

**Elementos para la construcción de un modelo de clasificación celular de la citología
cervicovaginal.**

Bernardo Meléndez Álvarez

Asesor

Isaac Esteban Camargo Freile

Universidad Nacional Abierta y a Distancia UNAD
Escuela de Ciencias Básicas, Tecnología e Ingeniería ECBTI
Especialización Ciencias de Datos y Analítica

2024

Director: Isaac Esteban Camargo Freile

Jurado: Luis Angel Anillo Arrieta

2024

Dedicatoria

A las células por darnos la vida y también por quitarla cuando su maquinaria falla.

Agradecimientos

Sincero agradecimiento a los profesores de la especialización, cuya dedicación y valiosos aportes desde cada una de sus disciplinas han enriquecido mi formación y por lo tanto este trabajo.

Un agradecimiento especial al profesor Isaac Esteban Camargo Freile por su guía y dirección durante el desarrollo de este proyecto. Su experiencia y orientación han sido fundamentales para alcanzar los objetivos.

A mi familia por ser mi impulso, motor y apoyo.

Resumen

Este trabajo se centra en mejorar la accesibilidad y precisión de los servicios de citología cervicovaginal a través del desarrollo de un modelo de clasificación celular específico. En el marco de un sistema de tele microscopía de bajo costo, se busca superar las limitaciones económicas y geográficas que restringen el acceso a servicios de calidad. El problema central reside en la carencia de modelos económicos, eficientes y precisos para la interpretación remota de muestras citológicas. La clasificación celular, crucial en la detección de anomalías cervicales y la prevención del cáncer de cuello uterino, se aborda mediante un modelo diseñado para su implementación en un sistema de tele microscopía accesible. El propósito principal es facilitar la detección temprana de lesiones cervicales y mejorar la accesibilidad a servicios de salud en países de ingresos limitados. La interpretación experta de las muestras es esencial, y la variabilidad en su calidad, junto con la diversidad en las variantes celulares, presenta desafíos, especialmente en entornos con recursos limitados. Este trabajo busca contribuir a cerrar esta brecha mediante el desarrollo de un modelo de clasificación celular que optimice la precisión diagnóstica, eficiencia y accesibilidad de los servicios citológicos. Al enfocarse en la citología cervicovaginal y aprovechar la tele microscopía de bajo costo, se espera proporcionar una solución práctica y efectiva para la detección temprana de lesiones cervicales, especialmente en áreas con limitaciones de recursos, marcando así una diferencia significativa en la salud de las mujeres.

Palabras clave: Cáncer de cuello uterino, citología, Papanicolau, Modelo automatizado de clasificación celular, Telemicroscopía.

Abstract

This work focuses on improving the accessibility and accuracy of cervicovaginal cytology services through the development of a specific cell classification model. Within the framework of a low-cost tele microscopy system, it seeks to overcome economic and geographical limitations that restrict access to quality services. The central problem lies in the lack of economical, efficient, and accurate models for the remote interpretation of cytological samples. Cell classification, crucial in detecting cervical abnormalities and preventing cervical cancer, is addressed through a model designed for implementation in an accessible tele microscopy system. The main purpose is to facilitate the early detection of cervical lesions and improve access to healthcare services in developing countries. Expert interpretation of samples is essential, and variability in their quality, along with diversity in cell variants, presents challenges, especially in resource-limited settings. This work aims to close this gap by developing a cell classification model that optimizes diagnostic accuracy, efficiency, and accessibility of cytological services. By focusing on cervicovaginal cytology and leveraging low-cost tele microscopy, it is expected to provide a practical and effective solution for the early detection of cervical lesions, especially in resource-constrained areas, thus making a significant difference in women's health.

Keywords: Uterine Cervical Neoplasms, Papanicolaou Test, Telemicroscopy, Machine Learning Model.

Tabla de contenido

Resumen.....	5
Abstract.....	6
Glosario.....	13
Introducción.....	15
Planteamiento del problema.....	16
Justificación.....	18
Objetivos.....	22
Objetivo general.....	22
Objetivos específicos.....	22
Marco conceptual.....	23
Modelo de Clasificación Celular.....	23
Tecnologías de imagen adaptadas al modelo.....	24
Métodos de clasificación celular.....	24
Validación y evaluación del modelo:.....	28
Desafíos clave que deben abordarse al desarrollar un sistema de clasificación celular automatizado en la detección del cáncer cervical.....	34
Marco teórico.....	37
La citología cervicovaginal.....	38

Sistemas automatizados de clasificación celular en la citología cervical.	41
Telemicroscopía y acceso a Salud.....	43
Tecnologías de hardware y software libres y abiertos en salud.....	45
Metodología y desarrollo de los objetivos	50
Conclusiones.....	74
Recomendaciones.....	75
Referencias.....	77
Apéndices.....	92

Lista de Tablas

Tabla 1. <i>Frecuencia y porcentaje de las categorías normal y anormal.</i>	65
Tabla 2. <i>Frecuencia y porcentaje de las 5 categorías del sistema Bethesda.</i>	65
Tabla 3. <i>Frecuencia y porcentaje de las 3 categorías del sistema Bethesda resumido...</i>	66
Tabla 4. <i>Resultado de las métricas de los modelos K-nn, SVM, Random Forest, XGBoost y Ensemble, con el uso de las clases normal y anormal.</i>	71
Tabla 5. <i>Resultado de las métricas de los modelos K-nn, SVM, Random Forest, XGBoost y Ensemble, con el uso de las clases Normal, Low-grade-Lesion y High-grade-Lesion.</i>	71
Tabla 6. <i>Resultado de las métricas de los modelos K-nn, SVM, Random Forest, XGBoost y Ensemble, con el uso de las 5 clases del sistema Bethesda.</i>	73

Lista de Figuras

Figura 1. <i>Imagen que resume los resultados obtenidos en la base de datos SCOPUS. ..</i>	50
Figura 2. <i>Imagen que resume los resultados obtenidos en la base de datos MEDLINE a través de PubMed.</i>	51
Figura 3. <i>Imagen que resume los resultados después de unir los resultados de las bases de datos SCOPUS y MEDLINE.</i>	51
Figura 4. <i>Se muestra la evolución de la producción bibliográfica a lo largo del tiempo. Se observa el considerable aumento en la producción anual durante los últimos 10 años.</i>	52
Figura 5. <i>Frecuencia absoluta según tipo celular. Dataset Herlev.....</i>	55
Figura 6. <i>Scatterplot que muestra la relación de las variables del dataset Herlev según su categoría de clasificación.</i>	57
Figura 7. <i>Scatterplots que muestran las relaciones de las variables del dataset Herlev según solo dos categorías de clasificación celular.....</i>	58
Figura 8. <i>Diagrama de caja que muestra la distribución de los datos del perímetro del núcleo en las categorías celulares Normal y Anormal del dataset Herlev.....</i>	59
Figura 9. <i>Muestra de uno de los archivos “.json” con la información para una de las imágenes del dataset CRIC donde se muestran las coordenadas calculadas para uno de los núcleos celulares.....</i>	61
Figura 10. <i>Se muestra una de las imágenes del dataset CRIC con la visualización de las etiquetas calculadas y el proceso de ajuste de estas etiquetas en el software labelme.</i>	62
Figura 11. <i>Matriz de correlación de las características nucleares y las diferentes clases</i>	67

Figura 12. <i>Diagramas de cajas de las principales variables en relación con las tres clases del sistema Bethesda resumido.</i>	68
Figura 13. <i>Diagramas de cajas de las principales variables en relación con las clases normal y anormal.</i>	68
Figura 14. <i>Diagramas de cajas de las principales variables en relación con las 5 clases del sistema Bethesda.</i>	69

Lista de Apéndices

Apéndice A. *Conjunto de datos CRIC con datos complementarios. (CRIC_New_label.csv)...*91

Glosario

Citología cervicovaginal o Test de Papanicolau: La prueba consiste en tomar muestras de células del cuello uterino y extenderlas en una placa que posteriormente será analizada bajo el microscopio en busca de células anormales o precancerosas.

Cáncer de cuello uterino: Cáncer que se origina en las células del cuello del útero. También se conoce como cáncer de cérvix o cáncer cervicouterino.

Machine Learning: Disciplina del campo de la Inteligencia Artificial que, a través de algoritmos, dota a los ordenadores de la capacidad de identificar patrones en datos masivos y elaborar predicciones. Este aprendizaje permite a los computadores realizar tareas específicas de forma autónoma.

Sistemas automatizados de clasificación celular: Sistemas que usando técnicas de Machine Learning y/o Deep learning permiten automatizar procesos de detección y clasificación celular.

Telemicroscopía: Consiste en crear y transmitir imágenes de muestras macroscópicas y/o microscópicas con diversos propósitos tales como el telediagnóstico, las consultas de segundas opiniones y con fines educativos.

Deep Learning: Métodos de aprendizaje profundo, como las redes neuronales convolucionales (CNN). Han demostrado ser gran promesa en la clasificación celular automatizada en citología cervical. Estos métodos pueden aprender automáticamente características de las imágenes, eliminando la necesidad de extracción de características hechas a mano.

Transfer learning: Es una técnica que implica el uso de un modelo pre-entrenado en un conjunto de datos grande, y posteriormente ajustarlo en un conjunto de datos más pequeño, como imágenes de citología cervical. Este enfoque aprovecha el conocimiento y las características aprendidas por el modelo pre-entrenado en tareas generales de reconocimiento de imágenes y las aplica a una tarea más específica, como la clasificación de células cervicales anormales.

Segmentación: La segmentación es uno de los problemas generales del campo de la visión artificial y consiste en dividir una imagen digital en varias regiones (grupos de píxeles) denominadas segmentos. Estos segmentos son de interés en una tarea específica.

Sistema de Clasificación Bethesda: Sistema de clasificación para la citología cervical. El sistema de Bethesda modificado (TBS) divide en dos grupos principales las células: células no neoplásicas y células neoplásicas. Dentro del grupo de células no neoplásicas, hay tres categorías principales: células escamosas, células glandulares y células endocervicales. Dentro del grupo de células neoplásicas, hay dos categorías principales: células escamosas anormales y células glandulares anormales. Las células escamosas anormales se dividen en tres subcategorías: células escamosas atípicas de significado indeterminado (ASC-US), células escamosas atípicas que no pueden excluir lesión de alto grado (ASC-H) y células escamosas que sugieren lesión de alto grado (HSIL). Las células glandulares anormales se dividen en dos subcategorías: células glandulares atípicas de significado indeterminado (AGC) y células glandulares que sugieren lesión de alto grado (AGC-H).

Introducción

La citología cervicovaginal representa un examen de tamizaje fundamental en la detección precoz de lesiones cervicales y la prevención del cáncer de cuello uterino, una de las principales causas de mortalidad entre mujeres a nivel mundial. Sin embargo, en numerosas regiones, el acceso a servicios de citología de calidad se ve restringido por barreras económicas y geográficas. En este contexto, la implementación de sistemas de tele microscopía de bajo costo puede ser una solución al permitir el análisis citológico de manera remota y automatizada.

A pesar de los avances en telemedicina, persiste el desafío de desarrollar modelos de clasificación celular económicos, eficientes y precisos para la citología cervicovaginal. La interpretación experta de muestras citológicas demanda gran número de horas de trabajo de personal capacitado que no siempre se encuentra disponible en países y regiones de bajos recursos ([Mustafa et al., 2020](#); [William et al., 2018](#)).

La variabilidad en la calidad de las muestras y la variabilidad interobservador pueden reducir la precisión de los diagnósticos citológicos, situación que se agrava en entornos con recursos limitados donde la capacitación y supervisión pueden resultar insuficientes ([Catarino, 2015](#)).

El presente trabajo se propone abordar esta brecha al investigar y precisar los elementos esenciales para implementar un modelo de clasificación celular específico para la citología cervicovaginal dentro del contexto de un sistema de tele microscopía de bajo costo. Este sistema busca mejorar la precisión diagnóstica, la eficiencia y la accesibilidad de los servicios citológicos, especialmente en áreas con recursos limitados, contribuyendo así a la prevención y detección temprana de patologías cervicales.

Planteamiento del problema

La citología cervicovaginal desempeña un papel fundamental en la detección temprana de lesiones cervicales y la prevención del cáncer de cuello uterino ([Swanson & Pantanowitz, 2024](#)). Sin embargo, en muchas regiones, el acceso a servicios de citología de calidad se ve limitado por barreras económicas y geográficas.

A pesar de los avances en telemedicina, la falta de modelos de clasificación celular eficientes y precisos en el contexto de la citología cervicovaginal, sigue siendo un desafío significativo ([Jiang et al., 2023](#); [Alias et al., 2022](#)). La interpretación de las muestras citológicas requiere una atención minuciosa y experta, lo que plantea la necesidad de desarrollar un modelo de clasificación celular robusto que pueda operar en un sistema de telemicroscopía de bajo costo.

La variabilidad en la calidad de las muestras citológicas se refiere a las posibles diferencias en la preparación y conservación, lo cual puede deberse a diversos factores como la técnica utilizada para la toma de muestras, la experiencia del profesional que realiza el procedimiento y la calidad de los materiales utilizados. Estas variaciones pueden afectar la integridad de las células presentes en la muestra y, por lo tanto, influir en la precisión de los resultados de la citología.

La disparidad en la interpretación entre observadores se refiere a las diferencias en la forma en que diferentes profesionales de la salud analizan y diagnostican las muestras citológicas. La interpretación de muestras citológicas requiere un alto nivel de experiencia y habilidad para identificar y clasificar adecuadamente las células anormales, y la falta de capacitación o supervisión adecuadas puede conducir a errores en el diagnóstico. Por lo tanto, la variabilidad en la calidad de las muestras y la variabilidad Inter observador pueden afectar la precisión de los diagnósticos citológicos. Este problema se agrava en entornos con recursos

limitados, donde la capacitación y supervisión de profesionales de la salud pueden ser insuficientes.

Este trabajo se propone abordar el vacío existente al indagar y precisar los elementos fundamentales para implementar un modelo de clasificación celular específico para la citología cervicovaginal en el contexto de un sistema de tele microscopía de bajo costo. El sistema busca mejorar la precisión diagnóstica, la eficiencia y la accesibilidad de los servicios citológicos, especialmente en áreas con recursos limitados, contribuyendo así a la prevención y detección temprana de patologías cervicales.

Justificación

El cáncer de cuello uterino constituye una preocupación global, siendo la cuarta causa más común de cáncer entre las mujeres a nivel mundial. En el caso específico de Colombia, el país se encuentra catalogado dentro de la franja de mayor riesgo en cuanto a la incidencia de esta enfermedad, además de estar clasificado entre aquellos con un alto riesgo de mortalidad asociada ([Jun et al., 2008](#); [Olmos, 2007](#)).

Numerosos estudios han puesto de manifiesto que la detección precoz y el tratamiento oportuno desempeñan papel crucial en la reducción significativa del riesgo de muerte por cáncer cervical invasivo. En este sentido, los programas de tamizaje implementados en países desarrollados han logrado reducir de manera notable las tasas de mortalidad asociadas a esta enfermedad ([Jóhannesson et al., 1978](#); [Jun et al., 2008](#); [Christopherson et al., 1976](#)). Sin embargo, es importante destacar que los países en vías de desarrollo no han alcanzado niveles equiparables de éxito en la implementación de programas de detección y prevención ([Cendales et al., 2010](#); [Wiesner et al., 2009](#)).

La importancia de estudiar esta problemática radica en su impacto significativo en la salud pública, así como en el bienestar y la calidad de vida de las mujeres a nivel global. La detección temprana y el tratamiento oportuno no solo pueden salvar vidas, sino también reducir la carga económica asociada al tratamiento de casos avanzados de cáncer cervical.

Este trabajo busca contribuir a la investigación y desarrollo de herramientas y estrategias innovadoras para la detección precoz y el manejo efectivo del cáncer de cuello uterino en contextos particulares como el caso de Colombia. A través de la implementación de modelos y algoritmos automatizados, que faciliten la implementación de políticas públicas que favorezcan

el acceso equitativo a programas de prevención y control, se espera contribuir a mejorar los resultados de salud y la reducción de la carga de esta enfermedad.

En Colombia, aunque el programa de detección para el cáncer de cuello uterino cuenta con una cobertura aceptable, en 2005 el 75,6% de las mujeres reportaron haberse realizado la citología en los últimos tres años, y 61% reportó realizarla anualmente en 2010, la mortalidad por esta causa no muestra una tendencia clara al descenso. Dentro de los aspectos críticos resaltan el insuficiente control de calidad de los laboratorios de citología y el recurso humano limitado en algunas regiones como el departamento de Boyacá, que con una extensión de 23.189 km² y 123 municipios, tan solo tiene 6 médicos patólogos y 16 histotecnólogos. Otro factor relevante es el bajo cubrimiento del programa de detección en zonas de difícil acceso como los departamentos de Guainía, Vichada, Amazonas y Vaupés donde se registran cifras superiores al 25% de mujeres que solo se han realizado la citología una vez en la vida o nunca ([Wiesner et al., 2009](#); [Cendales et al., 2010](#)).

La clasificación automatizada de lesiones cervicales es una tarea desafiante en el campo del aprendizaje automático. Varios estudios han empleado técnicas computacionales para mejorar la precisión diagnóstica de la citología cervical. Por ejemplo, Kuko y Pourhomayoun investigaron métodos tradicionales de aprendizaje automático, proponiendo un enfoque basado en Bosques Aleatorios, extrayendo 33 características morfológicas para clasificar las células ([Kuko & Pourhomayoun, 2019](#)).

Hussain et al. utilizaron varias arquitecturas de CNN, incluyendo AlexNet, VGGNet, ResNet y GoogLeNet, para clasificar lesiones cervicales con éxito ([Hussain et al., 2020](#)). Lin et al. propusieron un método que combina la apariencia y la morfología celular mediante CNN como GoogLeNet, AlexNet, ResNet y DenseNet para la clasificación, reportando cifras de

exactitud de 94.5%, 71.3% y 64.5% para problemas de clasificaciones de 2, 4 y 7 clases respectivamente ([Lin et al., 2019](#)).

Además, autores como Ghoneim et al. emplearon CNN como Shallow, VGG-16 y CaffeNet para extraer características, junto con técnicas como Máquinas de Aprendizaje Extremo para la clasificación celular aplicadas al dataset Herlev ([Ghoneim et al., 2020](#)). N. Diniz, D. et al. usan métodos tradicionales de aprendizaje automático para realizar una clasificación jerárquica sobre los conjuntos de datos Herlev y CRIC, usando un número variable de clases durante el proceso. Reportan, que la clasificación jerárquica funcionó mejor al utilizar Random Forest como el clasificador clave, en comparación con los árboles de decisión y k-NN ([Diniz et al., 2021](#)).

Los trabajos comentados anteriormente se caracterizan por usar métodos de aprendizaje profundo que, aunque pueden lograr cifras elevadas en las métricas de sensibilidad y especificidad, requieren de alto costo computacional que pueden limitar su implementación. El presente trabajo se enfoca en la clasificación automática de células cervicales utilizando el menor número de características con el fin de conseguir un costo computacional bajo, pero con buenas métricas de exactitud y precisión. Se busca maximizar la efectividad de lograr resultados verdaderos positivos y minimizar los resultados falsos negativos. Para ello, se exploran metodologías de aprendizaje automático y métodos de ensamble que sin elevar el costo computacional puedan lograr rendimientos aceptables.

El sustento de este modelo de clasificación celular radica en la necesidad de mejorar y optimizar los recursos disponibles en países en vías de desarrollo, donde la detección temprana del cáncer de cuello uterino es fundamental, pero enfrentan limitaciones económicas y de infraestructura. La implementación del modelo en un sistema de telemicroscopía de bajo costo

permitiría superar las barreras geográficas y brindar acceso a la detección temprana a mujeres que viven en áreas rurales o remotas, mejorando los tiempos en que se realiza la lectura y disminuyendo la carga de trabajo al escaso personal capacitado.

La lectura de la citología cervicovaginal mediante sistemas automatizados es potencialmente útil en países con dificultades de acceso y zonas dispersas en donde podría permitir ahorrar costos y mejorar la calidad de la lectura de la citología dentro de los programas de detección.

Por ello, el presente trabajo busca establecer los componentes necesarios para implementar un modelo eficiente y económicamente accesible para la clasificación celular en la citología cervicovaginal, que pueda ser implementado ampliamente y sin restricciones.

Objetivos

Objetivo general

Establecer los componentes necesarios para implementar un modelo altamente eficiente y económicamente accesible destinado a la clasificación celular de la citología cervicovaginal, utilizando un sistema de tele microscopía de bajo costo.

Objetivos específicos

1. Realizar una revisión exhaustiva de la literatura científica sobre la citología cervicovaginal y la detección automatizada de anomalías cervicales, con el fin de establecer una sólida base teórica y contextual.
2. Recolectar y preprocesar un conjunto de datos de alta calidad que contenga imágenes de células cervicales y metadatos pertinentes, garantizando la uniformidad de las características a emplear en la construcción del modelo.
3. Plantear un prototipo de modelo de aprendizaje automático que sea capaz de clasificar células cervicales en categorías relevantes, como células normales, células precancerosas o células cancerosas.
4. Evaluar el desempeño del modelo utilizando métricas de evaluación específicas para la clasificación.

Marco conceptual

Modelo de Clasificación Celular.

El modelo de clasificación celular en la citología cervicovaginal tiene como objetivo principal automatizar el proceso de identificación y clasificación de células en muestras microscópicas. Su propósito es mejorar la eficiencia del diagnóstico, permitiendo una identificación más rápida y precisa de células normales y anormales ([Jiang et al., 2023](#)).

Algunos aspectos clave en la definición del objetivo del modelo incluyen:

Identificación de anomalías celulares.

El modelo debe ser capaz de detectar y clasificar células que presenten anomalías indicativas de posibles condiciones patológicas, como células pre-cancerosas o cancerosas. Debe tener la capacidad de identificar diferentes tipos de células presentes en las muestras cervicovaginales, como células escamosas, células glandulares y células inflamatorias.

Optimización del tiempo de diagnóstico.

Una de las principales ventajas de utilizar un modelo de clasificación celular automatizado es la optimización del tiempo de diagnóstico. Con la lectura manual de citologías, los resultados pueden tardar días e incluso semanas en obtenerse, lo que prolonga el tiempo de espera de los pacientes y retrasa la toma de decisiones médicas. Sin embargo, con un modelo automatizado, el tiempo de diagnóstico se reduce significativamente, ya que el software puede analizar rápidamente las muestras y proporcionar resultados precisos en cuestión de minutos ([Lozano, 2007](#)).

Mejora de la precisión diagnóstica.

El modelo aspira a mejorar la precisión del diagnóstico al minimizar errores asociados con la interpretación humana, proporcionando así resultados más consistentes, repetibles y confiables.

Tecnologías de imagen adaptadas al modelo.

Microscopía digital de bajo costo. El prototipo de telemicroscopía, previamente desarrollado por el autor de este trabajo, posee características que facilitan la creación de un modelo de clasificación celular utilizando imágenes de células cervicales obtenidas a través de un telemicroscopio digital de bajo costo, lo que lo hace accesible para su implementación en comunidades con recursos limitados. Las resoluciones de desplazamiento en los ejes X y Y del prototipo son de 1.3 μm y 2.3 μm respectivamente, mientras que la resolución en el eje Z, que determina la capacidad de enfoque, es de 0.1 μm . La velocidad máxima de desplazamiento para los ejes X y Y es de 3mm/seg, y la densidad de píxeles por micra que se puede lograr con el prototipo es de 0.78125 $\mu\text{m}/\text{px}$ para un objetivo de 10X y de 0.1953 $\mu\text{m}/\text{px}$ para un objetivo de 40X ([Meléndez-Álvarez et al., 2017](#)). Estas características hacen que sea una opción ideal para la captura de imágenes de células cervicales con calidad óptima para su análisis y clasificación celular.

El empleo de esta tecnología de imagen garantiza una captura con calidad, que permita precisión y confiabilidad del modelo de clasificación celular, a la vez que reduce el costo.

Métodos de clasificación celular.

Los métodos usados en la clasificación automatizada pueden agruparse en:

Clasificación basada en reglas: La clasificación basada en reglas es uno de los métodos más antiguos utilizados en la detección del cáncer cervical ([William et al., 2018](#)). Este método utiliza reglas heurísticas para identificar células anormales, como la presencia de núcleos grandes o la falta de cohesión celular. Sin embargo, este enfoque basado en reglas tiene sus limitaciones, ya que depende de reglas predefinidas que pueden no ser suficientemente precisas o adaptativas a diferentes tipos de células cervicales. Además, requiere de una gran cantidad de conocimiento experto para establecer y mantener estas reglas. Este método es limitado en su capacidad para detectar anomalías sutiles y puede llevar a errores de diagnósticos.

Clasificación basada en características morfológicas: La clasificación basada en características morfológicas utiliza características, como la forma y el tamaño de las células, para identificar células anormales. Este enfoque se basa en algoritmos de extracción de características morfológicas, que analizan la forma, el tamaño y otros aspectos visuales de las células para determinar su clasificación ([Plissiti et al., 2011](#)). Estos algoritmos pueden ser entrenados previamente con muestras etiquetadas para aprender a reconocer patrones específicos de células anormales. La ventaja de este método es que no depende de reglas predefinidas, lo que lo hace más flexible y adaptable a diferentes tipos de células cervicales ([Plissiti et al., 2011](#)).

Este método es más preciso que la clasificación basada en reglas, pero aún puede ser limitado en su capacidad para detectar anomalías sutiles ([Dey et al., 2018](#)).

Clasificación basada en características de textura: La clasificación basada en características de textura utiliza características, como la rugosidad y la homogeneidad, para identificar células anormales. Este enfoque se basa en análisis de textura que examinan las variaciones en la intensidad y la distribución de los píxeles para identificar características distintivas en las células cervicales. Las técnicas de procesamiento de imágenes se utilizan para

extraer estas características de textura y luego se aplican algoritmos de clasificación para distinguir entre células normales y anormales.

La ventaja de la clasificación basada en características de textura es que puede capturar detalles sutiles en las células cervicales que pueden ser indicativos de anomalías. Este método es más preciso que la clasificación basada en características morfológicas y puede detectar anomalías sutiles ([Fekri-Ershad, 2019](#)).

Clasificación basada en aprendizaje automático: La clasificación basada en aprendizaje automático utiliza algoritmos de aprendizaje automático para identificar células anormales. Este método es el más preciso y eficiente de todos los métodos de clasificación celular y puede detectar anomalías sutiles. Para implementar la clasificación basada en aprendizaje automático, se utilizan algoritmos como redes neuronales, máquinas de vectores de soporte o árboles de decisión. Estos algoritmos se entrenan previamente con un conjunto de datos etiquetados que contienen tanto células normales como anormales, para que puedan aprender a reconocer patrones y características específicas asociadas con células anormales ([Dey et al., 2018](#); [William et al., 2018](#)).

En cuanto a los métodos basados en aprendizaje automático son importantes:

Métodos de aprendizaje automático tradicionales: Los métodos de aprendizaje automático tradicionales, como las máquinas de vectores de soporte (SVM) y los bosques aleatorios, se han utilizado para la clasificación celular automatizada en citología cervical. Estos métodos requieren la extracción de características hechas a mano de las imágenes, lo que puede ser lento y puede no capturar toda la información relevante. Sin embargo, aún pueden lograr una

alta precisión cuando se combinan con técnicas adecuadas de selección y preprocesamiento de características ([William et al., 2018](#)).

Métodos de aprendizaje profundo: Los métodos de aprendizaje profundo, como las redes neuronales convolucionales (CNN), han demostrado ser gran promesa en la clasificación celular automatizada en citología cervical. Estos métodos pueden aprender automáticamente características de las imágenes, eliminando la necesidad de extracción de características hechas a mano. Esto los hace más eficientes y rápidos en comparación con los métodos tradicionales. Las redes neuronales convolucionales han demostrado su efectividad en la clasificación de imágenes dermatoscópicas de lesiones cutáneas, lo cual sugiere que también pueden ser útiles en la clasificación de células cervicales anormales ([Maurya et al., 2023](#)).

Las CNN se han utilizado para diversas tareas, como la identificación, detección, segmentación y diagnóstico a nivel de células. Han logrado una alta precisión y han superado a los métodos de aprendizaje automático tradicionales en muchos casos ([Khamparia et al., 2020](#); [Alias et al., 2022](#); [Moral-Argumedo et al., 2023](#); [Attallah, 2023](#)).

Transfer learning: Transfer learning es una técnica que implica el uso de un modelo CNN pre-entrenado en un conjunto de datos grande, como ImageNet, y posteriormente ajustarlo a un conjunto de datos más pequeño, como imágenes de citología cervical ([Khamparia et al., 2020](#)). Este enfoque aprovecha el conocimiento y las características aprendidas por el modelo pre-entrenado en tareas generales de reconocimiento de imágenes y las aplica a una tarea más específica, como la clasificación de células cervicales anormales.

El uso de transfer learning ha demostrado ser especialmente efectivo en la clasificación de células cervicales, ya que los modelos pre-entrenados pueden capturar patrones y características abstractas que son relevantes en diversas tareas de reconocimiento de imágenes. Este enfoque puede reducir la cantidad de datos necesarios para el entrenamiento y mejorar el rendimiento de generalización del modelo ([Jiang et al., 2023](#); [Attallah, 2023](#)).

Métodos de ensamble: Los métodos de ensamble implican la combinación de múltiples modelos para mejorar la precisión y la robustez de la clasificación. Por ejemplo, se puede utilizar una combinación de CNN y SVM para lograr una alta precisión en la clasificación celular. ([Diniz et al., 2021](#); [Sabeena & Gopakumar, 2022](#)).

Validación y evaluación del modelo:

Conjunto de datos de validación

Se debe realizar la recolección y preprocesamiento de un conjunto de datos de alta calidad que contenga imágenes de células cervicales y metadatos pertinentes. La uniformidad y calidad de los datos son fundamentales para garantizar la validez y confiabilidad del modelo de aprendizaje automático. Este conjunto de datos debe ser representativo de la variabilidad y diversidad de las células cervicales para garantizar que el modelo sea capaz de generalizar y clasificar con precisión diferentes tipos de células cervicales. Además, se deben considerar las particularidades y desafíos específicos de la citología cervicovaginal en el diseño y desarrollo del modelo de aprendizaje automático.

Los principales conjuntos de datos o dataset, disponibles para entrenar y evaluar modelos de aprendizaje automático en la clasificación de células cervicales, incluyen el SIPaKMeD dataset, que contiene imágenes de citología de células cervicales clasificadas en cinco clases

diferentes ([Plissiti et al., 2018](#)). También se pueden utilizar otros conjuntos de datos, como el Herlev dataset, que se ha utilizado en estudios anteriores para clasificar células cervicales anormales ([Martin, 2003](#)).

Los principales conjuntos de datos de acceso público se listan a continuación describiendo sus principales características:

Conjunto de datos de Herlev: Este conjunto de datos se recopiló en el Hospital Herlev en Dinamarca y contiene 917 imágenes de células cervicales en siete clases. Las imágenes fueron anotadas manualmente por citotecnólogos y patólogos experimentados. El conjunto de datos se ha utilizado para tareas de clasificación y segmentación celular. Todas las imágenes se recogen mediante el uso de un microscopio conectado a un capturador de fotogramas con una resolución de 0,201 $\mu\text{m}/\text{píxel}$. Cada imagen celular se segmenta manualmente del fondo, el citoplasma y el núcleo para extracción adicional de características ([Martin, 2003](#)).

ISBI 2014: Este conjunto de datos se publica para el primer desafío de segmentación de imágenes de citología cervical superpuesta bajo los auspicios del Simposio internacional IEEE sobre imágenes biomédicas. El conjunto de datos consta de 16 citologías y 945 imágenes sintéticas. Cada imagen consta de 20 a 60 células cervicales teñidas con Papanicolaou con diferentes grados de superposición. Este conjunto de datos fue creado por la Universidad de Adelaida en Australia. La resolución de las imágenes es de aproximadamente 0,185 $\mu\text{m}/\text{píxel}$ ([Lu et al., 2015](#)).

ISBI 2015: Consiste en una colección de 17 volúmenes de células cervicales superpuestas. La principal diferencia entre ISBI 2015 e ISBI 2014 es que los datos de entrada consisten en un volumen de citología multicapa, lo que significa que los datos de entrada ahora

son un volumen que consta de un conjunto de imágenes multifocales adquiridas de la misma muestra. Este conjunto de datos de entrada más rico puede proporcionar más información sobre la tarea de detectar y segmentar las células cervicales, lo que permite una detección y segmentación citoplasmática y nuclear más precisa de las células cervicales ([Lu et al., 2017](#)).

Conjunto de datos de SIPaKMeD: Esta base de datos consta de 4049 imágenes de células cervicales aisladas que se adquieren en la Universidad de Loánina, Grecia, a través de una cámara adaptada a un microscopio óptico. Las células son anotadas por citopatólogos experimentados en cinco clases diferentes (células superficiales intermedias, parabasales, coilocitóticas, disqueratósicas y metaplásicas), según su apariencia citológica y morfología. Entre estas cinco clases, las superficiales-intermedias y las parabasales son células normales. Los coilocitos y las disqueratósicas son células anormales, pero no malignas, mientras que las metaplásicas pertenecen a las benignas. En cada imagen de la base de datos SIPaKMeD se definen manualmente las áreas del citoplasma y el núcleo ([Plissiti et al., 2018](#)).

CERVIX93: Este conjunto de datos consta de 93 imágenes proporcionadas por Moffitt Cancer Center (Tampa, FL). Todas las imágenes se adquieren mediante un sistema de microscopio integrado (Stereologer, SRC Biosciences, Tampa, FL) con un aumento de 40x. Cada una de las pilas tiene entre 10 y 20 imágenes y todas las imágenes tienen un tamaño de 1280×960 píxeles. Todas las imágenes son examinadas por citólogos y clasificadas con tres categorías (Negativo, LSIL, HSIL). Un total de 2705 núcleos se anotan manualmente con cuadros delimitadores según todas las categorías ([Phoulady & Mouton, 2018](#)).

BHS: Esta base de datos recopila 194 pruebas de Papanicolaou convencionales del Sistema Brasileño de Salud. Cada imagen tiene una resolución de 0,255 mm/píxel con un tamaño de 1392×1040 . Las imágenes están etiquetadas en dos clases (normal/anormal) y las imágenes anormales contienen 5 tipos diferentes de células pre-cancerosas (carcinoma, HSIL, LSIL, ASCUS y ASCH) ([Araújo et al., 2019](#)).

CRIC: La colección del conjunto de datos del CRIC tiene 400 imágenes de pruebas de Papanicolaou cervicales convencionales y 11.534 células clasificadas. Las pruebas de Papanicolaou se recolectan de 118 pacientes en la región sudeste de Brasil y se preparan y analizan en el Laboratorio de Citología de la Facultad de Farmacia de la Universidad Federal de Ouro Preto, Minas Gerais, Brasil. El tamaño de la imagen es 1376×1020 y la resolución es $0,228 \mu\text{m}/\text{píxel}$. La colección CRIC cubre seis tipos según la nomenclatura TBS: NILM (6779), ASC-US (606), LSIL (1360), ASC-H (925), HSIL (1703) y SCC (161) ([Rezende et al., 2021](#)).

Detector de comparación. Esta base de datos es recopilada por la Universidad Central del Sur en China. Consta de 7410 imágenes cervicales recortadas de placas de citología totalmente digitalizadas (Whole Slide Imaging). Hay un total de 48.587 cuadros delimitadores de instancias de objetos etiquetados por citopatólogos experimentados. Según las categorías de TBS, los objetos anotados pertenecen a 11 categorías: ASC-US, ASC-H, LSIL, HSIL, SCC, AGC, trichomonas (TRICH), candida (CAND), flora, herpes y actinomyces (ACTIN) ([Liang et al., 2021](#)).

RepoMedUNM Esta base de datos es publicada por Universitas Nusa Mandiri en Indonesia y consta de 6168 imágenes de células de frotis de Papanicolaou recopiladas de 24 placas, incluidas imágenes de la prueba de Papanicolaou convencional y de la prueba de Papanicolaou en medio líquido (ThinPrep). Contiene 3083 imágenes convencionales ThinPrep,

etiquetadas en dos categorías, normal y LSIL. Las imágenes ThinPrep se dividen en tres categorías: células normales (1513), células de coilocitos (434) y HSIL (410) ([Jiang et al., 2023](#)).

CCEDD: Este conjunto de datos recopila 686 imágenes de cuello uterino con un tamaño de 2048×1536 del Liaoning Cancer Hospital & Institute. Las imágenes capturadas contienen células cervicales superpuestas en varios fondos complejos y están etiquetadas por 6 citólogos experimentados para delinear los contornos del citoplasma y el núcleo. Las imágenes originales se dividen en conjuntos de entrenamiento, validación y prueba utilizando una proporción de 6:1:3. Todas las imágenes sin procesar se cortan en 512×384 píxeles y se obtienen 33.614 imágenes ([Liu et al., 2022](#)).

Cx22. Este conjunto de datos es una extensión del conjunto de datos CCEDD publicado por el Laboratorio de Procesamiento de Información Optoelectrónica de la Academia China de Ciencias, Shenyang, en el que se anotan citoplasma y núcleo. Se recopilan un total de 14,946 instancias en 1320 imágenes con un tamaño de 512×512 ([Liu et al., 2022](#)).

Métricas de evaluación.

Una vez que se ha entrenado el modelo de aprendizaje automático, es importante utilizar métricas de evaluación para medir su rendimiento y precisión ([Zhang et al., 2022](#)).

Sensibilidad: La sensibilidad se refiere a la capacidad del sistema de clasificación celular para detectar células anormales ([Lantz, 2019](#)). La sensibilidad es un indicador importante de la eficacia del sistema de clasificación celular en la detección del cáncer cervical. Los clasificadores celulares basadas en aprendizaje profundo han demostrado resultados prometedores al aprender a afrontar mejor la variabilidad de la imagen y la estructura celular. Sin embargo, el uso actual en sistemas de asistencia en la práctica diaria aún enfrenta varios desafíos,

no sólo porque se trata de enfoques que consumen más tiempo y recursos, sino también porque suelen presentar una complejidad significativamente mayor ([Conceição et al., 2019](#)).

Especificidad: La especificidad se refiere a la capacidad del sistema de clasificación celular para identificar células normales correctamente. La especificidad es importante para reducir el número de falsos positivos y minimizar la necesidad de pruebas de seguimiento innecesarias. Los sistemas de clasificación celular deben tener una especificidad alta y consistente ([Allahqoli et al., 2022](#)).

Precisión: La precisión se refiere a la capacidad del sistema de clasificación celular para identificar correctamente tanto las células anormales como las normales ([Bonaccorso, 2018](#)). La precisión es importante para garantizar que el sistema de clasificación celular sea confiable y efectivo.

Eficiencia: La eficiencia se refiere a la capacidad del sistema de clasificación celular para procesar grandes cantidades de muestras de manera rápida. La eficiencia es importante para reducir el tiempo y los costos asociados con la detección del cáncer cervical.

El tamizaje es una forma de identificar a personas aparentemente sanas que pueden tener un riesgo aumentado de una condición particular. Las pruebas de tamizaje, como la citología cervicovaginal, deben ser sensibles, específicas, eficientes y precisas. Una prueba de tamizaje debe, teóricamente tener una sensibilidad que supere el 95% si la especificidad es menor o igual al 95%, y viceversa. El objetivo es detectar más casos verdaderos positivos que casos falsos positivos cuando la prevalencia de la enfermedad es menor o igual al 5%. La mayoría de las pruebas de tamizaje no cumplen con este alto estándar ([Allahqoli et al., 2022](#)). Estudios sobre la citología cervicovaginal convencional reportan cifras variables de sensibilidad que se encuentran

entre el 30% y el 87%, en relación con la especificidad los resultados se encuentran en el rango del 86% y al 100% ([Nanda et al., 2000](#); [Smith et al., 2023](#)). Estas cifras tan variables son consecuencia principalmente de la variación inter e intraobservador, la carga de trabajo y la calidad de la muestra. A pesar de estos datos, la prueba de Papanicolau ha sido un método efectivo en la estrategia de detección temprana de las lesiones pre-cancerosas del cuello uterino ([Sharif, 2022](#); [Sachan et al., 2018](#); [Patil &, 2022](#)), esto por particularidades del desarrollo lento que tiene este tipo de cáncer y los esquemas usados en los programas de detección que permiten su efectividad.

De igual manera, el objetivo final de cualquier sistema automatizado de clasificación celular es apoyar efectivamente la detección de pacientes potencialmente enfermos. De este modo es fundamental lograr un buen equilibrio entre una alta sensibilidad y especificidad. La detección automatizada puede y debe mejorar las pruebas manuales aumentando la sensibilidad y la especificidad, evitando la variación inter e intraobservador presentes en el tamizaje manual. Se debe lograr disminuir la carga de trabajo de los citotecnólogos, reducir el costo de los programas de detección temprana y lograr reducir la probabilidad de incidencia de fases avanzadas de la enfermedad que logren reducir la tasa de mortalidad consecuencia de la enfermedad.

Desafíos clave que deben abordarse al desarrollar un sistema de clasificación celular automatizado en la detección del cáncer cervical.

Variabilidad en la apariencia de las células: La apariencia de las células puede variar significativamente debido a factores como la edad, el ciclo menstrual y la presencia de infecciones. Esto puede dificultar la identificación de células anormales y puede llevar a errores de diagnóstico. Para abordar este desafío, los sistemas de clasificación celular automatizados deben ser capaces de identificar y adaptarse a la variabilidad en la apariencia de las células.

Presencia de artefactos: Los artefactos, como las manchas y las burbujas, pueden aparecer en las muestras de células y pueden dificultar la identificación de células anormales. Los sistemas de clasificación celular automatizados deben ser capaces de identificar y eliminar los artefactos para mejorar la precisión del diagnóstico ([Bengtsson & Malm, 2014](#)).

Necesidad de una validación rigurosa: Los sistemas de clasificación celular automatizados deben ser validados rigurosamente para garantizar su precisión y fiabilidad. La validación debe incluir pruebas en una amplia variedad de muestras y condiciones para garantizar que el sistema sea efectivo en situaciones del mundo real, realizar pruebas clínicas en ambientes de la práctica diaria real es esencial.

Necesidad de una gran cantidad de datos: Los sistemas de clasificación celular automatizados basados en aprendizaje automático requieren una gran cantidad de datos para entrenar los algoritmos de clasificación. La recopilación de datos puede ser costosa y llevar mucho tiempo, lo que puede limitar la capacidad de los sistemas para mejorar la eficiencia del proceso de detección ([Jiang et al., 2023](#)).

Dificultad para identificar células anormales sutiles: Las células anormales pueden ser sutiles y difíciles de identificar, lo que puede llevar a errores de diagnóstico. Los sistemas de clasificación celular automatizados deben ser capaces de identificar células anormales sutiles y distinguirlas de las células normales, esto requiere de muestras y conjuntos de datos con suficiente número para poder entrenar de manera adecuada los sistemas.

Limitaciones de recursos: El desarrollo e implementación de sistemas de clasificación celular automatizados puede requerir una inversión significativa en términos de recursos técnicos, financieros y humanos. Estas limitaciones pueden dificultar la accesibilidad y adopción

generalizada de estos sistemas, especialmente en entornos con recursos limitados. La clasificación celular automatizada en la citología cervicovaginal presenta varios desafíos y consideraciones teóricas y conceptuales que deben abordarse para garantizar la eficiencia y accesibilidad del proceso de detección. Los avances en este campo pueden mejorar la eficiencia y precisión del proceso de detección de células anormales, lo que a su vez puede tener un impacto significativo en la detección temprana del cáncer de cuello uterino.

Marco teórico

Cáncer de cuello uterino

El cáncer cervicouterino es una de las principales causas de muerte en mujeres a nivel mundial. Según datos del Global Cancer Observatory, se estima que hubo 604,000 nuevos casos y 342,000 muertes por cáncer de cuello uterino en todo el mundo en 2020 ([Pradit et al., 2022](#); [Woldetsadik et al., 2020](#); [Araújo et al., 2019](#)). Este tipo de cáncer es la cuarta causa de cáncer en mujeres ([Pangtey et al., 2022](#)), lo que lo convierte en una enfermedad de alto interés en salud.

La variación en la tasa de incidencia de este cáncer entre países es sorprendentemente amplia. En muchos países es el cáncer más común entre las mujeres, mientras que en algunos países ha descendido al décimo lugar. Alrededor del 86% de los casos ocurren en países en desarrollo. En África, la tasa de incidencia estandarizada por edad es de 25 por 100.000 habitantes por año; en algunos países de ese continente es más del doble de esa tasa. En la India la tasa de incidencia es de 27, mientras que en EE.UU. es de 5,7 y en Finlandia de 3,7 por 100.000 habitantes ([WHO Information on Human papillomavirus \(HPV\) and cervical cancer – Canadian Cancer Survivor Network., 2023](#)). Se observan variaciones de más de diez puntos en las tasas de incidencia del cáncer de cuello uterino entre los diferentes países. Si bien, una parte de esta variación puede atribuirse a variaciones generales en las condiciones de vida y a la diferente propagación del virus del papiloma humano (VPH) en la población, la mayor parte se atribuye al éxito de la detección mediante la prueba de Papanicolaou en los países que cuentan con programas de detección y garantizan acceso a estos. Colombia, se ubica en la franja de más alto riesgo para incidencia y en el grupo de países con riesgo alto para mortalidad por cáncer de cuello uterino. Los programas de tamizaje en países desarrollados, usando la citología cervicovaginal, como examen de primera línea, han logrado reducir las tasas de mortalidad; sin

embargo, los países en desarrollo no han logrado un descenso equiparable ([Jun et al., 2008](#); [Olmos et al., 2007](#)).

Este cáncer es prevenible y tratable si se detecta a tiempo, la principal herramienta para detectar el cáncer de cuello uterino es la citología cervicovaginal, también conocida como prueba de Papanicolaou. Este examen es una herramienta fundamental en la detección temprana y prevención. Para llevar a cabo un adecuado tamizaje y detección temprana, es importante implementar programas de salud que promuevan la realización periódica de la prueba de Papanicolaou ([Jóhannesson et al., 1978](#); [Aristizabal et al., 1984](#); [Christopherson et al., 1976](#); [Lăără et al., 1987](#)). Esta prueba ha demostrado ser eficaz en la detección de células anormales, lo que permite un diagnóstico precoz y un tratamiento oportuno en etapas tempranas de la enfermedad.

La citología cervicovaginal

La citología consiste en la toma y análisis de muestras celulares del cuello uterino y la vagina para detectar posibles anomalías o células precancerosas ([Sanabria et al., 2022](#)). Esta prueba ha demostrado ser efectiva en la reducción de la incidencia y mortalidad por cáncer de cérvix, especialmente cuando se implementa dentro de programas de tamizaje ([Zappacosta & Rosini, 2008](#); [Joshi et al., 2021](#); [Zea et al., 2007](#); [Han et al., 2012](#)).

La prueba consiste en tomar muestras de células del cuello uterino y extenderlas en una placa que posteriormente será analizada bajo el microscopio en busca de células anormales o precancerosas. Estas células se clasifican de acuerdo con su apariencia y características, y se reporta el resultado como normal, anormal o sospechoso. La clasificación de las células cervicales en la citología cervicovaginal es crucial para determinar si existe alguna anomalía que pueda indicar la presencia de cáncer. Un sistema de informe estandarizado es necesario para

ayudar a superar las barreras geográficas y culturales, mejorar la eficiencia del diagnóstico y facilitar la investigación científica. El primer sistema de clasificación numérica para la citología cervical fue el sistema de clasificación de Papanicolaou. Este sistema, desarrollado por el Dr. George Papanicolaou en la década de 1940, clasificó las células cervicales en cinco categorías numéricas, desde la ausencia de células anormales hasta la presencia de células malignas. Posteriormente otros sistemas de clasificación se han desarrollado, como el sistema de Bethesda, que se introdujo en 1988 y se ha actualizado varias veces siendo hoy el sistema de clasificación patológica de la citología cervicovaginal más ampliamente usado ([Poomtavorn et al., 2013](#)).

El sistema de Bethesda modificado (TBS) divide en dos grupos principales las células: células no neoplásicas y células neoplásicas. Dentro del grupo de células no neoplásicas, hay tres categorías principales: células escamosas, células glandulares y células endocervicales. Dentro del grupo de células neoplásicas, hay dos categorías principales: células escamosas anormales y células glandulares anormales. Las células escamosas anormales se dividen en tres subcategorías: células escamosas atípicas de significado indeterminado (ASC-US), células escamosas atípicas que no pueden excluir lesión de alto grado (ASC-H) y células escamosas que sugieren lesión de alto grado (HSIL). Las células glandulares anormales se dividen en dos subcategorías: células glandulares atípicas de significado indeterminado (AGC) y células glandulares que sugieren lesión de alto grado (AGC-H). Su implementación ha demostrado ser de gran utilidad en la detección de lesiones cervicales y ha contribuido a la reducción de la mortalidad por cáncer cervicouterino en muchos países ([Jiang et al., 2023](#)).

Sin embargo, la detección manual y visual de células cervicales en las muestras citológicas es un proceso laborioso y propenso a errores debido a la variabilidad inter observador y la subjetividad en la interpretación de las imágenes. Además, la citología cervicovaginal

presenta limitaciones en términos de tiempo y costo, lo que dificulta su implementación y acceso a gran escala en regiones con recursos limitados ([Catarino, 2015](#)).

En países de bajos ingresos no se cuenta con el suficiente personal capacitado, esto sumado a que la población se ubica en zonas dispersas de difícil acceso, limita la oportunidad para realizar de manera eficiente estas pruebas de detección temprana. Si se consideran las defunciones por todas las neoplasias malignas (excluidas las de piel) en mujeres en las Américas, la mortalidad proporcional por cáncer cervicouterino varía desde menos de 2,5% en América del Norte hasta 49,2% en Haití ([Lewis, 2004](#)). La tasa de médicos disponibles en E.U es de 3.6 por cada 1.000 habitantes, mientras que Haití cuenta con una tasa de 0.2 médicos por cada 1.000 habitantes ([Bank, 2024](#)). Estas diferencias en disponibilidad de personal capacitado y oportunidad para realizar las pruebas de detección tienen un impacto significativo en los resultados de salud en diferentes regiones.

Según la OMS para el 2015, del total de muertes por cáncer, las dos terceras partes ocurrieron en países de bajos ingresos, en los que tan solo el 35% disponía de servicios de patología en el sector público. Mientras que el 98% de los países de altos ingresos los tienen funcionando adecuadamente ([Garwood, 2017](#)). La falta de acceso a servicios de patología y la escasez de personal capacitado en países de bajos ingresos representa un gran desafío para la detección temprana y el manejo adecuado de las lesiones cervicales.

La importancia de la citología vaginal radica en su eficacia y eficiencia para la detección precoz del cáncer cervicouterino. Esta técnica ha sido ampliamente utilizada durante más de 40 años y se estima que reduce en aproximadamente un 80% la mortalidad por este tipo de cáncer. La sensibilidad de la citología cervicovaginal en países en desarrollo varía entre el 44% y el 78%, mientras que su especificidad se sitúa entre el 91% y el 99% ([Catarino, 2015](#)).

Sistemas automatizados de clasificación celular en la citología cervical.

En los últimos años, el uso de técnicas de aprendizaje automático y visión por computadora ha demostrado ser prometedor para mejorar la eficiencia y precisión en la clasificación de células cervicales en la citología cervicovaginal.

En la década de 1960 se desarrollaron los primeros sistemas de clasificación celular automatizados, estos se basaban en reglas heurísticas y utilizaban reglas simples para identificar células anormales, pero eran limitados en su capacidad para detectar anomalías sutiles. La introducción de la microscopía automatizada en la década de 1970 permitió la captura de imágenes digitales de células individuales, lo que llevó al desarrollo de sistemas de clasificación celular automatizados más avanzados. Estos sistemas utilizaban características morfológicas y de textura para identificar células anormales y se basaban en algoritmos de clasificación supervisada.

En la década de 1980, se desarrollaron sistemas de clasificación celular automatizados más sofisticados que utilizaban técnicas de procesamiento de imágenes y aprendizaje automático. Estos sistemas podían identificar células anormales con mayor precisión y eficiencia que los sistemas anteriores. La década de 1990, produjo un aumento en la investigación sobre la automatización de la clasificación celular en la detección del cáncer cervical, y se desarrollaron varios sistemas comerciales. Estos sistemas utilizaban una variedad de técnicas de procesamiento de imágenes y aprendizaje automático para identificar células anormales. En la década de 2000, se produjo un aumento en la investigación sobre la integración de múltiples técnicas de clasificación celular automatizada para mejorar la precisión y eficiencia del proceso de detección ([Bengtsson & Malm, 2014; Jiang et al., 2023](#)).

La detección automatizada de anomalías cervicales en la citología cervicovaginal ha surgido como una prometedora alternativa para mejorar la eficiencia y precisión de este proceso ([Jiang et al., 2023](#); [William et al., 2018](#); [Chitra & Kumar, 2021](#)). Se basa en el uso de algoritmos y técnicas de aprendizaje automático que pueden analizar y clasificar las células cervicales de manera rápida y precisa brindando apoyo a los citotecnólogos, reduciendo cargas de trabajo y permitiendo un tamizaje más eficiente y económico a gran escala. Además, la detección automatizada también puede ayudar a reducir los falsos negativos y falsos positivos que pueden ocurrir durante la interpretación manual de las muestras. Los métodos basados en visión por computador tienen el potencial de aumentar el grado de objetividad y reproducibilidad de los análisis, así como su velocidad ([Matias et al., 2021](#)).

Estos métodos automatizados de lectura de citologías cervicales uterinas han demostrado ser una herramienta prometedora en el cribado y detección precoz del cáncer cervical. Al utilizar algoritmos y técnicas de procesamiento de imágenes, estos sistemas pueden analizar grandes volúmenes de citologías de manera más rápida y precisa que la lectura manual.

Uno de los beneficios clave de la lectura automatizada es su capacidad para reducir el error humano y mejorar la consistencia en la interpretación de los resultados. Los sistemas automatizados utilizan criterios objetivos y basados en evidencia para clasificar las células, lo que reduce la variabilidad Inter observador y mejora la precisión del diagnóstico. Además, la lectura automatizada de citologías cervicales también puede ayudar a identificar células anormales que pueden pasar desapercibidas durante la lectura manual ([Alias et al., 2022](#); [Lozano, 2007](#); [Zhang et al., 2017](#)).

Telemicroscopía y acceso a Salud

La telepatología, una subárea de la telemedicina, consiste en crear y transmitir imágenes de muestras macroscópicas y/o microscópicas con diversos propósitos tales como el tele diagnóstica, las consultas de segundas opiniones y con fines educativos. Ha sido aplicada en todas las subespecializaciones de la patología; incluyendo la patología quirúrgica, la citopatología, la anatomía patológica y la patología clínica. Se ha usado, además, como instrumento para el control de la calidad de laboratorios y en actividades de investigación ([Dudas et al., 2014](#)).

El desarrollo de la Tele patología ha estado ligado de manera estrecha al desarrollo tecnológico. No obstante, su alto costo generado en parte por la tecnología necesaria restringe su disponibilidad en países de bajos ingresos donde paradójicamente es más necesaria. Los costos de sistemas de tele patología comerciales, diseñados bajo modelos de software y hardware propietario, se encuentran por encima de los US\$10000, condicionando su uso a centros que cuenten con capacidad económica ([Marín & Romero, 2011](#)).

Frente a este problema han surgido opciones como el sistema de interconsulta descrito por Sirintrapun et al ([Sirintrapun & Cimic, 2012](#)), usando el software skype®, que logra un 89% de concordancia diagnóstica. Pero que tiene el inconveniente de necesitar la interacción sincrónica de personal capacitado. Otros autores como Bellina y Missoni en el año 2009 ([Bellina & Missoni, 2009](#)), plantearon el uso de imágenes tomadas con teléfonos celulares directamente del microscopio, como un método de tele microscopia estática, que puede usarse en caso de requerirse confirmación diagnóstica desde zonas apartadas. Alfaro y Roca en 2008 ([Alfaro & Roca, 2008](#)), utilizan el término de telepatología portátil al implementar un sistema de tele patología dinámica, usando software libre para la transmisión de vídeo e imágenes, restringiendo

los requerimientos de hardware a una cámara y adaptadores para unir esta al microscopio. Dudas et al en 2014 ([Dudas et al., 2014](#)), usan una tarjeta raspberry® para construir un sistema de telecitología de bajo costo. Sin embargo, ninguna de las soluciones mencionadas permite el control robotizado del microscopio.

Guo et al en 2015 ([Guo et al., 2015](#)), presentan un sistema de placas virtuales de bajo costo, pero con el inconveniente de depender del software Image Composite Editor® para crear las imágenes fusionadas, que aunque de distribución gratuita, no es multiplataforma y necesita que la gran cantidad de imágenes sean incorporadas manualmente.

En un trabajo anterior el autor del presente trabajo desarrolló un prototipo de telemicroscopia de bajo costo basado en hardware y software libre y abierto. Los resultados indican que este prototipo es viable para desarrollar los diferentes tipos de tele patología. Su buena resolución facilita la toma de imágenes que requieran desplazamientos muy pequeños (algo necesario a grandes aumentos). Su excelente precisión, facilita la ubicación exacta de elementos en las muestras histopatológicas, dando la posibilidad de marcar zonas o incluso implementar a partir de visión por computador, la detección de alteraciones sin necesidad de operarios calificados. Además, este prototipo tiene la capacidad de realizar reconstrucciones a partir de tomas a 40X, tal como lo hacen la mayoría de los equipos comerciales. Este dispositivo puede ser alternativa para implementar sistemas de placas virtuales, programas de tele microscopia en la práctica clínica o con fines educativos. El prototipo es aplicable en las diferentes formas de tele patología y una alternativa de bajo costo, construida dentro de los principios de software y hardware libres, que favorecen su implementación, adecuación y desarrollo ([Meléndez-Álvarez et al., 2017](#)).

En el contexto de la citología cervicovaginal, la tele microscopía se convierte en una herramienta fundamental al ofrecer la capacidad de analizar muestras a distancia, superando las limitaciones geográficas y mejorando el acceso a la atención médica. Además de su aplicación en el campo de la citología, la tele microscopía tiene amplias implicaciones en diversas especialidades médicas, donde la rápida evaluación de muestras microscópicas puede ser crucial para el diagnóstico temprano y el tratamiento oportuno.

Tecnologías de hardware y software libres y abiertos en salud.

Las tecnologías de hardware y software libres y abiertos en salud se refieren al uso de recursos tecnológicos que puedan ser accedidos, modificados y distribuidos de manera libre. Esto tiene un gran impacto en el ámbito de la salud, ya que permite el acceso a tecnologías de manera más inclusiva y a bajo costo. El uso de hardware libre y abierto permite la creación de dispositivos médicos asequibles y personalizables proporcionando mayor flexibilidad y adaptabilidad en el desarrollo de aplicaciones médicas.

Las tecnologías de hardware y software libres y abiertos en salud están revolucionando la forma en que se brinda atención médica. Consecuencia de los altos costos de los instrumentos médicos creados bajo el modelo de patentes, los sistemas de asistencia sanitaria tienen un acceso limitado a las mejores herramientas para la prevención, diagnóstico y tratamiento de muchas enfermedades curables. Esto resulta en un obstáculo para lograr el acceso de la mayoría de la población a la tecnología médica y un serio problema para la salud pública, especialmente en los países en desarrollo ([Sinha & Barry, 2011](#)). La adopción de tecnologías de hardware y software libres y abiertos en el campo de la salud ofrece una solución a este desafío.

Estas tecnologías permiten el acceso a herramientas médicas de alta calidad al tiempo que fomentan la colaboración entre investigadores, médicos y desarrolladores. Al ser tecnologías

abiertas, permiten la adaptación y personalización de los dispositivos y software de acuerdo con las necesidades y recursos locales facilitando el mantenimiento y evitando la dependencia de una única casa comercial. En el contexto del proyecto de clasificación celular en la citología cervicovaginal, la utilización de tecnologías de hardware y software libres y abiertos puede proporcionar una solución rentable y accesible.

Ejemplo de los esfuerzos de la Organización Mundial de la Salud (OMS) por poder garantizar un acceso a la mayoría de la población a la tecnología en salud, es su publicación titulada: “Compendium of innovative health technologies for low-resource settings” ([WHO, 2015](#)), cuyo objetivo es proporcionar una plataforma neutral de tecnologías que puedan ser adecuadas para su uso en entornos de recursos limitados. Esta publicación presenta varias tecnologías de salud que tienen el potencial de mejorar los resultados de salud y la calidad de vida, o de ofrecer una solución a una necesidad de tecnología médica no satisfecha. Esta iniciativa, busca también fomentar una mayor interacción entre los ministerios de salud, los funcionarios de adquisiciones, los donantes, los desarrolladores de tecnología, los fabricantes, los médicos, los académicos y el público en general para asegurar una mayor inversión en tecnología sanitaria y avanzar hacia el acceso universal a las tecnologías sanitarias esenciales.

El mayor acceso a estas tecnologías sanitarias garantizará la prevención de enfermedades, el diagnóstico precoz, el tratamiento efectivo y la adecuada asistencia, acercando a los países que más lo necesitan a las metas de la estrategia de Cobertura Universal en Salud y fortaleciendo la equidad de los sistemas sanitarios ([Ayala-García, 2014](#)).

La innovación tecnológica tiene el potencial de ampliar y mejorar la equidad en la atención sanitaria a nivel global. Al mismo tiempo, las patentes de dispositivos y la estricta legislación de protección de la propiedad intelectual pueden ser barreras a la innovación para los

países en desarrollo. La regulación y legislación de la propiedad intelectual es necesaria para estimular la invención de nuevas tecnologías. Sin embargo, la protección a la propiedad intelectual en los lugares de escasos recursos ha sido históricamente escasa.

La Organización Mundial de la Propiedad Intelectual reporta que, en 2012, los países con altos ingresos tenían el 64.5% del total del número de patentes, los países de ingresos medios-bajos únicamente el 2.9% y los países de bajos ingresos tan solo el 0.4%. Esta disparidad demuestra el apoyo y desarrollo limitado de la propiedad intelectual en los países de ingresos bajos y medios, lo que los hace dependientes de tecnologías creadas en otros lugares y con costos muy altos que limitan su accesibilidad. De forma paradójica, los inventores en las zonas de bajos recursos están frente a una abundante e importante cantidad de necesidades clínicas y sociales que son espacio para la innovación tecnológica. Además, muchos de ellos están interesados en mantener sus desarrollos bajo licencia libre y abierta para maximizar el impacto y acceso a sus trabajos ([Chao & Mody, 2015](#)).

En las últimas décadas los avances de las tecnologías de la información ([Chao & Mody, 2015](#)) han tenido un fuerte impacto en la vida cotidiana de la mayoría de las personas. Estos logros se han obtenido no solo mediante la investigación y el desarrollo de nuevas tecnologías, sino también gracias a la adopción de las llamadas licencias de código abierto y libre. Las licencias de código libre y abierto están concebidas para permitir a los usuarios utilizar, copiar, modificar y redistribuir libremente bajo términos y condiciones definidas un software y su código fuente. El ejemplo más notable de software abierto y libre lo representa el sistema operativo GNU/Linux ([Linux and GNU - GNU Project - Free Software Foundation, 2021](#)), que se puede considerar probablemente como el mayor proyecto de software colaborativo de la

historia. Actualmente este sistema es seguro, ampliamente usado en servidores web, ordenadores personales, tabletas y teléfonos.

El software libre se fundamenta en respetar la libertad del usuario de utilizarlo, ejecutarlo, estudiarlo y modificarlo, además de distribuir copias con o sin modificaciones de este. El hardware y el software abiertos hacen públicos sus diseños para que cualquier persona pueda estudiar, modificar, distribuir, fabricar y también vender las innovaciones basadas en esos diseños. Este enfoque ofrece ventajas únicas como la reducción de costos y que la generación de desarrollos e innovaciones se haga de forma más rápida ([Niezen et al., 2016](#); [Pearce, 2015](#)). Además, el enfoque abierto y libre permite una mayor transparencia, seguridad y robustez al permitir que más personas inspeccionen y mejoren los diseños. ([Niezen et al., 2016](#)).

Los primeros pasos en la fabricación de tecnología médica con el enfoque del software y hardware libres ya se están dando. Una organización sin ánimo de lucro dedicada a la promoción del uso de tecnología médica abierta ha sido pionera en este campo. Han desarrollado un prototipo de bomba de infusión de código abierto que utiliza componentes fácilmente accesibles y de bajo costo. La bomba es fabricada por el 5 – 10% del costo de una bomba comercial de características similares ([Wijnen et al., 2014](#)). MyOpen, es un proyecto que quiere reducir los costos de una prótesis de brazo a \$US250, frente al costo de \$US30000 que es el de una similar de venta en el comercio actualmente. La adopción de hardware y software abiertos en el campo de la medicina y la salud puede tener un impacto importante en la accesibilidad y democratización de la atención médica.

En vista de que los enfoques convencionales en el desarrollo de tecnología en salud, bajo el modelo de patentes, no han obtenido los resultados apropiados para lograr el desarrollo, innovación y disponibilidad de acceso necesarios. Se debería considerar promover el desarrollo

de tecnología de código abierto para así facilitar alcanzar el objetivo de cubrimiento universal en salud ([Aufieri et al., 2015](#)). El uso de la impresión 3D junto a los microcontroladores de bajo costo pueden ayudar a reducir los costos de la fabricación de prototipos y facilitar la construcción de soluciones innovadoras en salud. Estas soluciones son necesarias especialmente en los países de bajos ingresos, pero no ajenas a las realidades de los países con altos ingresos.

Los médicos y en general el personal en salud, debe comprometerse con el desarrollo e investigación en estas áreas. Al colaborar con ingenieros y científicos, pueden fomentar la innovación y generar soluciones que sean accesibles y asequibles para todas las comunidades. De la misma forma que el desarrollo de fármacos es labor de los médicos, incursionar en el área de desarrollo e investigación en tecnología médica debería ser papel de todos los que integran el cuerpo del personal sanitario y en especial del que se dedica a la atención primaria en salud en países de ingresos bajos.

Metodología y desarrollo de los objetivos

Realizar una revisión exhaustiva de la literatura científica sobre la citología cervicovaginal y la detección automatizada de anomalías cervicales, con el fin de establecer una sólida base teórica y contextual.

Se realizó una revisión bibliográfica consultando las bases de datos bibliográficas SCOPUS, y PubMed (**Figuras 1, 2**). Se usaron los términos de consulta siguientes: “(pap smear AND automated) OR (pap smear AND dataset) OR (pap smear AND machine learning)”, se filtraron los periodos de años de 1974 al 2024 en SCOPUS y de 1990 al 2024 en PubMed, no se filtró por tipos de trabajos. La búsqueda arrojó un total de 610 documentos en la base SCOPUS y 266 en PubMed. Se juntaron los resultados de las dos bases de datos y se eliminaron los datos que se sobreponen o repiten entre las dos. Para realizar este procedimiento se usó el software R ([Buitinck et al., 2013](#); [Team, 2023](#)) y la librería bibliometrix y biblioshiny ([Aria & Cuccurullo, 2017](#)). Después de eliminar los datos repetidos se obtuvieron 760 documentos (**Figura 3**).

Figura 1.

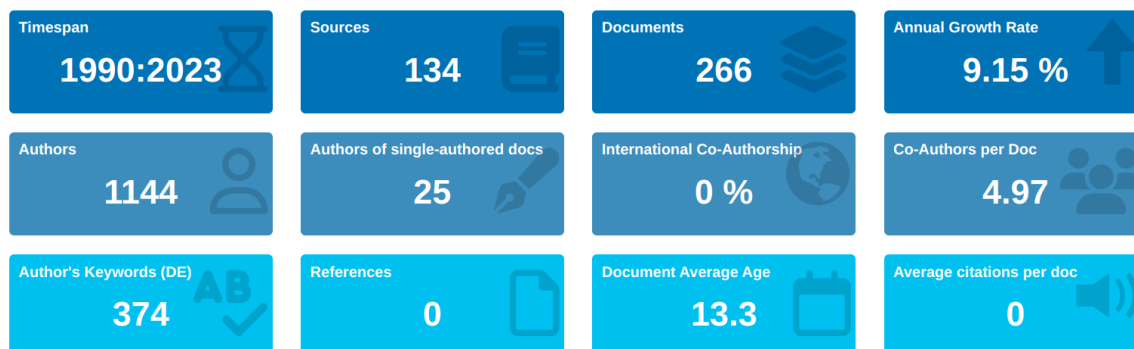
Imagen que resume los resultados obtenidos en la base de datos SCOPUS.



Fuente. Autor (creado en software R.)

Figura 2.

Imagen que resume los resultados obtenidos en la base de datos MEDLINE a través de PubMed.



Fuente. Autor (creado en software R.)

Figura 3.

Imagen que resume los resultados después de unir los resultados de las bases de datos SCOPUS y MEDLINE.



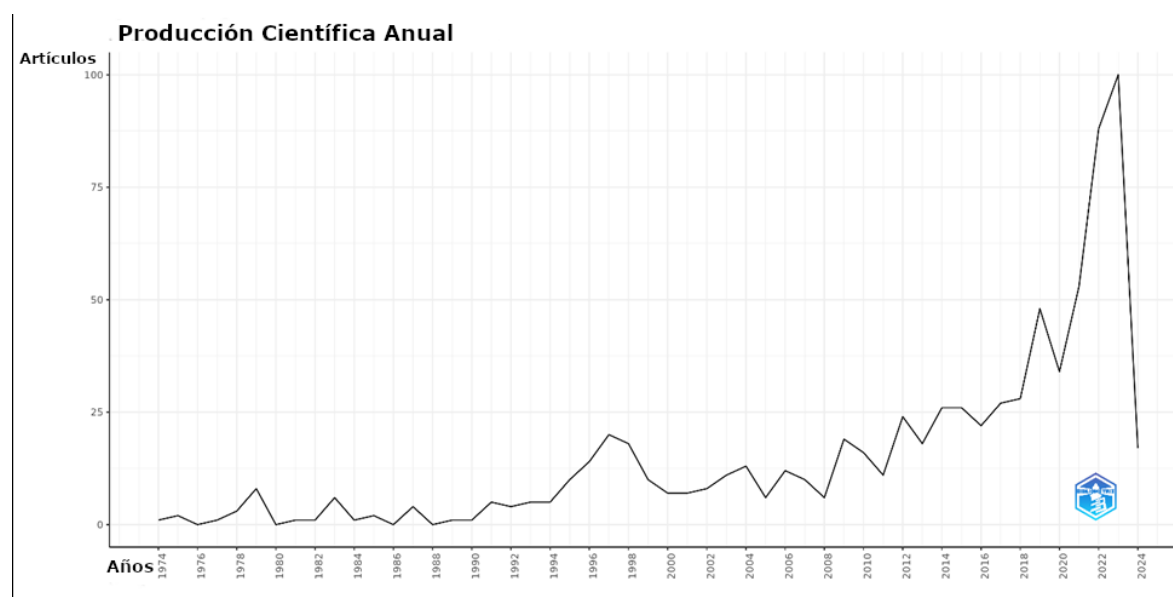
Fuente. Autor (creado en software R.)

De los documentos de la búsqueda se eligieron para revisión los 30 artículos con mayor impacto, tomando el número de citas que ha generado cada documento. De estos 30 artículos, 27 corresponden a artículos originales, 1 a documentos de conferencias y 2 a trabajos de revisión. Los países donde la producción es más importante son USA, China, India, Suecia, Brasil y Australia, respectivamente.

Posteriormente se revisaron artículos que fueron seleccionados por tema de interés y se incluyeron 11 artículos más, que son artículos de revisión que fueron publicados en los últimos 3 años, para incluir estos trabajos no se tuvo en cuenta el número de citas generadas, esto teniendo en cuenta la creciente producción respecto a este tema en los últimos años (**Figura 4**).

Figura 4.

Se muestra la evolución de la producción bibliográfica a lo largo del tiempo. Se observa el considerable aumento en la producción anual durante los últimos 10 años.



Fuente. Autor (creado en software R.)

Recolectar y preprocesar un conjunto de datos de alta calidad que contenga imágenes de células cervicales y metadatos pertinentes, garantizando la uniformidad de las características a emplear en la construcción del modelo.

Posterior a la revisión de la literatura y la búsqueda explícita de dataset públicos de citología cervicovaginal se logró acceder a 13 conjuntos de datos. Estos se examinaron buscando que cumplieran con el uso de la clasificación de Bethesda, que es la usada de manera habitual y más frecuentemente por los patólogos para etiquetar las células citológicas y que contaran con metadatos suficientes para poder evaluar características de los núcleos celulares que puedan ser tenidas en cuenta al desarrollar el modelo de clasificación.

De estos tan solo 2 cumplen con el criterio de usar la clasificación de Bethesda para etiquetar las células, Herlev que es el más antiguo y el más usado en estudios de clasificación y segmentación. Este dataset es una colección de 917 células etiquetadas en las 7 clases que usa la clasificación Bethesda. Cuenta con la medición de 20 características celulares que son:

1. El área del núcleo. (Kerne_A)
2. El área del citoplasma. (Cyto_A)
3. La razón entre el área del núcleo y el área del citoplasma. (K.C)
4. La intensidad del brillo del núcleo. (Kerne_Ycol)
5. La intensidad del brillo del citoplasma. (Cyto_Ycol)
6. El diámetro menor del núcleo. (KerneShort)
7. La elongación del núcleo (KerneElong): este parámetro se calcula como la relación entre el diámetro más corto y el diámetro más largo del núcleo.
8. La redondez del núcleo (KerneRund): se calcula como la relación entre el área real y el área delimitada por el círculo dado por el diámetro más largo del núcleo.

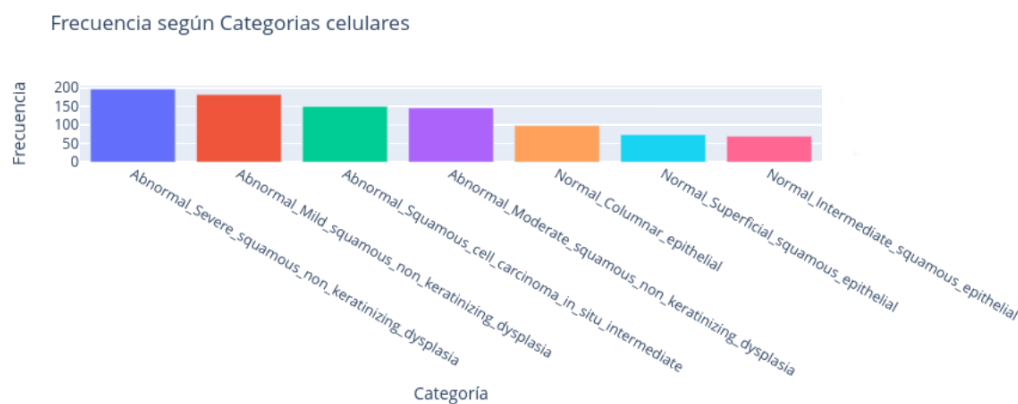
9. El diámetro menor del citoplasma. (CytoShort)
10. El diámetro mayor del citoplasma. (CytoLong)
11. La elongación del citoplasma (CytoElong): se calcula como la relación entre el diámetro más corto y el diámetro más largo del citoplasma.
12. La redondez del citoplasma (CytoRound): se calcula como la relación entre el área real y el área delimitada por el círculo dado por el diámetro más largo del citoplasma.
13. El perímetro del núcleo. (KernePeri)
14. El perímetro del citoplasma. (CytoPeri)
15. La posición del núcleo (KernePos): esta es una medida de que tan bien está centrado el núcleo en el citoplasma, se calcula encontrando la distancia entre el centro del núcleo y el centro del citoplasma.
16. El kernelMax: el número de píxeles cuyo valor es máximo o mínimo dentro de un radio de 3 píxeles alrededor de cada punto en la imagen. Este tipo de operación es común en el procesamiento de imágenes y se utiliza para realizar operaciones locales o vecinales. Para cada píxel en la imagen se examina un área circular de 3 píxeles alrededor de él, y se cuenta cuántos de esos píxeles tienen el valor máximo o mínimo.
17. El KernelMin: en relación a KernelMax pero midiendo el mínimo.
18. CytoMax: el número de píxeles cuyo valor es máximo o mínimo dentro de un radio de 3 píxeles alrededor de cada punto en la imagen. Este tipo de operación es común en el procesamiento de imágenes y se utiliza para realizar operaciones locales o vecinales. Para cada píxel en la imagen se examina un área circular de 3 píxeles alrededor de él, y se cuenta cuántos de esos píxeles tienen el valor máximo o mínimo.
19. CytoMin: en relación a CytoMax pero midiendo el mínimo.

20. La categoría celular. (Class_catég)

El segundo dataset es CRIC que, aunque usa la clasificación de Bethesda, tiene etiquetados los núcleos celulares tan solo por un punto aproximado al centro del núcleo. No cuenta con datos de identificación del citoplasma ni otras características celulares salvo la categoría de clasificación.

Figura 5.

Frecuencia absoluta según tipo celular. Dataset Herlev.



Fuente. Autor

La figura 5, muestra que la categoría de células más común es la de Displasia escamosa no queratinizada severa, con un total de 197 células. Le sigue la categoría de Displasia escamosa no queratinizada leve, con un total de 182 células. Carcinoma in situ es la tercera categoría más común, con un total de 150 células.

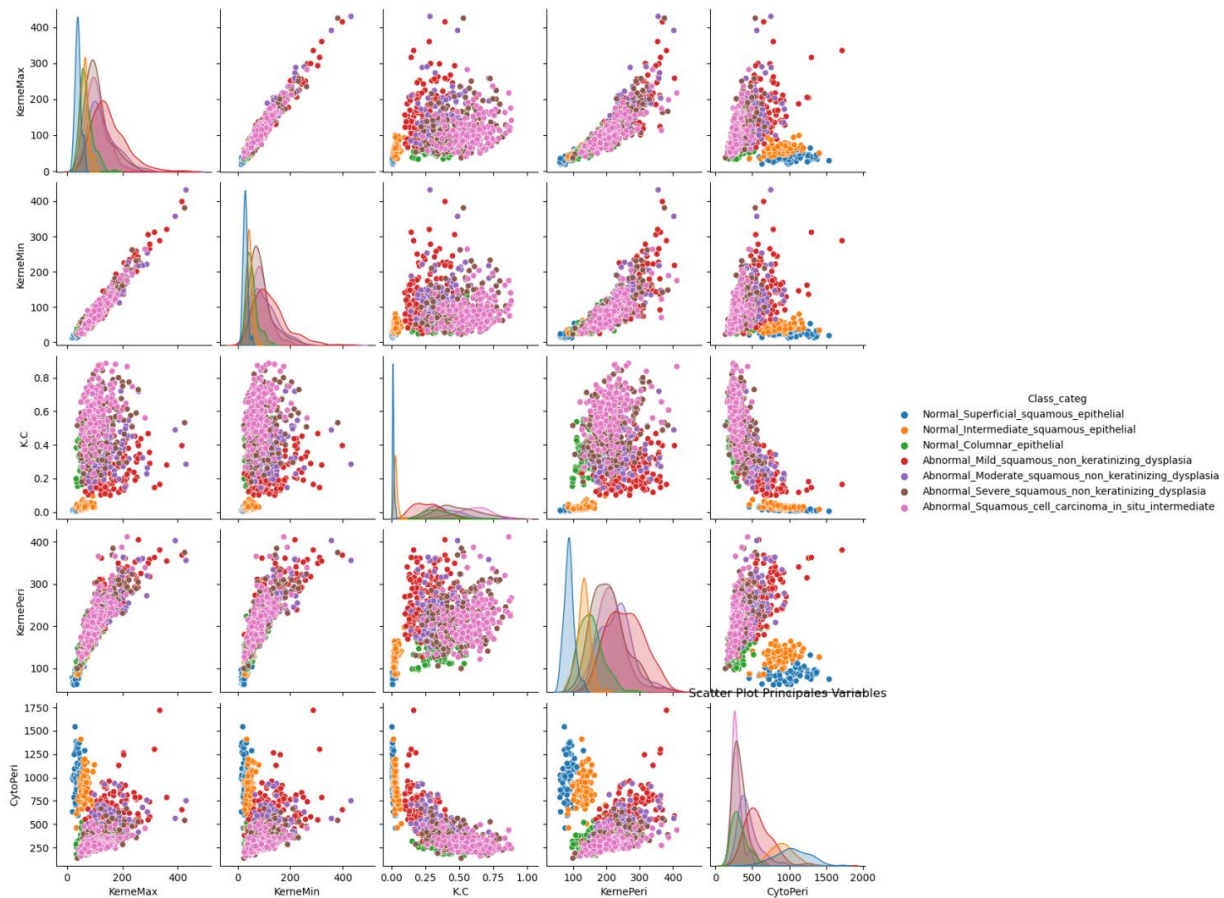
Las categorías de células con menor número es la de Células intermedias escamosas epiteliales y Células escamosas superficiales. Cada una de estas categorías tiene 70 y 74 células respectivamente.

Se evidencia que el dataset tiene un mayor porcentaje de células en las categorías consideradas anormales. Este grupo, de células anormales, tiene entre 146 y 197 células en las categorías de Displasia no queratinizada Moderada y Displasia no queratinizada Severa respectivamente. Las categorías celulares consideradas normales, tienen entre 70 y 98 células, pertenecientes a el grupo de Células escamosas epiteliales intermedias y Columnar epitelial normal respectivamente.

La figura 6, muestra la relación de las variables del dataset Herlev en gráficos scatterplot, se puede evidenciar que una de las variables que mejor puede permitir clasificar las células es el perímetro del núcleo, (**Figuras 6, 7**). Aunque se sobreponen bastante los datos de las siete categorías evidenciando la dificultad para encontrar una característica que diferencie fácilmente las clases.

Figura 6.

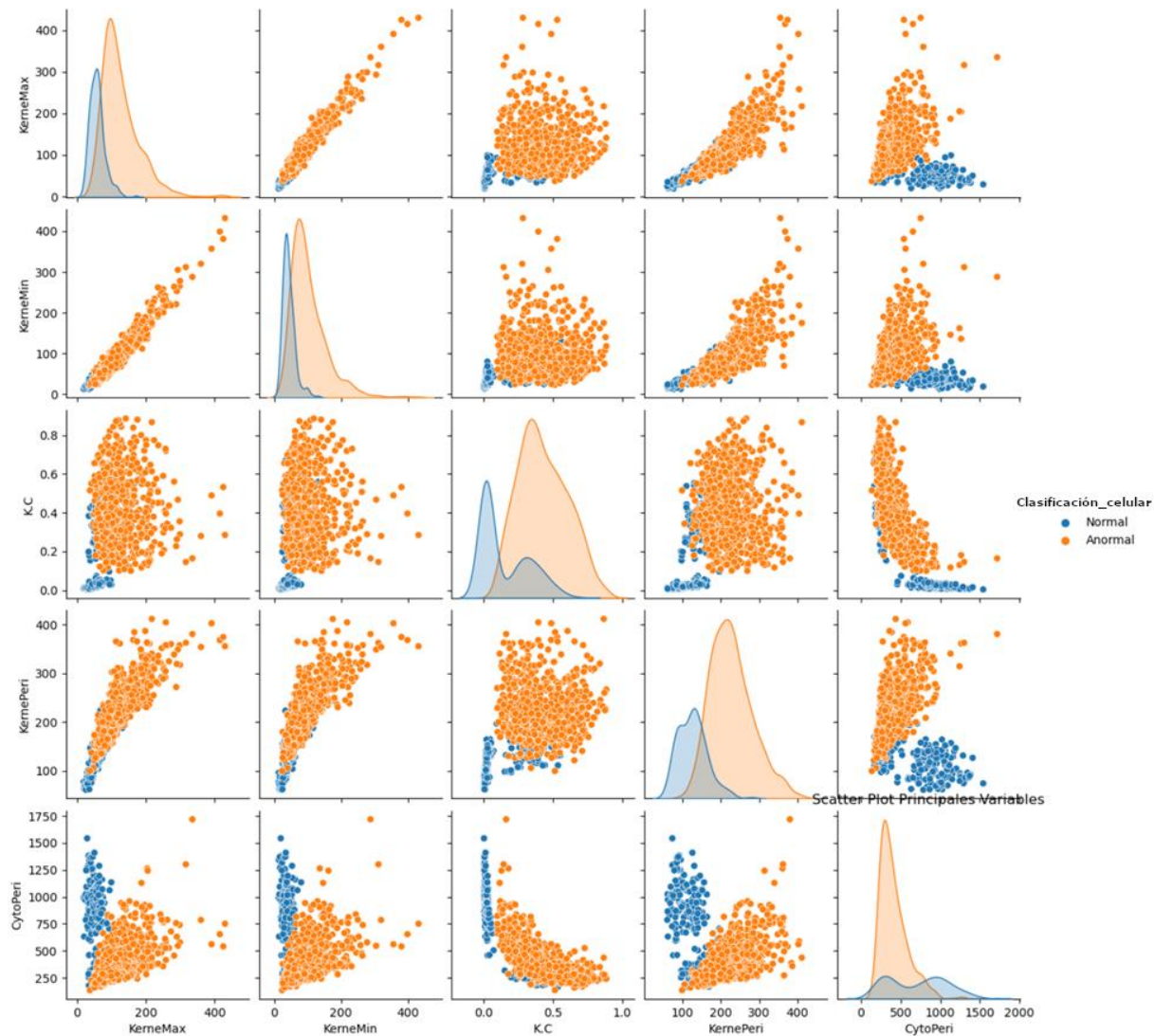
Scatterplot que muestra la relación de las variables del dataset Herlev según su categoría de clasificación.



Fuente. Autor (elaborado con el software Python.)

Figura 7.

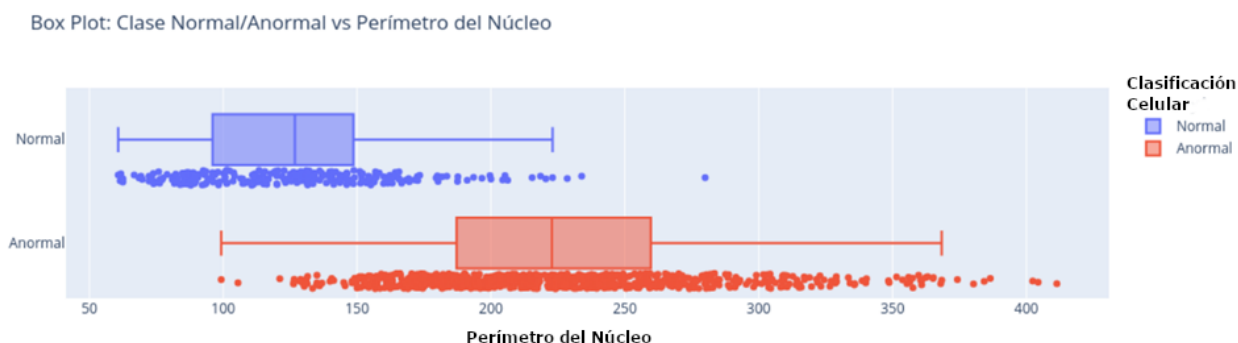
Scatterplots que muestran las relaciones de las variables del dataset Herlev según solo dos categorías de clasificación celular.



Fuente. Autor (elaborado con el software Python.)

Figura 8.

Diagrama de caja que muestra la distribución de los datos del perímetro del núcleo en las categorías celulares Normal y Anormal del dataset Herlev.



Fuente. Autor (elaborado con el software Python.)

La **Figura 8**, muestra diagramas de caja del perímetro del núcleo celular en relación con las categorías normal y anormal. Se evidencia que la mediana y la concentración del 50% de los datos de cada categoría están distantes entre cada categoría.

Con el propósito de consolidar un conjunto de datos de alta calidad a partir de los dataset CRIC y Herlev y teniendo presente lo analizado de las características relevantes para la clasificación celular en el dataset Herlev y la importancia de delimitar el núcleo celular de la manera más precisa. Con el propósito de poder calcular características relevantes como el perímetro del núcleo y teniendo presente la ausencia de datos que permitan etiquetar los núcleos en el dataset CRIC de mejor manera que con la técnica de “bounding box” o cuadro delimitador. Se calcularon los radios de los núcleos para las categorías normal y anormal en el dataset Herlev, estos radios se usaron para ser extrapolados a el dataset CRIC y generar etiquetas con polígonos de 8 puntos alrededor del dato del centro del núcleo y la clase celular. Este procedimiento se realizó implementando un script que realiza los cálculos en Python ([Rossum & Drake, 2009](#)) para generar los polígonos aproximados en los núcleos del conjunto de datos CRIC.

Posteriormente estos datos fueron incorporados a el software labelme, software abierto y gratuito que permite etiquetar imágenes con las técnicas de bounding box o polygon ([Kentaro, 2016](#)), y exportar estas etiquetas para ser usadas en diferentes modelos de redes neurales convolucionales como YOLO ([Serrano, 2017](#)). Para poder incorporar los datos de etiquetado del dataset CRIC a sus 400 imágenes fue necesario crear 400 archivos “.json” con la estructura sintáctica requerida e incorporar los datos de los puntos calculados de los polígonos para cada una de las 11.534 células de este dataset.

Esto permitió generar etiquetas aproximadas de los núcleos celulares para la totalidad del dataset CRIC. Posteriormente se realizó la corrección manual de todas las etiquetas de los núcleos en el dataset, rectificando a partir de las etiquetas preliminares nuevas etiquetas que son reescritas mediante el software labelme. Este proceso permitió que en el dataset CRIC se calcularan variables morfológicas como el perímetro del núcleo y de textura como el color y el brillo en diferentes espacios de color como HSV y RBG utilizando para esto la librería OpenCV ([Bradski, 2000](#)).

Este proceso se convierte en un gran aporte al consolidar un dataset con etiquetas que permitan realizar mejores y más precisas mediciones, que van desde la búsqueda de características de clasificación celular hasta su uso en la creación de sistemas y modelos de segmentación nuclear.

La **Figura 9**, muestra la estructura de los archivos “.json” creados con los puntos para delimitar el polígono que rodea al núcleo celular y conforman los datos para la etiqueta calculada, estos polígonos de 8 puntos permiten ser adaptados fácilmente a los diferentes contornos y tamaños de los núcleos celulares. La **Figura 10**, muestra el proceso de re-etiquetado en el software labelme y permite visualizar las etiquetas preliminares calculadas. Los puntos de

los polígonos fueron calculados a partir del radio aproximado asignado a cada célula del dataset CRIC, de acuerdo con la etiqueta de normal o patológica y de la estimación de este a partir de los datos promedio registrados en el conjunto de datos Herlev para cada categoría. Este proceso fue desarrollado mediante un script en Python ([Rossum & Drake, 2009](#)).

Figura 9.

Muestra de uno de los archivos “json” con la información para una de las imágenes del dataset CRIC donde se muestran las coordenadas calculadas para uno de los núcleos celulares.

The screenshot shows a Geany text editor window with a JSON file open. The file path is `3f30a996dcac3c_650fe874b5.json`. The JSON content is as follows:

```

1 {
2   "version": "4.5.7",
3   "flags": {},
4   "shapes": [
5     {
6       "label": "4260",
7       "points": [
8         [
9           622.7692307692307,
10          65.38461538461539
11         ],
12         [
13           652.8386179965449,
14           118.61783557599418
15         ],
16         [
17           656.6153846153846,
18           151.53846153846155
19         ],
20         [
21           633.5326582159443,
22           169.46875848420754
23         ],
24         [
25           593.5384615384615,
26           174.61538461538458
27         ],
28         [
29           584.3076923076923,
30           148.46153846153845
31         ],
32         [
33           584.3076923076923,
34           128.76923076923077
35         ],
36         [
37           593.5384615384615,
38           90.0
39         ]
40       ],
41       "group_id": null,
42       "shape_type": "polygon",
43       "flags": {}
44     }
45   ]
46 }

```

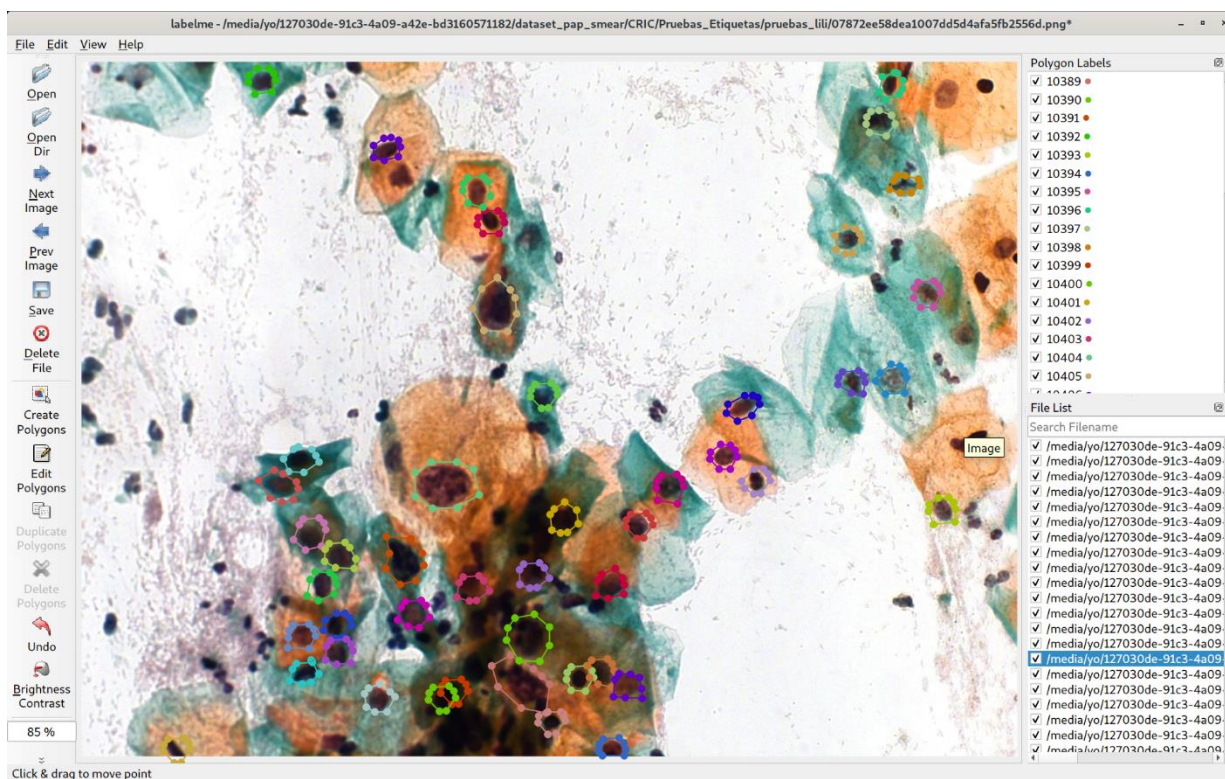
The status bar at the bottom indicates the file is open and the editor version is Geany 1.37.1.

Fuente. Autor

La **Figura 10**, muestra una de las imágenes del conjunto de datos CRIC, visualizada en el software lableme, en ella se ven los polígonos creados para cada núcleo. El proceso de delimitación de los núcleos celulares por los polígonos se rectificó para cada imagen de las 400 y para todos los núcleos del dataset CRIC de manera manual posterior a la creación automatizada de dichos polígonos con el script creado.

Figura 10.

Se muestra una de las imágenes del dataset CRIC con la visualización de las etiquetas calculadas y el proceso de ajuste de estas etiquetas en el software labelme.



Fuente. Autor

Una vez se marcaron con los polígonos todos los núcleos celulares del dataset CRIC, se procedió a realizar la medición de características para cada núcleo y consolidar esa información en un nuevo conjunto de datos que complementó el original. Se midió el brillo, calculado como la sumatoria de los promedios de cada color por el factor así: $\text{Rojo} * 0.22 + \text{Verde} * 0.587 + \text{Azul} * 0.114$, además se calcularon los valores promedio para el rojo, verde y azul. En el modelo de color HSV (Matiz, Saturación, Valor) se calcularon los valores promedio para cada componente. Finalmente se calculó el perímetro a partir de los puntos que conforman cada polígono.

Como resultado se consolidó un nuevo dataset que incluye esta nueva información calculada, y que también contiene las coordenadas de los polígonos que delimitan cada núcleo celular. Se adicionaron, además, columnas con categorías de normal y anormal, una versión resumida de la clasificación de Bethesda que solo tiene 3 clases: Lesiones de bajo grado, Lesiones de alto grado y Normal. El nuevo dataset quedó configurado con las 18 siguientes variables:

1. **imagePath:** corresponde a la ruta de la imagen que contiene la célula.
2. **image_id:** es el número de identificación de cada imagen del dataset.
3. **image_doi:** es un identificador para cada imagen que es asignado por los creadores del dataset CRIC.
4. **cell_id:** es el número de identificación de cada célula en el conjunto de datos.
5. **bethesda_system:** corresponde a la categoría en la que la célula se encuentra dentro de las 5 categorías del sistema Bethesda.
6. **nucleus_x:** representa las coordenadas en el eje x del punto que identifica el centro del núcleo celular.
7. **nucleus_y:** representa las coordenadas en el eje y del punto que identifica el centro del núcleo celular.
8. **Class_NyA:** corresponde a la clasificación de normal o anormal para cada célula.
9. **Short_Bethesda:** es la categoría a la que corresponde de acuerdo a la clasificación Bethesda con solo 3 clases.
10. **points:** contiene las coordenadas de los 8 puntos que conforman el polígono que delimita cada núcleo celular.

11. **perimetro:** representa el cálculo del perímetro en micras calculado para cada núcleo.
12. **Hue:** es el valor promedio calculado para el matiz, dentro del área delimitada para el núcleo, en el modelo de color HSV.
13. **Saturation:** es el valor promedio calculado para la saturación, dentro del área delimitada para el núcleo, en el modelo de color HSV.
14. **Value:** es el valor promedio calculado para el Value, dentro del área delimitada para el núcleo, en el modelo de color HSV.
15. **R:** es el valor promedio calculado para el color rojo, dentro del área delimitada para el núcleo, en el modelo de color RGB.
16. **G:** es el valor promedio calculado para el color verde, dentro del área delimitada para el núcleo, en el modelo de color RGB.
17. **B:** es el valor promedio calculado para el color azul, dentro del área delimitada para el núcleo, en el modelo de color RGB.
18. **Brillo:** es el valor promedio calculado para el brillo, dentro del área delimitada para el núcleo, en el modelo de color RGB según el siguiente cálculo: $R*0.22 + G*0.587 + B*0.114$

La **Tabla 1**, contiene los datos de las frecuencias y porcentajes en el nuevo dataset de acuerdo con la categoría Normal y Anormal. Se observa que predominan las células normales con el 58% de los datos correspondientes a 6779 células. La categoría anormal representa el 41% de los datos con 4755 células.

Tabla 1.

Frecuencia y porcentaje de las categorías normal y anormal.

Clasificación_celular	Porcentaje	Frecuencia
Normal	0.587741	6779
Anormal	0.412259	4755

Nota. La tabla presenta la distribución de las clases Normal y Anormal. *Fuente.* Autor

La **Tabla 2**, presenta la frecuencia y el porcentaje de las cinco clases celulares según el sistema Bethesda. Se observa que la clase normal (Negative for intraepithelial lesion) es la que contiene la mayoría de los datos correspondiendo al 58% (6779) de los datos. La siguen en orden descendente las categorías HSIL con el 14% (1703), LSIL con el 11% (1360), ASC-H con el 8% (925), ASC-US con el 5% (606) y SCC con el 1% (161) de los datos.

Tabla 2.

Frecuencia y porcentaje de las 5 categorías del sistema Bethesda.

Clasificación_celular	Porcentaje	Frecuencia
Negative for intraepitelial lesion	0.587741	6779
HSIL	0.147650	1703
LSIL	0.117912	1360
ASC-H	0.080198	925
ASC-US	0.052540	606
SCC	0.013959	161

Nota. La tabla presenta la distribución de las seis clases del sistema Bethesda. *Fuente.* Autor

Tabla 3.

Frecuencia y porcentaje de las 3 categorías del sistema Bethesda resumido.

Short_bethesda_system	Porcentaje	Frecuencia
Normal	0.587741	6779
High-grade-Lesion	0.241807	2789
Low-grade-Lesion	0.170453	1966

Nota. La tabla presenta la distribución de las tres clases del sistema Bethesda resumido. *Fuente.*

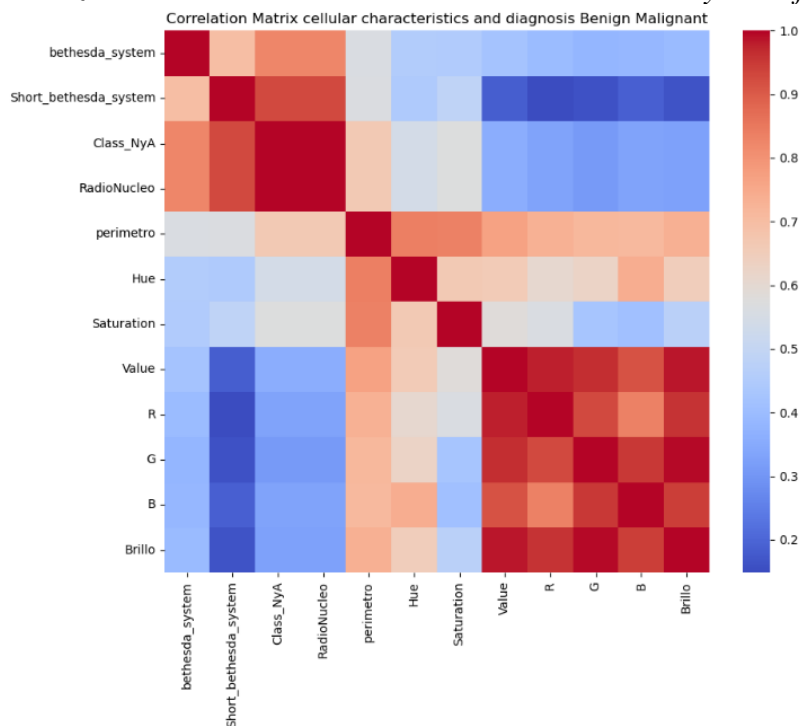
Autor

La **Tabla 3**, contiene las frecuencias y porcentaje de los datos de las tres categorías del sistema de Bethesda resumido. La clase normal es la que mayor número de datos contiene con 6779 que corresponden al 58% del total de datos. Las clases High-grade-Lesion (Lesión de alto grado) y Low-grade-Lesion (Lesión de bajo grado) contienen el 24% y 17 % respectivamente. En esta clasificación se han agrupado las clases ASC-US, LSIL como Low-grade-Lesion y las clases HSIL, SCC y ASC-H como High-grade-Lesion.

La **Figura 11**, muestra la matriz de correlación de las características del núcleo y las diferentes clases diagnósticas. Resalta la asociación directa que presenta el perímetro del núcleo con la clasificación de normal y anormal. En menor correlación en esta misma clasificación aparecen las variables Value, R, G, B y brillo. La clasificación del Bethesda resumido presenta mayor asociación con la variable perímetro y correlaciones más bajas y homogéneas con las variables Value, R, G, B y brillo. La clasificación del sistema Bethesda presenta correlaciones bajas con la mayoría de las variables siendo la que mayor correlación presenta la del perímetro.

Figura 11.

Matriz de correlación de las características nucleares y las diferentes clases diagnósticas.

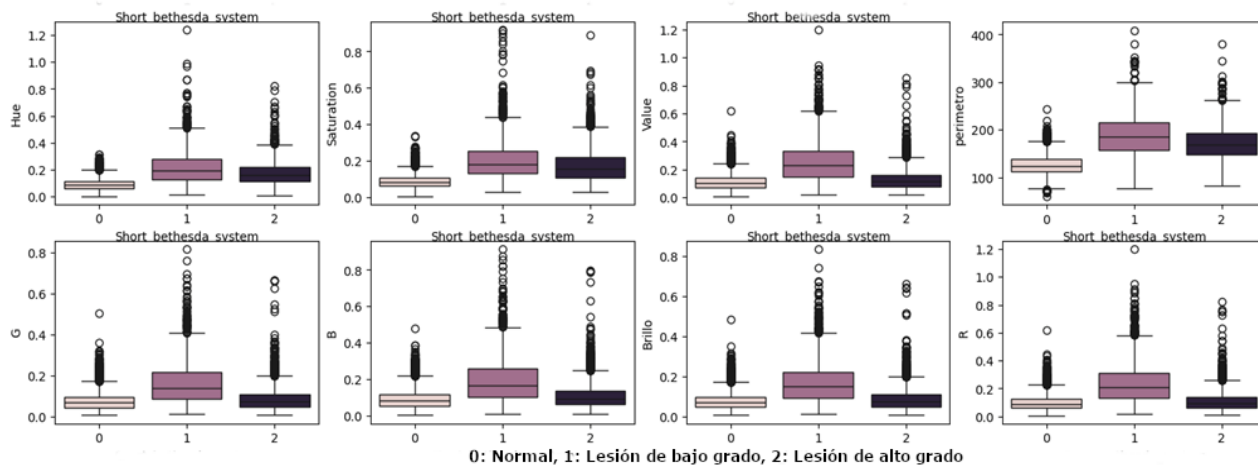


Fuente. Autor

La **Figura 12**, presenta diagramas de caja con las distribuciones de las principales variables en relación a las tres categorías del sistema Bethesda resumido, 0 indica la categoría Normal (No intraepithelial lesion), 1 las lesiones de bajo grado (Low-grade-Lesion) y 2 las lesiones de alto grado (High-grade-Lesion). Podemos observar que la variable que mejor diferencia las tres clases es el perímetro. Con las demás variables las clases 0 y 2 se superponen bastante.

Figura 12.

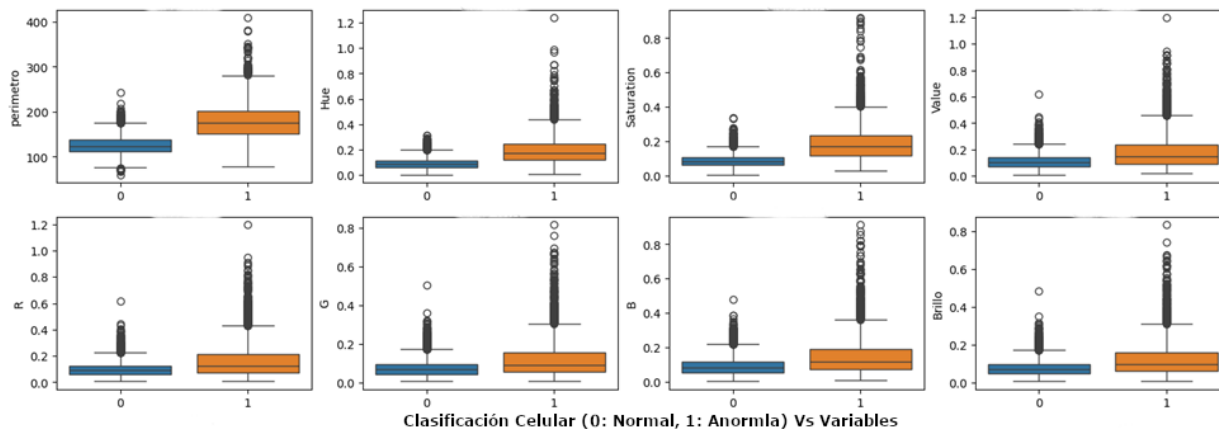
Diagramas de cajas de las principales variables en relación con las tres clases del sistema Bethesda resumido.



Fuente. Autor

Figura 13.

Diagramas de cajas de las principales variables en relación con las clases normal y anormal.



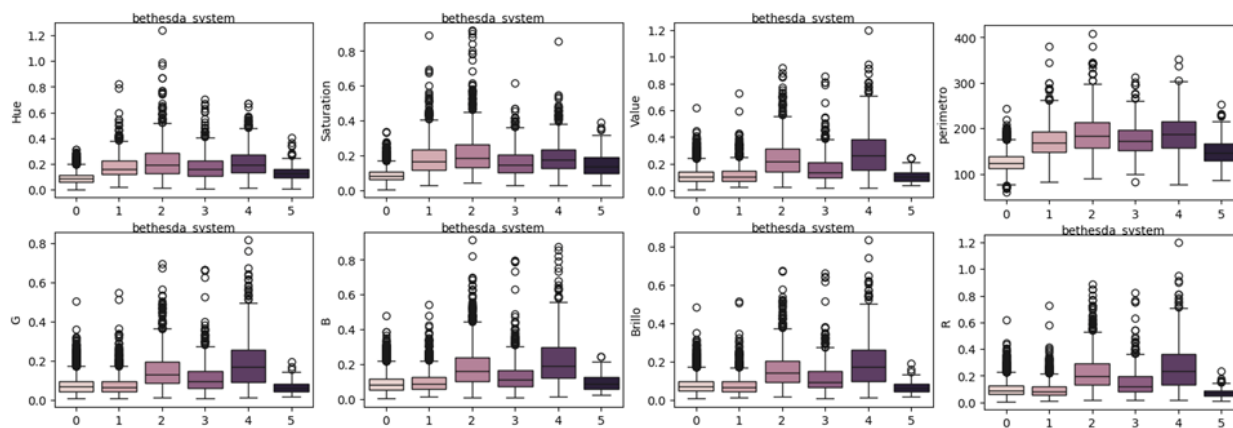
Fuente. Autor

La **Figura 13**, presenta diagramas de cajas para las variables en relación con la clasificación en dos clases normal y anormal. Se observa que las mejores variables que diferencian las clases son el perímetro y la saturación.

La **Figura 14**, presenta diagramas de caja para las principales variables en relación con las 5 clases del sistema Bethesda. Se evidencia que no existe una variable que diferencie fácilmente las 5 clases. Las clases se sobrepone en su distribución en la mayoría de los casos. Sin embargo, las variables perímetro y saturación parecen ser las que más se diferencian en las 5 clases. En la gráfica, el 0 representa la clase negativa, 1 la clase HSIL, 2 la clase LSIL, 3 a la clase ASCH, 4 a la clase ASC-US y 5 a la clase SCC.

Figura 14.

Diagramas de cajas de las principales variables en relación con las 5 clases del sistema Bethesda. 0: Normal, 1: HSLI, 2: LSLI, 3:ASC-H, 4:ASC-US, 5: SCC.



Fuente. Autor

Plantear un prototipo de modelo de aprendizaje automático que sea capaz de clasificar células cervicales en categorías relevantes, como células normales, células precancerosas o células cancerosas.

Evaluar el desempeño del modelo utilizando métricas de evaluación específicas para la clasificación.

Utilizando los datos consolidados en el nuevo conjunto de datos, se procedió a entrenar modelos de aprendizaje automático mediante diversos algoritmos, tales como K-nn, SVM, Random Forest y XGBoost. Se emplearon tanto todas las variables disponibles como aquellas seleccionadas que demostraron una mayor capacidad para diferenciar entre los distintos grupos de clase utilizados.

Estos modelos fueron entrenados con el objetivo de abordar problemas de clasificación que implicaban 2, 3 y 5 clases. Además, se construyeron modelos de ensamble utilizando el método de votación dura, combinando los modelos individuales que arrojaron los mejores resultados.

Para llevar a cabo la evaluación de los modelos, el conjunto de datos se dividió en conjuntos de entrenamiento y prueba en una proporción de 7:3. Este proceso se aplicó de manera consistente a todos los modelos entrenados y validados.

La **Tabla 4**, muestra los resultados de las métricas para los diferentes modelos entrenados con dos clases normal y anormal. Los mejores resultados los tiene el método de ensamble, con resultados discretamente mayores que los métodos de manera individual. Las cifras de sensibilidad son del 87%, las de especificidad de 94% y las de exactitud de 87%. Para entrenar los modelos en resolver este problema de clasificación binaria se usaron cuatro características del

núcleo: perímetro, valor promedio de la saturación, el brillo y hue (matiz). Este grupo de características mostró los mejores resultados.

Tabla 4.

Resultado de las métricas de los modelos K-nn, SVM, Random Forest, XGBoost y Ensemble, con el uso de las clases normal y anormal.

Modelo	VARIABLES USADAS	Accuracy	Precision	Recall	F1 Score	Specificity
K-nn		0.86	0.87	0.86	0.86	0.94
SVM	Perímetro, Saturación, Brillo, Hue.	0.87	0.87	0.87	0.87	0.91
Random Forest		0.87	0.87	0.87	0.86	0.92
XGBoost		0.86	0.86	0.86	0.86	0.95
Ensamble		0.87	0.87	0.87	0.86	0.94

Nota. K-nn: k-nearest neighbors, método de los k vecinos más cercanos. SVM: support-vector machines, máquinas de vectores de soporte o máquinas de vector soporte. Random Forest: Bosque aleatorio, algoritmo que se basa en árboles de decisión. XGBoost: eXtreme Gradient Boosting, es una biblioteca de software de código abierto usada en problemas de aprendizaje supervisado. Ensemble: Método de ensamble, consiste en combinar múltiples modelos base.
Fuente. Autor

La **Tabla 5**, reporta los resultados de las métricas aplicadas a los diferentes modelos. Nuevamente el método de ensamble logra los mejores valores en exactitud con un 81%, especificidad del 97% y sensibilidad del 81%. En este caso se usaron las 8 características medidas ya que los mejores resultados se lograron con esta combinación de variables para esta clasificación con tres clases.

Tabla 5.

Resultado de las métricas de los modelos K-nn, SVM, Random Forest, XGBoost y Ensemble, con el uso de las clases Normal, Low-grade-Lesion y High-grade-Lesion.

Modelo	VARIABLES USADAS	Accuracy	Precision	Recall	F1 Score	Specificity
K-nn		0.80	0.80	0.80	0.80	0.73
SVM	Perimetro, Saturación, Brillo, Hue, Value, R,G,B	0.80	0.81	0.80	0.81	0.86
Random Forest		0.81	0.80	0.81	0.80	0.77
XGBoost		0.80	0.80	0.80	0.80	0.75
Ensamble		0.81	0.80	0.81	0.80	0.97

Nota. K-nn: método de los k vecinos más cercanos. SVM:máquinas de vectores de soporte. Random Forest: Bosque aleatorio. XGBoost: biblioteca de software usada en problemas de aprendizaje supervisado. Ensemble: Método de ensamble, consiste en combinar múltiples modelos base. *Fuente.* Autor

La **Tabla 6**, tiene los resultados de las métricas para la clasificación de las 5 clases del sistema Bethesda. En esta ocasión los modelos tienen resultados similares en todos los casos, levemente superior el modelo de ensamble. La eficiencia de todos los modelos en este caso es inferior al 80%, en todas las métricas.

Tabla 6.

Resultado de las métricas de los modelos K-nn, SVM, Random Forest, XGBoost y Ensemble, con el uso de las 5 clases del sistema Bethesda.

Modelo	VARIABLES USADAS	Accuracy	Precision	Recall	F1 Score	Specificity
K-nn		0.73	0.69	0.73	0.69	0.68
SVM		0.73	0.69	0.73	0.69	0.68
Random Forest	Perimetro, Saturación, Brillo, Hue, Value, R, G, B	0.73	0.70	0.73	0.70	0.73
XGBoost		0.73	0.68	0.73	0.69	0.72
Ensamble		0.73	0.70	0.73	0.70	0.71

Nota. K-nn: k-nearest neighbors, método de los k vecinos más cercanos. SVM: support-vector machines, máquinas de vectores de soporte o máquinas de vector soporte. Random Forest: Bosque aleatorio, algoritmo que se basa en árboles de decisión. XGBoost: eXtreme Gradient Boosting, es una biblioteca de software de código abierto usada en problemas de aprendizaje supervisado. Ensemble: Método de ensamble, consiste en combinar múltiples modelos base.
Fuente. Autor

Conclusiones

Los modelos de clasificación binaria son los que mejores resultados tienen en las métricas, el método de ensamble sobresale en términos de rendimiento, superando ligeramente a los modelos individuales en términos de sensibilidad, especificidad y exactitud. Estos resultados sugieren que la combinación de varios modelos puede mejorar la capacidad de clasificación en problemas de dos clases, destacando la utilidad de esta estrategia en la detección de anomalías cervicales.

El uso de cuatro características del núcleo, a saber, perímetro, valor promedio de la saturación, brillo y hue (matiz), ha demostrado ser efectivo en la clasificación binaria. Esta selección de características mostró los mejores resultados en términos de métricas de rendimiento, lo que resalta la relevancia de estas variables en la detección precisa de anomalías cervicales.

Los modelos de clasificación multiclase también muestran que el método de ensamble es el más efectivo en términos de exactitud, especificidad y sensibilidad, especialmente en la clasificación a tres clases. Esto sugiere que la capacidad de generalización del método de ensamble es notable, incluso cuando se enfrenta a problemas de clasificación más complejos.

La clasificación de cinco clases del sistema Bethesda presenta los mayores desafíos. La eficiencia de todos los modelos es inferior al 80% en todas las métricas. Esto indica que la clasificación de múltiples clases presenta desafíos adicionales en comparación con la clasificación binaria o de tres clases, y destaca la necesidad de mejorar la precisión y robustez de los modelos en este contexto.

Recomendaciones

Al iniciar el trabajo, el principal desafío fue obtener un conjunto de datos suficiente y completamente etiquetado que representara adecuadamente todas las clases celulares. Para abordar este problema, se complementó el mejor y más reciente conjunto de datos existente mediante el etiquetado manual de los núcleos celulares. Este largo proceso permitió crear un dataset de mayor calidad, lo cual fue fundamental para plantear y construir el modelo presentado.

La limitación principal radica en las bajas métricas que el modelo obtiene en el problema de clasificación multiclase del sistema Bethesda. Para futuros trabajos, se sugiere incorporar el balanceo de datos como una estrategia para mejorar la robustez y la generalización de los modelos de clasificación celular. El balanceo de datos podría ayudar a abordar desafíos relacionados con la desproporción entre clases y mejorar la capacidad del modelo para detectar con precisión las diferentes clases del sistema usadas por el sistema Bethesda.

La colaboración interdisciplinaria entre profesionales de la salud, científicos de datos y desarrolladores de tecnología será fundamental para avanzar en la implementación efectiva de sistemas de telemicroscopía de bajo costo y mejorar la accesibilidad y precisión de los servicios citológicos.

Los resultados obtenidos en las clasificaciones binaria y de tres clases muestran un desempeño favorable, lo cual sugiere su relevancia para aplicaciones clínicas. Se recomienda llevar a cabo estudios adicionales para evaluar su efectividad en contextos clínicos reales. Estos estudios podrían proporcionar una comprensión más profunda de cómo estas clasificaciones pueden integrarse en la práctica clínica, así como su capacidad para mejorar la precisión diagnóstica, guiar decisiones terapéuticas y mejorar los resultados del paciente. Además, se

sugiere realizar análisis comparativos con métodos convencionales para validar su superioridad y establecer su viabilidad en entornos clínicos diversos.

Un gran desafío que puede ser abordado en trabajos futuros es la segmentación del núcleo celular. Obtener los núcleos celulares de manera adecuada, como se demostró en el presente trabajo, permite crear un modelo con métricas adecuadas de precisión y exactitud. Sin embargo, realizar la segmentación adecuada de los núcleos celulares es un desafío complejo que requiere atención continua y métodos avanzados.

Referencias

- Alfaro, L., & Roca, M J. (2008). Portable telepathology: methods and tools. *Diagnostic pathology, 3 Suppl 1* (Suppl 1), S19. <https://doi.org/10.1186/1746-1596-3-s1-s19>
- Alias, N A., Mustafa, W A., Jamlos, M A., Alquran, H., Hanafi, H F., Ismail, S., & Rahman, K S A. (2022). Pap Smear Images Classification Using Machine Learning: A Literature Matrix. *Diagnostic, 12(12)*, Article 12. <https://doi.org/10.3390/diagnostics12122900>
- Allahqoli, L., Laganà, A S., Mazidimoradi, A., Salehiniya, H., Günther, V., Chiàntera, V., Goghari, S K K., Ghiasvand, M M., Rahmani, A., Momenimovahed, Z., & Alkatout, İ. (2022). Diagnosis of Cervical Cancer and Pre-Cancerous Lesions by Artificial Intelligence: A Systematic Review. *Diagnostic (Basel, Switzerland)*, 12(11), 2771. <https://doi.org/10.3390/diagnostics12112771>
- Araújo, F H D., Silva, R., Ushizima, D., Rezende, M T., Carneiro, C M., Bianchi, A G C., & Medeiros, F N S D. (2019). Deep learning for cell image segmentation and ranking. *Computerized medical imaging and graphics: the official journal of the Computerized Medical Imaging Society*, 72, 13-21. <https://doi.org/10.1016/j.compmedimag.2019.01.003>
- Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis., 11(4), 959-975. <https://www.sciencedirect.com/science/article/pii/S1751157717300500>
- Aristizabal, N., Cuello, C., Correa, P., Collazos, T., & Haenszel, W. (1984). The impact of vaginal cytology on cervical cancer risks in Cali, Colombia. *International journal of cancer*, 34(1, 5-9). <https://doi.org/10.1002/ijc.2910340103>

- Attallah, O. (2023). CerCan·Net: Cervical cancer classification model via multi-layer feature ensembles of lightweight CNNs and transfer learning. *Expert Systems with Applications*, 229, 120624. <https://doi.org/10.1016/j.eswa.2023.120624>
- Aufieri, R., Picone, S., & Paolillo, P. (2015). Collaborative development of open source-appropriate technologies: a way to reduce the global access gap? *BMJ Innovations*, 1(2). <https://doi.org/10.1136/bmjinnov-2014-000034>
- Ayala-García, J. (2014). La salud en Colombia: más cobertura pero menos acceso. (pp. 249-279). Banco de la República de Colombia. <https://econpapers.repec.org/paper/bdrregion/204.htm>
- Bank, W. (2024). World Bank Open Data. <https://data.worldbank.org>
- Bellina, L., & Missoni, E. (2009). Mobile cellphones (M-phones) in telemicroscopy: increasing connectivity of isolated laboratories. *Diagnostic pathology*, 4, 19. <https://doi.org/10.1186/1746-1596-4-19>
- Bengtsson, E., & Malm, P. (2014). Screening for Cervical Cancer Using Automated Analysis of PAP-Smears. *Computational and mathematical methods in medicine*, 2014, 842037. <https://doi.org/10.1155/2014/842037>
- Bonaccorso, G. (2018). Machine Learning Algorithms: Popular Algorithms for Data Science and Machine Learning, 2nd Edition. Packt Publishing.
- Bradski, G. (2000). The OpenCV library., 25, 120-125. <https://ci.nii.ac.jp/naid/10028167478/en/>
- Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Müller, A., Grisel, O., Niculae, V., Prettenhofer, P., Gramfort, A., Grobler, J., Layton, R., Vanderplas, J., Joly, A., Holt, B

- D., & Varoquaux, G. (2013). API design for machine learning software: experiences from the scikit-learn project. En *arXiv e-prints*. <https://arxiv.org/abs/1309.0238>
- Carr, K C., & Sellors, J W. (2004). Cervical cancer screening in low resource settings using visual inspection with acetic acid. *Journal of midwifery & women's health*, 49(4), 329-337. <https://onlinelibrary.wiley.com/doi/10.1016/j.jmwh.2004.02.015>
- Catarino, R. Petingnat, P., Dongui, G., & Vassilakos, P. (2015). Cervical cancer screening in developing countries at a crossroad: Emerging technologies and policy choices. *World Journal of Clinical Oncology*. 6(6), 281-290. <https://doi.org/10.5306/wjco.v6.i6.281>
- Cendales, R., Wiesner, C., Murillo, R., Piñeros, M., Tovar, S., & Mejía, J C. (2010). La calidad de las citologías para tamización de cáncer de cuello uterino en cuatro departamentos de Colombia: un estudio de concordancia. *National Institute of Health*, 30(1), 107-107. <https://doi.org/https://doi.org/10.7705/biomedica.v30i1.158>
- Chao, T E., & Mody, G N. (2015). The impact of intellectual property regulation on global medical technology innovation. *BMJ Innovations*, 1(2). 49-50. <http://dx.doi.org/10.1136/bmjinnov-2014-000033>
- Chitra, B., & Kumar, S S. (2021). Recent advancement in cervical cancer diagnosis for automated screening: a detailed review. *J Ambient Intell Human Comput* 13, 351-269 (2022). <https://doi.org/10.1007/s12652-021-02899-2>
- Christopherson, W M., Lundin, F E., Mendez, W M., & Parker, J. (1976). Cervical cancer control: A study of morbidity and mortality trends over a twenty-one-year period. *Cancer*, 38(3), 1357-1366. [https://doi.org/10.1002/1097-0142\(197609\)38:3<1357::aid-cnrcr2820380340>3.0.co;2-a](https://doi.org/10.1002/1097-0142(197609)38:3<1357::aid-cnrcr2820380340>3.0.co;2-a)

- Conceição, T., Braga, C., Rosado, L., & Vasconcelos, M J M. (2019). A Review of Computational Methods for Cervical Cells Segmentation and Abnormality Classification. Multidisciplinary Digital Publishing Institute. *International journal of molecular sciences*, 20(20), 5114-5114. <https://doi.org/https://doi.org/10.3390/ijms20205114>
- Dey, N., Ashour, A S., & Borra, S. (2018). Classification in BioApps. <https://doi.org/10.1007/978-3-319-65981-7>
- Diniz, D N., Rezende, M., Bianchi, A G C., Carneiro, C M., Ushizima, D., Medeiros, F N S D., & Souza, M J F. (2021). A Hierarchical Feature-Based Methodology to Perform Cervical Cancer Classification. Multidisciplinary Digital Publishing Institute, 11(9), 4091-4091. <https://doi.org/https://doi.org/10.3390/app11094091>
- Diniz, D N., Vitor, R., Bianchi, A G C., Delabrida, S., Carneiro, C M., Ushizima, D., Medeiros, F N S D., & Souza, M J F. (2021). An ensemble method for nuclei detection of overlapping cervical cells. *Expert Syst. Appl.*, 185(C). <https://doi.org/10.1016/j.eswa.2021.115642>
- Dudas, R., VandenBussche, C., Baras, A., Ali, S Z., & Olson, M T. (2014). Inexpensive telecytology solutions that use the Raspberry Pi and the iPhone. *Journal of the American Society of Cytopathology*, 3(1), 49-55. <https://doi.org/10.1016/j.jasc.2013.09.005>
- Fekri-Ershad, S. (2019). Pap smear classification using combination of global significant value, texture statistical features and time series features. <https://doi.org/10.1007/s11042-019-07937-y>
- Garwood, P. (2017). El diagnóstico temprano del cáncer salva vidas y reduce los costos de tratamiento. Organización Mundial de la Salud. Retrieved November 13, 2023 from

<https://www.who.int/es/news/item/03-02-2017-early-cancer-diagnosis-saves-lives-cuts-treatment-costs>

Ghoneim, A., Muhammad, G., & Hossain, M S. (2020). Cervical cancer classification using convolutional neural networks and extreme learning machines. *Future Generation Computer Systems*, 102, 643-649.

<https://doi.org/https://doi.org/10.1016/j.future.2019.09.015>

Guo, K., Liao, J., Bian, Z., Heng, X., & Zheng, G. (2015). InstantScope: a low-cost whole slide imaging system with instant focal plane detection. *Biomedical optics express*, 6(9), 3210-3216. <https://doi.org/10.1364/boe.6.003210>

Gupta, R., Gupta, S., Mehrotra, R., & Sodhani, P. (2017). Cervical Cancer Screening in Resource-Constrained Countries: Current Status and Future Directions. *Asian Pacific journal of cancer prevention : APJCP*, 18(6), 1461-1467.

<https://doi.org/10.22034/apjcp.2017.18.6.1461>

Han, S., Li, L., Jia, X., Ou, W., Ma, J., Wang, H., Zhao, J., & Zhu, Q. (2012). A Molecular Beacon-Based Method for Screening Cervical Cancer. *Journal of nanoscience and nanotechnology*, 12(11), 8282-8286. <https://scite.ai/reports/10.1166/jnn.2012.6789>

Hanson, T., & Kuniholm, J. (2008). Leveraging larger markets to encourage innovation in prosthetics: design of myopen, an open myoelectric signal processor for use as an input device for video gaming and hobby robotics. *Proceedings of the MEC*, 8.

<https://dukespace.lib.duke.edu/dspace/bitstream/handle/10161/2781/Hanson2%202008.pdf?sequence=3&isAllowed=y>

- Hussain, E., Mahanta, L B., Das, C., & Talukdar, R K. (2020). A comprehensive study on the multi-class cervical cancer diagnostic prediction on pap smear images using a fusion-based decision from ensemble deep convolutional neural network. *Tissue & cell*, 65, 101347-101347. <https://doi.org/https://doi.org/10.1016/j.tice.2020.101347>
- Jantzen, J., Norup, J., Dounias, G., & Bjerregaard, B. (2005). Pap-smear Benchmark Data For Pattern Classification. *Nature inspired smart information systems (NiSIS 2005)*, 1-9. <https://www.researchgate.net/publication/265873515>
- Jiang, P., Li, X., Shen, H., Chen, Y., Wang, L., Chen, H., Feng, J., & Liu, J. (2023). A systematic review of deep learning-based cervical cytology screening: from cell identification to whole slide image analysis. *Artif. Intell. Rev.*, 56(Suppl 2), 2687-2758. <https://doi.org/10.1007/s10462-023-10588-z>
- Jóhannesson, G., Geirsson, G., & Day, N. (1978). The effect of mass screening in iceland, 1965–74, on the incidence and mortality of cervical carcinoma. *International journal of cancer*, 21(4), 418-425. <https://doi.org/10.1002/ijc.2910210404>
- Joshi, S., Muwonge, R., Kulkarni, V., Lucas, E., Kulkarni, S., Kand, S., Mandolkar, M., Baig, M., Wankhede, S., Surwase, K., Pardeshi, D B., Basu, P., & Sankaranarayanan, R. (2021). Mobile Screening Unit (MSU) for the Implementation of the ‘Screen and Treat’ Programme for Cervical Cancer Prevention in Pune, India. *Asian Pacific journal of cancer prevention: APJCP*, 22(2), 413-418. <https://scite.ai/reports/10.31557/apjcp.2021.22.2.413>
- Jun, J K., Choi, K S., Jung, K W., Lee, H Y., Gapstur, S M., Park, S., & Yoo, K Y. (2009). Effectiveness of an organized cervical cancer screening program in Korea: Results from a

- cohort study. *International journal of cancer*, 124(1), 188-193.
<https://doi.org/10.1002/ijc.23841>
- Kentaro, W. (2016). labelme. <https://github.com/wkentaro/labelme>
- Khamparia, A., Gupta, D., Albuquerque, V H C D., Sangaiah, A K., & Jhaveri, R H. (2020). Internet of health things-driven deep learning system for detection and classification of cervical cells using transfer learning. *J. Supercomput.*, 76(11), 8590-8608.
<https://doi.org/10.1007/s11227-020-03159-4>
- Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J B., Grout, J., Corlay, S., Ivanov, P., Ávila, D V., Abdalla, S., & Willing, C. (2016). Jupyter Notebooks – a publishing format for reproducible computational workflows., 87-90. <http://doi.org/10.3233/978-1-61499-649-1-87>
- Kuko, M., & Pourhomayoun, M. (2019). An Ensemble Machine Learning Method for Single and Clustered Cervical Cell Classification. *2019 IEEE 20th International Conference on Information Reuse and Integration for Data Science (IRI)*, 216-222.
<https://doi.org/https://doi.org/10.1109/iri.2019.00043>
- Läärä, E., Day, N., & Hakama, M. (1987). Trends in mortality from cervical cancer in the nordic countries: association with organised screening programmes. *The Lancet (London, England)*, 1(8544), 1247/1249. [https://doi.org/10.1016/s0140-6736\(87\)92695-x](https://doi.org/10.1016/s0140-6736(87)92695-x)
- Lantz, B. (2019). Machine Learning with R Third Edition. Packt Publishing
- Lewis, M. (2004). Análisis de la situación del cáncer cervicouterino en América Latina y el Caribe. *Pan American Health Organization*. <https://iris.paho.org/handle/10665.2/762>

- Liang, Y., Tang, Z., Yan, M., Chen, J., Liu, Q., & Xiang, Y. (2021). Comparison detector for cervical cell/clumps detection in the limited data scenario. . *Neurocomputing*, *437*, 195-205. <https://doi.org/10.1016/j.neucom.2021.01.006>
- Lin, H., Hu, Y., Chen, S., Yao, J., & Zhang, L. (2019). Fine-Grained Classification of Cervical Cells Using Morphological and Appearance Based Convolutional Neural Networks. *Institute of Electrical and Electronics Engineers*, *7*, 71541-71549. <https://doi.org/https://doi.org/10.1109/access.2019.2919390>
- Linux and GNU - GNU Project - Free Software Foundation. (2021). Retrieved November 13, 2023 from <https://www.gnu.org/gnu/linux-and-gnu.en.html>
- Liu, G., Ding, Q., Luo, H., Sha, M., Li, X., & Ju, M. (2022). Cx22: A new publicly available dataset for deep learning-based segmentation of cervical cytology images. *Computer in biology and medicine*, *150*, 106194. <https://doi.org/10.1016/j.combiomed.2022.106194>
- Liu, J., Fan, H., Wang, Q., Li, W., Tang, Y., Wang, D., Zhou, M., & Chen, L. (2022). Local Label Point Correction for Edge Detection of Overlapping Cervical Cells. *Frontiers in neuroinformatics*, *16*, 895290. <https://doi.org/10.3389/fninf.2022.895290>
- Lozano, R L. (2007). Comparison of computer-assisted and manual screening of cervical cytology. *Gynecologic oncology*, *104*(1), 134-138. <https://doi.org/10.1016/j.ygyno.2006.07.025>
- Lu, Z., Carneiro, G., & Bradley, A P. (2015). An Improved Joint Optimization of Multiple Level Set Functions for the Segmentation of Overlapping Cervical Cells. *IEEE Transactions on Image Processing*, *24*(4), 1261-1272. <https://doi.org/10.1109/tip.2015.2389619>

- Lu, Z., Carneiro, G., Bradley, A P., Ushizima, D., Nosrati, M S., Bianchi, A G C., Carneiro, C M., & Hamarneh, G. (2017). Evaluation of Three Algorithms for the Segmentation of Overlapping Cervical Cells. *IEEE Journal of Biomedical and Health Informatics*, 21(2), 441-450. <https://doi.org/10.1109/jbhi.2016.2519686>
- Marín, D., & Romero, E. (2011). Sistemas de microscopía virtual: análisis y perspectivas. *Biomédica*, 31(1), Article 1. <https://doi.org/10.7705/biomedica.v31i1.345>
- Martin, E. (2003). Pap-Smear Classification
- Matias, A V., Amorim, J G A., Macarini, L A B., Cerentini, A., Onofre, A S C., Onofre, F B D M., Daltoé, F P., Stemmer, M R., & Wangenheim, A V. (2021). What is the state of the art of computer vision-assisted cytology? A Systematic Literature Review. *Computerized Medical Imaging Society*, 91, 101934. <https://doi.org/10.1016/j.compmedimag.2021.101934>
- Maurya, S., Tiwari, S., Mothukuri, M C., Tangeda, C M., Nandigam, R N S., & Addagiri, D C. (2023). A review on recent developments in cancer detection using Machine Learning and Deep Learning models. *Biomedical Signal Processing and Control*, 80, 104398. <https://doi.org/10.1016/j.bspc.2022.104398>
- Meléndez-Álvarez, B F., Robayo, O., Gil-Guillén, V F., & Carratalá-Munuera, M C. (2017). Design and Validation of a Low-Cost Telepathology System. *Telemedicine and e-Health*, 23(12), 976-982. <https://doi.org/10.1089/tmj.2017.0030>
- Moral-Argumedo, M J D., Ochoa, A., Posada-Gómez, R., & Aguilar-Lasserre, A A. (2023). A Deep Learning approach for automated Cytoplasm and Nuclei cervical segmentation.

Biomedical Signal Processing and Control, 81, 104483.

<https://doi.org/10.1016/j.bspc.2022.104483>

Mustafa, W A., Halim, A., & Rahman, K S A. (2020). A Narrative Review: Classification of Pap Smear Cell Image for Cervical Cancer Diagnosis. *Oncologie*, 22(2), 53-63.

<https://doi.org/10.32604/oncologie.2020.013660>

Nanda, K., McCrory, D C., Myers, E R., Bastian, L A., Hasselblad, V., Hickey, J D., & Matchar, D B. (2000). Accuracy of the Papanicolaou test in screening for and follow-up of cervical cytologic abnormalities: a systematic review. *Annals of internal medicine*, 132(10), 810-

819. <https://doi.org/10.7326/0003-4819-132-10-200005160-00009>

Niezen, G., Eslambolchilar, P., & Thimbleby, H. (2016). Open-source hardware for medical devices. *BMJ innovations*, 2(2), 78-83. <https://doi.org/10.1136/bmjinnov-2015-000080>

Olmos, M. (2007). Recomendaciones para la tamización de neoplasias del cuello uterino en mujeres sin antecedentes de patología cervical (preinvasora o invasora) en Colombia., *Instituto Nacioanal de Cancerología*. 1(1), 64. https://doi.org/ISBN_978-958-98496-4-4

Pangtey, P., Shamsunder, S., Malik, S., & Ahluwalia, C. (2022). Efficacy of SEE and TREAT in a Cervical Screening Program With Visual Inspection With Acetic Acid (VIA).

<https://scite.ai/reports/10.21203/rs.3.rs-1398700/v1>

Patil, D S B., & . (2022). Outcome of cervical cancer screening using Pap smear test on a target population. *European Journal of Molecular & Clinical Medicine*, 09(2), 122-126

Pearce, J M. (2015). Quantifying the Value of Open Source Hard-ware Development. *Modern Economy*, 6, 1-11. <https://doi.org/10.4236/me.2015.61001>

- Phoulady, H A., & Mouton, P R. (2018). A New Cervical Cytology Dataset for Nucleus Detection and Image Classification (Cervix93) and Methods for Cervical Nucleus Detection. (*arXiv:1811.09651*). *arXiv*. <http://arxiv.org/abs/1811.09651v1>
- Plissiti, M E., Dimitrakopoulos, P., Sfikas, G., Nikou, C., Krikoni, O., & Charchanti, A. (2018). Sipakmed: A New Dataset for Feature and Image Based Classification of Normal and Pathological Cervical Cells in Pap Smear Images. *2018 25th IEEE International Conference on Image Processing (ICIP)*, 3144-3148. <https://doi.org/10.1109/icip.2018.8451588>
- Plissiti, M E., Nikou, C., & Charchanti, A. (2011). Combining shape, texture and intensity features for cell nuclei extraction in Pap smear images. *Pattern Recognition Letters*, 32(6), 838-853. <https://doi.org/10.1016/j.patrec.2011.01.008>
- Plissiti, M E., Nikou, C., & Charchanti, A. (2011). Automated Detection of Cell Nuclei in Pap Smear Images Using Morphological Reconstruction and Clustering. *IEEE transactions on information technology in biomedicine : a publication of the IEEE Engineering in Medicine and Biology Society*, 15(2), 233–241. <https://doi.org/10.1109/titb.2010.2087030>
- Poomtavorn, Y., Himakhun, W., Suwannarurk, K., Thaweekul, Y., & Maireang, K. (2013). Cytohistologic Discrepancy of High-Grade Squamous Intraepithelial Lesions in Papanicolaou Smears. *Asian Pacific Journal of Cancer Prevention: APJCP*, 14(1), 599-602. <https://scite.ai/reports/10.7314/apjcp.2013.14.1.599>
- Pradit, L., Tantipalakorn, C., Charoenkwan, K., Suprasert, P., Srisomboon, J., & Muangmool, T. (2022). Efficacy of Listening to Music on Pain Reduction during Colposcopy-Directed

- Cervical Biopsy: A Randomized, Controlled Trial. *Medicina (Kaunas, Lithuania)*, 58(3), 429. <https://scite.ai/reports/10.3390/medicina58030429>
- Rezende, M T., Silva, R., Bernardo, F D O., Tobias, A H G., Oliveira, P H C., Machado, T M., Costa, C S., Medeiros, F N S D., Ushizima, D., Carneiro, C M., & Bianchi, A G C. (2021). Cric searchable image database as a public platform for conventional pap smear cytology data. *Scientific Data*, 8(1), 151. <https://doi.org/10.1038/s41597-021-00933-8>
- Rossum, G V., & Drake, F L. (2009). Python 3 reference manual. <https://dl.acm.org/citation.cfm?id=1593511>
- Sabeena, K., & Gopakumar, C. (2022). A hybrid model for efficient cervical cell classification. *Processing and Control*, 72, 103288. <https://doi.org/10.1016/j.bspc.2021.103288>
- Sachan, P L., Singh, M., Patel, M L., & Sachan, R. (2018). A Study on Cervical Cancer Screening Using Pap Smear Test and Clinical Correlation. *Medknow*, 5(3), 337-341. https://doi.org/https://doi.org/10.4103/apjon.apjon_15_18
- Sanabria, M L V., Álvarez-Najar, J P., Loba-Rodríguez, N J., & Peña-Pita, A P. (2022). Incertidumbre en la toma de citología cervicouterina. *Revista Ciencia y Cuidado*, 19(2), 31-39. <https://scite.ai/reports/10.22463/17949831.3219>
- Serrano, A S I. (2017). YOLO Object Detector for Onboard Driving Images
- Sharif, Y H. (2022). Clinical correlation of cervical cancer screening using Pap smear test. *Journal of Population Therapeutics and Clinical Pharmacology = Journal De La Therapeutique Des Populations Et De La Pharmacologie Clinique*, 29(1), e1-e8. <https://doi.org/10.47750/jptcp.2022.892>

Sinha, S., & Barry, M. (2011). Health Technologies and Innovation in the Global Health Arena.

The New England Journal of Medicine, 365(9), 779-782.

<https://doi.org/10.1056/nejmp1108040>

Sirintrapun, S J., & Cemic, A. (2012). Dynamic nonrobotic telemicroscopy via skype: A cost effective solution to teleconsultation. *Journal of Pathology Informatics*, 3, 28.

<https://doi.org/10.4103/2153-3539.100150>

Smith, S K., Nwosu, O., Edwards, A., Zerihun, M., Chung, M H., Suvada, K., & Ali, M K.

(2023). Performance of screening tools for cervical neoplasia among women in low- and middle-income countries: A systematic review and meta-analysis. *Public Library of Science*, 3(2), e0001598-e0001598.

<https://doi.org/https://doi.org/10.1371/journal.pgph.0001598>

Swanson, A A., & Pantanowitz, L. (2024). The evolution of cervical cancer screening. *Journal of the American Society of Cytopathology*, 13(1), 10-15.

<https://doi.org/https://doi.org/10.1016/j.jasc.2023.09.007>

WHO Information on Human papillomavirus (HPV) and cervical cancer – Canadian Cancer Survivor Network. (2023). WHO. Retrieved November 10, 2023, from

<https://survivornet.ca/learn/health-concerns-for-cancer-patients/hpv-and-cancer/who-information-on-human-papillomavirus-hpv-and-cervical-cancer/>

WHO. (2015). WHO compendium of innovative health technologies for low-resource settings.

<https://www.who.int/publications/i/item/9789241509992>

Wiesner, C., Tovar, S., Piñeros, M., Cendales, R., & Murillo, R. (2009). La oferta de servicios para tamización de cáncer de cuello uterino en Colombia. *Revista Colombiana de*

- Cancerología*. 13(3), 134-144. [https://doi.org/https://doi.org/10.1016/s0123-9015\(09\)70132-5](https://doi.org/https://doi.org/10.1016/s0123-9015(09)70132-5)
- Wijnen, B., Hunt, E J., Anzalone, G C., & Pearce, J M. (2014). Open-Source Syringe Pump Library. *PLOS ONE*, 9(9), e107216. <https://doi.org/10.1371/journal.pone.0107216>
- William, W., Ware, A., Basaza-Ejiri, A H., & Obungoloch, J. (2018). A review of image analysis and machine learning techniques for automated cervical cancer screening from pap-smear images. *Computer Methods and Programs in Biomedicine*, 164, 15-22. <https://doi.org/10.1016/j.cmpb.2018.05.034>
- William, W., Ware, A., Basaza-Ejiri, A H., & Obungoloch, J. (2018, October 1). A review of image analysis and machine learning techniques for automated cervical cancer screening from pap-smear images. <https://doi.org/10.1016/j.cmpb.2018.05.034>
- WIP, W I P. (2013). World intellectual property indicators. https://www.wipo.int/edocs/pubdocs/en/intproperty/941/wipo_pub_941_2013.pdf
- Woldetsadik, A B., Amhare, A F., Bitew, S T., Pei, L., & Lei, J. (2020). Socio-demographic characteristics and associated factors influencing cervical cancer screening among women attending in St. Paul's Teaching and Referral Hospital, Ethiopia. *BMC Women's Health*, 20(1), 70. <https://scite.ai/reports/10.1186/s12905-020-00927-5>
- Zappacosta, R., & Rosini, S. (2008). Cervical Cancer Screening: From Molecular Basis to Diagnostic Practice, Going through New Technologies. *Technology in Cancer Research & Treatment*, 7(3), 161-174. <https://scite.ai/reports/10.1177/153303460800700302>

Zea, J H R., Mejía, A M., Arredondo, M V., Acevedo, L S T., & Zapata, W S. (2007).

Accesibilidad a la citología cervical en Medellín, Colombia en 2006. *Revista Española de Salud Pública*, 81(6), 657-666. <https://scite.ai/reports/10.1590/s1135-57272007000600009>

Zhang, L., Lü, L., Nogues, I., Summers, R M., Liu, S., & Yao, J. (2017). DeepPap: Deep

Convolutional Networks for Cervical Cell Classification. *IEEE Journal of Biomedical and Health Informatics*, 21(6), 1633-1643. <https://doi.org/10.1109/jbhi.2017.2705583>

Zhang, Y., Zall, Y., Nissim, R., Satyam., & Zimmermann, R. (2022). Evaluation of a new dataset

for visual detection of cervical precancerous lesions. *Expert Systems with Applications*, 190, 116048. <https://doi.org/10.1016/j.eswa.2021.116048>

Apéndices

Apéndice A.

Conjunto de datos CRIC con datos complementarios. (CRIC_New_label.csv)