo es la tolerancia al error) del clasificador SVM se ajustaron utilizando un marco de validación cruzada de 5 pliegues en el conjunto de datos de entrenamiento. El resultado fue una exactitud general del 96.10% utilizando una combinación de todas las características y con una partición de los datos del 80-20% para el entrenamiento y la prueba respectivamente.

Por otro lado, (T. Hoang et al., 2018) combina Kvasir y Nerthus (Pogorelov, Randel, et al., 2017) para clasificar 16 diferentes clases. Se implementa una ResNet con 101 capas para extraer características del conjunto de datos original, pero ampliado con instrumentos. Después de pasar por ResNet 101, las imágenes de salida clasificadas como clases especiales se convierten en la entrada de la red Faster R-CNN (B. L. Chen et al., 2021), que se entrena para detectar instrumentos en las imágenes. Finalmente, esta configuración obtuvo una exactitud de 99.33% y un F1 score del 94.6%, demostrando que la utilización de CNNs pre-entrenadas y un esquema de aumento de datos logran buenos resultados en la clasificación de imágenes endoscópicas del tracto GI.

La investigación de (Takiyama et al., 2018), presenta la implementación de una CNN para identificar la ubicación de puntos de referencia anatómicos en imágenes de Esofagogastroduodenoscopia (EDG). El conjunto de datos consta de 27,335 imágenes para el entrenamiento y 17,801 imágenes independientes para la validación. La CNN está basada en GoogLeNet y se redimensionó cada imagen a 244×244 píxeles, además, se utilizó Adam como optimizador (Alzubaidi et al., 2021). En conclusión, el sistema obtuvo una exactitud del 97%, demostrando el potencial de las CNNs en la clasificación de imágenes.

El trabajo de (Cogan et al., 2019), propone un marco para el Procesamiento Modular y Automático de Imágenes del Tracto GI (MAPGI), que aplica la eliminación de bordes, la mejora del contraste, el filtrado, el mapeo de color y el escalado de cada imagen del conjunto de datos, en este caso, Kvasir. Se utiliza una corrección gamma (Mcreeynolds & Blythe, 2005) para compensar las características de visualización no lineales, es decir, se calculan los valores para las imágenes individuales, de forma que el valor medio de los píxeles de cada imagen se normaliza a 90 ± 1 en un rango de valores de 0-255 píxeles. Posteriormente, se entrenan tres CNNs, Inception-ResNet-V2, Inception-V4 y NASNet (Zoph et al., 2018) con el 85% de los datos, mientras que el otro 15% se reservó para la prueba y se obtuvo una exactitud del 98.45%, 98.48% y 97.35% respectivamente.

En la investigación de (Luo et al., 2019), se propone una validación cruzada de 10 pliegues para entrenar 10 modelos de CNNs pre-entrenados en ImageNet, DenseNet 161, ResNet 101 e Inception V4 son algunos de ellos, y se emplea un esquema de aumento de datos para aliviar el problema de sobre ajuste. Cada uno de estos modelos se utilizó para obtener la probabilidad de los mapas de predicción, que luego se combinaron y utilizaron como datos para el aprendizaje de un modelo de conjunto adaptativo (técnica muy utilizada para mejorar la precisión de diversas tareas de aprendizaje automático). Por último, se verificó el rendimiento de la técnica propuesta y se obtuvo una exactitud de 99.33% y una puntuación de 94.80% en el Coeficiente de Correlación de Matthews (MCC) sin embargo, la velocidad fue solo de 41.51 cuadros por segundo (FPS).

Por otra parte, (Chang et al., 2019) se basó en el aprendizaje de diferentes representaciones de características para imágenes de etiquetas múltiples utilizando modelos basados en CNNs, incluyendo ResNet-34, SE-ReNeXt (Xie et al., 2017) y attention-Inception V3 (Szegedy, Vanhoucke, et al., 2014). Los modelos fueron entrenados utilizando técnicas de fusión multi-época (una época es el tiempo que tarda el modelo en procesar cada partición del conjunto de datos) y de umbralización adaptativa (permite establecer un valor en base a las características locales del entorno) con un esquema automático de aumento de datos. De acuerdo con la configuración anterior, se obtuvo una exactitud de 99.46%, un F1 score de 90.07% y un MCC de 95.20%.

Actualmente, el análisis de imágenes durante la colonoscopia empieza a lograr una creciente aceptación como herramienta para predecir con precisión la histología de lesiones precancerosas, de modo que estas lesiones podrían ser tratadas en un procedimiento post-colonoscopia. En este sentido, (Byrne et al., 2019) presenta una CNN profunda basada en una arquitectura Inception para la evaluación en tiempo real de imágenes de videos endoscópicos de pólipos colorrectales. Para el entrenamiento se utilizaron videos endoscópicos no alterados de pólipos, también se utilizaron videos de mucosa normal sin pólipos y se implementó SDG como método de optimización, junto con normalización por mini lotes construidos a partir de 128 fotogramas seleccionados aleatoriamente. Para cada fotograma se aplicó un esquema de aumento de datos para obtener mayor diversidad de ejemplos. En conclusión, el modelo no generó suficiente confianza para predecir la histología de 19 pólipos del conjunto de prueba, que representan el 15% del total de pólipos, para los 106 pólipos restantes el modelo presentó una exactitud del 94% y una especificidad del 83%.

De igual manera, (Xu et al., 2019) propone una red de Detección de Anatomía Multitarea en tiempo real (MT-AD) basada en CNN para la inspección de gastroscopia. MT-AD puede realizar dos tareas simultáneamente: (1) clasificar tres tipos de imágenes, incluyendo imágenes informativas y no informativas bajo gastroscopia de Luz Blanca (WL) (F. Habib & Jobe, 2019) e Imagen de Banda Estrecha (NBI) (Li et al., 2012). Las imágenes no informativas son aquellas que presentan elementos específicos, como el desenfoque por movimiento, los reflejos especulares y burbujas. (2) detectar anatomías gástricas en tiempo real, incluyendo el esófago, la línea dentada, el píloro, el bulbo duodenal entre otras. Para la detección se implementa un Detector Multi-Caja de un Solo Disparo (SSD) (Liu et al., 2015) con una tasa de aprendizaje del 0.0005 y un tamaño de 8 para el número de lotes como parámetros iniciales. Para la clasificación se utilizó una arquitectura de siete capas convolucionales, una capa de agrupación y una capa totalmente conectada y los parámetros se inicializaron aleatoriamente. Dando como resultado una exactitud general del 98.77% y un F1-score para imágenes informativas, no informativas y NBI del 99.01, 99.13 y 98.18% respectivamente.

Por otro lado, (Igarashi et al., 2020) emplea una arquitectura AlexNet para clasificar un total de 85,246 imágenes crudas obtenidas del Hospital Universitario de Hirosaki. Las imágenes fueron clasificadas manualmente en 14 categorías según la clasificación de patrones principales por órganos anatómicos. Para entrenar el modelo se utilizaron 49,174 imágenes de pacientes con cáncer gástrico que fueron sometidos a endoscopias superiores del tracto GI, asimismo, para evaluar el rendimiento del modelo se utilizaron 36,072 imágenes. Finalmente, el modelo obtuvo una exactitud general de 96.5% y, de acuerdo con los autores, el sistema puede ser utilizado en endoscopias rutinarias para la clasificación de imágenes.

El trabajo de (Borgli et al., 2020), presenta HyperKvasir, una base de datos de uso libre, la base de datos cuenta con un total de 110,079 imágenes y 374 videos de diferentes exámenes del tracto GI. El estudio fue aprobado por la Autoridad de Protección de Datos de Noruega, asimismo, se eliminaron todos los metadatos y se renombraron todos los archivos aleatoriamente, además, los datos no fueron preprocesados ni se le aplicaron métodos de aumento de datos. Los archivos son imágenes etiquetadas, imágenes segmentadas, imágenes sin etiquetar y videos etiquetados con un total de 40 clases, 16 clases para el tracto GI superior y 24 para el tracto GI inferior. Para probar la calidad técnica del conjunto de datos se realizaron diferentes experimentos con modelos CNNs para clasificar las imágenes y se midió el rendimiento con diferentes métricas para dar idea de las cualidades estadísticas del conjunto de datos. El enfoque con mejores resultados fue la combinación entre ResNet-50 y DenseNet-161 ambas pre-entrenadas, se utilizó el promedio de ambos modelos para clasificar al conjunto de imágenes etiquetadas, que cuenta con 23 clases y 10,662 imágenes. Ambas CNNs fueron entrenadas con 50 épocas y un tamaño de lote de 32, asimismo, se utilizó SGD como método de optimización. Por último, se utilizó el micro y macro promedio para la evaluación de los modelos, además, se utilizaron métricas de clasificación estandarizadas, lo cual, entregó un resultado de una exactitud de 91% para el micro y 63.3% para el macro promedio.

Otro trabajo es el presentado por (Thambawita et al., 2020), donde se propone la implementación de 5 diferentes enfoques de ML para la clasificación de 16 condiciones del tracto GI, que van desde enfoques de Características Globales (GF) (características calculadas sobre toda la imagen o que cubren una subsección regular de una imagen), hasta

enfoques simples de redes neuronales profundas (DNN) y enfoques complejos con DNN con aprendizaje de transferencia. El conjunto de datos se dividió en 70% para el entrenamiento y 30% para la validación, también, se implementó la técnica de validación cruzada para garantizar que los datos de entrenamiento y prueba son independientes, asimismo, se utiliza un conjunto de métricas para evaluar los modelos, la exactitud, especificidad y F1-score son algunas de ellas. Finalmente, el modelo que mostró mejores resultados fue el uso de dos CNNs en paralelo, que son ResNet-152 y DenseNet-161 combinadas con una multicapa perceptrón (MLP), el cual obtuvo una exactitud general de 99.32% y un MCC de 94.21%.

Por otra parte, (Smedsrud et al., 2021) presenta Kvasir-Capsule, un conjunto de datos de Video-Cápsula Endoscópica (VCE), el cual consiste en 117 videos recolectados de exámenes endoscópicos con un total de 14 clases diferentes de hallazgos y 47,238 imágenes etiquetadas. Una VCE se compone de una pequeña cápsula que contiene una cámara de gran angular, fuentes de luces, baterías y otros componentes electrónicos. Se entrenaron dos CNNs, DenseNet-161 y ResNet-152 para llevar a cabo la validación técnica del conjunto de datos etiquetados, se implementó una validación cruzada utilizando la pérdida de entropía cruzada categórica (Amorós Becerra, 2021) con y sin ponderación de clase, asimismo, se utilizó el muestreo ponderado, que equilibra el conjunto de datos añadiendo y eliminando imágenes para cada clase. Por último, se calculan la precisión, la exhaustividad, F1-score y el MCC, donde el mejor resultado fue el promedio de ambas CNNs con 73.66% y 29.94% de exactitud para el micro y macro promedio. De acuerdo con los resultados obtenidos, los autores mencionan que la clasificación de VCE es todo un reto, además, existe un gran margen de mejora en la implementación de modelos de DL para la clasificación automática de hallazgos patológicos.

El trabajo de (Gómez-Zuleta et al., 2021), presenta una metodología de aprendizaje profundo para la detección automática de pólipos en procedimientos de colonoscopia, Inception-V3, ResNet-50 y VGG-16 fueron los modelos asignados para esta tarea. Para la clasificación, se utiliza un enfoque de aprendizaje por transferencia, es decir, las redes fueron pre-entrenadas con ImageNet y los pesos resultantes son utilizados para iniciar el nuevo proceso de entrenamiento con imágenes de colonoscopia empleando la técnica de ajuste fino. El esquema de entrenamiento fue una separación de los datos del 70% para

entrenar y 30% para validar, en el cual, se combinaron 5 diferentes bases de datos, con un total de 23,831 y 47,013 cuadros con y sin pólipos para la validación. Se implementaron diferentes métricas para medir los resultados, exactitud, F1-score, curva ROC, son algunas de ellas. Por último, los modelos Inception-V3, ResNet-50 y VGG-16 obtuvieron una exactitud de 81%, 77% y 73% respectivamente. De acuerdo con los autores, es notable que estos modelos generalizan bien la alta variabilidad de los videos de colonoscopia, además, este método, puede servir como apoyo a futuras generaciones de gastroenterólogos.

Por otro lado, (Yogapriya et al., 2021) implementa el aprendizaje por transferencia para entrenar tres redes, VGG-16, ResNet-18 y GoogLeNet para clasificar 8 diferentes clases de la base de datos Kvasir con un total de 6,702 imágenes. Se utiliza una técnica de aumento de datos para compensar aquellas clases que contengan una menor cantidad de imágenes, además, se emplea Adam como optimizador, así como el enfoque de pérdida por entropía cruzada categórica, con el objetivo de minimizar el error. Después de 30 épocas, la red que mejor comportamiento presentó en la clasificación fue la VGG-16 con una exactitud de 96.3%. Con base a los resultados, los modelos CNN muestran que pueden ayudar a mejorar el avance de la tecnología en el análisis de imágenes médicas, ya que permiten generalizar y adquirir características muy valiosas mapeando la información en imágenes similares.

De igual manera, (Hmoud Al-Adhaileh et al., 2021) utiliza tres redes para evaluar su potencial en la clasificación de imágenes médicas en Kvasir como base de datos. Primeramente, se aplica un preprocesamiento para eliminar ruido de las imágenes y mejorar su calidad, al igual que una técnica de aumento de datos para mejorar el proceso de entrenamiento y una técnica de abandono (Dropout) para no sufrir sobreajuste, sin embargo, los autores mencionan que con esta técnica el tiempo de entrenamiento se duplicó. También, se utiliza Adam como optimizador para reducir la pérdida o el error, asimismo, una técnica de aprendizaje por transferencia y ajuste fino. Finalmente, se implementan para clasificar un total de 5,000 imágenes con un total de 5 clases y una división del 80% de la base de datos para entrenamiento y 20% para validación. Los modelos obtuvieron una exactitud de 96.7%, 95% y 97% para GoogLeNet, ResNet-50 y AlexNet, respectivamente.

De acuerdo con lo anterior, se puede observar una clara preferencia por el uso de técnicas de DL para la clasificación de lesiones gástricas, así como la diversidad de enfoques que existen para abordar la clasificación, sin embargo, sus resultados varían de un enfoque a otro. Las CNNs muestran resultados sólidos y una gran capacidad de adaptación para extraer características importantes de las imágenes de lesiones gástricas, además, con la implementación de técnicas de optimización se puede mejorar el desempeño significativamente de las mismas. No obstante, es necesario comparar diferentes arquitecturas para poder distinguir la arquitectura que mejor modele el problema de clasificación de lesiones del tracto GI. En este sentido, el presente trabajo presenta un estudio comparativo entre AlexNet, DenseNet-201, Inception-V3 y ResNet-101, seleccionadas de acuerdo con el comportamiento significativo y a sus resultados reportados.

A continuación, la **Tabla 1** muestra un resumen del estado del arte de las CNNs en la clasificación de lesiones gástricas.

Tabla 1. Comparación de los diferentes estudios de clasificación de lesiones gástricas, aquí, SGD = Stochastic Gradient Descent, ACC = Accuracy, MCC = Matthews Correlation Coefficient, SPEC = Specificity, PREC = Precision, SEN = Sensitivity.

Autores	Arquitectura (s)	Optimizador	GPU/CPU	Métrica (s)	Resultado
Pogorelov et al., 2017	Inception-V3	SGD	-	PREC, SPEC, ACC, MCC, F1- score	0.924
Petschnig et al., 2017	GoogLeNet	-	GeForce GTX Titan X	PREC, SPEC, ACC, MCC, F1- score	0.939
Agrawa et al., 2017	Inception-V3, VGGNet	SGD	-	ACC, MCC, F1-score	0.961
Hoang et al., 2018	ResNet-101, Faster R-CNN	Adam	Tesla K80	PREC, SPEC, ACC, MCC, F1- score	0.993

Takiyama et al., 2018	GoogLeNet	Adam	-	ACC	0.970
Cogan et al., 2019	Inception-ResNet-V2, Inception-V4, NASNet	-	-	ACC	0.985
Luo et al., 2019	10 CNN pre-entrenadas en ImageNet	SGD	RTX 2080 Ti	PREC, SPEC, ACC, MCC, F1-score	0.994
Chang et al., 2019	ResNet-34, SE-ReNeXt, attention-Inception V3	Adam	Tesla P100	PREC, SPEC, ACC, MCC, F1-score	0.995
Byrne et al., 2019	Inception	SGD	NVIDIA	ACC, SPEC, SEN	0.940
Xu et al., 2019	CNN personalizada	-	-	PREC, ACC	0.987
Igarashi et al., 2020	AlexNet	-	-	ACC	0.965
Borgli et al., 2020	ResNet-152, DenseNet-161	SGD	-	PREC, ACC, MCC, F1-score	0.910
Thambawita et al., 2020	ResNet-152, DenseNet-161	SGD	GeForce 1080	PREC, SPEC, ACC, MCC, F1-score	0.993
Smedsrud et al., 2021	DenseNet-161, ResNet-152	-	-	PREC, MCC, F1-score	0.736

Gómez-Zuleta et al., 2021	Inception-V3, ResNet-50, VGG-16	-	-	ACC, SEN, SPEC, PREC, F1- score	0.810
Yogapriya et al., 2021	VGG-16, ResNet-18, GoogLeNet	Adam	Quadro RTX6000	ACC, F1- score, PREC	0.963
Hmoud Al-Adhaileh et al., 2021	GoogLeNet, ResNet-50 y AlexNet	Adam	4G NVIDIA	ACC, SEN, SPEC	0.970

2.3 Modelo o esquema general de investigación

El presente trabajo tiene como fundamento un enfoque cuantitativo no experimental de acuerdo con (Hernández Sampieri Roberto; Méndez Valencia Sergio; Mendoza Torres Christian Paulina & Cuevas Ana, 2017), debido a que implica la investigación sin la manipulación de ningún tipo de variable, en cambio, se analizan hechos o fenómenos ya existentes, en este caso, se estudia la implementación de CNNs en la clasificación de imágenes de lesiones gastrointestinales. Además, es de orden cuantitativo por el hecho de que la selección de la arquitectura que mejor modela el problema de clasificación es a través de la implementación de un conjunto de métricas estandarizadas, las cuales podemos medir y comparar con facilidad.

Capítulo 3. Método y propuesta de investigación

El presente capítulo describe la metodología utilizada para la clasificación de lesiones del tracto GI. De manera general, el capítulo comprende la descripción del modelo de investigación, desde la adquisición y el contenido de la base de datos, hasta la configuración de los hiperparámetros, el proceso de entrenamiento y la aplicación de diferentes métricas para evaluar el rendimiento de los modelos.

2.1 Modelo de investigación

El modelo de investigación (**Figura 19**) consta de 3 etapas, la primera etapa consiste en la adquisición de los datos, aquí se describe la construcción de la base de datos y sus elementos, así como su validación. La segunda etapa es el entrenamiento de las CNNs, contiene la configuración de los hiperparámetros, como el optimizador, la tasa de aprendizaje, el tamaño del lote y el número de épocas. También se describe el enfoque de entrenamiento y el proceso de clasificación de las imágenes, así como las herramientas que se utilizaron para la implementación de las CNNs. La tercera etapa es la evaluación de los modelos, donde se mide el rendimiento a través de la aplicación de las diferentes métricas ya mencionadas en el capítulo dos.

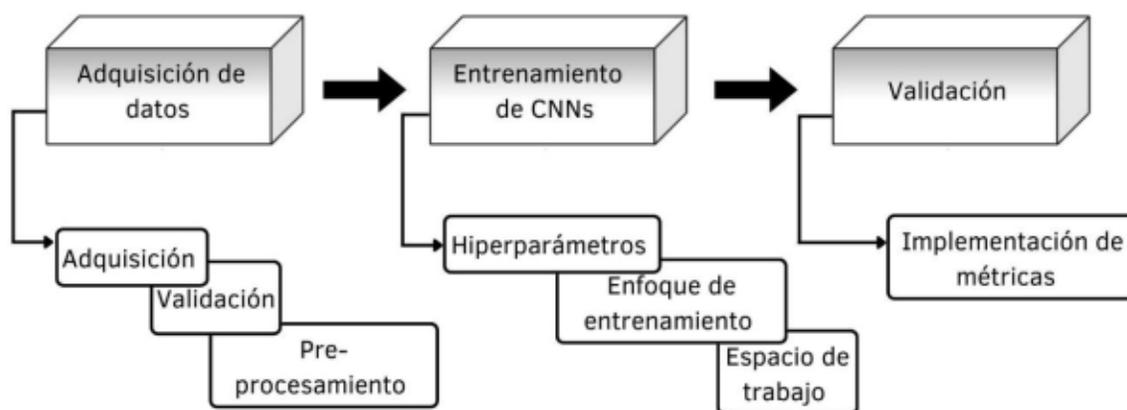


Figura 19. Diagrama de la metodología propuesta.

3.2 Descripción de la propuesta

3.2.1 Adquisición de los datos

Las imágenes fueron recolectadas usando equipo estándar de endoscopia de Olympus (<https://www.olympus-europa.com/>) y Pentax (<https://www.pentaxmedical.com/pentax/>) en el departamento de Gastroenterología del hospital de Bærum, Vestre Viken Hospital Trust, Noruega. Las anotaciones fueron realizadas por al menos un gastroenterólogo experimentado del hospital de Bærum, del centro de investigación de Registro de Cáncer en Noruega, del Hospital Universitario de Karolinska en Suecia, y una o más personas con experiencia en el campo de la medicina.

El estudio fue aprobado por la autoridad noruega de protección de datos personales y quedó exento del consentimiento del paciente debido a que los datos son totalmente anónimos. Se eliminaron todos los metadatos y se renombraron todos los archivos con nombres de archivo generados aleatoriamente antes de que el departamento de informática del hospital de Bærum exportara los archivos desde un servidor central. Así mismo, el estudio fue eximido de la aprobación del Comité Regional de Ética de la Investigación Médica y Sanitaria ya que la recopilación de los datos no interfirió con la atención prestada al paciente. Dado que los datos son anónimos, la base de datos se puede compartir públicamente según las leyes de noruega y el Reglamento General de Protección de Datos.

Queda claro que uno de los grandes desafíos de la IA en el campo de la medicina es la disponibilidad de datos ya que recuperar la información de los sistemas de atención a la salud es una tarea difícil, así como conseguir las aprobaciones de los comités médicos. En este sentido, la base de datos HyperKvasir pretende facilitar el desarrollo de IA en aplicaciones médicas.

HyperKvasir contiene un total de 110,079 imágenes (10,662 etiquetadas y 99,417 no etiquetadas) y 374 videos de diferentes exámenes gastrointestinales. Los registros se encuentran separados en 4 categorías, imágenes etiquetadas, imágenes segmentadas, imágenes sin etiquetar y videos etiquetados. La base de datos se encuentra disponible en el siguiente enlace: <https://datasets.simula.no/hyper-kvasir/>. En total, el conjunto de datos contiene 10,662 imágenes etiquetadas con un formato JPEG, del cual cuenta con 23 clases diferentes y están estructuradas

de acuerdo con la locación en el tracto GI y al tipo de hallazgo. En general, las 23 clases se encuentran separadas en 4 principales categorías: lugares anatómicos, calidad de las vistas de la mucosa, hallazgos patológicos e intervenciones terapéuticas. Sin embargo, para fines de la investigación solo se utilizan 10 diferentes clases, seleccionadas con respecto al mayor número de ejemplos por clase, ya que por lo general son las que se presentan con mayor frecuencia en los procesos de endoscopia de acuerdo con (Borgli et al., 2020).

La **Figura 20** muestra un ejemplo de cada clase del conjunto de datos utilizado, etiquetadas como: (1) ciego, (2) pólipos teñidos y levantados, (3) esofagitis grado a, (4) heces impactadas, (5) pólipos, (6) píloro, (7) retroflexión en el recto, (8) retroflexión en el estómago, (9) colitis ulcerativa de grado 3, (10) línea z.

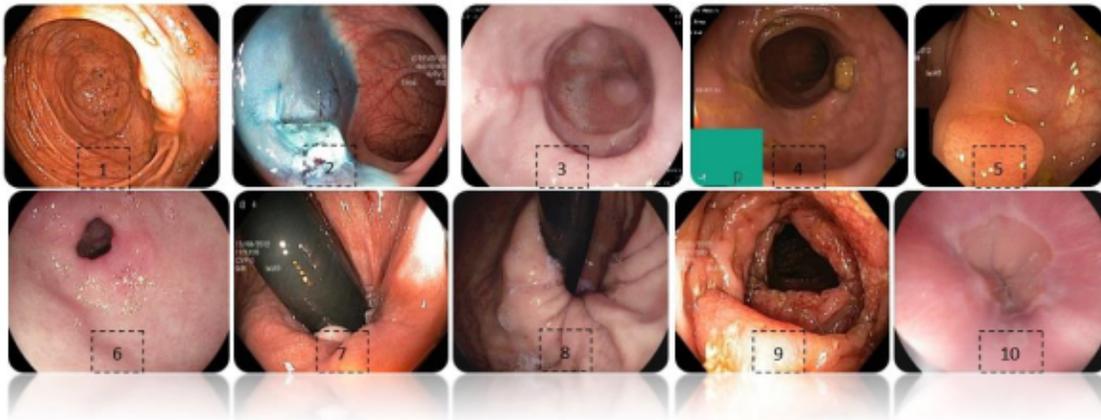


Figura 20. Diferentes clases del conjunto de datos.

Algunas imágenes presentan un recuadro en la esquina inferior izquierda de color verde, en realidad es la representación topográfica del colon. También, el número de imágenes por clase no está balanceado debido al hecho de que algunas lesiones se presentan más que otras, lo cual representa todo un reto en el ámbito médico. La **Figura 21** representa a manera de gráfico la cantidad de imágenes por clase del conjunto de datos.

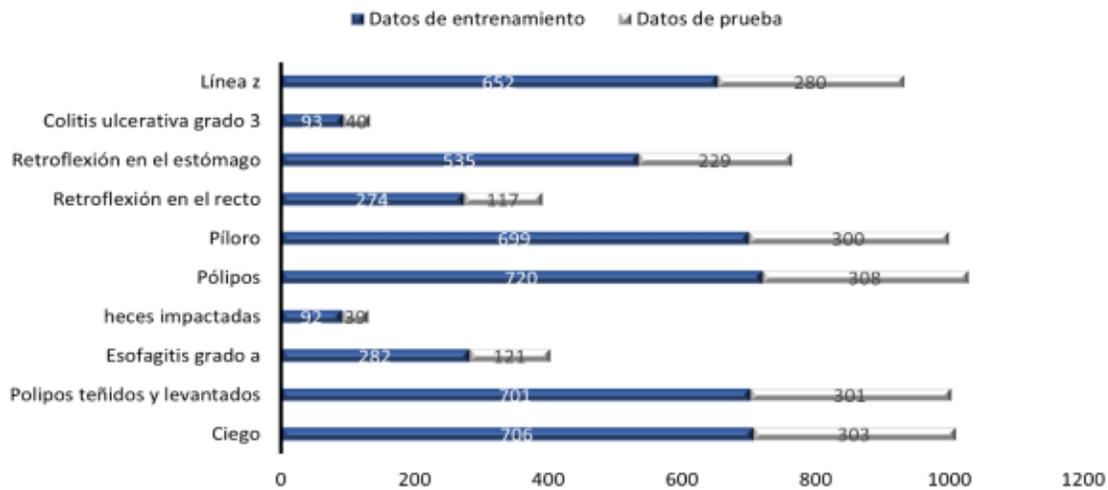


Figura 21. Número de imágenes por clase del conjunto de datos.

En total, se utilizan 6,792 imágenes para entrenar y probar el rendimiento de las CNNs, con una configuración del 70% de los datos para entrenamiento y el 30% para pruebas, de acuerdo con el trabajo de (Thambawita et al., 2020) y sus resultados reportados. La mayoría de las imágenes tienen una resolución de 768 píxeles de anchura y 576 de altura.

Sin embargo, la entrada de imagen de las CNNs tiene otro tamaño por lo que se aplica un redimensionamiento para ajustar la imagen al tamaño de entrada de la red neuronal, esto se logra por medio de una interpolación bilineal (Assad & Kiczales, 2020).

3.2.2 Entrenamiento de las CNNs

De acuerdo con la revisión no sistemática de literatura, se seleccionaron cuatro modelos de CNNs para evaluar su rendimiento en la clasificación de imágenes del tracto GI, la **Tabla 2** muestra las características de cada arquitectura.

Tabla 2. Características generales de las arquitecturas.

Arquitectura	Profundidad	Tamaño	Parámetros (millones)
AlexNet	8	227 MB	61
DenseNet-201	201	77 MB	20
Inception-v3	48	89 MB	23.9
ResNet-101	101	167 MB	44.6

Para comparar el rendimiento de las CNNs los hiperparámetros fueron estandarizados para cada uno de los modelos. Los hiperparámetros establecidos en la **Tabla 3** fueron seleccionados de acuerdo con los trabajos relacionados que mostraron mejor desempeño en la clasificación de lesiones del tracto GI.

Tabla 3. Hiperparámetros implementados. *Stochastic Gradient Descent with Momentum (SGDM).

Hiperparámetros	Valor
Tasa de aprendizaje	0.001
Tamaño del lote	16
Épocas	50
Optimizador	SGDM*
Función de pérdida	Softmax

Los modelos han sido pre-entrenados en ImageNet (Alex Krizhevsky, Ilya Sutskever, 2012), una base de datos con más de 15 millones de imágenes con alta resolución. Una vez que los modelos ya cuentan con el sesgo y los pesos ajustados se asigna a los modelos la nueva tarea de clasificación, a este proceso se le conoce como aprendizaje por transferencia, el cual ya fue descrito en la sección 2. Para poder realizar la nueva tarea de clasificación se utiliza la técnica de ajuste fino, consiste en retirar la última capa totalmente conectada de la red, y agregar una nueva que tenga el mismo número de neuronas que de clases, en este caso es igual a 10. La **Figura 22** representa este proceso de manera esquemática.

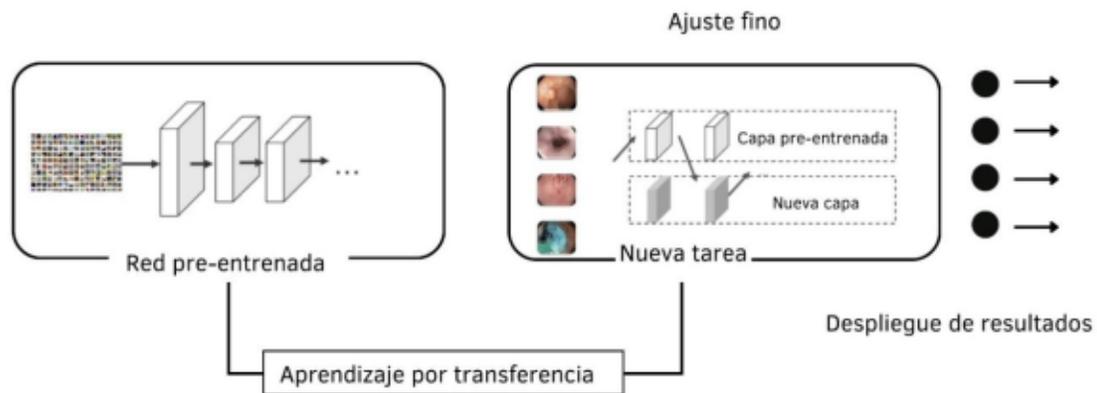


Figura 22. Representación esquemática del proceso de aprendizaje por transferencia.

Finalmente, en la etapa de clasificación se utiliza la función Softmax para calcular el error predicho sobre las muestras. Este error revela la diferencia entre la salida actual y la salida esperada.

La experimentación se realizó en MATLAB 2021a, utilizando el ToolBox de Deep Learning, el cual provee un marco de trabajo para diseñar e implementar CNNs. Los detalles del espacio de trabajo se muestran en la **Tabla 4**.

Tabla 4. Características del área de trabajo.

Hardware y Software	Descripción
Memoria RAM	16 gb
Procesador	AMD Ryzen 9 5900HX with Radeon Graphics @ 3.30 GHz
GPU	NVIDIA GeForce RTX 3060 @ 6 gb

Capítulo 4. Resultados y Limitaciones

En esta sección se presentan y discuten los resultados obtenidos por cada una de las arquitecturas. La **Tabla 5** muestra los resultados obtenidos de las métricas para cada arquitectura.

Tabla 5. Rendimiento general de las arquitecturas.

Métricas	AlexNet	DenseNet-201	Inception-v3	Resnet-101
Exactitud	0.949	0.971	0.959	0.964
Precisión	0.941	0.964	0.951	0.951
Sensibilidad	0.921	0.963	0.948	0.949
Especificidad	0.994	0.997	0.995	0.996
F1-score	0.924	0.963	0.949	0.953
AUC	0.902	0.949	0.929	0.945
Tiempo (min)	99	1005	389	338

En general, las cuatro arquitecturas muestran un rendimiento global estadísticamente aceptable, no obstante, DenseNet-201 sobresale en la mayoría de las métricas, por ejemplo, obtuvo el 97% de exactitud, lo cual indica que tiene un alto grado de confiabilidad con respecto al número de predicciones correctamente clasificadas en relación con el número total de predicciones realizadas.

En el ámbito médico, la precisión es un parámetro muy importante, ya que mide el porcentaje de las muestras positivas correctamente clasificadas. DenseNet-201 obtuvo el 96.4%, mientras que Inception-v3 y ResNet-101 obtuvieron el 95.1% y, por último, AlexNet obtuvo el 94.1% de precisión.

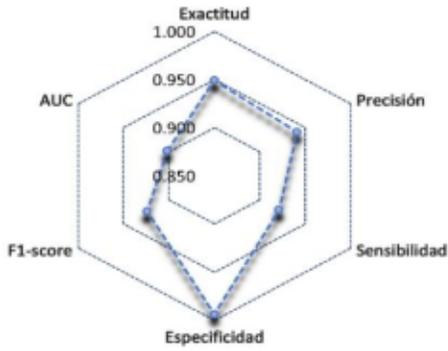
En términos de sensibilidad, DenseNet-201 obtuvo el 96.3%, ResNet-101 logró un 94.9%, Inception-v3 obtuvo el 94.8%, mientras que el menor rendimiento fue el de AlexNet con un 92.1% de sensibilidad.

En cuanto a la métrica de especificidad todas las arquitecturas lograron más del 99%, lo cual indica que tan bien clasificaron las muestras negativas.

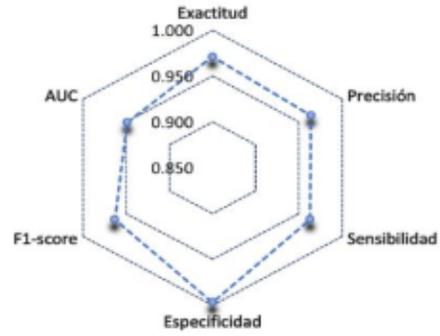
En función del F1-score, DenseNet-201 logró un 96.3%, lo que indica un buen radio entre exactitud y precisión, por lo tanto, se puede decir que la arquitectura DenseNet-201 tiene un buen balance para clasificar correctamente las muestras positivas. ResNet-101 obtuvo un 95.3%, Inception-v3 obtuvo un 94.9%, finalmente AlexNet logró un 92.4%.

En la **Figura 23** se muestra el resumen del rendimiento de las arquitecturas, se puede observar que AlexNet fue la arquitectura con el menor rendimiento en todas las métricas, a excepción de la especificidad donde todas las arquitecturas clasifican por encima del 99%. Inception-v3 y ResNet-101 casi tuvieron el mismo rendimiento, sin embargo, en términos de AUC ResNet-101 fue superior. DenseNet-201 la arquitectura con mayor número de capas obtuvo el mejor rendimiento en cada una de las métricas, sin embargo, una desventaja de trabajar con este tipo de arquitecturas es el tiempo de entrenamiento, ya que fue la arquitectura que mayor tiempo consumió, el entrenamiento de la red duró 1005 minutos. No obstante, el experimento de la presente investigación consistió en una sola implementación de las CNNs para la clasificación de lesiones del tracto GI, por lo tanto, el tiempo de entrenamiento tiene menor peso en comparación con las demás métricas. Por otro lado, sería interesante analizar el tiempo de respuesta de los modelos cuando una nueva imagen sea introducida.

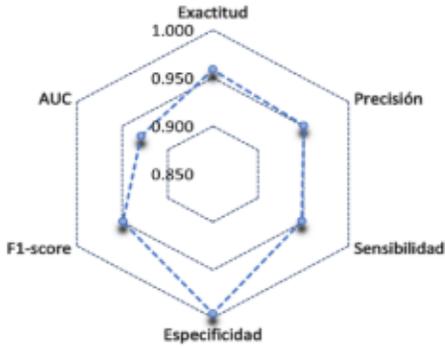
RESUMEN DEL RENDIMIENTO DE ALEX-NET



RESUMEN DEL RENDIMIENTO DE DENSENET-201



RESUMEN DEL RENDIMIENTO DE INCEPTION-V3



RESUMEN DEL RENDIMIENTO DE RESNET-101

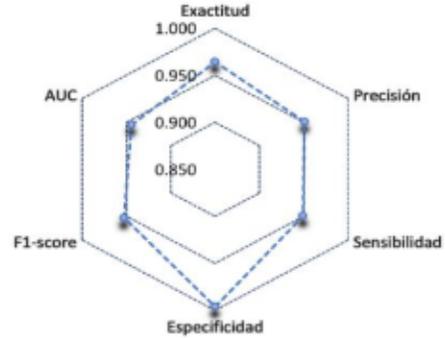


Figura 23. Resumen del rendimiento de las arquitecturas.

Para un análisis más detallado del comportamiento de las arquitecturas, se calcularon las matrices de confusión. La **Figura 24** muestra la matriz de confusión de AlexNet, está claro que el modelo tuvo dificultades para clasificar la clase 3, que corresponde a esofagitis grado-a, solo clasificando el 56.19% de las instancias correctamente.

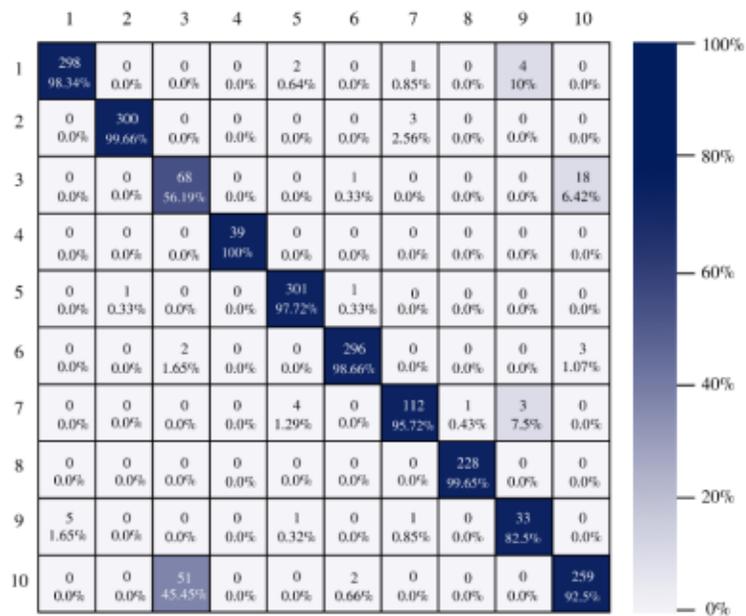


Figura 24. Matriz de confusión de AlexNet.

La matriz de confusión del modelo DenseNet-201 muestra su capacidad para clasificar correctamente las instancias; 7 de cada 10 clases fueron clasificadas en un 100%.

La diagonal principal muestra el número de muestras clasificadas correctamente. Se puede observar en la **Figura 25** que las clases con mayor conflicto fueron la clase 3 identificada como esofagitis grado-a y la clase 10 que corresponde a la línea z.

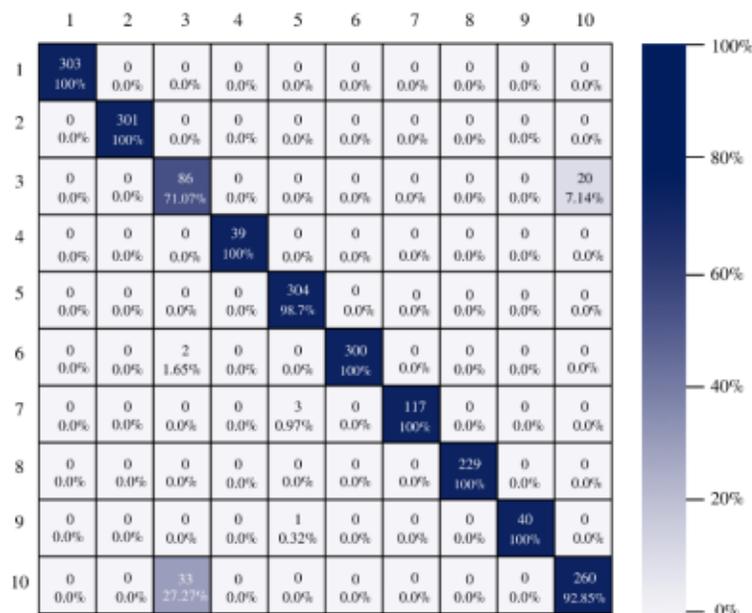


Figura 25. Matriz de confusión de DenseNet-201.

Específicamente para la clase 5, correspondiente a pólipos, DenseNet-201 logró clasificar el 98.7% correctamente. Lo cual indica un error del 1.3%, en comparación con el error del 25% que se obtiene habitualmente en un proceso de endoscopia, el modelo DenseNet-201 logra disminuir considerablemente el margen de error en la clasificación e identificación de pólipos.

Por otro lado, Inception-v3 clasificó correctamente todas las instancias de sólo 3 clases. Aun así, obtuvo buenos porcentajes en las clases restantes. Se puede observar en la **Figura 26** que la clase 3 sigue siendo la que genera más conflictos en las arquitecturas.

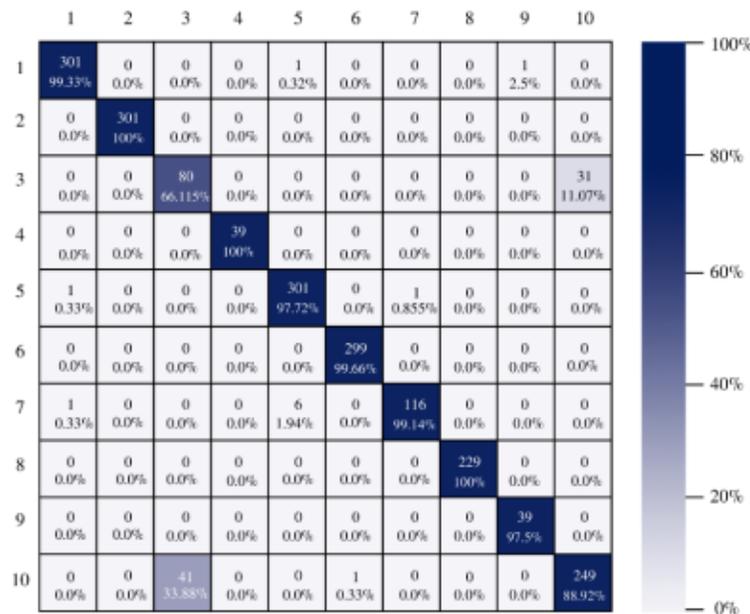


Figura 26. Matriz de confusión de Inception-v3.

La **Figura 27** muestra la matriz de confusión de ResNet-101. La arquitectura muestra buenos resultados en la clasificación de las instancias de cada clase, hay un buen equilibrio entre la referencia y la predicción. No obstante, sólo clasificó correctamente todas las instancias de dos clases, que son: heces impactadas y colitis ulcerosa-grado-3.

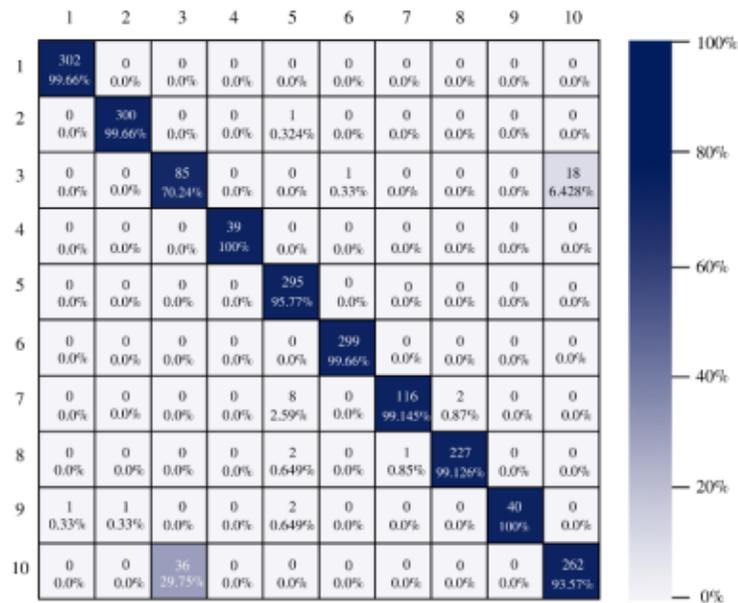


Figura 27. Matriz de confusión de ResNet-101.

En términos de AUC, la arquitectura DenseNet-201 alcanzó un 94.9%. La Figura 28 muestra la habilidad para clasificar los valores verdaderos, es decir, la relación entre especificidad y sensibilidad es estadísticamente confiable.

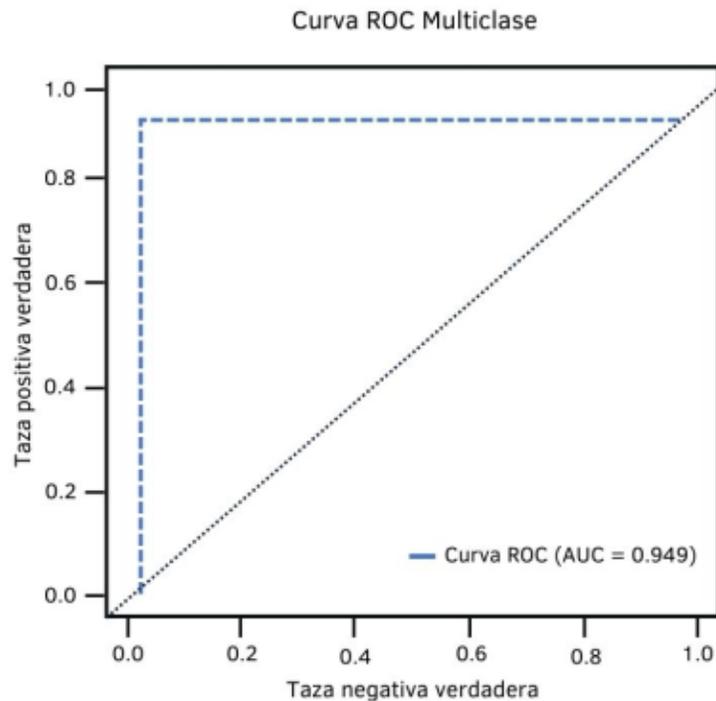


Figura 28. Curva ROC de la arquitectura DenseNet-201.

Capítulo 5. Conclusiones

En la presente investigación, se seleccionaron 4 arquitecturas para comparar su rendimiento en la clasificación de lesiones gastrointestinales. DenseNet-201 fue la arquitectura que mejor modeló el problema de clasificación de lesiones en el tracto GI, sobresalió en cada una de las métricas alcanzando una exactitud general del 97.11% y clasificando 7 de las 10 clases correctamente en su totalidad. Esto demuestra que utilizar un enfoque de aprendizaje por transferencia logra resultados estadísticamente significativos. Sin embargo, una de las limitaciones en la implementación de las redes neuronales convolucionales fue la falta de imágenes para ciertas clases, lo cual generó que el rendimiento general de las arquitecturas disminuyera. Se observó que el comportamiento de las CNN es mucho mejor cuando hay un gran número de imágenes para entrenar los modelos.

Como trabajo futuro se propone la implementación de la red neuronal DenseNet-201 como sistema de apoyo en el proceso de endoscopia para la identificación de lesiones del tracto gastrointestinal.

5.1. Objetivos alcanzados

La implementación de las redes neuronales convolucionales se llevó a cabo en MatLab, donde se inicializaron las arquitecturas con los hiperparámetros establecidos en la **Tabla 3**. Una vez que las arquitecturas fueron entrenadas utilizando el enfoque de aprendizaje por transferencia, se sometieron a pruebas donde se midió y comparó el rendimiento de cada una.

5.2. Hipótesis / Propositiones demostradas

Gracias a la comparación de las arquitecturas a través de diferentes métricas. Se logró realizar una selección de la arquitectura que mejor modela el problema de clasificación de lesiones del tracto GI. Por otro lado, con respecto a la tasa de reducción en la detección de pólipos, DenseNet-201 clasificó con un error de 1.3%.

5.3. Contribuciones de la investigación

La principal contribución de este trabajo es mostrar el potencial de cuatro arquitecturas de redes neuronales convolucionales, comparando su rendimiento mediante la implementación de diferentes métricas.

Referencias

1. Abraham, T., Todd, A., Orringer, D. A., & Levenson, R. (2021). Applications of artificial intelligence for image enhancement in pathology. *Artificial Intelligence and Deep Learning in Pathology*, 119–148. <https://doi.org/10.1016/B978-0-323-67538-3.00007-5>
2. Adam, K., Mohd, I. I., & Younis, Y. M. (2021). The impact of the soft errors in convolutional neural network on GPUs: Alexnet as case study. *Procedia Computer Science*, 182, 89–94. <https://doi.org/10.1016/J.PROCS.2021.02.012>
3. Agrawa, T., Gupta, R., Sahu, S., & Wilson, C. E. (2017). SCL-UMD at the medico task-mediaeval 2017: Transfer learning based classification of medical images. *CEUR Workshop Proceedings*, 1984, 3–5.
4. Alex Krizhevsky, Ilya Sutskever, G. E. H. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Handbook of Approximation Algorithms and Metaheuristics*, 1–1432. <https://doi.org/10.1201/9781420010749>
5. Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data 2021 8:1*, 8(1), 1–74. <https://doi.org/10.1186/S40537-021-00444-8>
6. Amin, J., Sharif, M., Anjum, M. A., Raza, M., & Bukhari, S. A. C. (2020). Convolutional neural network with batch normalization for glioma and stroke lesion detection using MRI. *Cognitive Systems Research*, 59, 304–311. <https://doi.org/10.1016/J.COGSYS.2019.10.002>
7. Amorós Becerra, P. (2021). *Desbalanceo de datos en redes de clasificación binaria*.
8. Anil, R., Gupta, V., Brain, G., Koren, T., & Singer, Y. (2019). *Memory-Efficient Adaptive Optimization*.
9. Assad, M. B., & Kiczales, R. (2020). Deep Biomedical Image Classification Using Diagonal Bilinear Interpolation and residual network. *International Journal of Intelligent Networks*, 1, 148–156. <https://doi.org/10.1016/J.IJIN.2020.11.001>
10. Avila-Tomás, J. F., Mayer-Pujadas, M. A., & Quesada-Varela, V. J. (2020). Artificial intelligence and its applications in medicine I: introductory background to AI and robotics. *Atención Primaria*, 52(10), 778–784. <https://doi.org/10.1016/j.aprim.2020.04.013>
11. Badr, E., Almotairi, S., Salam, M. A., & Ahmed, H. (2021). New Sequential and Parallel Support Vector Machine with Grey Wolf Optimizer for Breast Cancer Diagnosis. *Alexandria Engineering Journal*. <https://doi.org/10.1016/J.AEJ.2021.07.024>
12. Barcenilla Laguna, J., Goyeneche Gracia, M. L., Madrigal Domínguez, R. E., & Díez Redondo, P. (2000). Esofagitis. *Medicine - Programa de Formación Médica Continuada Acreditado*, 8(1), 14–22. [https://doi.org/10.1016/S0304-5412\(00\)70003-5](https://doi.org/10.1016/S0304-5412(00)70003-5)
13. Boden, M. A., & Pérez Parra, I. (2017). *Inteligencia artificial*.
14. Bohmrah, M. K., & Kaur, H. (2021). Classification of Covid-19 patients using efficient Fine-tuned Deep learning DenseNet Model. *Global Transitions Proceedings*. <https://doi.org/10.1016/J.GLTP.2021.08.003>
15. Borgli, H., Thambawita, V., Smedsrud, P. H., Hicks, S., Jha, D., Eskeland, S. L., Randel, K. R., Pogorelov, K., Lux, M., Nguyen, D. T. D., Johansen, D., Griwodz, C., Stensland, H. K., Garcia-Ceja, E., Schmidt, P. T., Hammer, H. L., Riegler, M. A., Halvorsen, P., & de Lange, T. (2020). HyperKvasir, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy. *Scientific Data*, 7(1), 1–14. <https://doi.org/10.1038/s41597-020-00622-y>
16. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality

- worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 68(6), 394–424. <https://doi.org/10.3322/CAAC.21492>
17. Brown, M. J., Hutchinson, L. A., Rainbow, M. J., Deluzio, K. J., & De Asha, A. R. (2017). A comparison of self-selected walking speeds and walking speed variability when data are collected during repeated discrete trials and during continuous walking. *Journal of Applied Biomechanics*, 33(5), 384–387. <https://doi.org/10.1123/jab.2016-0355>
 18. Byrne, M. F., Chapados, N., Soudan, F., Oertel, C., Pérez, M. L., Kelly, R., Iqbal, N., Chandelier, F., & Rex, D. K. (2019). Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model. *Gut*, 68(1), 94–100. <https://doi.org/10.1136/gutjnl-2017-314547>
 19. Cepeda-Vásquez, R. (2021). Inteligencia artificial en la detección de pólipos colónicos: qué dicen los estudios. *Revista Colombiana de Gastroenterología*, 36(1), 2–6. <https://doi.org/10.22516/25007440.726>
 20. Chang, Y., Chen, W., Huang, Z., & Shen, Q. (2019). Gastrointestinal tract diseases detection with deep attention neural network. *MM 2019 - Proceedings of the 27th ACM International Conference on Multimedia*, 2568–2572. <https://doi.org/10.1145/3343031.3356061>
 21. Chauhan, T., Palivela, H., & Tiwari, S. (2021). Optimization and fine-tuning of DenseNet model for classification of COVID-19 cases in medical imaging. *International Journal of Information Management Data Insights*, 1(2), 100020. <https://doi.org/10.1016/J.JJIMEI.2021.100020>
 22. Chen, B. L., Wan, J. J., Chen, T. Y., Yu, Y. T., & Ji, M. (2021). A self-attention based faster R-CNN for polyp detection from colonoscopy images. *Biomedical Signal Processing and Control*, 70, 103019. <https://doi.org/10.1016/J.BSPC.2021.103019>
 23. Chen, J., Wan, Z., Zhang, J., Li, W., Chen, Y., Li, Y., & Duan, Y. (2021). Medical image segmentation and reconstruction of prostate tumor based on 3D AlexNet. *Computer Methods and Programs in Biomedicine*, 200, 105878. <https://doi.org/10.1016/J.CMPB.2020.105878>
 24. Choi, H. N., Kim, H. H., Oh, J. S., Jang, H. S., Hwang, H. S., Kim, E. Y., Kwon, J. G., & Jung, J. T. (2014). Factors Influencing the Miss Rate of Polyps in a Tandem Colonoscopy Study. *The Korean Journal of Gastroenterology*, 64(1), 24–30. <https://doi.org/10.4166/KJG.2014.64.1.24>
 25. Cogan, T., Cogan, M., & Tamil, L. (2019). MAPGI: Accurate identification of anatomical landmarks and diseased tissue in gastrointestinal tract using deep learning. *Computers in Biology and Medicine*, 111, 103351. <https://doi.org/10.1016/J.COMPBIOMED.2019.103351>
 26. Da-Chuan Cheng, Wen-Chien Ting, Yung-Fu Chen, and X. J. (2008). Detection of Colorectal Polyps Using Texture Features and Support Vector Machine. *Advances in Mass Data Analysis of Images and Signals in Medicine, Biotechnology, Chemistry and Food Industry - Third International Conference, MDA 2008 Leipzig, Germany, July 14, 2008 Proceedings: Preface, 5108 LNAI(July)*. <https://doi.org/10.1007/978-3-540-70715-8>
 27. Damaševič, R. (2021). *Analysis of Features of Alzheimer ' s Disease : Detection of Early Stage from Functional Brain Changes in Magnetic Resonance Images Using a Finetuned ResNet18 Network*.
 28. Ding, Z., Shi, H., Zhang, H., Meng, L., Fan, M., Han, C., Zhang, K., Ming, F., Xie, X., Liu, H., Liu, J., Lin, R., & Hou, X. (2019). Gastroenterologist-Level Identification of Small-Bowel Diseases and Normal Variants by Capsule Endoscopy Using a Deep-Learning Model. *Gastroenterology*, 157(4), 1044-1054.e5. <https://doi.org/10.1053/J.GASTRO.2019.06.025>
 29. Fernández-Esparrach, G., & García-Rodríguez, A. (2020). Inteligencia artificial en

- endoscopia: una puesta al día. *Acta Gastroenterológica Latinoamericana*, 50(s4). <https://doi.org/10.52787/epak9041>
30. Glissen Brown, J. R., Mansour, N. M., Wang, P., Chuchuca, M. A., Minchenberg, S. B., Chandnani, M., Liu, L., Gross, S. A., Sengupta, N., & Berzin, T. M. (2021). DEEP LEARNING COMPUTER-AIDED POLYP DETECTION REDUCES ADENOMA MISS RATE: A U.S. MULTI-CENTER RANDOMIZED TANDEM COLONOSCOPY STUDY (CADeT-CS Trial). *Clinical Gastroenterology and Hepatology*. <https://doi.org/10.1016/j.cgh.2021.09.009>
 31. Gómez-Zuleta, M. A., Cano-Rosales, D. F., Bravo-Higuera, D. F., Ruano-Balseca, J. A., Romero-Castro, E., Gómez-Zuleta, M. A., Cano-Rosales, D. F., Bravo-Higuera, D. F., Ruano-Balseca, J. A., & Romero-Castro, E. (2021). Detección automática de pólipos colorrectales con técnicas de inteligencia artificial. *Revista Colombiana de Gastroenterología*, 36(1), 7–17. <https://doi.org/10.22516/25007440.471>
 32. González-Huix Lladó, F., Figa Francesch, M., & Huertas Nadal, C. (2010). Criterios de calidad que deben exigirse en la indicación y en la realización de la colonoscopia. *Gastroenterología y Hepatología*, 33(1), 33–42. <https://doi.org/10.1016/j.gastrohep.2009.02.014>
 33. Habib, F., & Jobe, B. A. (2019). Novel Diagnostic Technologies. *Shackelford's Surgery of the Alimentary Tract, 2 Volume Set*, 141–156. <https://doi.org/10.1016/B978-0-323-40232-3.00010-8>
 34. Habib, G., & Qureshi, S. (2020). Optimization and acceleration of convolutional neural networks: A survey. *Journal of King Saud University - Computer and Information Sciences*. <https://doi.org/10.1016/j.jksuci.2020.10.004>
 35. Hand, D. J., & Till, R. J. (2001). A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems. *Machine Learning 2001 45:2*, 45(2), 171–186. <https://doi.org/10.1023/A:1010920819831>
 36. Hernández Sampieri Roberto; Méndez Valencia Sergio; Mendoza Torres Christian Paulina & Cuevas Ana. (2017). Fundamentos de Investigación. In *Fundamentos Investigación*. <https://es.scribd.com/document/454886785/2017-Fundamentos-de-investigacion-Sampieri1-pdf>
 37. Hernández-Rocha, C., Ibáñez, P., elena Molina, M., Klaassen, J., Valenzuela, A., Candia, R., Bellolio, F., Zúñiga, Á., Migueles, R., FranCisCo Miquel, J., CHianale, J., & ÁLVarez-lobos, M. (2017). Diagnóstico y manejo de colitis ulcerosa grave. Una mirada actualizada. *Revista Medica de Chile*.
 38. Hiriyannaiah, S., Srinivas, A. M. D., Shetty, G. K., G.M., S., & Srinivasa, K. G. (2020). A computationally intelligent agent for detecting fake news using generative adversarial networks. *Hybrid Computational Intelligence*, 69–96. <https://doi.org/10.1016/B978-0-12-818699-2.00004-4>
 39. Hmoud Al-Adhaileh, M., Mohammed Senan, E., Alsaade, W., Aldhyani, T. H. H., Alsharif, N., Abdullah Alqarni, A., Uddin, M. I., Alzahrani, M. Y., Alzain, E. D., & Jadhav, M. E. (2021). Deep Learning Algorithms for Detection and Classification of Gastrointestinal Diseases. *Complexity*, 2021, 1–12. <https://doi.org/10.1155/2021/6170416>
 40. Hoang, T. H., Nguyen, H. D., Nguyen, T. A., Nguyen, V. T., & Tran, M. T. (2018). An application of residual network and faster - RCNN for medico: Multimedia task at MediaEval 2018. *CEUR Workshop Proceedings*, 2283, 2–4.
 41. Hoang, T., Nguyen, H., Nguyen, T., MediaEval, V. N., & 2018, U. (2018). An Application of Residual Network and Faster-RCNN for Medico: Multimedia Task at MediaEval 2018. *Eurecom.Fr*. <https://doi.org/10.1109/cvpr.2016.90>
 42. Hong, H., Tsangaratos, P., Ilija, I., Loupasakis, C., & Wang, Y. (2020). Introducing a novel multi-layer perceptron network based on stochastic gradient descent optimized by a meta-heuristic algorithm for landslide susceptibility mapping. *Science of The Total Environment*,

742, 140549. <https://doi.org/10.1016/J.SCITOTENV.2020.140549>

43. Igarashi, S., Sasaki, Y., Mikami, T., Sakuraba, H., & Fukuda, S. (2020). Anatomical classification of upper gastrointestinal organs under various image capture conditions using AlexNet. *Computers in Biology and Medicine*, 124. <https://doi.org/10.1016/J.COMPBIOMED.2020.103950>
44. Iliev, A., Kyurkchiev, N., & Markov, S. (2017). On the approximation of the step function by some sigmoid functions. *Mathematics and Computers in Simulation*, 133, 223–234. <https://doi.org/10.1016/J.MATCOM.2015.11.005>
45. Indolia, S., Goswami, A. K., Mishra, S. P., & Asopa, P. (2018). Conceptual Understanding of Convolutional Neural Network- A Deep Learning Approach. *Procedia Computer Science*, 132, 679–688. <https://doi.org/10.1016/J.PROCS.2018.05.069>
46. J, P., Y, H., JH, Y., MG, P., J, K., YJ, L., & HJ, C. (2019). Recent Development of Computer Vision Technology to Improve Capsule Endoscopy. *Clinical Endoscopy*, 52(4), 328–333. <https://doi.org/10.5946/CE.2018.172>
47. Jha, D., Ali, S., Hicks, S., Thambawita, V., Borgli, H., Smedsrud, P. H., de Lange, T., Pogorelov, K., Wang, X., Harzig, P., Tran, M. T., Meng, W., Hoang, T. H., Dias, D., Ko, T. H., Agrawal, T., Ostroukhova, O., Khan, Z., Atif Tahir, M., ... Halvorsen, P. (2021). A comprehensive analysis of classification methods in gastrointestinal endoscopy imaging. *Medical Image Analysis*, 70, 102007. <https://doi.org/10.1016/j.media.2021.102007>
48. Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. *Journal of Big Data*. <https://doi.org/10.1186/s40537-019-0192-5>
49. Kaiming He, X. Z. (2016). Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1–9.
50. Kaminski, M. F., Regula, J., Kraszewska, E., Polkowski, M., Wojciechowska, U., Didkowska, J., Zwierko, M., Rupinski, M., Nowacki, M. P., & Butruk, E. (2010). Quality Indicators for Colonoscopy and the Risk of Interval Cancer. *Http://Dx.Doi.Org/10.1056/NEJMoa0907667*, 362(19), 1795–1803. <https://doi.org/10.1056/NEJMoa0907667>
51. Kaminski, M. F., Thomas-Gibson, S., Bugajski, M., Bretthauer, M., Rees, C. J., Dekker, E., Hoff, G., Jover, R., Suchanek, S., Ferlitsch, M., Anderson, J., Roesch, T., Hultcranz, R., Racz, I., Kuipers, E. J., Garborg, K., East, J. E., Rupinski, M., Seip, B., ... Rutter, M. D. (2017). Performance measures for lower gastrointestinal endoscopy: a European Society of Gastrointestinal Endoscopy (ESGE) Quality Improvement Initiative. *Endoscopy*, 49(04), 378–397. <https://doi.org/10.1055/S-0043-103411>
52. Kazemi, Y. (2017). *A Deep Learning Pipeline for Classifying Different Stages of Alzheimer 's Disease from fMRI Data*.
53. Kochhar, G. S., Carleton, N. M., & Thakkar, S. (2021). Assessing perspectives on artificial intelligence applications to gastroenterology. *Gastrointestinal Endoscopy*, 93(4), 971–975.e2. <https://doi.org/10.1016/J.GIE.2020.10.029>
54. Komeda, Y., Suzuki, N., Sarah, M., Thomas-Gibson, S., Vance, M., Fraser, C., Patel, K., & Saunders, B. P. (2013). Factors associated with failed polyp retrieval at screening colonoscopy. *Gastrointestinal Endoscopy*, 77(3), 395–400. <https://doi.org/10.1016/J.GIE.2012.10.007>
55. Kwak, G. H., & Hui, P. (2019). *DeepHealth: Review and challenges of artificial intelligence in health informatics. ML*.
56. Landwehr, N., Hall, M., & Frank, E. (2005). Logistic model trees. *Machine Learning*, 59(1–2), 161–205. <https://doi.org/10.1007/s10994-005-0466-3>
57. Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
58. Lee, J., Jun, S., Cho, Y., Lee, H., Kim, G. B., Seo, J. B., & Kim, N. (2017). Deep Learning in Medical Imaging. *Korean Journal of Radiology*, 18(4), 570–584.

59. Levin, B., Lieberman, D. A., McFarland, B., Smith, R. A., Brooks, D., Andrews, K. S., Dash, C., Giardiello, F. M., Glick, S., Levin, T. R., Pickhardt, P., Rex, D. K., Thorson, A., & Winawer, S. J. (2008). Screening and Surveillance for the Early Detection of Colorectal Cancer and Adenomatous Polyps, 2008: A Joint Guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology*†. *CA: A Cancer Journal for Clinicians*, 58(3), 130–160. <https://doi.org/10.3322/CA.2007.0018>
60. Li, H. Y., Dai, J., Xue, H. B., Zhao, Y. J., Chen, X. Y., Gao, Y. J., Song, Y., Ge, Z. Z., & Li, X. B. (2012). Application of magnifying endoscopy with narrow-band imaging in diagnosing gastric lesions: a prospective study. *Gastrointestinal Endoscopy*, 76(6), 1124–1132. <https://doi.org/10.1016/J.GIE.2012.08.015>
61. Lin, G., & Shen, W. (2018). Research on convolutional neural network based on improved Relu piecewise activation function. *Procedia Computer Science*, 131, 977–984. <https://doi.org/10.1016/J.PROCS.2018.04.239>
62. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., & Berg, A. C. (2015). SSD: Single Shot MultiBox Detector. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9905 LNCS, 21–37. https://doi.org/10.1007/978-3-319-46448-0_2
63. Luo, Z., Chen, S., Qian, Y., & Hou, Y. (2021). Multi-stage stochastic gradient method with momentum acceleration. *Signal Processing*, 188, 108201. <https://doi.org/10.1016/J.SIGPRO.2021.108201>
64. Luo, Z., Wang, X., Xu, Z., Li, X., & Li, J. (2019). Adaptive ensemble: Solution to the Biomedica ACM MM grand challenge 2019. *MM 2019 - Proceedings of the 27th ACM International Conference on Multimedia*, 2583–2587. <https://doi.org/10.1145/3343031.3356078>
65. Macaulay, B. O., Aribisala, B. S., Akande, S. A., Akinnuwesi, B. A., & Olabanjo, O. A. (2021). Breast cancer risk prediction in African women using Random Forest Classifier. *Cancer Treatment and Research Communications*, 28, 100396. <https://doi.org/10.1016/J.CTARC.2021.100396>
66. Maeda Gutiérrez, V., & Galván Tejada, C. E. (2020). Comparison of Convolutional Neural Network Architectures for Classification of Tomato Plant Diseases. *Applied Sciences 2020, Vol. 10, Page 1245, 10(4)*, 1245. <https://doi.org/10.3390/APP10041245>
67. Maguolo, G., Nanni, L., & Ghidoni, S. (2021). Ensemble of convolutional neural networks trained with different activation functions. *Expert Systems with Applications*, 166. <https://doi.org/10.1016/j.eswa.2020.114048>
68. McCreynolds, T., & Blythe, D. (2005). Color, Shading, and Lighting. *Advanced Graphics Programming Using OpenGL*, 35–56. <https://doi.org/10.1016/B978-155860659-3.50005-6>
69. Mocsari, E., & Stone, S. S. (2017). Densely Connected Convolutional Networks. *American Journal of Veterinary Research*, 39(9), 1442–1446.
70. Mohanty, H., Bhuyan, P., & Chentai, D. (2015). Deep Learning Techniques for Biomedical and Health Informatics. In *Springer* (Vol. 23, p. 195).
71. Monshi, M. M. A., Poon, J., Chung, V., & Monshi, F. M. (2021). CovidXrayNet: Optimizing data augmentation and CNN hyperparameters for improved COVID-19 detection from CXR. *Computers in Biology and Medicine*, 133, 104375. <https://doi.org/10.1016/J.COMPBIOMED.2021.104375>
72. Moreira, L., Castells, A., & Castelví, S. (2021). *Pólipos y poliposis colorrectales*.
73. Nandini, G. S., Kumar, A. P. S., & K, C. (2021). Dropout technique for image classification based on extreme learning machine. *Global Transitions Proceedings*, 2(1), 111–116. <https://doi.org/10.1016/J.GLTP.2021.01.015>
74. Niyas, S., Pawan, S. J., Kumar, M. A., & Rajan, J. (2021). *Medical Image Segmentation*

using 3D Convolutional Neural Networks: A Review.

75. Olga de Conte Vila, C. B. E. (2003). Clínica y tratamiento. *Farmacia Preventiva*, 17.
76. Pannala, R., Krishnan, K., Melson, J., Parsi, M. A., Schulman, A. R., Sullivan, S., Trikudanathan, G., Trindade, A. J., Watson, R. R., Maple, J. T., & Lichtenstein, D. R. (2020). Artificial intelligence in gastrointestinal endoscopy. *VideoGIE*, 5(12), 598–613. <https://doi.org/10.1016/J.VGIE.2020.08.013>
77. Petschornig, S., Schoffmann, K., & Lux, M. (2017). An inception-like CNN architecture for GI disease and anatomical landmark classification. *CEUR Workshop Proceedings*, 1984, 0–2.
78. Pogorelov, K., Randel, K. R., de Lange, T., Eskeland, S. L., Griwodz, C., Johansen, D., Spampinato, C., Taschwer, M., Lux, M., Schmidt, P. T., Riegler, M., & Halvorsen, P. (2017). Nerthus: A bowel preparation quality video dataset. *Proceedings of the 8th ACM Multimedia Systems Conference, MMSys 2017*, 170–174. <https://doi.org/10.1145/3083187.3083216>
79. Pogorelov, K., Ranheim Randel, K., Griwodz, C., Hospital, B., Thomas de Lange, N., Johansen, D., Spampinato, C., Dang-Nguyen, D.-T., Lux, M., Thelin Schmidt Karolinska Institutet, P., Karolinska Hospital, S., Michael Riegler, S., Halvorsen, P., Losada Eskeland, S., de Lange, T., Thelin Schmidt, P., & Riegler, M. (2017). Kvasir: A Multi-Class Image Dataset for Com-puter Aided Gastrointestinal Disease Detection. *Proceedings of the 8th ACM on Multimedia Systems Conference*, 6, 6. <https://doi.org/10.1145/3083187>
80. Pohl, H., & Robertson, D. J. (2010). Colorectal Cancers Detected After Colonoscopy Frequently Result From Missed Lesions. *Clinical Gastroenterology and Hepatology*, 8(10), 858–864. <https://doi.org/10.1016/J.CGH.2010.06.028>
81. R., M. A. R., & G., F. F. (2015). ESÓFAGO DE BARRETT: REVISIÓN DE LA LITERATURA. *Revista Médica Clínica Las Condes*, 26(5), 557–564. <https://doi.org/10.1016/J.RMCLC.2015.09.002>
82. Reimers, C., & Requena-Mesa, C. (2020). Deep Learning – an Opportunity and a Challenge for Geo- and Astrophysics. *Knowledge Discovery in Big Data from Astronomy and Earth Observation*, 251–265. <https://doi.org/10.1016/B978-0-12-819154-5.00024-2>
83. Rodríguez de Santiago, E., Hernanz, N., Marcos-Prieto, H. M., De-Jorge-Turrión, M. Á., Barreiro-Alonso, E., Rodríguez-Escaja, C., Jiménez-Jurado, A., Sierra-Morales, M., Pérez-Valle, I., Machado-Volpato, N., García-Prada, M., Núñez-Gómez, L., Castaño-García, A., García García de Paredes, A., Peñas, B., Vázquez-Sequeiros, E., & Albillos, A. (2019). Rate of missed oesophageal cancer at routine endoscopy and survival outcomes: A multicentric cohort study. *United European Gastroenterology Journal*, 7(2), 189–198. <https://doi.org/10.1177/2050640618811477>
84. Rusk, N. (2015). Deep learning. *Nature Methods 2016 13:1*, 13(1), 35–35. <https://doi.org/10.1038/nmeth.3707>
85. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., & Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, 115(3), 211–252. <https://doi.org/10.1007/s11263-015-0816-y>
86. Russell, S. (2004). *¿Qué es la Inteligencia Artificial? - Iberdrola*.
87. Salama, W. M., & Aly, M. H. (2021). Deep learning in mammography images segmentation and classification: Automated CNN approach. *Alexandria Engineering Journal*, 60(5), 4701–4709. <https://doi.org/10.1016/J.AEJ.2021.03.048>
88. Sánchez-Montes, C., Bernal, J., García-Rodríguez, A., Córdova, H., & Fernández-Esparrach, G. (2020). Revisión de métodos computacionales de detección y clasificación de pólipos en imagen de colonoscopia. *Gastroenterología y Hepatología*, 43(4), 222–232. <https://doi.org/10.1016/J.GASTROHEP.2019.11.004>
89. Sastre Lozano, V. M. (2018). Aproximación a las causas del cáncer colorrectal de intervalo

- en colonoscopia en el área II del Servicio Murciano de Salud. *Proyecto de Investigación*: <https://digitum.um.es/digitum/handle/10201/59559>
90. Senior, A., & Lei, X. (2014). Fine context, low-rank, softplus deep neural networks for mobile speech recognition. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, Cd*, 7644–7648. <https://doi.org/10.1109/ICASSP.2014.6855087>
 91. Shajun Nisha, S., & Nagoor Meeral, M. (2021). Applications of deep learning in biomedical engineering. *Handbook of Deep Learning in Biomedical Engineering*, 245–270. <https://doi.org/10.1016/B978-0-12-823014-5.00008-9>
 92. Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 1–14.
 93. Smedsrud, P. H., Thambawita, V., Hicks, S. A., Gjestang, H., Nedrejord, O. O., Næss, E., Borgli, H., Jha, D., Berstad, T. J. D., Eskeland, S. L., Lux, M., Espeland, H., Petlund, A., Nguyen, D. T. D., Garcia-Ceja, E., Johansen, D., Schmidt, P. T., Toth, E., Hammer, H. L., ... Halvorsen, P. (2021). Kvasir-Capsule, a video capsule endoscopy dataset. *Scientific Data*, 8(1), 1–10. <https://doi.org/10.1038/s41597-021-00920-z>
 94. Song, Y., & Cai, W. (2021). Visual feature representation in microscopy image classification. *Computer Vision for Microscopy Image Analysis*, 73–100. <https://doi.org/10.1016/B978-0-12-814972-0.00004-7>
 95. Subasi, A. (2020). Other classification examples. *Practical Machine Learning for Data Analysis Using Python*, 323–390. <https://doi.org/10.1016/B978-0-12-821379-7.00005-9>
 96. Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, 71(3), 209–249. <https://doi.org/10.3322/CAAC.21660>
 97. Szegedy, C., Reed, S., Sermanet, P., Vanhoucke, V., & Rabinovich, A. (2014). *Going deeper with convolutions*. 1–12.
 98. Szegedy, C., Vanhoucke, V., & Shlens, J. (2014). *Rethinking the Inception Architecture for Computer Vision*.
 99. Takiyama, H., Ozawa, T., Ishihara, S., Fujishiro, M., Shichijo, S., Nomura, S., Miura, M., & Tada, T. (2018). Automatic anatomical classification of esophagogastroduodenoscopy images using deep convolutional neural networks. *Scientific Reports*, 8(1), 1–8. <https://doi.org/10.1038/s41598-018-25842-6>
 100. Terese Wislow, I. N. del C. (2005). *Definición de gastroscopia - Diccionario de cáncer del NCI*. <https://www.cancer.gov/espanol/publicaciones/diccionarios/diccionario-cancer/def/gastroscopia>
 101. Thambawita, V., Jha, D., Hammer, H. L., Johansen, H. D., Johansen, D., Halvorsen, P., & Riegler, M. A. (2020). An Extensive Study on Cross-Dataset Bias and Evaluation Metrics Interpretation for Machine Learning applied to Gastrointestinal Tract Abnormality Classification. *ACM Transactions on Computing for Healthcare*, 1(3), 1–29. <https://doi.org/10.1145/3386295>
 102. Wang, C., Chen, D., Hao, L., Liu, X., Zeng, Y., Chen, J., & Zhang, G. (2019). Pulmonary image classification based on inception-v3 transfer learning model. *IEEE Access*, 7, 146533–146541. <https://doi.org/10.1109/ACCESS.2019.2946000>
 103. Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2017). Deep learning. *Data Mining*, 417–466. <https://doi.org/10.1016/B978-0-12-804291-5.00010-6>
 104. Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. (2017). Aggregated residual transformations for deep neural networks. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017-Janua*, 5987–5995. <https://doi.org/10.1109/CVPR.2017.634>

105. Xu, Z., Tao, Y., Wenfang, Z., Ne, L., Zhengxing, H., Jiquan, L., Weiling, H., Huilong, D., & Jianmin, S. (2019). Upper gastrointestinal anatomy detection with multi-task convolutional neural networks. *Healthcare Technology Letters*, 6(6), 176–180. <https://doi.org/10.1049/htl.2019.0066>
106. Yadav, R. K., & Anubhav. (2020). PSO-GA based hybrid with Adam Optimization for ANN training with application in Medical Diagnosis. *Cognitive Systems Research*, 64, 191–199. <https://doi.org/10.1016/J.COGSYS.2020.08.011>
107. Yang, L., Li, Z., Ma, S., & Yang, X. (2021). Artificial intelligence image recognition based on 5G deep learning edge algorithm of Digestive endoscopy on medical construction. *Alexandria Engineering Journal*. <https://doi.org/10.1016/J.AEJ.2021.07.007>
108. Yogapriya, J., Chandran, V., Sumithra, M. G., Anitha, P., Jenopaul, P., & Suresh Gnana Dhas, C. (2021). Gastrointestinal Tract Disease Classification from Wireless Endoscopy Images Using Pretrained Deep Learning Model. *Computational and Mathematical Methods in Medicine*, 2021. <https://doi.org/10.1155/2021/5940433>
109. Yu, H., Singh, R., Shin, S. H., & Ho, K. Y. (2021). Artificial intelligence in upper GI endoscopy - current status, challenges and future promise. In *Journal of Gastroenterology and Hepatology (Australia)* (Vol. 36, Issue 1, pp. 20–24). <https://doi.org/10.1111/jgh.15354>
110. Zemouri, R., Zerhouni, N., & Racoceanu, D. (2019). Deep learning in the biomedical applications: Recent and future status. *Applied Sciences (Switzerland)*, 9(8). <https://doi.org/10.3390/app9081526>
111. Zhao, S., Wang, S., Pan, P., Xia, T., Chang, X., Yang, X., Guo, L., Meng, Q., Yang, F., Qian, W., Xu, Z., Wang, Y., Wang, Z., Gu, L., Wang, R., Jia, F., Yao, J., Li, Z., & Bai, Y. (2019). Magnitude, Risk Factors, and Factors Associated With Adenoma Miss Rate of Tandem Colonoscopy: A Systematic Review and Meta-analysis. *Gastroenterology*, 156(6), 1661-1674.e11. <https://doi.org/10.1053/J.GASTRO.2019.01.260>
112. Zoph, B., Vasudevan, V., Shlens, J., & Le, Q. V. (2018). Learning Transferable Architectures for Scalable Image Recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 8697–8710. <https://doi.org/10.1109/CVPR.2018.00907>