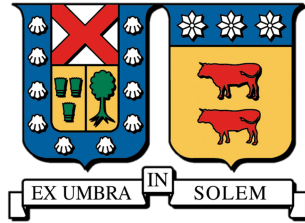


Universidad Técnica Federico Santa María
Departamento de Electrónica
Ingeniería Civil Telemática



Estimación del índice de proliferación en células cancerígenas basado en imágenes de microscopio mediante técnicas de procesamiento digital de imágenes

Memoria para optar al título de Ingeniero Civil Telemático

Autor:
Ignacio Vicente Pérez Núñez

Profesor Guía:
Marcos Zúñiga Barraza PhD

Profesor Correferente:
Nicolás Jara Carvallo PhD



CONSTANCIA DE VALIDACIÓN Y CONFIDENCIALIDAD DE MONOGRAFÍA A REPOSITORIO ACADÉMICO

1.- IDENTIFICACIÓN DEL TRABAJO ACADÉMICO

Tipo de monografía (marcar una opción): ☒ Memoria o trabajo de título ☐ Tesis de Postgrado

Título del trabajo: Estimación del índice de proliferación en células cancerígenas basado en imágenes de microscopio mediante técnicas de procesamiento digital de imágenes.

Nombre del candidato(a): Ignacio Vicente Pérez Núñez

Carrera / Grado: Ingeniería Civil Telemática

Campus: Casa Central Departamento: Electrónica

2.- VALIDACIÓN DEL PROFESOR GUÍA/DIRECTOR DE TESIS

Yo, Marcos David Zúñiga Barraza, en mi calidad de profesor guía del trabajo académico mencionado anteriormente DEJO CONSTANCIA que:

- He revisado esta versión del documento y corresponde a la versión final aprobada del trabajo.
- El trabajo cumple con los requisitos académicos y de formato establecidos por la institución.

3.- EVALUACIÓN DE CONFIDENCIALIDAD POR PROPIEDAD INDUSTRIAL (marcar una opción)

☒ El trabajo NO contiene información que amerite confidencialidad y puede ser publicado de inmediato en repositorio con acceso abierto.

☐ El trabajo CONTIENE información con potenciales implicancias de propiedad industrial o intelectual y requiere un período de confidencialidad (embargo) por (marcar una opción):

☐ 6 meses

☐ 12 meses

☐ 2 años

☐ 3 años

☐ 5 años

☐ 10 años

Fundamentación de la necesidad de confidencialidad (obligatorio si se solicita embargo):

4.- FIRMAS

Profesor(a) guía o director(a) de memoria o tesis:

Fecha: 27/11/2025 Firma: _____

Estudiante o Candidato(a):

Fecha: 27/11/2025 Firma: _____

Este formulario debe ser insertado como página 2 de la memoria o tesis, completado y firmado por estudiante y profesor(a) antes de la entrega en portal PRISMA de Biblioteca USM.

Agradecimientos

Quiero tomar este pequeño espacio para agradecer a todos aquellos que hicieron mi vida dentro de la universidad un poquitín mejor.

A todos mis compañeros que me apoyaron y trabajamos durante los ramos. A mis compañeros de memoria con los que tuve el agrado y el honor de trabajar, junto con su apoyo pudimos llevar a cabo la realización de este trabajo y espero que, pudiera haber sido de apoyo para ustedes como ustedes lo fueron conmigo.

A mis amigos que conocí durante esta etapa, no tengo palabras para expresar cuánto valoro su amistad. Ustedes se convirtieron en una familia para mí, siempre estuvieron ahí para apoyarme y alegrar mis días. Su compañía fue esencial, y siempre los llevaré en mi corazón con gratitud y cariño.

A los profesores guías quienes me apoyaron con su tiempo, paciencia y sabiduría durante el proceso del programa de memorias multidisciplinarias. Todos los comentarios y sugerencias que me dieron fueron importantes en la realización de esta memoria.

A mi familia en especial por todo, quienes siempre me brindaron su apoyo, confianza y amor durante todo el proceso. Siempre estuvieron conmigo y siempre estaré agradecido por todo.

Y finalmente lo más importante, este trabajo está dedicado a mi madre, quien fue el pilar fundamental de mi vida y de mis razones para llegar a esta universidad. Espero que ella pueda ver en lo que se convirtió su hijo y esté orgullosa de lo que pude lograr gracias a su apoyo y amor, te quiero mucho estés donde estés.

Contenidos

Agradecimientos	iii
Figuras	vii
Tablas	ix
1 Introducción	1
1.1 Problema por resolver	1
1.2 Acercamiento a la solución	2
1.3 Hipótesis	3
1.4 Objetivos	3
1.4.1 Objetivos Específicos	3
1.5 Estructura	4
2 Estado del Arte y de la técnica	5
2.1 Estado del arte	5
2.2 Estado de la técnica	7
2.3 Marco Teórico	8
2.4 Discusión	11

<i>CONTENIDOS</i>	vi
3 Desarrollo de la solución	13
3.1 Análisis del contexto del desafío	14
3.2 Desafíos	17
3.3 Diseño de la solución	20
3.3.1 Selección de ROI	21
3.3.2 Implementación algoritmo de thresholding y segmentación basada en colores	25
3.3.3 Detección de bordes	29
3.3.4 Algoritmo de Canny	33
3.4 Implementación	38
3.5 Resultados y validación	44
4 Conclusiones	62
4.1 Conclusiones	62
4.2 Trabajo a futuro	63
Bibliografía	66

Figuras

3.1	Arquitectura de Solución	14
3.2	Proceso de atención y tratamiento de un paciente	15
3.3	Diagrama general del sistema	16
3.4	Imágenes de muestras de biopsia	18
3.5	Sobrelapamiento de células	19
3.6	Células Proliferantes Falsas	20
3.7	Selección de ROI	22
3.8	ROI en área de oscuridad	23
3.9	Selección de ROI manual	24
3.10	Muestra de biopsia HSV	26
3.11	Máscara gris de células proliferantes	27
3.12	Imagen HSV de células no proliferantes	28
3.13	Imagen binaria de muestra biopsia	29
3.14	Detección de contornos primera versión.	31
3.15	Conteo células proliferantes versión 2	32
3.16	Conteo células no proliferantes versión 2	33
3.17	Fases del algoritmo de Canny en muestra Biopsia	34
3.18	Fases Algoritmo Canny	36
3.19	Detección de colores mediante tresholding	39
		vii

<i>FIGURAS</i>	viii
3.20 Segmentación cromática	40
3.21 Detección células proliferantes con operador Canny	41
3.22	42
3.23 Implementación de algoritmo sobre una muestra de biopsia . .	43
3.24 Ejemplo clasificación de tamaños	45
3.25 Ejemplo de primera iteración	48
3.26 Clasificación de tamaños	49
3.27 Muestra 1	51
3.28 Muestra 2	52
3.29 Muestra 3	53
3.30 Muestra 4	54
3.31 Muestra 5	55
3.32 Muestra de solapamiento de células 1	57
3.33 Muestra de solapamiento de células 2	58
3.34 Muestra de solapamiento de células 3	59
3.35 Muestra de solapamiento de células 4	60

Tablas

3.1	Comparación Muestra 1	51
3.2	Comparación Muestra 2	52
3.3	Comparación Muestra 3	53
3.4	Comparación Muestra 4	54
3.5	Comparación Muestra 5	55
3.6	Detección de solapamiento	57
3.7	Detección de solapamiento	58
3.8	Detección de solapamiento	59
3.9	Detección de solapamiento	60
3.10	Comparación muestras de solapamiento	61

CAPÍTULO 1

Introducción

1.1 Problema por resolver

En la actualidad, el cáncer es la principal causa de muerte en el mundo, en donde aproximadamente 1 de cada 6 personas fallecen debido a esta condición. Se proyecta que para el año 2040, habrá un aumento del 60% en los casos, situación más complicada para países más inestables económicamente, pues el porcentaje podría aumentar hasta un 81%. Chile no se queda ajeno a esta realidad y es por esta razón que surgen distintas organizaciones para frenar el avance de esta enfermedad.

La Fundación Arturo López Pérez (FALP) es uno de los varios ejemplos de entidades preocupadas por la investigación y tratamiento de enfermedades oncológicas, por lo que demuestran un constante interés en desarrollar herramientas que faciliten el trabajo de sus especialistas a la hora de evaluar a sus pacientes.

Para este caso en particular, la FALP busca desarrollar una solución tecnológica, que permita a los doctores patólogos contar con un método de estandarización en apoyo a la estimación del índice de proliferación de las células cancerígenas dado una muestra de biopsia. Este índice señala la velocidad de proliferación de una célula, por lo que se encuentra relacionado con el estado de gravedad del cáncer que se busca investigar. Cabe destacar que los doctores en la FALP realizan los cálculos de estimación manualmente, esto quiere decir que se requiere estar observando las muestras a través de sus microscopios, y realizar el conteo individual de células en una muestra.

1.2 Acercamiento a la solución

Como se mencionó anteriormente, la actual problemática que existe es la carencia de un método de estandarización en la toma de muestras y en los cálculos que son realizados por parte de los doctores patólogos.

Es así que se propone abordar esta situación mediante la implementación de una herramienta tecnológica destinada a proporcionar asistencia y eficiencia al médico en su labor. La herramienta consistirá en una aplicación de escritorio, principalmente basada en la toma y manipulación de imágenes de muestras de biopsia. Los resultados se verán mostrados en una interfaz y los datos de importancia serán guardados en una base de datos. En esta memoria se dará enfoque en lo que es el trabajo de manipulación de imágenes mediante técnicas de visión por computador.

Un acercamiento inicial consistía en el uso de técnicas de Deep Learning (DP), y un entrenamiento de modelos para la identificación automática de muestras de biopsia con tecnología de Machine Learning (ML). Sin embargo, estas propuestas son descartadas debido a que en la actualidad, la FALP no cuenta con la cantidad de muestras necesarias para poder desarrollar un

algoritmo de ML. Es así que se optará por trabajar mediante un acercamiento a técnicas de segmentación de imágenes y conteo de contornos.

1.3 Hipótesis

- Es posible desarrollar un algoritmo de procesamiento digital de imágenes que, utilizando parámetros dados, sea capaz de clasificar las muestras de biopsia con una precisión comparable a la de un patólogo humano.

1.4 Objetivos

El objetivo general del proyecto consiste en la creación de una aplicación de escritorio que permita el cálculo del índice de proliferación, esto mediante imágenes que serán entregadas por capturas de microscopios en tiempo real.

1.4.1 Objetivos Específicos

OE1 - Definir un esquema general y análisis de tecnologías de visión por computador.

OE2 - Investigar estado del arte y estado de la técnica para el cálculo del índice de proliferación.

OE3 - Desarrollar un programa capaz de leer, analizar y segmentar imágenes de microscopios.

1.5 Estructura

En primer lugar se hablará del estado del arte, identificando cuáles técnicas existen y se están empleando en problemas de alcance similar, el contexto en el que fueron aplicadas y si están siendo adaptables al desafío abordado en esta memoria. Posteriormente se describirá el proceso del desarrollo para solucionar el desafío; donde en primer lugar se hace un análisis en profundidad del contexto del problema, las indicaciones entregadas por la contraparte, las reuniones realizadas con los expertos en el área y cómo sus indicaciones fueron estructurando el modelo de la aplicación y las técnicas ocupadas para el desarrollo de esta. Se contará con un capítulo sobre el acercamiento hacia las técnicas utilizadas para llegar a la solución planteada en la sección de objetivos; así como la parte de prueba y la validación de la hipótesis planteada. Por último, se hace un análisis y conclusiones de los resultados obtenidos, primero para la validación de la hipótesis y posterior discusión acerca de la utilidad para la contraparte y el trabajo a futuro que tendrá como alcance este proyecto.

Estado del Arte y de la técnica

2.1 Estado del arte

Enfocándose en el trabajo con células utilizando tecnología de visión por computador, existe una gama bastante amplia de información. Cabe destacar que la mayoría de estas aproximaciones utilizan técnicas de Deep Learning (DL), siendo las más destacadas *Segment Anything in Medical Images* [1] la cual trabaja con segmentaciones de imágenes basadas con un dataset entregado, y por otra parte *MaskR-CNN* [2], el cual propone la creación de un framework para la segmentación de instancias de objetos en imágenes, basado en una arquitectura de red neuronal convolucional(CNN). Dado el contexto en el que se desarrolla esta memoria, el uso de Machine Learning (ML) está descartado, por lo que no se ahondará en la investigación de documentos que propongan soluciones con este tipo de acercamiento.

Si bien no existen documentos que trabajen la detección de células sin

el uso de DL, existen varios que se centran mayoritariamente en técnicas de segmentación de imágenes preprocesadas para la detección de contornos celulares. Por ejemplo, la tesis *Segmentación de células mediante técnicas de Procesamiento Digital de Imágenes para el rastreo de células cancerosas* [3] propone como una de las soluciones una ecualización de histogramas para la mejora de contrastes de imágenes, así como el uso de un operador morfológico de erosión para la detección de límites de estructuras que correspondan a células. A pesar de que esta solución fue descartada para enfocarse en una solución de DL, para el caso de este proyecto sí correspondería como una opción a tomar en cuenta, en cuanto a las técnicas de CV que aplica.

Por otro lado, el documento *Segmentación de imágenes de células cervicales y evaluación de características de lesiones neoplásticas*[4] propone un método de segmentación para clasificar el fondo, el citoplasma y el núcleo en imágenes de células aisladas mediante interacciones de un usuario en una región de interés (ROI). Explica el uso de técnicas de detección de bordes mediante la reconstrucción de la imagen original en niveles de grises, la aplicación de un filtro de media y la extracción de bordes con el método Canny. Este documento considera los valores del perímetro y el área de las imágenes, centrándose en el conteo de los píxeles de una célula para lograr la detección de su núcleo, considerando el tamaño del ROI. Este documento es el que más se acerca a la solución propuesta para el proyecto, con la diferencia principal de que estará detectando múltiples células de una muestra general, destacando aquellas marcadas por la proteína KI-67 entre varias.

2.2 Estado de la técnica

En cuanto a las tecnologías actuales en el mercado, existen varias aplicaciones cuya función principal consiste en el procesamiento digital de imágenes y de documentos médicos. Un ejemplo es el programa ImageJ[5], un software de código abierto y de dominio público desarrollado por el National Institutes of Health (NIH). ImageJ es utilizado principalmente por médicos para analizar imágenes de células y detectar anomalías. Una de sus características destacadas es la posibilidad de seleccionar un ROI dentro de una imagen dada.

El proyecto planteado se diferencia de ImageJ en que tendrá como principal función la detección de células dentro de un ROI y el cálculo del índice de proliferación.

En el documento "Methods for Cell and Particle Tracking"[6] se mencionan diferentes herramientas tecnológicas en el mercado para la detección y rastreo de células. Entre ellas se encuentran ImageJ y Qupath[7], un software de código abierto creado por la Universidad de Ulster en Irlanda del Norte. La principal diferencia entre Qupath e ImageJ es que la primera está desarrollada principalmente para el análisis de imágenes biomédicas. Esto significa que incluye una serie de funciones que son útiles para este tipo de análisis, como la segmentación de imágenes de biopsias e integración con bases de datos.

En un sentido técnico, Qupath es la tecnología más cercana a la que se desea acercar el desarrollo del proyecto, dada sus funciones y visualización interactiva.

Las tecnologías mencionadas son las bases para la visión que se tiene para el desarrollo. Lo que diferenciará este proyecto será el fin que se le dará, es decir, mientras que los softwares de código abierto proporcionan diversas

características amplias para distintas áreas de la medicina, el proyecto desarrollado tendrá un enfoque principal en el área patológica, siendo este una herramienta de ayuda para un doctor de fácil acceso y uso.

2.3 Marco Teórico

Tal como se mencionó en los incisos anteriores, se dará un enfoque principal en la solución del problema a las técnicas de segmentación en el análisis de imágenes. El objetivo de la segmentación es diferenciar entre los aspectos que interesan analizar y aquellos que no son de interés. Una de las suposiciones de estas técnicas es que los objetos o regiones de interés poseen diferentes características en contraste con el fondo, la cuales se pueden extraer utilizando procedimientos estadísticos. De este modo, es posible separar o segmentar regiones de interés en base a esas diferencias. En un caso ideal, el fondo de la imagen de muestras tiene una intensidad de píxeles uniforme y bien diferenciada del objeto en cuestión, pero esto no correspondería en la mayoría de casos. Además, la segmentación es en gran medida un proceso cualitativo, debido a que la mayoría de los casos no se tiene una referencia de datos empíricos (*groundtruth*) con el que comparar el resultado y determinar la efectividad de un algoritmo.

Según la tesis "*Desarrollo de técnicas de segmentación para detección de microcalcificaciones en imágenes mamográficas*"[8], se pueden clasificar los métodos de segmentación en supervisados y no supervisados. Los supervisados son aquellos en los que se dispone de un conjunto de datos etiquetados, es decir, los datos ya se encuentran asignados a un grupo (clase) específico. Por otro lado, la segmentación no supervisada no parte de una clase específica, sino más bien de un conjunto de datos sin etiquetar.

A su vez, los métodos de segmentación no supervisados son los más

usados para detectar las anomalías en imágenes médicas; y se pueden dividir en segmentación por regiones y segmentación por detección de bordes.

Las técnicas de segmentación por regiones fraccionan una imagen en regiones homogéneas que corresponden a diferentes objetos dentro de ella. Se tienen:

- **Umbralización:** La umbralización [9] es un método de segmentación de imágenes en escala de grises que divide cada píxel en dos categorías: blanco o negro. El valor de umbral se selecciona para separar los píxeles que pertenecen al fondo de los que pertenecen al primer plano. Los píxeles con valores de intensidad superiores al umbral se asignan al blanco, mientras que los píxeles con valores de intensidad inferiores o iguales al umbral se asignan al negro. De esta forma, se puede resaltar la forma y el contorno de los objetos de interés en la imagen.
- **Crecimiento de regiones:** Este método extrae de una imagen regiones que están conectadas entre sí según un criterio predefinido. A menudo, este criterio está basado en información de intensidades y/o bordes de la imagen. En su forma más simple, este método requiere un punto semilla que es seleccionado manualmente por el usuario e iterativamente extrae todos los píxeles conectados a la semilla que tengan un valor de intensidad similar. Algunas técnicas utilizadas son, por ejemplo el algoritmo de Watershed [10], el cual se basa en segmentación basada en la topología de una imagen, dividiéndose en regiones separadas por líneas de drenaje. Otra opción es el algoritmo de "conectividad regional". En concreto, se basa en la conectividad entre píxeles adyacentes. Estas técnicas son efectivas para la detección de células en imágenes, pero pueden presentar dificultades en casos de solapamiento de células o estructuras de interés, lo que puede reducir

su precisión. Además, el crecimiento de regiones requiere un alto coste computacional, por lo que esta metodología no siempre es la más adecuada.

- **Segmentación por agrupamiento:** Se basa en la idea de que los objetos que posean características similares deberían agruparse en un mismo grupo o región (clusters). En este enfoque, no se necesita información previa sobre la forma o la apariencia de los objetos en una imagen. En cambio, se buscan similitudes en las propiedades de los píxeles o regiones de la imagen y son agrupados. Como ejemplo, se tiene el algoritmo "K-Means"[11], el cual divide un conjunto de datos en una cantidad determinada de grupos; cada uno representado por un centroide, el cual es el punto medio de los datos en el grupo. Otra opción que se tiene es el algoritmo DBSCAN (Density-Based Spatial Clustering of Applications with Noise)[12], el cual está enfocado en encontrar clusters en función de la densidad de los puntos en el espacio. A diferencia de "K-Means", que se basa en la geometría y distancia, DBSCAN lo hace por medio de la proximidad y densidad de los puntos. Es particularmente útil para encontrar clusters de formas arbitrarias y puede identificar puntos ruidosos o atípicos. Estas técnicas tienen bastantes ventajas, como robustez al ruido y poder detectar grupos de cualquier forma. Sin embargo, para aprovecharlos al máximo se necesitaría el uso de una cantidad grande de datos, o imágenes en este contexto; por lo que dado el caso se descartan estas técnicas.

Por otro lado, se encuentra la **segmentación basada en la detección de bordes**, que se enfoca en identificar discontinuidades en la imagen, es decir, puntos donde se producen cambios significativos en los niveles de gris. En algunos casos, la detección se logra mediante el uso de gradientes y

operadores laplacianos. La detección de bordes a través de operadores de gradiente es efectiva en situaciones donde la imagen presenta transiciones de intensidad bien definidas y niveles de ruido relativamente bajos. No obstante, en imágenes de biopsia con múltiples células cercanas, la aplicación de esta técnica puede ser complicada debido a la alta densidad de bordes y discontinuidades presentes. Por otro lado, el operador laplaciano es altamente susceptible al ruido, lo que puede resultar en errores en la detección de bordes.

Un enfoque alternativo es el uso del algoritmo de Canny[13], el cual se basa en la identificación de máximos locales en la magnitud del gradiente de la imagen. El proceso incluye el suavizado de la imagen con un filtro gaussiano, el cálculo del gradiente para determinar la dirección y magnitud del cambio de intensidad, la supresión de no máximos, la aplicación de umbrales para detectar máximos y descartar aquellos que no se consideran parte del borde, seguido de la conexión de los puntos de gradiente que forman un contorno. En el contexto de la segmentación de células en imágenes, el uso de este algoritmo puede ser una opción apropiada, especialmente cuando las células están bien definidas y contrastan significativamente con su entorno. En situaciones donde se busca detectar células marcadas, el algoritmo Canny ofrece ventajas, ya que incorpora múltiples etapas para refinar la detección y eliminar detecciones falsas, lo que lo convierte en una elección particularmente útil cuando los bordes son nítidos y la imagen presenta ruido.

2.4 Discusión

Después de recopilar la información y sostener conversaciones con la contraparte, se determina la necesidad de desarrollar una aplicación con la función principal de calcular el índice de proliferación. Esto se logrará mediante

la detección de las células proliferantes que expresan la proteína KI67. El proyecto implica la creación de un programa con una interfaz y herramientas similares a las de Qupath, especialmente en lo que concierne a la selección de un Área de Interés (ROI) y la posterior cuantificación de los elementos relevantes dentro del ROI.

Considerando las tecnologías seleccionadas para el desarrollo de una propuesta de solución, es necesario tener en cuenta varios aspectos. El primero se refiere al tipo de umbralización a utilizar en el conteo de células, que puede ser global, implicando un umbral único para toda la imagen, o local, lo que implica la selección de umbral distinto para cada píxel de una imagen según la intensidad de sus vecinos.

Es importante destacar que la umbralización, por sí sola, no es un método infalible. Esto se debe a que la variación en la iluminación de las imágenes, el contraste y la presencia de ruido pueden dificultar la elección del umbral adecuado.

En este caso, se opta por emplear una combinación de técnicas de segmentación, específicamente la umbralización y la detección de bordes a través del algoritmo Canny. Esta elección se basa en la complementariedad de estas técnicas, ya que permiten resaltar células cancerígenas con intensidades distintas al fondo de la imagen y lograr una detección precisa de los bordes.

CAPÍTULO 3

Desarrollo de la solución

El desarrollo abarcó tres secciones. La primera se refiere a la interfaz de la aplicación web, que actuará como la capa de presentación para la interacción de los usuarios con el sistema. Esta contará con un diseño claro y estructurado, incorporando elementos visuales y gráficos con un enfoque central en la interactividad, mediante controles de usuario y opciones de entrada. La interfaz se diseñará principalmente para ser compatible como una aplicación de escritorio para los equipos de la FALP. Luego se contempla una sección de arquitectura de la base de datos relacional. Esta sección abordará el almacenamiento eficiente y la gestión de datos. Se selecciona un sistema de base de datos relacional como PostgreSQL. La arquitectura permitirá la normalización de datos para garantizar la consistencia e integridad del programa, diseñando un esquema que refleje las relaciones entre entidades y optimice las consultas de datos.

Finalmente, se incorpora la sección de visión por computador, en la

cual se llevará a cabo el desarrollo del algoritmo de detección de contornos celulares. Esta sección también se encargará de proporcionar el cálculo del índice de proliferación.

En la siguiente figura se muestra la arquitectura de software de la solución planteada:

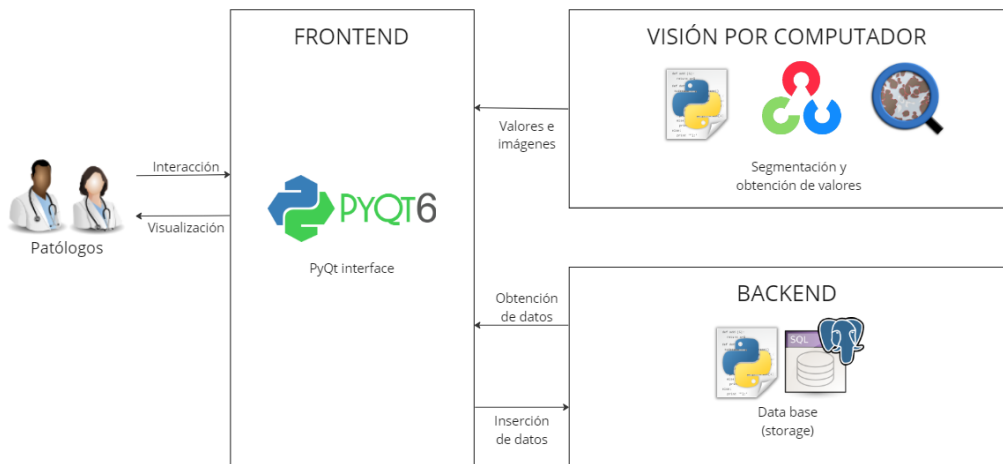


FIGURA 3.1: Arquitectura de Solución

En el marco de esta memoria, se focalizó el desarrollo de la solución en las técnicas de visión por computador, detallando el proceso de creación del algoritmo de segmentación de imágenes.

3.1 Análisis del contexto del desafío

Es necesario comprender el proceso de toma de muestras de biopsia realizado por los patólogos. Esto con el fin de comprender bajo que contexto se

encuentra el estado de desarrollo de una solución y cómo se debe abordar para su posterior implementación y en qué aspectos se puede mejorar.

En la siguiente figura se muestra el proceso de atención que un paciente llega a pasar al realizar una consulta médica y a su posterior tratamiento.

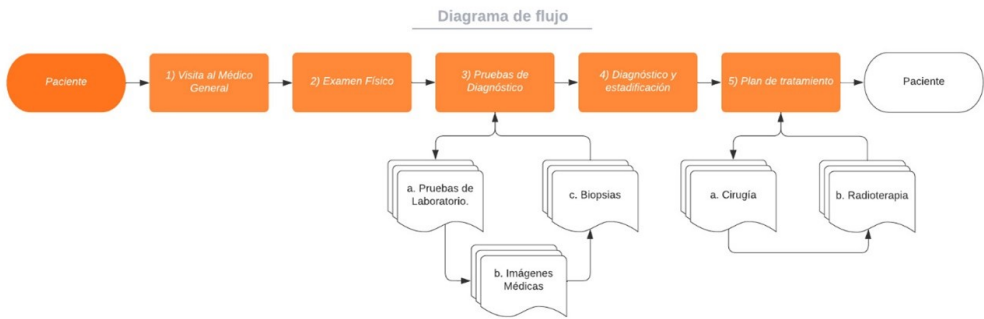


FIGURA 3.2: Proceso de atención y tratamiento de un paciente

Al observar la Figura 3.1, el proceso del paciente implica varias etapas, pero nuestro enfoque se dirige hacia las **pruebas de diagnóstico**, específicamente en las secciones de "Imágenes médicas" y "Biopsia". En caso de detectarse una masa o un tumor sospechoso, se procede a realizar una biopsia con el propósito de obtener una muestra de tejido. Es en esta sección donde se concentrará el desarrollo de la herramienta encargada de llevar a cabo el conteo de células.

El siguiente diagrama explica el flujo de datos con el que se estará trabajando en las secciones mencionadas:

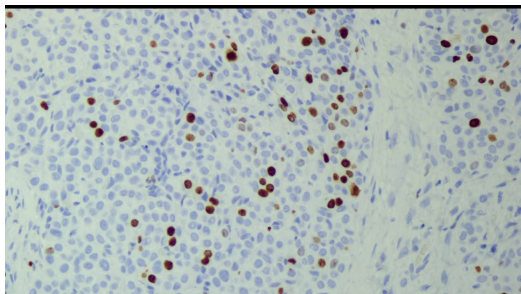


FIGURA 3.3: Diagrama general del sistema

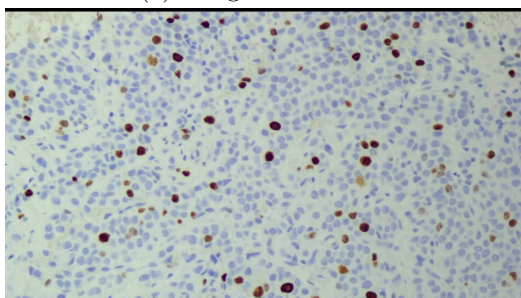
- **Captura de imagen:** Mediante una cámara integrada en el microscopio se obtiene la imagen de la muestra de tejido en tiempo real que posteriormente será enviadas al computador local para su preprocesamiento.
- **Interacción mediante interfaz:** Los usuarios pueden manipular la imagen por medio de una tableta o mouse para la selección de un ROI.
- **Procesamiento de imágenes:** Esta sección será en donde se desarrollaron los algoritmos propuestos para la segmentación de imágenes y detección de bordes para la obtención de los elementos de interés y su posterior conteo celular.
- **Almacenamiento de datos:** La información se almacena en los servidores FALP, siendo accesible en la interfaz de usuario a solicitud del mismo.

3.2 Desafíos

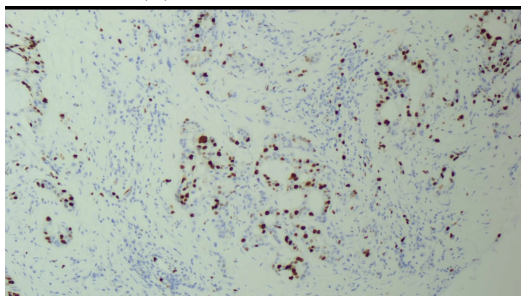
La implementación de algoritmos de segmentación de imágenes conlleva diversos desafíos, siendo la calidad de las imágenes obtenidas por parte de FALP uno de los más significativos. Para el desarrollo de este desafío, las imágenes no fueron adquiridas directamente desde la cámara del microscopio, sino que se obtuvieron a través de un dispositivo externo de captura de imágenes, lo que impacta negativamente en su calidad y resolución durante el proceso de trabajo. A continuación, se presentan algunas muestras de biopsias procedentes de los laboratorios de FALP:



(a) Imagen muestra 1



(b) Imagen muestra 2



(c) Imagen Muestra 3

FIGURA 3.4: Imágenes de muestras de biopsia

En algunas imágenes se observa ruido y baja resolución. Sin embargo, a pesar de la baja resolución, se puede distinguir el contraste entre las células

marcadas por KI67 y el resto de elementos presentes en la biopsia.

Por lo tanto, las células marcadas (negras) no deberían presentar problemas para ser destacadas mediante una segmentación en escala de grises, ya que contrastan bien con el fondo de la muestra.

El principal problema es la detección y separación de los elementos no proliferantes, como las células no marcadas por KI67 (celestes), ya que estas son más abundantes y ocupan una mayor superficie en las muestras.

Otro aspecto importante a tener en cuenta es el solapamiento entre células. Puede ocurrir que dos células se muestren en la imagen a una distancia tan cercana que parezcan una sola. Al realizar una detección de bordes, es posible que dos o más células sean contadas como una.

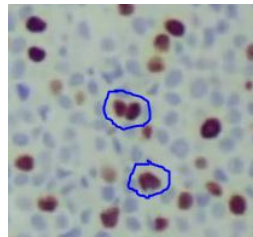


FIGURA 3.5: Sobrelapamiento de células

Para corregir el solapamiento de imágenes, se puede utilizar el algoritmo de Canny para detectar bordes. Para ello, se debe ajustar el umbral para que sea sensible a los bordes finos. Posteriormente, se puede utilizar un filtro de suavizado para reducir el ruido presente en la imagen.

De igual manera, se debe tener precaución con las células proliferantes falsas. En palabras de un doctor patólogo, por diversos motivos, una muestra puede tener células proliferantes no marcadas completamente o con un tinte

muy tenue, lo que puede llevar a que no sean consideradas como células cancerígenas.

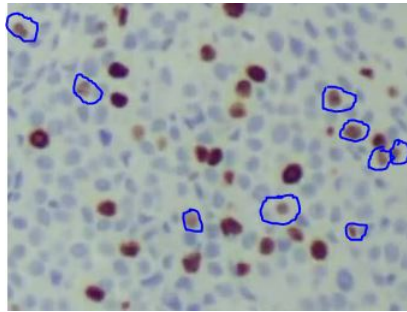


FIGURA 3.6: Células Proliferantes Falsas

Si se llega a dar este caso durante la segmentación de imagen y detección de borde, es posible que se cuenten estas células al momento del cálculo del índice de proliferación, afectando el resultado final.

Una posible solución sería establecer un umbral de opacidad del tinte para las células. Dado el caso en que se detecten células cancerígenas falsas, si estas poseen un nivel inferior al umbral, serían descartadas al momento de realizar el cálculo del porcentaje de proliferación.

3.3 Diseño de la solución

La solución se desarrolló en un entorno de trabajo basado en la tecnología de la biblioteca de visión por computadora OpenCV, empleando los parámetros específicos del lenguaje de programación Python 3.11. Este entorno proporcionó las herramientas necesarias para la implementación eficiente de algoritmos y técnicas de visión por computadora, fundamentales para el procesamiento de imágenes en el contexto del proyecto.

3.3.1 Selección de ROI

Después de importar las muestras en el sistema, se procedió con la primera fase del proceso, que implicó la manipulación de las imágenes con miras a su posterior segmentación.

Se implementó una solución focalizada en la segmentación de un ROI en lugar de realizar el proceso en la imagen completa. Esta estrategia emula el enfoque que un patólogo emplearía al seleccionar un área específica para el cálculo de índices, optimizando así el análisis y reduciendo la complejidad del procesamiento de imágenes.

En el primer caso, se implementó una solución que empleaba una selección aleatoria de regiones de interés. Esto se hizo para evaluar los límites de la imagen y la calidad del análisis en función del tamaño del ROI.

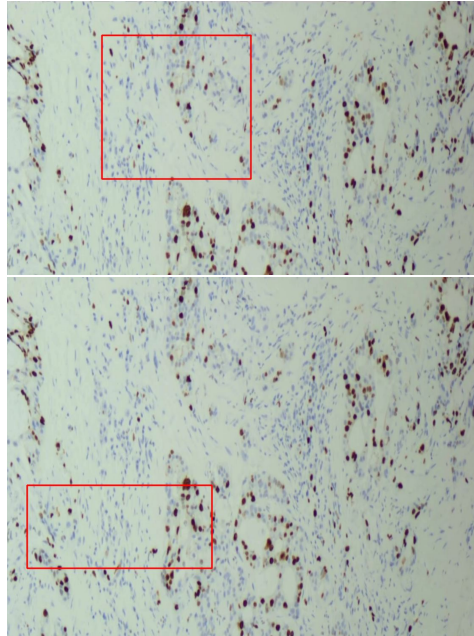


FIGURA 3.7: Selección de ROI

Debido a la forma en que la capturadora tomó las imágenes, existen ciertos límites de espacio llamados áreas de oscuridad. Estas son las áreas en las que la cámara no pudo captar la luz que reflejaba. En algunos casos, el ROI puede formar parte de estas áreas. Esto podría causar que las células proliferantes no se capturen correctamente.

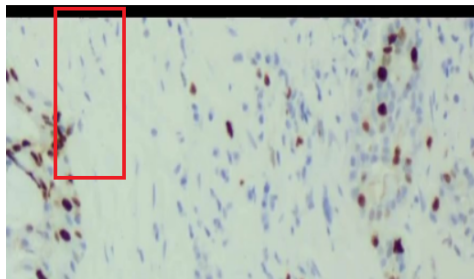


FIGURA 3.8: ROI en área de oscuridad

A primera vista, se observa que la generación aleatoria de ROI no es eficiente en virtud de estas situaciones. Además, existe la probabilidad de que se genere un ROI en una ubicación no relevante para analizar por el doctor patólogo. Este inconveniente resalta la necesidad de un enfoque más preciso en la generación de estas regiones, considerando la relevancia clínica en el proceso.

La opción óptima en este contexto es llevar a cabo la selección manual de una ROI. Este enfoque aborda de manera efectiva la problemática asociada a evitar zonas no deseadas, al tiempo que proporciona al patólogo un control total sobre el análisis de áreas específicas con células proliferantes. La selección manual de ROIs se traduce en una mayor precisión y afinidad con los requerimientos clínicos, optimizando así el proceso de segmentación y análisis de imágenes.

Así, se avanza en la implementación de una selección manual del ROI. Se adopta una solución que implica la selección de puntos en la imagen de muestra, los cuales se conectarán mediante líneas rectas hasta configurar una figura cerrada que resulte conveniente para el usuario. Este enfoque garantiza un mayor nivel de control y precisión en la delimitación del ROI, mejorando la adaptabilidad a diversas formas celulares y optimizando el

proceso de análisis de imágenes en el contexto clínico.

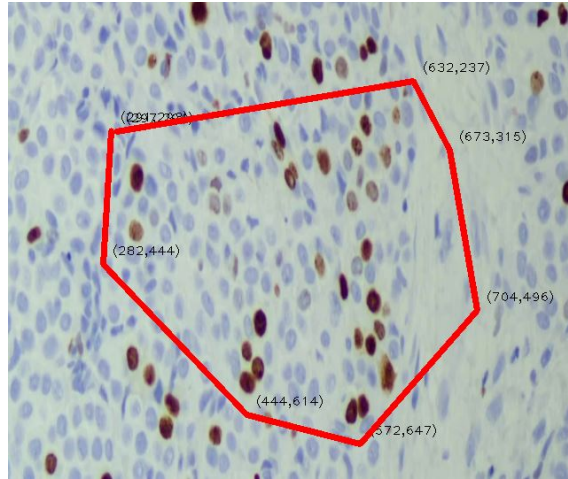


FIGURA 3.9: Selección de ROI manual

Como se evidencia en la figura 3.21, el ROI se delimita en una región donde se concentra una cantidad significativa de células proliferantes, destinada al estudio y análisis por parte del patólogo. Cada intersección de líneas está marcada por coordenadas de píxeles, facilitando el uso de la interfaz y el posterior almacenamiento en la base de datos. Se ha discutido con la contraparte la posibilidad de modificar estos puntos en tiempo real, de manera sincronizada con la actualización del índice de proliferación. La adaptación de elementos en la interfaz y sus detalles específicos no se profundizan en esta memoria, ya que el foco reside en la capacidad de realizar la segmentación de imagen y la selección de contornos celulares dentro del ROI.

3.3.2 Implementación algoritmo de thresholding y segmentación basada en colores

Se inició el primer enfoque de la solución mediante la técnica de umbralización (thresholding). Conforme se detalló en la sección de estado del arte(2), se aplicó umbralización en escala de grises para generar una imagen binaria. El objetivo principal es destacar y extraer los elementos de interés para el cálculo del índice de proliferación.

Una estrategia adoptada para abordar el problema fue la creación de diversas imágenes binarias que segmentan los distintos elementos de la muestra original. Los elementos clave para el cálculo del índice de proliferación incluyen las células marcadas por la proteína KI67 (negras) y las células no proliferativas (azules).

A la par, se emplearon técnicas de segmentación basadas en colores, las cuales consisten en dividir la imagen en segmentos según sus características cromáticas. En este contexto, se emplea el espacio de color HSV (Tono, Saturación y Valor) para establecer los intervalos cromáticos y generar máscaras que resalten los elementos de interés en la imagen original.

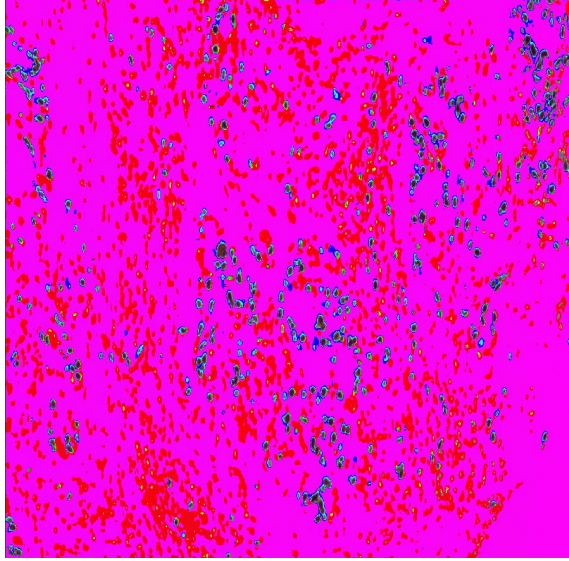


FIGURA 3.10: Muestra de biopsia HSV

Se observa la muestra de una biopsia bajo los parámetros del filtro HSV con el objetivo principal de destacar las células marcadas como proliferantes y las no proliferantes del fondo. Para facilitar una distinción más efectiva de los elementos para un posterior conteo celular, se procede a convertir la imagen en una máscara binaria, enfocada específicamente en resaltar las células proliferantes.

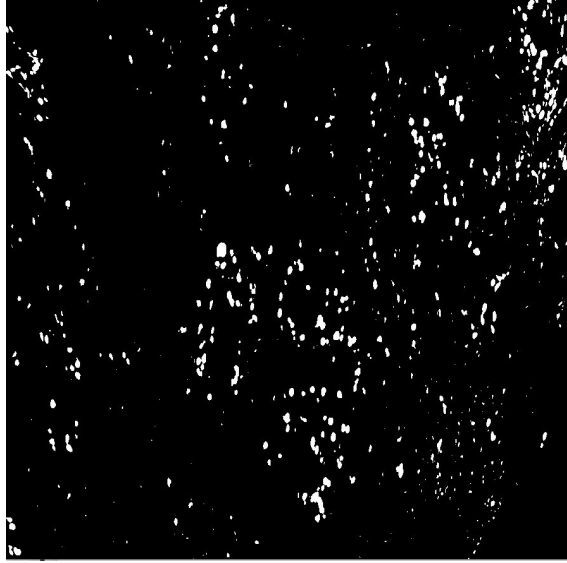


FIGURA 3.11: Máscara gris de células proliferantes

En base a la imagen proporcionada, se observa que las secciones en blanco corresponden a las células proliferantes. Estas están distribuidas de manera dispersa en toda la muestra y presentan variabilidad en su tamaño. El objetivo principal de un patólogo es seleccionar un ROI y llevar a cabo el conteo de las células marcadas (en este caso, las blancas), en comparación con los demás elementos presentes dentro del área.

De esta manera, se logró aislar el primer elemento de la muestra de biopsia. Para los siguientes elementos, se procede a trabajar con un umbral dirigido a las células no proliferativas (en este caso, las azules).

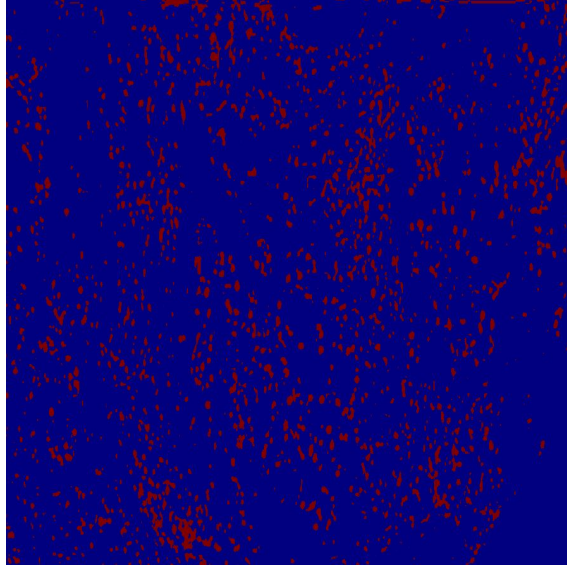


FIGURA 3.12: Imagen HSV de células no proliferantes

Se aplica un umbral a la imagen en escala de grises para obtener una imagen binaria. Los píxeles con intensidad mayor que 128 se establecen en 255 (blanco), mientras que los píxeles con intensidad menor o igual a 128 se establecen en 0 (negro). Esto crea una imagen binaria donde los objetos de interés (células proliferantes) están resaltados.

Así se obtienen los dos elementos clave para calcular el índice de proliferación. Ahora deben ser procesados mediante la segmentación en imágenes binarias para identificar cada elemento.

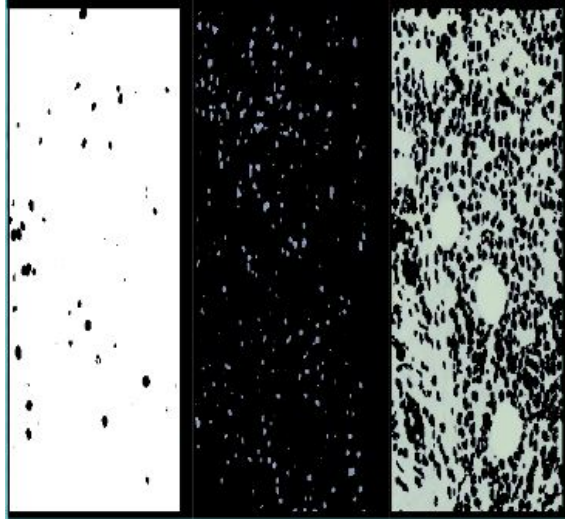


FIGURA 3.13: Imagen binaria de muestra biopsia

En la figura se muestra el resultado de aplicar una técnica de umbralización para segmentar tres componentes dentro del ROI seleccionada. El lado izquierdo contiene las células proliferantes marcadas, que son las que presentan un potencial cancerígeno. El centro muestra las células no proliferantes, que son las que no tienen dicha característica. El lado derecho representa el tejido celular, que corresponde al espacio donde no hay células. Este componente es importante de considerar, ya que puede afectar el análisis de la densidad y la distribución celular.

3.3.3 Detección de bordes

La detección de bordes es una técnica que permite identificar los cambios de intensidad en una imagen, lo que se corresponde con las fronteras de los objetos. Los bordes son importantes para el análisis de imágenes, ya que

permiten segmentar la imagen en regiones de interés, extraer características y reconocer formas.

El proceso de detección de bordes para este caso consiste en aplicar una umbralización para transformar la imagen a escala de grises y luego hallar los contornos en la imagen binaria resultante.

A partir de los elementos resaltados en la imagen binaria, se realiza el conteo de los contornos, utilizando la función 'findContours()' de la biblioteca OpenCV, que detecta los píxeles que forman los bordes.

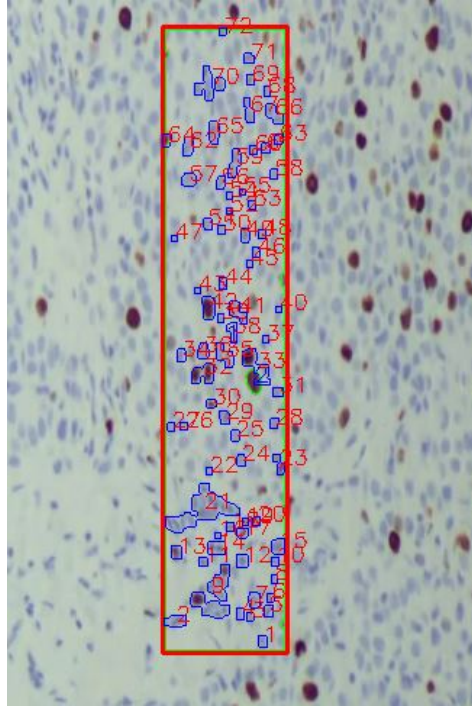
Primera versión

FIGURA 3.14: Detección de contornos primera versión.

En la figura se muestra una primera versión del algoritmo cuya principal función era seleccionar y contar los bordes de las células proliferantes (negras).

Se aprecia que existe un error en la selección de contornos, debido a que el umbral establecido para las células marcadas por la KI67 no es el adecuado. Este caso se analiza en la sección de desafíos (3.2) y corresponde a un falso positivo.

Existe la presencia de células demasiado pequeñas que no son contabi-

lizadas. En este caso, también se requiere ajustar el parámetro de selección para considerar un área mínima que permita la detección e inclusión de dichas células en el análisis.

Segunda versión

Estableciendo un área mínima de detección y ajustando el umbral tanto para las células proliferantes como para las no proliferantes, se implementa una segunda versión del algoritmo de detección. En esta versión, los elementos se separan en capas distintas, cada una con su respectivo conteo.

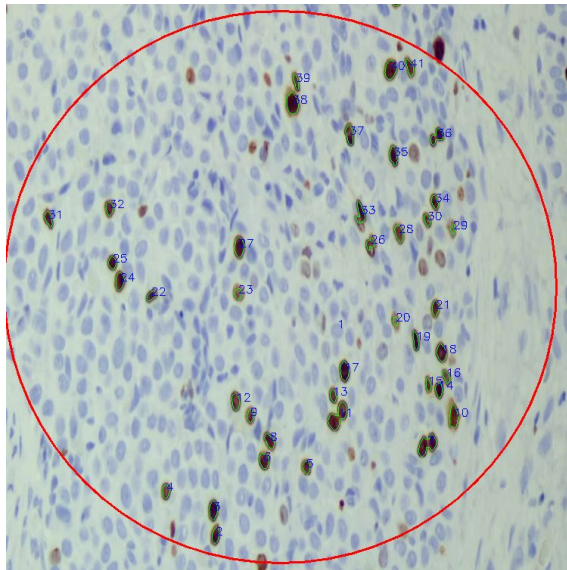


FIGURA 3.15: Conteo células proliferantes versión 2

Para el mismo caso es con las células no proliferantes:

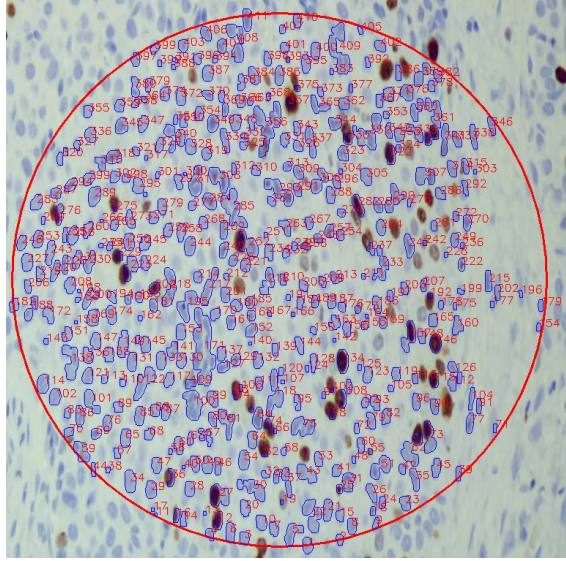


FIGURA 3.16: Conteo células no proliferantes versión 2

3.3.4 Algoritmo de Canny

En la sección de hipótesis, se propuso una alternativa que consiste en la implementación de las técnicas específicas del algoritmo de Canny para llevar a cabo la detección de bordes en un ROI. Esta propuesta se fundamenta en la premisa de utilizar el enfoque preciso y sofisticado del algoritmo de Canny para identificar de manera óptima los límites y contornos dentro de la región de interés definida.

Canny analiza la continuidad de los cambios de intensidad para identificar los bordes coherentes. Para ello, utiliza un proceso de seguimiento de bordes que conecta los puntos de borde detectados. este proceso comienza con un punto de borde de alto contraste y se va expandiendo en ambas direcciones agregando puntos de bordes que van cumpliendo con un criterio menos

exigente.

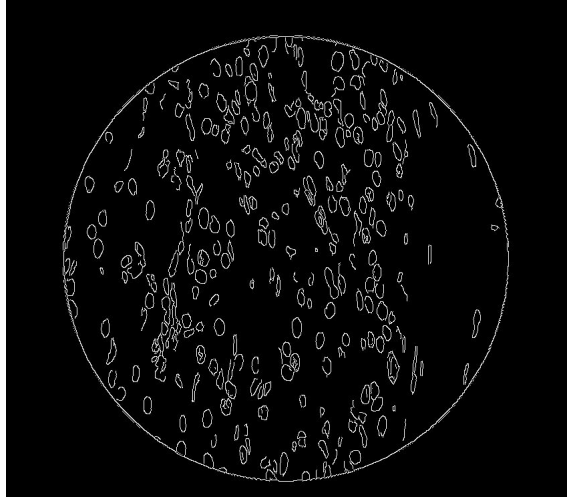


FIGURA 3.17: Fases del algoritmo de Canny en muestra Biopsia

Algoritmo para la detección de bordes

El algoritmo de Canny se basa generalmente en tres pasos:

- Obtención del gradiente: Se calcula la magnitud y orientación del vector gradiente de cambio de intensidad en cada píxel. Se realiza la aplicación de un filtro gaussiano a la imagen original con el objetivo de suavizar la imagen y tratar de eliminar el posible ruido existente.
- Supresión de no máximos: Se logra el adelgazamiento del ancho de los bordes obtenidos con el gradiente, hasta lograr bordes de un píxel de ancho. Se consideran cuatro direcciones identificadas por las orientaciones 0° , 45° , 90° o 135° respecto al eje horizontal. Para cada píxel se

encuentra la dirección que mejor se aproxime a la dirección del ángulo de gradiente.

- Umbralización por histéresis: Se aplica una función de histéresis basada en dos umbrales, reduciendo la posibilidad de aparición de contornos falsos. Consiste en tomar la imagen obtenida del paso anterior y tomar dos umbrales, el primero más pequeño que el segundo. Para cada punto de la imagen se debe localizar el siguiente punto de borde no explorado que sea mayor al segundo umbral. A partir de dicho punto, seguir las cadenas de máximos locales conectados, siempre que sean mayores al primer umbral. Así se marcan todos los puntos explorados y se almacena la lista de todos los puntos en el contorno conectado.

Se pueden apreciar los pasos para una detección de bordes Canny en una imagen de muestra de biopsia dentro de una región de interés para células proliferantes:

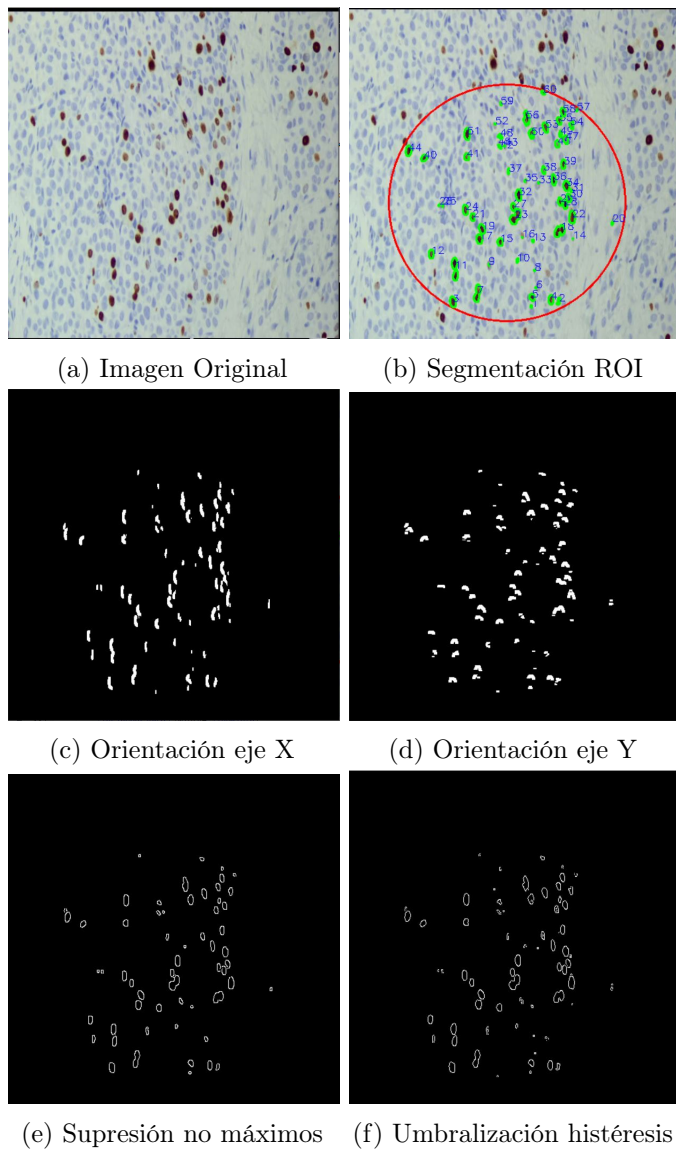


FIGURA 3.18: Fases Algoritmo Canny

La representación gráfica exhibe el proceso de identificación de los límites celulares en un ROI. No obstante, se presenta una eventualidad en la cual se logra identificar inadvertidamente algunos elementos no deseados.

Es relevante señalar que el algoritmo de Canny, cuando se implementa de manera independiente, presenta diversas desventajas significativas, especialmente en lo que respecta a su sensibilidad al ruido y a la intensidad de las imágenes. Esta susceptibilidad puede generar la detección potencial de células cancerígenas falsas positivas.

3.4 Implementación

Implementación de algoritmos

En el marco de la problemática planteada, se busca la capacidad de segmentar los elementos de interés en capas individuales. En este escenario, se observa que el algoritmo de detección de bordes de Canny cumple con dicha tarea de manera parcial.

Por lo discutido en la sección de discusión (2), se propone la implementación de los métodos de segmentación discutidos con el objetivo de integrar de manera sinérgica tecnologías complementarias en el desarrollo de un algoritmo destinado principalmente al conteo diferenciado de células basado en sus características cromáticas.

En el ámbito de conteo células cancerígenas en imágenes de biopsias de microscopio, la convergencia de tres tecnologías avanzadas —thresholding, segmentación de colores y detección de bordes con el operador Canny— emerge como una estrategia integral y poderosa. Esta combinación sinérgica no sólo aprovecha las fortalezas de cada técnica individual, sino que también aborda desafíos específicos asociados con la variabilidad en colores celulares.

Para la función de contar células cancerígenas y discriminar entre los distintos elementos cromáticos, se comentan algunos puntos importantes al momento de implementar las tecnologías:

- Diferenciación de colores mediante thresholding: Al establecer umbrales específicos, permite la diferenciación efectiva entre colores celulares. Esta primera etapa de la estrategia asegura una selección precisa de regiones de interés basadas en las características cromáticas de las células cancerígenas.

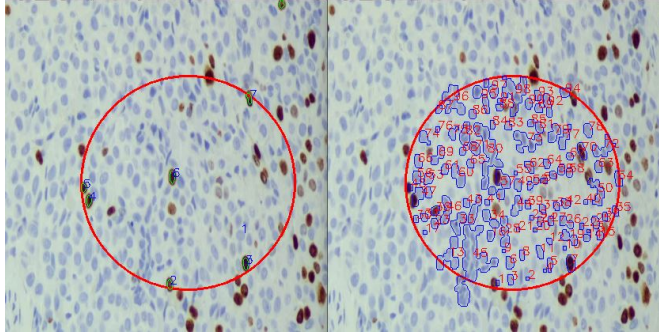
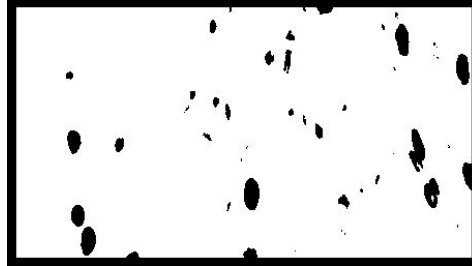
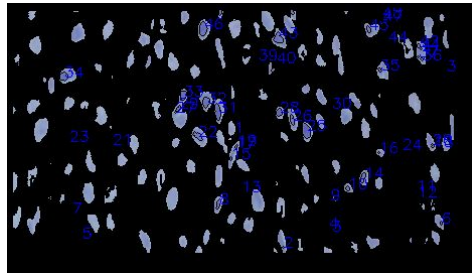


FIGURA 3.19: Detección de colores mediante tresholding

- Segmentación de colores para identificación específica: La segmentación de colores añade una capa adicional de precisión al identificar y delimitar áreas de interés basadas en los colores particulares asociados con células cancerígenas. Este enfoque refinado mejora la especificidad de la detección.



(a) Células Proliferantes



(b) Células NO proliferantes

FIGURA 3.20: Segmentación cromática

- Detección de bordes con el operador de Canny para detalles finos: La detección de bordes mediante el operador de Canny agrega una dimensión extra al resaltar los detalles finos y las transiciones de intensidad. Esto resulta especialmente beneficioso para capturar bordes celulares y proporcionar información detallada sobre las estructuras.

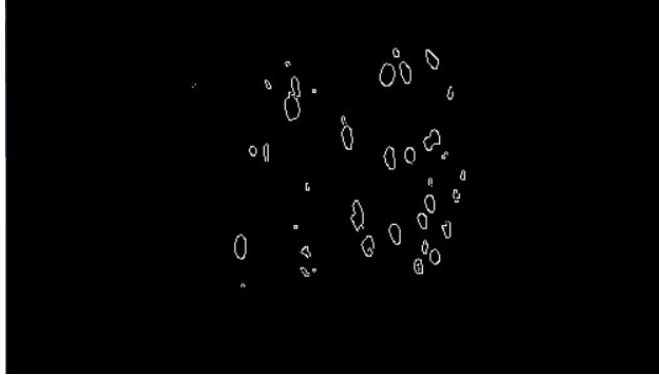


FIGURA 3.21: Detección células proliferantes con operador Canny

- Reducción de falsos positivos y negativos: La combinación de thresholding, segmentación de colores y detección de bordes con Canny permite una reducción significativa de falsos positivos y negativos.

Esto fue realizado por medio de la determinación de un ROI para contar células dentro de los límites establecidos del operador y luego definir un rango de color para la identificación de los píxeles negros en la imagen de muestra. Con la función de la biblioteca OpenCV `"cv2.inRange()"` se crea una máscara (en este caso una máscara negra) para aquellos píxeles que entren dentro del rango.

Como se observa en la figura 3.22, se crea una máscara circular blanca del mismo tamaño que la imagen original, con el centro y el radio coincidiendo con el círculo rojo de referencia. Esta máscara circular se utiliza para restringir la atención sólo en las regiones dentro del círculo.

Se utiliza la función `"cv2.bitwise-and"` para aplicar la máscara circular a la máscara de píxeles negros, generando así una imagen en la que sólo

se conservan los píxeles negros dentro del círculo. Luego, el algoritmo de Canny se aplica a esta imagen enmascarada para resaltar los bordes. Posteriormente se encuentran los contornos en la imagen resultante de aplicar Canny y se verifica si el centro de cada contorno está dentro del círculo de referencia. Esto se hace calculando la distancia entre el centro del contorno y el centro del círculo: si la distancia del contorno es menor a la del centro del círculo, se considera como una célula negra válida.

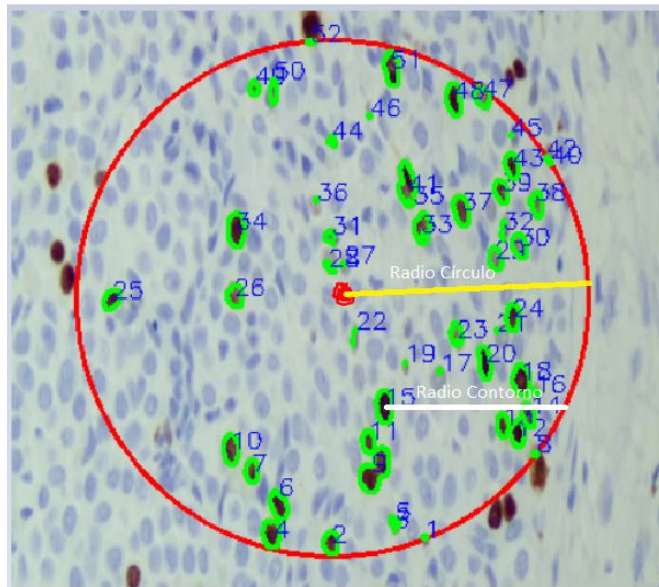


FIGURA 3.22

Los contornos encontrados dentro del círculo se dibujan sobre la imagen original en verde y el número de cada contorno se muestra cerca de su centro. Además, la imagen blanca se utiliza para visualizar sólo los

contornos dentro del círculo. Finalmente, el número total de células negras dentro del círculo es obtenido. Esto proporciona una medida cuantitativa del conteo de células.

Esta metodología garantiza que sólo se cuenten y visualicen las células negras dentro de esta región, contribuyendo a una detección más precisa y específica.

En la figura 3.23 se muestra el resultado de la implementación de los algoritmos de tresholding, segmentación cromática y posterior uso del algoritmo Canny para la obtención de una imagen con las células proliferantes marcadas.

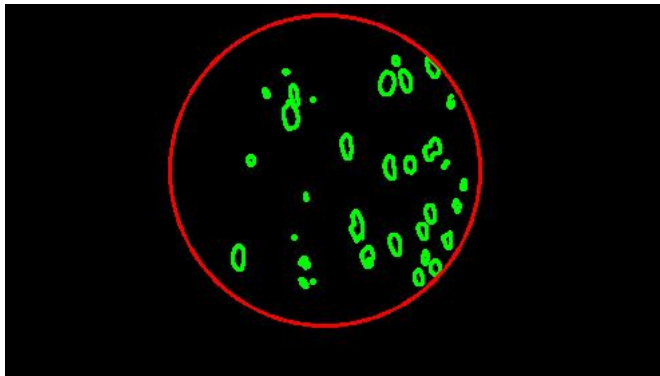


FIGURA 3.23: Implementación de algoritmo sobre una muestra de biopsia

Es importante destacar que, en numerosas instancias, la sinergia de enfoques puede conllevar a resultados óptimos. Se sugiere la ejecución de una segmentación cromática inicial para identificar las regiones de interés, seguido de la aplicación de técnicas de detección de bordes mediante un operador Canny, con el propósito de refinar de manera más precisa los elementos de interés de una muestra de biopsia previamente identificadas

o capturar detalles adicionales de relevancia. Esta estrategia multimodal, integrando la capacidad discriminativa de la segmentación cromática con la sensibilidad a contornos de la detección de bordes, se revela como una metodología robusta para abordar la complejidad de las imágenes biomédicas y maximizar la precisión en la identificación de características específicas.

3.5 Resultados y validación

Clasificación de tamaños de células

Derivado de la fase de implementación, surge la necesidad de identificar situaciones en las cuales se produzca un solapamiento entre células. Este escenario se presenta cuando se visualizan dos células muy cercanas, pero con la particularidad de que el algoritmo implementado las interprete como una entidad única. Es crucial abordar y corregir este tipo de casos para garantizar una segmentación celular precisa y evitar posibles distorsiones en el análisis cuantitativo.

Se requiere una metodología que pueda diferenciar y categorizar cada tamaño de célula de interés, con el propósito de segmentarlas en subconjuntos según sus dimensiones y abordar situaciones de solapamiento. Es importante destacar que, dado el carácter diverso de las imágenes y las células, estas presentarán variaciones significativas tanto en tamaño como en forma, lo cual descarta la viabilidad de implementar algoritmos basados en la clasificación por similitud de formas.

La aplicación de tecnologías fundamentadas en el aprendizaje automático (Deep Learning o Machine Learning) también se enfrenta a desafíos en este contexto, como se expone detalladamente en la sección correspondiente del estado del arte (Sección 2). La variabilidad inherente en los tamaños y

formas de las células hace que métodos tradicionales y enfoques convencionales basados en clasificación por forma sean insuficientes para abordar la complejidad de las imágenes y sus contenidos celulares.

Una estrategia integral para enfrentar esta problemática consiste en realizar una clasificación de células basada en su área promedio. Esta clasificación permite distinguir entre células pequeñas, medianas y grandes, siendo estas últimas las que potencialmente presentan solapamiento. Es crucial destacar que, al generalizar para áreas de células grandes, se está hablando de una simplificación del fenómeno, siendo no obstante, aplicable en la mayoría de los casos, esto discutido con la contraparte en un intento de encontrar un método óptimo. Este enfoque proporciona una base para abordar el análisis de imágenes celulares, facilitando la identificación y cuantificación de fenómenos biológicos relevantes.

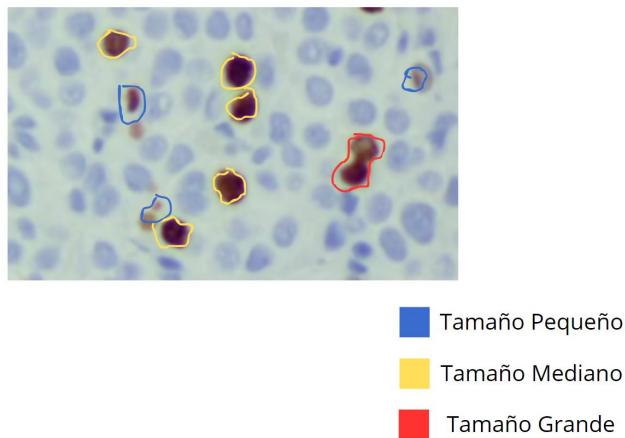


FIGURA 3.24: Ejemplo clasificación de tamaños

En la figura 3.24, se presentan los tamaños promedio que serán considera-

dos para iniciar la clasificación de células en función de sus áreas respectivas. Para abordar este análisis, se ha tomado la decisión de utilizar unidades de píxeles por milímetro (píxeles/mm), dada la naturaleza de las imágenes y considerando la resolución proporcionada en cada una de ellas.

Esta elección se fundamenta en la necesidad de normalizar las mediciones, facilitando así la comparación y clasificación de las células en un contexto más preciso y objetivo. La adopción de la unidad de píxeles/mm permite una representación más consistente y escalable de las dimensiones celulares, brindando mayor robustez a la clasificación basada en áreas específicas. Es necesario señalar que las unidades utilizadas están basadas en las imágenes de prueba entregadas por FALP.

Resultados clasificación

Es así, como definiremos los estándares para clasificar células por tamaño:

- Las células pequeñas están definidas para todas aquellas cuya área sea menor a $50[\text{píxeles}/\text{mm}^2]$
- Las células medianas están definidas para todas aquellas cuya área sea entre $50[\text{píxeles}/\text{mm}^2]$ y $100[\text{píxeles}/\text{mm}^2]$
- Las células grandes están definidas para todas aquellas cuya área sea superior a los $100[\text{píxeles}/\text{mm}^2]$

En este escenario, la elección de los intervalos específicos se basa en su eficacia al clasificar células por tamaño. Estos rangos demostraron ser los óptimos durante el proceso de clasificación. En situaciones donde algunas células pequeñas no eran detectadas, se descartaron para garantizar una detección más precisa. Por otro lado, se observó que en algunos casos, células

grandes y solapadas eran clasificadas de manera similar. Es importante señalar que estas situaciones son excepcionales y poco frecuentes. En caso de que se presenten, el doctor patólogo realizará una verificación manual para asegurar la precisión del resultado.

Se establecieron umbrales de área para la categorización del tamaño de las células, delineando distintas clases según su magnitud. Se procedió a iterar sobre las áreas de las células identificadas en la imagen, cuantificando la cantidad que se encuentra dentro de cada rango predefinido.

Ejemplo de clasificación

Se muestra un ejemplo de aplicación a una muestra de biopsia. En este contexto, se delimita un ROI específico en la muestra seleccionada. A continuación, se emplea un algoritmo de detección de bordes para identificar y segmentar las células cancerígenas presentes en la región delimitada. Se implementa un proceso de filtrado basado en umbral para descartar aquellas células que no cumplen con los criterios definidos, evitando así posibles casos de falsos positivos.

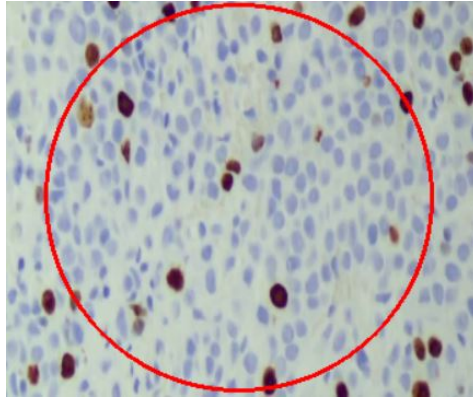


FIGURA 3.25: Ejemplo de primera iteración

Se procede a la obtención de los bordes detectados, los cuales son sometidos a un proceso de clasificación mediante el cálculo del área promedio de las células identificadas. La clasificación se realiza considerando la variación en las áreas celulares, proporcionando así una perspectiva más detallada de la distribución de tamaños dentro de la muestra analizada.

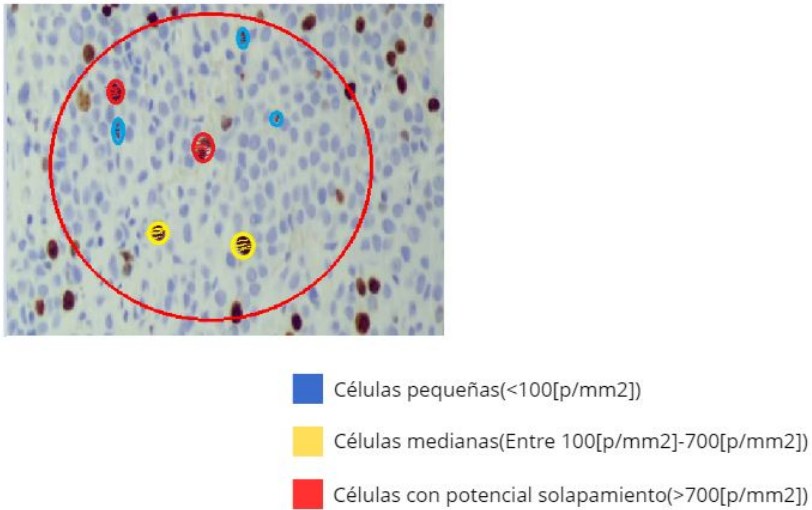


FIGURA 3.26: Clasificación de tamaños

Es así como se obtienen para este caso:

- Cantidad de células pequeñas: 3
- Cantidad de células medianas: 2
- Cantidad de potencial solapamiento: 2

Segundo Desarrollo : Cálculo del índice de proliferación

Para el cálculo del índice de proliferación, se deberá tener en consideración la razón entre las células proliferantes negras con el total de células en la muestra:

$$\text{Índice de prol.} = \frac{\text{C. proliferantes}}{\text{C. no proliferantes} + \text{C. proliferantes}} * 100$$

A continuación, se presentan las comparaciones efectuadas en cinco muestras de biopsia, llevadas a cabo tanto por el patólogo como por el algoritmo, con el objetivo de verificar el alcance y la coherencia en el conteo celular. Es relevante destacar que la evaluación comparativa se realiza con el propósito de garantizar la fiabilidad y precisión del algoritmo en su capacidad para identificar y cuantificar células en el contexto de muestras de biopsia. Este enfoque busca fortalecer la confianza en las capacidades diagnósticas del algoritmo, respaldando así su utilidad en aplicaciones clínicas y patológicas.

Muestra 1

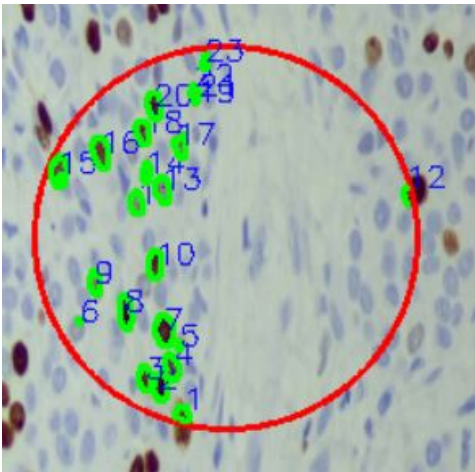


FIGURA 3.27: Muestra 1

Clasificación Célula	Diagnóstico Médico	Algoritmo Detección de Bordes
Pequeña	14	14
Mediana	8	10
Posible Solapamiento	1	1
Índice de Proliferación	20%	19.8%

Tabla 3.1: Comparación Muestra 1

Muestra 2

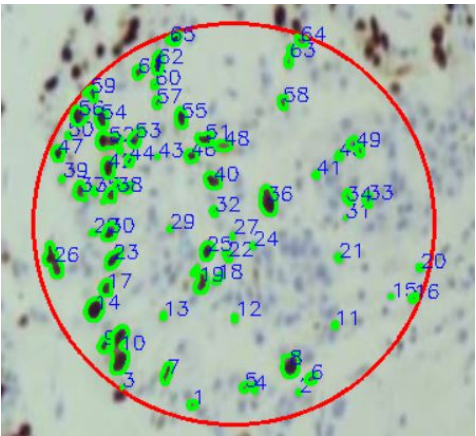


FIGURA 3.28: Muestra 2

Clasificación Célula	Diagnóstico Médico	Algoritmo Detección de Bordes
Pequeña	No específica	36
Mediana	15	14
Posible Solapamiento	15	15
Índice de Proliferación	30%	24%

Tabla 3.2: Comparación Muestra 2

Muestra 3

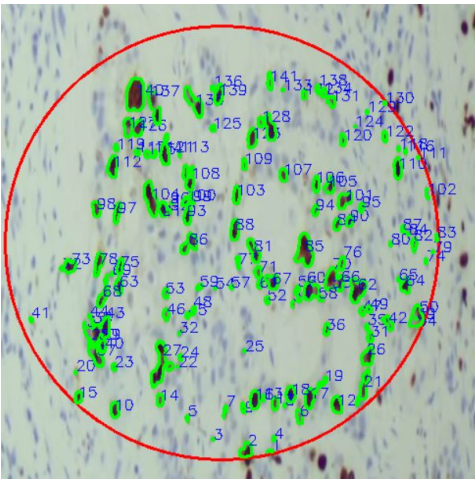


FIGURA 3.29: Muestra 3

Clasificación Célula	Diagnóstico Médico	Algoritmo Detección de Bordes
Pequeña	80	98
Mediana	20	19
Posible Solapamiento	20	21
Índice de Proliferación	30%	22.5%

Tabla 3.3: Comparación Muestra 3

Muestra 4

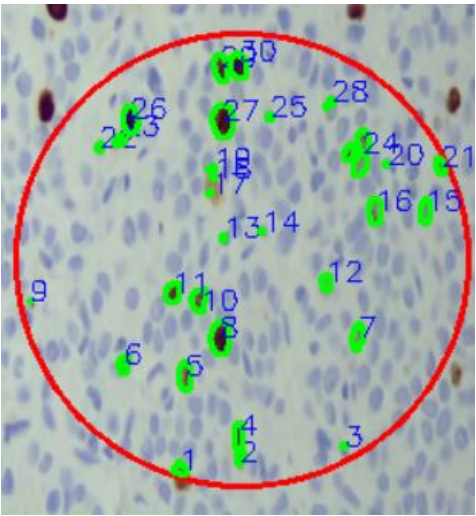


FIGURA 3.30: Muestra 4

Clasificación Célula	Diagnóstico Médico	Algoritmo Detección de Bordes
Pequeña	20	20
Mediana	10	7
Posible Solapamiento	2	2
Índice de Proliferación	20%	14.2%

Tabla 3.4: Comparación Muestra 4

Muestra 5

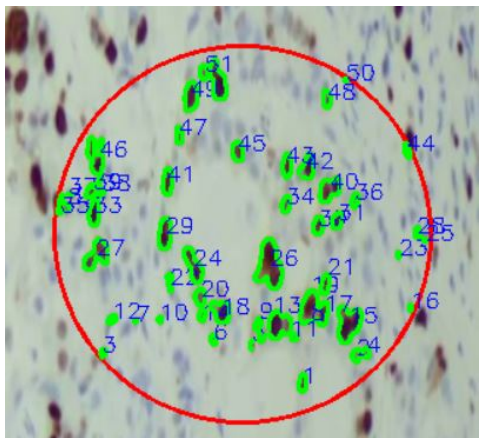


FIGURA 3.31: Muestra 5

Clasificación Célula	Diagnóstico Médico	Algoritmo Detección de Bordes
Pequeña	no específica	36
Mediana	12	8
Posible Solapamiento	10	7
Índice de Proliferación	25%	22.4%

Tabla 3.5: Comparación Muestra 5

Los resultados obtenidos de las muestras corroboran la eficacia del algoritmo de detección al identificar la mayoría de las células presentes en una muestra. Además, el algoritmo demuestra su capacidad al estimar

con precisión la cantidad de células no especificadas en ciertas muestras, como se ilustra en las Figuras 3.28 y 3.31.

Además, se entrega un resultado del índice de proliferación más específico que los seleccionados manualmente por los doctores, ya que ellos no cuentan con un método de estandarización para dichos cálculos y son realizados por medio de aproximaciones y conteos manuales. En contraste, el algoritmo implementado garantiza una mayor especificidad en el conteo de células y el cálculo de la proliferación celular. Al prescindir de métodos subjetivos, el algoritmo se basa en criterios definidos y procesamiento automatizado, eliminando la posibilidad de sesgos humanos y proporcionando una solución más consistente y confiable.

Tercer Desarrollo: detección de solapamientos

En concordancia con los planteamientos efectuados en la iteración previa, es posible respaldar las labores del doctor llevando a cabo una evaluación cuantitativa de la cantidad de células en casos de solapamiento, ya que actualmente no cuentan con alguna herramienta que sea capaz de identificar dichos casos.

A través de las evaluaciones efectuadas mediante el algoritmo de detección, se establecen las métricas para la selección óptima de los parámetros para los distintos casos:

- Solapamiento entre 1 a 4 células: De 100 a 150 [píxeles/ mm^2]
- Solapamiento entre 5 a 10 células: De 150 a 200 [píxeles/ mm^2]
- Solapamiento entre más de 10 células: A partir de más de 200 [píxeles/ mm^2]

Es así como se obtienen los resultados de las muestras bajo el criterio del doctor, que sólo considera el caso de los solapamientos de célula singular:

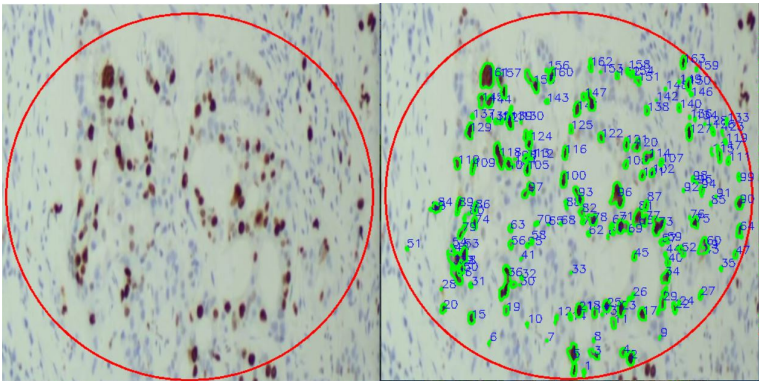


FIGURA 3.32: Muestra de solapamiento de células 1

Muestra 1:

Casos de solapamiento	26
Entre 2-4 células	16
Entre 5-10 células	5
Más de 10 células	5

Tabla 3.6: Detección de solapamiento

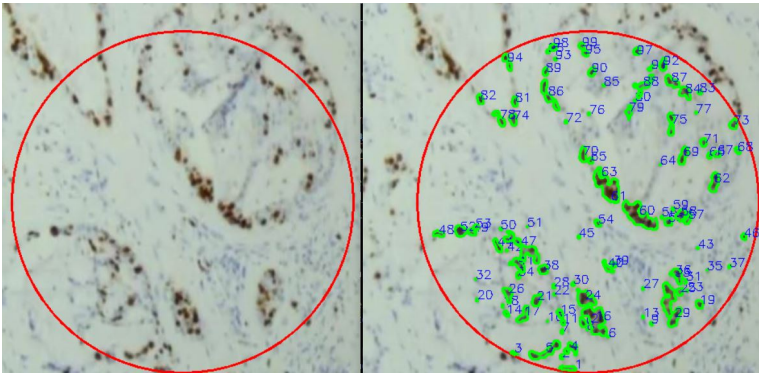


FIGURA 3.33: Muestra de solapamiento de células 2

Muestra 2:

Casos de solapamiento	20
Entre 2-4 células	9
Entre 5-10 células	6
Más de 10 células	5

Tabla 3.7: Detección de solapamiento

Muestra 3:

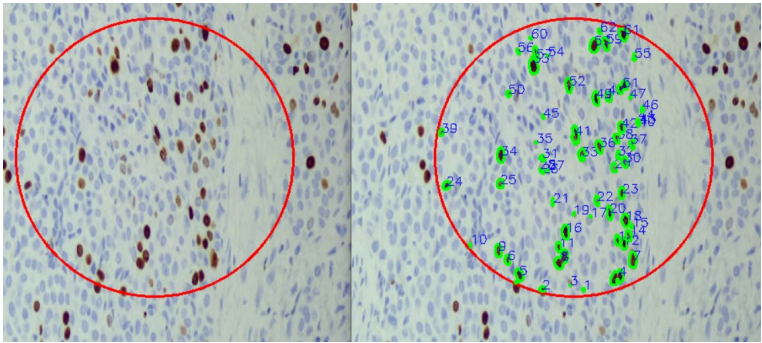


FIGURA 3.34: Muestra de solapamiento de células 3

Casos de solapamiento	7
Entre 2-4 células	7
Entre 5-10 células	0
Más de 10 células	0

Tabla 3.8: Detección de solapamiento

Muestra 4:

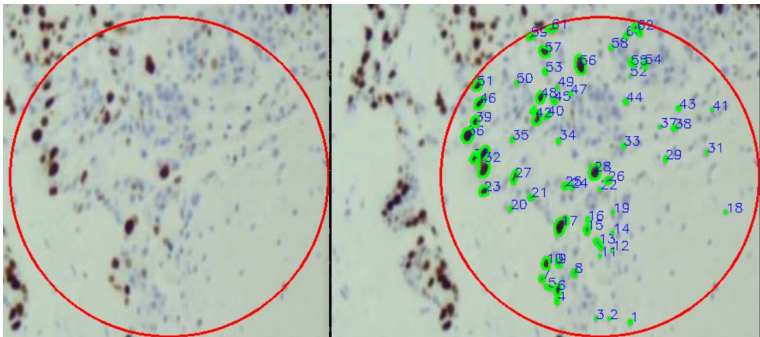


FIGURA 3.35: Muestra de solapamiento de células 4

Casos de solapamiento	9
Entre 2-4 células	4
Entre 5-10 células	2
Más de 10 células	3

Tabla 3.9: Detección de solapamiento

Conforme a lo desarrollado, se procede a verificar si los casos de solapamiento inciden en la precisión del cálculo del índice de proliferación. Este proceso implica la comparación de las muestras con el índice de proliferación y la evaluación de su error medio. Este análisis se lleva a cabo con el propósito de determinar la posible influencia de las situaciones de solapamiento en la exactitud de la medición del índice de proliferación.

Con base en los resultados observados en la tabla 3.10, se aprecia que el índice de proliferación en cada muestra exhibe una variación moderada, manteniéndose dentro de una diferencia aproximada de cinco unidades porcentuales, tal como fue estimado por los profesionales médicos.

Muestra	Ind. Prol. (Sola- pamiento)	Ind. Prol. (No Sola- pamiento)	Error Medio
Muestra 1	21.8%	20%	5.93
Muestra 2	23.8%	25%	12.93
Muestra 3	17.4%	22.1%	5.34
Muestra 4	36.6%	39.8%	7.09

Tabla 3.10: Comparación muestras de solapamiento

Esto da a entender que la evaluación de la cantidad de células en un caso de solapamiento no genera una variación sustancial en el recuento para el patólogo. Por lo tanto, les resulta conveniente omitir este paso y abordar los casos de solapamiento de manera general, simplificando así el proceso de análisis.

Una vez adoptada esta perspectiva, la identificación de células en situaciones de superposición no aporta significativamente al cálculo del índice de proliferación celular. En este sentido, se decide que no se incluirá en los criterios establecidos por el profesional médico para la determinación porcentual de dicho índice.

Conclusiones

4.1 Conclusiones

En el contexto de esta investigación, se ha concebido un algoritmo que fusiona tres tecnologías de visión por computador: thresholding, segmentación cromática (ver sección 3.3.2), y la aplicación del algoritmo de detección Canny (ver sección 3.3.4). Este algoritmo se ha diseñado con el propósito fundamental de proporcionar una herramienta altamente especializada a los doctores patólogos, con el objetivo de estandarizar un proceso que, hasta ahora, carecía de un método sistemático para calcular el índice de proliferación y clasificar células en muestras de biopsias.

La esencia de esta solución radica en la convergencia de estas técnicas. La aplicación de thresholding permite establecer límites claros en la imagen, facilitando la identificación precisa de regiones de interés. La segmentación cromática, amplía la capacidad del algoritmo al considerar características

específicas de color, enriqueciendo la precisión en la diferenciación celular.

Adicionalmente, el método propuesto incorpora el algoritmo de detección Canny, lo que agrega una capa adicional de precisión en la identificación de bordes y contornos celulares. Este enfoque integral no sólo mejora la robustez del algoritmo, sino que también eleva su desempeño a niveles comparables con la experiencia de un patólogo en situaciones del mundo real.

La meta última es proporcionar a los doctores patólogos una herramienta que no sólo mejore la eficiencia en el cálculo del índice de proliferación, sino que también proporcione una clasificación celular precisa y confiable en muestras de biopsias, contribuyendo así a la excelencia en el diagnóstico médico y a la investigación avanzada en patología.

Es relevante destacar que, aunque los resultados del índice de proliferación generados por el médico y el algoritmo pueden mostrar variaciones, estas no muestran necesariamente alteraciones sustanciales. Según las observaciones del patólogo, el índice proliferativo constituye simplemente una estimación aproximada de la proporción de células cancerígenas en estados activos de proliferación. Por ende, diferencias de hasta cinco unidades porcentuales no indicarían un cambio relevante en la realidad clínica.

4.2 Trabajo a futuro

Como se comentó en la sección de conclusiones, el desarrollo del algoritmo de esta memoria dejó bastantes casos hipotéticos y bastantes posibilidades de escalamiento de tecnologías. A continuación se presentan los puntos más importantes para tener en consideración en un trabajo a futuro:

- **Validación de criterios:** Se requiere llevar a cabo una validación de los criterios definidos en la sección de *Resultados y Validación*(3.5),

ampliando la muestra de casos y la participación de diversos profesionales médicos. Actualmente, el trabajo se ha desarrollado en situaciones específicas y está limitado por las imágenes de muestra de biopsias proporcionadas por el equipo de patología de los laboratorios de FALP. En este sentido, la colaboración con expertos en patología se presenta como un paso para evaluar y ajustar los parámetros del algoritmo de acuerdo con las características particulares de diversas muestras y tipos de tejidos. Este enfoque garantizará una mayor robustez y aplicabilidad del algoritmo en diferentes contextos clínicos.

- **Más pruebas:** Asimismo, se destaca la necesidad de realizar un mayor número de pruebas en un conjunto más amplio de imágenes de biopsias. La diversidad de las muestras contribuirá a la robustez del algoritmo, permitiendo su adaptabilidad a diferentes contextos y condiciones. La inclusión de variabilidad en las características histológicas de las muestras enriquecerá la capacidad del algoritmo para generalizar su desempeño en entornos clínicos diversos.

Los parámetros actualmente están siendo trabajados desde los casos cancerígenos de mama, por lo que se debería considerar a futuro el trabajo con otros tipos de cáncer y verificar si los criterios aplicarían de forma similar o, en caso contrario, modificarlos y ajustarlos según la necesidad de los doctores patólogos.

- **Prueba de casos bajo metodologías de Machine Learning:** Tal como se especificó durante el desarrollo de esta memoria, particularmente en la sección de *estado del arte* (2), debido a que la contraparte FALP no contaba con una cantidad mínima de datos, se dió por descartada la posibilidad de implementar una metodología de Deep Learning

para el desarrollo de un algoritmo de detección de bordes automático. En su lugar se optó por un enfoque basado en técnicas de vision por computadora para el procesamiento de imágenes individuales.

Como perspectiva futura, se plantea la necesidad de recopilar una cantidad mínima suficiente de datos e imágenes para llevar a cabo una evaluación más robusta. Se propone verificar la viabilidad de la implementación de una metodología de aprendizaje automático a través de los casos de prueba en una escala considerablemente ampliada. Este enfoque permitirá evaluar la eficacia del modelo en condiciones más diversas y representativas.

Este modelo podría incorporar características específicas y criterios de diagnóstico definidos por los doctores para mejorar la precisión y la confiabilidad en la identificación de células con potencial maligno. Este enfoque avanzado en el campo de la detección temprana del cáncer podría tener implicaciones significativas en la mejora de los métodos diagnósticos y, dado un umbral de índice de proliferación, se podría señalar la gravedad de una enfermedad y priorizar los casos para su tratamiento.

Bibliografía

- [1] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. Segment anything, 2023.
- [2] Piotr Dollár Kaiming He, Georgia Gkioxari and Ross Girshick. Mask r-cnn. <https://arxiv.org/abs/1703.06870>, 2018.
- [3] David Marín Soto. Segmentación de células mediante técnicas de procesamiento digital de imágenes para el rastreo de células cancerosas. <https://repositoriotec.tec.ac.cr/handle/2238/10383>, 2018.
- [4] Marco Alzate Marcela Mejia, Astrid Rubiano. Segmentación de imágenes de células cervicales y evaluación de características para detección de lesiones neoplásicas. <https://www.scielo.cl/pdf/ingeniare/v24n2/art10.pdf>, 2015.
- [5] Sitio web principal de imagej. <https://imagej.nih.gov/ij/download.html>, 2018.

- [6] Ihor Smal Erik Meijering, Oleh Dzyubachyk. Methods for cell and particle tracking. <https://www.sciencedirect.com/science/article/abs/pii/B9780123918574000094?via>
- [7] Sitio web principal de qupath. <https://qupath.readthedocs.io/en/0.4/>.
- [8] Patricia Moreno B. Methods for cell and particle tracking. <https://oa.upm.es/51995/>, 2018.
- [9] Andrés E. Gutiérrez Rafael M. Rodríguez César A. Triana, Nayid Jaramillo. Técnicas de umbralización para el procesamiento digital de imágenes de gem-foils, 2016. ISSN 0122-1701. URL <https://www.redalyc.org/articulo.oa?id=84950881009>.
- [10] Algoritmo de watershed. URL <https://unipython.com/segmentacion-imagenes-algoritmo-watershed/>.
- [11] Algoritmo k-means. URL <https://www.themachinelearners.com/k-means/>.
- [12] Algoritmo dbscan. URL <https://datascientest.com/es/machine-learning-clustering-dbscan>.
- [13] Algoritmo de canny. URL https://www.researchgate.net/publication/267240432_Deteccion_de_bordes_mediante_el_algoritmo_de_Canny