

Aplicación de inteligencia artificial mediante algoritmos de aprendizaje profundo para clasificar imágenes de citologías en un programa de tamizaje de cáncer de cuello uterino

Jose Said Manzano Chaya

**Trabajo de grado para optar por el título de
Especialista en Patología**

Director

Ernesto García Ayala

Médico Patólogo

Asesora Epidemiológica

Tania Mendoza Herrera

Médica internista

Magíster en Epidemiología

Universidad Industrial de Santander

Facultad de Salud

Escuela de Medicina

Especialización en Patología

Bucaramanga

2025

Agradecimientos

A Dios, en quien todo se puede lograr.

A mis padres Said y María Josefina, por su apoyo incondicional.

A mis profesores del Departamento de Patología de la Universidad Industrial de Santander, quienes siempre estuvieron ahí para guiarme y enseñarme.

Al personal de Bienestar Universitario de la Universidad Industrial de Santander, por su ayuda en la captación de participantes para la realización de este proyecto. Sin su colaboración no hubiera sido posible llevar a la realidad este trabajo.

A cada una de las participantes del estudio, gracias por su aporte a la investigación en nuestra universidad.

Gracias al equipo del Centro para Reconocimiento e Inspección Celular (CRIC) y al Centro de Bioingeniería Molecular y Celular de Dresden, Alemania, por mantener el conocimiento científico y la investigación al alcance de todos.

Tabla de contenido

Introducción	12
1. Planteamiento y justificación	14
2. Objetivos	16
2.1 Objetivo general	16
2.2 Objetivos específicos	16
3. Marco teórico y estado del arte	18
3.1 Epidemiología	18
3.2 Inteligencia artificial aplicada al diagnóstico médico	20
3.3 Ejemplos del uso de la IA en patología	22
3.4 Fisiopatología.	23
4. Metodología	26
4.1 Pregunta de investigación	26
4.2 Hipótesis nula	26
4.3 Hipótesis alterna	26
4.4 Tipo de estudio	26
4.5 Población	27
4.5.1 Población de referencia	27
4.5.2 Población blanco	27
4.5.3 Población de estudio (accesible)	27
4.6 Tamaño de la muestra	28
4.7 Selección de participantes	29
4.7.1 Criterios de inclusión	29

4.7.2 <i>Criterios de exclusión+</i>	30
4.8 Variables	31
4.8.1 <i>Operacionalización de Variables</i>	31
4.8.2 <i>Definición de variables</i>	32
4.9 Procedimientos para la recolección de la información	38
4.10 Materiales y métodos	39
4.10.1 <i>Fase Pre-analítica</i>	39
4.10.2 <i>Fase analítica</i>	39
4.11 Procesamiento y control de calidad de los datos	40
4.12 Análisis estadístico	40
4.12.1 <i>Análisis univariado</i>	40
4.12.2 <i>Análisis bivariado</i>	41
4.13 Análisis de Inteligencia Artificial	41
4.13.1 <i>Transformación de datos y entrenamiento de los modelos de IA</i>	47
5. Consideraciones éticas	50
6. Tratamiento de datos personales	52
7. Cronograma de actividades	54
8. Presupuesto	55
9. Resultados y discusión	56
9.1 Análisis descriptivo de datos clínicos	56
9.2 Análisis bivariado de datos sociodemográficos con cambios preneoplásicos en citología	58
9.3 Análisis bivariado entre variables de la muestra de citología y cambios preneoplásicos	62
9.4 Resultados de Inteligencia Artificial (IA)	64

<i>9.4.1 Desempeño de los modelos de IA con la base de datos de entrenamiento</i>	64
<i>9.4.2 Desempeño de los modelos de IA con la base de datos de prueba</i>	66
10. Conclusiones	76
11. Trabajos futuros	78
12. Conflicto de interés	79
Bibliografía	80

Lista de figuras

Figura 1. <i>Número de casos nuevos de cáncer en población femenina de todas las edades para el año 2020.</i>	18
Figura 2. <i>Incidencia, mortalidad y prevalencia por sitio de cáncer.</i>	19
Figura 3. <i>Bloque de cinco capas que usa la arquitectura DenseNet, mostrando cómo cada capa toma los mapas de características anteriores</i>	43
Figura 4. <i>Modelo inicial de Inception V1.</i>	44
Figura 5. <i>Arquitectura de InceptionV3.</i>	45
Figura 6. <i>Ejemplo del uso de MobileNet para detectar objetos con probabilidades correspondientes.</i>	46
Figura 7. <i>Ejemplos de imágenes de células normales (fila superior) y con alteraciones (fila inferior) de la base de datos usada para entrenar los modelos de IA</i>	48
Figura 8. <i>Ejemplo de células normales clasificadas correctamente por DenseNet</i>	65
Figura 9. <i>Ejemplo de células con alteraciones, clasificadas correctamente por DenseNet</i>	65
Figura 10. <i>Curva ROC para DenseNet</i>	69
Figura 11. <i>Curva ROC para InceptionV3</i>	70
Figura 12. <i>Curva ROC para VGG19</i>	71
Figura 13. <i>Curva ROC para MobileNet</i>	72

Lista de tablas

Tabla 1. <i>Edad actual, edad de menarquia y edad de inicio de relaciones sexuales en nuestra población</i>	56
Tabla 2. <i>Características sociodemográficas y clínicas</i>	57
Tabla 3. <i>Análisis bivariado de alteraciones en citología y variables sociodemográficas</i>	61
Tabla 4. <i>Análisis bivariado de alteraciones en citología y variables de la muestra.</i>	63
Tabla 5. <i>Desempeño de los modelos de IA con la base de datos de entrenamiento</i>	64
Tabla 6. <i>Desempeño de los cuatro modelos de IA con la base de datos de prueba</i>	67
Tabla 7. <i>Matriz de confusión de los cuatro modelos de IA con la base de datos de prueba, comparados con el Gold Standard (lectura manual)</i>	67
Tabla 8. <i>Tiempos de clasificación para cada modelo</i>	68

Resumen

Título: Aplicación de inteligencia artificial mediante algoritmos de aprendizaje profundo para clasificar imágenes de citologías en un programa de tamizaje de cáncer de cuello uterino*

Autor: Jose Said Manzano Chaya**

Palabras clave: Diagnóstico Precoz del Cáncer, Inteligencia Artificial, Aprendizaje Profundo, Prueba de Papanicolaou, Neoplasias del Cuello Uterino

Introducción: La citología cervicouterina convencional continúa siendo un pilar del tamizaje del cáncer cervicouterino en Colombia, pero su utilidad se ve opacada por una gran carga laboral y bajo rendimiento diagnóstico. La aplicación de inteligencia artificial (IA) puede proveer una solución a este problema, sin embargo, no hay estudios que evalúen su utilidad en nuestra población.

Objetivo: Evaluar y comparar la capacidad discriminativa de cuatro modelos de inteligencia artificial basados en aprendizaje profundo para identificar alteraciones en citología cervicouterina.

Materiales y métodos: Se entrenaron cuatro modelos de inteligencia artificial basados en aprendizaje profundo (DenseNet, InceptionV3, MobileNet y VGG19) con una base de datos de casi 7000 imágenes, para después determinar la capacidad discriminativa de cada modelo, el tiempo de computación y su tasa de acuerdo con la lectura manual convencional en nuestra población.

Resultados: Se incluyeron en el estudio un total de 80 participantes. Los cuatro modelos mostraron un área bajo la curva superior a 0.9. MobileNet tuvo la mejor especificidad (0.97) con valores predictivos positivo y negativo de 0.91, así como la mejor tasa de acuerdo interobservador Kappa

* Trabajo de Grado

** Facultad de Salud, Escuela de Medicina, Departamento de Patología. Director: Ernesto García Ayala, MD. Esp. Patología. Asesora epidemiológica: Tania Mendoza Herrera, MD. Esp. Medicina Interna, MSc Epidemiología

(0.777) en un tiempo de computación menor. Por otro lado, InceptionV3 tuvo la mejor sensibilidad (0.93) a costa de una especificidad (0.77) y tasa de acuerdo interobservador inferiores (0.617).

Conclusiones: La aplicación de inteligencia artificial tiene el potencial de disminuir la carga laboral de los patólogos y citólogos, sin alterar el rendimiento diagnóstico. En nuestro estudio, MobileNet mostró ofrecer el mejor valor costo-beneficio y mayor eficiencia a largo plazo, al tiempo que su menor demanda de poder computacional hace que sea más factible integrar su uso al flujo de trabajo convencional. Su aplicación podría ayudar a la lectura convencional, o incluso servir como método de tamizaje primario, sin embargo, se requieren estudios adicionales que evalúen esta posibilidad en el laboratorio de patología.

Abstract

Title: Application of Artificial Intelligence Using Deep Learning Algorithms for the Classification of Cytology Images in a Cervical Cancer Screening Program *

Author: Jose Said Manzano Chaya **

Keywords: Early Diagnosis of Cancer, Artificial Intelligence, Deep Learning, Papanicolaou Test, Cervical Neoplasms

Introduction: Conventional cervical cytology remains a cornerstone of cervical cancer screening in Colombia. However, its effectiveness is hindered by a high workload and low diagnostic performance. The application of artificial intelligence (AI) could provide a solution to this issue; however, no studies have evaluated its utility in our population.

Objective: To assess and compare the discriminative capacity of four deep learning-based artificial intelligence models in identifying abnormalities in cervical cytology.

Materials and Methods: Four deep learning-based artificial intelligence models (DenseNet, InceptionV3, MobileNet, and VGG19) were trained on a database of approximately 7,000 images. The discriminative capacity of each model, computation time, and agreement rate with conventional manual reading in our population were then determined.

Results: A total of 80 participants were included in the study. All four models demonstrated an area under the curve (AUC) greater than 0.9. MobileNet exhibited the highest specificity (0.97), with positive and negative predictive values of 0.91, as well as the highest interobserver agreement rate (Kappa = 0.777) with a lower computation time. Conversely, InceptionV3 had the highest

* Degree Work

** Faculty of Health, School of Medicine, Department of Pathology. Director: Ernesto García Ayala, MD. Esp. Pathology. Epidemiological advisor: Tania Mendoza Herrera, MD. Esp. Internal Medicine, MSc Epidemiology

sensitivity (0.93) at the expense of lower specificity (0.77) and interobserver agreement (Kappa = 0.617).

Conclusions: The application of artificial intelligence has the potential to reduce the workload of pathologists and cytologists without compromising diagnostic performance. In our study, MobileNet demonstrated the best cost-effectiveness and long-term efficiency, while its lower computational power requirements make it more feasible for integration into conventional workflows. Its application could assist conventional cytology reading or even serve as a primary screening method. However, additional studies are needed to evaluate this possibility in pathology laboratories.

Introducción

A nivel mundial, el cáncer de cérvix ocupa el cuarto lugar entre los cánceres en mujeres, con un estimado de casos nuevos en el 2020 de 604.000 y 342.000 muertes directamente atribuibles a esta patología, tendencia que ha ido en ascenso según cifras dadas por Globocan (1). Se considera entonces que, si se mantienen las tendencias actuales, el número de muertes en las Américas aumentará en un 45% en el 2030, por lo que, según la Organización Panamericana de la Salud, esta entidad representa un problema de salud pública de primer orden que debe ser abordado mediante una estrategia integral e interprogramática (2).

A diferencia de otros tipos de cáncer, esta entidad tiene un agente causal en la mayoría de los casos: el virus del VPH, así como una historia natural de la enfermedad reconocida en la cual se produce progresión de los cambios citológicos desde displasia de bajo grado hasta llegar a cáncer, por lo que la citología cervicouterina, como estrategia de tamizaje, incide en la carga de enfermedad de esta patología, al brindar la posibilidad de visualizar, clasificar y estadificar los cambios celulares previos a la progresión tumoral. Sin embargo, en muchos países en desarrollo incluyendo países de América Latina y el Caribe, su aplicación no se ha visto asociada a una reducción de la mortalidad similar a la de los países desarrollados, debido principalmente a limitaciones en la organización de los sistemas comunitarios, aspectos culturales, la baja sensibilidad para la detección de neoplasia intraepitelial cervical (NIC) o cáncer invasivo y su alto costo relativo, por lo que su uso como estrategia de tamizaje primario ha sido desplazada a favor de las pruebas ADN-VPH. Sin embargo, en mujeres jóvenes (menores de 30 años), la alta prevalencia de infecciones transitorias y baja incidencia de cáncer de cuello uterino ha conducido

a una sobrerreferencia a colposcopia, motivo por el cual en Colombia la citología cervicouterina se reserva aún como prueba de triage (3).

La inteligencia artificial (IA) es una rama de la tecnología que busca simular artificialmente el cerebro humano en sistemas computacionales, creando programas que piensen, tomen decisiones y actúen como lo harían los humanos en situaciones similares. La patología computacional basada en IA es una rama emergente de la patología que ha mostrado gran potencial al incrementar el desempeño diagnóstico y la disponibilidad de tecnologías diagnósticas a gran escala, y estudios que incluyen ensayos clínicos multicéntricos han demostrado que la lectura por IA de citologías cervicouterinas puede tener áreas bajo la curva mayores a 0.8, lo cual derivaría en mejores porcentajes de sensibilidad diagnóstica al aplicar estas tecnologías en el diagnóstico citológico manual convencional (4, 5, 6, 7).

1. Planteamiento y justificación

El cáncer de cuello uterino es el cuarto tipo de cáncer más frecuente en mujeres a nivel mundial. En Colombia, el cáncer de cuello uterino es la primera causa de muerte por cáncer entre mujeres de 30 a 59 años (3) y ocupa el segundo puesto en tumores invasivos más prevalentes en mujeres (1). Pero, a diferencia de otros tipos de cáncer, este tiene un agente causal en la mayoría de los casos: el virus del papiloma humano (VPH). Los serotipos 16 y 18 son causantes de aproximadamente el 50% de los cánceres y las lesiones precancerosas del cuello uterino. En torno al 90% de los nuevos casos y muertes en el año 2020 tuvieron lugar en países de ingresos bajos y medianos, siendo las tasas de mortalidad en América Latina y el Caribe 3 veces más altas que en Norteamérica, evidenciando enormes desigualdades en salud (2).

En Colombia, en el plan de Beneficios de Salud (3) se encuentran disponibles tres pruebas de tamizaje para cáncer de cuello uterino, las cuales son obligatorias a partir de la expedición de la Resolución 3280 de 2018:

- Citología cervicouterina para mujeres de 25 a 29 años
- Test de ADN - VPH en mujeres de 30 a 65 años
- Técnicas de inspección visual con ácido acético y lugol, dependiendo de la edad de la mujer y su lugar de residencia (población rural dispersa entre 30 y 50 años)

A pesar que se ha reportado una sensibilidad menor del tamizaje con citología en comparación con las pruebas de ADN-VPH, la citología cervicouterina tiene amplia evidencia científica sobre desenlaces relevantes que respaldan su uso (3). Sin embargo, existe gran variabilidad interobservador e intraobservador (8), y su utilidad se ve opacada al requerir de

personal experto que revise manualmente cientos de células de forma individual, un proceso laborioso, que lleva tiempo y es vulnerable a errores humanos, lo que se traduce en mayores demoras y costos para el sistema de salud.

El creciente avance de la patología digital y la investigación en inteligencia artificial (IA) plantean una posible solución a este problema. Diferentes estudios (4, 7, 9) han evaluado la posibilidad de utilizar programas de inteligencia artificial para mejorar la sensibilidad en la detección de lesiones premalignas en tamizaje de cáncer de cuello uterino basado en citología cervicouterina, al usar software de reconocimiento de imágenes para priorizar aquellas muestras que puedan tener patrones citológicos de riesgo, con el objetivo de mejorar la oportunidad en el diagnóstico.

En Colombia son escasos los estudios que han investigado las potenciales aplicaciones de la IA en salud pública, y lo que ha sido publicado se centra en usos en el campo de la ingeniería de fluidos (10), análisis de patrones de comportamiento humano en psicología (11), y predicción de duración de procedimientos quirúrgicos (12), sin proyectos referentes que evalúen la aplicación de sistemas de IA en el tamizaje de cáncer de cuello uterino.

Adicionalmente, el campo de la investigación en IA ha captado el interés de la comunidad científica en los últimos años, y hay una gran cantidad de algoritmos de IA siendo desarrollados continuamente, lo que dificulta el proceso de escoger el mejor para esta tarea, por lo que diseñamos el presente estudio para evaluar y comparar la capacidad discriminativa de cuatro diferentes modelos diagnósticos basados en IA para diagnosticar anomalías en citología cervicouterina en un programa de tamizaje de cáncer de cuello uterino de la ciudad de Bucaramanga, Colombia.

2. Objetivos

2.1 Objetivo general

Determinar la capacidad discriminativa de cuatro algoritmos de inteligencia artificial basados en aprendizaje profundo para identificar células anormales en citología cervicouterina obtenidas en un programa de tamizaje de la Universidad Industrial de Santander, entre el 1 de enero al 31 de diciembre del 2023.

2.2 Objetivos específicos

- Caracterización clínica de la población a estudio
- Establecer la prevalencia de ASCUS en la población a estudio
- Establecer prevalencia de lesiones intraepiteliales de bajo y alto grado en la población a estudio
- Establecer prevalencia de carcinoma escamoso in situ en la población a estudio
- Determinar la relación entre vacunación contra VPH y lesiones premalignas en citología cervicouterina
- Determinar la relación entre las lesiones premalignas en citología y el uso de métodos de barrera en la población a estudio.
- Determinar la relación entre las lesiones premalignas en el cuello uterino y la presencia de microorganismos patógenos.

-Comparar la capacidad discriminativa del software con la evaluación realizada por el Gold Standard (Médico patólogo experto)

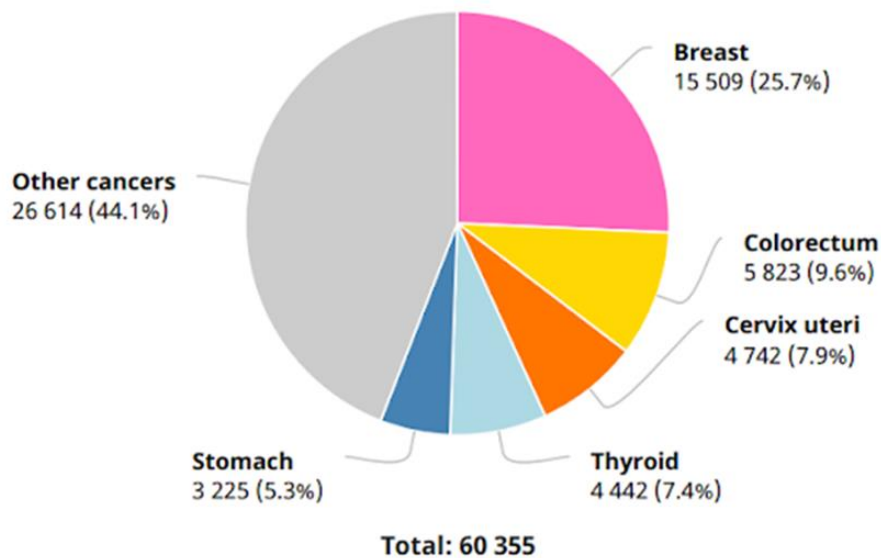
3. Marco teórico y estado del arte

3.1 Epidemiología

De acuerdo con el reporte del observatorio nacional de cáncer de Colombia, el cual se basa en las estadísticas de Globocan 2020, el territorio colombiano tiene una población de 50 millones de personas, con una incidencia de 113.221 casos de todos los tipos de cáncer, así como una prevalencia a 5 años de 293.524 casos (figura 1). De estos, el cáncer de cuello uterino representa el tercer lugar en incidencia en población femenina, solo superado por el cáncer de mama y de colon y recto.

Figura 1.

Número de casos nuevos de cáncer en población femenina de todas las edades para el año 2020.



Fuente: International Agency for Research on Cancer (GCO). 2020.

Al analizar incidencia, mortalidad y prevalencia por sitio del cáncer, se encuentra que en la población colombiana para el año 2020, el cáncer de cuello uterino tuvo una incidencia de 4.742 casos, siendo el número 7 en incidencia total para ambos sexos. La mortalidad fue de 2.409 casos, también en número 7 a nivel nacional. La prevalencia a 5 años fue de 12.472 casos, con una proporción de 48.16 casos por 100.000 habitantes, lo cual posiciona a este tipo de cáncer en el puesto número 3 (figura 2), después del cáncer de mama y el cáncer de próstata, datos que reflejan la carga de enfermedad que genera en la población esta patología.

Figura 2.

Incidencia, mortalidad y prevalencia por sitio de cáncer.

Cancer	New cases				Deaths				5-year prevalence (all ages)	
	Number	Rank	(%)	Cum.risk	Number	Rank	(%)	Cum.risk	Number	Prop. (per 100 000)
Breast	15 509	1	13.7	5.21	4 411	3	8.0	1.43	52 025	200.88
Prostate	14 460	2	12.8	6.10	3 846	5	7.0	0.96	49 172	196.81
Stomach	8 214	3	7.3	1.45	6 451	1	11.7	1.07	11 611	22.82
Colon	7 579	4	6.7	1.34	4 048	4	7.4	0.64	18 654	36.66
Lung	6 876	5	6.1	1.24	6 090	2	11.1	1.06	7 314	14.37
Thyroid	5 304	6	4.7	0.94	489	21	0.89	0.08	17 523	34.44
Cervix uteri	4 742	7	4.2	1.53	2 490	7	4.5	0.80	12 472	48.16
Non-Hodgkin lymphoma	4 242	8	3.7	0.77	2 004	10	3.6	0.35	11 996	23.58
Leukaemia	3 367	9	3.0	0.53	2 419	8	4.4	0.37	9 661	18.99
Rectum	2 721	10	2.4	0.50	1 265	13	2.3	0.21	7 302	14.35
Pancreas	2 693	11	2.4	0.47	2 639	6	4.8	0.46	2 040	4.01
Corpus uteri	2 635	12	2.3	0.99	576	18	1.0	0.21	8 403	32.45
Kidney	2 466	13	2.2	0.47	932	15	1.7	0.17	6 452	12.68
Ovary	2 391	14	2.1	0.82	1 485	12	2.7	0.51	6 344	24.50
Liver	2 289	15	2.0	0.39	2 220	9	4.0	0.38	2 149	4.22
Bladder	1 995	16	1.8	0.34	699	17	1.3	0.09	5 487	10.78
Brain, central nervous system	1 901	17	1.7	0.33	1 650	11	3.0	0.29	5 154	10.13
Melanoma of skin	1 805	18	1.6	0.31	490	20	0.89	0.08	5 268	10.35
Multiple myeloma	1 376	19	1.2	0.27	1 035	14	1.9	0.20	3 340	6.56
Testis	1 369	20	1.2	0.38	195	25	0.35	0.05	4 895	19.59
Larynx	1 000	21	0.88	0.20	552	19	1.0	0.10	2 893	5.69
Lip, oral cavity	914	22	0.81	0.16	378	23	0.69	0.06	2 495	4.90
Oesophagus	867	23	0.77	0.15	842	16	1.5	0.13	867	1.70
Hodgkin lymphoma	825	24	0.73	0.13	251	24	0.46	0.04	2 739	5.38
Gallbladder	680	25	0.60	0.12	459	22	0.83	0.08	794	1.56
Penis	550	26	0.49	0.19	162	27	0.29	0.06	1 638	6.56
Oropharynx	530	27	0.47	0.09	189	26	0.34	0.03	1 313	2.58
Anus	483	28	0.43	0.09	104	31	0.19	0.02	1 314	2.58
Vulva	447	29	0.39	0.14	134	28	0.24	0.03	1 354	5.23
Salivary glands	436	30	0.39	0.08	108	30	0.20	0.02	1 328	2.61
Kaposi sarcoma	396	31	0.35	0.06	57	33	0.10	0.01	1 130	2.22
Nasopharynx	155	32	0.14	0.03	80	32	0.15	0.01	448	0.88
Vagina	150	33	0.13	0.05	46	34	0.08	0.01	386	1.49
Mesothelioma	144	34	0.13	0.03	126	29	0.23	0.03	175	0.34
Hypopharynx	113	35	0.10	0.02	36	35	0.07	0.01	183	0.36
All cancer sites	113 221	-	-	18.39	54 987	-	-	8.66	293 524	576.9

Fuente: International Agency for Research on Cancer (GCO). 2020.

En Colombia existen 318 territorios considerados dentro de la categoría de rural y rural dispersa, con un área que representa el 84.7% del total nacional. Para el año 2014, según el Departamento Nacional de Planeación (13), aproximadamente el 30.4% de población de Colombia vive en zonas rurales, siendo los departamentos con mayor densidad de área rural Meta, Vichada, Guainía, Casanare, Guaviare, Vaupés, Amazonas, Caquetá y Putumayo. De manera similar, se observa que, de acuerdo al reporte de cáncer de cérvix emitido por el fondo colombiano de cuentas de alto costo, los departamentos con tasas de incidencia ajustada más altas para esta entidad fueron en los departamentos de Putumayo (14.6 x 100.000 per.), Meta (11.5 x 100.000 per.), Guaviare (10.9 x 100.000 per), y Cesar (9.2 x 100.000 per.), con los departamentos de Amazonas, Guainía y Vaupés con incidencia de 0.0 (14), lo cual puede ser explicado por falta de reporte al sistema de salud, o falta de afiliación de su población al mismo. Podemos concluir por lo tanto, que en Colombia las regiones con mayor área rural dispersa tienen tasas más altas de incidencia de cáncer de cuello uterino, una tendencia vista también en otros países de América Latina y el Caribe (3).

3.2 Inteligencia artificial aplicada al diagnóstico médico

La inteligencia artificial (IA) es un concepto que se refiere a la simulación de la mente humana en sistemas computacionales mediante algoritmos y procesos con el fin de crear programas que tengan la capacidad de pensar como humanos. La IA plantea la posibilidad de realizar tareas que normalmente requieren de inteligencia humana, tales como percepción visual, toma de decisiones y comunicación.

Para lograr estos objetivos, fue necesario crear una forma en la que las máquinas pudieran simular la forma en que los humanos adquieren el conocimiento, es decir, aprender, por lo que surgió el término “aprendizaje automático”.

El aprendizaje automático es la rama de la inteligencia artificial que se encarga de construir sistemas computacionales que mejoren o “aprendan” automática y progresivamente mediante su propia experiencia, usando tres acercamientos de aprendizaje: bajo supervisión, sin supervisión, y aprendizaje semi-supervisado (15). Es uno de los campos de más rápido crecimiento, y de vital importancia para nuestra actual definición de lo que representa una inteligencia artificial. Su desarrollo se ha visto impulsado en esta época de globalización por una disponibilidad cada vez mayor de información en línea y su potencial de aplicación en las diferentes áreas de las ciencias médicas, educación, comercio, políticas y mercadeo, entre otros (16).

Uno de los pilares del aprendizaje automático es el aprendizaje profundo (en inglés, deep learning); una técnica basada en el uso de redes artificiales neuronales, las cuales se componen por múltiples capas funcionales de neuronas artificiales, llamadas nodos, distribuidas en una capa de entrada, múltiples capas ocultas, y una capa de salida. Cada una de estas capas está interconectada, y la fuerza de sus conexiones (que establece la relevancia de los datos que aportan) se denomina “peso”. Este peso se ajusta y optimiza durante el proceso de entrenamiento de una red neuronal mediante la aplicación de diferentes métodos estadísticos; los más comunes son el método de gradiente descendiente (incluyendo sus variantes como el gradiente descendiente estocástico), retropropagación (backpropagation) y métodos de regularización (17), entre otros. Mediante el uso de estos métodos estadísticos, las neuronas que componen el algoritmo y sus conexiones asociadas adquieren diferentes pesos en el proceso de entrenamiento, que representan la importancia de la información que se transmite a través de la red.

El objetivo de utilizar estos algoritmos estadísticos es resolver el problema de cuáles características de una imagen dada son las más relevantes para categorizarla apropiadamente, ya que lo que se logra es extraer automáticamente las características imagenológicas significativas en múltiples niveles a la vez, gracias al uso de diferentes capas de nodos interconectados. El aprendizaje profundo ha emergido como una solución prometedora a problemas en distintos campos de la inteligencia artificial, tales como el reconocimiento de imágenes (18), reconocimiento de voz (19), traducción de lenguajes (20), reconocimiento facial (21), diagnóstico médico (22), entre otros.

Este sistema de aprendizaje automático usa una gran cantidad de datos iniciales, también llamados set de entrenamiento, para construir algoritmos estadísticos que le permitan interpretar y analizar datos nuevos (23). Al día de hoy se ha experimentado con diferentes modelos de aprendizaje automático en patología para asistir al diagnóstico patológico, usando la morfología de las células benignas, malignas, la morfología del núcleo celular, las divisiones normales y anormales de las células, así como el número y la forma de ductos, vasos sanguíneos, etc (24). }

3.3 Ejemplos del uso de la IA en patología

Los resultados iniciales son prometedores, y demuestran que los sistemas asistidos por IA tienen el potencial de clasificar diferentes patologías:

Campanella et al (5) validaron un algoritmo basado en aprendizaje profundo de alta capacidad para clasificar 44732 imágenes de 3 tipos de cáncer, incluyendo cáncer de próstata, carcinoma basocelular y metástasis de cáncer de mama a nódulos linfáticos axilares, obteniendo un área bajo la curva superior a 0.98 para todos los tipos de cáncer. Su aplicación clínica podría

permitir que los patólogos excluyeran hasta un 75% de las láminas de histología manteniendo un 100% de sensibilidad diagnóstica.

Korbar et al. (25) desarrollaron múltiples algoritmos de aprendizaje profundo para clasificar un set de 2074 imágenes de cinco tipos de pólipos colorrectales, incluyendo pólipos hiperplásicos, aserrados sésiles y tradicionales, tubulares y tubulovelloso. La sensibilidad diagnóstica obtenida en este estudio fue del 93% (CI 95%, 89.0-95.9%). En este mismo campo de estudio, Bychkov et al (26) combinaron redes convolucionales neuronales para predecir el pronóstico médico de 420 pacientes de cáncer colorrectal basados en análisis de micromatrices de tejidos (TMA), logrando un área bajo la curva de 0.69 (hazard ratio, 2.3; CI 95%, 1.79-3.03), un resultado mejor que el logrado por un consenso de expertos en patología (hazard ratio 1.67; IC 95%, 1.28-2.19; AUC 0.58).

En un ensayo multicéntrico evaluando diferentes algoritmos de diagnóstico por IA para identificar metástasis a nódulos linfáticos, el mejor algoritmo mostró un AUC de 0.994, resultados similares a los logrados por un panel de 11 patólogos diferentes, con un AUC de hasta 0.998 (6). De forma similar, un estudio multicéntrico en China con una población de más de 700 mil pacientes demostró una tasa de acuerdo diagnóstico entre la lectura manual y la lectura por IA de citologías de cuello uterino del 94.7%, y al usar la IA para asistir al diagnóstico citológico manual, se demostró una mejora en la sensibilidad de detección de lesiones intraepiteliales cervicales del 5.8% (4), con una ligera disminución en la especificidad diagnóstica.

3.4 Fisiopatología.

Actualmente se cuenta con información relevante sobre los factores de riesgo del cáncer de cuello uterino gracias a la identificación del Virus de Papiloma Humano (VPH) como causa de la mayoría de casos, con raras excepciones. Aunque las infecciones por VPH pueden transmitirse a través de rutas no sexuales, la mayoría son resultado del contacto sexual (2).

El VPH es un virus perteneciente a la familia *Papillomaviridae*, sin envoltura lipídica, constituido por ADN circular bicatenario con una longitud de aproximadamente 8.000pb. Su genoma está constituido por una región codificante de ocho genes en un mARN policistrónico y una región no codificante. La región no codificante o reguladora contiene la región promotora (P97 en VPH16) y una secuencia intensificadora que controla la transcripción de los genes virales. La región codificante está dividida en dos regiones de expresión, una temprana y una tardía. Los primeros genes (E1, E2, E4, E5, E6 y E7) están relacionados con la modulación de los procesos de transformación celular, replicación y transcripción viral. Los productos de los genes E6 y E7 son considerados elementos claves en la oncogénesis del cuello uterino, con estudios que han documentado su participación en la inmortalización de queratinocitos primarios *in vitro* y su actividad tumorigénica *in vivo* en monos transgénicos (27).

Entre los más de 200 tipos de VPH descritos y entre los más de 30 que causan infecciones en el epitelio genital, 14 tienen la capacidad de inducir el desarrollo de una neoplasia maligna de cérvix, por lo que se denominan variantes de alto riesgo. Estudios prospectivos han mostrado que un 15 a 28% de mujeres en las que se detectó ADN de VPH desarrollaron lesiones intraepiteliales *in situ* en 2 años, comparados con sólo 1 a 3% de aquellas en las que no se detectó el ADN. El riesgo de progresión de las variantes HPV-16 y 18 es mayor (aproximadamente un 40%) que para otros tipos de VPH (28).

Las etapas más importantes en la carcinogénesis del cáncer son: primero, la infección del epitelio metaplásico (particularmente en la zona de transformación), el cual posiblemente requiera abrasión o microtrauma de la epidermis, segundo la persistencia viral, tercero la progresión clonal de la infección persistente en el epitelio hasta la progresión a pre-cáncer (lesiones de alto grado o neoplasia intraepitelial cervical de grados 2 y 3), asociada a la interferencia de las proteínas E6 y E7 en los genes p53 y pRB, que altera la función normal de estas proteínas celulares, y finalmente la invasión, que toma entre 10 a 20 años en promedio (28).

4. Metodología

4.1 Pregunta de investigación

¿Cuál es la capacidad discriminativa de cuatro diferentes modelos de inteligencia artificial basados en aprendizaje profundo para detectar células cervicales anormales de citologías referidas de un programa de tamizaje de cáncer de cuello uterino, comparado con la observada por el médico patólogo experto?

4.2 Hipótesis nula

La capacidad discriminativa de los modelos basados en inteligencia artificial para identificar células anormales en citología cervicouterina es similar a la del médico patólogo.

4.3 Hipótesis alterna

La capacidad discriminativa de un software basado en inteligencia artificial para identificar células anormales en citología cervicouterina es diferente a la dada por el médico patólogo.

4.4 Tipo de estudio

Estudio observacional analítico de evaluación de método diagnóstico (AUC, evaluación de especificidad y sensibilidad, valor predictivo positivo, valor predictivo negativo y tasa de acuerdo

interobservador Kappa) con muestreo no probabilístico concurrente, con población del programa de tamizaje de cáncer de cuello uterino de la división de Bienestar Universitario de la Universidad Industrial de Santander.

4.5 Población

4.5.1 Población de referencia

Pacientes femeninas mayores de 18 años de edad residentes en la ciudad de Bucaramanga, Colombia.

4.5.2 Población blanco

Pacientes femeninas mayores de 18 años de edad que se encuentren en programas de tamizaje de cáncer de cérvix en la ciudad de Bucaramanga, Colombia.

4.5.3 Población de estudio (accesible)

Pacientes femeninas mayores de 18 años de edad que acudan al programa de tamizaje de cáncer de cérvix de la división de Bienestar Universitario de la Universidad Industrial de Santander ubicada en la ciudad de Bucaramanga, Colombia, del 1 de enero al 31 de diciembre del 2023.

4.6 Tamaño de la muestra

Para el cálculo del tamaño de la muestra en este estudio se utilizó un software de código abierto para salud pública: OPENEPI65 y se tomaron los resultados expuestos en un estudio observacional multicéntrico donde se determinó el probable diagnóstico citológico obtenido por la lectura manual vs la obtenida por inteligencia artificial de citologías cervicouterinas en base a alteraciones morfológicas y prevalencia de VPH, factor de riesgo para patología neoplásica, presente en el estudio de Bao, H., Bi, H., Zhang, X., Zhao, Y., Dong, Y., & Luo, X. et al. (2020). Artificial intelligence-assisted cytology for detection of cervical intraepithelial neoplasia or invasive cancer: A multicenter, clinical-based, observational study. *Gynecologic Oncology*, 159(1), 171-178. doi: 10.1016/j.ygyno.2020.07.099 (7).

Bao, H., Bi, H., Zhang, X.,	Casos: 1762	Relación no expuesto: expuesto
Zhao, Y., Dong, Y., & Luo, X.		1:1
Grupo no expuesto (VPH -)		Probabilidad grupo no expuesto: 5%
Grupo expuesto (VPH +)		Probabilidad grupo expuesto: 26%
Intervalo de confianza		95%
Odds ratio		6,7%

Tamaño muestral: transversal, de cohorte, y ensayo clínico			
Nivel de significación de dos lados(1-alpha)			95
Potencia (1-beta,% probabilidad de detección)			80
Razón de tamaño de la muestra, Expuesto/No Expuesto			1
Porcentaje de No Expuestos positivos			5
Porcentaje de Expuestos positivos			26
Odds Ratio:			6.7
Razón de riesgo/prevalencia			5.2
Diferencia riesgo/prevalencia			21
	Kelsey	Fleiss	Fleiss con CC
Tamaño de la muestra - Expuestos	47	46	55
Tamaño de la muestra- No expuestos	47	46	55
Tamaño total de la muestra	94	92	110
Referencias			
Kelsey y otros, Métodos en Epidemiología Observacional 2da Edición, Tabla 12-15			
Fleiss, Métodos Estadísticos para Relaciones y Proporciones, fórmulas 3.18&, 3.19			
CC= corrección de continuidad			
Los resultados se redondean por el entero más cercano			

Se establecieron un total de 110 citologías con presencia o no de alteraciones en una relación de 1:1 más un 10% adicional en caso de pérdidas, para un total de tamaño de muestra de 121 citologías.

4.7 Selección de participantes

4.7.1 Criterios de inclusión

Pacientes femeninas mayores o iguales a 18 años de edad que asistan al programa de tamizaje de cáncer de cuello uterino ofrecido por la división de Bienestar Universitario de la Universidad Industrial de Santander, las cuales cuentan con:

- Historia clínica debidamente rotulada e identificada según protocolo de realización del examen.
- Datos de historia clínica, incluyendo lugar de residencia, fórmula obstétrica, edad de inicio de relaciones sexuales, estado actual de embarazo, tratamiento hormonal actual, menopausia, citología previa con resultado de la misma, vacuna contra VPH y/o prueba de ADN VPH.
- Datos de procedimiento realizados, que incluyan estado del cuello uterino al momento de la inspección para toma de la muestra y observaciones, de ser necesarias.
- Datos de resultado del estudio, incluyendo calidad de la muestra y categorización general.
- Datos específicos en caso de que el estudio sea anormal, incluyendo los hallazgos (neoplásicos o no) encontrados.
- Firma del médico patólogo que realizó lectura de la muestra.
- Láminas de citología con coloración de Papanicolau, que cuenten con una celularidad adecuada estimada en aproximadamente 8000 a 12000 células.
- Representación de células exocervicales y células endocervicales adecuadamente visualizables en la lámina de citología.

4.7.2 Criterios de exclusión+

- Lámina de citología no identificada y/o rotulada, oscura, rota, fragmentada, o con escasa celularidad, definido por el patólogo evaluador de la muestra.
- Extendidos de citología en base líquida, u otra coloración diferente a la convencional.

4.8 Variables

4.8.1 Operacionalización de Variables

Variables extraídas de la historia clínica:

1. Residencia: Urbano / rural (1 urbano, 0 rural)
2. Número de gestaciones, partos, pérdidas gestacionales, cesáreas y nacidos vivos (1 variable por cada una, con números enteros)
3. Edad en años cumplidos
4. Edad de la menarquia (variable continua discreta)
5. Edad de inicio de relaciones sexuales
6. Planifica (si 1/no 0)
7. Método de planificación (si aplica según respuesta anterior)
8. Embarazada actualmente (si 1/no 0)
9. Tratamiento hormonal (si 1 / no 0)
10. Menopausia (si 1 / no 0)
11. Citología previa (si 1/no 0)
12. Resultado de citología previa (Alterado 1/ normal 0)
13. Vacuna contra VPH (Si 1 / no 0)
- 14 Número de dosis de vacuna contra VPH (si aplica)
15. Prueba de ADN VPH
16. Resultado de prueba ADN VPH (si aplica)
17. Procedimientos quirúrgicos realizados en el cuello uterino

18. Características del cuello uterino a la observación (reportadas por quien realice la citología)

VARIABLES EXTRAÍDAS DE LAS CITOLOGÍAS:

1. Calidad de la muestra obtenida (satisfactoria, insatisfactoria, rechazada)
2. Categorización general de la muestra
3. Clasificación de las células por el patólogo
4. Clasificación de las células por el software
5. Células categorizadas como normales, identificadas como normales (verdaderos negativos)
6. Células categorizadas como anormales, identificadas como anormales (verdaderos positivos)
7. Células categorizadas como normales, identificadas como anormales (falsos positivos)
8. Células categorizadas como anormales, identificadas como normales (falsos negativos)

4.8.2 Definición de variables

Nombre de la variable	Definición conceptual	Definición operacional	Escala de medición
Residencia	Establece en qué zona la paciente objeto del examen tiene su residencia usual	Cualitativa nominal	1 urbano 0 rural

Nombre de la variable	Definición conceptual	Definición operacional	Escala de medición
Gestaciones	Número de gestaciones de la paciente	Cuantitativa discreta	Números enteros
Partos	Número de partos de la paciente	Cuantitativa discreta	Números enteros
Pérdidas gestacionales	Número de gestaciones de la paciente finalizados en aborto, ectópico o mortinato	Cuantitativa discreta	Números enteros
Cesáreas	Número de cesáreas a las que ha sido sometida la paciente	Cuantitativa discreta	Números enteros
Nacidos vivos	Número de hijos nacidos vivos de la paciente	Cuantitativa discreta	Números enteros
Edad	Número de años cumplidos por la paciente al momento de la toma de muestra	Cuantitativa discreta	Números enteros
Edad de la menarquia	Número de años cumplidos por la paciente en el momento de la menarquia	Cuantitativa discreta	Números enteros
Edad de inicio de relaciones sexuales	Edad en años cumplidos en la que la paciente inició relaciones sexuales	Cuantitativa discreta	Números enteros
Gestación actual	Paciente se encuentra actualmente en estado de embarazo	Cualitativa nominal	0 = no 1 = si 2 = duda
Planifica	Establece si la paciente se encuentra recibiendo algún tipo de	Cualitativa nominal	0 = no 1 = si

Nombre de la variable	Definición conceptual	Definición operacional	Escala de medición
	tratamiento de planificación para prevenir embarazos no deseados		
Método de planificación	Si la paciente planifica, qué método de planificación se encuentra usando al momento de la toma del examen	Cualitativa nominal	0 = Ninguno 1 = Anovulatorio 2 = Dispositivo intrauterino 3 = Quirúrgico 4 = Condón 5 = Otro
Embarazada actualmente	Establece si la paciente se encuentra en estado de gestación al momento de realizar la toma del examen	Cualitativa nominal	1 = si 0 = no
Tratamiento hormonal	Establece si la paciente se encuentra tomando algún tipo de tratamiento hormonal	Cualitativa nominal	1 = si 0 = no
Menopausia	Establece si la paciente ha cesado permanentemente la menstruación después de 12 meses consecutivos de amenorrea	Cualitativa nominal	1 = si 0 = no
Citología previa	La paciente contaba con un resultado previo de citología, antes de la realización del examen	Cualitativa nominal	1 = si 0 = no

Nombre de la variable	Definición conceptual	Definición operacional	Escala de medición
Resultado de citología previa	Si la paciente contaba con un resultado previo de citología, este fue normal o anormal	Cualitativa nominal	1 = Anormal 0 = Normal 2 = No sabe
Vacuna contra VPH	Establece si la paciente ha recibido la vacunación contra el Virus del Papiloma Humano	Cualitativa nominal	1 = Si 0 = no
Número de dosis de vacuna contra VPH	Si la paciente ha recibido la vacunación contra el VPH, numera cuántas dosis de vacuna ha recibido	Cuantitativa discreta	Números enteros
Prueba de ADN VPH	Establece si la paciente se ha realizado previamente la prueba en sangre para detectar ADN de VPH	Cualitativa nominal	1 = si 0 = no
Resultado de prueba ADN VPH	En caso de que la paciente se haya realizado previamente la prueba en sangre para detectar ADN de VPH, establece cuál fue el resultado obtenido	Cualitativa nominal	0 = Negativo 1 = Positivo
Procedimientos quirúrgicos realizados en el cuello uterino	La paciente contaba o no con algún tipo de procedimiento quirúrgico realizado previamente en el cuello uterino	Cualitativa nominal	1 = Si 0 = No
Características del cuello uterino a la observación	Establece cuáles son las características del cuello uterino observadas por el realizador del examen	Cualitativa nominal	0 = Sano 1 = Congestivo 2 = Erosionado 3 = Ulcerado

Nombre de la variable	Definición conceptual	Definición operacional	Escala de medición
			4 = Sangrante 5 = Coliflor 6 = Condiloma 7 = Prolapsado 8 = Ausente 9 = Leucorrea
Calidad de la muestra	Establece si la citología tiene una calidad satisfactoria o no, según criterios establecidos por el sistema Bethesda 2015 para reportar citologías cervicouterinas.	Cualitativa nominal	0 = Insatisfactoria 1 = Satisfactoria
Categorización general	Establece si el diagnóstico final dado por el patólogo categorizó la citología como negativa para lesión intraepitelial o malignidad, o la clasificó en el grupo de anormalidades de células epiteliales	Cualitativa nominal	0 = Negativa para lesión intraepitelial o malignidad 1 = Anormalidades de células epiteliales
Clasificación de las células por el patólogo	Establece cuál fue la clasificación final dada por el patólogo para cada una de las células analizadas	Cualitativa nominal	0 = normal 1 = Alterada
Clasificación de las células por el modelo de inteligencia artificial	Establece cuál fue la clasificación final dada por el modelo de inteligencia artificial para cada una de las células analizadas	Cualitativa nominal	0 = normal 1 = Alterada

Nombre de la variable	Definición conceptual	Definición operacional	Escala de medición
Células categorizadas como normales, identificadas como normales	Establece si las células categorizadas inicialmente como normales por el médico patólogo fueron categorizadas como normales por el modelo de inteligencia artificial	Cualitativa nominal	0 = identificada 1 = no identificada
Células categorizadas como anormales (con o sin VPH), identificadas como anormales	Establece si las células categorizadas inicialmente como anormales por el médico patólogo fueron categorizadas como anormales por el modelo de inteligencia artificial	Cualitativa nominal	0 = identificada 1 = no identificada
Células categorizadas como normales, identificadas como anormales	Establece si las células categorizadas inicialmente como normales por el médico patólogo fueron categorizadas como anormales (falsos positivos) por el modelo de inteligencia artificial	Cualitativa nominal	0 = identificada 1 = no identificada
Células categorizadas como anormales, identificadas como normales	Establece si las células categorizadas inicialmente como anormales por el médico patólogo fueron categorizadas como normales (falsos negativos) por el modelo de inteligencia artificial	Cualitativa nominal	0 = identificada 1 = no identificada

4.9 Procedimientos para la recolección de la información

Se realizó recolección de reportes de extendidos de Papanicolaou enviados por el servicio de Bienestar Universitario de la Universidad Industrial de Santander que cumplan con requisitos de inclusión. A las pacientes que cumplieron criterios de inclusión se les ofreció información en forma escrita y verbal acerca del propósito y la utilidad del estudio, se les permitió realizar preguntas y se les aseguró que participar o no en el estudio no afectaría de ninguna manera su diagnóstico ni su tratamiento, y que podrían retirarse del estudio en cualquier momento si así lo deseaban. El personal se aseguró que las pacientes entendieran la información indicada. Luego, se obtuvo consentimiento informado escrito de las pacientes que desearan voluntariamente participar en el estudio.

Posteriormente, se analizaron los formatos de solicitud adjuntos a cada estudio de tamizaje de citología cervicouterina de manera individual, los cuales habían sido previamente identificados con número interno extrayendo los datos correspondientes a las variables de interés de cada caso, los cuales se tabularon en una tabla de excel con su respectivo código. Cada uno de los casos tuvo total anonimato, usando únicamente el número interno asignado a cada caso. Después del diagnóstico y reporte final, se tomaron fotografías digitales de las células representativas de cada caso con un microscopio óptico de luz de marca Leica DM750 conectado a una cámara digital con referencia ICC50W y resolución de 5.0 megapíxeles, usando un aumento de 40x.

4.10 Materiales y métodos

4.10.1 Fase Pre-analítica

Esta fase incluye los acontecimientos que transcurren entre la solicitud de citología y el análisis de la muestra en el laboratorio de patología. Una muestra adecuadamente preparada conlleva preparación e información a la usuaria, identificación, toma de la muestra, manejo, transporte de láminas y preparación del espécimen para su análisis.

Estas actividades se prestan en el Servicio en cumplimiento del Anexo técnico 1 de la Resolución 1043 de 2006, con inscripción y habilitación del servicio ante las autoridades sanitarias cumpliendo los estándares requeridos para la toma de muestras para citologías cervicouterinas, incluidas en el código 3.28 de la misma resolución (29).

4.10.2 Fase analítica

El procesamiento de los extendidos citológicos de cuello uterino comprende la asignación de un número de registro interno, distribución y organización de las láminas, preparación y manejo de la batería de coloración y el montaje. Las láminas al estar adecuadamente teñidas presentan una buena definición de los detalles nucleares, transparencia en el citoplasma y diferenciación celular. Para garantizar una buena calidad de las muestras se ejecutan las siguientes actividades, en concordancia con el Anexo Técnico N°1 de la resolución 1043 del 2006 (29):

- Identificación y marcación de las muestras
- Preparación y manejo de la batería de tinción

- Coloración de las muestras con tinción de Papanicolaou
- Montaje de las muestras
- Distribución de las láminas en bandejas para su análisis, lectura e interpretación por el médico patólogo

4.11 Procesamiento y control de calidad de los datos

La información obtenida de todos los instrumentos de recolección de datos se digitaron por duplicado de forma independiente por dos digitadores diferentes entrenados previamente en el manejo de la hoja electrónica Excel y se compararon para detectar errores de digitación por medio de los comandos de duplicación. Se estableció la estrategia de comparación de los formatos escritos y electrónicos, con un chequeo de consistencia interna de los datos luego de lo cual se consideraron como finales y se archivaron en medio magnético por duplicado. Posterior a esta verificación se exportaron y analizaron los datos con STATA 12.0 para proceder al análisis estadístico.

4.12 Análisis estadístico

4.12.1 Análisis univariado

Las variables recolectadas fueron evaluadas según su nivel de medición (medias o medianas para las variables continuas con medidas de dispersión como la desviación estándar para variables de distribución normal y mediana, recorrido para las de distribución no normal, y las

categorías fueron analizadas mediante proporciones), con reporte de los respectivos intervalos de confianza del 95%.

4.12.2 Análisis bivariado

Para determinar la relación entre cada una de las variables independientes con la presencia de alteraciones en citología, se realizó un análisis bivariado, utilizando la regresión logística, que precise el cálculo de medidas de efecto como el ODDS RATIO (OR).

4.13 Análisis de Inteligencia Artificial

Uno de los pasos más importantes a la hora de diseñar un estudio de inteligencia artificial en patología es escoger una base de datos que tenga la cantidad suficiente de casos con la patología en estudio para ser usada como set de entrenamiento (23). Con este fin, se escogió la base de datos de citología cervicouterina convencional en coloración de Papanicolaou desarrollada por Diniz et al. (30) la cual es de libre acceso, y contiene imágenes de células previamente balanceadas y clasificadas en dos, tres y seis categorías diagnósticas diferentes, tomando imágenes de la base de datos del Centro para Reconocimiento e Inspección Celular (en inglés, Center for Recognition and Inspection of Cells, CRIC) disponible en línea en: <https://database.cric.com.br> (31). Para el set de entrenamiento del presente estudio, usamos el set de dos categorías (células normales y células con alteraciones), que contiene casi 7000 imágenes en total de células procesadas independientemente, seleccionando el centro del núcleo y cortando un área de 90 x 90 píxeles a su alrededor. Esta metodología permite evaluar cada célula por separado, incluyendo el núcleo y parte

del citoplasma en la imagen final, y evita la aparición de dos o más núcleos en la misma imagen. Se tomó el 80% de imágenes de esta base de datos como set de entrenamiento, y el 20% restante como set de validación.

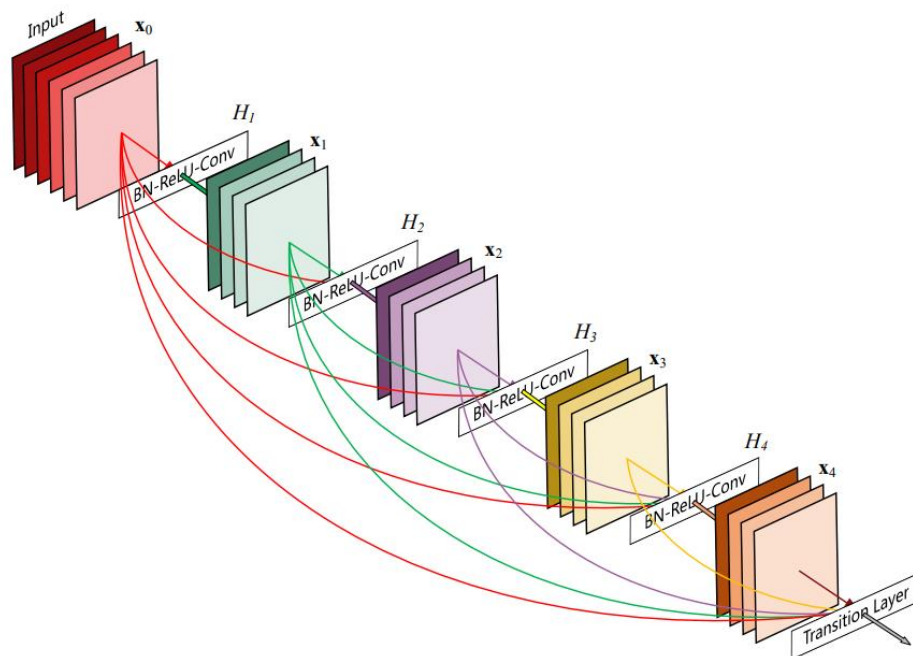
A continuación, las células correspondientes al set de entrenamiento fueron cargadas a un software de inteligencia artificial de libre acceso desarrollado por Kräter et al (33) del Centro de Bioingeniería Molecular y Celular de Dresden, Alemania, llamado “AIDeveloper”, el cual permite entrenar y evaluar diferentes modelos de IA con una interfaz visual. Este software integra diferentes modelos matemáticos y estadísticos pre-entrenados a un set nuevo de datos, e implementa la posibilidad de realizar transformación digital de imágenes de forma independiente y semi-automatizada, cambiando su altura, ancho y rotándolas en eje vertical u horizontal, una propiedad que en otros estudios (34) ha sido usada en imágenes histopatológicas para evitar el fenómeno de “overfitting” el cual sucede cuando el modelo se adapta adecuadamente a los datos de entrenamiento, pero tiene mal desempeño con los datos de prueba (17). Se usaron las imágenes obtenidas para evaluar cuatro diferentes modelos de aprendizaje automático con arquitectura basada en aprendizaje profundo, los cuales han mostrado desempeño prometedor en patología digital: DenseNet, Inception V3, MobileNet y VGG19, con el fin de definir el mejor desempeño comparado con el Gold Standard dado por el médico patólogo. Las ventajas y desventajas, así como diferencias entre estos modelos serán mencionadas a continuación:

DenseNet: Abreviación de “Densely connected Convolutional Networks”, es un modelo cuya arquitectura de redes neuronales está basada en el principio de usar múltiples capas interconectadas entre sí de forma que aquellas más cercanas a las capas de entrada y de salida tienen conexiones más cortas. Mientras que una red neural tradicional con L capas tiene L conexiones (una entre cada capa y sus capas subsecuentes), DenseNet tiene $\frac{L(L+1)}{2}$ conexiones, las

cuales se interconectan de forma que para cada capa se usa el mapa de características de todas las capas predecesoras, lo que permite evitar que características menores desaparezcan progresivamente (vanishing-gradient), reusar las variables previas y reducir el número de parámetros necesarios, logrando desempeño igual o mejor que otras arquitecturas neuronales con menor requerimiento de poder computacional (35).

Figura 3.

Bloque de cinco capas que usa la arquitectura DenseNet, mostrando cómo cada capa toma los mapas de características anteriores



Nota: Adaptado de Huang et al. (35)

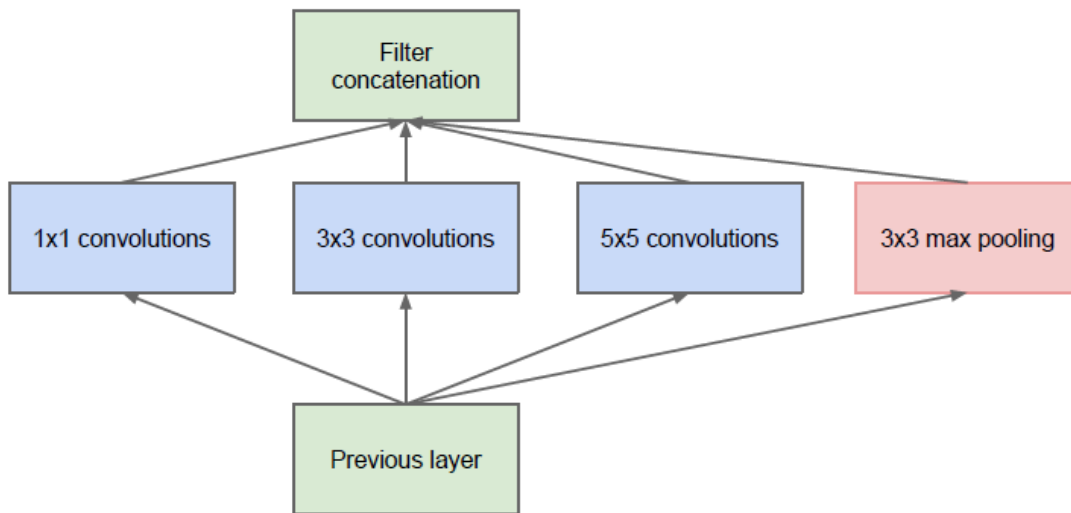
Inception V3: Modelo desarrollado a partir de Inception V1, originalmente llamado GoogLeNet en 2014, el cual fue desarrollado por un equipo de Google. En su primera versión, el

modelo Inception implementó el uso de múltiples filtros de diferentes tamaños en el mismo nivel, por lo que las capas se analizan en paralelo en lugar de en sentido descendente y progresivo como en otros modelos, con el objetivo de evitar el fenómeno de “overfitting”, que sucede cuando el modelo logra un porcentaje de error bajo en el set de datos de entrenamiento pero mayor porcentaje de error en datos de prueba (17).

Cada capa usaba tanto convolución (proceso de transformar una imagen aplicando una máscara a cada píxel para aumentar su nitidez, difuminarla, detectar sus ejes u otros realces basados en el kernel) como pooling (reducir las dimensiones del mapa de características mediante funciones matemáticas, usualmente máximo y promedio), en capas de 1x1, 3x3 y 5x5, para capturar diferentes tamaños de información en la imagen.

Figura 4.

Modelo inicial de Inception V1.



Nota: Adaptado de Szegedy et al. (2014) (36)

Al momento de su lanzamiento, el modelo de inception V1 mostró un desempeño mejor que muchos otros modelos disponibles con un margen de error de apenas el 6.67%.

En el 2015 fue lanzada la versión 3 del modelo Inception, la cual mejoró diferentes características incluyendo convoluciones asimétricas con factorización espacial, utilización de clasificadores auxiliares y factorización para lograr convoluciones menores, con lo cual se logró mejorar su rendimiento y disminuir su margen de error.

Figura 5.

Arquitectura de InceptionV3.

type	patch size/stride or remarks	input size
conv	3×3/2	299×299×3
conv	3×3/1	149×149×32
conv padded	3×3/1	147×147×32
pool	3×3/2	147×147×64
conv	3×3/1	73×73×64
conv	3×3/2	71×71×80
conv	3×3/1	35×35×192
3×Inception	As in figure 5	35×35×288
5×Inception	As in figure 6	17×17×768
2×Inception	As in figure 7	8×8×1280
pool	8 × 8	8 × 8 × 2048
linear	logits	1 × 1 × 2048
softmax	classifier	1 × 1 × 1000

Nota: Adaptado de Szegedy et al (37)

El modelo final cuenta con 42 capas, donde los datos de salida de cada capa constituyen los datos de entrada de la siguiente, y cuenta con múltiples filtros y variaciones de técnicas de

reducción para ajustar tamaños de bloques de datos entre bloques, como se evidencia en la figura 5. Este modelo logró un desempeño considerablemente mejor que los modelos anteriores, con un margen de error de apenas el 3.58%.

MobileNet: Modelo desarrollado por A. Howard et al (38) basado en convoluciones separables en profundidad, las cuales tienen la ventaja de que, al contrario de modelos estándar que filtran y combinan datos de entrada en un nuevo set de datos de salida en cada paso, estas logran separar dos capas, una para filtrado y otra para combinación de datos, con lo cual se reduce drásticamente el tamaño del modelo y la capacidad de computación requerida. Este modelo fue diseñado específicamente para aplicaciones en plataformas móviles como tablets y celulares, por lo que su uso podría ayudar al diseño de programas de smartphone que ayudaran a clasificar células en extendidos de Papanicolaou.

Figura 6.

Ejemplo del uso de MobileNet para detectar objetos con probabilidades correspondientes.



Nota: Adaptado de A. Howard et al (38).

VGG19: Modelo desarrollado a partir de AlexNet, uno de los primeros modelos de aprendizaje automático basado en redes neuronales convolucionales que salió al mercado en el 2012. VGG19 fue creado por el grupo de Geometría Visual de la universidad de Oxford (Visual Geometry Group, VGG) de donde derivan las siglas de su nombre. Este modelo tiene 19 capas de profundidad, y se basa en el uso de filtros de convolución pequeños (3x3) para mejorar la efectividad y disminuir el margen de error comparado con otros modelos.

4.13.1 Transformación de datos y entrenamiento de los modelos de IA

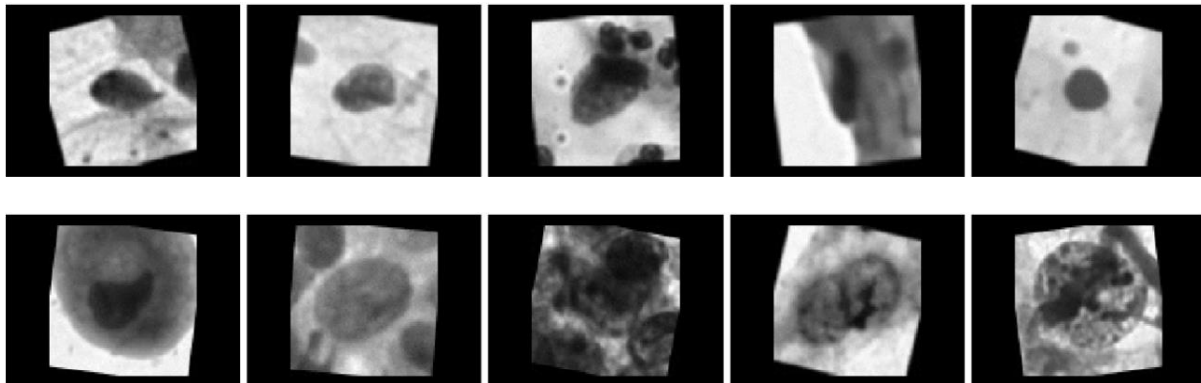
Los modelos de IA previamente descritos fueron pre-entrenados en más de un millón de imágenes de la base de datos ImageNet. Esto confiere la ventaja de poder hacer transfer-learning en nuestro estudio, un proceso que permite mejorar el desempeño y la velocidad de aprendizaje de distintos modelos de IA al crear una relación lógica entre las bases de datos a las que ha sido previamente expuesto el modelo y la base de datos actual (39) pero, debido a que las imágenes son tomadas de fotografías de entornos naturales (como carreteras, casas o personas caminando), tienden a mantener su verticalidad con respecto al eje horizontal, lo que puede causar que los modelos pre-entrenados de esta forma pueden sobre estimar la verticalidad de las imágenes a las que sean expuestas (24, 34). Con el propósito de evitar este fenómeno y mejorar el desempeño de los modelos con la base de datos de prueba (es decir, reducir el overfitting), las imágenes de la base de datos de entrenamiento fueron transformadas agregando ruido estadístico, modificando su enfoque ligeramente y rotandolas de forma aleatoria. Cada una de las transformaciones elegidas fueron aplicadas cada dos ciclos de entrenamiento logarítmico del conjunto de datos (epoch), y se monitorizó el entrenamiento de cada set de forma individual, evaluando el desempeño con el set

de entrenamiento y el set de validación, y deteniendo el proceso de forma temprana cuando los ciclos adicionales no aumentaran más el rendimiento.

Adicionalmente, las imágenes fueron convertidas a escala de grises, dado que, por experiencia de los investigadores en el campo, este paso permite mejorar considerablemente el desempeño del modelo, probablemente al disminuir el ruido estadístico que causan diferentes colores dentro de una imagen de citología. La figura 7 muestra ejemplos de las imágenes de células usadas para entrenar los modelos luego de las modificaciones realizadas.

Figura 7.

Ejemplos de imágenes de células normales (fila superior) y con alteraciones (fila inferior) de la base de datos usada para entrenar los modelos de IA



Al finalizar el entrenamiento y verificar que cada modelo hubiera logrado la mejor capacidad discriminativa posible, se procedió a transformar las células de cada una de las citologías recogidas en nuestro estudio (tomadas en su totalidad como set de prueba), seleccionando el núcleo y cortando un área de 90 x 90 píxeles a su alrededor, de manera que fueran comparables con el set de referencia. Posteriormente, las imágenes fueron clasificadas por un patólogo experto en dos grupos: uno de células normales (sin alteraciones morfológicas) y otro que incluyó todas aquellas

que presentaran alteraciones asociadas a lesiones premalignas o malignas, de acuerdo con los criterios del sistema Bethesda (32), para obtener así nuestro set de prueba, dividido también en dos categorías diagnósticas: Células normales y células con alteraciones.

Todo el proceso se llevó a cabo en un computador portátil con 16 GB de memoria RAM, procesador Core i7-9750H, tarjeta gráfica NVIDIA GeForce GTX 1660Ti de 6 GB y sistema operativo Windows 10. Para cuantificar y comparar los datos obtenidos, se realizó un análisis de especificidad y sensibilidad con un intervalo de confianza del 95%, utilizando para mostrar la certeza diagnóstica el valor predictivo positivo (VPP), valor predictivo negativo (VPN), odds ratio (OR) y área bajo la curva (AUC). Se comparó el nivel de certeza diagnóstica de los modelos basados en IA con la lectura manual usando índice de Kappa.

5. Consideraciones éticas

El presente trabajo está clasificado como una “investigación sin riesgo”, de acuerdo al literal “A” del artículo 11 de la Resolución 008430 de 1993 del Ministerio de Salud de Colombia: “Son estudios que emplean técnicas y métodos de investigación documental retrospectivos, y aquellos en los que no se realiza ninguna intervención o modificación intencionada de las variables biológicas, fisiológicas, psicológicas o sociales de los individuos que participan en el estudio, entre los que se consideran: revisión de historias clínicas, entrevistas, cuestionarios, y otros en los que no se le identifique ni se traten aspectos sensitivos de su conducta”.

El presente trabajo de investigación se rige por los principios éticos de:

- **Beneficencia:** Los investigadores protegerán a las personas y sus datos personales de cualquier daño. El objetivo del estudio de investigación es trabajar para mejorar beneficios potenciales tanto de las participantes como de la población en general. Se respetará este principio al prevenir la identificación del paciente y al omitir procesos no validados que pudieran generar conclusiones que llegaran a causar daño o perjuicio sobre futuros tamizajes de pacientes.

- **No maleficencia:** Los investigadores no harán daño bajo ninguna circunstancia a las participantes del estudio. Es prudente mencionar que la presente investigación no confiere riesgo adicional, dado que no se realiza ninguna intervención sobre las participantes diferente a la realizada en una toma de citología cervicouterina de rutina, por lo cual se considera que los beneficios potenciales del estudio superan los potenciales riesgos del mismo. Adicionalmente, se asegurará que el personal que forme parte del estudio tenga una formación teórica y práctica rigurosa, con actualización permanente certificada.

- Justicia: Los investigadores serán equitativos al distribuir riesgos y beneficios potenciales de la investigación, ya que la selección de participantes será únicamente basada en los criterios de inclusión y exclusión mencionados en el apartado “selección de participantes”, sin miramientos sobre su condición social o étnica. Cabe mencionar que, para los propósitos del presente estudio, no fue posible incluir mujeres embarazadas, ya que algunos autores como Li Ma et al. (40) han descrito que algunos de los cambios a los que está sometido el cuello uterino durante el proceso de embarazo pueden ser confundidos con lesiones malignas, lo cual podría interferir con la validez de los datos obtenidos.

- Autonomía: Se refiere al respeto por los derechos de cada participante a tomar sus propias decisiones, que en el presente estudio se concreta con la obtención del consentimiento informado, comprensivo, competente y voluntario.

- Respeto: Los investigadores reconocen en todo momento la dignidad personal y autonomía de las participantes del estudio. Se respetará este principio al confirmar la existencia del formato de consentimiento informado de la citología recibida en el laboratorio de patología, donde conste la autorización para realizar el procedimiento diagnóstico, confirmando el uso de una codificación especial de cada muestra para emplearla de manera anónima.

6. Tratamiento de datos personales

Las participantes del estudio tienen derecho a conocer, actualizar y rectificar en cualquier momento la información que sobre ellas se haya recogido en la base de datos, de acuerdo al artículo 1 de la Ley 1581 del 2012. Adicionalmente, se respetaron los principios para el tratamiento de datos personales que consigna el artículo 4 de la misma ley, los cuales incluyen el principio de legalidad en materia de tratamiento de datos: El principio de finalidad, bajo el cual todos los datos obtenidos obedecerán a una finalidad legítima (proyecto de investigación); principio de libertad, según el cual el tratamiento de datos solo puede ejercerse con consentimiento previo expreso e informado del titular, y jamás se divulgarán datos sin previa autorización, o en ausencia del mandato legal o judicial que releve el consentimiento; principio de veracidad o calidad, según el cual la información consignada en la base de datos será veraz, completa, exacta, actualizada, comprobable y comprensible; principio de transparencia, según el cual cualquier participante del estudio podrá obtener, en cualquier momento y sin restricciones, información acerca de la existencia de datos que le conciernan; principio de acceso y circulación restringida, según el cual los datos personales serán salvaguardados por una única persona (investigador principal), y no estarán disponibles en internet u otros medios de divulgación ni comunicación masiva; principio de seguridad, según el cual se manejarán con medidas técnicas, humanas y administrativas los datos obtenidos para otorgar seguridad a los registros y evitar su adulteración, pérdida, consulta, uso o acceso no autorizado; y principio de confidencialidad, según el cual los investigadores estarán obligados a garantizar la reserva de la información, inclusive después de finalizada su relación con la investigación, pudiendo sólo realizar suministro o comunicación de datos

personales cuando ello corresponda al desarrollo de las actividades autorizadas por la Ley 1581 del 2012.

Adicionalmente, debido a que el presente estudio incluye datos sensibles según el Artículo 5 de la Ley 1581 del 2012, el uso de estos se hará única y exclusivamente en casos donde la Titular (o participante del estudio) haya dado su autorización explícita y escrita a dicho tratamiento. De forma similar, el tratamiento de los datos socio-demográficos, clínicos y hallazgos microscópicos es efectuado con finalidad estadística, académica y científica, adoptando las medidas conducentes a la supresión de identidad de las participantes (titulares). La identidad de las pacientes y sus datos personales no son del interés del grupo de investigación y no están incluidos dentro de las bases de datos utilizadas para el presente estudio. Por lo tanto, la información obtenida es completamente anónima. Solamente se utiliza el número consecutivo del extendido citológico y el año en que se realizó la toma de la misma, correspondiente a la nomenclatura asignada por protocolo institucional por el Departamento de Patología de la Universidad Industrial de Santander. En ningún momento de la investigación fue necesario contactar a las pacientes, lo que garantiza la protección de los datos personales.

La información obtenida dentro de la investigación fue salvaguardada por el investigador principal, y los datos obtenidos fueron custodiados durante el tiempo que abarcó la duración del proyecto de investigación bajo llave en un gabinete del Departamento de Patología de la Universidad Industrial de Santander.

7. Cronograma de actividades

ACTIVIDADES	Año 2022												
	Mes	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
1. Selección del tema		■											
2. Delimitación del tema			■										
3. Pregunta de investigación				■									
4. Objetivos de investigación					■								
5. Hipótesis						■							
6. Diseño de investigación							■						
7. Diseño del marco teórico								■	■				
8. Entrega y socialización del proyecto										■			
9. Entrega y aprobación del comité de ética											■	■	■
		Año 2023											
10. Recolección de datos		■	■	■	■	■	■	■	■	■	■	■	■
11. Trabajo de campo													■
12. Análisis de información													■
13. Entrega de resultados													
14. Correcciones													
15. Segunda entrega de resultados													
16. Sustentación del proyecto													

	Año 2024												
	Mes	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
11. Trabajo de campo		■											
12. Análisis de información		■	■	■									
13. Entrega de resultados				■									
14. Correcciones					■								
15. Segunda entrega de resultados						■							
16. Sustentación del proyecto								■	■	■	■	■	■

8. Presupuesto

Presupuesto global de la propuesta											
NOMBRE PROPUESTA:		INTELIGENCIA ARTIFICIAL APLICADA PARA LA DETECCIÓN DE ANORMALIDADES EN CITOLOGÍA CERVICO UTERINA EN UN CENTRO DE ATENCIÓN PRIMARIA DE BUCARAMANGA									
CONVOCATORIA:		INVESTIGACIÓN APLICADA									
Ítem	RUBRO	FINANCIACIÓN UIS 2022		FINANCIACIÓN UIS 2023		FINANCIACIÓN OTRA(S) INSTITUCIÓN(ES) 2022		FINANCIACIÓN OTRA(S) INSTITUCIÓN(ES) 2023		TOTAL	
		Efectivo	Especie	Efectivo	Especie	Efectivo	Especie	Efectivo	Especie		
1	Honorarios	\$	-	\$	-	\$	87.840.000	\$	-	\$	87.840.000
2	Auxiliares estudiantiles	\$	-	\$	-	\$	-	\$	-	\$	-
3	Riesgos laborales estudiantiles	\$	-	\$	-	\$	-	\$	-	\$	-
4	Servicios Técnicos	\$	-	\$	-	\$	2.000.000	\$	-	\$	2.000.000
5	Pasajes	\$	-	\$	600.000	\$	-	\$	-	\$	600.000
6	Máticos	\$	-	\$	-	\$	-	\$	-	\$	-
7	Gastos de viaje	\$	-	\$	300.000	\$	-	\$	-	\$	300.000
8	Equipos de laboratorio	\$	-	\$	-	\$	-	\$	-	\$	-
9	Maquinaria	\$	-	\$	-	\$	-	\$	-	\$	-
10	Prácticas docentes y salidas de campo	\$	-	\$	2.100.000	\$	-	\$	-	\$	2.100.000
11	Reparación y mantenimiento	\$	-	\$	-	\$	-	\$	-	\$	-
12	Reactivos y materiales de laboratorio	\$	-	\$	-	\$	-	\$	-	\$	-
13	Licencias de software	\$	-	\$	-	\$	4.469.581	\$	-	\$	4.469.581
14	Papelaria y útiles de escritorio	\$	-	\$	-	\$	750.000	\$	-	\$	750.000
15	Libros y material bibliográfico	\$	-	\$	-	\$	3.000.000	\$	-	\$	3.000.000
16	Equipo audiovisual	\$	-	\$	-	\$	-	\$	-	\$	-
17	Equipo de oficina	\$	-	\$	-	\$	-	\$	-	\$	-
18	Equipo de cómputo	\$	-	\$	-	\$	-	\$	-	\$	-
19	Muebles y enseres	\$	-	\$	-	\$	-	\$	-	\$	-
20	Elementos de laboratorio	\$	-	\$	-	\$	-	\$	-	\$	-
21	Herramientas	\$	-	\$	-	\$	-	\$	-	\$	-
22	Montaje e instalaciones	\$	-	\$	-	\$	-	\$	-	\$	-
23	Adecuaciones	\$	-	\$	-	\$	-	\$	-	\$	-
24	Arrendamientos	\$	-	\$	-	\$	-	\$	-	\$	-
25	Seguros	\$	-	\$	-	\$	-	\$	-	\$	-
26	Capacitación de personal	\$	-	\$	-	\$	-	\$	-	\$	-
27	Portes y fletes	\$	-	\$	-	\$	-	\$	-	\$	-
28	Acarreos	\$	-	\$	-	\$	-	\$	-	\$	-
29	Avisos e impresos	\$	-	\$	-	\$	-	\$	-	\$	-
30	Semovientes	\$	-	\$	-	\$	-	\$	-	\$	-
31	Gastos de im portación	\$	-	\$	-	\$	-	\$	-	\$	-
TOTAL		\$	-	\$	3.000.000	\$	98.059.581	\$	-	\$	101.059.581

9. Resultados y discusión

9.1 Análisis descriptivo de datos clínicos

Desde el 1 de enero hasta el 31 de diciembre del año 2023, se incluyeron en nuestro estudio un total de 80 participantes, de las cuales se recolectaron muestras y reportes de citologías. El resumen de los hallazgos poblacionales se muestra en las tablas 1 y 2. La media de edad fue de 21.64 ± 2.45 años. En la caracterización demográfica, el 95% provenían de zona urbana, reportando en un 65% (52) algún método de planificación, dentro de los cuales los más usados fueron pastillas anovulatorias (31 casos, 38.75%) seguidos por método de barrera (18 casos, 22.5%) y dispositivo intrauterino (2 casos, 2.5%).

Tabla 1.

Edad actual, edad de menarquia y edad de inicio de relaciones sexuales en nuestra población

Variable	Media	Rango
Edad actual	21.64 años	± 2.45 años
Edad de la menarquia	12.26 años	± 0.15 años
Edad de inicio de relaciones sexuales	17.21 años	± 2.14 años

Tabla 2.

Características sociodemográficas y clínicas

Variable	Categorías	No. Casos	Porcentaje
Residencia	Urbana	77	95%
	Rural	3	5%
Gestaciones	Ninguna	71	88.75%
	1	8	10%
	2	1	1.25%
Partos	Ninguno	78	97.5%
	1 o más	2	2.5%
Abortos	Ninguno	74	92.5%
	1 o más	6	7.5%
Método de planificación	Ninguno	28	35%
	Pastillas anovulatorias	31	38.75%
	Dispositivo intrauterino (DIU)	2	2.5%
	Quirúrgico	1	1.25%
	Barrera (condón)	18	22.5%
Citología previa	Si	41	51.25%
	No	34	42.5%
Vacuna contra VPH	Si	64	80%
	No	16	20%
Número de dosis de vacuna contra VPH	1	15	19%
	2	35	44%
	3	8	10%

El 51% reportaron haberse realizado citología cervicouterina previamente, dentro de las cuales sólo un 14.6% presentaron anomalías en resultados de citología previa. En cuanto al

esquema de vacunación contra el VPH inmerso en el programa ampliado de inmunización (PAI) del Ministerio de Salud de Colombia desde el año 2013, el 80% (64 pac.) reportaron haber sido vacunadas, con un 10% (8 pac) prevalente para esquema completo de 3 dosis, un 44% (35 pac.) prevalente para 2 dosis, y un 19% (15 pac) con reporte de haber recibido sólo 1 dosis. Sólo 1 participante reportó haberse realizado prueba de ADN para VPH, con resultado negativo.

Al momento de la toma de la muestra, en el 28.75% (23 pac) de los especímenes se reportaron alteraciones macroscópicas en el cérvix, que correspondieron a cérvix de aspecto congestivo, con erosiones y/o con sangrado; y en un 17% (4 pac) se reportó leucorrea macroscópica.

Todas las muestras recibidas contaron con una calidad satisfactoria de acuerdo con los criterios del sistema Bethesda (32). Entre estas, el 17.5% (14 lam.) fueron categorizadas por el patólogo como alteradas, de las cuales, el 93% (13 lam.) correspondieron a ASCUS, y sólo 1 (7%) a lesión intraepitelial de bajo grado. Se reportó la presencia de microorganismos en un 28.75% (23 lam.), siendo los más frecuentes bacterias (82.6%), seguido por candida (26%). Se observó también coinfección bacterias/candida en 2 casos. Los hallazgos no neoplásicos más frecuentes fueron la presencia de cambios relacionados con inflamación (19 casos, 23.75%) seguidos por cambios de metaplasia escamosa (9 casos, 11.25%).

9.2 Análisis bivariado de datos sociodemográficos con cambios preneoplásicos en citología

La tabla 3 resume los hallazgos entre las variables sociodemográficas y su relación con la variable desenlace (cambios preneoplásicos, incluyendo ASCUS y lesión intraepitelial de bajo grado).

Al analizar la edad de la menarquia, se esboza una asociación entre la menarquia temprana y cambios preneoplásicos en citología cervicouterina. Se estableció que, por cada año de retraso en la edad de la menarquia, el riesgo relativo de desarrollar lesiones preneoplásicas disminuyó en forma representativa (OR 0.58, IC 95%: [0.3595±0.9365] con valor de $p=0.02$). La asociación estadística más fuerte fue encontrada en casos con menarquia igual o menor a 10 años, para los cuales el riesgo incrementó 4.65 veces con respecto a la población general ($p=0.06$), aunque el intervalo de confianza fue amplio debido a la escasa representación de este grupo (IC 95%: [0.9027, 23.9522]). Esta tendencia ha sido observada en diferentes cohortes a nivel mundial, en las cuales se han encontrado asociaciones fuertes entre una edad de menarquia más temprana y riesgo de desarrollar cáncer de cérvix, endometrio, glándula mamaria y ovario, entre otros (41, 42, 43). En particular, la asociación entre una menarquia más temprana y el riesgo de desarrollar cáncer de cérvix ha sido observada en diferentes cohortes de Turquía (44) e India (45).

Al analizar las variables de factores sexuales y reproductivos, se encontró que en nuestra población un 35% (28 pac.) no usan ningún método de planificación, y sólo un 22.5% (18 pac.) reportaron usar método de barrera. Adicionalmente, la edad promedio de inicio de relaciones sexuales es de 17 años (± 2.14 años). La prevalencia de maternidad previa fue de un 11.25% (9 pacientes) dentro de lo cual 6 (7.5%) reportaron haber tenido abortos previos. En cohortes a nivel mundial (44, 45), la historia de maternidad y abortos previos, así como el inicio de relaciones sexuales antes de los 18 años, han sido asociados con un incremento en el riesgo de cáncer de cuello uterino. Estos datos sugieren que las mujeres con vida sexual activa están en riesgo mayor de desarrollar cáncer de cérvix, debido al riesgo de exposición a VPH, sin embargo, en nuestra población no se encontraron asociaciones significativas entre estas variables y aumento de riesgo de tener alteraciones en citología. Es posible que esto se deba a que nuestra cohorte está compuesta

de mujeres jóvenes, en las cuales la literatura ha demostrado que hasta el 90% de infecciones por VPH desaparecen espontáneamente entre 18 meses y 5 años después de su inicio, y las alteraciones en citología son expresión de la infección transitoria (3).

Al evaluar la relación entre vacunación para VPH y cambios en citología, se encontró que la historia de vacunación, así como el número de dosis de vacuna contra el VPH no se asocian de forma significativa con disminución o aumento en el riesgo de alteraciones en citología en nuestra población. La literatura respecto a este tema plantea que en mujeres en edad reproductiva con vacunación contra VPH, las alteraciones citológicas pueden deberse a exposición a VPH antes de la vacunación y/o efectos de diferentes tipos de VPH no cubiertos por la vacuna (46). Un estudio de genotipificación de VPH realizado a 810 mujeres de la comuna Norte de Bucaramanga, llevado a cabo por investigadores de la Universidad Industrial de Santander, mostró que el genotipo de alto riesgo más común fue VPH-59 (17%) seguido por VPH-58 (14%) y de bajo riesgo fueron VPH-62/81 (30%) y VPH44/55 (17%) (47), lo que contrasta con lo reportado por la IARC en Colombia, quienes reportan que el VPH-16 es el más prevalente (17.6%), seguido de VPH-31 (8.5%), VPH-18 (7%) y VPH-33 (5.9%) (48). En nuestra población está disponible la vacuna nonavalente que protege contra los serotipos 6, 11, 16, 18, 31, 33, 45, 52 y 58 (49). Sin embargo, esta vacuna no protege contra el serotipo de alto riesgo VPH-59 ni contra los serotipos de bajo riesgo circulantes en nuestro medio, como el VPH-62/81 y el VPH-44/55, lo que puede explicar que la vacunación no sea un factor protector significativo en nuestra población.

Tabla 3.

Análisis bivariado de alteraciones en citología y variables sociodemográficas

VARIABLES	CATEGORÍAS	CITOLOGÍA normal	CITOLOGÍA alterada	OR	Valor de p	IC 95%
Residencia	Urbana	63	14			
	Rural	3	0		0.41	
Gestaciones	0	60	12	0.2	0	0.1076-0.3717
	1	6	1	1.66	0.56	0.2995-9.2716
Abortos	0	63	12			
	1	4	2	2.62	0.29	0.4312-15.9783
Edad de menarquia	Media de edad por desenlace	Edad media de 12.42+1.35 años	Edad media de 11.46+1.19 años	0.58	0.02	0.3595+0.9365
	≤10 años			4.65	0.06	0.9027-23.9522
Método de planificació n	Ninguno	23	5	1.10	0.88	0.2963-4.1121
	Anovulatorios	25	6			
	DIU	2	0			
	Quirúrgico	1	0			
	Condón	16	2	0.57	0.53	0.0989-3.3407
Resultado de citología previa	Normal	23	3			
	Alterada	4	1	1.91	0.61	0.1573-23.3466
Vacuna contra VPH	Ninguna dosis	12	3	0.72	0.66	0.1734-3.0493
	1 o más dosis	55	10			

9.3 Análisis bivariado entre variables de la muestra de citología y cambios preneoplásicos

La tabla 4 resume los hallazgos entre las variables tomadas de la muestra de citología y los cambios preneoplásicos reportados. Los hallazgos a la evaluación macroscópica del cérvix mostraron una asociación fuerte y estadísticamente significativa con el resultado de la citología cervicouterina. Se observó una probabilidad del 82.20% y un riesgo relativo 4.62 veces mayor (IC 95%: [1.3861, 15.4128], con valor de $p=0.01$) que las alteraciones macroscópicas se asocien con cambios preneoplásicos en citología. Entre estas alteraciones, la erosión del cérvix se asoció con una mayor probabilidad (89.55%) y un riesgo 8.57 veces mayor (IC 95%: [2.3178, 31.6976] con valor de $p=0.001$) de alteraciones preneoplásicas. De forma similar, la presencia de sangrado en el cérvix se asoció a un riesgo 8.75 veces mayor (IC 95%: [1.9744, 38.7771] con valor de $p=0.004$) de alteraciones. Estos resultados enfatizan la importancia de la evaluación macroscópica del cérvix, incluso cuando el resultado de la citología sea negativo. Un estudio descriptivo realizado en 231 mujeres (50) mostró que los cambios macroscópicos del cérvix se asociaron significativamente con displasia en la biopsia, mientras que la citología sólo mostró una sensibilidad de 55.5% y especificidad de 75%.

Tabla 4.

Análisis bivariado de alteraciones en citología y variables de la muestra.

Variables	Categorías	Citología	Citología	OR	Valor de p	IC 95%
		normal	alterada			
Características macroscópicas del cérvix	Normal	52	6			
	Alterado	15	8	4.62	0.009	1.3861-15.4128
Cérvix erosionado	No	60	7			
	Si	7	7	8.57	0.001	2.3178-31.6976
Cérvix con sangrado	No	63	9			
	Si	4	5	8.75	0.004	1.9744-38.7771
Cérvix con leucorrea	No	54	12	0.69	0.65	0.1377-3.4800
	Si	13	2			
Muestra con microorganismos	No	49	9			
	Si	18	5	1.51	0.50	0.4467-5.1195
Presencia de cándida	No	63	12			
	Si	4	2	2.62	0.29	0.4312-15.9783
Presencia de vaginosis bacteriana	No	52	10			
	Si	15	4	1.38	0.62	0.3801-5.0577

9.4 Resultados de Inteligencia Artificial (IA)

9.4.1 Desempeño de los modelos de IA con la base de datos de entrenamiento

Los resultados del entrenamiento de los cuatro modelos de IA usando la base de datos de referencia se muestran en la tabla 5. La mejor sensibilidad fue obtenida por InceptionV3, (0.96), seguido por VGG19, DenseNet y MobileNet con un 0.95, 0.94 y 0.92 respectivamente, con mejor desempeño en especificidad para MobileNet, de 0.92. En general, todos los modelos contaron con un excelente valor de área bajo la curva (0.96 - 0.98), lo que resalta su capacidad como clasificadores binarios para diferenciar células normales y células con alteraciones.

Tabla 5.

Desempeño de los modelos de IA con la base de datos de entrenamiento

Modelo	Sensibilidad	Especificidad	VPP	VPN	AUC (Curva ROC)
DenseNet	0.94	0.90	0.90	0.95	0.98
InceptionV3	0.96	0.89	0.88	0.96	0.98
VGG19	0.95	0.84	0.82	0.96	0.96
MobileNet	0.92	0.92	0.92	0.92	0.97

Debido a que sólo vemos el resultado del entrenamiento, es difícil entender cómo los modelos de IA llegaron a tomar las decisiones necesarias para clasificar las células como normales o alteradas, y si esos métodos fueron similares o diferentes a los que usan los patólogos. Una forma de lograr entender este proceso es mediante el uso de mapas de activación de clase ponderada por gradiente (Grad-CAM), los cuales resaltan las regiones más importantes de la imagen para su

clasificación usando un gradiente de colores (51). La figura 8 muestra un ejemplo de células normales clasificadas como normales, y la figura 9 muestra células con alteraciones citológicas clasificadas adecuadamente por DenseNet.

Figura 8.

Ejemplo de células normales clasificadas correctamente por DenseNet

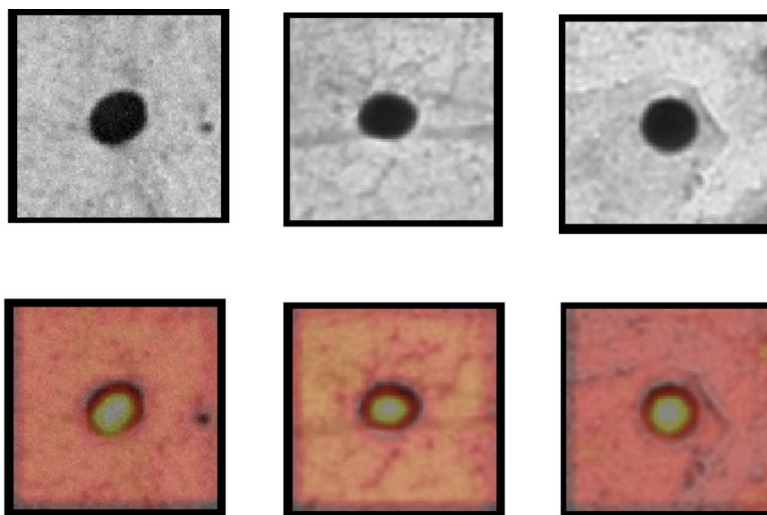
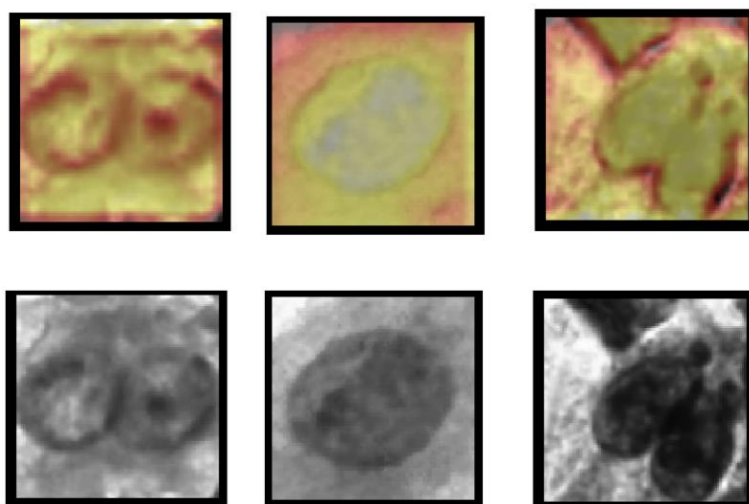


Figura 9.

Ejemplo de células con alteraciones, clasificadas correctamente por DenseNet



Estas gráficas se crean usando un gradiente de colores que va desde negro y rojo para las áreas menos importantes, hasta amarillo y blanco para las áreas más relevantes. Al comparar estas dos figuras, podemos observar que el tamaño y la forma del núcleo cobran una gran importancia a la hora de clasificar una célula. En la figura 8 la mayor parte de la activación está centrada en el núcleo y en su contorno, que, al ser redondeado, pequeño y regular, lleva a clasificar esta célula como normal. En la figura 9 podemos observar una activación más heterogénea, con puntos de mayor relevancia estadística centrados en alteraciones de la cromatina nuclear, el contorno nuclear, y el citoplasma. Por tanto, podemos concluir que las variaciones en la densidad de la cromatina, alteraciones en el contorno nuclear y una menor cantidad de citoplasma (que podría traducirse en alteraciones en la relación núcleo-citoplasma) llevan a clasificar estas células como alteradas. Estos parámetros están en concordancia con los usados por los patólogos para diagnosticar anomalías en citología cervicouterina, estandarizados por el sistema Bethesda (32).

9.4.2 Desempeño de los modelos de IA con la base de datos de prueba

De las 800 células individuales de la muestra total (10 por cada caso), 554 (69%) fueron clasificadas por el patólogo experto y el residente de patología como células normales, y las 246 restantes (31%) fueron clasificadas dentro del grupo de atipia citológica, usando los criterios del sistema Bethesda (32).

Tras finalizar el entrenamiento de los modelos y verificar que tuvieran un adecuado desempeño con la base de datos de referencia, probamos su desempeño en la base de datos de nuestra población. La totalidad de las células fueron usadas como set de prueba, y los resultados de sensibilidad, especificidad, valor predictivo positivo, valor predictivo negativo, área bajo la

curva y tasa de acuerdo interobservador calculado usando el coeficiente Kappa se muestran en la tabla 6. La matriz de confusión de cada modelo se muestra en la tabla 7, y las curvas ROC-AUC para cada modelo se muestran en las figuras 10-13. Finalmente, el tiempo de clasificación de cada modelo para la totalidad de los datos se muestra en la tabla 8.

Tabla 6.

Desempeño de los cuatro modelos de IA con la base de datos de prueba

Modelo	Sensibilidad	Especificidad	VPP	VPN	AUC (Curva ROC)	Coeficiente Kappa
DenseNet	0.83	0.91	80%	92%	0.931	0.736
Inception V3	0.93	0.77	64%	96%	0.929	0.617
VGG19	0.85	0.89	77%	93%	0.919	0.714
MobileNet	0.78	0.97	91%	91%	0.952	0.777

Tabla 7.

Matriz de confusión de los cuatro modelos de IA con la base de datos de prueba, comparados con el Gold Standard (lectura manual)

Modelo de Inteligencia Artificial		Gold Standard	
		Alteración	Normal
DenseNet	Alteración	205	50
	Normal	41	504
InceptionV3	Alteración	228	129
	Normal	18	425

Modelo de Inteligencia		Gold Standard	
Artificial		Alteración	Normal
VGG19	Alteración	208	62
	Normal	38	492
MobileNet	Alteración	192	19
	Normal	54	535

Tabla 8.

Tiempos de clasificación para cada modelo

Modelo	Tiempo de clasificación - células alteradas	Tiempo de clasificación - células normales	Tiempo de clasificación total
DenseNet	42 segundos	50 segundos	1 minuto, 32 segundos
InceptionV3	30 segundos	50 segundos	1 minuto, 20 segundos
VGG19	1 minuto, 57 segundos	3 minutos, 58 segundos	5 minutos, 55 segundos
MobileNet	20 segundos	26 segundos	46 segundos

Figura 10.

Curva ROC para DenseNet

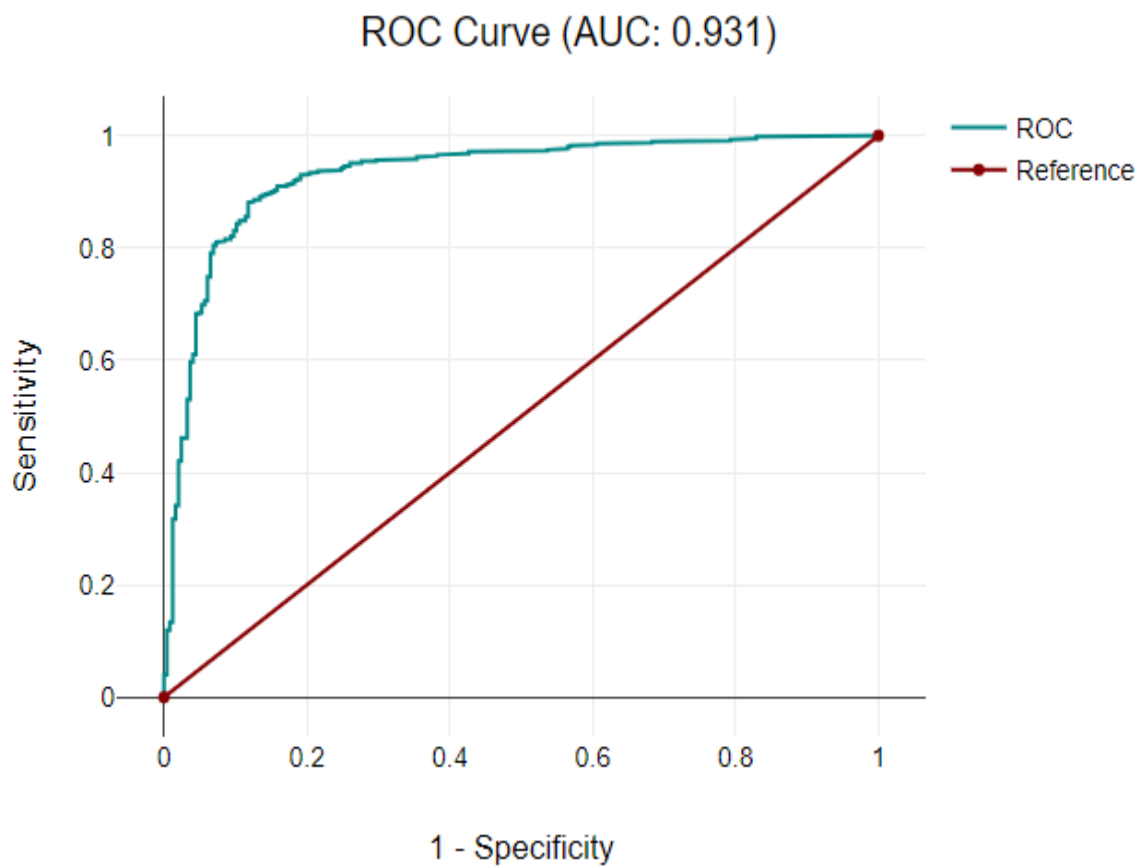


Figura 11.

Curva ROC para InceptionV3

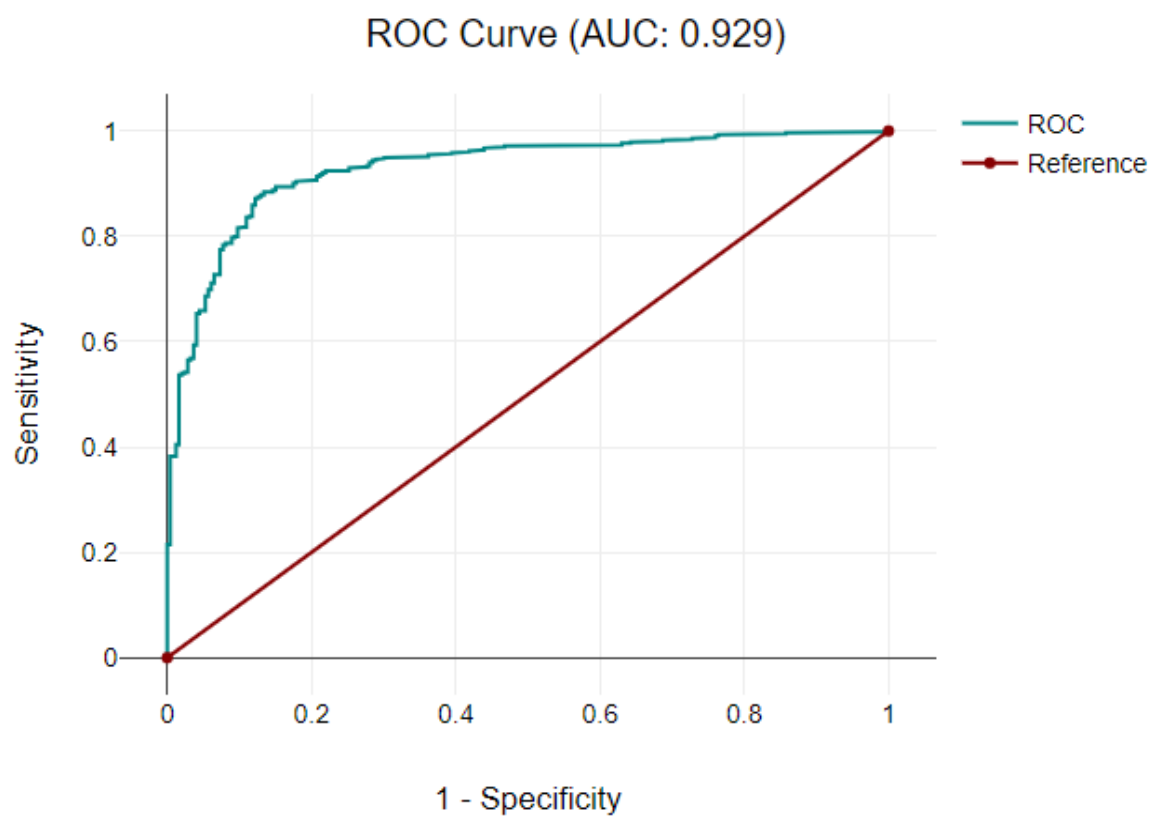


Figura 12.

Curva ROC para VGG19

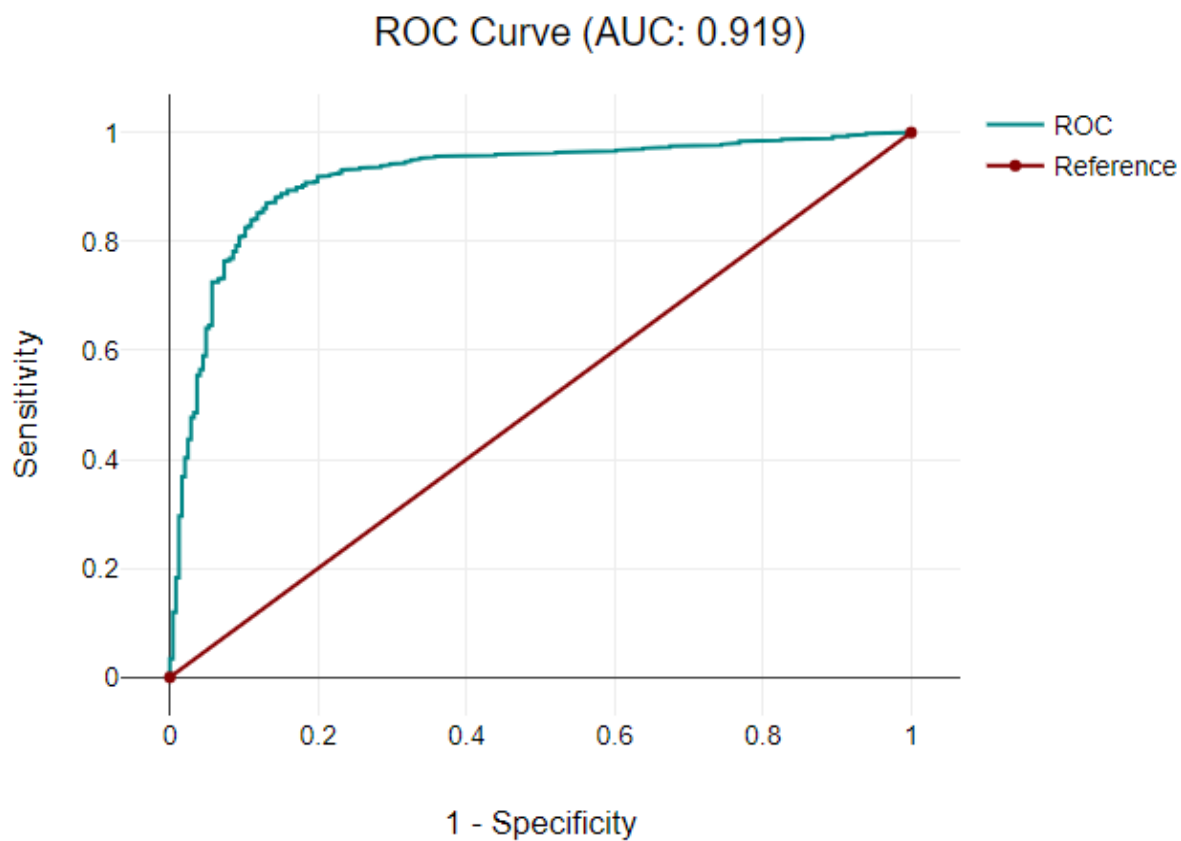
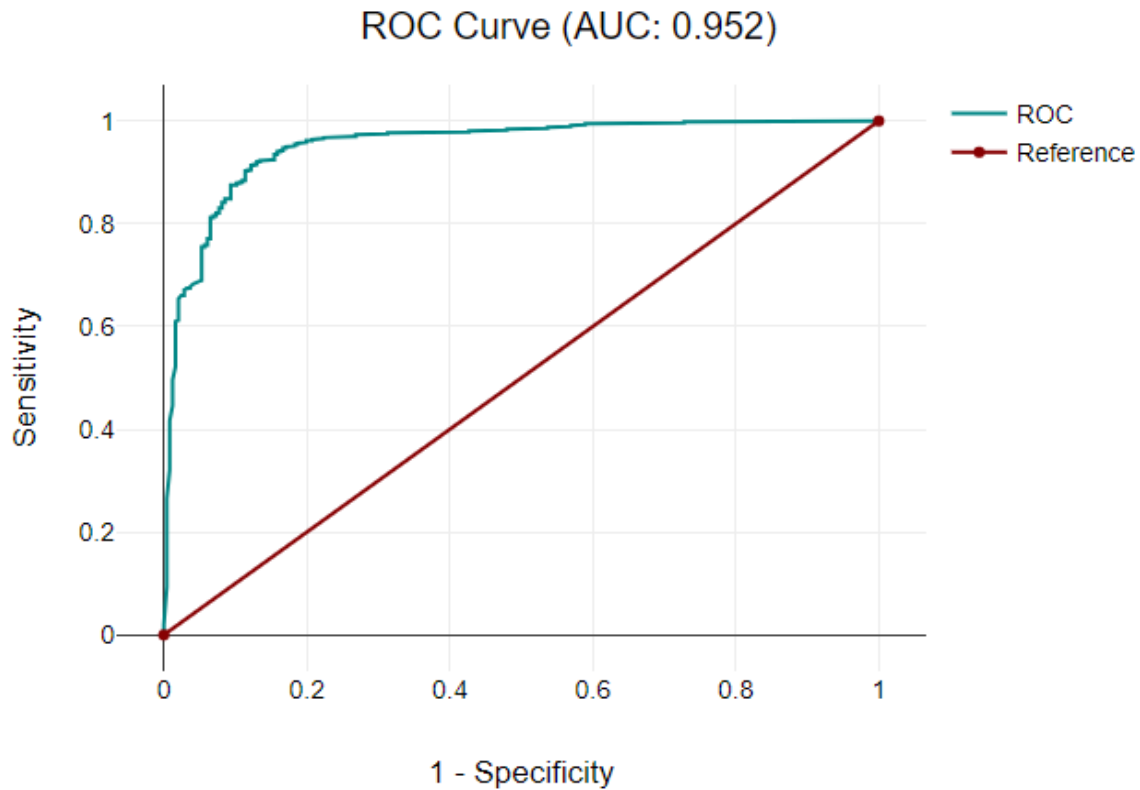


Figura 13.

Curva ROC para MobileNet



Los cuatro modelos mostraron un desempeño excelente como clasificadores binarios, evidenciado por sus valores en la curva de característica operativa del receptor (ROC), el cual fue superior a 0.9 en todos los casos. Sin embargo, el modelo que tuvo la mejor área bajo la curva fue MobileNet, de 0.952. Este modelo también se destacó por tener la mejor especificidad, de 0.97; sin embargo, tuvo la sensibilidad más baja de todos los modelos, de 0.78. Esto puede deberse a que MobileNet fue diseñada con convoluciones separables en profundidad, que permiten disminuir el número de parámetros y la carga computacional, optimizando la eficiencia del modelo. Esta optimización se logra al evitar sobrecargar el dispositivo con predicciones erróneas. Sin embargo,

este proceso puede causar que se pierdan características complejas de una imagen que pueden ser críticas para clasificar correctamente los casos positivos, lo que disminuye su sensibilidad.

Debido a que la prueba de citología cervicouterina funciona como una prueba de tamizaje diagnóstico, la prueba elegida debería tener una sensibilidad mayor, con el fin de poder detectar la mayor cantidad de pacientes con resultados verdaderamente positivos. Para este propósito el mejor modelo de IA fue InceptionV3, el cual mostró una mejor sensibilidad tanto en el set de entrenamiento como en el set de prueba (0.96 y 0.93, respectivamente), a costa de una especificidad y un valor predictivo positivo menores que otros modelos. Esto se debe a la arquitectura de InceptionV3, la cual es una de las más complejas hasta el momento, principalmente gracias al uso de capas más profundas y con mayor número de parámetros, que le permiten aprender características y relaciones más complejas entre los datos. Adicionalmente, InceptionV3 está mejor diseñada para evitar el overfitting (17, 36) lo que permite generalizar mejor a nuevos datos, como en el caso de nuestro set de prueba. Sin embargo, el tener una arquitectura tan compleja puede causar que el modelo sea más propenso a identificar características sutiles que no son tan relevantes a la hora de clasificar las células de nuestro estudio, por lo que el modelo interpreta erróneamente datos poco relevantes como indicativos de la clase positiva, reduciendo la especificidad.

En cuanto a la tasa de acuerdo interobservador, medida con el coeficiente Kappa, todos los modelos mostraron una tasa de acuerdo sustancial, sin embargo, InceptionV3 tuvo el índice más bajo, de 0.617. En la práctica, este valor puede ser problemático, debido a que significaría que casi un 40% de los casos estarían errados al compararse con la lectura manual (52). La mejor tasa de acuerdo interobservador fue lograda por el modelo MobileNet, con un 0.777. La aplicación práctica de este modelo tendría la ventaja de lograr una mayor especificidad, lo que ayudaría a

disminuir la cantidad de falsos positivos, y por tanto disminuiría los potenciales costos que las pruebas adicionales podrían acarrear al sistema de salud (como referencia a especialista y colposcopia), sin embargo, una menor sensibilidad podría traducirse en una mayor tasa de falsos negativos.

En cuanto al tiempo de computación necesario para el diagnóstico, el mejor modelo fue nuevamente MobileNet. La arquitectura de este modelo reduce drásticamente la capacidad de computación requerida, ya que fue diseñado para aplicaciones en plataformas móviles. Teniendo en cuenta que este tiempo fue empleado para clasificar un total de 800 células, y que, de acuerdo al sistema Bethesda, una citología normal tiene entre 8000 y 20.000 células (32), este modelo podría tardar entre 460 y 1150 segundos (o entre 7 y 19 minutos) para clasificar la totalidad de las células encontradas en un extendido rutinario. En comparación, los modelos de U-net (55) y SegNet (56) toman 4370 y 4024 segundos respectivamente para procesar y segmentar la totalidad de un extendido escaneado (Whole-Slide imaging), mientras que modelos más rápidos, como el desarrollado por Araujo et al. (57) y Wang et al. (53) toman 219 y 210 segundos respectivamente en realizar la misma tarea.

En este estudio se realizó la toma de fotografías y segmentación celular de forma manual, un paso que toma tiempo y limita la practicidad que este sistema puede tener en la práctica diaria. Para solventar este problema, diferentes autores han desarrollado dos metodologías principales: una es escanear la totalidad de la lámina (whole-slide imaging, WSI por sus siglas en inglés) para posteriormente separar y clasificar las células mediante IA. Esta metodología es la más usada en diferentes estudios (53) y fue aplicada por Bao et al. (4) en un estudio de cohorte a gran escala que incluyó 0.7 millones de mujeres en China, logrando mejoría en la sensibilidad diagnóstica con un índice de acuerdo entre el diagnóstico por lectura manual y por IA del 99%. El segundo método

consiste en el uso de realidad aumentada con clasificación de células por IA en tiempo real para asistir a la lectura manual. Esta metodología tiene la ventaja de ser fácilmente aplicable a la práctica diaria, y un estudio conducido por Tang et al (54), demostró que podía aumentar la sensibilidad y la tasa de acuerdo interobservador en citología cervicouterina, con un modelo de IA cuya sensibilidad fue del 0.90 y un área bajo la curva de 0.81, valores ligeramente inferiores a los obtenidos en nuestro estudio.

10. Conclusiones

En nuestra población, el cáncer de cérvix continúa siendo un problema de salud pública de primer orden, a pesar del amplio conocimiento con el que contamos sobre la etiología, fisiopatología, cambios premalignos e historia natural de esta enfermedad. Colombia se enfrenta a múltiples problemas a la hora de establecer planes de tamizaje efectivos que logren disminuir la incidencia y mortalidad por esta causa, debido principalmente a limitaciones en la organización de los sistemas de salud, aspectos culturales, alta proporción de personas que viven en zonas rurales y difícil acceso a sistemas de salud.

A pesar del avance de la tecnología y el desarrollo de pruebas diseñadas para la detección y genotipificación de VPH, la citología cervicouterina convencional sigue siendo uno de los pilares principales para el tamizaje de cáncer de cérvix, sobretodo en mujeres menores de 30 años. Sin embargo, su utilidad se ve limitada debido a la alta carga laboral que enfrentan los citólogos y patólogos que evalúan estas muestras, junto al bajo rendimiento diagnóstico reportado en la literatura. En este estudio planteamos el uso de inteligencia artificial para solventar este problema.

Los resultados obtenidos muestran que los distintos modelos de clasificación de imágenes por IA usados tienen un área bajo la curva superior a 0.9 para diferenciar células normales y células con alteraciones morfológicas, con valores de sensibilidad y especificidad comparables con los reportados para la lectura convencional. De los modelos evaluados, MobileNet logró la mejor área bajo la curva, los mejores valores predictivos positivo y negativo, así como la mejor tasa de acuerdo interobservador en menos tiempo que otros modelos, por lo que consideramos que podría ofrecer el mejor valor costo-beneficio y mayor eficiencia a largo plazo, al tiempo que su menor

demanda de poder computacional hace que sea más factible integrar su uso al flujo de trabajo convencional.

La forma en la que puede aplicarse depende de los medios con los que cuente el laboratorio de patología. En laboratorios que cuenten con medios para escanear la totalidad de la lámina de citología, se podrían clasificar la totalidad de las células y obtener un resultado por cada caso, lo que podría incluso llegar a servir como un método de tamizaje primario para descartar casos negativos, disminuyendo la carga laboral de los patólogos y citólogos, sin alterar el rendimiento diagnóstico. La segunda posibilidad sería integrar los modelos de IA dentro de un software de realidad aumentada, lo que podría servir para asistir a la lectura manual, mejorando la sensibilidad a un costo relativamente bajo.

A pesar que los resultados son prometedores, todavía se necesitan estudios adicionales para evaluar estos escenarios en la práctica diaria del laboratorio de patología.

11. Trabajos futuros²

El campo de investigación en inteligencia artificial y su aplicación en patología se encuentra aún en una etapa inicial, y muchos son los estudios que pueden ser realizados para avanzar nuestro conocimiento en esta área, ya que tiene un gran potencial en casi todos los campos de la patología.

En cuanto al uso de inteligencia artificial en citología cervicouterina, consideramos que futuros trabajos podrían explorar también la clasificación de imágenes de campo completo (por ejemplo, escaneando la totalidad de la lámina) y con varios conjuntos de células, incluyendo no sólo casos positivos y negativos, sino también todas las categorías del sistema Bethesda, así como identificación de microorganismos, con el fin de poder acercarse más al trabajo real del patólogo.

De manera similar, podría explorarse la aplicación de inteligencia artificial en otro tipo de estudios de citología. Por ejemplo, citología de glándula mamaria, de tiroides, de páncreas, de orina o de médula ósea (mielograma). El campo de la citología es ideal para los estudios de inteligencia artificial gracias al alto volumen de trabajo y la facilidad de adquisición de las imágenes, que permite entrenar modelos con sets de datos grandes y completos.

Adicionalmente, y con el avance de la patología digital, futuros estudios podrían explorar la aplicación de inteligencia artificial en la lectura e interpretación de biopsias para detectar diferentes tipos de cáncer, por ejemplo cáncer de mama, de pulmón o de próstata, en los cuales importa tanto la arquitectura del órgano como el detalle celular, por lo que se podría investigar el uso de IA para detectar tanto tipo histológico como grado de diferenciación.

Finalmente, hay un gran potencial de investigación en cuanto al uso de la IA para la interpretación de estudios de inmunohistoquímica (por ejemplo, la interpretación de marcadores como HER2 o PD-L1).

12. Conflicto de interés

Los autores de este estudio no tienen ningún conflicto de interés que declarar.

Bibliografía

- 1: World Health Organization (WHO). Global Health Estimates 2020: Deaths by Cause, Age, Sex, by Country and by region, 2000-2019. WHO; 2020.
- 2: Cáncer cervicouterino - OPS/OMS | Organización Panamericana de la Salud. (2022). Retrieved 13 June 2022, from <https://www.paho.org/es/temas/cancer-cervicouterino>
- 3: Minsalud. Guía de práctica clínica para la detección y manejo de lesiones precancerosas de cuello uterino. 2014; 44: 26-34.
- 4: Bao, H., Sun, X., Zhang, Y., Pang, B., Li, H., & Zhou, L. et al. (2020). The artificial intelligence-assisted cytology diagnostic system in large-scale cervical cancer screening: A population-based cohort study of 0.7 million women. *Cancer Medicine*, 9(18), 6896-6906. doi: 10.1002/cam4.3296
- 5: Campanella, G., Hanna, M.G., Geneslaw, L. et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat Med* 25, 1301–1309 (2019). <https://doi.org/10.1038/s41591-019-0508-1>
- 6: Ehteshami Bejnordi B, Veta M, Johannes van Diest P, et al. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. *JAMA*. 2017;318(22):2199–2210. doi:10.1001/jama.2017.14585
- 7: Bao, H., Bi, H., Zhang, X., Zhao, Y., Dong, Y., & Luo, X. et al. (2020). Artificial intelligence-assisted cytology for detection of cervical intraepithelial neoplasia or invasive cancer: A multicenter, clinical-based, observational study. *Gynecologic Oncology*, 159(1), 171-178. doi: 10.1016/j.ygyno.2020.07.099

8. DeMay RM. Common problems in Papanicolaou smear interpretation. Arch Pathol Lab Med. 1997; 121(3):229-38
9. Rezende, M. T., Tobias, A. H. G., Silva, R., Oliveira, P., Sombra de Medeiros, F., Ushizima, D., ... Bianchi, A. G. C.. (2020). CRIC Cervix Cell Classification. doi: 10.6084/m9.figshare.c.4960286.v2
10. Arteaga-Arteaga HB, Mora-Rubio A, Florez F, et al. Machine learning applications to predict two-phase flow patterns. PeerJ Comput Sci. 2021;7:e798. Published 2021 Nov 29. doi:10.7717/peerj-cs.798
11. Vélez JI. Machine Learning based Psychology: Advocating for A Data-Driven Approach. Int J Psychol Res (Medellin). 2021;14(1):6-11. doi:10.21500/20112084.5365
12. Martinez O, Martinez C, Parra CA, Rugeles S, Suarez DR. Machine learning for surgical time prediction. Comput Methods Programs Biomed. 2021;208:106220. doi:10.1016/j.cmpb.2021.106220
13. Departamento Nacional de Planeación. Dirección de Desarrollo Rural Sostenible – DDRS Equipo de la Misión para la Transformación del Campo. Definición de categorías de Ruralidad. Disponible en: <https://colaboracion.dnp.gov.co/CDT/Estudios%20Economicos/2015ago6%20Documento%20de%20Ruralidad%20-%20DDRS-MTC.pdf>
14. Fondo Colombiano de Enfermedades de alto costo, 2015. Situación del cáncer de la población atendida en el SGSSS en Colombia, 191-206
15. Dharmadhikari, S. C., Ingle, M., and Kulkarni, P. (2011) “Empirical studies on machine learning based text classification algorithms.” Advanced Computing 2 (6): 161

16. Jordan, M., & Mitchell, T. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255-260. doi: 10.1126/science.aaa8415
17. Cao, C., Liu, F., Tan, H., Song, D., Shu, W., Li, W., Zhou, Y., Bo, X., & Xie, Z. (2018). Deep Learning and Its Applications in Biomedicine. *Genomics, proteomics & bioinformatics*, 16(1), 17–32. <https://doi.org/10.1016/j.gpb.2017.07.003>
18. Xin, M., Wang, Y. Research on image classification model based on deep convolution neural network. *J Image Video Proc.* 2019, 40 (2019). <https://doi.org/10.1186/s13640-019-0417-8>
19. Kumar, L., Renuka, D., Rose, S., Shunmuga priya, M., & Wartana, I. (2022). Deep learning based assistive technology on audio visual speech recognition for hearing impaired. *International Journal Of Cognitive Computing In Engineering*, 3, 24-30. doi: 10.1016/j.ijcce.2022.01.003
20. Popel, M., Tomkova, M., Tomek, J. et al. Transforming machine translation: a deep learning system reaches news translation quality comparable to human professionals. *Nat Commun* 11, 4381 (2020). <https://doi.org/10.1038/s41467-020-18073-9>
21. K B, P., & J, M. (2020). Design and Evaluation of a Real-Time Face Recognition System using Convolutional Neural Networks. *Procedia Computer Science*, 171, 1651-1659. doi: 10.1016/j.procs.2020.04.177
22. Cui, M., Zhang, D.Y. Artificial intelligence and computational pathology. *Lab Invest* 101, 412–422 (2021). <https://doi.org/10.1038/s41374-020-00514-0>
23. Komura D, Ishikawa S. Machine learning Methods for histopathological image analysis. *Comput Struct Biotechnol J.* 2018;9;34-42

24. Komura D, Ishikawa S. Machine learning approaches for pathologic diagnosis. *Virchows Arch.* 2019;475:131-8
25. Korbar B, Olofson AM, Miraflor AP, et al. Deep Learning for Classification of Colorectal Polyps on Whole-slide Images. *J Pathol Inform.* 2017;8:30. Published 2017 Jul 25. doi:10.4103/jpi.jpi_34_17
26. Bychkov, D., Linder, N., Turkki, R. et al. Deep learning based tissue analysis predicts outcome in colorectal cancer. *Sci Rep* 8, 3395 (2018). <https://doi.org/10.1038/s41598-018-21758-3>
27. López Jacqueline et al, integración viral y cáncer de cuello uterino. *Rev Col Cienc Quim Farm.* Vol 35 (1), 5-32
28. Burd, E. (2003). Human Papillomavirus and Cervical Cancer. *Clinical Microbiology Reviews*, 16(1), 1-17. doi: 10.1128/cmr.16.1.1-17.2003
29. Anexo Técnico No. 1 de la resolución N° 1043 de abril 3 de 2006, “Por la cual se establecen las condiciones que deben cumplir los Prestadores de Servicios de Salud para habilitar sus servicios e implementar el componente de auditoria para el mejoramiento de la calidad de la atención y se dictan otras disposiciones”
30. N. Diniz, D., T. Rezende, M., G. C. Bianchi, A., M. Carneiro, C., J. S. Luz, E., & J. P. Moreira, G. et al. (2021). A Deep Learning Ensemble Method to Assist Cytopathologists in Pap Test Image Classification. *Journal Of Imaging*, 7(7), 111. doi: 10.3390/jimaging7070111
31. Rezende, M. T., Tobias, A. H. G., Silva, R., Oliveira, P., Sombra de Medeiros, F., Ushizima, D., ... Bianchi, A. G. C.. (2020). CRIC Cervix Cell Classification. doi: 10.6084/m9.figshare.c.4960286.v2
32. The Bethesda System for Reporting Cervical Cytology. (2015). doi: 10.1007/978-3-319-11074-5

33. Kräter, M., Abuhattum, S., Soteriou, D., Jacobi, A., Krüger, T., Guck, J., & Herbig, M. (2021). AIDeveloper: Deep Learning Image Classification in Life Science and Beyond. *Advanced Science*, 8(11), 2003743. doi: 10.1002/advs.202003743
34. Mikolajczyk, A., & Grochowski, M. (2018). Data augmentation for improving deep learning in image classification problem. 2018 International Interdisciplinary Phd Workshop (Iiphdw). doi: 10.1109/iiphdw.2018.8388338
35. Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700-4708). Modelo disponible en: <https://github.com/liuzhuang13/DenseNet>.
36. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA, 7–12 June 2015; pp. 1–9
37. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826
38. Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.
39. Hosna A, Merry E, Gyalmo J, Alom Z, Aung Z, Azim MA. Transfer learning: a friendly introduction. *J Big Data*. 2022;9(1):102. doi: 10.1186/s40537-022-00652-w. Epub 2022 Oct 22. PMID: 36313477; PMCID: PMC9589764.

40. Ma, L., Bian, M. L., Liu, J., Wang, X. H., Pang, C. H., & Chen, Y. (2011). Pregnancy related cervical cytological changes and clinical management, 46(2), 84–87.
41. Fuhrman, B. J., Moore, S. C., Byrne, C., Makhoul, I., Kitahara, C. M., Berrington de González, A., Linet, M. S., Weiderpass, E., Adami, H. O., Freedman, N. D., Liao, L. M., Matthews, C. E., Stolzenberg-Solomon, R. Z., Gaudet, M. M., Patel, A. V., Lee, I. M., Buring, J. E., Wolk, A., Larsson, S. C., Prizment, A. E., ... Hoover, R. N. (2021). Association of the Age at Menarche with Site-Specific Cancer Risks in Pooled Data from Nine Cohorts. *Cancer research*, 81(8), 2246–2255. <https://doi.org/10.1158/0008-5472.CAN-19-3093>
42. Fujita, M., Tase, T., Kakugawa, Y., Hoshi, S., Nishino, Y., Nagase, S., Ito, K., Niikura, H., Yaegashi, N., & Minami, Y. (2008). Smoking, earlier menarche and low parity as independent risk factors for gynecologic cancers in Japanese: a case-control study. *The Tohoku journal of experimental medicine*, 216(4), 297–307. <https://doi.org/10.1620/tjem.216.297>
43. Gong, TT., Wang, YL. & Ma, XX. Age at menarche and endometrial cancer risk: a dose-response meta-analysis of prospective studies. *Sci Rep* 5, 14051 (2015). <https://doi.org/10.1038/srep14051>
44. Reis, N., Beji, N. K., & Kilic, D. (2011). Risk factors for cervical cancer: results from a hospital-based case-control study.
45. Pragati Sharma, Sanjay M. Pattanshetty (2018) A study on risk factors of cervical cancer among patients attending a tertiary care hospital: A case-control study, *Clinical Epidemiology and Global Health*, Volume 6, Issue 2, 2018, Pages 83-87, ISSN 2213-3984, <https://doi.org/10.1016/j.cegh.2017.10.001>.
(<https://www.sciencedirect.com/science/article/pii/S2213398417300702>)

46. Heley, S., & Brotherton, J. (2009). Abnormal Pap tests after the HPV vaccine. *Australian family physician*, 38(12), 977–979.
47. Torrado G, L. M., Rincón Orozco, B., & Martínez-Vega, R. A. (2018). Genotipificación del Virus de Papiloma Humano en mujeres de la comuna norte de Bucaramanga. *Revista de la Universidad Industrial de Santander. Salud*, 50(3), 225–232.
<https://doi.org/10.18273/revsal.v50n3-2018007>
48. Bruni L, Barrionuevo-Rosas L, Albero G, Serrano B, Mena M, Gómez D, et al. ICO/IARC Information Centre on HPV and Cancer (HPV Information Centre). Human Papillomavirus and Related Diseases in the World. Summary Report; 2017
49. Prospecto Gardasil 9 suspensión inyectable en jeringa precargada. Vacuna nonavalente frente al Virus del Papiloma Humano (Recombinante, adsorbida). Agencia Española de Medicamentos y Productos Sanitarios. Disponible en:
https://cima.aemps.es/cima/dochtml/p/1151007002/P_1151007002.html
50. Nkwabong E, Laure Bessi Badjan I, Sando Z. Pap smear accuracy for the diagnosis of cervical precancerous lesions. *Tropical Doctor*. 2019;49(1):34-39. doi:10.1177/0049475518798532
51. Korbar, B. et al. Looking under the hood: Deep neural network visualization to interpret whole-slide image analysis outcomes for colorectal polyps. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, 821–827 (2017)
52. McHugh M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia medica*, 22(3), 276–282.
53. Wang, CW., Liou, YA., Lin, YJ. et al. Artificial intelligence-assisted fast screening cervical high grade squamous intraepithelial lesion and squamous cell carcinoma diagnosis and

- treatment planning. *Sci Rep* 11, 16244 (2021). <https://doi.org/10.1038/s41598-021-95545-y>
54. Tang, H., Cai, D., Kong, Y., Ye, H., Ma, Z., Lv, H., Tuo, L., Pan, Q., Liu, Z., & Han, X. (2021). Cervical cytology screening facilitated by an artificial intelligence microscope: A preliminary study. *Cancer Cytopathology*, 129(9), 693–700. <https://doi.org/10.1002/cncy.22425>
55. Falk, T. et al. U-net: Deep learning for cell counting, detection, and morphometry. *Nat. Methods* 16, 67–70 (2019).
56. Badrinarayanan, V., Kendall, A. & Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 2481–2495 (2017).
57. Araújo, F. H. et al. Deep learning for cell image segmentation and ranking. *Comput. Med. Imaging Graph.* 72, 13–21 (2019).