



# Software para la Clasificación de Fibras Musculares en Imágenes Histológicas

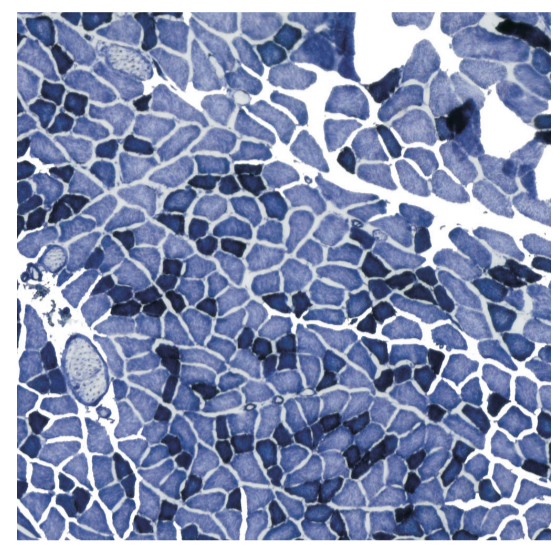


Chávez Fragoso Gonzalo Adán<sup>1</sup>, López Ortega Omar<sup>2</sup>, Rodríguez Torres Erika Elizabeth<sup>3</sup>, López García Kenia<sup>4</sup> y Jiménez Estrada Ismael<sup>4</sup>  
<sup>1</sup>Licenciatura en Matemáticas Aplicadas, Universidad Autónoma del Estado de Hidalgo. <sup>2</sup>Centro de Investigación en Tecnologías de Información y Sistemas, Universidad Autónoma del Estado de Hidalgo. <sup>3</sup>Centro de Investigación en Matemáticas, Universidad Autónoma del Estado de Hidalgo, <sup>4</sup>Departamento de Fisiología, Biofísica y Neurociencias, CINVESTAV-IPN.

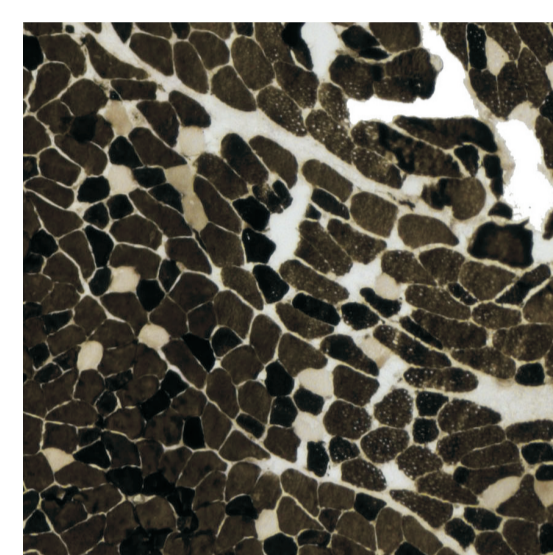
## 1 Planteamiento del problema

Para establecer la organización y distribución de los tipos de fibras presentes en los músculos esqueléticos es necesario identificar, contar y establecer su posición relativa dentro del músculo. Esta información se obtiene de las imágenes de los cortes histológicos. Tal procedimiento conlleva en la mayoría de las ocasiones, un considerable gasto de tiempo y esfuerzo por parte del observador. Ante ello, se propone el desarrollo de un procedimiento computacional que facilita la determinación del número y de la posición de los distintos tipos de fibras musculares mediante el empleo de cómputo inteligente y de la interacción con el usuario.

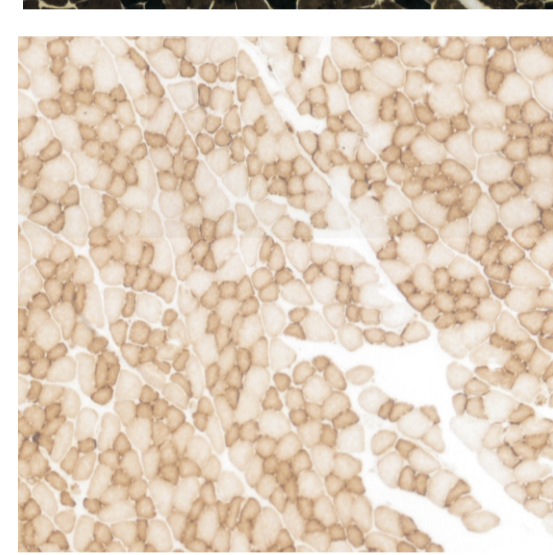
• La tinción NADH-TR permite la clasificación de las fibras musculares en 2 tipos, según su metabolismo. Las fibras de tono oscuro corresponden a fibras con metabolismo oxidativo y las de tono claro a fibras glucolíticas.



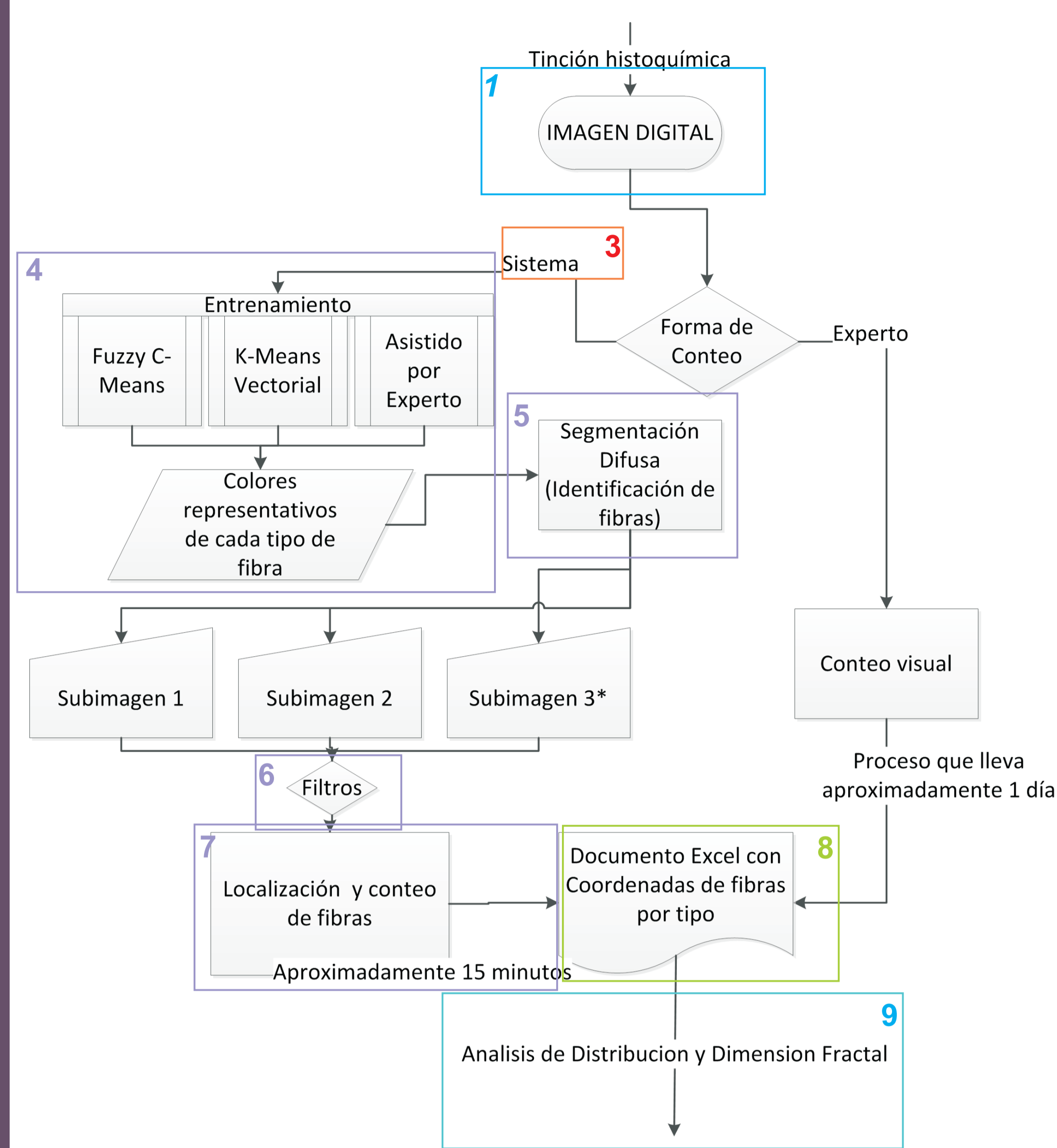
• La tinción ATPasa básica (pH 9.4) permite identificar a las fibras musculares en tres grupos: Rápidas (tono gris), Intermedias (tono oscuro) y Lentas (tono claro).



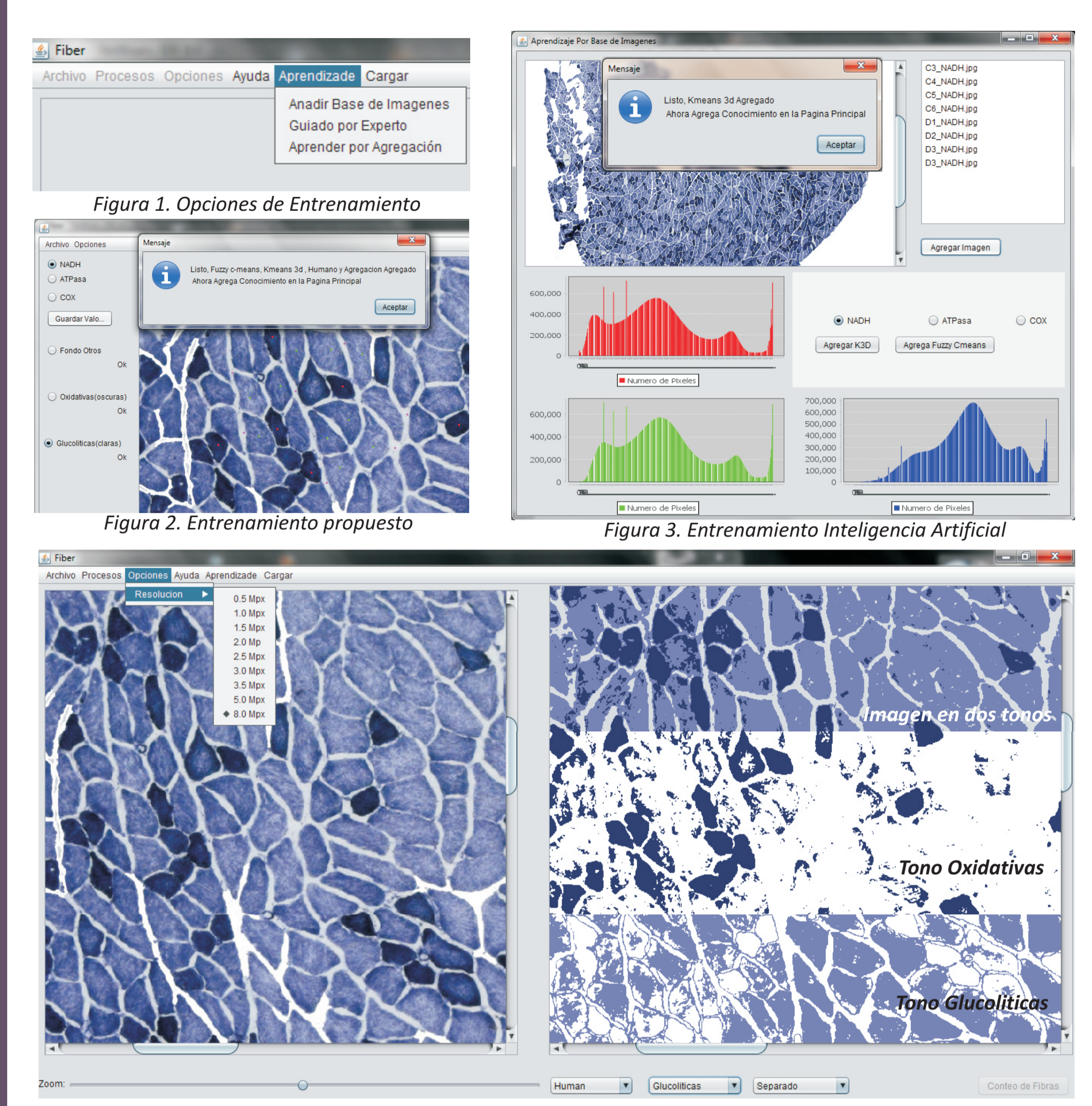
• La tinción COX, al igual que la NADH-TR diferencia a las fibras según su metabolismo en Oxidativas o Glucolíticas



## 2 Diagrama de flujo del proceso



## 3 Presentación de la Interfaz de Usuario



Los algoritmos fueron implementados en el lenguaje de programación Java debido a que es un lenguaje multiplataforma, es decir es posible ejecutarlo en los sistemas operativos más populares.

En la primera etapa se utilizan algoritmos de minería de datos para localizar los valores de referencia así como los umbrales para la separación de la imagen bajo el supuesto de que los píxeles de fibras de distinto tipo se clasifiquen en distintos grupos.

**Fuzzy C-Means**  
Es la versión difusa de K-means, se reemplaza la distancia por una medida de similitud, en forma de función de pertenencia un dato dado a cada uno de los clústeres (fórmula 1), la cual tiene la propiedad de estar siempre entre cero y uno y además la suma de la pertenencia de un dato a todos los clústeres es uno. Los centroides se actualizan según la fórmula 2.

$$\mu_i(x) = \frac{1}{\sum_{j=0}^k \left( \frac{\|x - c_j\|}{\|x - c_i\|} \right)^{\frac{2}{m-1}}}$$

Formula 1: Pertenencia del dato x al cluster i

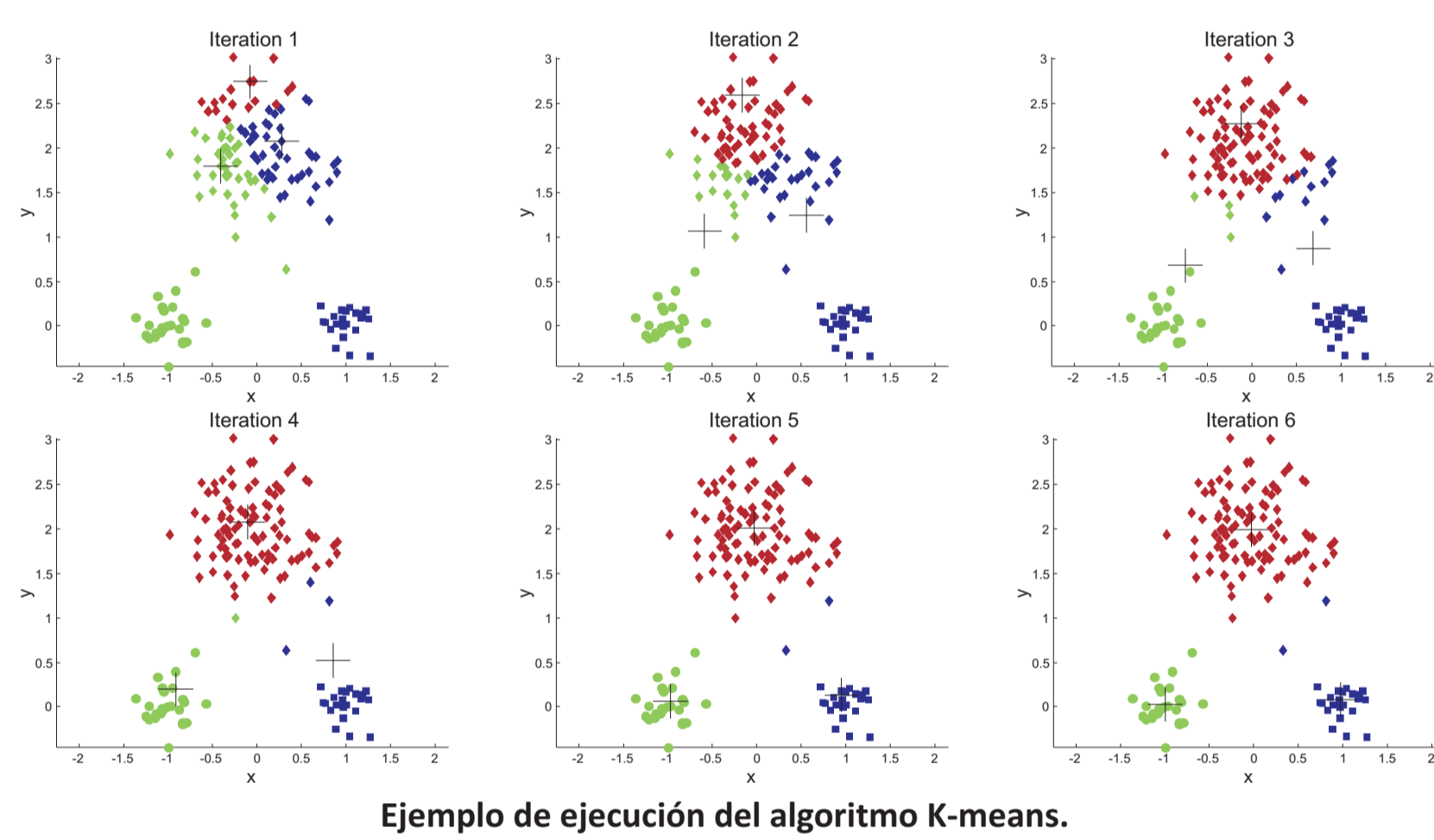
$$c_j = \frac{\sum_{x \in X} \mu_j(x)^m \cdot x}{\sum_{x \in X} \mu_j(x)^m}$$

Formula 2: Cálculo del nuevo centroide del cluster j.

**Propuesta: Asistido por un experto**  
El algoritmo K-means es de tipo isodato, es decir está basado en semejanza estricta de los datos. En la práctica algunas imágenes presentan objetos similares (desde el punto de vista de minería de datos) que sin embargo biológicamente pertenecen a fibras distintas, lo que genera una agrupación imprecisa pues datos distintos son colocados en un mismo clúster. El presente software permite al usuario indicar manualmente los valores de referencia de cada fibra, seleccionando 10 datos por tipo de fibra para calcular los valores de referencia de los clústeres. Con dichos valores se lleva a cabo el proceso de segmentación.

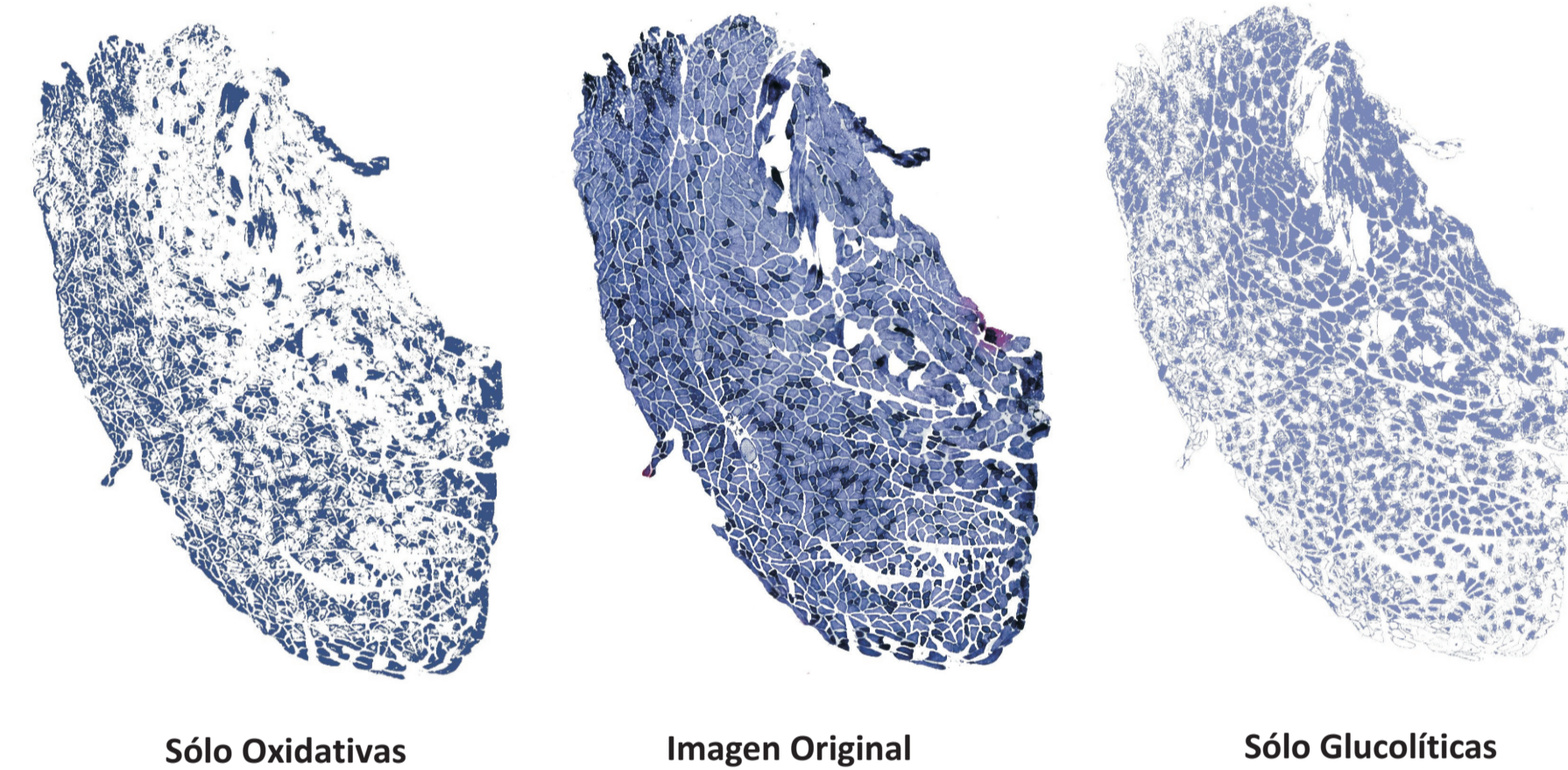
## 4 Métodos de Clasificación

**K-Means Vectorial**  
Es un algoritmo de agrupamiento por particiones, el número de clústeres es conocido K, cada clúster tiene asociado un centroide. Los datos (colores de píxeles) se asignan al clúster cuyo centroide está más cerca. Iterativamente, se van actualizando los centroides de cada uno de los clústeres reemplazándolos por el promedio, hasta que los centroides dejen de cambiar.

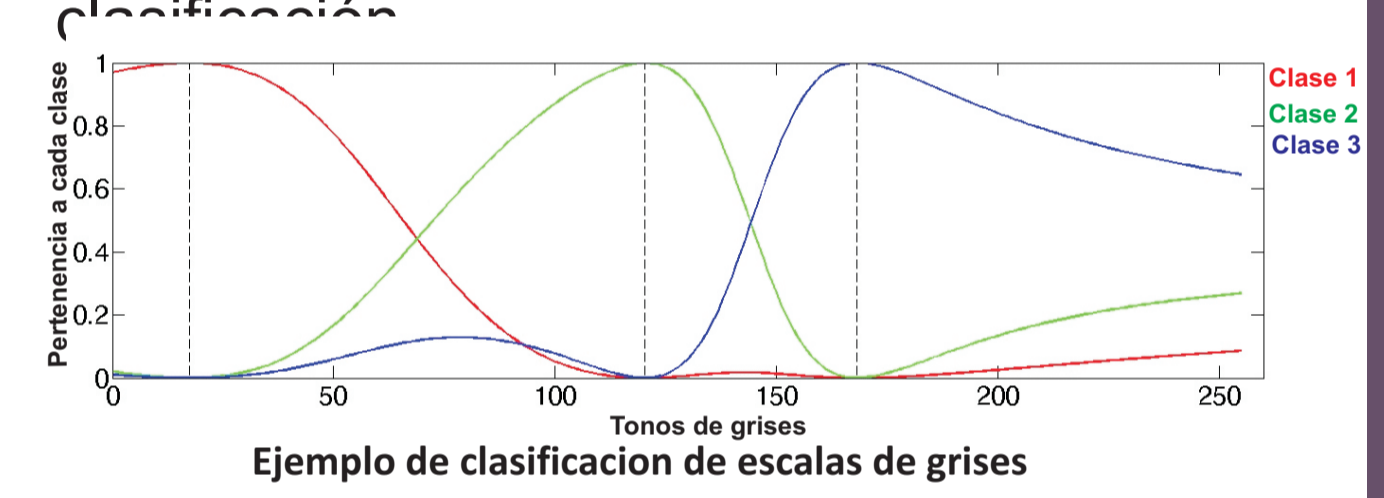


Ejemplo de ejecución del algoritmo K-means.

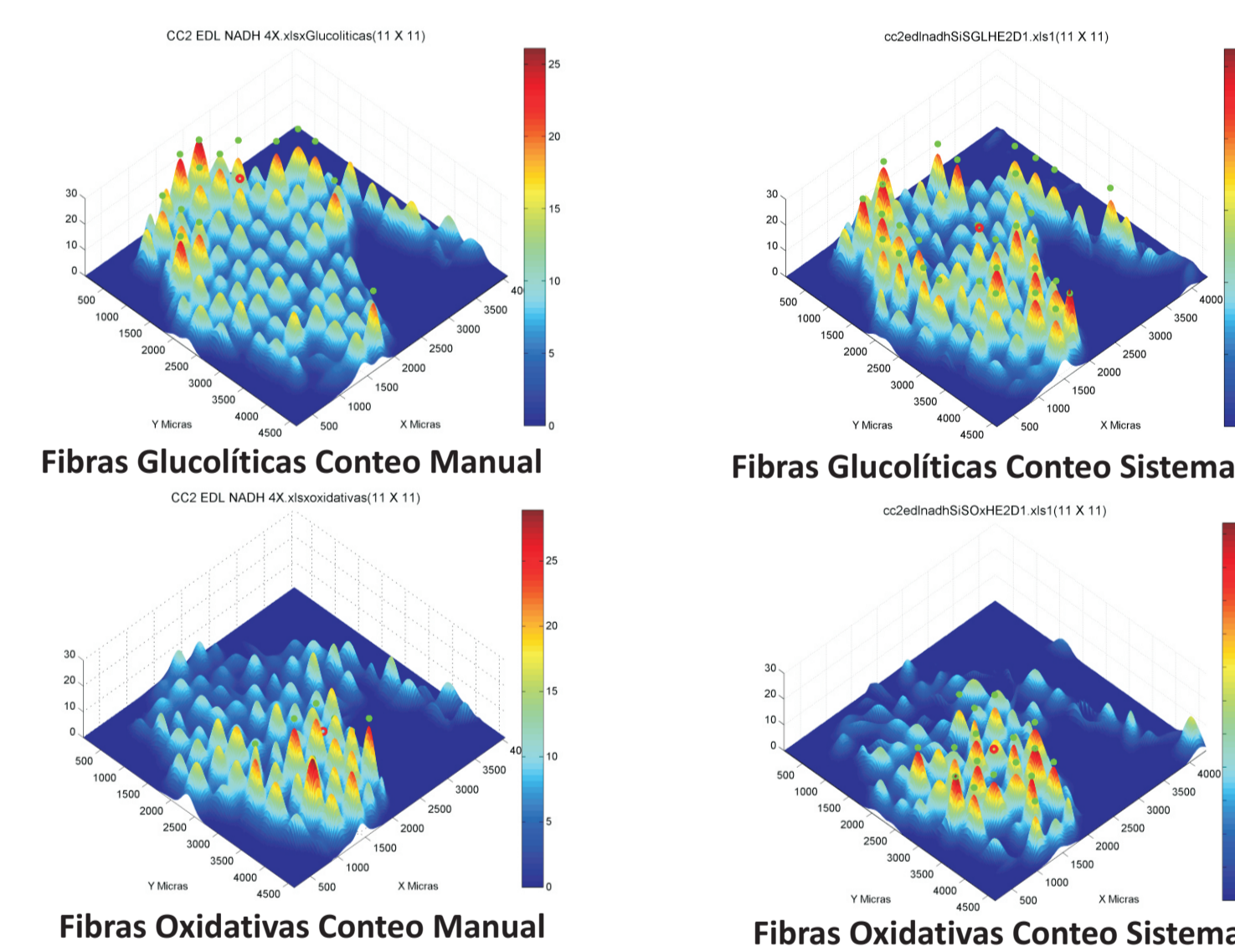
## 5 Segmentación de Imágenes



Se usa la función de pertenencia difusa de cada pixel a los clústeres definidos en el paso anterior para determinar a cual clúster pertenece y así generar subimágenes en las cuales realizar el conteo de cada tipo de fibra. Se considera que si un dato tiene una pertenencia a un clúster mayor que 0.45 entonces pertenece a este, dado que los datos varían en un continuo de valores, lo que permite una mejor clasificación.



## Determinación de la distribución de las fibras.



## 8 Resultados

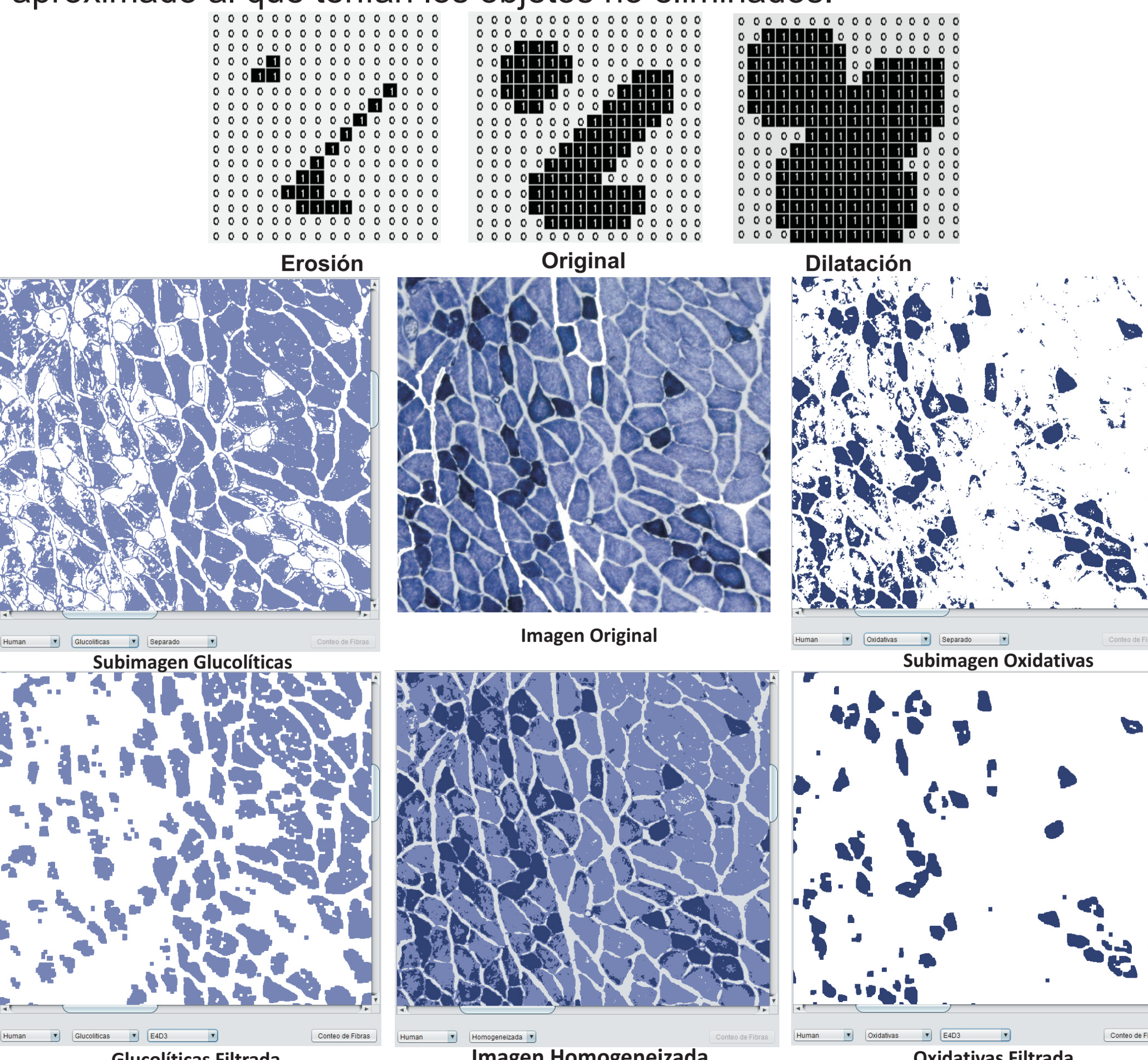
Imagen	Metodo	Tipo Fibra	Conteo	Porcentajes	Dim Fractal
C67 mPc NADH	Experto	Oxidativas	1467	49.327505	1.847
	Sistema	Oxidativas	1261	41.8936877	1.78
	Experto	Glucolíticas	1507	50.672495	1.847
	Sistema	Glucolíticas	1749	58.1063123	1.87
C66 mPc NADH	Experto	Oxidativas	1068	36.0688956	1.806
	Sistema	Oxidativas	1672	45.3362256	1.794
	Experto	Glucolíticas	1893	63.9311044	1.827
	Sistema	Glucolíticas	2016	54.6637744	1.843
Primipara	Experto	Lentas	7	0.37614186	1.569
	Sistema	Lentas	40	2.42718447	1.484
CA EDL ATP	Experto	Intermedias	737	39.6023643	1.791
	Sistema	Intermedias	927	56.25	1.831
	Experto	Rápidas	1117	60.0214938	1.81
	Sistema	Rápidas	681	41.3228155	1.764

Tabla comparativa de resultados

## 6 Aplicación de filtros personalizados

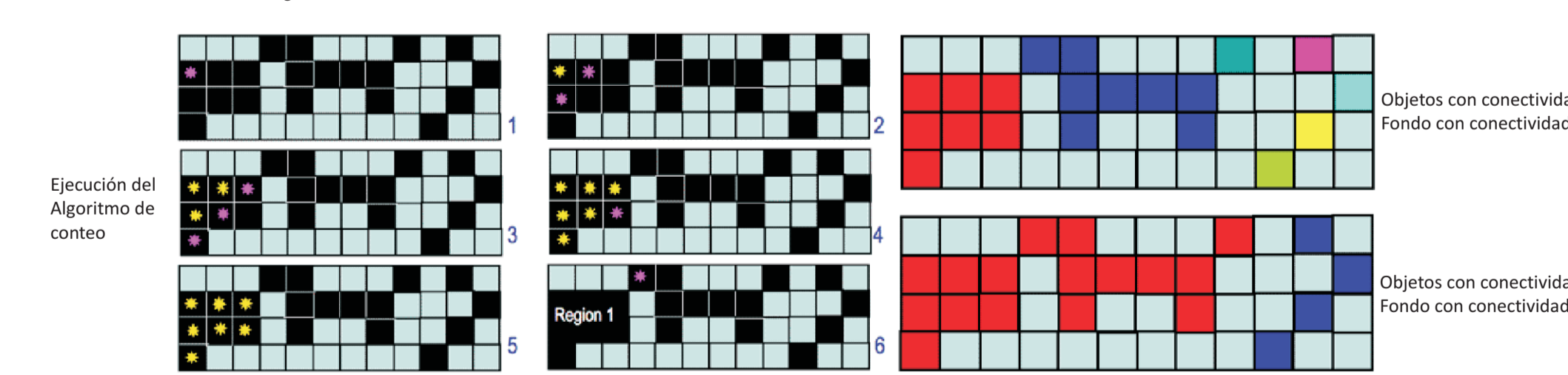
### Aplicación del filtro apropiado

De acuerdo con el criterio del experto, se aplica un filtro para eliminar la mayor cantidad de ruido pero sin eliminar las fibras menos definidas y más pequeñas. Se usa el filtro de erosión para eliminar objetos muy pequeños en la imagen y la dilatación para regresar a un tamaño aproximado al que tenían los objetos no eliminados.

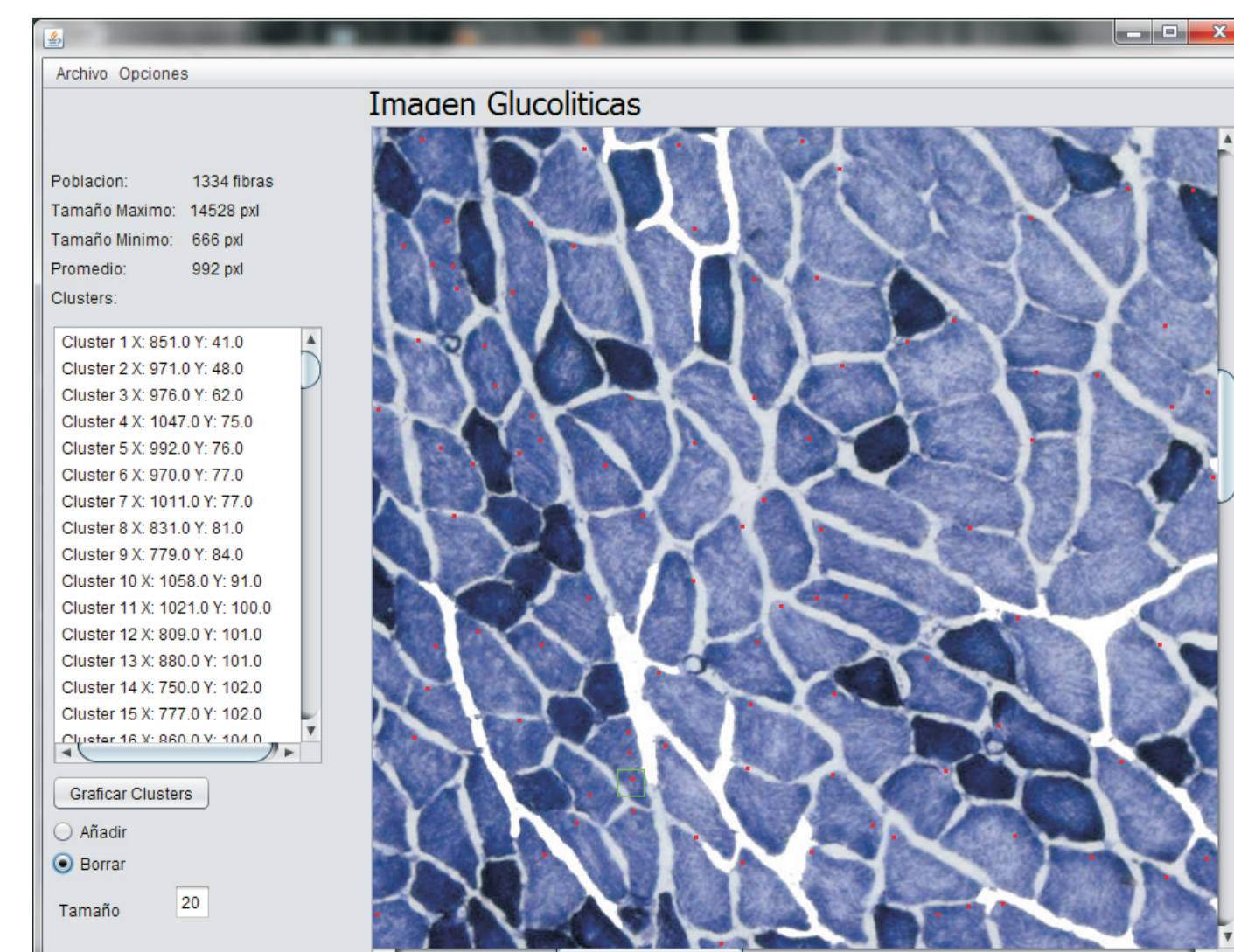


## 7 Conteo de Fibras

En cada una de las imágenes separadas se procede al conteo de fibras siguiendo el supuesto que cada cúmulo de píxeles del color seleccionado corresponde a una fibra del tipo requerido. Después, prosigue un etiquetado de componentes conexas con conectividad 8 para la localización y conteo de las fibras musculares.



**Limitaciones**  
Como se puede observar en la figura de conteo de fibras, el algoritmo de conteo de cúmulos, que se espera que corresponda uno a cada fibra, arroja un conteo mayor dado que la segmentación no es perfecta dado que las imágenes no son completamente homogéneas.



Por otra parte, en algunas ocasiones los cortes teñidos presentan dobleces o no están apropiadamente teñidos que en algunos casos se presentan como regiones de tono oscuro.

## 9 Conclusiones

- Este sistema permite determinar el número y la posición relativa de cada una de las fibras presentes en los músculos esqueléticos, así como su identificación de acuerdo a sus características metabólicas por tipo de ATPasa.
- La eficiencia del sistema es de aproximadamente un 95%.
- El tiempo de procesamiento de las imágenes histológicas se reduce hasta en un 90%.
- El sistema permite exportar los datos en hojas de cálculo para ser usados en diferentes tipos de análisis (Distribución, Reconstrucción de imágenes binarias, etc)
- Ofrece una interfaz amable con el usuario, familiarizado con un entorno de escritorio.

### Propuestas:

Se pretende integrar y comparar el desempeño de diferentes algoritmos de clasificación, que se emplearán para la etapa de minería de datos aplicada al color con los que se determinarán los colores de referencia para la separación. Además se deben incluir criterios de clasificación de píxeles que tomen en cuenta la vecindad de cada pixel, en un proceso de eliminación de ruido. Por otra parte se planea crear un algoritmo que permita determinar el tamaño de cada una de las fibras y discriminar los datos atípicos generados en la etapa de adquisición mediante un análisis

El método de Funciones de Distribución empleado para visualizar la distribución de las fibras fue sugerido por el Dr. Jorge Viveros Rogel del CIMA, UAEH.  
Estudio apoyado parcialmente por: CONACyT 265483 beca posdoctoral a KLG, SNI a ERT, KLG e IJE, donativo PROMEP-1006 a ERT y Premio Sofia Kovalévskaya a ERT.