



**UNIVERSIDAD
TORCUATO DI TELLA**

Escuela de Negocios

Licenciatura en Tecnología Digital

Clasificación de Células Cervicales

TD8: Informe Final

Autor: Azul Noguera, Rocio Gonzalez, Natalia Massri

Profesor: Pía Torres

Universidad Torcuato Di Tella, [29/11/24]

Índice

Índice.....	2
Resumen Ejecutivo.....	3
Introducción.....	3
Justificación del tema.....	4
Objetivos de Proyecto.....	5
Métricas de Impacto.....	5
Planteamiento del problema.....	6
Plan de Trabajo.....	7
Hipótesis.....	12
Marco teórico.....	13
Marco metodológico.....	16
Métricas a utilizar.....	21
Resultados.....	23
Discusión.....	29
Conclusiones.....	32
Investigación Futura.....	33
Aprendizaje.....	35
Referencias bibliográficas.....	35

Resumen Ejecutivo

El cáncer cervical representa una amenaza significativa para la salud de las mujeres a nivel mundial. La prueba de Papanicolaou se ha consolidado como un método preventivo crucial, ya que permite identificar cambios celulares tempranos en el cuello uterino, reduciendo significativamente la mortalidad mediante intervenciones oportunas. Sin embargo, el análisis actual de estas pruebas, realizado manualmente por patólogos, es un proceso lento y propenso a errores humanos. Este estudio propone la implementación de modelos de aprendizaje profundo para automatizar el análisis de dichas pruebas. Los resultados obtenidos demuestran que los modelos propuestos superan la precisión del análisis manual, destacándose como herramientas prometedoras para asistir a los patólogos y mejorar los flujos de trabajo clínicos.

Introducción

El cáncer cervical constituye una preocupación sanitaria global significativa debido a su prevalencia y potencial letalidad. A pesar de los avances en métodos de detección y tratamiento, sigue siendo una de las principales causas de muerte por cáncer entre mujeres en muchas partes del mundo. En Argentina, cada año se diagnostican aproximadamente 5000 casos nuevos de cáncer cervical y 2000 mujeres mueren debido a esta enfermedad.[1]

La prueba de Papanicolaou (PAP), un examen ginecológico que permite detectar anomalías en las células cervicales, juega un papel crucial en la detección temprana de cáncer de cuello uterino. Su importancia radica en la identificación oportuna de cambios en las células cervicales que podrían indicar riesgos futuros. No obstante, la efectividad de esta prueba depende en gran medida de la precisión con la que se clasifiquen las alteraciones celulares observadas en las imágenes microscópicas obtenidas durante el procedimiento.

Actualmente, la clasificación y análisis de las imágenes de Papanicolaou se realiza manualmente por citopatólogos entrenados. Este enfoque presenta varias desventajas significativas, incluyendo un alto consumo de tiempo y la necesidad de una amplia experiencia y especialización, lo que eleva los costos operativos. Además, el análisis manual es propenso a errores debido a la fatiga y subjetividad del citopatólogo, aumentando el riesgo de diagnósticos incorrectos.

Frente a estos desafíos, este proyecto propone la implementación de un modelo de clasificación automatizado basado en técnicas de aprendizaje profundo para mejorar la eficacia de la distinción de la enfermedad. Al automatizar el análisis de las imágenes de Papanicolaou, se busca reducir el tiempo de evaluación, disminuir los costos asociados con el análisis manual y aumentar la precisión de los diagnósticos, reduciendo así los riesgos de falsos negativos y positivos.

Justificación del tema

El cáncer cervical representa no solo un problema de salud pública, sino también un desafío multidimensional que involucra aspectos médicos, sociales y económicos. Su alta incidencia y mortalidad oncológica femenina, subraya la necesidad de implementar estrategias más eficaces para su detección y manejo. La prueba de Papanicolaou (PAP) constituye un pilar fundamental en las estrategias de salud pública, ya que permite detectar lesiones celulares en etapas tempranas, reduciendo significativamente la mortalidad. Sin embargo, el análisis manual de las muestras presenta desafíos significativos: es un proceso lento, meticuloso y costoso, que depende de especialistas calificados. Esto no solo incrementa los costos totales del procedimiento, sino que también introduce un riesgo considerable de errores humanos debido a la fatiga y la subjetividad, comprometiendo la fiabilidad de los resultados. Este impacto en la precisión se refleja en la tasa de falsos negativos que puede alcanzar hasta un 28% según algunos estudios.[2] Estas cifras subrayan las limitaciones inherentes a los métodos de detección manuales, donde factores como la falta de atención y el cansancio pueden llevar a diagnósticos críticos omitidos. Tales tasas no sólo señalan la necesidad de mejorar la formación y las técnicas de los profesionales implicados sino también subrayan la importancia de integrar tecnologías avanzadas y sistemas de apoyo que puedan reducir estos errores y mejorar la eficacia general del cribado.

Desde una perspectiva académica, el desarrollo de modelos de clasificación automatizada basados en aprendizaje profundo para el análisis de muestras de PAP representa una contribución significativa al avance del campo de la bioinformática y la inteligencia artificial aplicada a la salud. Este proyecto no solo se alinea con las tendencias actuales de digitalización y automatización en el diagnóstico médico, sino que también permite explorar nuevas fronteras en términos de precisión diagnóstica, optimización de recursos y la implementación de tecnologías avanzadas en la práctica clínica.

En la práctica, la automatización en la clasificación celular tiene el potencial de transformar el manejo del cáncer cervical, abordando las limitaciones del análisis manual. Este enfoque ofrece una evaluación más rápida con ventajas evidentes. Al reducir la influencia de la subjetividad humana, el modelo puede mejorar la precisión diagnóstica, disminuyendo significativamente los errores como los falsos negativos. Además, al disminuir la carga de trabajo sobre los citopatólogos y mejorar la consistencia y exactitud del diagnóstico, se espera que el proyecto contribuya de manera significativa al manejo clínico de los pacientes, evitando tratamientos innecesarios y enfocando recursos hacia casos que requieran atención especializada. De esta forma, no solo mejora la eficacia del diagnóstico, sino que también contribuye a reducir la mortalidad asociada al cáncer cervical, integrando tecnología avanzada para transformar las estrategias de detección y tratamiento en beneficio de la salud pública.

Objetivos de Proyecto

Objetivo General

El propósito principal del proyecto es optimizar el proceso de diagnóstico del cáncer cervical mediante la implementación de un modelo de clasificación automatizado que utilice técnicas avanzadas de aprendizaje profundo para analizar imágenes de Papanicolaou. Este modelo busca mejorar la eficiencia operativa y la calidad de los diagnósticos, reduciendo significativamente los costos temporales del cribado de cáncer cervical.

Objetivos Específicos

Los objetivos específicos del proyecto son:

1. Desarrollar un modelo de clasificación binaria para diferenciar entre células con lesiones y sin lesiones.
2. Desarrollar un modelo para clasificar las células según el grado de lesión: alto, bajo, y sin lesión.
3. Desarrollar un modelo de clasificación de cinco clases centrado exclusivamente en células lesionadas.
4. Desarrollar un modelo de clasificación de seis clases que incluya células sin lesiones y todos los subtipos de lesiones.
5. Desarrollar un modelo de clasificación en dos fases: primero una clasificación binaria y luego una de cinco clases.

Métricas de Impacto

El procedimiento actual de detección de células de cáncer cervical enfrenta el desafío significativo de analizar eficazmente grandes volúmenes de pruebas de Papanicolaou, un procedimiento tradicionalmente lento y propenso a errores.

En respuesta a esta situación, se implementarán Indicadores Clave de Desempeño (KPIs) que permitirán evaluar y comparar el estado actual del proceso con las propuestas de mejora proyectadas. Estos KPIs se establecerán como referencias esenciales para la optimización continua y el aumento de la exactitud operativa en el proyecto.

Los Indicadores de Desempeño (KPIs) son:

1. *Tiempo de análisis:* Implementar un sistema de clasificación automatizado que *disminuya* el tiempo requerido para evaluar cada muestra de Papanicolaou, aumentando la capacidad de procesamiento de los laboratorios. Un estudio realizado por Cao et al, en el que se comparó el desempeño de cuatro patólogos con experiencia laboral de entre 4 y 10 años, reveló que la evaluación manual requiere aproximadamente **14.83** segundos por imagen, lo que establece un estándar de referencia para mejorar la eficiencia mediante la adopción de tecnologías de inteligencia artificial.[3] Además en algunos países un citopatólogo puede analizar hasta 100 frotis diarios, lo cual justifica aún más la aplicación de una metodología de cribado que apoye, facilite y mejore sus decisiones.[4]
2. *Precisión diagnóstica:* *Mejorar* la exactitud de los resultados de la detección del cáncer cervical mediante la reducción de errores humanos como la fatiga y la subjetividad, asegurando diagnósticos más confiables y consistentes.

Un estudio llevado a cabo en Tailandia por la Sociedad Tailandesa de Citología entre 2010 y 2014, incluyó a 124 laboratorios en 76 provincias, con el objetivo de evaluar la eficacia del cribado cervical nacional mediante la revisión aleatoria del 10% de frotis inicialmente reportados como negativos. El estudio reveló una tasa de falsos negativos promedio de **13.8%** para lesiones LSIL o más severas.[5]

Asimismo, en otro estudio realizado por Wergeland Sørbye, se evaluaron cien muestras de Papanicolaou por cuatro patólogos en tres hospitales en Noruega. Los casos analizados incluyeron diferentes clasificaciones como normal, LSIL, HSIL, ASC-US y ASC-H. La precisión para detectar lesiones varió entre **74.1% y 83.8%**, demostrando la variabilidad en la exactitud diagnóstica entre profesionales y destacando la potencial mejora con la implementación de sistemas automatizados.[6]

Planteamiento del problema

En la actualidad, el análisis de imágenes de Papanicolaou se lleva a cabo de forma manual por citopatólogos entrenados, un método que presenta desventajas significativas. El proceso manual no solo es lento y laborioso, lo que resulta en altos costos temporales y crea cuellos de botella en laboratorios que manejan un alto volumen de muestras, sino que también implica un costo monetario considerable debido al empleo de especialistas calificados. Además, la precisión del diagnóstico depende estrechamente de la habilidad y el estado del citopatólogo, siendo susceptible a errores por fatiga y factores humanos, lo cual puede afectar negativamente el manejo clínico de los pacientes.

Ante estos desafíos, este proyecto sugiere la implementación de un sistema automatizado de clasificación de células cervicales, empleando tecnologías avanzadas de aprendizaje profundo. Proponemos un enfoque estratificado y sistemático para la clasificación de las células cervicales, basado en la presencia y gravedad de lesiones celulares.

En primer lugar, se desarrollará un modelo de clasificación binaria diseñado para discernir entre células lesionadas y no lesionadas. Este modelo se aplicará para aportar una distinción inicial.

Adicionalmente, se plantea la creación de un segundo modelo que clasifique las células según el grado de lesión. Este modelo tiene como objetivo proporcionar una evaluación más granular del estado patológico de las células, contribuyendo así a una mejor comprensión de la progresión de la enfermedad.

Asimismo, se implementará un tercer modelo específico para clasificar las células lesionadas de acuerdo con el tipo específico de lesión. Este modelo se centrará exclusivamente en las células previamente identificadas como lesionadas, excluyendo de su análisis a las células no lesionadas, con el fin de mejorar la precisión en la determinación del tipo específico de patología celular.

Finalmente, se propone comparar dos enfoques metodológicos distintos para la clasificación general de células cervicales. El primero consiste en el desarrollo de un modelo integrado que clasifique todas las células directamente por tipo de lesión, tratando a las células no lesionadas como una categoría más dentro del espectro de clasificación (clasificación en 6 clases). El segundo enfoque, que se plantea como una mejora potencial del primero, divide el proceso de clasificación en dos fases secuenciales: una clasificación binaria inicial que determina la presencia o ausencia de lesiones en las células, seguida de una fase de clasificación más detallada en la que solo las células identificadas como lesionadas son analizadas para determinar el tipo específico de lesión que presentan. Este enfoque bifásico tiene el potencial de optimizar el rendimiento del sistema de clasificación al reducir la variabilidad inicial y enfocar los recursos computacionales y analíticos en las muestras que requieren una evaluación más compleja.

Plan de Trabajo

Primer Sprint (04/08 al 17/08)

Durante el primer sprint, se llevó a cabo un análisis exhaustivo de material científico y artículos académicos relacionados con técnicas de aprendizaje profundo utilizadas para la detección y clasificación de células de Papanicolaou. El objetivo principal fue profundizar en la comprensión de la problemática y explorar metodologías efectivas para abordarla. Se realizó una reunión inicial (kick-off) con el propósito de establecer objetivos claros y organizar los pasos a seguir.

No obstante, debido a discrepancias entre las partes involucradas, el proyecto inicialmente planteado no pudo ser continuado.

Segundo Sprint (18/08 al 31/08)

En el segundo sprint, se llevó a cabo una reunión con los propietarios del proyecto, quienes presentaron una nueva propuesta para desarrollar. Posteriormente, el equipo organizó una reunión interna para estructurar la nueva metodología de trabajo, definir los objetivos para cada sprint y dar inicio al nuevo proyecto. Se retomó la investigación sobre la temática, intentando proponer nuevos objetivos alineados con los requerimientos establecidos.

Tercer Sprint (01/09 al 14/09)

Durante el tercer sprint, se actualizó el brief del proyecto incorporando la nueva información relativa a los objetivos, metodología, pasos a seguir y aspectos organizativos esenciales para su desarrollo. Además, se llevó a cabo una nueva reunión con los propietarios del proyecto para asegurar una alineación en cuanto a los requerimientos, objetivos y metas alcanzables.

Cuarto Sprint (15/09 al 28/09)

En el cuarto sprint, se inició el desarrollo práctico del proyecto. Se realizó el análisis y preprocesamiento de los datos disponibles y se comenzó a investigar posibles modelos y arquitecturas de aprendizaje profundo que pudieran ser utilizados para la tarea de clasificación de células cervicales.

Quinto Sprint (01/10 al 12/10)

Durante el quinto sprint, se definieron dos posibles enfoques para abordar la clasificación:

1. Utilización de imágenes completas (WSI): Este enfoque consistía en emplear las imágenes de alta resolución (Whole Slide Images) en su totalidad para la clasificación, utilizando modelos pre entrenados capaces de detectar y clasificar directamente a partir de estas imágenes.
2. Clasificación basada en recortes de núcleos celulares: Se centraba en la clasificación de las células a partir de imágenes WSI recortadas, enfocándose exclusivamente en los núcleos celulares con un tamaño estándar de 90x90 píxeles.

Tras seleccionar el segundo enfoque, realizamos el recorte de los núcleos de las imágenes para crear un nuevo conjunto de datos, dividido en tres subconjuntos: entrenamiento (train), validación (validation) y prueba (test). Esta preparación permitió iniciar la evaluación del estado del arte en la clasificación de objetos de pequeño tamaño, con el objetivo de implementar las mejores técnicas en el siguiente sprint del proyecto.

Sexto Sprint (13/10 al 26/10)

El sexto sprint se dedicó exclusivamente al entrenamiento de modelos de clasificación utilizando las imágenes recortadas de los núcleos celulares. Se desarrollaron diversos modelos para cada tipo de clasificación:

- Clasificación binaria: Distinguir entre células sin lesión y células con lesión.
- Clasificación de tres clases: Clasificar el grado de lesión en tres categorías.
- Clasificación de cinco clases: Enfocada únicamente en las células con lesión, clasificándolas en cinco subtipos según el Sistema Bethesda.
- Clasificación de seis clases: Incluyendo tanto las células sin lesión como las cinco categorías de células con lesión.

Para abordar el desbalance de clases presente en el conjunto de datos, se implementaron técnicas de balanceo y aumentación de datos, mejorando así el rendimiento de los modelos. La disparidad significativa en la distribución de clases, tanto entre positivas y negativas como dentro de las clases con lesión, requirió estrategias específicas para garantizar una representación adecuada de todas las categorías.

Séptimo Sprint (27/10 al 09/11)

En el séptimo sprint, se obtuvieron resultados satisfactorios, lo que permitió definir el mejor modelo para cada tipo de clasificación. Se llevó a cabo una optimización de hiperparámetros con el objetivo de maximizar el rendimiento de los modelos. Adicionalmente, se propuso e implementó un nuevo enfoque para mejorar la precisión en la clasificación: desarrollar un pipeline de clasificación en dos etapas. Este pipeline conectaba la salida del modelo binario y utilizaba las células clasificadas con lesión como entrada para el modelo de clasificación de cinco clases.

Paralelamente, se iniciaron las labores de redacción de algunas secciones del informe final, como la introducción y la justificación del tema.

Octavo Sprint (10/11 al 22/11)

En el último sprint, el equipo se enfocó en la elaboración y terminación del informe final. Se llevó a cabo una última reunión con los propietarios del proyecto para presentar los resultados obtenidos y verificar que estos cumplieran con las expectativas y requerimientos establecidos.

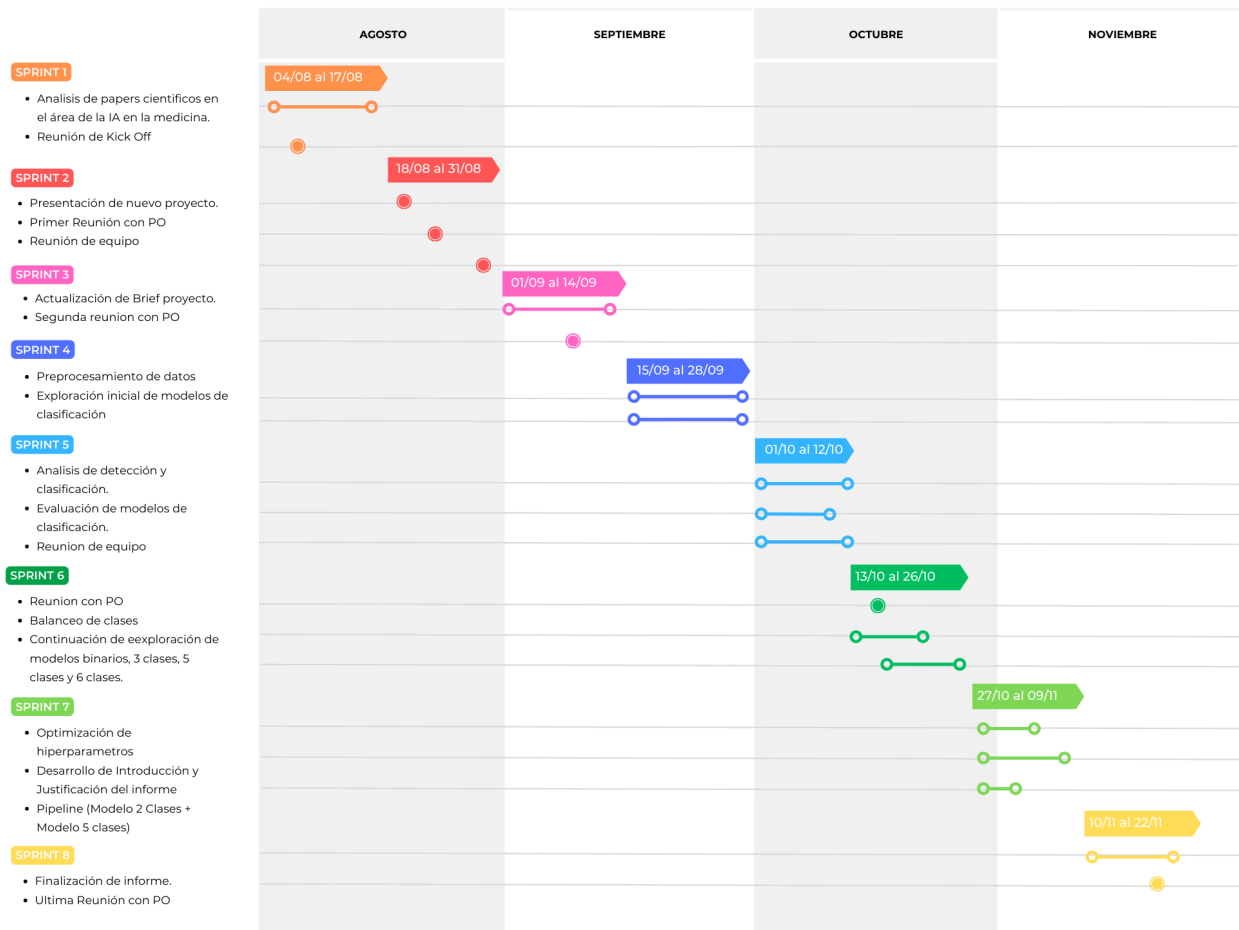


Figura 1: Diagrama de Gantt

Herramientas empleadas para gestión del proyecto

Para la gestión y organización del proyecto, se emplearon las siguientes herramientas, que facilitaron la planificación, colaboración y seguimiento de las tareas:

- GitHub:** Utilizada como plataforma principal para el control de versiones del código y la colaboración entre los integrantes del equipo. Permitió mantener un registro de los cambios realizados en el proyecto, realizar revisiones de código mediante pull requests y documentar avances de manera ordenada. Además, el uso de Git LFS (Large File Storage) fue clave para gestionar los archivos grandes, como los modelos guardados, sin comprometer la integridad del repositorio.

- ClickUp: Herramienta utilizada para la gestión de tareas y planificación de los sprints. Se organizaron tableros que reflejaban las etapas del flujo de trabajo, como "Pendiente", "En progreso" y "Completado". Además, permitió asignar tareas a los miembros del equipo, establecer prioridades y realizar un seguimiento continuo del avance de cada iteración.
- Google Meet: Utilizado como medio principal para la comunicación y coordinación remota. Las reuniones periódicas permitieron revisar el progreso de los sprints, discutir bloqueos y definir los objetivos de las siguientes etapas del proyecto.

Etapas del proyecto

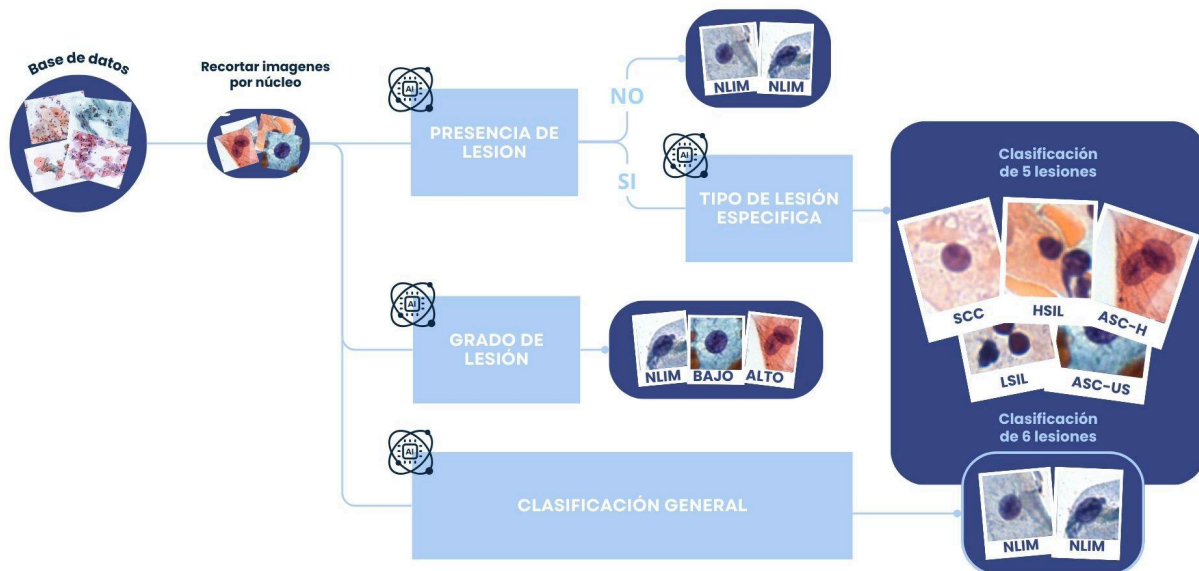


Figura 2: Diagrama de flujo de Trabajo

El flujo de trabajo del proyecto comenzó con el procesamiento de imágenes WSI (Whole Slide Images), en las cuales se realizaron recortes específicos de los núcleos celulares. Este preprocesamiento aseguró que el análisis se enfocara exclusivamente en las áreas de interés, eliminando información irrelevante y optimizando los datos para las siguientes etapas.

A partir de ello, de manera simultánea se desarrollaron varios modelos con diferentes niveles de clasificación. En primer lugar, se diseñó un modelo de clasificación binaria para discernir entre células lesionadas y no lesionadas, sirviendo como un filtro inicial que aportó una distinción crucial para estructurar el análisis. Paralelamente, se implementó un modelo triclase que clasificó las células en tres categorías según el grado de lesión: sin lesión, grado bajo y grado alto, permitiendo

así una evaluación más detallada del estado patológico de las células y contribuyendo a una mejor comprensión de la progresión de las lesiones. A su vez, se desarrolló un modelo orientado exclusivamente a la clasificación de células lesionadas en cinco categorías basadas en el Sistema Bethesda (ASC-US, LSIL, ASC-H, HSIL y SCC), focalizándose en las células previamente identificadas como lesionadas y excluyendo las no lesionadas, con el objetivo de mejorar la precisión en la clasificación.

A partir de estos modelos, se construyó un pipeline de clasificación que combinó dos etapas secuenciales. En la primera etapa, el modelo binario clasificó las células por su presencia de lesión, y en la segunda etapa, las células identificadas como lesionadas se analizaron con el modelo de cinco clases para determinar el tipo específico de lesión. Este enfoque progresivo y jerarquizado optimizó el análisis, mejoró la precisión del diagnóstico y ofreció una solución más robusta al problema de la clasificación de células cervicales.

De manera paralela, se desarrolló un modelo alternativo de clasificación de seis clases, diseñado para categorizar directamente todas las células, incluyendo aquellas sin lesión, dentro de las seis categorías del Sistema Bethesda. Este modelo trató las células no lesionadas como una categoría más dentro del espectro general de clasificación.

Finalmente, se compararon ambos enfoques, el pipeline y el modelo de seis clases, para evaluar su rendimiento en términos de precisión y eficiencia, determinando cuál ofrecía mejores resultados para la clasificación general de células cervicales.

Hipótesis

El uso de algoritmos de aprendizaje automático en la clasificación de células cervicales a partir de imágenes de pruebas de papanicolaou, mejorará la precisión en la categorización de datos de la salud y reducirá significativamente el tiempo requerido para el análisis.

Hipótesis Secundaria

El enfoque de clasificación en dos fases secuenciales (pipeline), que consiste en un modelo inicial para diferenciar entre células con y sin lesión, seguido por un segundo modelo especializado en clasificar los tipos específicos de lesiones, ofrecerá una mayor precisión en la clasificación de células cervicales en comparación con el enfoque integrado de un modelo único para la clasificación en seis clases.

Marco teórico

Metodologías de Desarrollo de Software

Para el desarrollo de este proyecto, se empleó la metodología Agile, específicamente adaptando prácticas de la metodología Scrum para estructurar y guiar el proceso. Este enfoque permitió un desarrollo iterativo y flexible, crucial para adaptarse a los requisitos cambiantes y a las complejidades inherentes a un proyecto de clasificación de imágenes médicas usando aprendizaje profundo.

Implementación de Agile con estructura de Scrum:

1. Organización en Sprints:
 - El trabajo se organizó en un total de 8 sprints, cada uno con una duración de dos semanas durante la cual se establecieron y completaron objetivos específicos. Esta estructura de sprints facilitó un enfoque modular y ágil al desarrollo, permitiendo ajustes rápidos a medida que evolucionaban las necesidades del proyecto y se obtenían retroalimentaciones.
2. Roles:
 - Scrum Master: Se asignó un Scrum Master, que facilitó todas las reuniones de Scrum, aseguró que el equipo siguiera las prácticas ágiles, y ayudó a resolver impedimentos que podrían afectar la productividad del equipo.
 - Product Owner: La figura del Product Owner fue clave para mantener la alineación con los objetivos del proyecto. Esta persona fue responsable de definir los requisitos del proyecto.
3. Reuniones Regulares:
 - Se mantuvo una comunicación constante y estructurada a través de reuniones semanales que involucraban al equipo de desarrollo y al Scrum Master. Estas reuniones, conocidas como Daily Scrums en el contexto diario, se adaptaron a un formato semanal debido a las características del proyecto, permitiendo discutir progresos, desafíos enfrentados y estrategias para abordar problemas emergentes.
 - Las revisiones de sprint y las planificaciones de sprint se realizaron al inicio y al final de cada sprint, respectivamente, para evaluar los resultados y planificar los siguientes pasos con claridad y consenso.
4. Iteración y Retroalimentación:
 - Al final de cada sprint, se llevó a cabo una reunión de revisión con el product owner para demostrar las funcionalidades desarrolladas. Esto no sólo proporcionó una oportunidad vital para la retroalimentación directa, sino que también ayudó a realinear los siguientes sprints con las expectativas del proyecto.

- Las retrospectivas de sprint permitieron al equipo reflexionar sobre su dinámica de trabajo y procesos, identificar áreas de mejora e implementar ajustes para incrementar la eficacia en sprints futuros.

Paradigmas de Programación Utilizados en el Proyecto

El proyecto se desarrolló bajo el paradigma declarativo, utilizando frameworks como TensorFlow y Keras, que simplificaron significativamente la implementación de arquitecturas de redes neuronales preentrenadas, como EfficientNet y XceptionNet. Este paradigma permitió enfocar el desarrollo en la definición de qué se quería lograr, delegando al framework la tarea de realizar los cálculos subyacentes y optimizar los procesos de entrenamiento.

El paradigma declarativo fue esencial para configurar funciones de pérdida adecuadas, optimizadores y métricas clave enfocadas en maximizar el recall, una métrica crítica para reducir los falsos negativos en el contexto del diagnóstico médico. Además, este enfoque permitió realizar experimentos iterativos para ajustar hiperparámetros clave, como la tasa de aprendizaje, el tamaño del batch y el número de épocas, de forma eficiente. Esto hizo posible adaptar las arquitecturas preentrenadas para las necesidades específicas del proyecto mediante el ajuste de las capas superiores, a través de fine-tuning en algunos casos.

Tecnologías y Frameworks Utilizados en el Proyecto

Este proyecto incorpora una variedad de tecnologías y frameworks avanzados para el desarrollo de un sistema automatizado de clasificación de imágenes de células cervicales. A continuación, se detallan las tecnologías implementadas y se proporcionan justificaciones para su selección, así como una descripción de cómo se integran para formar una solución cohesiva.

Selección de Tecnologías y Frameworks

1. TensorFlow y Keras:

- Componentes Utilizados: *tensorflow.keras* con diversas capas, modelos y funciones como *MobileNetV2*, *MobileNetV3*, *EfficientNet (B0, B2, B4, B5, B6)*, *ResNet*, *Xception* y *InceptionV3*, además de optimizadores como *Adam* y callbacks como *EarlyStopping* y *ReduceLROnPlateau*.
- Justificación: *TensorFlow* y *Keras* son elegidos por su robustez, amplia adopción en la comunidad científica y soporte para redes neuronales profundas. Facilitan la experimentación rápida y la implementación de arquitecturas de red neural avanzadas. Estos frameworks son ideales para el procesamiento de imágenes médicas debido a su capacidad para manejar grandes volúmenes de datos y su eficiencia en el entrenamiento de modelos complejos.

2. Librerías Auxiliares:

- Componentes Utilizados: *numpy* para la manipulación numérica eficiente de datos, *matplotlib.pyplot* y *seaborn* para visualización de datos y resultados, y herramientas de *scikit-learn* como *compute_class_weight*, *classification_report*, y *confusion_matrix* para la evaluación y manejo de desequilibrios en clases.
- Justificación: Estas librerías complementan los frameworks de aprendizaje profundo al proporcionar funcionalidades esenciales para el preprocesamiento de datos, la visualización gráfica y el análisis estadístico. Facilitan una evaluación detallada del rendimiento del modelo y permiten una visualización intuitiva de los resultados obtenidos.

Integración de Tecnologías

- Flujo de Trabajo de Modelado: Los modelos de *TensorFlow* y *Keras* se utilizan para la arquitectura principal de clasificación, aprovechando arquitecturas preentrenadas que se ajustan para la tarea específica utilizando el *ImageDataGenerator* para la aumentación de datos. Esto es crucial para mejorar la generalización del modelo en datos médicos variados.
- Visualización y Análisis: *matplotlib* y *seaborn* se integran para generar representaciones visuales de la distribución de datos y la confusión de clasificaciones, facilitando la interpretación de los resultados y la comunicación de los hallazgos a un público más amplio, incluidos stakeholders no técnicos.

Investigaciones Previas Relevantes

En este apartado, queremos destacar un estudio reciente que fue crucial para definir el enfoque de nuestro proyecto: “A Deep Learning Ensemble Method to Assist Cytopathologists in Pap Test Image Classification” [4]. Este trabajo propone un método de aprendizaje profundo basado en un ensamble de modelos para la clasificación de imágenes del Pap Test, con un enfoque en el análisis automatizado de imágenes citológicas. Resalta cómo las tecnologías avanzadas pueden complementar el diagnóstico humano, optimizando la identificación de lesiones celulares.

Sin embargo, el aspecto más distintivo de este estudio radica en la premisa de que **el núcleo celular es suficiente para clasificar una célula según el tipo de lesión**. Este hallazgo nos llevó a reconsiderar nuestro enfoque, adoptando una metodología centrada en el análisis de núcleos celulares recortados en lugar de trabajar con imágenes completas.

Marco metodológico

Base de datos

El conjunto de datos utilizado en este proyecto contiene 400 imágenes de alta resolución (1,376 x 1,020 píxeles) y cuenta con la clasificación manual de 11,534 células cervicales. Cada célula fue meticulosamente clasificada por un equipo de citopatólogos experimentados siguiendo la nomenclatura del Sistema Bethesda, que es el estándar global para la terminología en citología cervicovaginal. Las imágenes han sido clasificadas en seis categorías distintas según la presencia y el tipo de lesión cervical: células normales (NILM), células escamosas atípicas de significado indeterminado (ASC-US), lesión intraepitelial escamosa de bajo grado (LSIL), células escamosas atípicas que no pueden excluir una lesión de alto grado (ASC-H), lesión intraepitelial escamosa de alto grado (HSIL) y carcinoma de células escamosas (SCC).

La base de datos se encuentra disponible públicamente en la plataforma CRIC de Imágenes Buscables (<https://database.cric.com.br>)

Preprocesamiento de los datos

Durante la investigación inicial en el área, descubrimos que varios autores, como [7,8,9,10], argumentaban que el núcleo celular era suficiente para clasificar una célula según su grado de lesión. Esto nos llevó a considerar adoptar una metodología basada en el análisis de núcleos celulares recortados, centrando un recorte de tamaño $n \times n$ en el núcleo demarcado en la base de datos, donde $n = 90$ se identificó como el tamaño ideal para incluir completamente el núcleo sin capturar múltiples núcleos en una misma imagen. La Figura 3 representa un ejemplo de recorte de núcleo de cada una de las clases en el dataset.

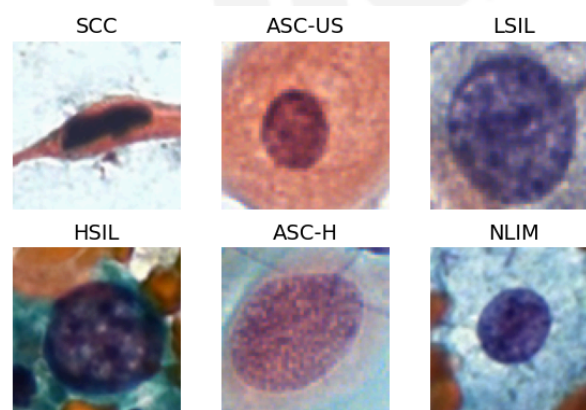


Figura 3: Recortes de núcleos de cada clase

La base de datos muestra un desbalance significativo, eso significa que algunas clases están sobre-representadas mientras que otras tienen muy pocos ejemplos. Como se evidencia en la **Figura 4**, la clase **NILM** tiene 6779 ejemplos, mientras que **SCC** tiene solo 161 muestras. Este desbalance puede inducir un sesgo en los modelos de aprendizaje automático, favoreciendo la predicción de clases mayoritarias y reduciendo la eficacia en la detección de clases minoritarias críticas. Al centrarse en los núcleos celulares y aplicar técnicas de balanceo como el sobremuestreo de las clases minoritarias o el submuestreo de las mayoritarias, se puede combatir este desbalance de manera más efectiva. Lo que ofrece una gran ventaja en comparación a la clasificación de imágenes enteras, donde aumentar datos aplicando transformaciones a imágenes que incluyen células de múltiples clases podría perpetuar el sesgo existente.

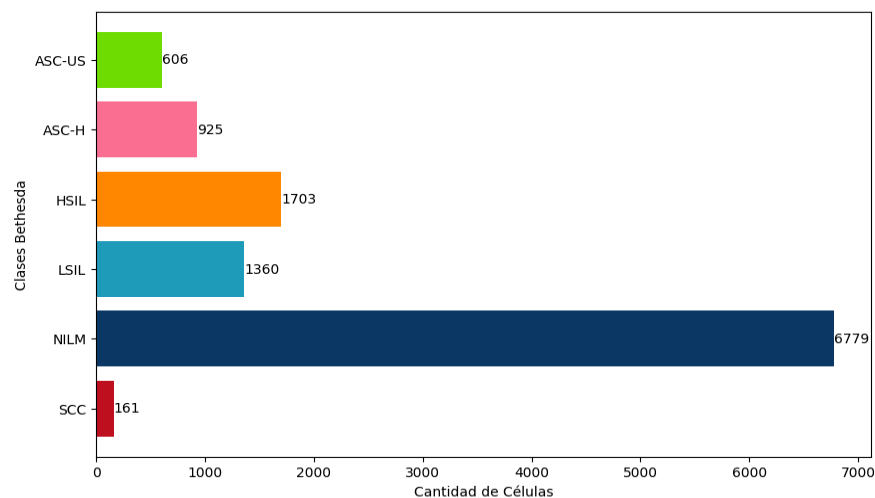


Figura 4: Distribución de Clases

División de Datos

Se dividieron los datos en tres conjuntos distintos, como se muestra en la **Tabla 1**, para optimizar el proceso de entrenamiento y evaluación de nuestro modelo. Esta división se realizó de manera estratificada, asegurando que la proporción de las clases en cada subconjunto reflejara la distribución original del conjunto de datos. Este enfoque es particularmente importante debido al desbalance de clases presente en nuestra base de datos. Al garantizar que todas las clases estén representadas de manera proporcional en los subconjuntos, se busca promover un entrenamiento más equilibrado y una evaluación más representativa del desempeño del modelo.

Utilizamos el 70% de la base de datos para el conjunto de entrenamiento, empleado para ajustar los parámetros del modelo. Un 15% adicional se destina al conjunto de validación, utilizado para afinar

la configuración del modelo y prevenir el sobreajuste. El restante 15% se asigna al conjunto de prueba, que empleamos para evaluar la capacidad de generalización del modelo en datos no vistos.

Optamos por este enfoque para equilibrar el entrenamiento y la validación, permitiendo ajustes precisos del modelo mientras aseguramos suficientes datos para pruebas significativas y una evaluación robusta de la generalización.

Clases	NLIM	HSIL	LSIL	ASC-H	ASC-US	SCC	Total
Train	4745	1192	952	647	424	112	8072
Validation	1017	255	204	139	91	24	1730
Test	1017	256	204	139	91	25	1732
Total por Clase	6779	1703	1360	925	606	161	11534

Tabla 1: Proporción de imágenes de cada conjunto de datos

Balanceo y Aumentación de datos

En problemas de clasificación, el desbalance de las clases en un conjunto de datos presenta desafíos significativos. Cuando las proporciones de datos entre las clases varían ampliamente, los modelos tienden a adaptarse excesivamente a las clases mayoritarias, aprendiendo de manera desproporcionada sus características. Esto compromete la capacidad del modelo para reconocer y clasificar correctamente las clases minoritarias, cuyas características son menos representadas en el conjunto de entrenamiento. Como resultado, la precisión y la generalización del modelo disminuyen, limitando su eficacia en entornos de aplicación real donde la distribución de las clases puede ser más equilibrada o desconocida.

Para mitigar el problema del desbalance de clases, una estrategia efectiva es la aumentación de datos. Esta técnica consiste en enriquecer el conjunto de entrenamiento mediante la generación de variaciones artificiales de los datos existentes. Estas manipulaciones ayudan a incrementar la diversidad dentro de las clases subrepresentadas, promoviendo así un aprendizaje más equilibrado y mejorando la capacidad de generalización del modelo frente a nuevas instancias de datos que presenten una distribución de clases diferente a la observada en el entrenamiento. Este enfoque no solo equilibra la representación de las clases minoritarias, sino que también contribuye a un desempeño más robusto y consistente del modelo en aplicaciones prácticas.

Dada la notable disparidad en la distribución de clases en la base de datos CRIC, se implementó una estrategia de aumentación de datos para mitigar las limitaciones asociadas con el desbalance significativo. Se conservó la proporción de la clase predominante, NILM, mientras que se incrementó la representación de las clases menos frecuentes. Se emplearon técnicas avanzadas como rotaciones, desplazamientos, volteos y aplicaciones de zoom, buscando enriquecer la diversidad en las imágenes y fortalecer la generalización del modelo en situaciones prácticas. Estas intervenciones están orientadas a asegurar una representación más equitativa y robustecer la precisión del modelo en la clasificación.

Tipos de clasificaciones

En el marco de nuestro proyecto de clasificación de células cervicales, se desarrollaron diversas estrategias de clasificación para optimizar el análisis y la interpretación de los datos:

Clasificación de 2 clases:

Se propuso inicialmente diferenciar entre células sin lesiones (NILM) y células lesionadas (que incluyen ASC-US, LSIL, ASC-H, HSIL, SCC), proporcionando una visión binaria y simplificada del conjunto de datos.

Clasificación de 3 clases:

Este esquema profundiza al categorizar las células en sin lesiones (NILM), lesiones de bajo grado (ASC-US, LSIL) y lesiones de alto grado (ASC-H, HSIL, SCC), permitiendo una distinción más detallada y clínicamente relevante.

Clasificación de 5 clases:

Focalizada exclusivamente en células lesionadas, esta clasificación prescinde de la categoría NILM para concentrarse en las diferencias entre ASC-US, LSIL, ASC-H, HSIL y SCC, buscando detallar aún más los tipos específicos de lesiones detectadas.

Clasificación de 6 clases:

Para alcanzar una evaluación integral, se desarrolló una clasificación exhaustiva que determina el tipo de lesión en células cervicales. Este modelo no solo identifica células sin lesiones, sino que también clasifica todos los subtipos de lesiones, lo cual potencia la precisión diagnóstica y aumenta la relevancia clínica del modelo.

Clasificación de 2 clases → Clasificación de 5 clases:

En la búsqueda de maximizar la precisión y eficiencia en la clasificación de lesiones celulares, se implementó una estrategia de clasificación en dos fases. Inicialmente, se realiza una clasificación binaria para discernir entre células sin lesiones (NILM) y células lesionadas. Posteriormente, las células clasificadas como lesionadas se analizan más detalladamente para identificar el tipo específico de lesión (ASC-US, LSIL, ASC-H, HSIL, SCC). Un aspecto clave de esta estrategia es que, al excluir las células de la clase NILM en la segunda fase, se logra reducir significativamente el desbalance entre las clases restantes, generando un conjunto más equilibrado. Suponemos que esto podría mejorar el rendimiento del modelo al mitigar la tendencia a favorecer las clases dominantes. Este enfoque bifásico busca reducir la complejidad inicial del modelo y optimizar los recursos computacionales para una discriminación precisa de las categorías críticas.

Implementación de Modelos

Arquitecturas

En el desarrollo de nuestros modelos para la clasificación de células cervicales, se exploraron diversas arquitecturas de redes neuronales convolucionales para determinar cuál ofrecería la mejor precisión y generalización en la detección de lesiones. Se evaluaron múltiples arquitecturas, incluyendo *MobileNetV2*, *MobileNetV3*, *EfficientNet (B0, B2, B4, B5, B6)*, *ResNet*, *Xception* y *InceptionV3*. Este estudio se enfocó en identificar las que mejor se adaptan a las particularidades de las imágenes cervicales y sus patrones de lesiones.

Configuración

En nuestro estudio, se adoptaron diversas estrategias para configurar los modelos de clasificación, optimizando su desempeño según las características del conjunto de datos:

- **Función de pérdida:** Dependiendo del tipo de clasificación, utilizamos *binary_crossentropy* para clasificaciones binarias y *categorical_crossentropy* para clasificaciones multiclase, adecuando así la función de pérdida a la naturaleza del problema.
- **Métricas de evaluación:** Se emplearon métricas como accuracy, precisión, recall, F1-score y área bajo la curva ROC (AUC), las cuales se explicarán en detalle en un apartado posterior para ilustrar su impacto en la evaluación del modelo.
- **Fine Tuning:** Implementamos la congelación de las primeras capas en modelos pre entrenados con la base de datos ImageNet para mantener las características generales aprendidas y permitir el ajuste fino de capas superiores, específicamente alineadas a nuestro conjunto de datos.

- **Dropout:** Se utilizó para promover la generalización del modelo, reduciendo el riesgo de sobreajuste mediante la desactivación aleatoria de neuronas durante el entrenamiento.
- **Early Stopping:** Este mecanismo detiene el entrenamiento cuando una métrica monitoreada, como el recall, deja de mejorar, ayudando a prevenir el sobreajuste y asegurando que el modelo retenga la capacidad de generalización.
- **Reduce LR On Plateau:** Se aplicó para reducir la tasa de aprendizaje cuando las métricas de pérdida dejan de mejorar, permitiendo afinamientos más finos del modelo sin estancamientos en mínimos locales durante el entrenamiento.

Optimización de Hiperparámetros

Durante nuestra investigación, descubrimos que la arquitectura *XceptionNet* proporcionaba los mejores resultados en todos los tipos de clasificación, lo que nos motivó a enfocarnos en la optimización de hiperparámetros para esta arquitectura específica. La optimización de hiperparámetros apunta a ajustar la configuración del modelo para maximizar su rendimiento y eficacia, involucrando la calibración del learning rate para mejorar la convergencia durante el entrenamiento, la tasa de dropout para controlar el sobreajuste y fomentar una mejor generalización, la elección de capas para el ajuste fino y la determinación del número de unidades en las capas densas para captar de manera eficiente características complejas.

Métricas a utilizar

Para evaluar la efectividad de los métodos de diagnóstico del cáncer cervical, se utilizarán varias métricas críticas que proporcionarán una visión integral sobre la precisión y fiabilidad de las técnicas empleadas, en la **Figura 5** se visualizan los posibles casos en la clasificación:

- **Tasa de Falsos Negativos (FNR):** Mide la proporción de resultados negativos que en realidad son casos positivos, omitidos por el diagnóstico. Es crucial para enfermedades graves como el cáncer cervical donde un diagnóstico perdido puede tener consecuencias fatales.

$$\frac{FN}{TP + FN}$$

- **Exactitud (Accuracy):** Calcula la proporción de verdaderas predicciones (tanto verdaderos positivos como verdaderos negativos) en relación con el total de casos. Esta métrica ayuda a entender la eficacia general del método de diagnóstico utilizado.

$$\frac{TP + TN}{TP + FP + TN + FN}$$

- **Precisión (Precision):** Evalúa la proporción de verdaderos positivos frente a todos los resultados que el sistema ha clasificado como positivos. Esta métrica es importante para determinar la eficacia de un método en la identificación correcta de lesiones cancerosas.

$$\frac{TP}{TP + FP}$$

- **Sensibilidad (Recall):** Se refiere a la capacidad del método para identificar correctamente todos los casos positivos. Es especialmente vital en el contexto médico para asegurar que todos los pacientes con posibles lesiones cancerosas sean identificados para tratamientos adicionales.

$$\frac{TP}{TP + FN}$$

- **Puntuación F1 (F1 Score):** Combina la precisión y la sensibilidad en una sola métrica mediante su media armónica. Es útil para evaluar la precisión y la exhaustividad del método de diagnóstico, proporcionando un balance entre ambas.

$$2 \times \frac{Precision \times recall}{precision + recall}$$

Etiqueta Real	Con Lesión	TN	FP
	Sin Lesión	FN	TP
		Sin Lesión	Con Lesión
		Predicción	

Figura 5: Matriz de Confusión.

Resultados

Tras un periodo intensivo de investigación y desarrollo en la aplicación de técnicas de Inteligencia Artificial para la detección de cáncer cervical, se han logrado avances significativos. La integración de algoritmos avanzados y análisis de imágenes ha permitido desarrollar soluciones robustas y eficaces para la identificación de lesiones celulares. Este logro constituye un avance crucial en la mejora de procesos médicos. A continuación se presentan los resultados obtenidos.

Precisión en la Clasificación

Se presenta inicialmente una tabla en donde se pueden ver los resultados de forma general y comparativa.

Modelos	Accuracy	Precision	Recall	F1-Score	AUC
Clasificación 2 Clases	93%	93%	93%	93%	98%
Clasificación 3 Clases	88%	88%	88%	88%	97%
Clasificación 5 Clases	83%	83%	83%	83%	97%
Clasificación 6 Clases	83%	84%	83%	83%	98%
Clasificación Pipeline	87%	87%	87%	87%	98%

Tabla 2: Resultados de todos los modelos

Los resultados de nuestro modelo de clasificación binario revelan que logra identificar correctamente la presencia o ausencia de lesiones en las células en el **93%** de los casos. De las células diagnosticadas como lesionadas, el **93%** efectivamente presentaba lesiones. Además, el modelo acertó el **93%** de las células verdaderamente lesionadas. El F1-Score alcanzó un **93%**, evidenciando un equilibrio óptimo entre precisión y sensibilidad (recall), mientras que el valor de área bajo la curva de **98%** subraya la capacidad del modelo para distinguir eficazmente entre las categorías.

La tasa de falsos negativos (FNR) del modelo es del **7.13%**. Esto indica que el modelo no logró identificar correctamente el 7.13% de las células que efectivamente presentaban una lesión, clasificándolas erróneamente como no lesionadas.

La matriz de confusión de este modelo se detalla en la **Figura 6**.

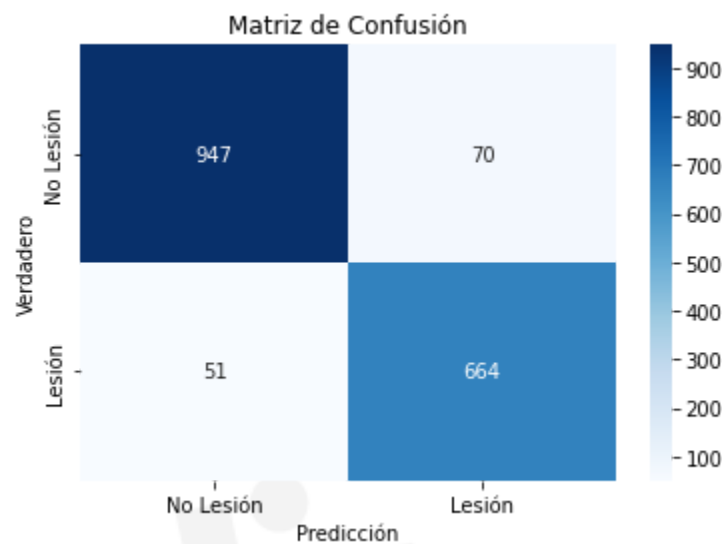


Figura 6: Matriz de Confusión de Modelo de Clasificación Binaria

Nuestro modelo de clasificación por grado de lesión ha demostrado una capacidad para identificar las categorías de lesión eficazmente. Alcanzó una precisión general del **88%**, indicando que la mayoría de las veces el modelo predijo correctamente el estado de las células. Logró identificar un **88%** de las células como pertenecientes a su grado específico de lesión. Además, el modelo fue un **88%** efectivo en identificar la presencia de lesiones cuando existían. El F1-Score de **88%** refleja un equilibrio sólido entre la precisión y la sensibilidad, crucial para determinar tratamientos adecuados en contextos clínicos. Por su parte, el valor de área bajo la curva alcanzó **97%**, destacando la capacidad del modelo para diferenciar entre categorías.

La matriz de confusión de este modelo se detalla en la **Figura 7**.

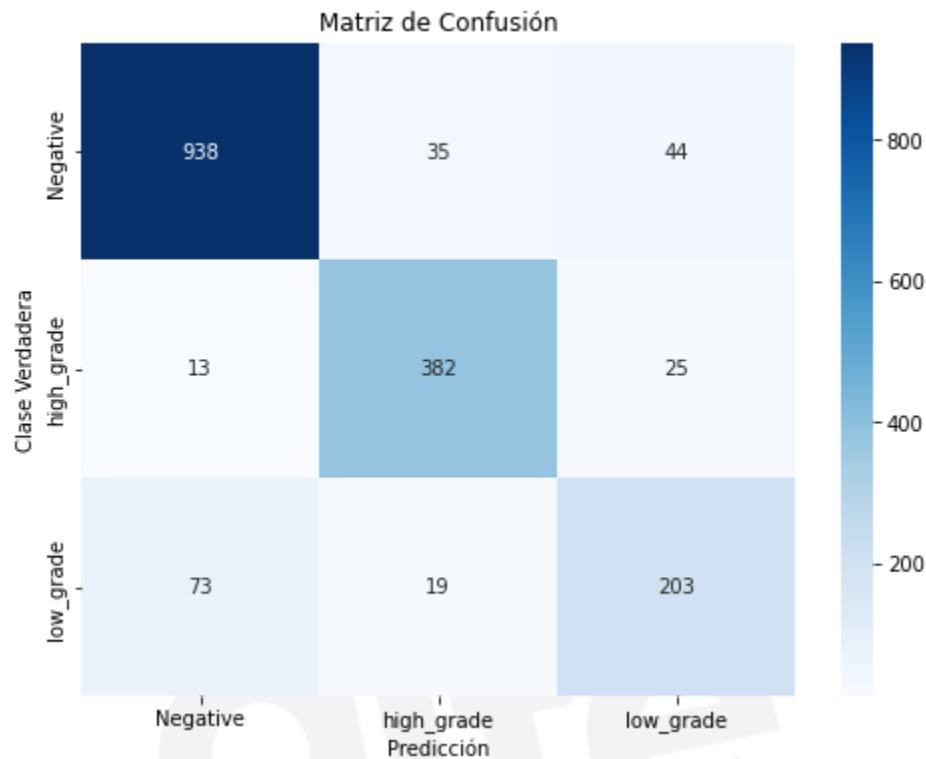


Figura 7: Matriz de Confusión de Modelo de Clasificación por Grado de Lesión

Para el modelo de clasificación enfocado exclusivamente en células lesionadas, los resultados muestran un notable desempeño en la distinción de tipos específicos de lesiones. Logró una precisión general del **83%**, lo que indica que identifica correctamente el tipo de lesión en la mayoría de los casos. La precisión del **83%** indica que, de todas las predicciones de un tipo específico de lesión hechas por el modelo, la mayoría son acertadas. La sensibilidad o recall del **83%** refleja que el modelo identifica correctamente una alta proporción de todas las lesiones verdaderas presentes en las muestras. El F1-Score de **83%** sugiere un equilibrio eficaz entre la precisión y la sensibilidad, mientras que el valor de área bajo la curva de **97%** subraya la capacidad del modelo para distinguir eficazmente entre las categorías.

La matriz de confusión de este modelo se detalla en la **Figura 8**.

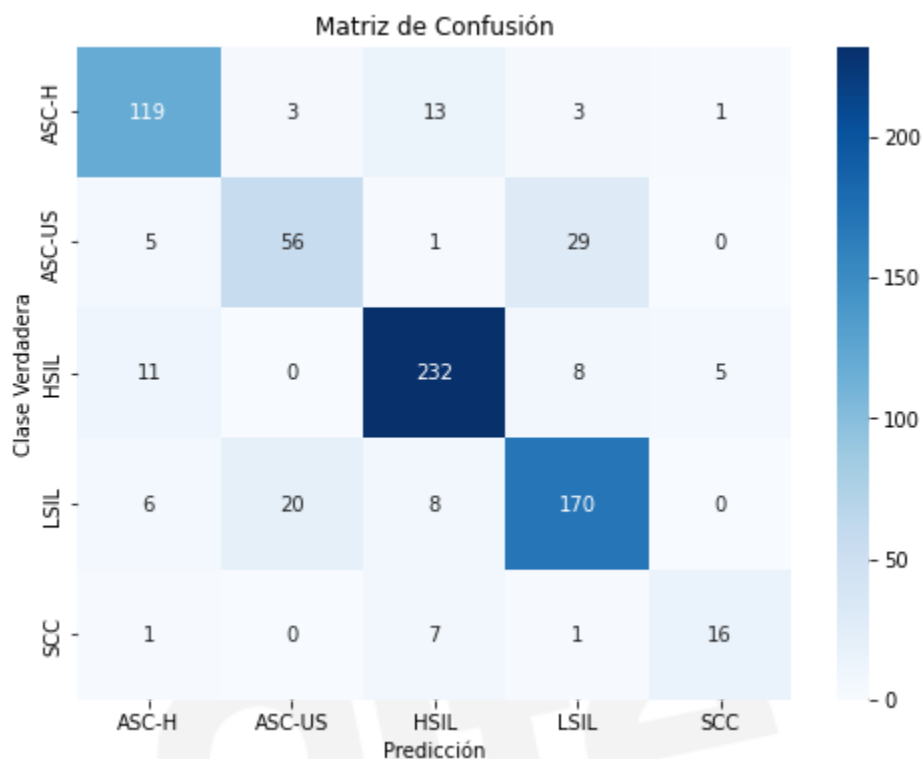


Figura 8: Matriz de Confusión de Modelo de Clasificación por Tipo de Lesión

En este segmento se detallan los resultados de dos enfoques de clasificación diseñados para realizar un análisis integral de las células cervicales. Ambos modelos no sólo distinguen entre células sin lesiones y las lesionadas, sino que también diferencian entre los distintos subtipos de lesiones. Esto permite una clasificación exhaustiva que mejora la precisión diagnóstica y eleva la relevancia clínica de los resultados obtenidos en el proyecto, contribuyendo significativamente al manejo y tratamiento médico.

Los resultados del modelo de clasificación de seis clases muestran que el modelo identifica correctamente si una célula presenta alguna lesión en el **83%** de los casos. Logra una precisión del **84%** en clasificar los tipos específicos de lesión. Además, detectó correctamente el **83%** de todas las lesiones verídicas presentes en las muestras. El F1-Score alcanzó un **83%**, demostrando un balance adecuado entre la precisión y el recall, mientras que un valor de Roc-Auc de **98%** destaca la capacidad excepcional del modelo para diferenciar entre clases de lesión con alta efectividad.

La matriz de confusión de este modelo se detalla en la **Figura 9**.

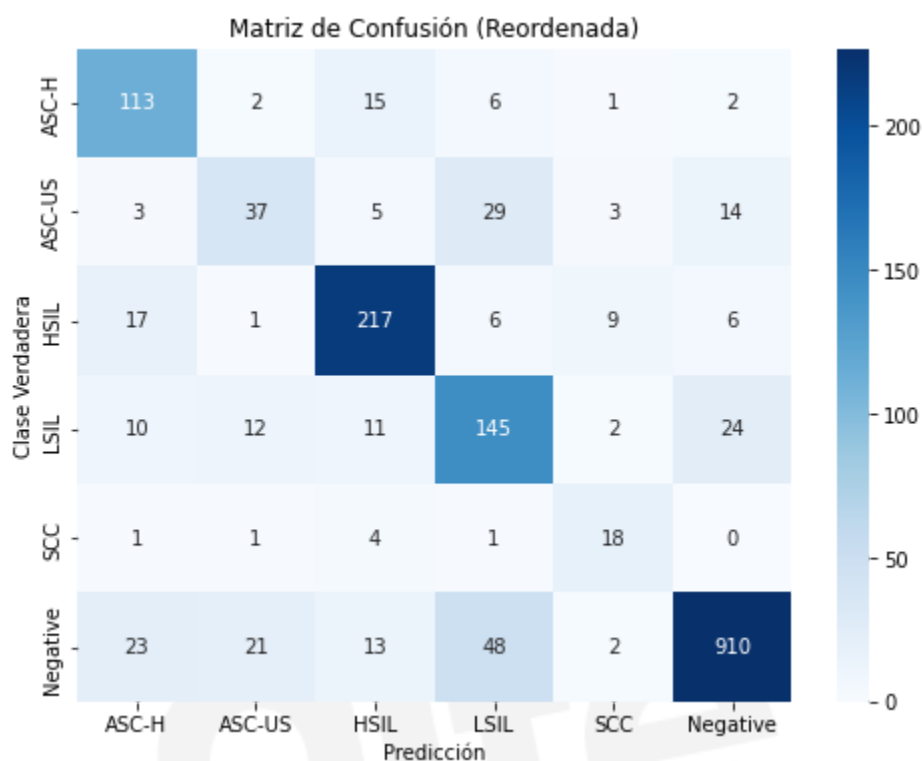


Figura 9: Matriz de Confusión de Modelo de Clasificación de 6 Clases

En el desarrollo de nuestro segundo enfoque de clasificación mediante dos modelos, el binario seguido por el clasificador específico de lesiones, hemos observado resultados alentadores. El proceso se inicia con la distinción binaria de las células lesionadas y no lesionadas, seguido por la identificación detallada del tipo de lesión en las células previamente clasificadas como lesionadas. Este enfoque ha demostrado una precisión general del **87%**, con la misma tasa para la precisión, el recall y el F1-score, indicando un balance óptimo entre la detección y la clasificación precisa de los tipos de lesión. Además, un área bajo la curva de **98%** subraya la capacidad del modelo para diferenciar efectivamente entre los diversos subtipos de lesiones en un contexto clínico.

La matriz de confusión final de este enfoque se detalla en la **Figura 10**.

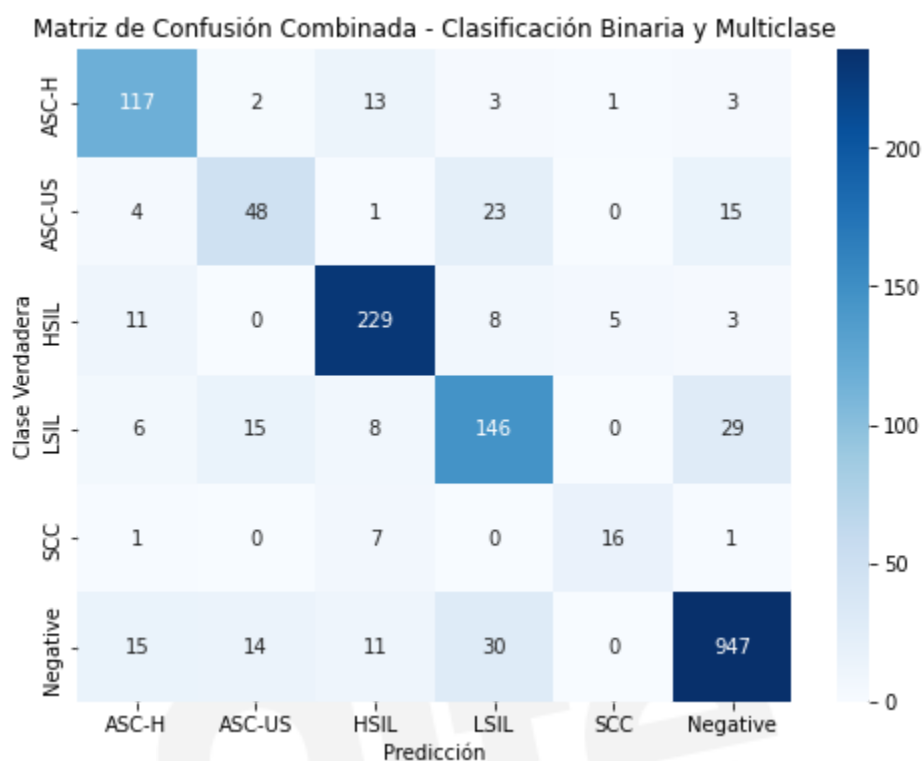


Figura 10: Matriz de Confusión del Pipeline

En la **Tabla 3** se detalla, de los dos últimos enfoques, la Tasa de Falsos Negativos (FNR), que corresponde a la proporción de células erróneamente clasificadas como no lesionadas cuando en realidad sí presentaban algún tipo de lesión. Asimismo, se especifican las Tasas de Falsos Negativos para cada tipo específico de lesión, reflejando las incidencias en que células de una categoría particular fueron incorrectamente asignadas a otra categoría de lesión. Esta desagregación permite una evaluación más precisa de la sensibilidad del modelo frente a diversas patologías celulares.

Modelos	FNR	FNR ASC-H	FNR ASC-US	FNR HSIL	FNR LSIL	FNR SCC	FNR NLIM
Clasificación 6 Clases	6.88%	19%	59%	15%	29%	10%	28%
Clasificación Pipeline	7.13%	16%	47%	10%	28%	36%	7%

Tabla 3: Tasa de Falsos Negativos

Tiempo en de Análisis y Clasificación

Los tiempos de análisis y clasificación de las imágenes enteras, evaluados en una computadora personal convencional, se desagregan de la siguiente forma:

- **Tiempo de Cropeo Promedio por Imagen:** Este proceso fue aplicado a las 400 imágenes e implica **0.0782** segundos por imagen.
- **Tiempo de Predicción Promedio por Imagen:** Refleja el tiempo requerido por el modelo para emitir una predicción por imagen.
- **Tiempo Total de Predicción Promedio por Imagen:** Es la suma del tiempo de cropeo y el tiempo de predicción promedio por imagen. Contabiliza el tiempo total que le lleva al modelo a predecir una solución final.

En la **Tabla 4** se detallan los resultados.

Modelos	Tiempo de Cropeo Promedio por Imagen	Tiempo de Predicción Promedio por Imagen	Tiempo Total por Imagen
Clasificación 2 Clases	0.0782 seg	0.1006 seg	0.1788 seg
Clasificación 3 Clases	0.0782 seg	0.0949 seg	0.1731 seg
Clasificación 5 Clases	0.0782 seg	0.0982 seg	0.1764 seg
Clasificación 6 Clases	0.0782 seg	0.1050 seg	0.1832 seg
Clasificación Pipeline	0.0782 seg	0.1988 seg	0.2770 seg

Tabla 4: Tiempos de Cropeo y Predicción

Discusión

En el siguiente apartado se realiza un análisis crítico de los resultados obtenidos. Se comparan los resultados con las metas bases de detección manual por patólogos y se evalúa si las expectativas fueron cumplidas.

A partir de un estudio realizado por Wergeland Sørbye, en el que se evaluaron cien muestras de Papanicolaou por cuatro patólogos en tres hospitales de Noruega, se observó una variación en la

precisión para detectar lesiones cervicales, que en promedio resultó en **78.95%**. Al comparar estos resultados con los de nuestros modelos automatizados, se evidencia una clara superioridad en términos de eficacia. El modelo de clasificación de seis clases alcanzó una precisión del **83%**, mientras que el modelo basado en un pipeline de dos fases logró una precisión del **87%**, superando de manera significativa la precisión obtenida manualmente. Estos resultados subrayan la ventaja de implementar técnicas avanzadas de aprendizaje automático en la clasificación de células cervicales.

En la comparación del rendimiento de los modelos implementados, el enfoque basado en un pipeline jerárquico (modelo binario seguido de un modelo de clasificación de cinco clases) demostró un desempeño superior en comparación con el modelo de clasificación directa en seis clases. Este enfoque secuencial evidenció mejoras consistentes en todas las métricas evaluadas, con incrementos de aproximadamente un 4% en *accuracy*, *recall* y *F1-score*, y un 3% en *precision*. Estos resultados destacan la efectividad del pipeline jerárquico en la optimización del proceso de clasificación, subrayando su capacidad para abordar de manera más eficiente la complejidad inherente a la tarea de clasificación multiclase.

Otro estudio realizado por la Sociedad Tailandesa de Citología mostró una tasa de falsos negativos del **13.8%**, indicando que ese porcentaje de lesiones significativas no fue identificado en la primera evaluación. En contraste, nuestros modelos redujeron esta tasa a un **6.88%** y un **7.13%**, logrando una mejora del 50% y 48% respectivamente en la detección de lesiones. Esto demuestra un notable avance en la precisión diagnóstica comparado con los métodos convencionales.

La **Tabla 5** muestra la comparación en términos de precisión y tasa de falsos negativos entre los métodos de clasificación automática y la clasificación manual realizada por patólogos.





Modelos	Precision	FNR
Situación Actual	78.95%	13.8%
Clasificación 6 Clases	83%  5%	6.88%  -50%
Clasificación Pipeline	87%  10%	7.13%  -48%

Tabla 5: Precisión y Tasa de Falsos Negativos de Clasificación Automatizada y Manual

En cuanto al tiempo de análisis y clasificación de células lesionadas, tomando como referencia el estudio de Cao et al., se estableció que el tiempo promedio requerido por citopatólogos experimentados para clasificar una célula es de aproximadamente **14.83 segundos** por imagen. En comparación, nuestros modelos automatizados demostraron una notable mejora en la eficiencia. El tiempo total de procesamiento, que incluye tanto el recorte de imágenes de núcleos como su clasificación, fue de **0.1832 segundos** por imagen para el modelo de clasificación de 6 clases, y de **0.2770 segundos** por imagen para el enfoque basado en un pipeline de dos modelos (modelo clasificación binaria seguida del modelo de 5 clases).

Sin embargo, es importante señalar que para automatizar completamente el proceso de clasificación de cáncer cervical no solo se debe considerar el recorte de las imágenes de núcleos y su clasificación, sino también la detección de los núcleos celulares en las imágenes originales de cada muestra de Papanicolaou. Por esta razón, los tiempos presentados en la **Tabla 4** no representan una referencia absoluta del tiempo total necesario para automatizar este trabajo.

De acuerdo con el estudio publicado en Current Oncology, un modelo Faster R-CNN-FPN aplicado a imágenes de la base de datos CRIC presentó un tiempo promedio de detección de núcleos de 0.277 segundos por imagen, derivado de su tasa de 3.6 cuadros por segundo (FPS).[11] Al incorporar este tiempo de detección a los resultados obtenidos en nuestro proyecto, se calcula que el tiempo total requerido para el flujo completo es de **0.4602 segundos** por imagen para el modelo de clasificación de seis clases y de **0.554 segundos** por imagen para el enfoque basado en un pipeline de dos modelos.

A pesar de este incremento, los tiempos totales estimados, que incluyen tanto la detección como la clasificación de las células, siguen siendo considerablemente menores que los 14.83 segundos necesarios para que un citopatólogo clasifique manualmente una célula. En la **Tabla 6** se puede visualizar esta comparación de tiempos, lo que resalta no solo la viabilidad de los sistemas automatizados, sino también su impacto potencial para optimizar los flujos de trabajo en la detección de cáncer cervical, reduciendo significativamente el tiempo requerido sin comprometer la precisión diagnóstica.

En relación con los modelos empleados cabe destacar que, si bien el enfoque de dos fases secuenciales dio mejores resultados, requiere un tiempo computacional mayor en comparación con el modelo de clasificación directa. No obstante, al contextualizar esta diferencia con el considerable tiempo que demanda el análisis manual por citopatólogos, ambos enfoques presentan mejoras significativas en términos de eficiencia. Estos resultados posicionan a ambos modelos como herramientas prometedoras para la automatización de tareas en la detección de lesiones cervicales, optimizando los flujos de trabajo clínicos sin comprometer la precisión diagnóstica.





Modelos	Tiempos
Situación Actual	14.83 seg
Clasificación 6 Clases 	0.46 seg 
Clasificación Pipeline 	0.55 seg 

Tabla 6: Tiempos de Clasificación Automatizada y Manual

Por último, en relación con la adopción de la metodología basada en el análisis de núcleos celulares recortados, compartimos la premisa de que el núcleo celular proporciona suficiente información para clasificar una célula según su grado de lesión. Este enfoque no solo ofrece una ventaja significativa frente a la clasificación de imágenes completas, al facilitar un manejo más eficiente del desbalance de clases presente en la base de datos, sino que también representa una alternativa más sencilla y práctica para la implementación de modelos de clasificación. Al reducir el espacio de búsqueda y centrarse en las características más relevantes, este método optimiza tanto los recursos computacionales como la precisión del modelo.

Conclusiones

Los resultados obtenidos respaldan ambas hipótesis planteadas. En primer lugar, se confirma la hipótesis principal, que sostiene que el uso de algoritmos de aprendizaje automático en la clasificación de células cervicales mejorará la precisión y reducirá el tiempo de análisis. Al comparar los modelos automatizados con los métodos manuales, se observó no solo una mejora significativa en la precisión, sino también una notable reducción en el tiempo de procesamiento, cumpliendo así con los objetivos de optimización del análisis.

En segundo lugar, los resultados validan la hipótesis secundaria, que postulaba que el enfoque de pipeline jerárquico (modelo binario seguido de un modelo especializado en cinco clases) sería más preciso que un modelo único para la clasificación en seis clases. El análisis de las métricas de rendimiento mostró que el enfoque secuencial logró mejoras significativas en comparación con el modelo integrado, lo cual demuestra la efectividad del pipeline en mejorar la clasificación de células cervicales.

A partir de estos resultados, proponemos que los modelos desarrollados sean utilizados como herramientas complementarias al trabajo de los citopatólogos. En este contexto, es importante señalar que gran parte de las pruebas analizadas manualmente corresponden a células normales (sin lesión), lo que genera una carga de trabajo innecesaria para los especialistas. La automatización de estas tareas repetitivas optimiza el flujo de trabajo y permite a los especialistas

centrarse en los casos que realmente lo requieren. Si bien la automatización aporta mejoras significativas en términos de eficiencia, la supervisión y validación humana continúan siendo esenciales para garantizar la fiabilidad diagnóstica. Este enfoque reduce la cantidad de citopatólogos necesarios para analizar una misma muestra, permitiendo que solo uno se encargue de la supervisión. Así, se incrementa la eficiencia global del proceso sin comprometer la precisión del diagnóstico.

En resumen, los resultados obtenidos demuestran que tanto la implementación de algoritmos de aprendizaje automático como el uso de un enfoque secuencial basado en un pipeline han cumplido con las expectativas de mejora en precisión, eficiencia y reducción de errores diagnósticos. Estos avances posicionan a la solución como una opción viable y altamente eficaz para la clasificación automatizada de células cervicales, mostrando como la integración de nuevas tecnologías en el proceso diagnóstico tiene el potencial de transformar el manejo del cáncer cervical, facilitando una detección más temprana y precisa, lo que podría marcar una diferencia crucial en el tratamiento y pronóstico de los pacientes.

Investigación Futura

Este proyecto abre múltiples posibilidades para la investigación futura y la implementación práctica en el ámbito clínico, permitiendo una mejora continua del sistema propuesto. Una dirección prometedora sería desarrollar un **modelo de detección** automatizado que identifique de forma precisa los núcleos celulares relevantes en imágenes completas de Papanicolaou, eliminando así la necesidad de detección manual y aumentando la escalabilidad del proceso. Además, se podrían explorar técnicas de **clustering** para identificar patrones en los tipos de lesiones, especialmente en casos como los que se muestran en la **Figura 11**, donde las cajas de detección se solapan. En situaciones así, los núcleos que se encuentran cercanos entre sí generalmente pertenecen a la misma clase. En los ejemplos, las cajas rosas corresponden a núcleos de la clase HSIL y las marrones corresponden a la clase ASC-H, que tienden a agruparse en "manada" dentro de las imágenes. Dado que la agrupación es un comportamiento común en ciertas clases, el clustering optimizaría el rendimiento del modelo al enfocarse en las agrupaciones naturales de los núcleos, en lugar de tratar cada detección de manera aislada, lo que podría mejorar la precisión.

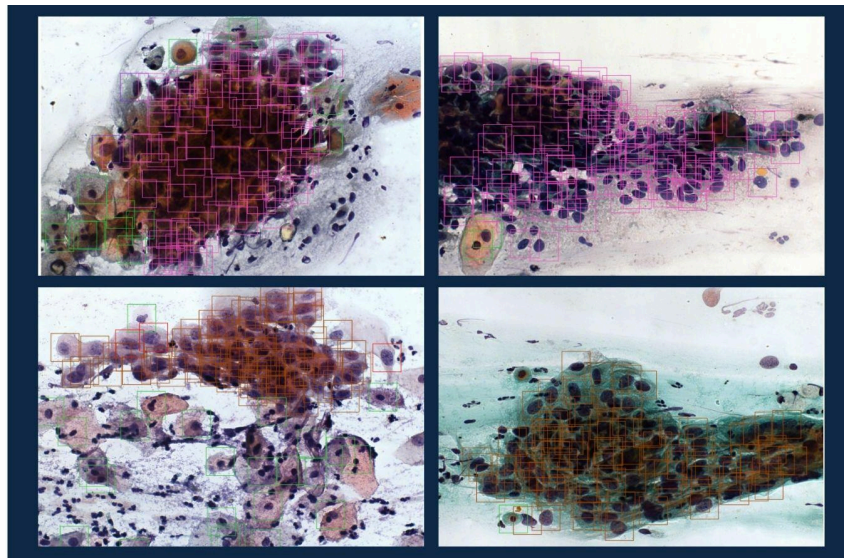


Figura 11: Visualización de Aglomeración Celular en Muestras

Asimismo, para facilitar la integración en entornos médicos reales, sugerimos la creación de una **interfaz de usuario intuitiva** que permita a los médicos utilizar los modelos de detección y clasificación de manera eficiente y accesible, optimizando el flujo de trabajo clínico. También consideramos que podría ser interesante implementar **otro pipeline**, que primero realice una clasificación inicial en tres categorías (negative, high-grade y low-grade), seguido de dos modelos especializados que clasifiquen más detalladamente las lesiones en cada categoría. En este enfoque, las bounding boxes clasificadas como “low-grade” por el modelo de 3 clases serán utilizadas como entrada para un modelo que clasifica entre ASC-US y LSIL, mientras que las clasificadas como “high-grade” servirán como entrada para un modelo que clasifica entre ASC-H, HSIL y SCC. Este diseño modular podría mejorar la precisión y lograr un mayor rendimiento.

Otra línea de investigación podría ser explorar el impacto de **ajustar los tamaños de las bounding boxes**. Aunque este proyecto utilizó tamaños de 90x90 píxeles, investigar otros tamaños podría ayudar a determinar la configuración óptima para capturar suficiente información relevante del núcleo sin incluir áreas irrelevantes. Esta exploración podría mejorar tanto la precisión como la eficiencia del modelo, especialmente si se encuentra un tamaño que balancee adecuadamente estos factores. Una alternativa interesante sería probar **bounding boxes de tamaños variables** que se adapten al tamaño específico de cada núcleo (por ejemplo, de 60x50, 70x70, etc.), en lugar de mantener un tamaño fijo. Este enfoque podría evitar que núcleos de otras células ingresen en la bounding box a clasificar, mejorando la precisión del modelo. Aunque exploramos esta idea en las etapas iniciales del proyecto, no llegamos a desarrollarla por completo, pero consideramos que sería valioso retomarla en futuras investigaciones.

A su vez, consideramos que sería relevante probar la performance del modelo en **otros conjuntos de datos** para evaluar su capacidad de generalización y adaptabilidad a diferentes escenarios clínicos. Comparar los resultados en bases de datos con diferentes características, distribuciones y resoluciones podría proporcionar información clave sobre las fortalezas y limitaciones del enfoque actual, así como posibles ajustes necesarios para su implementación práctica.

Estas propuestas no solo amplían el alcance del proyecto, sino que también consolidan su aplicabilidad práctica al ofrecer soluciones adaptadas a las necesidades reales del ámbito clínico. La implementación de un pipeline modular, en combinación con técnicas avanzadas de detección, clustering y optimización de bounding boxes, no solo busca mejorar la precisión de los modelos, sino también garantizar su eficiencia y facilidad de uso en entornos médicos reales. Al abordar las limitaciones actuales y ofrecer herramientas que optimicen tanto el análisis automatizado como la interacción usuario-modelo, estas líneas de trabajo sientan las bases para una integración efectiva de la inteligencia artificial en la práctica médica diaria, favoreciendo así una mejor toma de decisiones clínicas y una atención más precisa y oportuna para los pacientes.

Aprendizaje

A lo largo del desarrollo de nuestra investigación, enfrentamos desafíos relacionados con la claridad y el enfoque de nuestro proyecto. Estos desafíos nos dejaron un aprendizaje significativo: la importancia de establecer objetivos claros desde el inicio. Es importante que el equipo esté alineado con el propósito del proyecto en todo momento. Además, trabajar con un project owner abierto y receptivo a nuevas ideas nos permitió explorar múltiples perspectivas, lo cual fue muy enriquecedor, pero también nos enseñó que la flexibilidad debe equilibrarse con una dirección definida para evitar la dispersión. Este proceso nos enseñó la necesidad de encontrar un equilibrio entre la flexibilidad para innovar y la necesidad de tener una planificación sólida para avanzar de manera efectiva hacia nuestras metas.

Referencias bibliográficas

Lista de todas las fuentes consultadas a lo largo del trabajo, utilizando el formato bibliográfico adecuado.

1. Ministerio de la Salud, Provincia de Córdoba, Cáncer de Cuello Uterino: prevención y detección precoz, (Mar 2024)[[MinisteriodeSalud](#)]
2. Naryshkin Sonya, The false-negative fraction for Papanicolaou smears (Mar 1997) [[ProQuest](#)]

3. Cao L, Yang J, Rong Z, Li L, Xia B, You C, Lou G, Jiang L, Du C, Meng H, et al. A novel attention-guided convolutional network for the detection of abnormal cervical cells in cervical cancer screening. 2021. [[Google Scholar](#)] [[CrossRef](#)] [[PubMed](#)]
4. Débora N Diniz, Mariana T Rezende, Andrea G C Bianchi, Claudia M Carneiro, Eduardo J S Luz, Gladston J P Moreira, Daniela M Ushizima, Fátima N S de Medeiros, Marcone J F Souza, A Deep Learning Ensemble Method to Assist Cytopathologists in Pap Test Image Classification (Jul 2021) [[PubMed](#)]
5. Supinda Koonmee; Andrey Bychkov; Somruetai Shuangshoti; Kanyaprin Bhumichitra; Wanwisa Himakhun; Anant Karalak; Samreung Rangdaeng; False-Negative Rate of Papanicolaou Testing: A National Survey from the Thai Society of Cytology; (Nov 2017) [[Karger](#)]
6. Leila Allahqoli, Antonio Simone Laganà, Afroz Mazidimoradi, Hamid Salehiniya, Veronika Günther, Vito Chiantera, Shirin Karimi Goghari, Mohammad Matin Ghiasvand, Azam Rahmani, Zohre Momenimovahed, Ibrahim Alkatout, Diagnosis of Cervical Cancer and Pre-Cancerous Lesions by Artificial Intelligence: A Systematic Review, (Nov 2022) [[Mdpi](#)]
7. Diniz D.N., Souza M.J.F., Carneiro C.M., Ushizima D.M., de Medeiros F.N.S., Oliveira P.H.C., Bianchi A.G.C. An iterated local search-based algorithm to support cell nuclei detection in Pap smears test. (2019) [[Google Scholar](#)]
8. Moshavegh R., Bejnordi B.E., Mehnert A., Sujathan K., Malm P., Bengtsson E. Automated segmentation of free-lying cell nuclei in pap smears for malignancy-associated change analysis; Proceedings of the 2012 Annual International Conference of Engineering in Medicine and Biology Society (EMBC) (Sep 2012) [[PubMed](#)]
9. Samsudin N.A., Mustapha A., Arbaiy N., Hamid I.R.A. Extended local mean-based nonparametric classifier for cervical cancer screening; Proceedings of the International Conference on Soft Computing and Data Mining; Bandung, Indonesia. (Aug 2016) [[Google Scholar](#)]
10. Débora N Diniz, Mariana T Rezende, Andrea G C Bianchi, Claudia M Carneiro, Eduardo J S Luz, Gladston J P Moreira, Daniela M Ushizima, Fátima N S de Medeiros, Marcone J F Souza, A Deep Learning Ensemble Method to Assist Cytopathologists in Pap Test Image Classification (Jul 2021) [[PubMed](#)]
11. Xia Li, Zhenhao Xu, Xi Shen, Yongxia Zhou, Binggang Xiao, Tie-Qiang Li, Detection of Cervical Cancer Cells in Whole Slide Images Using Deformable and Global Context Aware Faster RCNN-FPN (Sep 2021) [[PubMed](#)]