

**UNIVERSIDAD POLITÉCNICA SALESIANA**

**SEDE CUENCA**

**CARRERA DE: INGENIERÍA DE SISTEMAS**

**TRABAJO DE TITULACIÓN PREVIO  
A LA OBTENCIÓN DEL TÍTULO DE  
INGENIERO DE SISTEMAS**

**PROYECTO TÉCNICO:  
DISEÑO Y DESARROLLO DE UN SISTEMA  
INTELIGENTE PARA LA SEGMENTACIÓN Y  
EXTRACCIÓN DE INFORMACIÓN DE IMÁGENES  
DE BIOMATERIALES OBTENIDAS A TRAVÉS DE  
MICROSCOPIA ELECTRÓNICA**

**Autor:**

Daniel Eduardo Calle López

**Tutor:**

Ing. Vladimir Espartaco Robles Bykbaev

Cuenca – Ecuador  
2018

## **CESIÓN DE DERECHOS DE AUTOR**

Yo, Daniel Eduardo Calle López con C.I.: 010574421-3, manifiesto mi voluntad y cedo a la Universidad Politécnica Salesiana, la titularidad sobre los derechos patrimoniales en virtud de que soy autor del trabajo de titulación “DISEÑO Y DESARROLLO DE UN SISTEMA INTELIGENTE PARA LA SEGMENTACIÓN Y EXTRACCIÓN DE INFORMACIÓN DE IMÁGENES DE BIOMATERIALES OBTENIDAS A TRAVÉS DE MICROSCOPIA ELECTRÓNICA”, mismo que ha sido desarrollado para optar por el título de Ingeniero de Sistemas, en la Universidad Politécnica Salesiana, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente.

En aplicación a lo determinado en la Ley de Propiedad Intelectual, en mi condición de autor, me reservo los derechos morales de la obra antes citada. En concordancia, suscribo este documento en el momento que hago entrega del trabajo final en formato impreso y digital a la Biblioteca de la Universidad Politécnica Salesiana.

Cuenca, Diciembre 2018

A handwritten signature in black ink, appearing to read 'D. Calle', with a vertical line to the left and a small 'c' at the end.

Daniel Eduardo Calle López

010574421-3

# CERTIFICACIÓN

Yo, declaro que bajo mi tutoría fue desarrollado el trabajo de titulación: “DISEÑO Y DESARROLLO DE UN SISTEMA INTELIGENTE PARA LA SEGMENTACIÓN Y EXTRACCIÓN DE INFORMACIÓN DE IMÁGENES DE BIOMATERIALES OBTENIDAS A TRAVÉS DE MICROSCOPIA ELECTRÓNICA”, realizado por el autor Daniel Eduardo Calle López, obteniendo el proyecto técnico, que cumple con todos los requisitos estipulados por la Universidad Politécnica Salesiana.

Cuenca, Diciembre 2018



Vladimir Espartaco Robles Bykbaev

Tutor del trabajo de titulación

## **DECLARATORIA DE RESPONSABILIDAD**

Yo, Daniel Eduardo Calle López con C.I.: 010574421-3, autor del trabajo de titulación “DISEÑO Y DESARROLLO DE UN SISTEMA INTELIGENTE PARA LA SEGMENTACIÓN Y EXTRACCIÓN DE INFORMACIÓN DE IMÁGENES DE BIOMATERIALES OBTENIDAS A TRAVÉS DE MICROSCOPIA ELECTRÓNICA”. Certifico que el total contenido de este proyecto técnico es de mi exclusiva responsabilidad y autoría.

Cuenca, Diciembre 2018

A handwritten signature in black ink, appearing to read 'D. Calle', with a stylized flourish above the name.

Daniel Eduardo Calle López

010574421-3

## **Agradecimientos**

Quisiera expresar mi más sincero agradecimiento a mi tutor de proyecto, Ing. Vladimir Robles; quien ha sido un excelente guía durante el desarrollo de este proyecto, apoyándome durante todas las etapas del mismo, mostrando siempre toda la predisposición para solucionar los problemas y minucias que se fueran presentando durante el desarrollo de esta tesis. También, quisiera agradecer al Instituto de Investigación Biomédica de A Coruña, por haber facilitado material indispensable para el desarrollo de este trabajo.

Finalmente, extendiendo mis agradecimientos para todas las personas que de una u otra manera intervinieron en el proceso de formación académica durante toda mi trayectoria universitaria.

## **Dedicatoria**

Esta tesis va dedicada a mi, que de no haber procrastinado tanto no hubiera necesitado prórroga.

## **Abstract**

This type of analysis based on the observation of images does not demonstrate through robust techniques the quantification of their properties. These procedures, depending mainly on the criteria and expertise of the professional on their visual interpretation of the processes related to the analysis of biomaterials, involve a great time consumption and are subject to errors since the criterion is almost exclusively subjective.

In response to the limitations presented, it is necessary to implement support systems for the segmentation and analysis of the images supported by the criteria and feedback that the professional considers given the needs of the research.

The segmentation process is one of the most critical stages in the field of image analysis and processing. Fundamentally, since it affects all subsequent processes such as the measurement of attributes and even higher-level tasks such as the classification of objects. Then, the process of segmentation of images ends up being the most crucial stage of the analysis procedure, since an incorrect identification or measurement implies that the subsequent statistical analyzes do not conform to reality and therefore, will lead to erroneous or inaccurate conclusions.==

These image segmentation processes are commonly based on color filtering, intensity, edge detection in groups of pixels and even specific optimization models. These, to be effective, require that the region of interest within the samples obtained have a clear differentiation in colors, edges and intensity. For cases where these segmentation techniques or

derivative procedures are applied, they are incomplete, since the properties of the classes to be identified are not of a single type and require a richer identification. It is necessary to consider that humans have superior knowledge to perform manual segmentations, since the fact of segmenting images under the criterion and direct observations of the professional about key elements of the image will allow refining the classification to generate probability maps with minimum error.

# Índice general

<b>Índice de figuras</b>	<b>XIII</b>
<b>Índice de cuadros</b>	<b>XV</b>
<b>1. Resumen</b>	<b>1</b>
<b>2. Introducción</b>	<b>3</b>
<b>3. Objetivos</b>	<b>5</b>
3.1. General . . . . .	5
3.2. Específicos . . . . .	5
<b>4. Fundamentos y trabajo relacionado</b>	<b>7</b>
4.1. Biomateriales . . . . .	7
4.2. Microscopía electrónica . . . . .	8
4.3. Procesamiento digital de imágenes . . . . .	9
4.3.1. Imágenes digitales . . . . .	9
4.3.1.1. Tipos de imágenes . . . . .	10

<i>ÍNDICE GENERAL</i>	x
4.4. Métodos de Segmentación de imágenes . . . . .	12
4.4.1. Clasificadores . . . . .	12
4.4.2. Región en crecimiento . . . . .	12
4.4.3. Clustering . . . . .	13
4.4.4. Redes neuronales . . . . .	14
4.4.5. Modelos deformables . . . . .	15
4.5. Tipos de Clasificadores . . . . .	16
4.5.1. J48 . . . . .	16
4.5.2. LMT . . . . .	16
4.5.3. Bosques aleatorios . . . . .	16
4.5.4. Decision Stump (DS) . . . . .	17
4.5.5. Hoeffding tree . . . . .	17
<b>5. Diseño y desarrollo del sistema inteligente</b>	<b>18</b>
5.1. Requerimientos para el desarrollo del sistema . . . . .	18
5.1.1. Identificación de clases segmentables dentro de biomateriales . . . . .	19
5.2. Especificación de plataforma de desarrollo . . . . .	20
5.3. Diseño y construcción de arquitectura . . . . .	20
5.4. Especificaciones técnicas del sistema . . . . .	21
5.4.1. Especificaciones de hardware . . . . .	22
5.4.2. Especificaciones de software . . . . .	22
5.5. Diagramas . . . . .	23

<i>ÍNDICE GENERAL</i>	XI
5.5.1. Diagrama de secuencias . . . . .	23
5.5.2. Diagrama de clases . . . . .	26
5.6. Diseño e implementación del plugin . . . . .	26
5.6.1. Pre-procesamiento de las imágenes . . . . .	26
5.6.2. Extracción de atributos . . . . .	27
5.6.2.1. Capas HSV . . . . .	28
5.6.2.2. Imagen original BN . . . . .	30
5.6.2.3. Tensor de estructura . . . . .	31
5.6.2.4. Entropía . . . . .	32
5.6.2.5. Desenfoque Gaussiano . . . . .	33
5.6.2.6. Operador Sobel . . . . .	33
5.6.3. Fase de entrenamiento . . . . .	37
5.6.4. Fase de clasificación . . . . .	37
5.6.5. Extracción de atributos y resultados . . . . .	41
<b>6. Experimentación, modelos y resultados</b>	<b>44</b>
6.1. Preparación de corpus de imágenes . . . . .	44
6.2. Aproximaciones alternativas . . . . .	45
6.3. Selección de modelo de clasificación . . . . .	45
6.3.0.1. Modelo logístico de árboles LMT . . . . .	46
6.3.0.2. Modelo de árbol Hoeffding . . . . .	47
6.3.0.3. Modelo de árbol J48 . . . . .	48

<i>ÍNDICE GENERAL</i>	XII
6.3.0.4. Modelo de árbol REPTree . . . . .	49
6.3.0.5. Modelo de reglas con tabla de decisión . . . . .	50
6.3.0.6. Modelo de función: SMO . . . . .	52
6.4. Resultados . . . . .	53
<b>Conclusiones</b>	<b>57</b>
<b>Recomendaciones</b>	<b>59</b>
<b>Trabajo futuro</b>	<b>60</b>
<b>Bibliografía</b>	<b>61</b>

# Índice de figuras

4.1. Ejemplo de imagen raster . . . . .	11
5.1. Formulación de biopolímeros a base de PLA y ácido glicólico . . . . .	19
5.2. Arquitectura general del sistema . . . . .	21
5.3. Secuencia de la fase de entrenamiento . . . . .	24
5.4. Secuencia de la fase de clasificación . . . . .	25
5.5. Diagrama de clases . . . . .	26
5.6. Interfaz de selección de atributos . . . . .	27
5.7. Imagen base para ejemplificación . . . . .	28
5.8. Capa de color . . . . .	29
5.9. Capa de saturación . . . . .	29
5.10. Capa de valor o brillo . . . . .	30
5.11. Capa base en escala de grises . . . . .	31
5.12. Tensor de estructura . . . . .	32
5.13. Desenfoque Gaussiano . . . . .	33
5.14. Operador de sobel . . . . .	34

5.15. Interfaz para la fase de clasificación . . . . .	38
5.16. Imagen original para calcular resultados . . . . .	38
5.17. Mapa de probabilidad de colágeno . . . . .	39
5.18. Mapa de probabilidad de matriz . . . . .	39
5.19. Mapa de probabilidad de núcleos . . . . .	40
5.20. Mapa de probabilidad de espacio . . . . .	40
5.21. Imagen resultante segmentada . . . . .	41
6.1. Comparativa de resultados . . . . .	55

# Índice de cuadros

5.1. Atributos extraídos de los segmentos resultantes . . . . .	42
6.1. Sumario de modelo LMT . . . . .	46
6.2. Matriz de confusión modelo LMT . . . . .	47
6.3. Sumario de modelo Hoeffding . . . . .	48
6.4. Matriz de confusión modelo Hoeffding . . . . .	48
6.5. Sumario de modelo J48 . . . . .	49
6.6. Matriz de confusión modelo J48 . . . . .	49
6.7. Sumario de modelo REPTree . . . . .	50
6.8. Matriz de confusión modelo REPTree . . . . .	50
6.9. Sumario de modelo DecisionTable . . . . .	51
6.10. Matriz de confusión modelo DecisionTable . . . . .	51
6.11. Sumario de modelo SMO . . . . .	52
6.12. Matriz de confusión modelo SMO . . . . .	53
6.13. Resumen de experimentación . . . . .	54

# Capítulo 1

## Resumen

El tipo de análisis basado en la observación de imágenes no demuestran mediante técnicas robustas la cuantificación de sus propiedades. Estos procedimientos al depender principalmente del criterio y la experticia del profesional sobre su interpretación visual de los procesos relacionados con el análisis de biomateriales implican un gran consumo de tiempo y están sujetos a errores ya que el criterio es casi exclusivamente de tipo subjetivo.

En respuesta de las limitaciones presentadas, es necesaria la implementación de sistemas de soporte para la segmentación y análisis de las imágenes respaldado en el criterio y feedback que el profesional considere dadas las necesidades de la investigación.

El proceso de segmentación es una de las etapas más críticas en el campo del análisis y procesamiento de imágenes. Fundamentalmente, ya que éste afecta todos los procesos subsecuentes como la medición de atributos e incluso tareas de nivel superior como la clasificación de objetos. Entonces, el proceso de segmentación de imágenes termina siendo la etapa más crucial del procedimiento de análisis, ya que una identificación o medición errada implica que los análisis estadísticos subsecuentes no se apegan a la realidad y por ende, llevarán a tomar conclusiones erróneas o poco acertadas.

“Estos procesos de segmentación de imágenes son comúnmente basados en filtrado de colores, intensidad, detección de bordes en grupos de píxeles e incluso modelos específicos de optimización. Éstos, para ser efectivos, requieren que la región de interés dentro de las muestras obtenidas posea una clara diferenciación en colores, bordes e intensidad”=. Para los casos donde se apliquen estas técnicas de segmentación o procedimientos derivados, resultan incompletos, ya que las propiedades de las clases que se desean identificar no son de un único tipo y requieren una identificación más rica. Es necesario considerar que los humanos poseen un conocimiento superior para realizar segmentaciones manuales, pues el hecho de segmentar imágenes bajo el criterio y observaciones directas del profesional acerca de elementos clave de la imagen, permitirán afinar la clasificación a generar mapas de probabilidad con mínimo error.

# Capítulo 2

## Introducción

Actualmente, la visión por computador nos permite determinar el comportamiento biológico de las células, como el crecimiento y diferenciación celular así como la degradación de biomateriales (colágeno tipo I). Bajo esta línea, la segmentación de imágenes es una metodología muy importante utilizada para alcanzar este objetivo, “Es un problema central en muchos estudios, y a menudo considerado como la piedra angular del análisis de imágenes, es su segmentación” [21].

Aunque la segmentación es conceptualmente simple, “carece de generalidad, y por lo tanto no se puede implementar de manera confiable y sin esfuerzo en todas las líneas celulares sin antes haber pre-procesado las imágenes” [21].

La ausencia de un procedimiento de segmentación universal no es una sorpresa, sin embargo, habrá que aprovechar las facilidades que tenemos hoy en día en capacidades de procesamiento así como nuevas metodologías e invenciones para solventar en cierto grado estos problemas, por tanto en el presente trabajo se brindará una solución para hacer posible el análisis de fotografías sobre imágenes de biomateriales. Dando apertura a estudios tales como

el análisis de degradación de materiales, crecimiento secuencial y diferenciación celular, entre otros.

# Capítulo 3

## Objetivos

### 3.1. General

Diseñar y desarrollar de un sistema inteligente para la segmentación y extracción de información de imágenes de biomateriales obtenidas a través de microscopía electrónica

### 3.2. Específicos

- **OE1.** Investigar y conocer las principales características de imágenes de biomateriales obtenidas por microscopía electrónica.
- **OE2.** Estudiar las principales técnicas de segmentación y análisis de características morfológicas de imágenes de biomateriales
- **OE3.** Ensamblar un corpus de imágenes de biomateriales obtenidas por microscopía electrónica y establecer las clases de biomateriales que se segmentarán.

- **OE4.** Diseñar y desarrollar un conjunto de módulos basados en técnicas de visión por computador y aprendizaje de máquina a fin de realizar las siguientes tareas:
  - Pre-procesamiento de las imágenes.
  - Segmentación en base a las clases de biomateriales establecidas.
  - Extracción de atributos para realizar el entrenamiento del clasificador.
  - Entrenar el clasificador que permitirá realizar la identificación de clases.
  - Extracción de atributos correspondientes a cada clase a fin de realizar análisis posteriores.
  
- **OE5.** Diseñar y ejecutar un plan de experimentación a fin de validar la propuesta desarrollada

# Capítulo 4

## Imágenes de biomateriales:

## Fundamentos y trabajo relacionado

### 4.1. Biomateriales

Los biomateriales se reconocen más usualmente como una estructura, que son capaces de realizar funciones potencialmente útiles como:

- Promover la unión, supervivencia, proliferación y diferenciación de las células mientras que poseen toxicidad mínima en las formas originales.
- Permitir el transporte o suministro de gases, nutrientes y factores de crecimiento.
- Ofrece suficiente soporte estructural mientras sea degradable/absorbible a tasas apropiadas para la regeneración de tejidos.

La utilidad de la mayoría de los biomateriales depende mucho de los comportamientos de degradación y absorción, ya que dichos comportamientos pueden afectar una gama de sucesos,

tales como crecimiento celular, regeneración tisular, liberación del fármaco, respuesta del huésped y función del material [6].

Existen biomateriales degradables que son utilizados para la construcción de implantes médicos que proporcionan eventos clínicos como la curación, dichos implantes sirven para reemplazar una función o parte del cuerpo de una manera confiable y segura [2]. “Los biomateriales se pueden definir como materiales naturales o sintéticos diseñados para interactuar con sistemas biológicos que se utilizan para el tratamiento médico.”

La gran variedad de biomateriales existentes hoy en día tienen un gran impacto dentro de las tecnologías de materiales [17].

- Los biomateriales poliméricos ofrecen una ventaja principal sobre los metales y las cerámicas en cuanto a su facilidad de fabricación para formar varias formas
- Los biomateriales de cerámica brindan inercia, alta resistencia a la compresión y apariencia estética
- Los biomateriales compuestos se pueden hacer con metales, polímeros o cerámicas como matriz, y se pueden reforzar con uno de estos materiales

## 4.2. Microscopía electrónica

La microscopía ha ido de la mano con el estudio de los sistemas vivos, un espectro de tintes y sondas ahora permite la localización de moléculas de interés dentro de células vivas mediante microscopía de fluorescencia. Con la Microscopía Electrónica (*EM*), se ha revelado la ultra-estructura celular. Superando estas dos modalidades, la Microscopía de Luz Correlacionada y *EM* (*CLEM*) abre nuevas vías. Los estudios de dinámica de proteínas con Proteínas Fluorescentes (*FP*), que dejan al investigador “en la oscuridad” en relación con

el contexto celular, pueden ser seguidos por un examen EM. Los eventos raros se pueden preseleccionar en el nivel de microscopía óptica antes del análisis EM. El desarrollo continuo, que incluye sondas dedicadas, microscopios integrados, microscopía de fluorescencia de supersolución y EM a gran escala y tridimensional.

La Microscopía de Fluorescencia (FM) permite a los investigadores identificar moléculas específicas y estudiar sus funciones biológicas. Sin embargo, debido a que una gran parte de las moléculas permanecen sin etiquetar y, por lo tanto, “en la oscuridad”, el contexto de la localización se pierde. Además, la resolución de la Microscopía de Luz (*LM*) suele ser submicrométrica y, por lo tanto, no coincide con el tamaño de las biomoléculas, que suelen oscilar entre 0,1 y 10 nm. La forma de analizar las moléculas tanto en su contexto biológico como en alta resolución es a través de la microscopía electrónica. Sin embargo, con la microscopía electrónica, el análisis ultraestructural se realiza en imágenes en escala de grises, en las cuales las moléculas son difíciles de definir; las muestras biológicas están en un estado fijo; y encontrar eventos raros en el espacio y el tiempo es casi imposible [4].

La resolución ahora no solo se ajusta mejor entre las dos modalidades, sino que el análisis de microscopía electrónica ahora también se puede realizar en volúmenes más grandes.

## **4.3. Procesamiento digital de imágenes**

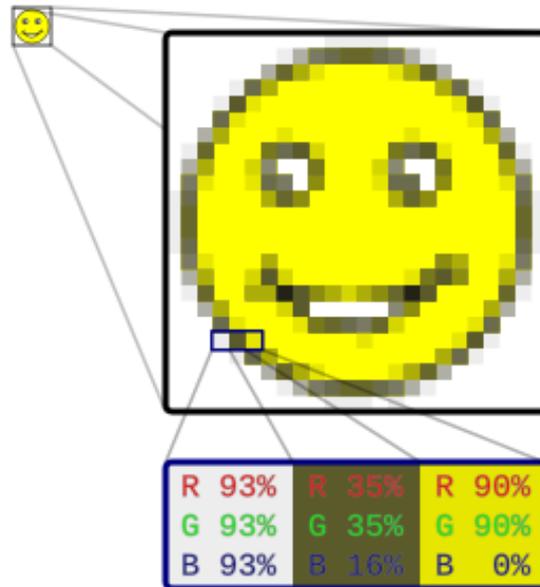
### **4.3.1. Imágenes digitales**

Es necesario entender primeramente ¿Qué es una imagen digital?. Se define una imagen digital como una representación numérica, generalmente binaria de una imagen de dos dimensiones.

#### 4.3.1.1. Tipos de imágenes

**Raster** Una imagen raster está formada por una matriz de celdas (cada celda representa un píxel) organizadas en filas y columnas, cada una de las celdas contiene cierta información acerca de la imagen en dicha posición. Una imagen puede tener varias matrices de celdas, representando varias capas de información. Generalmente, las imágenes se componen de las 3 capas de color básicas RGB (Rojo, Verde y Azul), aunque pueden representar también fenómenos de temperatura, elevación o datos espectrales.

El área (o superficie) representada por cada celda contiene el mismo ancho y altura, y es una parte equivalente a toda la superficie representada por el ráster [5]. Dicho eso, dependiendo de la escala de los elementos que se requieran analizar, se necesitará que cada píxel represente más o menos área de la realidad; entre más pequeño sea el tamaño representado por cada píxel, más detallada será la imagen en cuestión. Si el área representada por la celda resulta demasiado grande, se podría llegar a perder información importante de la imagen raster, dando lugar a que no se pueda delimitar bien el espacio representado por cada entidad dentro de la imagen.



**Figura 4.1** Imagen “smiley face” agrandada, donde se pueden apreciar los píxeles que la conforman, de [3]. Al acercarse, los píxeles individuales aparecen como cuadrados, se puede apreciar su composición al combinar los valores para éste píxel de las capas Rojo, Verde y Azul que lo conforman.

Por ejemplo: Si tuviéramos una imagen de microscopía electrónica que cubra un área de 100x100 micras, suponiendo que la imagen raster se componga de 100x100 celdas; cada celda representaría un área de 1x1 micra en la realidad.

**Vectoriales** Este tipo de imágenes se encuentran conformadas principalmente de formas geométricas (segmentos de línea, arcos, círculos, polígonos, etc), cada elemento asimismo tiene propiedades asignadas, definiendo su tamaño, color y propiedades específicas de su forma. Este tipo de imágenes no corresponde con los formatos obtenibles por microscopía electrónica, por tanto se descarta su uso para los intereses de este trabajo.

## 4.4. Métodos de Segmentación de imágenes

### 4.4.1. Clasificadores

Los métodos de segmentación basados en clasificadores, por lo general se utilizan cuando se pretende identificar espacios con características semejantes que componen una imagen, utilizando información extraída de dichas imágenes previamente etiquetado. Los clasificadores se conocen como métodos supervisados porque requieren datos de entrenamiento que se segmentan manualmente y luego se usan como referencias para segmentar automáticamente datos nuevos.

En el proceso de segmentación de las imágenes mediante éste método puede realizarse de distintas maneras. Uno de los ejemplos más básicos es “*k-means*”, bajo esta aproximación, se conglomeran los píxeles bajo cada capa de color con los valores más cercanos. El clasificador “*k-means*” más cercano es una generalización de este enfoque, en el que el píxel se clasifica en la misma etiqueta que los datos de entrenamiento más cercanos a  $k$ . El clasificador de vecino más cercano  $k$  se considera un clasificador no paramétrico porque no hace ninguna suposición subyacente sobre la estructura estadística de los datos. Otro clasificador no paramétrico es la ventana de Parzen, en la que la clasificación se realiza mediante un proceso de decisión ponderado dentro de una ventana predefinida del espacio de características, centrado en la intensidad de píxeles sin etiquetar.

### 4.4.2. Región en crecimiento

Esta técnica se utiliza para extraer regiones específicas de imágenes. Éstas que se conectan en función de algunos criterios predefinidos. Los criterios de selección son generalmente intensidad de color, brillo, atributos de capas, formas (bordes), entre otros. Es necesario establecer manualmente puntos iniciales desde donde se comienzan a expandir las etiquetas

hasta cubrir los criterios de selección. Por ejemplo, un posible criterio podría ser hacer crecer la región hasta que se cumpla un borde en la imagen.

Uno de los problemas principales en el uso de este método es la necesidad de establecer manualmente los puntos iniciales. Por lo tanto, para cada región que necesita ser extraída, una semilla debe ser plantada. *Splitand-merge* es un algoritmo relacionado con el crecimiento de regiones, pero no requiere un punto de semilla. Las regiones que suponen ruido en la imagen pueden afectar en gran medida el proceso de crecimiento de la región, esto puede producir que los segmentos identificados se corten o se identifiquen mal, sin haber cubierto el área que corresponden. A la inversa, el volumen parcial puede hacer que regiones separadas se conecten. Para ayudar a aliviar estos problemas, realizan un algoritmo homotópico de crecimiento de regiones que preserva la topología entre una región inicial y una región extraída.

### 4.4.3. Clustering

Los algoritmos basados en “*clustering*” tienen una gran semejanza con los métodos que ocupan modelos clasificadores, en gran medida ya que ciertas técnicas se pueden adaptar entre éstos. Se diferencian una de otra ya que el “*clustering*” es no-supervisado, es decir, no se tiene información previa con la que se puedan comparar los resultados. Es por eso que los métodos de validación de éstas van a ser diferentes; dentro de las metodologías no-supervisadas no existen datos que puedan ensamblar un corpus de pruebas. El proceso es, grosso modo, alternar varias veces el proceso de entrenamiento del sistema con el de aplicación, generando más información de entrenamiento que al final se terminará ajustando al modelo. Po supuesto, esto ha demostrado generar cierto nivel de sobre-ajuste con los modelos, pero el desempeño general es bastante efectivo.

Este tipo de clasificación no precisa tener un gran cuerpo de información inicial, sino cierta cantidad de imágenes segmentadas de donde se puedan extraer atributos de entrenamiento. El algoritmo EM ha demostrado una mayor sensibilidad a la inicialización que el algoritmo “*K-means*” o “*fuzzy c-means*”. Volviendo a la comparativa con la metodología basada en clasificación, éstas metodologías no han sido diseñadas con el procesamiento de imágenes en mente, es por eso que requieren cierto pre-procesamiento para poder ser implementadas; cabe señalar que no es precisamente una desventaja, ya que éstos algoritmos una vez cubran su funcionalidad base, pueden ser adaptados para que funcionen de la manera más óptima, ajustándose al problema en cuestión. Ésto resulta muy útil en las implementaciones menos comunes.

#### 4.4.4. Redes neuronales

Se puede realizar también segmentación de imágenes basada en “Redes Neuronales Artificiales” (*ANN*). Éstas se encuentran conformadas con unidades de procesamiento que trabajan simultáneamente, son un modelo más específico que tiene origen en redes más simples, sea el caso del “*perceptrón*”. Las unidades de procesamiento que componen estas redes, luego de cierta cantidad de iteraciones y dependiendo de el nivel de la capa en la que opere, se van especializando en detectar cierto tipo de elementos que conforman la imagen; de esta manera, los nodos de las primeras capas se dedicarán a la detección de figuras básicas como líneas, curvas, planos y colores, mientras que las capas subsecuentes comienzan a especializarse en elementos más complejos tales como polígonos y figuras propias de los elementos que se pretenda identificar.

El uso de este tipo de redes se ha popularizado de gran manera en los últimos años, esto puede deberse a la gran capacidad de procesamiento y velocidades que han adquirido los equipos en los pasados años así como por ser un campo de investigación muy popular. Este

tipo de redes así como las anteriores, no está diseñado específicamente para el proceso de identificación de elementos morfológicos en imágenes, por tanto las implementaciones que surjan de ésta serán en gran medida muy apegadas al problema de segmentación específico que se pretenda solucionar. Estas redes también se pueden utilizar como una metodología no supervisada. Debido a las muchas interconexiones utilizadas en una red neuronal, la información dentro de los campos de imágenes pueden ser reconocidas fácilmente mediante modelos bien definidos dentro de este campo. A pesar que esta metodología ocupe unidades de procesamiento que operan paralelamente, su procesamiento generalmente se simula en una computadora serial estándar, de esta manera reduciendo la potencial ventaja computacional.

#### **4.4.5. Modelos deformables**

Los modelos deformables son técnicas basadas en modelos y motivados físicamente para delinear los límites de las regiones mediante el uso de curvas definidas de manera paramétrica, éstas pueden ser cerradas o superficie. Este tipo de elementos pierden su forma bajo la efectos de agentes externos e internos. Si se pretende identificar segmentos de imágenes que representen bordes de elementos se deberá inicialmente superponer una superficie cerrada ó figura sobre el límite que se pretende segmentar para entonces someterlo a un proceso iterativo de relajación. Concluido lo anterior, se calculan nuevamente las características matemáticas del elemento que se dibujó sobre el límite [15]. Esto es necesario, ya que deberá permanecer suave mientras dure el proceso de deformación.

## **4.5. Tipos de Clasificadores**

### **4.5.1. J48**

“Un árbol de decisión J48 es un modelo predictivo de aprendizaje automático que decide el valor objetivo de una nueva muestra en función de los diferentes valores de los atributos de los datos disponibles. los diferentes atributos denotados por los nodos internos de un árbol de decisión, las ramas entre los nodos nos dicen los posibles valores que estos atributos pueden tener en las muestras experimentales, mientras que los nodos terminales nos dicen el valor final de la variable dependiente” [16].

### **4.5.2. LMT**

Un modelo de clasificación con un algoritmo de entrenamiento supervisado asociado que combina la predicción logística y el aprendizaje en árbol de decisiones es el Árbol de Modelo Logístico (LMT). Los árboles de modelos logísticos utilizan un árbol de decisión que tiene modelos de regresión lineal en sus hojas para proporcionar un modelo de regresión lineal por secciones.

### **4.5.3. Bosques aleatorios**

Es un método de aprendizaje conjunto para la clasificación, regresión y otras tareas, que funcionan mediante la construcción de una multitud de árboles de decisión en el momento del entrenamiento y la salida de la clase que es el modo de la clasificación o predicción media de los árboles individuales. “Los bosques aleatorios corrigen el hábito de los árboles de decisión de ajustarse a su conjunto de entrenamiento. Son una forma de promediar múltiples árboles de decisión profunda, entrenados en diferentes partes del mismo conjunto de entrenamiento, con

el objetivo de reducir la varianza. Esto se produce a expensas de un pequeño aumento y cierta pérdida de interpretabilidad, pero en general aumenta considerablemente la presentación del modelo final” [24].

#### 4.5.4. Decision Stump (DS)

Se utiliza para generar un árbol de decisión con una sola división. “El árbol resultante se puede usar para clasificar ejemplos invisibles. El tocón de decisión puede ser muy eficiente cuando se refuerza con operadores como el operador *AdaBoost*.” Las muestras se asignan a las diferentes clases identificadas de manera binaria, lo que es lo mismo, una unidad de imagen puede o no pertenecer a una sola clase. “Los nodos hoja de un árbol de decisión contienen el nombre de la clase, mientras que un nodo no hoja es un nodo de decisión.” El nodo de decisión es una prueba de atributo en la que cada rama del nodo es un valor viable del atributo. Los gestos de decisión pueden parecer muy simples, cuando se aumentan, producen un buen algoritmo de clasificación en la práctica [10].

#### 4.5.5. Hoeffding tree

Es una inducción incremental del árbol de decisión que es capaz de aprender de flujos de datos muy grandes, suponiendo que los ejemplos de generación de distribución no se transformen con el tiempo [16]. “Los árboles Hoeffding aprovechan el hecho de que una pequeña muestra a menudo puede ser suficiente para elegir un atributo de división óptimo. Esta idea está respaldada matemáticamente por el límite de Hoeffding, que cuantifica el número de observaciones necesarias para estimar algunas estadísticas dentro de una precisión prescrita” [8].

# **Capítulo 5**

## **Diseño y desarrollo del sistema inteligente para segmentación de imágenes de biomateriales**

### **5.1. Requerimientos para el desarrollo del sistema**

Al desarrollar el Sistema inteligente para la segmentación y extracción de información de imágenes de biomateriales, se pretende dar soporte a los profesionales del área de las ciencias de la salud, mediante técnicas de aprendizaje de máquina conforme a las observaciones y entradas convenientes para el tipo de materiales a procesar.

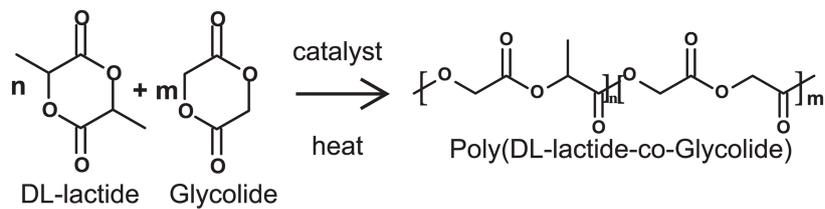
Se requiere entonces un conjunto de módulos que nos permitan segmentar los biomateriales que se especifican en la siguiente subdivisión, dicha segmentación se realizará con técnicas de aprendizaje de máquina.

Es necesario que se puedan cargar imágenes de biomateriales para que, en base a dichas entradas, se puedan producir imágenes con el mismo formato, apiladas en base a las clases

segmentables identificadas. Además de eso, se podrá generar corpus de datos a partir de los atributos de forma, posición, intensidad, excentricidad y demás atributos propios de las denominadas partículas de imagen que conformen cada elemento resultante del análisis.

### 5.1.1. Identificación de clases segmentables dentro de biomateriales

Acerca del tipo de biomateriales a segmentar, tomando en cuenta los conceptos expuestos en la sección de fundamentos teóricos será el copolímero: Poly(D,L-lactic-co-glycolic acid) 5.1, “PDLGA, un material sintético biodegradable muy importante para varios procesos biomédicos. Sus aplicaciones biomédicas son: implantes quirúrgicos, fijaciones, suturas, administración controlada de fármacos, entre otras dentro del dominio de ingeniería de tejidos” [23].



**Figura 5.1** Formulación de biopolímeros a base de PLA y ácido glicólico (Cortesía: PURAC-CORBION, 2010) [23] .

Con base en [23] se identifica que los tipos de elementos que conforman el copolímero PDLGA son:

- Colágeno (tipo I)
- Matriz extracelular
- Núcleos
- Fondo

## 5.2. Especificación de plataforma de desarrollo

El sistema se desarrolló en el lenguaje de programación Java versión 1.8. Java es un componente central en los sitios y entornos web de alta ingeniería de hoy en día. Se han desarrollado con éxito docenas de aplicaciones que aplican métodos de aprendizaje de máquina, inteligencia artificial y visión por computador basadas en Java. Se escogió esta plataforma gracias a que es un lenguaje que implementa interfaces y es común con las siguientes herramientas y/o librerías:

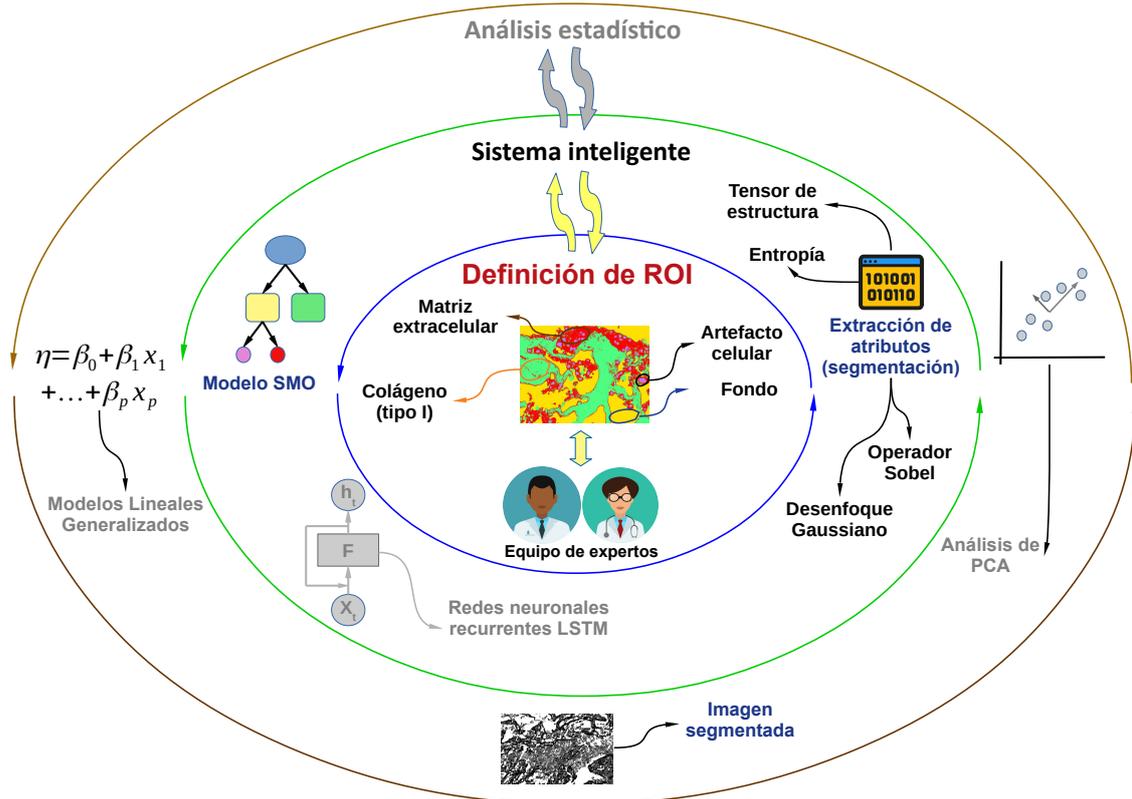
- Waikato Environment for Knowledge Analysis (WEKA)
- ImageJ
- Bio-Formats
- Fiji Is Just ImageJ (FIJI)

## 5.3. Diseño y construcción de arquitectura

Acercas del diseño, se pretende brindar módulos que permitan realizar paso a paso el proceso de segmentación de colágeno tipo-I a través del uso de imágenes previamente identificadas mediante un equipo de expertos. Habiendo ellos previamente identificado las áreas de interés de las imágenes de entrenamiento.

Con base en dichas áreas de interés, se procederá a la fase de extracción de atributos para el entrenamiento de clasificadores. En esta fase se aplicarán diferentes filtros y operaciones sobre las imágenes, dando como resultado un corpus de datos y cabeceras. Estas se utilizarán para ensamblar un corpus que pueda ser ocupado bajo las especificaciones del software *WEKA*.

Para la fase de clasificación, los atributos definidos en el corpus de datos actuarán como entradas y las clases definidas como áreas de interés como salidas (clases segmentadas). Bajo esa premisa, se realiza un proceso de clasificación por píxel, y se le asigna un porcentaje de pertenencia a cada una de las áreas de interés.



**Figura 5.2** Arquitectura general del sistema [21]. En gris los componentes implementados (fuera de alcance) del presente documento

## 5.4. Especificaciones técnicas del sistema

A continuación se detallan los requisitos de hardware necesarios para ejecutar el programa, cabe tener en cuenta que la memoria RAM necesaria variará en base a la resolución de las imágenes que se procesen. Para un proceso óptimo éstas no deberán sobrepasar resoluciones de 1000 píxeles por lado.

### 5.4.1. Especificaciones de hardware

Especificaciones mínimas:

- Procesador a 2 GHz
- Memoria RAM: 4 GB
- Espacio en disco: 1 GB

Especificaciones recomendadas:

- Procesador a 3 GHz
- Memoria RAM: 8 GB
- Espacio en disco: 1 GB

### 5.4.2. Especificaciones de software

El sistema se desarrolló con software multiplataforma, por tanto no debieran haber inconvenientes relacionados. En todo caso el ambiente de desarrollo fue el siguiente:

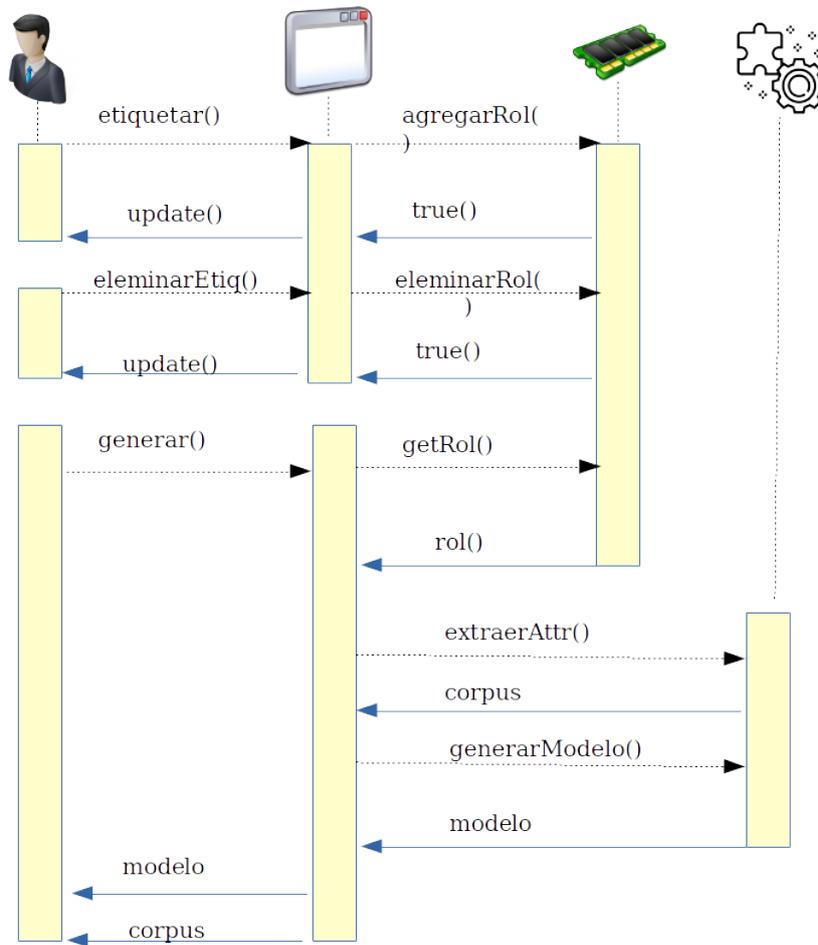
- **Sistema operativo:** Arch Linux x86\_64
- Java(TM) SE Runtime Environment (build 1.8.0\_181-b13)
- Apache Maven 3.5.4
- Eclipse 4.9 (Photon)

## **5.5. Diagramas**

### **5.5.1. Diagrama de secuencias**

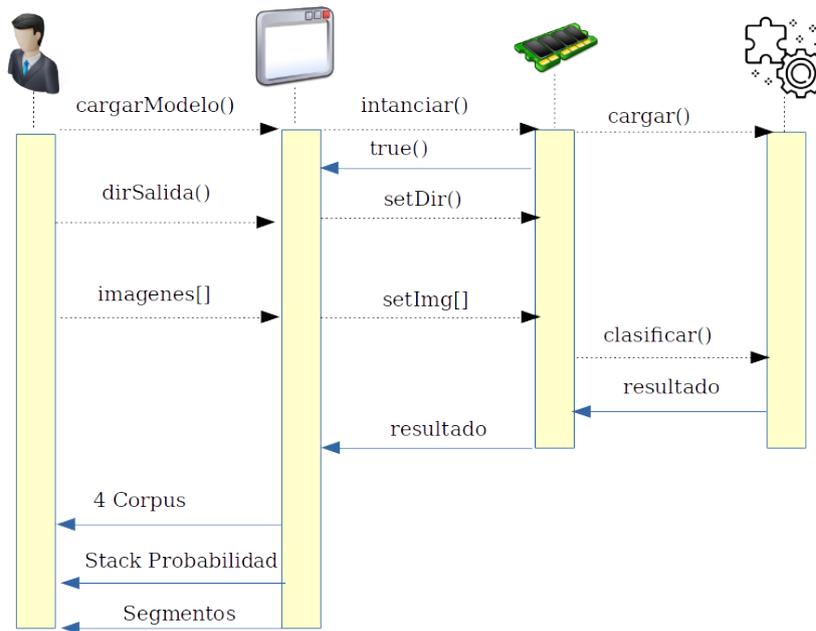
Para la fase de entrenamiento 5.3, el usuario inicialmente brinda la información al sistema, ésta información está en forma de etiquetas sobre secciones de una imagen de entrada. El sistema actualizará la interfaz de usuario para brindar control sobre las áreas de interés ya definidas. De la misma manera el usuario eliminará las áreas que erróneamente hayan sido seleccionadas y el sistema notificará al usuario sobre el cambio. Internamente, el sistema se encargará de administrar las áreas de interés definidas con sus respectivas etiquetas para las siguientes fases.

Una vez el usuario haya definido suficientes muestras al sistema, el sistema comenzará el proceso de generación de un modelo para clasificación; para ello extraerá atributos de imagen correspondientes a cada región y ensamblará un corpus de entrenamiento. Finalmente el usuario debe guardar el modelo y corpus generados para proceder a la siguiente fase.



**Figura 5.3** Secuencia de la fase de entrenamiento.

Para la fase de clasificación 5.4, el usuario deberá tener el un archivo modelo, resultado de haber completado la fase 5.3. El usuario procede a definir un directorio para almacenar los resultados del análisis. Luego, procede a cargar una o más imágenes a ser procesadas, el sistema entonces procede a extraer atributos de la imagen a ser clasificada y los clasifica con el modelo cargado. Producto de este proceso, se generarán: cuatro archivos de cuerpo de datos (uno por cada clase segmentable), un conjunto de imágenes representando mapas de probabilidad y una imagen semejante a la original salvo que lleva resaltada de los segmentos que la componen.



**Figura 5.4** Secuencia de la fase de clasificación

### 5.5.2. Diagrama de clases

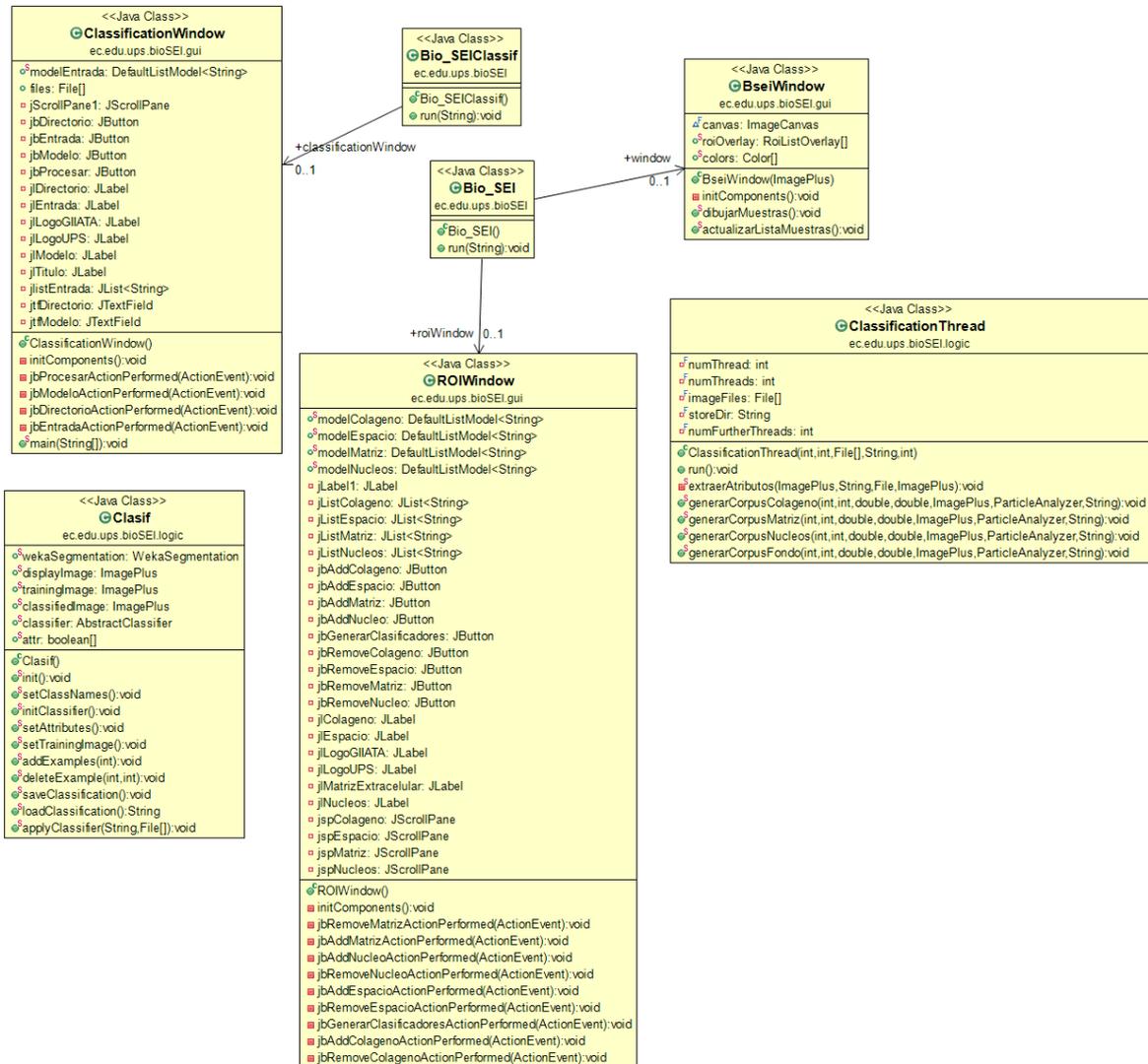


Figura 5.5 Diagrama de clases.

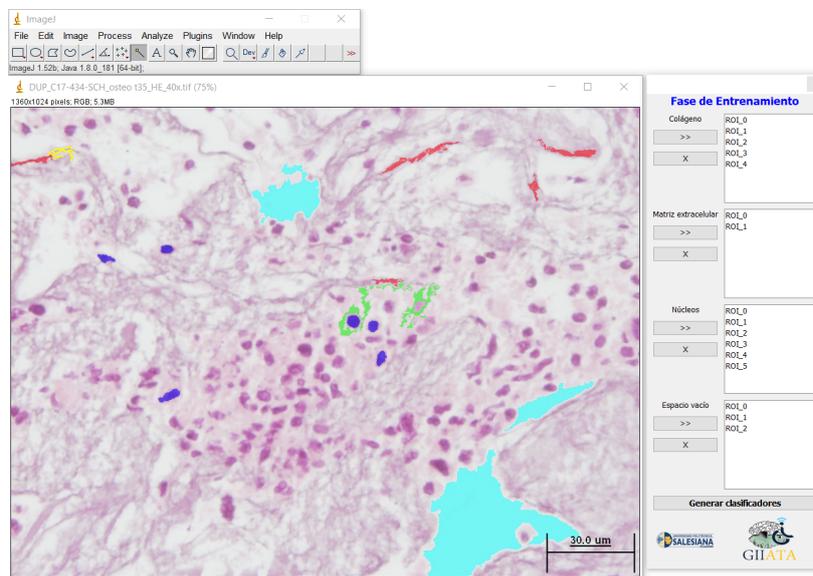
## 5.6. Diseño e implementación del plugin

### 5.6.1. Pre-procesamiento de las imágenes

Permite realizar ciertas operaciones pertinentes a al mejoramiento de las imágenes previo a ser procesadas a fin de eliminar distorsiones no deseadas o aclarar ciertos aspectos para

su uso. Por otra parte en esta fase es necesaria la intervención de un profesional a cargo de identificar muestras base de las clases segmentables definidas con el fin de ensamblar un cuerpo de datos lo suficientemente grande como para generar los clasificadores (expuestos a continuación).

Para ello deberá realizar selecciones sobre las regiones de interés y etiquetarlas mediante la interfaz provista por la herramienta desarrollada. En caso de cometer un error dispone de la utilidad de eliminar regiones.



**Figura 5.6** Interfaz de usuario para la fase de etiquetado de atributos y generación del corpus de datos para el clasificador

### 5.6.2. Extracción de atributos

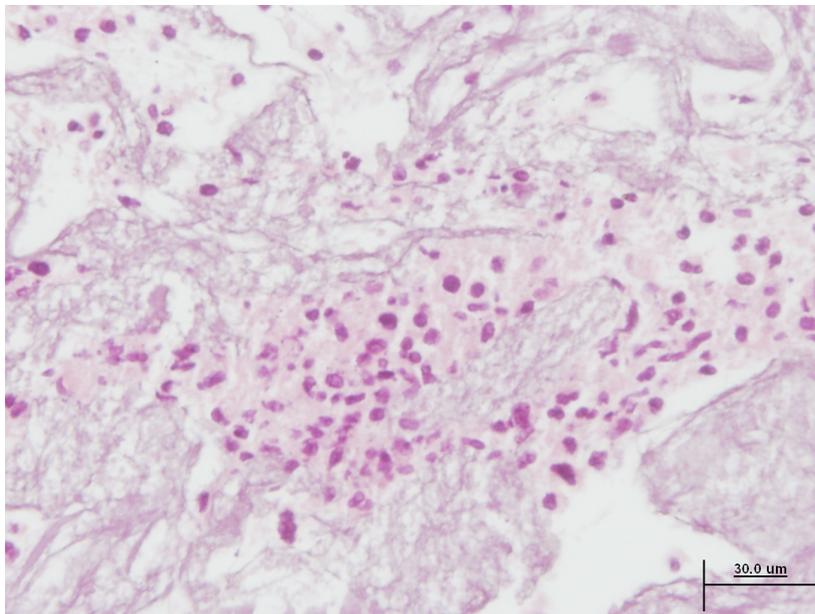
En este módulo, basado en las características y tipo de imagen a segmentar, se definen los tipos de elementos que la componen y se extrae información en base a procesos y algoritmos aplicados sobre los píxeles de la imagen.

Atributos segmentados:

- Colágeno (tipo I)

- Matriz extracelular
- Núcleos
- Fondo

La segmentación de atributos se ejemplificará en torno a la siguiente imagen de correspondiente a un cultivo de células C17 en un medio osteogénico con un zoom de 40x. Escala de referencia en la imagen: 30.0um.

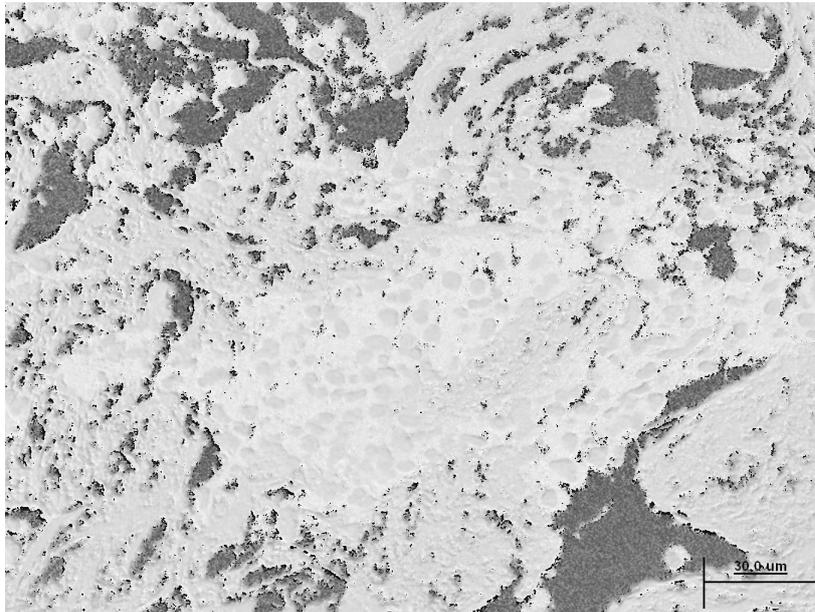


**Figura 5.7** Imagen base utilizada para ejemplificar un caso guía del proceso de segmentación [21].

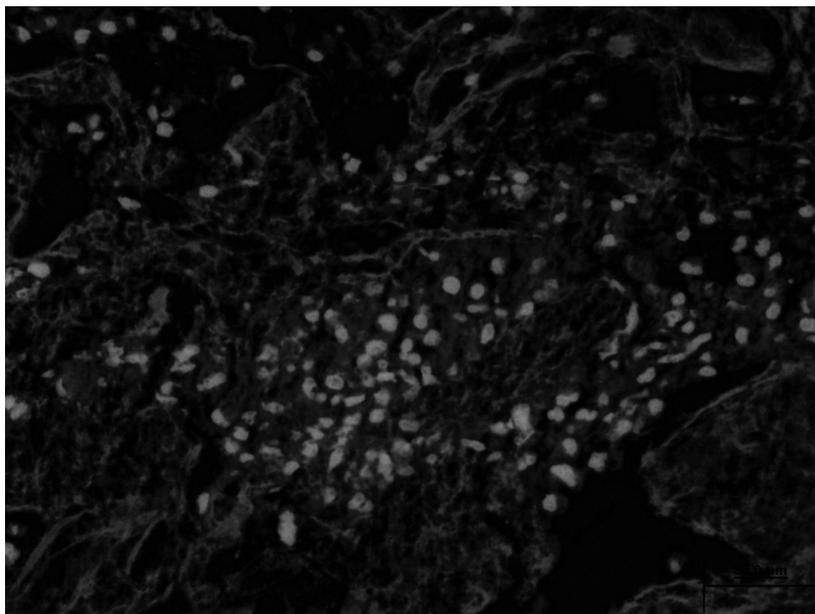
Antes de considerar el entrenamiento del sistema, debemos establecer claramente los atributos, por tanto es necesario extraer la siguiente información de cada atributo segmentable:

#### 5.6.2.1. Capas HSV

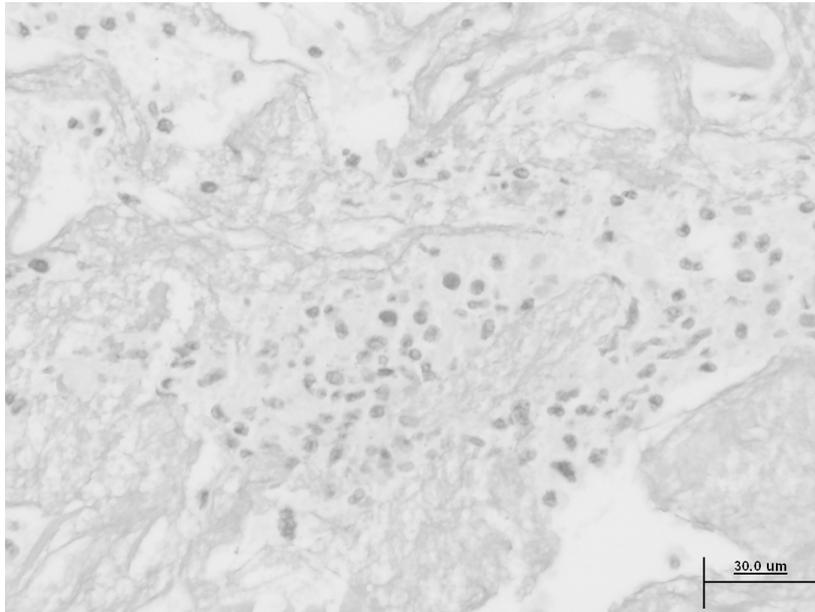
Inicialmente obtenemos los valores del modelo de color HSB de la imagen, nos da como resultado los primeros tres atributos del corpus. Por otra parte se obtiene la intensidad de color al convertir la imagen a blanco y negro.



**Figura 5.8** Resultado de la separación de capas HSV, siendo esta imagen la correspondiente al *color*.



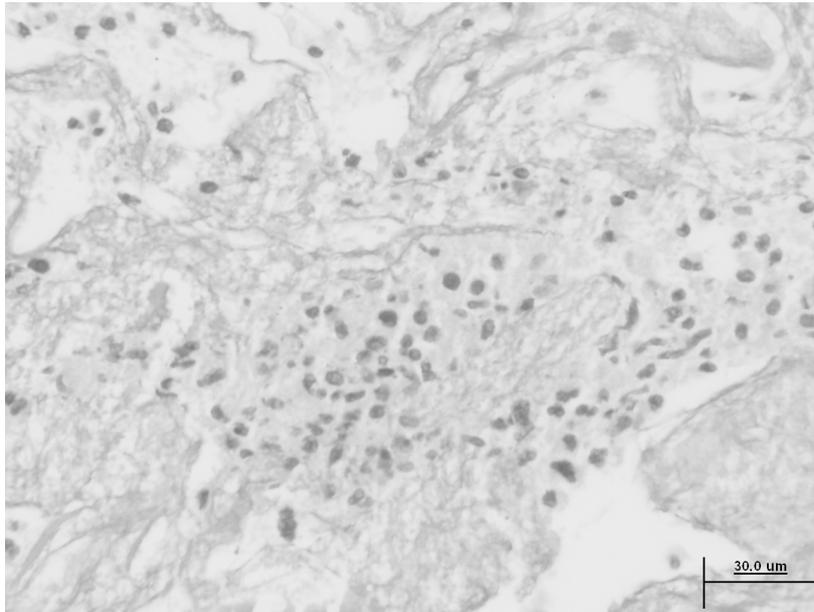
**Figura 5.9** Resultado de la separación de capas HSV, siendo esta imagen la correspondiente a la *saturación*.



**Figura 5.10** Resultado de la separación de capas HSV, siendo esta imagen la correspondiente al *valor*, también identificable como *brillo*.

#### 5.6.2.2. Imagen original BN

Otro atributo para componer el corpus de entrenamiento es la conversión de la imagen original a una en Blanco y Negro. Dando como resultado una imagen con un valor único por píxel en el dominio de 0 a 255.



**Figura 5.11** Resultado de la conversión de la imagen original a una equivalente de una sola capa, con valores en escala de grises que varían de 0 a 255.

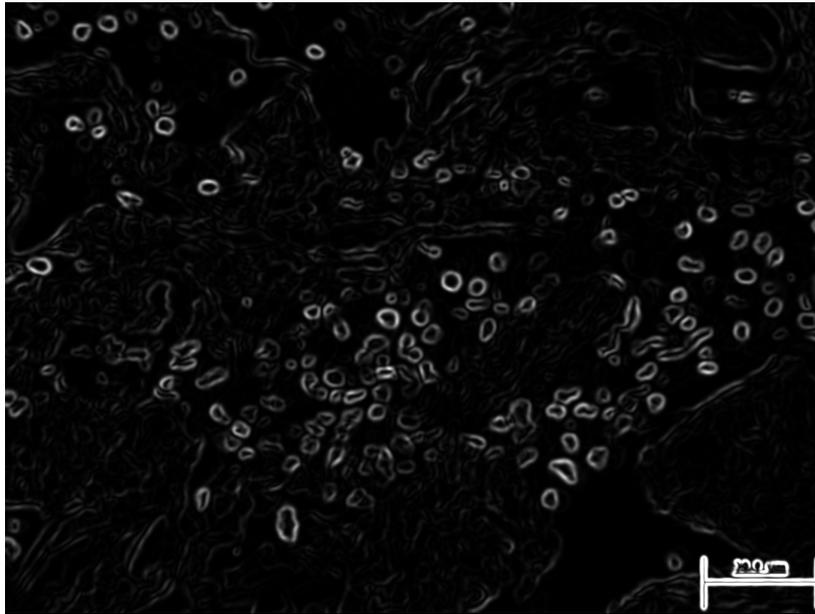
### 5.6.2.3. Tensor de estructura

Luego de calcular el tensor de estructura, los auto-valores resultantes (mayor, medio y menor) son ordenados para cada elemento de la imagen (píxel) [20]. “Este proceso resulta en una representación matricial por los mayor y menor auto-valores calculados y sus valores nos servirán para componer parte del corpus de entrenamiento.”

La matriz resultante se define mediante:

$$J = \begin{bmatrix} \langle f_x, f_x \rangle_w & \langle f_x, f_y \rangle_w \\ \langle f_x, f_y \rangle_w & \langle f_y, f_y \rangle_w \end{bmatrix}$$

donde  $f_x$  y  $f_y$  son las imágenes de las derivadas parciales espaciales,  $\frac{\partial f}{\partial x}$  y  $\frac{\partial f}{\partial y}$  respectivamente [21].



**Figura 5.12** Resultado de aplicación de filtro de estructura, se ha realizado un ligero proceso de realze de colores por motivos de presentación.

#### 5.6.2.4. Entropía

Se dibuja un círculo de radio  $r$  alrededor de cada elemento de la imagen, posteriormente se obtiene el mapa de intensidades para todos los elementos de dicho círculo aparte del resto de la imagen generando fragmentos de imágenes circulares, luego se calcula la entropía como:

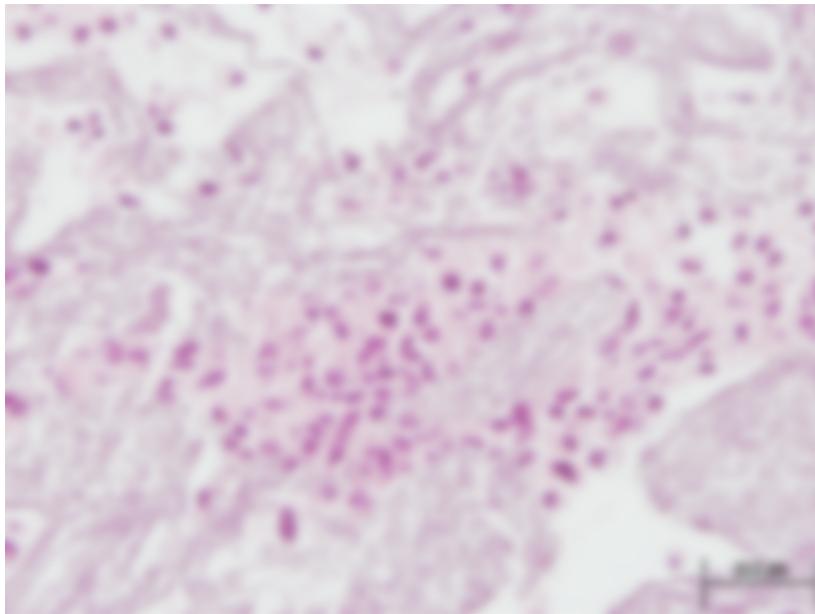
$$\sum_{p \text{ en hist}} - p * \log_2(p)$$

para cada partícula, donde  $p$  es la probabilidad de cada fragmento en el histograma de las 3 capas de la imagen, tanto RGB como HSB y  $r$  es igual a 1, 2, 4, 8 y 16. Este proceso resulta en 5

### 5.6.2.5. Desenfoque Gaussiano

Para obtener los atributos relacionados con el desenfoque, se realizan convoluciones con la función Gaussiana para suavizar la imagen, para esto se definen las siguientes funciones:

- $\sigma$ : Es el radio de decaimiento  $\exp(-0,5) \sim 61\%$ . Esto es, la desviación estandar  $\sigma$  de la gaussiana.

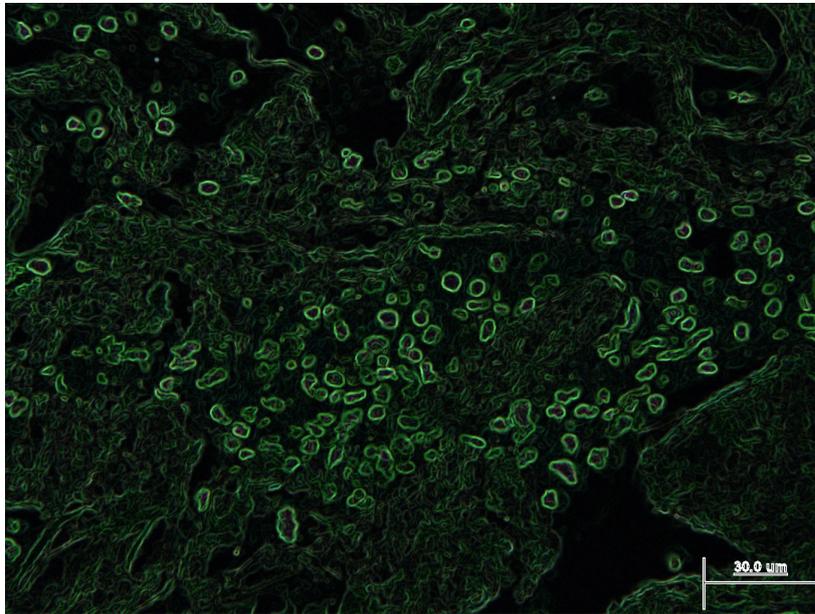


**Figura 5.13** Resultado de aplicación del filtro de desenfoque Gaussiano con un radio  $\sigma = 8$ . Este filtro se utiliza para reducir los detalles y el ruido en las imágenes.

### 5.6.2.6. Operador Sobel

El operador de Sobel, a veces denominado “Operador Sobel-Feldman” o “Filtro de Sobel” es utilizado en procesamiento de imágenes y visión por computador, tiene un desempeño especialmente optimizado para los problemas donde se requiere identificar bordes. “Este operador realiza una medición del gradiente espacial de una imagen para lograr resaltar las zonas con alta frecuencia espacial que corresponden a los bordes” [1]. El análisis numérico

nos muestra que para ciertas clases de superficies se puede obtener una estimación aún mejor utilizando los pesos promedio de tres de tales diferencias centrales [25].



**Figura 5.14** Resultado de aplicación del operador de sobel. Este filtro generalmente se ocupa para la detección de bordes en las imágenes.

Como resultado de la aplicación de cada uno de los filtros especificados, se procede a ensamblar el cuerpo de entrenamiento del clasificador. Siendo éste un formato de archivo atributo-relación “ARFF”. Éstos archivos se componen de dos secciones diferentes. La primera sección almacena información de **cabeceras** mientras la segunda almacena un cuerpo de **datos**.

La cabecera del documento contiene el nombre de la relación y una lista de atributos (las columnas de datos) y sus tipos, como se muestra a continuación:

```
@relation segment
```

```
@attribute original numeric
```

```
@attribute Hue numeric
```

@attribute Saturation numeric  
@attribute Brightness numeric  
@attribute Sobel\_filter\_0.0 numeric  
@attribute Gaussian\_blur\_1.0 numeric  
@attribute Sobel\_filter\_1.0 numeric  
@attribute Structure\_largest\_1.0\_1.0 numeric  
@attribute Structure\_smallest\_1.0\_1.0 numeric  
@attribute Structure\_largest\_1.0\_3.0 numeric  
@attribute Structure\_smallest\_1.0\_3.0 numeric  
@attribute Entropy\_1\_32 numeric  
@attribute Entropy\_1\_64 numeric  
@attribute Entropy\_1\_128 numeric  
@attribute Entropy\_1\_256 numeric  
@attribute Gaussian\_blur\_2.0 numeric  
@attribute Sobel\_filter\_2.0 numeric  
@attribute Structure\_largest\_2.0\_1.0 numeric  
@attribute Structure\_smallest\_2.0\_1.0 numeric  
@attribute Structure\_largest\_2.0\_3.0 numeric  
@attribute Structure\_smallest\_2.0\_3.0 numeric  
@attribute Entropy\_2\_32 numeric  
@attribute Entropy\_2\_64 numeric  
@attribute Entropy\_2\_128 numeric  
@attribute Entropy\_2\_256 numeric  
@attribute Gaussian\_blur\_4.0 numeric  
@attribute Sobel\_filter\_4.0 numeric  
@attribute Structure\_largest\_4.0\_1.0 numeric

@attribute Structure\_smallest\_4.0\_1.0 numeric  
@attribute Structure\_largest\_4.0\_3.0 numeric  
@attribute Structure\_smallest\_4.0\_3.0 numeric  
@attribute Entropy\_4\_32 numeric  
@attribute Entropy\_4\_64 numeric  
@attribute Entropy\_4\_128 numeric  
@attribute Entropy\_4\_256 numeric  
@attribute Gaussian\_blur\_8.0 numeric  
@attribute Sobel\_filter\_8.0 numeric  
@attribute Structure\_largest\_8.0\_1.0 numeric  
@attribute Structure\_smallest\_8.0\_1.0 numeric  
@attribute Structure\_largest\_8.0\_3.0 numeric  
@attribute Structure\_smallest\_8.0\_3.0 numeric  
@attribute Entropy\_8\_32 numeric  
@attribute Entropy\_8\_64 numeric  
@attribute Entropy\_8\_128 numeric  
@attribute Entropy\_8\_256 numeric  
@attribute Gaussian\_blur\_16.0 numeric  
@attribute Sobel\_filter\_16.0 numeric  
@attribute Structure\_largest\_16.0\_1.0 numeric  
@attribute Structure\_smallest\_16.0\_1.0 numeric  
@attribute Structure\_largest\_16.0\_3.0 numeric  
@attribute Structure\_smallest\_16.0\_3.0 numeric  
@attribute Entropy\_16\_32 numeric  
@attribute Entropy\_16\_64 numeric  
@attribute Entropy\_16\_128 numeric

```
@attribute Entropy_16_256 numeric  
@attribute class {colageno,matriz,nucleo,espacio}
```

### 5.6.3. Fase de entrenamiento

Esta fase incluye los procesos necesarios para aplicar algoritmos de aprendizaje de máquina con el fin de entrenar el clasificador. Aquí se realizan los procedimientos necesarios dependiendo del tipo de algoritmo que se implemente y de las entradas u operaciones específicas que este proceso requiera.

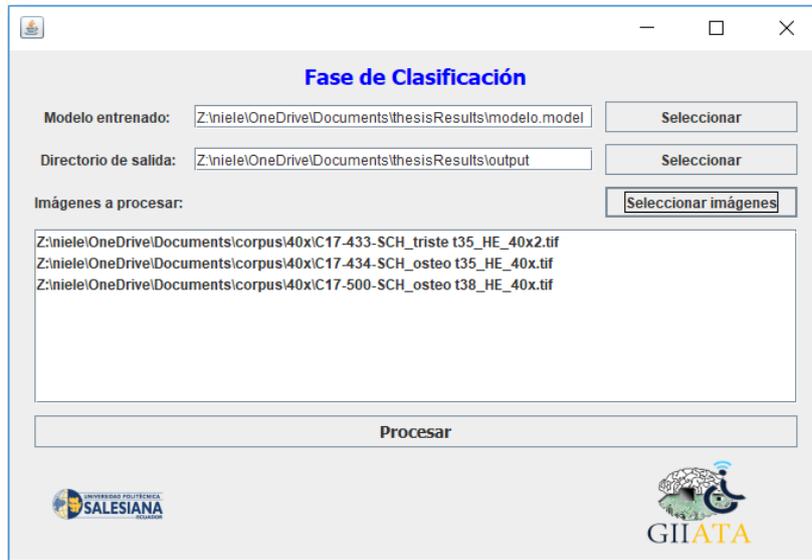
Con base en el documento de cabeceras obtenidas de la fase anterior, es necesario establecer un clasificador que maneje los datos de la mejor manera.

Para ello, con base en el resumen de resultados individual expuesto en la sección 6.2, específicamente bajo el porcentaje de instancias clasificadas correctamente, se ha escogido al modelo *SMO* como el modelo con mayor precisión para este caso específico.

El programa desarrollado nos devolverá entonces dos archivos, el corpus de datos antes mencionado en formato ARFF y un archivo binarizado \*.model con el modelo entrenado para su posterior aplicación.

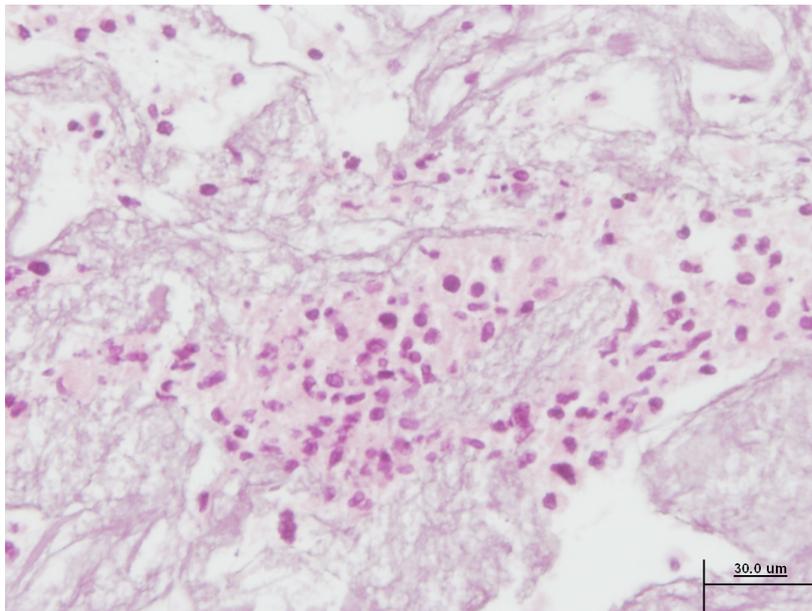
### 5.6.4. Fase de clasificación

En base a la identificación correcta de los componentes en las imágenes de entrenamiento, el sistema inteligente se encarga de recibir los atributos pertinentes dichas clases para posteriormente clasificar cada píxel de todas las imágenes del corpus. “La correcta identificación de estas clases permite generar mapas de probabilidad para cada clase de objeto de células CMMh3A6 y extraer atributos tales como el área para su posterior análisis” [21].

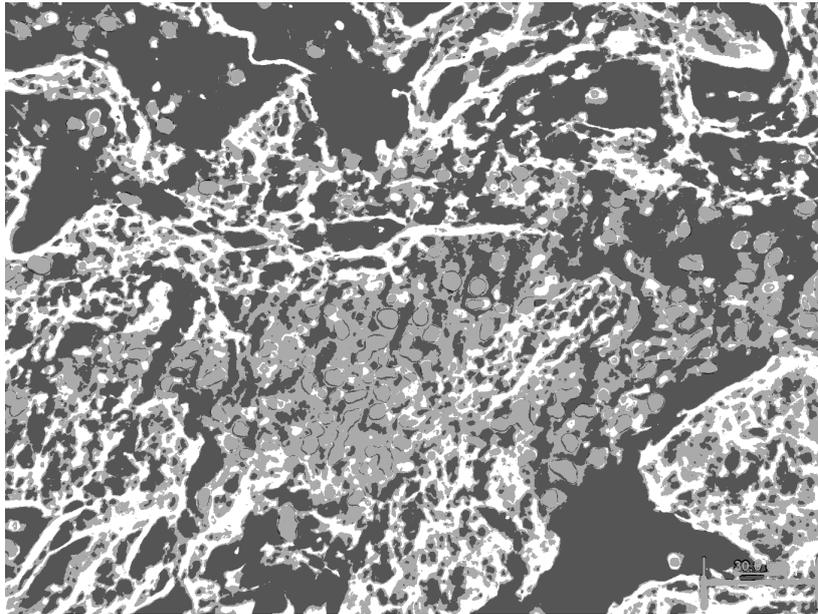


**Figura 5.15** Interfaz de usuario para la fase de clasificación, aquí deberá seleccionar el modelo previamente generado, establecer un directorio de salida y seleccionar una o más imágenes del mismo tipo para proceder a generar los mapas de probabilidad con la información de clasificación.

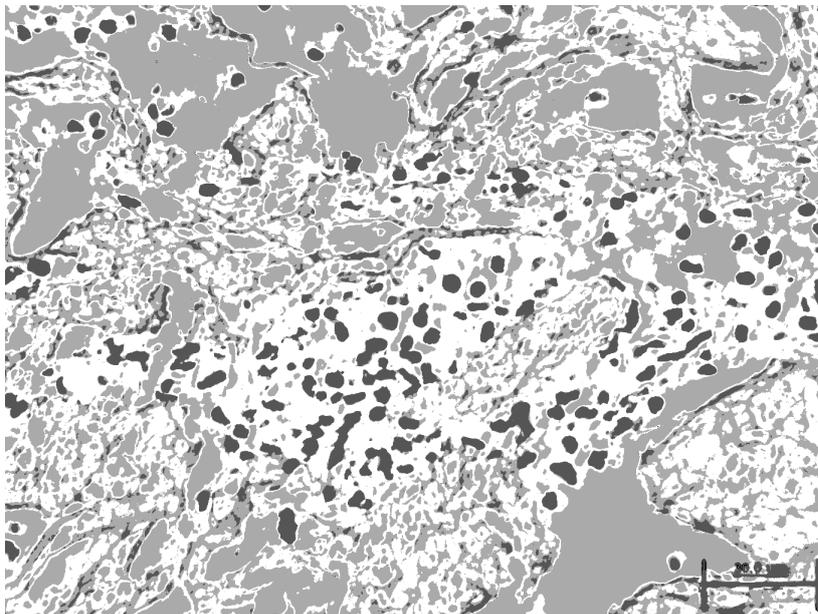
Los resultados se calculan tomando a la figura 5.16 como entrada. A partir de esta se identifican los elementos segmentables y se generan los mapas de probabilidad como se muestra en las figuras 5.17, 5.18, 5.19 y 5.20.



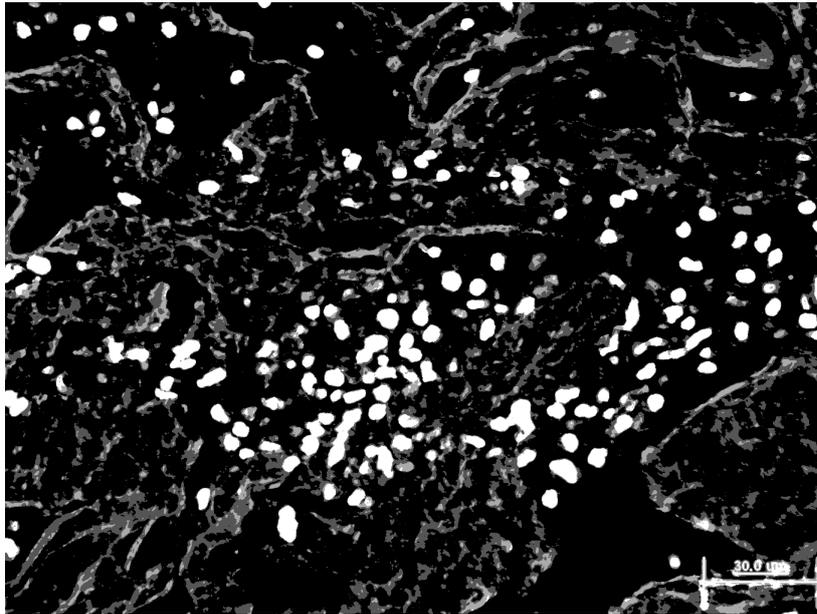
**Figura 5.16** Imagen de muestra para ser clasificada, ésta es la imagen base sobre la que se identificarán los atributos segmentables [21].



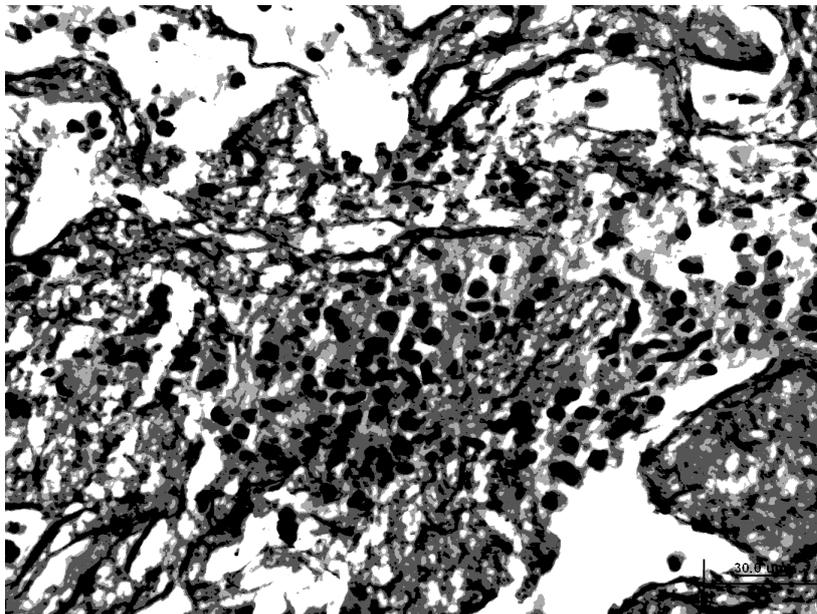
**Figura 5.17** Mapa de probabilidad representando el atributo clasificado *colágeno* con base en la imagen 5.16.



**Figura 5.18** Mapa de probabilidad representando el atributo clasificado *matriz* con base en la imagen 5.16.



**Figura 5.19** Mapa de probabilidad representando el atributo clasificado *núcleos* con base en la imagen 5.16.

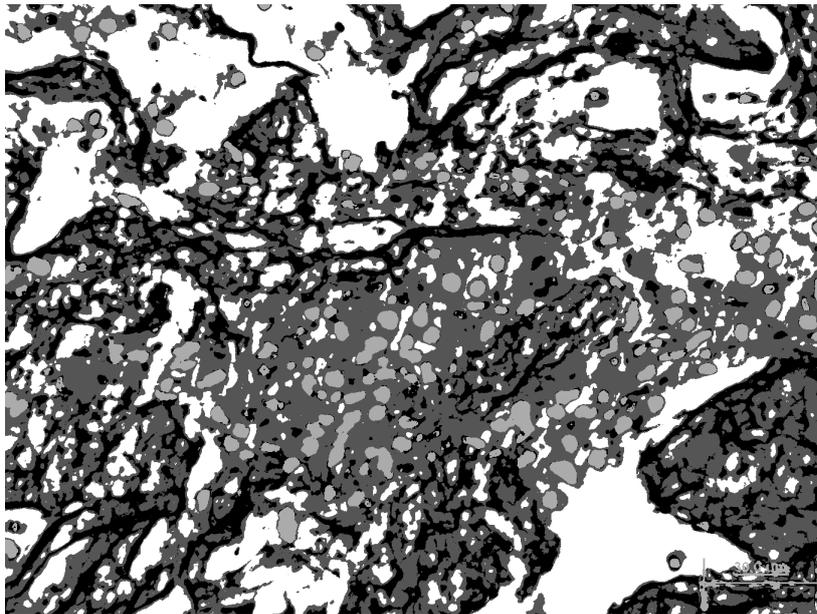


**Figura 5.20** Mapa de probabilidad representando el atributo clasificado *espacio* con base en la imagen 5.16.

### 5.6.5. Extracción de atributos y resultados

En base a los resultados arrojados el módulo de entrenamiento y recomendación del clasificador, este módulo se encarga de realizar las respectivas segmentaciones sobre la imagen original así como se el cálculo de los atributos propios de cada una de las clases, estos atributos están relacionados con propiedades de la imagen para cada una de las clases segmentadas.

A partir de los 4 mapas de probabilidad generados, se realiza un proceso de selección para escoger la etiqueta final de cada elemento de la imagen como una clase clasificada. Para ello, por cada píxel se consulta el valor de probabilidad en cada uno de los 4 mapas y se selecciona el mayor valor. Hecho esto se genera una imagen final con secciones claramente identificadas, tal como se aprecia en la figura 5.21.



**Figura 5.21** Resultado de la selección de las clases con mayor probabilidad por cada píxel. Se puede identificar en color negro el colágeno, gris oscuro la matriz extracelular, gris claro los núcleos y de blanco el fondo.

Con base en la imagen 5.21 se procede a extraer los atributos de la imagen base de la siguiente manera:

- Sabiendo que, la imagen se encuentra conformada por unidades de imagen con valores de 0 a 3
- Se procede a realizar segmentación mediante los valores de umbral definidos, definiendo como mínimo y máximo a cada valor de clase [0-3]
- Por cada una de las cuatro clases segmentables establecidas, se crea una máscara de selección
- La máscara de selección se aplica sobre la imagen base y se procede a extraer los atributos descritos en la tabla 5.1

**Cuadro 5.1** Atributos extraídos de los segmentos resultantes

<b>Atributos resultantes (por clase)</b>	
Area	Promedio de valores gris
Desviación estándar	Moda de valor gris
Valor gris mínimo y máximo	Centroide
Centro de masa	Perimetro
Rectangulo delimitador	Forma de elipse
Descriptores de forma	Diámetro de Feret
Densidad integrada	Mediana
Asimetría estadística	Kurtosis
Fracción de área	Posición en la pila

Finalmente el sistema nos devuelve cuatro corpus de datos en archivos, representando las 4 clases segmentables. Cada archivo contiene la información expuesta en el cuadro 5.1 solamente con datos de las partículas en la imagen que representan su clase.

Este corpus de datos puede ser utilizado entonces para un análisis cuantitativo de cada componente identificado, en todo caso, existen alternativas al análisis estadístico clásico

basadas en métodos de clasificación, análisis descriptivo, estadísticas de implicación y estadísticas de cuasi-implicación, entre otras [21].

# Capítulo 6

## Experimentación, modelos y resultados

### 6.1. Preparación de corpus de imágenes

Como resultado de una investigación realizada en el Instituto de Investigación Biomédica de A Coruña gracias al trabajo conjunto de Yaroslava Robles-Bykbaev, Silvia Díaz-Prado, Clara Sanjurjo y Francisco J. Blanco. Se realizó un trabajo enfocado en el estudio de la degradación (en vitro) de polímeros PDLGA y copolímeros debido a la hidrólisis en condiciones de cuerpo humano [23].

En dicha investigación se diseñó un esquema de preparación de muestras para modelar la degradación causada por hidrólisis de diferentes compuestos PDLG. Habiendo mantenido todas las muestras a una temperatura de 37°C con un pH inicial de 7.4, asimilando las condiciones propias de un cuerpo humano. A partir de este esquema, se extrajeron muestras de manera periódica (cada 3 días) bajo condiciones adecuadas, dando como resultado un conjunto de imágenes que contienen información sobre la trayectoria de degradación de los polímeros analizados [23]. Para los propósitos del presente trabajo se nos ha facilitado el

corpus de imágenes producto de ésta investigación y serán nuestro punto de partida para la evaluación de los modelos desarrollados.

## 6.2. Aproximaciones alternativas

Con el fin de realizar el proceso de clasificación de los componentes de imágenes de este proyecto, se consideró el uso de cuadernos Jupyter, basados en el proyecto “*SciJava Jupyter Kernel*”. La idea de dicho proyecto es utilizar las librerías desarrolladas por SciJava (componente base en los proyectos ImageJ y FIJI ocupados para el desarrollo del presente trabajo) para brindar una interfaz interactiva para los proyectos que requieren análisis de datos, actuando como un vínculo entre éstas plataformas.

Esta alternativa resultó ser poco viable, ya que si bien en principio el núcleo tiene soporte para varios lenguajes (Groovy, Python, BeanShell, Clojure, Java, JavaScript, Ruby y Scala), todavía no cuenta con la robustez necesaria ó la documentación suficiente para poder explotar todo el potencial de sus librerías, útiles para el tratamiento y procesamiento de imágenes en general. No cabe duda que en tanto el proyecto “*SciJava Jupyter Kernel*” alcance cierto nivel de madurez, tendrá el potencial de brindarnos un enfoque mucho más favorable en tanto a los objetivos de brindar una herramienta que facilite el la enseñanza y aprendizaje de los métodos y algoritmos presentes en este trabajo.

## 6.3. Selección de modelo de clasificación

La precisión y confiabilidad del sistema en el momento de realizar la clasificación de los elementos celulares dependerá principalmente de la calidad del modelo de clasificación que se defina. Es por eso que se ha realizado un proceso de selección con base en los resultados individuales de diferentes modelos sobre el mismo cuerpo de datos.

Para verificar la eficacia de los clasificadores aplicados, los modelos de clasificación han sido validados mediante el método de *validación cruzada* ó *estimación de rotación* con 10 dobleces. Se ha escogido esta técnica ya que demuestra su eficiencia al momento de probar modelos donde el propósito es la predicción [13], donde se pretende medir con cuánta precisión desempeñará el modelo en la práctica. Los resultados individuales se detallan a continuación.

### 6.3.0.1. Modelo logístico de árboles LMT

Los LMT *Logistic Model Tree* “son modelos de clasificación de entrenamiento supervisado. Éste combina regresiones logísticas (LR) y aprendizaje por árboles de decisión” [14]. Esquema aplicado: `weka.classifiers.trees.LMT -I -1 -M 15 -W 0.0`.

**Cuadro 6.1** Sumario de modelo LMT

<b>Sumario</b>	
Instancias clasificadas correctamente	99.9788 %
Instancias clasificadas incorrectamente	0.0212 %
Coefficiente de Kappa	0.9996
Error medio absoluto (MAE)	0.25
Error cuadrático medio	0.3111
Error relativo absoluto	92.8312 %
Error cuadrático relativo	84.7792 %
Número total de instancias	61259

**Cuadro 6.2** Matriz de confusión modelo LMT

<b>Matriz de confusión</b>					
a	b	c	d	←	clasificado como
10866	12	0	0		a = colageno
0	8494	0	0		b = matriz
0	0	2728	0		c = nucleo
0	1	0	39158		d = espacio

### 6.3.0.2. Modelo de árbol Hoeffding

“Un árbol Hoeffding (VFDT) es un algoritmo de inducción incremental, en cualquier momento, del árbol de decisiones que es capaz de aprender de flujos de datos masivos, asumiendo que los ejemplos de generación de distribución no cambian con el tiempo.”

[8] Este tipo de moldeamiento tiene una gran ventaja de desempeño ya que no requiere de muchos espacios muestrales para obtener buenos resultados. Esquema aplicado:  
`weka.classifiers.trees.HoeffdingTree -L 2 -S 1 -E 1.0E-7 -H 0.05 -M 0.01  
 -G 200.0 -N 0.0.`

**Cuadro 6.3** Sumario de modelo Hoeffding

<b>Sumario</b>	
Instancias clasificadas correctamente	99.9037 %
Instancias clasificadas incorrectamente	0.0963 %
Coefficiente de Kappa	0.9982
Error medio absoluto (MAE)	0.001
Error cuadrático medio	0.0219
Error relativo absoluto	0.3715 %
Error cuadrático relativo	5.9669 %
Número total de instancias	61259

**Cuadro 6.4** Matriz de confusión modelo Hoeffding

<b>Matriz de confusión</b>					
a	b	c	d	←	clasificado como
10850	17	11	0		a = colageno
27	8465	2	0		b = matriz
2	0	2726	0		c = nucleo
0	0	0	39159		d = espacio

### 6.3.0.3. Modelo de árbol J48

El árbol de decisión J48 “es la implementación del algoritmo ID3 (Iterative Dichotomiser 3), éste algoritmo utiliza un enfoque codicioso al seleccionar el mejor atributo para dividir el conjunto de datos en cada iteración” (una mejora que se puede hacer en el algoritmo puede ser el uso del seguimiento durante la búsqueda del árbol de decisiones óptimo) [19], implementado por el equipo de desarrollo de “WEKA”. Esquema aplicado: `weka.classifiers.trees.J48 -C 0.25 -M 2.`

**Cuadro 6.5** Sumario de modelo J48

<b>Sumario</b>	
Instancias clasificadas correctamente	99.9592 %
Instancias clasificadas incorrectamente	0.0408 %
Coefficiente de Kappa	0.9992
Error medio absoluto (MAE)	0.0002
Error cuadrático medio	0.014
Error relativo absoluto	0.0905 %
Error cuadrático relativo	3.8147 %
Número total de instancias	61259

**Cuadro 6.6** Matriz de confusión modelo J48

<b>Matriz de confusión</b>					
a	b	c	d	←	clasificado como
10871	7	0	0		a = colageno
17	8476	1	0		b = matriz
0	0	2728	0		c = nucleo
0	0	0	39159		d = espacio

#### 6.3.0.4. Modelo de árbol REPTree

El modelo RepTree ocupa una logica semejante a la de los árboles de regresión, creando nuevos nodos durante su ejecución. Posee ciertas métricas para evaluar el desempeño de cada iteración y de esa manera seleccionar los mejores candidatos modelo. A partir del mejor modelo seleccionado; “los valores que faltan se tratan con el método de *C4.5*” [19], usando instancias fraccionarias. [9]. Para el caso actual no se ha aplicado la característica

de “podar” ya que es un caso de clasificación, mas no de regresión. Esquema aplicado:

```
weka.classifiers.trees.REPTree -M 2 -V 0.001 -N 3 -S 1 -L -1 -I 0.0.
```

**Cuadro 6.7** Sumario de modelo REPTree

<b>Sumario</b>	
Instancias clasificadas correctamente	99.9755 %
Instancias clasificadas incorrectamente	0.0245 %
Coficiente de Kappa	0.9995
Error medio absoluto (MAE)	0.0002
Error cuadrático medio	0.011
Error relativo absoluto	0.0704 %
Error cuadrático relativo	2.9913 %
Número total de instancias	61259

**Cuadro 6.8** Matriz de confusión modelo REPTree

<b>Matriz de confusión</b>					
a	b	c	d	←	clasificado como
10871	6	1	0		a = colageno
8	8486	0	0		b = matriz
0	0	2728	0		c = nucleo
0	0	0	39159		d = espacio

### 6.3.0.5. Modelo de reglas con tabla de decisión

Modelo para la construcción y uso de un clasificador simple basado en tabla de decisiones, mediante el criterio de *mayoría*. Las tablas de decisión son generalmente los espacios de hipótesis más simples. En [12] se estipula que: “*Dada una muestra de entrenamiento que contiene instancias etiquetadas, un algoritmo de inducción construye*

una hipótesis en alguna representación. La representación que investigamos aquí es una tabla de decisiones con una regla predeterminada que se asigna a la clase de la mayoría, que abreviamos como DTM. Un DTM tiene dos componentes: un esquema y un cuerpo". Dada una instancia no etiquetada  $I$ , se asigna una etiqueta a la instancia mediante el clasificador DTM antes mencionado y se calcula, grosso modo, estimando a qué grupo de atributos etiquetados corresponden mayoritariamente los atributos de la instancia actual. Esquema aplicado: `weka.classifiers.rules.DecisionTable -X 1 -S "weka.attributeSelection.BestFirst -D 1 -N 5"`.

**Cuadro 6.9** Sumario de modelo DecisionTable

<b>Sumario</b>	
Instancias clasificadas correctamente	99.6915 %
Instancias clasificadas incorrectamente	0.3085 %
Coefficiente de Kappa	0.9943
Error medio absoluto (MAE)	0.0052
Error cuadrático medio	0.0388
Error relativo absoluto	1.9158 %
Error cuadrático relativo	10.5717 %
Número total de instancias	61259

**Cuadro 6.10** Matriz de confusión modelo DecisionTable

<b>Matriz de confusión</b>					
a	b	c	d	←	clasificado como
10765	86	8	19		a = colageno
34	8452	0	8		b = matriz
22	0	2694	12		c = nucleo
0	0	0	39159		d = espacio

### 6.3.0.6. Modelo de función: SMO

“Implementa el algoritmo de optimización mínimo secuencial (*SMO*) de John Platt para entrenar un clasificador de vectores de soporte” [18]. La ventaja de utilizar este tipo de modelo es que compensa la información de las clases que no poseen suficiente data y binarizando atributos con valores en un campo discreto. (En ese caso, los coeficientes en la salida se basan en los datos normalizados, no en los datos originales, esto es importante para interpretar el clasificador). “Para obtener estimaciones de probabilidad adecuadas, se utiliza la opción que ajusta los modelos de calibración a las salidas de la máquina de vectores de soporte (*SVM*)” [18][11].

En el caso de varias clases (en nuestro caso 4), las probabilidades predichas se acoplan utilizando el método de acoplamiento a pares de Hastie y Tibshirani [7]. Esquema aplicado:

```
weka.classifiers.functions.SMO -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1
-W 1 -K "weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C
250007calibrator "weka.classifiers.functions.Logistic -R 1.0E-8 -M -1
-num-decimal-places 4"
```

**Cuadro 6.11** Sumario de modelo SMO

<b>Sumario</b>	
Instancias clasificadas correctamente	99.9951 %
Instancias clasificadas incorrectamente	0.0049 %
Coefficiente de Kappa	0.9999
Error medio absoluto (MAE)	0.001
Error cuadrático medio	0.0045
Error relativo absoluto	0.0258 %
Error cuadrático relativo	1.2227
Número total de instancias	61259

**Cuadro 6.12** Matriz de confusión modelo SMO

<b>Matriz de confusión</b>					
a	b	c	d	←	clasificado como
10877	1	0	0		a = colageno
2	8492	0	0		b = matriz
0	0	2728	0		c = nucleo
0	0	0	39159		d = espacio

## 6.4. Resultados

El modelo desarrollado ha servido como base para la realización de ciertos análisis estadísticos, análisis PCA, modelos lineales generalizados, redes neuronales LSTM [21], así como también el desarrollo de una plataforma educativa que incluye objetos de aprendizaje para MOODLE donde se explican los modelos de segmentación descritos en el presente trabajo así como modelos estadísticos validados sobre dicha segmentación [22].

Para comprobar la validez del sistema, se tomaron 20 ejemplos de imágenes y se realizó el proceso de segmentación a mano. Cada una de dichas imágenes fueron entonces procesadas según lo descrito en la fase de entrenamiento del modelo.

Los 20 modelos resultantes fueron evaluados mediante dos técnicas, la primera mitad mediante separación del cuerpo de datos en proporciones de 70% para entrenamiento y 30% para pruebas. La segunda mitad de imágenes fueron evaluadas mediante validación cruzada, con 10 dobleces. Los resultados de dicho experimento se muestran en el cuadro 6.13.

**Cuadro 6.13** Resumen de experimentación ejecutado en la herramienta "WEKA Explorer" con base en los resultados individuales de imágenes segmentadas a mano.

Resumen de experimentación								
N	CORR	INC	Kappa	MAE	ECM	ERA	ECR	NUM
1	95.6973 %	4.3027 %	0.9415	0.2537	0.3177	68.950 %	74.021 %	5299
2	99.3323 %	0.6677 %	0.99	0.2506	0.3126	74.924 %	76.438 %	3145
3	95.9969 %	4.0031 %	0.9425	0.2535	0.3169	72.671 %	75.911 %	20884
4	98.971 %	1.029 %	0.9847	0.2509	0.313	74.862 %	76.288 %	1069
5	95.0332 %	4.9668 %	0.9227	0.2544	0.3188	78.867 %	79.415 %	21221
6	99.9512 %	0.0488 %	0.9955	0.25	0.3119	470.422 %	188.504 %	20491
7	95.9387 %	4.0613 %	0.9447	0.2535	0.3173	68.906 %	73.986 %	15586
8	99.9784 %	0.0216 %	0.9979	0.25	0.3118	475.085 %	192.260 %	60268
9	97.2811 %	2.7189 %	0.9604	0.2525	0.3154	73.223 %	75.953 %	3347
10	90.3638 %	9.6362 %	0.867	0.2582	0.3247	71.040 %	76.194 %	9319
11	98.4568 %	1.5432 %	0.9539	0.2515	0.3142	151.980 %	108.593 %	25790
12	98.4378 %	1.5622 %	0.9541	0.2515	0.3142	150.652 %	107.638 %	15171
13	99.6143 %	0.3857 %	0.9918	0.2505	0.3122	107.520 %	91.052 %	1815
14	90.0265 %	9.9735 %	0.8625	0.2585	0.3251	71.067 %	76.216 %	15842
15	90.1854 %	9.8146 %	0.8648	0.2583	0.3248	71.028 %	76.175 %	46594
16	95.0813 %	4.9187 %	0.9235	0.2544	0.3187	78.831 %	79.342 %	12483
17	97.891 %	2.109 %	0.9686	0.2521	0.3146	73.363 %	76.441 %	1138
18	95.9544 %	4.0456 %	0.9419	0.2535	0.3169	72.655 %	75.868 %	12285
19	99.6032 %	0.3968 %	0.9914	0.2504	0.3121	108.132 %	91.745 %	9073
20	97.4913 %	2.5087 %	0.9624	0.2522	0.3152	75.900 %	77.111 %	8929

Donde:

**N:** Número de prueba

**CORR:** Porcentaje de instancias clasificadas correctamente

**INC:** Porcentaje de instancias clasificadas incorrectamente

**Kappa:** Coeficiente de Kappa

**MAE:** Error Medio Absoluto

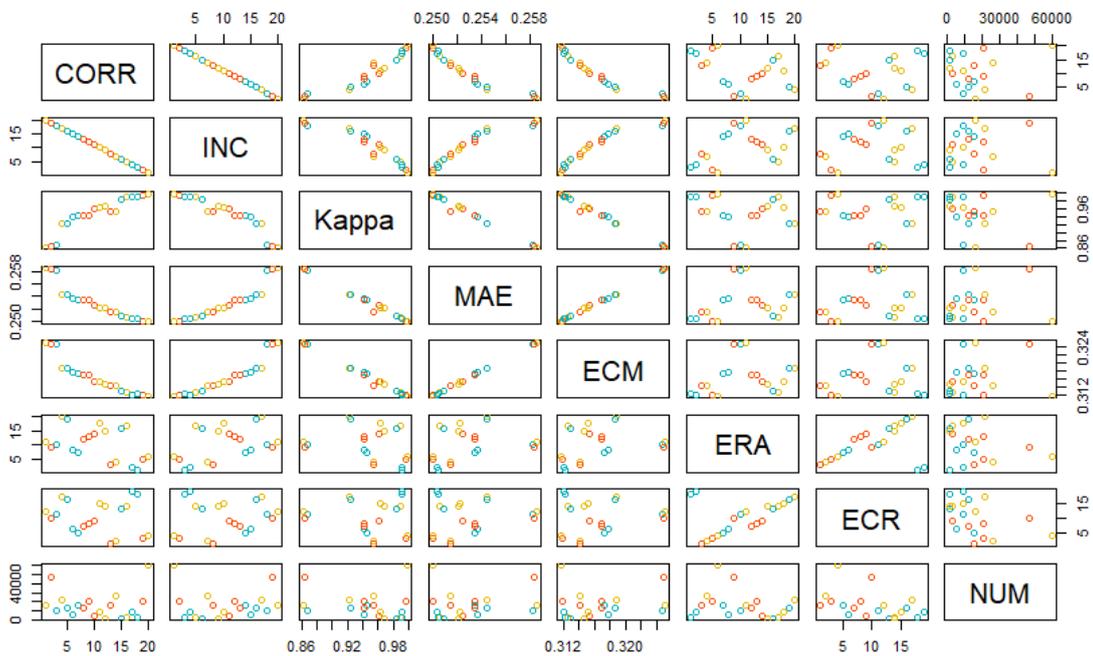
**ECM:** Error Cuadrático Medio

**ERA:** Error Relativo Absoluto

**ECR:** Error Cuadrático Relativo

**NUM:** Número total de instancias

Como se puede observar en el cuadro 6.13, se ha conseguido un promedio de 96.56% de unidades de imagen segmentadas correctamente. Podemos concluir entonces, que los elementos clasificados con los modelos aplicados son muy significativos en tanto que las etiquetas que haya definido el usuario correspondan a los elementos existentes en las imágenes.



**Figura 6.1** Se muestra un *scatterplot* comparativo que muestra las relaciones que existen entre las propiedades resultantes de la experimentación detalladas en el cuadro 6.13.

Por otra parte, cada imagen segmentada da como resultado cuatro corpus de datos, uno por cada clase segmentada, calculando para toda partícula (grupo de píxeles contiguos del mismo tipo) todos los atributos definidos en 5.1 así como un conjunto de mapas de probabilidad como se muestra en 5.17, 5.18, 5.19 y 5.20.

# Conclusiones

El desarrollo de este complemento de análisis de imágenes representa una herramienta base que puede ayudar tanto a investigadores como estudiantes a iniciar un estudio sobre imágenes de microscopía electrónica, ya sea que su propósito sea realizar un estudio de tipo investigativo o para explorar en el aprendizaje de los métodos de segmentación de imágenes en general. Este proyecto forma parte de un estudio sobre la degradación de colágeno en células de colágeno tipo-I, en donde se han realizado análisis estadísticos, análisis de componentes principales, modelos lineales generalizados y el uso de redes neuronales LSTM. Todas ellas demostrando la importancia y validez del presente trabajo.

Durante el proceso de desarrollo de este sistema fue fundamental contar con el apoyo de expertos en el área de la microscopía de biomateriales, primero por el trabajo de la obtención de las imágenes sobre los cultivos celulares y también por las guías y direcciones que representan. Su criterio es uno de los puntos más importantes al momento de definir e identificar los elementos que componen los biomateriales utilizados el hecho de considerar errores en el proceso de identificación de los materiales que componen los cultivos celulares, implicaría arrastrar dichos errores en las siguientes fases que se desarrollaron, por lo tanto cabe realzar la importancia que conlleva realizar una segmentación brindando resultados que se acerquen lo más posible a lo que representan las muestras celulares.

En cuanto a los resultados obtenidos en la etapa de experimentación, se pudo observar con gran satisfacción que el desempeño general de la segmentación de imágenes basada

en clasificadores es muy alta. Un análisis algo más exhaustivo puede revelar cuáles son los atributos más importantes o tal vez no relevantes para lograr la segmentación de cada elemento y así mejorar el proceso.

Por otra parte, con base en los resultados de segmentación se pudo identificar la dependencia del área que ocupan los grupos de colágeno con respecto al tiempo, ayudando a establecer un modelo matemático que explica la degradación del área que ocupa el colágeno en las muestras bajo condiciones ambientales claramente definidas [21]. Por otra parte se han logrado definir redes neuronales LSTM que permiten realizar estimaciones presuntivas de los valores que definen los componentes de las imágenes segmentadas (tales como áreas, posiciones, atributos de forma y demás) en los instantes de tiempo donde no existen registros de experimentación.

# Recomendaciones

Es recomendable que el sistema se ejecute en una máquina con buenos recursos, tanto de RAM como de procesamiento, idealmente se prefiere trabajar sobre un sistema operativo Linux ya que no tiene costo, se encuentra (generalmente) bajo licencias libres o permisivas y su óptimo manejo de procesos.

En cuanto al uso de memoria necesario, ya que todo atributo que se extrae de las imágenes es cargado a memoria, sería muy favorable evitar cargar imágenes de alta resolución ya que generan grandes volúmenes de datos y la máquina virtual de java pueda devolver un error, ya que limita el uso de memoria a un parámetro predefinido.

Cabe tener en cuenta que al momento de definir las etiquetas iniciales, previo a la generación de modelos, éstas deben ser lo más precisas posible. De seleccionar clases que no correspondan con su tipo de elemento, se generará ruido en el modelo entrenado, y es probable que no realice la clasificación de manera precisa.

# Trabajo futuro

Inicialmente puede mejorarse los tiempos de respuesta de la aplicación así como el uso de memoria. Revisar qué alternativas se encuentran disponibles para realizar clasificación de corpus de datos semejantes a los que se generan en este proyecto y comparar su eficacia y desempeño.

Al ser un importante inconveniente el tratamiento de múltiples imágenes así como el uso de imágenes demasiado grandes en resolución, realizar algoritmos que nos permitan segmentar dichas imágenes en equipos con recursos limitados. Pudiendo estas alternativas ser controlar de mejor manera los hilos donde se ejecuta la segmentación así como segmentar las imágenes grandes en varias más pequeñas y manejables, para luego ensamblar los resultados parciales en uno solo. Para ello fuera ideal medir el uso de recursos y establecer rangos de resolución mínimos y máximos manejables por la memoria y procesamiento del equipo desde donde se ejecute el programa.

Por ser un procedimiento relativamente pesado, fuera ideal proporcionar una medida del porcentaje del proceso completado, o una estimación del tiempo restante tanto para el proceso de generación de clasificadores como en la clasificación en si.

# Bibliografía

- [1] Aldalur, B. and Santamaría, M. (2002). Realce de imágenes: filtrado espacial. *Revista de teledetección*, 17:31–42.
- [2] Ballester, J. and Sueiro, J. (2011). Biomateriales y sustitutos óseos en traumatología y cirugía ortopédica. *Cádiz: Servicio de Publicaciones de la Universidad de Cádiz*.
- [3] Commons, W. (2018). File:rgb-raster-image.svg — wikimedia commons, the free media repository. [Online; accessed 27-August-2018].
- [4] De Boer, P., Hoogenboom, J. P., and Giepmans, B. N. (2015). Correlated light and electron microscopy: ultrastructure lights up! *Nature methods*, 12(6):503.
- [5] ESRI (2018). ¿qué son los datos ráster?
- [6] Guo, K. W. (2017). Biomaterials degradation and bioabsorbability: Biomedical potentials of marine enzymes. *Handbook of Ecomaterials*, pages 1–28.
- [7] Hastie, T. and Tibshirani, R. (1998). Classification by pairwise coupling. In Jordan, M. I., Kearns, M. J., and Solla, S. A., editors, *Advances in Neural Information Processing Systems*, volume 10. MIT Press.
- [8] Hulten, G., Spencer, L., and Domingos, P. (2001). Mining time-changing data streams. In *ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining*, pages 97–106. ACM Press.
- [9] Kalmegh, S. (2015). Analysis of weka data mining algorithm reptree, simple cart and randomtree for classification of indian news. *Int. J. Innov. Sci. Eng. Technol*, 2(2):438–446.
- [10] Kawaguchi, S. and Nishii, R. (2007). Hyperspectral image classification by bootstrap adaboost with random decision stumps. *IEEE Transactions on Geoscience and Remote Sensing*, 45(11):3845–3851.
- [11] Keerthi, S., Shevade, S., Bhattacharyya, C., and Murthy, K. (2001). Improvements to platt’s smo algorithm for svm classifier design. *Neural Computation*, 13(3):637–649.
- [12] Kohavi, R. (1995). The power of decision tables. In *8th European Conference on Machine Learning*, pages 174–189. Springer.
- [13] Kohavi, R. et al. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai*, volume 14, pages 1137–1145. Montreal, Canada.

- [14] Landwehr, N., Hall, M., and Frank, E. (2005). Logistic model trees. *Machine learning*, 59(1-2):161–205.
- [15] McAuliffe, M. J., Lalonde, F. M., McGarry, D., Gandler, W., Csaky, K., and Trus, B. L. (2001). Medical image processing, analysis and visualization in clinical research. In *Computer-Based Medical Systems, 2001. CBMS 2001. Proceedings. 14th IEEE Symposium on*, pages 381–386. IEEE.
- [16] Nagar, A. (2017). A comparative study of data mining algorithms for decision tree approaches using weka tool.
- [17] Nasution, A. K. and Hermawan, H. (2016). Degradable biomaterials for temporary medical implants. In *Biomaterials and Medical Devices*, pages 127–160. Springer.
- [18] Platt, J. (1998). Fast training of support vector machines using sequential minimal optimization. In Schoelkopf, B., Burges, C., and Smola, A., editors, *Advances in Kernel Methods - Support Vector Learning*. MIT Press.
- [19] Quinlan, R. (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, San Mateo, CA.
- [20] Rao, A. R. and Schunck, B. G. (1991). Computing oriented texture fields. *CVGIP: Graphical Models and Image Processing*, 53(2):157–185.
- [21] Robles-Bykbaev, Y., Naya, S., Prado, S. D., Calle-López, D., Robles-Bykbaev, V., Garzón-Muñoz, L., Rodríguez, C. S., and Saavedra, J. T. (2018a). An expert system based on computer vision and statistical modelling to support the analysis of collagen degradation. In *Intelligent System*. IntechOpen.
- [22] Robles-Bykbaev, Y., Naya, S., Tarrío, J., Díaz-Prado, S., Sanjurjo-Rodríguez, C., Blanco, F., Cornejo-Reyes, J., Garzón, L., Calle-López, D., and Robles-Bykbaev, V. (2018b). An educational environment based on digital image processing to support the learning process of biomaterials degradation in stem cells. In *IEEE XXV International Conference on Electronics, Electrical Engineering and Computing*.
- [23] Robles-Bykbaev, Y., Tarrío-Saavedra, J., Quintana-Pita, S., Díaz-Prado, S., García-Sabán, F. J., and Naya, S. (2018c). Statistical degradation modelling of poly(d,l-lactide-co-glycolide) copolymers for bioscaffold applications. *PLOS ONE*, 13(10):1–30.
- [24] Sewaiwar, P. and Verma, K. K. (2015). Comparative study of various decision tree classification algorithm using weka. *International Journal of Emerging Research in Management & Technology*, 4:2278–9359.
- [25] Sobel, I. (1990). An isotropic  $3 \times 3$  image gradient operator. *Machine vision for three-dimensional scenes*, pages 376–379.