

**DETECCIÓN AUTOMÁTICA DE LA MUTACIÓN GERMINAL
TP53 R337H POR PCR-RFLP MEDIANTE VISIÓN POR
COMPUTADORA**

Juan José Dos Santos

Orientadores: Prof. Horacio Legal Ayala, Dr.
Prof. José Luis Vázquez N., M.Sc.
Prof. Edith Falcon, M.Sc.

Tesis presentada a la Facultad Politécnica, Universidad Nacional de Asunción,
como requisito para la obtención del Grado de Máster en Ciencias de la
Computación.

ASUNCIÓN - PARAGUAY
Marzo - 2018

**DETECCIÓN AUTOMÁTICA DE LA MUTACIÓN GERMINAL
TP53 R337H POR PCR-RFLP MEDIANTE VISIÓN POR
COMPUTADORA**

Juan José Dos Santos

Aprobado en Marzo de 2018 por:

Prof. Examinador 1, D.Sc. Sebastian Grillo

Prof. Examinador 2, D.Sc. Magna Monteiro

Prof. Examinador 3, D.Sc. Horacio Legal Ayala

Prof. Examinador 4, D.Sc. Edith Falcón

Prof. Examinador 5, M.Sc. José Luís Vazquez Noguera

Datos Internacionales de Catalogación en la Publicación (CIP)
DE BIBLIOTECA CENTRAL DE LA UNA

Dos Santos, Juan José

Detección automática de la mutación germinal TP53 R337H por PCR-RFLP mediante visión por computadora/Juan José Dos Santos. – Asunción, 2018.

70 p. : il.

Tesis (Maestría en Ciencias de la Computación) – Facultad Politécnica, 2018.

Bibliografía.

1. Visión por computador. 2. Mutación germinal. 3. Segmentación. 4. Morfología matemática. 5. Diagnóstico de mutaciones. 6. Cáncer en el córtex adrenal. I. Título.

CDD 519.4

Dedicatoria

Dedicado a mis padres Martín Dos Santos y Petrona Díaz, gracias por el apoyo incondicional que en todo momento me brindan en esta vida.

Dedicado a mis hermanas Lilian Dos Santos y Gladys Dos Santos quienes me dieron fuerzas para seguir siempre adelante.

Dedicado a mis compañeros de trabajo y de la maestría quienes me dieron sus alientos y ayuda para culminar esta etapa exitosamente.

Agradecimientos

En primera instancia agradezco a Dios y a la Virgen por todas las bendiciones
que recibo día a día.
Gracias a mi familia por la comprensión, el apoyo y el aliento que me brindaron
para el logro de mis objetivos.
Gracias a mis orientadores Prof. Horacio Legal, Prof. Edith Falcón y Prof. José
Luís Vázquez que con sus correcciones me ayudaron a realizar de la mejor
manera el presente trabajo de Tesis.
Agradezco a la Prof. Bióloga Elis Rosane Sade, del Instituto de Pesquisa Pele
Pequeno Príncipe (Curitiba, Brasil) por el curso de capacitación dictado sobre
biología molecular y la técnica de detección de la mutación del gen TP53R337H.
Agradezco al CONACYT por la beca otorgada que me dió la oportunidad para
culminar esta maestría.

DETECCIÓN AUTOMÁTICA DE LA MUTACIÓN GERMINAL TP53 R337H POR PCR-RFLP MEDIANTE VISIÓN POR COMPUTADORA

Autor: Juan José Dos Santos
Orientadores: Horacio Legal Ayala, Dr.
José Luis Vázquez N., M.Sc.
Edith Falcon, M.Sc.

RESUMEN

El PCR_RFLP (Polymerase Chain Reaction, Restriction Fragment Length Polymorphism - Reacción en cadena de polimerasa, polimorfismos en la longitud de los fragmentos de restricción) es una técnica de amplio uso en la actualidad para el estudio del ADN en el área de genética molecular. Esta técnica es utilizada para la detección de la mutación del R337H del gen TP53 para el diagnóstico precoz del cáncer en el córtex adrenal en humanos que tiene una incidencia mayor en el sur del Brasil y región este del Paraguay, específicamente en los departamentos de Concepción, Itapúa, Alto Paraná, Amambay y Canindeyú.

La detección de la mutación se realiza mediante el análisis visual de imágenes digitales resultantes de la electroforesis. Las imágenes son analizadas visualmente con el fin de identificar muestras con casos positivos de la mutación del R337H del gen TP53, donde un caso positivo es detectado en una fila de muestras con 3 bandas identificadas, los casos negativos son identificados con 2 bandas.

Esta tesis propone un método automático, rápido, sensible y confiable que determina la correcta identificación de las muestras y permite auxiliar en el diagnóstico de la mutación TP53R337H en imágenes de electroforesis en gel mediante técnicas de visión por computadora, agilizando el proceso de identificación en comparación con las técnicas del experto.

A partir del análisis de un conjunto de 1993 muestras, previamente analizadas por especialistas del área, el algoritmo de procesamiento automático propuesto consigue una tasa de aciertos de 93.4%, y del 6,6% de la tasa de error. Esto representa una contribución muy importante para auxiliar a los especialistas en la detección de la mutación y en la toma de decisiones.

Palabras claves: PCR-RFLP, Mutación, DNA, imágenes de electroforesis en geles, visión por computadora, procesamiento, segmentación.

AUTOMATIC DETECTION OF TP53R337H GERMLINE MUTATION BY PCR-RFLP THROUGH COMPUTER VISION

Author: Juan José Dos Santos

Advisors: Horacio Legal Ayala, Dr.

José Luis Vázquez N., M.Sc.

Edith Falcon, M.Sc.

SUMMARY

PCR RFLP (Polymerase Chain Reaction, Restriction Fragment Length Polymorphism) is technique widely used at the present time for the study of DNA in the area of molecular genetics. This technique is used to detect the R537H mutation of the TP53 gene for the early diagnosis of cancer in the adrenal cortex in humans, which has a higher incidence in southern Brazil and in the eastern region of Paraguay, specifically in the departments of Concepción, Itapúa, Alto Paraná, Amambay and Canindeyu.

The detection of the mutation is carried out through the visual analysis of digital images resulting from electrophoresis. The images are analyzed visually in order to identify samples with positive cases of the R337H mutation of the TP53 gene, where a positive case is detected in a row of samples with 3 identified bands, the negative cases are identified with 2 bands.

This thesis proposes an automatic, fast, sensitive and reliable method that determines the correct identification of the samples and allows auxiliary in the diagnosis of the TP53R337H mutation in gel electrophoresis images by computer vision techniques, speeding up the process of identification in comparison with the techniques of the expert.

From the analysis of a set of 1993 samples, previously analyzed by area specialists, the proposed automatic processing algorithm achieves a success rate of 93.4 %, and 6.6 % of the error rate. This represents a very important contribution to help the specialists in mutation detection and decision making.

Keywords: PCR-RFLP, Mutation, DNA, gel electrophoresis images, computer vision, processing, segmentation.

ÍNDICE GENERAL

LISTA DE FIGURAS	VII
LISTA DE TABLAS	X
LISTA DE SÍMBOLOS	XI
LISTA DE ABREVIATURAS	XII
1 INTRODUCCIÓN	1
1.1 Objetivo General	2
1.2 Objetivos específicos	2
1.3 Relevancia	3
1.4 Organización	3
2 CONCEPTOS BÁSICOS SOBRE LA MUTACIÓN TP53 R337H	5
2.1 Estudio PCR RFLP para detección de la mutación TP53 R337H .	5
2.2 Adquisición, Criterios y Análisis de las Imágenes de Electroforesis en Gel	6
3 CONCEPTOS BÁSICOS DE VISIÓN POR COMPUTADORA	11
3.1 Imagen Digital	11
3.2 Umbralización de imagen	15
3.2.1 Umbralización global y local	15
3.3 Morfología matemática	16
3.3.1 Elemento estructurante	16
3.3.2 Morfología binaria	17
3.3.3 Erosión y dilatación	17
3.3.4 Apertura y cierre	20
3.3.5 Reconstrucción morfológica	22
3.4 Morfología matemática en imágenes a niveles de gris	23

3.4.1	Erosión y dilatación	23
4	ABORDAJE PROPUESTO	27
4.1	Pseudocódigo	27
4.2	Propuesta general	29
5	RESULTADOS	36
5.1	Prueba experimental 1: Métodos de umbralización para detección de bandas	36
5.1.1	Resultados	36
5.2	Prueba experimental 2: Diagnóstico de la mutación germinal TP53 R337H.	37
5.2.1	Resultados	39
5.3	Ejemplos de detección.	41
5.3.1	Identificación y detección de la mutación R337H del gen TP53.	41
5.3.2	Casos erróneos de identificación de muestras	42
6	CONCLUSIONES Y TRABAJOS FUTUROS	44
6.1	Evaluación de los objetivos	44
6.2	Trabajos a futuro	45
	REFERENCIAS BIBLIOGRÁFICAS	46

LISTA DE FIGURAS

2.1	Proceso de Adquisición de las imágenes de electroforesis en gel . .	7
2.2	Posiciones de las muestras en una imagen de electroforesis en gel.	8
2.3	Criterios para las imágenes de electroforesis en gel.	9
2.4	Imagen de electroforesis en gel con bajo contraste y contaminado por objetos indeseados.	10
3.1	Imagen digital en 4 tamaños de resoluciones (a) imagen con resolución 128×128 (b) imagen con resolución 64×64 (c) imagen con resolución 32×32 (d) imagen con resolución 16×16	12
3.2	Representación de una imagen digital en forma matricial.	12
3.3	(a) Imagen digital binaria original, (b) segmento de la imagen original, (c) valores de intensidad que componen los píxeles del segmento.	13
3.4	(a) Imagen digital en niveles de gris original, (b) segmento de la imagen original, (c) valores de intensidad que componen los píxeles del segmento.	13
3.5	Representación una imagen de modelo R.G.B.	14
3.6	(a) Imagen resultante de la electroforesis en gel original en formato RGB, (b) Canal rojo de la imagen, (c) Canal verde de la imagen (d) canal azul de la imagen	14
3.7	(a) Imagen original en niveles de gris, (b) Histograma de la imagen con un umbral y (c) Imagen binaria a partir de la umbralización.) canal azul de la imagen	15
3.8	(a) Imagen original en niveles de gris, (b) Histograma de la imagen con un umbral en 127 (c) Imagen binaria a partir de la umbralización.	16
3.9	Elementos estructurantes con las siguientes formas, (a) forma circular, (b) forma cuadrada, (c) forma romboide (d) forma hexagonal. Los puntos rojos indican el centro del elemento.	17
3.10	Erosión de una imagen I por un elemento estructurante B	18
3.11	a) Imagen de electroforesis en gel original b) Resultado de la erosión	18

3.12	Dilatación de una imagen I por un elemento estructurante B . . .	19
3.13	(a) Imagen de electroforesis en gel original (b) imagen resultante de la dilatación.	19
3.14	Apertura de una imagen I por un elemento estructurante B . . .	20
3.15	a) Imagen binaria de electroforesis en gel original b) imagen resultante de la apertura.	21
3.16	Cierre de una imagen I por un elemento estructurante B	21
3.17	(a) Imagen binaria de electroforesis en gel original (b) imagen resultante del cierre morfológico.	22
3.18	(a) Imagen binaria de electroforesis en gel. (b) Imagen resultante luego de una reconstrucción.	23
3.19	a) Imagen en niveles de gris resultante de la electroforesis en gel. b) Imagen resultante de la erosión.	24
3.20	(a) Imagen en niveles de gris resultante de la electroforesis en gel. (b) Imagen resultante de la dilatación.	24
3.21	(a) Imagen en niveles de gris resultante de la electroforesis en gel. (b) Imagen resultante de la apertura.	25
3.22	(a) Imagen en niveles de gris resultante de la electroforesis en gel. (b) Imagen resultante del cierre morfológico.	26
4.1	Carriles horizontales que contienen las bandas de cada muestra . .	32
4.2	Estimación del promedio del tamaño horizontal de las bandas . .	32
4.3	Estimación del promedio de tamaño del espacio vertical de las bandas	32
4.4	Máscara estimada desde los promedios calculados	33
4.5	a) imagen original, (b) canal Rojo, (c) canal rojo erosionado, (d) imagen substraída, (e) imagen binarizada, f) imagen binaria erosionada, imagen sementada y filtrada	34
4.6	Descripción de pasos principales de la propuesta abordada	35
5.1	Histograma correspondiente a las métricas de Sensibilidad, Especificidad, exactitud y precisión de las imágenes	40
5.2	Imagen original con un caso positivo de la mutación.	41
5.3	Imagen_15 con un caso positivo de la mutación.	42
5.4	Imagen Original Imagen_90 con bajo contraste en las muestras. . .	42
5.5	Imagen Segmentada Imagen_90 con errores de identificación en las bandas de algunas muestras.	43
1	Segmentación de la Imagen_20	51
2	Segmentación de la Imagen_22	52

3	Segmentación de la Imagen_24	53
4	Segmentación de la Imagen_49	54

LISTA DE TABLAS

5.1	Tabla de métricas de los algoritmos de umbralización	37
5.2	Resultados de sensibilidad, especificidad, exactitud y precisión de muestras detectadas.	39
5.3	Matriz de confusión obtenidos del algoritmo propuesto.	40
5.4	Tabla de métricas de la detección de bandas	41

LISTA DE SÍMBOLOS

I	Representación de una imagen digital.....
$\mathbb{N}$	Representación de los números naturales.....
$\mathbb{Z}$	Representación de los números enteros.....
$\mathbb{R}$	Representación de los números Reales.
R	Rango de conjunto de valores.....
p	píxel de una imagen.....
B	Elemento Estructurante.
M	Total de filas en una imagen.....
N	Total de columnas en una imagen.....
$e_B(I)$	Erosión de una imagen I por un elemento estructurante B
$\delta_B(I)$	Dilatación de una imagen I por un elemento estructurante B
$\varphi_B(I)$	Cierre de una imagen I por un elemento estructurante B
$\gamma_B(I)$	Apertura de una imagen I por un elemento estructurante B
(x, y)	Par de coordenadas x e y en una imagen.....
O	Conjunto perteneciente al objeto de la imagen.
F	Conjunto perteneciente al fondo de la imagen.
μ	Función de pertenencia de un elemento.
K	Máscara estimada para la identificación de las bandas de una imagen de electroforesis en gel.
P	Número de muestras en la imagen de electroforesis en gel.
$bandInf$	Valor de píxel del extremo inferior de una banda.....
$bandSup$	Valor de píxel del extremo superior de una banda.....
$cantBan$	Cantidad de bandas detectadas.
$promTamBand$	Promedio del tamaño de bandas.
$Diagnosticok$	Resultado del diagnóstico de la mutación germinal.....
$promEspHor$	Promedio de los espacios horizontales entre bandas.
$\delta_I(X)$	Reconstrucción morfológica de una Imagen I por una imagen X .

LISTA DE ABREVIATURAS

ACC: Cáncer del Córtex Adrenal (ACC).

CEBIOTEC: Centro de Biotecnología e Informática.

DNA: Deoxyribonucleic Acid (Ácido desoxirribonucleico).

FN : Falsos Negativos.

FP: Falso Positivo.

VP: Verdadero Positivo.

Verdadero Negativo: Verdadero Negativo

HLA: Antígeno leucocitario humano

PB: Par de Base

PCR-RFLP: Polymerase Chain Reaction - Restriction Fragment Length Polymorphism. (Reacción en cadena de polimerasa- polimorfismos en la longitud de los fragmentos de restricción)

RGB: Red, Green, Blue (Rojo, Verde y Azul)

RN: Recien Nacidos

UV: Ultravioleta

CPU: Central Processing Unit (Unidad Central de Procesamiento)

GB: Gigabyte

RAM: Random Access Memory (Memoria de Acceso Aleatorio)

Capítulo 1

INTRODUCCIÓN

El cáncer en la actualidad sigue siendo una importante causa de mortalidad en la población mundial [FJ14]. Los procesos en el desarrollo del cáncer son complejos y multifactoriales e involucran factores ambientales y alteraciones en genes implicados en la proliferación celular. Entre las alteraciones o variaciones genéticas se encuentra comprometida la pérdida o alteración en la función de genes supresores de tumores. El gen supresor de tumores TP53, a diferencia de otros genes de la misma familia, participa en diferentes funciones fisiológicas de la célula para mantener la integridad y estabilidad genómica, referido así como un gen maestro [DP.92]. El gen TP53 también es el más comúnmente mutado en el cáncer humano y el más estudiado en la genética del cáncer [HMC91]. Las alteraciones en TP53 ocurren en la línea germinal (las cuales se heredan de la madre o del padre y, por lo tanto, están presentes en todas las células del organismo) y se encuentran en síndromes con predisposición al cancer familiar como el de Li-Fraumeni [JM.03]. Este síndrome se caracteriza por el inicio temprano de múltiples cánceres incluyendo carcinomas adrenocortical infantiles, sarcomas y tumores cerebrales.

En diversos estudios realizados en el Brasil, fue demostrada la alta incidencia de Cáncer Adrenocortical (ACC), principalmente en la región sur del país (10 a 15 veces mayor que la de la ACC pediátrica en todo el mundo) [Rib01], [CA00], [Guh17], [Sed98], [Her12].

Estos hallazgos motivaron a realizar estudios de prevalencia en la identificación de una posible mutación del TP53 R337H en niños recién nacidos (RN) en la región este del Paraguay, zona fronteriza Paraguay-Brasil, la cual esta mutación fue identificada en 5 casos positivos en 10.000 estudios de DNA de RN [EFdL15].

Los estudios biológicos son realizados mediante las técnicas de procesos en PCR, en la cual, posteriormente las muestras colocadas sobre geles de agarosa,

son sometidos al proceso de electroforesis. Los resultados son visualizados mediante luz UV, con el uso de una cámara fotográfica los resultados son obtenidos en formato digital. Las imágenes de electroforesis en gel, en ocasiones presentan dificultades a los profesionales biólogos en la lectura de los datos, debido a la contaminación por objetos indeseados y bajo contraste en las muestras procesadas, lo que podrían concluir en lecturas y diagnósticos erróneos de los resultados. En este trabajo se presenta el diseño e implementación de un algoritmo basado en técnicas de visión por computador, para la identificación de bandas, detección y diagnóstico automático de las muestras estudiadas. Las imágenes de electroforesis en gel fueron obtenidas de una base de datos realizadas en ciudades del Paraguay zona fronteriza con ciudades de Brasil [EFdL15].

1.1 Objetivo General

Desarrollar una herramienta para el análisis de datos de imágenes resultantes del estudio de PCR RFLP que permita la detección automática de la mutación TP53 R337H.

1.2 Objetivos específicos

- Comparar técnicas de umbralización automática para la segmentación de las imágenes de electroforesis en gel con el fin de obtener la mayor cantidad de muestras.
- Implementar técnicas de visión por computadora para la segmentación de las imágenes de electroforesis en geles que permita agilizar el proceso del experto en la toma de decisiones.
- Comparar el proceso automático y manual con respecto al tiempo sobre detección e identificación de las muestras para obtener diagnósticos más precisos de los casos de mutación R337H en el gen TP53.
- Evaluar el proceso automático con respecto a la precisión del diagnóstico de la mutación R337H utilizando técnicas de visión por computador Según Especificidad, Sensibilidad, Precisión y Exactitud.

1.3 Relevancia

Las herramientas de visión por computadora para procesar imágenes de PCR-RFLP en gel de agarosa no es nuevo. Existen paquetes de software de análisis comerciales tales como GelCompar [Gel] (Applied Maths, Belgium) o Advanced Quantifier (Bio Image Systems, MI, USA) [bio]; así como otras aplicaciones ya propuestas sobre imágenes PCR-RFLP, por ejemplo [MC10]. Sin embargo, no hemos encontrado una herramienta de visión por computadora para el auxilio en la detección de la mutación del gen TP53R337H por PCR-RFLP para la detección precoz del tumor de suprarrenal.

Los estudios sobre gel de agarosa, objeto del presente trabajo, son iluminados con rayos ultravioleta (UV) y observados a través de una cámara digital fotográfica. La detección de las mutaciones suelen dificultarse debido a la contaminación de objetos indeseados en estas imágenes, lo que pueden inducir a diagnósticos erróneos. El estudio de imágenes se limita a otros tipos de mutación germinal que no sea TP53 R337H.

En este trabajo será presentada una técnica para la detección automática de la mutación germinal TP53 R337H que tiene como relevancia los siguientes puntos:

- Es automática: La imagen a analizar es sometida al algoritmo para su análisis arrojando los resultados como diagnósticos finales.
- Es rápida: Cada imagen de dimensión 4276 x 2276 es analizada en un intervalo de 2.43 segundos a partir de una CPU de cuatro núcleos y 2 gb de RAM; el promedio que al biólogo le toma analizar cada muestra es de 2 minutos.
- Es sensible: Cualquier muestra que presente tres bandas no sea nítida en la imagen para la percepción visual, el algoritmo lo detecta y lo diagnostica como un caso sospechoso.
- Es confiable: Los resultados obtenidos de los experimentos realizados al algoritmo en comparación con el experto son similares, incluso en algunos casos el algoritmo detecta bandas que el experto no logra identificar debido a que las intensidades de estas bandas es cercano al fondo de la imagen.

1.4 Organización

- En el capítulo 2 los fundamentos teóricos son presentados en cuanto a los conceptos básicos de la mutación R337H y lo que envuelve el estado del

arte.

- En el capítulo 3 los conceptos básicos sobre imágenes y visión por computador son detallados, especificando las técnicas utilizadas para la segmentación de la imagen.
- En el capítulo 4 el algoritmo propuesto es detallado para la lectura, identificación y diagnósticos de las muestras en busca de la mutación R337H.
- En el capítulo 5 se detallan los experimentos realizados junto con los resultados obtenidos de los experimentos.
- En el capítulo 6 las discusiones y trabajos a futuros planteados.

Capítulo 2

CONCEPTOS BÁSICOS SOBRE LA MUTACIÓN TP53 R337H

En este capítulo son detallados los conceptos básicos sobre la mutación TP53 R337H, con el fin de conocer sobre los detalles que envuelven en la obtención de las imágenes de electroforesis en gel.

2.1 Estudio PCR RFLP para detección de la mutación TP53 R337H

La técnica disponible, conocida como PCR-RFLP, consiste en el procesamiento en diferentes y numerosas etapas de análisis para cada muestra que serán estudiadas y testadas con la intención de diagnosticar mutaciones en el gen humano.

Una placa corresponde al objeto donde las muestras son ubicadas para el estudio de PCR, la cual contiene 96 pozos (cada pozo corresponde a una muestra), cada una de ellas contenían sangre de dos recién nacidos.

Un fragmento del ADN de la muestra es amplificado por el estudio de PCR. Para la confirmación de todas las muestras neonatales que resultan positivas, los exones son secuenciados de vuelta para esas muestras [EFdL15].

Luego del proceso en el estudio de PCR RFLP, los resultados de la electroforesis en los geles de agarosa, son sometidos a luz Ultravioleta, (debido a que el reactivo es fluorescente solo bajo luz UV), dentro de una caja tipo prisma cuadrangular piramidal, en la cual, una cámara fotográfica profesional ubicada en la parte superior es utilizada para la visualización de las muestras y el posterior diagnóstico.

2.2 Adquisición, Criterios y Análisis de las Imágenes de Electroforesis en Gel

Los siguientes pasos corresponden al proceso de adquisición de las imágenes de electroforesis en gel.

- 1 La muestra es obtenida mediante la toma de sangre desde el talón del recién nacido, que es colocado en un disco de papel.
- 2 Los discos de papel son perforados y agregados a una placa para ser extraídos las muestras de ADN.
- 3 El lavado de los discos de papel, que están contenidas en una placa, es realizado para eliminar todo tipo de partículas.
- 4 Los discos de papel son llevados a una estufa para el secado.
- 5 Los discos son procesados dentro de sus respectivas placas en una máquina de PCR utilizando el termociclador.
- 6 Las muestras son extraídas de las placas y puestas en gel de agarosa para ser sometidas a una corriente eléctrica. Este proceso se realiza mediante la técnica de la electroforesis que permite separar moléculas y amplificar muestras de ADN buscando mutaciones.
- 7 Los resultados son expuestas en una caja de forma cuadrangular piramidal con superficie transiluminar para la fotodocumentación. La fotodocumentación es realizada con una cámara digital profesional que captura los resultados que contienen los geles de agarosa.

En la Figura 2.1 se pueden observar los pasos que fueron detallados anteriormente, que tiene como entrada la toma de sangre del paciente y un retorno correspondiente a la imagen de electroforesis en gel.

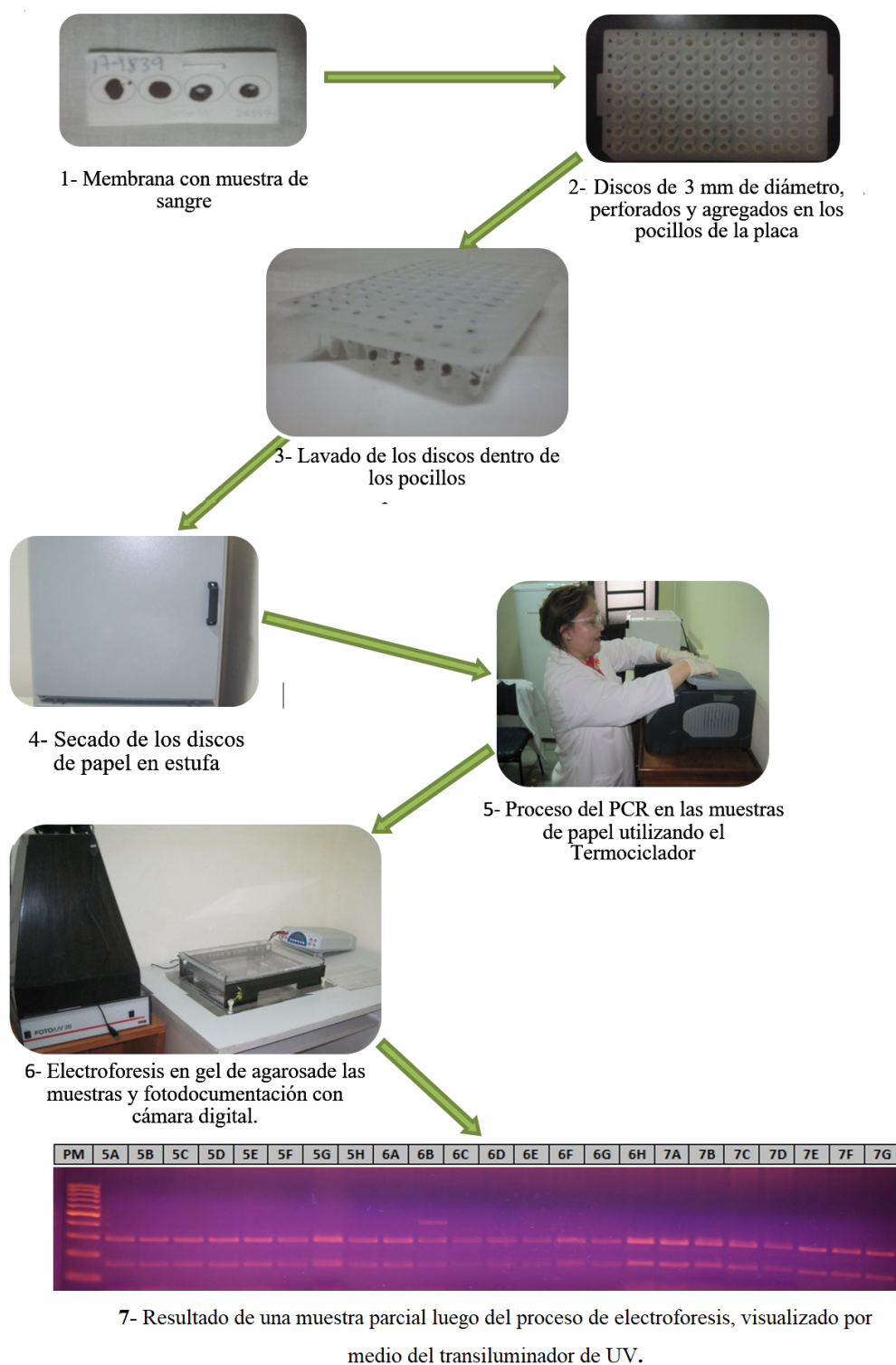


Figura 2.1: Proceso de Adquisición de las imágenes de electroforesis en gel

En la Figura 2.2 se puede observar las muestras por medio de una imagen de electroforesis en gel, en la cual P1, P2, ..., P44 corresponden a los números de muestras analizadas, las mismas se encuentran ubicadas dentro de los cuadros

específicos de la imagen.

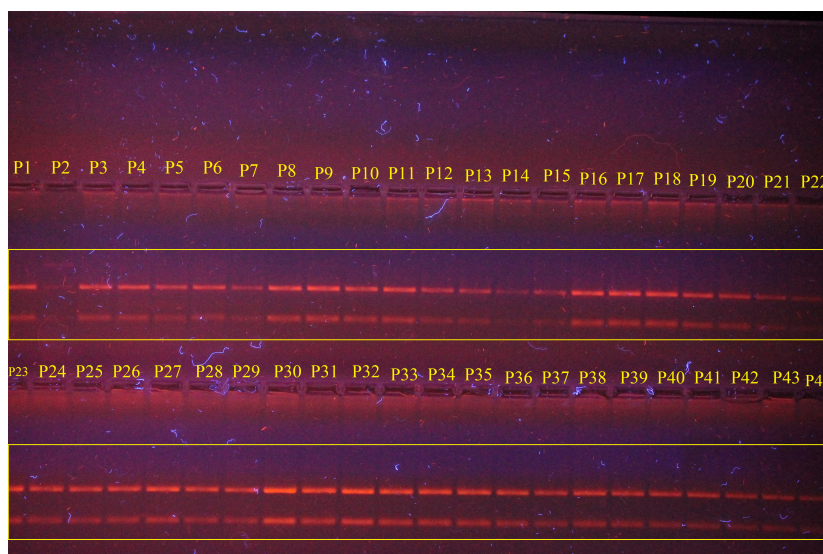


Figura 2.2: Posiciones de las muestras en una imagen de electroforesis en gel.

En la imagen obtenida, cada muestra es analizada en busca de la mutación germinal. Las muestras son identificadas correctamente con las dos últimas bandas en donde se encuentran ubicadas en la fila correspondiente. La línea ubicada por encima de las dos últimas bandas, corresponde a la detección de la mutación. Los casos estudiados de manera general para el análisis de la mutación son los siguientes:

- **Caso positivo:** Una muestra contiene tres bandas, donde la primera banda contiene un grosor más ancho que las dos bandas de abajo.
- **Caso negativo:** Una muestra contiene dos bandas identificadas correctamente por debajo de la línea que identifica a la mutación.
- **Caso sospechoso:** Una muestra contiene 3 bandas, donde la primera banda no posee un grosor con mayor anchura que las siguientes bandas de la muestra.
- **Caso erróneo:** Una muestra tiene resultado erróneo en cualquiera de los siguientes casos:
 - a- Posee una sola banda en cualquiera de las tres posiciones
 - b- Posee dos bandas, donde la primera banda se encuentra ubicada en la posición que identifica a la mutación y la siguiente en una de las dos posiciones restantes.

La Figura 2.3 muestra un ejemplo de una imagen típica de electroforesis en gel a ser analizada.

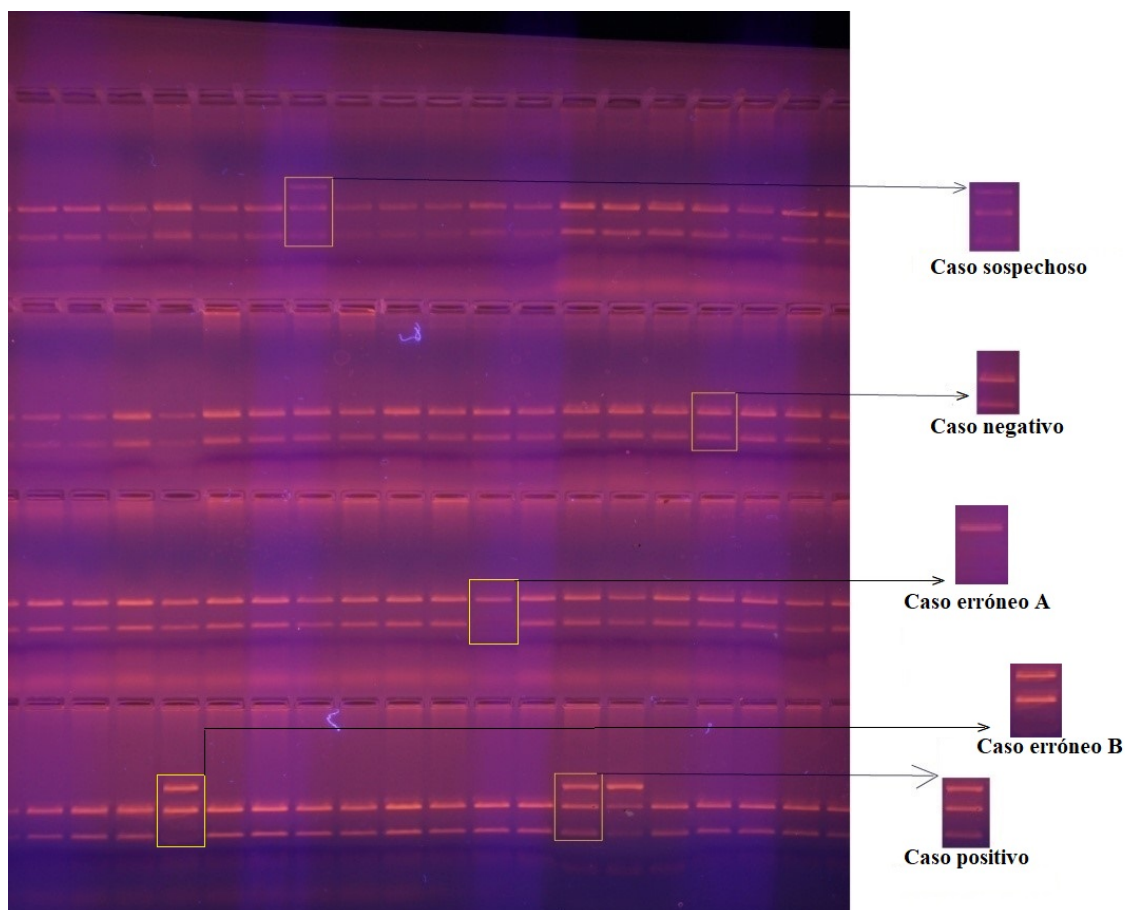


Figura 2.3: Criterios para las imágenes de electroforesis en gel.

El problema con algunas imágenes de electroforesis en gel, consiste en la contaminación por objetos indeseados y el bajo contraste, como consecuente, esto dificulta la percepción de las muestras al ojo humano, que produce limitaciones en el estudio y análisis de las imágenes, que en el peor de los casos, podría concluir con un diagnóstico erróneo. La Figura 2.4 muestra un ejemplo de una imagen de electroforesis en gel con bajo contraste y contaminación por objetos indeseados.

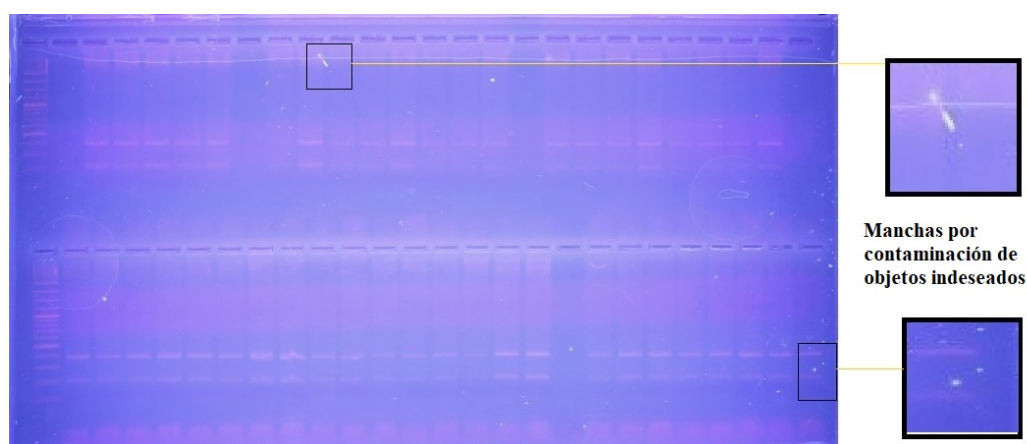


Figura 2.4: Imagen de electroforesis en gel con bajo contraste y contaminado por objetos indeseados.

La contaminación de la imagen por objetos indeseados suele propagarse por razones como el intercambio de iluminaciones entre la luz ultravioleta y el flash de la cámara emitidos en la captura de la imagen, así también por otra razón como la contaminación misma de los objetos dentro del gel.

En ocasiones, las muestras analizadas, no logran identificarse a simple vista, esto podría ser causa de varios factores en procesos anteriores, algunas como error de las muestras en la amplificación, el ADN presenta contaminación o es de mala calidad, fragmentos de PCR mal purificados. A consecuencia generan dificultades en la interpretación y posterior conclusión en el diagnóstico.

Resumen

En este capítulo se detallaron aspectos básicos sobre la obtención de las imágenes de electroforesis en gel, el estudio y el análisis de las mismas. El proceso del estudio de las muestras fue definido junto con la técnica de PCR RFLP a partir de la toma de sangre por cada muestra. Por último se detallaron los criterios y el análisis que se tuvieron en cuenta para el diagnóstico de los resultados.

Capítulo 3

CONCEPTOS BÁSICOS DE VISIÓN POR COMPUTADORA

Visión por computadora persigue asimilar al sistema visual humano, incorporando a los sistemas capacidades de aprendizaje, reconocimiento de patrones, inferencia y toma de decisiones basándose en el contenido de imágenes. En este capítulo se detallan las bases teóricas visión por computadora, las técnicas que son utilizadas para la segmentación de las imágenes resultantes del estudio de PCR RFLP para la mutación R337H en tp53, incluye como la morfología matemática en niveles de gris, binarización mediante una técnica de agrupamiento difuso conocida como Fuzzy C-Means Algorithm y la Morfología matemática en imágenes binarias.

3.1 Imagen Digital

Una imagen digital a niveles de gris puede definirse como una función $I : \mathbb{Z}^2 \rightarrow \mathbb{Z}$, en la cual (x, y) son coordenadas espaciales y los valores en estos puntos son cantidades enteras, discretas y están acotados.

Los valores de I en cualquier par de coordenadas (x, y) son las intensidades de la imagen en esos puntos. El número de píxeles de ancho y alto define la resolución de la imagen [GW06].

La Figura 3.1 muestra el ejemplo de una imagen digital en niveles de gris con una resolución de 256×256 transformado en 3 resoluciones distintas (la Figura 3.1.(a) con una resolución de 128×128 , la Figura 3.1.(b) con una resolución de 64×64 , la Figura 3.1.(c) con una resolución de 32×32 , la Figura 3.1.(d) con una resolución de 16×16).

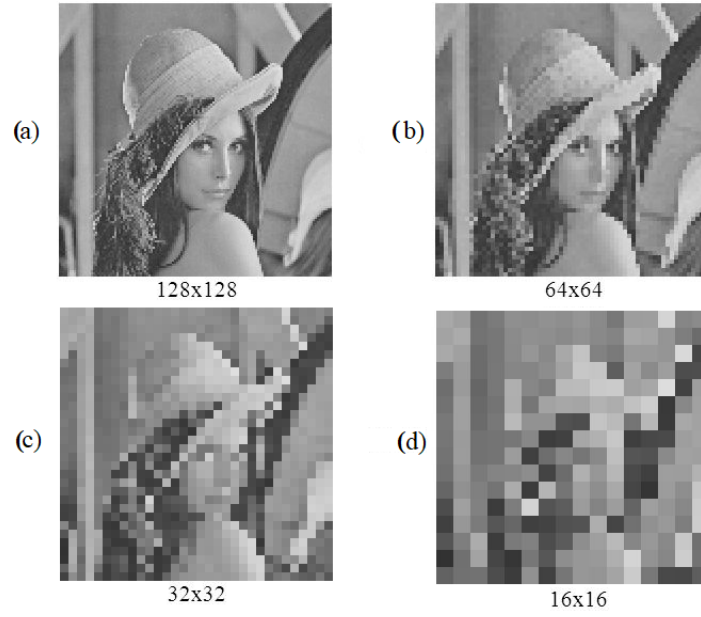


Figura 3.1: Imagen digital en 4 tamaños de resoluciones (a) imagen con resolución 128×128 (b) imagen con resolución 64×64 (c) imagen con resolución 32×32 (d) imagen con resolución 16×16 .

Una imagen digital en niveles de gris de $M \times N$, puede ser representada de manera matricial tal que M representa a las columnas, N representa a las filas y cada elemento de $I(x, y) \in \mathbb{Z}$ es un valor discreto que es conocido como la intensidad o nivel de gris en ese punto.

En la Figura 3.2 muestra la representación de una imagen digital en forma matricial.

$$I(x, y) = \begin{bmatrix} I(0, 0) & I(0, 1) & \cdots & I(0, N - 1) \\ I(1, 0) & I(1, 1) & \cdots & I(1, N - 1) \\ \vdots & \vdots & & \vdots \\ I(M - 1, 0) & I(M - 1, 1) & \cdots & I(M - 1, N - 1) \end{bmatrix}$$

Figura 3.2: Representación de una imagen digital en forma matricial.

La codificación de un píxel es definido por la cantidad de bits utilizados para almacenar información del nivel de gris de cada píxel. Así, una imagen binaria utiliza 1 bit para almacenar 2 posibles valores de intensidad, el valor 0 que corresponde al color negro o el valor 1 que corresponde al color blanco. La Figura 3.3 muestra un ejemplo de una imagen binaria.

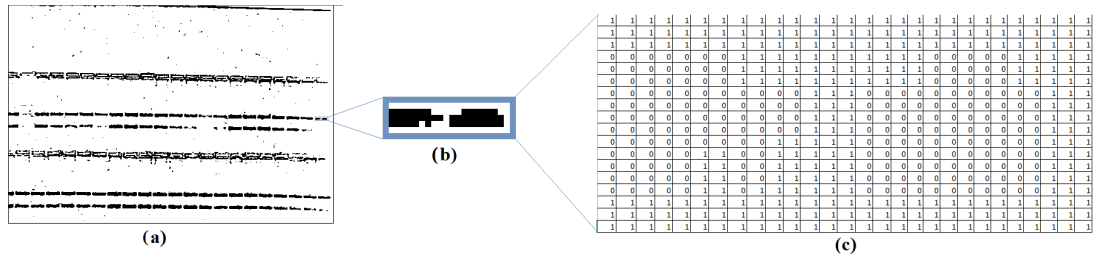


Figura 3.3: (a) Imagen digital binaria original, (b) segmento de la imagen original, (c) valores de intensidad que componen los píxeles del segmento.

Comúnmente, una imagen en niveles de gris emplea 1 byte (8 bits), capacidad para almacenar información en un píxel hasta 256 valores de intensidades o niveles de gris, donde el primer valor corresponde al 0 con un color oscuro hasta el valor 255 al color más claro.

La Figura 3.4 muestra el ejemplo de una imagen digital en niveles de gris con 8 bits de almacenamiento en cada píxel, que parte de la imagen es tomada para observar los valores de los píxeles.

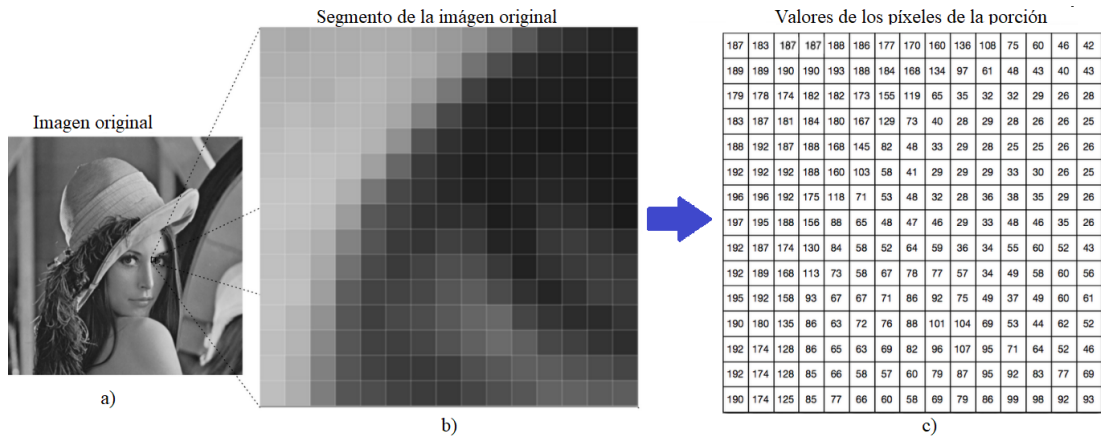


Figura 3.4: (a) Imagen digital en niveles de gris original, (b) segmento de la imagen original, (c) valores de intensidad que componen los píxeles del segmento.

Una imagen en color RGB combina los tres colores primarios identificados por el rojo, verde y azul, la cual es representada como una matriz $M \times N \times 3$. Un píxel de color de este modelo corresponde a la combinación de los píxeles de los canales de color Rojo, Azul y Verde (R, G, B), la cual conforman una tripleta de sus componentes como muestra la Figura 3.5

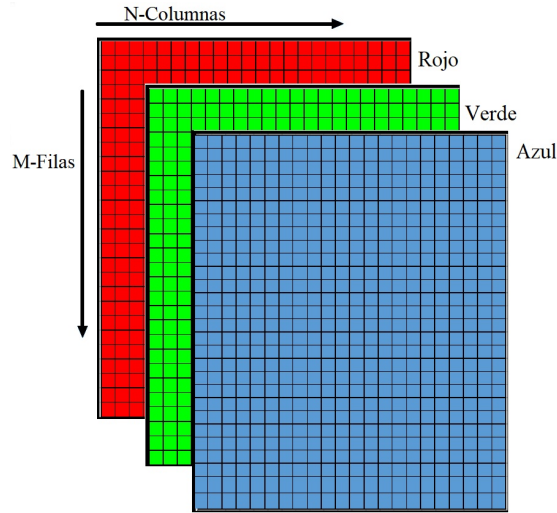


Figura 3.5: Representación una imagen de modelo R.G.B.

Para una imagen, un píxel de 24 bits en total, denota un color en una imagen RGB, en que el número total de colores corresponde a $(2^8)^3 = 16.777.216$. Cada componente está representado por una matriz con intensidades en escalas de grises entre los valores $0, 1, \dots, 255$ [GW06].

La Figura 3.6 muestra el ejemplo de una imagen de electroforesis en gel en el formato RGB con sus 3 canales correspondientes.

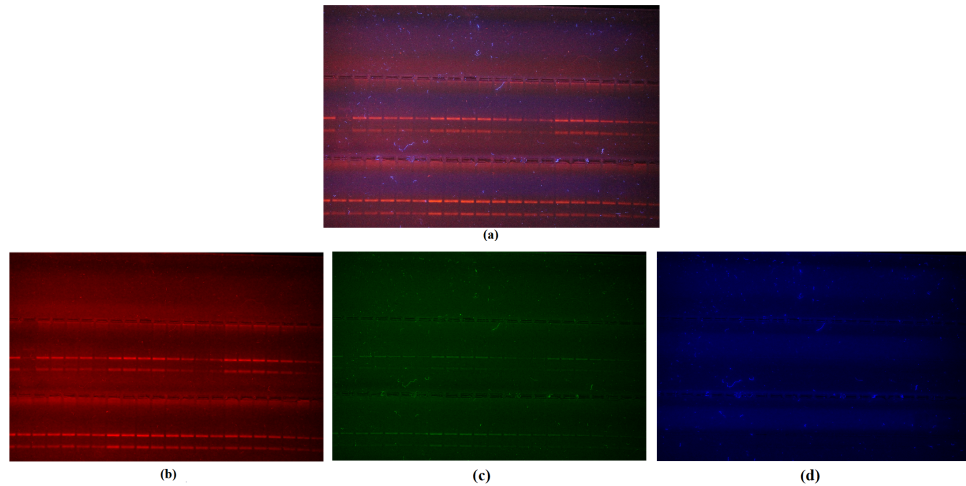


Figura 3.6: (a) Imagen resultante de la electroforesis en gel original en formato RGB, (b) Canal rojo de la imagen, (c) Canal verde de la imagen (d) canal azul de la imagen

Con el fin de transformar una imagen a una más significativa, la misma es convertida a una imagen binaria. Con esto se consigue separar los objetos de interés del fondo. Los métodos de valor umbral consisten en algoritmos que tiene

como principal objetivo separar objetos de mayor interés en una imagen. En el caso de las imágenes de electroforesis en gel, con los métodos umbrales es posible detectar las muestras y separarlos del fondo.

3.2 Umbralización de imagen

Las imágenes con objetos claros sobre un fondo oscuro resultan sencillos de encontrar el umbral a partir del histograma ya que ocupan áreas comparables en tamaño, de tal modo que todos los píxeles que se encuentren por debajo del umbral corresponderán al fondo y todos los píxeles que se encuentren por encima del umbral consistirán en los objetos de la imagen. Sin embargo, en algunas imágenes el histograma no resulta óptimo para ser binarizada con la técnica convencional, existen técnicas de umbralización avanzadas con métodos globales y locales que son detalladas a continuación.

3.2.1 Umbralización global y local

Las umbralizaciones globales consisten en buscar un único valor umbral para la binarización de la imagen. Este tipo de umbralización se muestra efectivo cuando el histograma es bimodal. Las técnicas de umbralizaciones globales son efectivas con imágenes que contienen mucho contraste, donde el histograma resulta bimodal (cuando contiene dos picos de intensidades que diferencian el fondo y los objetos de la imagen).

En la Figura 3.7 muestra una imagen en niveles de gris con su correspondiente histograma, en la misma puede observarse claramente un histograma bimodal, que a partir de un umbral es posible llevar a cabo la binarización de la imagen.

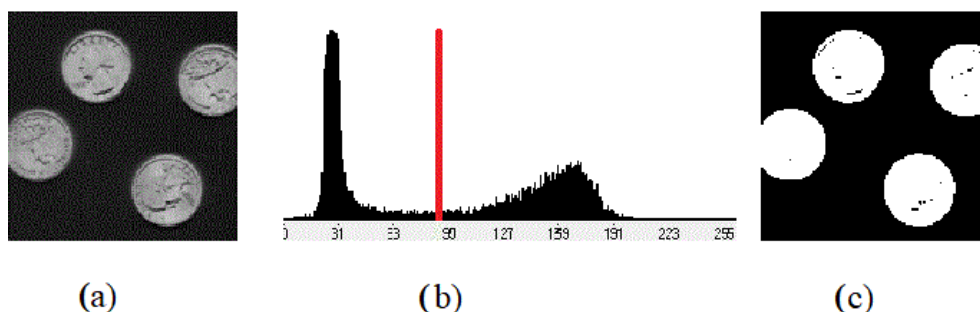


Figura 3.7: (a) Imagen original en niveles de gris, (b) Histograma de la imagen con un umbral y (c) Imagen binaria a partir de la umbralización.) canal azul de la imagen

Cuando un histograma tiene una distribución no bimodal, los algoritmos globales no resultan buenos y no es trivial encontrar un umbral óptimo. La Figura 3.8 muestra una imagen en niveles de gris con su correspondiente histograma, con un nivel de umbral en 127 se puede observar que no es posible separar el fondo de la moneda.

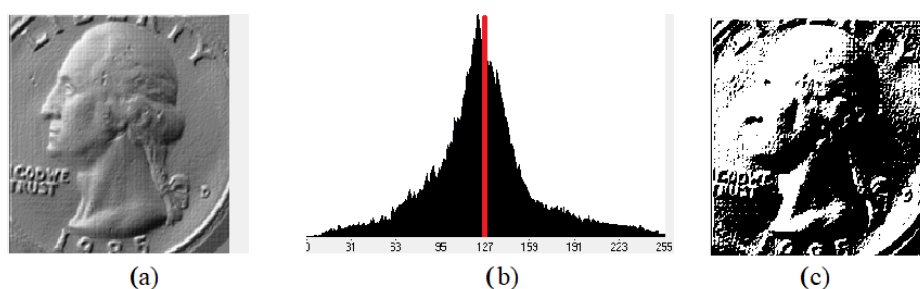


Figura 3.8: (a) Imagen original en niveles de gris, (b) Histograma de la imagen con un umbral en 127 (c) Imagen binaria a partir de la umbralización.

Las umbralizaciones locales consisten en dividir una imagen en partes y calcular el valor umbral para cada una de las divisiones. El método es denominado adaptativo, debido a que el umbral es calculado en cada división o región de la imagen, es decir, cada sub imagen contiene su propio umbral, que es seleccionado de los píxeles que lo componen.

La umbralización permite binarizar las imágenes a partir de uno o varios umbrales logrando separar los objetos de interés del fondo de la imagen, para obtener una imagen limpia de toda contaminación de objetos no deseados para procesarla, es necesario llevar a cabo a filtros morfológicos que se describen a continuación.

3.3 Morfología matemática

La morfología matemática consiste en extraer informaciones, como las geometrías y las topologías de conjuntos desconocidos de una imagen por la transformación a partir de otro conjunto completamente definido denominado elemento estructurante. A partir de su estructura y tamaño, podemos probar y cuantificar, de qué manera el elemento estructurante está o no está incluido en la imagen [Pei83].

3.3.1 Elemento estructurante

Un elemento estructurante es un conjunto binario completamente definido en tamaño y forma, conocido por el computador, la cual, está definido por los píxeles

que lo forman. Todo elemento estructurante tiene un centro definido, conocido y que dependiendo de la operación realizada, es el centro que se encarga de pintar el píxel de la imagen con el valor correspondiente.

El centro del elemento estructurante dentro de la imagen, definirá (dependiendo de la operación que se realice) la nueva forma que tendrá la imagen cuando las regiones de la imagen son alcanzados por los puntos del elemento estructurante.

La Figura 3.9 muestra algunos ejemplos de elementos estructurante con sus tamaños y formas definidas.

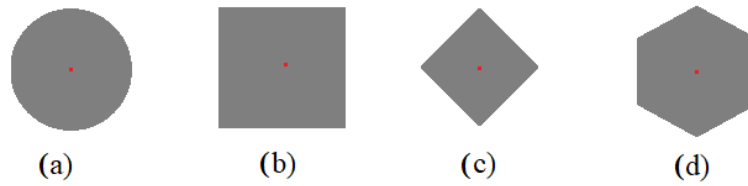


Figura 3.9: Elementos estructurantes con las siguientes formas, (a) forma circular, (b) forma cuadrada, (c) forma romboide (d) forma hexagonal. Los puntos rojos indican el centro del elemento.

Las operaciones básicas de morfología fueron introducidas por Matheron y Serra [Ser91], [Mat67], para examinar la estructura de imágenes binarias que pueden ser representados matemáticamente como conjuntos.

3.3.2 Morfología binaria

La idea básica de la morfología matemática en imágenes binarias, es analizar un conjunto con una forma predefinida simple conocido como elemento estructurante sobre otro conjunto correspondiente a la imagen a ser procesada, que son aquellas que tienen dos niveles de intensidad blanco y negro. Las operaciones de morfología están definidas generalmente por la erosión, dilatación, apertura y cierre que en esta sección serán detalladas.

3.3.3 Erosión y dilatación

Sea I una imagen binaria dada y el conjunto B el elemento estructurante, la erosión de I por B denotada por $e_B(I)$ y se define como:

$$e_B(I) = \bigcap_{(b \in B)} I_{(-b)}, \quad (3.1)$$

donde la erosión de I por B puede comprenderse como la intersección de los puntos alcanzados por el centro de B cuando es desplazado dentro de I y $-b$ es la reflexión de b en I . Si el valor del píxel de la imagen I que es alcanzado por el centro de B tiene valor 0 (negro), la erosión modifica el valor en 1 (blanco).

La Figura 3.10 muestra un ejemplo la erosión de una imagen I por un elemento estructural B , la cual es una elemento estructural simétrico (Cuando B es igual a su reflexión $-B$). Puede observarse que con la operación de la erosión, el objeto disminuye en tamaño y forma pintando los píxeles negros en blanco cuando el centro de B es alcanzado por un píxel en la imagen I .

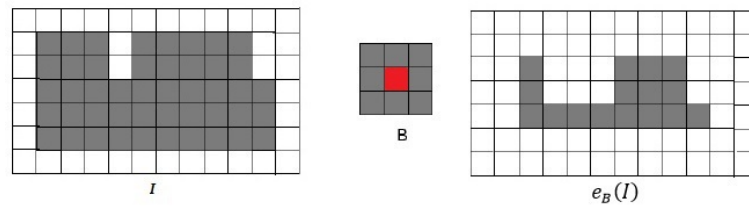


Figura 3.10: Erosión de una imagen I por un elemento estructural B

La erosión tiene como finalidad eliminar zonas de la imagen donde los píxeles que son alcanzados por el centro del elemento estructural intersectan entre sí.

La Figura 3.11 muestra un ejemplo de una imagen de electroforesis en gel en binario, la cual es erosionada por un elemento estructural de 12 píxeles horizontales. Así en la imagen puede observarse como los objetos más pequeños al elemento estructural son eliminados y los objetos son reducidos de igual manera 12 píxeles menos.

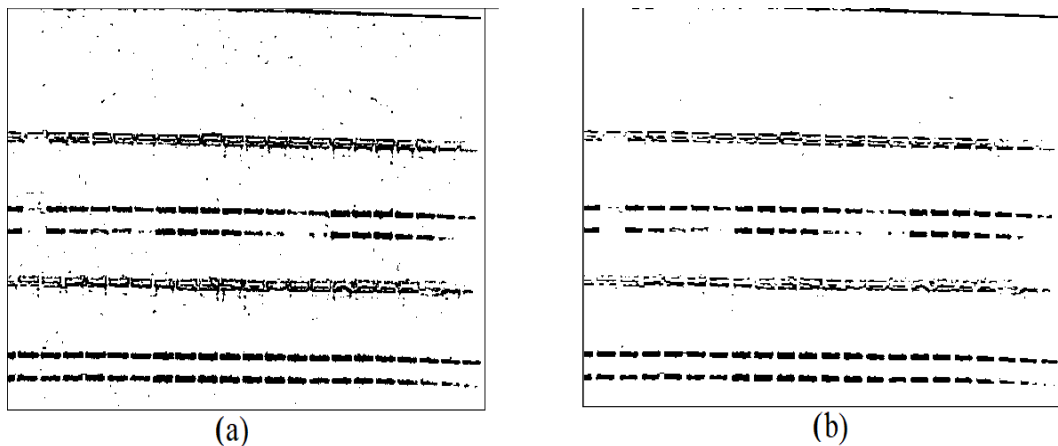


Figura 3.11: a) Imagen de electroforesis en gel original b) Resultado de la erosión

En la operación binaria, con el elemento estructural se reduce la cantidad de píxeles de los objetos que contienen valores 0 (negro) en sus píxeles pintándolos

en 1 (blanco). La operación de la dilatación opera de manera inversa.

Sea I una imagen binaria dada y B el elemento estructurante, la dilatación de I por B es denotada por $\delta_B(I)$ y se define como:

$$e_B(I) = \bigcup_{(b \in B)} I_{(b)}. \quad (3.2)$$

Cuando el centro de B es desplazado en I y encuentra un píxel con valor 0 y uno de los píxeles de B es alcanzado con algún punto de I , entonces el valor de I en la posición del centro de B cambia en valor 1.

En la Figura 3.12 muestra el ejemplo de una imagen I dilatada por un elemento estructurante B .

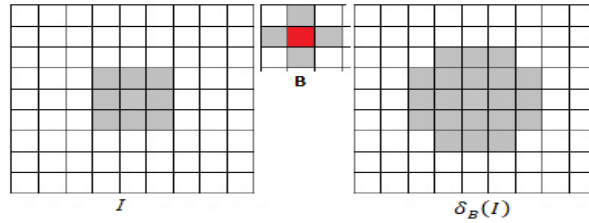


Figura 3.12: Dilatación de una imagen I por un elemento estructurante B

La finalidad de la operación de dilatación binaria es agregar píxeles pintados al objeto de la imagen cuando el centro del elemento estructurante es alcanzado por los píxeles del contorno del objeto de la imagen.

La Figura 3.13 muestra la imagen binaria de electroforesis en gel original y el resultado de la dilatación que fue realizada por un elemento estructurante tipo cruz con 5 píxeles. Se puede observar que en la imagen dilatada elementos aumentaron significativamente sus tamaños.

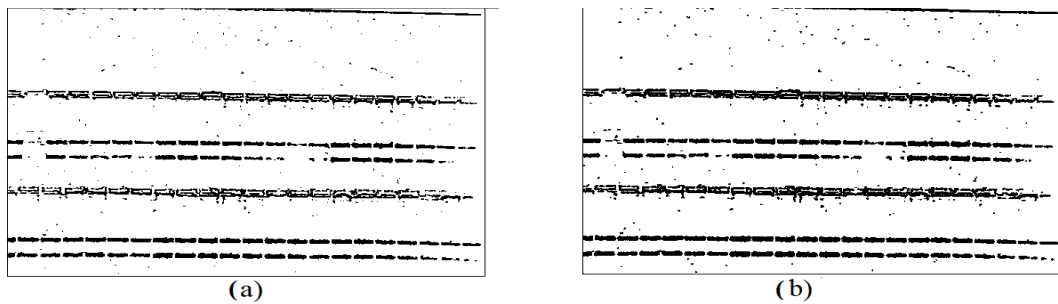


Figura 3.13: (a) Imagen de electroforesis en gel original (b) imagen resultante de la dilatación.

En las operaciones morfológicas la erosión y la dilatación, los objetos pierden

su forma y tamaño, sin embargo existen operaciones más complejas a partir de la erosión y la dilatación como la apertura y el cierre que están más centradas en las partículas dentro de la imagen.

3.3.4 Apertura y cierre

Sea I la imagen y B el elemento estructurante, la siguiente operación de apertura es denotada por $\gamma_B(I)$ y se define como:

$$\gamma_B(I) = \delta_B(e_B(I)). \quad (3.3)$$

La apertura es la dilatación (ver ecuación 3.2) de la erosión (ver ecuación 3.1) de un conjunto I con el mismo elemento estructurante B . La apertura de una imagen binaria es utilizada para eliminar partículas como objetos indeseables sin modificar el tamaño de los otros conjuntos.

En la Figura 3.14 muestra la apertura de una imagen I por un elemento estructurante B .

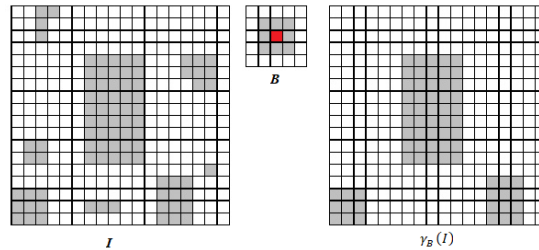


Figura 3.14: Apertura de una imagen I por un elemento estructurante B

La Figura 3.15 muestra la imagen binaria de electroforesis en gel original y la imagen resultante de la operación apertura con un elemento estructurante cuadrada de 9 píxeles, se puede observar como los pequeños elementos de las imágenes fueron eliminados, quedando así solamente los elementos con mayor tamaño al elemento estructurante.

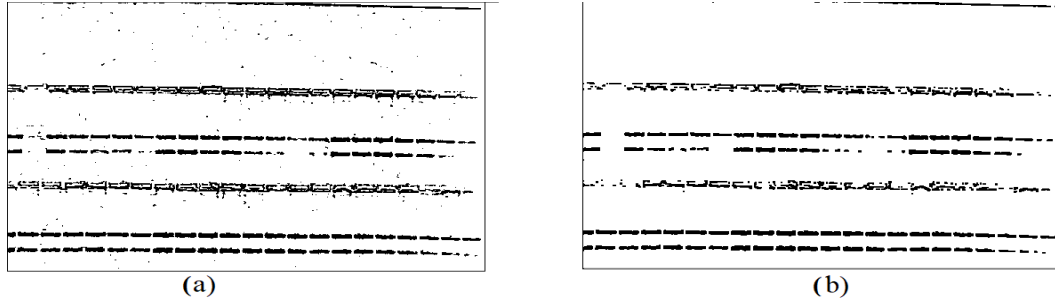


Figura 3.15: a) Imagen binaria de electroforesis en gel original b) imagen resultante de la apertura.

La operación de apertura elimina las partículas de la imagen, mientras que la operación de cierre morfológico opera de manera inversa.

Sea I la imagen y B el elemento estructurante, la siguiente operación de cierre morfológico es denotado por $\phi_B(I)$ y se define como:

$$\phi_B(I) = e_B(\delta_B(I)). \quad (3.4)$$

El cierre es la erosión 3.1 de la dilatación (ver ecuación 3.2) de una imagen I con el mismo elemento estructurante B . El cierre de una imagen binaria es utilizado para rellenar cavidades sin modificar el tamaño de los otros conjuntos.

En la Figura 3.16 muestra el cierre de una imagen I por un elemento estructurante B .

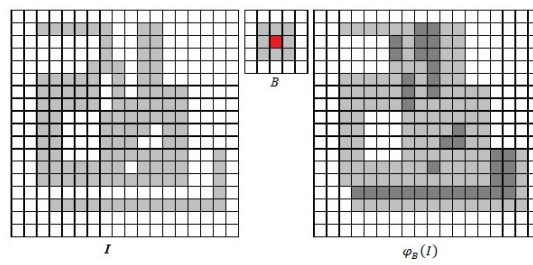


Figura 3.16: Cierre de una imagen I por un elemento estructurante B .

El operador de cierre morfológico es utilizado para llenar cavidades sin modificar el tamaño de los otros conjuntos.

La Figura 3.17 muestra una imagen binaria de electroforesis en gel con el resultado de una operación morfológica denominada cierre, así la imagen resultante de la operación morfológica, los objetos contienen más píxeles pintados con un grosor mayor.

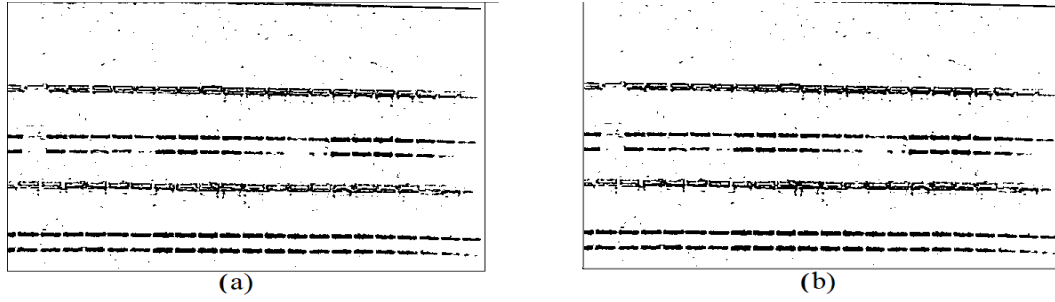


Figura 3.17: (a) Imagen binaria de electroforesis en gel original (b) imagen resultante del cierre morfológico.

3.3.5 Reconstrucción morfológica

La reconstrucción es un operador que sirve para extraer informaciones conectadas a partir de una imagen [Ser91] .

La reconstrucción es realizada utilizando el operador de dilatación (ver ecuación 3.2) a una imagen I a partir de un elemento estructurante B sobre una imagen X hasta la idempotencia.

$$\delta_I(X)^{i+1} = \delta X_B^{(i)} \wedge I, \quad \delta_I(X)^{i=0} = X, \quad (3.5)$$

donde $\wedge$ es el operador de intersección, I es la imagen original (llamado máscara), B el elemento estructurante, X la imagen a ser dilatada (llamado marcador) hasta la idempotencia, es decir $\delta_I(X)^{i+1} = \delta X^{(i)}$.

La Figura 3.18.(a) muestra una imagen de electroforesis en gel luego de una erosión con pequeños fragmentos y la Figura 3.18.(b) muestra la imagen resultante de una reconstrucción morfológica que fue dilatada condicionalmente a la Figura 3.18.(a).

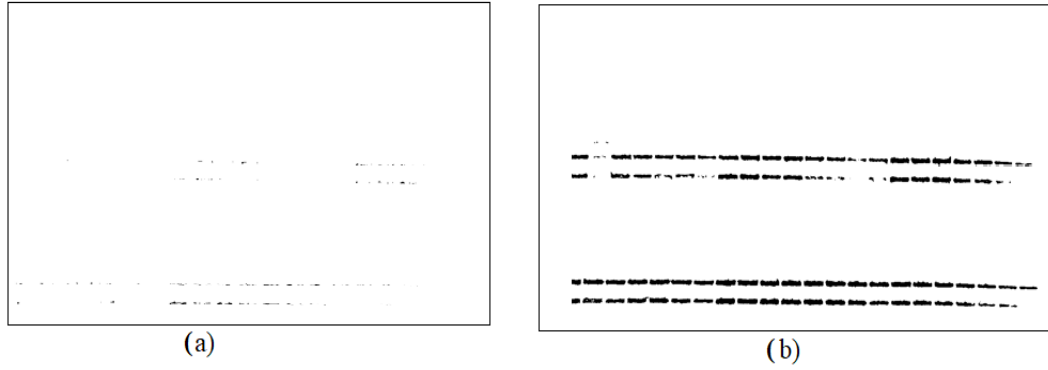


Figura 3.18: (a) Imagen binaria de electroforesis en gel. (b) Imagen resultante luego de una reconstrucción.

3.4 Morfología matemática en imágenes a niveles de gris

En esta sección se detallan algunas de las técnicas de la morfología matemática en niveles de gris, como la erosión, dilatación, apertura y cierre. La morfología matemática en imágenes con intensidades en niveles de gris es una generalización de las imágenes binarias.

3.4.1 Erosión y dilatación

Sea I una imagen dada y B el elemento estructurante, la erosión de I por B es denotada por $e_B(I)$ y la definición está dada por:

$$e_B(I)_{(p)} = \min\{I(p + b) | b \in B\}, \quad (3.6)$$

donde $p = I(x, y)$ corresponde al pixel erosionado por el centro de B en I y $\min$ corresponde al valor mínimo del conjunto de píxeles ubicados en B y que pertenecen a I . Con la operación morfológica en escalas de grises, la imagen resultante suele ser más oscura que la original.

La Figura 3.19 muestra una imagen de electroforesis en gel en niveles de gris y el resultado de la operación morfológica denominada erosión.

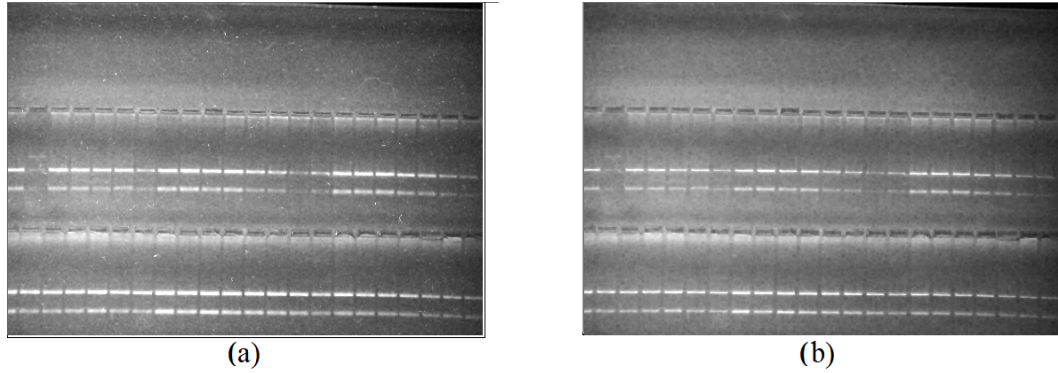


Figura 3.19: a) Imagen en niveles de gris resultante de la electroforesis en gel. b) Imagen resultante de la erosión.

La dilatación en imágenes de niveles de gris, es la operación morfológica dual de erosión.

Sea I una imagen en escala de grises y la imagen B el elemento estructurante, la dilatación de I por B es denotada por:

$$\delta_B(I)_{(p)} = \max\{I(p+b)|b \in B\}, \quad (3.7)$$

donde p corresponde al pixel dilatado por el centro de B en I y $\max$ corresponde al valor máximo del conjunto de píxeles ubicados en B y que pertenecen a I .

En la dilatación en escalas de grises, los valores de la imagen dilatada, resultan tener tonos más claros que los valores de los píxeles de la imagen original.

La Figura 3.20 muestra una imagen de electroforesis en gel en niveles de gris y el resultado de la operación morfológica denominada dilatación, así en la imagen operada puede observarse que los pequeños píxeles tienen un grosor superior y los valores de píxeles de los objetos tienen una intensidad más clara.

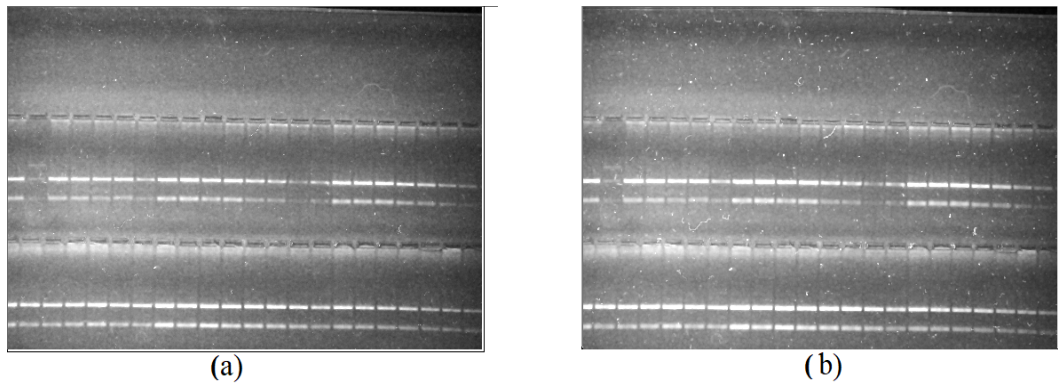


Figura 3.20: (a) Imagen en niveles de gris resultante de la electroforesis en gel. (b) Imagen resultante de la dilatación.

La apertura y el cierre morfológico en niveles de gris puede utilizar la misma definición que los operadores de apertura y cierre binario (Ver ecuación 3.3 para apertura y ecuación 3.4 para cierre en imágenes binarias)

Para la operación morfológica apertura, la definición es la misma que apertura morfológica en imágenes binarias

La apertura morfológica de una imagen en niveles de gris consiste en erosionar una imagen I (ver ecuación 3.6) por un elemento estructurante B , seguidamente dilatar (ver ecuación 3.7) la imagen erosionada por el mismo elemento estructurante B . Las imágenes operadas por la apertura morfológica permiten eliminar partículas que las contaminan.

La Figura 3.20 muestra una imagen de electroforesis en gel en escalas de grises junto con la imagen resultante de la apertura morfológica.

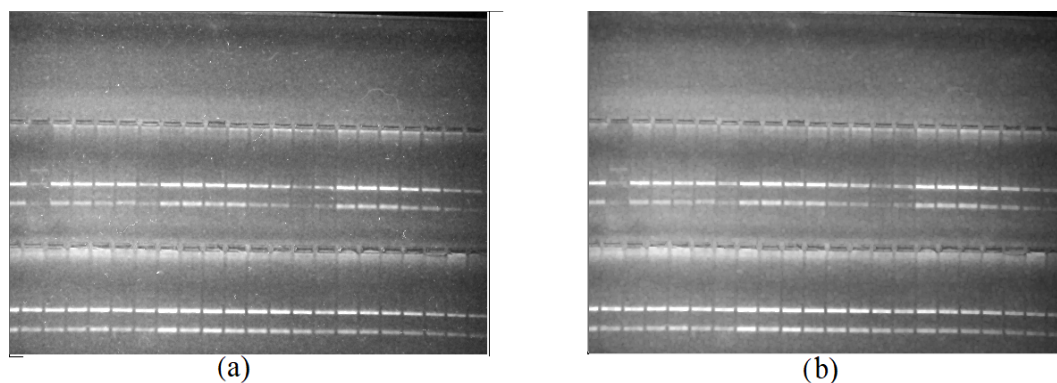


Figura 3.21: (a) Imagen en niveles de gris resultante de la electroforesis en gel. (b) Imagen resultante de la apertura.

El cierre morfológico de una imagen en niveles de gris consiste en dilatar (ver ecuación 3.7) una imagen I por un elemento estructurante B , seguidamente erosionar (ver ecuación 3.6) la imagen dilatada por el mismo elemento estructurante B . Las imágenes operadas por el cierre morfológico permiten rellenar los espacios claros con los valores más oscuros del elemento estructurante.

La Figura 3.22 muestra una imagen de electroforesis en gel en escalas de grises junto con la imagen resultante del cierre morfológico.

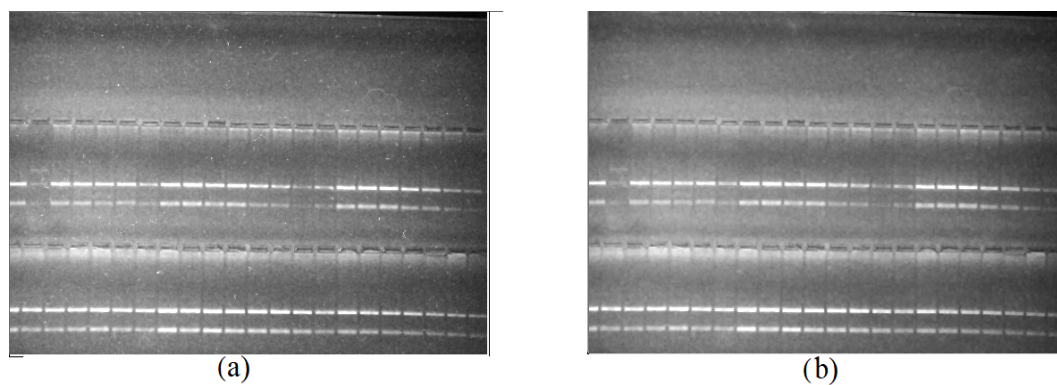


Figura 3.22: (a) Imagen en niveles de gris resultante de la electroforesis en gel. (b) Imagen resultante del cierre morfológico.

Resumen

En este capítulo se presentaron las técnicas de visión por computador, tales como umbralización de imágenes entre locales y globales, morfología matemática en niveles binarios y niveles de gris que serán utilizados en el procesamiento de las imágenes de electroforesis en gel en la propuesta.

Capítulo 4

ABORDAJE PROPUESTO

En este capítulo se describen detalladamente el abordaje y las técnicas utilizadas para lograr el diagnóstico de la mutación tp53 R337H a partir de las imágenes en gel por electroforesis. Con el abordaje se pretende llegar a detectar casos de la mutación germinal tp53 R337H.

4.1 Pseudocódigo

En esta sección es detallado el pseudocódigo de la propuesta, la cual recibe como parámetro una imagen en formato RGB y retorna un vector de 4 posiciones correspondiente al diagnóstico final (positivo, negativo, sospechoso y erróneo).

El algoritmo 1 tiene como entrada una imagen de electroforesis en gel en formato RGB y retorna una imagen segmentada.

Algoritmo 1 Segmentación de la imagen

Entrada: Imagen inicial en RGB X

- 1: $R = X[R]$ Se obtiene el canal rojo
 - 2: $E = erosionar(R, B)$ se erosiona la imagen R con un elemento estructurante cuadrangular B (consultar erosión 3.6)
 - 3: $obj = substraer(E, R)$; Se realiza la substracción de las imagenes $E - R$
 - 4: $FilApe = apertura(obj, B)$; Se procesa la apertura de la imagen obj con un elemento estructurante cuadrangular B (consultar ecuación 3.3)
 - 5: $ImagBin = binarizar(FilApe)$; Se binariza la imagen
 - 6: $Fragmentos = ErosionBinaria(ImagBin, B)$; Se filtra la imagen $ImagBin$ con un elemento estructurante cuadrangular B (consultar ecuación 3.1)
 - 7: $Imagen = Reconstruccion(Fragmentos, ImagBin, B)$; Se procede a la reconstrucción del marcador $Fragmentos$ mediante la máscara $ImagBin$ por un elemento estructurante B (Consultar ecuación 3.5)
 - 8: **devolver** $Imagen$
-

El siguiente algoritmo 2 es una función que recibe la imagen segmentada y

retorna las posiciones de los carriles que contienen las bandas.

Algoritmo 2 Detección de los carriles

Entrada: *Imagen*

```

     $carInf_k = 0$ ; valor fila de la posición inicial del carril
2:  $carSup_k[2] = 0$ ; valores fila de la posición final del carril
     $aux = 0$ ; auxiliar para detectar carriles
4:  $findCarril = 0$ ; contador de carriles
    para  $y = 0$  a  $Imagen.tamanoFila$  hacer
6:     para  $x = 0$  a  $Imagen.tamanoColumna$  hacer
            si  $Imagen(x, y) == 0$  entonces
8:                 Se obtienen los píxeles  $I(y, x)$  para cada carril
                    si  $CarInf_k == 0$  entonces
10:                         $CarInf_k = y$  coordenada espacial para la fila inicial del carril
                            fin si
12:                         $findCarril++$ ; aumenta el número de carriles
                            fin si
14:     fin para
        fin para
16: devolver  $CarInf_k, CarSup_k$ 

```

El algoritmo 3 detecta las bandas a partir de la imagen recibida como parámetro y las coordenadas de los carriles que contienen a las bandas.

El algoritmo 4 es el principal que recibe una imagen como parámetro, llama a sus respectivas funciones y retorna el resultado del diagnóstico de la detección de casos de mutación.

Algoritmo 3 Detección de bandas

Entrada: $Imagen, CarInf_k, CarSup_k$,
si Parámetros no están nulos **entonces**
 para $y = 0$ a $CarInf$ **hacer**
3: **para** $x = 0$ a $CarSup$ **hacer**
 si $Imagen(x, y) = 0$ **and** $Imagen(proxCol) = 0$ **entonces**
 si $bandinf_k = 0$ **entonces**
6: Almacena primera posición $bandinf_k$
 si $(bandsup_k - 1) \neq 0$ **entonces**
 almacenar tamaño de espacio entre bandas $PromEspHor$
9: **fin si**
 fin si
 si $bandinf_k \neq 0$ **entonces**
12: $cantBand++$ y Almacena posición $bandsup_k$ y calcula tamaño de
 banda $SumtamBanda = SumtamBanda + (bandsup_k - bandinf_k)$
 fin si
 fin si
15: **fin para**
 fin para
 devolver $catBand, bandsup, sumtamBanda$
18: **si no**
 devolver TRUE
 fin si

4.2 Propuesta general

La propuesta está basada en la detección de la mutación germinal tp53 R337H siguiendo unas instrucciones que son detallados en esta sección.

Los geles de agarosa, que contienen las muestras analizadas, son posicionados sobre un transiluminador con luz UV para obtener las imágenes de electroforesis en gel mediante una cámara digital que captura imágenes en formato RGB. Como muestra La Figura 4.5.(a), es necesario separar el fondo y los objetos que lo conforman.

1- Procesamiento: Las imágenes de electroforesis en gel que fueron obtenidas de una base de datos están con el formato RGB. El color de las muestras en las imágenes es la misma en su totalidad, pueden ser observadas que el rojo es superpuesta al fondo de la imagen.

El canal con el mayor resaltado de las muestras que son los objetos de interés a obtener, es seleccionado (Canal Rojo), las mismas son observadas con más detalles para ser procesadas como puede observarse en la Figura 4.5.(b).

Algoritmo 4 Algoritmo principal

Entrada: X
 $Diagnosticos[4]$; Vector de 4 valores para el diagnóstico final

 $PromEspHor = 0$; promedio tamaño del espacio entre bandas

 $cantBand = 0$; Cantidad de bandas detectadas en la imagen

4: $SumtamBanda = 0$; Sumatoria de tamaño de bandas de la imagen

 $Imagen = Segmentacion(X)$; Llama a la función, (Ver algoritmo 1)

 $Imagen = DetectarCarril(Imagen)$; Llama a la función, retorna $CarInf, CarSup$ (ver algoritmo 2)

 $DetectarBandas(Imagen, CarInf, CarSup)$; Llama a la función y retorna $SumtamBanda, canBand$, (ver algoritmo 3)

8: $PromptamBand = \frac{\sum cantBand}{sumtamBanda}$

Se estima una máscara k desde los promedios de $PromEspHor$ y $PromptamBand$
para $x = 0$ a $Imagen.height$ **hacer**
para $y = 0$ a $Imagen.width$ **hacer**

12: **si** $Imagen(x, y) = k(x, y)$ **entonces**

Detectar bandas y diagnosticar casos $Diagnosticos$,

fin si
fin para

16: **fin para**
Salida:
 $Diagnosticos$ Un vector con 4 valores resultados del diagnóstico final

La separación de los objetos y el fondo de la imagen son realizados mediante técnicas de morfología matemática. La técnica específica es la erosión (ecuación 3.6). Que debido a la forma de los objetos de la imagen (las muestras), un elemento estructurante cuadrado de 9×9 es seleccionado, donde el elemento es desplazado sobre la imagen con 8 iteraciones. Con la erosión (ecuación 3.7) se consigue eliminar los objetos de la imagen, obteniendo solo el fondo como muestra en la Figura 4.5.(c).

La operación de substracción es realizada entre la Figura 4.5.(c) y la imagen original de canal rojo Figura 4.5.(b), con la misma técnica los objetos y el fondo de la imagen son separados tal como muestra en la Figura 4.5.(d).

La Figura 4.5.(d) es procesada con la operación de apertura morfológica (ecuación 3.3) en niveles de gris con un elemento estructurante cuadrado de 9×9 en 2 iteraciones, el proceso es realizado para la disminución de los objetos indeseados que generan contaminación en las imágenes.

2- Segmentación: El problema de la umbralización para las imágenes de electroforesis ADN en gel, consiste en la contaminación de objetos indeseados, y como consecuencia, sus histogramas son difíciles de analizar, en su mayoría, no poseen una marcada separación de regiones.

La solución está basado en la identificación de un umbral óptimo para la separación de los objetos y el fondo de la imagen. La binarización para imágenes de electroforesis en gel es realizada con el método del algoritmo de Clustering Fuzzy C-Means (anexo 6.2) con 2 agrupamientos correspondientes al objeto y al fondo, el resultado de la binarización se muestra en la Figura 4.5.(e).

3- Filtro morfológico: Debido a la contaminación de objetos indeseados que aparecen, la imagen binarizada es erosionada (ecuación 3.1) con 15 iteraciones con un elemento estructurante de tipo cruz de acuerdo a la forma de los objetos de interés. Con la operación se logra eliminar todos los elementos dejando pequeños fragmentos, las cuales corresponden a las muestras analizadas como muestra la Figura 4.5.(f).

La dilatación condicional a partir de la imagen original Figura 4.5.(b) es procesada con un elemento estructurante cruz a fin de reconstruir la imagen obteniendo solo las muestras a ser procesadas. La figura 4.5.(g) muestra la imagen segmenta y filtrada.

- 4- Estimación de máscara:** El número de muestras y sus posiciones correspondientes pueden ser determinados buscando los píxeles que lo contienen. Sin embargo, debido a que algunos pozos del gel, pueden estar vacíos (sin muestras analizadas) o también mal procesadas, lo que las mismas bandas pueden no ser identificados en las imágenes de electroforesis en gel, esto podría llevar a conclusiones no deseadas con el diagnóstico.

La imagen es escaneada de izquierda a derecha en cada fila en busca de píxeles marcados por las bandas de cada muestra para almacenar las coordenadas *INF* inferior y *SUP* superior para obtener el carril que las contiene, ejemplo en la Figura 4.1.

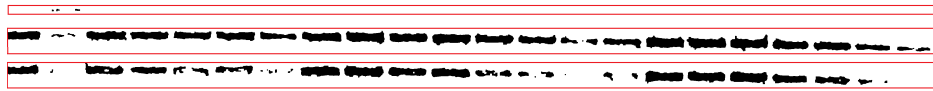


Figura 4.1: Carriles horizontales que contienen las bandas de cada muestra

La imagen es escaneada de arriba hacia abajo dentro de cada carril para calcular el tamaño promedio de las bandas Figura 4.2, así como las distancias de los espacios que separan cada banda de forma horizontal, Figura 4.3.

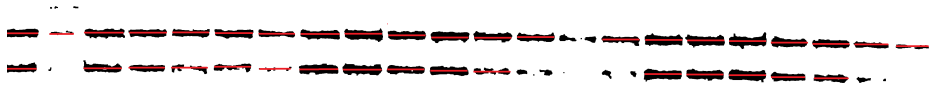


Figura 4.2: Estimación del promedio del tamaño horizontal de las bandas

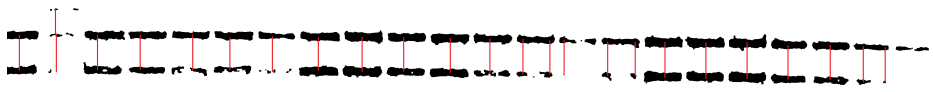


Figura 4.3: Estimación del promedio de tamaño del espacio vertical de las bandas

Identificando las coordenadas de la primera banda para cada carril horizontal, se lleva a cabo la estimación de la máscara. La estimación de la máscara K está dada por el cálculo de las posiciones del primer carril, con el uso del tamaño promedio para las bandas y los espacios que los separan, Figura 4.4.

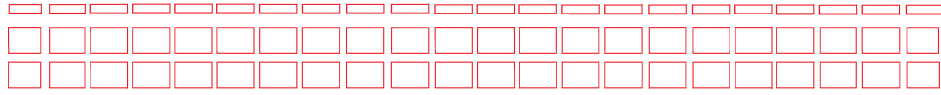


Figura 4.4: Máscara estimada desde los promedios calculados

La máscara es proyectada sobre la imagen segmentada, donde la búsqueda de los pixeles es realizada en los intervalos $[I(i, j) + K(i, j), I(x, y) + K(x, y)]$ correspondientes a las bandas de las muestras, donde I corresponde a la imagen segmentada y K la máscara estimada. Si al menos un píxel se encuentra marcado en aquellos intervalos, la banda es detectada en su respectiva posición.

5- Detección de bandas: La máscara “Figura 4.4”, es proyectada sobre la Figura 4.5.(f), la detección de bandas se realiza recorriendo cada píxel en la imagen con la máscara proyectada, cuando un píxel es detectado en uno de los pocillos que corresponde a una banda, el algoritmo toma como una banda activa.

6- Diagnóstico de casos: Para el diagnóstico de casos positivos, el algoritmo debe comprobar si tres carriles horizontales, se encuentran en un promedio de los intervalos calculados en alguna columna específica, lo cual la detección de una o más muestras, son identificadas en la posición específica.

Los casos positivos, que el algoritmo detecta, son validados por un biólogo profesional volviendo a procesar la muestra mediante la técnica de PCR-RFLP, por lo cual, el resultado de la segunda adquisición de la imagen es sometida de vuelta al algoritmo para su detección. El algoritmo si vuelve a detectar como un caso positivo, el profesional biólogo es el encargado de confirmar dicho diagnóstico de la mutación.

La Figura 4.5 muestra los resultados de cada paso que fueron detallados de los ítems principales del proceso.

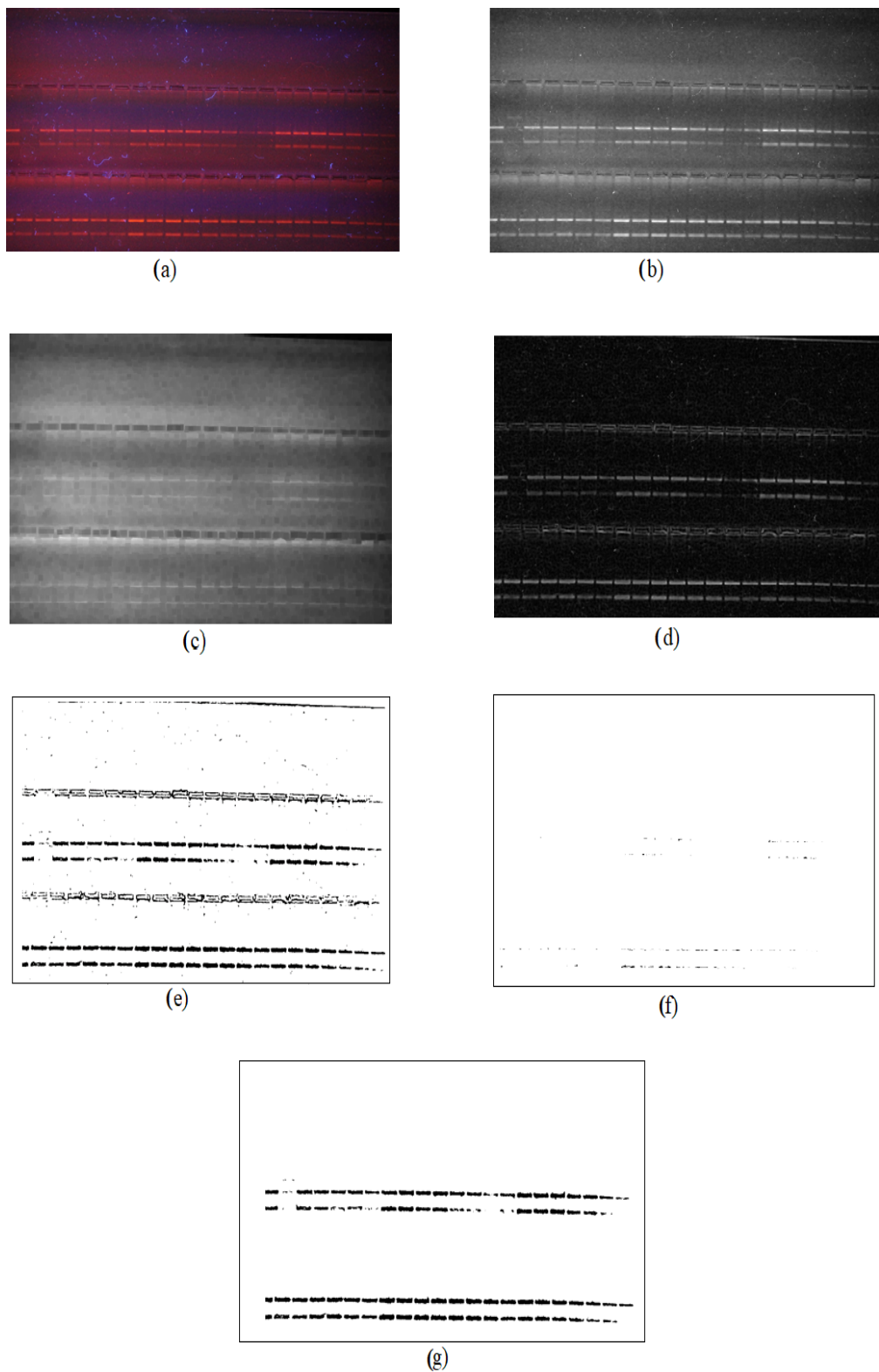


Figura 4.5: a) imagen original, (b) canal Rojo, (c) canal rojo erosionado, (d) imagen substraída, (e) imagen binarizada, f) imagen binaria erosionada, imagen sementada y filtrada

La Figura 4.6 muestra el diagrama de los pasos que fueron realizados para diagnosticar casos de mutación tp53 R337H.

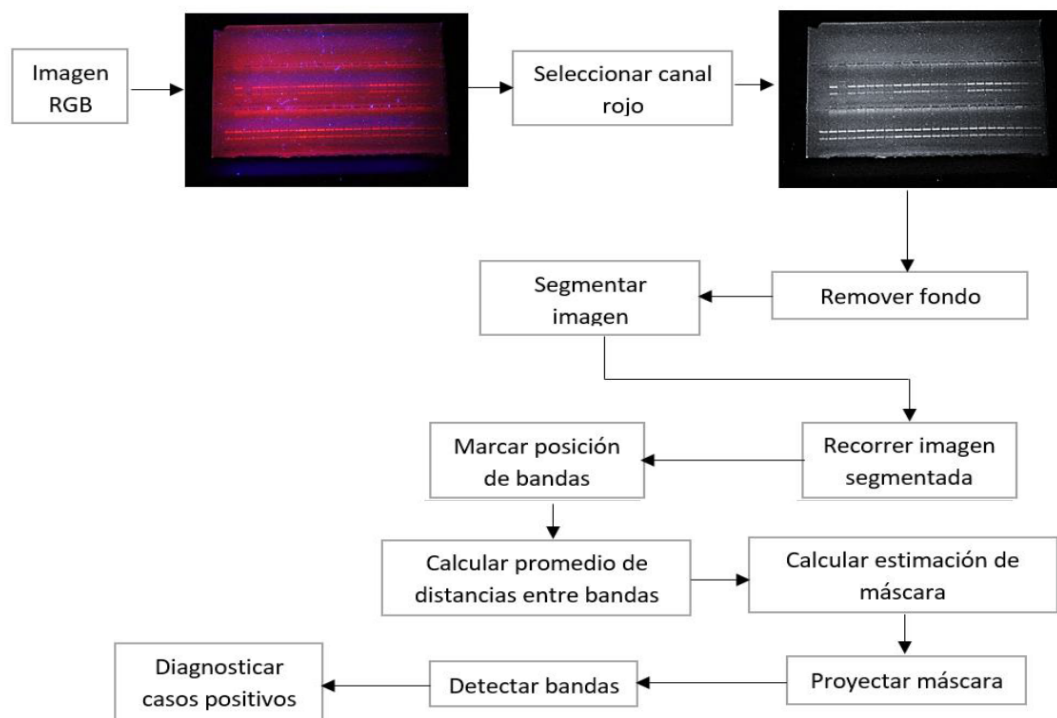


Figura 4.6: Descripción de pasos principales de la propuesta abordada

Resumen

En este capítulo se detallaron los procesos que fueron llevados a cabo para ésta propuesta, en la misma se describieron los pasos que lo conforman como el pre-procesamiento, la segmentación, filtro morfológico, estimación de una máscara, detección de bandas y diagnósticos de casos.

Capítulo 5

RESULTADOS

En este capítulo son detallados las pruebas realizadas a la propuesta, así también los resultados obtenidos de las mismas.

5.1 Prueba experimental 1: Métodos de umbralización para detección de bandas

En la umbralización de las imágenes se ejecutaron 20 algoritmos, de las cuales se tomaron en cuenta las cantidades de bandas detectadas por cada algoritmo.

El algoritmo de umbralización Fuzzy C-Means presentó mejores resultados en la binarización de las imágenes de electroforesis en gel.

5.1.1 Resultados

La Tabla 5.1 muestra los resultados obtenidos de los algoritmos de umbralización que fueron utilizados para el experimento. Las muestras correctamente identificadas fueron validadas por un profesional experto.

El motivo de los errores de los algoritmos “con menores cantidades de muestras identificadas”, fue a causa de que las imágenes contenían variaciones en las intensidades luminosas, contaminaciones por objetos indeseados o en otros casos las bandas contenían valores de intensidades cercanos al fondo de la imagen. Por tal motivo, al ser procesados en los filtros morfológicos en niveles binarios, producían pérdidas en las bandas.

El algoritmo utilizado para la umbralización de imágenes de electroforesis en gel está basado en Fuzzy C-Means, con el cual se obtuvo mejores resultados en la detección de las bandas.

Tabla 5.1: Tabla de métricas de los algoritmos de umbralización

Algoritmo	Nro. De bandas identificadas
Algoritmo de Bernsen [Ber86]	965
Algoritmo de Niblack [Nib00]	1133
Algoritmo de Chow-Kaneko [yTK79]	1212
Algoritmo de Eikvil [LE88]	881
Algoritmo de Sauvola [Sau00]	1012
Algoritmo de entropía Sahoo[Sah88]	935
Algoritmo de Fuzzy de Yager [R.R79]	1093
Algoritmo de Anisotropía de Pun[Pun81]	415
Algoritmo de Otsu-Yan [Yan96]	1757
Algoritmo de Jawahar [Jaw96]	1243
Algoritmo de Li [Kit85]	1802
Algoritmo de Gamil Valverde [Val00]	1513
Algoritmo de Otsu [Ots79]	1659
Algoritmo de Brink [Bri92]	1736
Algoritmo de Chehikian [Che99]	567
Algoritmo de Yan [Yan96]	1055
Algoritmo de Thrussel [Thr79]	1211
Algoritmo de Huang [Hua95]	977
Algoritmo Fuzzy C-Means [Pei83]	1993
Algoritmo de Shanoon [MS04]	1835

5.2 Prueba experimental 2: Diagnóstico de la mutación germinal TP53 R337H.

Las pruebas realizadas para la metodología abordada, fueron cuantificadas con respecto a la fiabilidad diagnóstica de estas pruebas en función de los valores de sensibilidad, especificidad, exactitud y precisión.

La sensibilidad o tasa de verdaderos positivos de una prueba puede verse como la probabilidad de que una muestra analizada, es detectada con la mutación en el gen tp53 R337H visualizándose con 3 bandas activas. Corresponde a los casos que en cierta edad puede poseer el cáncer en el córtex adrenal.

La especificidad o tasa de verdaderos negativos, puede verse como la probabilidad de que una muestra analizada, no posee la mutación R337H en el gen tp53, identificándose en la imagen solo con dos bandas activas. Corresponden a los casos realmente sanos.

- Los VP (Verdaderos Positivos) corresponden a aquellas muestras diagnosticadas con la mutación R337H en el gen TP53. Las mismas son identificadas con tres bandas tanto en el resultado del algoritmo como en el diagnóstico

del profesional experto.

- Los FP (Falsos Positivos) son aquellas muestras que son visualizadas con 3 bandas en el diagnóstico pero que no poseen la mutación. Corresponden a las muestras que son detectados con la mutación en el algoritmo, pero el diagnóstico final del profesional concluye con un resultado negativo.
- Los verdaderos negativos (VN) corresponden a aquellas muestras que no poseen la mutación en el gen TP53. Son aquellas muestras obtuvieron un resultado negativo tanto en el resultado del algoritmo como en el diagnóstico final del profesional experto.
- Los falsos negativos (FN) son aquellas muestras que no son diagnosticadas pero pueden contener la mutación. Corresponden a aquellas muestras que en el algoritmo no fueron detectados con la mutación, pero en el resultado final del diagnóstico del profesional experto concluye con un caso positivo de la mutación.

De tal manera que la sensibilidad o tasa de los verdaderos positivos está definida por:

$$sensibilidad = \frac{VP}{VP + FN} \quad (5.1)$$

Los VP son los verdaderos positivos y los FN son los falsos negativos

Mientras que el valor de especificidad o tasa de verdaderos negativos refiere a la probabilidad de que una muestra no tenga la mutación en el gen tp53 R337H.

$$Especificidad = \frac{VN}{VN + FP} \quad (5.2)$$

Donde los VN son los verdaderos negativos y los FP son los Falsos positivos.

La exactitud está dada por:

$$Exactitud = \frac{VP + VN}{VP + FP + FN + VN} \quad (5.3)$$

La precisión está dada por:

$$Precision = \frac{VP}{VP + FP} \quad (5.4)$$

Otra metodología de prueba realizada a la propuesta está basada en el conteo de bandas detectadas correctamente. Los resultados obtenidos del algoritmo fueron validados en comparación a los resultados del diagnóstico del experto profesional médico.

5.2.1 Resultados

Los resultados obtenidos de las pruebas realizadas al algoritmo propuesto, fueron evaluados mediante imágenes de gel en electroforesis para detectar mutaciones R337H del gen TP53. Estas imágenes fueron obtenidas de una base de datos del laboratorio del Centro de Biotecnología e Informática (CEBIOTEC), resultados de una investigación realizada en 5 departamentos del Paraguay zona fronteriza con Brasil [EFdL15]. Fueron analizadas 2134 muestras, de los cuales 6 fueron diagnosticados con la mutación R337H del gen TP53.

La Tabla 5.2 muestra algunos resultados de las imágenes analizadas, donde son visualizados los mejores resultados según la sensibilidad (Columna 5) y la especificidad (Columna 6), así también los resultados de exactitud (Columna 7) y precisión (Columna 8).

Tabla 5.2: Resultados de sensibilidad, especificidad, exactitud y precisión de muestras detectadas.

	VP	FP	VN	FN	Se.	Es.	Ex	Pr
imagen_24	3	0	73	1	0,7500	1,0000	0,9870	1,0000
Imagen_15	1	4	30	0	1,0000	0,8824	0,8857	0,2000
Imagen_30	2	2	40	0	1,0000	0,9523	0,9545	0,5000
Imagen_73	2	1	42	0	1,0000	0,9767	0,9777	0,6666
Imagen_39	7	7	16	0	1,0000	0,6956	0,7666	0,5000
Imagen_21	2	6	30	0	1,0000	0,8333	0,8421	0,2500

VP = Verdadero positivo. FP = Falso positivo. VN.= Verdadero negativo. FN = Falso negativo. Ex= exactitud. Se= sensibilidad. Es= especificidad. Pr= Precisión.

Debido a que las imágenes contienen varias muestras, las métricas fueron evaluadas en cada imagen. La Figura 5.1 muestra el histograma correspondiente a las métricas realizadas a todas las imágenes evaluadas.

Tabla 5.3: Matriz de confusión obtenidos del algoritmo propuesto.

		Salida del Algoritmo	
		Positivo	Negativo
Mutación tp53 R337H	Positivo	5	0
	Negativo	55	1.933

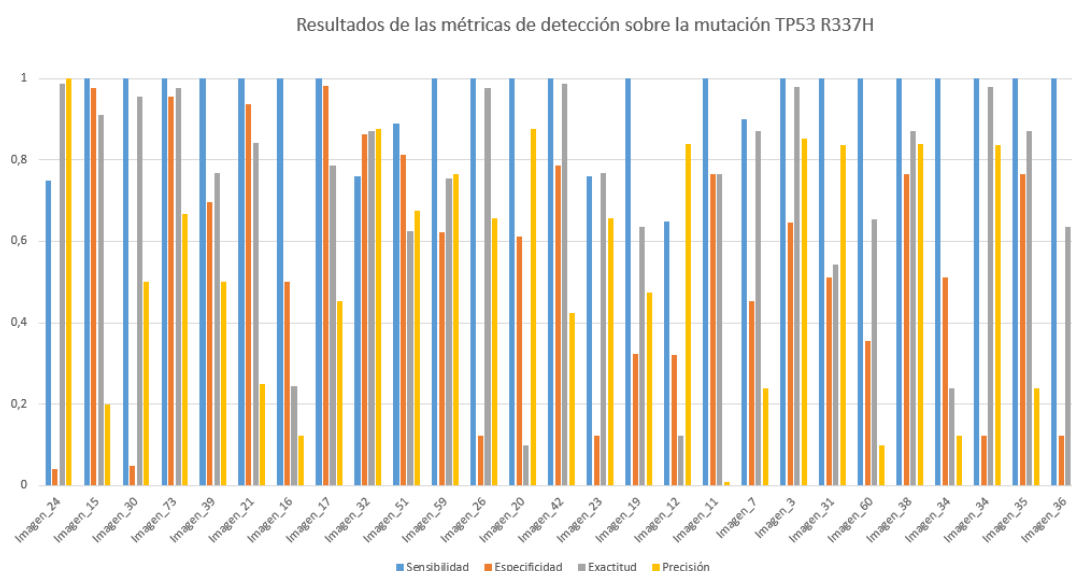


Figura 5.1: Histograma correspondiente a las métricas de Sensibilidad, Especificidad, exactitud y precisión de las imágenes

De 1993 muestras identificadas correctamente, 5 casos fueron detectados correctamente con la mutación R337H en el gen tp53. No fueron identificados muestras correspondientes a los falsos negativos. Mientras que 55 casos fueron detectados con la mutación en el algoritmo, sin embargo en el resultado del diagnóstico del experto profesional resultaron negativos. Los resultados correspondientes a los verdaderos negativos, 1993 muestras fueron identificados correctamente como se muestra en la Tabla 5.3.

La Tabla 5.4 muestra los resultados de la eficiencia del algoritmo con respecto a la tasa de aciertos y tasa de error evaluados desde la identificación y conteo de muestras analizadas por el algoritmo y los resultados del conteo de muestras observadas.

Entre las muestras analizadas, fueron un total de 1993 las muestras correctamente identificadas y un total de 141 muestras fueron erróneamente identificadas. En comparación a los resultados del experto profesional, esto lleva a una tasa de aciertos del 93,4% y una tasa de error del 6,6%.

Tabla 5.4: Tabla de métricas de la detección de bandas

Muestra Analizadas	Muestras Identificadas Correctamente	Muestras no Identificadas	Tasa de Error	Tasa de Aciertos
2.134	1.993	141	6,6%	93,4%

5.3 Ejemplos de detección.

En esta sección se detallan algunos casos de identificación y detección de bandas, donde se muestran algunos ejemplos de una correcta identificación de las bandas, así también casos en la cual resultan erróneas en la identificación.

5.3.1 Identificación y detección de la mutación R337H del gen TP53.

El algoritmo logra detectar la mutación R337H del gen TP53 en la muestra visualizado con 3 bandas y donde la primera banda posee un promedio mayor de anchura que las 2 bandas por debajo de la misma. En el siguiente ejemplo se puede observar una imagen con una muestra positiva (Mutación R337H del gen TP53 detectada) ubicada en la segunda fila de las muestras analizadas, columna 8.

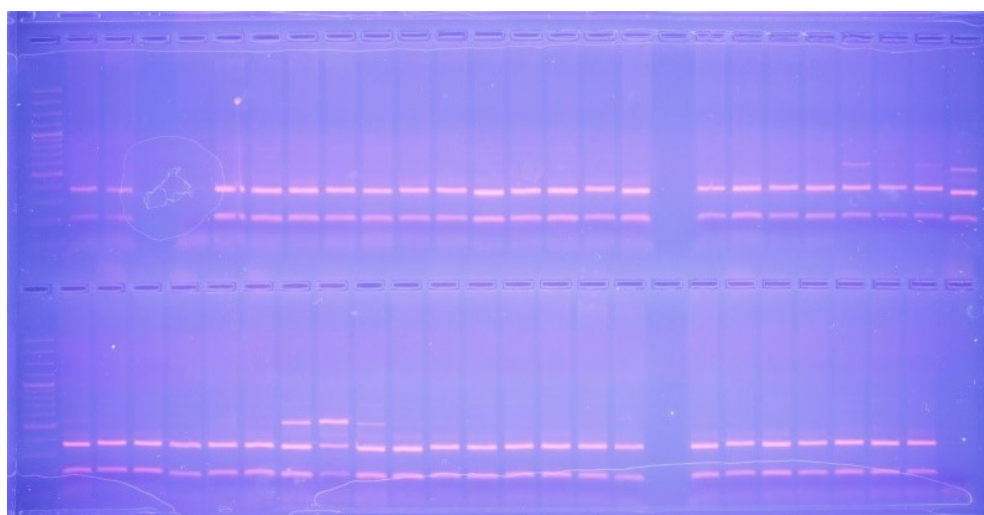


Figura 5.2: Imagen original con un caso positivo de la mutación.

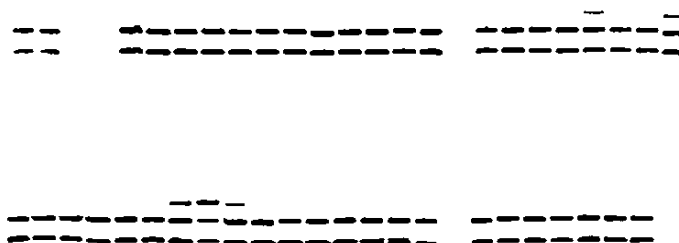


Figura 5.3: Imagen_15 con un caso positivo de la mutación.

5.3.2 Casos erróneos de identificación de muestras

Los casos que generan error en la identificación de las imágenes de electroforesis en gel, son resultados de imágenes obtenidos con poca iluminación o contraste, lo que dificultan en la visualización e identificación de las muestras.

En la Figura 5.4, muestra como una imagen con bajo contraste en las muestras, puede no identificarse a la hora de segmentar la imagen, lo cual genera una mala lectura en el algoritmo con la identificación de las muestras.

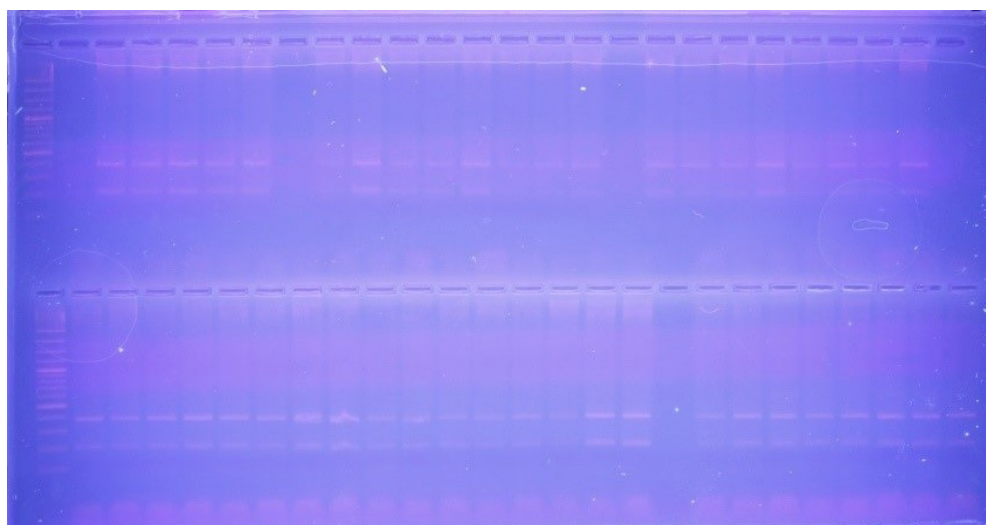


Figura 5.4: Imagen Original Imagen_90 con bajo contraste en las muestras.

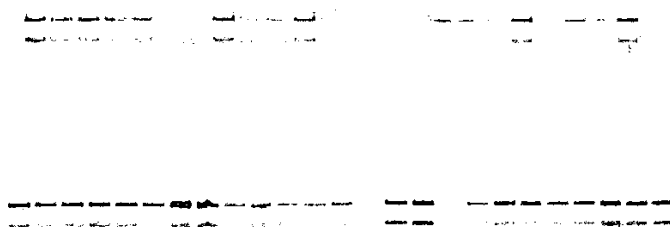


Figura 5.5: Imagen Segmentada Imagen_90 con errores de identificación en las bandas de algunas muestras.

Capítulo 6

CONCLUSIONES Y TRABAJOS FUTUROS

En este capítulo son descritas las conclusiones finales en base a los objetivos del trabajo, así también los trabajos a futuro.

Se presentó una propuesta para la identificación automática de la mutación R337H del gen TP53 mediante técnicas de visión por computadora.

La propuesta presenta ventajas en la correcta identificación de las bandas, agilizando el proceso manual con respecto a la detección de la mutación, que tiene dificultades de ser analizados a simple vista, ayudando de tal manera a la factibilidad y fiabilidad de los resultados y diagnósticos de las muestras detectadas con la mutación.

Una base de datos compuesta por 150 imágenes de electroforesis en gel específicamente Fueron testadas, se analizaron en busca de la detección de la mutación R337H en el gen TP53, los cuales fueron proveídas por investigadores y profesionales del CEBIOTEC.

6.1 Evaluación de los objetivos

Objetivo 1: Comparar técnicas de umbralización automática para la segmentación de las imágenes de electroforesis en gel con el fin de obtener la mayor cantidad de muestras. Se compararon 20 técnicas de umbralización (ver tabla 5.1), Fuzzy C-Means fue el algoritmo con mayor cantidad de muestras detectadas dando una tasa de aciertos 93.4% y una tasa de error del 6,6%.

Objetivo 2: Implementar técnicas de visión por computador para

la segmentación de las imágenes de electroforesis en geles que permita agilizar el proceso del experto profesional en la toma de decisiones. Se lograron implementar técnicas de visión por computador que permite agilizar el proceso en la toma de decision del profesional médico.

Objetivo 3: Comparar el proceso automático y manual con respecto al tiempo sobre detección e identificación de las muestras para obtener diagnósticos más precisos de los casos de mutación R337H en el gen TP53. Se compararon los procesos automático y manual para el diagnóstico de las muestras, que en comparación al proceso de análisis visual del profesional biólogo, el algoritmo resulta preciso en el diagnóstico de los casos.

Objetivo 4: Evaluar el proceso automático con respecto a la precisión del diagnóstico de la mutación R337H utilizando técnicas de visión por computador Según Especificidad, Sensibilidad, Precisión y Exactitud. Se lograron detectar 5 muestras con casos positivos, ningún caso positivo no detectado por el algoritmo, 55 casos sospechosos y 1933 casos negativos. Por cada imagen fueron utilizados las métricas con alto valores en porcentaje de sensibilidad, especificidad, exactitud y precisión.

Los resultados obtenidos para medir la eficacia algoritmo fueron validados mediante los métodos de sensibilidad, especificidad, exactitud y precisión. También se obtuvieron los resultados de medición según las tasas de error (6,6 %) y de aciertos (93,4 %) mostrando una alta eficiencia con respecto a la identificación de las muestras evaluadas.

6.2 Trabajos a futuro

Basados en el estudio de las técnicas utilizadas en la propuesta, algunas consideraciones son expuestas que pueden ser investigados:

- Otras técnicas de segmentación a ser utilizadas, tomando como base otros modelos de color y mejoras del contraste.
- Nuevas técnicas para la identificación y diagnóstico de las muestras mediante algoritmos de aprendizaje automático para imágenes de electroforesis en gel con la mutación R337H.
- También se sugiere realizar modificaciones de la propuesta para la detección y diagnóstico de otras mutaciones que generan cáncer en humanos.

REFERENCIAS BIBLIOGRÁFICAS

- [Ber86] J. Bernsen. Dynamic thresholding of gray level images. *Pattern recognition*, 28(1):1251–1255, 1986.
- [bio] Gel imaging and analysis software and systems. <http://www.bioimage.net/AQ.htm>. [Web; accedido el 17/03/2017].
- [Bri92] A. D. Brink. Thresholding of digital images using two-dimensional entropies. *Pattern Recognition*, 25(8):803–808, 1992.
- [CA00] Chompret A, Brugieres L, Ronsin M, Gardes M, Dessarps-Freichay F, Abel A, et al.. P53 germline mutations in childhood cancers and cancer risk for carrier individuals. *Br J. Cancer*, 69(1932):82, 2000.
- [Che99] A. Chehikian. Image segmentation by contours and regions cooperation. *Signal processing*, 78(3):329–347, 1999.
- [DP.92] Lane DP. Cancer. p53, guardian of the genome. *Nature*. 1992, 358(15-6):6381, 1992.
- [EFdL15] Edith Falcon-de Legal, Marta Ascurra, Gislaine Custódio Horacio Legal Ayala et al. Prevalence of an inherited cancer predisposition syndrome associated with the germ line tp53 r337h mutation in paraguay. *Cancer epidemiology*, 39:166–169, April 2015.
- [FJ14] Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. Cancer incidence and mortality worldwide: sources, methods and major patterns in globocan 2012. *International Journal of Cancer*, page 7, 2014.
- [Gel] Analysis and comparison of electrophoresis patterns. <http://www.applied-maths.com/gelcompar-ii>. [Web; accedido el 17/03/2017].
- [Guh17] T. Guha. Inherited tp53 mutations and the li fraumeni syndrome. cold spring harbor perspectives in medicine. *Cold Spring Harb Perspect Med*, 4(a026187):7, 2017.

- [GW06] Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing (3rd Edition)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2006.
- [Her12] N. Hahner Hermann, L. J., Heinze B. Fassnacht M. Willenberg H. S. Quinkler M. Reisch Tp53 germline mutations in adult patients with adrenocortical carcinoma. the journal of clinical endocrinology and metabolism. *Journal of Clinical Endocrinology*, 97(3):E476–E485, 2012.
- [HMC91] Vogelstein Bert Hollstein-M., Sidransky D and Harris Curtis C. *p53 mutations in human cancers*. 1991.
- [Hua95] Wang M. J. J. Huang, L. K. Image thresholding by minimizing the measures of fuzziness. *Pattern recognition*, 28(1):41–51., 1995.
- [Jaw96] C. V. Jawahar. Investigations on fuzzy thresholding based on fuzzy clustering. *Image Processing*, 29(12):2025 2032, 1996.
- [JM.03] Varley. JM. Germline tp53 mutations and lifraumeni syndrome. *Human.Mutat*, 20(21):313, 2003.
- [Kit85] J. I. Kittler. Threshold selection based on a simple image statistic. *Image Processing*, 30(8):771 776, 1985.
- [LE88] T. T. (s.f.). Line Eikvil. une méthode adaptive pour segmentation d’images de documents. *Pattenr Recogn.*, 1988.
- [Mat67] Georges Matheron. *Éléments pour une théorie des milieux poreux*. Masson et Cie, 1967, 1967.
- [MC10] Delopoulos A. Maramis C., Minga E. *An Application for Semi-automatic HPV Typing of PCR-RFLP Images*. In: Campilho A., Kamel M. (eds) Image Analysis and Recognition. ICIAR 2010., 2010. Lecture Notes in Computer Science, vol 6112. Springer, Berlin, Heidelberg.
- [MS04] B. Sankur, M. Sezgin. Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging*, 13(1):146-165, 2004.
- [Nib00] W. Niblack. An introduction to digital image processing. *Pattenr Recogn.*, 33(3):225 236, 2000.

- [Ots79] N. Otsu. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1):62–66, 1979.
- [Pei83] W. Peizhuang. Pattern recognition with fuzzy objective function algorithms (james c. bezdek). *SIAM Review*, 25(3):442–442, 1983.
- [Pun81] T. Pun. Entropic thresholding, a new approach. computer graphics and image processing. *Image Processing*, 16(3):210–239, 1981.
- [Rib01] Ribeiro-RC, Sandrini F, Figueiredo B, Zambetti GP, Michalkiewicz E, Lafferty AR, et al. An inherited p53 mutation that contributes in a tissue-specific manner topediatic adrenal cortical carcinoma. *Proc-NatlAcadSci U S A*, (5):98, 2001.
- [R.R79] Yager R.R. On the measures of fuzziness and negation.part.1: Membership in the unit interval. *Int. Journalof Gen*, 5:221–229, 1979.
- [Sah88] P. K. Sahoo. A survey of thresholding techniques. computer vision, graphics, and image processing. *Image Processing*, 41(2):233–260, 1988.
- [Sau00] M. Sauvola, J. & Pietaksinen. Adaptive document image binarization. *Pattern Recognition.*, 33:225–236, 2000.
- [Sed98] Goetz Sedlacek, Z., Kodet R. Kriz V. Seemanova E. Vodvarka P. Wilgenbus P. Two li-fraumeni syndrome families with novel germline p53 mutations: loss of the wild type p53 allele in only of tumours. *British j. of cancer*, 4(29):1034–1039, 1998.
- [Ser91] Jean Serra. Introduction to mathematical morphology. *Computer vision, graphics, and image processing*, 35(3):283–305, 1986/9/1.
- [Thr79] H. Thrussel. Picture threshold using an iterative selection method. *IEEE Transactions on Systems, Man and Cybernetics*, 9(5):311, 1979.
- [Val00] J. S. Valverde. Optimum binarization of technical document images. *International Conference*, 3:985–988, 2000.
- [Yan96] H. Yan. Unified formulation of a class of image thresholding techniques. pattern recognition. *Image Processing*, 29(12):2025–2032, 1996.
- [yTK79] Chow C.K. y T. Kaneko. Automatic detection of the left ventricle from cineangiograms. *Computers and biomedical research.*, 5:348–410, 1979.

ANEXO A

Algoritmo de umbralización Fuzzy C-Means

Los algoritmos de agrupamiento difuso permiten que un elemento esté incluido a más de un grupo, esto se da mediante un grado de pertenencia. La teoría que envuelve a los conjuntos difusos, es una extensión de la teoría de conjuntos clásicos, que a diferencia del conjunto clásico, un conjunto difuso relaciona la pertenencia de la clase de un objeto con un cierto valor de grado, donde un elemento puede pertenecer a más de un conjunto, sin embargo, en un conjunto clásico un elemento puede pertenecer o no a un cierto conjunto.

El método de Fuzzy C-Means es un algoritmo de agrupamiento difuso presentado en 1973 [Pei83], está basado en funciones objetivo con los de la lógica difusa, en la que un patrón podría contener grados de pertenencia a los diferentes subgrupos resultantes. Consideremos un conjunto conformado por n patrones $X = X_1, X_2, X_3, \dots, X_n \subset \mathbb{R}^2$, con p atributos cada elemento, El objetivo es agrupar estos n elementos en c clusters.

La función de pertenencia de un elemento del grupo, está dado en el rango $[0, 1]$, un elemento puede pertenecer a todas las clases, a uno o a ninguno.

En el caso de la Umbralización de imágenes digitales puede darse mediante la visualización de objetos O y fondo F como conjuntos difusos, en que cada píxel muestra un valor de pertenencia parcial para cada región, dependiendo de sus niveles de gris, siendo μ la función de pertenencia, el valor para cada píxel $p = I(x, y)$ puede pertenecer a un grupo ya sea para $\mu_{O(p)} \in [0, 1]$ o para $\mu_{f(p)} \in [0, 1]$ [Pei83].

Cuando $\mu = (\mu_1, \mu_2, \dots, \mu_k)$, el algoritmo de Fuzzy C-Means minimiza la siguiente función objetivo, que es la suma de los errores cuadráticos:

$$S_f = \sum_{i,j \in n} \sum_{c=1}^z \mu_c^k \|x(i, j) - v_c\|^2 \quad (1)$$

Donde la matriz de pertenencia μ representa la probabilidad de que el píxel p pertenezca al clúster c , k representa la cantidad de cluster establecido, la expresión de actualización para esta matriz es denotada por:

$$\mu_{c(x,y)} = \left[\sum_{c=1}^Z \left(\frac{\|p - v_c\|}{\|p - v_n\|} \right)^{\frac{1}{(q-1)}} \right]^{-1} \quad (2)$$

Y para el centroide en el c -ésimo agrupamiento:

$$v_c = \frac{\sum_{x,y \in n} \mu_{c(x,y)}^q p}{\sum_{i,j \in n} \mu_{c(x,y)}^q} \quad (3)$$

Donde $\mu_c(x, y)$ corresponde a la función de pertenencia, q controla la cantidad de difusión, $x(i, j) \in 1, 2, \dots, L - 1$ pertenece al valor del píxel.

ANEXO B

Resultados de segmentación con imágenes de electroforesis en gel

Resultado para la Imagen_20

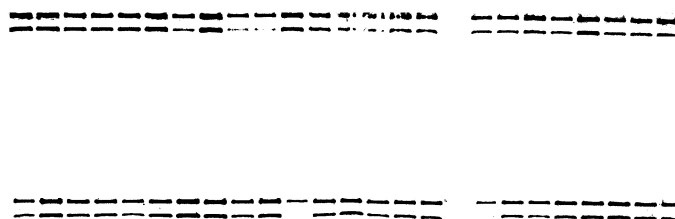
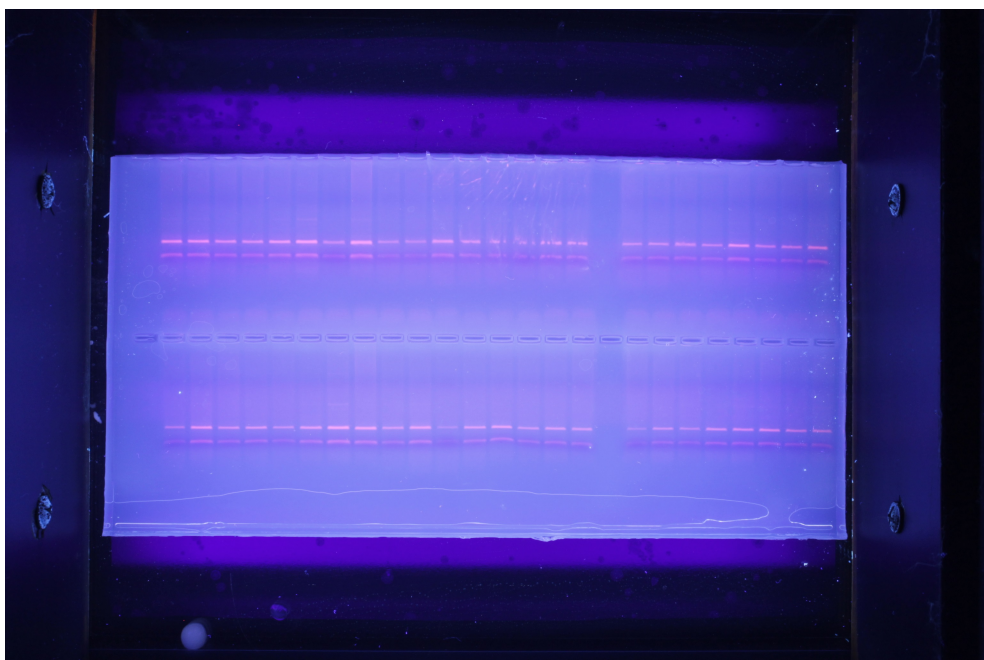


Figura 1: Segmentación de la Imagen_20

Resultado para la Imagen_22

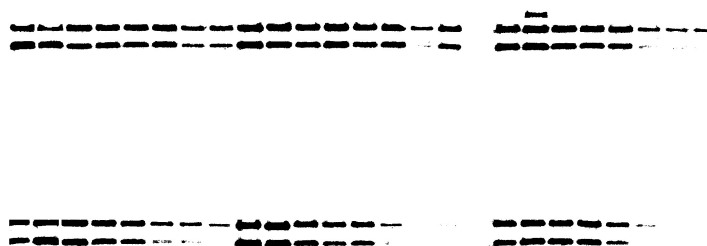
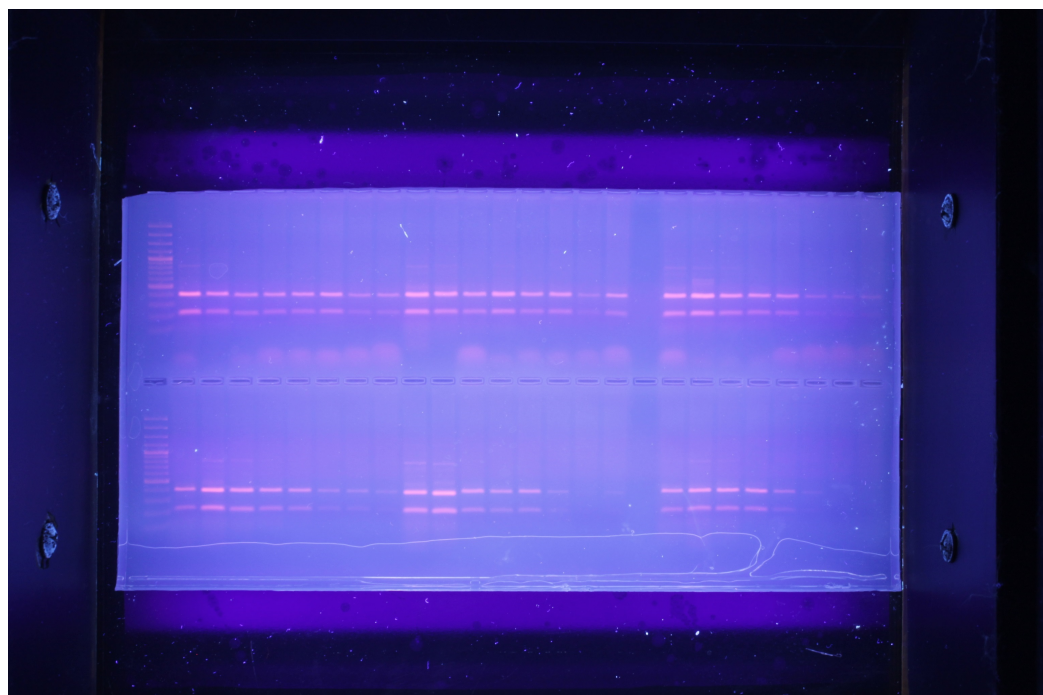


Figura 2: Segmentación de la Imagen_22

Resultado para la Imagen_24

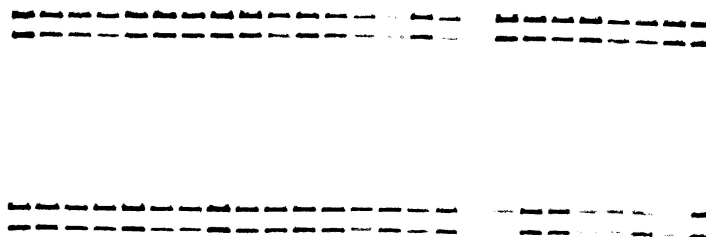
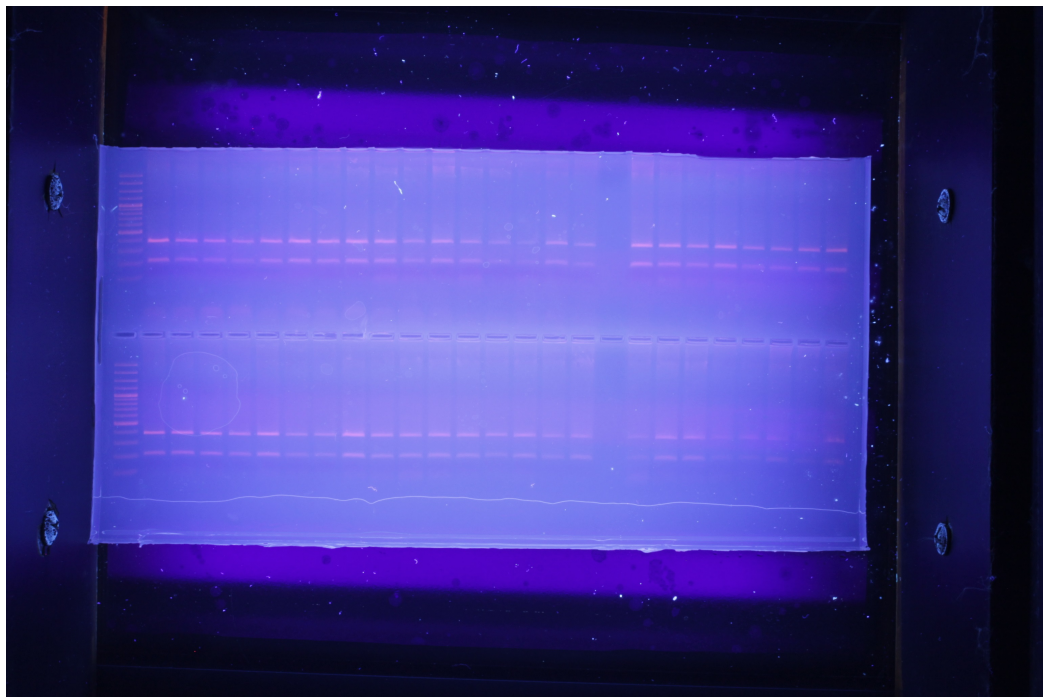


Figura 3: Segmentación de la Imagen_24

Resultado para la Imagen_49

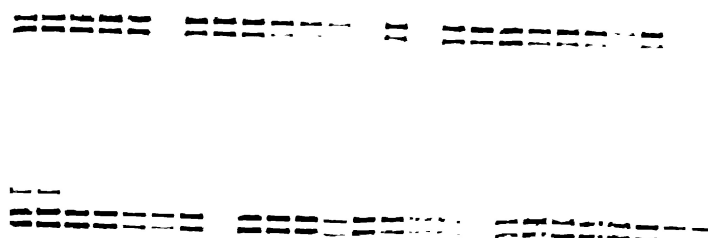
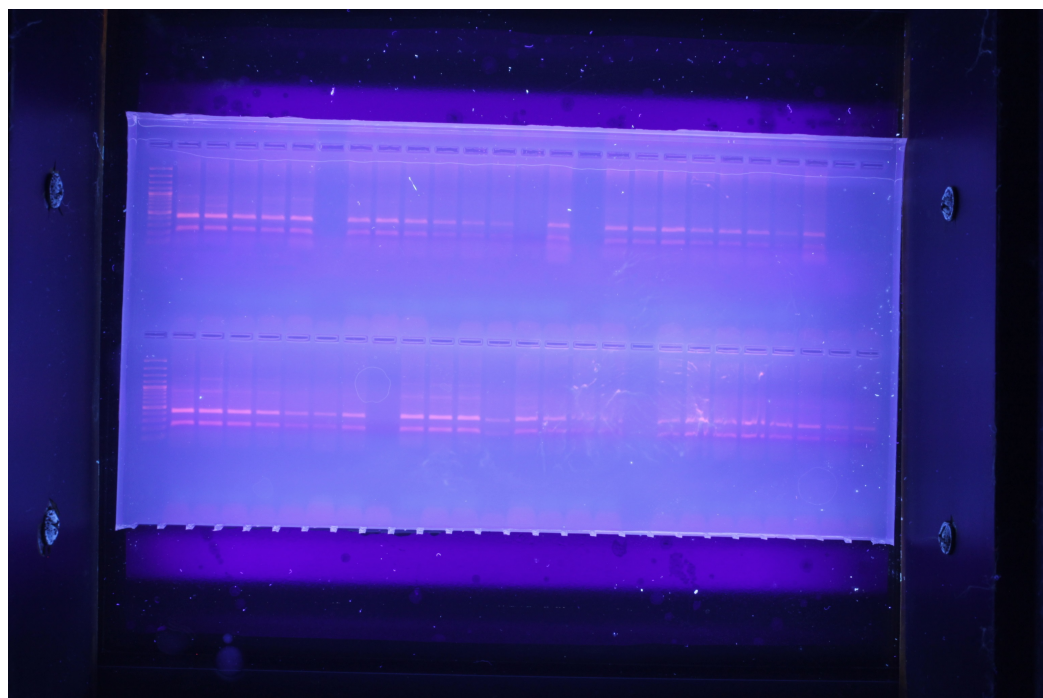


Figura 4: Segmentación de la Imagen_49