



Universidad Autónoma del Estado de Hidalgo  
Instituto de Ciencias Básicas e Ingeniería  
Centro de Investigación en Matemáticas

---

# Software para la Clasificación de Fibras Musculares en Imágenes Histológicas

Tesis que para obtener el título de

Licenciado en Matemáticas Aplicadas

presenta

Gonzalo Adán Chávez Fragoso

bajo la dirección de

Erika Elizabeth Rodríguez Torres

Omar López Ortega

PACHUCA, HIDALGO. Enero de 2017



**UNIVERSIDAD AUTÓNOMA DEL ESTADO DE HIDALGO**

**Instituto de Ciencias Básicas e Ingeniería**  
*Institute of Basic Sciences and Engineering*

**Área Académica de Matemáticas y Física**  
*Mathematics and Physics Department*

Gonzalo Adán Chávez Fragoso  
**PRESENTE**

Por este conducto le comunico que el Jurado que le fue asignado a su trabajo de tesis titulado *Software para la clasificación de fibras musculares en imágenes histológicas*, después de revisarlo en reunión han decidido autorizar la impresión del mismo, hechas las correcciones que fueron acordadas.

A continuación se anotan las firmas de conformidad de los integrantes del Jurado:

Presidente: Mtra. Margarita Tetlalmatzi Montiel

Secretario: Dr. Roberto Avila Pozos

Primer Vocal: Dra. Liliaha Peralta Hernández

Segundo Vocal: Dra. Erika Elizabeth Rodríguez Torres

Tercer Vocal: Dr. Omar López Ortega

Atentamente  
"Amor, Orden y Progreso"  
Mineral de la Reforma, Hidalgo, a 20 de enero de 2017.

Dr. Rafael Villarroel Flores  
Secretario del Comité de Titulación  
de la Licenciatura en Matemáticas Aplicadas



Ciudad del Conocimiento  
Carretera Pachuca – Tulancingo Km. 4.5  
Colonia Carboneras  
Mineral de la Reforma, Hidalgo, México, C.P. 42184  
Tel. +52 771 7172000 exts. 6164, Fax 2109  
aamyf\_icbi@uaeh.edu.mx



[www.uaeh.edu.mx](http://www.uaeh.edu.mx)

## Resumen

Para la caracterización de los tipos de fibras presentes en músculos esqueléticos de los organismos se emplean tinciones histoquímicas. La identificación y clasificación de esas fibras se lleva a cabo visualmente por un experto investigador, lo que conlleva una considerable inversión de recursos y tiempo. Por lo anterior, se desarrolló un sistema computacional que permite la automatización de la tarea de clasificación de las fibras musculares en un menor tiempo y alto grado de eficiencia. Para tal propósito, se emplean herramientas computacionales tales como algoritmos de minería de datos, inteligencia artificial y de reconocimiento de patrones, implementados en el lenguaje de programación Java, lo que garantiza un soporte multiplataforma. Los algoritmos usados son *K-means*, *Fuzzy c-means*, Mapas auto organizados de Kohonen y un algoritmo de aprendizaje supervisado por el experto. Los resultados obtenidos de dicho sistema pueden ser almacenados en hojas de cálculo pues dada su alta compatibilidad permiten alimentar otros sistemas que realicen otro tipo de análisis, por ejemplo determinar la dimensión fractal de un grupo de fibras del mismo tipo para distinguir si se encuentran distribuidas aleatoriamente o no. Este sistema se utiliza para posteriormente hacer posible la determinación de patrones de distribución y organización de las fibras de los músculos, tanto en condiciones normales como patológicas. Los resultados obtenidos a partir de los distintos análisis realizados, permiten establecer que el sistema computacional desarrollado realiza un análisis de los tipos de fibras con un buen grado de eficiencia, reduciendo hasta un 90 % el tiempo de procesamiento, pero que al ser asistido por un experto mejora notoriamente su eficacia.

## Abstract

Histochemical stains are used for the characterization of the fibers types present in skeletal muscles of organisms. Usually, the identification and classification of these fibers is carried out visually by an expert researcher, which entails a considerable investment of resources and time. Because of that, a computational system was developed which allows the automation of the classification of the muscle fibers in a shorter time and with high efficiency. For that purpose we used computational tools such as data mining algorithms, artificial intelligence and pattern recognition, implemented in the programming language Java, which guarantees multiplatform support. The algorithms used are K-means, Fuzzy c-means, Kohonen self-organized maps and an algorithm supervised by an expert. Because of its high compatibility with other platforms, the results obtained by this computational system can be stored in spreadsheets to perform other types of analysis, for example to compute the fractal dimension of a group of fibers of the same type and thus determine whether they are distributed randomly or not. The spreadsheet data is used later to determine distribution patterns and organization of muscle fibers, both under normal and pathological conditions. The results obtained from the different analyzes of the developed computer system indicated a good degree of efficiency, reducing processing time by up to 90 %, although expert assistance improves its effectiveness.

A mis padres.



# Agradecimientos

Quiero agradecer a mis revisores de tesis, a la Maestra Margarita Tetlalmatzi, al Doctor Roberto Ávila, a la Doctora Liliana Peralta por su tiempo y disposición para ayudarme a concluir este trabajo. A mis asesores, la Doctora Erika Elizabeth Rodríguez Torres y al Doctor Omar López Ortega por guiarme en el desarrollo de este trabajo.

Agradezco a todos mis profesores del área académica de Matemáticas y Física por su contribución en mi desarrollo tanto profesional como personal. Agradezco al Doctor Jorge Viveros por la sugerencia del método de Funciones de Distribución que fue de importancia para la validación del presente trabajo.

Agradezco al Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional por su colaboración, observaciones y sugerencias para el mejoramiento de este trabajo.

También quiero agradecer a mis compañeros y amigos que me acompañaron a lo largo de la carrera.

Quiero agradecer a mi familia por apoyarme en la realización de mis estudios en la Licenciatura en Matemáticas aplicadas, así como en la realización de este trabajo.

En especial agradezco a Chely por todo su apoyo y motivación para mejorar mi persona, por todas las alegrías compartidas.

Estudio apoyado parcialmente por: SNI a ERT, donativo PROMEP-1006 a ERT y Premio Sofia Kovalévskaya a ERT.



# Índice general

<b>1. Introducción</b>	<b>1</b>
<b>2. Antecedentes</b>	<b>5</b>
2.1. Músculo . . . . .	5
2.1.1. Clasificación de fibras del músculo esquelético . . . . .	6
2.1.2. Plasticidad muscular . . . . .	6
2.2. Procesamiento de imágenes . . . . .	8
2.2.1. Pasos fundamentales en el procesamiento de imágenes . . . . .	10
2.2.2. Componentes de un sistema de procesamiento de imágenes . . . . .	12
2.3. Descubrimiento de conocimiento y Minería de datos . . . . .	13
2.3.1. Análisis de conglomerados (Clustering) . . . . .	16
2.3.2. Clasificación . . . . .	17
2.4. Estado del Arte . . . . .	19
<b>3. Planteamiento del problema</b>	<b>23</b>
3.1. Justificación . . . . .	24
3.2. Hipótesis . . . . .	24
3.3. Objetivos . . . . .	25

3.3.1. Objetivo general . . . . .	25
3.3.2. Objetivos específicos . . . . .	25
<b>4. Metodología</b>	<b>27</b>
4.1. Analisis histoenzimático . . . . .	27
4.1.1. Obtención de imágenes . . . . .	28
4.2. Requerimientos del usuario . . . . .	29
4.3. Diseño . . . . .	30
4.3.1. Aprendizaje . . . . .	35
4.3.2. Segmentación . . . . .	42
4.3.3. Filtrado . . . . .	46
4.3.4. Conteo . . . . .	50
4.3.5. Edición de resultados . . . . .	54
4.4. Funciones de distribución . . . . .	57
4.4.1. Función de distribución para una sola imagen . . . . .	57
4.4.2. Funciones de distribución para un lote de imágenes . . . . .	62
<b>5. Resultados</b>	<b>65</b>
5.1. Preliminares . . . . .	65
5.1.1. Comparación de resolución y filtros . . . . .	65
5.1.2. Funciones de distribución . . . . .	71
5.2. Comparación con un experto . . . . .	74
5.2.1. Resultados con la tinción COX . . . . .	74
5.2.2. Resultados con la tinción NADH . . . . .	79
5.2.3. Resultados con la tinción ATPasa pH 9.4 . . . . .	83

*ÍNDICE GENERAL* XI

**6. Conclusiones** **93**

**7. Apéndice** **95**

    7.1. Manual del programa . . . . . 96



# Capítulo 1

## Introducción

La interacción entre diferentes disciplinas, como la Biología y las Matemáticas, ha permitido la realización de análisis más profundos y detallados de los fenómenos biológicos desde la perspectiva analítica de las matemáticas, lo que ha facilitado en gran medida la adecuada interpretación de los mismos. En la presente tesis se plantea la interacción de disciplinas afines a la computación y la histología, para analizar los tipos de fibras presentes en músculos esqueléticos, que han sufrido alteraciones metabólicas o lesiones derivadas de procesos biológicos, tales como la desnutrición o la multiparidad (múltiples partos), respectivamente.

De acuerdo a la función contráctil que desempeñan, los músculos contienen distintas proporciones de fibras musculares que se clasifican de acuerdo a sus características contráctiles como: Tipo I (lentas), las cuales desarrollan una contracción de poca fuerza y larga duración, son resistentes a la fatiga y presentan un metabolismo exclusivamente oxidativo; Tipo IIb (rápidas), desarrollan una contracción breve, de gran fuerza y se fatigan rápidamente y contienen un metabolismo predominantemente glucolítico; y finalmente, las del tipo IIa/IIID (intermedias) cuyas características contráctiles y metabólicas las sitúan entre las fibras lentas y las rápidas, esto es, desarrollan una fuerza contráctil media, son resistentes a la fatiga y poseen un metabolismo dual, tanto oxidativo como glucolítico.

Dependiendo de las condiciones fisiológicas de los organismos (tanto intrínsecas como extrínsecas), el fenotipo de las fibras puede cambiar de un tipo a otro. Lo anterior confiere a las fibras musculares la capacidad “plástica” de modificar sus propiedades contráctiles y/o metabólicas, dependiendo de la influencia de factores intrínsecos (es decir enfermedades) o extrínsecos (como lesiones o por una inadecuada

alimentación).

Una forma de caracterizar las influencias que los factores intrínsecos y/o extrínsecos ejercen sobre los músculos esqueléticos es establecer los cambios en la composición y distribución en los fenotipos de las fibras que los componen con respecto a músculos sanos. Las técnicas que se utilizan con mayor frecuencia para analizar la composición de los tipos de fibras presentes en los músculos son de índole histoquímico, las cuales facilitan su clasificación debido a que la actividad generada en cada reacción forma un precipitado que les confiere una coloración particular de acuerdo al nivel de actividad de enzimas metabólicas o de proteínas con capacidad enzimática presentes en los músculos.

Para el análisis de las fibras en un corte histológico teñido, se toman microfotografías por secciones del corte completo, las cuales se unen en una sola imagen para posteriormente proceder a la clasificación y conteo manual. El tiempo invertido en este análisis es muy significativo dado que dependiendo de las condiciones que se manejen en el diseño experimental, el número de individuos de cada grupo y la cantidad de fibras que componga el músculo de interés (que va de 2000 a 5000 fibras) se toman hasta 3 horas por imagen, lo que representa un importante esfuerzo visual, lo que podría resultar en errores en la identificación de fibras, el retardo en el avance de otros experimentos y análisis, o inclusive daños a la capacidad visual.

Debido a lo anterior, el objetivo de este trabajo es apoyar en dicha tarea a los investigadores y estudiantes proponiendo un sistema computacional basado en algoritmos de minería de datos, lógica difusa y procesamiento de imágenes. Implementando los paradigmas de programación orientada a objetos además de orientación a eventos por parte de un usuario, desarrollado en el lenguaje Java por su portabilidad a distintos sistemas operativos. El desarrollo y las pruebas del programa se hicieron utilizando datos de músculos bajo diferentes condiciones experimentales como la multiparidad (de la línea de Investigación de la Dra. Margarita Martínez-Gómez, Investigadora del Instituto de Investigaciones Biomédicas de la UNAM, sede Tlaxcala) en colaboración con la Dra. Kenia López-García; y la desnutrición crónica en colaboración con el Dr. Ismael Jiménez Estrada del Departamento de Fisiología, Biofísica y Neurociencias del CINVESTAV-IPN.

En el Capítulo 2 se presentan los antecedentes de índole biológica, se presentan conceptos clave del músculo esquelético y de su clasificación; también se presenta la teoría involucrada en la solución al problema planteado, es decir los conocimientos de computación en cuanto análisis de imágenes y minería de datos; además se ilustra el estado del arte en cuanto a software que están encaminados a la misma proble-

mática. En el Capítulo 3 se describe el problema planteado en el presente trabajo, tanto general como específico, se da una justificación y se presentan objetivos. En Capítulo 4 se ilustra la metodología que se siguió para el diseño y desarrollo de este trabajo. En el Capítulo 5 se presentan resultados de la utilización del software y se presentan comparaciones de su eficacia. Finalmente en el Capítulo 6 se presentan algunas conclusiones de este trabajo.



# Capítulo 2

## Antecedentes

### 2.1. Músculo

El tejido muscular es de gran importancia para la vida animal, pues es responsable del movimiento en su entorno en busca de alimento para su supervivencia y además para el transporte interno de nutrientes y procesos orgánicos. Debido a esto el músculo es el tejido más abundante en la mayoría de los animales y representa la mayor parte del consumo de energía en un animal activo.

En general, los diferentes tipos de músculos son categorizados como estriados y lisos. El músculo liso, también llamado involuntario, es controlado por el sistema nervioso autónomo y está presente en el aparato reproductor, excretor, en los vasos sanguíneos y órganos internos. Dentro de los tipos que entran en la clasificación de músculo estriado está el esquelético y el cardíaco. El músculo esquelético es caracterizado por dos fenotipos generales, de contracción rápida y contracción lenta, mientras que el músculo cardíaco se puede diferenciar entre el músculo del ventrículo y la aurícula.

El músculo esquelético representa cerca del 45 % del peso total del cuerpo, es responsable del mantenimiento de la posición del cuerpo y del movimiento del esqueleto y está constituido por células llamadas fibras musculares cuyo calibre varía de 10 a 100  $\mu\text{m}$  en un mismo músculo con una longitud de hasta 15 cm, las cuales están unidas unas a otras mediante tejido conjuntivo (para más información consulte [27]).

### 2.1.1. Clasificación de fibras del músculo esquelético

Inicialmente los músculos fueron clasificados como rápidos o lentos basados en la velocidad de acortamiento, esta división también correspondía a una diferencia morfológica. Los músculos rápidos presentan una coloración blanca en algunas especies particularmente en las aves y los músculos lentos una coloración roja debido al resultado de altas cantidades de mioglobina.

La mayor cantidad de mioglobina contenida en los músculos con fibras lentas contribuye a la mayor capacidad oxidativa en comparación con los músculos blancos. El análisis histológico muestra que existe una relación entre la actividad de ATPasa de la miosina y la velocidad de acortamiento del músculo, este análisis histoquímico llevó a la división original de las fibras musculares en tipo I (lento), con baja hidrólisis de ATP, y tipo II (rápido), que presentan una mayor actividad respecto a las fibras de tipo I.

El desarrollo de los procedimientos para la ATPasa de la miosina en histoquímica enzimática permitió la identificación de dos poblaciones de fibras, llamadas tipo IIa y IIb, que son abundantes en músculos de contracción rápida, y distintas de las fibras tipo I, predominantes en músculos de contracción lenta.

Así en la actualidad la clasificación de fibras musculares más ampliamente aceptada es: fibras tipo I (lentas), las cuales se contraen con poca fuerza y larga duración, son resistentes a la fatiga y presentan metabolismo oxidativo; fibras del tipo IIb (rápidas), desarrollan una contracción breve, de gran fuerza, se fatigan rápidamente y, por contraparte, tienen metabolismo glucolítico; y fibras de los tipos IIa/IIc (intermedias) con características contráctiles entre las lentas y las rápidas, resistentes a la fatiga y con un metabolismo dual, oxidativo-glucolítico (véase [27]). Sin embargo estos fenotipos pueden modificarse por diversos factores, tanto internos como externos, fenómeno que es llamado plasticidad muscular.

### 2.1.2. Plasticidad muscular

El músculo esquelético de mamíferos es un tejido extremadamente heterogéneo, compuesto por una gran variedad de tipos de fibras. Estas fibras, sin embargo, no son fijas, sino que representan unidades altamente versátiles, capaces de responder a las demandas funcionales y a una variedad de señales cambiando su fenotipo en términos de tamaño, composición y propiedades metabólicas, lo que se conoce como

plasticidad. La gama completa de la capacidad de adaptación se extiende de rápido a lento. Sin embargo, ahora está claro que las transiciones de tipo de fibra no proceden en saltos inmediatos de un extremo a otro, si no que se producen de una manera secuencial, gradual y ordenada.

El ejercicio estimula profundamente los mecanismos celulares. Por ejemplo, realizar repetidas sesiones de larga duración y baja intensidad contráctil (ejercicio aeróbico) resulta en una mejora de la fatiga y resistencia, en contraste, los episodios repetidos de exposición a fatiga intensa y alta intensidad contráctil (ejercicio de resistencia), induce a incrementar el área de la sección transversal de la fibra así como la fuerza de contracción.

La plasticidad muscular puede ser diferenciada en dos: 1) cambios en la distribución de las fibras musculares (rápidas y lentas) y 2) cambio en el tamaño de las fibras (hipertrofia y atrofia). Los principales factores para inducir la respuesta adaptativa son la actividad contráctil, las condiciones metabólicas y la influencia hormonal (véase [34]). Prácticamente toda la hipertrofia muscular se debe a un aumento del número de filamentos de actina y miosina en cada fibra muscular, dando lugar a un aumento en el tamaño de las fibras musculares individuales, a esto se le denomina hipertrofia de las fibras. La hipertrofia aparece en un grado mucho mayor cuando el músculo está sometido a carga durante el proceso contráctil. La velocidad de síntesis de las proteínas contráctiles del músculo es mucho mayor cuando se está produciendo la hipertrofia lo que da lugar también a cantidades cada vez mayores de filamentos tanto de actina como de miosina en las miofibrillas aumentando con frecuencia hasta un 50 %.

Durante la desnutrición crónica se produce atrofia de las fibras musculares, las cuales disminuyen de tamaño y de diámetro debido a que la tasa catabólica es mucho mayor que la tasa anabólica. Lo anterior resulta de una baja producción de hormonas tiroideas así como del aporte reducido de los elementos nutricionales necesarios para la formación de las proteínas. En la fase final de la atrofia por denervación la mayor parte de las fibras musculares son destruidas y sustituidas por tejido fibroso y adiposo (el lector interesado puede consultar [31]).

## 2.2. Procesamiento de imágenes

Con la llegada de las cámaras fotográficas en el siglo *XIX* fue posible la captura y almacenamiento de imágenes para distintos propósitos, comúnmente la captura de sucesos importantes y en algunas ocasiones para usos industriales. Sin embargo su uso, en el último de los casos, estaba limitado por la calidad final según las condiciones en el momento de captura o el proceso de revelado y poco se podía hacer con una fotografía en su estado final, es decir no era posible modificaciones posteriores de una imagen.

El procesamiento digital de imágenes está íntimamente relacionado con la utilización de computadoras, y su desarrollo ha dependido de la potencia de procesamiento disponible (para más información sobre este tema consultar [10]). Las primeras computadoras lo suficientemente potentes para llevar a cabo procesamiento digital de imágenes aparecieron a principios de los años 1960's, utilizadas en el programa espacial, lo que se puede considerar como el nacimiento de la disciplina. Para esto fue necesario el desarrollo de métodos de digitalización de imágenes: se utilizó la primera matriz de elementos sensores discretos, diseñada por Eugene F. Lally del *Jet Propulsion Laboratory* (Pasadena, California), en el *Ranger 7* para transmitir imágenes de la luna en 1964. Fue necesaria la creación de algoritmos corregir distorsiones inherentes del método adquisición recién creado, estos algoritmos sirvieron como base para métodos mejorados de restauración y mejoramiento de imágenes para misiones espaciales posteriores y que son aun usados hoy en día en muchos otros campos.

Una imagen puede ser definida como una función  $f(x, y)$ , donde  $x$  y  $y$  son coordenadas espaciales, y el valor de  $f$  en cualquier par de coordenadas  $(x, y)$  puede ser la intensidad de nivel de gris o una representación de color. Cuando  $x, y$  y  $f$  tienen valores finitos y discretos, se dice que es una imagen digital. El campo de procesamiento digital de imágenes se refiere a la utilización de computadoras digitales para procesar imágenes digitales. Notese que una imagen digital está compuesta de una cantidad finita de elementos, cada uno con localización y valor particular. Estos elementos son llamados pixels. El estándar actual en cuanto a color es el modelo RGB de 24 bits, el cual está compuesto de 3 componentes de color: rojo, verde y azul, con intensidades representadas por un número entero de 8 bits, representando un rango de 0 a 255, representando 0 el color negro y 255 la máxima intensidad del color respectivo.

Los límites entre el procesamiento digital de imágenes y otras disciplinas afines como la Visión por Computadora o el Análisis de Imágenes no están bien definidos, sin embargo un paradigma útil es considerar tres tipos de procesos incluidos en sus

límites:

1. *Procesos de bajo nivel*: involucra operaciones primitivas como preprocesamiento de imágenes para la reducción de ruido, mejora de contraste y mejora de enfoque. Un proceso de bajo nivel es caracterizado por el hecho de que tanto la entrada como la salida son imágenes.
2. *Proceso de medio nivel*: incluye tareas como la segmentación (partición de una imagen en regiones u objetos), descripción de esos objetos para reducirlos a una forma adecuada para su procesamiento por computadora, y clasificación (reconocimiento) de objetos individuales. Este tipo de procesos es caracterizado por que las entradas son generalmente imágenes, pero la salida son atributos extraídos de éstas (como bordes, contornos, y la identidad de objetos individuales).
3. *Procesos de alto nivel*: involucra "hacer sentido" de un conjunto de objetos reconocidos, como en análisis de imágenes, y en los límites de la disciplina, llevar a cabo funciones cognitivas normalmente asociadas con visión.

### 2.2.1. Pasos fundamentales en el procesamiento de imágenes

- *Adquisición*: es el almacenamiento en un archivo digital de la representación de la energía recibida por un sensor, se transforma la información perceptible por el ojo humano o por cualquiera de las fuentes energéticas mencionadas previamente. Además también se considera adquisición la generación de estos archivos en un dispositivo electrónico mediante software de asistencia en el diseño.
- *Mejora*: es el proceso de manipular una imagen de manera que el resultado es más adecuado que la imagen original para una aplicación específica. No hay teoría general del mejoramiento de imágenes. Cuando una imagen es procesada para interpretación visual, el observador es el juez definitivo de que tan bien funciona un método.
- *Restauración*: es un área que también se ocupa del mejoramiento de la apariencia de una imagen. Sin embargo, a diferencia del mejoramiento, el cual es subjetivo, la restauración de imágenes es objetiva, en el sentido de que las técnicas de restauración tienden a basarse en modelos matemáticos o probabilísticos de degradación de imágenes.
- *Procesamiento del color*: es un área que ha estado ganando importancia por el incremento significativo en el uso de imágenes digitales en Internet. El color es muy usado como descriptor que frecuentemente simplifica la identificación de objetos y extracción de una escena.
- *Transformación Wavelet*: es el fundamento para representar imágenes en varios grados de resolución. Es usado para la compresión y para la representación piramidal, en la cual las imágenes son subdivididas sucesivamente en regiones más pequeñas.
- *Compresión*: se ocupa de técnicas para reducir el almacenamiento requerido para guardar una imagen, o el ancho de banda para transmitirla. La compresión de imágenes es familiar (quizás inadvertidamente) para la mayoría de usuarios de computadoras en forma de extensiones de archivos de imagen, como la extensión jpg usada en el estándar de compresión de imágenes JPEG (Joint Photographics Experts Group).
- *Procesamiento morfológico*: trata las herramientas para la extracción de una imagen que son útiles en la representación y descripción de la forma. Es una

transición entre procesos donde la salida es una imagen y procesos en los cuales la salida son atributos de la imagen.

- *Segmentación*: procesos que llevan a cabo la partición de una imagen en las partes u objetos que la constituyen. En general la segmentación autónoma es una de las tareas más difíciles en procesamiento de imágenes digitales.
- *Representación y descripción*: casi siempre se realiza después de la etapa de segmentación, donde los datos son píxeles crudos que constituyen ya sea la frontera de una región o todos los puntos de la región en sí. En cualquier caso es necesario convertir los datos a una forma conveniente para el procesamiento en la computadora. También se debe especificar un método para describir los datos tal que las características de interés sean resaltadas. La descripción, también llamada selección de características, trata de la extracción de atributos que resulta en información cuantitativa de interés o es básica para la diferenciación de una clase de objetos de otra.
- *Reconocimiento*: es el proceso que asigna una etiqueta a un objeto basado en sus descriptores. Desarrolla métodos para el reconocimiento de objetos individuales.

Una parte importante y de gran utilidad a lo largo de estas etapas es una base de conocimiento. Puede contener información simple como el detalle de las regiones de una imagen donde se sabe que se encuentra la información de interés, y así limitar la búsqueda de la información. La base de conocimiento puede ser bastante compleja, como una lista interrelacionada de todos los mayores posibles defectos un problema de inspección de materiales o una base de imágenes que contiene imágenes satelitales en alta resolución de una región en conexión con aplicaciones de detección de cambios. Además de guiar la operación de cada modulo de proceso, la base de datos también controla la interacción entre estos.

Aunque no se menciona explícitamente la visualización de las imágenes, es importante mantener en mente que se pueden observar los resultados del procesamiento de una imagen al final de cada etapa. También se debe señalar que no todas las aplicaciones de procesamiento de imágenes requieren seguir la complejidad de las etapas mostradas. De hecho en muchos casos no se necesitan todos las etapas.

### 2.2.2. Componentes de un sistema de procesamiento de imágenes

Los componentes de un sistema de procesamiento de imágenes de propósito general son:

- *Dispositivo de adquisición*: el sensor es un dispositivo sensible a la energía radiada por el objeto del que se desea obtener una imagen, la cual es proporcional a la intensidad de la luz. Además es compuesto por un digitalizador, que es un dispositivo que convierte la energía obtenida por el sensor a una forma digital.
- *Hardware especializado en procesamiento de imágenes*: usualmente consiste del digitalizador ya mencionado además de hardware que ejecuta operaciones primitivas, como una unidad lógica (ALU por sus siglas en inglés), que realiza operaciones aritméticas y lógicas en paralelo en imágenes enteras. Por ejemplo se puede usar para promediar imágenes tan rápido como son digitalizadas, con el propósito de eliminar ruido.
- *Computadora*: es una computadora de propósito general que puede ir desde una computadora personal hasta una supercomputadora. Casi cualquier computadora personal bien equipada es adecuada para tareas de procesamiento de imágenes fuera de línea.
- *Software*: consiste en módulos especializados que ejecutan tareas específicas. Un paquete bien diseñado también incluye la capacidad para el usuario de escribir código que como mínimo utiliza los módulos especializados. Los paquetes de software más sofisticados permiten la integración de esos módulos y comandos de software de propósito general para, por lo menos, un lenguaje de computadora.
- *Almacenamiento masivo*: una imagen con una intensidad de 8 bits por pixel (escala de grises) requiere un espacio de almacenamiento de un MegaByte por Megapixel, es decir una imagen de color real de 24 bits por pixel, una resolución de 12 Megapíxeles, que es muy popular actualmente, sin comprimir requeriría un almacenamiento de 36 MegaBytes. Al manejar miles o inclusive millones de imágenes, proveer almacenamiento adecuado puede ser un reto.
- *Visualizador de imágenes*: como monitores, que son alimentados por las salidas de imágenes y tarjetas gráficas, las cuales son una parte integral de un sistema

computacional. Raramente se encuentran requerimientos para aplicaciones de visualización de imágenes que no pueden ser cubiertas por las tarjetas gráficas disponibles en el mercado.

- *Dispositivos de copia fuente*: incluyen impresoras láser, cámaras de filmación, dispositivos sensibles al calor, unidades de inyección de tinta, y unidades digitales. Las últimas son ampliamente usadas en la actualidad.
- *Disponibilidad de redes*: dada la gran cantidad de datos inherente en aplicaciones de procesamiento de imágenes, la consideración clave en transmisión de imágenes es el ancho de banda. En redes dedicadas, esto no es típicamente un problema, pero la comunicación con sitios remotos vía Internet no es siempre eficiente.

## 2.3. Descubrimiento de conocimiento y Minería de datos

El descubrimiento de conocimiento en bases de datos (KDD por sus siglas en inglés) es definido como el proceso no trivial de identificar patrones válidos, nuevos, potencialmente útiles, y a la larga comprensibles en información. El proceso en general consiste en convertir información de bajo nivel en conocimiento de alto nivel (en [21] el lector interesado puede obtener más información sobre minería de datos). El proceso de KDD es interactivo e iterativo y envuelve, más o menos, los siguientes pasos:

1. *Comprensión del dominio de aplicación*: incluye conocimiento previo relevante y los objetivos de la aplicación.
2. *Extracción del conjunto de datos objetivo*: la selección de un conjunto de datos o el enfoque en un subconjunto de variables, usando jerarquización de características y técnicas de selección.
3. *Preprocesamiento de los datos*: es requerido para mejorar la calidad de los datos actuales para la minería. Este también incrementa la eficiencia de la minería reduciendo el tiempo requerido para minar los datos preprocesados. Preprocesamiento de datos involucra:

- Limpieza de datos.  
Consiste en algunas operaciones básicas como normalización, eliminación de ruido y manipulación de datos faltantes, reducción de redundancia, etc.
  - Integración de datos.  
Incluye integrar conjuntos de datos múltiples y heterogéneos, generados de diferentes fuentes.
  - Reducción y proyección de datos.  
Incluye el hallazgo de características útiles para representar los datos y usar métodos de reducción de dimensionalidad, discretización y extracción de características.
4. *Minería de datos*: constituye una o varias de las siguientes funciones: clasificación, regresión, agrupación (análisis de clústeres o conglomerado), resumido, recuperación de imágenes, descubrimiento de reglas de asociación y dependencias funcionales, extracción de reglas, etc.
  5. *Interpretación*: incluye la interpretación de los patrones descubiertos, visualización lo mejor posible (en pocas dimensiones) de los patrones extraídos.
  6. *Uso del conocimiento descubierto*: incluye incorporar este conocimiento en el desempeño del sistema y tomar acciones basados en el conocimiento.

En otras palabras, dados grandes volúmenes de datos heterogéneos, el objetivo es extraer eficientemente patrones significativos que pueden ser de interés y por lo tanto útiles para el usuario.

La minería de datos es un paso del proceso de KDD consistente en una enumeración particular de patrones sobre los datos, sujeto de varias limitaciones computacionales. El término patrón va más allá de su sentido tradicional para incluir modelos o estructuras en los datos. Datos históricos son usados para descubrir regularidades y mejorar decisiones futuras. Los datos puede consistir en una colección de descripciones de series de tiempo que pueden ser aprendidas para predecir eventos posteriores en las series.

La minería de datos involucra ajuste de modelos o determinación de patrones de los datos observados. Los modelos ajustados juegan el papel de conocimiento inferido. Decidir si el modelo refleja conocimiento útil o no es parte del proceso principal del KDD para lo cual se requiere usualmente un juicio subjetivo humano. Típicamente, un algoritmo de minería de datos constituye alguna combinación de los siguientes tres componentes:

1. *El modelo*: la función del modelo (por ejemplo análisis de conglomerados o clasificación) y su forma figurativa (por ejemplo discriminantes lineales, árboles de decisión). Un modelo contiene parámetros que se van a determinar de los datos por la función escogida usando la forma figurativa o herramienta particular.
2. *El criterio de preferencia*: una base para la preferencia de un modelo o conjunto de parámetros sobre otros, dependiendo en los datos dados. El criterio es usualmente alguna forma de mejor ajuste del modelo a los datos, quizás templado por un término de suavización para evitar el sobre-ajuste, o generar un modelo con muchos grados de libertad restringidos por los datos dados.
3. *El algoritmo de búsqueda*: las especificaciones de un algoritmo para encontrar modelos particulares o patrones y parámetros, dados los datos, modelo(s) y un criterio de preferencia.

Un algoritmo particular de minería de datos es usualmente una instancia de los componentes modelo-preferencia-búsqueda. Algunos de las funciones modelo comunes en la actual práctica de minería de datos incluyen:

1. *Clasificación*: esta función modelo clasifica un elemento de los datos en una de varias clases categóricas predefinidas.
2. *Regresión*: el propósito de esta función modelo es mapear un elemento de los datos una variable de predicción con valore reales.
3. *Análisis de conglomerados*: esta función mapea un elemento de los datos en uno de varios conglomerados, donde los conglomerados son agrupaciones naturales de elementos de los datos basadas en métricas de similaridad o modelos de probabilidad de densidad.
4. *Generación de reglas*: aquí se minan o generan reglas de los datos. Minería de reglas por asociación se refiere a descubrir relaciones de asociación entre diferentes atributos. El modelado de dependencias corresponde a extraer dependencias significativas entre variables.
5. *Resumido o condensación*: esta función provee una descripción compacta para un subconjunto de datos. La compresión de datos podría jugar un rol significativo aquí, particularmente para datos multimedia, por las ventajas que ofrece para representar compactamente los datos con un reducido número de bits, así incrementar el ancho de banda de la base de datos de almacenamiento.

6. Análisis secuencial: este modela patrones secuenciales, como análisis de series de tiempo, secuencias de genes, etc. El objetivo es modelar los estados del proceso que genera la secuencia, o extraer y reportar las desviaciones y tendencias sobre el tiempo.

Dos tipos de algoritmos usados en este trabajo y que se describirán más ampliamente más adelante son: análisis de conglomerados y clasificación.

### 2.3.1. Análisis de conglomerados (Clustering)

Un *cluster* (traducido del inglés como agrupación o conglomerado) es una colección de datos que son similares unos a otros en el mismo *cluster* pero diferentes a los datos de otros *clusters*. El análisis de conglomerados se refiere a la agrupación de un conjunto de datos en conglomerados. El análisis de conglomerados también es un tipo de clasificación no supervisada, donde no hay asignadas clases predefinidas.

Un buen método de análisis de conglomerados producirá conglomerados de alta calidad con alta similitud intraclase y baja similitud interclase. La calidad del resultado del análisis de conglomerados depende tanto de la medida de similitud usada por el método y de su implementación. Esta es medida por la habilidad del sistema para descubrir algunos de todos los patrones ocultos.

Los enfoques del análisis de conglomerados pueden ser ampliamente clasificados como:

- Particional: crea una partición inicial y después usa una estrategia de control iterativo para optimizar un objetivo.
- Jerárquico: crea una descomposición jerárquica (dendograma) del conjunto de datos (u objetos) usando algún criterio de terminación.
- Basado en densidad: usa conectividad y funciones de densidad.
- Basado en mallas: crea una estructura granular con múltiples niveles, por la cuantización del espacio de características en términos de celdas finitas.

### 2.3.2. Clasificación

La clasificación es también descrita como aprendizaje supervisado. Sea una base de datos de  $n$ -adas, cada una con una etiqueta de clase. El objetivo es desarrollar un modelo o perfil para cada clase.

El objetivo es predecir la clase  $C_i = f(x_1, \dots, x_n)$ , donde  $x_1, \dots, x_n$  son los atributos de entrada. La entrada del algoritmo de clasificación es, típicamente, un conjunto de datos de registros de entrenamiento con varios atributos. Hay un atributo distinguido llamado el atributo dependiente. Los atributos sobrantes de predicción pueden ser de naturaleza numérica o categórica. Si el atributo dependiente es categórico, el problema es llamado clasificación con este atributo siendo denominado la etiqueta de la clase. Sin embargo, si el atributo dependiente es numérico, el problema es determinado regresión. El objetivo de la clasificación y la regresión es construir un modelo conciso de la distribución del atributo dependiente en términos de los atributos de predicción. El modelo resultante es usado para asignar valores a una base de datos de registros de prueba, donde los valores de los atributos de predicción son conocidos pero el atributo independiente debe determinarse.

Los métodos de clasificación pueden clasificarse como:

- Árboles de decisión: dividen el espacio de decisión en regiones constantes a trozos. Típicamente, una medida de información teórica es usada para evaluar el poder discriminatorio de los atributos a cada nivel del árbol.
- Modelos probabilísticos o generativos: calculan las probabilidades para las hipótesis basados en el Teorema de Bayes.
- Clasificadores de vecino más cercano: calculan la mínima distancia a instancias o prototipos.
- Regresión: pueden ser lineales o no lineales, de la forma  $a + bx + cx^2 + \dots = C_i$ .
- Redes neuronales: son sistemas de procesamiento de señales que tratan de emular el comportamiento de sistemas nerviosos biológicos al proveer un modelo matemático para la combinación de numerosas neuronas conectadas en una red. Pueden ser formalmente definidos como “interconexiones masivamente paralelas de elementos de procesamiento simples que interaccionan con objetos del mundo real en una manera similar a los sistemas biológicos”. Particionan con fronteras no lineales. Estas incorporan aprendizaje en un ambiente “rico

en datos”, de tal manera que toda la información es codificada en una manera de distribución entre los pesos de conexión.

## 2.4. Estado del Arte

La clasificación de fibras musculares en imágenes histoquímicas ha sido, desde el siglo pasado, fuente para una gran cantidad de estudios (véase por ejemplo [8], [30], [3] y [5]). En la actualidad, diversos grupos de trabajo realizan estudios en los que se analiza la composición de distintos músculos esqueléticos de diferentes especies en cuanto a los tipos de fibras musculares que contienen, por lo que éste es un problema de actualidad.

Entre estos estudios es posible encontrar aquellos sobre la calidad de la carne (véase [2], [25] y [13]). Además existen otros de diversa índole como en los que se analiza la causa de la atrofia ligada al envejecimiento en humanos (véase [17]). En [34] y [31] se pueden encontrar aquellos relacionados a los efectos de la desnutrición en crías.

En esta tesis se utilizan los datos de [32] para analizar los efectos de la desnutrición en ratas; con los datos del trabajo [19], también en ratas, se revisa la desnutrición y primiparidad y finalmente de [1] se tomaron los datos que fueron obtenidos de conejas.

Dado lo laborioso que es realizar este tipo de análisis de manera manual por un experto, ha surgido la necesidad de apoyarse en la computación, muestra de esto es que el equipo de trabajo con el que se colaboró para la realización de este trabajo realizaba sus conteos manualmente utilizando el software *ImageJ* [24]. Además se pueden encontrar trabajos en los que se aborda este problema, usando diversos enfoques.

Por ejemplo en el trabajo de Lefaucheur de 1992 (ver [16]) se describe un software que se basa en la localización del tejido que separa a las fibras musculares, es decir el tejido conectivo o, como lo llaman, red, por la forma que toma. Seguido de un paso de localización de fibras delimitadas por esta red para posteriormente proceder a la clasificación de acuerdo a la coloración asignada por la tinción utilizada. Sin embargo está diseñado para imágenes con una resolución de  $512 \times 512$  píxeles (0.25 Mpx) en escala de grises, por lo que no podría ser utilizada en los protocolos para los que se plantea puesto que las imágenes pueden tener una resolución mayor a 16 Mpx y tal reducción de dimensionalidad conlleva a pérdida de información, además que este software está pensado para el análisis de pequeñas regiones de una imagen histoquímica, por lo que un análisis completo requeriría gran cantidad de análisis de secciones. Cabe mencionar que este software permite la clasificación de 200 a 300

fibras en un tiempo que va de 46 a 51 minutos según es reportado.

También se puede observar en lo expuesto por Miller en 1994 (véase [20]) la descripción de una metodología que se basa en el ajuste de la iluminación y la ganancia de brillo en el momento de la adquisición (guardando varias posibilidades), modificar los umbrales de saturación y negros, promediado de imágenes y detección de bordes, todo en imágenes también de resolución de  $512 \times 512$  píxeles y además en escala de grises.

En el trabajo de Buche (ver referencia [7]) publicado en el año de 1997 también se sigue la metodología de detección de bordes para la identificación de tejido conectivo en una imagen de referencia con una tinción elegida y además un algoritmo para parear dicho tejido conectivo con otras imágenes de cortes transversales contiguos con otras tinciones para un conteo múltiple. Sin embargo entre sus limitaciones se observa, como en los casos anteriores, una limitación en la resolución de las imágenes a analizar, siendo esta de  $512 \times 512$  píxeles, además que se reporta la clasificación de entre 150 y 300 fibras en un tiempo que va de 14 a 17 minutos. Se puede observar un ejemplo en 2.1.

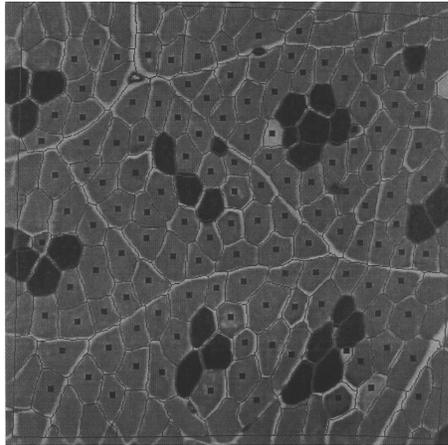


Figura 2.1: Imágen procesada con el método descrito en el trabajo de Buche.

Se presenta un trabajo que tiene como característica más resaltante de ser completamente automatizado en el trabajo de Brox publicado en 2006 (ver [6]), que sin embargo solo menciona tener como propósito imágenes con tinción H&E (Hematoxilina y Eosina) la cual se puede observar en la figura 2.2, usada para la realización de biopsias. Es usada para la identificación de las fibras musculares y además la determinación de su tamaño y forma, pero sin clasificarlas de acuerdo a sus propiedades.

Por otro lado, las imágenes que analiza tienen una resolución de  $640 \times 480$ , que son de zonas del músculo lo menos ruidosas posibles, elegidas visualmente, permitiendo contar un promedio de 19 fibras por imagen.

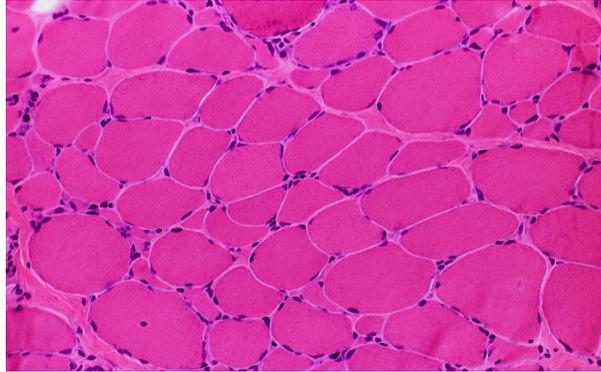


Figura 2.2: Fibra muscular teñida con H& E

Al igual que Buche (ver [7]), en el trabajo de Karen de 2009 (consultar [12]) se presenta una metodología basada en el conteo de múltiples imágenes de cortes contiguos utilizando una técnica de superposición, con lo que solo se indican las fibras presentes en una imagen guía y estas seorean con las demás imágenes, y que posteriormente clasifica las fibras de cada imagen según la tinción aplicada. Sin embargo se necesita delinear manualmente los bordes de las fibras en la imagen guía con un software externo, lo que significa un trabajo manual extenso, que es lo que se busca evitar, además que el propósito del presente trabajo no es la clasificación de cortes contiguos, si no de distintos especímenes de grupos experimentales específicos.

Un trabajo más reciente publicado por Sertel en 2011 (véase [26]) permite la clasificación de fibras musculares mediante el uso de la tinción histoquímica ATPasa con un pH de 4.5 y a su vez con un pH de 10.4. Menciona que permite procesar imágenes de hasta  $11000 \times 9000$  píxeles, dividiéndolas en varias regiones y procesandolas por separado. Se basa en la segmentación del tejido conectivo para identificar las fibras musculares sin importar su coloración, posteriormente, siguiendo una metodología similar a la usada por Buche y Karen, usa algoritmos para alinear imágenes con las dos tinciones mencionadas. Finalmente usa el algoritmo K-means para la clasificación de las fibras musculares según el promedio de colores (en escala de grises) que presentan, formando un vector donde cada entrada es el promedio de color en cada una de las imágenes histoquímicas. Es decir los datos para el algoritmo K-means son vectores de la forma  $(gray(i)_{pH4.5}, gray(i)_{pH10.4})$  para la muestra del individuo

*i.* Muestra tener alto grado de eficiencia, sin embargo no menciona el tiempo que toma el análisis de cada par de imágenes ni los requisitos computacionales para su ejecución. Además muestra tener un desempeño en relación directa con la calidad de las imágenes obtenidas.

## Capítulo 3

# Planteamiento del problema

El Departamento de Fisiología, Biofísica y Neurociencias del Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional en colaboración con investigadores del Centro de Investigación en Matemáticas y del Centro de Investigación en Tecnologías de la Información y Sistemas, ambos de la Universidad Autónoma del Estado de Hidalgo, realizan un estudio para determinar los cambios producidos en músculo en distintas condiciones experimentales tales como la multiparidad en los músculos pubococcígeo (mPc) y bulboesponjoso (mBe), uno pélvico y otro perineal en la coneja doméstica nulípara y múltipara (véase [1]), otro estudio sobre los efectos de la desnutrición y la primiparidad sobre el músculo pubococcígeo de la rata (en el trabajo [19]), además del estudio del músculo extensor largo del dedo de la pata de la rata con desnutrición posnatal (ver [32] y [33]).

Para cuantificar los cambios de la distribución de las fibras es necesario clasificarlas mediante la utilización de tinciones histoquímicas, las cuales facilitan su caracterización gracias a que los tejidos reaccionan a estas adoptando una coloración dependiendo de sus características. La tinción NADH-Tr caracteriza a las fibras de acuerdo a su metabolismo, las fibras oxidativas adquieren una coloración azul oscuro, mientras que las fibras glucolíticas se tornan de color azul claro. Por otro lado la tinción ATPasa facilita la distinción de las fibras por su capacidad contráctil en: Rápidas (Tono medio), intermedias (tono oscuro) y lentas (tono claro).

### 3.1. Justificación

La identificación de las fibras se hace tradicionalmente por un investigador de manera manual, lo que resulta en un alto consumo de tiempo dado que cada imagen histológica a analizar contiene entre 2000 y 5000 fibras musculares, además de representar un esfuerzo visual mayor al tener que estar frente a un monitor un prolongado periodo, lo que podría resultar en errores en la identificación o inclusive daños visuales.

Además se requiere determinar si la distribución espacial de las fibras musculares tiene una organización aleatoria o fractal, y a la vez describir y cuantificar los cambios sufridos entre grupos control y grupos experimentales.

### 3.2. Hipótesis

Es posible el diseño, implementación e integración de software que:

- Apoye la tarea de clasificación de fibras musculares en imágenes histoquímicas.
- Ofrezca una interfaz amigable.
- Compatibilidad con los sistemas operativos de mayor uso.
- Tenga requerimientos de recursos computacionales promedio, es decir pueda ejecutarse en computadoras portátiles de la actualidad.
- Permita el almacenamiento de los resultados en la forma más conveniente para sus análisis posteriores.

## 3.3. Objetivos

### 3.3.1. Objetivo general

Proveer de herramientas computacionales de clasificación, localización y conteo así como la descripción y cuantificación de cambios en la distribución de cada uno de los tipos de fibras musculares en imágenes histológicas, que apoyen a los expertos en el análisis de los cambios producidos por la multiparidad en los músculos pubococcígeo y bulboesponjoso de ratas y conejas además de los efectos de la desnutrición en el músculo extensor digitorum longus de la pata de la rata. Lo anterior ofreciendo un entorno de escritorio amigable e intuitivo, integrado en un sólo paquete, alto grado de confiabilidad, compatibilidad con formatos ampliamente utilizados y capacidad de ejecución en distintas plataformas.

### 3.3.2. Objetivos específicos

- Descomponer automáticamente y agrupar píxeles a partir de imágenes histoquímicas obtenidas en el laboratorio.
- Clasificar automáticamente en los grupos pertinentes de manera eficaz mediante técnicas de minería de datos.
- Almacenar resultados en formatos compatibles con otros sistemas para análisis posteriores.
- Proveer a expertos herramientas de edición y manipulación de imágenes separadas.
- Validar resultados de acuerdo con la opinión de los expertos.
- Proveer datos que sirvan para calcular la dimensión fractal.
- Establecer una función de distribución de densidad de los distintos tipos de fibras musculares para determinar los posibles patrones de cada tipo de fibra.



# Capítulo 4

## Metodología

### 4.1. Análisis histoenzimático

Los cortes del músculo fueron removidos rápidamente, medidos, pesados y sumergidos en 2-metilbutano, para ser congelados con nitrógeno líquido. Se obtuvieron series de cortes transversales de un grosor de 8 a 10  $\mu\text{m}$  en un criostato a  $-20\text{ }^{\circ}\text{C}$ . Estas secciones fueron montadas en un portaobjetos para su tinción.

#### Histoquímica para NADH-TR

Para determinar el tipo de metabolismo se siguió el método modificado de Nachlas (ver [22]). Las secciones de músculo se incubaron 1 hora a  $37\text{ }^{\circ}\text{C}$  en una solución 1:1 (v/v) de NBT-NADH (1.2 mM nitroazil de tetrazolio) diluido en 50 mM de buffer Tris, pH 7.6; 2.25 mM de nicotinamida adenina dinucleótido (NADH) diluido en 50 mM de buffer Tris. Posteriormente, las secciones de tejido se lavaron tres veces con agua desionizada. El exceso de la solución NBT-NADH se removió lavando las secciones con concentraciones crecientes y decrecientes (30, 60 y 90 %) de acetona. Finalmente las preparaciones fueron montadas con un cubreobjetos para su posterior análisis.

### Histoquímica para ATPasa pH 9.4

En secciones seriales subsecuentes de tejido, la actividad miofibrilar de la ATPasa a pH 9.4 fue visualizada usando la técnica de la ATPasa alcalina (pH 9.4; Modificado de Guth y Samaha [11]). Las secciones de músculo se sumergieron en una solución de preincubación de 0.01 M de Tris base y 0.018 M de  $CaCl_2$  (pH 9.4) por 20 minutos. Las muestras se lavaron tres veces por 5 minutos con agua desionizada y posteriormente se incubaron a 37 °C por 1 hora en 1.5 % p/v de adenosin-5'-trifosfato en un solución de preincubación a pH 9.4. Después la muestras fueron incubadas en 2% de  $CaCl_2$  por 5 minutos, se transfirieron a una solución 2% p/v de CoCl por otros 5 minutos y se transfirieron a una solución 10% v/v de  $(NH_4)_2HSO_4$  por 3 minutos. Posteriormente las secciones se lavaron, deshidrataron y se montaron con glicerol sobre un cubreobjetos.

#### 4.1.1. Obtención de imágenes

A cada sección de fascículo se les tomó microfotografías mediante cámara fotográfica digital Nikon DS-R1U3 de 12.7 Megapixeles (4,076 x 3,116) montada en un microscopio de luz con un objetivo de 10X y se unieron con el uso del programa Adobe Photoshop; se almacenan en formato JPEG para su la clasificación de fibras. Utilizando la técnica de la ATPasa alcalina (pH 9.4), las fibras claras, grises y oscuras se identifican como lentas, rápidas e intermedias, respectivamente. Las fibras en las secciones de fascículos teñidas por la técnica de NADH-TR se clasifican de acuerdo al siguiente criterio: las fibras oscuras se identificaron como oxidativas; mientras que las claras se consideraron como glucolíticas.

En la figura 4.1 se muestran ejemplos de secciones de músculo teñidas con la tinción NADH y ATPasa.

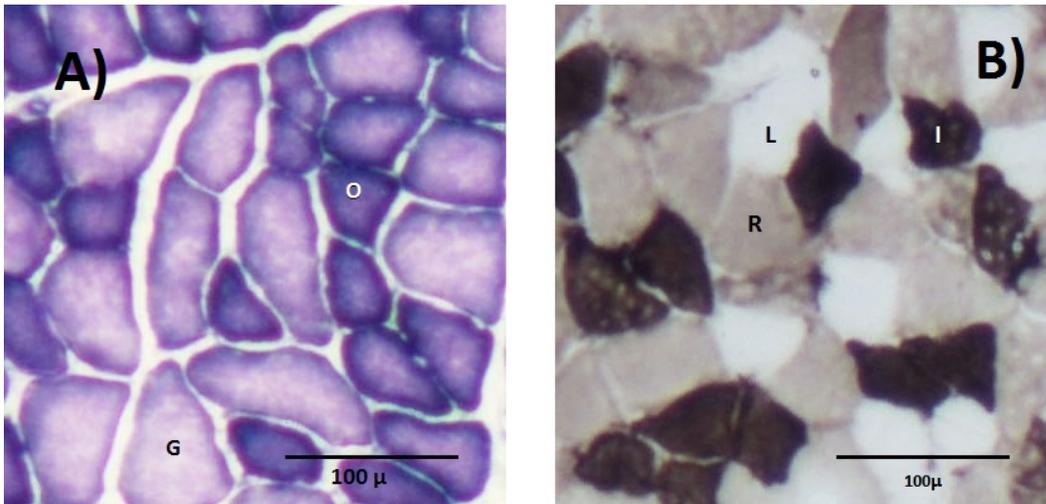


Figura 4.1: A) Segmento de imagen con tinción NADH. O: Oxidativas, G: Glucolíticas  
 B) Segmento de imagen con tinción ATPasa. L: Lentas, I: Intermedias, R: Rápidas.

## 4.2. Requerimientos del usuario

El departamento de Neurociencias, Biofísica y Fisiología, del Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, unidad Zacatenco requiere un sistema de procesamiento de imágenes histológicas que, de acuerdo con la Sección 2.2.2 se pretende completar con los siguientes componentes:

- Dispositivo de adquisición: Cámara fotográfica digital Nikon DS-R1-U3 de 12.7 Megapíxeles (4,076 x 3,116) capaz de obtener imágenes de un microscopio, propiedad del CINVESTAV.
- Hardware especializado en procesamiento de imágenes: comprende desde la cámara ya mencionada incluyendo los dispositivos necesarios para adaptar ésta al microscopio y a la computadora encargada de recibir, preprocesar, seleccionar y almacenar las imágenes.
- Las computadoras a utilizar van desde equipos de cómputo de estudiantes, comúnmente portátiles de potencia media, hasta computadoras de escritorio de altas potencias o Workstations, es decir que las prestaciones y los sistemas operativos serán muy variados.

- Software: es el tema principal de este trabajo, se requiere que permita la clasificación de fibras musculares de manera automática en una imagen histológica teñida con tinciones NADH, ATPasa y COX de acuerdo a su coloración. También que cuente con la capacidad de conteo de las fibras musculares de cada uno de los tipos de tinción, así como su localización dentro de la imagen en formato de coordenadas (x,y). Es importante la visualización y edición de los resultados sobre la imagen analizada y el almacenamiento en formatos de hoja de cálculo par análisis posteriores. Debe soportar su ejecución sobre los sistemas operativos más utilizados en la actualidad, con la optimización necesaria para que las computadoras que se pretenden utilizar puedan ejecutarlo sin problemas.
- Los dispositivos de almacenamiento son en principio los contenidos en las computadoras utilizadas y posteriormente se integrara todo en discos duros externos, los cuales también fungirán como dispositivos de copia fuente.
- Actualmente la disponibilidad de redes no es un problema, este sistema usa herramientas como Dropbox.

### 4.3. Diseño

Siguiendo el modelo presentado en la Sección 2.2.1 se incluyen los siguientes pasos en este sistema:

- La adquisición es llevada a cabo por la cámara mencionada en la Sección 4.2, además del software que la acompaña.
- La mejora es llevada a cabo mediante el uso del programa Photoshop, para la eliminación de ruido, homogeneización del fondo y unión de varias tomas en un único archivo.
- Las restauraciones más simples y en la mayoría de los casos suficientes consisten en la eliminación de ruido mediante máscaras de medias o de medianas. Sin embargo son optativas, pues no todas las imágenes las requieren con la misma intensidad.
- Para el procesamiento del color se usan algoritmos de minería de datos para la determinación de los colores que pertenecen a cada uno de los tipos de fibra

presentes en una imagen histológica. Se implementan tres algoritmos: K-means, Fuzzy C-means y Mapas autoorganizados de Kohonen.

- Las técnicas de compresión utilizadas son las estándar, es decir, según extensiones de archivo, sin embargo ya están implementadas en el lenguaje java.
- Para la segmentación de las imágenes, según los tipos de fibra encontrados de acuerdo a su color, se propone la utilización del algoritmo de análisis de conglomerados difuso Fuzzy C-means.
- Las técnicas de procesamiento morfológico implementadas son las más básicas: erosión, dilatación y combinaciones de éstas. Son utilizadas para la discriminación de objetos tan pequeños obtenidos después de la etapa de segmentación que no pertenecen a fibras musculares.
- El reconocimiento, en este caso de las fibras musculares, se lleva a cabo mediante la modificación del algoritmo de etiquetado de componentes conexas, lo que le añade la capacidad del cálculo del centroide y la medición de áreas de cada objeto reconocido.

Se observa que, como fue planteado, no es necesaria la inclusión de toda la complejidad de interacción de los pasos del procesamiento, en este caso no fue implementado el análisis mediante Wavelets, además la etapa de representación y descripción es cubierta durante la segmentación y el reconocimiento, pues las características de interés son el color, la localización y el tamaño. También cabe mencionar que el orden de las etapas de procesamiento morfológico y de segmentación fueron intercambiados, pues para nuestros propósitos este orden funciona de mejor manera. Finalmente la base de datos consiste únicamente de los colores resultado de la etapa de procesamiento de color, es decir los colores representativos de cada tipo de fibra.

De acuerdo con los requerimientos de los usuarios de este sistema se propone que el procedimiento para llegar al objetivo presentado se divida en distintos procesos específicos, como se muestra en la figura 4.2.

Se distinguen cuatro principales fases: aprendizaje, segmentación, filtrado y conteo. Dichas fases están inspiradas en la identificación de las fibras mediante la visión humana: la característica principal es el color, por lo que en la etapa de entrenamiento se pretende determinar los colores representativos de cada uno de los tipos de fibras. Otro aspecto importante es el tamaño y la forma, por lo que se aplica una etapa de segmentación para analizar esta característica y finalmente se realiza el conteo

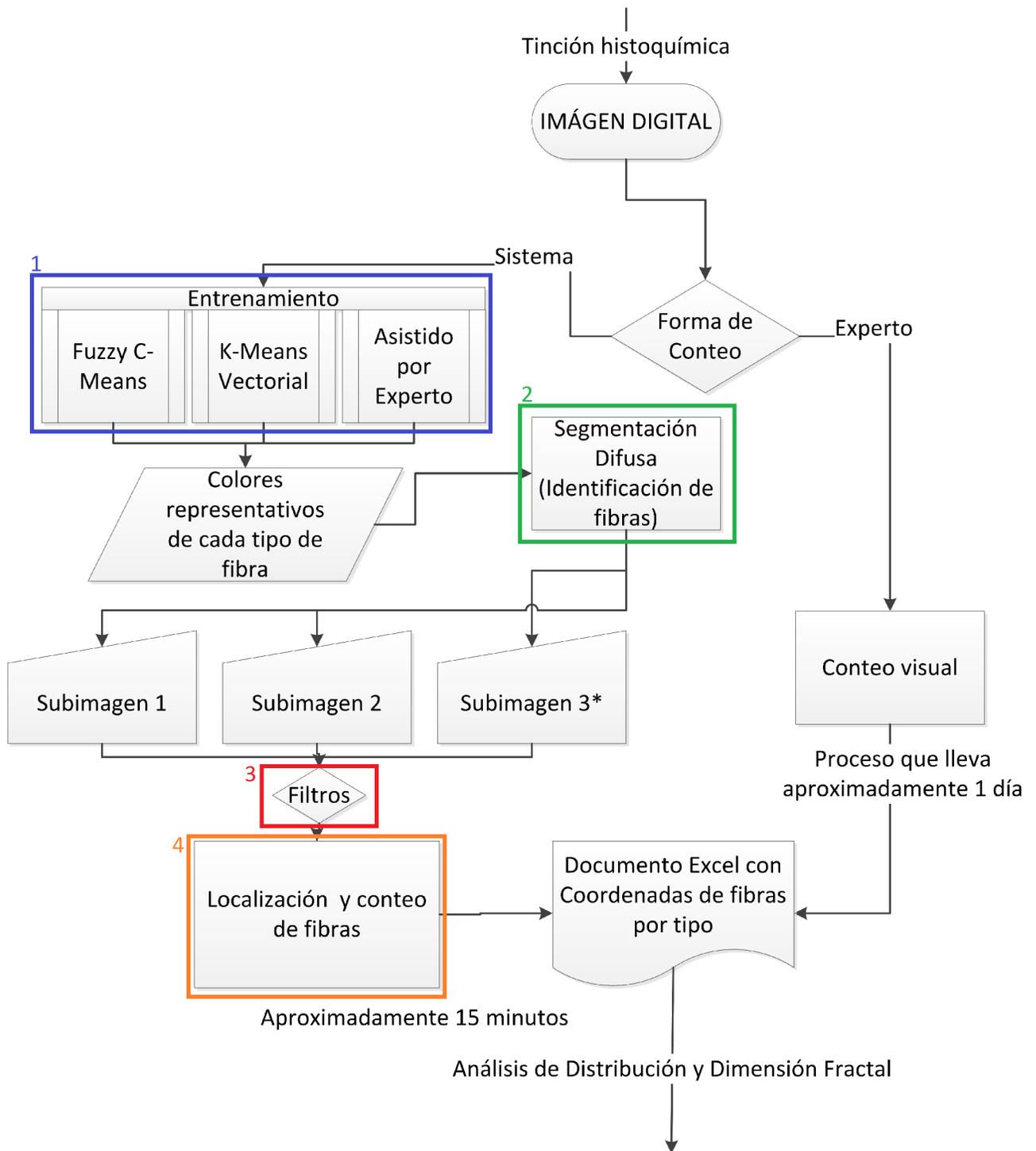


Figura 4.2: Flujo del proceso seguido.

y ubicación de fibras dentro de la imagen. Sin embargo es necesario la inclusión de una etapa de filtrado para mejorar los resultados de la etapa del conteo por parte del sistema.

Cabe mencionar que esta separación resulta más evidente pues para la ejecución de cada una de estas etapas son necesarias todas las anteriores, por lo tanto este software se estructuró de manera que cada una de estas tareas fueran realizadas por una ventana en específico.

Las ventanas diseñadas se muestran en la figura 4.4. Como su nombre lo indica, la ventana Main 4.3 es la ventana principal, es la primera que se muestra al ejecutar el programa, contiene todos los botones para abrir las demás ventanas.

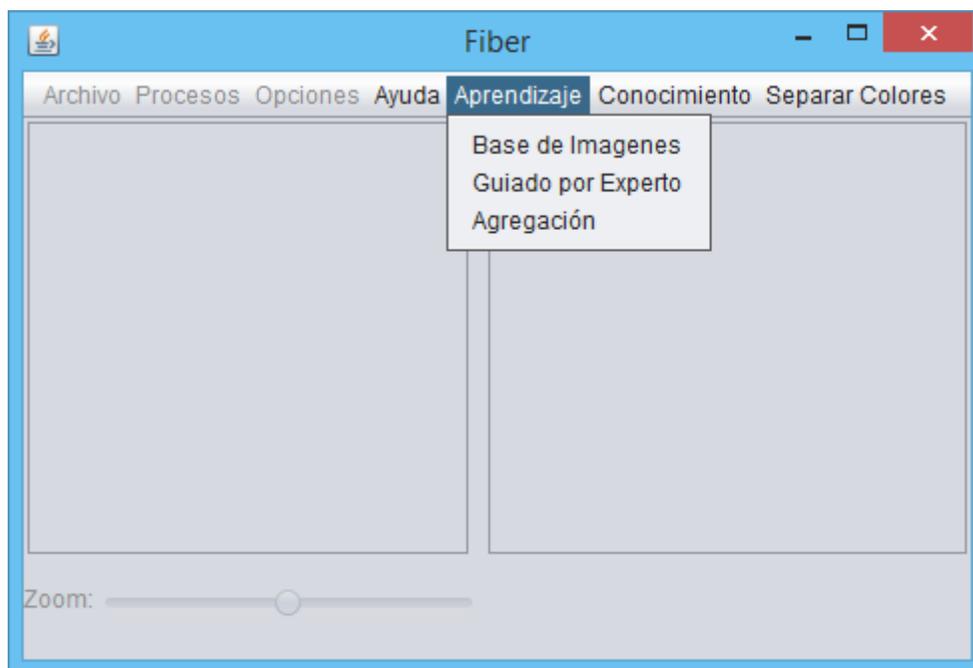


Figura 4.3: Captura de pantalla de la interfaz gráfica principal.

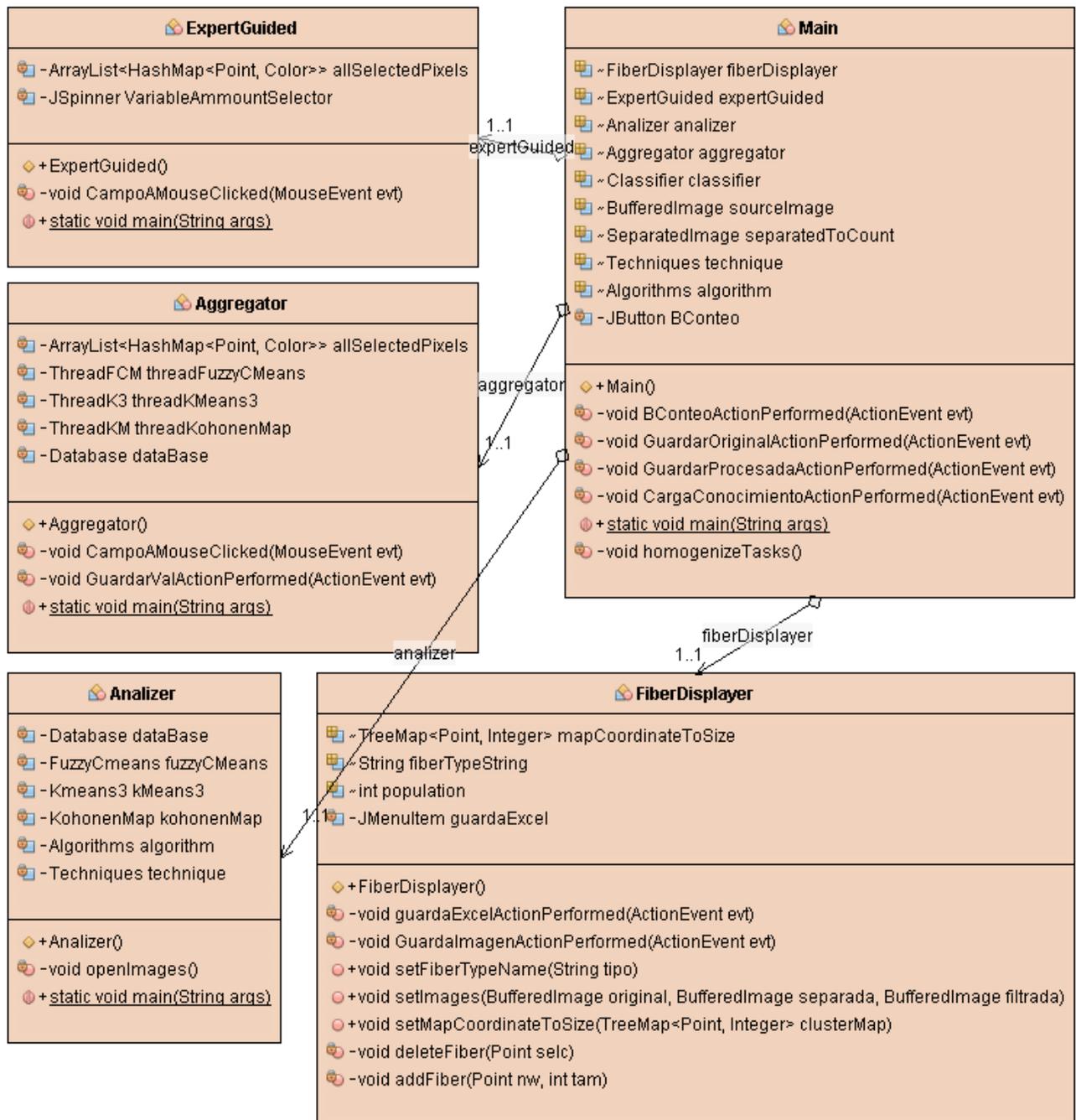


Figura 4.4: Diseño de ventanas.

### 4.3.1. Aprendizaje

Las ventanas *Analyzer*, *ExpertGuided* y *Aggregator* cubren la primera fase del proceso: el aprendizaje. Siguiendo el esquema presentado en la Sección 2.3, esta fase del proceso se compone de la siguiente manera:

- El conocimiento previo relevante es que cada imagen contiene un corte transversal de tejido muscular, teñida con NADH, ATPasa o COX y el objetivo es encontrar un valor representativo de cada tipo de fibra presente.
- El conjunto de datos de interés son todos los colores de los píxeles en formato RGB que componen a una imagen, los cuales son tratados como vectores de tres componentes enteras, aunque si la imagen es demasiado grande se tiene que reducir en dimensiones conservando proporciones y utilizando herramientas disponibles en Java, lo último con el propósito de reducir el espacio de memoria requerido, así como el tiempo de procesamiento, pero conservando la mayor cantidad de información posible.
- El preprocesamiento de los datos incluye la edición de las imágenes en programa Photoshop para eliminar el ruido generado en el momento de adquisición, se elimina el fondo, se borran objetos indeseados y se unen varias tomas parciales de un músculo.
- La etapa de minería de datos comprende la utilización de los algoritmos de análisis de conglomerados K-means, Fuzzy C-means y Mapas autoorganizados de Kohonen.
- La interpretación no tiene lugar, dado que los resultado de los algoritmos usados son los colores que representan a cada tipo de fibra.
- El conocimiento obtenido se usará para la segmentación de una imagen histológica en subimágenes, cada una conteniendo solo un tipo de fibra.

#### *Analyzer*

En esta ventana es posible agregar un conjunto de imágenes de la misma tinción para ejecutar los algoritmos de minería de datos, además permite la visualización de los histogramas de color de cada uno de los canales rojo, verde y azul, que son las componentes del modelo de color RGB, como se muestra en la figura 4.5.

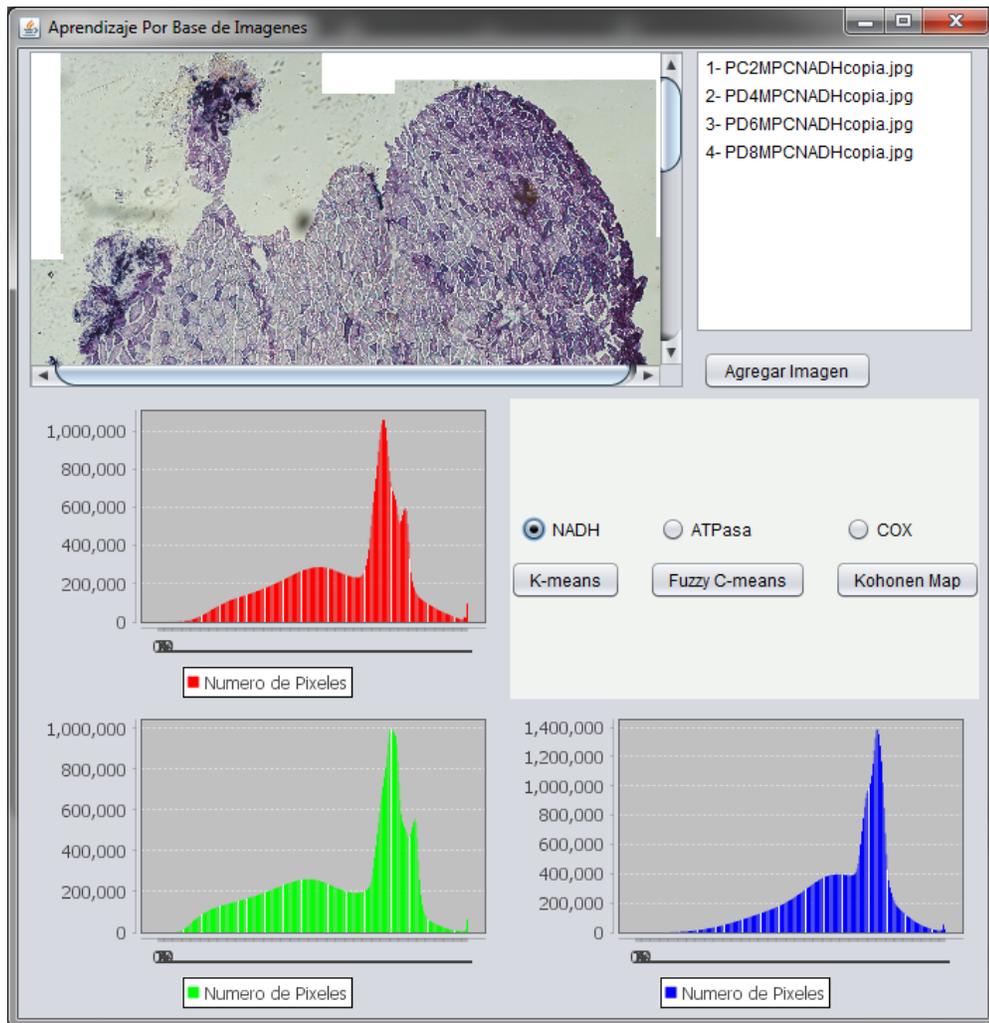


Figura 4.5: Ventana Analyzer para aprendizaje por minería de datos.

Los algoritmos utilizados se presentan a continuación, se observa que para su ejecución, la cantidad de conglomerados debe ser identificada de antemano, lo cual en esta aplicación es posible pues se sabe cuantos grupos se deben buscar: para las imágenes con tinción NADH y COX se ejecuta el algoritmo con  $k = 3$  y para la tinción ATPasa  $k = 4$ ; la razón por la que se considera un grupo más que la cantidad de tipos de fibras presentes es que también se debe considerar el color de fondo y del tejido conectivo, pues inherentemente formarán un grupo pues su área dentro de cada imagen es considerable.

El algoritmo *K-means* algoritmo de búsqueda es un modelo de análisis de conglomerados, de enfoque particional, agrupa los datos en  $k$  grupos diferentes, dando como resultado los representantes de cada grupo (centroide). El procedimiento se muestra a continuación (véase [18]).

1. Se eligen los valores de los centroides  $c_1, c_2, \dots, c_k$ . de partida, suelen ser de manera aleatoria.
2. En cada iteración se reparten los datos de aprendizaje (colores)  $x$  entre los  $k$  conglomerados  $\bar{C}_j$ . Cada dato se asigna al conglomerado cuyo centroide dista menos.
3. Se calculan los nuevos centroides de cada conglomerado como promedio de los datos de aprendizaje asignados en el paso 2. Así llamando  $N_j$  al número de datos que han correspondido al conglomerado  $j$  en el reparto, se tiene:

$$c_j = \frac{1}{N_j} \sum_{x \in \bar{C}_j} x.$$

4. Si los valores de los centroides varían respecto de la iteración anterior en menor medida que cierto umbral  $\epsilon$  determinado previamente, el algoritmo ya ha convergido. Si no es así, volver al paso 2.

En este trabajo, dado que los datos que se manejan son colores en formato RGB que van de 0 a 255, el valor del umbral  $\epsilon$  que se utiliza se escogió igual a 0.1, pues se considera que un cambio de esa magnitud es poco significativo. Además, si este valor se disminuye, los resultados son casi idénticos a los obtenidos con el valor seleccionado, pero sin embargo la cantidad de iteraciones crece desde un promedio de 9 hasta incluso más de 20, con lo que el tiempo de procesamiento se ve afectado.

El algoritmo *Fuzzy C-Means* es la versión difusa del algoritmo *K-Means* (para más información consultar [21]). Particiona un conjunto de  $N$  datos en  $C$  conglomerados minimizando la función objetivo:

$$J = \sum_{n=1}^N \sum_{i=1}^C \mu_i(x_n)^{m'} \|x_n - m_i\|^2,$$

donde  $1 \leq m' < \infty$  es un parámetro de difusión,  $m_i$  es el  $i$ -ésimo centro de conglomerado, y  $\mu_i(x_n) \in (0, 1)$  es la membresía del  $n$ -ésimo dato al  $i$ -ésimo conglomerado y se calculan:

$$m_i = \frac{\sum_{n=1}^N \mu_i(x_n)^{m'} \cdot x_n}{\sum_{n=1}^N \mu_i(x_n)^{m'}}, \quad (4.1)$$

$$\mu_i(x_n) = \frac{1}{\sum_{j=1}^C \left( \frac{\|x_n - m_i\|}{\|x_n - m_j\|} \right)^{\frac{2}{m'-1}}}. \quad (4.2)$$

Resaltando las propiedades  $\sum_{i=1}^C \mu_i(x_n) = 1, \forall n$  y  $\sum_{n=1}^N \mu_i(x_n) > 0, \forall i$ ; las cuales cobran gran relevancia en la etapa de segmentación. El algoritmo procede de la siguiente manera:

1. Asignar los centroides iniciales  $m_i$ . Elegir el valor del difusor  $m'$  y el umbral  $\epsilon$ . Asignar el contador de iteración  $t = 0$ .
2. Repetir pasos 3 y 4 incrementando  $t$ , hasta que  $\|m_i(t) - m_i(t-1)\| < \epsilon$ .
3. Calcular  $\mu_i(x_n)$  según la fórmula anterior
4. Calcular los centroides  $m_i$ .

Para su ejecución en el presente software, el parámetro  $m'$  se asignó igual a 2, ya que su elección es propuesta en el intervalo  $[1.5, 2.5]$  en la referencia [23], además que al modificarlo no se encontraron diferencias en el resultado obtenido. Aquí el umbral  $\epsilon$  se asigna como 0.3, más alto que en el algoritmo K-means pues cada iteración tarda más en completarse en este algoritmo, además que de igual manera, al elegirlo más pequeño no se encontraron diferencias significativas en los resultados por tratarse los datos colores en formato RGB.

Finalmente el algoritmo de los mapas autoorganizados de Kohonen presenta una red neuronal de tipo competitivo que lleva a cabo un entrenamiento no supervisado (véase [21]), por lo que en esencia son parecidos a los algoritmos de análisis de conglomerados. Se inspiran en el comportamiento observado en neuronas del corteza, las cuales al activarse ejercen una acción sobre las neuronas a su alrededor en función de su distancia. Cada neurona tiene un vector modelo, y los datos de entrada se clasificarán de acuerdo a la neurona que tenga mayor activación, esta es la neurona que tiene su vector modelo más parecido al dato. Entonces el algoritmo de entrenamiento consiste en el ajuste de los vectores modelo para que las activaciones sean las apropiadas en el momento de la clasificación.

Inicialmente se deben ordenar las neuronas en un arreglo, comúnmente de dos dimensiones, aunque también puede ser de una sola dimensión o inclusive hasta de tres. También se debe determinar como se medirá la distancia entre las neuronas en el arreglo, se puede usar distintas métricas como la Euclidiana o la de Manhattan (la cual define la distancia entre  $X = (x_1, x_2, \dots, x_n)$  y  $Y = (y_1, y_2, \dots, y_n)$  como  $d(X, Y) = \sum_{i=1}^n |x_i - y_i|$ ). Sin embargo en esta aplicación se utilizan arreglos de una dimensión, y por lo tanto ambas métricas son equivalentes. Véase el algoritmo 1.

### ***Expert Guided***

En la práctica algunas imágenes presentan retos importantes para los algoritmos de minería de datos, por una parte existen datos similares desde el punto de vista de minería de datos, que sin embargo biológicamente pertenecen a fibras distintas, en su mayoría en las imágenes de tinción histoquímica ATPasa, donde es difícil diferenciar las fibras de tono intermedio de aquellas de tono oscuro, por otra parte, completamente opuesta, una parte importante de las fibras no contienen un solo color homogéneo, sino un degradado que puede ir de un color que representa un tipo de fibra hasta otro color de un tipo de fibra diferente. Ambas situaciones llevan a que los algoritmos dan resultados que no están de acuerdo con la opinión de los expertos.

Por esta razón fue necesario el desarrollo de un método alternativo de asignación de color representativo a cada uno de los tipos de fibra, además de una interfaz gráfica amigable que permita incluir estos resultados en el flujo de ejecución del software como alternativa de aquellos que arrojaron los algoritmos sin perder su disponibilidad.

En la figura 4.6 se muestra la ventana diseñada, en el panel de la izquierda se selecciona el tipo de imagen histológica elegida y posteriormente para cada tipo de

**Algoritmo 1:** Mapas de Kohonen.**Datos:**  $k, \alpha, \eta, v, error_{maximo}, x_j, j \in \{1, 2, \dots, T\}$ **Resultado:** Vectores modelo entrenados  $m_i, i \in 1, 2, \dots, k$ **inicio**

Iniciar vectores modelo aleatoriamente

**repetir** $error \leftarrow 0$ Guardar vectores modelo actuales en  $w_i \leftarrow m_i, i \in \{1, 2, \dots, N\}$ **para**  $x_j, j \in \{1, 2, \dots, T\}$  **hacer** $d_{min} \leftarrow 512$ **para**  $m_i, i \in \{1, 2, \dots, k\}$  **hacer** $d_{ij} \leftarrow d(m_i, x_j)$ **si**  $d_{ij} < d_{min}$  **entonces** $d_{min} \leftarrow d_{ij}$  $g \leftarrow j$  $m_g \leftarrow m_g + \alpha(x_j - m_g)$  $m_k \leftarrow m_k + \alpha\eta(x_j - m_k)$  para  $k \neq g$  y  $k$  dentro de la región de vecindad. $error \leftarrow 0$  **para**  $m_i, i \in \{1, 2, \dots, k\}$  **hacer** $error_i \leftarrow d(m_i, w_i)$ **si**  $error_i > error$  **entonces** $error \leftarrow error_i$ Reducir  $\alpha, \eta$  y el radio de vecindad  $v$ **hasta que**  $error \geq error_{maximo}$ 

fibra, se selecciona en el panel abajo a la izquierda y se hace clic sobre la imagen que se mostrará en el panel de la derecha. Se eligen 10 fibras de cada tipo y posteriormente se hace clic en el botón "Guardar Valores" que se habilitará cuando se haya terminado el proceso para todos los tipos de fibra. El valor asignado a cada tipo de fibra es simplemente el promedio de los colores de los diez píxeles sobre los que se hizo clic, que sin embargo pese a su sencillez logra mejorar los resultados del software.

### ***Aggregation***

Las ventanas anteriores son útiles cuando se desea agregar conocimiento con un método específico, ofreciendo la comodidad de elegir alguno y que sus resultados coexistan con los demás. Sin embargo, cuando se requiere tener disponibilidad de todos los resultados para su comparación este proceso puede aumentar considerablemente el tiempo necesario para este propósito: se debe ejecutar cada uno de los algoritmos por separado en la ventana de *Analizer* y, posteriormente, abrir la ventana de *ExpertGuided* para agregar el conocimiento manualmente.

Por esta razón, aprovechando las capacidades de procesamiento paralelo de Java, se diseñó la ventana de *Aggregation*, con la misma interfaz gráfica que *ExpertGuided*, pero en esta se ejecutan todos los algoritmos de minería de datos durante la indicación manual de los colores de cada fibra por parte del experto. Además este método conforma una fuente de conocimiento más, la agregación de los resultados obtenidos como promedio ponderado, dando más peso a los resultados del experto, pero tomando en cuenta a los de los algoritmos.

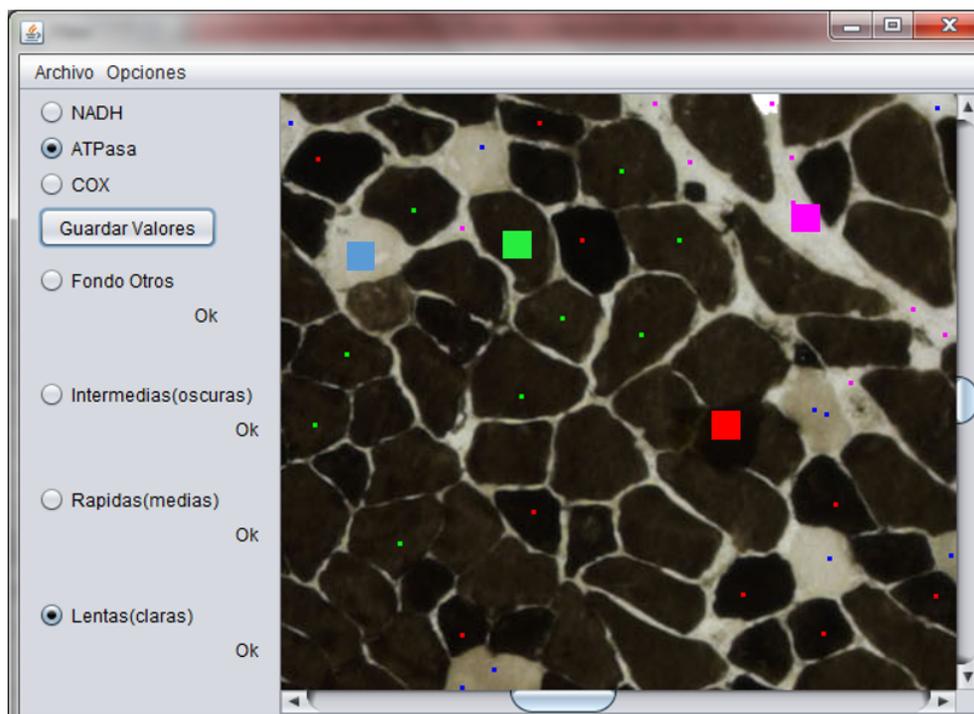


Figura 4.6: Interfaz Gráfica Entrenamiento Asistido.

### 4.3.2. Segmentación

La segunda etapa del flujo de procesamiento propuesto consiste en la preparación de la imagen histoquímica para el conteo de cada uno de los tipos de fibra existentes, utilizando la información previamente obtenida en la etapa de aprendizaje, con la capacidad de seleccionar cualquier de las fuentes descritas. Este proceso se realiza en la ventana *Main*, mostrada en la figura 4.7.

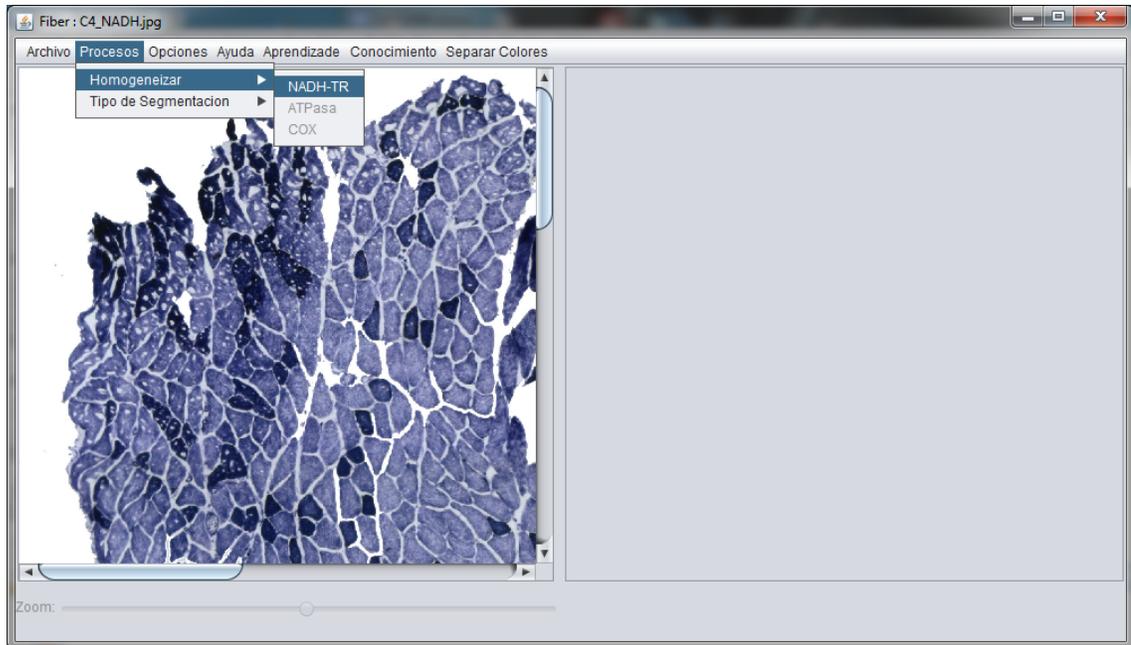


Figura 4.7: Interfaz para la separación.

Los algoritmos de conteo de objetos más comunes realizan esta tarea únicamente sobre imágenes binarias, o equivalentemente eligiendo un umbral entre dos clases de color. En este sistema se utiliza la primera técnica, pues adicionalmente resulta útil para la validación de la segmentación por parte del experto.

Por lo tanto cada una de las imágenes histoquímicas se separa en subimágenes que contienen únicamente un tipo de fibra y color blanco de fondo. La primera aproximación para llevar esto a cabo es clasificando cada uno de los píxeles en el grupo al que se asemeje más.

Los algoritmos de análisis de conglomerados devuelven como resultado los valores representativos de cada uno de los conglomerados, es decir una partición del conjunto

de datos inicial, y más aún, de su dominio. Por esta razón, inherentemente inducen la existencia de un Diagrama de Voronoi (ver [4]) sobre el dominio de los datos, el cual es el espacio de color RGB.

Los Diagramas de Voronoi sobre un espacio euclidiano generan, dado un conjunto de puntos  $P$  (en este caso los centroides de los conglomerados), un conjunto de polígonos tales que el polígono  $V_i$  contiene solamente al punto  $p_i \in P$  y además todos los puntos  $x$  del espacio tales que  $d(x, p_i) < d(x, p_j)$  con  $j \neq i$ . Esto produce que las aristas de estos polígonos sean las mediatrices de los segmentos que unen a puntos  $p_l$  vecinos. En la figura 4.8 se observa un ejemplo de diagrama de Voronoi.

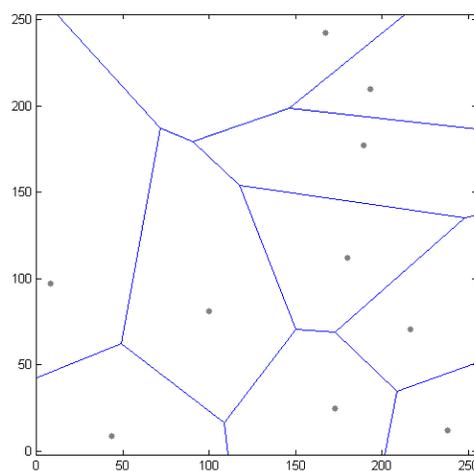


Figura 4.8: Diagrama de Voronoi de 10 datos aleatorios.

Sin embargo surge el problema de que existe la posibilidad de que haya píxeles en una imagen que no deban pertenecer a ningún grupo (manchas) y que sin embargo son clasificadas y contadas como de un tipo de fibra, lo que genera errores en el conteo. Para contrarrestar este problema se busca una forma de clasificación en la que no se asigne al conglomerado más cercano o al que sea más semejante, sino al que dicho valor cumpla criterios más específicos.

La propuesta de este trabajo es la siguiente: Se utiliza la función de pertenencia difusa (fórmula 4.2):

$$\mu_i(x_n) = \frac{1}{\sum_{j=1}^c \left( \frac{\|x_n - m_i\|}{\|x_n - m_j\|} \right)^{\frac{2}{m'-1}}},$$

y sus propiedades  $\sum_{i=1}^K \mu_i(x_n) = 1, \forall n$  y  $\sum_{n=1}^N \mu_i(x_n) > 0, \forall i$ , donde  $K$  es la cantidad

de grupos utilizados para la etapa de aprendizaje.

Dado un dato  $x_n$  se determina su pertenencia a cada uno de los centroides dados y se considera su asignación a un grupo si su pertenencia es mayor que  $\frac{0.9}{K-1}$  (valor propuesto en este trabajo y que experimentalmente genera los resultados deseados). Dicho criterio permite que en los casos donde un píxel tiene pertenencia muy alta a  $y$  grupos ( $y = 1, 2, \dots, k-1$ ) y muy baja a otros  $k-y$  se asigne a los  $y$  primeros grupos simultáneamente ya que la decisión es más complicada pues se podría considerar que es un término alrededor de las fronteras de las zonas de pertenencia de los  $n$  valores. En la práctica, los centroides se encuentran esparcidos muy cercanos al eje R=G=B, como se muestra en la figura 4.9, donde se puede observar que las regiones de clasificación que generan los polígonos de Voronoi podrían resultar excesivamente alejadas de los centroides de cada grupo, además se muestra el comportamiento buscado en la propuesta dada.

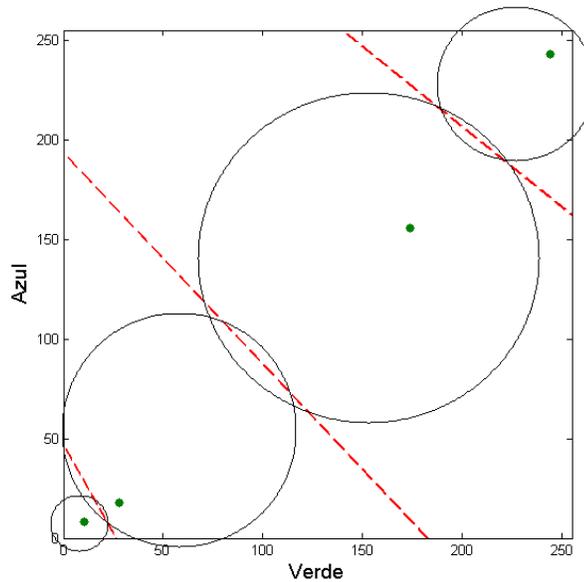


Figura 4.9: Vista de separación sobre el plano Verde-Azul, de las regiones de clasificación según los algoritmos de análisis de conglomerados (líneas punteadas rojas) y aproximación de como se comportaría la propuesta presentada.

#### 4.3.2.1. Primer Conteo

Sobre las imágenes resultantes de la separación se cuentan los cúmulos de píxeles del mismo color, pues se espera que cada uno corresponda a una fibra muscular. Se utiliza un algoritmo de etiquetado de componentes conexas, descrito en la Sección 4.3.4:

El resultado de este algoritmo es una lista de las coordenadas de los píxeles que fueron guardados, es decir la ubicación del primer píxel encontrado correspondiente a una fibra. Además este sistema permite corroborar los resultados mediante la superposición de las coordenadas obtenidas sobre la imagen procesada.

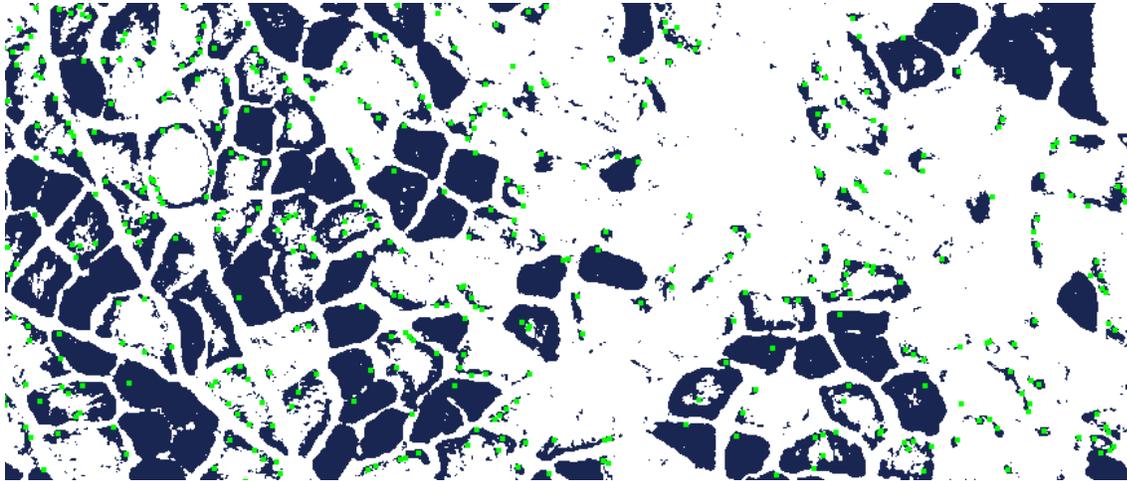


Figura 4.10: Coordenadas obtenidas del primer conteo, después de la separación.

Como se puede observar en la figura 4.10, el algoritmo de conteo implementado da como resultado una mayor cantidad de fibras musculares (puntos color verde) que las que un humano percibe, pues como las fibras musculares presentan una gama de tonalidades, algunos píxeles que se encuentran dentro de una fibra, se clasifican dentro de un grupo distinto al que pertenece la fibra y en las imágenes separadas aparecen como puntos aislados, pero que sin embargo el algoritmo de conteo toma en cuenta y los agrega a los resultados.

Es por esta situación, dada por la naturaleza homogénea de las imágenes que se necesita un paso más en el proceso propuesto en el presente trabajo: la aplicación de filtros.

1	1	1
1	1	1
1	1	1

Figura 4.11: Elemento estructural utilizado.

### 4.3.3. Filtrado

Basándose en 2.2.1 se recurre a algoritmos de procesamiento morfológico con el propósito de extraer objetos relevantes para nuestro problema, en este caso objetos aislados que correspondan a fibras musculares, con el propósito de corregir los resultados del conteo.

Se requiere que las componentes conexas que no correspondan a fibras musculares sean eliminadas de la imagen, pero que sin embargo las que realmente representan fibra sean conservadas para la etapa de conteo. Se propone la utilización de los algoritmos de procesamiento morfológico más básicos: la dilatación y la erosión así como su combinación en tiempo de procesamiento de acuerdo a los resultados obtenidos por cada uno.

Ambos algoritmos se basan en la superposición y recorrido sobre la imagen de máscaras de píxeles, llamadas elementos estructurales, que de acuerdo al algoritmo generaran un efecto u otro. Cabe mencionar que los elementos estructurales usados son los más básicos, que se pueden diseñar elementos estructurales para propósitos específicos, pero se trató de ser lo más general posible.

Los algoritmos de procesamiento morfológico quedan en la línea divisoria entre proceso de medio y de bajo nivel (véase [10]), pues comúnmente son utilizados para extraer características, pero en el caso de la aplicación presentada, su propósito también es la reducción de ruido. El elemento estructural usado se puede ver en la figura 4.11.

#### Erosión

Dada una imagen binaria, erosiona los objetos que se encuentran en la imagen, desgastándolo y disminuyendo su tamaño en píxeles. La aplicación de este algoritmo es con el propósito de poder ser utilizado las veces necesarias para que se eliminen los cúmulos pequeños que no deberían pertenecer a fibras. Como se muestra en la

figura 4.12, erosiona los objetos más pequeños y conserva una parte de los grandes.

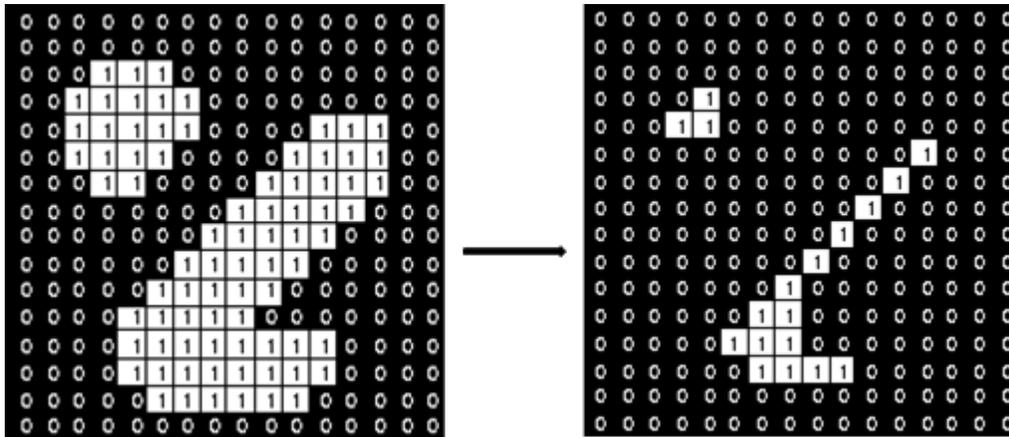


Figura 4.12: Filtro de erosión.

El seudocódigo es el siguiente:

---

**Algoritmo 2:** Seudocódigo erosión.

---

**Datos:** ImagenBinaría

**Resultado:** Imagen Resultado de la erosión

**inicio**

    Crear *ImageResultado*

**para** todos los píxeles  $IB(i, j)$  que no sean bordes **hacer**

        Centrar Elemento Estructural

**si** la *Elemento Estructural* interseca por lo menos un píxel blanco

**entonces**

                píxel  $IR(i, j) \leftarrow$  blanco

**en otro caso**

            píxel  $IR(i, j) \leftarrow$  color

Otra manera de entender la erosión es determinar si el elemento estructural está completamente contenido en un conjunto de píxeles, si es así entonces el píxel central de la imagen resultado debe ser incluido. Por lo tanto si se encuentran cúmulos pequeños, se propone aplicar erosiones gradualmente, para eliminarlas, pero conser-

vando en mayor medida posible los cúmulos que sí pertenecen a fibras musculares reales.

En la figura 4.13 se observan los resultados de la aplicación consecutiva del filtro de erosión sobre una imagen de fibras glucolíticas de la tinción NADH.

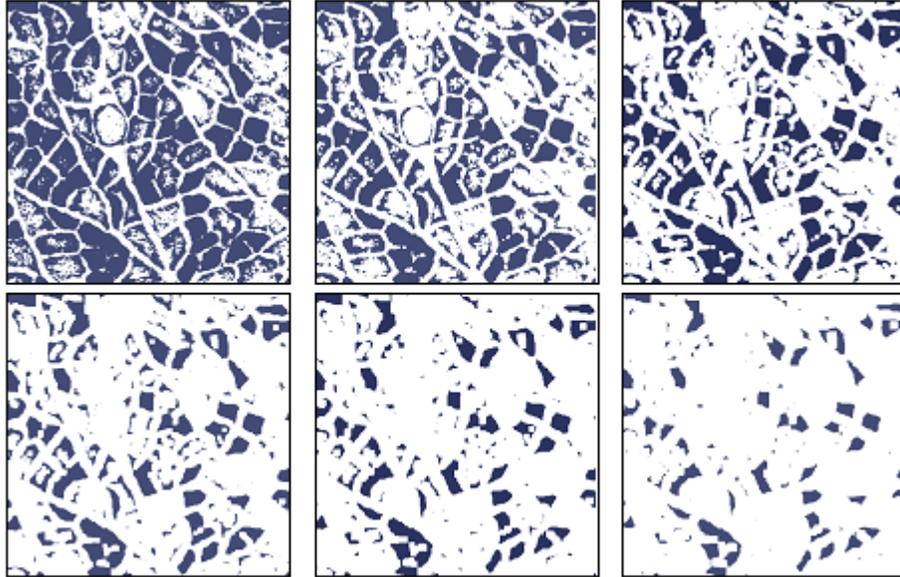


Figura 4.13: Visualización del efecto del filtro de erosión.

## Dilatación

El proceso de dilatación es análogo a la erosión, puede ser vista como una erosión al fondo de la imagen, es decir tiene el efecto de hacer más grandes a los objetos contenidos en ésta, lo que lo hace de interés pues tenemos como objetivo determinar el tamaño de las fibras identificadas, tamaño en píxeles que se ve afectado bajo la erosión, por lo que se usa el proceso inverso para regresar a un tamaño cercano al original a los cúmulos que no fueron eliminados en la erosión, es decir aquellos que tienen altas posibilidades de representar realmente una fibra muscular.

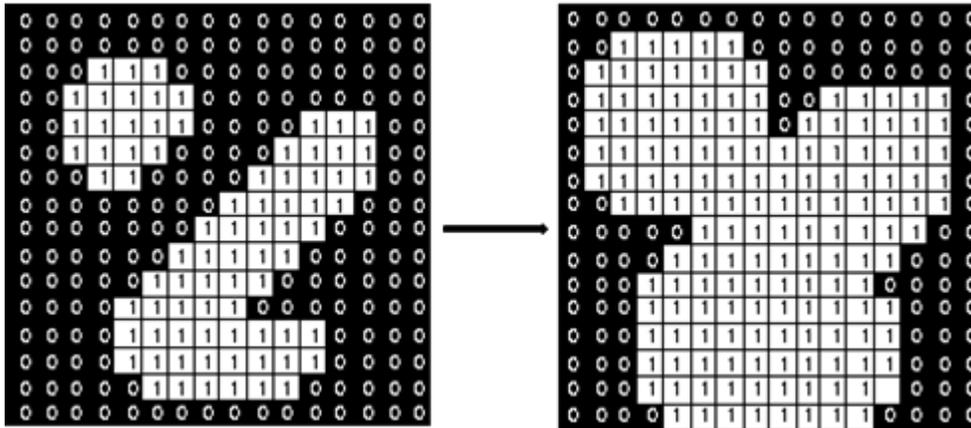


Figura 4.14: Filtro de dilatación.

El pseudocódigo es muy similar a la erosión: se recorre el elemento estructural de la figura 4.11, pero como ya se mencionó, el cambio en la condición necesaria para la inclusión de un píxel en la imagen resultado, marca grandes diferencias aunque se trata del mismo elemento estructural.

---

**Algoritmo 3:** Seudocódigo eilatación.
 

---

**Datos:** ImagenBinaria

**Resultado:** Imagen Resultado de la dilatación

**inicio**

  Crear *ImageResultado*

**para** todos los píxeles  $IB(i, j)$  que no sean bordes **hacer**

    Centrar Elemento Estructural

**si** la *Elemento Estructural* intersecta por lo menos un píxel de color

**entonces**

        píxel  $IR(i, j) \leftarrow color$

**en otro caso**

      píxel  $IR(i, j) \leftarrow blanco$

---

En la figura 4.15 se observan los el efecto de dilatar consecutivamente sobre la imagen resultante de la ultima erosión de la figura 4.13. Es notable que el tamaño de los cúmulos que no fueron eliminados se acerca mucho al tamaño original, pero sin embargo se nota una eliminación de ruido con respecto a la imagen original.

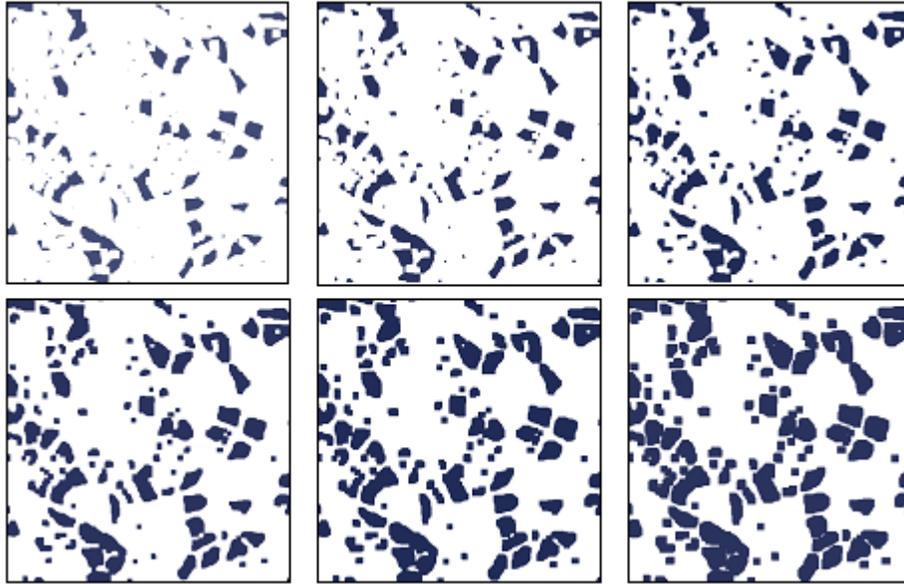


Figura 4.15: Ilustración del filtro de dilatación después de erosionar.

Este software permite la elección de filtros predefinidos en la ventana Main, los cuales se eligieron de acuerdo a los resultados que arrojaron cada uno en la etapa de desarrollo para distintas imágenes a diferentes resoluciones.

#### 4.3.4. Conteo

Dentro de los objetivos principales de este sistema se encuentra la localización y conteo de las fibras musculares de cada tipo presentes en una imagen. Los pasos anteriores sirvieron para la clasificación de acuerdo al color, además del filtrado de acuerdo al tamaño con el propósito de facilitar el proceso de conteo.

Se implementó el algoritmo de etiquetado de componentes conexas (descrito en [10]) y además se adaptó para la determinación del centroide y la medición del área. Como el nombre lo sugiere, este algoritmo identifica cuántos objetos se encuentran una imagen, entendiendo como objeto un grupo de píxeles del mismo color interconectados entre sí, en el caso de imágenes binarias se decide cual es el color de fondo (en este caso blanco) y cual el de los objetos de interés (el color de las fibras

musculares en imágenes separadas, resultado de la segmentación).

---

**Algoritmo 4:** Etiquetado de componentes conexas.

---

**Datos:** ImagenBinaria

**Resultado:** Lista de coordenadas  $LC$  donde comienza un objeto

**inicio**

```

para Todos los píxeles  $p(i, j)$  con  $i \in \{1, 2, \dots, ancho\}$  y
   $j \in \{1, 2, \dots, alto\}$  hacer
  si  $p(i, j)$  no está marcado entonces
    Guardar  $(i, j)$  en  $LC$ 
    5 Recorrido( $i, j$ )

```

---

Como se observa en el seudocódigo 4 se recorre toda la imagen en busca de píxeles de color, se marcan y se buscan sus vecinos, es decir los píxeles que lo rodean y por lo tanto pertenecen al mismo objeto dentro de la imagen.

Existen varios conceptos de vecindad de píxeles, las cuales se pueden ver en la figura 4.16 y se describen a continuación:

- 4-vecindad : Los vecinos de un píxel con coordenadas  $(x, y)$  son los píxeles  $(x, y - 1)$ ,  $(x - 1, y)$ ,  $(x + 1, y)$ ,  $(x, y + 1)$ .
- 4-vecindad diagonal : Los vecinos de un píxel con coordenadas  $(x, y)$  son los píxeles  $(x - 1, y - 1)$ ,  $(x + 1, y - 1)$ ,  $(x - 1, y + 1)$ ,  $(x + 1, y + 1)$ .
- 8-vecindad : Los vecinos del píxel con coordenadas  $(x, y)$  son los que pertenecen a la 4-vecindad y los de la 4-vecindad diagonal.

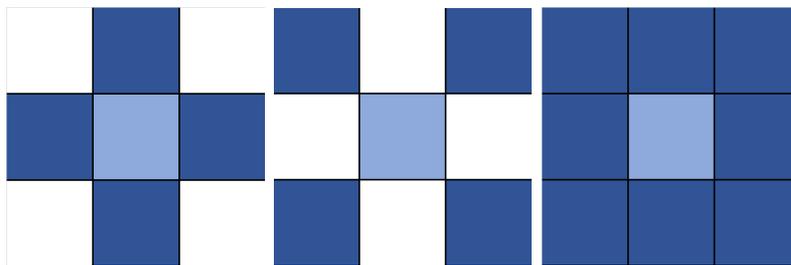


Figura 4.16: Tipos de vecindad. De izquierda a derecha: 4-vecindad, 4-vecindad diagonal y 8-vecindad.

Donde el método recorrido se puede ver en el algoritmo 5.

---

**Algoritmo 5:** Recorrido.

---

**Datos:**  $(i, j)$

**inicio**

    | Marcar  $(i, j)$

    | **para** *Todos los píxeles vecinos PV* **hacer**

        | **si** *PV no está marcado* **entonces**

          |   | Recorrido(*PV*)

    |

---

Además puede observarse que el método 5 es recursivo, para cada píxel encontrado se marca y se buscan sus vecinos.

En la figura 4.17 se muestra el conteo de fibras después de la aplicación de la combinación de filtros elegida por el usuario comparado con el resultado del conteo justo después de la segmentación.

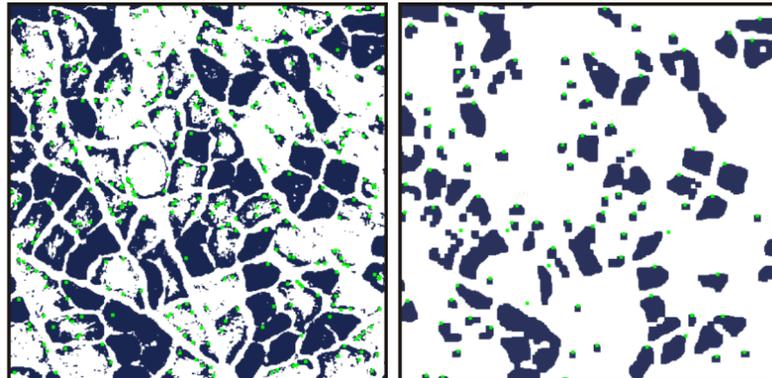


Figura 4.17: Comparación de conteos.

Este algoritmo tiene el inconveniente que solo guarda el primer píxel encontrado de cada componente conexa, sin embargo uno de los objetivos de la aplicación es encontrar los centroides de las fibras musculares, para que los análisis posteriores que se realicen con los datos obtenidos de este sistema sean confiables. Se propone la siguiente modificación para el calculo de centroides así como del tamaño de cada

fibra.

---

**Algoritmo 6:** Etiquetado de componentes conexas modificado.

---

**Datos:** ImagenBinaria

**Resultado:** Lista de coordenadas  $LC$  de centroides de cada objeto y lista de tamaños  $LT$  con índices correspondientes

**inicio**

    Crear lista  $Temporal$

**para** Todos los píxeles  $p(i, j)$  con  $i \in \{1, 2, \dots, ancho\}$  y

$j \in \{1, 2, \dots, alto\}$  **hacer**

        Vaciar  $Temporal$

**si**  $p(i, j)$  no está marcado **entonces**

            7 RecorridoModificado( $i, j$ )

$Centroide \leftarrow Promedio(Temporal)$

$Tamaño \leftarrow Tamaño(Temporal)$  Guardar  $Centroide$  en  $LC$

            Guardar  $Tamaño$  en  $LT$

---

Siendo el algoritmo RecorridoModificado el siguiente:

---

**Algoritmo 7:** RecorridoModificado.

---

**Datos:**  $(i, j)$

**inicio**

    Marcar  $(i, j)$

    Guardar  $(i, j)$  en  $Temporal$

**para** Todos los píxeles vecinos  $PV$  **hacer**

**si**  $PV$  no está marcado **entonces**

            RecorridoModificado( $PV$ )

---

Lo obtenido del conteo de fibras después de la aplicación de la combinación de filtros elegida por el usuario, comparada con el conteo con el algoritmo modificado se muestra en la figura 4.18.

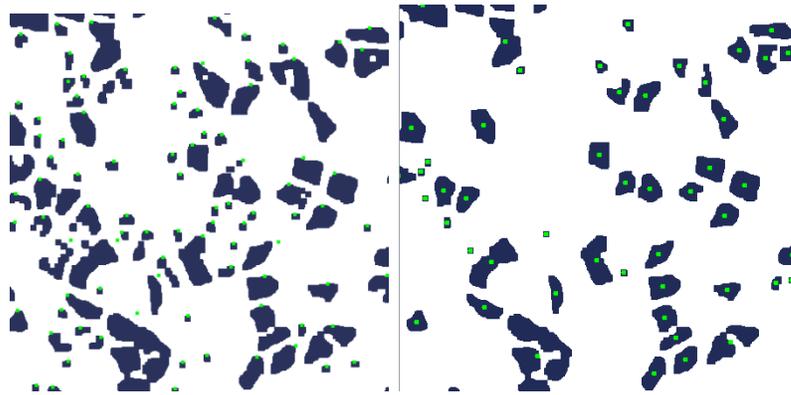


Figura 4.18: Comparación de resultados del conteo con un filtro más fino y el algoritmo modificado.

### 4.3.5. Edición de resultados

En el proceso de desarrollo del presente trabajo se encontraron casos en los que la metodología anterior no es suficiente para la correcta clasificación de fibras musculares, teniendo como razón principal la calidad de las imágenes proporcionadas. Por ejemplo, en la figura 4.19 se observa la presencia de regiones claras dentro de fibras de tono intermedio y oscuro, lo cual provoca que, al aplicar filtros de erosión, las fibras musculares se separen o incluso sean descartadas. Por otra parte algunas fibras no están bien delimitadas y es necesaria la inspección de un experto.

Por lo anterior, se diseñó una ventana (figura 4.20) en la que se muestran la localización de las fibras identificadas por el presente software, sobrepuestas sobre la imagen original, además en formato de lista, lo que permite una evaluación inmediata de los datos obtenidos antes de su almacenamiento. Permite cambiar la imagen sobre la que se muestran las coordenadas obtenidas, como se muestra en la figura 4.20, entre la imagen original, con la posibilidad de cambiar la ampliación (*zoom*), así como la imagen separada y la imagen filtrada.

También tiene opciones para borrar coordenadas que el software agrega por los casos anteriormente descritos, o agregar nuevas coordenadas ya sea con o sin área, como se muestra en la Sección 7.1.

Finalmente en el menú *Archivo* ofrece opciones para almacenamiento tanto en formato de hoja de cálculo de las coordenadas, así como la imagen de fondo con las coordenadas resultantes sobrepuestas.

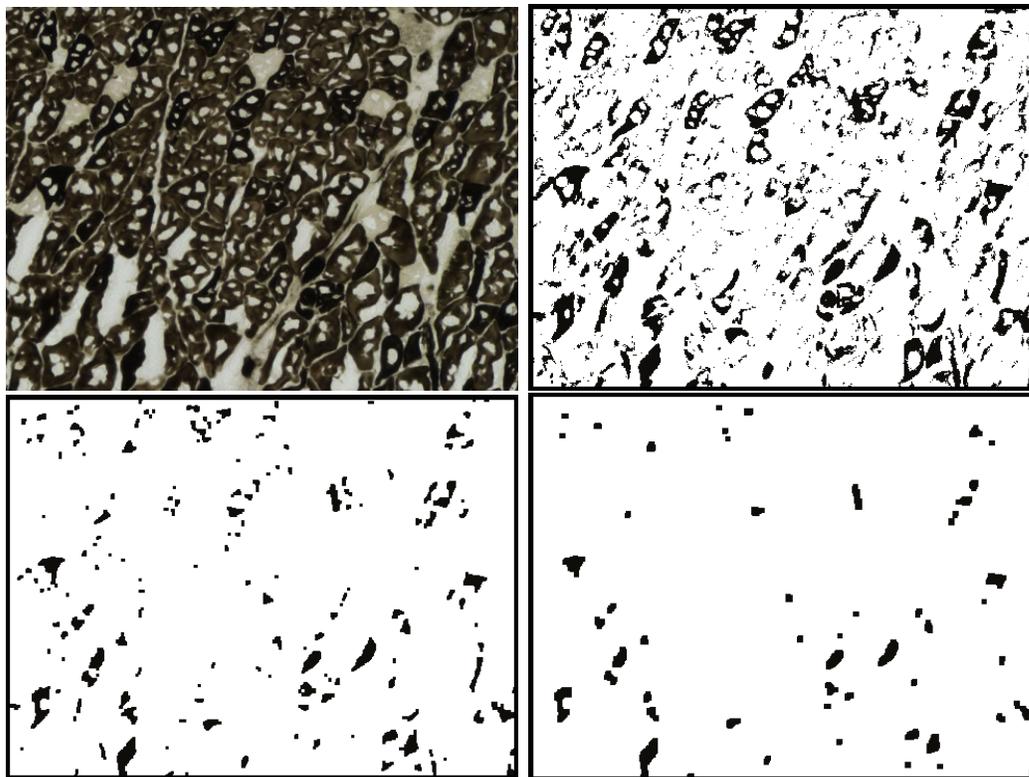


Figura 4.19: Región de la imagen C6\_Atp; su segmentación de fibras Intermedias; la aplicación de filtro de dos erosiones y una dilatación y aplicación de filtro de tres erosiones y dos dilataciones.

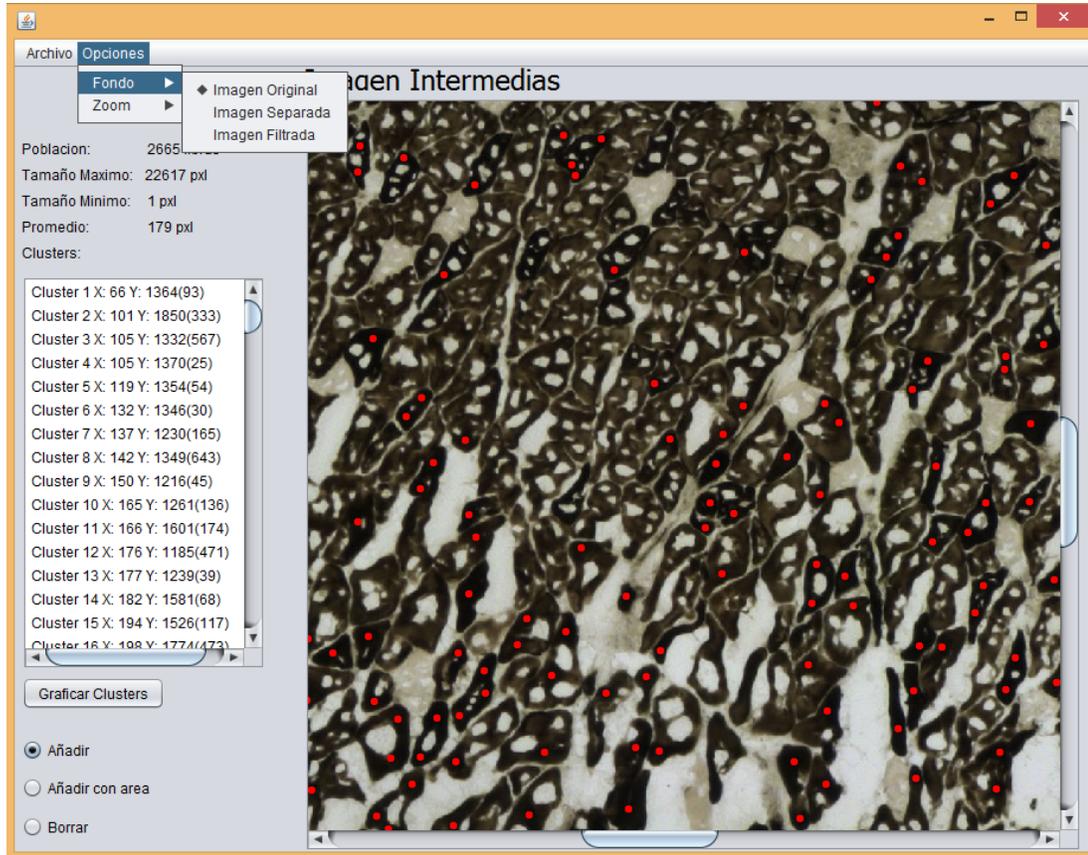


Figura 4.20: Interfaz gráfica para mostrar y editar conteos.

## 4.4. Funciones de distribución

En esta sección se describe una de las metodologías usadas en la sección de resultados.

El problema exacto abordado es ajustar una función de distribución a un conjunto finito de puntos en el plano. Por función de distribución se entenderá una suavización de un histograma de datos centrados, construida sobre una malla de  $N \times N$ . Siendo  $N$  un parámetro a determinar por el usuario. Las funciones de distribución son una superposición lineal de funciones Gaussianas de núcleo, básicamente una función de núcleo para cada uno de los cuadros de la malla. Las funciones Gaussianas de núcleo son un tipo de función de base radial, las cuales son utilizadas en problemas de aproximación e interpolación. Se muestra que las funciones de distribución pertenecen a un espacio con producto interno, el cual se puede completar a un *Espacio de Hilbert Reproductor de Núcleos* bajo la norma inducida por el producto interno. Trabajar en un espacio con producto interno permite medir distancias entre distribuciones y, en el fondo, entre los conjuntos de datos asociados con esas funciones.

Esto es útil para la comparación de los resultados del software presentado con los resultados proporcionados por los expertos, y por lo tanto la evaluación de lo presentado en este trabajo. Esta metodología fue diseñada por Viveros-Rogel y enviada a una revista para su publicación y se encuentra en proceso.

### 4.4.1. Función de distribución para una sola imagen

Sea  $\{\tilde{z}_1, \dots, \tilde{z}_\nu\} \subset \mathbb{R}^2$ ,  $\tilde{z}_j = (\tilde{x}_j, \tilde{y}_j)$  una colección finita de pares de coordenadas de puntos dentro de una imagen, cada una de las cuales representa una fibra muscular de la misma categoría que todas las demás en la colección. Sea  $\bar{z} = \nu^{-1} \sum_{j=1}^{\nu} \tilde{z}_j$  el centroide de la colección y  $S = \{z_1, \dots, z_\nu\}$ , con  $z_j = \tilde{z}_j - \bar{z} = (x_j, y_j)$ , la colección de pares de coordenadas centrados.  $S$  es nuestro conjunto de datos.  $S$  está contenido en una caja compacta  $\mathcal{B} = [a, A] \times [b, B] \subset \mathbb{R}^2$ , donde  $a = -\varepsilon + \text{mín } x_j$  y  $A = \varepsilon + \text{máx } x_j$ , con  $\varepsilon$  un número pequeño, positivo elegido ( $b$  y  $B$  se definen similarmente).  $\mathcal{B}$  también podría ser determinada por las dimensiones de la imagen, y así evitar el “corte” descrito.

Escojamos un entero positivo fijo  $N$  y consideremos particiones uniformes  $P = \{a_0, \dots, a_N\}$  y  $Q = \{b_0, \dots, b_N\}$  de  $[a, A]$  y  $[b, B]$ , respectivamente. Por tanto  $a_k = a + k(A - a)/N$  ( $b_k$  definido similarmente).  $P$  y  $Q$  definen una malla rectangular sobre

$\mathcal{B}$ , compuesta de  $N^2$  celdas,  $C_{ij} = [a_{i-1}, a_i] \times [b_{j-1}, b_j]$ . Más adelante se describe un método para montar una función Gaussiana bivariada sobre cada celda  $C_{ij}$ . Tales funciones después son superpuestas linealmente para formar una versión suave del histograma definido sobre la malla.

Consideremos una celda  $C_{ij}$  y sea  $S_{ij}$  la sub-colección de puntos que esta contiene,  $S_{ij} = \{\mathbf{z}_k \in C_{ij}\} \subseteq S$ . Sea  $|S_{ij}|$  la cardinalidad de  $S_{ij}$  y  $\bar{\mathbf{z}}_{ij} = |S_{ij}|^{-1} \sum_{S_{ij}} \mathbf{z}_k$  el centroide de  $S_{ij}$ .

Considere la matriz de covarianza muestral de puntos en  $C_{ij}$  (véase [9]),

$$\Sigma_{ij} = (|S_{ij}| - 1)^{-1} \sum_{S_{ij}} (\mathbf{z}_k - \bar{\mathbf{z}}_{ij})(\mathbf{z}_k - \bar{\mathbf{z}}_{ij})^t. \quad (4.3)$$

Asumamos por el momento que  $\Sigma_{ij}$  es invertible (el caso cuando  $\Sigma_{ij}$  es singular es tratado de otra manera) y sea

$$\phi_{ij}(\mathbf{z}) = \exp\left(-2^{-1}(\mathbf{z} - \bar{\mathbf{z}}_{ij})^t \Sigma_{ij}^{-1} (\mathbf{z} - \bar{\mathbf{z}}_{ij})\right), \quad (4.4)$$

esta es la función Gaussiana bivariada referida arriba, también llamada *función Gaussiana de núcleo* (ver más adelante). Luego asociamos a la colección  $S$  la siguiente *función de distribución*, donde  $R_{ij} = |S_{ij}|$  (la cardinalidad de  $S_{ij}$ ):

$$\Phi(\mathbf{z}) = \sum_{i=1}^N \sum_{j=1}^N R_{ij} \phi_{ij}(\mathbf{z}). \quad (4.5)$$

$\Phi$  es el principal objeto de estudio. Ahora se mencionan tres situaciones que deben ser consideradas cuando se construye  $\Phi$ .

- (i) Si  $C_{ij}$  es vacío, se asigna  $\phi_{ij} = 0$ .
- (ii) Si  $C_{ij}$  contiene sólo un punto  $\mathbf{z}_p = (x_p, y_p)$ , calculamos su distancia  $\delta = \max\{a_j - x_p, b_j - x_p\}$  a la frontera de  $C_{ij}$ , redefinimos  $\Sigma_{ij} = (\delta/3)^2 Id_2$  donde  $Id_2$  es la matriz identidad  $2 \times 2$ , y construimos  $\phi_{ij}$  como en la fórmula (4.4), esto es  $\phi_{ij}(\mathbf{z}) = \exp(-9 \cdot 2^{-1} \delta^{-2} \|\mathbf{z} - \mathbf{z}_p\|^2)$ . El caso improbable en el que  $C_{ij}$  contiene un solo punto el cual se encuentra exactamente sobre su frontera, puede ser tratado en varias maneras. Una manera es asociar este punto a la celda adyacente de la frontera compartida, asumiendo que esa celda tiene por lo menos otro punto. Otra manera es asignar a  $\delta$  arbitrariamente un valor positivo muy pequeño predefinido, de modo que la contribución de  $\phi_{ij}$  a la suma  $\Phi$  es altamente localizada.

(iii) Si  $\Sigma_{ij}$  es singular o cercana a serlo (su determinante es menor que un número preestablecido  $\epsilon_0 > 0$ ), aplicamos los siguientes cuatro pasos:

(a) sea  $\lambda_1 = \max_{S_{ij}} \|\mathbf{z}_k - \bar{\mathbf{z}}_{ij}\|^2$ , y  $\lambda_2 = \lambda_1/9$  con el propósito de hacer el eje menor de la función de base radial igual a la tercera parte del eje mayor (aquí  $\|\cdot\|$  denota distancia euclidiana).

(b) escoger cualquier  $\mathbf{z}_k$  en  $C_{ij}$  y definir  $\mathbf{u}_1 = (\mathbf{z}_k - \bar{\mathbf{z}}_{ij})\|\mathbf{z}_k - \bar{\mathbf{z}}_{ij}\|^{-1}$ , después si  $\mathbf{u}_1 = (v, w)$  sea  $\mathbf{u}_2 = (-w, v)$ , por lo tanto  $\mathbf{u}_1^t \mathbf{u}_2 = 0$ .

(c) Sea  $M = [\mathbf{u}_1 \ \mathbf{u}_2]$ , una matriz  $2 \times 2$  con columnas  $\mathbf{u}_1$  y  $\mathbf{u}_2$ , en ese orden. Sea  $\Lambda = \text{diag}(\lambda_1, \lambda_2)$  una matriz diagonal.

(d) Redefinimos  $\Sigma_{ij} = M\Lambda M^t$  y construimos  $\phi_{ij}$  como en la fórmula (4.4).

A continuación se establece que la función  $\Phi$  en la fórmula (4.5) pertenece a un espacio con producto interno, así en consecuencia permitiendo el uso de herramientas para el cuantificador de similaridad entre dos funciones dadas.

Se ha estructurado esta sección en una serie de observaciones con la intención de ganar gradualmente entendimiento dentro de las funciones de distribución definidas como en la fórmula (4.5).

**1.** Si una matriz  $\Sigma$  es positiva definida (simétrica y  $\xi^t \Sigma \xi \geq 0$ , con igualdad si y sólo si  $\xi$  es el vector cero), su inversa  $\Sigma^{-1}$  existe y es positiva definida también. Las matrices de covarianza  $\Sigma_{ij}$  usadas en este trabajo son siempre positivas definidas (fórmula (4.3) y puntos (ii) y (iii) de arriba), esto garantiza que todas las funciones  $\phi_{ij}$  tienen gráficas con forma de campana.

**2. Lema 1:** si  $\Sigma$  es una matriz positiva definida, la forma cuadrática  $d : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}$ , definida por

$$d^2(\bar{\mathbf{z}}, \mathbf{z}) = (\mathbf{z} - \bar{\mathbf{z}})^t \Sigma^{-1} (\mathbf{z} - \bar{\mathbf{z}}), \quad (4.6)$$

es una métrica en  $\mathbb{R}^2$ .

**3. Lema 2.:** si  $d$  es una métrica, entonces

$$\kappa(\bar{\mathbf{z}}, \mathbf{z}) = \exp(-2^{-1} d^2(\bar{\mathbf{z}}, \mathbf{z})), \quad (4.7)$$

es una *función de núcleo* (ver [29]). Cuando  $d$  es una distancia euclidiana  $\kappa$  es básicamente una *función Gaussiana de base radial*, (véase [28]), es decir,

(i)  $\kappa$  es simétrica:  $\kappa(\bar{\mathbf{z}}, \mathbf{z}) = \kappa(\mathbf{z}, \bar{\mathbf{z}})$ , y

(ii) para cualquier colección de vectores  $\{\mathbf{u}_1, \dots, \mathbf{u}_N\}$  en  $\mathbb{R}^2$ ,  $\mathbf{b}^t \mathbb{K} \mathbf{b} \geq 0$  para todo  $\mathbf{b}$  en  $\mathbb{R}^N$ , donde  $\mathbb{K}$  es una matriz con componentes  $\mathbb{K}_{ij} = \kappa(\mathbf{u}_i, \mathbf{u}_j)$ .

**4:** Sea  $\kappa_{ij} : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}$  definida de la siguiente manera:

$$\kappa_{ij}(\mathbf{y}, \mathbf{z}) = \exp \left( -2^{-1}(\mathbf{z} - \mathbf{y})^t \Sigma_{ij}^{-1} (\mathbf{z} - \mathbf{y}) \right), \quad (4.8)$$

donde  $\Sigma_{ij}$  es como antes la matriz de covarianza muestral de la celda  $C_{ij}$  y  $\mathbf{z}$  y  $\mathbf{y}$  son cualquier vectores en  $\mathbb{R}^2$ . Ahora definimos el espacio vectorial real

$$\mathcal{L}_{ij} = \text{span}\{\kappa_{ij}(\mathbf{y}, \cdot) : \mathbf{y} \in \mathbb{R}^2\}. \quad (4.9)$$

**Lema 3:**  $\mathcal{L}_{ij}$  es un espacio con producto interno. Para  $f$  y  $g$  en  $\mathcal{L}_{ij}$  dado por

$$f(\mathbf{z}) = \sum_{k=1}^m D_k \kappa_{ij}(\mathbf{u}_k, \mathbf{z}) \quad \text{y} \quad g(\mathbf{z}) = \sum_{\ell=1}^n E_\ell \kappa_{ij}(\mathbf{v}_\ell, \mathbf{z}), \quad (4.10)$$

su producto interno  $(f, g)_{ij}$  es definido por

$$(f, g)_{ij} = \sum_{k=1}^m \sum_{\ell=1}^n D_k E_\ell \kappa_{ij}(\mathbf{u}_k, \mathbf{v}_\ell). \quad (4.11)$$

**5.** Siendo un espacio con producto interno,  $\mathcal{L}_{ij}$  es un espacio normado y por lo tanto un espacio métrico: para cada  $f$  en  $\mathcal{L}_{ij}$  como arriba, podemos definir su norma  $\|f\|_{ij}$ , como sigue,

$$\|f\|_{ij}^2 = (f, f)_{ij} = \sum_{k=1}^m \sum_{\ell=1}^m D_k D_\ell \kappa_{ij}(\mathbf{u}_k, \mathbf{u}_\ell). \quad (4.12)$$

y si  $g$  en  $\mathcal{L}_{ij}$  es como arriba podemos definir la distancia entre  $f$  y  $g$ ,  $d(f, g)$ , como

$$d_{ij}^2(f, g) = (f - g, f - g) = \|f\|_{ij}^2 + \|g\|_{ij}^2 - 2(f, g)_{ij}. \quad (4.13)$$

Note que en particular  $\phi_{ij}(\mathbf{z}) = \kappa_{ij}(\bar{\mathbf{z}}_{ij}, \mathbf{z})$  y por lo tanto es un elemento de  $\mathcal{L}_{ij}$ , más aún

$$\|\phi_{ij}\|_{ij}^2 = \kappa_{ij}(\bar{\mathbf{z}}_{ij}, \bar{\mathbf{z}}_{ij}) = 1. \quad (4.14)$$

**6.** Por la desigualdad de Cauchy-Schwarz, el coseno del ángulo  $\theta$  entre dos funciones puede ser definido como:

$$\cos \theta = \frac{(f, g)_{ij}}{\|f\|_{ij} \|g\|_{ij}}; \quad (4.15)$$

por lo tanto también se puede establecer el concepto de perpendicularidad entre funciones.

7. Ahora consideremos  $\mathcal{L}_N = \bigoplus\{\mathcal{L}_{ij} : i, j = 1, \dots, N\}$ , el espacio vectorial hecho por las combinaciones lineales de funciones de los espacios  $\mathcal{L}_{ij}$ . Ahora se muestra que  $\mathcal{L}_N$  es un espacio con producto interno. Sean  $F$  y  $G$  elementos en  $\mathcal{L}_N$

$$F(\mathbf{z}) = \sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^{m_{ij}} D_k^{ij} \kappa_{ij}(\mathbf{u}_k^{ij}, \mathbf{z}) \quad \text{y} \quad G(\mathbf{z}) = \sum_{i=1}^N \sum_{j=1}^N \sum_{\ell=1}^{n_{ij}} E_\ell^{ij} \kappa_{ij}(\mathbf{v}_\ell^{ij}, \mathbf{z}), \quad (4.16)$$

donde  $\{\mathbf{u}_k^{ij} : k = 1, \dots, m_{ij}; i, j = 1, \dots, N\}$  y  $\{\mathbf{v}_\ell^{ij} : \ell = 1, \dots, n_{ij}; i, j = 1, \dots, N\}$  son colecciones arbitrarias de vectores en  $\mathbb{R}^2$ . Definimos su producto interno como:

$$\begin{aligned} (F, G)_N &= \sum_{i=1}^N \sum_{j=1}^N \left( \sum_{k=1}^{m_{ij}} D_k^{ij} \kappa_{ij}(\mathbf{u}_k^{ij}, \mathbf{z}), \sum_{\ell=1}^{n_{ij}} E_\ell^{ij} \kappa_{ij}(\mathbf{v}_\ell^{ij}, \mathbf{z}) \right)_{ij} = \\ &= \sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^{m_{ij}} \sum_{\ell=1}^{n_{ij}} E_\ell^{ij} D_k^{ij} \kappa_{ij}(\mathbf{u}_k^{ij}, \mathbf{v}_\ell^{ij}). \end{aligned} \quad (4.17)$$

La simetría, linealidad y no degeneración de este producto interno son consecuencia directa del hecho que esas propiedades ya fueron establecidas para los productos internos individuales  $(\cdot, \cdot)_{ij}$ . También note que, por esta construcción, los espacios factor  $\mathcal{L}_{ij}$  que componen  $\mathcal{L}_N$  son subespacios ortogonales.

8. Similarmente a  $\mathcal{L}_{ij}$ , siendo un espacio con producto interno  $\mathcal{L}_N$  es un espacio normado y también un espacio métrico. En efecto, si  $F$  y  $G$  están definidos como en la fórmula (4.16) entonces

$$\|F\|_N^2 = (F, F)_N = \sum_{i=1}^N \sum_{j=1}^N \left\| \sum_{k=1}^{m_{ij}} D_k^{ij} \kappa_{ij}(\mathbf{u}_k^{ij}, \mathbf{z}) \right\|_{ij}^2 = \sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^{m_{ij}} \sum_{\ell=1}^{m_{ij}} D_k^{ij} D_\ell^{ij} \kappa_{ij}(\mathbf{u}_k^{ij}, \mathbf{u}_\ell^{ij}), \quad (4.18)$$

y

$$\begin{aligned} d_N^2(F, G) &= (F - G, F - G)_N = \sum_{i=1}^N \sum_{j=1}^N d_{ij}^2 \left( \sum_{k=1}^{m_{ij}} D_k^{ij} \kappa_{ij}(\mathbf{u}_k^{ij}, \mathbf{z}), \sum_{\ell=1}^{n_{ij}} E_\ell^{ij} \kappa_{ij}(\mathbf{v}_\ell^{ij}, \mathbf{z}) \right) \\ &= \|F\|_N^2 + \|G\|_N^2 - 2(F, G)_N. \end{aligned} \quad (4.19)$$

También, similarmente como en la fórmula (4.15), el coseno del ángulo  $\theta$  entre  $F$  y  $G$  es definido como

$$\cos \theta = \frac{(F, G)_N}{\|F\|_N \|G\|_N}. \quad (4.20)$$

9. La función de distribución  $\Phi$  definida en la fórmula (4.5), pertenece a  $\mathcal{L}_N$ .  $\Phi$  sólo toma un elemento  $\mathbf{u}^{ij} = \bar{\mathbf{z}}_{ij}$  por celda  $C_{ij}$  (i.e, por cada par  $(i, j)$ ), y dado que las funciones  $\phi_{ij}$  son ortogonales a pares y unitarias (véase la fórmula (4.14)), se tiene que (recuerde que  $R_{ij} = |S_{ij}|$ )

$$\|\Phi\|_N^2 = \sum_{i=1}^N \sum_{j=1}^N R_{ij}^2. \quad (4.21)$$

De hecho, para nuestros propósitos, sea en adelante  $\Psi(\mathbf{z}) = \sum_{i=1}^N \sum_{j=1}^N T_{ij} \psi_{ij}(\mathbf{z})$ , donde los coeficientes  $T_{ij}$  son números reales.  $\psi_{ij}(\mathbf{z}) = \kappa_{ij}(\bar{\mathbf{z}}'_{ij}, \mathbf{z})$  y  $\bar{\mathbf{z}}'_{ij}$  son vectores en  $\mathbb{R}^2$ .  $\Psi$  es por lo tanto una función como  $\Phi$  (sólo toma un vector  $\mathbf{u}_{ij}$  por cada par  $(i, j)$ ). En este caso particular

$$d^2(\Phi, \Psi) = \|\Phi\|_N^2 + \|\Psi\|_N^2 - 2(\Phi, \Psi)_N = \sum_{i=1}^N \sum_{j=1}^N (R_{ij}^2 + T_{ij}^2 - 2R_{ij}T_{ij}\kappa_{ij}(\bar{\mathbf{z}}_{ij}, \bar{\mathbf{z}}'_{ij})) \quad (4.22)$$

y la fórmula (4.20) en este caso se convierte en

$$\cos \theta = \frac{\sum_{i=1}^N \sum_{j=1}^N R_{ij}T_{ij}\kappa_{ij}(\bar{\mathbf{z}}_{ij}, \bar{\mathbf{z}}'_{ij})}{\left(\sum_{i=1}^N \sum_{j=1}^N R_{ij}^2\right)^{1/2} \left(\sum_{i=1}^N \sum_{j=1}^N T_{ij}^2\right)^{1/2}} \quad (4.23)$$

#### 4.4.2. Funciones de distribución para un lote de imágenes

Una vez que una función de distribución ha sido obtenida para cada imagen, deseamos calcular la distancia o ángulo entre pares de estas y ver si esas medidas reflejan la clasificación independientemente establecida.

Se debe asegurar que las funciones de distribución de las imágenes a comparar *pertenecen al mismo espacio con producto interno*  $\mathcal{L}_N$ . Para alcanzar esto sólo necesitamos cambiar ligeramente nuestra definición de funciones  $\phi_{ij}$  en la fórmula (4.4). Más precisamente, necesitamos modificar la definición de la matriz de covarianza  $\Sigma_{ij}$  en la fórmula (4.3) como se describe a continuación.

Similarmente como en la sección anterior dividimos cada imagen en  $N \times N$  celdas. Sea  $C_{ij}^b$  la celda  $(i, j)$  de la imagen  $b$  y  $\mathbf{z}_k^b$  el  $k$ -ésimo punto en esa imagen. Sea  $S_{ij}^b = \{\mathbf{z}_k^b \in C_{ij}^b\}$  (el conjunto de puntos de la imagen  $b$  en su celda  $C_{ij}^b$ ), y sea  $|S_{ij}^b|$  el número de puntos en  $C_{ij}^b$ . Definimos la *matriz de covarianza global para la celda*

$(i, j)$  como:

$$\Sigma_{ij} = (\mathcal{S}_{ij} - 1)^{-1} \sum_{b=1}^r \sum_{S_{ij}^b} (\mathbf{z}_k^b - \bar{\mathbf{z}}_{ij})(\mathbf{z}_k^b - \bar{\mathbf{z}}_{ij})^t, \quad (4.24)$$

donde  $\mathcal{S}_{ij} = \sum_{b=1}^r |S_{ij}^b|$  y  $\bar{\mathbf{z}}_{ij}$  es el *centroide global para la celda*  $(i, j)$ ,

$$\bar{\mathbf{z}}_{ij} = \mathcal{S}_{ij}^{-1} \sum_{b=1}^r \sum_{S_{ij}^b} \mathbf{z}_k^b. \quad (4.25)$$

Por lo tanto si

$$\phi_{ij}^b(\mathbf{z}) = \exp(2^{-1}(\mathbf{z} - \bar{\mathbf{z}}_{ij}^b)^t \Sigma_{ij}^{-1} (\mathbf{z} - \bar{\mathbf{z}}_{ij}^b)), \quad (4.26)$$

entonces

$$\Phi^b(\mathbf{z}) = \sum_{i=1}^N \sum_{j=1}^N R_{ij}^b \phi_{ij}^b(\mathbf{z}), \quad (4.27)$$

es la función de distribución para la imagen muestra  $b$ , con  $R_{ij}^b = |S_{ij}^b|$ . Podemos adaptar los casos *(i)*, *(ii)* y *(iii)* para ocuparse de escenarios en los que  $\Sigma^{-1}$  es singular o cercana a serlo.

Este método fue implementado en el lenguaje de programación de Matlab, en un conjunto de *scripts* para su utilización desde la línea de comandos, con los cuales, además de las mediciones de distancias y ángulos entre distribuciones, hace posible la generación y almacenamiento de distintos tipos de gráficos como reconstrucción binaria, mapas de distribución con paletas de colores en tres dimensiones y conjuntos de nivel, las cuales son utilizadas en el Capítulo 5.



# Capítulo 5

## Resultados

### 5.1. Preliminares

Para el proceso de desarrollo del software presentado se proporcionó un conjunto de imágenes de tinciones histoquímicas por parte de investigadores, con el propósito de la validación del progreso obtenido. Entre éstas imágenes se encuentran imágenes con tinciones NADH, ATPasa y COX. A continuación se presentan los resultados obtenidos.

#### 5.1.1. Comparación de resolución y filtros

La imagen de nombre *C5\_NADH* se proporcionó con un conteo por parte de un experto de 3420 fibras de las cuales 1347 son fibras Oxidativas representando el 39 % y 2073 Glucolíticas que representan el restante 61 % de las fibras.

En el cuadro 5.1 se presentan los conteos de esta imagen realizados con el programa y los distintos filtros disponibles, utilizando el conocimiento adquirido por el algoritmo K-Means y una resolución de 5 Megapíxeles (Mpx), comparados con el conteo obtenido por un experto.

Se puede observar que, desde el filtro de erosión, hasta el filtro de 3 erosiones y 2 dilataciones (E3D2) la cantidad de fibras contadas excede el conteo del experto. Regularmente si un conteo es mayor al real es porque no se eliminó todo el ruido presente en la imagen, por lo que es necesaria la aplicación de filtros con más iteraciones.

5 Mpx	Oxi	Glu	Total	% Oxi	% Glu	Dif Oxi	Dif Glu	Dif Total
Erosión	5029	6232	<b>11261</b>	45 %	55 %	3682	4159	<b>7841</b>
Apertura	3681	4087	<b>7768</b>	47 %	53 %	2334	2014	<b>4348</b>
E2D1	2636	2531	<b>5167</b>	51 %	49 %	1289	458	<b>1747</b>
E3D2	1570	2058	<b>3628</b>	43 %	57 %	223	-15	<b>208</b>
E4D3	1016	1825	<b>2841</b>	36 %	64 %	-331	-248	<b>-579</b>
E5D4	544	1530	<b>2074</b>	26 %	74 %	-803	-543	<b>-1346</b>
E6D5	212	1159	<b>1371</b>	15 %	85 %	-1135	-914	<b>-2049</b>
E7D6	66	744	<b>810</b>	8 %	92 %	-1281	-1329	<b>-2610</b>
<b>Objetivo</b>	<b>1347</b>	<b>2073</b>		<b>39 %</b>	<b>61 %</b>			

Cuadro 5.1: Tabla comparativa C5\_NADH a 5 Mpx, los valores objetivo son los obtenidos por un experto. Glu representa Glucolíticas y así como Oxi se refiere a Oxidativas. Dif significa Diferencia.

Sin embargo, no solo se debe tomar en cuenta la cantidad total en el conteo de fibras, pues existen casos en los que las proporciones de las fibras no sean las adecuadas para una imagen dada, es decir los porcentajes de cada fibra contada mediante el software difieran de los que devuelve el conteo por un experto.

Por lo anterior es importante poner atención a los porcentajes de las fibras en nuestros resultados y compararlos con los reales, como lo muestran la columnas de porcentaje. Podemos considerar como consistentes en este aspecto los resultados de los filtros E3D2 y E4D3, pues sólo difieren en 4% y 3% respectivamente, además de que la diferencia entre fibras que sobran o que faltan no son muy grandes en cada caso. La decisión de qué conteo usar la debe de tomar el experto de acuerdo a su visualización de resultados sobre la imagen original, en la etapa de edición, y borrar los sobrantes o agregar los faltantes, según preferencia.

8 Mpx	Oxi	Glu	Total	% Oxi	% Glu	Dif Oxi	Dif Glu	Dif Total
Erosión	7508	12636	<b>20144</b>	37 %	63 %	6161	10563	<b>16724</b>
Apertura	5524	7898	<b>13422</b>	41 %	59 %	4177	5825	<b>10002</b>
E2D1	3946	3917	<b>7863</b>	50 %	50 %	2599	1844	<b>4443</b>
E3D2	2559	2134	<b>4693</b>	55 %	45 %	1212	61	<b>1273</b>
E4D3	1721	1644	<b>3365</b>	51 %	49 %	374	-429	<b>-55</b>
E5D4	1233	1408	<b>2641</b>	47 %	53 %	-114	-665	<b>-779</b>
E6D5	830	1224	<b>2054</b>	40 %	60 %	-517	-849	<b>-1366</b>
E7D6	487	1066	<b>1553</b>	31 %	69 %	-860	-1007	<b>-1867</b>
<b>Objetivo</b>	<b>1347</b>	<b>2073</b>		<b>39 %</b>	<b>61 %</b>			

Cuadro 5.2: Tabla comparativa C5\_NADH a 8 Mpx, mismos símbolos que 5.1.

También se muestran los resultados análogos a la tabla anterior pero con la imagen *C5\_NADH* reducida a 8 Mpx en el cuadro 5.2. Se puede notar la dependencia que existe de los algoritmos morfológicos a la resolución de la imagen procesada. En este caso los filtros más apropiados resultan ser E4D3 y E5D4 para las fibras oxidativas y para las fibras glucolíticas E3D3 y E4D3, sin embargo el software presentado está diseñado para que la elección de un filtro sea independiente de los otros, por lo que es posible la edición de resultados con distinto filtro, lo que permite la reducción del tiempo de edición, escogiendo los resultados más acordes y editar lo mínimo posible.

En la figura 5.1 se presenta el procesamiento de la imagen *C4\_ATPasa*. Aunque los algoritmos de minería de datos fueron suficientes para la separación de la imagen *C5\_NADH*, al tratar de utilizarlos para *C4\_ATPasa*, éstos no son capaces de realizar una separación razonable de acuerdo a la figura 4.1 pues como se puede observar en la figura 5.1 B) y C) no se logran distinguir las fibras rápidas de las intermedias, por lo que el conteo resulta erróneo.

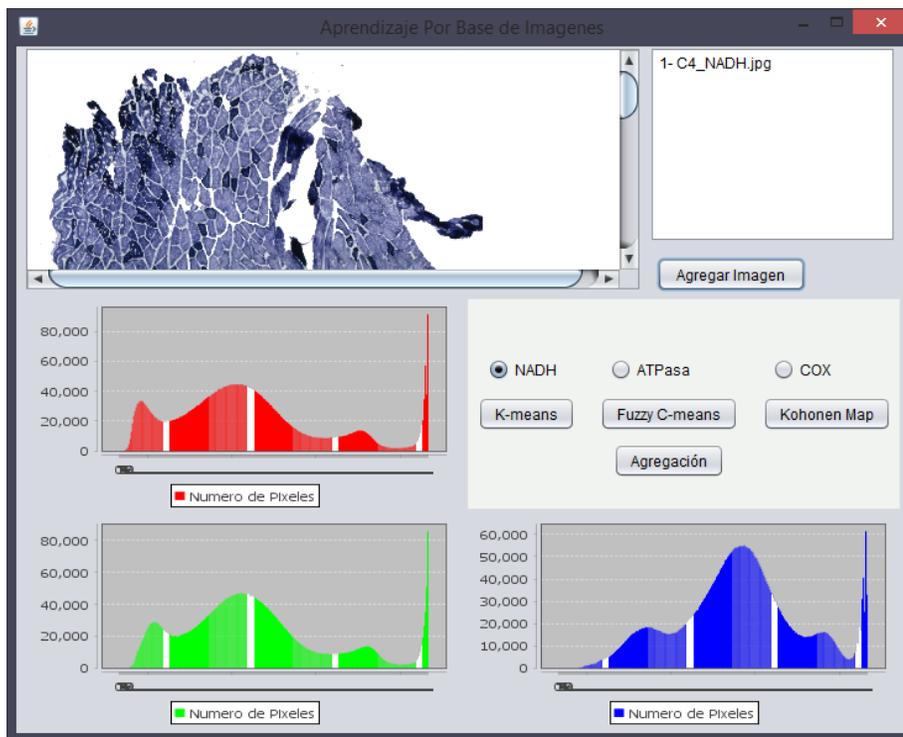


Figura 5.2: Ventana *Analyzer* al agregar la imagen *C4\_NADH*.

Analizando la figura 5.2, donde se muestran los histogramas de color obtenidos

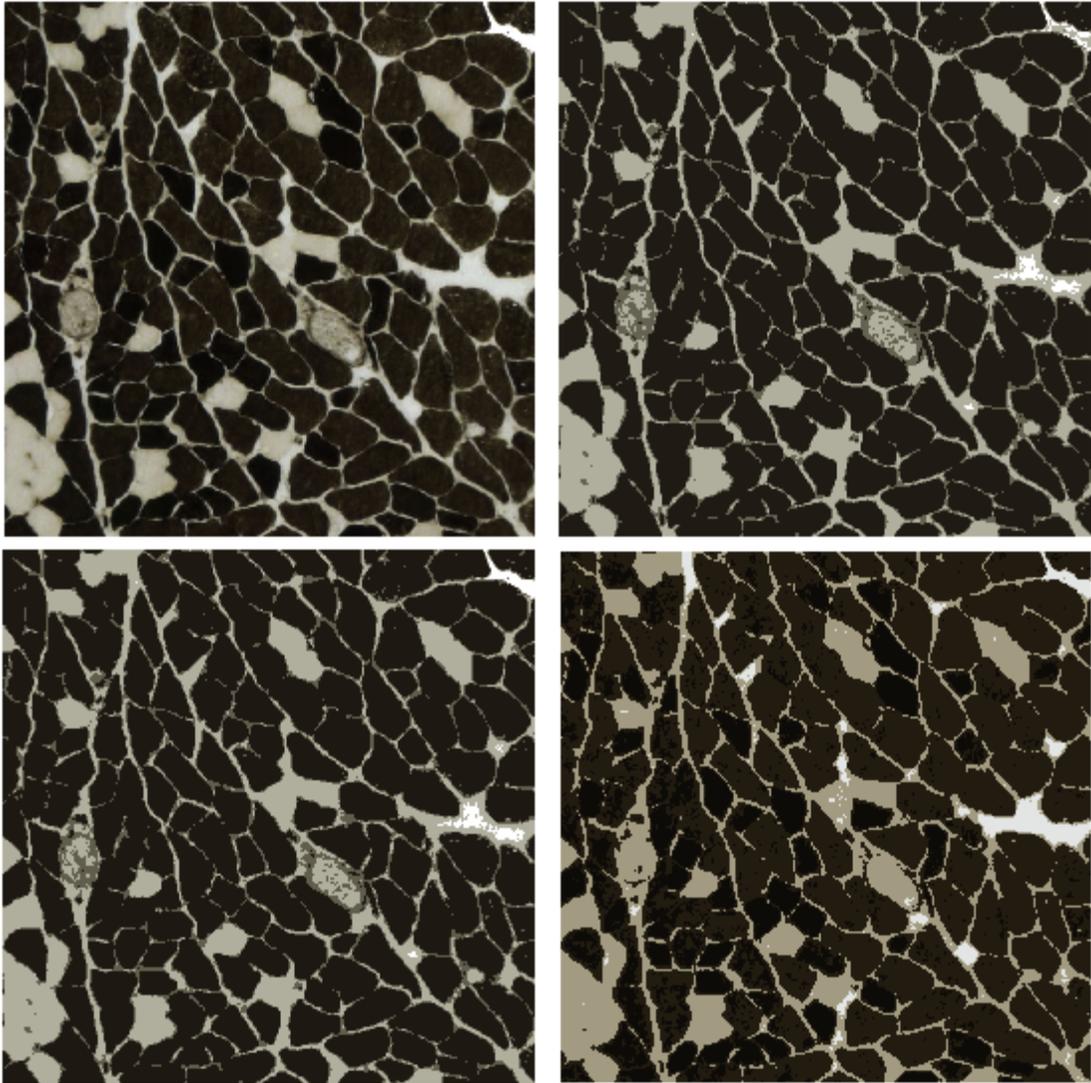


Figura 5.1: Comparación de resultados de la separación de la imagen C4 ATP, A) Original B) K-means C) Fuzzy C-means D) Experto.

de la imagen *C4\_NADH*, es posible identificar, en cada uno de estos, los valores de color que se obtienen con mayor frecuencia, es decir los valores correspondientes a cada tipo de fibra, los cuales sobresalen como crestas. En cada histograma se observan 4 crestas de frecuencia, sin embargo las últimas corresponden a valores muy cercanos, y como los algoritmos buscan únicamente 3 grupos (recordar  $k = 3$  para imágenes NADH), estos valores son asignados al mismo grupo.

Sin embargo con la figura 5.3, de los histogramas correspondientes a la imagen *C4\_ATPasa*, no es posible una identificación visual de grupos análoga, pues se identifica una gran cresta en valores alrededor de la primera cuarta parte del histograma (valores menores que 64) y únicamente otras dos pequeñas crestas en valores mayores. En estas condiciones los algoritmos no obtienen valores satisfactorios desde el punto de vista de la histología, pues estadísticamente los grupos de fibras oscuras y de tono medio no se pueden diferenciar, como se muestra en la figura 5.4.

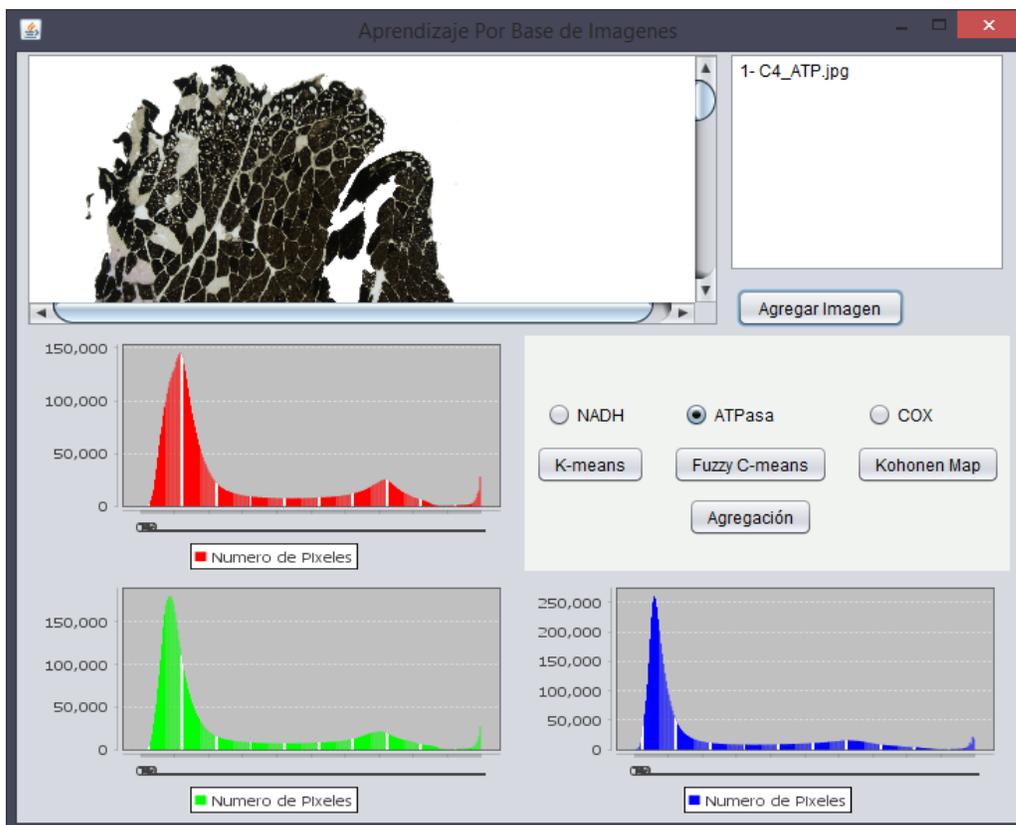


Figura 5.3: Ventana *Analyzer* al agregar la imagen *C4\_ATPasa*.

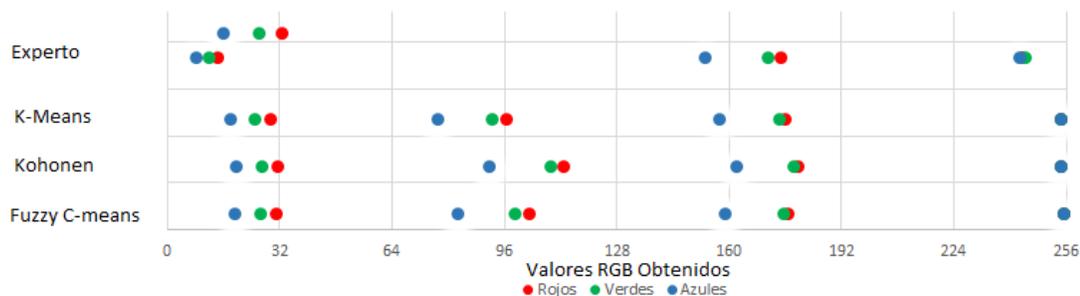


Figura 5.4: Se muestran los valores obtenidos (en el eje horizontal que va de 0 a 255) por cada uno de los algoritmos disponibles en la ventana *Analizer* para la imagen *C4\_ATPasa*, indicados por color. En los algoritmos (en los cuales los resultados son muy parecidos) es posible identificar los 4 grupos, sin embargo los valores del segundo grupo no concuerdan con el obtenido por el experto.

Es en estos casos en los que se debe recurrir a la separación guiada por experto descrita en la Sección 4.3.1. En la figura 5.1 D) se muestran los resultados de la separación con el método recomendado. A continuación se muestra la comparativa de resolución y filtros morfológicos con respecto a los conteos proporcionados por expertos que son de 1780 fibras de las cuales 765 son fibras Intermedias representando el 43% y 783 Rápidas que representan 44% de las fibras y por último 232 fibras Lentas siendo el 13% del total.

En el cuadro 5.3 se muestran los resultados del procesamiento de la imagen *C4\_ATPasa* con una resolución de 5 Mpx, se observa que para las fibras intermedias, los filtros que devuelven conteos más cercanos a los de referencia son E3D2 y E4D3, en cambio para las fibras rápidas y lentas los más apropiados son E5D4 y E5D5.

5 MPX	Inter	Rap	Len	<b>Total</b>	% I	% R	% L	Dif I	Dif R	Dif L	<b>Total</b>
Erosión	3653	6049	3795	<b>13497</b>	27%	45%	28%	2888	5266	3563	<b>11717</b>
E2D1	1562	2944	1321	<b>5827</b>	27%	51%	23%	797	2161	1089	<b>4047</b>
E3D2	938	1703	586	<b>3227</b>	29%	53%	18%	173	920	354	<b>1447</b>
E4D3	680	1125	346	<b>2151</b>	32%	52%	16%	-85	342	114	<b>371</b>
E5D4	508	863	262	<b>1633</b>	31%	53%	16%	-257	80	30	<b>-147</b>
E6D5	371	682	209	<b>1262</b>	29%	54%	17%	-394	-101	-23	<b>-518</b>
E7D6	245	575	181	<b>1001</b>	24%	57%	18%	-520	-208	-51	<b>-779</b>
<b>Objetivo</b>	<b>765</b>	<b>783</b>	<b>232</b>		<b>43%</b>	<b>44%</b>	<b>13%</b>				

Cuadro 5.3: Conteos C4\_ATP a 5 Mpx.

También se muestra el procesamiento a 8 Mpx en el cuadro 5.4, el cual muestra que los filtro más apropiados para las fibras intermedias en este caso son E4D3 y E5D4 y que para las fibras de tipo rápido y lento el filtro de con más iteraciones E7D6 aún devuelve conteos superiores a los del experto.

8 MPX	Inter	Rap	Len	Total	% I	%	% L	Dif I	Dif R	Dif L	Total
Erosión	4971	6860	5007	<b>16838</b>	30 %	41 %	30 %	4206	6077	4775	<b>15058</b>
E2D1	2224	4543	2146	<b>8913</b>	25 %	51 %	24 %	1459	3760	1914	<b>7133</b>
E3D2	1274	2566	1038	<b>4878</b>	26 %	53 %	21 %	509	1783	806	<b>3098</b>
E4D3	913	1714	536	<b>3163</b>	29 %	54 %	17 %	148	931	304	<b>1383</b>
E5D4	674	1279	370	<b>2323</b>	29 %	55 %	16 %	-91	496	138	<b>543</b>
E6D5	551	999	291	<b>1841</b>	30 %	54 %	16 %	-214	216	59	<b>61</b>
E7D6	436	816	242	<b>1494</b>	29 %	55 %	16 %	-329	33	10	<b>-286</b>
<b>Objetivo</b>	<b>765</b>	<b>783</b>	<b>232</b>		<b>43 %</b>	<b>44 %</b>	<b>13 %</b>				

Cuadro 5.4: Conteos C4\_ATP a 8 Mpx.

### 5.1.2. Funciones de distribución

En los casos anteriores no se tenían disponibles los archivos de coordenadas, sólo los conteos de cada fibra, por lo que el análisis se hizo en torno a estos. Sin embargo se facilitó el archivo del conteo manual de la imagen CC2\_EDL\_NADH para propósitos de comparación con los resultados del software.



Figura 5.5: Imagen CC2 EDL NADH.

Los resultados de un experto fueron un conteo de 1088 fibras oxidativas y 1087 fibras glucolíticas, que llevaron alrededor de 3 horas en concluirse. Los resultados

usando el software con resolución de 8 Mpx, con aprendizaje guiado por experto, y filtros más convenientes según inspección visual, fueron de 648 fibras oxidativas y 951 fibras glucolíticas en aproximadamente 15 minutos de procesamiento.

A pesar de que las cifras se alejan de los valores reales hasta en 570 fibras, se llegó a estas cifras teniendo un alto grado de confiabilidad, es decir la gran mayoría de las 648 fibras identificadas como oxidativas, realmente lo son, pues con filtros menos rigurosos se llegaba a conteos más cercanos pero gran parte de las coordenadas obtenidas no correspondían a fibras reales, más bien eran ruido, o distintas fibras interconectadas, que resultan más difíciles de corregir manualmente, pues se tiene que borrar y volver a agregar fibras, según reportan los expertos.

En la figura 5.6 se comparan las funciones de distribución de los resultados obtenidos del conteo del experto. Se puede observar que a pesar de la diferencia en los conteos, como se menciona en el párrafo anterior, la forma general de la distribución no presenta cambios drásticos además que también se pueden identificar hacia que parte del músculo se agregan las fibras, como es el caso de las fibras oxidativas que tienden hacia la parte inferior, y en menor medida se identifica la agregación de las fibras glucolíticas hacia arriba de la imagen.

Usando la metodología descrita en la Sección 4.4, usando una malla de  $11 \times 11$  se obtienen los resultados en el cuadro 5.5 al comparar los resultados obtenidos por un experto y los obtenidos por el software presentado para la imagen CC2 EDL NADH. La distancia entre distribuciones es influida por el tamaño de la imagen, por lo que el dato más fiable es el ángulo entre las estas.

CC2 EDL NADH	Oxidativas	Glucolíticas
Distancia	77.5183	58.0222
Angulo	28.6645°	23.2755°

Cuadro 5.5: Comparativa de distribución.

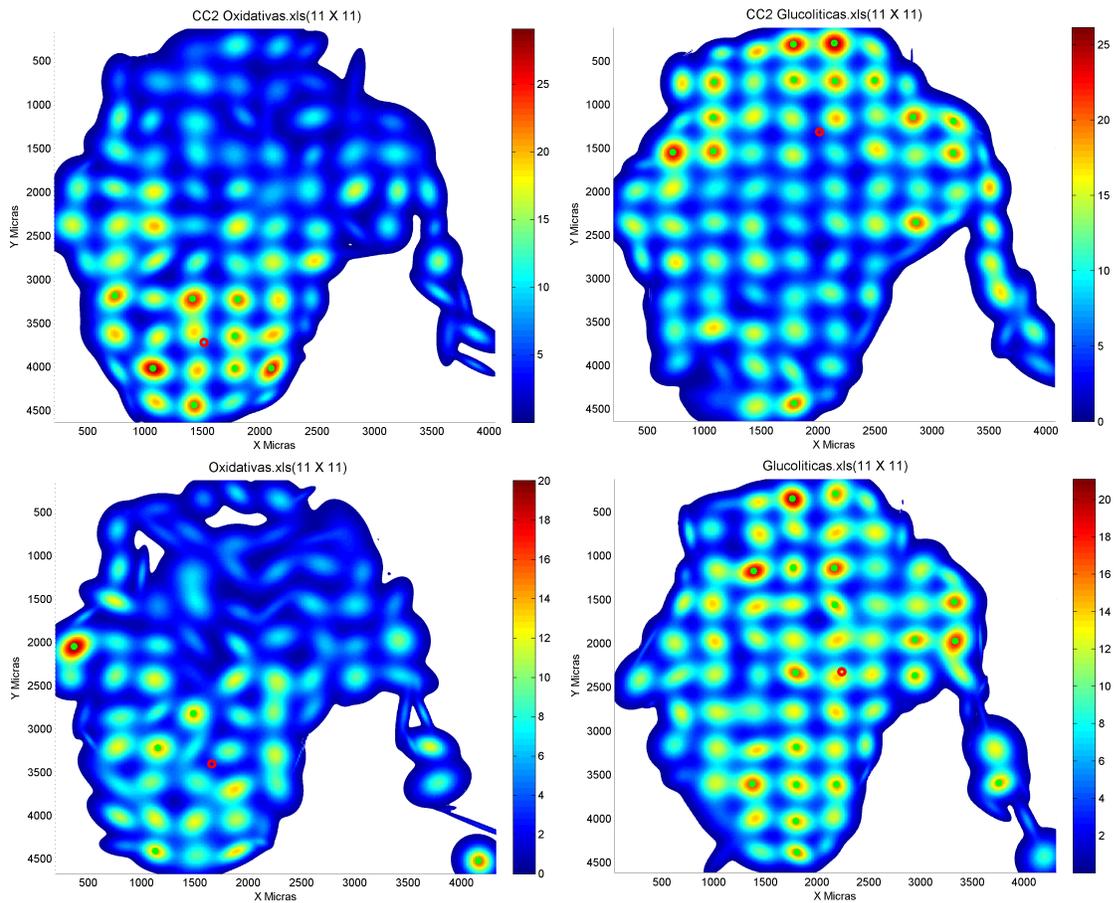


Figura 5.6: Imagen CC2 EDL NADH comparación de la distribución de fibras musculares. En la parte superior la reconstrucción de las funciones de distribución con el conteo de un experto y en la parte inferior el conteo con solo el software. En la izquierda fibras oxidativas y a la derecha las glucolíticas.

## 5.2. Comparación con un experto

### 5.2.1. Resultados con la tinción COX

Se utilizaron imágenes del músculo gastrocnemio de ratones teñidos con la tinción COX con el propósito de estudiar la capacidad del ácido clorogénico como protector contra la enfermedad de Huntington.

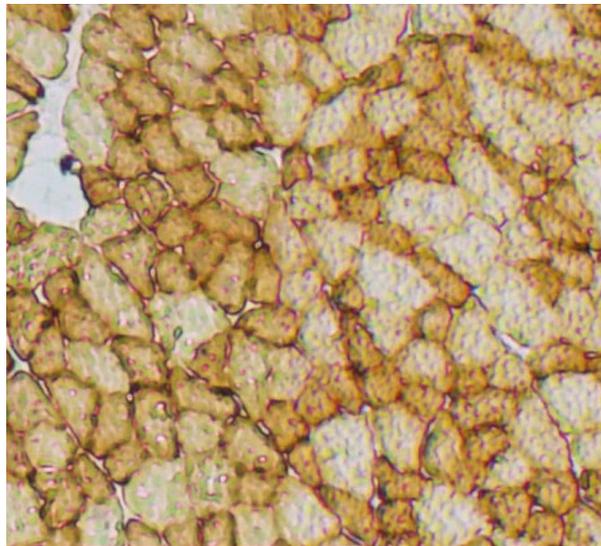


Figura 5.7: Muestra de una imagen con tinción COX del conjunto usado para los siguientes resultados.

En la figura 5.7, se muestra una de las imágenes con tinción histoquímica COX, los tipos de fibra identificados, en este caso por su metabolismo como glucolíticas y oxidativas. Cada uno de estos tipos fue contado manualmente por un experto utilizando el software *ImageJ*, identificado, en la figura 5.8, en la columna Experto, y contado mediante el Software presentado en este trabajo sin realizar ninguna modificación en la etapa de edición.

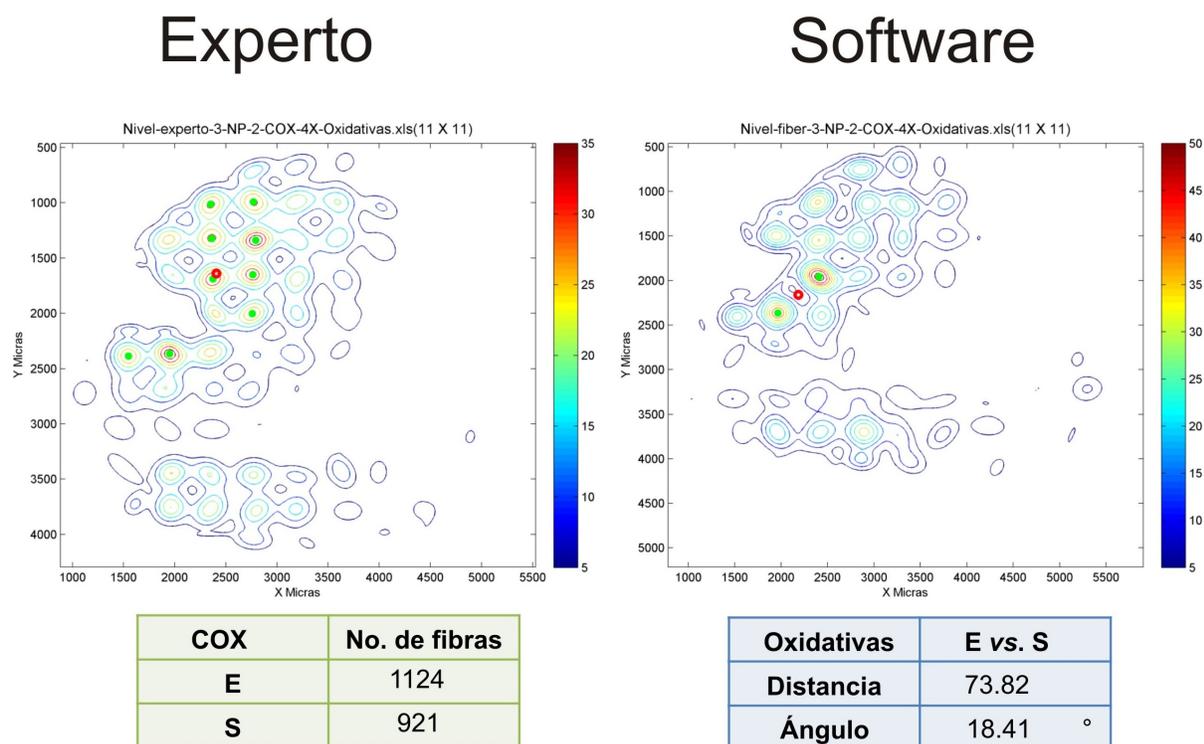


Figura 5.8: Proporción, distancia y ángulo con la tinción histoquímica COX comparando Experto (E) contra software (S).

En la figura 5.9 se muestran los resultados de los conteos de imágenes con tinción COX. Se realizaron pruebas estadísticas pareadas (se comparan todos los resultados de los conteos para todos pares de maneras de conteo por tipo de fibra, es decir, se comparan las población T-Experto vs. T-Software, T-Experto vs. T-Software+Experto y T-Software vs. T-Software+Experto donde T se refiere al tipo de fibras, es decir Glucolíticas u Oxidativas) y sin asumir distribución Normal (prueba de Wilcoxon) utilizando el software *GraphPadPrism* versión 6.01 para *Windows*; en cada gráfica se muestra la significancia para la cual cada par de distribuciones es diferente, en caso de serlo: \* para  $p < 0.05$ , \*\* indica  $p < 0.01$ , \*\*\* se usa para  $p < 0.001$  y \*\*\*\* señala  $p < 0.0001$ . Esta misma prueba se realizó para todas las comparaciones de cantidades de fibras siguientes en este Capítulo. En este caso se encontraron diferencias significativas entre las fibras oxidativas que fueron contadas por el experto comparadas con las contadas por el software por \*  $p = 0.03$ ,  $p < 0.05$ . En el caso de las fibras musculares glucolíticas fueron entre las contadas por el experto comparadas con las contadas por el software+experto, \*  $p = 0.0313$ ,  $p < 0.05$ . Cabe mencionar

en este caso fueron expertos distintos en las comparaciones *Experto* comparado con *Software+experto*.

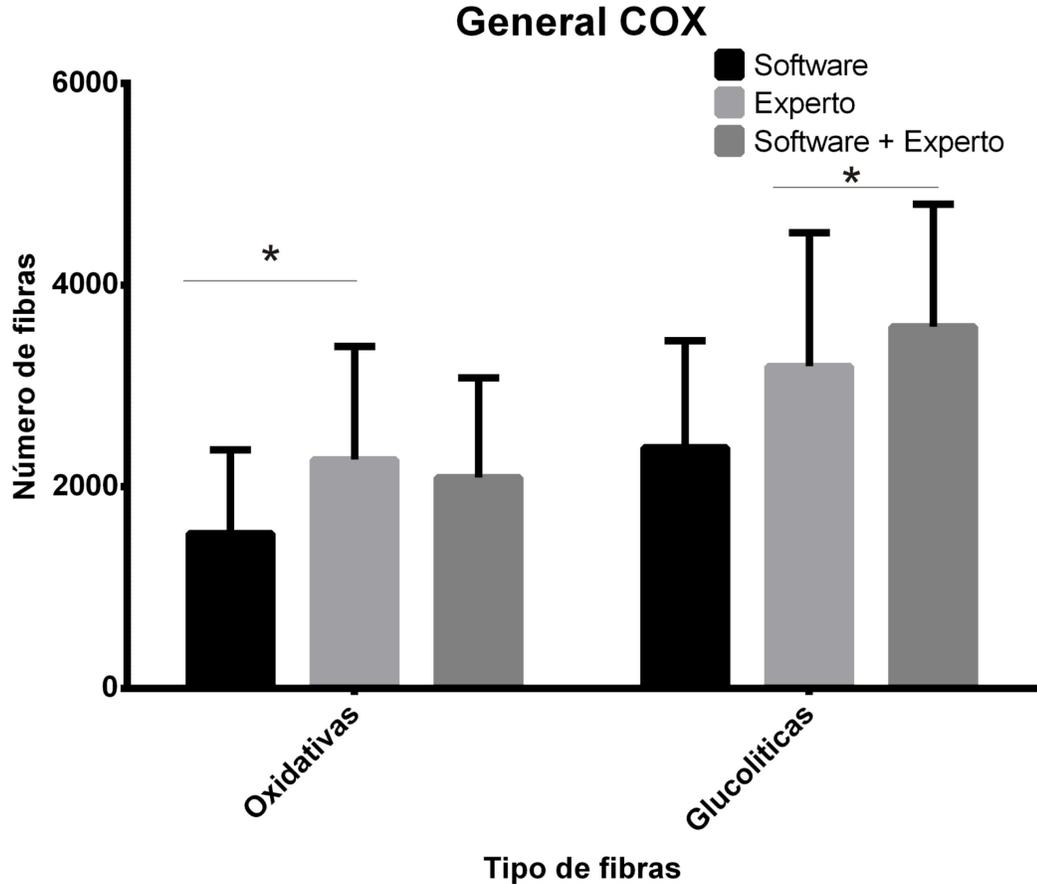


Figura 5.9: Proporción con la tinción histoquímica COX comparando el número de fibras musculares contadas por solo el Software, Experto y el Software guiado por experto (Software + experto).

Para para medir la eficacia en los conteos de los métodos propuestos por el software se compararon sus conteos con los del experto, es decir los conteos del software comparado con los del experto y los del software guiado por experto también comparado con experto. Se calcula la diferencia en los conteos, pero como la cantidad de fibras difiere en cada imagen, esta diferencia se divide entre la cantidad de fibras resultado del conteo del experto, con lo que se pretende tener una medida normalizada. Entonces se calcula el porcentaje de error como  $e_m = |c_e - c_m|/c_e$ , donde  $e_m$

es el error para el método  $m$ , que puede ser software o software guiado por experto (software + experto),  $c_m$  es el conteo del método  $m$  y  $c_e$  es el conteo del experto.

En la figura 5.10 se muestran los porcentajes de error para las fibras oxidativas de imágenes con tinción COX, y análogamente para fibras glucolíticas en la figura 5.11.

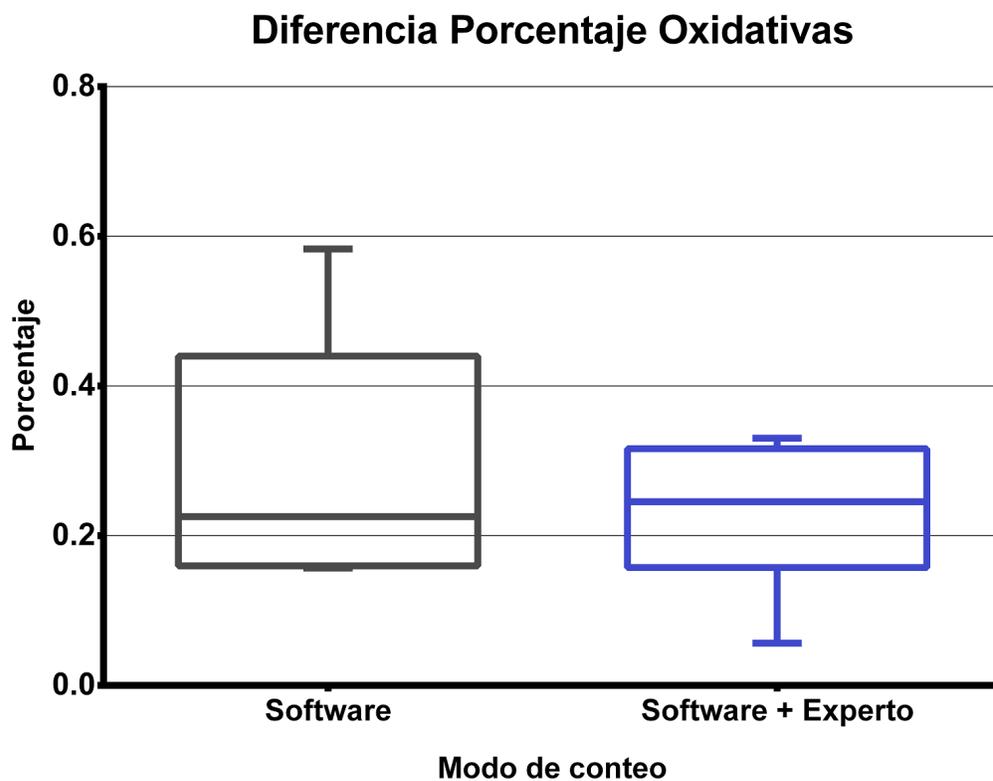


Figura 5.10: Porcentajes de diferencia para las fibras Oxidativas las imágenes de tinción COX.

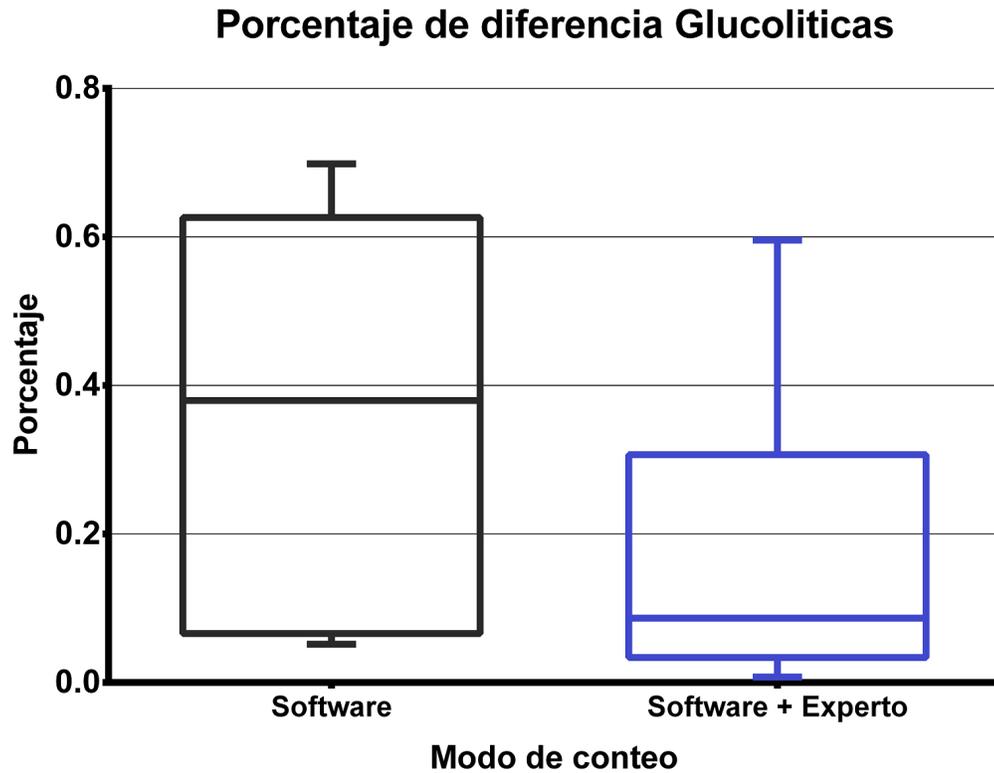


Figura 5.11: Porcentajes de diferencia para las fibras Glucolíticas las imágenes de tinción COX.

En la figura 5.12 se pueden ver los resultados del ángulo obtenido con imágenes con la tinción histoquímica COX, comparando experto contra software, software+experto contra software y experto contra software+experto. Para todos los casos en ángulo fue menor de 45 grados indicando que son similares las distribuciones obtenidas por las todas las comparaciones. No hubo diferencias significativas en este caso.

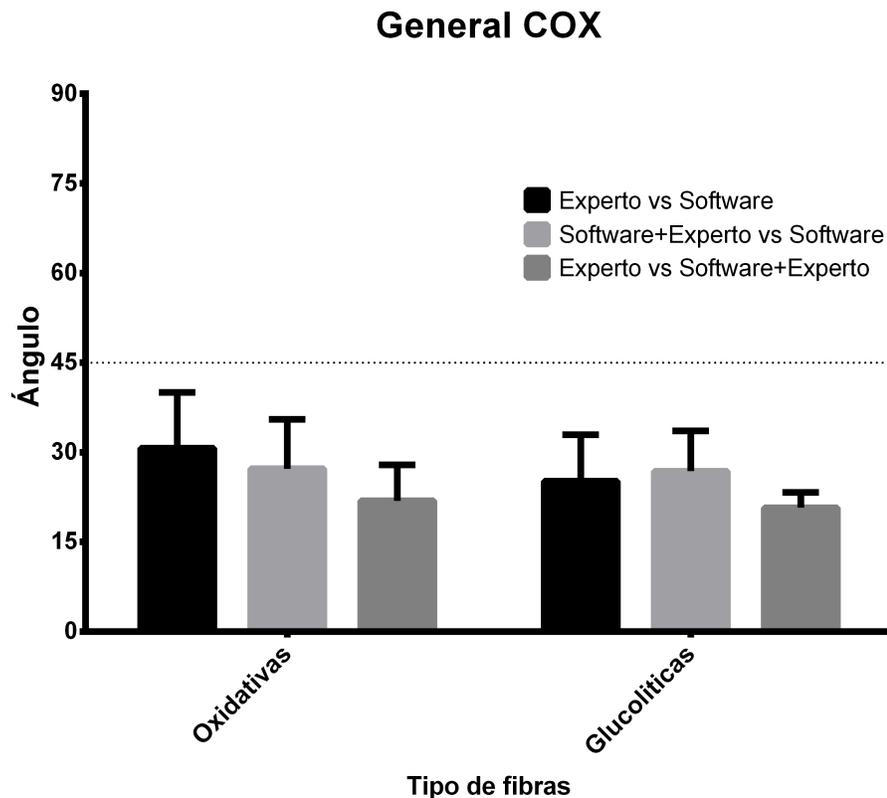


Figura 5.12: El ángulo con la tinción histoquímica COX. Más detalles en el texto.

### 5.2.2. Resultados con la tinción NADH

Se presentan los resultados del procesamiento de imágenes histoquímicas del músculo *extensor digitorum longus* de la rata utilizando la tinción histoquímica NADH, en la figura 5.13 se muestra un segmento de una imagen contada en este conjunto de prueba.

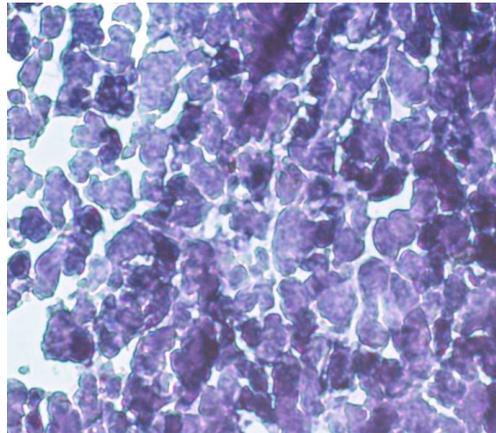


Figura 5.13: Muestra de una imagen con tinción NADH del conjunto usado para los siguientes resultados.

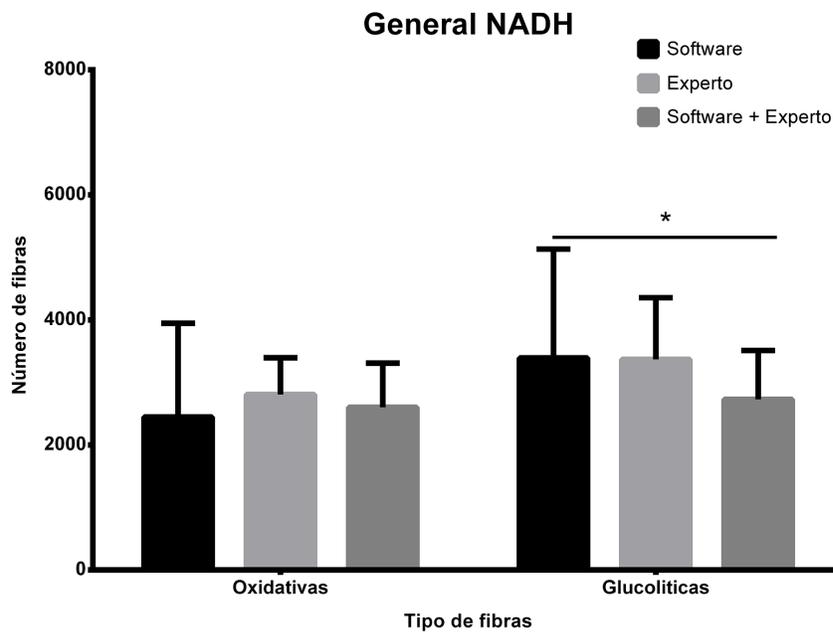


Figura 5.14: Número de fibras musculares identificadas en la tinción histoquímica NADH contadas por solo el software, solo el experto y software guiado por experto. Realizando la misma prueba que descrita en la Sección 5.2.1, las diferencias significativas fueron encontradas en las fibras glucolíticas entre software y software guiado por experto,  $p < 0.0273$ . Cabe mencionar que fueron 2 expertos distintos.

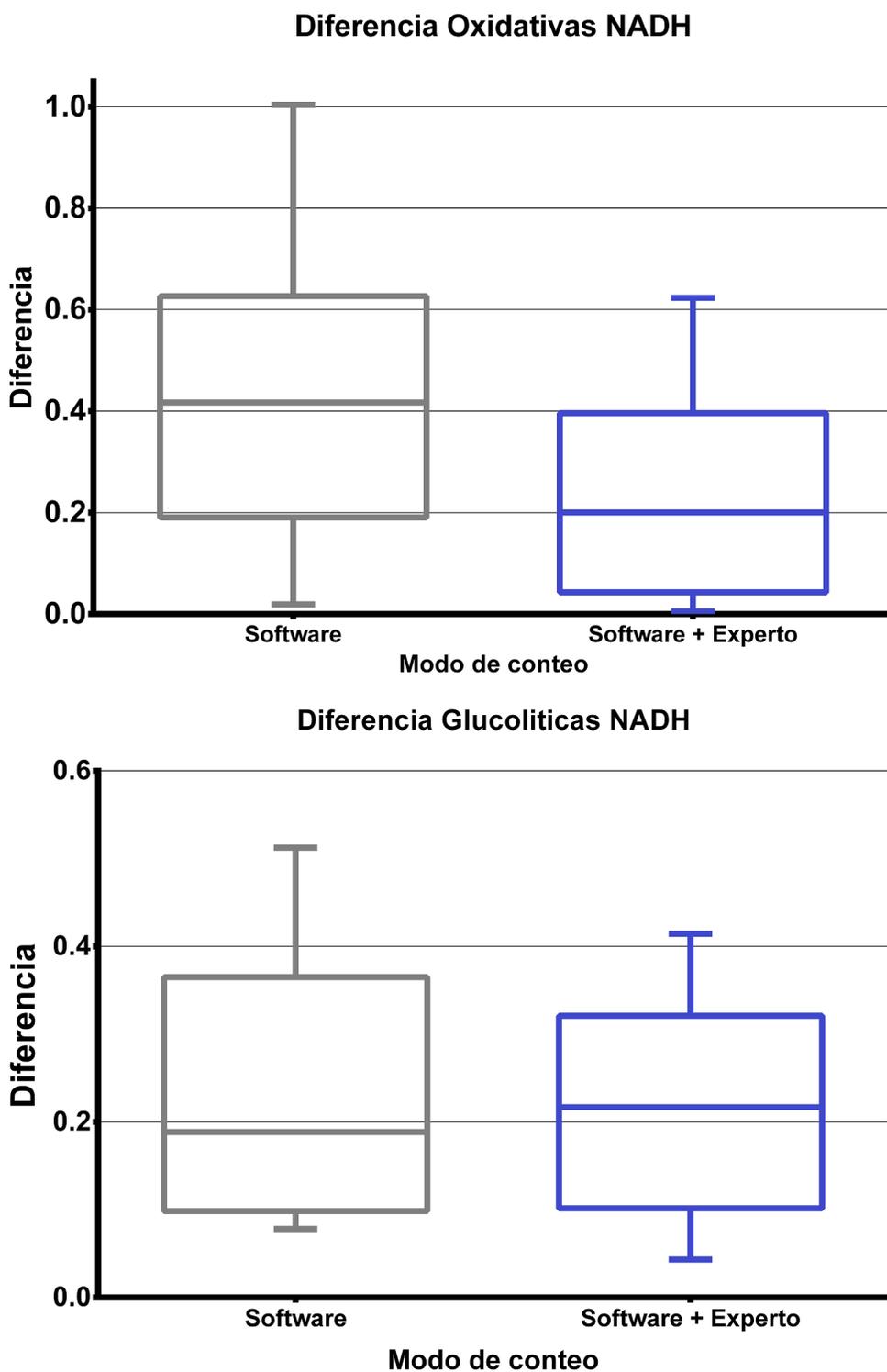


Figura 5.15: Diagrama de caja del porcentaje de error en imágenes de tinción NADH, mostrando mediana, desviación estándar, mínimo y máximo.

En la figura 5.14 se muestran los conteos de cada uno de los tipos de fibras contados con los tres métodos. Además en la figura 5.15 se muestran los porcentajes de error o diferencia de los métodos de conteo con el software y software+experto comparados con experto. Se muestra que en el caso de las fibras oxidativas la mediana baja de alrededor de 0.41 con solo software a 0.2 con software guiado. También se puede observar que los porcentajes se comportan muy parecido, por lo que sugiere que el software por si solo arroja buenos resultados.

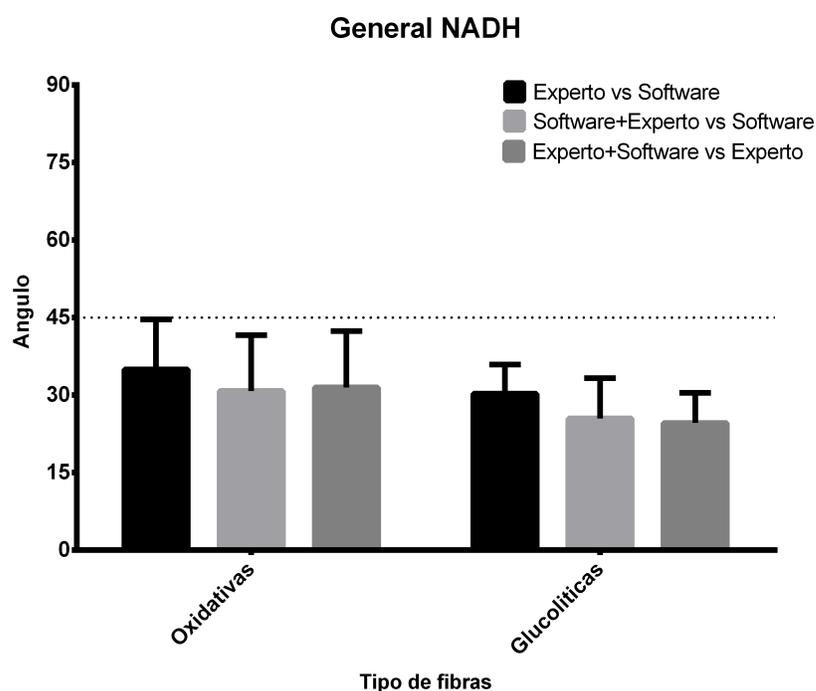


Figura 5.16: Ángulos obtenidos de las imágenes histoquímicas en la tinción NADH. En las comparaciones ilustradas el ángulo fue menor a 45 grados indicado que son similares las distribuciones encontradas en todos los casos.

En la figura 5.16 se muestran los ángulos obtenidos del conjunto de imágenes para prueba. Se observa que en todos los casos los ángulos caen son menores que 45°.

### 5.2.3. Resultados con la tinción ATPasa pH 9.4

Se presentan los resultados del procesamiento de imágenes histoquímicas del músculo *extensor digitorum longus* de la rata utilizando la tinción histoquímica ATPasa pH 9.4, se comparan las coordenadas de las fibras identificadas por un experto, software y los resultados obtenidos por el software guiado por experto presentado mediante el método de funciones de distribución.

En la figura 5.17 se muestra una imagen con la tinción ATPasa P.H. 9.4, se puede observar que la imagen presenta pequeñas perforaciones y en algunas fibras los tonos no son uniformes. Esto es común en los cortes histológicos. Para determinar la capacidad del software para detectar los distintos tipos de fibras musculares se llevó a cabo el siguiente análisis con 47 imágenes.

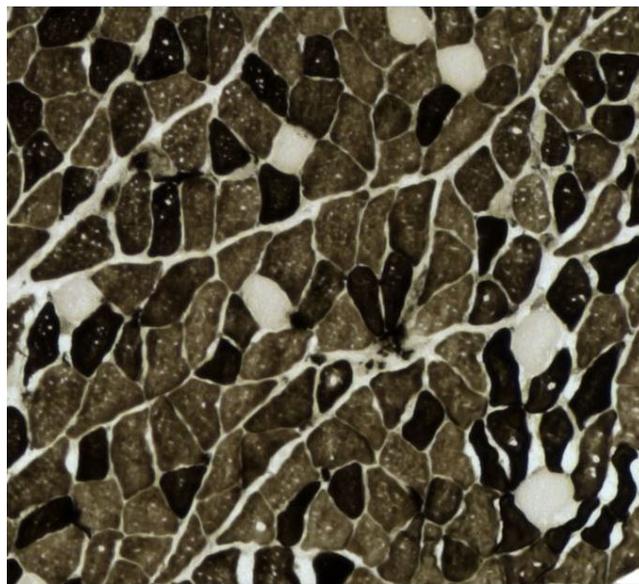


Figura 5.17: Imagen histoquímica de un corte de músculo con la tinción de ATPasa pH 9.4 en donde se muestran 3 tipos de fibras musculares rápidas (claras), intermedias (oscuras) y lentas (color blanco).

En las figuras 5.18, 5.19 y 5.20 se muestran las reconstrucciones de las funciones de distribución para hacer una comparación cuantitativa, de los tres tipos de fibra musculares identificados. En este caso rápidas, intermedias y lentas al tratarse de imágenes analizadas usando la técnica ATPasa pH 9.4. Cada uno de estos tipos contado manualmente por un experto utilizando el software *ImageJ*, identificado en

la columna Experto (E). Además, contado mediante el software presentado en este trabajo, utilizando el método de aprendizaje guiado por experto (SE) descrito en 4.3.1, y seleccionando el mejor filtro posible antes del proceso de conteo, y realizando modificaciones en la etapa de edición.

Para la figura 5.18, usando la misma prueba que en la Sección 5.2.1, se observa que el número de fibras contadas por el experto (E) es significativamente diferente \*\*\* ( $p < 0.001$ ) que las contadas con el software guiado por el experto (SE). Además, los cuantificadores de disimilaridad dan una distancia pequeña y el ángulo entre las fibras contadas por ambos casos es menor que 45 grados indicando que son similares los conteos.

## Rápidas

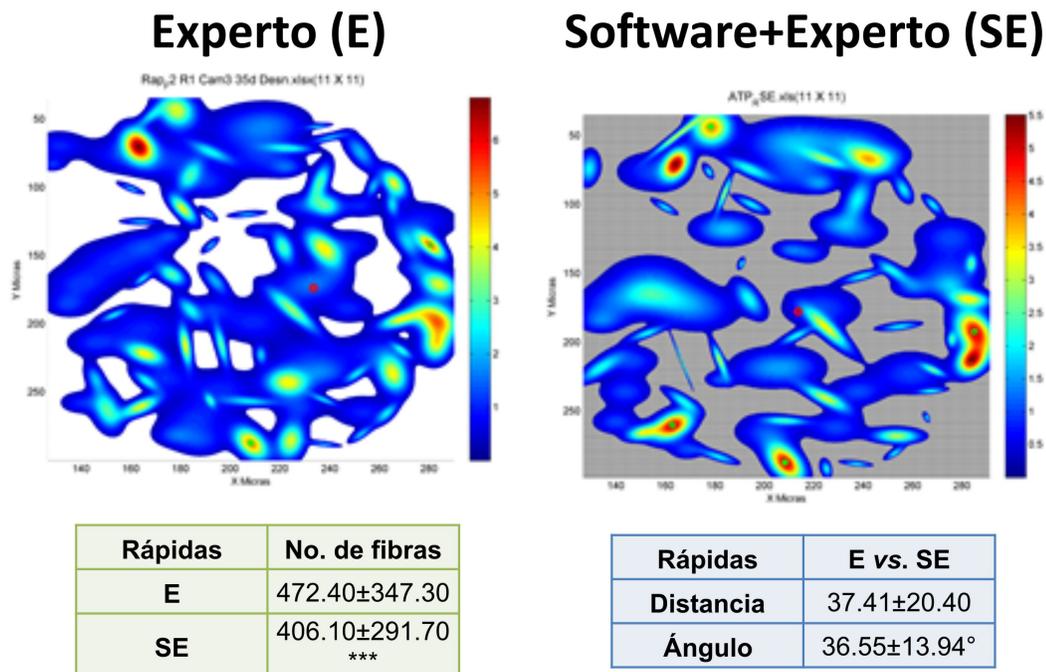


Figura 5.18: Resultados de comparar fibras musculares rápidas analizadas por un Experto y Software guiado por experto.

En el caso de la fibras intermedias tenemos como se ilustra en la figura 5.19 donde los conteos no muestran diferencias significativas. Se puede observar que la distancia es pequeña y el ángulo es menor de 45 grados indicando que son similares ambos conteos.

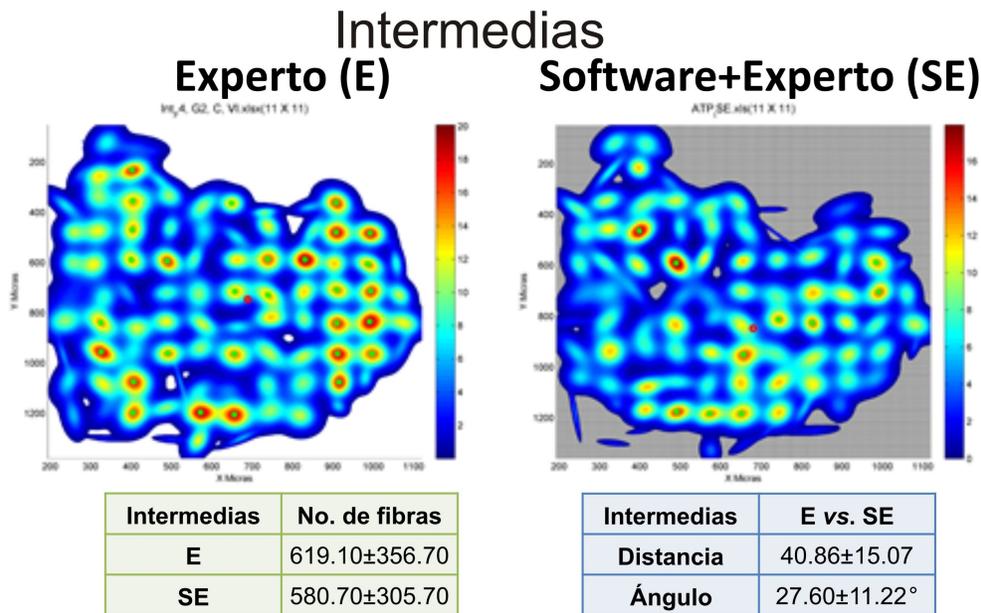


Figura 5.19: Resultados de comparar fibras musculares intermedias analizadas por un Experto y Software guiado por experto.

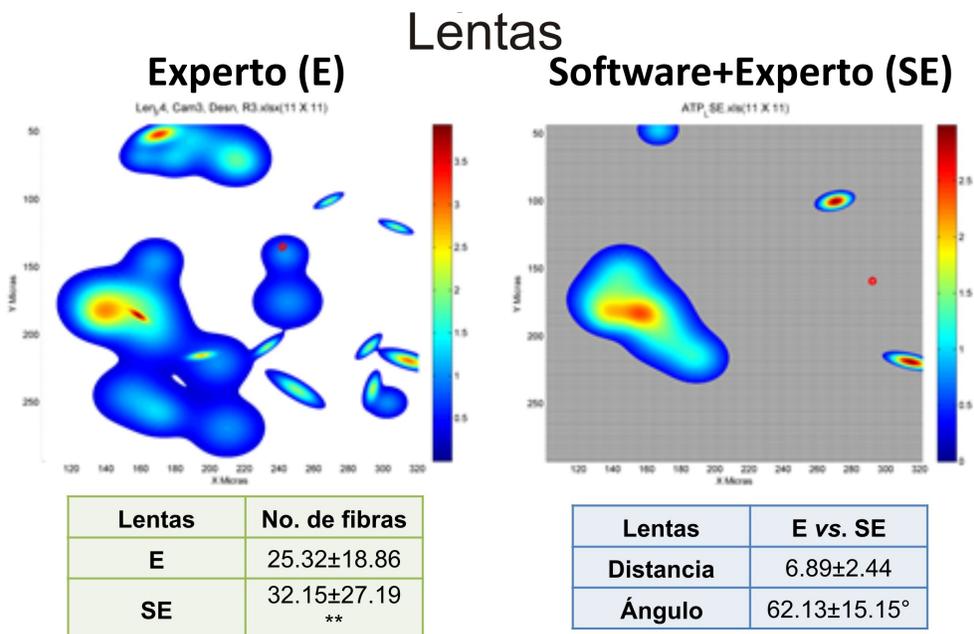


Figura 5.20: Resultados de comparar fibras musculares lentas analizadas por un Experto y Software guiado por experto.

En el caso de las fibras lentas que se presentan muy pocas en el corte histológico, como se ilustra en la figura 5.20. A pesar de ser pocas al ser contadas solo por el experto (E) y el software guiado por el experto (SE) existe diferencia significativa ( $p < 0.01$  con la prueba descrita en 5.2.1). Además, a pesar que la distancia es pequeña, el ángulo demuestra que las distribuciones en ambas fibras son diferentes. Por lo tanto cuando son pequeñas en número las fibras musculares el software no es necesario.

Las siguientes comparaciones software contra software guiado por experto 5.21 y experto contra software, 5.22 fueron realizadas con propósitos de probar la eficiencia del programa. Los resultados parecen ser muy alejados de la realidad, pues en la mayoría de los casos los resultados obtenidos con el software son mayores a los obtenidos por el experto, sin embargo cabe mencionar que el conteo por el experto tarda en promedio 120 minutos por imagen, y que los resultados del conteo por el software se obtuvieron en un promedio de 8 minutos por imagen, desde que se agrega conocimiento, hasta que se almacenan todos los archivos excel de resultados.

Esto sugiere que los resultados pueden ser mejorados si el software fuera guiado por un experto, que además éstos conteos serán más cercanos a los del método manual usado anteriormente si se usará la edición de los resultados, que tardaría menos pues ya existe un porcentaje de fibras contadas.

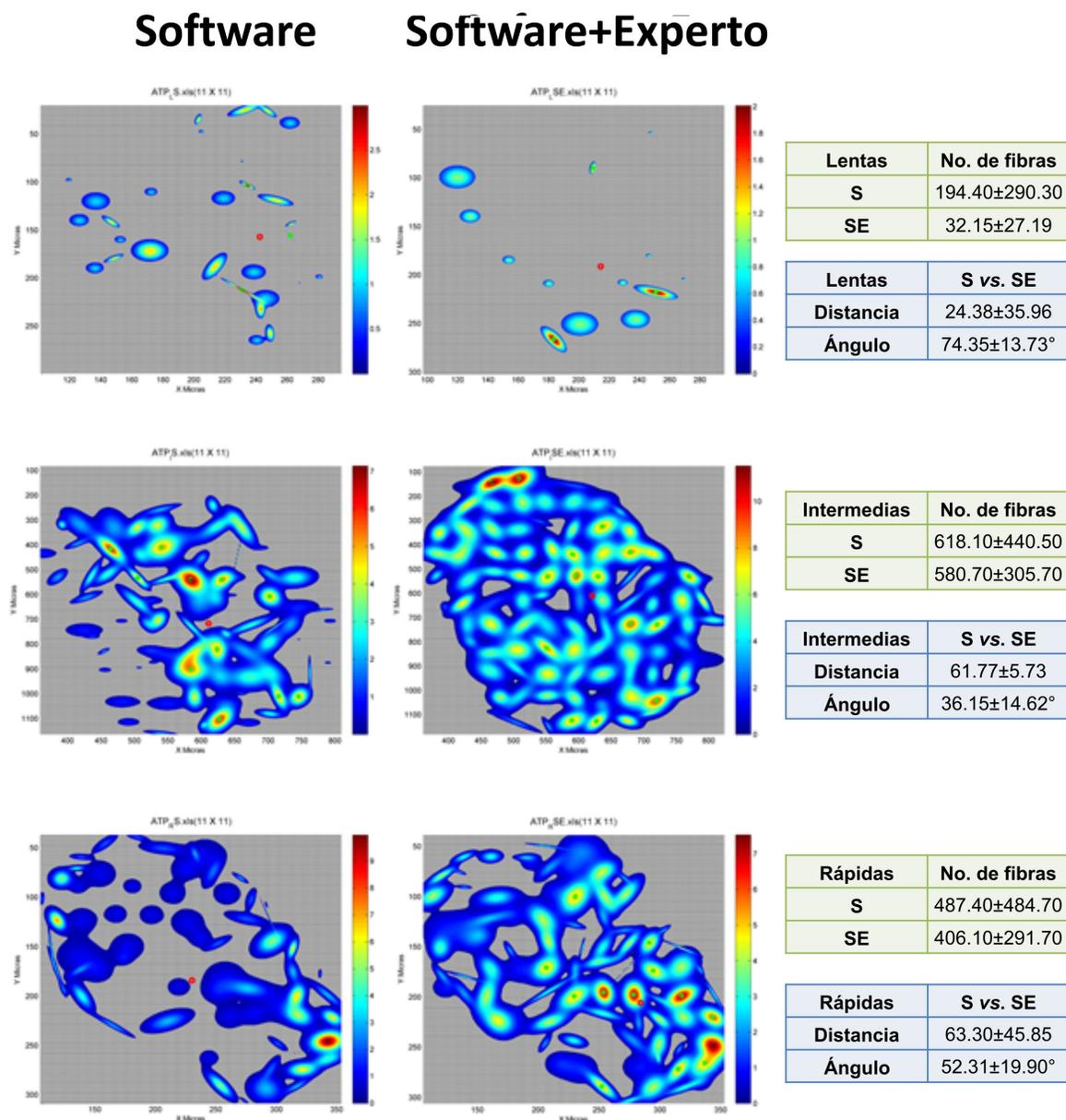


Figura 5.21: Proporción, distancia y ángulo con la tinción ATPase pH 9.4 comparando Software (S) contra software guiado por experto (SE).

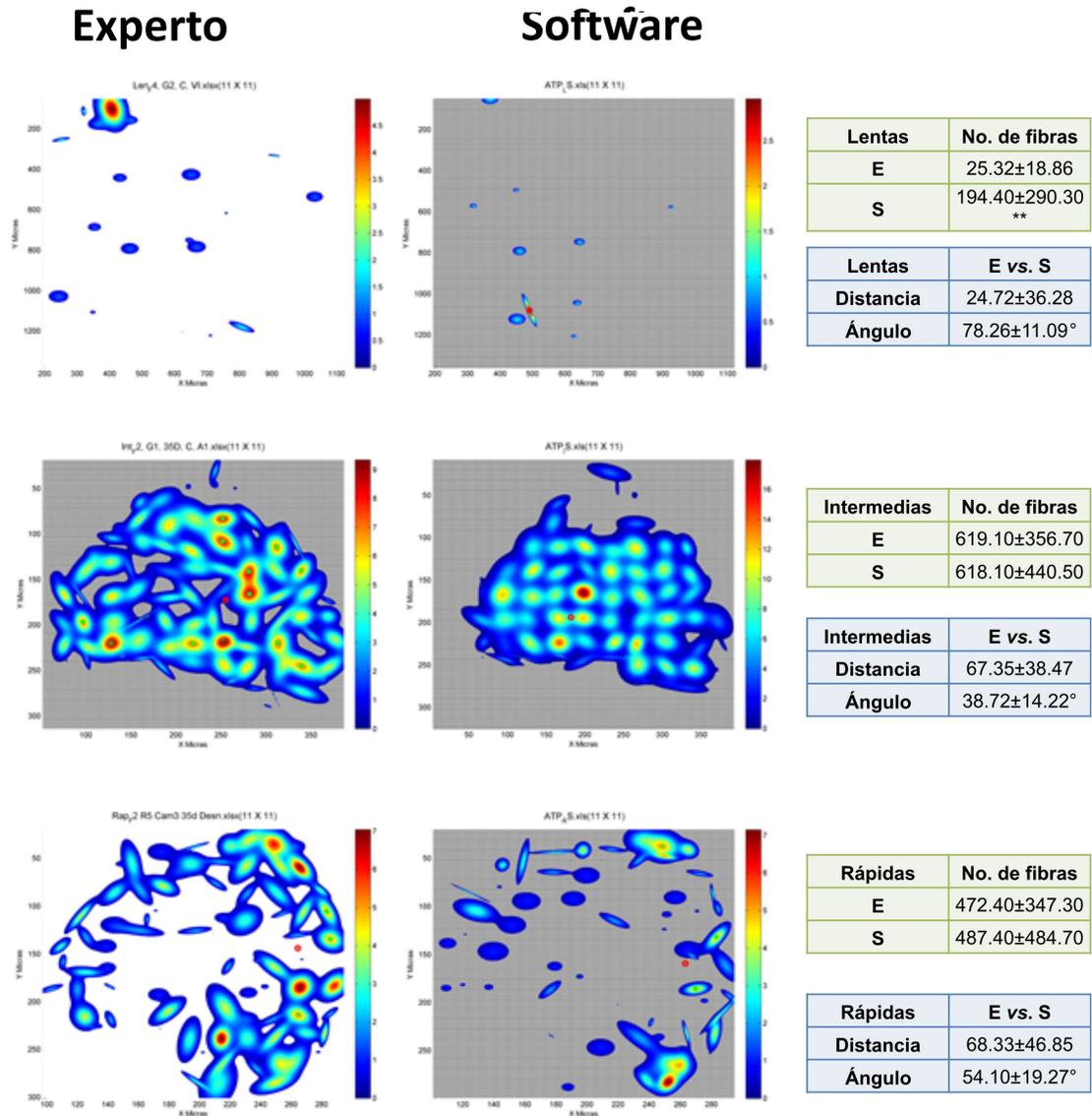


Figura 5.22: Proporción, distancia y ángulo con la tinción ATPasa pH 9.4 comparando Experto (E) contra software (S).

Los resultados generales comparando el conteo de fibras musculares se pueden consultar en la figura 5.23, hecho por solo el experto, solo el software o bien el software guiado por experto muestran diferencias estadísticas solo en las fibras musculares lentas y rápidas. En el caso de las fibras lentas debido a que son muy pocas el software cuenta un número considerablemente mayor. Y para las fibras musculares rápidas también el software contabiliza más que solo el experto o el software guiado por el experto. En el caso de las fibras intermedias el software cuenta la misma cantidad de fibras musculares que el experto.

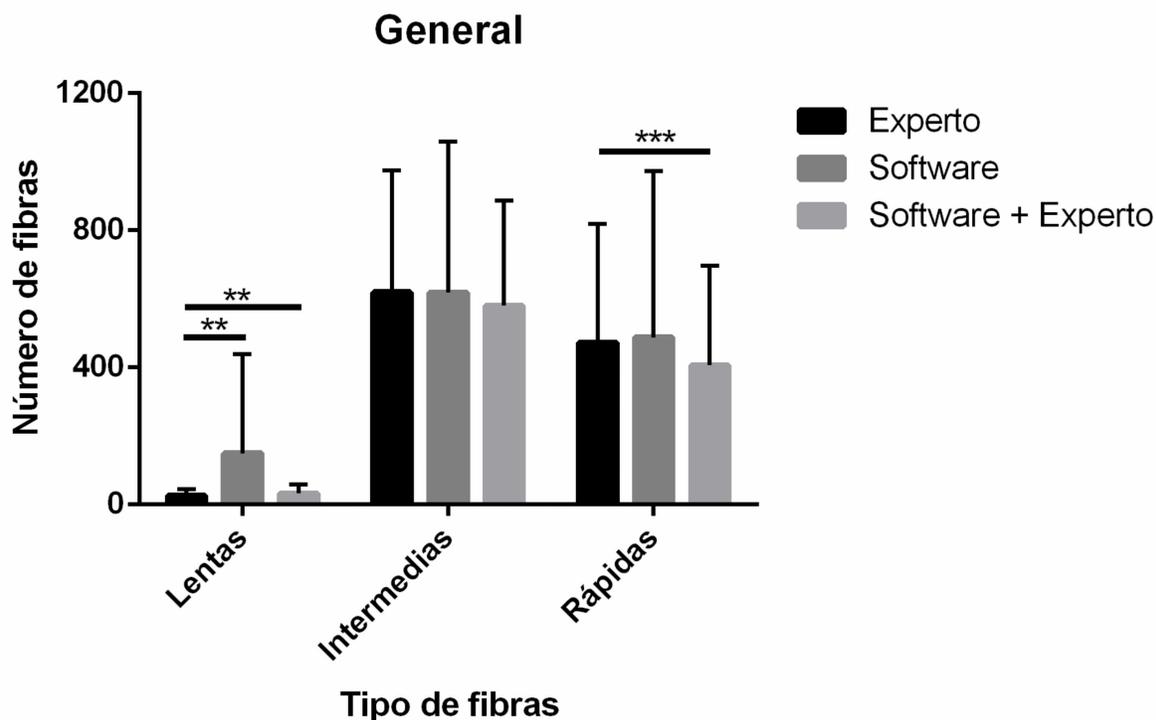


Figura 5.23: Número de fibras musculares usando imágenes con tinciones histoquímicas ATPasa pH 9.4. Se ilustra el conteo de 47 imágenes, ya sea solo por Experto, solo software o bien software guiado por experto. Las diferencias significativas en las pruebas estadísticas \*\* para  $p < 0.01$  y \*\*\*  $p < 0.001$ . Más detalles en el texto.

En la figura 5.24 se muestran los porcentajes de diferencia de los conteos entre Software+Experto y Experto. Se muestran diferencias muy bajas, un 0.14 en el caso de las fibras intermedias y un 0.2 en el caso de las fibras rápidas, es decir una eficacia

de alrededor de 86% y 80% respectivamente. No se incluyen resultados de fibras lentas pues estos son muy dispares, pueden superar un valor de 20, pues como las fibras lentas suelen ser pocas dentro de la imagen y, además, su color es muy similar al del tejido conectivo, por lo que el software realiza un mal conteo.

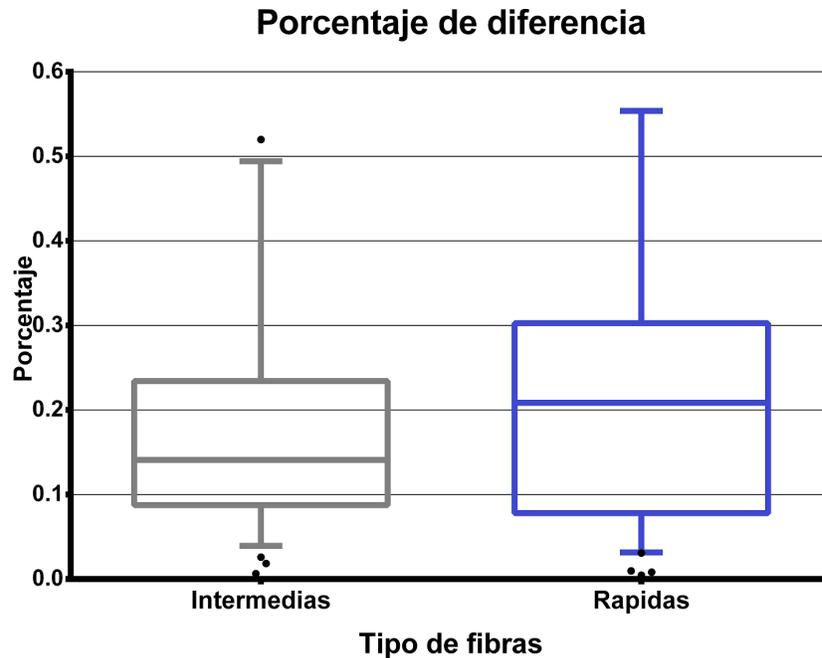


Figura 5.24: Porcentaje de diferencia para imágenes con tinción ATPasa entre Software+Experto y solo Experto, además no se incluyen fibras lentas pues el porcentaje es muy alto.

En la figura 5.25 se ilustran los detalles de comparar cada función de distribución como se indicó anteriormente (figuras 5.18, 5.19, 5.20, 5.22 y 5.21). Como se puede observar en la gráfica, las fibras musculares lentas tienen diferencias significativas en todos los casos. Se puede notar que al tener un ángulo menor de 45 grados se considera que las funciones de distribución son similares y esto ocurre para las fibras intermedias en todos los casos y para las fibras musculares rápidas se observa que los valores de los cuantificadores de disimilitud indican que son iguales las funciones de distribución cuando se comparan entre el experto contra software guiado por experto. las diferencias estadísticas están detalladas en la figura.

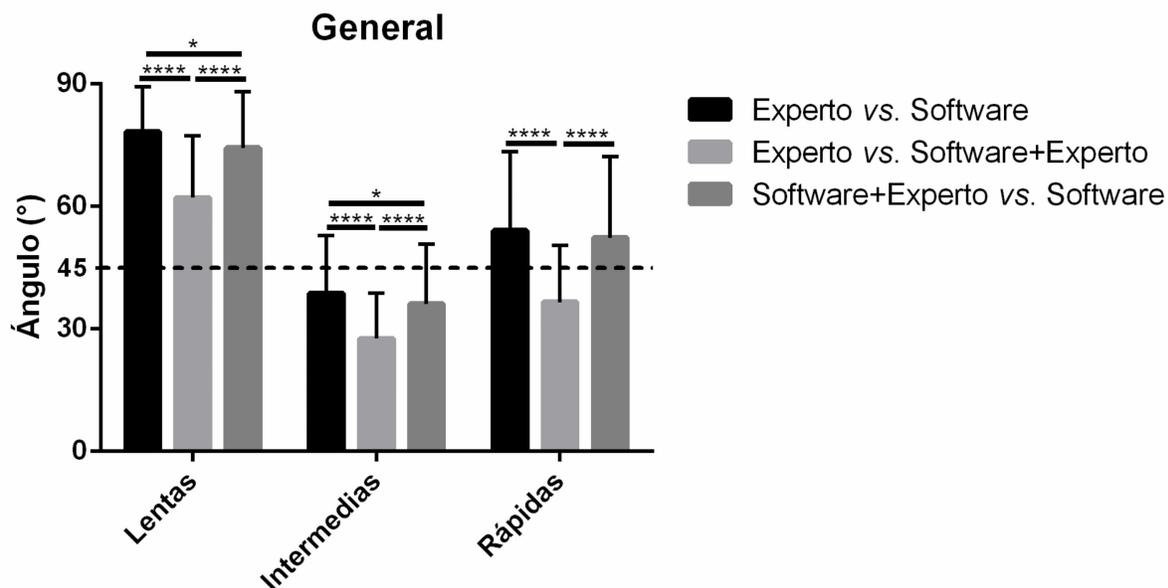


Figura 5.25: Cálculo de ángulos entre 47 imágenes con tinciones ATPasa pH 9.4. Las diferencias estadísticas entre cada una de las maneras de conteo del mismo tipo de fibra tienen la notación \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$  y \*\*\*\*  $p < 0.0001$  usando la prueba descrita en la Sección 5.2.1.

Los resultados en general con las 3 tinciones histoquímicas COX, NADH y ATPasa pH 9.4 utilizando el software con las opciones de los filtros y edición, demuestran que existen ventajas de usarlo.



# Capítulo 6

## Conclusiones

Se desarrolló un software que facilita la tarea de la clasificación de fibras musculares en imágenes histoquímicas utilizando tanto herramientas de inteligencia artificial como procesamiento de imágenes; se implementó en el lenguaje de programación Java, lo que permite su ejecución en múltiples plataformas con recursos computacionales fácilmente cumplidos en la actualidad, además de ofrecer un entorno amigable.

Entre sus características se encuentran las siguientes:

- Permite determinar el número y la posición relativa de cada una de las fibras presentes en los músculos esqueléticos, así como su identificación.
- Incluyen algoritmos de aprendizaje, clasificación así como filtros de ruido para los distintos tipos de fibras musculares.
- Usando el algoritmo modificado se observa una mejor aproximación al centroide de cada fibra en comparación con el algoritmo original de conteo. Además con la combinación de filtros se permite el cálculo del área de cada fibra que es útil para determinar la fuerza ejercida por el músculo estudiado.
- Tiene una interfaz amable con el usuario, familiarizado con un entorno de escritorio, ofreciendo capacidades de edición y de interacción con el usuario para la mejora de resultados.
- Optimiza recursos computacionales, tanto en memoria como potencia de procesamiento.

- Es flexible, pues permite regresar a cualquier punto del proceso y modificarlo.
- Permite exportar los datos en hojas de cálculo para ser usados en diferentes tipos de análisis (Distribución, Reconstrucción de imágenes binarias, etc.).

Cabe mencionar que no fue posible la integración de una herramienta de validación en el propio sistema de Java, la cual (funciones de distribución) fue implementada en el lenguaje de programación de Matlab, sin embargo es totalmente compatible mediante la exportación de los resultados en hojas de cálculo.

La comparación de los resultados del uso del software con aquellos proporcionados por los expertos sugieren que:

- La eficacia del sistema varía de acuerdo a la calidad de la imagen.
- Se pueden obtener resultados con una confiabilidad de aproximadamente 70 % la cual se puede aumentar significativamente al ser guiado por un experto hasta un 90 %.
- El tiempo de procesamiento de las imágenes histológicas se reduce hasta en un 90 %, desde 120 minutos, tiempo que tarda un investigador experimentado, hasta alrededor de 10 a 15 minutos utilizando este software, y alrededor de 30 minutos al editar los resultados.

# Capítulo 7

## Apéndice

## 7.1. Manual del programa

### Manual Fiber

#### Tabla de contenido

Manual Fiber .....	1
Introducción .....	2
Requisitos.....	3
Ejecución .....	3
Desde línea de comandos: .....	3
Utilizando NetBeans .....	4
En el programa.....	5
Aprendizaje .....	5
Segmentación .....	8
Filtrado .....	10
Conteo y Corrección.....	11

## Introducción

Dado que actualmente, en muchos estudios relacionados con fibras musculares, la clasificación y conteo de las fibras se realiza manualmente por un experto, lo que conlleva a una alta inversión de tiempo y esfuerzo, aunado a la disponibilidad de equipos computacionales con altas capacidades de procesamiento, surge la necesidad de la creación de software que apoye en dicha tarea, que reduzca los requerimientos de supervisión por parte de un experto.

Utilizando herramientas de inteligencia artificial, este software fue desarrollado como una herramienta de apoyo para la clasificación, localización y conteo de fibras musculares en imágenes histológicas, así como para su almacenamiento en formatos compatibles que permitan estudios posteriores como lo es cálculo de dimensión fractal o comparación de distintos especímenes del mismo músculo. Además, al ser desarrollado en el lenguaje de programación Java, ofrece capacidades multiplataforma.

## Requisitos computacionales

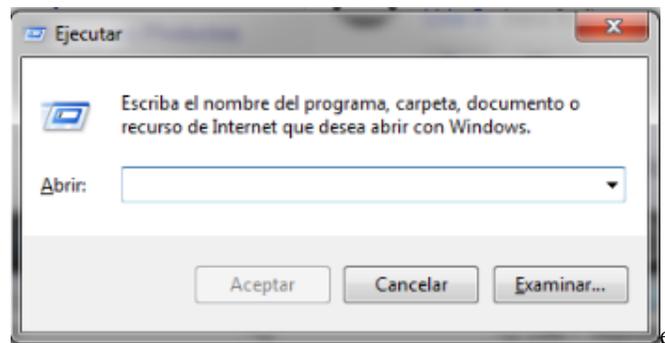
- Procesador Doble núcleo de 2 Ghz.
- Sistema operativo Microsoft Windows 7 o posterior o Apple Mac OS.
- Memoria RAM de 4GB
- Tener instalado Java

## Ejecución

Para ejecutar el software Fiber es necesario tener una copia del proyecto, además de las librerías en disco duro, de preferencia en C:/, el directorio raíz del sistema operativo.

Desde línea de comandos:

Presionar botón de Windows + R.



Escribir cmd y presionar la tecla enter o click en aceptar.

### *Buscar el programa en almacenamiento*

Navegar en la línea de comandos hacia la carpeta donde está el programa.

La línea de comandos se inicia en el directorio:

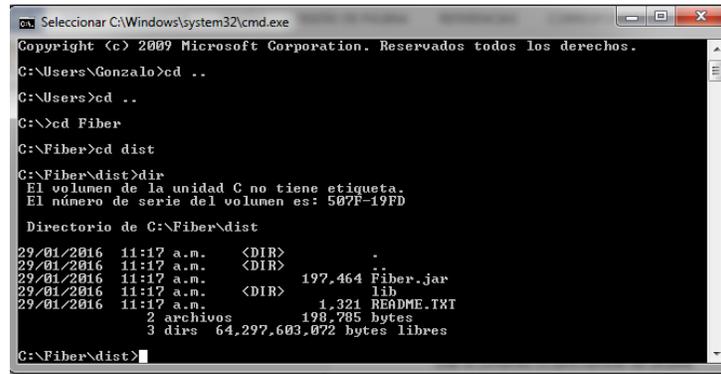
```
C:\Users\(Nombre)>
```

Usar el comando dir para visualizar contenido de carpeta actual

Usar el comando cd para cambiar de carpeta.

```
C:\Users\(Nombre)> cd (nombre carpeta).
```

Usar cd .. para ir una carpeta atrás.



```
Selecionar C:\Windows\system32\cmd.exe
Copyright (c) 2009 Microsoft Corporation. Reservados todos los derechos.
C:\Users\Gonzalo>cd ..
C:\Users>cd ..
C:\>cd Fiber
C:\Fiber>cd dist
C:\Fiber\dist>dir
El volumen de la unidad C no tiene etiqueta.
El número de serie del volumen es: 507F-19FD

Directorio de C:\Fiber\dist

29/01/2016  11:17 a.m.    <DIR>          .
29/01/2016  11:17 a.m.    <DIR>          ..
29/01/2016  11:17 a.m.                197,464 Fiber.jar
29/01/2016  11:17 a.m.                1,321 lib
29/01/2016  11:17 a.m.                1,321 README.TXT
                2 archivos          198,785 bytes
                3 dirs          64,297,603,072 bytes libres
C:\Fiber\dist>
```

### Parámetros de ejecución óptima

En la carpeta dist introducir el comando:

```
java -jar -Xms512m -Xmx1792m -XX:+UseG1GC Fiber.jar
```

Los números dependen de la cantidad de memoria RAM disponible en el sistema, a mayor cantidad mejor, pero algunas ocasiones el sistema operativo tiene un máximo para este valor, pero depende de los programas que se estén usando y muchos otros factores.

Por lo tanto se debe de probar hasta que se pueda correr el programa.

El número después de `-Xms` indica la cantidad mínima de memoria RAM con la que inicia el sistema, no olvidar la `m` (Megabytes).

El número después de `-Xmx` indica la cantidad máxima de memoria RAM que usara el programa.

En pruebas realizadas la cantidad recomendada permite usar imágenes de 8 Mpx sin problemas.

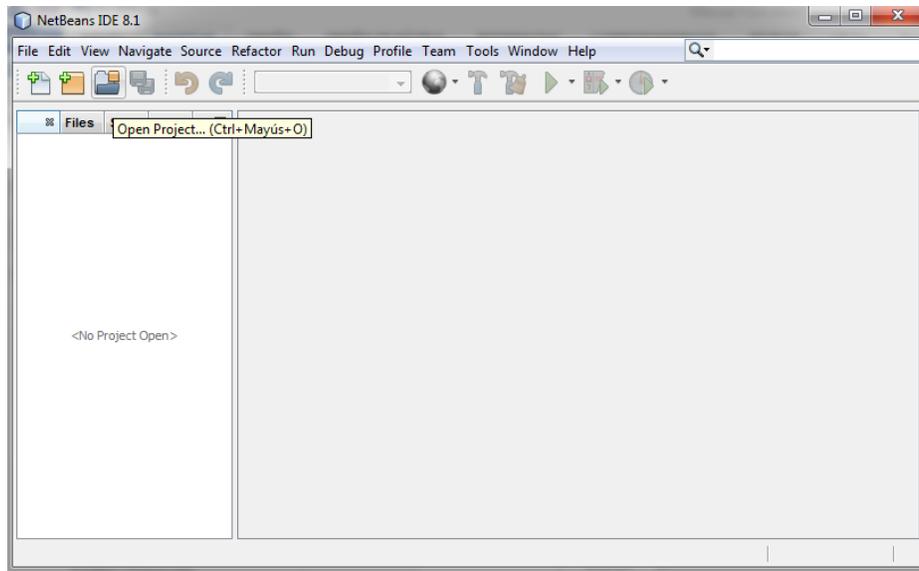
### Utilizando NetBeans

NetBeans es un entorno de desarrollo para varios lenguajes de programación como HTML, PHP, JavaScript, C, C++, y por supuesto Java, entre otros. Ofrece autocompletado de código, organización de aplicaciones en proyectos, herramientas para el diseño visual y un mejor entorno para correr aplicaciones.

### Instalación

Se necesita descargar e instalar el Java Development Kit (JDK). Además es recomendado descargar e instalar la última versión de NetBeans IDE.

Al abrir por primera vez NetBeans aparecerá una ventana como la que se muestra en la siguiente figura:



Para agregar el proyecto se presiona el botón Open Project y se navega a la carpeta donde se encuentra el proyecto, será fácilmente identificable porque la carpeta cambia de ícono:

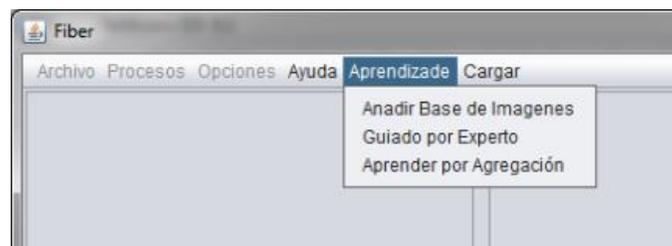
Una vez abierto el proyecto se habilitaran varios botones de la ventana de NetBeans, el más importante es Run Project, una flecha de color verde, para ejecutar Fiber se da click en este botón.

### En el programa

El procesamiento de imágenes histológicas con el programa se divide principalmente en 5 etapas: Aprendizaje, Segmentación, Filtrado, Conteo y Corrección.

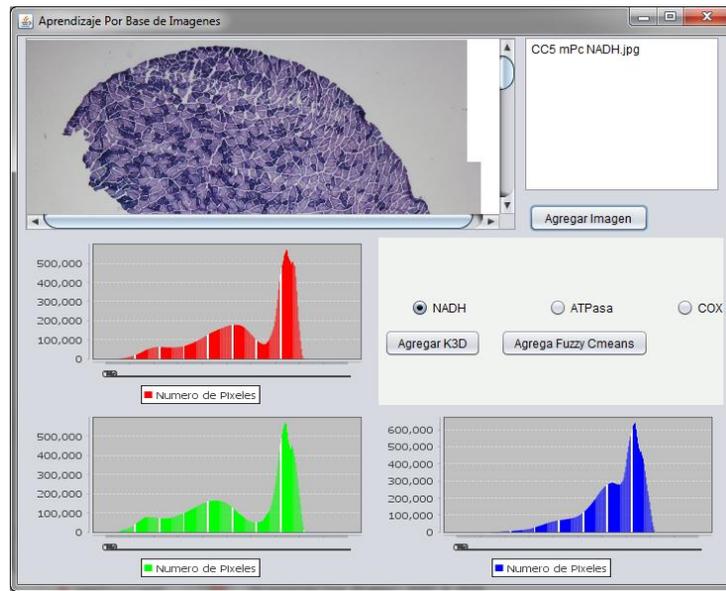
### Aprendizaje

En esta etapa se determinaran las gamas de colores de cada una de las fibras presentes en una imagen histológica, es decir se clasifican por su color. Se soportan 3 tipos de aprendizaje: Por Minería de datos en una base de Imágenes, Guiado por un experto y Agregación, que es una combinación de los anteriores.



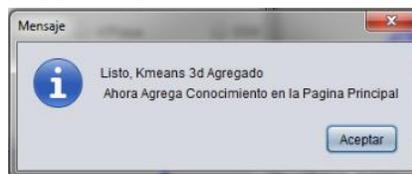
## Base de Imágenes

Se utiliza un conjunto de imágenes de la misma tinción para determinar los colores de cada uno de los tipos de fibra presentes. Para agregar imágenes se presiona el botón Agregar Imagen y se abre un selector de archivos, se presiona abrir y se pre visualizara la imagen y además se muestran histogramas de cada uno de los tonos presentes en la imagen por componente de color: rojo, verde y azul, como se muestra en la siguiente figura.



Cuando se termine de agregar todas las imágenes elegidas se selecciona el tipo de tinción y se presiona el botón perteneciente al algoritmo seleccionado; este programa soporta los algoritmos K-means en 3 componentes y Fuzzy C-Means.

Al terminar de ejecutarse cada algoritmo mostrara un mensaje:

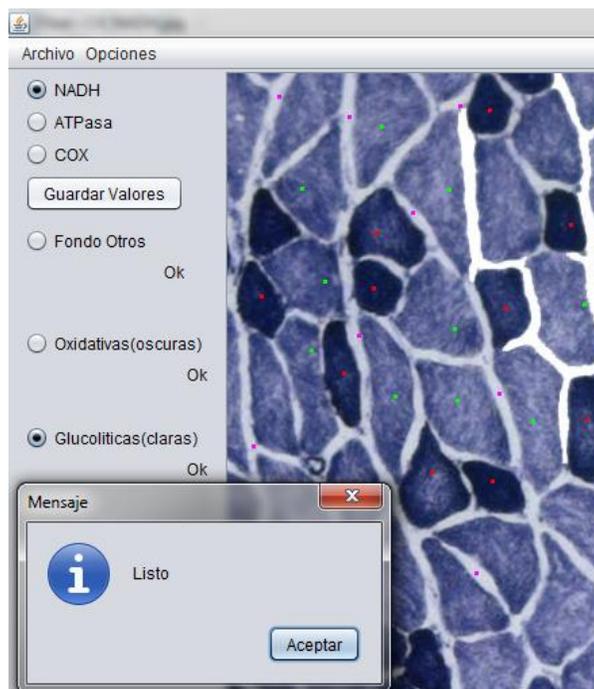


Como se indica, al finalizar de agregar conocimiento se debe agregar en la ventana principal en el menú Cargar -> Carga Conocimiento, con lo que se habilitara la elección de archivos para su análisis en el siguiente paso: Segmentación.

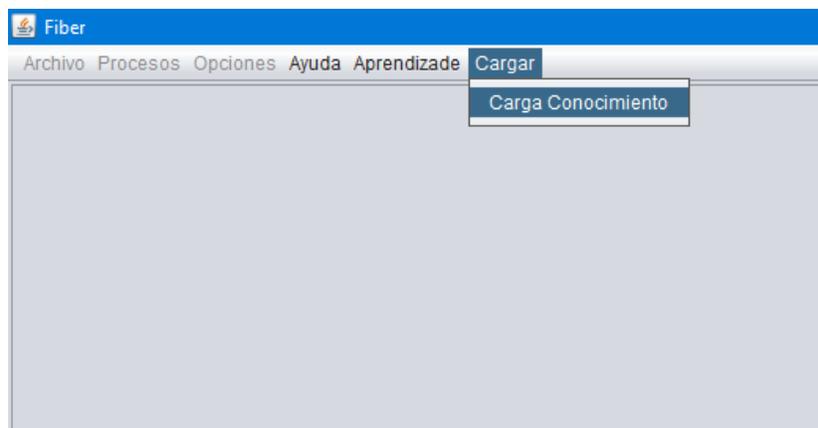
### Guiado por experto

Cabe destacar que dadas las características de los colores de las imágenes de la tinción ATPasa, los algoritmos de minería de datos en muchas ocasiones no son capaces de clasificarlos correctamente, por lo tanto se propone un aprendizaje guiado por un experto, es decir se indican manualmente los tonos de cada tipo de fibra.

En la ventana que se abrirá, se selecciona la técnica de tinción (NADH, ATPasa o COX) de la imagen que se utilizará, posteriormente se va a Archivo para abrir la fotografía (en formato JPG, GIF, BMP, TIFF o PNG). Ya abierta la imagen, se colocan 10 puntos para marcar el fondo (y tejido conectivo) y los tipos de fibras (lentas, intermedias y rápidas para ATPasa o glucolíticas y oxidativas para NADH y COX), al finalizar esto se da clic en Guardar Valores y aparecerá un mensaje que confirmará la acción.



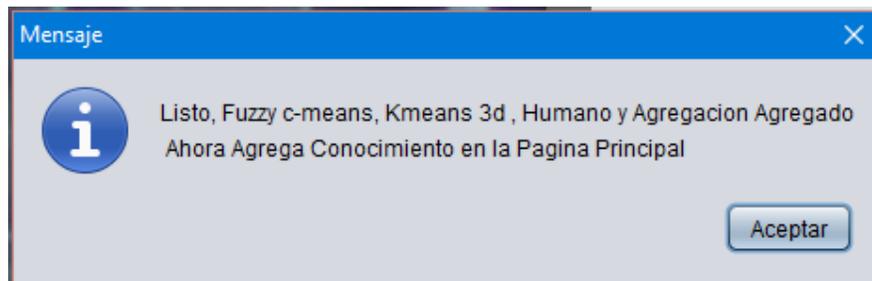
Después de dar clic en Aceptar en el mensaje volver a la página principal y dar clic en Cargar y seleccionar Carga Conocimiento.



## Agregación

Es una combinación de los métodos anteriores: mientras el experto indica manualmente los colores de cada tipo de fibra los algoritmos se ejecutan en segundo plano y arrojan un resultado que se guarda para las etapas posteriores además de una combinación de datos que se propone. Es el método recomendado pues se tiene información disponible de todas las fuentes y se puede cambiar entre una y otra fácilmente en la etapa de segmentación.

El procedimiento es el mismo que el de Guiado por experto, con la adición que se debe esperar a que se terminen de ejecutar los algoritmos antes de agregar conocimiento.

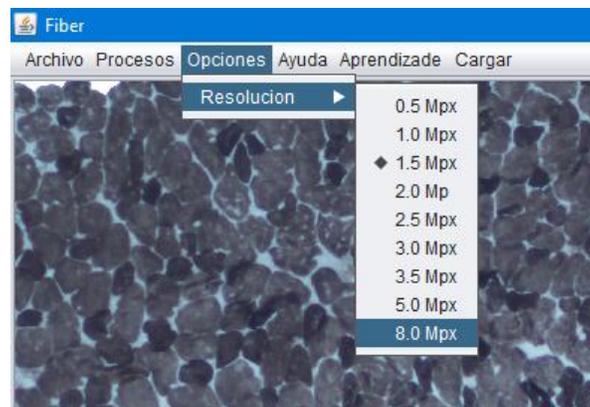


## Segmentación

En esta etapa cada imagen se separa en imágenes con un solo tipo de fibra.

Al agregar conocimiento de cualquiera de los tipos de aprendizaje realizados activará Archivo y Procesos. En Archivo dar clic en Abrir imagen y se selecciona la imagen con la que se trabajó en la ventana anterior.

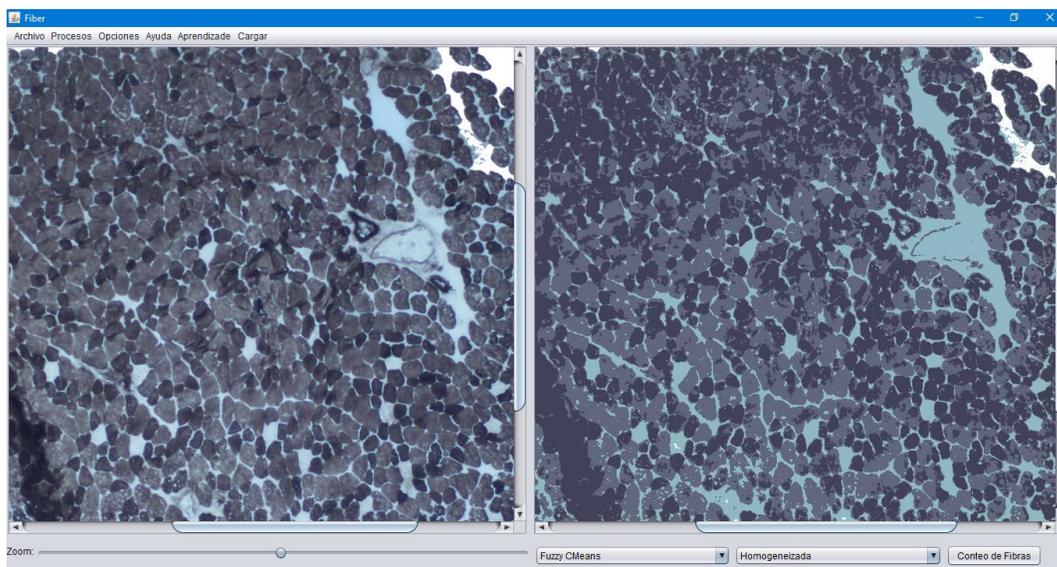
Es posible elegir la resolución de la imagen a procesar, pues en pruebas realizadas al tratar de procesar en la resolución original de la imagen, la memoria RAM resultaba insuficiente, además que se requiera gran capacidad de procesamiento. Cuenta con opciones desde 0.5 Mpx hasta 8 Mpx, pero inicialmente se encuentra en 1.5 Mpx. Se puede cambiar este valor en el Menú Opciones → Resolución).



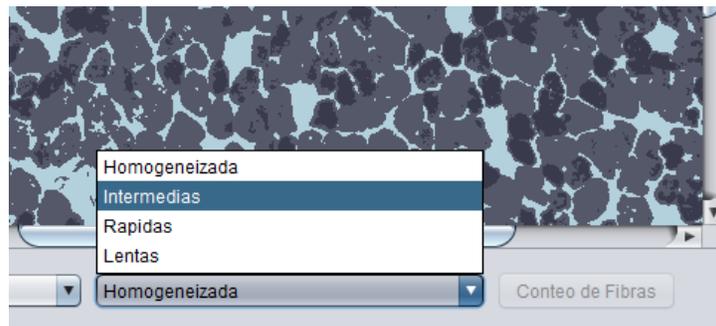
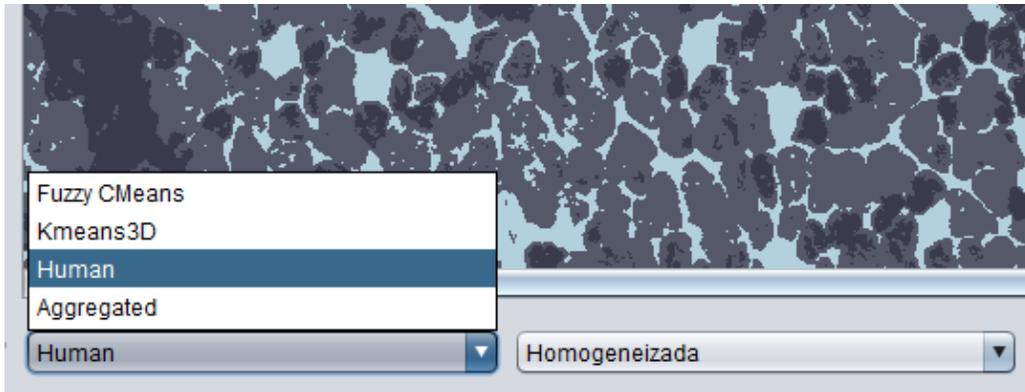
Una vez seleccionada la resolución de procesamiento ir a Procesos, luego en Homogeneizar y ahí elegir la técnica de tinción.



Una vez terminado el proceso, en la parte derecha aparecerá la imagen homogeneizada.



En la parte inferior se encuentran dos menús desplegables, el primero para elegir el algoritmo de aprendizaje del que proceden los datos para la segmentación, donde estarán los que se agregaron en el paso de aprendizaje y el segundo para cambiar la imagen que se ve en el panel derecho, con opción de seleccionar la imagen homogeneizada o imágenes con un solo tipo de fibra.



Al seleccionar un tipo de fibra aparecerá un tercer menú desplegable que corresponde a la siguiente etapa: Filtrado.

## Filtrado

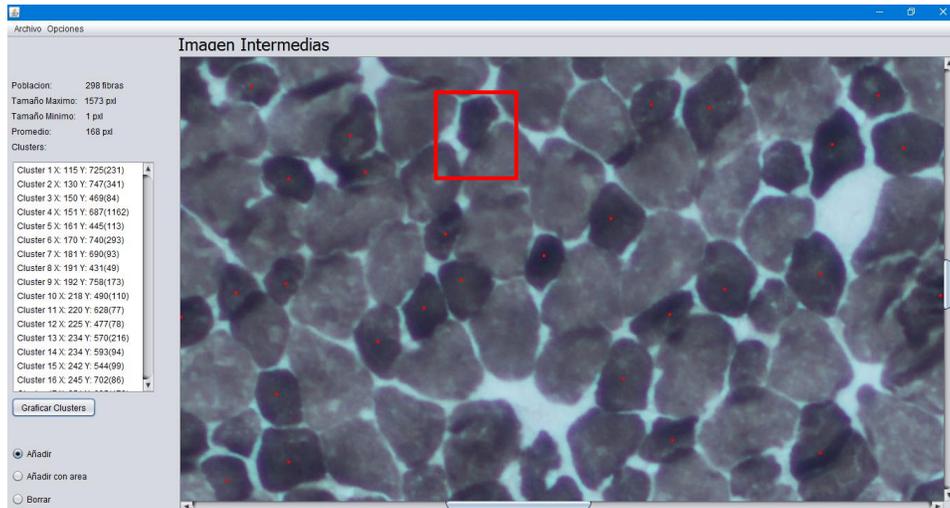
En esta etapa se pretende preparar la imagen de cada fibra para el algoritmo de conteo dado que en pruebas iniciales este arrojaba resultados erróneos pues las imágenes presentan ruido o la separación por color no es perfecta por la homogeneidad de los colores al cambiar de una fibra a otra.

Como se mencionó anteriormente en el tercer menú desplegable aparecen opciones de filtros personalizados que han funcionado en etapas de prueba del programa, sin embargo se deja a elección del experto el mejor posible según distintos factores como son el tipo de tinción, el tamaño del musculo, el Zoom del objetivo, la resolución de la imagen de procesamiento, ruido al momento de adquisición entre otros.

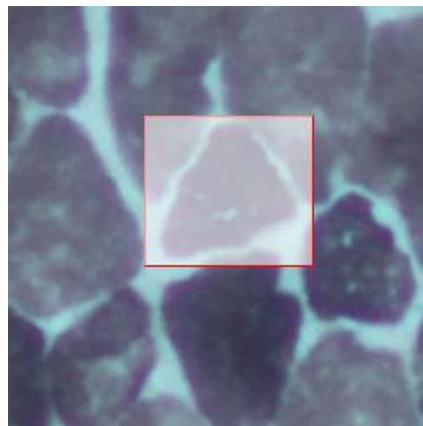


## Conteo y Corrección

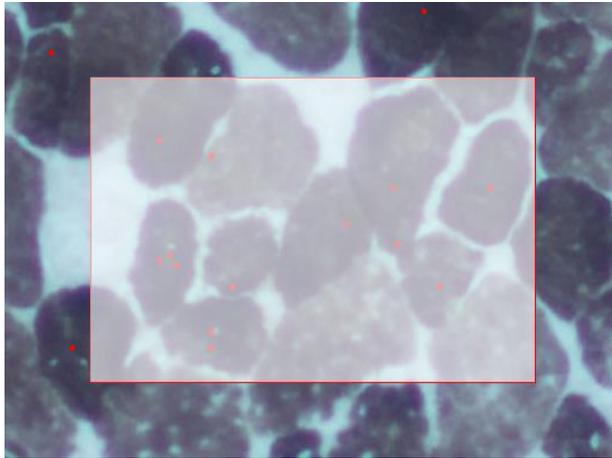
Una vez seleccionado el filtro óptimo se da clic en Conteo de Fibras, lo cual abrirá una ventana donde mostrará el número de fibras contadas así como una imagen sobre la cual se muestran los centroides de las fibras contadas.



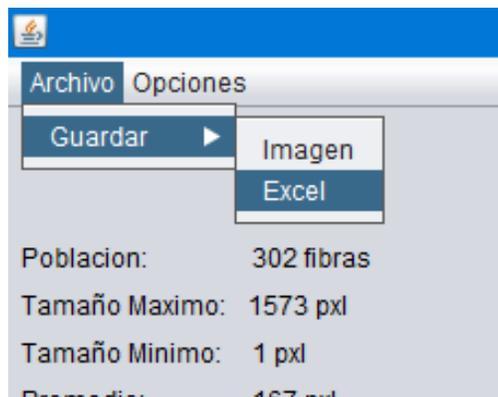
En caso de que el programa no haya contado alguna fibra (como la del recuadro en la imagen anterior) se puede usar la opción Añadir (ubicada en la parte inferior izquierda de la ventana), esto hará que al dar clic sobre la imagen se añada una fibra más al conteo, así como sus coordenadas X, Y. Además de esta opción se encuentra la de Añadir con área, la cual permite que el programa coloque el punto en un sitio más preciso al centroide; para esto se requiere marcar con el cursor a la fibra.



Si el programa contó una fibra de un tipo diferente al que se está trabajando o donde no hay fibra alguna el programa tiene la opción de Borrar, con la cual sólo hace falta dar clic sobre el punto que se desea quitar y éste se eliminará del conteo. Asimismo, esta opción de borrar permite borrar un área, con sólo dar clic y arrastrar el puntero por la imagen para elegir el tamaño y área donde se quiera eliminar los puntos.



Finalmente, se va a Archivo → Guardar y se da clic en Excel, lo cual abrirá una ventana para elegir la carpeta nombrar al archivo.



Para realizar el conteo de otro tipo de fibra sólo se regresa a la ventana principal y en se cambia el tipo de fibra y se vuelve a elegir el filtro óptimo y se procede a hacer el conteo.



# Bibliografía

- [1] Antonio Argüelles, E., Viveros Rogel, J., Rodríguez-Torres, E. E., López-García, K., Segura-Alegría, B. & Jiménez-Estrada, I. (2015). *Efecto de la Multiparidad sobre la Distribución y Organización Estructural de las Fibras de Músculos Pélvicos y Perineales de la Coneja (Oryctolagus cuniculus)*. Congreso Nacional, Sociedad Mexicana de Ciencias Fisiológicas 2015, San Miguel Allende, Guanajuato México.
- [2] Ashmore, C. R. (1974). *Phenotypic expression of muscle fiber types and some implications to meat quality*. Journal of Animal Science, 38(5), 1158-1164.
- [3] Barnard, R. J., Edgerton, V. R., & Peter, J. B. (1970). *Effect of exercise on skeletal muscle. I. Biochemical and histochemical properties*. Journal of applied physiology, 28(6), 762-766.
- [4] de Berg, M., Cheong, O., van Kreveld, M., Overmars, M.(2008) *Computational Geometry: Algorithms and Applications*, 3rd edn. Springer, Heidelberg
- [5] Booth, F. W., & Kelso, J. R. (1973). *Effect of hind-limb immobilization on contractile and histochemical properties of skeletal muscle*. Pfluegers Archiv, 342(3), 231-238.
- [6] Brox, T., Kim, Y. J., Weickert, J., & Feiden, W. (2006). *Fully-automated analysis of muscle fiber images with combined region and edge-based active contours*. In Bildverarbeitung für die Medizin 2006 (pp. 86-90). Springer Berlin Heidelberg.
- [7] Buche, P., & Mauron, D. (1997). *Quantitative characterization of muscle fiber by image analysis*. Computers and electronics in agriculture, 16(3), 189-217.
- [8] Dubowitz, V., & Pearse, A. E. (1960). *A comparative histochemical study of oxidative enzyme and phosphorylase activity in skeletal muscle*. Histochemistry and Cell Biology, 2(2), 105-117.

- [9] Duda, R. O. Hart, PE, & Stork, D. G. (2001). *Pattern Recognition (2nd edition)*. New York: John Wiley and Sons Inc.
- [10] Gonzalez, R. C., Woods, R. E., & Eddins, S. L. (2008) *Morphological image processing. Digital Image Processing.*, Prentice-Hall, Inc.
- [11] Guth, L., & Samaha, F. J. (1970). *Procedure for the histochemical demonstration of actomyosin ATPase*. *Experimental neurology*, 28(2), 365-367.
- [12] Karen, P., Števanec, M., Smerdu, V., Cvetko, E., Kubínová, L., & Eržen, I. (2009). *Software for muscle fibre type classification and analysis*. *European journal of histochemistry: EJH*, 53(2).
- [13] Karlsson, A. H., Klont, R. E., & Fernandez, X. (1999). *Skeletal muscle fibres as factors for pork quality*. *Livestock Production Science*, 60(2), 255-269.
- [14] Klemencic, A., Kovacic, S., & Pernus, F. (1998). *Automated segmentation of muscle fiber images using active contour models*. *Cytometry*, 32(4), 317-326.
- [15] Klont, R. E., Brocks, L., & Eikelenboom, G. (1998). *Muscle fibre type and meat quality*. *Meat science*, 49, S219-S229.
- [16] Lefaucheur, L., Buche, P., Ecolan, P., & Lemoing, M. (1992). *Classification of pig myofibres and assessment of postmortem glycogen depletion according to fibre type by computerized image analysis*. *Meat science*, 32(3), 267-278.
- [17] Lexell, J., Taylor, C. C., & Sjöström, M. (1988). *What is the cause of the ageing atrophy?: Total number, size and proportion of different fiber types studied in whole vastus lateralis muscle from 15-to 83-year-old men*. *Journal of the neurological sciences*, 84(2), 275-294.
- [18] Martín del Brío B. & Sanz Molina A. (2001). *Redes Neuronales y Sistemas Borrosos*. RA-MA.
- [19] Mendoza Quintana, A., Rodríguez Torres, E.E, Jiménez-Estrada, I. & López-García, K.(2015) *Efecto de la desnutrición y la primiparidad sobre el músculo pubococcígeo de la rata*. Congreso Nacional, Sociedad Mexicana de Ciencias Fisiológicas 2015, San Miguel Allende, Guanajuato México.
- [20] Miller, G. R., & Stauber, W. T. (1994). *Use of computer-assisted analysis for myofiber size measurements of rat soleus muscles from photographed images*. *Journal of Histochemistry & Cytochemistry*, 42(3), 377-382.

- [21] Mitra, S., & Acharya, T. (2005). *Data mining: multimedia, soft computing, and bioinformatics*. John Wiley & Sons.
- [22] Nachlas, M. M., D. G. Walker y A. M. Seligman. (1958). *A histochemical method for the demonstration of DNP-diaphorase*. The Journal of Cell Biology 4: 29-38.
- [23] Pal, N.R. , Bezdek, J.C. (1995), On cluster validity for fuzzy c-means model, IEEE Trans. Fuzzy Syst. 1 370–379.
- [24] Rasband, W.S., *ImageJ*, U. S. National Institutes of Health, Bethesda, Maryland, USA, <http://imagej.nih.gov/ij/>, 1997-2016
- [25] Ryu, Y. C., & Kim, B. C. (2005). *The relationship between muscle fiber characteristics, postmortem metabolic rate, and meat quality of pig longissimus dorsi muscle*. Meat Science, 71(2), 351-357.
- [26] Sertel, O., Dogdas, B., Chiu, C. S., & Gurcan, M. N. (2011). *Microscopic image analysis for quantitative characterization of muscle fiber type composition*. Computerized Medical Imaging and Graphics, 35(7), 616-628.
- [27] Schiaffino, S., & Reggiani, C. (2011). *Fiber types in mammalian skeletal muscles*. Physiological reviews, 91(4), 1447-1531.
- [28] Schölkopf, B., Tsuda, K., & Vert, J. P. (2004). *Kernel methods in computational biology*. MIT press.
- [29] Shawe-Taylor J., & Cristianini N.(2000). *An Introduction to Support Vector Machines and other kernel based methods*. Cambridge: Cambridge University Press.
- [30] Stein, J. M., & Padykula, H. A. (1962). *Histochemical classification of individual skeletal muscle fibers of the rat*. American Journal of Anatomy, 110(2), 103-123.
- [31] Toscano, A. E., R. Manhães-de-Castro & F. Canon. (2007). *Effect of a low-protein diet during pregnancy on skeletal muscle mechanical properties of offspring rats*. Nutrition 24: 270-278.
- [32] Vázquez Mendoza E., Hinojosa C. X., Rodríguez-Torres E. E., Mariscal Tovar S., Jiménez-Estrada I., & López-García K. *Composición y organización de las fibras en los fascículos del músculo extensor digitorum longus de rata*. Congreso Nacional, Sociedad Mexicana de Ciencias Fisiológicas 2015, San Miguel Allende, Guanajuato México.

- [33] Whitaker, R. H., & Borley, N. R. (2016). *Instant anatomy*. John Wiley & Sons.
- [34] Zhu, M. J., Ford, S. P., Means, W. J., Hess, B. W., Nathanielsz, P. W., & Du, M. (2006). *Maternal nutrient restriction affects properties of skeletal muscle in offspring*. *The Journal of physiology*, 575(1), 241-250.