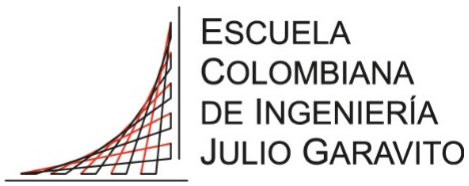

Identificación de Neoplasia Intraepitelial Cervical mediante el uso de Aprendizaje de Máquina

Camilo Antonio Tenjo Castaño

Trabajo Dirigido presentado como requisito para aplicar al título de:
Magister en Ingeniería Biomédica

Tutor:
Ph.D. Oscar Julián Perdomo Charry
Ph.D. Álvaro David Orjuela Cañón

UNIVERSIDAD DEL ROSARIO
UNIVERSIDAD ESCUELA COLOMBIANA DE INGENIERÍA JULIO GARAVITO
PROGRAMA DE MAESTRÍA EN INGENIERÍA BIOMÉDICA
BOGOTÁ, COLOMBIA
2024

No estoy desanimado porque cada intento equivocado descartado es un paso adelante.
- Thomas Edison

Agradecimientos

Infinitas gracias a mis tutores PhD. Oscar Julián Perdomo Charry y PhD Álvaro David Orjuela Cañón, por su paciencia, apoyo y la oportunidad de adentarme mucho más en el campo de la inteligencia artificial mediante este trabajo.
Agradezco a mis padres, mis tías y mi abuela QEPD 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, por mostrarme que el conocimiento existe para cultivarlo y usarlo en pro de la humanidad.
Aquellas personas que conocieron este proceso y de una u otra forma se involucraron, les agradezco, el éxito de este trabajo también es de todos ellos.

Resumen

Los diagnósticos incorrectos de Neoplasia Intraepitelial Cervical (NIC), impactan directamente en el aumento de la tasa de mortalidad por cáncer cervical. Específicamente, América Latina ha estado entre las regiones con mayores tasas de incidencia y mortalidad en los últimos años. Actualmente existen investigaciones que se enfocan en su prevención teniendo como objetivo el diagnóstico temprano y seguimiento de su lesión predecesora, la Neoplasia Intraepitelial Cervical, también llamada Displasia Cervical. Por tanto, las metodologías basadas en visión computacional y aprendizaje de máquina son vitales, para el desarrollo de herramientas de asistencia diagnóstica temprana para el apoyo de especialistas. El objetivo de esta propuesta de trabajo de grado de maestría es la aplicación de arquitecturas de Aprendizaje Profundo y Transformadores de Visión para clasificar los grados de avance de la Neoplasia Intraepitelial Cervical usando imágenes de colposcopía obtenidas de la base de datos libre generada para el reto *Intel & Mobile ODT Cervical Cancer Screening*.

Palabras clave: Aprendizaje automático, Aprendizaje profundo, Clasificación, Colposcopía, Displasia Cervical, Neoplasia Intraepitelial Cervical, Transformadores de Visión.

Tabla de Contenidos

Índice de Figuras

Índice de Tablas

1 Introducción

1.1 Justificación

En la actualidad, el cáncer de cérvix uterino (También conocido como cáncer cervical o de cuello uterino), es uno de los tipos de cáncer que mayor impacto tienen a nivel mundial. Es considerado el tercero con mayor tasa de incidencia y el segundo en mortalidad por complicaciones asociadas [10]. En 2020, la investigación llevada a cabo por Arbyn M. y otros llamada *"Estimates of incidence and mortality of cervical cancer in 2018: a worldwide analysis"* demostró que a lo largo del 2018 surgieron al menos 570000 casos nuevos, donde se dieron más de 311000 fallecimientos por este tipo de cáncer y complicaciones asociadas al mismo [11]. Entre sus diversas acciones, la Organización Mundial de la Salud (OMS), recopila y valida la mayor cantidad de datos de la mayoría de países a lo largo del tiempo. Esto mediante la división *Observatorio Global del Cáncer* (Globocan) [12]. En otras investigaciones, se ha estimado que al menos el 80% de los nuevos casos detectados corresponden a países en vías de desarrollo [13]. A partir de esto es posible evidenciar el impacto de la enfermedad en diversas regiones del mundo. En 2020, América Latina y la Región Caribe fueron ranqueadas en el segundo lugar en incidencia y mortalidad, superando los 59000 casos nuevos y 31000 fallecimientos, esto reportado por Globocan *"Día Mundial del Cáncer de Cérvix"* [14].

En Colombia, se evidencia una alta escasez de información, debido a que en las bases de datos de Globocan, los datos de incidencia encontrados son del año 2012 y los de mortalidad de 2017. Considerando los datos más recientes, se constató que el número total de fallecimientos superó los 1.200 casos en 2013, llegando casi a 1.400 en 2017, esto referente a pacientes mayores de 20 años [15]. Por otro lado, Osorio-Castaño J. y otros, realizaron un estudio tomando datos de pacientes en la ciudad de Medellín, en el que se reportó que en 2018, el 33% de las participantes no se realizaron seguimientos posteriores a una prueba de citología alterada (Presencia de células anormales en el epitelio del cérvix uterino) [13, 16]. Esto evidencia la necesidad de implementar políticas de mayor calidad para el diagnóstico y seguimiento para aquellas pacientes que presenten tanto cáncer cervical, como la Neoplasia Intraepitelial Cervical.

Algunas de las políticas de salud pública enmarcadas en los Objetivos de Desarrollo Sostenible para reducir el impacto del NIC en la población Colombiana son [17]:

- Vacunación contra el VPH (Niñas entre 9 a 14 años y adultas menores de 26 años).
- Tamizaje y detección oportuna de lesiones precancerosas.
- Tratamiento de lesiones precancerosas o cáncer cervicouterino de manera oportuna.

También existen otras políticas con un enfoque social, en términos de educación y concientización de la salud sexual y reproductiva, equidad de prevención y control (De VPH, NIC y Cáncer, en todas las poblaciones), así como replanteamiento e integración de estrategias [17]. Sin embargo, muchas de estás metodologías no han tenido un gran impacto, muchas veces por barreras sociales y de aplicación de éstas políticas.

1.2 Estado de arte

1.2.1 Neoplasia Intraepitelial Cervical

La Neoplasia Intraepitelial Cervical (NIC), también conocida como Displasia Cervical, es la enfermedad predecesora del cáncer cervical [18]. Esta lesión afecta el sistema reproductor femenino, específicamente tejido del cérvix uterino; sin embargo, en la mayoría de los casos, está lesión es asintomática en estadios tempranos [18, 19]. En 2014 la OMS construyó el *"Manual de Imagenología de Cáncer Cervical"*, en el que se declararon las causas y comportamiento de dicha enfermedad [18]. Generalmente, el NIC es causado principalmente por el Virus del Papiloma Humano (VPH), sin embargo, otras causas se asocian a: Factores hormonales, consumo de tabaco y alcohol, entre otras [20].

Figura 1-1: Anatomía del cérvix, disponible en [1]

Entidades como *The American Cancer Society* recomiendan realizar pruebas de VPH a mujeres con edades alrededor de los 25 años (Tendiendo en cuenta factores como inicio de la vida sexual, entre otros) [21]. Esta prueba se lleva a cabo como complemento al procedimiento diagnóstico denominado citología, el cual se utiliza para analizar el tejido

vaginal y del cuello uterino, principalmente con el propósito de evaluar la presencia de lesiones precancerosas. [22]. En caso de que el diagnóstico sea *Lesión escamosa intraepitelial de alto grado*, es necesario realizar una colposcopía. Mediante este procedimiento diagnóstico no invasivo es posible analizar la anatomía del cérvix uterino (Ver figura **1-1**), a través del canal vaginal. A partir de esto, los especialistas que realizan dicho procediento analizan dos áreas de estás estructuras anatómicas [1]:

- La **Unión Escamosa-Columnar (SCJ)**, para el análisis del epitelio columnar y el epitelio escamoso, también conocidos como endocérvix y exocérvix, respectivamente.
- La **Zona de Transformación** en la que colindan la nueva SCJ y antigua SCJ.

Los procedimientos diagnósticos como la colposcopía se fundamentan en la imagenología médica, este tipo de procedimientos buscan la detección temprana y el avance de lesiones precancerosas [23]. A partir de esto y de una primera clasificación (Lesión Escamosa Intraepitelial de Bajo y Alto Grado), es posible especificar el grado de avance de la lesión (Grados I, II y III, ver figura **1-2**) [1, 24].

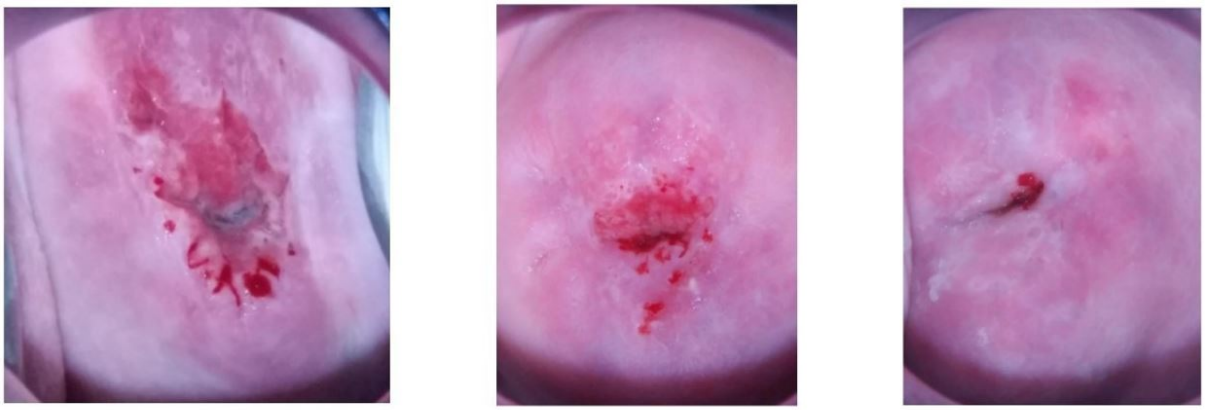

Figura 1-2: Tipos de NIC en imágenes de colposcopía, disponible en [1]

El NIC I corresponde a lesión de bajo grado, los grados II y III corresponden a lesión de alto grado; las lesiones de alto se consideran precursoras de cáncer invasivo [24]. A partir de esto, se dice que lesiones de grado I se observan para evaluar el desarrollo; mientras que los grados II y III son tratados [25]. Este tratamiento inicia con la evaluación de la presencia de tejido canceroso mediante biopsia [18, 26]. Esta clasificación es de suma importancia, puesto que define el tratamiento a seguir y las decisiones que se deberán tomar para minimizar los riesgos de la enfermedad.

1.2.2 Colposcopía

Este procedimiento diagnóstico se define como *"Un examen de baja potencia, que ilumina el tracto epitelial del bajo genital"* [18]. Dicho procedimiento implica el uso de una cámara especializada llamada colposcopio y un espéculo que expande la cavidad vaginal para obtener imágenes [27]. La principal ventaja de este tipo de procedimientos es que al ser basado en imágenes, es mínimamente invasivo. Sin embargo, pese al tener un nivel de eficacia mayor en comparación a otros procedimientos tradicionales como la citología o la inspección visual directa con ácido acético o solución de Yodo de Lugol, aún existen barreras en términos del seguimiento que se les da a las pacientes que padecen la lesión [28]. La experiencia del especialista en colposcopia es un factor clave para la calidad de las imágenes obtenidas y, por ende, para la precisión del diagnóstico, por está razón La Sociedad Americana para la Colposcopia y Patología Cervical (ASCCP) desarrolló 11 indicadores de calidad para el desarrollo de dicho procedimiento diagnóstico [29]. Algunos de los indicadores de mayor importancia son: La obtención de imágenes de alta calidad, identificación de Lesiones precancerosas y cancerosas, reducción de falsos positivos, entre otros.

1.2.3 Inteligencia Artifical

Una de las primeras definiciones de Inteligencia Artificial (IA) describió está disciplina como *"la ciencia e ingeniería que desarrolla máquinas inteligentes"* [30]. Posteriores avances en hardware permitieron que nuevos sitemas con mayor poder de cómputo pudieran correr los modelos desarrollados años antes. Formalmente la Real Academia de la Lengua define la IA como *"Una disciplina científica que se ocupa de crear programás informáticos que ejecutan operaciones comparables a las que realiza la mente humana, como el aprendizaje o el razonamiento lógico"* [31]. Finalmente, una de las definiciones más específicas se encuentra según la IEEE, está la define como *"Máquinas capaces de realizar tareas que requieren inteligencia humana"* [32]. En los últimos años, está tecnología tiene aplicaciones para toma de decisiones en industrias, procesamiento de textos, análisis del entorno, etc [33].

En el campo de la salud, algunas ramás de la inteligencia artificial han tenido un impacto más significativo. Entre las disciplinas más usadas se encuentran: Aprendizaje de Máquina, Aprendizaje Profundo, Procesamiento del Lenguaje Natural y la Visión por Computadora [34]. El desarrollo de aplicativos basados en estás tecnologías ha permitido mejorar la calidad y eficiencia de procedimientos médicos, administrativos y burocráticos en diversas áreas de la salud. Estos aplicativos presentan una nueva ola de herramientas de asistencia en áreas clínicas, diagnósticas, de rehabilitación, entre otras [35]. El alto rendimiento de estás herramientas se debe gracias su alta capacidad de procesamiento y análisis de grandes volúmenes de datos de diversas fuentes, para detección de enfermedades y asistir en decisiones clínicas [36]. Dichos aplicativos son conocidos como Herramientas de Asistencia Diagnóstica.

La IA a lo largo del tiempo ha evolucionado, gracias a los constantes avances e innovaciones de Hardware que permiten el entrenamiento de nuevas arquitecturas y algoritmos. La IA inició con el diseño de Redes Neuronales Artificiales, fundamentadas en emular el funcionamiento del cerebro con neuronas artificiales que procesaran información en capas [37]. El Aprendizaje de Máquina (*Machine Learning*), mezcla estos conceptos, con otros estádísticos y de algorítmica para el reconocimiento de patrones sobre datos. El Aprendizaje Profundo (*Deep Learning*), tomó dichas metodologías y las combinó con el diseño de capas de Redes Neuronales Convolucionales adicional a la interconección de una gran cantidad de estás para aprender de datos sin la necesidad de ser preprocesados como en el Aprendizaje de Máquina. Finalmente, los Transformadores de Visión (*Vision Transformers* ViT), combinaron las redes del Aprendizaje Profundo con conceptos de Procesamiento del Lenguaje Natural para generar modelos que prestan atención de forma localizada a ciertos patrones de información [2, 34, 38]. En la figura **1-3** se evidencia la evolución desde IA hasta Transformadores de Visión, esto ha permitido desarrollos que suplen las necesidades del análisis de imágenes médicas (Ver tabla **1-1**) [2].

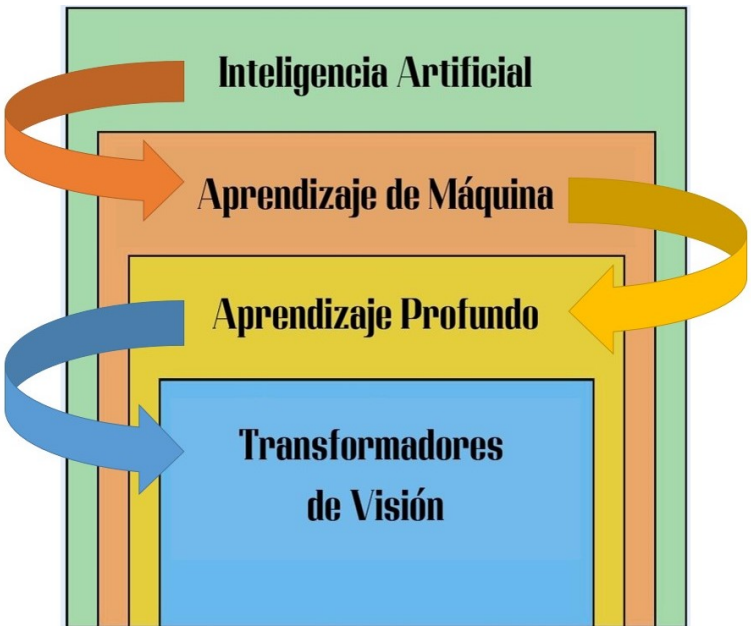

Figura 1-3: Evolución Inteligencia Artificial, Aprendizaje de Máquina, Aprendizaje profundo y Transformadores de Visión, Disponible en [2]

Pese a tener mayor disponibilidad de datos en los últimos años, uno de los problemás más grandes es el desbalance de las bases de datos (Categorías subrepresentadas o con menor cantidad de datos) [39]. El conjunto de metodologías de mayor uso para atacar este problema es el Desaprendizaje de Máquina (*Machine Unlearning*), formalmente se basa en *"Remover datos de influencia reentrenando modelos sin estos"* [40]. Un ejemplo es el entrenamiento

Adquisición	Preprocesamiento
Segmentación	Clasificación
Registro	Extracción de características
Visualización	Localización

Tabla 1-1: Procedimientos del Análisis de Imágenes Médicas, Disponible en [2]

SISA (*Shared, Isolated, Sliced and Aggregated training*), está metodología consiste en entrenar modelos con secciones aisladas de la base de datos, estos modelos serán reentrenados posteriormente [41].

1.3 Trabajos relacionados

En los últimos años han surgido diversos estudios para el desarrollo de herramientas de asistencia diagnóstica basados en aprendizaje de máquina. Estos se basan en el mismo análisis que realiza la mayoría de especialistas para clasificar NIC por medio de colposcopía [42]:

1. Detección del cérvix uterino.
2. Extracción de características como la zona de transformación y análisis de vasos sanguíneos.
3. Interpretración de dichas características y emisión de diangóstico (Grado de avance de la lesión).

Algunos de ejemplos basados aprendizaje profundo y transformadores de visión son descritos a continuación. Todos estos bajo la premisa de la velocidad y eficacia de estos modelos a la hora de generar una respuesta.

Yao Yu, y otros [43], generaron un modelo de clasificación aplicando una arquitectura combinada CNN-GRU, el cual mostró un mejor rendimiento en comparación con los modelos GRU y CNN por separado. Este rendimiento se demostró mediante métricas de exactitud (96.87%), sensibilidad (95.68%) y especificidad (98.72%). Sin embargo, para descartar la presencia de cáncer, es necesario analizar el tejido extraído mediante una biopsia. En este caso, se identificó una falencia en la interpretación de los resultados.

Payette J, y otros [44], implementaron un modelo de clasificación basado en Redes Neuronales Convolucionales (CNN) residuales de 32 capas, utilizando transfer learning y comparando entre otras arquitecturas como InceptionV3. La metodología empleada incluyó la modificación de la arquitectura y el aumento de datos para mejorar el aprendizaje, lo cual resultó exitoso. Esto se evidenció al analizar métricas de exactitud (87.3%) y F1-Score (55.1%). Se

demostró una respuesta favorable a esta tarea por parte de redes como ResNet e InceptionV3.

Bravo, M, y otros [45], desarrollaron dos modelos de clasificación basados en transfer learning utilizando las arquitecturas VGG16 y VGG19. Además de aplicar técnicas de aumento de datos y segmentación semiautomática del cérvix. Posteriormente, se modificó la clasificación de multiclase (Grados I, II o III) a clasificación binaria (Lesión intraepitelial de bajo y alto grado) para abordar el desbalance de la base de datos (Intel & MobileODT Cervical Cancer Screening). En este caso, se evaluaron métricas de precisión en ambas redes: VGG16 obtuvo un 97.36% y VGG19 un 97.1%. Aunque modificar el tipo de clasificación de multiclase a binario mejoró el rendimiento, se generaron problemas de interpretación. A partir del Grado II, se sospecha de cáncer, lo que puede dificultar la interpretación de los resultados.

Darwish, M y otros [46], implementaron un modelo basado en Transformadores de Visión utilizando la metodología de tokenización por cambio de parches. Se realizó un análisis de métricas de exactitud global (91.02%), precisión (91%) y F1-Score (94%) sobre el problema de clasificación multiclase (Tres grados de avance de NIC).

1.3.1 Base de datos: Intel & MobileODT Cervical Cancer Screening

Basados en diversos desarrollos y en la necesidad de datos confiables, surgieron plataformas e iniciativas que se dedican a la recopilación y liberación de datos para propósitos investigativos. Este es el caso de la plataforma Kaggle, la cual en conjunto con Intel y Mobile ODT [1], generaron una competencia conocida como *Cervical Cancer Screening*, en la cual se recopilaron cientos de imágenes de cérvix uterino tomadas mediante colposcopía [3]. Según la descripción del reto *"Intel se alió con MobileODT para retar a los Kagglers a desarrollar un algoritmo para identificar con precisión el tipo de cérvix en imágenes"* [3]; fundamentado en la efectividad que permite la detección temprana de lesiones precancerosas ante el riesgo latente de cáncer de cuello uterino.

2 Objetivos

2.1 General

Implementar algoritmos de aprendizaje profundo y transformadores de visión para clasificar los diferentes grados de avance de la Neoplasia Intraepitelial Cervical (NIC) a partir del análisis de imágenes de colposcopía.

2.2 Específicos

1. Implemetar diferentes modelos de aprendizaje profundo y transformadores de visión para clasificar los grados de NIC.
2. Identificar los modelos de clasificación más adecuados basados en sus métricas de desempeño.
3. Evaluar el rendimiento de los mejores clasificadores e incrementar su rendimiento mediante el uso de técnicas de ajuste fino.

3 Metodología

El desarrollo de esta propuesta contempló diferentes fases para un estudio retroespectivo. Por un lado, la base de datos con imágenes de colposcopía se encuentra con libre acceso y disponbile en la sección *Dataset Description* como parte del reto *Intel & MobileOTD: Cervical Cancer Screening*, en la plataforma Kaggle [3, 47, 48]. Esta base de datos surgió en 2017 para fomentar el desarrollo de algoritmos que detectaran el tipo de lesión de cérvix en imágenes. Es importante mencionar, que mediante citología es posible identificar NIC de grado I, sin embargo por si solo no permite clasificar entre los grados II y III [49].

Dependiendo de la localización de la lesión en la zona de transformación (Ver figura **3-4**), es posible generar 3 tipos de clasificación en dichas imágenes obtenidas al realizar el procedimiento de colposcopía [48].

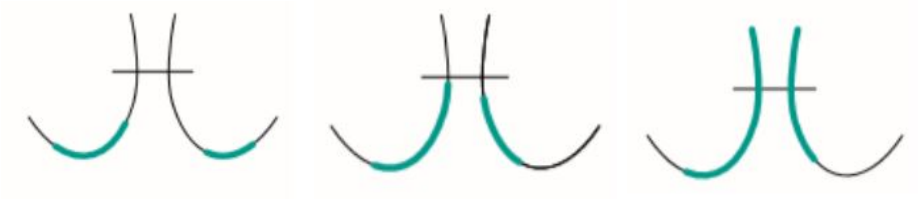

Figura 3-1: I **Figura 3-2**: II **Figura 3-3**: III

Figura 3-4: Grados de NIC en cérvix uterino, disponible en [3]

La base de datos cuenta con más de 1000 imágenes de colposcopía previamente etiquetadas según su clasificación entre los diferentes grados de avance de la enfermedad (I, II o III) [3]. Además, se establecieron tres fases que enmarcaron procedimientos de experimentación, los cuales abordaron cada uno de los objetivos específicos planteados, como se detalla a continuación.

3.1 Fase 1: Pre-procesamiento de datos e implementación de modelos de clasificación basados en aprendizaje profundo y transformadores de visión

Inicialmente se identificaron los datos/imágenes disponibles en la base de datos. Al realizar una inspección preliminar, se identificó que la base de datos fue dividida previamente por los investigadores que recopilaron los datos para el reto *Cervical Cancer Screening* [3]. Dicha división en una sección de entrenamiento y prueba.

La sección de entrenamiento a su vez en dos secciones (Original y adicional). La sección adicional constó de duplicados e imágenes de baja calidad de la sección original, por esta razón esta sección fue descartada para favorecer la calidad del entrenamiento de los modelos. La sección original estaba compuesta por 1480 imágenes. Estas fueron clasificadas desde el grado I al III por los desarrolladores del reto *Intel & MobileOTD: Cervical Cancer Screening*, de las cuales 249 corresponden al grado I, 781 al grado II y 450 al grado III [48].

El conjunto de datos de prueba contó con 512 imágenes definidas con sus etiquetas correspondientes también asignadas desde el diseño del reto [48].

Previo al entrenamiento de las redes se realizó la estandarización de muestras en tamaño a 224x224, así como valores de pixeles entre 0 y 1; esto dado que estas redes fueron configuradas como modelos de clasificación con los pesos del proyecto ImageNet, este estándar requiere que los datos de entrada tengan dichas dimensiones para favorecer el entrenamiento de los modelos [50].

Se exploró el entrenamiento de diferentes arquitecturas de redes neuronales convolucionales, para clasificar entre los grados I, II y III de la lesión. Todo esto, partiendo de dos arquitecturas reportadas en la literatura: InceptionResNetV2 y MobileNet [51].

InceptionResNetV2: Esta CNN es una combinación de la red Inception, adicionando conceptos de redes residuales (ResNet). Actualmente esta ha presentado alta eficacia en tareas de extracción de características de diferentes escalas, óptima en términos de recursos computacionales y de alta adaptabilidad [52]. Esta arquitectura consta de 5 fases o bloques (Ver figura **3-5**), en la cual, las fases más cercanas a la entrada extraen características de bajo nivel, mientras que las finales extraen características de alto nivel [4, 52].

MobileNet: Originalmente diseñada para uso en dispositivos móviles gracias al uso de bloques residuales invertidos que mejoran la eficiencia computacional, en conjunto a capas convolucionales profundas y estrechas (ver figura **3-6**) [53].

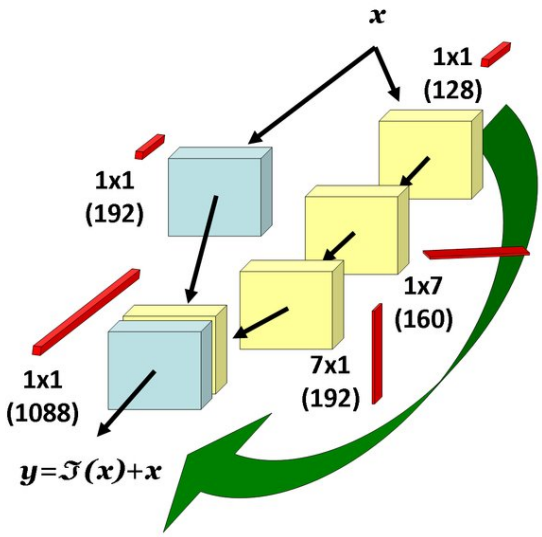

Figura 3-5: Arquitectura InceptionResNetV2, disponible en [4]

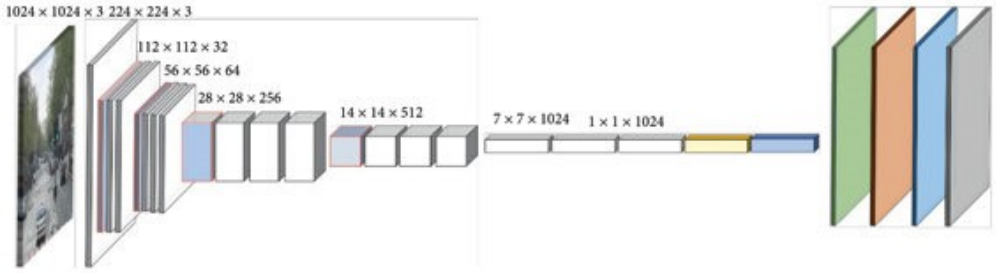

Figura 3-6: Arquitectura MobileNet, disponible en [5]

3.2 Fase 2: Identificación de mejores de modelos de clasificación de imágenes con NIC

3.2.1 Aprendizaje profundo

Teniendo en cuenta revisiones de literatura, en los últimos años el desarrollo de modelos basados en aprendizaje profundo ha permitido obtener buenos resultados en términos de clasificación del grado de avance de la NIC.

Las redes MobileNet e InceptionResNetV2 fueron modificadas; Inicialmente congelando sus pesos para evitar sobreajuste de los datos y posteriormente agregando cuatro capas adicionales a la salida del modelo: *Global Average Pooling 2D, Batch Normalization, Dropout* a una tasa de 0.2 y *Dense* con función de activación *softmax*. Los modelos fueron compilados usando un optimizador *Adam* con tasa de aprendizaje de 0.001, pérdida basada en entropía categórica cruzada y exactitud como métrica.

Posterior a esto, se hizo uso de técnicas de aumento de datos, entre estas fueron aplicadas: Rotaciones, zoom, cambio de posición horizontal y vertical, giros horizontales y verticales, cambios de tamaño horizontal y vertical, entre otras. Aplicadas aleatoriamente a las imágenes para favorecer el rendimiento del entrenamiento de las redes.

El entrenamiento de los modelos se realizó en tres pasos o subentrenamientos. En el primer entrenamiento se implementaron dos monitores:

1. **Detención temprana (*Early Stopping*):** Evitar el sobreajuste, detendiendo el entrenamiento en caso de que pasadas 10 épocas, la exactitud de validación no mejorara.
2. **Punto de control del modelo (*Model Checkpoint callback*):** Guardar únicamente los mejores pesos del modelo de acuerdo al mínimo valor de pérdida en validación.

El primer entrenamiento se realizó con un tamaño de lote de 128 en 10 épocas, utilizando los datos de prueba para validación. Antes de iniciar el entrenamiento los pesos de las capas de los diferentes modelos fueron congeladas, esto con la finalidad de evitar que los pesos en una capa determinada se actualice durante el entrenamiento [54].

El segundo entrenamiento tuvo casi la misma configuración que el primero. Sin embargo para este caso, el Monitor de Detención temprana se ajustó a 14 épocas de evaluación. Este entrenamiento se realizó en 20 épocas con un tamaño de lote de 32. Además, a partir de este, los pesos de los modelos fueron descongelados.

Este proceso fue replicado aplicando metodología entrenamiento SISA para Desaprendizaje de Máquina. Fueron tomadas secciones de 50, 156 y 90 imágenes por clase (Grado I, II y III respectivamente). Como resultado, se obtuvieron 5 subsecciones de la base de datos original para dicha aproximación.

3.2.2 Transformadores de visión

Según revisiones de literatura, los modelos basados en transformadores de visión (ViT) tienen una capacidad mayor que los modelos basados en aprendizaje profundo para tareas de visión computacional [7]. Adicional a que presentan una eficiencia computacional superior.

Para estos modelos de clasificación el proceso de aumento de datos se construyó a partir de: Giros horizontales aleatorios, Rotaciones aleatorias y Zoom. Posterior a esto se implementaron dos arquitecturas generales para clasificación basadas en ViT.

El primero entrenado con ***Vanilla Vit***; En este tipo de tokenización se dividen las imágenes en forma de cuadrícula (ver figura **3-7**), a partir de esta división se generan los parches que serán codificados para calcular la importancia de cada uno durante el entrenamiento del modelo [55].

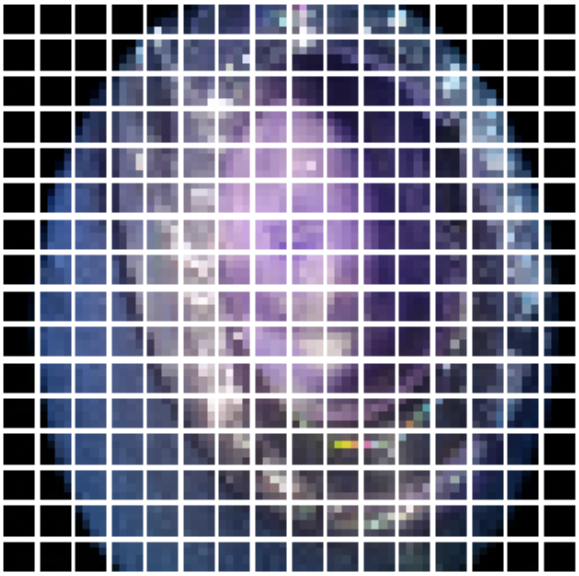

Figura 3-7: Original

Para la segunda arquitectura, se implementó la **Tokenización por Cambio de Parches (*Shifted Patch Tokenization - SPT)*** en la que se extraen parches de las imágenes modificadas que posteriormente se proyectan linealmente como tokens (Ver figura **3-8**). En la figura **3-13** se observa una muestra del resultado de este procedimiento.

Las imágenes resultantes fueron posteriormente codificadas para implementar **Atención Automática Local (*Locality Self Attention - LSA)***. Esto para calcular la relación de similitud entre los tokens de una misma región [7].

Para generar la atención se tienen en cuenta 3 entradas: Consulta (*Query*), llave (*Key*) y valor (*Value*). La consulta corresponde la característica en la que se centra la atención (ver figura **3-14**). La llave representa las posibles características relacionadas con la consulta. Finalmente, el valor representa la importancia de la relación entre la consulta y la llave [7, 56]. Este modelo se representa mediante la ecuación 3-1. En este caso se hace uso de la función ***softmax*** para generar las relaciones entre los mismos tokens (Intra-token). En este caso se enmascaró la diagonal del producto punto entre consulta y clave, forzando a que el modelo priorice la atención entre tokens (Inter-tokens). Finalmente se hizo uso de un módulo ***Multi Head Attention*** para computar dichos cálculos recibiendo las entradas

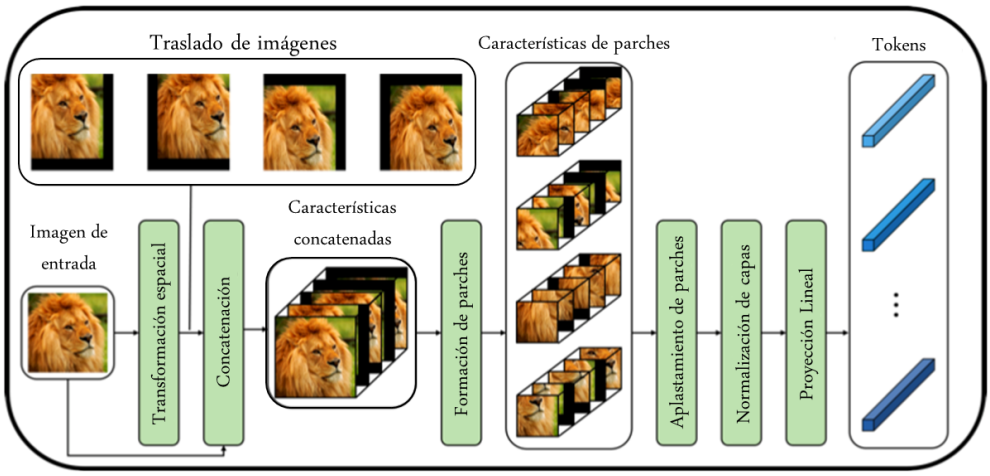

Figura 3-8: Tokenización por cambio de parches, disponible en [6]

(Consulta, llave y valor), en paralelo.

$$\text{Atención}(Q, K, V) = \text{softmax}\left(-\frac{QK^T}{\sqrt{d_k}}\right)V \tag{3-1}$$

Para el entrenamiento de ambos modelos, como en los modelos de aprendizaje profundo, se implementó un Punto de Control del modelo, para guardar los mejores pesos de acuerdo al mínimo valor de pérdida en validación.

Los modelos entrenados fueron sometidos al segmento del dataset destinado a pruebas, con el fin de identificar el rendimiento de cada red de acuerdo a métricas estadísticas. Con esto, se logró identificar el mejor modelo/arquitectura. Dado que se generó una clasificación multiclase, fueron usadas métricas como: ***Precisión*** (pre), ***Recall*** (rec) y ***F1-score*** (Fs) [57]. La precisión para verificar la relación entre verdaderos positivos (vp) y falsos positivos (fp): $\frac{vp}{vp+fp}$. El Recall para verificar la relación entre verdaderos positivos (vp) y falsos negativos (fv): $\frac{vp}{vp+fn}$. Finalmente, la media armónica o F1-score, que relaciona los valores de precisión y recall: $\frac{2(pre)(rec)}{pre+rec}$. Estas métricas fueron calculadas con configuración micro, dada la naturaleza imbalanceada de la base de datos [58].

Una representación adicional del rendimiento de los modelos es la Matriz de Confusión (ver figura **3-15**), en este caso se evidencian la concordancia entre las etiquetas obtenidas con la predicción de los modelos y las etiquetas reales de los datos de prueba [59]. Con esto, fue posible calcular el índice Kappa (ver tabla **3-1**), el cual es una medida de evaluación

Figura 3-9: Izquierda y Arriba

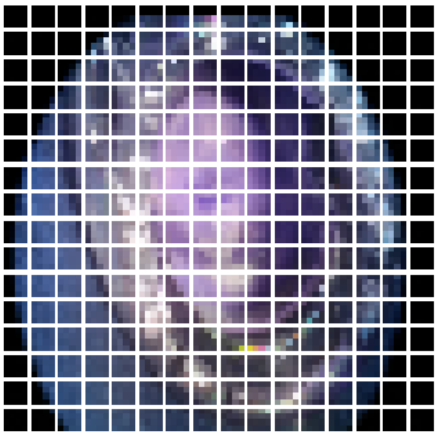

Figura 3-10: Izquierda y Abajo

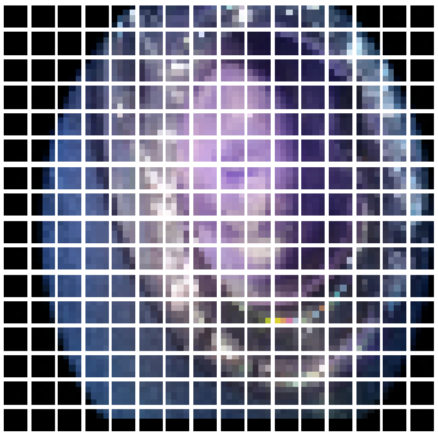

Figura 3-11: Derecha y Arriba

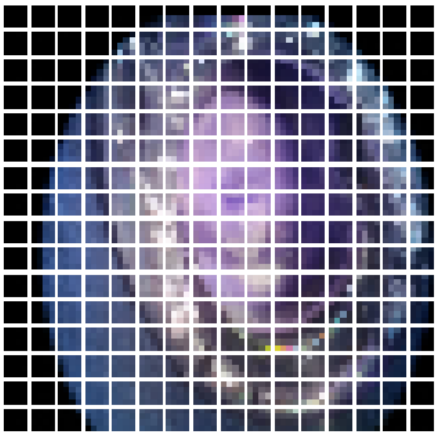

Figura 3-12: Derecha y Abajo

Figura 3-13: Resultado proceso de aumento de datos y tokenización por cambio de parches

intermedia para medir la fiabilidad del modelo [9]. Finalmente, se calculó la exactitud global; métrica que muestra la probabilidad de realizar una predicción individual correcta [60].

3.3 Fase 3: Evaluación y ajuste de mejores modelos obtenidos

Los mejores modelos obtenidos en la fase previa fueron ajustados mediante diferentes técnicas para incementar su desempeño en la tarea de clasificación de los tres grados de NIC. Inicialmente, se implementaron técnicas de aumento de datos sintéticos con diferentes transformaciones aplicadas al conjunto de entrenamiento. Por otro lado, se realizó la exploración aleatoria, sistemática y combinada de los hiperparámetros de la tasa de aprendizaje y del

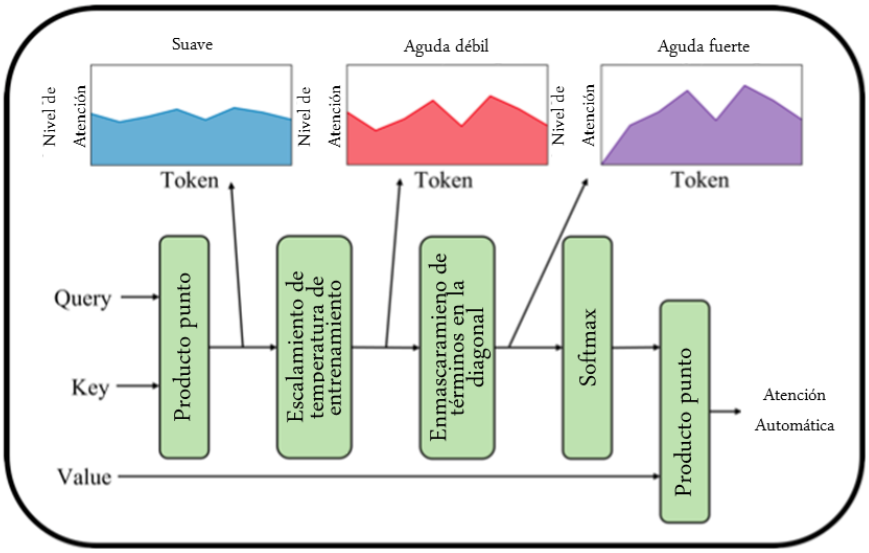

Figura 3-14: Modelo de Atención Automática Local, Disponible en [7]

		Predicción	
		Positivos	**Negativos**
Observación	**Positivos**	Verdaderos Positivos (VP)	Falsos Negativos (FN)
	Negativos	Falsos Positivos (FP)	Verdaderos Negativos (VN)

Figura 3-15: Visualización de una Matriz de Confusión, Disponible en [8]

tamaño del lote durante el entrenamiento para incrementar el desempeño de los mejores modelos de aprendizaje profundo. En el caso de los mejores modelos de transformadores de visión fueron modificados los hiperparámetros de tasa de aprendizaje, tamaño de parches y cantidad de épocas de aprendizaje.

Específicamente se implementó un tercer entrenamiento de los modelos de aprendizaje profundo. Configurado en 30 épocas, adicional a la asignación de pesos específicos para las clases. Esto con la finalidad de darle un mayor grado de prioridad a las clases con pocos datos (Clase I) y reducir el impacto del desbalance de datos en el entrenamiento de los modelos. El proceso de asignación de pesos específicos para las clases fue replicado para los modelos de transformadores de visión.

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 3-1: Interpretación del índice kappa, disponible en [9]

3.4 Actividades por fase

- **Fase 1**:
  1. Inspección visual para la identificación de NIC y limpieza de la base de datos.
  2. Carga de imágenes/datos limpios de entrenamiento y prueba.
  3. Estandarización de dimensiones de las muestras (224x224), normalización de las muestras (valores de 0 a 1) y aplicación de aumento de datos.
- **Fase 2**: Esta será implementada en cada uno de los modelos
  1. Selección de modelos de aprendizaje profundo y transformadores de visión.
  2. Entrenamiento de modelos de clasificación con la subsección de la base de datos destinada a entrenamiento.
  3. Análisis de parámetros de entrenamiento (Precisión, Recall, F1-Score, Matriz de Confusión, Índice Kappa y Exactitud Global).
- **Fase 3**:
  1. Realizar nuevas predicciones de etiquetas sobre la subsección de la base de datos destinada a validación.
  2. Analizar el nuevo rendimiento de los modelos usando las métricas de: Precisión, Recall, F1-Score, Matriz de Confusión, Índice Kappa y Exactitud Global.
  3. Ajuste y evaluación final de los modelos para mejorar su rendimiento.
  4. Selección del mejor modelo.

4 Resultados

En este capítulo son expuestos los resultados obtenidos durante la implementación de la metodología, cada uno de estos asociados a las fases II y III. En el caso de la fase I, no se presentan resultados tangibles.

En términos de la clasificación, fue necesario recategorizar las etiquetas de las imágenes; De esta forma los grados I, II y III, serían asignados como 0, 1 y 2 respectivamente. Esto argumentado por la naturaleza en la que los modelos reciben los datos para entrenamiento y generan posteriormente clasificación de los datos de prueba/validación.

Se guardaron los pesos de los diferentes modelos en formato **.h5**, este formato ofrece una alta eficiencia de almacenamiento y velocidad de carga, compatibilidad y flexibilidad.

4.0.1 Resultados Clasificación Modelos basados en Aprendizaje Profundo

Una vez preprocesadas las imágenes (Ajuste a tamaño 224x224 y estandarización de valores de pixeles de 0 a 1). Se implementaron las siguientes transformaciones aleatorias para el aumento de datos:

- Rotación de 0 a 10 grados.
- Zoom de 0% a 15%.
- Cambio de ancho de 0% a 15%.
- Cambio de altura de 0% a 10%.
- Giros horizontales.

El proceso de entrenamiento de modelos para clasificación basado en Aprendizaje Profundo se implementó bajo los siguientes ajustes:

1. Entrenamiento con la base de datos completa (1480 imágenes correspondientes a 249 de Grado I, 781 de Grado II y 450 de Grado III).

2. Entrenamiento reduciendo a 550 imágenes de Grado II para reducir el impaco del desbalance de datos, resultado en 1249 imágenes de entrenamiento.

3. Entrenamiento aplicando pesos ajustados.

A partir de este punto surgieron 2 clases de modelos; Aquellos que fueron entrenados **únicamente** con el numeral 1 y 3, anteriormente descrito y los que fueron entrenados con los numerales 2 y 3.

La mejor configuración se obtuvo a partir de la modificación del optimizardor en el segundo entrenamiento, específicamente la tasa de aprendizaje, para esta se tomaron valores entre *1e-3* hasta *1e-6*. En este caso se obtuvo el mejor rendimiento al asignar un valor de *1e-5*. Para el caso de la selección pesos para cada clase se experimentaron con las siguientes configuraciones (ver tabla **4-1**):

	Clases		
Configuración	I	II	III
A	1.4	1	1.1
B	1.5	1	1.2
C	1.5	1.05	1.2
D	1.4	1	1.2

Tabla 4-1: Configuraciones de pesos clases

En este caso, la mejor configuración de pesos fue conseguida a partir de la configuración C descrita en la tabla **4-1**. Este primer mejor modelo entrenado con la base de datos completa (Numerales 1 y 3), sus métricas de rendimiento global y matriz de confusión se muestran en la tabla **4-2** y figura **4-1**. Así como sus métricas de rendimiento por clase se muestran en la tabla **4-3**.

Precisión	Recall	F-Score	Exactitud Global	Índice Kappa
64%	67%	65%	66.6%	0.468

Tabla 4-2: Métricas modelo 1

El segundo mejor modelo entrenado con la base de datos segmentada (Numerales 2 y 3), sus métricas de rendimiento y matriz de confusión se muestran en la fila **Original** de la tabla **4-4** y figura **4-4**. A este último se le realizó un reentrenamiento adicional, seleccionando datos de la sección de base de datos entrenada para pruebas como datos de validación durante el entrenamiento, estas métricas corresponden a la fila **Ajustado** de la tabla **4-4**, su matriz

Clase	Precisión (%)	Recall(%)	F-Score (%)
I	66.6	45.6	54.1
II	65.6	73.7	69.4
III	68.8	73.1	70.9

Tabla 4-3: Métricas por clase modelo 1

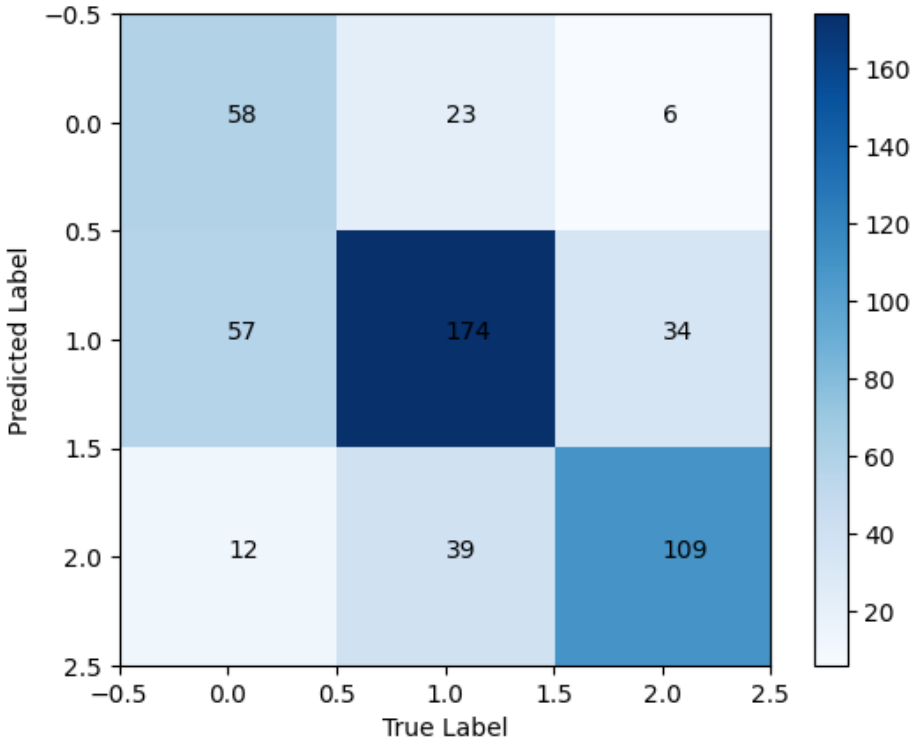

Figura 4-1: Matriz de confusión modelo 1 entrenado con la base de datos completa

de confusión se muestra en la figura **4-4**. Así como sus métricas de rendimiento por clase se muestran en la tabla **4-5**.

Seguido de esto, se implementó la metodología SISA de entrenamiento, con las 5 secciones de la base de datos (Secciones de 50, 156 y 90 muestras de los grado I, II y III respectivamente). Una vez entrenados, se implementó una arquitectura en la que los 5 modelos recibían la misma entrada, posterior a esto se aplicó la moda estadística a la salida de los 5 modelos, esta sería la respuesta del modelo (ver figura **4-5**). Sus métricas de rendimiento y matriz de confusión se muestran en la tabla **4-6** y figura **4-6**. Así como sus métricas de rendimiento por clase se muestran en la tabla **4-7**.

	Precisión	Recall	F-Score	Exactitud Global	Índice Kappa
Original	54%	50%	57.8%	62.3%	0.248
Ajustado	62%	59%	60%	64%	0.39

Tabla 4-4: Métricas modelo 2

Clase	Precisión (%)	Recall(%)	F-Score (%)
Original			
I	25.2	44	32
II	78.1	59.1	67.3
III	41.1	59.8	48.7
Ajustado			
I	45.9	48.2	47
II	73.6	65.2	69.1
III	58.1	71.5	64.1

Tabla 4-5: Métricas por clase modelo 2

4.0.2 Resultados Clasificación Modelos Transformadores de Visión

Para los modelos basados en Transformadores de Visión se preprocesaron las imágenes de la misma forma que para los modelos basados en Aprendizaje Profundo. Sin embargo, el proceso de aumento de datos aleatorio consistió en:

- Giros horizontales.
- Rotaciones con un factor de 0.02.
- Zoom vertical en factor 0.2.
- Zoom horizontal en factor 0.2.

En este caso, el ajuste de entrenamientos se fundamentó en la modificación de la tasa de aprendizaje (1e-2 hasta 1e-6), la cantidad de épocas de entrenamiento (50 hasta 150), el tamaño de parches (4 a 8) y la selección del mejor ajuste de pesos de clases.

Precisión	Recall	F-Score	Exactitud Global	Índice Kappa
77.9%	75.5%	76.7%	62.3%	0.305

Tabla 4-6: Métricas Arquitectura combinada de modelos entrenados con metodología SISA

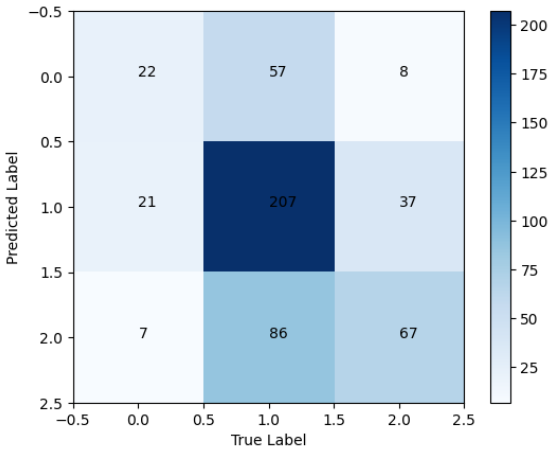

Figura 4-2: Original

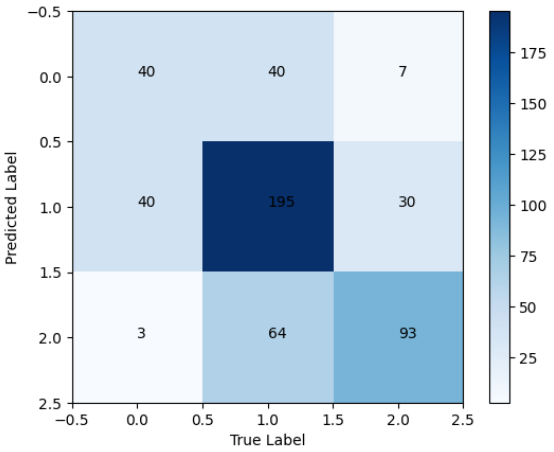

Figura 4-3: Ajustado

Figura 4-4: Matriz de confusión modelo 2

Clase	Precisión (%)	Recall(%)	F-Score (%)
I	13.8	63.1	22.6
II	86.8	61.3	71.6
III	48.1	65.2	55.4

Tabla 4-7: Métricas por clase arquitectura combinada

La primera configuración exitosa constó de un tamaño de parche de 6, tasa de aprendizaje de 1e-2 y un entrenamiento en 50 épocas (ver tabla **4-8**). Sus matrices de confusión se muestran en la figura **4-9**). Sus métricas de rendimiento por clase se muestran en la tabla **4-9**.En esta primera configuración no se fijaron los pesos de las clases durante el entrenamiento de los modelos.

	Precisión	Recall	F-Score	Exactitud Global	Índice Kappa
Vanilla	51%	57%	51%	54.8%	0.259
SPT - LSA	46%	47%	46%	48.3%	0.172

Tabla 4-8: Métricas obtenidas con la configuración 1 ViT

La segunda configuración exitosa constó de un tamaño de parche de 6, tasa de aprendizaje de 1e-2 y un entrenamiento en 100 épocas (ver tabla **4-10**). Sus matrices de confusión se muestran en la figura **4-12**). Así como sus métricas de rendimiento por clase se muestran en la tabla **4-3**. En este caso los valores de pesos por clase fueron dados como 1.4, 1 y 1.2 para las clases I, II y III respectivamente. Sobre esta configuración se aplicó ajuste fino.

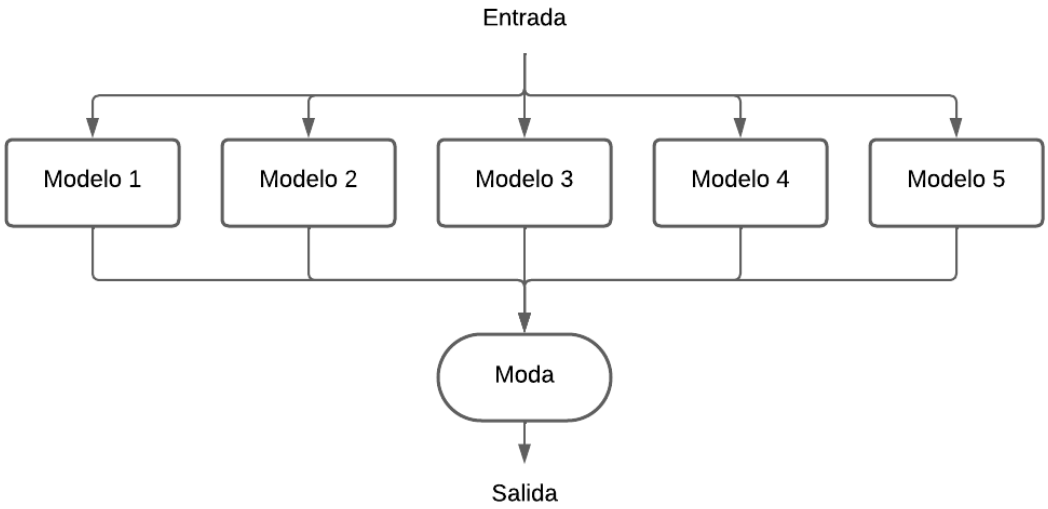

Figura 4-5: Arquitectura combinada de modelos entrenados con metodología SISA

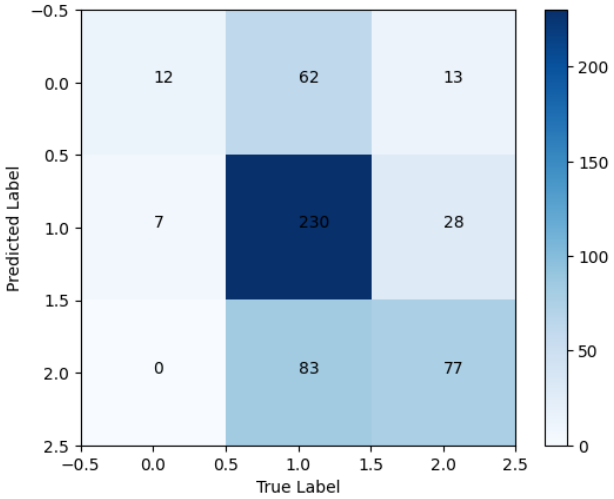

Figura 4-6: Matriz de Confusión Arquitectura combinada de modelos entrenados con metodología SISA

Clase	Precisión (%)	Recall(%)	F-Score (%)
Vanilla			
I	35.6	40.2	37.8
II	56.2	59.6	57.9
III	63.1	54.6	58.5
SPT - LSA			
I	34.4	36.1	35.2
II	46	55.7	50.4
III	59	45.4	51.3

Tabla 4-9: Métricas por clase configuración 1 ViT

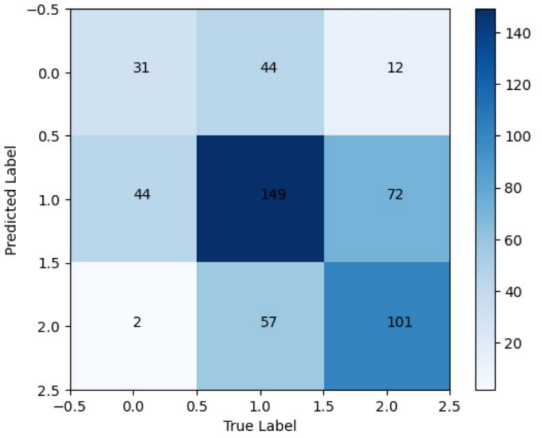

Figura 4-7: Vanilla

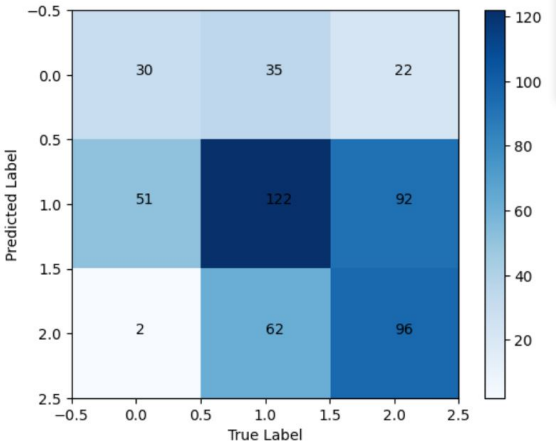

Figura 4-8: SPT - LSA

Figura 4-9: Matriz de confusión configuración 1 ViT

Finalmente, la última configuración exitosa constó de un tamaño de parche de 4, tasa de aprendizaje de 1e-4 y un entrenamiento en 150 épocas (ver tabla **4-12**). Sus matrices de confusión se muestran en la figura **4-15**). Así como sus métricas de rendimiento por clase se muestran en la tabla **4-13**.

	Precisión (%)	Recall	F-Score	Exactitud Global	Índice Kappa
Vanilla	48%	49%	48%	52.3%	0.221
SPT - LSA	47%	49%	48%	52.9%	0.248

Tabla 4-10: Métricas obtenidas con la configuración 2 ViT

Clase	Precisión (%)	Recall(%)	F-Score (%)
Vanilla			
I	31	31.7	31.3
II	53.2	56.6	54.8
III	62.5	56.1	59.1
SPT - LSA			
I	39	36.9	37.9
II	47.1	57.3	51.7
III	60	47.5	53

Tabla 4-11: Métricas por clase configuración 2 ViT

	Precisión	Recall	F-Score	Exactitud Global	Índice Kappa
Vanilla	49%	50.1%	49%	51.8%	0.231
SPT - LSA	49%	51%	50%	52.1%	0.241

Tabla 4-12: Métricas obtenidas con la configuración 3 ViT

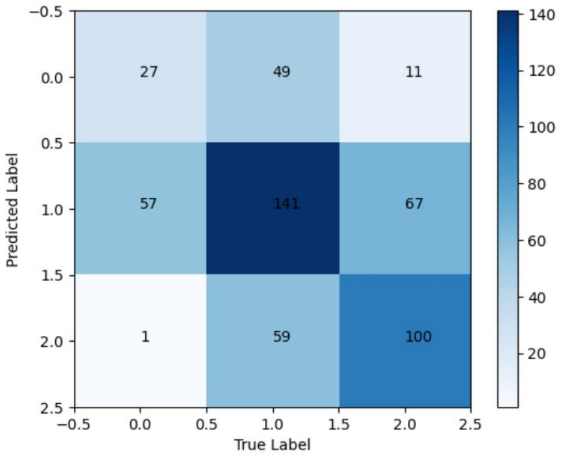

Figura 4-10: Vanilla

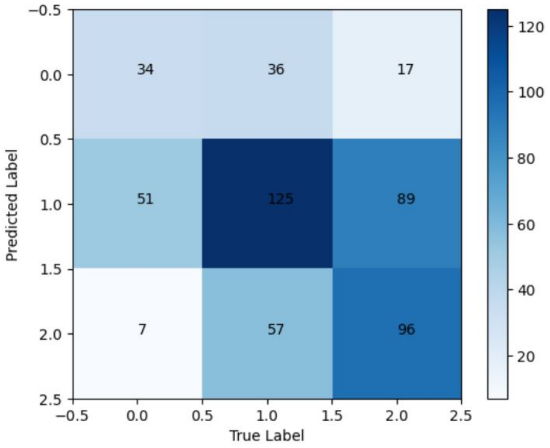

Figura 4-11: SPT - LSA

Figura 4-12: Matriz de confusión configuración 2 ViT

Clase	Precisión (%)	Recall(%)	F-Score (%)
Vanilla			
I	40.2	33.3	36.4
II	50.5	59.2	54.5
III	60	53	56.3
SPT - LSA			
I	43.6	32.4	37.2
II	52	61.6	56.4
III	56.8	53.2	54.9

Tabla 4-13: Métricas por clase configuración 3 ViT

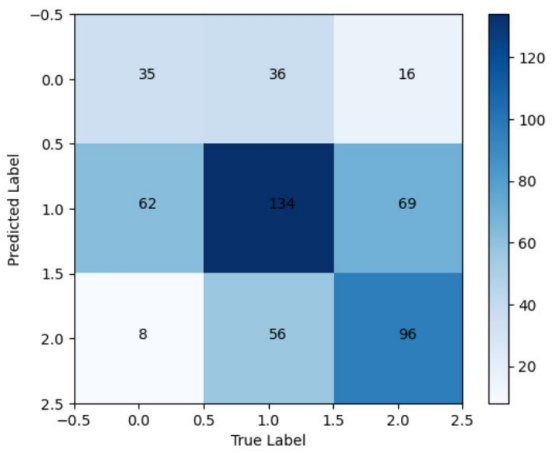

Figura 4-13: Vanilla

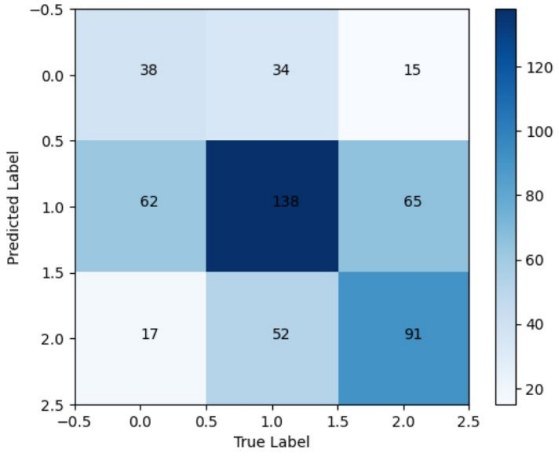

Figura 4-14: SPT - LSA

Figura 4-15: Matriz de confusión obtenida con la configuración 3 ViT

5 Discusión

Durante este proyecto, se comprobó la gran capacidad que tiene el desaprendizaje de máquina para generar modelos con un rendimiento similar al de investigaciones similares. En este caso esta arquitectura presentó el mayor rendimiento según las métricas evaluadas (Precisión, Recall, F-Score y Exactitud global de 77.9%, 75.5%, 76.7%, y 62.3% respectivamente), pese a que el nivel de acuerdos según el índice Kappa es Mínimo (0.305). Vale la pena resaltar que las arquitecturas basadas en MobileNet fueron descartadas, dado que su rendimiento no superó métricas de 25%, esto demuestra la necesidad de modelos robustos para lograr extraer correctamente características en este tipo de imágenes, que por la naturaleza del procedimiento diagnóstico, tienden a ser de baja resolución.

Los modelos basados en Transformadores de Visión tienen una mayor eficiencia de cómputo, en comparación los modelos basados en Aprendizaje profundo tardaron mucho más tanto en entrenamiento, como realizando tareas de clasificación en pruebas. Esta tecnología puede robustecerse con arquitecturas mucho más complejas y la implementación de procesos de aumento de datos más robustos. Una tendencia encontrada fue que las arquitecturas que mejoraban al ser entrenadas con SPT-LSA (Precisión, Recall, F-Score y Exactitud global de 49%, 50.1%, 49% y 51.8% respectivamente) tendían a tener un menor rendimiento con Vanilla (Precisión, Recall, F-Score y Exactitud global de 49%, 51%, 50% y 52.1% respectivamente).

En general, el mayor desafío de los modelos (tanto los basados en aprendizaje profundo como transformadores de visión) existió al momento de clasificar el grado I de la Neoplasia Intraepitelial Cervical. Esto se evidenció al analizar las tablas de rendimiento por clase. La mayor importancia de la detección temprana de esta enfermedad radica en la correcta clasificación de los grados I y II, dado que este es el referente para tratar la lesión como carcinoma; Sin embargo dado el comportamiento de la lesión (Ver figura **3-4**) dificultó en gran medida el diseño de modelos de clasificación.

6 Conclusiones

Se logró la implementación de algoritmos de Aprendizaje Profundo y Transformadores de Visión para clasificación de los grados de avance del NIC (Grado I, II y III). Teniendo mejor rendimiento aquellos modelos basados en Aprendizaje Profundo. El mayor desafío surgió a partir de la predisposición de clasificación de lesiones de grado II y III, producto del desbalance de cantidad de datos.

Tomar como base un modelo con buena relación entre eficiencia computacional y que a su vez fuera robusto (InceptionResNetV2) favoreció el desarrollo de esta propuesta, dado que se logró un alto rendimiento en un tiempo relativamente corto (Tanto entrenamiento como prueba) para la clasifiación de la lesión. El implementar modelos que puedan ser embebidos en aplicativos móbiles (MobileNet), podría requerir metodologías tipo Knowledge Distillation para transferir el aprendizaje y favorecer el rendimiento de estos modelos más livianos. Pese a que es una tecnología relativamente nueva los modelos basados en Transformadores de visión presentaron un buen rendimiento, lamentablemente se corroboró que estos modelos presentan la necesidad de una mayor cantidad de datos en comparación con modelos basados en aprendizaje profundo.

El uso de técnicas de desaprendizaje de máquina (Entrenamiento SISA) favoreció el rendimiento de los modelos de aprendizaje profundo. Sin embargo, se evidenció que la presencia de un desbalance tan alto puede limitar los beneficios de este tipo de metodologías. Se observó que más del 50% de los datos correspondían a la clase II, lo que pudo afectar la capacidad del modelo para generalizar correctamente y sesgar los resultados hacia esa clase mayoritaria.

En comparación con la mayoría de las investigaciones realizadas, este proyecto tuvo en cuenta la naturaleza de los datos para clasificar NIC, considerando que es un error clasificar el cáncer únicamente a partir de imágenes médicas. Para realizar esta clasificación, es necesario llevar a cabo una biopsia de tejido.

Finalmente, vale la pena ajustar para obtener un mejor rendimiento para el grado I, dado que este es el punto en el que se evalúa la necesidad de realizar biopsia del tejido de la paciente. Este bajo rendimiento se produjo principalmente por la baja cantidad de datos disponibles de esta clase (16% del total de la base de datos).

7 Trabajos futuros

El avance del Aprendizaje de Máquina ha permitido el diseño e implementación de modelos que permitan la clasificación a partir de segmentación. Permitiendo el análisis directo del tejido afectado, para esto es necesario generar máscaras para el entrenamiento de modelos como U-Net. Idealmente estas máscaras serían generadas por expertos o la aplicación de técnicas de procesamiento de imágenes. A partir de estos nuevos datos sería posible implementar un análisis de la geometría de la lesión; Puesto que, según literatura médica, estas características son de gran importancia al momento en que un especialista diagnostica a partir de la colposcopía. En conclusión, serían generados modelos que analicen tanto visualmente las imágenes, como datos asociados a estas.

Según estadísticos, en regiones latinoamericanas, es común que esta lesión suceda a una edad más temprana (Al rededor de los 18 a 20 años). Por esta razón, generar una base de datos con muestras de pacientes en estos rangos de edad, permitiría un análisis mucho más contextualizado al comportamiento de esta enfermedad en esta región. Finalmente, la aplicación de redes más livianas como EfficientNet o la DenseNet, mediante transfer learning partiendo de estos primeros modelos generados para la implementación en dispositivos móviles, podrían favorecer a que estos aplicativos lleguen a regiones alejadas y favorecer al mejoramiento de la calidad de vida de las personas (Pacientes y familiares).

Finalmente, continuar con la exploración de modelos basados en Transformadores de visión aplicando tokenizaciones y procesos de entrenamiento más robustos. Esta tecnología demostró su alta eficiencia y robustez en tareas de clasificación. Según literatura, presentan un alto rendimiento en tareas de detección de objetos. Sería posible aplicarlos junto a modelos basados en Yolo (You Only Look Once).

Bibliography

[1] MobileODT Kaggle, Intel. Cervix types classification, 2017. Disponible en: https://www.kaggle.com/c/intel-mobileodt-cervical-cancer-screening/overview/description.

[2] Yudong Zhang, Jiaji Wang, Juan Manuel Gorriz, and Shuihua Wang. Deep learning and vision transformer for medical image analysis, 2023.

[3] Kumar H Meg Risdal MRao Vadim Sherman Vipul Wendy Kan Yau Ben-Or BenO, jljones. Intel & mobileodt cervical cancer screening, 2017. https://kaggle.com/competitions/intel-mobileodt-cervical-cancer-screening.

[4] Jian-Feng Shi, Steve Ulrich, and Stéphane Ruel. Cubesat simulation and detection using monocular camera images and convolutional neural networks. In *2018 AIAA Guidance, Navigation, and Control Conference*, page 1604, 2018.

[5] Kalyani Dhananjay Kadam, Swati Ahirrao, Ketan Kotecha, et al. Efficient approach towards detection and identification of copy move and image splicing forgeries using mask r-cnn with mobilenet v1. *Computational Intelligence and Neuroscience*, 2022, 2022.

[6] Seung Hoon Lee, Seunghyun Lee, and Byung Cheol Song. Vision transformer for small-size datasets. *arXiv preprint arXiv:2112.13492*, 2021.

[7] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[8] Zelada C. Evaluación de modelos de clasificación, 2017. https://rpubs.com/chzelada/275494.

[9] Mary L McHugh. Interrater reliability: the kappa statistic. *Biochemia medica*, 22(3):276–282, 2012.

[10] Aamod Dhoj Shrestha, Dinesh Neupane, Peter Vedsted, and Per Kallestrup. Cervical cancer prevalence, incidence and mortality in low and middle income countries: a systematic review. *Asian Pacific journal of cancer prevention: APJCP*, 19(2):319, 2018.

[11] Marc Arbyn, Elisabete Weiderpass, Laia Bruni, Silvia de Sanjosé, Mona Saraiya, Jacques Ferlay, and Freddie Bray. Estimates of incidence and mortality of cervical cancer in 2018: a worldwide analysis. *The Lancet Global Health*, 8(2):e191–e203, 2020.

[12] World Health Organization: International Agency for Research on Cancer. Global Cancer Observatory. Disponible en: https://gco.iarc.fr.

[13] Alicia Azuaga Martinez, Manuela Undurraga Malinverno, Emily Manin, Patrick Petignat, and Jasmine Abdulcadir. A cross-sectional study on the prevalence of cervical dysplasia among women with female genital mutilation/cutting. *Journal of Lower Genital Tract Disease*, 25(3):210–215, 2021.

[14] Ministerio de Salud y Protección Social & Ministerio de Hacienda y Crédito Público. Cuenta de Alto Costo: Día mundial del cáncer de cérvix 2022. Disponible en: https://cuentadealtocosto.org/site/cancer/dia-mundial-del-cancer-de-cervix-2022/.

[15] International Agency for Research on Cancer. Absolute numbers *Colombia*, incidence and mortality, females, age [20-74].

[16] Jhon H Osorio-Castaño, Marjorie Pérez-Villa, Claudia P Montoya-Zapata, and Fernando A Cardona-Restrepo. Características citológicas previas al diagnóstico de cáncer de cérvix en mujeres de medellín (colombia). *Universidad y Salud*, 22(3):231–237, 2020.

[17] Jair Andrey Ruiz Arias and Daniela María Solano Torres. Análisis de las estrategias de prevención de cáncer de cuello uterino a partir de genotipos de alto riesgo del virus del papiloma humano en mujeres de colombia. 2023.

[18] 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.

[19] Zahra Javanbakht, Mastaneh Kamravamanesh, Roumina Rasulehvandi, Amirhossin Heidary, Mehdi Haydari, and Mohsen Kazeminia. Global prevalence of cervical dysplasia: A systematic review and meta-analysis. *Indian Journal of Gynecologic Oncology*, 21(3):62, 2023.

[20] Anna-Barbara Moscicki, Mark Schiffman, and Silva Franceschi. The natural history of human papillomavirus infection in relation to cervical cancer. In *Human papillomavirus*, pages 149–160. Elsevier, 2020.

[21] Elizabeth TH Fontham, Andrew MD Wolf, Timothy R Church, Ruth Etzioni, Christopher R Flowers, Abbe Herzig, Carmen E Guerra, Kevin C Oeffinger, Ya-Chen Tina Shih, Louise C Walter, et al. Cervical cancer screening for individuals at average risk: 2020 guideline update from the american cancer society. *CA: a cancer journal for clinicians*, 70(5):321–346, 2020.

[22] Paul N Staats, Diane Davis Davey, Benjamin L Witt, Mohiedean Ghofrani, Chengquan Zhao, Leslie G Dodd, Kelly Goodrich, Mujtaba Husain, Daniel FI Kurtycz, Donna K Russell, et al. Performance of specific morphologic features in distinguishing low-grade squamous intraepithelial lesions from high-grade squamous intraepithelial lesions in borderline cases: a college of american pathologists cytopathology committee multiobserver study. *Journal of the American Society of Cytopathology*, 11(2):102–113, 2022.

[23] Johanna Norenhag, Juan Du, Matts Olovsson, Hans Verstraelen, Lars Engstrand, and Nele Brusselaers. The vaginal microbiota, human papillomavirus and cervical dysplasia: a systematic review and network meta-analysis. *BJOG: An International Journal of Obstetrics & Gynaecology*, 127(2):171–180, 2020.

[24] John W Sellors and R Sankaranarayanan. La colposcopia y el tratamiento de la neoplasia intraepitelial cervical: Manual para principiantes. *Lyon, Francia: International Agency for Research on Cancer (IARC)*, 140, 2003.

[25] Ministerio de Salud: Observatorio Nacional del Cáncer. Ruta integral de atención en salud, cáncer de cuello uterino, 2024. https://www.sispro.gov.co/observatorios/oncancer/Paginas/ruta$_i$*ntegral*$_c$*uellouterino.aspx*.

[26] Roxana Elizabeth Zamora-Julca, Jorge Ybaseta-Medina, and Adrián Palomino-Herencia. Relación entre citología, biopsia y colposcopía en cáncer cérvico uterino. *Rev. méd. panacea*, pages 31–45, 2019.

[27] Prendiville W.and Sankaranarayanan R. *Colposcopy and reatment of cervical precancer*. International Agency for Research on Cancer, 2017.

[28] Óscar Gamboa, Mauricio González, Jairo Bonilla, Joaquín Luna, and Raúl Murillo. Visual techniques for cervical cancer screening in colombia. *Biomédica*, 39(1):65–74, 2019.

[29] Edward J Mayeaux Jr, Akiva P Novetsky, David Chelmow, Francisco Garcia, Kim Choma, Angela H Liu, Theognosia Papasozomenos, 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.

[30] Pavel Hamet and Johanne Tremblay. Artificial intelligence in medicine. *Metabolism*, 69:S36–S40, 2017.

[31] Real Academia Española. Inteligencia. Disponible en: https://dle.rae.es/inteligenciaLqtyoaQ.

[32] Asa B Simmons and Steven G Chappell. Artificial intelligence-definition and practice. *IEEE journal of oceanic engineering*, 13(2):14–42, 1988.

[33] Sebastian Raschka and Vahid Mirjalili. *Python machine learning: Machine learning and deep learning with Python, scikit-learn, and TensorFlow 2*. Packt Publishing Ltd, 2019.

[34] Vivek Kaul, Sarah Enslin, and Seth A Gross. History of artificial intelligence in medicine. *Gastrointestinal endoscopy*, 92(4):807–812, 2020.

[35] Silvana Secinaro, Davide Calandra, Aurelio Secinaro, Vivek Muthurangu, and Paolo Biancone. The role of artificial intelligence in healthcare: a structured literature review. *BMC medical informatics and decision making*, 21:1–23, 2021.

[36] Sobia Hamid. The opportunities and risks of artificial intelligence in medicine and healthcare. 2016.

[37] Fei Jiang, Yong Jiang, Hui Zhi, Yi Dong, Hao Li, Sufeng Ma, Yilong Wang, Qiang Dong, Haipeng Shen, and Yongjun Wang. Artificial intelligence in healthcare: past, present and future. *Stroke and vascular neurology*, 2(4), 2017.

[38] J Matthew Helm, Andrew M Swiergosz, Heather S Haeberle, Jaret M Karnuta, Jonathan L Schaffer, Viktor E Krebs, Andrew I Spitzer, and Prem N Ramkumar. Machine learning and artificial intelligence: definitions, applications, and future directions. *Current reviews in musculoskeletal medicine*, 13:69–76, 2020.

[39] D Ramyachitra and Parasuraman Manikandan. Imbalanced dataset classification and solutions: a review. *International Journal of Computing and Business Research (IJCBR)*, 5(4):1–29, 2014.

[40] Korbinian Koch and Marcus Soll. No matter how you slice it: Machine unlearning with sisa comes at the expense of minority classes. In *2023 IEEE Conference on Secure and Trustworthy Machine Learning (SaTML)*, pages 622–637. IEEE, 2023.

[41] Lucas Bourtoule, Varun Chandrasekaran, Christopher A Choquette-Choo, Hengrui Jia, Adelin Travers, Baiwu Zhang, David Lie, and Nicolas Papernot. Machine unlearning. In *2021 IEEE Symposium on Security and Privacy (SP)*, pages 141–159. IEEE, 2021.

[42] Jinhee Park, Hyunmo Yang, Hyun-Jin Roh, Woonggyu Jung, and Gil-Jin Jang. Encoder-weighted w-net for unsupervised segmentation of cervix region in colposcopy images. *Cancers*, 14(14):3400, 2022.

[43] 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.

[44] Jack Payette, Jake Rachleff, and C de Graaf. Intel and mobileodt cervical cancer screening kaggle competition: cervix type classification using deep learning and image classification. *Stanford University*, 2017.

[45] Bravo-Ortz Mario Alejandro, Arteaga-Arteaga Harold Brayan, Tabares-Soto Kl Reinel, Padilla-Buritic Jorge Ivn, and Orozco-Arias Simón. Clasificación de cáncer cervical usando redes neuronales convolucionales, transferencia de aprendizaje y aumento de datos. *Revista EIA*, 18(35):100–111, 2021.

[46] Manal Darwish, Mohamad Ziad Altabel, and Rahib H Abiyev. Enhancing cervical pre-cancerous classification using advanced vision transformer. *Diagnostics*, 13(18):2884, 2023.

[47] MobileODT Kaggle, Intel. Intel & mobileodt cervical cancer screening - dataset description, 2017. Disponible en: https://www.kaggle.com/competitions/intel-mobileodt-cervical-cancer-screening/data.

[48] MobileODT Kaggle, Intel. Intel & mobileodt cervical cancer screening, 2017. Disponible en: https://www.kaggle.com/competitions/intel-mobileodt-cervical-cancer-screening/overview.

[49] W Prendiville and R Sankaranarayanan. *Colposcopy and Treatment of Cervical Precancer*, volume 45 of *IARC Technical Report*. International Agency for Research on Cancer, Lyon (FR), 2017. `https://www.ncbi.nlm.nih.gov/books/NBK568361/`.

[50] Francois Chollet. *Deep learning with Python*. Simon and Schuster, 2021.

[51] Keras. Keras applications, 2023. https://keras.io/api/applications/.

[52] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 31, 2017.

[53] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.

[54] Tensorflow. Transfer learning and fine-tuning, 2022. https://www.tensorflow.org/tutorials/images/transfer$_l$*earning*?$hl = es - 419$.

[55] Yuxin Fang, Shusheng Yang, Shijie Wang, Yixiao Ge, Ying Shan, and Xinggang Wang. Unleashing vanilla vision transformer with masked image modeling for object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6244–6253, 2023.

[56] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[57] Aurélien Gerón. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly, 2019.

[58] Shantanu Godbole and Sunita Sarawagi. Discriminative methods for multi-labeled classification. In *Pacific-Asia conference on knowledge discovery and data mining*, pages 22–30. Springer, 2004.

[59] Narkhede S. Understanding confusion matrix, 2018. https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62.

[60] Anthony J Alberg, Ji Wan Park, Brant W Hager, Malcolm V Brock, and Marie Diener-West. The use of "overall accuracy" to evaluate the validity of screening or diagnostic tests. *Journal of general internal medicine*, 19(5p1):460–465, 2004.