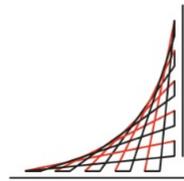




Universidad del  
**Rosario**



ESCUELA  
COLOMBIANA  
DE INGENIERÍA  
JULIO GARAVITO

# Identificación de neoplasia cervical intraepitelial para la predicción de cáncer cervical mediante el uso de aprendizaje profundo

Camilo Antonio Tenjo Castaño

Trabajo Dirigido presentado como requisito para aplicar al título de:  
**Ingeniero Biomédico**

Tutor:

Ph.D. Oscar Julián Perdomo Charry

UNIVERSIDAD DEL ROSARIO  
ESCUELA COLOMBIANA DE INGENIERÍA JULIO GARAVITO  
PROGRAMA DE INGENIERÍA BIOMÉDICA  
BOGOTÁ, COLOMBIA

2021

*That's the life, really, isn't it? You write. You  
record. You play. And it never grows old.*  
*- Rob Halford*

# Agradecimientos

Infinitas gracias a mi tutor: Dr. Oscar Julián Perdomo Charry por su paciencia, apoyo y la oportunidad de adentrarme mucho más el campo de la inteligencia artificial mediante este trabajo.

Agradezco a mis padres Javier Tenjo y Ana Castaño, mi abuela Nelly Forero y mis tías Cristina y Nelly Tenjo, por todo su apoyo, además de demostrarme, que ante cualquier adversidad solo importa el valor y la voluntad de nunca rendirse. Finalmente, a mi hermano Francisco, por ser esa inspiración y mostrarme que el conocimiento existe para cultivarlo y usarlo en pro de la humanidad.

Gracias a la Escuela Colombiana de Ingeniería: Julio Garavito y a la Universidad del Rosario finalmente por haber sido el lugar de aprendizaje y esfuerzo que hoy culmina en este proyecto de grado; Además de ser los lugares que me permitieron conocer tan increíbles colegas.

## Resumen

Hoy en día son de suma importancia los procedimientos de diagnóstico basados en imágenes gracias a su alto nivel de eficiencia y practicidad, lamentablemente existen limitantes en cuanto a su interpretación se trata; Este es el caso de la Neoplasia Intraepitelial Cervical o displasia cervical. Muchas mujeres alrededor del mundo padecen esta lesión y aunque en varias ocasiones no compromete su vida, aún existen casos en los que las pacientes desarrollan uno de los 5 tipos de cáncer con mayor tasa de mortalidad del mundo, el cáncer cervical. Para la displasia cervical es usual la implementación de colposcopia como método de diagnóstico inicial, dado que permite identificar la lesión y su grado de severidad; Pero se limita en la complejidad de los tejidos y en ocasiones, la falta de experiencia del especialista que lleva a cabo el diagnóstico. Pese a que en países desarrollados estas problemáticas de diagnóstico no son muy comunes, es importante el desarrollo de tecnologías de asistencia diagnóstica para regiones en las que se necesitan. Con este trabajo se propone el uso de modelos de clasificación basados en aprendizaje profundo, para comprobar la presencia de la lesión y el grado de severidad en imágenes de colposcopia.

**Palabras clave:** Aprendizaje profundo, Colposcopia, Modelos de Clasificación, Neoplasia Cervical Intraepitelial.

# Tabla de Contenidos

<b>Agradecimientos</b>	<b>iii</b>
<b>Resumen</b>	<b>iv</b>
<b>Índice de Figuras</b>	<b>vii</b>
<b>Índice de Tablas</b>	<b>viii</b>
<b>1 Introducción</b>	<b>1</b>
1.1 Motivación . . . . .	1
1.2 Estado del arte . . . . .	1
1.2.1 Neoplasia intraepitelial cervical (CIN) . . . . .	1
1.2.2 Colposcopia . . . . .	4
1.3 Inteligencia artificial . . . . .	5
1.3.1 Aprendizaje profundo . . . . .	7
1.3.2 Aplicaciones en Neoplasia Intraepitelial Cervical . . . . .	8
<b>2 Objetivos</b>	<b>10</b>
2.1 General . . . . .	10
2.2 Específicos . . . . .	10
<b>3 Metodología</b>	<b>11</b>
3.1 Fase 1: Preprocesamiento . . . . .	11
3.2 Fase 2: Entrenamiento . . . . .	12
3.2.1 ResNet-50 y ResNet-152 . . . . .	12
3.2.2 InceptionV3 . . . . .	13
3.3 Fase 3: Validación e implementación con mejores modelos . . . . .	16
3.3.1 Validación . . . . .	16
3.3.2 Implementación . . . . .	17
3.4 Actividades . . . . .	17
<b>4 Resultados</b>	<b>18</b>
4.1 ResNet-50 . . . . .	18
4.1.1 Exploración con 10 épocas . . . . .	18
4.1.2 Exploración de ResNet-50 con 25 épocas . . . . .	19

---

4.1.3	Exploración de ResNet-50 con 50 épocas . . . . .	20
4.1.4	Exploración de ResNet-50 con 100 épocas . . . . .	21
4.2	InceptionV3 . . . . .	22
4.2.1	Exploración de InceptionV3 con 10 épocas . . . . .	22
4.2.2	Exploración de InceptionV3 con 25 épocas . . . . .	23
4.2.3	Exploración de InceptionV3 con 50 épocas . . . . .	24
4.2.4	Exploración de InceptionV3 con 100 épocas . . . . .	25
4.3	ResNet-152 . . . . .	26
4.3.1	Exploración de ResNet-152 con 10 épocas . . . . .	26
4.3.2	Exploración de ResNet-152 con 25 épocas . . . . .	27
4.3.3	Exploración de ResNet-152 con 50 épocas . . . . .	28
4.3.4	Exploración de ResNet-152 con 100 épocas . . . . .	29
4.4	Mejores modelos con pesos ajustados . . . . .	30
4.4.1	ResNet-50 con 50 épocas . . . . .	30
4.4.2	InceptionV3 con 100 épocas . . . . .	31
4.4.3	ResNet-152 con 25 épocas . . . . .	32
<b>5</b>	<b>Discusión</b>	<b>33</b>
<b>6</b>	<b>Conclusiones</b>	<b>35</b>
<b>7</b>	<b>Trabajos futuros</b>	<b>36</b>
	<b>Bibliografía</b>	<b>37</b>

# Índice de Figuras

<b>1-1</b>	Anatomía del cervix, disponible en [1] . . . . .	2
<b>1-2</b>	Endocérvix y Exocérvix, disponible en [1] . . . . .	3
<b>1-3</b>	Cambios en el epitelio escamoso cervical por HPV, disponible en [2] . . . . .	3
<b>1-4</b>	Tipos de CIN, disponible en [1] . . . . .	4
<b>1-5</b>	Procedimiento de colposcopia, disponible en [3] . . . . .	5
<b>1-6</b>	Arquitectura de Redes neuronales artificiales y aprendizaje profundo, disponible en [4] . . . . .	8
<b>3-1</b>	Red residual profunda, disponible en [5] . . . . .	13
<b>3-2</b>	Arquitectura ResNet, disponible en [6] . . . . .	14
<b>3-3</b>	Arquitectura Inception V3, disponible en [7] . . . . .	14
<b>3-4</b>	Arquitectura general modelos . . . . .	15

# Índice de Tablas

3-1	Modelos seleccionados . . . . .	15
3-2	Interpretación del índice kappa, disponible en [8] . . . . .	16
4-1	Métricas entrenamiento ResNet-50 entrenada con 10 épocas . . . . .	18
4-2	Métricas prueba ResNet-50 entrenada con 10 épocas . . . . .	18
4-3	Matriz de confusión entrenamiento con 10 épocas ResNet-50 . . . . .	18
4-4	Matriz de confusión prueba con 10 épocas ResNet-50 . . . . .	19
4-5	Métricas entrenamiento ResNet-50 entrenada con 25 épocas . . . . .	19
4-6	Métricas prueba ResNet-50 entrenada con 25 épocas . . . . .	19
4-7	Matriz de confusión entrenamiento con 25 épocas ResNet-50 . . . . .	19
4-8	Matriz de confusión prueba con 25 ResNet-50 . . . . .	20
4-9	Métricas entrenamiento ResNet-50 entrenada con 50 épocas . . . . .	20
4-10	Métricas prueba ResNet-50 entrenada con 50 épocas . . . . .	20
4-11	Matriz de confusión entrenamiento con 50 épocas ResNet-50 . . . . .	20
4-12	Matriz de confusión prueba con 50 ResNet-50 . . . . .	21
4-13	Métricas entrenamiento ResNet-50 entrenada con 100 épocas . . . . .	21
4-14	Métricas prueba ResNet-50 entrenada con 100 épocas . . . . .	21
4-15	Matriz de confusión entrenamiento con 100 épocas ResNet-50 . . . . .	21
4-16	Matriz de confusión prueba con 100 ResNet-50 . . . . .	21
4-17	Métricas entrenamiento InceptionV3 entrenada con 10 épocas . . . . .	22
4-18	Métricas prueba InceptionV3 entrenada con 10 épocas . . . . .	22
4-19	Matriz de confusión entrenamiento con 10 épocas InceptionV3 . . . . .	22
4-20	Matriz de confusión prueba con 10 épocas InceptionV3 . . . . .	22
4-21	Métricas entrenamiento InceptionV3 entrenada con 25 épocas . . . . .	23
4-22	Métricas prueba InceptionV3 entrenada con 25 épocas . . . . .	23
4-23	Matriz de confusión entrenamiento con 25 épocas InceptionV3 . . . . .	23
4-24	Matriz de confusión prueba con 25 épocas InceptionV3 . . . . .	23
4-25	Métricas entrenamiento InceptionV3 entrenada con 50 épocas . . . . .	24
4-26	Métricas prueba InceptionV3 entrenada con 50 épocas . . . . .	24
4-27	Matriz de confusión entrenamiento con 50 épocas InceptionV3 . . . . .	24
4-28	Matriz de confusión prueba con 50 épocas InceptionV3 . . . . .	24
4-29	Métricas entrenamiento InceptionV3 entrenada con 100 épocas . . . . .	25
4-30	Métricas prueba InceptionV3 entrenada con 100 épocas . . . . .	25

<b>4-31</b> Matriz de confusión entrenamiento con 100 épocas InceptionV3 . . . . .	25
<b>4-32</b> Matriz de confusión prueba con 100 épocas InceptionV3 . . . . .	25
<b>4-33</b> Métricas entrenamiento ResNet-152 entrenada con 10 épocas . . . . .	26
<b>4-34</b> Métricas prueba ResNet-152 entrenada con 10 épocas . . . . .	26
<b>4-35</b> Matriz de confusión entrenamiento con 10 épocas ResNet-152 . . . . .	26
<b>4-36</b> Matriz de confusión prueba con 10 épocas ResNet-152 . . . . .	26
<b>4-37</b> Métricas entrenamiento ResNet-152 entrenada con 25 épocas . . . . .	27
<b>4-38</b> Métricas prueba ResNet-152 entrenada con 25 épocas . . . . .	27
<b>4-39</b> Matriz de confusión entrenamiento con 25 épocas ResNet-152 . . . . .	27
<b>4-40</b> Matriz de confusión prueba con 25 épocas ResNet-152 . . . . .	27
<b>4-41</b> Métricas entrenamiento ResNet-152 entrenada con 50 épocas . . . . .	28
<b>4-42</b> Métricas entrenamiento ResNet-152 entrenada con 50 épocas . . . . .	28
<b>4-43</b> Matriz de confusión entrenamiento con 50 épocas ResNet-152 . . . . .	28
<b>4-44</b> Matriz de confusión prueba con 50 épocas ResNet-152 . . . . .	28
<b>4-45</b> Métricas entrenamiento ResNet-152 entrenada con 100 épocas . . . . .	29
<b>4-46</b> Métricas prueba ResNet-152 entrenada con 100 épocas . . . . .	29
<b>4-47</b> Matriz de confusión entrenamiento con 100 épocas ResNet-152 . . . . .	29
<b>4-48</b> Matriz de confusión prueba con 100 épocas ResNet-152 . . . . .	29
<b>4-49</b> Métricas entrenamiento ResNet-50 con pesos ajustados entrenada con 50 épocas	30
<b>4-50</b> Métricas prueba ResNet-50 con pesos ajustados entrenada con 50 épocas . .	30
<b>4-51</b> Matriz de confusión entrenamiento con pesos ajustados y 50 épocas ResNet-50	30
<b>4-52</b> Matriz de confusión prueba con pesos ajustados y 50 épocas ResNet-50 . . .	30
<b>4-53</b> Métricas entrenamiento InceptionV3 con pesos ajustados entrenada con 100 épocas . . . . .	31
<b>4-54</b> Métricas prueba InceptionV3 con pesos ajustados entrenada con 100 épocas .	31
<b>4-55</b> Matriz de confusión entrenamiento con pesos ajustados y 100 épocas Incep- tionV3 . . . . .	31
<b>4-56</b> Matriz de confusión prueba con pesos ajustados y 100 épocas InceptionV3 .	31
<b>4-57</b> Métricas entrenamiento ResNet-152 con pesos ajustados entrenada con 25 épocas . . . . .	32
<b>4-58</b> Métricas prueba ResNet-152 con pesos ajustados entrenada con 25 épocas . .	32
<b>4-59</b> Matriz de confusión entrenamiento con pesos ajustados y 25 épocas ResNet-152	32
<b>4-60</b> Matriz de confusión prueba con pesos ajustados y 25 épocas ResNet-152 . .	32



# 1 Introducción

## 1.1 Motivación

La enfermedad precursora del cáncer cervical, también llamado cáncer de cérvix es la Neoplasia intraepitelial cervical (CIN por sus siglas en inglés, también conocida como displasia cervical) [9]. En 2018 Arbyn M. et al. llevaron a cabo un estudio en el que se tomaron datos de 185 países del Observatorio global de cáncer, encontrando aproximadamente 570000 casos y 311000 muertes. Siendo las mujeres de mediana edad las más afectadas [10]. En 2020 estas cifras aumentaron a 604000 nuevos casos y 320000 muertes, teniendo una mayor tasa de mortalidad en mujeres residentes en países en vías de desarrollo [11]. Se ha demostrado que el desarrollo de tecnologías de procesamiento de imágenes y tratamientos tempranos son métodos con alta eficacia para prevenir el desarrollo de la CIN hasta convertirse en cáncer cervical, dado que este tipo de cáncer en particular se encuentra en el cuarto puesto de mayor incidencia en mujeres a nivel mundial [12].

El segundo paso en la identificación de un paciente con neoplasia intraepitelial cervical (CIN), es un examen de colposcopia [13]. Usualmente las pacientes son sometidas a este procedimiento en caso de que la paciente en cuestión presente resultados anormales para ADN de virus del papiloma humano o precáncer; Estos procedimientos han sido de gran impacto en Europa [13], pero la aplicación de programas similares en Latino América y La Región Caribe, no han sido tan efectivos. En su mayoría por barreras sociales, económicas, educativas, entre otras; Así mismo como la falta de centros especializados en la aplicación de pruebas diagnósticas basadas en imágenes [14]. Por esta razón, es de vital importancia el desarrollo de nuevas tecnologías que funcionen como ayudas diagnósticas a los especialistas, con esto se podrían asegurar diagnósticos tempranos de CIN y así reducir el desarrollo de cáncer cervical.

## 1.2 Estado del arte

### 1.2.1 Neoplasia intraepitelial cervical (CIN)

Esta enfermedad afecta el sistema reproductor femenino desde el cérvix del útero; En 2014 la organización mundial de la salud generó un manual de imagenología en el cual expresaban que este tipo de lesión era generalmente causada por infecciones como la del virus del papiloma

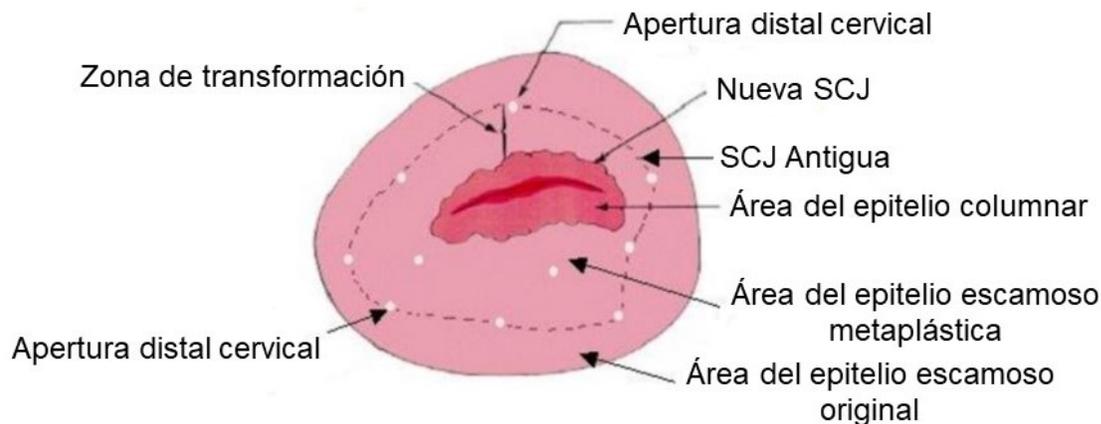
humano (HPV por sus siglas en inglés Human Papilloma Virus), así mismo como los síntomas que puede presentar en etapas tempranas y avanzadas de la lesión [15].

### Tempranas

- Sangrado anormal en periodos inter-menstruales o aumentos en el fluido usual.
- Aumento de la frecuencia de la necesidad urinaria y dolor de espalda.
- Presencia de olores extraños provenientes de los tejidos afectados.

### Avanzadas

- Dolor en zona pélvica y extremidades inferiores.
- Hinchazón de extremidades inferiores.
- Fallas renales.
- Formación de fistulas que generan el paso de orina y heces por el canal vaginal.



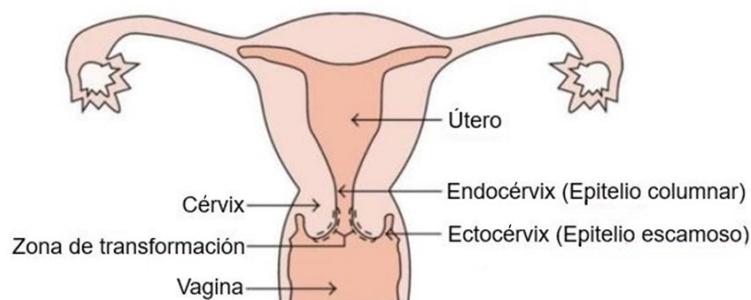
**Figura 1-1:** Anatomía del cervix, disponible en [1]

En la figura 1-1 se muestra la anatomía del cérvix visto desde el canal vaginal. Es importante resaltar 2 áreas específicas en las cuales sucede esta lesión [1]:

- La unión de escamosa-columnar (SCJ por sus siglas en inglés Squamo-Columnar Junction). Dado que en esta unión se encuentran las células del epitelio columnar (También conocido como endocérvix, que es el tejido más cercano al útero) y el epitelio escamoso (También conocido como ectocérvix o exocérvix).

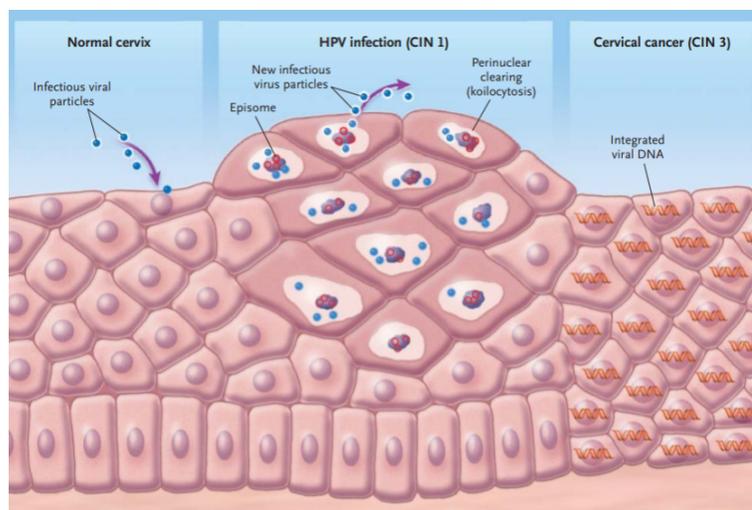
- La zona de transformación, dado que en esta se encuentra la nueva y antigua unión escamosa-columnar.

Esta lesión inicia en el exocérvix y a medida que avanza, se desplaza hacia el endocérvix, lo que dificulta su diagnóstico en etapas avanzadas (Ver figura 1-2). Sezgin I., et al. en 1989 definieron esta lesión como *Una serie de anomalías en el tejido escamoso epitelial del cérvix uterino que genera un riesgo de carcinoma en el cérvix*, de acuerdo del avance de la lesión esta se clasifica como grado I, II o III [16].



**Figura 1-2:** Endocérvix y Exocérvix, disponible en [1]

En 2003 Goodman A. et al. analizaron un caso clínico de una paciente de 37 años de edad que desarrolló cáncer cervical a partir de una infección por HPV, mostrando el proceso desde la inflexión en un cérvix sano (Ver figura 1-3), definiendo que la presencia de una displasia debía ser remitida a una biopsia para descartar la presencia de carcinomas [2].



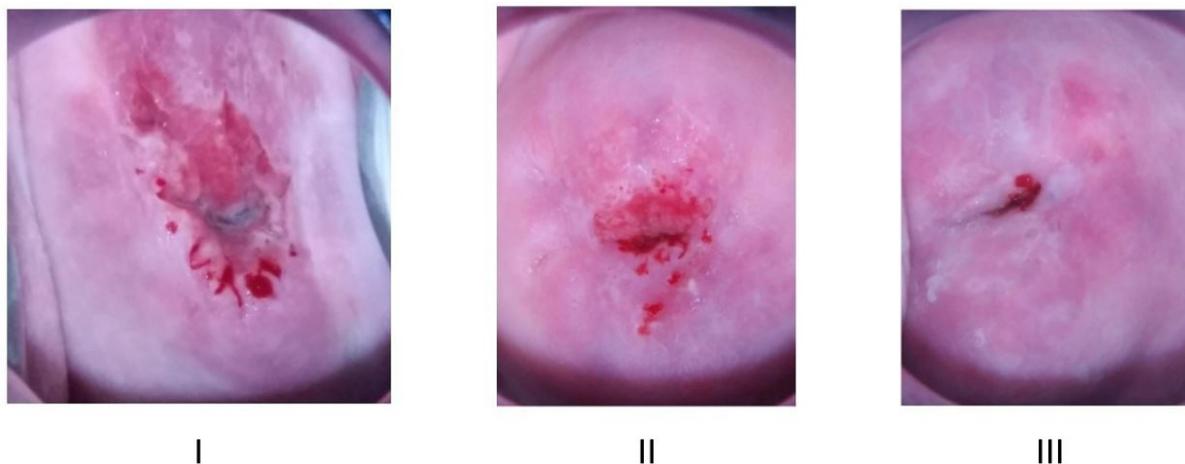
**Figura 1-3:** Cambios en el epitelio escamoso cervical por HPV, disponible en [2]

Así mismo, el diagnóstico se dificulta dependiendo del tipo de cérvix de la paciente, dado que esto genera que la zona de transformación presente una diferente ubicación; Además de que dependiendo del tipo, la CIN puede generar lesiones ocultas [17].

En la figura 1-4 se pueden evidenciar los 3 grados que puede tomar la CIN [1].

- Para el grado I, la lesión se encuentra solamente en el tejido del cérvix, afectando la zona de transformación.
- En el grado II, esta lesión ocupa parte del área del epitelio columnar.
- En el grado III la lesión deja totalmente el epitelio escamoso y se extiende por la pared del epitelio columnar.

A partir del grado II de la lesión es de suma importancia que se someta a la paciente a una biopsia, dado que, en estas etapas, es muy probable la presencia de cáncer cervical y no solo la lesión en tejido escamoso CIN [15].



**Figura 1-4:** Tipos de CIN, disponible en [1]

### 1.2.2 Colposcopía

Este es un procedimiento diagnóstico definido como: *Un examen de baja potencia iluminado del tracto epitelial del bajo genital* [18]. Tiene la finalidad de identificar anomalías tales como lesiones intraepiteliales escamosas de alto grado (HSIL por sus siglas en inglés high squamous intraepithelial lesion), CIN entre otras, en general identificación de lesiones a nivel cervical, para este procedimiento (Ver figura 1-5), se tienen los siguientes pasos [18]:

1. Ingreso del espéculo.
2. Evaluación inicial.

3. Vista de baja potencia, en esta fase se limpia la mayor cantidad de mucosa posible evitando daños en los tejidos. También es posible aplicar algunos filtros de luz; Usualmente se usa un filtro de luz blanca, es posible también el uso de un filtro de luz verde para analizar otros tejidos como los vasos sanguíneos en el área.
4. Examinación epitelial, para esta fase, se alterna entre una vista de baja y alta potencia, solo en caso de ser necesario, esto dependerá de la experiencia del especialista encargado de realizar el procedimiento.

Pese a ser un procedimiento relativamente sencillo, es común que genere cierto grado de estrés a los pacientes dado que puede generar molestia, cierto grado de dolor, temor dada la posibilidad de presentar cáncer, por ende, afectan el correcto desarrollo del procedimiento. Para atacar esta problemática, se han dado estudios en los que se evalúa el grado de estrés de los pacientes durante el procedimiento mientras se distrae su atención con música, vídeos u otras prácticas [19]. Adicional a esto, la falta de experiencia del especialista a la hora de realizar el procedimiento también puede afectar la calidad de las imágenes obtenidas y por ende, la precisión del diagnóstico. Por esta razón la sociedad americana para colposcopia y patología cervical (ASCCP por sus siglas en inglés American Society for Colposcopy and Cervical Pathology), realizó un estudio en el que se generaron 11 indicadores de calidad para la aplicación de este procedimiento [20]. Pese a que iniciativas como estas favorecen a un mejor manejo de los equipos, aún existe un gran vacío en la obtención de imágenes de buena calidad para dar un correcto diagnóstico.

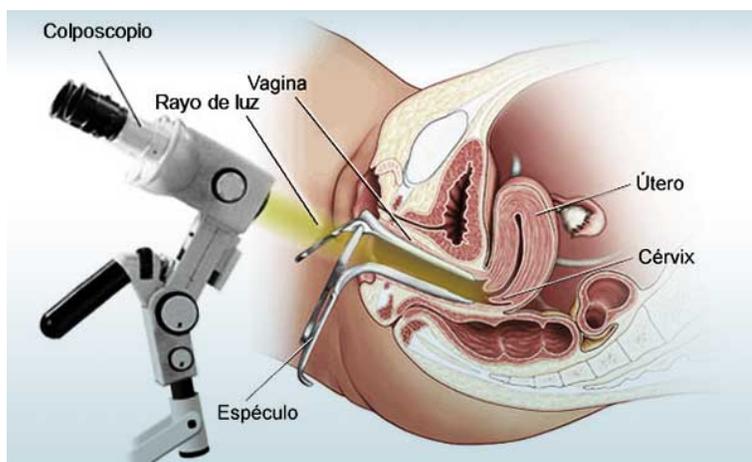


Figura 1-5: Procedimiento de colposcopia, disponible en [3]

## 1.3 Inteligencia artificial

A mediados del siglo XX surgió el concepto de inteligencia artificial. Existe una placa en el Dartmouth College que data de 1956 en la que se expresa el primer uso del término *Inteligencia*

*artificial* [21]. La real academia de la lengua española, presenta varias definiciones al concepto de inteligencia, entre ellas se destacan [22]:

- Capacidad de entender o comprender.
- Capacidad de resolver problemas.
- Habilidad, destreza y experiencia.

Así mismo, define la inteligencia artificial como *Una disciplina científica que se ocupa de crear programas informáticos que ejecutan operaciones comparables a las que realiza la mente humana, como el aprendizaje o el razonamiento lógico* [22]. Finalmente, una de las definiciones más específicas se encuentra en la IEEE *Máquinas capaces de realizar tareas que requieren inteligencia humana* [23]. Estos conceptos buscaban emular el aprendizaje de los seres humanos, a partir de experiencias previas reflejadas en generalizaciones a partir de dicho aprendizaje; A lo largo de los años, se han generado diferentes aplicativos aplicativos de estos conceptos [4]:

- Inteligencia computacional, en la que se desarrollan temas como Lógica difusa (Fuzzy logic). Redes neuronales artificiales (ANNs por sus siglas en inglés Artificial Neuronal Networks), las cuales han sido implementadas para problemas de clasificación de características biológicas, datos químico-cinéticos, problemáticas clínicas y hospitalarias, entre otros. Computación Evolutiva (Evolutionary Computation) e Inteligencia de Enjambre (Swarm Intelligence).
- Aprendizaje de máquina (ML por sus siglas en inglés Machine Learning), en este campo se desarrollan ANNs, Máquinas de Soporte Vectorial, Bosques Aleatorios (Random Forest), Modelos ocultos de Markov, entre otros.
- Aprendizaje profundo (DL por sus siglas en inglés Deep Learning), en este campo se desarrollan modelos basados en ANNs, aumentando el número de capas ocultas, concluyendo en ANNs mucho más robustos como se ve en la figura **1-6**. Esta gran cantidad de capas ocultas permite que estos modelos realicen el proceso de extracción de características de manera automática [24].

El aprendizaje de cualquier algoritmo de inteligencia artificial se ve reflejado en la obtención de una función matemática  $h(x)$  que soluciona el problema. A partir de esto, se selecciona un tipo de aprendizaje específico para el tipo de problema que se necesite solucionar, para este caso se tienen algunos tipos de aprendizaje [25]:

- *Supervisado*: Se da al generar el modelo matemático a partir de un conjunto de datos que contiene las entradas y salidas deseadas. Esto se aplica para problemas de clasificación y regresión.

- *No supervisado*: Se somete el modelo solo al conjunto de datos de entrenamiento sin etiquetas, a partir de esto el modelo genera agrupamiento (Clustering), de dichos datos mediante la identificación de patrones entre ellos.
- *Semi- supervisado* : Se da al someter el modelo a un conjunto de datos de entrenamiento dividido en datos que contienen y no contienen etiquetas.
- Adicional a estos, se tienen otras opciones dependiendo de las necesidades del investigador como lo son, Aprendizaje de Refuerzo, Aprendizaje Propio, Detección de anomalías, Asociaciones de reglas, entre otros.

Específicamente en el campo de la medicina se aplica en 4 campos [26]:

- Análisis de imágenes.
- Diseño de fármacos.
- Sistemas de asistencia diagnóstica.
- Análisis bioquímicos.

En el campo de la salud ha ganado fama en los últimos años al proveer herramientas de alta calidad para el procesamiento de datos en tiempo real, análisis de imágenes oncológicas, implementación en dispositivos portátiles, al mismo tiempo que van de la mano junto al campo de la interconexión de equipos electrónicos internet de las cosas (IoT por sus siglas en inglés Internet of Things) y el procesamiento de grandes cantidades de datos denominado Big Data [27].

### 1.3.1 Aprendizaje profundo

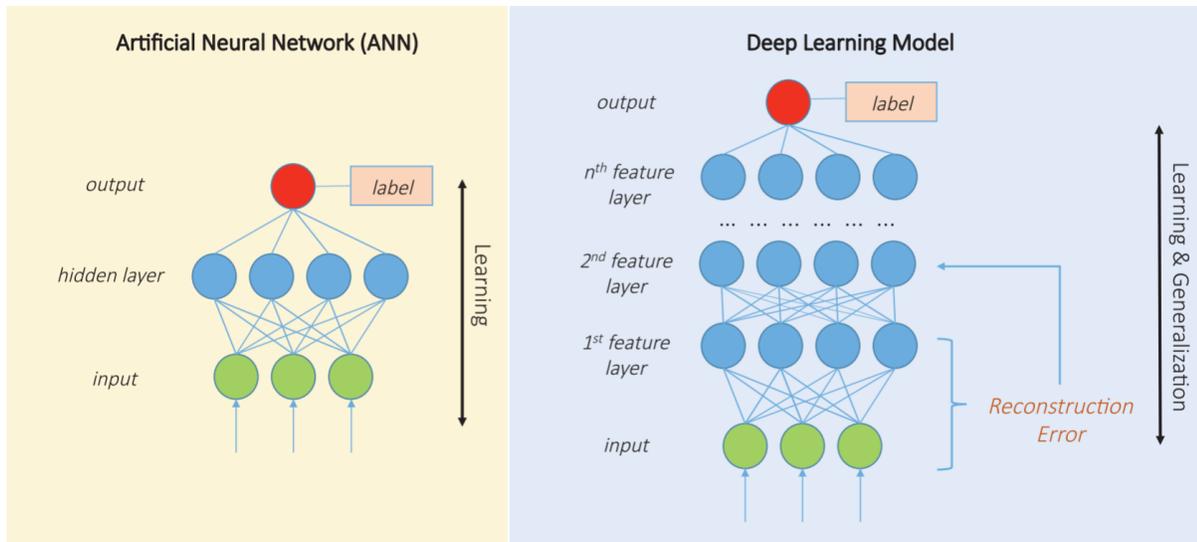
Los algoritmos de ML implementan 4 pasos generales, armonización de muestras, representación del aprendizaje, ajuste del modelo y evaluación del modelo [28]. La mayor ventaja que presenta el aprendizaje profundo es la capacidad de generar múltiples niveles de abstracción a partir de los datos de entrenamiento, esto se logra gracias a que en comparación a las ANNs, los modelos de aprendizaje profundo implementan capas de neuronas ocultas, esta comparación entre sus arquitecturas se evidencia en la figura 1-6.

En base a esto, se han generado diferentes arquitecturas, entre estas [4]:

- Redes neuronales convolucionales (CNNs por sus siglas en inglés Convolutional Neural Network).
- Redes neuronales recurrentes (RNNs por sus siglas en inglés Recurrent Neural Networks).

- Máquinas restringidas de Boltzmann (RBM por sus siglas en inglés Restricted Boltzmann Machine).
- Autoencoders (AE).

Para aplicaciones en imágenes, por lo general estas nuevas redes funcionan basadas en la implementación de series de capas convolucionales, seguidas de capas de *pooling*. Este proceso logra realizar una reducción de la dimensionalidad al mismo tiempo que las neuronas empiezan a realizar un proceso de extracción de características, finalmente en sus últimas capas, estos modelos implementan capas densas para realizar el proceso de clasificación [4].



**Figura 1-6:** Arquitectura de Redes neuronales artificiales y aprendizaje profundo, disponible en [4]

### 1.3.2 Aplicaciones en Neoplasia Intraepitelial Cervical

Muchos de los aplicativos de este tipo de desarrollos se basan en el análisis de imágenes obtenidas generalmente mediante el procedimiento de colposcopia. En algunos casos como lo es el de Oluwatomisin E, et al. 2021 llevaron a cabo su estudio *Classification of Cervical Intraepithelial Neoplasia (CIN) using fine-tuned Convolutional Neuronal Networks*, implementaron técnicas de transferencia de aprendizaje (Transfer learning) a partir de la red DenseNet, enfocado en generar un modelo de clasificación entre paciente sana, CIN I, II o III y cáncer [29].

Existen aplicaciones a datos desde otros enfoques como el caso de Sornapudi, Sudhir and Stanley, Ronald Joe, et al. en 2018 con su estudio *Deep learning nuclei detection in digitized histology images by superpixels*, enfocado a la segmentación de núcleos en células del epitelio

---

mediante la generación de superpíxeles utilizando algoritmos de agrupamiento linear simple para entrenar una red neuronal convolucional [30].

El mayor problema en los últimos tiempos de la inteligencia artificial es la falta de bases de datos adecuadas para implementar modelos. Por esta razón la plataforma Kaggle sube retos de diversos tipos, en un caso específico junto a Intel Y MobileODT, en 2017 generaron el reto *Intel & MobileODT Cervical Cancercer Screening*, buscando que se generara la mejor herramienta de asistencia para diagnóstico [1]. Así mismo como Yao Yu, et al. generaron una base de datos en 2021 a partir de 679 pacientes de 4753 imágenes [31]. En estos últimos 2 trabajos se evidenciaron lesiones de CIN de grados I, II y III, esto dado que una de las mayores problemáticas radica en la capacidad de discernimiento sobre si una lesión es de grado II, dado que ya es una señal de alerta para proceder a una biopsia para descartar si la lesión solo se trata de una displasia cervical o ya se trata de la formación/presencia de cáncer cervical. Impactando el posible tratamiento de la paciente en cuestión.

## **2 Objetivos**

### **2.1 General**

Generar un método basado en aprendizaje profundo para la detección de al menos CIN II en el banco de imágenes de colposcopia de Intel & MobileODT Cervical Cancer Screening.

### **2.2 Específicos**

1. Realizar un pre-procesamiento de la base de datos a fin de estandarizar todas las imágenes.
2. Implementar modelos de aprendizaje profundo para la extracción de características y clasificación de imágenes de colposcopia.
3. Comparar el rendimiento del modelo con los resultados obtenidos mediante inspección visual para la detección de CIN II.

## 3 Metodología

El desarrollo de este proyecto tendrá contempladas 3 fases generales cada una con sus actividades respectivas, para el uso de un modelo de DL [32]. Generando para cada fase un algoritmo aplicado como código en Python, dado que este es un lenguaje que presenta un óptimo rendimiento para tareas tanto de procesamiento de imágenes, como en proyectos de ML.

### 3.1 Fase 1: Preprocesamiento

El primer acercamiento de este proyecto es analizar la información de la base de datos, con esto fue posible identificar el tipo de paciente que se presenta en cada imagen de la base de datos. Dicho de otra forma, clasificar cada imagen de paciente dependiendo del grado de CIN mediante inspección visual y la información de la base de datos como:

- CIN I.
- CIN II.
- CIN III.

En total, la base de datos presentó 4753 imágenes (De tamaño 640x480 en formato RGB), obtenidas con un colposcopio de resolución 1050 TVL; A partir de 679 pacientes de nacionalidad China [31], separadas en secciones de imágenes entrenamiento, imágenes prueba, imágenes adicionales de entrenamiento e imágenes adicionales de prueba. Las secciones de entrenamiento y prueba presentaban imágenes obtenidas solo con filtro de luz blanca, mientras que las adicionales de entrenamiento y adicionales de prueba presentaban algunas implementando filtros de luz verde; Sin embargo, muchas de las imágenes de estas secciones adicionales se encontraban repetidas, por lo que no fueron usadas.

A partir de esto se estandarizaron las dimensiones de todas las imágenes de las secciones de entrenamiento y prueba a un tamaño de 224x224 y su representación en formato RGB. Esto se realizó de manera automática por código mediante las librerías Glob, PIL y Matplotlib, para tener imágenes suficientes tanto para el entrenamiento y prueba de las redes; Es decir, se segmentó la base de datos entre imágenes para entrenamiento y prueba. Contando con un total de 1480 imágenes para entrenamiento distribuidas como:

- 249 CIN I, etiquetadas como 001\_NúmeroDeImagen.
- 781 CIN II, etiquetadas como 010\_NúmeroDeImagen.
- 450 CIN III, etiquetadas como 100\_NúmeroDeImagen.

En esta fase se evidenció un problema de desbalance entre los grados de enfermedad, dado que las diferencias entre las cantidades para cada grado son bastante grandes. Para la sección de pruebas se contó con 512 muestras.

## 3.2 Fase 2: Entrenamiento

Dado que en los últimos años se ha dado un gran avance en diseño de modelos con óptimos resultados para tareas similares, es necesario revisar el rendimiento de algunos a fin de tener el mejor resultado posible. Gracias a bases de datos como *Keras Applications* es posible realizar dicha tarea dado que organiza y clasifica según valores como [33]:

- Tamaño de memoria.
- Exactitud a la hora de tomar decisiones de cada modelo.
- Profundidad de cada modelo.

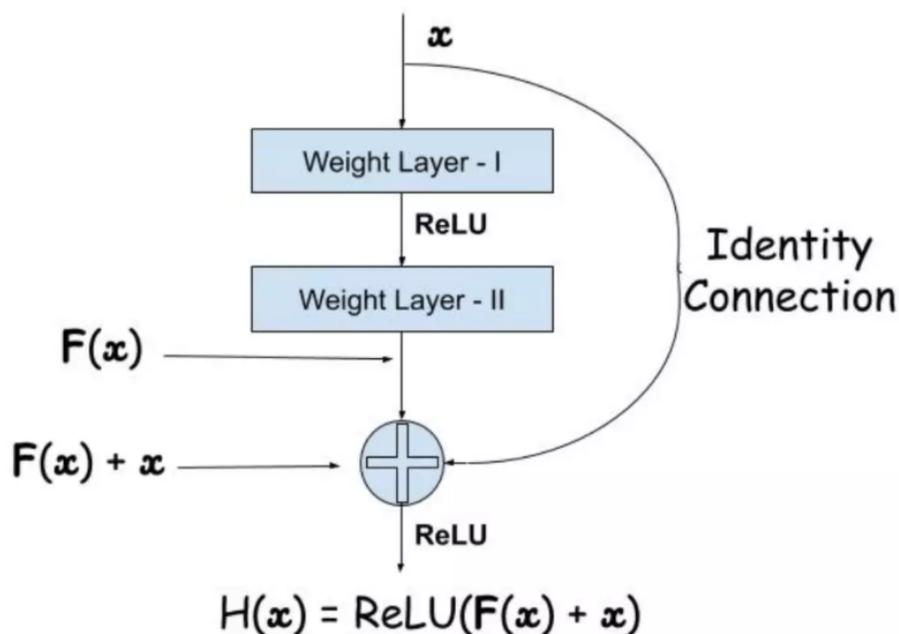
Es importante definir la arquitectura de los modelos implementados, todos estos son parte de la documentación de la librería *Keras Applications* [33].

### 3.2.1 ResNet-50 y ResNet-152

El modelo general para la ResNet consiste en 5 grupos de capas consecutivas. Adicional a esta arquitectura el modelo ResNet implementó un modelo de función residual (También conocido como mapeo residual), esta es una conexión similar a las redes que implementan una secuencia de capas de convolución, pooling, activación y conexión completa, para este caso se implementa la denominada *Identity connection* (Conexión identidad), entre capas conectando la entrada a un bloque con la salida en la función de activación [5], esta conexión se puede evidenciar en la figura 3-1.

Estas se encuentran en la siguiente distribución [6]:

- Para la primera etapa se tiene una capa convolucional, capa de normalización de lote, capa de activación con función ReLU (Rectificadora lineal unitaria) y una capa de max-pooling.



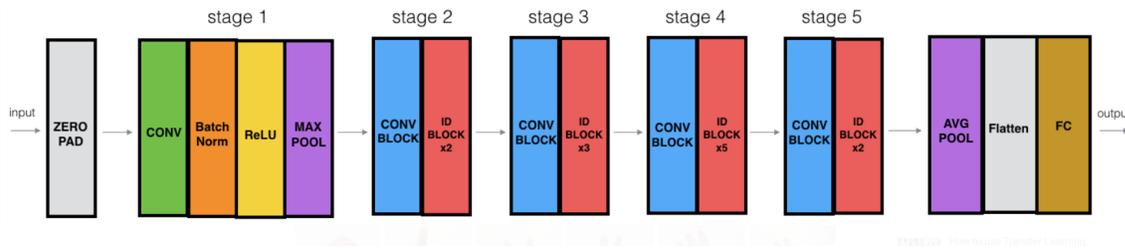
**Figura 3-1:** Red residual profunda, disponible en [5]

- Para las etapas 2 a 5 se tiene para cada una, una capa convolucional y un bloque de identidad.

En la figura 3-2 se muestra gráficamente la estructura de este modelo. Cabe resaltar que luego de sus primeras 5 etapas se encuentra una última etapa previa a la salida que consiste en una capa de Average pooling, una de Flattening y una de Conexión Completa (FC por sus siglas en inglés Fully Connected). A partir de este modelo inicial lo que diferencia la ResNet-50 a la ResNet-152 es la cantidad de capas que implementa de este modelo original, siendo 50 capas y 152 capas respectivamente. Finalmente se obtuvieron las características para la ResNet-50, estas son 98MB de peso, 25636712 parámetros de entrenamiento, Top 1 de exactitud de 0.749, Top5 de exactitud de 0.921 y un tiempo de inferencia en GPU de 4.55 ms. Para la ResNet-152 se tienen 232MB de peso, 60419944 parámetros de entrenamiento, Top 1 de precisión de 0.766, Top5 de precisión de 0.931 y un tiempo de inferencia en GPU de 6.54 ms [33].

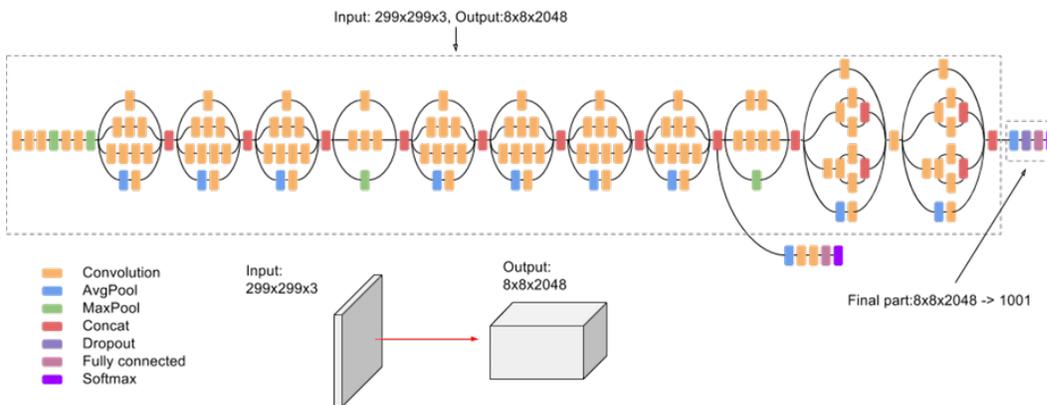
### 3.2.2 InceptionV3

Este modelo surgió luego de demostrar su gran rendimiento al ser entrenado a un conjunto de datos con 1 millón de imágenes, como es el conjunto ImageNet, demostrando un rendimiento superior a sus predecesores InceptionV1 e InceptionV2 [7]. Para este modelo Szegedy et al. [34] modificaron las anteriores versiones de este modelo, resultando en bloques que incluyeron



**Figura 3-2:** Arquitectura ResNet, disponible en [6]

capas de convolución, reducción promedio, concatenación, reducción máxima, retiro aleatorio (Dropout) y de conexión completa, aplicando la función de activación Softmax. Esta compleja arquitectura se puede evidenciar en la figura 3-3. Finalmente se obtiene un modelo que pesa 92MB, presenta 23851784 parámetros de entrenamiento, Top 1 de exactitud de 0.77, Top5 de exactitud de 0.937 y un tiempo de inferencia en GPU de 6.86 ms [33].



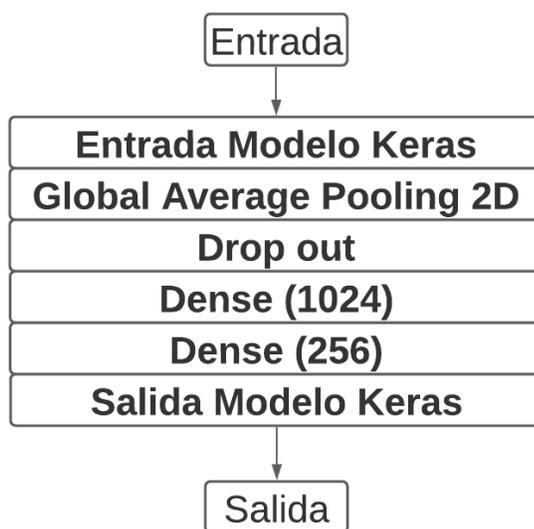
**Figura 3-3:** Arquitectura Inception V3, disponible en [7]

Una vez revisada esta base de datos y seleccionados los 3 modelos mostrados en la tabla 3-1. Se modificaron para entrenar con el segmento del dataset de imágenes seleccionado para este fin. Para esto, se tuvo en cuenta la arquitectura con capas adicionales a la salida del modelo mostrada en la figura 3-4. Todas estas capas como parte de la librería *Tensorflow* [35]. La capa de Global Average Pooling 2D se implementó para asegurar la dimensionalidad de las características extraídas por la red del modelo Keras, para este caso sus hiper-parámetros se mantuvieron por defecto, dado que al modificarlos afectó el rendimiento del modelo [33]. El paso por esta capa favorecía a que la clasificación realizada por las capas posteriores no presentara variaciones de dimensionalidad. Se definió una capa de Dropout para prevenir overfitting dado que, en experimentos previos sin esta capa, se lograba clasificar correctamente el grado II en las imágenes, pero los demás grados no presentaban buenos resultados. Adicional a esto, las funciones de activación para ambas capas Dense fueron definidas como

Relu. El modelo fue compilado con una pérdida asignada como entropía cruzada categórica al ser un problema muticlase, un optimizador Adam con un valor  $\epsilon$  de  $1e^{-4}$ .

Los experimentos llevados a cabo en esta sección se generalizaron a partir de 2 hiperparámetros del modelo:

- Para la capa de Dropout se hicieron iteraciones en su índice, variando dicho valor entre 0,1 hasta 0,5.
- Cantidad de épocas de entrenamiento se tomaron valores de 10, 25, 50 y 100 épocas.



**Figura 3-4:** Arquitectura general modelos

A partir de esto, se definió un punto de control (Checkpoint), para el entrenamiento del modelo, el cuál verificaba el valor de validación de pérdida para guardar los mejores pesos del modelo entrenado y tener el mejor modelo posible de entrenamiento. Dichos modelos con menor cantidad de épocas de entrenamiento, se usaron como base para el entrenamiento de los modelos con cantidad de épocas de entrenamiento consecuentes; Por ejemplo, los pesos del modelo con ResNet50 entrenado en 10 épocas, fueron la base para entrenar el modelo ResNet50 con 25 épocas.

ResNet50	InceptionV3	Resnet152
----------	-------------	-----------

**Tabla 3-1:** Modelos seleccionados

## 3.3 Fase 3: Validación e implementación con mejores modelos

### 3.3.1 Validación

Las redes entrenadas fueron probadas mediante el segmento del dataset destinado a pruebas, con el fin de identificar el rendimiento de cada red de acuerdo a métricas acordes a las características de la base de datos, estos cálculos mediante la librería Scikit-learn [36]. En este caso al ser problema de múltiples clases (CIN grado I, II o III), se tuvieron en cuenta las siguientes métricas:

- Precisión: Definida como la relación  $vp/(vp + fp)$  entre verdaderos positivos (vp) y falsos positivos (fp) [36].
- Recall: Definido como la relación  $vp/(vp + fn)$  entre verdaderos positivos (vp) y falsos negativos (fn) [36].
- F-score: Definido como la media armónica entre el recall y la precisión  $2 * (precision * recall)/(precision + recall)$ , este también es conocido como F1-score [36].

Adicional a esto, se implementó una matriz de confusión para tener una forma más evidente de los acuerdos y/o diferencias entre las etiquetas de respuesta del segmento de prueba de la base de datos y las predicciones realizadas por el modelo. Con esto, se calculó el índice Kappa, este es un método que funciona como un evaluador intermedio para medir la fiabilidad del modelo [8].

Valor de Kappa	Nivel de acuerdo	Porcentaje de fiabilidad
0-0.2	Nulo	0-4
0.21-0.39	Mínimo	4-15
0.40-0.59	Bajo	15-35
0.60-0.79	Moderado	35-63
0.80-0.90	Fuerte	64-81
Mayor a 0.9	Casi perfecto	82-100

**Tabla 3-2:** Interpretación del índice kappa, disponible en [8]

Esto con el fin de validar el índice y cantidad de épocas que generaban el mejor modelo posible, evitando problemas como el *underfitting* y *overfitting*.

### 3.3.2 Implementación

Para esta última parte, se sometieron los 3 mejores modelos, a un nuevo proceso de entrenamiento ajustando los pesos de cada clase, dado el problema presente de desbalance. A partir de esto repitió el proceso de validación.

## 3.4 Actividades

- Fase 1:
  1. Inspección visual para identificar la presencia de CIN en imágenes del dataset.
  2. Programación de un algoritmo de pre-procesamiento para dataset.
  3. Segmentación del dataset en entrenamiento y prueba.
- Fase 2: Esta será implementada en cada uno de los modelos
  1. Selección de modelos a partir de la documentación de la librería KERAS.
  2. Ajuste de parámetros y arquitectura para entrenamiento.
  3. Implementación del dataset de entrenamiento en el modelo, realizando pruebas de variación de hiper-parámetros (Cantidad de épocas, índice de Dropout, etc.).
- Fase 3:
  1. Análisis de rendimiento entre los modelos mediante sus métricas.
  2. Reentrenamiento de mejores modelos con pesos ajustados.
  3. Análisis de rendimiento entre los modelos reentrenados mediante sus métricas.
  4. Selección del mejor modelo.

## 4 Resultados

A continuación, se muestran los resultados obtenidos con el valor óptimo de índice de Dropout de 0.15, el cual fue explorado sistemáticamente para evitar presentar overfitting. Al sobrepasar el valor de 0.23 el modelo tendía a no clasificar las lesiones de grado I.

Para las matrices de confusión, las columnas representan las etiquetas predichas por el modelo sobre las imágenes seleccionadas para prueba y las filas representan las etiquetas reales de estas imágenes de prueba.

### 4.1 ResNet-50

#### 4.1.1 Exploración con 10 épocas

Métrica	Grado I	Grado II	Grado III
Precisión	0.793	0.765	0.533
Recall	0.526	0.644	0.78
F-score	0.632	0.699	0.633

Tabla 4-1: Métricas entrenamiento ResNet-50 entrenada con 10 épocas

Métrica	Grado I	Grado II	Grado III
Precisión	0.487	0.626	0.447
Recall	0.229	0.59	0.643
F-score	0.312	0.596	0.528

Tabla 4-2: Métricas prueba ResNet-50 entrenada con 10 épocas

	<b>I</b>	<b>II</b>	<b>III</b>
<b>I</b>	131	61	57
<b>II</b>	28	503	250
<b>III</b>	6	93	351

Tabla 4-3: Matriz de confusión entrenamiento con 10 épocas ResNet-50

	<b>I</b>	<b>II</b>	<b>III</b>
<b>I</b>	20	37	30
<b>II</b>	17	151	97
<b>III</b>	4	53	103

**Tabla 4-4:** Matriz de confusión prueba con 10 épocas ResNet-50

Para los índices o coeficientes de Kappa, se obtuvo 0.453 y 0.228 para entrenamiento y prueba, calculados a partir de las matrices de confusión de las tablas **4-3** y **4-4** respectivamente).

#### 4.1.2 Exploración de ResNet-50 con 25 épocas

Métrica	Grado I	Grado II	Grado III
Precisión	0.99	0.756	0.528
Recall	0.40	0.644	0.835
F-score	0.579	0.695	0.647

**Tabla 4-5:** Métricas entrenamiento ResNet-50 entrenada con 25 épocas

Métrica	Grado I	Grado II	Grado III
Precisión	0.777	0.613	0.459
Recall	0.080	0.539	0.775
F-score	0.145	0.574	0.576

**Tabla 4-6:** Métricas prueba ResNet-50 entrenada con 25 épocas

	<b>I</b>	<b>II</b>	<b>III</b>
<b>I</b>	102	89	58
<b>II</b>	0	503	278
<b>III</b>	1	73	376

**Tabla 4-7:** Matriz de confusión entrenamiento con 25 épocas ResNet-50

	<b>I</b>	<b>II</b>	<b>III</b>
<b>I</b>	7	54	26
<b>II</b>	2	143	120
<b>III</b>	0	36	124

**Tabla 4-8:** Matriz de confusión prueba con 25 ResNet-50

Para los índices o coeficientes de Kappa, se obtuvo 0.443 y 0.221, para entrenamiento y prueba, calculados a partir de las matrices de confusión de las tablas 4-7 y 4-8 respectivamente).

### 4.1.3 Exploración de ResNet-50 con 50 épocas

Métrica	Grado I	Grado II	Grado III
Precisión	0.979	0.873	0.839
Recall	0.951	0.918	0.777
F-score	0.965	0.895	0.807

**Tabla 4-9:** Métricas entrenamiento ResNet-50 entrenada con 50 épocas

Métrica	Grado I	Grado II	Grado III
Precisión	0.608	0.679	0.594
Recall	0.32	0.72	0.687
F-score	0.421	0.699	0.637

**Tabla 4-10:** Métricas prueba ResNet-50 entrenada con 50 épocas

	<b>I</b>	<b>II</b>	<b>III</b>
<b>I</b>	237	7	5
<b>II</b>	2	717	62
<b>III</b>	3	97	350

**Tabla 4-11:** Matriz de confusión entrenamiento con 50 épocas ResNet-50

	<b>I</b>	<b>II</b>	<b>III</b>
<b>I</b>	28	44	15
<b>II</b>	14	191	60
<b>III</b>	4	46	110

**Tabla 4-12:** Matriz de confusión prueba con 50 ResNet-50

Para los índices Kappa, se obtuvo 0.8 y 0.469, para entrenamiento y prueba, calculados a partir de las matrices de confusión de las tablas 4-11 y 4-12 respectivamente).

#### 4.1.4 Exploración de ResNet-50 con 100 épocas

Métrica	Grado I	Grado II	Grado III
Precisión	0.991	0.907	0.835
Recall	0.991	0.907	0.835
F-score	0.991	0.907	0.835

**Tabla 4-13:** Métricas entrenamiento ResNet-50 entrenada con 100 épocas

Métrica	Grado I	Grado II	Grado III
Precisión	0.615	0.641	0.495
Recall	0.18	0.6	0.737
F-score	0.283	0.619	0.592

**Tabla 4-14:** Métricas prueba ResNet-50 entrenada con 100 épocas

	<b>I</b>	<b>II</b>	<b>III</b>
<b>I</b>	247	0	2
<b>II</b>	0	709	72
<b>III</b>	2	72	376

**Tabla 4-15:** Matriz de confusión entrenamiento con 100 épocas ResNet-50

	<b>I</b>	<b>II</b>	<b>III</b>
<b>I</b>	16	47	24
<b>II</b>	10	159	96
<b>III</b>	0	42	118

**Tabla 4-16:** Matriz de confusión prueba con 100 ResNet-50

Para los índices o coeficientes de Kappa, se obtuvo 0.83 y 0.282, para entrenamiento y prueba, calculados a partir de las matrices de confusión de las tablas 4-15 y 4-16 respectivamente).

## 4.2 InceptionV3

### 4.2.1 Exploración de InceptionV3 con 10 épocas

Métrica	Grado I	Grado II	Grado III
Precisión	0.849	0.858	0.875
Recall	1	0.916	0.686
F-score	0.918	0.88	0.769

**Tabla 4-17:** Métricas entrenamiento InceptionV3 entrenada con 10 épocas

Métrica	Grado I	Grado II	Grado III
Precisión	0.446	0.652	0.654
Recall	0.482	0.701	0.543
F-score	0.464	0.673	0.593

**Tabla 4-18:** Métricas prueba InceptionV3 entrenada con 10 épocas

	<b>I</b>	<b>II</b>	<b>III</b>
<b>I</b>	249	0	0
<b>II</b>	21	716	44
<b>III</b>	23	118	309

**Tabla 4-19:** Matriz de confusión entrenamiento con 10 épocas InceptionV3

	<b>I</b>	<b>II</b>	<b>III</b>
<b>I</b>	16	47	24
<b>II</b>	10	159	96
<b>III</b>	0	42	118

**Tabla 4-20:** Matriz de confusión prueba con 10 épocas InceptionV3

Para los índices o coeficientes de Kappa, se obtuvo 0.767 y 0.358, para entrenamiento y prueba, calculados a partir de las matrices de confusión de las tablas 4-19 y 4-20 respectivamente).

### 4.2.2 Exploración de InceptionV3 con 25 épocas

Métrica	Grado I	Grado II	Grado III
Precisión	0.946	0.862	0.952
Recall	0.987	0.975	0.713
F-score	0.966	0.915	0.815

**Tabla 4-21:** Métricas entrenamiento InceptionV3 entrenada con 25 épocas

Métrica	Grado I	Grado II	Grado III
Precisión	0.428	0.592	0.592
Recall	0.344	0.686	0.5
F-score	0.382	0.636	0.542

**Tabla 4-22:** Métricas prueba InceptionV3 entrenada con 25 épocas

	<b>I</b>	<b>II</b>	<b>III</b>
<b>I</b>	246	0	3
<b>II</b>	6	762	13
<b>III</b>	8	121	321

**Tabla 4-23:** Matriz de confusión entrenamiento con 25 épocas InceptionV3

	<b>I</b>	<b>II</b>	<b>III</b>
<b>I</b>	30	50	7
<b>II</b>	35	182	48
<b>III</b>	5	75	80

**Tabla 4-24:** Matriz de confusión prueba con 25 épocas InceptionV3

Para los índices o coeficientes de Kappa, se obtuvo 0.826 y 0.264, para entrenamiento y prueba, calculados a partir de las matrices de confusión de las tablas **4-23** y **4-24** respectivamente).

### 4.2.3 Exploración de InceptionV3 con 50 épocas

Métrica	Grado I	Grado II	Grado III
Precisión	0.979	0.846	0.859
Recall	0.951	0.942	0.704
F-score	0.965	0.892	0.774

**Tabla 4-25:** Métricas entrenamiento InceptionV3 entrenada con 50 épocas

Métrica	Grado I	Grado II	Grado III
Precisión	0.378	0.583	0.548
Recall	0.16	0.728	0.49
F-score	0.225	0.647	0.519

**Tabla 4-26:** Métricas prueba InceptionV3 entrenada con 50 épocas

	<b>I</b>	<b>II</b>	<b>III</b>
<b>I</b>	237	4	8
<b>II</b>	1	736	44
<b>III</b>	4	129	317

**Tabla 4-27:** Matriz de confusión entrenamiento con 50 épocas InceptionV3

	<b>I</b>	<b>II</b>	<b>III</b>
<b>I</b>	14	62	11
<b>II</b>	18	193	54
<b>III</b>	5	76	79

**Tabla 4-28:** Matriz de confusión prueba con 50 épocas InceptionV3

Para los índices o coeficientes de Kappa, se obtuvo 0.781 y 0.219, para entrenamiento y prueba, calculados a partir de las matrices de confusión de las tablas **4-27** y **4-28** respectivamente).

#### 4.2.4 Exploración de InceptionV3 con 100 épocas

Métrica	Grado I	Grado II	Grado III
Precisión	0.972	0.875	0.943
Recall	0.995	0.975	0.744
F-score	0.984	0.923	0.832

**Tabla 4-29:** Métricas entrenamiento InceptionV3 entrenada con 100 épocas

Métrica	Grado I	Grado II	Grado III
Precisión	0.482	0.609	0.59
Recall	0.31	0.705	0.55
F-score	0.377	0.653	0.569

**Tabla 4-30:** Métricas prueba InceptionV3 entrenada con 100 épocas

	<b>I</b>	<b>II</b>	<b>III</b>
<b>I</b>	248	0	1
<b>II</b>	0	762	19
<b>III</b>	7	108	335

**Tabla 4-31:** Matriz de confusión entrenamiento con 100 épocas InceptionV3

	<b>I</b>	<b>II</b>	<b>III</b>
<b>I</b>	27	51	9
<b>II</b>	26	187	52
<b>III</b>	3	69	88

**Tabla 4-32:** Matriz de confusión prueba con 100 épocas InceptionV3

Para los índices o coeficientes de Kappa, se obtuvo 0.845 y 0.293, para entrenamiento y prueba, calculados a partir de las matrices de confusión de las tablas **4-31** y **4-32** respectivamente).

## 4.3 ResNet-152

### 4.3.1 Exploración de ResNet-152 con 10 épocas

Métrica	Grado I	Grado II	Grado III
Precisión	0.861	0.814	0.869
Recall	0.875	0.928	0.651
F-score	0.868	0.867	0.744

**Tabla 4-33:** Métricas entrenamiento ResNet-152 entrenada con 10 épocas

Métrica	Grado I	Grado II	Grado III
Precisión	0.403	0.609	0.614
Recall	0.264	0.754	0.487
F-score	0.319	0.674	0.543

**Tabla 4-34:** Métricas prueba ResNet-152 entrenada con 10 épocas

	<b>I</b>	<b>II</b>	<b>III</b>
<b>I</b>	218	23	8
<b>II</b>	20	725	36
<b>III</b>	15	142	293

**Tabla 4-35:** Matriz de confusión entrenamiento con 10 épocas ResNet-152

	<b>I</b>	<b>II</b>	<b>III</b>
<b>I</b>	23	55	9
<b>II</b>	25	200	40
<b>III</b>	9	73	78

**Tabla 4-36:** Matriz de confusión prueba con 10 épocas ResNet-152

Para los índices o coeficientes de Kappa, se obtuvo 0.718 y 0.28, para entrenamiento y prueba, calculados a partir de las matrices de confusión de las tablas **4-35** y **4-36** respectivamente).

### 4.3.2 Exploración de ResNet-152 con 25 épocas

Métrica	Grado I	Grado II	Grado III
Precisión	0.913	0.842	0.838
Recall	0.807	0.957	0.693
F-score	0.857	0.896	0.759

**Tabla 4-37:** Métricas entrenamiento ResNet-152 entrenada con 25 épocas

Métrica	Grado I	Grado II	Grado III
Precisión	0.5	0.644	0.686
Recall	0.229	0.822	0.575
F-score	0.314	0.723	0.625

**Tabla 4-38:** Métricas prueba ResNet-152 entrenada con 25 épocas

	<b>I</b>	<b>II</b>	<b>III</b>
<b>I</b>	201	18	30
<b>II</b>	3	748	30
<b>III</b>	16	122	312

**Tabla 4-39:** Matriz de confusión entrenamiento con 25 épocas ResNet-152

	<b>I</b>	<b>II</b>	<b>III</b>
<b>I</b>	20	56	11
<b>II</b>	16	218	31
<b>III</b>	4	64	92

**Tabla 4-40:** Matriz de confusión prueba con 25 épocas ResNet-152

Para los índices o coeficientes de Kappa, se obtuvo 0.819 y 0.369, para entrenamiento y prueba, calculados a partir de las matrices de confusión de las tablas **4-39** y **4-40** respectivamente).

### 4.3.3 Exploración de ResNet-152 con 50 épocas

Métrica	Grado I	Grado II	Grado III
Precisión	0.989	0.884	0.641
Recall	0.771	0.793	0.833
F-score	0.866	0.836	0.724

**Tabla 4-41:** Métricas entrenamiento ResNet-152 entrenada con 50 épocas

Métrica	Grado I	Grado II	Grado III
Precisión	0.5	0.666	0.47
Recall	0.09	0.558	0.818
F-score	0.155	0.60	0.602

**Tabla 4-42:** Métricas entrenamiento ResNet-152 entrenada con 50 épocas

	<b>I</b>	<b>II</b>	<b>III</b>
<b>I</b>	192	7	50
<b>II</b>	1	620	160
<b>III</b>	1	74	375

**Tabla 4-43:** Matriz de confusión entrenamiento con 50 épocas ResNet-152

	<b>I</b>	<b>II</b>	<b>III</b>
<b>I</b>	8	44	35
<b>II</b>	8	144	109
<b>III</b>	0	29	131

**Tabla 4-44:** Matriz de confusión prueba con 50 épocas ResNet-152

Para los índices o coeficientes de Kappa, se obtuvo 0.674 y 0.272, para entrenamiento y prueba, calculados a partir de las matrices de confusión de las tablas **4-43** y **4-44** respectivamente).

#### 4.3.4 Exploración de ResNet-152 con 100 épocas

Métrica	Grado I	Grado II	Grado III
Precisión	1	0.888	0.854
Recall	0.991	0.924	0.797
F-score	0.995	0.905	0.825

**Tabla 4-45:** Métricas entrenamiento ResNet-152 entrenada con 100 épocas

Métrica	Grado I	Grado II	Grado III
Precisión	0.642	0.672	0.558
Recall	0.206	0.705	0.718
F-score	0.313	0.688	0.628

**Tabla 4-46:** Métricas prueba ResNet-152 entrenada con 100 épocas

	<b>I</b>	<b>II</b>	<b>III</b>
<b>I</b>	247	0	2
<b>II</b>	0	722	59
<b>III</b>	0	91	359

**Tabla 4-47:** Matriz de confusión entrenamiento con 100 épocas ResNet-152

	<b>I</b>	<b>II</b>	<b>III</b>
<b>I</b>	18	46	23
<b>II</b>	10	187	68
<b>III</b>	0	45	115

**Tabla 4-48:** Matriz de confusión prueba con 100 épocas ResNet-152

Para los índices o coeficientes de Kappa, se obtuvo 0.828 y 0.358, para entrenamiento y prueba, calculados a partir de las matrices de confusión de las tablas 4-47 y 4-48 respectivamente).

## 4.4 Mejores modelos con pesos ajustados

### 4.4.1 ResNet-50 con 50 épocas

Métrica	Grado I	Grado II	Grado III
Precisión	0.98	0.84	0.94
Recall	0.983	0.978	0.671
F-score	0.981	0.904	0.783

**Tabla 4-49:** Métricas entrenamiento ResNet-50 con pesos ajustados entrenada con 50 épocas

Métrica	Grado I	Grado II	Grado III
Precisión	0.558	0.616	0.646
Recall	0.275	0.788	0.525
F-score	0.369	0.682	0.579

**Tabla 4-50:** Métricas prueba ResNet-50 con pesos ajustados entrenada con 50 épocas

	<b>I</b>	<b>II</b>	<b>III</b>
<b>I</b>	245	1	3
<b>II</b>	1	764	16
<b>III</b>	4	144	302

**Tabla 4-51:** Matriz de confusión entrenamiento con pesos ajustados y 50 épocas ResNet-50

	<b>I</b>	<b>II</b>	<b>III</b>
<b>I</b>	24	57	6
<b>II</b>	16	209	40
<b>III</b>	3	73	84

**Tabla 4-52:** Matriz de confusión prueba con pesos ajustados y 50 épocas ResNet-50

Para los índices o coeficientes de Kappa, se obtuvo 0.804 y 0.324, para entrenamiento y prueba, calculados a partir de las matrices de confusión de las tablas **4-51** y **4-52** respectivamente).

### 4.4.2 InceptionV3 con 100 épocas

Métrica	Grado I	Grado II	Grado III
Precisión	0.813	0.855	0.976
Recall	0.995	0.956	0.655
F-score	0.895	0.903	0.784

**Tabla 4-53:** Métricas entrenamiento InceptionV3 con pesos ajustados entrenada con 100 épocas

Métrica	Grado I	Grado II	Grado III
Precisión	0.403	0.619	0.616
Recall	0.574	0.626	0.462
F-score	0.473	0.622	0.528

**Tabla 4-54:** Métricas prueba InceptionV3 con pesos ajustados entrenada con 100 épocas

	<b>I</b>	<b>II</b>	<b>III</b>
<b>I</b>	248	0	1
<b>II</b>	28	747	6
<b>III</b>	29	126	295

**Tabla 4-55:** Matriz de confusión entrenamiento con pesos ajustados y 100 épocas InceptionV3

	<b>I</b>	<b>II</b>	<b>III</b>
<b>I</b>	50	29	8
<b>II</b>	61	166	38
<b>III</b>	13	73	74

**Tabla 4-56:** Matriz de confusión prueba con pesos ajustados y 100 épocas InceptionV3

Para los índices o coeficientes de Kappa, se obtuvo 0.783 y 0.295, para entrenamiento y prueba, respectivamente (Calculados a partir de las matrices de confusión de las tablas **4-55** y **4-56** respectivamente).

### 4.4.3 ResNet-152 con 25 épocas

Métrica	Grado I	Grado II	Grado III
Precisión	0.99	0.91	0.66
Recall	0.827	0.791	0.882
F-score	0.901	0.846	0.761

**Tabla 4-57:** Métricas entrenamiento ResNet-152 con pesos ajustados entrenada con 25 épocas

Métrica	Grado I	Grado II	Grado III
Precisión	0.5	0.663	0.519
Recall	0.126	0.581	0.837
F-score	0.201	0.619	0.641

**Tabla 4-58:** Métricas prueba ResNet-152 con pesos ajustados entrenada con 25 épocas

	<b>I</b>	<b>II</b>	<b>III</b>
<b>I</b>	206	8	35
<b>II</b>	2	618	161
<b>III</b>	0	53	397

**Tabla 4-59:** Matriz de confusión entrenamiento con pesos ajustados y 25 épocas ResNet-152

	<b>I</b>	<b>II</b>	<b>III</b>
<b>I</b>	11	53	23
<b>II</b>	10	154	101
<b>III</b>	1	25	134

**Tabla 4-60:** Matriz de confusión prueba con pesos ajustados y 25 épocas ResNet-152

Para los índices Kappa, se obtuvo 0.714 y 0.307, para entrenamiento y prueba, respectivamente (Calculados a partir de las matrices de confusión de las tablas **4-59** y **4-59** respectivamente).

## 5 Discusión

Durante la exploración sistemática de hiper-parámetros y arquitecturas, encontramos que la capa Dropout juega un papel vital en la tarea de clasificación. La variación de la tasa de Dropout presentó un óptimo valor de 0.15 para evitar overfitting durante los experimentos. Adicionalmente, valores de tasa de dropout mayores de 0.2 nos presentaban un *underfitting* durante la clasificación de las imágenes. Además, se evidenció, que este valor optimizaba el tiempo que tomó el entrenamiento al momento de tener las mayores cantidades de épocas (50 y 100). En experimentos previos se optó por utilizar un punto de control tipo *early stop*, para optimizar el tiempo de entrenamiento de los modelos; Este punto de control fue descartado, dado que impedía un correcto desarrollo del proceso de aprendizaje de los modelos.

Esta base de datos específicamente presentó grandes limitantes para el problema. Inicialmente el desbalance en las clases, dado que, pese a que favorecía a las imágenes de lesiones de grado II y III, limitaba bastante las imágenes de grado I, esto se puede evidenciar dado que en Precisión y Recall, rara vez superaron umbrales de 0,4; Evidenciando que aún existe una gran cantidad de falsos positivos y falsos negativos predichos por estos modelos. Pese a que existían muchas más imágenes que las usadas, estas imágenes adicionales eran en gran medida copias de las usadas o copias de las mismas, así que la depuración de estas imágenes adicionales hubiera tenido como repercusión el tiempo dedicado para experimentación con los modelos. Este problema generó que los modelos presentaran una predisposición por las lesiones de grado II y III. Este desbalance fue atacado con un reajuste de los pesos de cada clase, se evidenció específicamente para cada modelo que:

- Para la **ResNet-50 entrenada con 50 épocas**: Originalmente este modelo presentaba un buen rendimiento para las lesiones de grado II y III al ser sometido al set de imágenes de prueba como se muestra en la tabla 4-10. En este modelo se presentó el mayor índice kappa de todos, teniendo un 0,469. Comparando con la tabla 4-50 se evidencia que se redujo el rendimiento del modelo para las lesiones de grado I, Precisión y Recall bajaron de 0,60 y 0,32 a 0,55 y 0,27 respectivamente. En este caso, su índice kappa se redujo a 0,324 demostrando que el reajuste de pesos no fue beneficioso para este modelo.
- En el caso de la **InceptionV3 entrenada con 100 épocas**: Este modelo originalmente presentaba un buen rendimiento para las clases II y III al ser sometido al set de prueba. Para estos casos, ninguna de la métricas presentó valores menores a 0.5

**4-30**, el gran problema de este modelo se encontraba con los falsos negativos, evidenciado con un recall de 0,31 que a su vez, generaba 0,377 para su f-score, en este caso estos valores aumentaron considerablemente, mientras que se dió una reducción en la precisión, aumentando la cantidad de errores por falsos positivos (Ver tabla **4-54**). Al analizar, su valor de Recall aumentó por encima de 0,55 pero la precisión cayó a 0,4. Su índice kappa original fue de 0,293 y al modificar los pesos fue de 0,295; Esto puede ser un indicador que para este modelo, el reajuste de pesos fue beneficioso.

- En el caso de la **ResNet-152 entrenada con 25 épocas**: Este modelo fue el que tuvo un mejor rendimiento comparado con la cantidad de épocas que necesitó para obtener métricas buenas, este modelo inicialmente solo presentaba problemas con los falsos negativos, al presentar un recall de 0,229 (Ver tabla **4-38**), adicional a esto, presentaba un índice kappa de 0,272. Al implementar el ajuste de pesos, no se presentaron grandes mejorías, dado que, aunque hubo un aumento en el valor de recall para las imágenes de Grado III, esta métrica se redujo para las demás clases. Finalmente, su índice kappa aumentó a 0,307.

Adicional a este problema de desbalance, el rendimiento de los modelos podría haber cambiado si se hubieran contado con imágenes de pacientes sanas. Adicional a esto, solo se implementaron modelos de clasificación dado que no se presentaron máscaras para entrenar un modelo de segmentación o generar datos para un modelo tipo Yolo. Esto demuestra la necesidad de bases de datos de buena calidad y una depuración acorde a estas. Otra posible alternativa podría ser el generar una base de datos con muestras de la región (Latinoamérica), para evidenciar si existe un comportamiento diferente de dichos modelos. Sin embargo, se evidenció que es posible la aplicación de modelos de clasificación que presenten un buen acercamiento a la clasificación correcta de al menos el grado II de la lesión.

## 6 Conclusiones

El desarrollo de este trabajo dirigido permitió evidenciar que si es posible la identificación de neoplasia cervical intraepitelial, útil en la predicción cáncer cervical mediante el uso de modelos basados en aprendizaje profundo. Se evidenció que, para este problema en particular, el modelo InceptionV3 modificado, con sus pesos ajustados fue el modelo con el mejor rendimiento.

Existen muchas más combinaciones de hiper-parámetros y arquitecturas que podrían ser implementadas a este modelo para incrementar aún más su rendimiento. El problema de este y la mayoría de los demás modelos implementados fue su predisposición por predecir lesiones de Grado II y III, esto dado que eran las clases que más cantidad de muestras tenían. Pero es evidente la capacidad que tienen los modelos de aprendizaje profundo para resolver problemáticas en el campo clínico, así como se evidencia que la disponibilidad de datos es una gran barrera para permitir experimentos con muestras de clases relativamente balanceadas; Este fue el caso para estos datos, pese a que se logró la implementación de modelos de clasificación, la ausencia de máscaras de las imágenes limitó la implementación de modelos de segmentación. Por esto es importante la labor que llevan a cabo plataformas como Kaggle para liberar conjuntos de datos que le permitan a los investigadores desarrollar nuevas herramientas en diferentes campos.

El contar con una mayor cantidad de imágenes con otros tipos de filtros de luz al interior de la base de datos hubiera permitido un ajuste diferente de estos modelos de clasificación. Esto se evidencia en trabajos previos realizados sobre la misma base de datos (Yao Yu, et al. [31]), en las que se tienen en cuenta arquitecturas específicas para imágenes con filtro de luz verde y blanca. Adicional a que generan arquitecturas propias, en estos reportan exactitud del 96%, recall de 95% y especificidad del 98%.

Pese a que es un resultado favorable la correcta clasificación de lesiones de grado II, es indispensable que este tipo de herramientas sean correctamente ajustadas para clasificar lesiones de grado I, dado que entre estos grados de severidad de la neoplasia intraepitelial cervical radica la necesidad de someter a la paciente a otro tipo de procedimientos diagnósticos como sería una biopsia, con la finalidad de descartar si se trata solo de una displasia cervical o si se trata de células cancerígenas en desarrollo.

# 7 Trabajos futuros

Consideramos que existen dos ítems a explorar en otras tesis o trabajos dirigidos como son:

1. La implementación de modelos de segmentación como la UNet, mediante la generación de una base de datos de máscaras mediante la aplicación de técnicas de procesamiento de imágenes o la segmentación de un experto.
2. Aplicación de redes más livianas como la MobileNet, EfficientNet o la DenseNet, mediante transfer learning partiendo de estos primeros modelos generados para la implementación en dispositivos móviles.

# Bibliography

- [1] MobileODT Kaggle, Intel. Cervix types classification, 2017. Disponible en: <https://www.kaggle.com/c/intel-mobileodt-cervical-cancer-screening/overview/description>.
- [2] Annekathryn Goodman and David Wilbur. Case 32-2003: A 37-year-old woman with atypical squamous cells on a papanicolaou smear. *The New England Journal of Medicine*, 2003.
- [3] Colposcopia. información, procedimiento y aplicaciones. Disponible en: <https://www.luciaviorginecologa.org/colposcopia-asturias/>.
- [4] Riccardo Miotto, Fei Wang, Shuang Wang, Xiaoqian Jiang, and Joel T Dudley. Deep learning for healthcare: review, opportunities and challenges. *Briefings in bioinformatics*, 19(6):1236–1246, 2018.
- [5] Ankit Sachan. Detailed guide to understand and implement resnets, 2017. Disponible en: <https://cv-tricks.com/keras/understand-implement-resnets/>.
- [6] Dwivedi Priya. Understanding and coding a resnet in keras, 2019. Disponible en: <https://towardsdatascience.com/understanding-and-coding-a-resnet-in-keras-446d7ff84d33>.
- [7] Google Cloud. Guía avanzada de inception v3 para cloud. Disponible en: <https://cloud.google.com/tpu/docs/inception-v3-advancedintroduction>.
- [8] Mary L McHugh. Interrater reliability: the kappa statistic. *Biochemia medica*, 22(3):276–282, 2012.
- [9] Vicki B. Benard, Philip E. Castle, Steven A. Jenison, William C. Hunt, Jane J. Kim, Jack Cuzick, Ji-Hyun Lee, Ruofei Du, Michael Robertson, Scott Norville, Cosette M. Wheeler, and for the New Mexico HPV Pap Registry Steering Committee. Population-based incidence rates of cervical intraepithelial neoplasia in the human papillomavirus vaccine era. *JAMA Oncology*, 3, 2017.

- 
- [10] Weiderpass E, Bruni L, de Sanjosé S, Saraiya M, Ferlay J, Bray F, Arbyn, M. Estimates of incidence and mortality of cervical cancer in 2018: a worldwide analysis. *The Lancet Global Health*, 2019.
- [11] Hyuna Sung, Jacques Ferlay, Rebecca L Siegel, Mathieu Laversanne, Isabelle Soerjomataram, Ahmedin Jemal, and Freddie Bray. 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, 2021.
- [12] Nijman HW, Slangen BF, Van Gorp T, Kruse AJ, Koeneman MM, Kruitwagen RF. Natural history of high-grade cervical intraepithelial neoplasia: a review of prognostic biomarkers. *expert rev mol diagn*. 2015.
- [13] Ugo Indraccolo, Erica Santi, Piergiorgio Iannone, Chiara Borghi, and Pantaleo Greco. Number of colposcopic cervical biopsies and diagnosis of cervical intraepithelial neoplasia: a prospective study. *European Journal of Gynaecological Oncology*, 42(4):649–654, 2021.
- [14] Melissa S Lopez, Ellen S Baker, Mauricio Maza, Georgia Fontes-Cintra, Aldo Lopez, Juan M Carvajal, Fernanda Nozar, Veronica Fiol, and Kathleen M Schmeler. Cervical cancer prevention and treatment in latin america. *Journal of surgical oncology*, 115(5):615–618, 2017.
- [15] World Health organization. Cervical cancer screening manual, 2014. Disponible en: <https://www.moh.gov.bt/wp-content/uploads/ict-files/2014/11/Cervical-Cancer-screening-manual-2014.pdf>.
- [16] Sezgin M Ismail, Angela B Colclough, John S Dinnen, Douglas Eakins, DM Evans, Ernest Gradwell, Jerry P O’Sullivan, Joan M Summerell, and Robert G Newcombe. Observer variation in histopathological diagnosis and grading of cervical intraepithelial neoplasia. *British Medical Journal*, 298(6675):707–710, 1989.
- [17] Joseph Jordan, Albert Singer, Howard Jones, and Mahmood Shafi. *The cervix*. John Wiley & Sons, 2009.
- [18] Prendiville W. and Sankaranarayanan R. *Colposcopy and treatment of cervical precancer*. International Agency for Research on Cancer, 2017.
- [19] YM Chan, Peter WH Lee, TY Ng, Hextan YS Ngan, and LC Wong. The use of music to reduce anxiety for patients undergoing colposcopy: a randomized trial. *Gynecologic Oncology*, 91(1):213–217, 2003.
- [20] Edward J Mayeaux Jr, Akiva P Novetsky, David Chelmow, Francisco Garcia, Kim Choma, Angela H Liu, Theognosia Pappasozomenos, Mark H Einstein, L Stewart Massad,

- Nicolas Wentzensen, et al. Asccp colposcopy standards: colposcopy quality improvement recommendations for the united states. *Journal of lower genital tract disease*, 21(4):242, 2017.
- [21] Stephanie Dick. Artificial intelligence. 2019.
- [22] Real Academia Española. Inteligencia. Disponible en: <https://dle.rae.es/inteligenciaLqtyoaQ>.
- [23] Asa B Simmons and Steven G Chappell. Artificial intelligence-definition and practice. *IEEE journal of oceanic engineering*, 13(2):14–42, 1988.
- [24] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.
- [25] K Rajasekaran and P Saravanan. Conceptual methodology on machine learning and types of learning algorithms. *JAC: A JOURNAL OF COMPOSITION THEORY*, 13:233–249, 2020.
- [26] Lopez A Peña Vanhara P Hampl A Havel Amato F, Amato F. Artificial neural networks in medical diagnosis. *Applied Biomedicine*, 2013.
- [27] Gomez E. Márquez-Rivas J. Guerrero-Claro M. Fernández-Lizaranzu I. Relimpio-López M. I. ... Capitán-Morales L. Gómez-González, E. Artificial intelligence in medicine and healthcare: a review and classification of current and near-future applications and their ethical and social impact. 2020. arXiv preprint arXiv:2001.09778.
- [28] Mitchell TM Jordan MI. Machine learning: Trends, perspectives and prospects. *Science*, 2015. 349:255–60.
- [29] Oluwatomisin E Aina, Steve A Adeshina, Adeyinka P Adedigba, and Abiodun M Aibinu. Classification of cervical intraepithelial neoplasia (cin) using fine-tuned convolutional neural networks. *Intelligence-Based Medicine*, 5:100031, 2021.
- [30] Sudhir Sornapudi, Ronald Joe Stanley, William V Stoecker, Haidar Almubarak, Rodney Long, Sameer Antani, George Thoma, Rosemary Zuna, and Shelliane R Frazier. Deep learning nuclei detection in digitized histology images by superpixels. *Journal of pathology informatics*, 9, 2018.
- [31] Yao Yu, Jie Ma, Weidong Zhao, Zhenmin Li, and Shuai Ding. Msci: A multistate dataset for colposcopy image classification of cervical cancer screening. *International Journal of Medical Informatics*, 146:104352, 2021.
- [32] Ajay Shrestha and Ausif Mahmood. Review of deep learning algorithms and architectures. *IEEE Access*, 7, 2019.

- 
- [33] François Chollet et al. Keras. <https://github.com/fchollet/keras>, 2015.
- [34] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.
- [35] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean and Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from [tensorflow.org](http://tensorflow.org).
- [36] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.