

**Universidad Central “Marta Abreu” de Las Villas**

**Facultad de Ingeniería Eléctrica**

**Centro de Estudios de Electrónica y Tecnologías de la Información**



## **TRABAJO DE DIPLOMA**

### **Clasificación de leucocitos usando rasgos morfométricos del núcleo**

**Autor: Jorge E. Alfonso Dulzaides**

**Tutor: MSc. Alexander Falcón Ruiz**

**Santa Clara**

**2015**

**"Año 57 de la Revolución"**

**Universidad Central “Marta Abreu” de Las Villas**

**Facultad de Ingeniería Eléctrica**

**Centro de Estudios de Electrónica y Tecnologías de la Información**



## **TRABAJO DE DIPLOMA**

### **Clasificación de leucocitos usando rasgos morfométricos del núcleo**

**Autor: Jorge Enrique Alfonso Dulzaides**

jalfonso@uclv.edu.cu

**Tutor: MSc. Alexander Falcón Ruiz**

Profesor Auxiliar MSc., CEETI, Facultad de Ingeniería Eléctrica

afalcon@uclv.edu.cu

**Santa Clara**

**2015**

**"Año 57 de la Revolución"**



Hago constar que el presente trabajo de diploma fue realizado en la Universidad Central “Marta Abreu” de Las Villas como parte de la culminación de estudios de la especialidad de Ingeniería Biomédica, autorizando a que el mismo sea utilizado por la Institución, para los fines que estime conveniente, tanto de forma parcial como total y que además no podrá ser presentado en eventos, ni publicados sin autorización de la Universidad.

---

Firma del Autor

Los abajo firmantes certificamos que el presente trabajo ha sido realizado según acuerdo de la dirección de nuestro centro y el mismo cumple con los requisitos que debe tener un trabajo de esta envergadura referido a la temática señalada.

---

Firma del Tutor

---

Firma del Jefe de Departamento  
donde se defiende el trabajo

---

Firma del Responsable de  
Información Científico-Técnica

## PENSAMIENTO

*“La emoción más bella y profunda que puede sentir un hombre, es la sensación de lo misterioso en la ciencia. Quien no haya experimentado esta sensación, me parece, si no un muerto, por lo menos un ciego.”*

*Albert Einstein.*

## **DEDICATORIA**

A mi viejita linda, que aunque no esté presente sé que estaría muy orgullosa.

A mis padres, a mi hermano y a cutú, que son mi vida entera.

A mis amigos, siempre incondicionales.

## **AGRADECIMIENTOS**

A mis padres, que siempre me han ayudado en lo que he necesitado a lo largo de mis años de estudio.

Al mu y a la cutusuta, por confiar siempre en mí.

A mi tutor Alexander Falcón Ruiz, por apoyarme, guiarme y darme confianza para llegar hasta el final.

A mis profesores en la Facultad de Ingeniería Eléctrica, que durante estos 5 años supieron enseñarme con mucha serenidad y pasión, especialmente los profesores del CEETI.

A todas aquellas personas que de una forma u otra han contribuido a mi formación y a hacer realidad este sueño.

A mis amigos Osvaldo, Ameth, Tony, el Poty, Lisca, José Osvaldo, Lioni y a todos los demás que tantas batallas hemos librado juntos.

## TAREA TÉCNICA

- Realizar búsqueda sobre rasgos morfométricos y programarlos en Matlab
- Hacer revisión sobre técnicas de selección de rasgos y proponer estrategia a seguir
- Hacer revisión sobre técnicas de clasificación y seleccionar conjunto de clasificadores representativos dentro de varias técnicas para realizar las pruebas
- Realizar búsqueda de herramientas de software y código en Matlab, así como programar lo necesario para realizar la selección de rasgos y la clasificación
- Analizar los resultados obtenidos de manera crítica
- Escribir el informe

---

Firma del Autor

---

Firma del Tutor

## RESUMEN

El conteo de diferentes tipos de leucocitos provee información importante en el diagnóstico de diferentes enfermedades. La automatización de este proceso ahorra tiempo y evita errores en el conteo. En este trabajo exploramos cómo usando solo rasgos de forma del núcleo es suficiente para clasificar los leucocitos en sangre periférica. La segmentación del núcleo es más fácil que la segmentación de la célula entera. Las imágenes del núcleo fueron segmentadas manualmente. Filtros de procesamiento, seguidos de *wrappers*, fueron usados como métodos de selección de rasgos. Se usó la *cross-validation* en los experimentos en los cuales *Multi-Layer Perceptron*, *pair-wise Support Vector Machine*, *k-Nearest Neighbors*, *PART* and *C4.5* fueron aplicados como clasificadores. Los resultados muestran que usando sólo rasgos de forma del núcleo podemos lograr un porcentaje de clasificación de más del 96% para todos los clasificadores.

**Palabras claves:** clasificación automática de leucocitos, rasgos basados en contorno y basados en región, selección de rasgos.



## TABLA DE CONTENIDOS

PENSAMIENTO .....	i
DEDICATORIA .....	ii
AGRADECIMIENTOS .....	iii
TAREA TÉCNICA .....	iv
RESUMEN .....	v
INTRODUCCIÓN .....	1
Organización del informe .....	3
CAPÍTULO 1. REVISIÓN BIBLIOGRÁFICA.....	4
1.1 Microscopio.....	4
1.1.1 Tipos de microscopios de luz.....	4
1.1.2 Partes del microscopio y sus funciones .....	5
1.1.3 Otros tipos de microscopios empleados.....	7
1.1.4 Aplicaciones.....	8
1.1.5 Formación de la imagen a través del microscopio de luz .....	9
1.2 Cámara digital.....	9
1.3 Conexión de una cámara digital a un microscopio .....	10
1.4 Leucocitos .....	11
1.5 Conteo de leucocitos .....	13
CAPÍTULO 2. MATERIALES Y MÉTODOS.....	15
2.1 Adquisición y preprocesamiento de los datos.....	15
2.2 Extracción de rasgos .....	16
2.2.1 Rasgos basados en contorno .....	16

2.2.2	Rasgos basados en región .....	18
2.2.3	Propiedades geométricas.....	18
2.2.4	Momentos de Zernike .....	19
2.2.5	<i>Pecstrum</i> .....	20
2.3	Selección de rasgos .....	20
2.4	Clasificación.....	21
CAPÍTULO 3. RESULTADOS Y DISCUSIÓN .....		24
3.1	Selección de rasgos .....	24
3.2	Aplicación de los wrappers.....	25
3.3	Clasificación.....	27
CONCLUSIONES Y RECOMENDACIONES .....		28
Conclusiones .....		28
Recomendaciones .....		28
REFERENCIAS BIBLIOGRÁFICAS .....		29
ANEXOS .....		31
Anexo I	Programación en Matlab .....	31

## INTRODUCCIÓN

El conteo de diferentes tipos de leucocitos provee información importante en el diagnóstico de diferentes enfermedades como la leucemia, el cáncer y el SIDA. Tradicionalmente el experto selecciona una muestra de sangre periférica, y usando un microscopio, detecta los distintos tipos de leucocitos. Este proceso, realizado manualmente, resulta muy tedioso, consume mucho tiempo y es muy susceptible a errores. Un sistema de conteo automático ahorrará tiempo, a la vez que el experto puede realizar otros trabajos de mayor importancia.

La mayoría de los métodos utilizados para clasificar leucocitos usan rasgos del núcleo, el citoplasma y la célula. Rasgos computados de color, tamaño y textura del núcleo y el citoplasma son utilizados para clasificar los leucocitos. En [1] se usó un clasificador *Naïve Bayesian*, logrando errores de 10.93% a 25.91%.

Un vector de 26 rasgos de estadísticas de colores del núcleo y el citoplasma y rasgos de forma del núcleo, es clasificado por una máquina de vector soporte, logrando exactitudes entre 74.5% y 99.5%, dependiendo de la clase de leucocito [2].

En [3] se utilizan 57 vectores de rasgos de forma, color y textura del núcleo, citoplasma y la célula para clasificar 12 clases de leucocitos. Se usaron *k-Nearest Neighbor (k-NN)*, *Linear Vector Quantization (LVQ)*, *Multi-Layer Perceptron (MLP)* y *Support Vector Machine (SVM)*, logrando 80.76%, 83.33%, 89.74% y 91.03% de clasificación correcta respectivamente. En [4] los leucocitos son clasificados basados en mediciones cuantitativas de tamaño, curvatura y color del núcleo y el citoplasma con una proporción de éxito cerca del 91%. En [5] usa 7 componentes de *pecstrum* del modelo del núcleo, área de la célula normalizada y la proporción del núcleo-citoplasma para clasificar leucocitos, usando distancia Euclidiana, *k-NN*, red neuronal con *backpropagation* y *SVM*. Usando 36 modelos para entrenar y 18 para probar, fue obtenido un resultado del 87% de clasificación correcta.

En [6] 19 rasgos de forma y textura del núcleo y el citoplasma fueron computados para clasificar leucocitos. Se utiliza una red neuronal, logrando exactitud de clasificación de un 86%.

La segmentación de la célula y la extracción de rasgos pueden ser obstruidas por problemas en la preparación de la muestra y en los procesos de adquisición de la imagen. Las células se pueden dañar o estar incorrectamente teñidas debido a una mala manipulación en la preparación de la muestra. Por otro lado, en el proceso de adquisición de imágenes se pueden obtener imágenes borrosas, con ruido o con desplazamiento de colores. Estos factores negativos pueden invalidar la segmentación y algunos rasgos valiosos como el color y la textura. Sin embargo, en muchos casos la información del núcleo permanece intacta, y los especialistas tienen la información necesaria para clasificar la célula. Si la clasificación solo es basada en la información sobre la forma del núcleo, no se necesita segmentar la célula completa, solo con una segmentación del núcleo, que es mucho más fácil que la segmentación de la célula entera, se pueden obtener buenos resultados de clasificación.

En este trabajo se pretende clasificar leucocitos usando solo información del núcleo. Se extraen manualmente los rasgos de forma del núcleo. Se usó *MLP*, *pair-wise SVM*, *k-NN*, *PART* y *C4.5* como clasificadores para comparar los resultados con los obtenidos por el experto.

### **Objetivo general**

Determinar, de un conjunto de rasgos morfométricos basados en contorno y región, los que mejor tributan a la correcta clasificación de leucocitos en imágenes de microscopía óptica, empleando sólo información del núcleo.

### **Objetivos específicos**

- Obtener un conjunto de rasgos morfométricos basados en contorno y región.
- Crear una estrategia para la selección de rasgos que reduzca el conjunto inicial.
- Proponer un método de clasificación empleando los rasgos seleccionados.

**Organización del informe**

Este trabajo está estructurado de la siguiente forma: introducción, capitulario, conclusiones, recomendaciones, referencias bibliográficas y anexos. En el Capítulo 1 se tratan las características morfológicas de los leucocitos, los análisis de laboratorio que se realizan para estudiar los mismos, así como la descripción de las herramientas para la adquisición de las imágenes. El Capítulo 2 describe las herramientas utilizadas para la adquisición, los métodos para la adquisición de los datos y el pre-procesamiento de las imágenes. También se muestran los diferentes tipos de rasgos de forma del núcleo utilizados, el proceso de selección de rasgos y los clasificadores utilizados. En el Capítulo 3 se describen los resultados obtenidos en el transcurso del experimento con sus respectivas tablas. Además se muestran los gráficos de alguno de los rasgos calculados.

## CAPÍTULO 1. REVISIÓN BIBLIOGRÁFICA

En este capítulo se tratan las características morfológicas de los leucocitos, los análisis de laboratorio que se realizan para estudiar los mismos, así como la descripción de las herramientas para la adquisición de las imágenes.

### 1.1 Microscopio

El estudio detallado de los componentes de células y tejidos animales o vegetales, por el tamaño que poseen, requiere el uso de instrumentos que permitan ampliar muchas veces más la imagen de las estructuras que los constituyen. El instrumento que fue empleado por los primeros biólogos para estudiar la célula y los tejidos, es el microscopio. El nombre deriva etimológicamente de dos raíces griegas: *mikrós*, que significa pequeño y *skopéoo*, que significa observar. Es decir, el microscopio es un instrumento que sirve para observar objetos o estructuras pequeñas [7].

#### 1.1.1 Tipos de microscopios de luz

Existen dos tipos de microscopios que emplean la luz como fuente de energía para formar imágenes aumentadas y detalladas de objetos que a simple vista no es posible observar:

- a) Microscopio simple o lupa.
- b) Microscopio compuesto.

El microscopio simple o lupa es un instrumento de amplificación de imágenes que consiste en la utilización de una o más lentes convergentes en un solo sistema óptico. Dependiendo de la curvatura de la superficie de la(s) lente(s), las lupas pueden ampliar las imágenes de los objetos desde 5, 8, 10, 12, 20 y hasta 50 veces. Forman una imagen de mayor tamaño, derecha y virtual.

Los microscopios compuestos que se emplean actualmente tienen sus antecesores en los instrumentos ópticos desarrollados, en el periodo comprendido entre 1590 y 1610, por Hans (padre) y Zacarías (hijo) Janssen; quienes mediante el tallado cuidadoso de lentes biconvexas construyeron los primeros microscopios compuestos.

A partir de esa época, el microscopio es el instrumento más utilizado en el estudio de células y tejidos. Se fue perfeccionando, tanto en su parte óptica como en su parte mecánica, gracias a los adelantos tecnológicos aplicados a los componentes antes mencionados.

Se denominan compuestos porque la imagen se forma mediante la utilización de tres sistemas de lentes, cada uno de ellos constituidos por lentes convergentes y divergentes. Los sistemas de lentes son el condensador, los objetivos y los oculares.

En la actualidad, el microscopio es un instrumento de uso cotidiano en los laboratorios de investigación y de diagnóstico, así como en las aulas de enseñanza de Biología, Embriología, Histología, Microbiología y Patología [7].

### **1.1.2 Partes del microscopio y sus funciones**

(a) Componentes mecánicos: Son aquellos que sirven de sostén, movimiento y sujeción de los sistemas ópticos y de iluminación así como de los objetos que se van a observar. Éstos se muestran en la imagen del microscopio de la Figura 1.

- Base o pie. Es un soporte metálico, amplio y sólido en donde se apoyan y sostienen los otros componentes del microscopio.
- Brazo, estativo o columna. Permite la sujeción y traslado del microscopio. Soporta al tubo óptico, a la platina y el revólver.
- Platina. Superficie plana de posición horizontal que posee una perforación circular central. En ella se apoya la preparación (lámina portaobjetos que contiene a la muestra que se va a examinar) que se sujeta a la platina mediante pinzas o con un carrito o charriot que, mediante mandos especiales facilitan el movimiento de la preparación de derecha a izquierda y de adelante hacia atrás.
- Tubo óptico. Consiste en un cilindro metálico que suele medir 160mm o 170 mm de longitud (dependiendo del fabricante del microscopio) el cual en un

extremo, está conectado al revolver o portaobjetivos y en el otro se relaciona con el (los) ocular(es).

- Revolver o portaobjetivos. Es un componente que gira alrededor de un eje con la finalidad que los objetivos que sostiene coincidan de manera perpendicular con la perforación central de la platina. En su superficie inferior posee varios agujeros donde se atornillan los objetivos.
- Tornillos macrométrico y micrométrico. Generalmente están situados en la parte inferior del brazo o columna. Pueden estar separados (en los microscopios antiguos) o el tornillo micrométrico está incorporado en la circunferencia del tornillo macrométrico (microscopios actuales). Ambos tornillos permiten el desplazamiento de la platina hacia arriba y hacia abajo con la finalidad de acercar o alejar la preparación hacia los objetivos y así conseguir un enfoque óptimo de la imagen.
- Engranajes y cremallera. Constituyen mecanismos de desplazamientos de las diferentes partes del microscopio.
- Cabezal. Es un componente situado en relación con el tubo del microscopio que alberga principalmente prismas o espejos que sirven para acondicionar en él dos o más oculares, o sistemas mecánicos que soportan cámaras fotográficas, de vídeo o sistemas de proyección de la imagen [7].

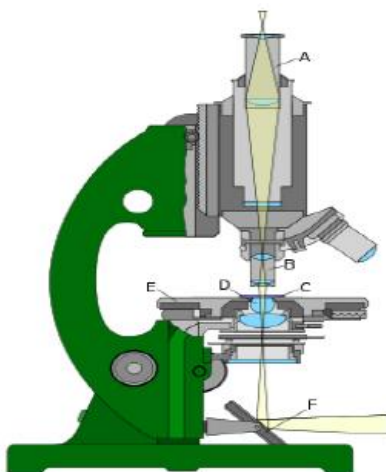


Figura 1. Microscopio: A) ocular, B) objetivo, C) portador del objeto, D) lentes de la iluminación, E) sujeción del objeto, F) espejo de la iluminación.



(b) Componentes ópticos: Son los objetivos, los oculares, el condensador y los prismas.

Los tres primeros están constituidos por sistemas de lentes positivos y negativos.

- Condensador: Es el componente óptico que tiene como función principal concentrar y regular los rayos luminosos que provienen de la fuente luminosa.
- Objetivos. Los objetivos están considerados los elementos más importantes en la formación de la imagen microscópica, ya que estos sistemas de lentes establecen la calidad de la imagen en cuanto a su nitidez y la capacidad que tiene para captar los detalles de la misma (poder de resolución). Están constituidos también por un juego de lentes, en este caso, convergente y divergente, para eliminar, en la medida de lo posible, una serie de aberraciones que afectarían la calidad de las imágenes formadas [7].
- Oculares. Están formados por dos lentes que se encuentran separadas por un diafragma. Para hacer más cómoda la observación, existen microscopios que tienen dos oculares en lugar de uno. Ello permite observar con los dos ojos a la vez. Actualmente existen cabezales múltiples, que permiten que varias personas puedan observar una misma muestra a la vez.
- Prismas. En los microscopios modernos monoculares o binoculares, se requiere el empleo de prismas, estructuras transparentes que, en el caso de los microscopios monoculares, sirven para desviar los rayos luminosos de la trayectoria rectilínea del eje óptico del objetivo y dirigirlos hacia el tubo óptico ligeramente inclinado y luego hacia el ocular [8].

### 1.1.3 Otros tipos de microscopios empleados

- Microscopio de transparencia o de campo claro. Se caracteriza porque emplea luz natural o luz artificial como energía luminosa para formar las imágenes del objeto que se observa. La imagen muestra puntos o áreas iluminadas (generalmente coloreados) sobre un fondo claro o transparente.
- Microscopio de campo oscuro. Se denomina así porque la imagen que se forma está constituida por una serie de estructuras brillantes sobre un fondo oscuro.

En el microscopio de transparencia o de campo claro, para que la imagen sea visible con nitidez, es necesario que el objeto examinado este coloreado o teñido, es decir que los componentes celulares y tisulares de la estructura se contrasten mediante colorantes específicos que absorban y transmitan determinadas longitudes de onda del espectro visible. Las longitudes de onda que no se absorban son transmitidas al ojo humano o a un material fotográfico sensible dando como resultado que la estructura teñida aparece de un determinado color. El campo microscópico aparece claro o transparente porque los rayos luminosos directos que provienen del condensador no encuentran en su camino ninguna estructura coloreada y entran como rayos de luz blanca hacia el objetivo. Si se examinan objetos sin colorear la imagen ofrecerá detalles poco contrastados, casi transparentes. Las secciones de tejidos deben ser delgadas con un grosor de 5  $\mu\text{m}$  a 10  $\mu\text{m}$ . Se puede emplear secciones de mayor espesor pero la imagen no mostrará un buen poder de resolución y tampoco exhibirá contornos nítidos, pues en el campo microscópico se observarán varios planos superpuestos de las imágenes.

El principio óptico del microscopio de campo oscuro es el de aprovechar un conjunto de rayos luminosos oblicuos que inicialmente no entran a la lente frontal del objetivo, observándose el campo microscopio totalmente oscuro. Para que esto ocurra se requiere en primer lugar que la apertura numérica del condensador sea mayor que la apertura numérica del objetivo. Esta condición facilita que los rayos luminosos directos provenientes de la fuente luminosa sean impedidos de entrar a la porción central de la lente del condensador y solo penetren y emerjan de él, los rayos periféricos que al refractarse se hacen oblicuos [7].

#### **1.1.4 Aplicaciones**

El microscopio de luz ha sido de gran utilidad en el campo de la medicina, sobre todo en los campos de la ciencia en donde la estructura y la organización microscópica es importante, incorporándose con éxito a investigaciones dentro del área de la química, la física, la geología y en el campo de la biología (en el estudio de estructuras microscópicas de la materia viva), por citar algunas disciplinas de la ciencia.

Hasta ahora se utilizan en los laboratorios de histología y anatomía patológica, donde la microscopía permite determinadas aplicaciones diagnósticas, entre ellas el diagnóstico con

certeza del cáncer, numerosas estructuras cristalinas, pigmentos, lípidos, proteínas, depósitos óseos, depósitos de amiloide, etcétera [7].

### **1.1.5 Formación de la imagen a través del microscopio de luz**

La imagen total del microscopio de luz se forma mediante las imágenes que generan sucesivamente el objetivo y el ocular. Para que se forme una imagen del objeto observado es indispensable que por lo menos dos rayos luminosos que inciden en el objeto iluminen una porción del mismo. Un punto iluminado del objeto trasmite dos rayos luminosos: uno, orientado paralelamente al eje principal, se refracta al atravesar la lente del objetivo, en tanto que el otro se traslada en dirección oblicua hacia el centro óptico de la lente y la atraviesa sin refractarse.

En el lugar donde los dos rayos luminosos se interceptan se formará la imagen del punto iluminado. Si del objeto se transmiten dos rayos luminosos por cada punto que lo constituyen igual número de puntos luminosos formarán la imagen. Con la diferencia que la imagen será ampliada  $N$  número de veces correspondiente al aumento propio que posea el objetivo.

Los otros rayos luminosos que llegan a la lente frontal del objetivo y que no iluminan ningún punto del objeto, al atravesar la lente, le confieren al campo microscópico una iluminación uniforme. A este tipo de iluminación se denomina de campo claro. La imagen generada por el objetivo es captada posteriormente por la lente de campo del ocular y la amplía el número de veces del aumento que aquel posee. La imagen formada por el ocular será de mayor tamaño, derecha con relación a la imagen formada por el objetivo y virtual.

La imagen total formada por ambas lentes será: aumentada de tamaño, invertida y virtual con relación al objeto [7].

## **1.2 Cámara digital**

Una cámara digital es una cámara fotográfica que captura y almacena fotografías digitalmente mediante un dispositivo electrónico, o en cinta magnética usando un formato analógico como muchas cámaras de video.

La resolución de una cámara fotográfica digital está limitada por su sensor, generalmente un *charge-coupled device* (CCD) o *complementary metal oxide semiconductor* (CMOS), que

responde a las señales de luz, substituyendo el trabajo de la película en la fotografía tradicional. El sensor se compone de millones de cubos que se cargan en respuesta a la luz. Generalmente, estos cubos responden solamente a una gama limitada de longitudes de onda ligeras, debido a un filtro de color sobre cada cubo. Cada uno de estos es un pixel, y se utiliza un algoritmo de mosaicismo e interpolación para unir la imagen de cada gama de longitud de onda por pixel en una imagen del RGB donde están las tres imágenes por pixel para representar un color completo. Los dispositivos CCD transportan la carga a través del chip hasta un conversor analógico-digital. Éste convierte el valor de cada uno de los pixeles en un valor digital midiendo la carga que le llega. Dependiendo del número de bits del conversor obtendremos una imagen con mayor o menor gama de color. Por ejemplo, si se utilizase un sólo bit tendríamos valores de 0 y 1, y sólo podríamos representar presencia o ausencia de luz, lo que supondría una imagen en blanco y negro puro [10].

Por otro lado, los aparatos CMOS contienen varios transistores en cada pixel. El proceso de conversión digital se produce en la propia estructura del sensor, por lo que no se necesita un conversor añadido. Su proceso de fabricación es más sencillo, y hace que las cámaras que utilizan esta tecnología resulten más baratas. La cantidad de pixeles resultante en la imagen determina su tamaño. Existen varios factores que afectan la resolución de un sensor como puede ser el tamaño del sensor, la calidad de la lente y la organización de los pixeles.

Los pixeles excesivos pueden incluso conducir a una disminución de la calidad de la imagen. Cada sensor del pixel es tan pequeño que recoge muy pocos fotones, y así la relación señal-ruido (SNR) disminuirá. Esta disminución conduce a cuadros ruidosos, calidad pobre en sombras y generalmente a imágenes de pobre calidad [10].

### **1.3 Conexión de una cámara digital a un microscopio**

Un adaptador óptico mecánico es importante en fotografía digital. Dicho adaptador sirve de enlace entre la cámara y el microscopio. Es importante que la conexión mecánica sea firme, pues cualquier movimiento mínimo o vibraciones de la cámara, reduciría la calidad de la imagen notablemente. Adicionalmente, se requiere un adaptador óptico para el trayecto de luz, con el que se logrará así que el sensor CCD/CMOS de la cámara proyecte una imagen de total nitidez e iluminación.

El procedimiento para adaptar la cámara al microscopio consiste en desmontar el lente más exterior del ocular del microscopio y colocar la cámara ensamblada con los accesorios adecuados en lugar de este, sujetándola con tornillos de presión.

Para enfocar la imagen se mueven manualmente las perillas de los micrómetros del microscopio hasta tener una visión clara de la imagen en el monitor. No obstante, existen variaciones en la iluminación que no se logran controlar [11].

#### 1.4 Leucocitos

Los leucocitos son células móviles que se encuentran en la sangre transitoriamente, así, forman la fracción celular de los elementos figurados de la sangre. Son los representantes hemáticos de la serie blanca. A diferencia de los eritrocitos (glóbulos rojos), no contienen pigmentos, por lo que se les califica de glóbulos blancos [12].

Son células con núcleo, mitocondrias y otros orgánulos celulares. Son capaces de moverse libremente mediante pseudópodos. Su tamaño oscila entre los 8 y 20  $\mu\text{m}$  (micrómetros). Su tiempo de vida varía desde algunas horas, meses y hasta años. Estas células pueden salir de los vasos sanguíneos a través de un mecanismo llamado diapédesis (prolongan su contenido citoplasmático), esto les permite desplazarse fuera del vaso sanguíneo y poder tener contacto con los tejidos del interior del cuerpo humano. Los leucocitos son las células encargadas de defender al organismo de las infecciones y ayudar a eliminar los residuos y desechos de los tejidos. Se producen y se almacenan en la médula ósea y salen a la sangre cuando el organismo los necesita. Son un componente importante de la sangre y una pieza clave en el sistema inmunológico del cuerpo. En individuos sanos, representan alrededor del 1% de la sangre. La proporción de leucocitos en la sangre se puede utilizar como un diagnóstico de muchos tipos de enfermedades [13].

Una elevación en leucocitos en la sangre puede indicar la presencia de infección o de una enfermedad subyacente. Esta condición se conoce como **leucocitosis**. Por otra parte, estas células se reducen en número debido a una enfermedad, exposición a la radiación, o deficiencia de la médula ósea. Esta condición se conoce como **leucopenia**.

La duración de la vida de los leucocitos maduros varía de aproximadamente unas pocas horas a varios días. La producción de células de la sangre a menudo se regula por las estructuras del cuerpo como los ganglios linfáticos, el bazo, el hígado y los riñones. Durante una infección o lesión, se producen más glóbulos blancos y están presentes en la sangre [14].

La cifra normal de leucocitos es de 5.000 a 10.000 por milímetro cúbico y hay cinco tipos distintos [14]:

- **Neutrófilos.** Son los leucocitos más numerosos y a los que nos referimos normalmente cuando hablamos de granulocitos. Constituyen cerca del 60-70% de leucocitos. Permanecen en la sangre unos pocos días, ya que su función consiste en localizar y neutralizar a las bacterias o células dañadas en los tejidos, de tal forma que cuando las encuentran en un tejido las digieren, se rompen y liberan sustancias que hacen que aumente la circulación de sangre en la zona y atraen a más neutrófilos, lo que provoca que la zona esté enrojecida y caliente. Son el primer tipo de célula inmune que responde y llega al sitio de la infección [15].
- **Linfocitos.** Constituyen alrededor del 20-40% del recuento total de leucocitos, e incluyen los linfocitos B, linfocitos T y células NK. Se forman en la médula ósea, pero luego se dirigen a los ganglios linfáticos, bazo, amígdalas, timo y en realidad a cualquier parte del cuerpo. Al contrario que los granulocitos, viven mucho tiempo y maduran y se multiplican ante estímulos determinados. No sólo luchan contra las infecciones sino que son células muy especializadas en el sistema inmune. Pueden defender el cuerpo contra las infecciones, ya que distinguen las células del propio cuerpo de las extranjeras [15].
- **Monocitos.** Constituyen de un 5% al 12% del total de glóbulos blancos en la sangre. Su función también es de defensa, destruyendo y digiriendo células infectadas o dañadas. Pero también tienen otras importantes funciones, pues al igual que los linfocitos se dirigen a los diferentes tejidos (la piel, los pulmones, el hígado o el bazo), en los que ejercen distintas funciones como macrófagos (células que engullen y procesan todos los desechos de células

moribundas) o se convierten en células especializadas, como los osteoclastos, que remodelan el tejido óseo envejecido [15].

- **Eosinófilos.** Son los encargados de responder a las reacciones alérgicas. Lo que hacen es inactivar las sustancias extrañas al cuerpo para que no causen daño, y también poseen gránulos tóxicos que matan a las células invasoras y limpian el área de inflamación. El porcentaje normal en sangre es del 2 al 10% [15].
- **Basófilos.** También intervienen en las reacciones alérgicas, liberando histamina, sustancia que aumenta la circulación sanguínea en la zona para que aparezcan otro tipo de glóbulos blancos y, además, facilitan que estos salgan de los vasos sanguíneos y avancen hacia la parte dañada. También liberan heparina, una sustancia que disuelve los coágulos. En la sangre representan menos del 2% en condiciones normales. Desempeñan un papel en la respuesta inmune [15].

En la Figura 2 se muestran los distintos leucocitos y sus respectivos núcleos.



Figura 2. De izquierda a derecha: linfocito, monocito, neutrófilo, basófilo y eosinófilo.

### 1.5 Conteo de leucocitos

Los recuentos diferenciales tienen como finalidad obtener información de la distribución relativa de los diferentes leucocitos en sangre periférica. El porcentaje obtenido, o más adecuadamente, el número absoluto de cada célula son de gran utilidad para el diagnóstico de múltiples enfermedades, orienta en la elección del tratamiento a seguir en los pacientes y pone de manifiesto la existencia de efectos secundarios en pacientes que reciben quimioterapia o radioterapia [12].

Los laboratorios clínicos utilizan diferentes procedimientos para efectuar el recuento diferencial de los leucocitos, los más empleados se basan en la preparación de extensiones sanguíneas sobre cubreobjetos o portaobjetos. Actualmente se dispone de técnicas semi-automatizadas para la preparación de extensiones sanguíneas con algunas ventajas sobre los métodos tradicionales [12] [16]. Como una alternativa para disminuir el error intrínseco de los métodos manuales, se ha hecho extensivo el uso de analizadores hematológicos, los cuales se basan en el conteo y la diferenciación de un gran número de leucocitos [14]. No obstante que estos analizadores producen resultados bastante precisos para todos los parámetros hematológicos, en muchas ocasiones, los resultados de la cuenta diferencial deben ser corroborados mediante una inspección visual de las células, especialmente en los casos donde se sospecha la existencia de una patología [17].

Estos procedimientos tradicionales empleados para el recuento diferencial de leucocitos, y que están basados en la preparación de extensiones sanguíneas, se caracterizan por un elevado error intrínseco, reflejado por una imprecisión e inexactitud significativa [18].

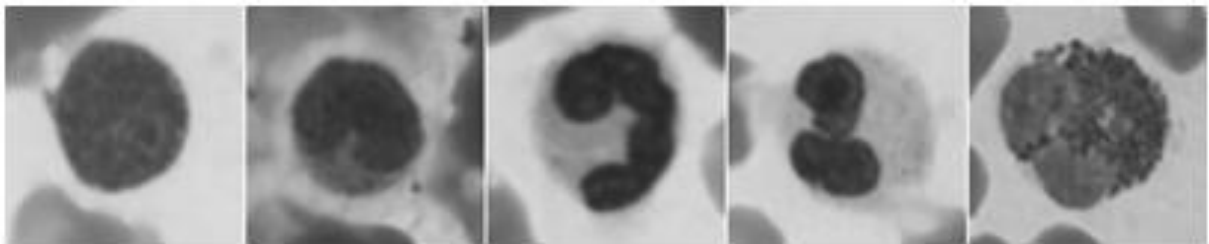


## CAPÍTULO 2. MATERIALES Y MÉTODOS

En este capítulo se describen los procesos para la adquisición de los datos y el preprocesamiento de las imágenes. También se muestran los diferentes tipos de rasgos de forma del núcleo utilizados, el proceso de selección de rasgos y los clasificadores implementados.

### 2.1 Adquisición y preprocesamiento de los datos.

Las imágenes usadas fueron adquiridas empleando una cámara digital Micrometrics 318CU CMOS, resultando en imágenes a color de 24 bits y de tamaño 2048H x 1536V. La cámara está acoplada a un microscopio trinocular Accuscope 3016PL con un objetivo en inmersión de aceite de 100x y ocular 10x. Las imágenes digitales fueron clasificadas por un hematólogo en las 5 clases de interés: linfocitos (L), monocitos (M), neutrófilos (N), basófilos (B) y eosinófilos (E). Imágenes recortadas de cada una de estas clases son mostradas en la Figura 3a. La base de datos contiene 224 imágenes de células, pero un subconjunto más pequeño fue considerado para tratar con el desbalance de los datos. Específicamente se trabajó con  $L = 40$ ,  $M = 33$ ,  $N = 40$ ,  $B = 38$  y  $E = 33$ .



(a)

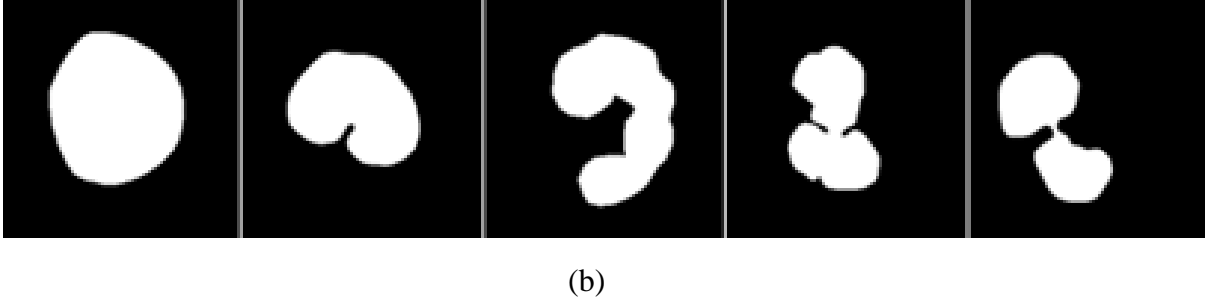


Figura 3. Leucocitos. (a) De izquierda a derecha: linfocito, monocito, neutrófilo, basófilo y eosinófilo; (b) Núcleos segmentados correspondientes.

Cada imagen fue recortada manualmente para formar una imagen de célula simple (como se muestra en la Figura 3a). Luego, cada núcleo de esta imagen de célula simple fue segmentado manualmente para obtener una imagen de máscara. En la Figura 3b se muestran los núcleos segmentados manualmente de las imágenes de la Figura 3a. Después de esto, para proporcionar invarianza a la traslación, el centroide del núcleo segmentado fue movido al centro de la imagen. Finalmente, todas las imágenes fueron redimensionadas al mismo tamaño (100 x 100 píxeles).

## 2.2 Extracción de rasgos

Rasgos de forma son categorizados, generalmente, en 2 grupos: basados en contorno y basados en región [19]. Basados en contorno, los rasgos solamente se concentran en las líneas de límite, mientras que los rasgos basados en región consideran el área de todo el objeto. En este trabajo se experimenta con rasgos de forma del núcleo en cada uno de estos grupos. Tanto preprocesamiento como extracción de rasgos, fueron realizados usando Matlab R2009a.

### 2.2.1 Rasgos basados en contorno

Para clasificar leucocitos se usaron los rasgos de forma usados en [20]. El límite del núcleo puede ser representado por secuencias de la distancia Euclidiana entre el centroide y todos los píxeles límites del núcleo [21]. Luego los  $N$  píxeles límites son descritos en orden par  $[x(i), y(i)]$ ,  $i = 1, 2, \dots, N$ . La distancia Euclidiana  $z(i)$  entre el centroide y el píxel límite  $[x(i), y(i)]$  está dada por:

$$z(i) = \sqrt{[x(i) - \bar{x}]^2 + [y(i) - \bar{y}]^2}, \quad (1)$$

donde  $[\bar{x}, \bar{y}]$  es el centroide del núcleo segmentado.

Basados en  $z(i)$  se pueden calcular los siguientes rasgos:

- (a) Desviación de la forma global: la desviación es una medida de la asimetría de los datos alrededor de la muestra. Si la desviación es negativa, los datos son arrojados más a la izquierda de la muestra. En caso de que la desviación sea positiva, pues entonces los datos son arrojados más a la derecha. La desviación de  $z(i)$  con respecto a la media es definida por:

$$F_1 = \frac{\frac{1}{N} \sum_{i=1}^N [z(i) - \bar{z}]^3}{\frac{1}{N} \sum_{i=1}^N z(i)^3} \quad (2)$$

donde  $\bar{z}$  es la media de  $z(i)$ .

- (b) Porción de mín/máx de  $z(i)$ . Esta está definida por:

$$F_2 = \frac{\min[z(i)]}{\max[z(i)]} \quad (3)$$

- (c) Aspereza del límite del núcleo: se define por:

$$F_3 = \frac{\text{Num}\{z(i) < \bar{z}\}}{N} \quad (4)$$

donde Num { } significa número de pixeles.

- (d) Redondez del área del núcleo segmentado: definido por:

$$F_4 = \frac{A_E}{A} \quad (5)$$

donde A es el área del contorno del núcleo y  $A_E$  es la región de intersección entre la parte fuera del círculo de radio  $\bar{z}$  y la parte de adentro del contorno del núcleo.

- (e) Varianza circular: representa como una forma, es similar a un círculo y está definida por la ecuación:

$$F_5 = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N [z(i) - \bar{z}]^2}}{\bar{z}} \quad (6)$$

- (f) Varianza elíptica: es un error de cartografía de forma de una elipse que tiene una matriz de covarianza igual a la de la forma. Está definida por:

$$F_6 = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N [\dot{z}(i) - \bar{\dot{z}}]^2}}{\bar{\dot{z}}} \quad (7)$$

donde  $\dot{z}(i) = \sqrt{V_i^T C^{-1} V_i}$ ,  $V_i = \begin{pmatrix} x(i) - \bar{x} \\ y(i) - \bar{y} \end{pmatrix}$  y  $C = \begin{pmatrix} C_{xx} & C_{xy} \\ C_{yx} & C_{yy} \end{pmatrix}$  es la matriz de covarianza del contorno de la forma.

### 2.2.2 Rasgos basados en región

Se probaron 3 grupos diferentes de rasgos basados en la región. El primero está basado en las propiedades geométricas de la imagen, el segundo está basado en el *pecstrum* del modelo morfológico de la región de interés y el tercero son los momentos de Zernike [22].

### 2.2.3 Propiedades geométricas

Una forma puede describirse por diferentes aspectos [20]. Estos parámetros son la excentricidad, solidez, extensión, factor de forma, convexidad, número de regiones disjuntas, entre otros.

En este grupo se consideraron los siguientes rasgos:

- (g) Excentricidad. Es la medida de proporción de aspecto. Su valor está entre 0 y 1 (0 es un círculo y 1 es un segmento). Se calcula como:

$$F_7 = \frac{D_F}{D_{MA}} \quad (8)$$

donde  $D_F$  es la distancia entre el foco de la elipse y  $D_{MA}$ , que es la longitud del eje mayor.

- (h) Solidez. La solidez describe hasta qué punto la forma es convexa o cóncava y se define por:

$$F_8 = \frac{A_R}{A_C} \quad (9)$$

donde  $A$  es el área de la forma de la región y  $A_C$  es el área del polígono convexo.

- (i) Extensión. Es la proporción de los pixeles del rectángulo que envuelve la figura que limita también la región, definida por el área de la región A dividido por el área del rectángulo límite de  $A_B$ .

$$F_9 = \frac{A_R}{A_B} \quad (10)$$

- (j) Factor de forma. Se define por (igual 1 para un objeto de círculo perfecto):

$$F_{10} = \frac{4\pi A_R}{P_R^2} \quad (11)$$

donde A y P son el área y el perímetro respectivamente de los núcleos.

- (k) Convexidad. Es la proporción de los perímetros del área envolvente del polígono  $P_C$  sobre la región de P.

$$F_{11} = \frac{P_C}{P_R} \quad (12)$$

- (l) El número de regiones disjuntas en el área del polígono envolvente que no está en la región, calculado encima de la imagen representada como:

$$F_{12} = \text{NumD}\{C \wedge \sim R\} \quad (13)$$

donde C es la imagen del polígono envolvente, R es la región de la imagen y NumD { } es el número de regiones disjuntas en la imagen resultante.

#### 2.2.4 Momentos de Zernike

Los momentos Zernike son momentos ortogonales. Los momentos Zernike complejos son derivados de los polinomios ortogonales de Zernike [23].

$$V_{nm}(x, y) = V_{nm}(r \cos \theta, \sin \theta) = R_{nm}(r) \exp(jm\theta) \quad (14)$$

donde  $R_{nm}(r)$  es el polinomio ortogonal radial:

$$R_{nm}(r) = \sum_{s=0}^{\frac{(n,|m|)}{2}} (-1)^s \frac{(n-s)!}{s! * \left(\frac{n-2s+|m|}{2}\right)! * \left(\frac{n-2s-|m|}{2}\right)!} * r^{n-2s} \quad (15)$$

$n = 0, 1, 2, \dots; 0 \leq |m| \leq n; n - |m|$  es igual.

Los polinomios Zernike son funciones ortogonales de valor complejo por encima del disco unitario.

El momento Zernike de orden  $n$  con  $m$  repeticiones de la región de forma  $f(x,y)$  está dado por la ecuación:

$$Z_{nm} = \frac{n+1}{\pi} \sum_r \sum_{\theta} f(r \cos \theta, r \sin \theta) * R_{nm}(r) \exp(jm\theta) \quad r \leq 1 \quad (16)$$

### 2.2.5 Pecstrum

El *pecstrum* de una imagen binaria  $I \subseteq R^2$ , relativo a una imagen binaria  $E \subseteq R^2$  llamado elemento estructurante, puede ser calculado como:

$$P(I, E, n) = \frac{M[I \circ nE] - M[I \circ (n+1)E]}{M[I]} \quad (17)$$

donde  $M$  representa el área medida en las operaciones intermedias,  $nE$  es las  $n$ -veces que dilata el elemento estructurante e  $I \circ nE$  indica la función de apertura (erosión seguida por una dilatación que usa el mismo elemento estructurante). El *pecstrum* descompone la imagen designada en los componentes morfológicos según la forma y el tamaño del elemento estructurante, proporcionando un análisis cuantitativo del volumen morfológico de la imagen. Este *pecstrum* tiene la propiedad de invarianza de rotación y traslación cuando es un elemento estructurante isotrópico. La balanza es determinada por el elemento estructurante.

Los momento *pecstrum*  $\mu^{(k)}(I)$ , son variables aleatorias denominadas momentos granulométricos. En casos discretos, los momentos granulométricos  $k^{\text{th}}$  pueden ser calculados por la fórmula:

$$\mu^{(k)}(I) = \frac{\sum_{n=1}^{\infty} n^k P(I, E, n)}{\sum_{n=1}^{\infty} P(I, E, n)} \quad (18)$$

Se consideraron 9 valores de *pecstrum* del modelo morfológico utilizando un disco de tamaño 5 como elemento estructurante.

## 2.3 Selección de rasgos

Se utilizaron algunos de los algoritmos de selección de rasgos implementados en Waikato Environment for Knowledge Analysis (WEKA) [24] para eliminar los rasgos no pertinentes

[25]. En todos los casos se desecharon los rasgos que se encontraban por debajo del 50% del valor máximo de mérito. Luego se utilizó un método de envoltura (*wrappers*) para reducir el tamaño del subconjunto, intentando con varios clasificadores que son lo que se emplean finalmente en la etapa de clasificación [26].

Los algoritmos de selección de rasgos utilizados fueron:

- **ChiSquaredAttributeEval.** Evalúa el valor de un atributo computando el valor de la estadística chi-cuadrada con respecto a la clase.
- **GainRatioAttributeEval.** Evalúa el valor de un atributo midiendo la proporción de ganancia con respecto a la clase.
- **InfoGainAttributeEval.** Evalúa el valor de un atributo midiendo la ganancia de información con respecto a la clase. Discretiza atributos numéricos que utilizan primero el método de discretización basado en *minimum description length* (MDL).
- **ReliefFAttributeEval.** Evalúa el valor de un atributo probando un caso repetidamente y considerando el valor del atributo dado para el caso más cercano de la misma y diferente clase. Opera en los datos de la clase discretos y continuos. Los parámetros especifican el número de casos para probar, el número de vecinos para verificar, si para pesar los vecinos por la distancia, y una función exponencial que gobierna el decaimiento de pesos rápidamente con la distancia.
- **OneRAttributeEval.** Este método evalúa el valor de un atributo usando el clasificador OneR. Puede usar los datos de entrenamiento para la evaluación, como lo hace OneR, o puede aplicar la validación cruzada interna. El número de pliegues es un parámetro. OneR adopta el método de discretización: el tamaño del cubo mínimo es un parámetro.
- **SymmetricalUncertAttributeEval.** Este método evalúa el valor de un atributo midiendo la incertidumbre simétrica con respecto a la clase [24].

## 2.4 Clasificación

Se probaron algunos clasificadores representativos de cada familia implementados en WEKA basados en las diferentes técnicas [24].

- **Multi-LayerPerceptron:** Red neuronal que usa la propagación para clasificar los casos. Se usaron los parámetros por defecto.

- **SMO:** Implementa un algoritmo de optimización mínimo secuencial para entrenar un clasificador de vector soporte.
- **IBk:** Clasificador k-NN. El valor apropiado de k se seleccionó automáticamente basado en la validación cruzada y la distancia euclidiana se usó como una función de distancia.
- **PART:** Construye un árbol parcial en cada iteración.
- **J48:** Usa escenas por defecto.

Multi-LayerPerceptron es una red neuronal que entrena utilizando *backpropagation*. Varios parámetros pueden estar listos desde el objeto editor para controlar su funcionamiento. Con *autoBuill* se agregan las capas fijas y las ocultas y son conectadas. El parámetro *hiddenLayers* define qué capas ocultas están presentes y cuántos nodos contiene cada una [24].

El SMO lleva a cabo el algoritmo de optimización mínimo secuencial para entrenar un clasificador de vectores soporte, usando funciones de kernel polinómicas o Gaussianas. Los coeficientes en la salida son basados en los datos normalizados. Esta normalización puede ser apagada, o la salida estandarizada a media 0 y varianza 1. La clasificación de *pair-wise* es utilizada para problemas multiclases. Los modelos de regresión lógica pueden ser atacados por la máquina de vector soporte para obtener las estimaciones de probabilidad. En el caso multiclases, la predicción probabilística se acoplarán. Cuando se trabaja con casos esparcidos, se apaga la normalización para un funcionamiento más rápido [24].

IBk es un clasificador de k vecinos cercanos. Una variedad de algoritmos diferentes de búsqueda pueden ser usados para acelerar la tarea de encontrar los vecinos más cercanos. El valor por defecto es igual que para IB1 que es la distancia euclidiana. El número de los vecinos más cercanos (k predefinido = 1) puede especificarse explícitamente en el editor del objeto o automáticamente determinado usando validación cruzada, dejando uno afuera sujeto a un límite superior dado por el valor especificado [24].

PART obtiene las reglas de los árboles de decisión parciales. Construye el árbol usando la heurística C4.5 con los mismos parámetros de usuario definidos como J48. Tres reglas son encontradas y son destinadas a ser procesadas en orden, la predicción generó, para cualquier caso de prueba, que es el resultado de la primera regla que dispara. La última regla, “*catch-*



*all'*, siempre dispara. Como con J48, los números en paréntesis que siguen cada regla dan el número de casos que se cubren por la regla seguida por el número que son los mal clasificados [24].

El algoritmo C4.5 (J48) se deriva del algoritmo simple dividir y conquistar para producir árboles de decisión. Este necesita ser extendido de varias maneras antes de estar listo para su uso en los problemas del mundo real. Primeramente se consideró cómo repartir con atributos numéricos y, después de esto, los valores perdidos. Luego, se observó un problema importantísimo de poda de árboles de decisión, porque aunque los árboles construidos por el algoritmo simple de dividir y conquistar realicen bien el entrenamiento, normalmente son la sobrecarga de los datos de entrenamiento y no generaliza bien las pruebas independientes.

Se consideró cómo convertir los árboles de decisión a las reglas de clasificación y explorar las opciones que provee el C4.5. Finalmente se encontró una estrategia de poda alternativa que se lleva a cabo con el famoso sistema CART para la clasificación de aprendizaje y los árboles de regresión [24].

## CAPÍTULO 3. RESULTADOS Y DISCUSIÓN

En este capítulo se describen los procesos de selección de rasgos y clasificación de las imágenes. Los datos y tablas plasmados fueron obtenidos a partir del WEKA.

### 3.1 Selección de rasgos

La Tabla 3.1 muestra los métodos de selección de rasgos utilizados. La segunda columna muestra el valor máximo de mérito de cada método, mientras que la tercera columna nos muestra los atributos rechazados, de acuerdo con el criterio de selección utilizado. Los rasgos 1 (asimetría), 3 (aspereza) y 7 (excentricidad) fueron considerados irrelevantes, ya que su valor quedaba por debajo del 50% del máximo valor de mérito.

Tabla 3.1 Resultados de la selección de rasgos

Métodos	Valor máx. de mérito	Rasgos rechazados
ChiSquaredAttributeEval	526,84	7 y 3
GainRatioAttributeEval	0,76	3
InfoGainAttributeEval	1,81	7 y 3
ReliefFAttributeEval	0,33	1, 7 y 3
OneRAttributeEval	79,35	3
SymmUncertSttributeEval	0,75	3

### 3.2 Aplicación de los wrappers

Luego de aplicar los *wrappers*, más rasgos fueron eliminados en cada caso, como se muestra en la Tabla 3.2. Los rasgos 10 (factor de forma), 12 (número de concavidades) y 8 (solidez) resultaron ser los más utilizados en todos los casos. En el caso de los momentos de Zernike, todos fueron rechazados tras haber aplicado los *wrappers*.

Tabla 3.2 Rasgos seleccionados por *wrappers*

Clasificador	Rasgos seleccionados
MLP	2, 4,6, 8,10, 11, 12, 13,14
SVM ( $p=2$ )	5, 6, 8, 11, 12, 13, 14
k-NN ( $k=1$ )	2, 4, 6, 10, 12
PART	2, 8, 9, 10, 12
C4.5	4, 8, 9, 10, 12

En los siguientes gráficos se muestra, por el eje de las abscisas, el valor del rasgo correspondiente y por el eje de las ordenadas se muestra el total clasificado dentro de un rango determinado.

El factor de forma (Figura 4) tiene un alto poder de discriminación para la mayoría de las clases, diferenciando completamente los linfocitos (color cian) del resto. Este rasgo también puede separar muy bien los monocitos (color gris) y basófilos (color azul) de eosinófilos (color rojo) y neutrófilos (color rosado).

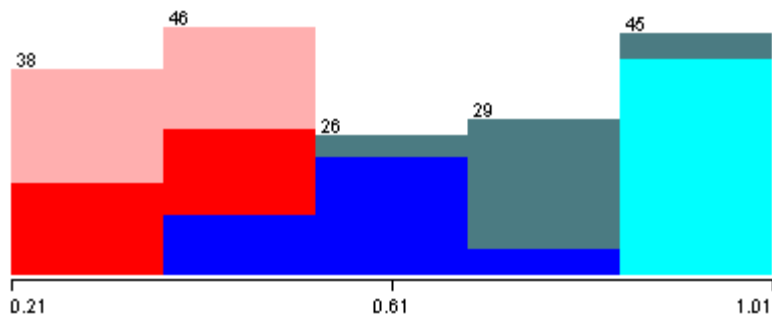


Figura 4. Poder de discriminación del rasgo factor de forma

El número de concavidades también resulta un rasgo valioso, separando monocitos de basófilos (Figura 5).

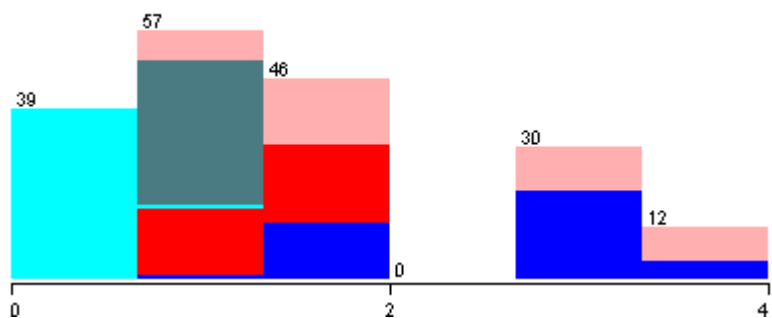


Figura 5. Poder de discriminación del rasgo número de concavidades.

La solidez (Figura 6) y extensión (Figura 7) ayudan a separar los eosinófilos del resto.

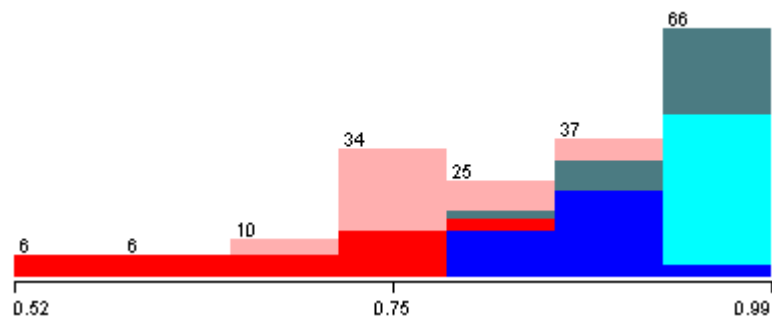


Figura 6. Poder de discriminación del rasgo solidez.

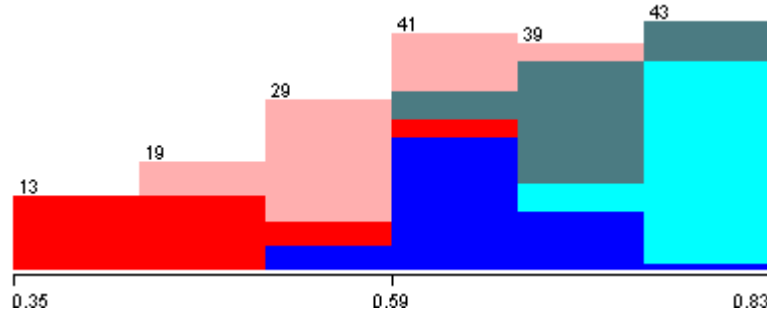


Figura 7. Poder de discriminación del rasgo extensión.

### 3.3 Clasificación

La Tabla 3.3 muestra los resultados de la clasificación para todos los clasificadores. En todos los casos, se usaron los atributos seleccionados en el proceso de selección de rasgos luego de emplear los *wrappers*. En cada caso se usaron los rasgos arrojados por el *wrapper* correspondiente. En todos los casos se usó crovalidación *5-fold* para obtener los porcentos de clasificación correcta.

Tabla 3.3 Resultados de la clasificación

Clasificador	% de clasificación correcta
MLP	97,3%
SVM ( $p=2$ )	96,7%
k-NN ( $k=1$ )	98,4%
PART	96,2%
C4.5	96,2%

Todos los clasificadores arrojan excelentes resultados, aunque MLP y SVM necesitan más rasgos para obtener un alto porcentaje de clasificación (9 y 7 respectivamente). PART y C4.5 son algoritmos más simples, no obstante arrojan resultados comparables con el resto. El algoritmo k-NN tiene el mejor desempeño de clasificación, al mismo tiempo que trata bien con el *overfitting* y permite el aprendizaje incremental.

## CONCLUSIONES Y RECOMENDACIONES

### Conclusiones

- Se obtuvo un extensivo conjunto de rasgos basados en contorno y región para un total de 42.
- La estrategia de selección de rasgos empleada permitió reducir el conjunto de rasgos inicial a menos de 10.
- El algoritmo de clasificación k-NN arroja los mejores resultados al mismo tiempo que trata mejor el problema del *overfitting* y permite el aprendizaje incremental.

### Recomendaciones

- Emplear un algoritmo de clasificación que emplee otros tipos de rasgos en dependencia de la calidad de los mismos.
- Aplicar los métodos obtenidos pero poniendo a punto antes un algoritmo de segmentación.

## REFERENCIAS BIBLIOGRÁFICAS

- [1] Sabino. D. M. U, Costa. L. F, Rizzatti. E. G, and Zago. M. A, "Toward Leukocyte Recognition using Morphometry, Texture and Color.," ed: IEEE International Symposium on Biomedical Imaging, 2004.
- [2] Ramoser. H, Laurian. V, Bischof. H, and Eckeret. R, "Leukocyte segmentation and classification in blood-smear images.," ed: 27th Annual International Conference of the Engineering in Medicine and Biology Society., 2006.
- [3] Ongun. G, Halici. U, Leblebicioglu. K, Atalay. V, Beksac. M, and Beksac. S, "Feature extraction and classification of blood cells for an automated differential blood count system.," ed: International Joint Conference on Neural Networks, 2001.
- [4] Bikheth. S. F, Darwish. A. M, Tolba. H. A, and Shaheen. S. I, "Segmentation and classification of white blood cells.," ed: IEEE International Conference on Acoustics, Speech, and Signal, 2000.
- [5] Gomez.-Gil. P, Ramirez.-Cortez. M, Gonzalez.-Bernal. J, Garcia.-Pedrero. A, Prieto.-Castro. C. I, Valencia. D, *et al.*, "A Feature Extraction Method Based on Morphological Operators for Automatic Classification of Leukocytes.," ed: Seventh Mexican International Conference on Artificial Intelligence, 2008.
- [6] Osowski. S, Sroic. R, Markiewicz. T, and Siwek. K, *Application of Support Vector Machine and Genetic Algorithm for Improved Blood Cell Recognition.*: IEEE Transactions on Instrumentation and Measurement, 2009.
- [7] Montalvo. Arenas. C. E, "Microscopia," ed, 2010.
- [8] Arraiza. N, Viguria. P. M, Navarro. J, and Ainciburu. A, "Manual de Microscopia. Historia, descripción y uso del microscopio óptico ", ed, 2012.
- [10] B. L. Kuyatt, R. Weaver, and P. Merlo, "Image Capture Methods," ed, 2005.
- [11] Maria del Rocio Ochoa Montiel, "Clasificación de Leucocitos utilizando Visión por Computadora," ed: Universidad Autónoma de Tlaxcala, Facultad de Ciencias Básicas, Ingeniería y Tecnología, 2006.
- [12] Bauer. DJ, *Métodos y diagnósticos del laboratorio clínico*, 8va ed., 2009.
- [13] Green. AE, Middleton. VL, Prentis. KG, and Sygny. AG, "A report on experience with an automated blood counting machine," ed, pp. 19-27, 2009.
- [14] Nelson. A. D and Morris. M. W, *Examen básico de la sangre*, 9na ed., 1993.

- [15] Lewis. S. M and Bentley. S. A, "Automated differential counting: the present state of the art," pp. 481-485, 2012.
- [16] Wintrobe. M. M, *Recuento celular diferencial: Hematología clínica*, 2008.
- [17] L. H. Staib and J. S. Duncan, " Left ventricular analysis from cardiac images using deformable models, IEEE Comput," pp. 427-430, 2012.
- [18] E. A. Geiser, D. A. Conetta, M. C. Limacher, V. O. Stockton, L. H. Olivier, and B. Jones, *A second-generation computer-based edge detection algorithm for short-axis two-dimensional echocardiographic images: Accuracy and improvement in interobserver variability* vol. 3, 2006.
- [19] C. deGraaf, A. Koster, K. Vincker, and M. Vierger, " A methodology for the validation of images segmetation algorithms," ed: Proc. IEEE Symp. Computer-base Medical Sytems, pp. 17-24 , 2011.
- [20] C. M., Kanade, T. P. D., and Rowely, "Anomaly Detection through registration, Pattern Recognition," ed, pp. 113-128, 2013.
- [21] S. S. and P. H., "Invariant feature extraction based on radial and distance function for automatic target recognition," ed: International Conference on Image Processing, 2010.
- [22] V. Chalana and Y. Kim, " A methodology for evaluation of boundary detection algorithms on medical images," ed: IEEE Transactions on Medical Imaging, pp. 642-652, 2013.
- [23] Theera-Umpon. N, "Morphological Granulometric Features of Nucleus in Automatic Bone Marrow White Blood Cell Classification," ed: IEEE Transactions on Information Technology in Biomedicine, 2007.
- [24] Witten. I. H., Frank. E, and H. M.A, *Data Mining. Practical Machine Learning Tools and Techniques*, 3ra ed., 2011.
- [25] Y. Minqiang, Kidiyo, K., Joseph, R, "A Survey of Shape Feature Extraction Techniques," ed, 2008.
- [26] Guyon. I. and Elisseeff. A., "An Introduction to Variable and Feature Selection. J. Machine Learning Research," ed, 2009.



## ANEXOS

Los códigos de los programas en Matlab empleados en la realización de este trabajo se muestran a continuación:

### Anexo I Programación en Matlab

#### **Programa principal**

```
srcDir = 'Nuc_seg';
dstDir = 'Nuc_proc';
filenamesId = '_n';
arffFilename = 'MyData.arff';
javaPath = fullfile('.', 'Weka', 'Weka.jar');
classNames = {'Lymphocyte', 'Monocyte', 'Basophil', 'Neutrophil', 'Eosinophil'};
classFilenameIds = {'l', 'm', 'b', 'n', 'e'};
imgSize = 100;

% ProcessMasks(srcDir, filenamesId, dstDir, imgSize);
[feat, featNames] = GetFeatures(dstDir, filenamesId, classNames, classFilenameIds);
javaaddpath(javaPath);
train = matlab2weka('MyData', featNames, feat, size(feat, 2));
saveARFF(arffFilename, train);
```

#### **Obtención de rasgos.**

```
function [feat, featNames] = GetFeatures(srcDir, filenamesId, classNames,
classFilenameIds)
imgFilenames = dir(fullfile(srcDir, ['*', filenamesId, '.*']));
l = length(imgFilenames);
if l == 0
    error(['There are not images in "', srcDir, '" directory.']);
end
disp('Getting features...');
```

```

disp('Processing image #1');
I = imread(fullfile(srcDir,imgFileNames(1).name));
[fAS, fASn] = AreaShape(I);
[fGM, fGMn] = GranulometricsMoments(I);
[fZM, fZMn] = ZernikeMoments(I);
[fC, fCn] = ContourFeat(I);
featNames = [fCn, fASn, fGMn,fZMn {'Class'}];
FAS = zeros(1, length(fASn)); FAS(1, :) = fAS;
FGM = zeros(1, length(fGMn)); FGM(1, :) = fGM;
FZM = zeros(1, length(fZMn)); FZM(1, :) = fZM;
FC = zeros(1, length(fCn)); FC(1, :) = fC;
Class = repmat({'', 1, 1}); Class{1} = getClassName(imgFileNames(1).name(1),
classNames, classFilenameIds);
for i = 2:l,
    disp(['Processing image #', num2str(i)]);
    I = imread(fullfile(srcDir,imgFileNames(i).name));
    FAS(i, :) = AreaShape(I);
    FGM(i, :) = GranulometricsMoments(I);
    FZM(i, :) = ZernikeMoments(I);
    FC(i, :) = ContourFeat(I);
    Class{i} = getClassName(imgFileNames(i).name(1), classNames, classFilenameIds);
end
FAS = num2cell(FAS); FGM = num2cell(FGM); FZM = num2cell(FZM);
FC = num2cell(FC);
feat = [FC, FAS, FGM, FZM, Class];

function className = getClassName(classId, classNames, classIds)
    className = '';
    for i = 1:length(classIds),
        if strcmp(classIds{i}, classId)
            className = classNames{i};
        end
    end
end

```

```

        break;
    end
end

Area Shape

function [feat, featNames] = AreaShape(I)
%AREASHAPE Get area and shape features
% Get area and shape features from image I.
%

rp = regionprops(I, {'Area', 'Eccentricity', 'Solidity', 'Extent', 'Perimeter', 'ConvexImage'});
FormFactor = 4*pi*rp.Area/rp.Perimeter^2;
rpci = regionprops(rp.ConvexImage, {'Perimeter'});
Convexity = rpci.Perimeter/rp.Perimeter;

[x, y] = find(I);
Ibb = I(min(x):max(x), min(y):max(y));
As = regionprops(rp.ConvexImage & ~Ibb, 'Area'); nc = 0;
for i = 1: length(As)
    if As(i).Area > 50
        nc = nc + 1;
    end
end

featNames = {'Eccentricity', 'Solidity', 'Extent', 'FormFactor', 'Convexity', 'NoCon'};
feat = [rp.Eccentricity, rp.Solidity, rp.Extent, FormFactor, Convexity, nc];

```

### **Momentos granulométricos.**

```

function [feat, featNames] = GranulometricsMoments(I)
%GRANULOMETRICSMOMENTS Get granulometrics moments features
% Get granulometrics moments features from image I.
%
% Alexander Falcón Ruiz, CEETI, UCLV, January 2010

P = pecstrum(I, 9);
u1 = granulometricmoment(P, 1);
u2 = granulometricmoment(P, 2);
%P(P < 0.05) = 0;
% featNames = {'P1', 'P2', 'P3', 'P4', 'P5', 'P6', 'P7', 'P8', 'P9', 'GM1', 'GM2'};

```

```
% feat = [P, u1, u2];
```

```
featNames = {'GM1', 'GM2'};
```

```
feat = [u1, u2];
```

```
function u = granulometricmoment(P, k)
```

```
l = length(P);
```

```
u = 0;
```

```
for n = 1:l,
```

```
    u = u+n^k*P(n);
```

```
end
```

```
u = u/sum(P);
```

### **Zernike moments**

```
function [Znm, ZnmNames, n, m] = ZernikeM(I, order)
```

```
%ZERNIKE Get Zernike moments from a binary image
```

```
% Get Zernike moments of a given orden from binary image I
```

```
%
```

```
[x, y] = meshgrid(-1:2/(size(I, 1)-1):1,-1:2/(size(I, 2)-1):1);
```

```
insideCircleR1 = find(x.^2+y.^2<=1);
```

```
Ioi = I(insideCircleR1);
```

```
Z = x+1i*y;
```

```
Zoi = Z(insideCircleR1);
```

```
p = abs(Zoi);
```

```
o = angle(Zoi);
```

```
k = 1;
```

```
ZnmNames = { };
```

```
for i = 0:order,
```

```
    for j = 0:i,
```

```
        if rem(i-j, 2) == 0
```

```
            n(k) = i; m(k) = j;
```

```
            ZnmNames = [ZnmNames, ['Zernike', num2str(i), num2str(j)]];
```

```
            Vnm = ZerComPoly(i, j, p, o);
```

```
            Znm(k) = (i+1)/pi*sum(Ioi.*conj(Vnm));
```

```
            k = k + 1;
```

```
        end
```

```
    end
```

```
end
```

```
function Vnm = ZerComPoly(n, m, p, o)
```

```
a = (n+abs(m))/2;
```

```
b = (n-abs(m))/2;
```

```

Rnm = zeros(size(p));
for s = 0:b,
    num = (-1)^s*fact(n-s)*(p.^(n-2*s));
    den = fact(s)*fact(a-s)*fact(b-s);
    Rnm = Rnm + num/den;
end
Vnm = Rnm.*exp(1i*m*o);

```

```

function f = fact(n)
factTable = [1, 2, 6, 24, 120, 720, 5040, 40320, 362880, 3628800];
f = factTable(n+1);

```

```

function [feat, featNames] = ZernikeMoments(I)
%ZERNIKEMOMENTS Get Zernike moments features
% Get Zernike moments features from image I.
%
% Alexander Falc3n Ruiz, CEETI, UCLV, January 2010

```

```

order = 9;
[feat, featNames] = ZernikeM(I, order);
feat = feat(3:end);
featNames = featNames(3:end);
feat = abs(feat);

```

### **Contour feat**

```

function [feat, featNames] = ContourFeat(I)
%CONTOURFEAT Get contour features
% Get contour features from image I.
%
featNames = {'Skew', 'RatioMinMax', 'Roughness', 'Roundness', 'CircVar', 'EllVar'};
s = regionprops(I, 'centroid');
xc = round(s.Centroid(2));
yc = round(s.Centroid(1));
BW = edge(I,'sobel');
[x, y] = find(BW);
z = sqrt((x-xc).^2+(y-yc).^2);
m = mean(z);
N = length(z);

```

```

skew = sum((z-m).^3)/sum(z.^3);

```

```

ratioMinMax = min(z) / max(z);

```

```

roughness = length(find(z<m))/N;

```

```

A = length(find(I));

```

---

```

[xi, yi] = find(I);
Ae = length(find(sqrt((xi-xc).^2+(yi-yc).^2)>m));
roundness = Ae/A;

CircVar = sqrt(1/N*sum((z-m).^2))/m;

cxx = 1/N*sum((x-xc).^2);
cxy = 1/N*sum((x-xc).*(y-yc));
cyx = cxy;
cyy = 1/N*sum((y-yc).^2);
C = [cxx, cxy; cyx, cyy];
d = zeros(1, N);
for i = 1:N,
    Vi = [x(i)-xc; y(i)-yc];
    d(i) = sqrt(Vi'*inv(C)*Vi);
end
ur = mean(d);
EllVar = sqrt(1/N*sum((d-ur).^2))/ur;

% Principal axis ratio
%a = cxx+cyy; b = sqrt(a^2-4*(cxx*cyy-cxy^2));
%PrAxRatio = (a-b)/(a+b);

feat = [skew, ratioMinMax, roughness, roundness, CircVar, EllVar];

```