



FACULTAD DE INGENIERÍA Y ARQUITECTURA  
ESCUELA PROFESIONAL DE INGENIERÍA ELECTRÓNICA

**DESARROLLO DE UN APLICATIVO DE SOFTWARE  
ORIENTADO A LA CUANTIFICACIÓN DEL MARCADOR KI-67  
BASADO EN TÉCNICAS DE PROCESAMIENTO DIGITAL DE  
IMÁGENES**

**PRESENTADA POR  
DAVID RONALD ATENCIA MONDRAGÓN  
IVAN FERNANDO ORTEGA AMPUERO**

**ASESOR  
GUILLERMO LEOPOLDO KEMPER VASQUEZ**

**TESIS**

**PARA OPTAR EL TÍTULO PROFESIONAL DE INGENIERO ELECTRÓNICO**

**LIMA – PERÚ**

**2018**



**CC BY-NC-SA**

**Reconocimiento – No comercial – Compartir igual**

El autor permite transformar (traducir, adaptar o compilar) a partir de esta obra con fines no comerciales, siempre y cuando se reconozca la autoría y las nuevas creaciones estén bajo una licencia con los mismos términos.

<http://creativecommons.org/licenses/by-nc-sa/4.0/>



**USMP**  
UNIVERSIDAD DE  
SAN MARTÍN DE PORRES

FACULTAD DE  
INGENIERÍA Y ARQUITECTURA

**ESCUELA PROFESIONAL DE INGENIERÍA ELECTRÓNICA**

**DESARROLLO DE UN APLICATIVO DE SOFTWARE  
ORIENTADO A LA CUANTIFICACIÓN DEL MARCADOR KI-67  
BASADO EN TÉCNICAS DE PROCESAMIENTO DIGITAL DE  
IMÁGENES**

**TESIS**

**PARA OPTAR EL TÍTULO PROFESIONAL DE INGENIERO  
ELECTRÓNICO**

**PRESENTADA POR**

**ATENCIA MONDRAGÓN, DAVID RONALD**

**ORTEGA AMPUERO, IVAN FERNANDO**

**ASESOR**

**KEMPER VASQUEZ, GUILLERMO LEOPOLDO**

**LIMA- PERÚ**

**2018**

Dedico este trabajo a mis padres, por el gran apoyo que me brindaron a lo largo de mi desarrollo; por la confianza depositada en mí y los ánimos que me otorgaron, los cuales fueron indispensables para el desarrollo de esta tesis.

David Atencia M.

Dedico este trabajo a los sueños que fueron dominados; a los pensamientos, que no fueron el único objetivo. Dedico este trabajo al Triunfo y al Desastre, porque ambos fueron tratados de la misma manera.

Ivan Ortega A.

Agradezco a los profesores, a nuestro asesor Guillermo Kemper, a los médicos e instituciones médicas que contribuyeron al desarrollo de esta tesis. A la universidad San Martín de Porres por todos estos años de enseñanza.

David Atencia M.

Agradezco a los profesores, a nuestro asesor Guillermo Kemper, a los médicos e instituciones médicas que contribuyeron al desarrollo de esta tesis. A la universidad San Martín de Porres por todos estos años de enseñanza.

Ivan Ortega A.

## ÍNDICE

	Página
<b>RESUMEN</b>	xx
<b>ABSTRACT</b>	xxi
<b>INTRODUCCIÓN</b>	xxii
<b>CAPÍTULO I. ANTECEDENTES Y DESCRIPCIÓN DEL PROBLEMA</b>	
1.1 Comentarios Iniciales	1
1.2 Justificación e Importancia de la Investigación	3
1.3 Definición del problema	4
1.4 Justificación Social	4
1.5 Justificación Ingenieril	5
1.6 Antecedentes	6
1.7 Hipótesis	7
1.8 Objetivos	7
1.9 Solución propuesta y Diagrama de bloques	8
1.10 Comentarios finales	9
<b>CAPÍTULO II. MARCO TEÓRICO Y CONCEPTUAL</b>	
2.1 Comentarios Iniciales	11
2.2 Ciclo Celular	11



2.3	Marcadores tumorales	13
2.4	Inmunohistoquímica	14
2.5	Fundamentos de procesamiento digital de imágenes	15
2.6	Teoría del Pixel	16
2.7	Modelos de color	19
2.8	Transformación de Intensidad	23
2.9	Filtros detectores de gradiente	24
2.10	Transformaciones morfológicas en imágenes	25
2.11	Momentos estadísticos	29
2.12	Histograma	30
2.13	Proyecciones sobre el eje	32
2.14	Mínimos locales o mínimos relativos	32
2.15	Segmentación	32
2.16	Comentarios Finales	34
<b>CAPÍTULO III. DESARROLLO DE LA INVESTIGACIÓN</b>		
3.1	Comentarios Iniciales	35
3.2	Fuentes de información e instrumentos utilizados	35
3.3	Técnica de Recolección de Base de datos	36
3.4	Metodología	37
3.5	Comentarios finales	65
<b>CAPÍTULO IV. RESULTADOS Y VALIDACIÓN</b>		
4.1	Comentarios iniciales	66
4.2	Costo computacional y tiempo de procesamiento	66
4.3	Validación de resultados	66
4.4	Métrica de validación: Índice <i>Kappa</i> de <i>Cohen</i>	67
4.5	Métrica de validación <i>F-score</i>	70
4.6	Informe económico	72

## **CAPÍTULO V. IMPLEMENTACIÓN DEL SOFTWARE**

5.1 Aplicaciones	74
<b>CONCLUSIONES</b>	85
<b>RECOMENDACIONES</b>	87
<b>FUENTES DE INFORMACIÓN</b>	88
<b>ANEXOS</b>	92

## ÍNDICE DE FIGURAS

	Página
Figura 1. Diagrama de bloques de la solución propuesta	10
Figura 2. Etapas de reproducción celular	12
Figura 3. Muestra de tejido lista para observación	15
Figura 4 Ejemplo de visualización de píxeles en una imagen digital aplicando aumento	17
Figura 5. Comparación de distintos niveles de cuantificación	18
Figura 6. Modelo de color escala de grises	19
Figura 7. Modelo de color RGB	20
Figura 8. Modelo de color HSV	22
Figura 9. Ejemplo de plano de color $Cb - Cr$ con luminancia $Y=0.5$	22
Figura 10. Representación de la función de corrección de intensidad variando $\gamma$	24
Figura 11. (a) Mascara de filtros columnas. (b) Mascara de filtro filas	25
Figura 12. Etapas del filtro Sobel. (a) Imagen Original, (b) imagen procesada por filtro Sobel solo en la vertical, (c) imagen procesada solo en la horizontal	25
Figura 13 (a) Imagen Lena original (b) Imagen Lena rotada $45^\circ$ sobre su centro en sentido antihorario	27
Figura 14. Ejemplos de elementos estructurante (a) tipo cruz, (b) tipo aspa	27

Figura 15. Ejemplo de dilatación: a) Imagen binaria original de texto	28
Figura 16. Ejemplo de erosión a) Imagen binaria original, b) Imagen erosionada con elemento estructurante 13x13	29
Figura 17. (a) Imagen en escala de grises (b) Histograma de la imagen	31
Figura 18. Demostración gráfica de vecindad (a) Vecindad 4 y (b) Vecindad 8	33
Figura 19. (a) Imagen blanco y negro antes de identificar regiones. (b) Equivalente de la imagen después de proceso de etiquetado	34
Figura 20. Equipo <i>BenchMark XT</i> de la marca Ventana	35
Figura 21. Imagen de microscopio digital Olympus BX53	36
Figura 22. Esquema del proceso de tinción de láminas	38
Figura 23. Imagen Original	39
Figura 24. Imagen Recortada y acercada	39
Figura 25 (a) Imagen $I(x, y)$ , (b) Componente V de la transformación al modelo de color HSV de $I(x, y)$ , (c) Imagen $IVg(x, y)$	41
Figura 26. (a) Espacio dimensional $ICb_{x, y}$ original (histograma) y (b) Espacio dimensional desplazado $ICb_{2x, y}$ (histograma).	41
Figura 27. (a) Imagen original RGB, (b) Imagen $ICb_{x, y}$ y (c) imagen $ICb_{2x, y}$ resultante de la aplicación de la corrección gamma	42
Figura 28. Imagen $IVg_{x, y}$ a ser sometida al proceso de segmentación de células proliferantes, en este ejemplo, en $p=5$	44
Figura 29. Procesamiento de los $i$ segmentos. Las barras verdes corresponden al valor medio de cada segmento y las barras rojas al valor del umbral elegido en cada segmento (a) Segmento 1 de $IVg_{x, y}$ (b) Segmento 2 de $IVg_{x, y}$ (c) Segmento 3 de $IVg_{x, y}$ (d) Segmento 4 de $IVg_{x, y}$ (e) Segmento 5 de $IVg_{x, y}$	45

Figura 30. Umbralización de células proliferantes. (a) Imagen $IVg(x,y)$ (b) Imagen binaria $Ipro(x,y)$ resultante del proceso de segmentación de células proliferantes Elaboración: los autores	46
Figura 31. Resultados de la segmentación de células proliferantes con umbrales adaptivos locales. (a) Imagen de prueba (b) Sin subdivisión (c) Dos subdivisiones (d) Cuatro subdivisiones (e) Cinco subdivisiones (f) Diez subdivisiones	47
Figura 32. (a) Imagen original. (b) Imagen resultante $ICb3x,y$	48
Figura 33. (a) Imagen original (b) Imagen resultante de aplicar filtro Sobel (c) Imagen final $ISobelx,y$ después de llenado de huecos	52
Figura 34. (a) Imagen original recortada, (b) Imagen $Icelx,y$ y (c) Imagen $Inprox,y$ Elaboración: los autores	54
Figura 35. (a) Muestras con juntura y (b) muestras después de proceso de separación Elaboración: los autores	57
Figura 36. Células etiquetadas incorrectamente debido a la superposición y cercanía de células (a) Células cercanas (b) Células superpuestas (c) Células cercanas (d) Células superpuestas	58
Figura 37. Giros de un objeto para separarlas adecuadamente (a) Primer giro (b) Segundo giro (c) Tercer giro (d) Cuarto giro (e) Quinto giro (f) Sexto giro	59
Figura 38. Proyecciones con respecto al eje de las abscisas para cada giro (a) Proyección del primer giro (b) Proyección del segundo giro (c) Proyección del tercer giro, la flecha roja indica el mínimo global encontrado (d) Proyección del cuarto giro (e) Proyección del quinto giro (f) Proyección del sexto giro	60
Figura 39. Resultados de la separación de células (a) Células juntas (b) Células separadas por el algoritmo propuesto	60
Figura 40. Separación de células (a) Prueba con 2 giros (b) Prueba con 3 giros (c) Prueba con 4 giros (d) Prueba con 6 giros (d) Prueba con 10 giros	61
Figura 41. Visualización de la ventana principal	75

Figura 42. Ubicación de botón ABRIR	75
Figura 43. Vista del cuadro de carga de imágenes	76
Figura 44. Barra de opciones de la imagen cargada	76
Figura 45. (a)Vista del aplicativo con la imagen cargada y (b) vista acercada	78
Figura 46. (a)Vista del botón ANALIZAR	78
Figura 47. Vista del proceso de clasificación y conteo finalizado	80
Figura 48. Vista de la opción de eliminar (a)Conteo actual de células. (b) Selección de células a eliminar (c) Vista actualizada con conteo actualizado	81
Figura 49. Vista de la ventana de guardar	82
Figura 50. Vista de la ventana para generar reporte	83
Figura 51. Vista del reporte generado en PDF	84

## ÍNDICE DE TABLAS

	Página
Tabla 1. Casos nuevos de cáncer en mujeres registrados en el INEN, periodo 2000-2014	5
Tabla 2. Categorización de los Marcadores Tumorales de Cáncer de mama	13
Tabla 3. Tiempo de procesamiento del software	67
Tabla 4. Informe economico	72
Tabla 5. Comparación Software vs. Especialista 1	73
Tabla 6. Comparación Software vs. Especialista 2	73
Tabla 7. Comparación Software vs. Especialista 3	73
Tabla 8. Comparación Software vs. Especialista 4	73
Tabla 9. Comparación Software vs. Especialista 5	73
Tabla 10. Comparación Especialista 1 vs. Especialista 2	73
Tabla 11. Comparación Especialista 2 vs. Especialista 4	73
Tabla 12. Comparación Especialista 1 vs. Especialista 5	73

## GLOSARIO

**Ki-67.** Proteína nuclear no histona, que está asociada al índice de proliferación de una célula.

**INEN:** Instituto Nacional de Enfermedades Neoplásicas se encuentra ubicado en Lima, Perú y es uno de los principales centros a nivel nacional donde se tratan enfermedades cancerígenas.

**HNERM:** El Hospital Nacional Edgardo Rebagliati Martins, antiguo Hospital del Empleado. Es un centro hospitalario público peruano situado en Lima, Perú y administrado por EsSalud.

**ER:** *Estrogen receptor*, es un marcador tumoral mamario utilizado para detección y estimación del avance de cáncer mamario.

**EP:** *Progesterone receptor*, también es un marcador tumoral mamario utilizado para detección y estimación del avance de cáncer mamario. Tanto el ER como PR son marcadores utilizados conjuntamente para poder realizar un análisis de la presencia y evolución de un carcinoma mamario.

**Patología:** Rama de la medicina que estudia los trastornos anatómicos y fisiológicos de los tejidos de órganos enfermos.

**Inmunohistoquímica:** Proceso especial de coloración con agentes cromógenos que se realiza sobre un tejido obtenido mediante biopsia. Se utiliza para determinar la presencia de ciertos antígenos o biomarcadores.

**PDI:** Procesamiento Digital de Imágenes, es el área de estudio de las ciencias matemáticas en el que se utilizan técnicas de programación para procesar imágenes digitales y obtener una solución frente a un problema.

**Umbralización:** Técnica de PDI que se basa en escoger valores digitales u objetos que cumplan ciertas condiciones dentro de un rango establecido.

**Segmentación:** Técnica de PDI que se basa en la separación de una imagen digital en varias partes u objetos, con el objetivo de simplificar su la representación de la imagen y acelerar su análisis.



**Cámara multispectral:** Cámara digital que posee un sensor que le permite percibir otras longitudes de onda distintas del espectro visible del ojo humano.

**Random Forest (RF):** es una técnica de agregación desarrollada por Leo Breiman, que mejora la precisión en la clasificación mediante la incorporación de aleatoriedad en la construcción de cada clasificador individual.

**Red Neuronal:** es una técnica de aprendizaje computacional que trata de emular el comportamiento de las redes neuronales biológicas, utilizado comúnmente para agilizar tareas de clasificación de información.

**Máquinas de soportes vectoriales:** del inglés *Support Vector Machine (SVM)*, son un conjunto de algoritmos de aprendizaje supervisado, utilizado para problemas de clasificación y regresión en las ciencias de la computación.

**Organelos:** en biología celular se les denomina a las diferentes estructuras contenidas dentro del citoplasma de las células, principalmente las eucariotas, que tienen forma determinada.

**Centrosoma:** es una estructura de la célula que no posee membrana y se encarga de la parte nuclear de la célula, en la etapa de proliferación es la que se divide para formar dos células posteriormente.

**Crominancia:** es la señal en la que los sistemas de videos transportan información de color de la imagen.

**Convolución:** es una operación matemática que transforma dos funciones  $f$  y  $g$  en una tercera función que representa la magnitud en la que se superponen dichas funciones de entrada.

**Morfología:** en PDI, es una técnica de procesamiento no lineal, caracterizada por realzar la geometría y forma de un objeto utilizando la base matemática de teoría de conjuntos.

**Micrótopo:** instrumento que permite el corte de objetos o tejidos en láminas muy finas, del orden de los micrómetros, para su observación en microscopios.

## LISTA DE VARIABLES MATEMÁTICAS

- *resolución*: cantidad de valores digitales máximos por cada píxel en una imagen
- *res*: cantidad de bits a la que es cuantificada una imagen
- *gris*: ecuación de conversión de modelo de color RGB a escala de grises
- *R*: componente de color Rojo del modelo de color RGB
- *G*: componente de color Verde del modelo de color RGB
- *B*: componente de color Azul del modelo de color RGB
- *H*: componente de Tono del modelo de color HSV
- *S*: componente de Saturación del modelo de color HSV
- *V*: componente de Valor del modelo de color HSV
- *Y*: componente de Luminancia del modelo de color YCbCr
- *C<sub>b</sub>*: componente de Saturación del modelo de color YCbCr
- *C<sub>r</sub>*: componente de Valor del modelo de color YCbCr
- *x*: valor de un punto en el eje de las abscisas en un plano cartesiano
- *y*: valor de un punto en el eje de las ordenadas en un plano cartesiano
- *e(x, y)*: función resultante de la transformada de intensidad
- *r(x, y)*: función de entrada al que se le aplicada transformada de intensidad
- *gamma(x, y)*: función de corrección gamma
- $\gamma$ : valor constante que se aplica en la ecuación de corrección gamma
- *Z(x, y)*: función de filtro *Sobel*
- *Img(x, y)*: función de imagen a ser tratada
- *h(x, y)*: función de la máscara que se aplica dentro del filtro Sobel
- *Giro(x, y)*: función de rotación de imagen
- *A*: Matriz a ser tratada por proceso morfológico
- *est*: Matriz de elemento estructurante
- $\sigma^2$ : Varianza
- *z<sub>i</sub>*: función de variable aleatoria
- $\bar{Z}^2$ : segundo momento con respecto a la media de la variable aleatoria *z*
- *num*: numero de datos de una muestra
- $\sigma$ : Desviación estándar

- $H(b)$ : función de histograma
- $P(x), P(y)$ : proyecciones con respecto a los ejes
- $m$ : filas de una imagen
- $n$ : columnas de una imagen
- $b$ : amplitud o intensidad de gris en un pixel
- $L$ : número de niveles de gris posibles
- $p$ : ecuación de 8-vecindad para una imagen de coordenadas  $(x, y)$
- $F(x, y)$ : función de imagen de entrada del algoritmo
- $F_R(x, y), F_G(x, y), F_B(x, y)$ : expresión de cada banda de color, rojo, verde y azul respectivamente; del modelo de color RGB de la función de imagen  $F(x, y)$
- $I(x, y)$ : imagen  $F(x, y)$  recortada
- $M$ : número de filas de una imagen
- $N$ : número de columnas de una imagen
- $I_R(x, y), I_G(x, y)$  e  $I_B(x, y)$ : expresión de cada banda de color, rojo, verde y azul respectivamente; del modelo de color RGB de la función de imagen  $I(x, y)$
- $I_V(x, y)$ : imagen  $I(x, y)$  expresada en su componente  $V$  del modelo de color HSV
- $I_{Vg}(x, y)$ : imagen resultante de aplicar la corrección gamma a la imagen  $I_V(x, y)$
- $I_{Cb}(x, y)$ : imagen  $I(x, y)$  expresada en su componente  $C_b$  del modelo de color  $YC_bC_r$
- $I_{Cb2}(x, y)$ : imagen resultada de aplicar corrección gamma a  $I_{Cb}(x, y)$
- $I_i(x, y)$ : segmentos de la imagen  $I(x, y)$  partido en  $i$  partes
- $\bar{x}_i$ : valor medio de cada segmento  $i$  normalizado
- $\sigma_i$ : desviación estándar de cada segmento  $i$
- $u_i$ : umbral de cada segmento  $i$
- $I_{pro}(x, y)$ : imagen resultante que posee solo las células proliferantes
- $I_{Cb3}(x, y)$ : imagen resultada de aplicar un umbral a  $I_{Cb2}(x, y)$
- $M_h$ : matriz de máscara horizontal de *Sobel*
- $M_v$ : matriz de máscara vertical de *Sobel*

- $I'_{Sobel}(x, y)$ : imagen resultante de aplicar el filtro Sobel a  $I_{Cb3}(x, y)$
- $u_s$ : constante para umbral post filtro *Sobel*
- $I_{Sobel}(x, y)$ : imagen resultante de aplicar el umbral  $u_s$  a  $I'_{Sobel}(x, y)$
- $B_h$ : elemento estructurante horizontal
- $B_v$ : elemento estructurante vertical
- $I_{clo}$ : imagen resultante de aplicar la técnica de 'cierre' a  $I_{Sobel}(x, y)$
- $I_{clo}^k$ : imagen resultante de aplicar técnica de llenado de hueco hecho  $k$  veces sobre  $I_{clo}$
- $k_0$ : cantidad de veces finales que se aplicó la técnica de llenado de hueco
- $D_p$ : elemento estructurante para aplicación de técnica de llenado de hueco
- $I_{SobelF}$ : imagen resultante final de aplicar el llenado de hueco sobre la imagen  $I_{Sobel}(x, y)$
- $F_R(x, y)$ : matriz resultante de filtrar los componentes  $I_R(x, y)$  e  $I_G(x, y)$
- $F_G(x, y)$ : matriz resultante de filtrar el componente  $I_G(x, y)$  e  $I_B(x, y)$
- $I_{cel}(x, y)$ : matriz que contienen las regiones de células proliferantes y no proliferantes
- $I_{nopro}(x, y)$ : matriz resultante que contiene solo las células no proliferantes
- $I_{total}(x, y)$ : matriz resultante que es la unión de  $I_{nopro}(x, y)$  e  $I_{pro}(x, y)$
- $j$ : numero de objeto que posee un tamaño de más de 1500 pixeles
- $I_j(x_j, y_j)$ : segmento de  $I_{total}(x, y)$  que contiene el objeto  $j$
- $I_j^\theta(x'_j, y'_j)$ : matriz de objeto  $j$  rotado  $\theta$  grados
- $P_j^\theta(y'_j)$ : proyección horizontal del objeto  $j$  girado  $\theta$
- $\hat{P}_j^\theta(y'_j)$ : zona central de la proyección  $P_j^\theta(y'_j)$
- $\hat{P}_j^\theta(y_0(\theta, j))$ : valor mínimo de la proyección  $\hat{P}_j^\theta(y'_j)$
- $y_0(\theta, j)$ : posición en  $\hat{P}_j^\theta(y'_j)$  donde se encuentra el valor mínimo
- $W_j$ : vector donde se almacenan todos los valores mínimos por cada objeto  $j$
- $Z_j(x'_j, y'_j)$ : imagen resultante del objeto  $j$  donde se ha eliminado el valor mínimo de la proyección horizontal, generado la separación de célula

- $Total_{pro}$ ,  $Total_{nopro}$ : contadores de células proliferantes y no proliferantes respectivamente
- $T_{obj}$ : número total de objetos o células que contiene  $I_{total}(x, y)$
- $I_k(x^*, y^*)$  : segmento extraído de  $I_{total}(x, y)$
- $f_k$ : variable que evalúa si aumenta el contador de  $Total_{pro}$  o de  $Total_{nopro}$
- $Apro_k$ : variable que contabiliza el área de cantidad proliferante de una célula
- $Anpro_k$ : variable que contabiliza el área que no es proliferante de una célula proliferante
- $R_k$ : factor de intensidad de cada célula
- $fp_k$ : contador de célula altamente proliferante
- $fm_k$ : contador de célula medianamente proliferante
- $fd_k$ : contador de célula débilmente proliferante
- *Diagnóstico*: porcentaje final del índice de proliferación
- $\tilde{K}$ : índice Kappa
- $\bar{T}$  : probabilidad relativa entre observadores
- $\bar{T}_e$ : probabilidad hipotética de acuerdo al azar
- F: Valor F o F-Score
- *Precisión*: variable que representa la tasa de acierto respecto un observador a una porción de muestra
- *Exhaustividad*: variable que representa la tasa de acierto respecto un observador al total de muestras
- $\beta_2$ : variable de balance entre precisión y exhaustividad

## RESUMEN

El objetivo de esta tesis es proponer un algoritmo diseñado para cuantificar la expresión de la proteína KI-67 en imágenes digitales; obtenidas a través de microscopio óptico de muestras de tejido de biopsia de mama. El mismo que permite obtener un reporte de la cantidad de células proliferantes y no proliferantes a través de la detección y cuantificación del marcador KI-67. El análisis de muestra vía software apunta a reducir el nivel de subjetividad por inspección visual por parte de los especialistas en el diagnóstico de enfermedades tal como el cáncer de mama. El algoritmo propuesto involucra la aplicación de técnicas estadísticas de procesamiento digital de imágenes, tales como umbrales adaptivos, segmentación de objetos y filtrado de color. Asimismo, para mejorar su eficiencia se propone un método de análisis y cuantificación de células sobrepuestas.

Para la etapa de validación, se realizó pruebas de conteo de muestras entre especialistas y se comparó con el resultado del software para diferentes muestras, dando resultados satisfactorios. Los resultados del método propuesto fueron altamente correlacionados con los resultados obtenidos por inspección visual de patólogos, obteniendo así un *F-score* de 0.8764 y un índice *Kappa* de 0.883; mientras que el obtenido por los especialistas era de 0.686.

**Palabras clave:** KI-67, cáncer de mama, inmunohistoquímica, células, proliferación, procesamiento digital de imágenes.

## **ABSTRACT**

This thesis proposes an algorithm aimed at quantifying the expression of KI-67 protein in digital images of breast biopsy tissue samples obtained through an optical microscope. The same that allows to obtain a report on the quantity of non-proliferating and proliferating cells through the detection and quantification of KI-67. The sample analysis via software aims to reduce the level of subjectivity in the diagnosis by visual inspection made by specialists in the diagnosis of diseases such as breast cancer. The algorithm proposed involves the application of statistical image processing techniques, such as adaptive thresholds, object segmentation and color filtering. Likewise, to improve its efficiency a method of analysis and quantification of overlapping cells is proposed.

For the stage of validation, a comparison between doctors and software counting results for different samples was made, the results of the method proposed were satisfactory, as they were highly correlated with those obtained by visual inspection by pathologists, obtaining a F-Score of 0.8764 and a Kappa index of 0.883, meanwhile the index obtained by the specialists was of 0.686.

Keywords: KI-67, breast cancer, immunohistochemistry, cells, proliferation, digital image processing.

## INTRODUCCIÓN

Según datos del Instituto Nacional de Enfermedades Neoplásicas (INEN) hasta el 2014, el número de casos de cáncer de mama detectados en el Perú ha aumentado, ubicándose en el segundo lugar de mortalidad. Una de las pruebas que se realizan para evaluar el crecimiento de esta enfermedad es la prueba de conteo del índice de proliferación KI-67. Esta prueba se realiza mediante la inspección visual de una muestra de biopsia tratada previamente.

El problema radica en la alta subjetividad a la hora de realizar el examen, puesto que el médico especialista hace una revisión rápida de la lámina y calcula visualmente qué porcentaje hay de tinción en la muestra. Esto se ha demostrado en publicaciones como *Akihiro K. et al* (2016), donde un médico da como resultado un porcentaje y otro médico discierne dando otro resultado.

Actualmente, existen soluciones completas de proceso de tinción y cálculo de porcentaje en el mercado, que poseen fiabilidad alto, aunque el costo de estas soluciones es elevado.

Hoy en día, los principales centros de salud donde se realizan pruebas del marcador KI-67 no disponen de una herramienta que realicen el cálculo del índice de proliferación de forma automática. Es por ello, que la presente tesis está orientada al desarrollo de una herramienta de software que realice la cuantificación del índice de proliferación de la proteína KI-67 en imágenes



de muestras de tejido mamario obtenidos vía microscopio óptico utilizando técnicas de procesamiento digital de imágenes.

Para desarrollar este software se trabajó con un médico patólogo que proporcionó los conceptos médicos necesarios para el entendimiento de los índices que se deben hallar y cuantificar. Para la etapa de validación, igualmente se tuvo el apoyo de cinco médicos patólogos de diferentes instituciones del cuidado de la salud.

## **CAPÍTULO I**

### **ANTECEDENTES Y DESCRIPCIÓN DEL PROBLEMA**

#### **1.1 Comentarios Iniciales**

En esta sección de la tesis se comentará el problema social e ingenieril que existe y se desea resolver. También se comentarán los estudios previos e investigaciones que se realizaron para intentar solucionar el problema presentado.

##### **1.1.1 Situación Problemática Médica**

El cáncer es un problema de salud pública a escala mundial, pues así lo demuestran sus altas tasas de incidencia y mortalidad.

En el Perú, por ejemplo, el Registro Poblacional de Cáncer de Lima (E. Paquimio Rojas en 2008) ha registrado aproximadamente solo en el año 2011 un total de 34000 nuevos casos de cáncer. Asimismo, según este autor, en el periodo 2006-2011, el cáncer de mama ocupó el segundo lugar en número de casos detectados en mujeres peruanas. A nivel de investigación médica, se han validado algunos biomarcadores para la detección y estimación del avance de la enfermedad. Entre ellos, se tiene el receptor de estrógeno (*estrogen receptor*, ER), el receptor de progesterona (*progesterone receptor*, PR), el HER2 y el KI67 (según W. C. Ramos et al. en 2013 y S. Coronato et al. en 2002). Este último, a través de un análisis cuantitativo, brinda una estimación confiable de la actividad proliferativa en un tejido. Su

aplicación en las biopsias de pacientes con cáncer de mama brinda al médico algunos elementos de juicio para atacar la enfermedad.

Para demostrar la presencia del antígeno de la proteína KI-67 en una muestra de biopsia, se utiliza un anticuerpo específico conjugado con una sustancia cromógena que permite su visualización. Finalmente, a través de inspección visual vía microscopio óptico, el médico especialista analiza la muestra (de manera subjetiva), a fin de emitir un informe respecto a la magnitud de la actividad proliferativa de las células Ki67 positivas.

Según el Consenso de *St. Gallen* (*S. Bustreo, S. Osella-Abate et al. 2016*), el número de células proliferativas con respecto al total de células en una región determinada de tejido, llamado desde ahora: índice de proliferación, se dividió en dos niveles de predicción para el Ki-67 a los cuales se llamará: Luminal A, para porcentajes menores a 20% y Luminal B para porcentajes mayores al 20%.

La evaluación visual es generalmente subjetiva y puede generar errores en el diagnóstico, esto se debe a varios factores: iluminación del ambiente donde se evalúa la muestra, calidad de la imagen, estrés laboral, fatiga, desconcentración, capacidad visual, experiencia del evaluador, etc. Esto genera muchas veces que los resultados proporcionados por los médicos patólogos sean no correlacionados y presenten una alta varianza en la cuantificación del índice de proliferación emitidos sobre una misma muestra de tejido mamario (*Arihiro, K. et al. en 2016; Fulawka L. & Halon A. en 2017*).

La evaluación visual de KI-67 podría tener dificultades para distinguir los contornos del núcleo de las células y en la clasificación de células debido a las distribuciones de color extremadamente desigual. Además, la cuantificación visual es un proceso de alta concentración y que consume un tiempo considerable dentro de las actividades de los patólogos si es que se hace con el rigor adecuado (*Fulawka L. & Halon A. en 2017; Shi, P. et al. en 2016*).

Como existen estudios previos que demuestran la alta probabilidad de cometer errores en la cuantificación del índice de proliferación

del marcador KI-67, y clasificación del mismo (Luminal A o Luminal B), la posibilidad de encontrar muestras que arrojen falsos positivos o falsos negativos es muy alta y como consecuencia se puede provocar un tratamiento inadecuado para un paciente.

### **1.1.2 Situación Problemática Ingenieril**

Diversas instituciones del cuidado de la salud y centros de investigación de ciencias médicas realizan pruebas de biopsia del marcador KI-67 en el Perú, como el Instituto Nacional de Enfermedades Neoplásicas (INEN), Instituto Arias Stella, HNERM. Cada institución utiliza sus propios instrumentos, procesos de tinción y análisis de las muestras; lo cual originan muestras de KI-67 no estandarizadas. Tomando en cuenta el reactivo que se utilizó para teñir las células proliferantes, la densidad del epitelio del paciente, las condiciones de iluminación del ambiente donde se realizan el análisis de las muestras y el dispositivo sobre el cual se hacen análisis de las pruebas, se tienen diferentes condiciones de color, forma, iluminación y ruido en una muestra, lo cual hace que los médicos analicen el índice de proliferación del KI-67 de una forma muy subjetiva; esta situación también se ve reflejado en un estudio hecho por *C. Focke, H. Bürger et al. (2017)*.

Al intentar solucionar este problema con procesamiento digital de imágenes, se encuentra con un panorama adverso para un proceso de segmentación y posterior etiquetado. Las características morfológicas de las células proliferantes y no proliferantes son similares y también los espacios de color que representan en el espectro de colores RGB. Además, las células presentan una distribución espacial no uniforme que origina regiones con pocas células y regiones muy cargadas de células. Las células demasiado cercanas o superpuestas provocan aún más confusiones en la segmentación de objetos con el uso de las técnicas convencionales.

### **1.2 Justificación e Importancia de la Investigación**

A continuación, se expondrá el problema que se quiere solucionar y las razones por las cuales la investigación es importante.

### **1.3 Definición del problema**

Alta subjetividad visual por parte de médicos patólogos respecto a la identificación y conteo de células cancerígenas con reacción al antígeno KI-67, en imágenes digitales de biopsias de mama adquiridas vía microscopio óptico.

### **1.4 Justificación Social**

Según datos del (INEN) hasta el 2014, el número de casos de cáncer de mama detectados en el Perú ha aumentado, ubicándose en el segundo lugar de mortalidad como muestra la Tabla 1.

Esta tabla evidencia la gran demanda de pruebas médicas que existen en el Perú. Tomando en cuenta que las buenas prácticas en la patología recomiendan la evaluación de pocas muestras en un día para cada especialista (para evitar la subjetividad), existe un excedente de pruebas que no se procesan diariamente. Este hecho adicionado a las pocas instituciones que realizan pruebas de marcadores tumorales se encuentra con un panorama de muchas pruebas inmunohistoquímicas; pruebas de laboratorio para la que se usan anticuerpos a fin de identificar ciertos antígenos (marcadores) en una muestra de tejido; por procesar y poco personal disponible a evaluarlo.

Tabla 1: Casos nuevos de cáncer en mujeres registrados en el INEN, periodo 2000-2014

LOCALIZACION	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
CERVIX	1319	1360	1402	1337	1378	1358	1532	1499	1621	1593	1567	1610	1635	1575	1485
MAMA	1018	1002	1015	1009	1016	1034	1160	1112	1105	1193	1236	1274	1337	1230	1212
ESTOMAGO	277	288	267	278	283	290	331	335	344	384	365	365	383	422	443
TIROIDES	208	176	188	217	243	233	232	217	249	256	284	350	392	402	472
LINFOMA NO HODGKIN	225	214	244	202	251	253	266	237	236	241	287	250	276	300	270
PIEL NO MELANOMA	199	185	189	195	193	208	225	224	225	246	257	269	300	276	303
PULMON	152	117	158	161	186	190	196	209	215	219	210	221	200	202	226
OVARIO	182	157	166	187	156	180	198	163	198	203	240	209	199	202	200
LEUCEMIA LINFOIDE	119	125	125	137	136	142	130	140	141	131	125	150	170	178	177
CUERPO UTERINO	81	78	93	101	105	101	124	116	122	135	145	171	140	175	165
PRIMARIO DESCONOCIDO	127	114	135	121	147	110	111	115	110	123	100	125	126	139	134
CAVIDAD ORAL	84	91	109	103	116	116	122	128	121	114	134	153	145	137	155
COLON	81	69	83	93	67	112	119	119	138	136	152	154	151	164	137
VESICULA BILIAR	77	85	73	90	94	93	111	114	111	133	118	128	158	152	146
SIST.NERVIOSO CENTRAL	69	93	83	105	92	80	98	84	79	83	119	120	131	161	176
TEJ.BLANDOS Y PERITONEO	62	83	88	84	100	91	112	113	89	84	101	107	93	104	110
RECTO	69	59	67	68	74	78	76	90	116	125	111	107	132	126	114
LEUCEMIA MIELOIDE	78	76	77	86	84	90	97	79	86	83	83	100	92	112	96
RIÑÓN	57	59	53	53	53	51	87	88	88	91	71	81	104	99	118
MELANOMA	54	61	53	56	68	67	87	93	68	75	57	78	91	86	104
PANCREAS	44	48	64	50	47	70	77	61	64	83	70	96	99	88	108
HIGADO	61	51	61	60	57	60	70	78	66	80	89	85	66	79	88
ANO	39	32	34	41	36	44	51	48	31	56	50	52	59	60	63
HUESOS Y CARTILAGO	26	26	39	35	33	46	45	44	41	51	36	39	44	43	52
VULVA	33	29	40	22	38	29	39	29	39	44	44	42	50	48	41
OJO	29	26	35	18	36	30	39	35	36	35	38	48	39	44	51
VEJIGA	29	31	26	31	32	37	26	30	28	34	28	43	41	34	55
VIAS BILIARES	19	15	18	17	29	14	28	34	35	35	35	40	39	43	42
CORIOCARCINOMA	37	29	27	45	34	27	36	38	27	30	19	20	23	18	16
LINFOMA HODGKIN	28	21	27	31	30	24	30	23	26	26	22	23	25	31	32
MIELOMA	20	23	17	21	30	15	29	26	23	26	23	27	35	37	46
OTRAS LEUCEMIAS	4	4	5	6	10	18	22	18	27	21	24	31	30	29	31
SENOS PARANASALES	14	13	14	19	14	20	25	15	13	22	13	26	15	21	16
ESOFAGO	10	11	15	17	9	10	13	19	28	25	25	19	18	24	14
FOSA NASAL	13	15	13	11	12	18	13	15	17	18	14	14	13	16	20
LARINGE	11	6	8	5	12	1	5	6	6	9	11	9	7	8	10
OTROS	99	90	75	83	94	99	106	94	95	89	85	103	119	113	124
<b>Total</b>	<b>5054</b>	<b>4962</b>	<b>5186</b>	<b>5195</b>	<b>5395</b>	<b>5439</b>	<b>6068</b>	<b>5888</b>	<b>6064</b>	<b>6332</b>	<b>6388</b>	<b>6739</b>	<b>6977</b>	<b>6978</b>	<b>7052</b>

Fuente: Datos epidemiológicos INEI (2000-2014)

Se concluye que es necesaria una herramienta que permita a los médicos especialistas poder cuantificar las muestras de tejido inmunohistoquímico de KI-67 de forma automática, puesto que, el método visual demanda de mucho tiempo y los médicos no se darían tiempo suficiente para analizar todas las muestras y obtener una baja subjetividad.

### 1.5 Justificación Ingenieril

- Se ha comprobado mediante varios estudios científicos y médicos, que el índice de proliferación del marcador KI-67 está relacionado con el cáncer de mama y en qué fase de desarrollo se encuentra. En el Perú, los análisis se hacen de forma visual subjetiva.

- El diseño del software mejorará la precisión de conteo de células proliferantes y no proliferantes, dando un porcentaje más acertado respecto a un análisis visual.
- La solución del problema a nivel ingenieril implica la aplicación de modelos matemáticos computacionales que permitan alcanzar un desempeño que pueda satisfacer las expectativas de los médicos patólogos.
- Se realizarán validaciones de los datos obtenidos por el software con distintos médicos patólogos para comprobar su eficiencia.

## 1.6 Antecedentes

Investigaciones anteriores sobre posibles soluciones a este problema se han propuesto en diversos trabajos; sin embargo, los principales problemas que afectan la detección de células en softwares similares son, por ejemplo, imperfección en la recolección de datos presentes en forma de ruido; distorsión cromática y deformidades en el material histopatológico causado en la etapa de pigmentación de la muestra. Anari, V. (2010) sostiene, por ejemplo, que los resultados del algoritmo de cuantificación fueron prósperos al analizar el KI-67 con distintas técnicas de procesamiento de imágenes de microscopio como umbralización, segmentación y etiquetado. Sin embargo, usando la técnica de ultra erosión se terminó deteriorando información valiosa para el proceso de evaluación de la muestra.

*M. Bouzid et al.* (2013) muestra por otro lado una interesante propuesta con la utilización de un hardware especial: cámara multiespectral con fuente de luz programable. Esto definitivamente elevó el costo y la disponibilidad de la solución.

*F. Kabir y N. Yusoff* en 2013 utilizaron el método de *Random Forest* (RF) (Bosques Aleatorios) consistentes en estructuras tipo árbol predictores, cada uno independiente de los demás; obteniendo sus datos de distintos vectores aleatorios, pero de misma distribución. RF ha demostrado ser un método con un desempeño mejorado a diferencia de otras técnicas de aprendizaje de máquinas, por ejemplo: Redes Neuronales, Máquinas de soporte vectorial; sin embargo, en los resultados obtenidos según sus

pruebas/exámenes/ensayos, la precisión obtenida fue de 72% en la diferenciación entre tejido maligno y benigno.

### **1.7 Hipótesis**

Una herramienta de software que utiliza procesamiento digital de señales reducirá el nivel de subjetividad en los diagnósticos y resultados de pruebas inmunohistoquímicas de la expresión KI-67 mediante un conteo no subjetivo de células proliferantes y no proliferantes.

### **1.8 Objetivos**

Se proponen a continuación los objetivos trazados en el desarrollo de la tesis de investigación.

#### **1.8.1 Objetivo General**

Desarrollar un aplicativo de software de procesamiento de imágenes, destinado a reducir el nivel de subjetividad que existe en el análisis por percepción visual de la expresión KI-67 en muestras de tejido adquiridos por biopsia, cuantificando las células proliferantes y no proliferantes de forma más acertada.

#### **1.8.2 Objetivos Específicos**

- Conocer los criterios y estándares que utilizan los patólogos al examinar la pigmentación y morfología de las células en muestras de tejido mamario.
- Investigar técnicas y algoritmos de segmentación para la identificación de células proliferantes.
- Aplicar técnicas de procesamiento de imágenes a las imágenes de tejido mamario para mejorar la cuantificación del índice de proliferación correspondiente a la evaluación de la proteína KI-67.
- Diseñar una interfaz gráfica de fácil uso para el algoritmo cuantificador con la finalidad de resultar atractivo para el usuario final (médico patólogo).



- Validar los datos obtenidos por el software desarrollado mediante reuniones periódicas con los especialistas y hacer una comparación de sus resultados con respecto a los resultados que otorgan los evaluadores.

### **1.9 Solución propuesta y Diagrama de bloques**

La solución propuesta consta de cinco secciones como muestra la Figura 1. Se explica a continuación un resumen de cada una de ellas:

La primera, es la etapa de adquisición de la imagen. Se explica la etapa de adquisición de las muestras tratadas por inmunohistoquímica (proceso de tinción) y así conseguir una estandarización de los niveles de color y contraste. También se explica el aumento del lente para tener un estándar de tamaños de células.

La segunda etapa, es la del pre-procesamiento de la imagen digitalizada. Esta es la sección en la que se adecua la imagen para los procesos posteriores de procesamiento. Etapa de filtraje de ruido y corrección de iluminación.

La tercera etapa, es la segmentación y etiquetado. En esta parte se utilizan diferentes criterios de umbralización para la detección de células proliferantes y no proliferantes. Además de la propuesta de un proceso de separación de células superpuestas por procesamiento digital de imágenes.

Cuarta etapa, es la clasificación de células. Luego de segmentar adecuadamente, las células son clasificadas en dos valores, proliferantes y no proliferantes, y contadas para calcular el índice de proliferación del marcador KI-67.

Finalmente, la quinta etapa es la entrega de resultados al médico especialista para la debida interpretación de los datos arrojados por él software desarrollado.

## **1.10 Comentarios finales**

En este capítulo se han tratado los aspectos introductorios para que el lector tenga la capacidad de entender los conceptos globales y los objetivos necesarios para acompañar de forma adecuada el desarrollo de esta tesis.

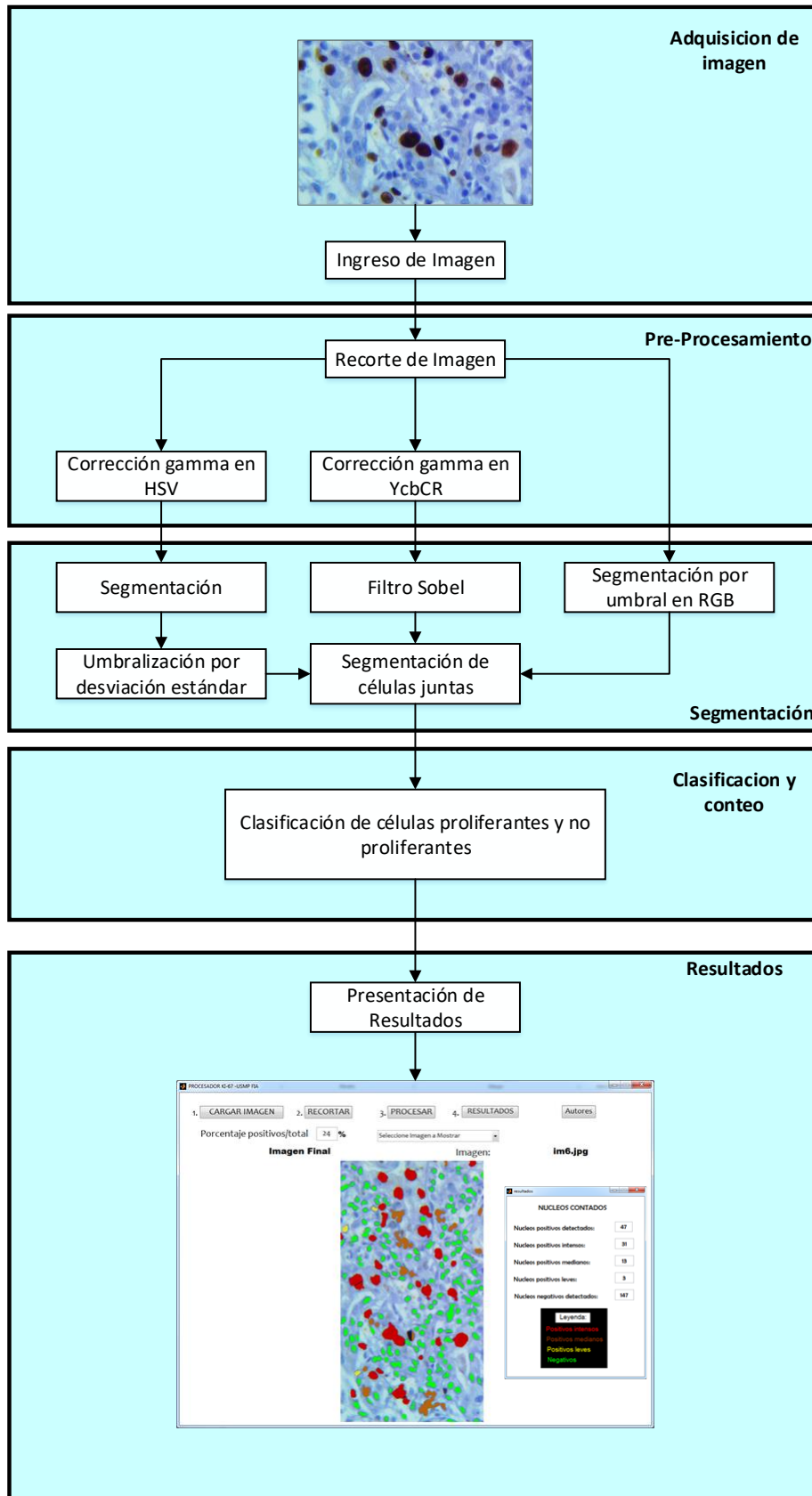


Figura 1. Diagrama de bloques de la solución propuesta  
Elaboración: los autores

## **CAPÍTULO II**

### **MARCO TEÓRICO Y CONCEPTUAL**

#### **2.1 Comentarios Iniciales**

En este capítulo se detallarán los conceptos y teorías necesarios para el desarrollo de la tesis. Está comprendido en dos partes importantes, la primera son los fundamentos médicos para entender parte del problema a resolver, en este caso el marcador KI-67, desde donde se origina, que es y cómo es que se trata químicamente para que sea visible al médico especialista; la segunda parte comprende todos los métodos matemáticos y computacionales utilizados para resolver el problema.

#### **2.2 Ciclo Celular**

El ciclo celular son los eventos que posee una célula para completar un ciclo vital, es decir la serie de etapas de crecimiento y desarrollo que experimenta la célula entre su nacimiento y su división en dos células. En la Figura 2 se aprecia las etapas del ciclo celular. Están son:

- Fase G1: del inglés “Gap 1” (Intervalo 1), la célula crece físicamente, copia sus organelos y crea componentes moleculares que necesitará para etapas posteriores
- Fase S: del inglés “Synthesis” (Síntesis), la célula sintetiza una copia completa del ADN en su núcleo. También duplica su

estructura de organización, llamado centrosoma. Estos ayudarán a separar el ADN en la fase M.

- Fase G2: vendría a ser el Gap 2 (intervalo 2), durante esta etapa, la célula crece más, generando nuevos organelos. Comienza también a reorganizarse en preparación para la mitosis, esta fase acaba cuando esta empieza.

A las fases G1, S y G2, en conjunto se les conoce como Interfase.

- Fase M: en esta última etapa, la célula divide su ADN duplicado y su citoplasma para hacer dos nuevas células.

También se puede apreciar en el gráfico una etapa que no está dentro del ciclo, este es el G0.

- Fase G0: en esta etapa se encuentran las células que poseen un ciclo de división celular más lento o inclusive de que carecen de este. Se puede decir que son las células que están en etapa de reposo, realizando otros trabajos como, por ejemplo, las neuronas.

Algunos tipos de células tienden a entrar rápidamente al ciclo celular. Por ejemplo, las células embrionarias cuando se está generando el feto. Otro claro ejemplo, son los tumores que tienden a tener un rápido crecimiento o proliferación.

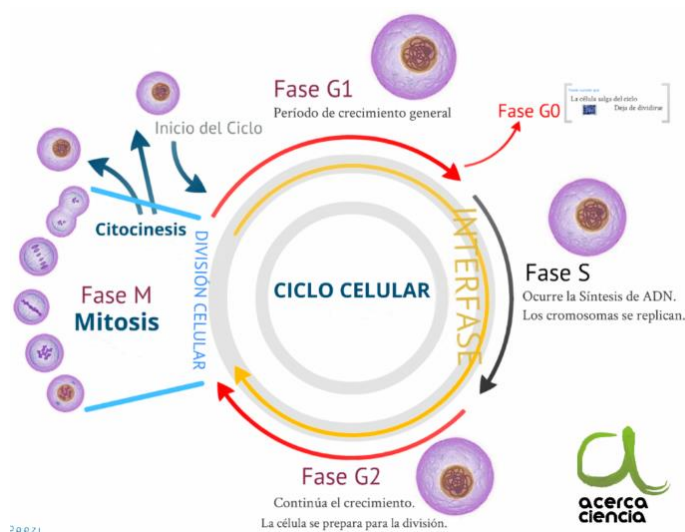


Figura 2. Etapas de reproducción celular  
Fuente: AcercaCiencia

## **2.3 Marcadores tumorales**

Uno de los campos de estudio de la patología es el cáncer, este proyecto se enfoca en el estudio de un marcador tumoral, sin embargo, existen diferentes tipos de marcadores que evalúan diferentes parámetros del ciclo celular.

En los últimos años se ha ampliado el espectro de marcadores en cáncer de mama, debido a los avances de la biología molecular que está estudiando en profundidad los eventos de la carcinogénesis, la progresión tumoral y los mecanismos de producción de las metástasis. Bajo el auspicio del Colegio de Patólogos Americanos se reunió, en 1999, un equipo interdisciplinario de clínicos, patólogos y expertos en estadística a fin de evaluar el uso de los marcadores pronósticos y predictivos en cáncer de mama y clasificarlos en categorías, en base a su utilidad clínica. Categorización de los marcadores tumorales (MT) en cáncer de mama (Tabla 2.1).

- Categoría I: marcadores cuya utilidad como factor pronóstico y de orientación terapéutica está indudablemente comprobada.
- Categoría II: marcadores ampliamente estudiados desde el punto de vista tanto clínico como biológico, pero cuya importancia debe ser validada con estudios estadísticos.
- Categoría III: otros marcadores aún no suficientemente estudiados como factores pronósticos. Este trabajo de categorización consta de una detallada exposición de los hallazgos y recomendaciones de los autores y proporciona una ayuda invaluable para los profesionales y técnicos dedicados a la evaluación clínica del cáncer de mama.

### **2.3.1 KI-67**

El foco de este proyecto radica en el KI-67, de la categoría 2 de los marcadores tumorales. El KI-67 es una proteína nuclear no histona, que está asociada al índice de proliferación celular. Está activamente presente en casi todas las etapas del ciclo celular (G1, S, G2, M), pero ausente en la etapa inicial o de reposo (G0).

Tabla 2: Categorización de los Marcadores Tumorales de Cáncer de mama.

CATEGORÍA I	CATEGORÍA II	CATEGORÍA III
Tamaño del tumor	Her-2/neu	Ploidia de ADN
Estado de los ganglios	P 53	Angiogénesis
Micrometastasis	Invasión vascular	EGFr
Ganglio centinela	KI-67	Bcl-2
Grado histológico	Síntesis de ADN	P S2
Tipo histológico		Catepsina
Conteo micótico		
Receptores hormonales		

Fuente: Marcadores tumorales en cáncer de mama. Elaboración: Los autores

Durante las etapas donde está presente el KI-67 puede ser encontrado dentro del núcleo y en la etapa de mitosis se le puede observar en la superficie de los cromosomas. En los estudios clínicos no se han encontrado presencia del KI-67 durante la etapa de reparación del DNA.

En la Figura 3 se aprecia una muestra de biopsia ya teñida para la detección del gen KI-67. Las células de color marrón son aquellas que reaccionaron a la proteína KI-67 y a las cuales se denominarán desde ahora “células proliferantes” y las células de color azul son aquellas que se encuentran en estado de reposo y se denominarán “células no proliferantes”.

Cuando los médicos evalúan el índice de proliferación, agarran una muestra, ya sea vista desde un microscopio o una muestra digitalizada, de ahí escogen una región de interés, donde consideren que se encuentra una vista clara, bien teñida, sin residuos de tinción y calculan de forma visual cuanto porcentaje hay de las células proliferantes (de color marrón) respecto al total.

## 2.4 Inmunohistoquímica

La inmunohistoquímica es un procedimiento histopatológico que se basa en la utilización de anticuerpos que se unen específicamente a una sustancia que se quiere identificar (anticuerpo primario). Estos anticuerpos pueden tener unida una enzima o esta puede encontrarse unida a un anticuerpo secundario que reconoce y se une al primario.

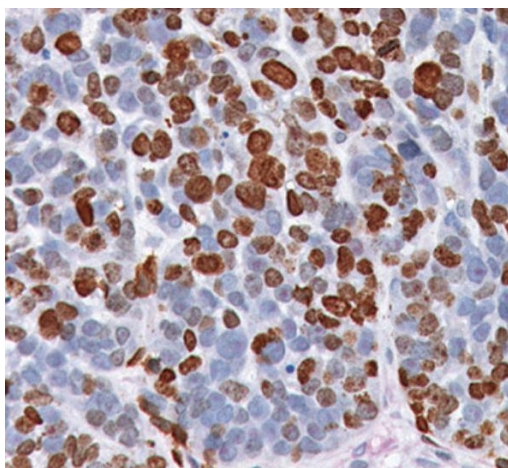


Figura 3. Muestra de tejido lista para observación  
Fuente BiocareMexico

Aplicado a un tejido orgánico, el anticuerpo primario se une específicamente al sustrato y se aprovecha la actividad enzimática para visualizar la unión. De esta manera se consigue un complejo sustrato-anticuerpos-enzima unido al lugar donde se encuentre el sustrato y mediante la activación de la enzima con la adición de su sustrato, se genera un producto identificable donde se encuentre el complejo.

Esta técnica permite identificar la localización de una sustancia específica permitiendo identificar su localización tisular o citológica, de esta manera se pueden identificar los marcadores antigénicos característicos de una línea celular, identificar células que secretan una proteína, receptores de membrana, gradientes de concentración tisulares o células que han respondido a una hormona (con anticuerpos específicos para las vías de señalización intracelular).

## **2.5 Fundamentos de procesamiento digital de imágenes**

Según R. Gonzales (2006), cuando se habla de procesamiento digital de imágenes (PDI), se refiere al procesamiento de estas mediante el uso de computación digital. Está establecido que la visión es el sentido más importante en la percepción humana y abarca un espectro electromagnético limitado, en cambio las computadoras pueden abarcar casi por completo el espectro electromagnético; pueden operar sobre imágenes generadas por sensores que comúnmente el humano no está acostumbrado a asociarlo a



una imagen, como por ejemplo ultrasonido, microscopio electrónico, señales de radar, etc.

Los procesos que se pueden desarrollar en PDI están categorizados en tres niveles: nivel bajo, medio y alto. En el nivel bajo se encuentran operaciones básicas como mejora de contraste, reducción de ruido, donde la entrada y la salida del sistema serán siempre imágenes. Del otro lado, se tiene el procesamiento de nivel medio, donde se pueden encontrar técnicas como segmentación, clasificación, etiquetado, en este nivel la entrada mayormente son imágenes y la salida pueden ser atributos de estas; como bordes, contornos e identidades individuales de un objeto. Finalmente, están los procesamientos de alto nivel que incluye la detección automática de objetos, visión computacional o aproximar el cómputo de un computador lo más cercano a función cognitiva asociada a la visión.

En el desarrollo de este proyecto se han utilizado técnicas de bajo y medio nivel para la detección de células; para empezar a procesar las imágenes, es necesario entender los conceptos que se presentan a continuación que fueron utilizados a lo largo del proyecto.

Los fundamentos teóricos de esta sección fueron mayormente basados en la literatura de Rafael Gonzales (2006 y 2008).

## **2.6 Teoría del Pixel**

Una imagen digital o gráfico digital es una representación bidimensional de una imagen a partir de una matriz numérica, frecuentemente en binario (unos y ceros). Dependiendo de si la resolución de la imagen es estática o dinámica, puede tratarse de una imagen matricial (o mapa de bits) o de un gráfico vectorial. En el caso de una imagen matricial, su composición está hecha por píxeles (R. Gonzales, Capítulo 1, 2006). En la Figura 4 se aprecia una imagen digital acercada al nivel que se aprecian los píxeles que lo componen.



Figura 4 Ejemplo de visualización de píxeles en una imagen digital aplicando aumento. Fuente: [píxeles-creativo.blogspot.com](http://píxeles-creativo.blogspot.com)

El píxel, proveniente de la unión de las palabras *PICTURE* y *ELEMENT*, es la unidad mínima de expresión de una imagen digital de 2 dimensiones, siendo esta capaz de representarse en dos formas: monocroma y policroma; y representa el nivel de intensidad luminosa de un punto de la imagen.

En el capítulo 2 de R. Gonzales (2006) se explica que la intensidad que pueda poseer cada unidad está determinada por el nivel de cuantificación de una imagen, la cuantificación es el método por el cual se le asigna una escala de valores finitos a cada píxel. Los niveles suelen ser dados en potencias de 2 para facilitar su almacenamiento o procesado en computadoras. Por ejemplo, una imagen binaria es aquella que tiene solo 1 nivel de cuantificación, o también llamado resolución radiométrica.

$$\begin{aligned} \text{resolución} &= 2^{\text{res}} \\ \text{res} = 1 &\rightarrow \text{resolución} = 2^1 = 2 \text{ (bits/píxel)} \end{aligned} \quad (2.1)$$

Esto significa que para  $\text{res} = 1$ , una imagen puede tener dos valores digitales, 0 (negro) o 1 (blanco).

Comúnmente se utiliza una resolución de 8 bits, que equivale a 1 byte de almacenamiento, que proporciona un rango de 256 valores, siendo, 0, el negro, y 255, el blanco, pues a este nivel, el cambio entre cada nivel de intensidad pasa imperceptible al ojo humano. En la Figura 5 se aprecia distintos niveles de cuantificación para una misma imagen.

Cuando se habla de imágenes monocromas, se suele denotar la intensidad del pixel mediante la función  $g(x,y)$ , donde “ $x$ ” e “ $y$ ” son coordenadas espaciales de la imagen y “ $g$ ” en un punto  $(x,y)$  es proporcional al brillo de la imagen o nivel de gris.

Siendo la imagen una matriz de pixeles de resolución espacial  $m \times n$ , como se observa a continuación:

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \quad (2.2)$$

Para imágenes polícromas, es decir, de varios colores, se tiene una función tridimensional,  $g(x,y,n)$ , donde “ $n$ ” representa la crominancia de cada pixel, correspondiente según cada modelo de color.



Figura 5. Comparación de distintos niveles de cuantificación  
Fuente: Carlos Platero, Apuntes de visión artificial.

## 2.7 Modelos de color

Los modelos de color son modelos matemáticos que permite representar mediante valores numéricos a los colores que el ser humano visualiza de distintas maneras según sea la aplicación. Estos suelen estar representado o en matrices unidimensionales (como el modelo escala de grises) o tridimensionales (ya sea el RGB o HSV), inclusive llegando hasta tetradimensionales como el caso del modelo CMYK).

- **Modelo Escala de Grises.** Maneja el canal negro y permite 256 tonos de gris entre el blanco y negro, como se ve en la Figura 6. Para la obtención del modelo de escala de grises, se suele hacer una conversión del modelo RGB hacia gris utilizando la siguiente ecuación:

$$gris = R \times 0.3 + G \times 0.59 + B \times 0.11 \quad (2.3)$$

donde  $R$ : es la banda de color ROJO.

$G$ : es la banda de color VERDE.

$B$ : es la banda de color AZUL.



Figura 6. Modelo de color escala de grises  
Fuente: Instituto Superior de Formación y Recursos en Red para el Profesorado.

- **Modelo RGB.** Los colores primarios son aquellos que no se generan en base a la combinación de otros colores. Este modelo se basa en el sistema de coordenadas cartesianas. Existen dos tipos de colores primarios, los de tipo químico y tipo de luz. El modelo de color RGB es del tipo primario de luz, cada color se forma por combinación de tres canales. Las imágenes representadas en el modelo RGB están compuestas de tres imágenes componentes principales, una por cada color primario: *Red* (rojo), *Green* (verde), y *Blue* (azul) (R. Gonzales, capítulo 2.6.1, 2006). Se le asigna un valor de intensidad a cada color que oscila entre 0 y 255 (resolución de 8 bits). De la combinación surgen hasta 16,7 millones de colores. Ejemplo: El valor R:255, G:0, B:0 representa al color rojo. En la Figura 7 se aprecia la imagen de la Figura 6 pero en modelo de color RGB. El modelo de color RGB se utilizó en el proyecto para realizar filtros de colores para separar las células proliferantes del fondo de la imagen.



Figura 7. Modelo de color RGB

Fuente: Instituto Superior de Formación y Recursos en Red para el Profesorado.

- **Modelo HSV.** Cada color surge de los valores de estos tres parámetros: **Hue** (Tono) que es el valor del color puro: rojo, azul, verde, etc. Se representa como un grado de ángulo de una circunferencia, con posibles valores en 0° y 360°, aunque algunas veces suelen normalizarlo entre 0 y 100%. **Saturation** (Saturación) que indica el porcentaje de pureza de un color puro diluido en luz blanca, su valor oscila entre 0% y 100%. **Value**

(Valor) referencia la intensidad de luz del color, es decir, la cantidad de negro o blanco que contiene estando su valor entre 0 (negro) y 100 (blanco) (R. Gonzales, capítulo 2.6.3, 2006).

Para convertir del modelo de color RGB a HSV se utiliza las ecuaciones desde 2.4 a 2.6, considerando  $MAX$  el valor máximo entre los componentes de  $R, G$  y  $B$ ,  $MIN$  el valor mínimo de estos y  $H, S$  y  $V$  como los componentes de *Hue*, *Saturation* y *Value*, respectivamente.

$$H = \begin{cases} 0^\circ, & \text{si } MAX = MIN \\ 60^\circ \times \frac{G - B}{MAX + MIN} + 0^\circ, & \text{si } MAX = R \text{ y } G \geq B \\ 60^\circ \times \frac{G - B}{MAX + MIN} + 360^\circ, & \text{si } MAX = R \text{ y } G < B \\ 60^\circ \times \frac{B - R}{MAX + MIN} + 120^\circ, & \text{si } MAX = G \\ 60^\circ \times \frac{B - R}{MAX + MIN} + 240^\circ, & \text{si } MAX = B \end{cases} \quad (2.4)$$

$$S = \begin{cases} 0, & \text{si } MAX = 0 \\ 1 - \frac{MIN}{MAX}, & \text{en otro caso} \end{cases} \quad (2.5)$$

$$V = MAX \quad (2.6)$$

Ejemplo: El color rojo tiene un código RGB como (255, 0, 0) y también un código HSV (0, 100, 100). En la mayoría de los programas de tratamiento de imágenes se puede elegir un color introduciendo su código RGB, es la opción más frecuente, o alternativamente su código HSV. En ambos casos la imagen maneja una paleta de colores de 24 bits. En la Figura 8 se aprecia un ejemplo de visualización en forma cónica.

En el proyecto se utilizó el modelo de color en la etapa de preprocesamiento para corregir la imagen conjuntamente con la corrección *gamma* para de ahí aplicar técnicas de desviación estándar para extraer las células no proliferantes del resto.

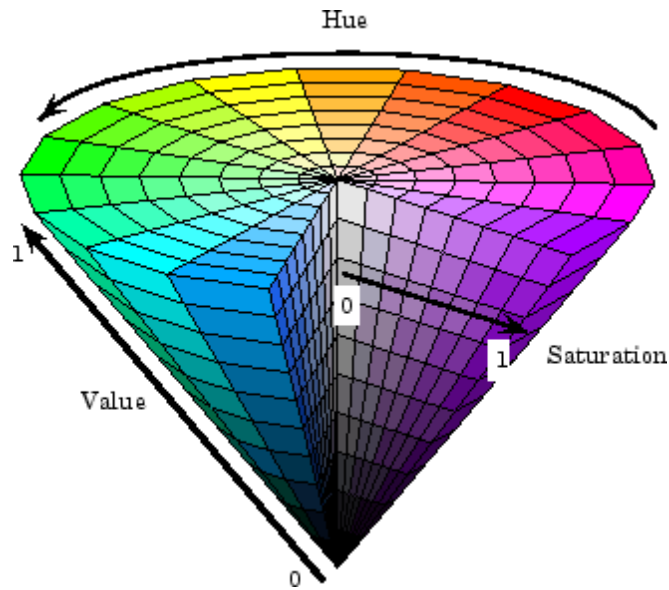


Figura 8. Modelo de color HSV  
Fuente: *Mathworks*

- **Modelo de color  $YCbCr$ .** es una familia de espacios de color utilizado mayormente en sistemas de vídeo digital y procesamiento de imágenes.

En este modelo, el color se define en términos de un componente de luminancia ( $Y$ ) y dos de crominancia ( $C_b$  y  $C_r$ ), estos últimos corresponden a los componentes de crominancia azul y rojo, respectivamente.  $C_b$  ubica el color en una escala de azul y amarillo, mientras que  $C_r$  ubica el color entre rojo y verde. Un ejemplo basado en  $Y = 0.5$  representado en un plano cartesiano se aprecia en la Figura 9.

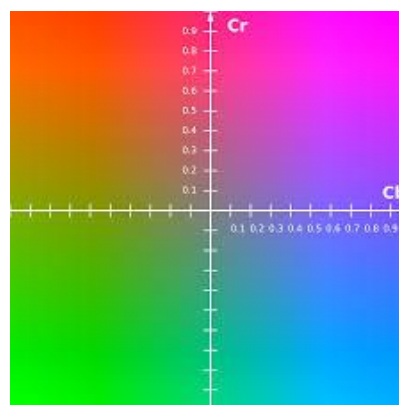


Figura 9. Ejemplo de plano de color  $C_b - C_r$  con luminancia  $Y=0.5$   
Fuente: *ResearchGate.net*

YCbCr es confundido a veces con el espacio de color YUV y generalmente este último término YCbCr son usados indistintamente dando lugar a cierta confusión. Cuando se refiere a las señales de vídeo analógicas o en formato digital, el término YUV significa YCbCr. Las ecuaciones para obtener cada componente de color, en base al modelo de color RGB con valores digitales entre 0 y 255 se muestran a continuación.

$$Y = 0 + 0.299 \times R + 0.587 \times G + 0.114 \times B \quad (2.7)$$

$$C_b = 128 - 0.169 \times R - 0.331 \times G + 0.5 \times B \quad (2.8)$$

$$C_r = 128 + 0.5 \times R - 0.419 \times G - 0.081 \times B \quad (2.9)$$

El modelo de color YCbCr es utilizado en el proyecto para corregir la imagen con corrección gamma para detectar células proliferantes y no proliferantes.

## 2.8 Transformación de Intensidad

La transformación de intensidad es una categoría del procesado en el dominio del espacio, esto se refiere a que opera directamente sobre el pixel de una imagen, la expresión común que se utiliza para el uso transformación está dada por:

$$e(x, y) = T[r(x, y)] \quad (2.10)$$

Donde  $r(x, y)$  es la imagen de entrada, en escala de grises,  $e(x, y)$  es la imagen de salida o procesada y  $T$  es el operador en  $r$  sobre una vecindad de pixeles en un punto  $(x, y)$ .

La forma más simple de transformación  $T$  es cuando la vecindad es de dimensiones  $(1 \times 1)$  (un único pixel). Para este caso,  $e(x, y)$  dependerá solamente del valor de intensidad de  $r(x, y)$ .

Como dependen solo de los valores de intensidad, y no explícitamente sobre  $(x, y)$ , estas funciones se suelen escribir en forma simplificada como



$c = T(d)$ . Donde  $d$  denota la intensidad de  $r$  y  $c$  la intensidad de  $e$ , ambas sobre un punto correspondiente  $(x, y)$  de la imagen.

• **Corrección gamma:**

La función de corrección gamma de R. Gonzales et al (2009) transforma los valores de intensidad de una imagen  $r(x, y)$  utilizando la expresión de la ecuación 2.7:

$$gamma(x, y) = 255 \left( \frac{r(x, y)}{255} \right)^\gamma \quad (2.11)$$

donde  $\gamma$  es el valor de gamma

En la Figura 10 se ilustra como la función transforma los valores de intensidad de la imagen de entrada  $r(x, y)$ , el parámetro gamma especifica la forma de la curva.

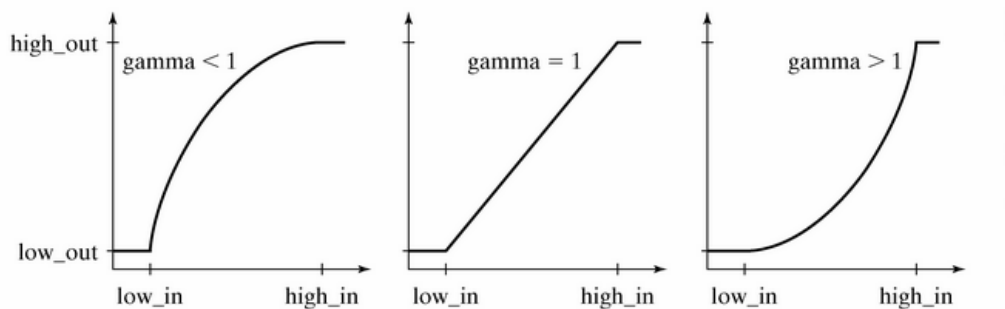


Figura 10. Representación de la función de corrección de intensidad variando  $\gamma$  Fuente: *Digital Image Processing using Matlab*

## 2.9 Filtros detectores de gradiente

• **Filtro Sobel:**

En base a la bibliografía de R. Gonzales et al. del capítulo 11 (2009) Es un tipo de filtro direccional, utilizado para detectar bordes en base a dos tipos de máscaras: columnas y filas, como se ve en la Figura 11. Estas matrices se aplican sobre la imagen deseada a procesar.

-1	0	1
-2	0	2
-1	0	1

(a)

1	2	1
0	0	0
-1	-2	-1

(b)

Figura 11. (a) Mascara de filtros columnas. (b) Mascara de filtro filas  
Elaboración: los autores

Para aplicar el filtro Sobel en imágenes digitales se aplica la fórmula de convolución discreta:

$$\begin{aligned}
 Z(x, y) &= \text{Img}(x, y) * h(x, y) \\
 &= \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} I(i, j) h(x - i, y - j)
 \end{aligned}
 \tag{2.12}$$

donde:

$\text{Img}(x, y)$ : es la imagen a ser tratada

$h(x, y)$ : es la máscara que se aplica sobre la imagen.

En la Figura 12 se muestran los resultados de procesar una imagen con las dos matrices de detección de borde.

## 2.10 Transformaciones morfológicas en imágenes

Las transformaciones morfológicas comentadas a continuación fueron recogidas de la bibliografía de Gonzales R. et al. del capítulo 9 (2006).

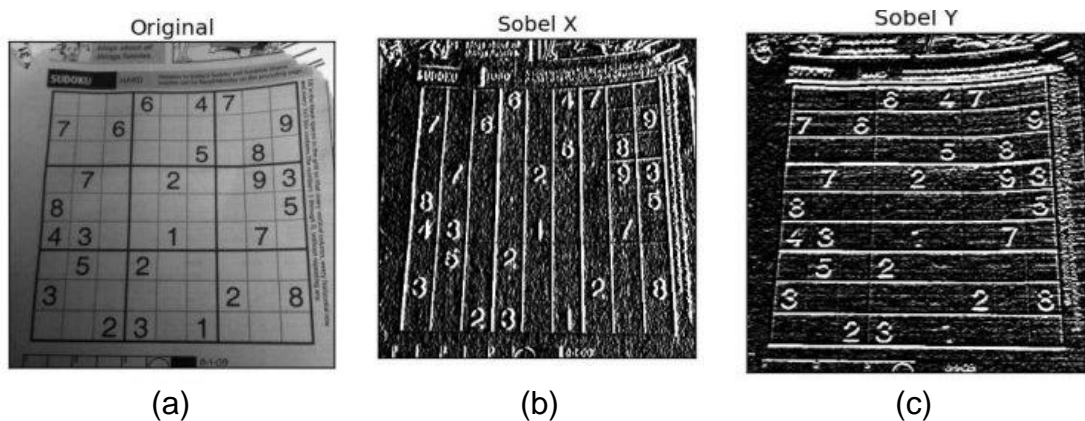


Figura 12. Etapas del filtro Sobel. (a) Imagen Original, (b) imagen procesada por filtro Sobel solo en la vertical, (c) imagen procesada solo en la horizontal

Fuente: Cold Vision Library

- **Geometría de la imagen:**

La rotación de una imagen  $I(x, y)$ , dado un cierto ángulo  $\theta$ , se logra aplicando una matriz de transformación a la imagen deseada, dada por:

$$Giro(x, y) = \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \times I(x, y) \quad (2.13)$$

donde *Giro*: Matriz resultante

$\theta$ : Ángulo de rotación deseado

$I(x, y)$ : la Imagen a rotar

Los resultados del proceso de giro son se observan en la Figura 13, siendo 13(a) la original y 13(b) la imagen rotada.

- **Dilatación:**

La dilatación es una operación morfológica que se encarga de “engordar” o “engrosar” un objeto en una imagen binaria (de fondo negro o tonalidad 0 y objetos blancos o tonalidad 255). Se puede comprender la dilatación como el mecanismo de una convolución espacial discreta, donde el elemento B recorre todas las ubicaciones distintas de 0 y va agregando espacio en base al elemento estructura.



(a) (b)  
 Figura 13 (a) Imagen Lena original (b) Imagen Lena rotada 45° sobre su centro en sentido antihorario  
 Elaboración: los autores

Estos elementos de estructura suelen ser representado por matrices binarias que representan una forma geométrica en particular (rombos, elipses, cuadrados, cruces, aspás, círculos, etc.), en una matriz de tamaño simétrico (3x3, 5x5, 7x7, etc.), según sea el tipo de dilatación que se desea darle, como se ve en la Figura 14.

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

(a)

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

(b)

Figura 14. Ejemplos de elementos estructurante (a) tipo cruz, (b) tipo aspá.  
 Elaboración: los autores

La ecuación de dilatación está denotada por el símbolo “ $\oplus$ ”, como se muestra en Gonzales R. et al. (2008):

$$A \oplus est = \{z \mid (\widehat{est})_z \cap A \neq \emptyset\} \quad (2.14)$$

donde: *est*: Elemento estructurante

*A*: matriz a dilatar

Los resultados del proceso de dilatación de muestran en la Figura 15.

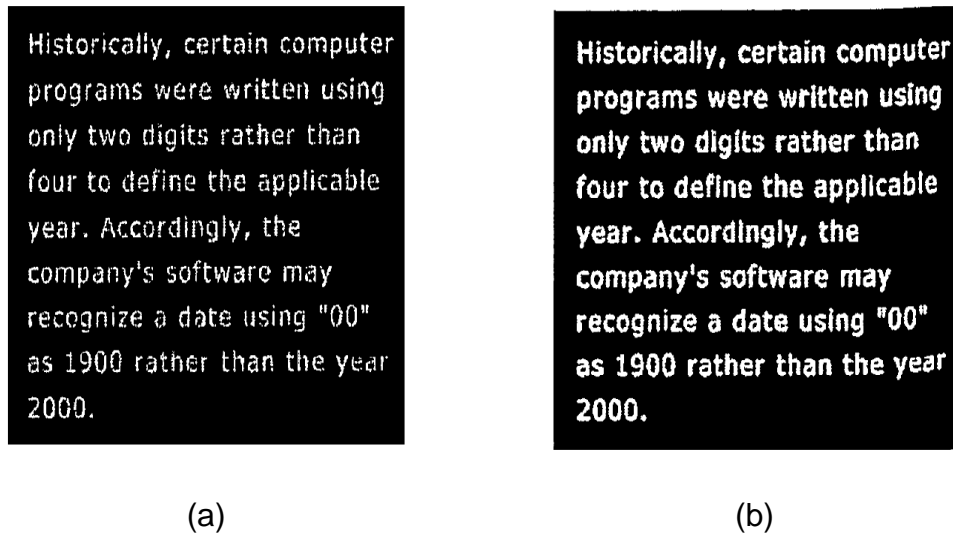


Figura 15. Ejemplo de dilatación: a) Imagen binaria original de texto  
 b) Imagen dilatada  
 Fuente: *Matlab*

**•Erosión**

Similar a la operación de dilatación, la erosión consiste en una convolución espacial discreta entre un elemento estructura con una imagen binaria, pero en este caso ocurre una disminución en el área del objeto erosionado.

La ecuación de erosión esta denotada por el símbolo "  $\ominus$  ":

$$A \ominus est = \{z \mid (est)_z \subseteq A \neq \emptyset\} \tag{2.15}$$

dónde: *est*: Elemento estructurante

*A*: matriz a erosionar

Los resultados del proceso de erosión se muestran en la Figura 16.

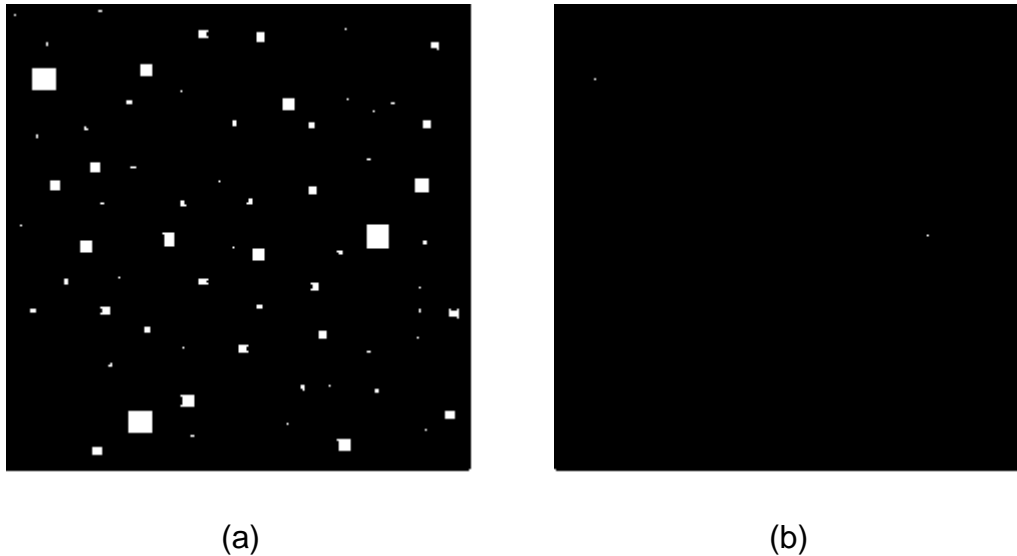


Figura 16. Ejemplo de erosión a) Imagen binaria original, b) Imagen erosionada con elemento estructurante 13x13  
Fuente: Matlab

Las técnicas de dilatación y erosión son utilizadas en las etapas de pre-procesamiento dentro de las correcciones en los modelos de color, con el fin de limpiar las impurezas o pixeles pequeños resultantes de los procesos que puedan alterar el conteo final de células. La técnica de geometría de la imagen se utiliza en la etapa de separación por juntura de células para calcular su proyección de histograma.

## 2.11 Momentos estadísticos

Son formulaciones matemáticas, que se definen como parámetros estadísticos. Algunos de ellos cuales tienen amplia connotación dentro del estudio de curvas de distribución de frecuencias y más específicamente respecto del sesgo o media.

### • Varianza

La varianza o variancia de una de una variable aleatoria es una medida de dispersión definida como la esperanza del cuadrado de la desviación de dicha variable respecto a su media.

$$\begin{aligned}\sigma_n^2 &= \frac{1}{num} \sum_{i=1}^{num} (z_i - \bar{Z})^2 = \left( \frac{1}{num} \sum_{i=1}^{num} z_i^2 \right) - \bar{Z}^2 \\ &= \frac{1}{num^2} \sum_{i=1}^{num} \sum_{j>i} (z_i - z_j)^2\end{aligned}\tag{2.16}$$

donde:  $z_i$ : variable aleatoria

$\bar{Z}^2$ : segundo momento con respecto al origen

$num$ : número de datos

Su unidad de medida corresponde al cuadrado de la unidad de medida de la variable: por ejemplo, si la variable mide una distancia en metros, la varianza se expresa en metros al cuadrado. La varianza tiene como valor mínimo 0.

#### • Desviación estándar

La desviación típica o desviación estándar (denotada con el símbolo  $\sigma$  o  $s$ , dependiendo de la procedencia del conjunto de datos) es una medida de dispersión para variables de razón (variables cuantitativas o cantidades racionales) y de intervalo. Se define como la raíz cuadrada de la varianza de la variable.

$$\sigma_{num} = \sqrt{\sigma_{num}^2}\tag{2.17}$$

Donde  $num$  representa el número de datos en la muestra de la variable aleatoria. Para conocer con detalle un conjunto de datos, no basta con conocer las medidas de tendencia central, sino que se necesita conocer también la desviación típica que presentan los datos en su distribución respecto de la media aritmética.

## 2.12 Histograma

El histograma representa el número de píxeles que tienen colores en cada una de las listas fijas de rangos de colores que se extienden sobre el espacio de color de la imagen; es decir, el conjunto de todos los colores posibles. Se define mediante la ecuación 2.18.

$$H(b_k) = U_k \quad (2.18)$$

donde:

$b_k$  = es el k-esimo nivel de intensidad en el intervalo [0, L-1]

$L$  = número de niveles de gris posibles

$H(b_k)$  = cantidad de pixeles de  $b$

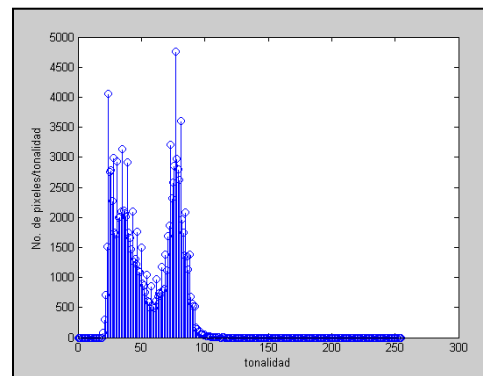
$U_k$  = es el número de pixeles en la imagen que su intensidad es  $b_k$

Cabe resaltar que el histograma no representa a una imagen específica, solo la cantidad de tonos de grises acumulado en toda la imagen. Pueden existir imágenes con igual distribución de tonos y no tener relación una de otra. En la Figura 17 se puede observar un ejemplo de histograma en base a una plaza con fondo.

Las medidas estadísticas presentadas anteriormente, juntamente con el histograma, se utilizaron en la etapa de umbralización por desviación estándar para separar las células proliferantes de las no proliferantes y del fondo de la imagen.



(a)



(b)

Figura 17. (a) Imagen en escala de grises (b) Histograma de la imagen  
Fuente: Material de clase sobre PDI-USMP



### 2.13 Proyecciones sobre el eje

Las proyecciones sobre el eje o integrales proyectivas son una técnica de reducción a sub-espacios lineales, donde cada valor de salida es la media aritmética de una fila o columna de píxeles de la entrada. En coordenadas rectangulares, se pueden proyectar de dos formas:

Con respecto al eje de las abscisas

$$P(y) = \sum_{x=0}^m f(x, y) \quad (2.19)$$

O con respecto al eje de las ordenadas

$$P(x) = \sum_{y=0}^n f(x, y) \quad (2.20)$$

Donde  $m$  es la cantidad de filas en la imagen y  $n$  es la cantidad de columnas en la imagen.

### 2.14 Mínimos locales o mínimos relativos

Los mínimos de una función, conocidos colectivamente como extremos de una función, son los valores más pequeños que toma una función en un punto situado dentro de una región en particular de la curva. Sea una función diferenciable  $g$ , el mínimo relativo cumplirá las siguientes condiciones:

$$g'(t) = 0 \quad (2.21)$$

$$g''(t) > 0 \quad (2.22)$$

Donde  $t$  es la variable independiente sobre la cual se analiza una función y el operador  $\varepsilon'$  representa la derivada de la función  $\varepsilon$ .

### 2.15 Segmentación

- **Etiquetado de imágenes:**

Vecindad de píxel:

Se denomina vecindad de un píxel a los píxeles que se encuentran contiguos a este. Para este caso se evalúa la 8-vecindad, en la Figura 18 se aprecian 2 casos de tipos de vecindad.

La representación de la 8-vecindad de un píxel  $p$  con coordenadas  $(x, y)$  se muestra a continuación:

$$\{(x + 1, y + 1), (x - 1, y - 1), (x - 1, y + 1), (x + 1, y - 1), \dots \dots (x + 1, y), (x - 1, y), (x, y + 1), (x, y - 1)\} \quad (2.23)$$

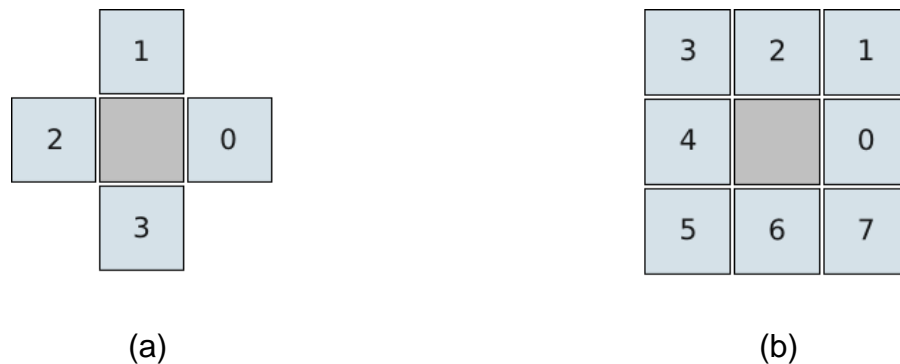


Figura 18. Demostración gráfica de vecindad (a) Vecindad 4 y (b) Vecindad 8  
Fuente: Nick Efford: *Digital Image Processing: A Practical Introduction Using Java*

Para poder identificar los píxeles que pertenecen a cada una de las regiones y, por tanto, asociar estas regiones a los objetos para extraer sus características, es necesario obtener las componentes conexas presentes en la imagen. Es obvio que, si en la imagen aparecen varias regiones, no se podrá obtener, por ejemplo, su área, hasta que estas no estén identificadas de forma individual. Este proceso se conoce como etiquetado y consiste en definitiva en asignar una etiqueta a cada uno de los píxeles que pertenecen a una componente conexa. La etiqueta será un valor numérico que permite identificar todos los píxeles que forman parte de una misma región. De esta forma, la imagen de etiquetas obtenida permitirá la extracción individualizada de características de cada una de las regiones conexas de la imagen.

La definición de un componente conexo dependerá del tipo de adyacencia utilizado. En este caso se utiliza la 8 vecindad. Por fines didácticos se explica el etiquetado con un ejemplo gráfico en la Figura 19.

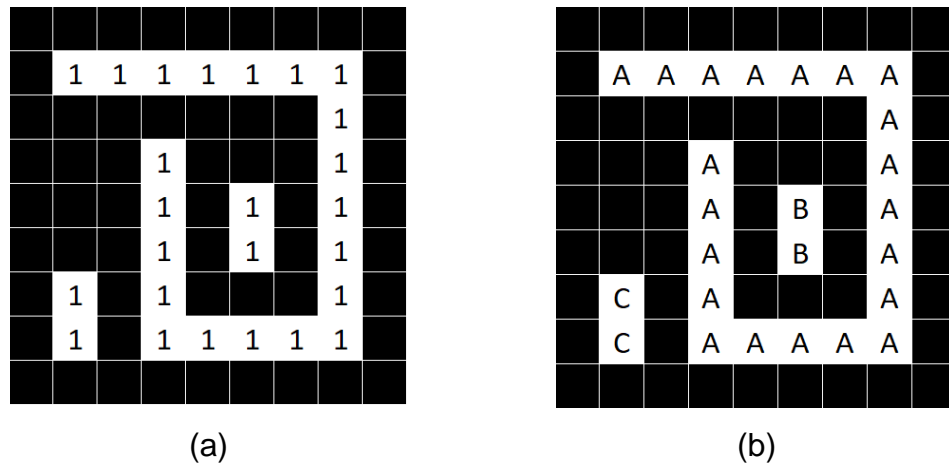


Figura 19. (a) Imagen blanco y negro antes de identificar regiones. (b) Equivalente de la imagen después de proceso de etiquetado  
Fuente: CodeProject.

## 2.16 Comentarios Finales

En este capítulo se ha explicado las bases médicas e ingenieriles sobre las que se trabajó el proyecto, abarcando desde las fases de la célula hasta inmunohistoquímica para entender qué es lo que se busca contabilizar y cómo es que se genera la muestra. También se cubrieron los conceptos de procesamiento digital de imágenes (PDI) que son usados a lo largo del capítulo 3 y 4 para el desarrollo del aplicativo de software.

## CAPÍTULO III

### DESARROLLO DE LA INVESTIGACIÓN

#### 3.1 Comentarios Iniciales

Este capítulo detallará los materiales utilizados en el desarrollo de la tesis, la metodología empleada y los resultados parciales.

#### 3.2 Fuentes de información e instrumentos utilizados

El proceso de tinción en el HNERM se hizo mediante el uso de la máquina de tinción automática Ventana *BenchMark XT*, especializada en estos procesos de inmunohistoquímica. El hospital se reservó el derecho de dar la información sobre los insumos químicos utilizados, solo especificó que utilizaron los provistos por la empresa distribuidora del equipo.



Figura 20. Equipo *BenchMark XT* de la marca Ventana  
Fuente: Empresa Roche

Para la etapa de digitalización de las muestras se utilizó un microscopio digital Marca OLYMPUS, modelo BX53, como el que se muestra en la Figura 21, utilizado en laboratorios médicos, debido a su afinidad para resaltar los colores mediante iluminación LED, teniendo métodos de observación de fluorescencia para excitar colores verdes y azules, especialmente para imágenes de tinciones tipo Hematoxicilina y Eosina, como corresponde a la tinción de KI-67.



Figura 21. Imagen de microscopio digital Olympus BX53  
Fuente: Olympus

### **3.3 Técnica de Recolección de Base de datos**

A continuación, se muestra el proceso de toma de muestra de tejido mamario y tinción. En la Figura 22 se puede apreciar gráficamente el proceso. La técnica de recolección está estandarizada para su uso en otras instituciones.

1. Recepción de la orden con las muestras en bloque de parafina.
2. Corte en láminas de 3 a 4  $\mu\text{m}$  del bloque usando un micrótomo y se lleva a un baño de flotación a 50 °C para deshacer la parafina, haciendo que se pegue el tejido a la lámina.

3. Secado de las láminas a temperatura ambiente y posterior deshidratación en una estufa a 48 °C.
4. Lavado en tres diferentes concentraciones de alcohol por 10 minutos cada una.
5. Se introduce en una máquina (PT LINK) que hará un baño con una solución estabilizadora de PH previamente calentada a 65°C y se deja reposar durante 20 minutos a 97 °C y se vuelve a dejar a que baje la temperatura en la máquina hasta los 65 °C.
6. Se lleva las láminas a un pequeño recipiente con solución *EnVision FLEX Wash Buffer*, que se encarga de llevar el PH de las láminas hasta un nivel de 7 durante 3-5 minutos.
7. Se procede a montar las láminas en racks especiales que se acomodan en la máquina *AUTOSTAINER Link 48* para comenzar el proceso de tinción; este dura aproximadamente 150 minutos (2h30min).
8. Una vez acabado el proceso de tinción se deja deshidratar en una estufa a 50 °C y de ahí se deja enfriar a temperatura ambiente.
9. Para finalizar, se procede a cubrir las láminas para su posterior observación.

El tiempo total de la tinción es un aproximado de 4h30min a 5 horas.

### **3.4 Metodología**

A continuación, se presenta el desarrollo de los pasos que se siguieron para el desarrollo del aplicativo de software.

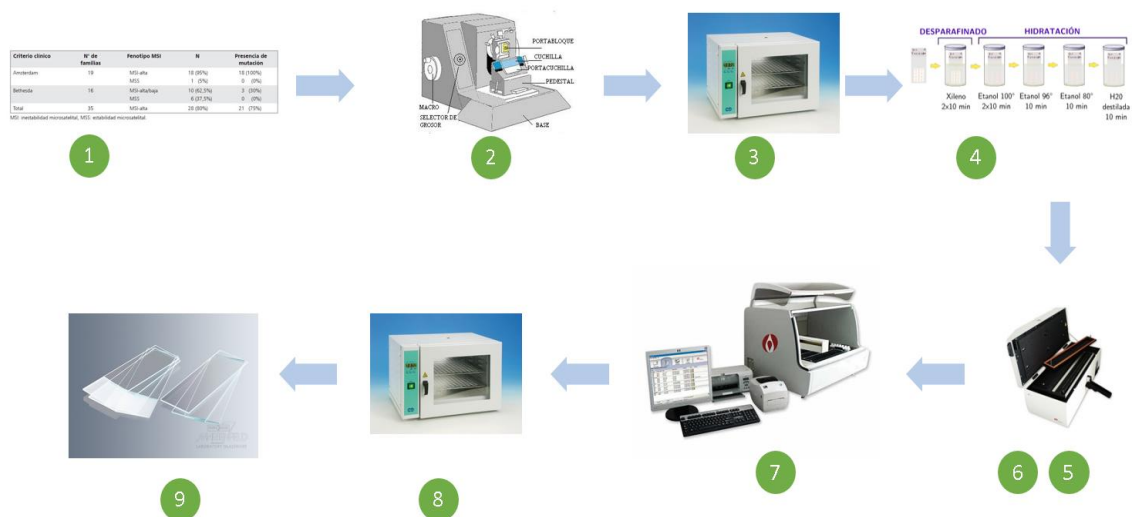


Figura 22. Esquema del proceso de tinción de láminas  
Elaboración: los autores

### 3.4.1 Adquisición de la imagen

Una vez acabado el proceso de tinción, la muestra pasa por un microscopio digital mencionado anteriormente con objetivo de aumento 40x (40 veces). En este caso la resolución de la cámara fue de 2748 filas por 3584 columnas y da como resultado a la imagen  $F(x, y)$ . El modelo de color como resultado de la adquisición es RGB con profundidad de 12bits/pixel por componente, a estas componentes o bandas se les llamará  $F_R(x, y)$ ,  $F_G(x, y)$ ,  $F_B(x, y)$  respectivamente.

### 3.4.2 Elección de la Región de Interés

No todas las regiones de la imagen  $F(x, y)$  son de interés para los patólogos, y esto se debe a las diferentes densidades de células. Es por ello que se da la posibilidad de elegir un área menor de la imagen  $F(x, y)$  en el que el especialista considere trascendente hacer el cálculo del índice de proliferación KI-67. La imagen recortada de  $F(x, y)$  tendrá resolución de  $M$  filas por  $N$  columnas (valores que varían según el recorte que aplica el especialista) y será llamada  $I(x, y)$ . El formato es también RGB con componentes primarias  $I_R(x, y)$ ,  $I_G(x, y)$  e  $I_B(x, y)$ .

Las Figuras 23 y 24, muestra una imagen original adquirida y una imagen recortada respectivamente.

### 3.4.3 Pre-Procesamiento

El pre-procesamiento apunta a mejorar la calidad de la imagen adquirida con miras a incrementar la eficiencia de los algoritmos de segmentación. Este proceso involucra los siguientes bloques de tratamiento de la imagen: corrección Gamma en el modelo de color HSV (*Hue*, *Saturation* y *Value*) y corrección Gamma en el modelo YCbCr.

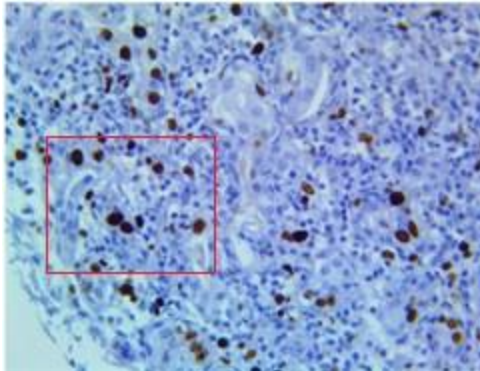


Figura 23. Imagen Original  
Elaboración: los autores

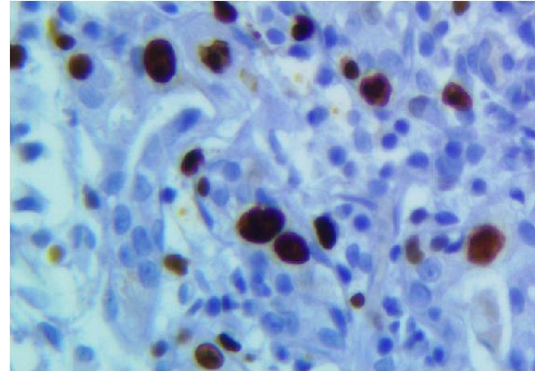


Figura 24. Imagen Recortada y acercada  
Elaboración: los autores

#### 3.4.3.1 Corrección Gamma en HSV

La imagen recortada expresada en sus componentes primarias  $I_R(x, y)$ ,  $I_G(x, y)$  e  $I_B(x, y)$  se transforma a la componente de valor  $I_V(x, y)$  del modelo de color HSV. Se utiliza la corrección gamma, visto en el capítulo 2.8, en la matriz  $I_V(x, y)$ , a fin de mejorar el contraste entre el fondo (epitelio) y las células proliferantes. Esto permitirá incrementar posteriormente la eficiencia de los algoritmos de segmentación. La transformación aplicada puede ser expresada como se indica en la ecuación 3.1.

$$I_V(x, y) = \max(I_R(x, y), I_G(x, y), I_B(x, y)) \quad (3.1)$$

donde la instrucción  $\max(a_1, a_2, a_3)$  devuelve el valor máximo entre los valores  $a_1$ ,  $a_2$  y  $a_3$ .

La corrección gamma aplicada a  $I_V(x, y)$  es expresada como:



$$I_{Vg}(x, y) = 255 \times \left( \frac{I_V(x, y)}{255} \right)^\gamma \quad (3.2)$$

donde  $\gamma = 0.9$

En la Figura 25a se muestra la imagen  $I(x, y)$  en formato RGB, en la Figura 25b, se muestra  $I_V(x, y)$  luego de aplicar la ecuación 3.1 a  $I(x, y)$  y en la Figura 25c, se tiene  $I_{Vg}(x, y)$  donde se aprecia cómo las células no proliferantes o de color azul adquieren una tonalidad parecida a la del fondo de la imagen, haciendo que se resalten las células proliferantes. Asimismo, en la Figura 25c, se hacen evidentes células parcialmente proliferantes, que son las células más oscuras comparadas al resto.

En paralelo al proceso anterior la imagen recortada  $I(x, y)$  se obtiene la componente  $C_b$  del modelo de color  $YC_bC_r$  visto en el capítulo 2.7, esto con el fin de poder apreciar los componentes en azul de la imagen para poder distinguir las células no proliferantes. La componente  $C_b$  se expresa como  $I_{Cb}(x, y)$ . Utilizando la ecuación 2.8, la transformación de modelo de color se expresa como:

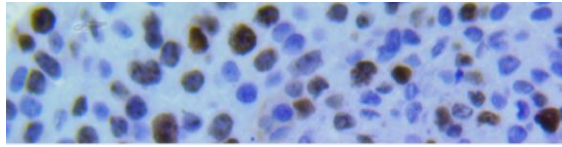
$$I_{Cb}(x, y) = 128 - 0.169 \times R - 0.331 \times G + 0.5 \times B \quad (3.3)$$

A la imagen  $I_{Cb}(x, y)$  se le aplica la corrección gamma con  $\gamma = 4$  (efecto exponencial). Esto permite desplazar las tonalidades de  $I_{Cb}(x, y)$  de las células no proliferantes respecto a las proliferantes. El desplazamiento en el espacio dimensional  $I_{Cb}(x, y)$  en su histograma es mostrado en la Figura 26, donde se observa un desplazamiento hacia la izquierda de las tonalidades, haciendo que se oscurezca la imagen. La imagen resultante del proceso es expresada como  $I_{Cb2}(x, y)$  y es mostrada en la Figura 27.

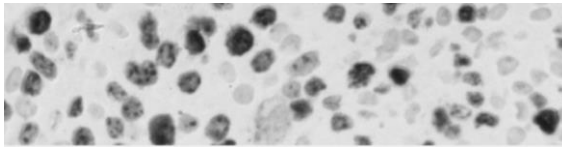
La corrección aplicada permite separar tonalidades y mejorar la segmentación futura de los dos tipos de células según su estado proliferativo.

### 3.4.4 Segmentación

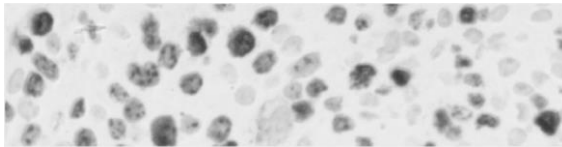
En este proceso se segmentan las células proliferantes y no proliferantes. La segmentación es realizada a partir de las imágenes corregidas  $I_{Vg}(x, y)$ ,  $I_{Cb2}(x, y)$  y de la imagen recortada sin procesar  $I(x, y)$ .



(a)



(b)



(c)

Figura 25 (a) Imagen  $I(x, y)$ , (b) Componente V de la transformación al modelo de color HSV de  $I(x, y)$ , (c) Imagen  $I_{Vg}(x, y)$   
Elaboración: los autores

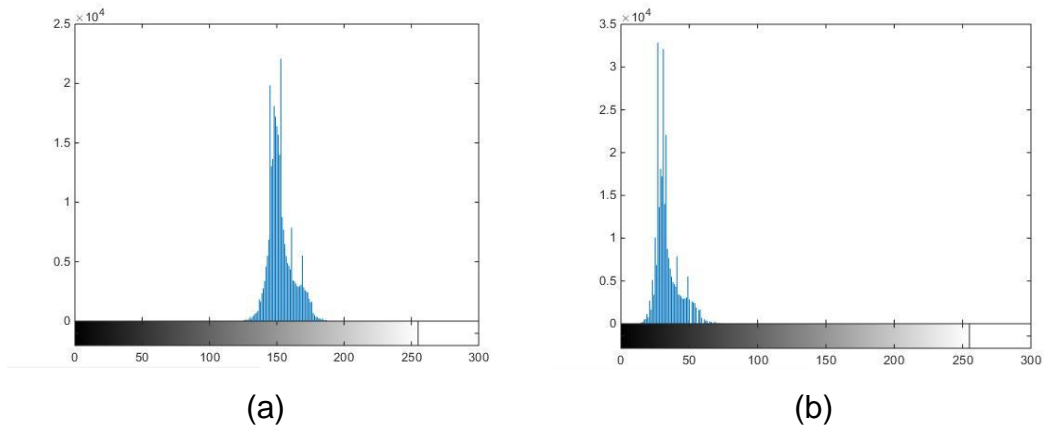


Figura 26. (a) Espacio dimensional  $I_{Cb}(x, y)$  original (histograma) y (b) Espacio dimensional desplazado  $I_{Cb2}(x, y)$  (histograma).  
Elaboración: los autores

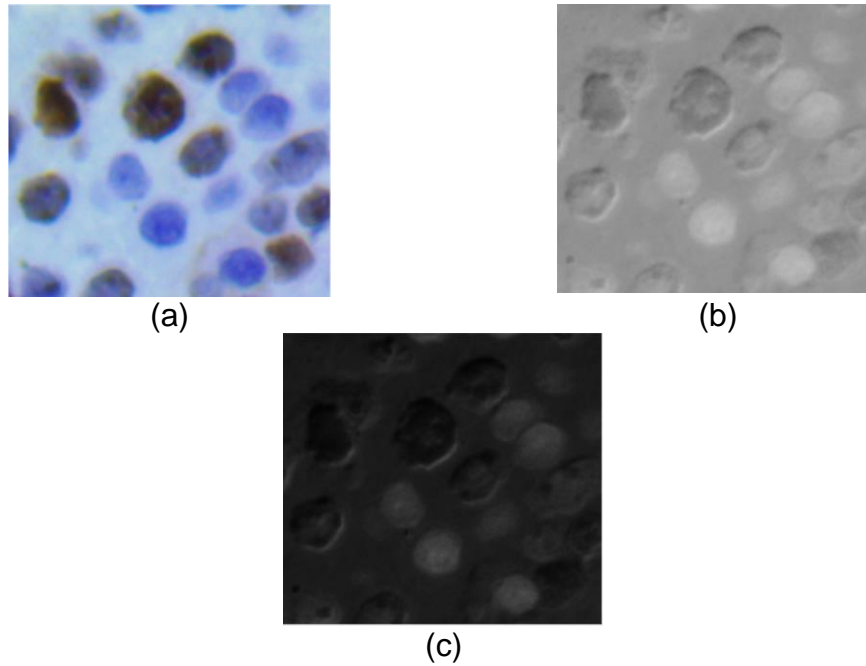


Figura 27. (a) Imagen original RGB, (b) Imagen  $I_{cb}(x, y)$  y (c) imagen  $I_{cb2}(x, y)$  resultante de la aplicación de la corrección gamma  
Elaboración: los autores

#### 3.4.4.1 Segmentación de Células proliferantes

Para separar el fondo (tejido celular que no es de interés para el diagnóstico) de las células proliferantes, se decide trabajar con umbrales locales adaptivos, los cuales se hallan a partir de la función de densidad de probabilidad (obtenida por histograma) y de medidas estadísticas de primer orden (media y varianza), vistos en el capítulo 2.11. El algoritmo se describe a continuación:

Paso 1:

Se procede a segmentar la imagen  $I_{vg}(x, y)$  (ver en la Figura 28) en  $p$  segmentos. Los segmentos de imagen tendrán resolución espacial de  $M$  filas por  $\frac{N}{p}$  columnas. Los segmentos  $I_i(x, y)$  de imagen  $I_{vg}(x, y)$  se denotan como:

$$I_i(x, y) = I_{vg}\left(x, y + i\left(\frac{N}{p}\right)\right), \frac{N}{p} \in \mathbb{Z} \quad (3.4)$$

donde  $x = 0, 1, \dots, M - 1$ ;  $y = 0, 1, \dots, \frac{N}{p} - 1$  y  $i = 0, 1, \dots, p - 1$ .

Paso 2:

A cada segmento  $i$  se le determina su histograma  $h_i(r)$ . En este caso  $r$  representa a cada una de las tonalidades de 0 a 255.

Paso 3:

Se obtiene el valor medio  $\bar{x}_i$  de cada segmento  $i$  normalizado

$$\bar{x}_i = \frac{p}{MN} \sum_{r=0}^{255} h_i(r) \cdot r \quad (3.5)$$

Paso 4:

Se obtiene la desviación estándar  $\sigma_i$  para cada segmento  $i$ :

$$\sigma_i = \sqrt{\frac{p}{MN} \sum_{r=0}^{255} (r - \bar{x}_i)^2 h_i(r) \cdot r} \quad (3.6)$$

Paso 5:

El umbral  $u_i$  que se utilizará de forma local para cada segmento de la imagen con la finalidad de separar el tejido celular de la región de células proliferantes será la resta de la media y la desviación estándar de cada segmento.

$$u_i = \bar{x}_i - \sigma_i \quad (3.7)$$

Paso 6:

Se umbraliza el segmento  $i$ :

$$I_{iium}(x, y) = \begin{cases} I_i(x, y) & ; \quad I_i(x, y) \geq u_i \\ 0 & ; \quad I_i(x, y) < u_i \end{cases} \quad (3.8)$$

Finalmente, cada segmento de la imagen umbralizada  $I_{ium}(x, y)$  es posicionado en el orden original a fin de reconstruir una imagen con resolución espacial de  $M$  filas y  $N$  columnas, a la cual se le llamara  $I_{pro}(x, y)$  y que contiene las regiones de células proliferantes como muestra en la Figura 29a y la Figura 30b muestra la imagen obtenida luego del proceso de umbralización.

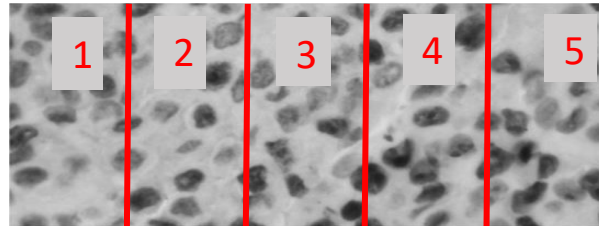
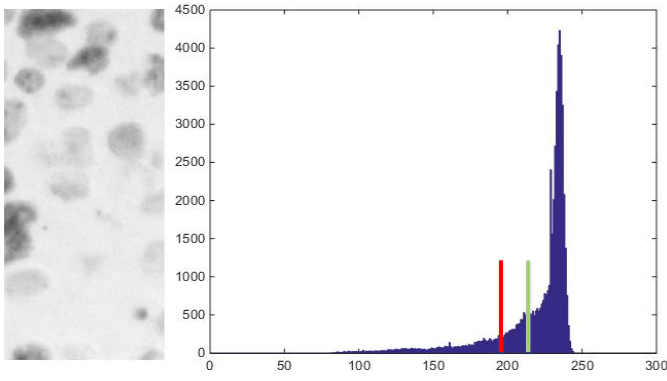


Figura 28. Imagen  $I_{vg}(x, y)$  a ser sometida al proceso de segmentación de células proliferantes, en este ejemplo, en  $p=5$   
Elaboración: los autores

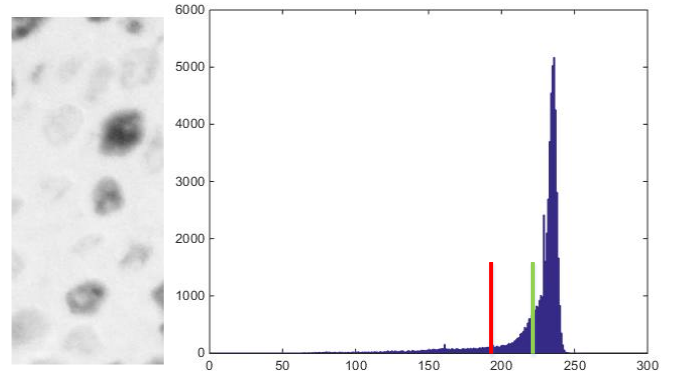
- **Desempeño de segmentación de células proliferantes**

En esta sección se muestra de forma visual el desempeño del filtro adaptivo local para la clasificación de células proliferantes. Es importante mencionar que para los fines de clasificación celular es necesaria una correcta segmentación de la forma completa de una célula y de cuanta área de cada una ha reaccionado a la proteína KI-67.

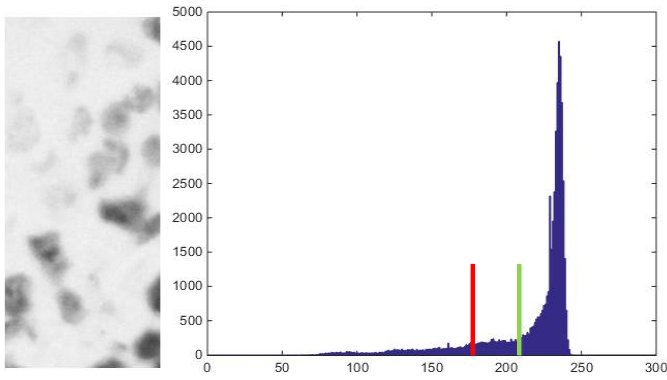
El fin de este proceso es segmentar de acuerdo con las estadísticas de cada porción de imagen, ya que la distribución en la ubicación de células proliferantes y no proliferantes es aleatoria. En la Figura 31a se muestra la imagen de prueba sobre la cual se desea encontrar las áreas que reaccionaron a la proteína KI-67. Y en la Figura 31b, c, d, e y f diferentes valores de división de la imagen para observar su desempeño.



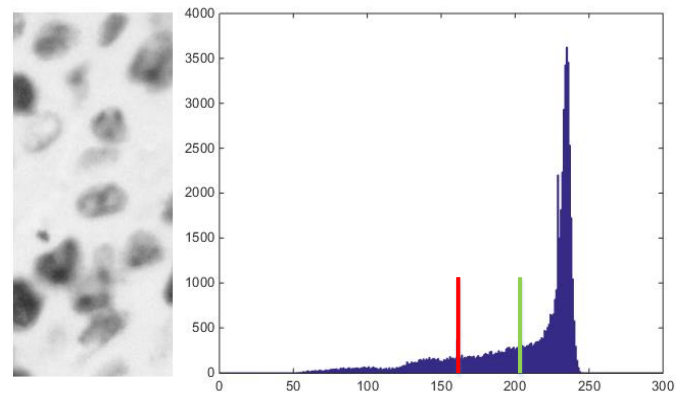
(a)



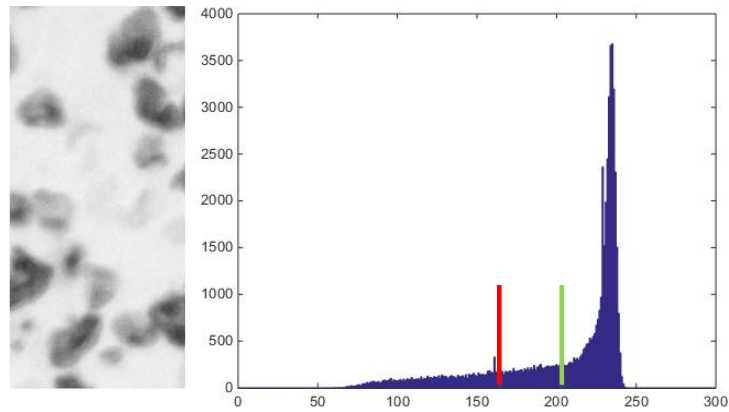
(b)



(c)



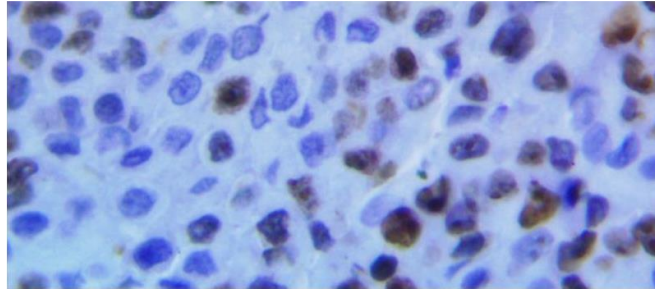
(d)



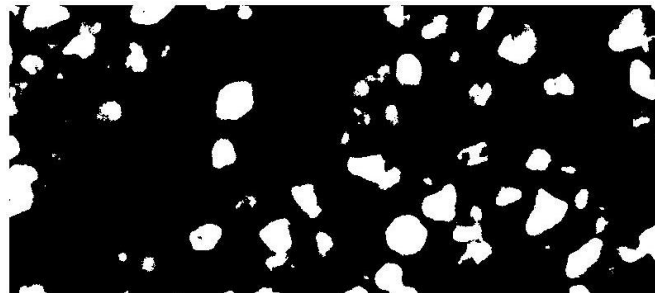
(e)

Figura 29. Procesamiento de los  $i$  segmentos. Las barras verdes corresponden al valor medio de cada segmento y las barras rojas al valor del umbral elegido en cada segmento (a) Segmento 1 de  $I_{Vg}(x,y)$  (b) Segmento 2 de  $I_{Vg}(x,y)$  (c) Segmento 3 de  $I_{Vg}(x,y)$  (d) Segmento 4 de  $I_{Vg}(x,y)$  (e) Segmento 5 de  $I_{Vg}(x,y)$

Elaboración: los autores



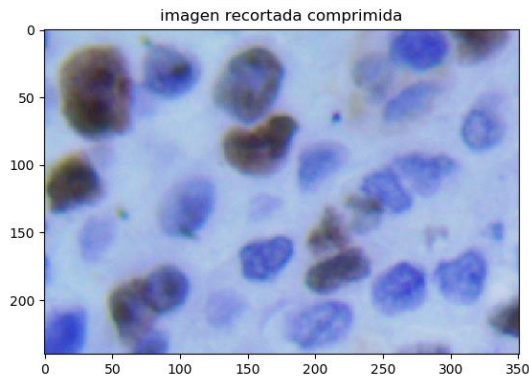
(a)



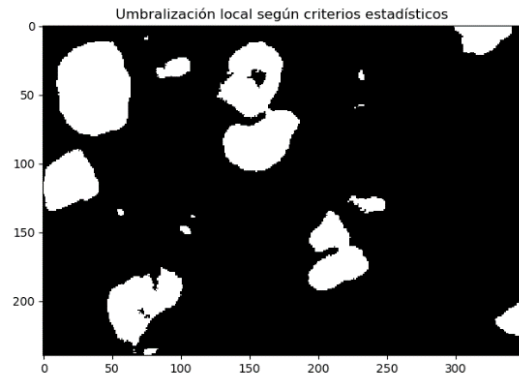
(b)

Figura 30. Umbralización de células proliferantes. (a) Imagen  $I_{vg}(x, y)$  (b) Imagen binaria  $I_{pro}(x, y)$  resultante del proceso de segmentación de células proliferantes  
Elaboración: los autores

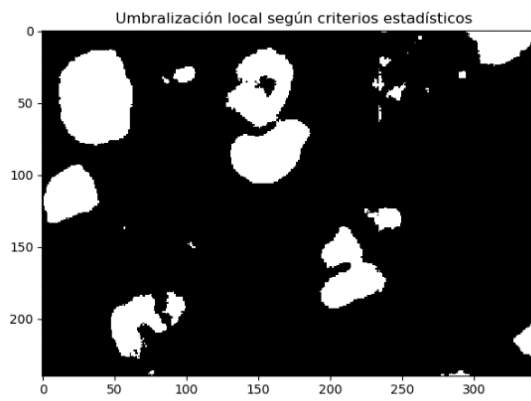
Se observa que para un solo segmento de imagen (que es lo mismo que utilizar un umbral global y no local) las regiones correspondientes a células proliferantes son las más grandes, tanto que abarcan zonas de células no proliferantes y, por lo tanto, no son resultados correctos. Para una división de dos y cuatro subsegmentos de imagen se obtuvieron regiones de células proliferantes más definidas; sin embargo, existen problemas de células parcialmente segmentadas debido a los cambios estadísticos de cada subsegmento. Para diez subdivisiones el problema se invierte, existen demasiados *objetos* y ruido. Por lo tanto, la cantidad de subdivisiones óptima que se eligió para esta tesis fue de cinco. Con esta cantidad de subdivisiones se obtuvieron los mejores resultados de tamaño y forma de células proliferantes.



(a)



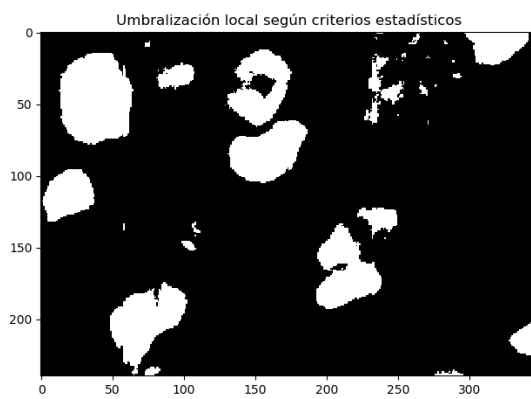
(b)



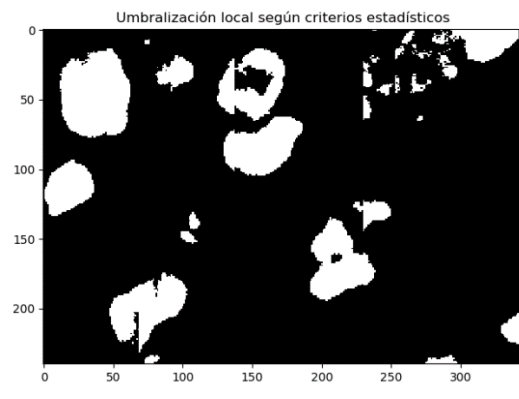
(c)



(d)



(e)



(f)

Figura 31. Resultados de la segmentación de células proliferantes con umbrales adaptivos locales. (a) Imagen de prueba (b) Sin subdivisión (c) Dos subdivisiones (d) Cuatro subdivisiones (e) Cinco subdivisiones (f) Diez subdivisiones  
Elaboración: los autores



### 3.4.4.2 Segmentación de Células no proliferantes:

Para la segunda etapa se realizaron algoritmos que detecten las células no proliferantes o células azules, esto con el fin de luego poder cuantificarlas y brindar un diagnóstico útil.

#### a) Segmentación por Filtrado de Sobel y Morfología

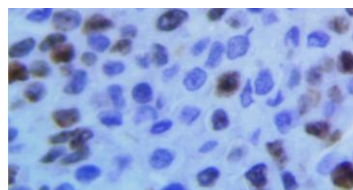
Luego de identificar las células proliferantes es necesario encontrar las células no proliferantes a fin de hallar el porcentaje de células con actividad proliferativa. Para este proceso se utiliza la imagen  $I_{cb2}(x, y)$  obtenida en la etapa de pre-procesamiento. El procedimiento es detallado a continuación:

Paso 1:

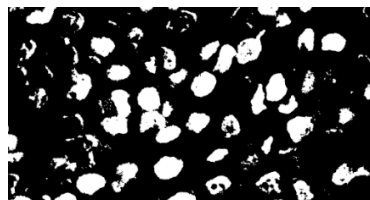
Se aplica un filtro por umbralización a  $I_{cb2}(x, y)$  a fin de segmentar las células no proliferantes. La umbralización puede ser expresada como:

$$I_{cb3}(x, y) = \begin{cases} 0 & ; \quad I_{cb2}(x, y) < 41 \\ 1 & ; \quad \text{otro caso} \end{cases} \quad (3.9)$$

El umbral fue obtenido a partir del estudio y evaluación de histogramas de imágenes  $I_{cb2}(x, y)$  correspondientes a varias muestras adquiridas. El resultado del proceso es mostrado en la Figura 32.



(a)



(b)

Figura 32. (a) Imagen original. (b) Imagen resultante  $I_{cb3}(x, y)$   
Elaboración: los autores

Paso 2:

A continuación, se aplica el filtro Sobel, del capítulo 2.9, para detección de bordes en imágenes binarias. En este proceso se convoluciona  $I_{Cb3}(x, y)$  con cada una de las máscaras de Sobel (horizontal y vertical) las cuales son definidas matricialmente como:

$$M_h = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$$

$$M_v = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$$

La imagen resultante se define como:

$$I_{Sobel}(x, y) = \begin{cases} 255 & ; I'_{Sobel}(x, y) > u_s \\ 0 & ; \text{otra forma} \end{cases} \quad (3.10)$$

donde:

$$I'_{Sobel}(x, y) = |I_{Cb3}(x, y) * M_h(x, y)| + |I_{Cb3}(x, y) * M_v(x, y)| \quad (3.11)$$

Como  $I'_{Sobel}(x, y)$  no es una imagen binaria, se binariza utilizando un umbral  $u_s$  que tiene un valor pequeño, en este caso igual a 10.

Paso 3:

$I_{Sobel}(x, y)$  es sometida a la operación morfológica de “cierre” la cual se expresa a través de la siguiente ecuación matricial:

$$I_{clo} = ((I_{Sobel} \oplus B_v) \oplus B_h) \ominus B_v) \ominus B_h \quad (3.12)$$

donde “ $\oplus$ ” indica la operación de dilatación y “ $\ominus$ ” la operación de erosión.

Los elementos estructurantes utilizados son:

$$B_h = [1 \ 1 \ 1]$$

$$\mathbf{B}_v = \begin{bmatrix} \mathbf{1} \\ \mathbf{1} \\ \mathbf{1} \end{bmatrix}$$

A fin de reformar las células detectadas y no distorsionar el resultado, se aplica llenado de huecos en las células a la matriz " $\mathbf{I}_{clo}$ ". Esta operación se obtiene aplicando:

$$\mathbf{I}_{clo}^k = (\mathbf{I}_{clo}^{k-1} \oplus \mathbf{D}_p) \cap \mathbf{I}_{clo}^c \quad , \quad k = 1, 2, 3, \dots, k_0 \quad (3.13)$$

donde  $k$ : indica el número de la iteración de relleno

$\cap$ : indica la operación de intersección.

$\mathbf{I}^c$ : expresa el complemento de la matriz  $\mathbf{I}_{clo}$ .

La operación de llenado es hecha a través de  $k_0$  iteraciones hasta que se satisfaga la condición  $\mathbf{I}_{clo}^k = \mathbf{I}_{clo}^{k-1}$ .

El elemento estructurante  $\mathbf{D}_p$  utilizado puede ser expresado como:

$$\mathbf{D}_p = \begin{bmatrix} \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{1} & \mathbf{1} & \mathbf{1} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} \end{bmatrix}$$

Finalmente, el resultado de esta etapa es colocado en la matriz  $\mathbf{I}_{Sobelf}$ :

$$\mathbf{I}_{Sobelf} = \mathbf{I}_{clo}^{K_0} \quad (3.14)$$

### **b) Segmentación por umbralización de componentes primarias:**

En esta etapa se aplica umbralización a las componentes primarias RGB, visto en los modelos de color del capítulo 2.7; para separar las células proliferantes y no proliferantes del epitelio (fondo). Los umbrales utilizados fueron obtenidos a partir de pruebas experimentales realizadas con imágenes de diferentes muestras de tejido. El procedimiento ayuda a eliminar la región del epitelio y resaltar las células de interés. A continuación, se detallan los pasos que conforman el procedimiento indicado:

Paso 1:

Se umbralizan las componentes primarias de la imagen original  $I(x, y)$  para obtener tres matrices:

$$F_R(x, y) = \begin{cases} 0 & , \quad I_R(x, y) < 128 \\ 255 & , \quad I_R(x, y) \geq 128 \vee I_G(x, y) \geq 128 \end{cases} \quad (3.15)$$

$$F_G(x, y) = \begin{cases} 0 & , \quad I_G(x, y) < 128 \\ 255 & , \quad \text{otra forma} \end{cases} \quad (3.16)$$

$$F_B(x, y) = \begin{cases} 0 & , \quad I_B(x, y) < 128 \\ 255 & , \quad I_B(x, y) \geq 128 \vee I_G(x, y) \geq 128 \end{cases} \quad (3.17)$$

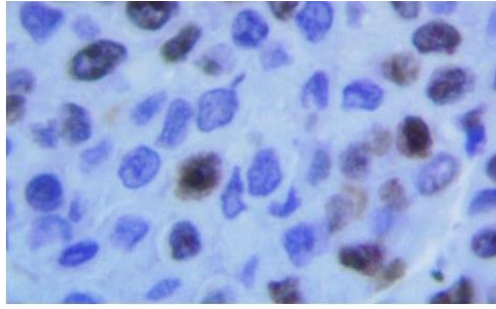
Paso 2:

Se genera una matriz  $I_{cel}(x, y)$  con la misma resolución espacial de  $I(x, y)$  que contendrá regiones de células proliferantes y no proliferantes.

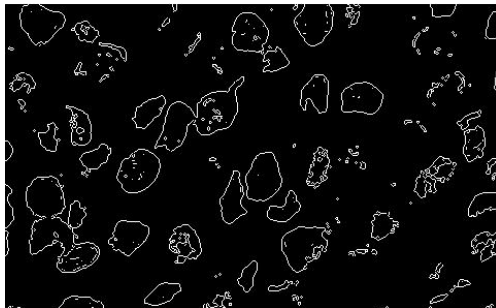
Para ello,  $I_{cel}(x, y)$  se obtiene partir de la siguiente condición:

$$I_{cel}(x, y) = \begin{cases} 255 & , \quad F_R(x, y) = 255 \wedge F_G(x, y) = 255 \wedge F_B(x, y) = 255 \\ 0 & , \quad \text{otra forma} \end{cases} \quad (3.18)$$

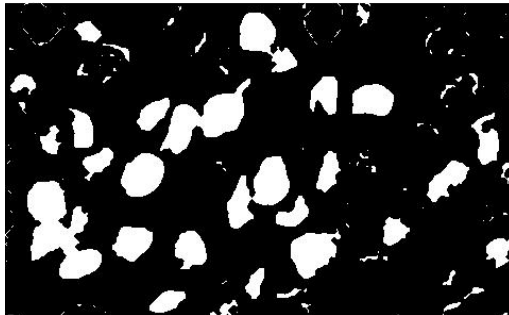
$I_{cel}(x, y)$  constituye finalmente una plantilla binaria, donde las regiones con valor igual a 255 corresponden a células y las regiones con valor igual a 0 es fondo o epitelio.



(a)



(b)



(c)

Figura 33. (a) Imagen original (b) Imagen resultante de aplicar filtro Sobel (c) Imagen final  $I_{Sobel}(x, y)$  después de llenado de huecos  
Elaboración: los autores

Paso 3:

Finalmente, la matriz de células no proliferantes es obtenida a partir de la siguiente expresión:

$$I_{nopro}(x, y) = \begin{cases} 1 & (I_{cel}(x, y) = 255 \vee I_{sobelf}(x, y) = 255) \dots \\ & \dots \wedge I_{pro}(x, y) \neq 1 \\ 0 & \text{otra forma} \end{cases} \quad (3.19)$$

Los resultados de este algoritmo se muestran en la Figura 34.

### 3.4.5 Separación por juntura de células

Unificando las áreas de células proliferantes y no proliferantes se obtiene la imagen binaria que contiene los dos tipos de células segmentadas, utilizando la técnica de etiquetado de imágenes del capítulo 2.7. Esta imagen es obtenida de la siguiente forma:

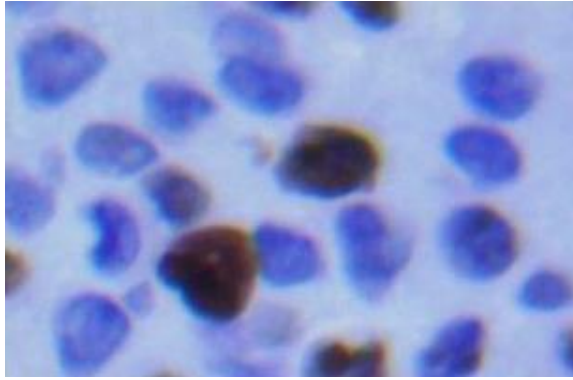
$$I_{total}(x, y) = \begin{cases} 1 & I_{pro}(x, y) = 1 \vee I_{nopro}(x, y) = 1 \\ 0 & \text{otra forma} \end{cases} \quad (3.20)$$

Sin embargo, uno de los problemas que aún se presenta al momento de contabilizar las células detectadas, es la identificación de objetos formados por dos o más células que se encuentran juntas o superpuestas.

En ese sentido, tomando en cuenta que para una adquisición con aumento de 40x el tamaño promedio de una célula se encuentra entre 200 y 1500 pixeles, se puede considerar que un objeto de  $I_{total}(x, y)$  cuyo tamaño se encuentra por encima de 1500 pixeles está conformado por más de una célula.

Esto definitivamente afectaría el índice de proliferación, ya que en principio se considera que un objeto etiquetado está conformado por una única célula.

Para solucionar este inconveniente se propone el siguiente algoritmo de separación de células, a fin de aminorar el error de contabilización.



(a)



(b)



(c)

Figura 34. (a) Imagen original recortada, (b) Imagen  $I_{cel}(x, y)$  y (c) Imagen  $I_{nopro}(x, y)$   
Elaboración: los autores

Paso 1:

Sea  $I_j(x_j, y_j)$  un segmento binario de  $I_{total}(x, y)$  que contiene exactamente al objeto  $j$  conformado por más de 1500 píxeles. En este caso se tiene  $x_j = 0, 1, 2, \dots, m_j - 1$  y  $y_j = 1, 2, \dots, n_j - 1$ ; es decir el tamaño del segmento es  $m_j \times n_j$  píxeles.

Paso 2:

Se realiza la rotación de la imagen para diferentes ángulos  $\theta$  ( $\theta_0 = 0^\circ, \theta_1 = 15^\circ, \theta_2 = 30^\circ, \theta_3 = 45^\circ, \theta_4 = 60^\circ, \theta_5 = 75^\circ, \theta_6 = 90^\circ$ ). En este caso el eje de rotación pasa por el centro de la imagen  $I_j(x_j, y_j)$ . El intervalo de  $15^\circ$  entre los ángulos de rotación fue elegido en base a los mejores resultados obtenidos en varias pruebas de separación de células, el sentido de giro es indistinto.

Cada giro genera una imagen que puede ser expresada como  $I_j^\theta(x'_j, y'_j)$ , donde " $\theta$ " indica el ángulo de giro. Las dimensiones espaciales de cada  $I_j^\theta(x'_j, y'_j)$  es expresada como  $m_j^\theta$  filas por  $n_j^\theta$  columnas.

Paso 3:

Se obtiene la proyección horizontal  $I_j^\theta(x'_j, y'_j)$ :

$$P_j^\theta(y'_j) = \sum_{x'_j=0}^{m_j^\theta-1} I_j^\theta(x'_j, y'_j) \quad (3.21)$$

Paso 4:

Solo es importante evaluar la zona central de la protección para cada  $P_j^\theta(y'_j)$ , es por ello que se eliminan las zonas de los extremos del vector:

$$\hat{P}_j^\theta(y'_j) = \begin{cases} P_j^\theta(y'_j) & ; \text{ si } y'_j > 0.3 \cdot n_j^\theta \vee y'_j < 0.7 \cdot n_j^\theta \\ 0 & ; \text{ otro caso} \end{cases} \quad (3.22)$$



Paso 5:

Se obtiene el valor  $\hat{P}_j^\theta(y_0(\theta, j))$ , el cual es definido como el valor mínimo de la proyección  $\hat{P}_j^\theta(y'_j)$ . En este caso " $y_0(\theta, j)$ " es la posición donde  $\hat{P}_j^\theta(y'_j)$  es mínima.

Paso 6:

Luego se conforma el vector  $\mathbf{W}_j$ :

$$\mathbf{W}_j = \begin{bmatrix} \hat{P}_j^{\theta_0}(y_0(\theta_0, j)) \\ \hat{P}_j^{\theta_1}(y_0(\theta_1, j)) \\ \vdots \\ \hat{P}_j^{\theta_5}(y_0(\theta_5, j)) \end{bmatrix} \quad (3.23)$$

Paso 7:

Luego se obtiene el valor mínimo del vector  $\mathbf{W}_j$ , el cual es definido como  $\hat{P}_j^{\theta_{k_0}}(y_0(\theta_{k_0}, j))$  donde  $\theta_{k_0}$  es el ángulo correspondiente al mínimo valor de  $\mathbf{W}_j$ .

Paso 8:

Luego se obtiene  $Z'_j(x'_j, y'_j)$  a partir de  $I_j^{\theta_{k_0}}(x'_j, y'_j)$ :

$$Z'_j(x'_j, y'_j) = \begin{cases} 0 & ; \quad i \quad y'_j = y_0(\theta_{k_0}, j) \\ I_j^{\theta_{k_0}}(x'_j, y'_j) & ; \quad \text{otro caso} \end{cases} \quad (3.24)$$

Paso 9:

Luego se aplica la rotación de  $-\theta_{k_0}$  sobre  $I_j^{\theta_{k_0}}(x'_j, y'_j)$  a fin de reestablecer la orientación original (aunque con distorsión debido a la interpolación aplicada en el proceso de rotación). Después de la rotación se obtiene la imagen  $Z_j(x'_j, y'_j)$ .

Paso 10:

Finalmente se inserta el bloque de imagen  $Z_j(x'_j, y'_j)$  en  $I_{total}(x, y)$  (en la región de donde se extrajo  $I_j(x_j, y_j)$ ).

$$I_{total}(xa_j + x'_j, ya_j + y'_j) = Z_j(x'_j, y'_j) \quad (3.25)$$

donde " $xa_j$ " e " $ya_j$ " son las coordenadas del primer pixel del segmento extraído  $I_j(x_j, y_j)$ .

La Figura 35a muestra regiones originales de  $I_{total}(x, y)$  que albergan objetos conformados por más de una célula. La Figura 35b muestra los resultados obtenidos después de la aplicación del algoritmo de separación de células.

- **Resultados de Separación de células superpuestas**

La separación de células superpuestas es un proceso trascendental en el conteo y clasificación de células. En la Figura 36 se muestran 4 imágenes, cada una muestra dos células que fueron segmentadas como un solo objeto. Como el índice de proliferación es igual a la razón de las células proliferantes sobre el total de células, el problema de las células cercanas o superpuestas va a ocasionar un conteo de células errado y, por ende, un índice de proliferación del KI-67 errado.

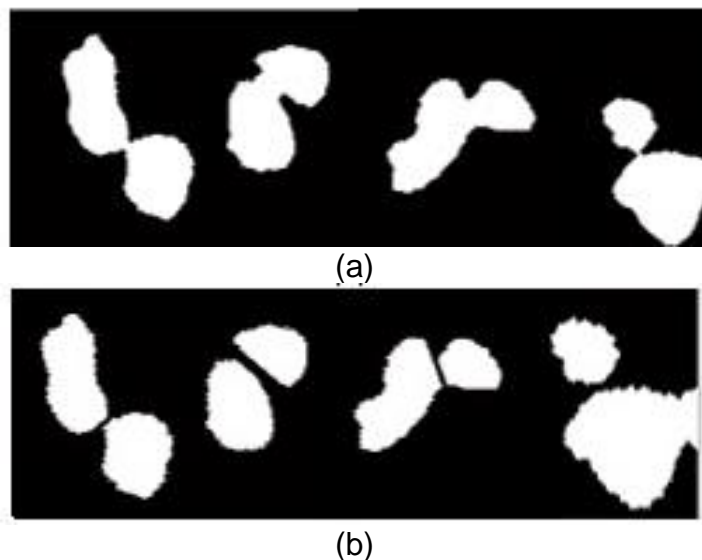


Figura 35. (a) Muestras con juntura y (b) muestras después de proceso de separación  
Elaboración: los autores

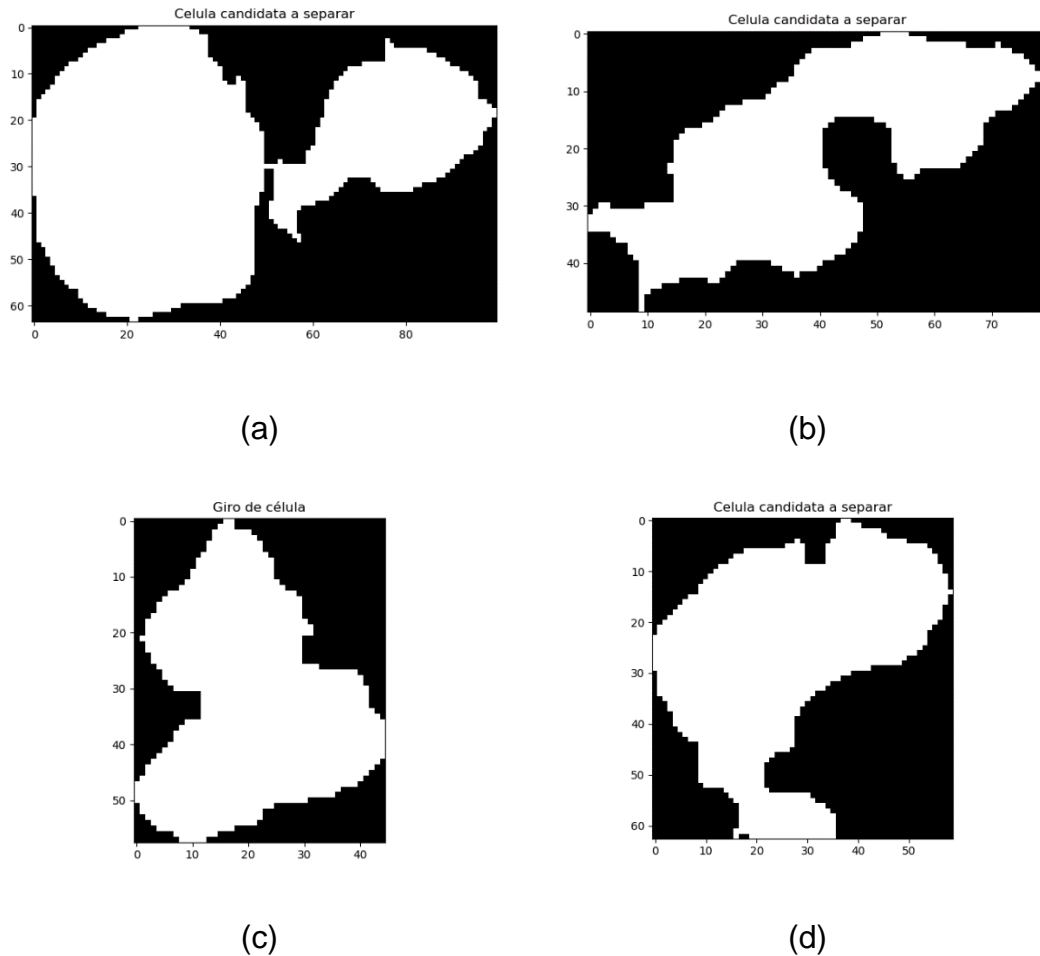


Figura 36. Células etiquetadas incorrectamente debido a la superposición y cercanía de células (a) Células cercanas (b) Células superpuestas (c) Células cercanas (d) Células superpuestas  
Elaboración: los autores

El algoritmo propuesto consta de giros sucesivos y la localización del mínimo global de las proyecciones con respecto al eje de las abscisas. Este proceso queda mejor explicado observando la Figura 37 donde se ven todos los giros del objeto etiquetado en la Figura 37c. En la Figura 38 se observan las proyecciones con el eje de las abscisas de cada giro. Luego, el algoritmo indica que el giro óptimo para la separación de las células es aquel en la que se halle el mínimo valor entre los mínimos globales de todas las proyecciones de los giros de una etiqueta. En este caso, el mínimo valor entre los mínimos globales se encuentra en la Figura 38c.

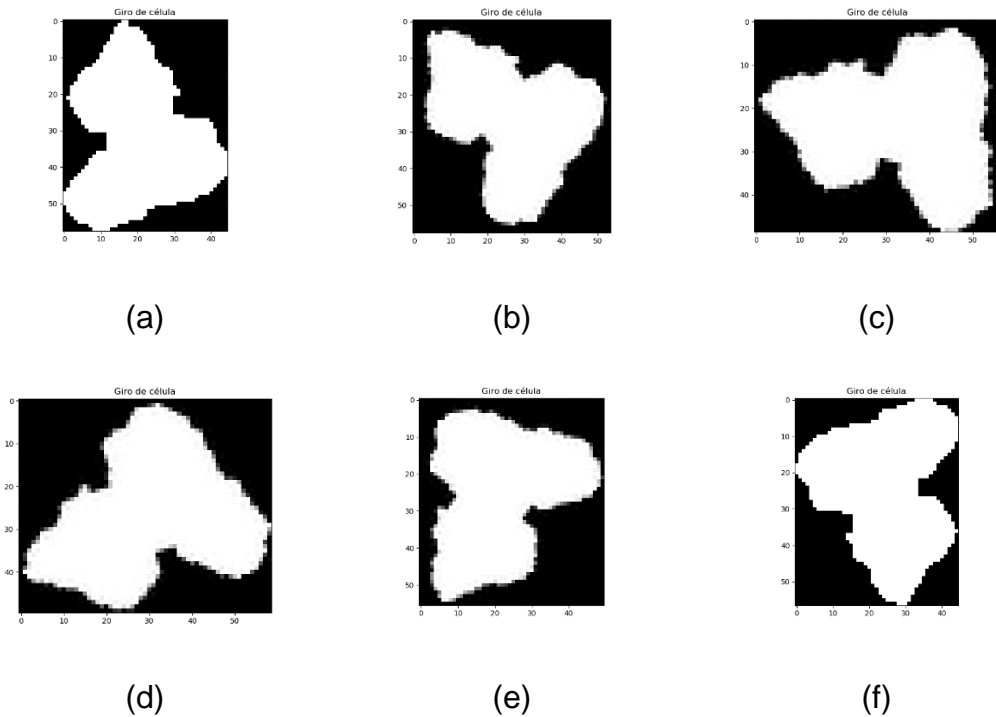


Figura 37. Giros de un objeto para separarlas adecuadamente (a) Primer giro (b) Segundo giro (c) Tercer giro (d) Cuarto giro (e) Quinto giro (f) Sexto giro  
Elaboración: los autores

Finalmente, el corte de las células se realiza en la posición correspondiente al mínimo local de la imagen y que corresponde a la célula girada en donde se encuentra el valor mínimo de los mínimos locales. Los resultados se muestran en la Figura 38. Se debe tomar en cuenta que la transformación morfológica de giro puede introducir pequeñas deformaciones en nuestro objeto segmentado.

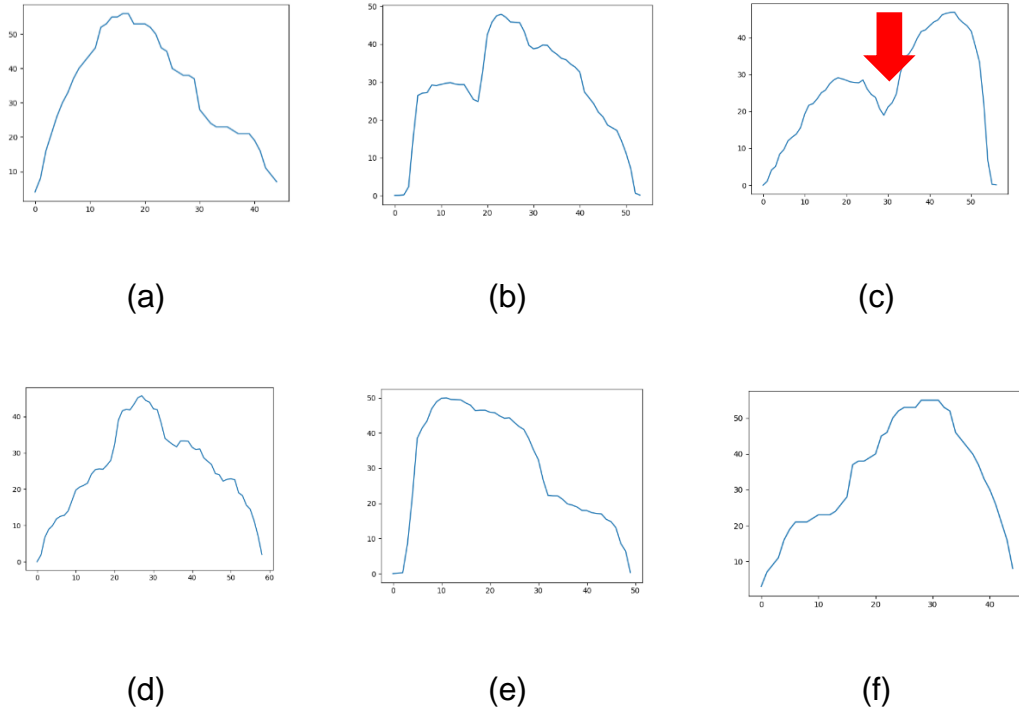


Figura 38. Proyecciones con respecto al eje de las abscisas para cada giro (a) Proyección del primer giro (b) Proyección del segundo giro (c) Proyección del tercer giro, la flecha roja indica el mínimo global encontrado (d) Proyección del cuarto giro (e) Proyección del quinto giro (f) Proyección del sexto giro  
Elaboración: los autores



Figura 39. Resultados de la separación de células (a) Células juntas (b) Células separadas por el algoritmo propuesto  
Elaboración: los autores

Como se observa en la Figura 37c, el giro óptimo sucede cuando la región más angosta por donde se unen las dos células es perpendicular al eje x. Si bien es cierto que una mayor cantidad de giros dará mejor fineza en el corte a realizar, este es un proceso computacionalmente

costoso y una cantidad de giros grande causará una demora en la entrega de resultados del software. Por ello, se hicieron pruebas con diferentes cantidades de giros (para hallar la mejor relación costo computacional-desempeño) y los resultados se muestran en la Figura 40.

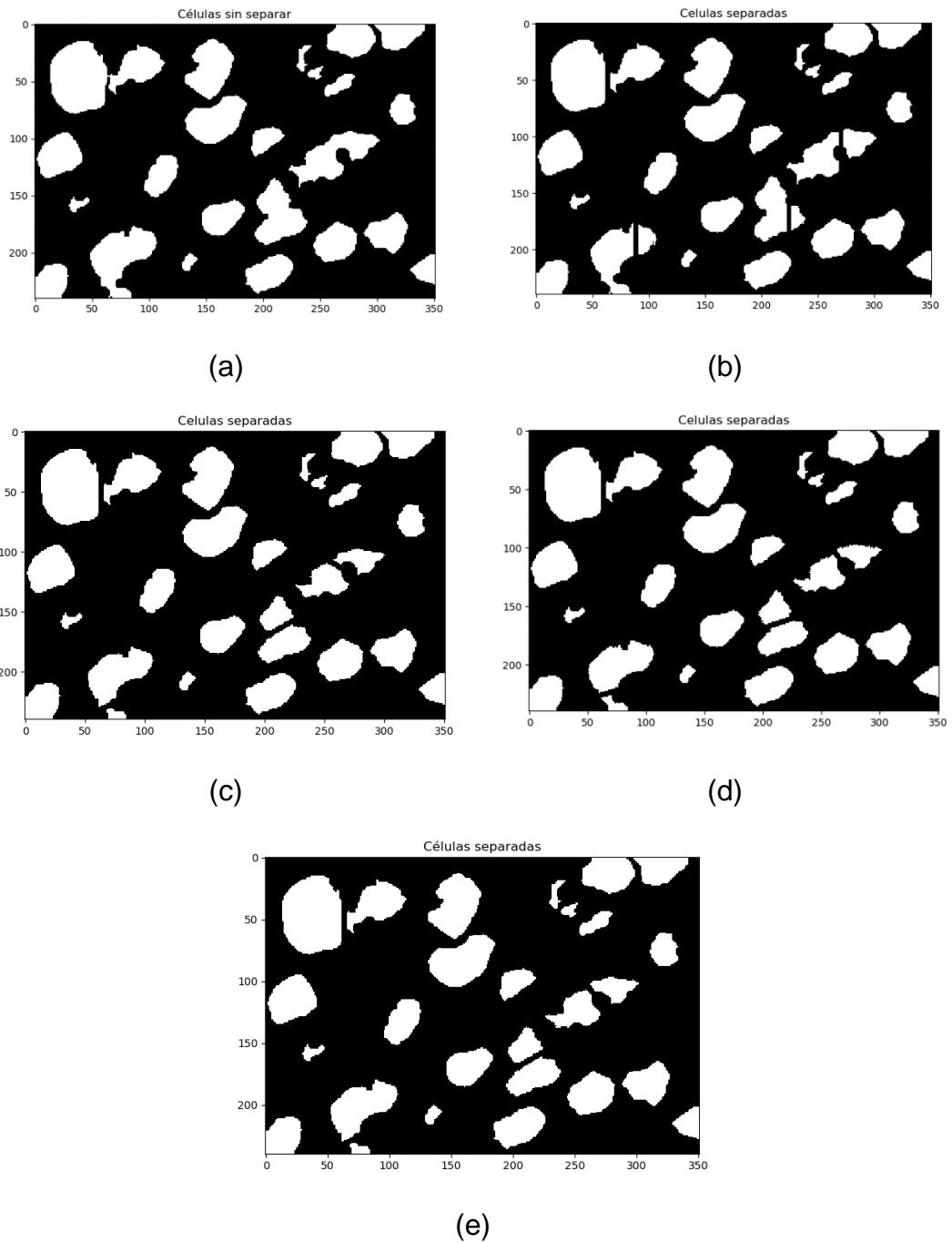


Figura 40. Separación de células (a) Prueba con 2 giros (b) Prueba con 3 giros (c) Prueba con 4 giros (d) Prueba con 6 giros (d) Prueba con 10 giros  
Elaboración: los autores

Como se esperaba, a partir de tres giros ya se obtienen resultados adecuados, sin embargo, se decidió trabajar con 6 giros para garantizar un corte fino en la mayoría de los casos de células etiquetadas con más de 1500 píxeles (más de 1 célula).

### **3.4.6 Clasificación y conteo de células proliferantes y no proliferantes**

A partir del total de células detectadas en la imagen  $I_{total}(x, y)$ , se procede a clasificar y contabilizar las células proliferantes y no proliferantes. Para ello se realizaron las coordinaciones con los médicos patólogos colaboradores a fin de que se instruya respecto al esquema de clasificación que requerían para poder brindar un mejor diagnóstico. En este caso, los médicos solicitaron clasificar las células proliferantes en tres subgrupos: fuertemente proliferantes, medianamente proliferantes y débilmente proliferantes. El criterio de clasificación es basado en el tamaño del área de célula comprometido con el marcador KI 67.

La imagen  $I_{total}(x, y)$  es sometida a un algoritmo de etiquetado que deberá asignar a cada célula detectada un índice  $k$ . En este caso  $k$  constituye el contador de objetos detectado en la imagen.

#### **3.4.6.1 Clasificación de células proliferantes**

Para la etapa de clasificación se inicializa en primera instancia los contadores  $Total_{pro} = 0$  (contador de células proliferantes) y el contador  $Total_{npro} = 0$  (contador de células no proliferantes).

Defínase  $T_{obj}$  como el número total de objetos o células que contiene  $I_{total}(x, y)$ , calculado mediante el uso de la técnica de etiquetado de imagen del capítulo 2.7. En ese contexto se lleva  $k$  a ser el número de objeto o célula a procesar. En este caso  $k = 0, 1, \dots, T_{obj} - 1$ .

Se hace  $k = 0$  y se ejecuta el siguiente procedimiento:

Paso 1:

Se extrae un segmento  $I_k(x^*, y^*)$  de la imagen  $I_{total}(x, y)$  conformado de  $M_k$  filas y  $N_k$  columnas, donde  $x^* = 0, 1, \dots, M_k - 1$  e  $y^* = 0, 1, \dots, N_k - 1$ . En este caso  $I_k(x^*, y^*)$  es el objeto de imagen que contiene exactamente al objeto  $k$  detectado en  $I_{total}(x, y)$ .

Paso 2:

Se actualiza una bandera  $f_k$  a través de la siguiente expresión:

$$f_k = \sum_{x^*=0}^{M_k-1} \sum_{y^*=0}^{N_k-1} I_k(x^*, y^*) \cdot I_{pro}(x_{0k} + x^*, y_{0k} + y^*) \quad (3.26)$$

donde  $x_{0k}$  e  $y_{0k}$  son las coordenadas del primer pixel de  $I_k(x^*, y^*)$  en  $I_{total}(x, y)$ .

Paso 3:

Si  $f_k \neq 0$  se considera que la célula evaluada es proliferante y se incrementa el contador  $Total_{pro}$  en una unidad. En caso contrario se incrementa el contador  $Total_{npro}$ .

Dentro de los requerimientos médicos se requiere sub-clasificar las células proliferantes en 3 niveles de proliferación: fuertemente proliferantes, medianamente proliferantes y débilmente proliferantes.

En tal sentido cuando  $f_k \neq 0$  se ejecutan los pasos del 4 al 9.

Paso 4:

Se genera una matriz temporal  $I_{kp}(x^*, y^*)$ , conformado de  $M_k$  filas y  $N_k$  columnas y se aplica:

$$I_{kp}(x^*, y^*) = \begin{cases} 1 & ; \text{ si } I_{pro}(x_{0k} + x^*, y_{0k} + y^*) \neq 0 \\ 0 & ; \text{ otro caso} \end{cases} \quad (3.27)$$



Paso 5:

Se contabiliza cada pixel que pertenece a un área que reaccionó a la proteína Ki 67:

$$Apro_k = \sum_{x^*=0}^{M_k-1} \sum_{y^*=0}^{N_k-1} I_{kp}(x^*, y^*) \quad (3.28)$$

Paso 6:

De la misma manera para el área no proliferante de la célula, se genera una matriz  $I_{knp}(x^*, y^*)$ , conformado de  $M_k$  filas y  $N_k$  columnas y se aplica:

$$I_{knp}(x^*, y^*) = \begin{cases} 1; & \text{si } I_{nopro}(x_{0_k} + x^*, y_{0_k} + y^*) \neq 0 \\ 0; & \text{otro caso} \end{cases} \quad (3.29)$$

Paso 7:

Se clasifica los pixeles de las regiones que no reaccionaron a la proteína Ki 67:

$$Anpro_k = \sum_{x^*=0}^{M_k-1} \sum_{y^*=0}^{N_k-1} I_{knp}(x^*, y^*) \quad (3.30)$$

Paso 8:

Luego se halla el factor de intensidad de cada célula:

$$R_k = \frac{Apro_k}{Anpro_k + Apro_k} \quad (3.31)$$

Paso 9:

Se actualizan los contadores de intensidad de proliferación  $fp_k$ ,  $fm_k$  y  $fd_k$  inicializados en 0 de la siguiente forma:

$$fp_k = \begin{cases} fp_k + 1 & ; \text{ si } R_k \geq 0.7 \\ fp_k & ; \text{ otro caso} \end{cases} \quad (3.32)$$

$$fm_k = \begin{cases} fm_k + 1 & ; \text{ si } R_k > 0.2 \wedge R_k < 0.7 \\ fm_k & ; \text{ otro caso} \end{cases} \quad (3.33)$$

$$fd_k = \begin{cases} fd_k + 1 & ; \text{ si } R_k \leq 0.2 \\ fd_k & ; \text{ otro caso} \end{cases} \quad (3.34)$$

Estos contadores tienen almacenada la cantidad de células subclasificadas. Nótese que  $fp_k + fm_k + fd_k = Total_{pro}$ .

Los datos son otorgados al médico patólogo para su respectiva evaluación y diagnóstico.

Este procedimiento se aplica para  $k = 1, 2, 3, \dots, T_{obj}$ .

#### 3.4.6.2. Conteo de células

Luego de evaluar todas las células se halla el porcentaje de células proliferantes aplicando:

$$Diagnostico = \left( \frac{Total_{pro}}{Total_{pro} + Total_{npro}} \right) \times 100 \quad (3.35)$$

### 3.5. Comentarios finales

En este capítulo se desarrolló la metodología utilizada en el desarrollo de la tesis y se mostraron los resultados parciales que sirvieron para hallar los parámetros específicos que optimizaron los resultados obtenidos. A continuación, se mostrarán los resultados globales de forma cuantitativa, haciendo comparativas entre el software y los especialistas.

## **CAPÍTULO IV**

### **RESULTADOS Y VALIDACIÓN**

#### **4.1 Comentarios iniciales**

En este capítulo se describirán los resultados y validaciones realizadas entre el software y los especialistas, quienes realizaron pruebas de forma visual.

#### **4.2 Costo computacional y tiempo de procesamiento**

Una variable importante en el análisis del software para su futura integración y utilización en los centros de salud es la velocidad de procesamiento. Si bien los médicos generalmente analizan una muestra de KI-67 entre 5 y 20 segundos. (resultados obtenidos mientras los médicos evaluaban las pruebas del ANEXO 1), los médicos especialistas están dispuestos a utilizar una herramienta de software siempre y cuando les otorgue una ventaja con respecto a la inspección visual que ellos realizan cotidianamente. Por ello, se presentan en la Tabla los tiempos de demora del software para diferentes imágenes de la base de datos utilizada. Estos tiempos corresponden a una computadora con procesador i7 de séptima generación y 8 gigas de memoria RAM.

#### **4.3 Validación de resultados**

Para medir el grado de acierto en los diagnósticos que el algoritmo propuesto entrega, se decidió medir el grado de concordancia “software-

especialista” haciendo uso del índice *Kappa* de *Cohen*. Por otro lado, para medir la sensibilidad y especificidad del software propuesto se hizo uso del F-score. A continuación, se detalla de forma más extensa las mediciones efectuadas:

#### 4.4 Métrica de validación: Índice *Kappa* de *Cohen*

También conocido como “Coeficiente *kappa* de *Cohen*” (*Josep L. Fleiss* et al. (1971)), es una medida estadística que refleja la concordancia inter-observador y puede ser calculado siempre y cuando el número de observadores sea como mínimo 2. El coeficiente *kappa* toma valores en -1 y +1. Se comprende que mientras más acercado a +1 sea el resultado, mayor es el grado de concordancia inter-observador. En el lado contrario, mayor sería el grado de discordancia. Si se obtiene un valor de 0, refleja que la concordancia observada es precisamente la que se espera a causa exclusivamente del azar.

Tabla 3. Tiempo de procesamiento del software

<b>IMAGEN</b>	<b>TIEMPO DE PROCESAMIENTO (segundos)</b>
Imagen1 (3584x2748 pixeles)	43.19
Imagen2 (3584x2748 pixeles)	95.93
Imagen3 (3584x2748 pixeles)	43.34
Imagen4 (3584x2748 pixeles)	45.92
Imagen5 (3584x2748 pixeles)	33.57
Imagen6 (3584x2748 pixeles)	32.38
Imagen7 (3584x2748 pixeles)	62.40

Elaboración: los autores

##### 4.4.1 Lugar y fecha de la prueba

Las pruebas se realizaron entre los meses de mayo y junio del 2017. Todos los especialistas fueron patólogos que laboran en instituciones de salud de la ciudad de Lima.

##### 4.4.2 Variable o parámetro a validar

Se medirá la concordancia entre el porcentaje de KI-67 proliferantes hallado por el software contra el porcentaje de KI-67 hallado por

cada uno de los especialistas. Asimismo, se calculará el índice *Kappa* entre especialistas para determinar el grado de concordancia que hay entre sus mediciones.

#### 4.4.3 Equipamiento y materiales

Se les entregó a los especialistas una lista de imágenes impresas para que las inspeccionaran visualmente e hicieran el cálculo del índice de proliferación de la expresión KI-67. También se les solicitó que calificaran de la siguiente manera: Luminal A, si el índice de proliferación es menor al 20% y Luminal B, si el índice de proliferación es mayor a 20%.

#### 4.4.4 Métrica de medición de desempeño

La ecuación que denota al índice *kappa* es la siguiente:

$$\tilde{K} = \frac{\bar{T} - \bar{T}_e}{1 - \bar{T}_e} \quad (4.1)$$

donde:  $\tilde{K}$ : índice *Kappa*

$\bar{T}$ = probabilidad relativa entre observadores

$\bar{T}_e$ = probabilidad hipotética de acuerdo con el azar

Definiendo  $\bar{T}$  como:

$$\bar{T} = \frac{1}{N_0} \sum_{i=1}^{N_0} \beta_i \quad (4.2)$$

donde  $N_0$ : es el número total de categorías o resultados esperados.

$\beta_i$  es definido como:

$$\beta_i = \frac{1}{\alpha(\alpha - 1)} \left[ \left( \sum_{j=1}^{N_0} \alpha_{ij}^2 \right) - \alpha \right] \quad (4.3)$$

donde  $\alpha$ : el número de ocurrencias de una categoría

$\alpha_{ij}$ : el número de veces que un observador escogió un resultado en la categoría  $i$  y el segundo observador escogió la categoría  $j$ .

Finalmente se tiene:

$$\bar{T}_e = \sum_{j=1}^{N_0} \varepsilon_j^2 \quad (4.4)$$

donde  $\varepsilon_j$  es calculado usando la siguiente expresión:

$$\varepsilon_j = \frac{1}{N_0 \alpha} \left( \sum_{i=1}^{N_0} \alpha_{ij} \right) \quad (4.5)$$

#### 4.4.5 Resultados de acuerdo al índice *Kappa*

Desde la Tabla 5 hasta la 9 se muestran los resultados obtenidos en la comparación del desempeño del software contra los especialistas utilizando el índice *Kappa*. Sin embargo, para hacer una comparación e interpretación más objetiva del índice *Kappa* calculado, también se realizó el cálculo del índice de concordancia entre algunos patólogos (elegidos de forma aleatoria). Estos resultados se encuentran en las Tablas 10, Tabla 11 y Tabla 12. El promedio de índice *Kappa* “software-especialista” es de 0.883, mientras que el índice de concordancia “especialista-especialista” es de 0.686.

#### 4.4.6 Comentarios y Conclusiones sobre el índice *Kappa*

Según *Landis J.R., Koch G.G. (1977)*, el índice *Kappa* promedio “software-especialista” de 0.883 está dentro del rango de concordancia casi perfecto mientras que el índice de concordancia “especialista-especialista” se encuentra dentro del rango sustancial. Como se esperaba, los especialistas varían en los porcentajes de células proliferantes de KI-67 debido a la subjetividad visual. Si bien el software presenta resultados que la mayoría de los médicos pueden aceptar como correctos, es necesario el análisis a detalle de la precisión que tiene el software para detectar correctamente cada célula. Es por ello que se hizo el cálculo del *F-score* que será explicado a continuación.

## 4.5 Métrica de validación *F-score*

El *F-Score*, conocido en español como Valor-F o medida-F, es una medida estadística que permite calcular la precisión que tiene una prueba experimental, explicado por Marina S. et al (2006). Se usa en determinar un valor único ponderado entre la precisión y exhaustividad. La precisión se define como la probabilidad de ocurrencia de que, en un grupo de muestras escogidas de una base de datos, estas sean las correctas, mientras que la exhaustividad es la probabilidad de ocurrencia de que las muestras sean las correctas de todas las posibles en la base de datos.

### 4.5.1 Lugar y fecha de la prueba

Las pruebas se realizaron entre los meses mayo y junio de 2017. La prueba se elaboró con la ayuda de un médico especialista del HNERM.

### 4.5.2 Variable o parámetro a validar

En esta sección se hace el cálculo de la capacidad del software para clasificar correctamente a las células en las categorías adecuadas. De esta manera se podrá saber qué tan confiable es el índice de proliferación proporcionado por el software.

### 4.5.3 Medición o prueba de referencia (*Gold Standard*)

En el ANEXO 2 se encuentra la lista de imágenes que el especialista examinó y clasificó. Esta fue impresa a color sobre papel blanco y contenía porciones de muestras de tejido mamario con células proliferantes y no proliferantes, las cuales el médico debía clasificar una a una para tener un análisis adecuado.

### 4.5.4 Métrica de Medición de desempeño

Se calcula mediante la siguiente ecuación:

$$F = \frac{(\beta_2^2 + 1) \times Precision \times Exhaustividad}{\beta_2^2 \times Precision + Exhaustividad} \quad (4.6)$$

donde precisión y exhaustividad se calcula mediante:

$$\text{Precisión} = \frac{tp}{tp + fp} \quad (4.7)$$

Y

$$\text{Sensibilidad} = \frac{tp}{tp + fn} \quad (4.8)$$

donde:

*tp*: verdaderos positivos, valor que el especialista y el software coincidieron en que una célula pertenece a una categoría.

*fp*: falsos positivos, valor que el especialista consideró que no pertenece a la categoría evaluada pero el software sí.

*fn*: falsos negativos, valor que el especialista consideró que pertenece a la categoría evaluada pero el software no.

El valor  $\beta_2$  denota el balance entre precisión y exhaustividad, si es mayor a 1, le da preferencia a exhaustividad, si es menor a 1, le da preferencia a precisión, si es 1, le da igual ponderación a ambos.

#### 4.5.5 Resultados del *F-score*

Se procedió a calcular el *F-score* para las categorías: Célula proliferante y célula no proliferante. Ponderando el *F-score* de ambas categorías con  $\beta_2 = 1$  se obtuvo un valor F igual a 0.8764.

#### 4.5.6 Comentarios y Conclusiones sobre la métrica *F-score*

Con el resultado obtenido se aprecia que hay una alta concordancia entre las células que detecta el médico y que el software también las considera detectadas, aunque el software termina detectando más células para dar un resultado más preciso. Para aumentar la precisión habría que realizar más pruebas con médicos sobre las mismas imágenes,



aunque debido a la subjetividad que se ha demostrado que hay entre especialistas, se obtendrían valores distintos, manteniéndose los mismos resultados de detección en el software.

#### 4.6 Informe económico

A continuación, se muestra en la Tabla 4 los equipos y programas utilizados para el desarrollo de la tesis y sus costos unitarios. Asimismo, se muestra el costo total utilizado para finalizar la tesis y los proveedores de los equipos y software que finalmente hicieron viable esta investigación.

Tabla 4. Informe económico

NOMBRE	USO	CANT.	COSTO UNITARIO	COSTO TOTAL	PROVEEDOR
Microscopio Digital	Captura de muestras digitales	1	S/. 18'380	S/. 18380	INEN
<i>Matlab</i>	Programación del prototipo	2	S/. 4000	S/. 8000	USMP
<i>Python</i>	Programación del software final	2	S/. 0	S/. 0	Libre
<i>Laptop</i>	Desarrollo de la tesis	2	S/. 2000	S/. 4000	Tesistas
Computadora de escritorio	Desarrollo de la tesis	2	S/. 1500	S/. 3000	USMP
Gastos de oficina	Impresiones y gastos otros	1	S/. 500	S/. 500	Tesistas
TOTAL				S/. 33880	

Elaboración: los autores

Tabla 5. Comparación Software vs. Especialista 1

Especialista 1: Kappa $\tilde{K} = 0.689$			
Especialista	Luminal A	Luminal B	TOTAL
Software			
Luminal A	7	3	10
Luminal B	0	9	9
TOTAL	7	12	19

Tabla 6. Comparación Software vs. Especialista 2

Especialista 2: Kappa $\tilde{K} = 0.689$			
Especialista	Luminal A	Luminal B	TOTAL
Software			
Luminal A	7	3	10
Luminal B	0	9	9
TOTAL	7	12	19

Tabla 7. Comparación Software vs. Especialista 3

Especialista 3: Kappa $\tilde{K} = 1$			
Especialista	Luminal A	Luminal B	TOTAL
Software			
Luminal A	10	0	10
Luminal B	0	9	9
TOTAL	0	9	19

Tabla 8. Comparación Software vs. Especialista 4

Especialista 4: Kappa $\tilde{K} = 1$			
Especialista	Luminal A	Luminal B	TOTAL
Software			
Luminal A	10	0	10
Luminal B	0	9	9
TOTAL	0	9	19

Tabla 9. Comparación Software vs. Especialista 5

Especialista 5: Kappa $\tilde{K} = 0.787$			
Especialista	Luminal A	Luminal B	TOTAL
Software			
Luminal A	10	0	10
Luminal B	2	7	9
TOTAL	12	7	19

Tabla 10. Comparación Especialista 1 vs. Especialista 2

Especialista 1 vs 2: Kappa $\tilde{K} = 0.774$			
Especialista 1	Luminal A	Luminal B	TOTAL
Especialista 5			
Luminal A	6	1	7
Luminal B	1	11	12
TOTAL	7	12	19

Tabla 11. Comparación Especialista 2 vs. Especialista 4

Especialista 2 vs 4: Kappa $\tilde{K} = 0.596$			
Especialista 2	Luminal A	Luminal B	TOTAL
Especialista 4			
Luminal A	7	0	7
Luminal B	4	8	12
TOTAL	11	8	19

Tabla 12. Comparación Especialista 1 vs. Especialista 5

Especialista 1 vs 5: Kappa $\tilde{K} = 0.689$			
Especialista 1	Luminal A	Luminal B	TOTAL
Especialista 5			
Luminal A	7	3	10
Luminal B	0	9	9
TOTAL	7	12	19

Elaboración: los autores

## CAPÍTULO V

### IMPLEMENTACIÓN DEL SOFTWARE

#### 5.1 Aplicaciones

El lenguaje de programación utilizado para la creación de la interfaz final para la implementación del algoritmo fue *Python*, en su versión 3.5. A continuación, se explicará en los siguientes pasos el funcionamiento del aplicativo de software realizado.

Paso 1:

Una vez ejecutado el aplicativo, en la Figura 40 se muestra la ventana principal, en la Figura 41 se observa la ubicación del botón ABRIR donde se presiona para cargar una nueva imagen.

Paso 2:

En la Figura 42 se selecciona la imagen a procesar desde la ventana del explorador; por defecto el software busca imágenes con extensión “.JPG”, se puede alternar entre formatos “.BMP”, “.PNG”, “.TIF”. También está la opción para ver todos los archivos contenidos en la carpeta. El usuario visualizará el nombre del archivo abierto a lado del botón ABRIR.

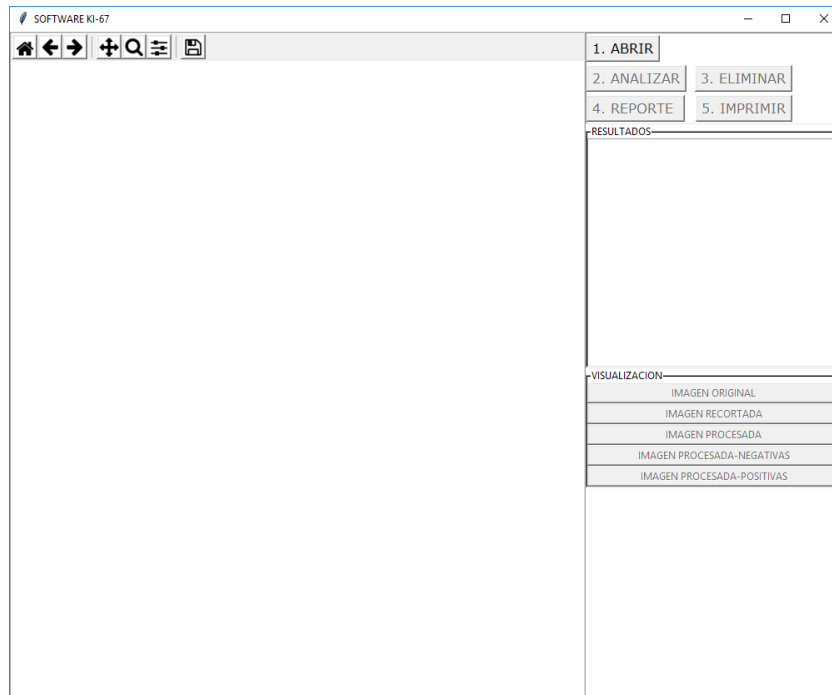


Figura 41. Visualización de la ventana principal  
Elaboración: los autores

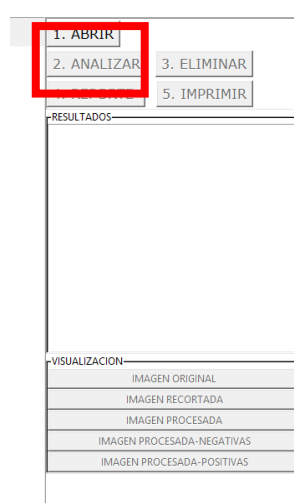


Figura 42. Ubicación de botón ABRIR  
Elaboración: los autores

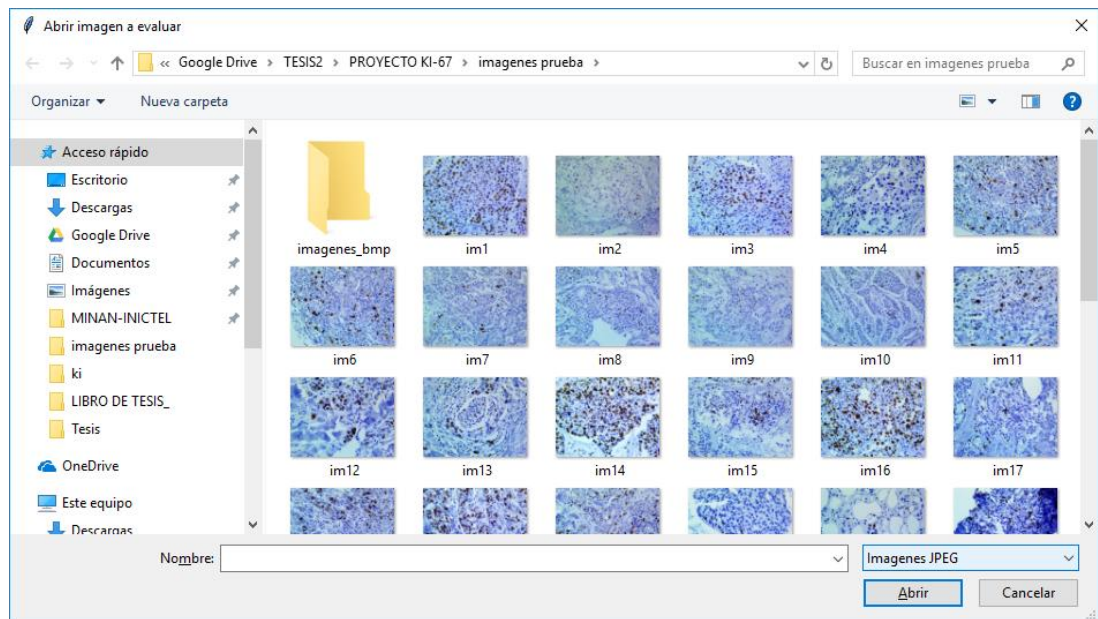


Figura 43. Vista del cuadro de carga de imágenes  
Elaboración: los autores

### Paso 3:

Una vez escogida la imagen a procesar, el usuario puede desplazarse a través de la imagen usando los botones de la barra de dibujo que se observa en la Figura 43.



Figura 44. Barra de opciones de la imagen cargada  
Elaboración: los autores

La función de los botones es la siguiente, en orden de aparición de izquierda a derecha:

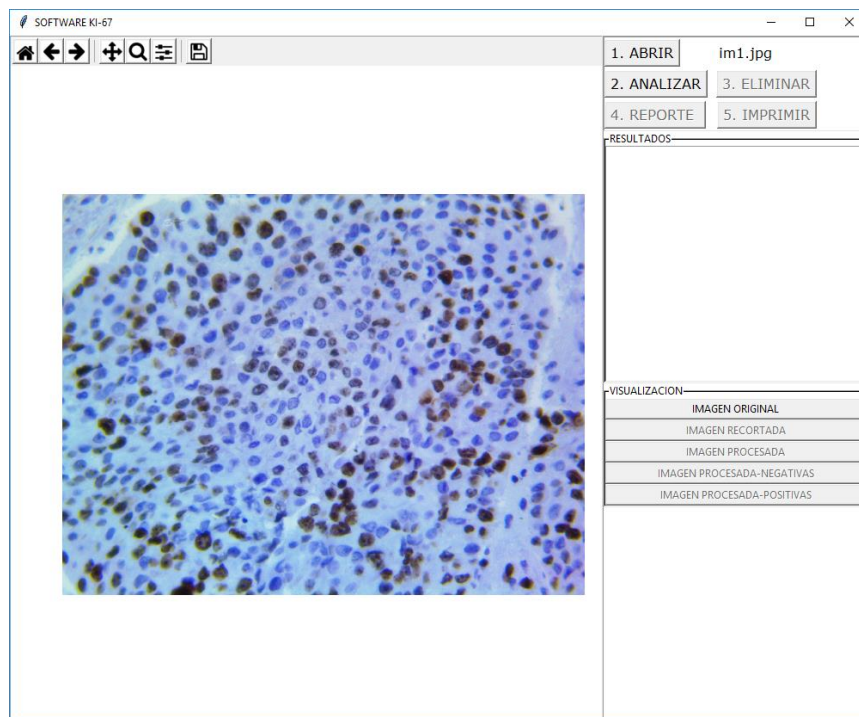
1. *Home*: regresa la vista de la imagen a la posición original.
2. Flecha izquierda: regresa a una vista anterior de la imagen.
3. Flecha derecha: avanza a una vista, en caso se haya retrocedido de vista.
4. Botón *Paneo/Zoom*: el usuario puede desplazarse por la imagen manteniendo presionado el botón izquierdo de la imagen y

moviendo hacia la dirección deseada. Si se presiona el botón derecho y mueve el mouse hacia arriba o a la derecha, realiza un acercamiento la imagen, si mueve hacia abajo o izquierda, realiza un alejamiento.

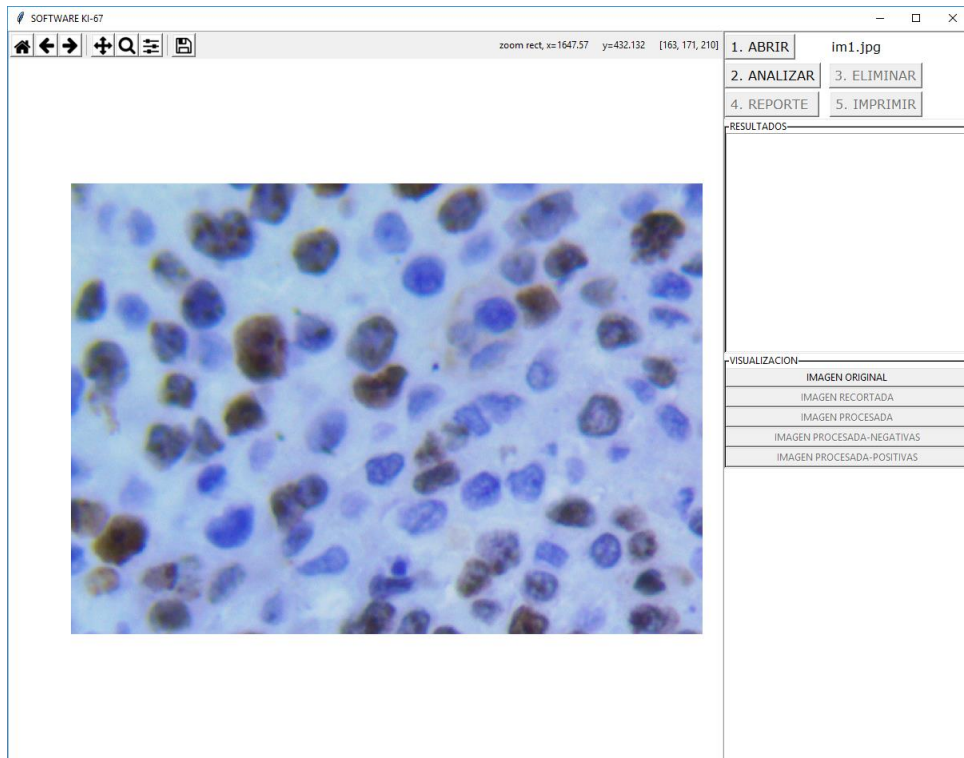
5. Botón Lupa: si se presiona el botón izquierdo del mouse realiza un acercamiento en base a un rectángulo que se dibuje, si se presiona el botón derecho, realiza un alejamiento en base al rectángulo dibujado.

Paso 4:

En la Figura 44a, el usuario debe posicionar la vista sobre el área a la que desea analizar usando las herramientas de la barra de dibujo, sino lo desea, puede pasar al siguiente paso, en la Figura 44b se observa ya la imagen acercada.



(a)



(b)

Figura 45. (a) Vista del aplicativo con la imagen cargada y (b) vista acercada  
Elaboración: los autores

Paso 5:

En la Figura 45, se observa dónde el usuario debe presionar el botón ANALIZAR y el aplicativo tomará la imagen que se visualiza en el momento para realizar el procesado.



Figura 46. (a) Vista del botón ANALIZAR  
Elaboración: los autores

#### Paso 6:

Una vez acabado el procesamiento, el software mostrará una nueva imagen, como se observa en la Figura 46, dónde se visualizará el resultado de las células pintadas de distinto color según la leyenda:

- Células negativas: **VERDE**
- Células Positivas Intensas: **ROJO**
- Células Positivas Medianas: **NARANJA**
- Células Positivas Débiles: **AMARILLO**

#### Paso 7:

Si el usuario desea, puede escoger ver entre varias imágenes de parte del proceso, presionando el botón del cuadro de Visualización según corresponda la imagen (estas se activarán una vez haya acabado el proceso):

- Imagen Original
- Imagen Recortada
- Imagen Procesada
- Imagen Procesada - Solo Negativas
- Imagen Procesada - Solo Positivas



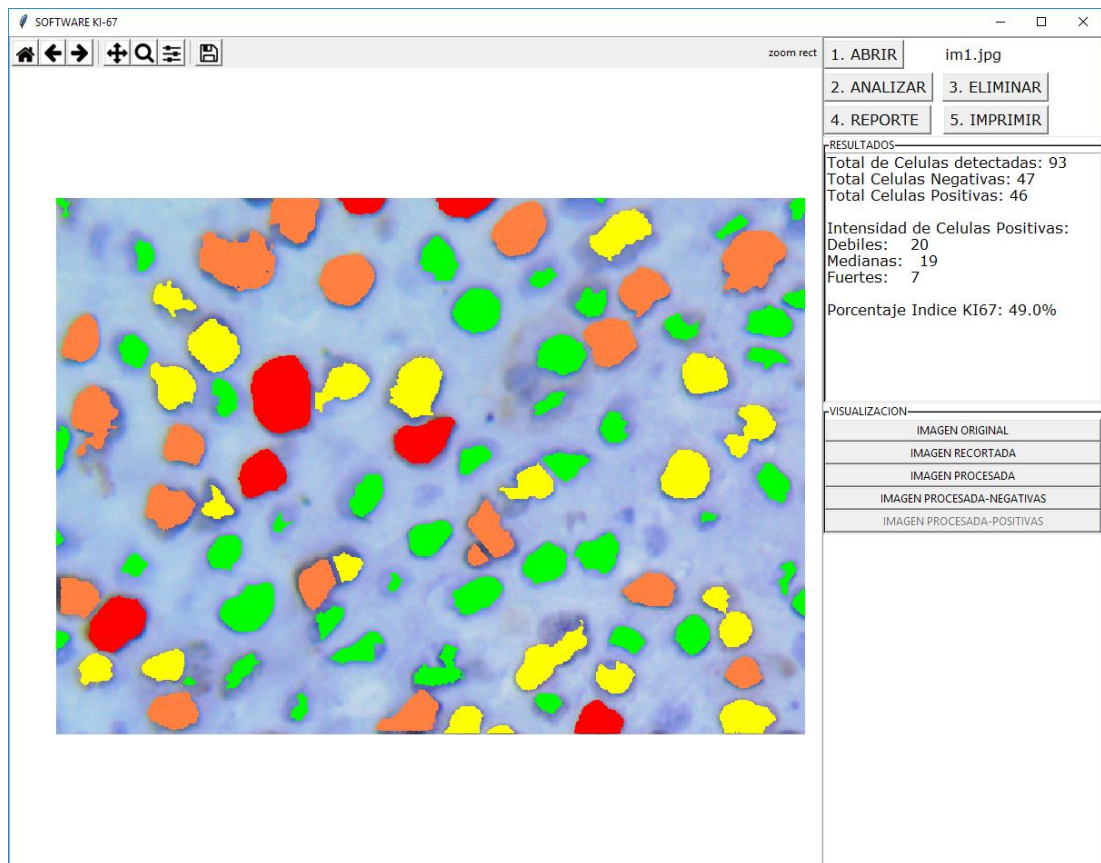


Figura 47. Vista del proceso de clasificación y conteo finalizado  
Elaboración: los autores

#### Paso 8: (Opcional)

Según criterio del médico usuario, puede decidir si hay células que no debieron ser contabilizadas y eliminarlas del conteo, mediante el uso del botón ELIMINAR; para esto, el usuario tendrá que presionar con el botón izquierdo del mouse sobre las células que desee eliminar y una vez acabado presionar la tecla ENTER y desaparecerán de la imagen y también de la estadística, en la Figura 47 se aprecian los pasos para eliminar las células no deseadas, la Figura 47a se observa el conteo total de células, en el Figura 47b se presiona click sobre el botón Eliminar y se aprecia dos células de color verde con una cruz roja encima, que son las células seleccionadas, finalmente en la Figura 47c se aprecia el cambio realizado, las células ya no pertenecen al conteo y se actualiza el porcentaje del resultado final.

Paso 9:

Una vez el médico usuario esté conforme con la imagen, puede primero a ver el resultado en la tabla de VISUALIZACIÓN, donde se muestra el conteo total, conteo por tipo de células y conteo según nivel de proliferación de las positivas; al igual que el porcentaje de índice de proliferación. Una vez conforme se procede a guardar la imagen para su posterior uso dando click en el botón guardar (mencionado en el paso 3). En ese momento aparecerá una nueva ventana que se observa en la Figura 48, donde puede decidir qué nombre poner a la imagen, el tipo de extensión de archivo y la ubicación donde será guardada.

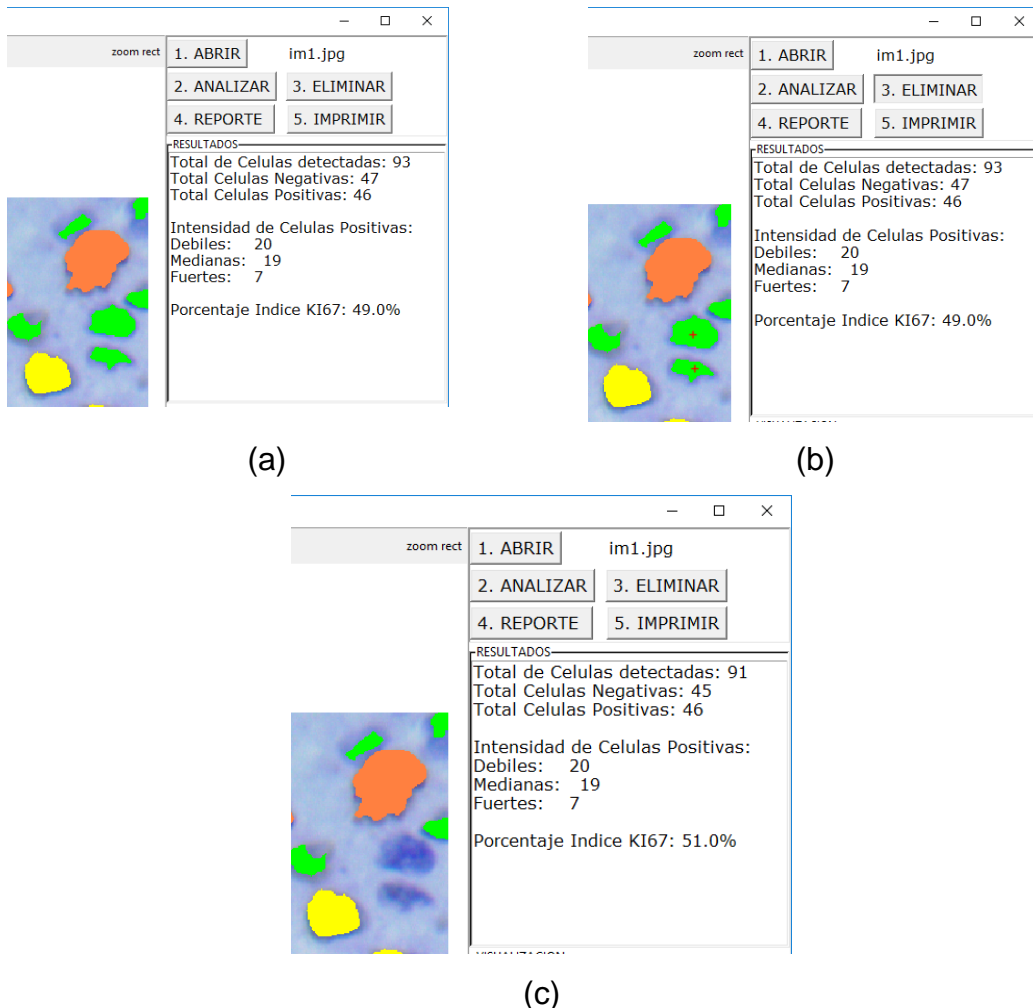


Figura 48. Vista de la opción de eliminar (a)Conteo actual de células. (b) Selección de células a eliminar. (c) Vista actualizada con conteo actualizado  
Elaboración: los autores

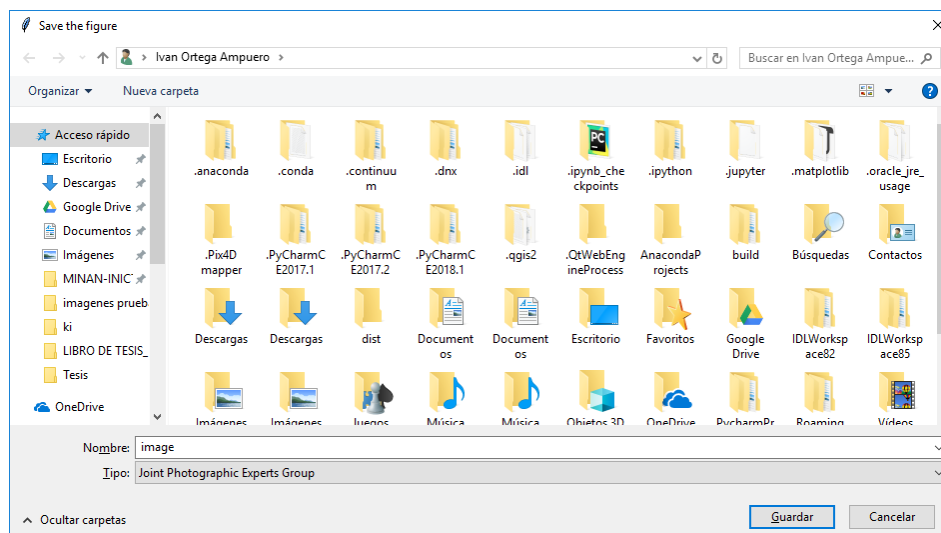


Figura 49. Vista de la ventana de guardar  
Elaboración: los autores

#### Paso 10:

El médico usuario tiene la facilidad de generar un reporte que le permita almacenar los resultados en un formato de PDF. El médico presiona en el botón REPORTE y se abre una nueva ventana, como se ve en la Figura 49, donde se ingresa los datos principales del médico y del paciente, una vez acabado presiona la tecla GENERAR REPORTE y saldrá una ventana para escoger donde guardar el reporte, que llevará de nombre Reporte KI-67 #####, donde "#####" es el número de DNI ingresado del paciente, una vista del reporte se puede apreciar en la Figura 50.

#### Paso 11:

Una vez acabado la generación del reporte, el médico usuario puede presionar el botón IMPRIMIR y automáticamente el software buscará la impresora conectada a la computadora e imprimirá el último reporte generado.

The image shows a software window titled "Información de Reporte" with standard window controls (minimize, maximize, close). The window contains a form with the following fields:

- Nombre del Paciente:
- Sexo:
- DNI:
- EDAD:
- #Historia Clínica:
- Tipo de Seguro:
- Fecha:
- Nombre del Doctor:
- Motivo:
- Observaciones:

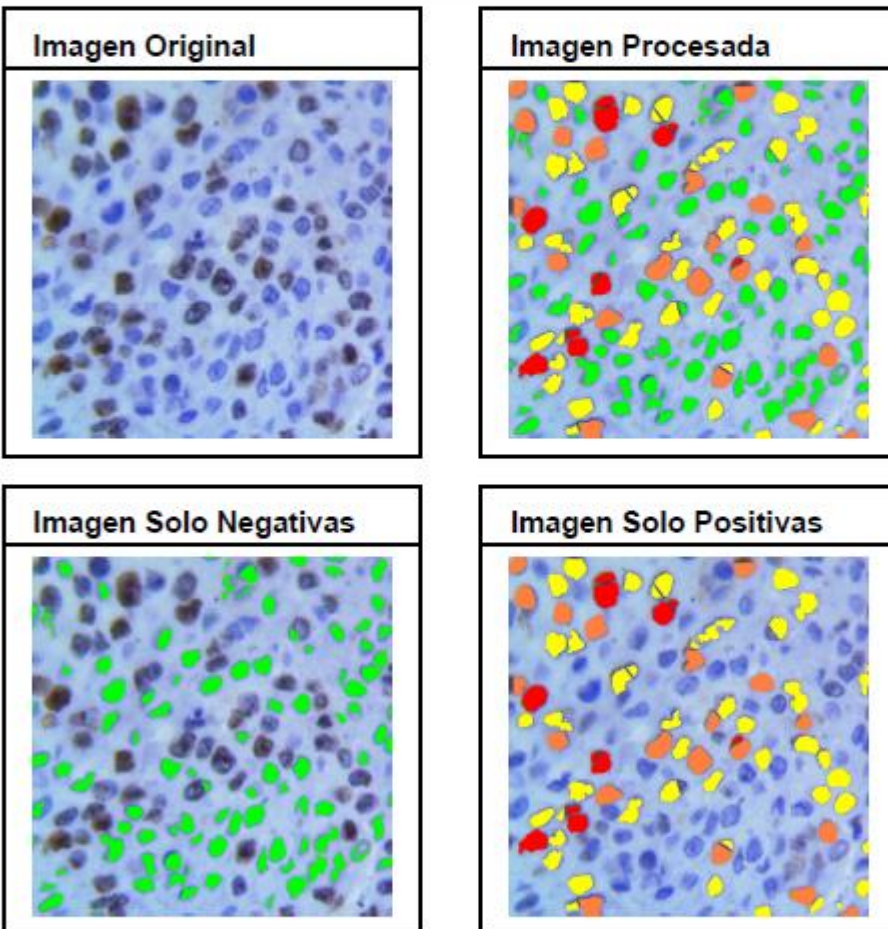
A button labeled "GENERAR REPORTE" is located to the right of the "Observaciones" field.

Figura 50. Vista de la ventana para generar reporte  
Elaboración: los autores

REPORTE DE ÍNDICE DE PROLIFERACIÓN KI-67		
Nombres: Juana Fernandez		
DNI: 70808794	#H.C.: 12345678	Sexo: Femenino
Seguro: Privado		Edad: 28
Doctor: D. Ortega		Fecha: 10/08/2018
Motivo:		
Observaciones:		

RESULTADO ANÁLISIS DE INDICE DE PROLIFERACIÓN KI-67

**IMÁGENES**



<p><b>Resultado de Conteo</b>            Porcentaje KI-67: 42.0%            Total de Celulas: 183            Total Celulas Positivas: 76            Total Celulas Negativas: 107</p>
--

Figura 51. Vista del reporte generado en PDF  
 Elaboración: los autores

## CONCLUSIONES

A partir de los resultados obtenidos y la investigación realizada en este trabajo se concluye lo siguiente:

1. Se consiguió implementar una herramienta de software para la cuantificación del marcador tumoral de la proteína KI-67 haciendo el uso de técnicas de procesamiento digital de imágenes, obteniendo un *F-score* de 0.876 y un promedio de índice *Kappa* contra cinco especialistas de 0.883. Estos resultados son bastante aceptables tomando en cuenta que el promedio de índice *Kappa* entre especialistas (especialista vs especialista) es de 0.686. Esto demuestra que la subjetividad es demasiado alta entre los médicos y que el diagnóstico entregado de una misma muestra puede variar según el especialista que la inspeccione.
2. Se consiguió hacer el cálculo del índice de proliferación en un promedio de 48 segundos, lo cual es aceptado por los médicos especialistas que probaron el software. Las actas de conformidad se encuentran en el ANEXO 3. Con respecto a la interfaz visual desarrollada, tuvo aceptación por parte de los patólogos.
3. Para garantizar que los resultados de precisión puedan ser replicados en diferentes instituciones de la salud, se requiere que se utilicen procesos similares a los que fueron utilizados en

la etapa de tinción de la base de datos utilizada en esta tesis. De no ser así es posible que se incurran en errores debido a la varianza de las nuevas muestras por procesar.

4. Se obtuvieron nuevos conocimientos en técnicas de procesamiento digital de imágenes mientras se satisfacía una necesidad en el rubro de la patología. Un proyecto multidisciplinario siempre necesita de una constante comunicación entre los desarrolladores y los usuarios finales para poder comprender plenamente la necesidad que se quiere subsanar y si se está haciendo de la forma correcta.

## RECOMENDACIONES

1. Analizar de manera más profunda de cuanto ha influenciado la subjetividad en la validación de resultados. Un análisis de la incertidumbre podría incrementar los índices de concordancia usados en esta tesis.
2. Proponer dos posibles caminos para la reducción de tiempo en el cálculo del índice de proliferación. La primera es la adaptación de la algoritmia para procesar con bloques en paralelo. Según la cantidad GPU y CPU, el proceso del cálculo del índice de proliferación podría reducirse exponencialmente. Otra solución a esto es el uso de redes neuronales para el proceso de segmentación y discriminación. La razón por la cual no se empleó esa técnica en el desarrollo de esta tesis es por la poca cantidad de imágenes como base de datos para poder conseguir un resultado generalizado de una red neuronal.
3. Se aconseja que los estándares y procesos de tinción sean más homogéneos en los centros de salud. Se considera que la poca correlación entre los resultados otorgados por los especialistas en nuestra base de datos está influenciada por lo diversas que son las imágenes que observan en diferentes laboratorios de inmunohistoquímica.



## FUENTES DE INFORMACIÓN

### Bibliográficas

Elbert Paquimio Rojas. (2008). Situación del cáncer de mama en el Perú. INEN, Departamento Estadística y Epidemiología.

Rafael C. Gonzalez, Richard E. Woods. (2006). Digital Image Processing (3rd Edition). Estados Unidos: Prentice-Hall.

Rafael C. Gonzalez. (2008). Digital Image Processing Using Matlab - 2nd edition. Estados Unidos: Gatesmark, Llc.

S. Coronato, G. E. Laguens, O. Spinelli, W. Di Girolamo. (2002). Marcadores tumorales en cáncer de mama. Cátedra de Patología B, Universidad de la Plata, 62, 73-82.

W. C. Ramos and D. R. Venegas. (2013). Análisis de la Situación del Cáncer en el Perú. Ministerio de Salud, Dirección General de Epidemiología.

### Hemerográficas

Ahmad, F. K., & Yusoff, N. (2013). Classifying breast cancer types based on fine needle aspiration biopsy data using random forest classifier. In Intelligent Systems Design and Applications (ISDA), 2013 13th International Conference on (pp. 121-125). IEEE.

- Anari, V., Mahzouni, P., & Amirfattahi, R. (2010). Computer-aided detection of proliferative cells and mitosis index in immunohistochemically images of meningioma. In *Machine Vision and Image Processing (MVIP), 2010 6th Iranian* (pp. 1-5). IEEE.
- Arihiro, K., Oda, M., Ohara, M., Kadoya, T., Osaki, A., Nishisaka, T., ... & Kobayashi, Y. (2016). Comparison of visual assessment and image analysis in the evaluation of Ki-67 expression and their prognostic significance in immunohistochemically defined luminal breast carcinoma. *Japanese journal of clinical oncology*, 46(12), 1081-1087.
- Banin Hirata, B. K., Oda, J. M. M., Losi Guembarovski, R., Ariza, C. B., Oliveira, C. E. C. D., & Watanabe, M. A. E. (2014). Molecular markers for breast cancer: prediction on tumor behavior. *Disease markers*.
- Bustreo, S., Osella-Abate, S., Cassoni, P., Donadio, M., Airoidi, M., Pedani, F., ... & Castellano, I. (2016). Optimal Ki67 cut-off for luminal breast cancer prognostic evaluation: a large case series study with a long-term follow-up. *Breast cancer research and treatment*, 157(2), 363-371.
- Focke, C. M., Bürger, H., van Diest, P. J., Finsterbusch, K., Gläser, D., Korsching, E., ... & Friedrich, K. (2017). Interlaboratory variability of Ki67 staining in breast cancer. *European Journal of Cancer*, 84, 219-227.
- Fulawka, L., & Halon, A. (2017). Ki-67 evaluation in breast cancer: The daily diagnostic practice. *Indian Journal of Pathology and Microbiology*, 60(2), 177.
- Goldhirsch, A., Ingle, J. N., Gelber, R. D., Coates, A. S., Thürlimann, B., Senn, H. J., & Panel members. (2009). Thresholds for therapies: highlights of the St Gallen International Expert Consensus on the primary therapy of early breast cancer 2009. *Annals of oncology*, 20(8), 1319-1329.

- He, X., Chen, Z., Fu, T., Jin, X., Yu, T., Liang, Y., ... Huang, L. (2014). Ki-67 is a valuable prognostic predictor of lymphoma but its utility varies in lymphoma subtypes: evidence from a systematic meta-analysis. *BMC Cancer*, 14, 153.
- Josep L. Fleiss et al. 'Measuring Nominal Scale Agreement Among Many Raters', *Psychological Bulletin*, Vol. 76, No. 5, pp 378-382 (1971).
- Landis J.R., Koch G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*. 33:159-174.
- M. Bouzid, A. Khalfallah, A. Bouchot, M. S. Bouhlel, F. S. Marzani (2013). "Automatic cell nuclei detection: a protocol to acquire multispectral images and to compare results between color and multispectral images," *Proc. SPIE 8587, Imaging, Manipulation, and Analysis of Biomolecules, Cells, and Tissues XI*.
- Marina S. et al. 'Beyond Accuracy, F-Score and ROC: a Family of Discriminant Measures for Performance Evaluation. Lecture notes in Computer Science. (2006)
- Shi, P., Zhong, J., Hong, J., Huang, R., Wang, K., & Chen, Y. (2016). Automated ki-67 quantification of immunohistochemical staining image of human nasopharyngeal carcinoma xenografts. *Scientific reports*, 6, 32127.
- Soledad Torres et. al. (2013). Estado del arte el diagnóstico y tratamiento del cáncer de mama. *Revista Médica Clínica Las Condes*, Vol. 24 (pp. 588-609).
- Tsai, C. M., & Lee, H. J. (2002). Binarization of color document images via luminance and saturation color features. *IEEE Transactions on Image Processing*, 11(4), 434-451.

## Electrónicas

Datos epidemiológicos INEI (2000-2014). Recuperado de: [https://www.inen.sld.pe/portal/documentos/pdf/estadistica/datos\\_estadisticos/02062016\\_DATOS\\_EPIDEMIOLOGICOS%20INEN\(2000-2014\).pdf](https://www.inen.sld.pe/portal/documentos/pdf/estadistica/datos_estadisticos/02062016_DATOS_EPIDEMIOLOGICOS%20INEN(2000-2014).pdf)

AcercaCiencia. Ciclo Celular. Recuperado de: <http://www.acercaciencia.com/2012/10/15/ciclo-celular/>

BiocareMexico. Ki 67. Recuperado de: <http://www.biocaremexico.com/contacto/noticias/ki-67>

Cold Vision Library. Image gradient with Sobel operator, OpenCV 3.X and CUDA. Recuperado de: <http://www.coldvision.io/2016/03/18/image-gradient-sobel-operator-opencv-3-x-cuda/>

Instituto Superior de Formación y Recursos en Red para el Profesorado. Modos de Color. Recuperado de: <http://www.ite.educacion.es/formacion/materiales/107/cd/imagen/imagen0104.html>

Mathworks. Convert from HSV to RGB Color Space. Recuperado de: [https://la.mathworks.com/help/images/convert-from-hsv-to-rgb-color-space.html?searchHighlight=hsv&s\\_tid=doc\\_srchtile](https://la.mathworks.com/help/images/convert-from-hsv-to-rgb-color-space.html?searchHighlight=hsv&s_tid=doc_srchtile)

Mathworks. Gamma Correction. Recuperado de: <https://la.mathworks.com/help/images/gamma-correction.html>

Nick Efford (2000). Digital Image Processing: A Practical Introduction Using Java. Recuperado de: <https://www.cs.auckland.ac.nz/courses/compsci773s1c/lectures/ImageProcessing-html/topic3.htm>

Torben Trindkaer Nielsen. An Implementation Of The Connected Component Labelling Algorithm. Recuperado de: <https://www.codeproject.com/articles/825200/an-implementation-of-the-connected-component-label>

## **ANEXOS**

	<b>Página</b>
Anexo N° 01 Imágenes evaluadas por los médicos	93
Anexo N° 02 Imágenes utilizadas para la validación	148
Anexo N° 03 Actas de conformidad	159