



CARTA DE AUTORIZACIÓN

CÓDIGO

AP-BIB-FO-06

VERSIÓN

1

VIGENCIA

2014

PÁGINA

1 de 1

Neiva, 22 de junio de 2018

Señores

CENTRO DE INFORMACIÓN Y DOCUMENTACIÓN

UNIVERSIDAD SURCOLOMBIANA

Ciudad

El suscrito:

Oscar Armando Quiroga Garcés, con C.C. No.1077871475 autor de la tesis y/o trabajo de grado titulado Sistema de clasificación automática de leucocitos usando características morfológicas del núcleo mediante visión por computador presentado y aprobado en el año 2018 como requisito para optar al título de Ingeniero Electrónico;

Autorizo al CENTRO DE INFORMACIÓN Y DOCUMENTACIÓN de la Universidad Surcolombiana para que, con fines académicos, muestre al país y el exterior la producción intelectual de la Universidad Surcolombiana, a través de la visibilidad de su contenido de la siguiente manera:

- Los usuarios puedan consultar el contenido de este trabajo de grado en los sitios web que administra la Universidad, en bases de datos, repositorio digital, catálogos y en otros sitios web, redes y sistemas de información nacionales e internacionales "open access" y en las redes de información con las cuales tenga convenio la Institución.
- Permita la consulta, la reproducción y préstamo a los usuarios interesados en el contenido de este trabajo, para todos los usos que tengan finalidad académica, ya sea en formato Cd-Rom o digital desde internet, intranet, etc., y en general para cualquier formato conocido o por conocer, dentro de los términos establecidos en la Ley 23 de 1982, Ley 44 de 1993, Decisión Andina 351 de 1993, Decreto 460 de 1995 y demás normas generales sobre la materia.
- Continúo conservando los correspondientes derechos sin modificación o restricción alguna; puesto que de acuerdo con la legislación colombiana aplicable, el presente es un acuerdo jurídico que en ningún caso conlleva la enajenación del derecho de autor y sus conexos.

De conformidad con lo establecido en el artículo 30 de la Ley 23 de 1982 y el artículo 11 de la Decisión Andina 351 de 1993, "Los derechos morales sobre el trabajo son propiedad de los autores", los cuales son irrenunciables, imprescriptibles, inembargables e inalienables.

EL AUTOR/ESTUDIANTE: Oscar Armando Quiroga Garcés

Firma:

Oscar Quiroga Garcés

Vigilada Mineducación



**TÍTULO COMPLETO DEL TRABAJO:** Sistema de clasificación automática de leucocitos usando características morfológicas del núcleo mediante visión por computador

**AUTOR O AUTORES:**

Primero y Segundo Apellido	Primero y Segundo Nombre
Quiroga Garcés	Oscar Armando

**DIRECTOR Y CODIRECTOR TESIS:**

Primero y Segundo Apellido	Primero y Segundo Nombre
Salgado Patrón	José de Jesús

**ASESOR (ES):**

Primero y Segundo Apellido	Primero y Segundo Nombre
----------------------------	--------------------------

**PARA OPTAR AL TÍTULO DE:** Ingeniero Electrónico

**FACULTAD:** Ingeniería

**PROGRAMA O POSGRADO:** Ingeniería Electrónica

**CIUDAD:** Neiva

**AÑO DE PRESENTACIÓN:** 2018

**NÚMERO DE PÁGINAS:** 49

**TIPO DE ILUSTRACIONES (Marcar con una X):**

Diagramas  Fotografías  Grabaciones en discos \_\_\_ Ilustraciones en general \_\_\_ Grabados \_\_\_ Láminas \_\_\_  
Litografías \_\_\_ Mapas \_\_\_ Música impresa \_\_\_ Planos \_\_\_ Retratos \_\_\_ Sin ilustraciones \_\_\_ Tablas o Cuadros

Vigilada mieducación

La versión vigente y controlada de este documento, solo podrá ser consultada a través del sitio web Institucional [www.usco.edu.co](http://www.usco.edu.co), link Sistema Gestión de Calidad. La copia o impresión diferente a la publicada, será considerada como documento no controlado y su uso indebido no es de responsabilidad de la Universidad Surcolombiana.



<b>CÓDIGO</b>	<b>AP-BIB-FO-07</b>	<b>VERSIÓN</b>	<b>1</b>	<b>VIGENCIA</b>	<b>2014</b>	<b>PÁGINA</b>	<b>2 de 3</b>
---------------	---------------------	----------------	----------	-----------------	-------------	---------------	---------------

**SOFTWARE** requerido y/o especializado para la lectura del documento: PDF Reader

**MATERIAL ANEXO:** Código y guía de usuario

**PREMIO O DISTINCIÓN** (*En caso de ser LAUREADAS o Meritoria*):

**PALABRAS CLAVES EN ESPAÑOL E INGLÉS:**

<u>Español</u>	<u>Inglés</u>	<u>Español</u>	<u>Inglés</u>
1. Leucocitos	Leukocytes	6. Mononuclear	Mononuclear
2. Linfocitos	Lymphocytes	7. Polimorfonuclear	Polymorphonuclear
3. Monocitos	Monocytes	8. K-means	K-means
4. Neutrófilos	Neutrophils	9. Análisis Discriminante	Discriminant Analysis
5. Eosinófilos	Eosinophils	10. Python	Python

**RESUMEN DEL CONTENIDO:** (Máximo 250 palabras)

Los glóbulos blancos también llamados leucocitos son las células del sistema inmunitario que están involucradas en la protección del cuerpo tanto contra enfermedades infecciosas como contra los invasores extraños. El análisis de los leucocitos proporciona información valiosa a los especialistas médicos, ayudándolos a diagnosticar diferentes enfermedades hemáticas importantes, como el SIDA y el cáncer de sangre (leucemia). En la actualidad, los patólogos anotan manualmente los leucocitos, lo que lleva mucho tiempo y es susceptible de error, debido a la naturaleza tediosa del proceso. Este proyecto tiene como objetivo la automatización del proceso del Conteo Diferencial de Sangre, a fin de aumentar la productividad y eliminar los errores humanos. El sistema propuesto toma las imágenes microscópicas de frotis de sangre como entrada para la detección y clasificación de leucocitos. En el marco propuesto, primero se segmenta el núcleo, seguido de la extracción de las características de textura y geométricas. Finalmente, los leucocitos detectados se clasifican en sus respectivas clases. Se usó la crovalidación en los experimentos en los cuales Support Vector Machine (SVM), Análisis discriminante lineal (LDA) y Análisis discriminante cuadrática (QDA) fueron aplicados como clasificadores. Los resultados indican un 92.3% de exactitud del sistema para clasificar las células.



**ABSTRACT:** (Máximo 250 palabras)

White blood cells, also called leukocytes, are cells of the immune system that are involved in protecting the body against both infectious diseases and foreign invaders. The leukocyte analysis provides valuable information to medical specialists, helping them to diagnose different important blood diseases, such as AIDS and blood cancer (leukemia). Currently, pathologists perform the leukocyte count manually, which is time-consuming and error-prone due to the tedious nature of the process. Automating the Differential Blood Count process offers a low-cost solution compared to high-tech medical equipment. Due to the multiple nature of these cells and the uncertainty in hematological imaging, leukocyte segmentation is one of the most important stages in this process. Scrupulous segmentation obviously reduces the errors of the following stages. In this article, we present the K-means grouping algorithm in the HSI color space to segment the cores. In addition, the performances of three classifiers, Support Vector Machine (SVM), Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA) for the recognition of leukocyte types are compared. The results show a 92.3% accuracy of the system for classifying cells.

**APROBACION DE LA TESIS**

Nombre Presidente Jurado:

Firma:

Nombre Jurado: Martín Diomedes Bravo Obando

Firma:

Nombre Jurado: Vladimir Mosquera Cerquera

Firma:

**SISTEMA DE CLASIFICACIÓN AUTOMÁTICA DE LEUCOCITOS USANDO  
CARACTERÍSTICAS MORFOLÓGICAS DEL NÚCLEO MEDIANTE VISIÓN  
POR COMPUTADOR**

**OSCAR ARMANDO QUIROGA GARCÉS**

**UNIVERSIDAD SURCOLOMBIANA  
FACULTAD DE INGENIERÍA  
PROGRAMA DE INGENIERÍA ELECTRÓNICA  
NEIVA  
2018**

**SISTEMA DE CLASIFICACIÓN AUTOMÁTICA DE LEUCOCITOS USANDO  
CARACTERÍSTICAS MORFOLÓGICAS DEL NÚCLEO MEDIANTE VISIÓN  
POR COMPUTADOR**

**OSCAR ARMANDO QUIROGA GARCÉS**

**Trabajo de grado para optar al título de Ingeniero Electrónico**

**Director**

**José de Jesús Salgado Patrón**

**Magister en Ingeniería Electrónica y de Computadores**

**UNIVERSIDAD SURCOLOMBIANA**

**FACULTAD DE INGENIERÍA**

**PROGRAMA DE INGENIERÍA ELECTRÓNICA**

**NEIVA**

**2018**

Nota de aceptación:

---

---

---

---

---

---

---

Firma del presidente del Jurado

---

Firma del Jurado

---

Firma del Jurado

Neiva, Marzo 01 de 2018

Dedico este proyecto de grado a mis padres, Luz Elena Garcés Parra y Miller Armando Quiroga Calderón. A mis hermanos por su apoyo incondicional y confianza.

## **AGRADECIMIENTOS**

En primera instancia agradezco a mi director de tesis, Eng. M. Sc. José Salgado, bajo cuya supervisión comencé la tesis. A Cristian Barrera por su colaboración y asesoría durante el proyecto. A Sandra Rojas, bacterióloga de la E.S.E. María Auxiliadora de Garzón por su tiempo en la toma de muestras y validación del proyecto.

## CONTENIDO

	Pág.
INTRODUCCIÓN	12
1. PLANTEAMIENTO DEL PROBLEMA	13
2. JUSTIFICACIÓN	14
3. OBJETIVOS	15
3.1 OBJETIVO GENERAL	15
3.2 OBJETIVOS ESPECÍFICOS	15
4. LEUCOCITOS	16
5. SISTEMA DE CLASIFICACIÓN AUTOMÁTICA DE LEUCOCITOS	17
5.1 ADQUISICIÓN DE DATOS Y PRE-PROCESAMIENTO	17
5.2 SEGMENTACIÓN	20
5.2.1 Selección del número óptimo de Clústers	20
5.2.2 Algoritmo de agrupamiento de K-means	21
5.2.3 Filtrado espacial	22
5.3 EXTRACCIÓN DE CARACTERÍSTICAS	22
5.3.1 Circularidad	23
5.3.2 Número de Lóbulos	24
5.4 CLASIFICADORES	24
5.4.1 Support Vector Machine (SVM)	25
5.4.2 Linear Discriminant Analysis (LDA) y Quadratic Discriminant Analysis (QDA)	27
5.4.3 Selección del clasificador y Cross Validation	29
6. IMPLEMENTACIÓN	31
6.1 DIAGRAMAS DE CASOS DE USO	31
6.2 INTERFAZ GRÁFICA	31
6.3 REQUISITOS	33
7. PRUEBAS Y RESULTADOS	34
7.1 SEGMENTACIÓN	34
7.2 CLASIFICACIÓN	34
8. CONCLUSIONES	37
9. TRABAJOS FUTUROS	39
BIBLIOGRAFÍA	40
ANEXOS	42

## LISTA DE TABLAS

	Pág.
Tabla 1. Conjunto de datos de imágenes	18
Tabla 2. Algoritmo de agrupamiento K-means	21
Tabla 3. Características extraídas	23
Tabla 4. 10 Fold Cross Validation expresado en porcentaje	30
Tabla 5. Requisitos del sistema	33
Tabla 6. Leucocitos detectados versus datos reales clasificados por un especialista	35
Tabla 7. Clasificación de los resultados obtenidos para cada tipo de leucocito	35
Tabla 8. Valores calculados de sensibilidad, especificidad y exactitud	36

## LISTA DE FIGURAS

	Pág.
Figura 1. Tipos de leucocitos	16
Figura 2. Sistema base del proyecto	17
Figura 3. Microscopio Primo Star	18
Figura 4. Histograma de la imagen original	19
Figura 5. Histograma Ecualizado	19
Figura 6. Imagen RGB	19
Figura 7. Imagen cromática rg	19
Figura 8. Modelo HSI	20
Figura 9. Elbow Curve	21
Figura 10. Aplicando el algoritmo de agrupación K-means	22
Figura 11. Calculo de Lóbulos	24
Figura 12. Ejemplo de funciones Kernel para SVM	25
Figura 13. Plot de parámetro Lineal para el tipo de núcleo	26
Figura 14. Plot de parámetro Lineal para las células mononuclear	26
Figura 15. Plot de parámetro Lineal para las células polimorfonuclear	27
Figura 16. Ejemplo LDA vs QDA	28
Figura 17. Plot LDA vs QDA para el tipo de núcleo	28
Figura 18. Plot LDA vs QDA para las células mononuclear	29
Figura 19. Plot LDA vs QDA para las células polimorfonuclear	29
Figura 20. Cross Validation	30
Figura 21. Diagrama de casos de uso	31
Figura 22. Interfaz gráfica del sistema	32
Figura 23. Ejemplo de visualización de leucocitos	32
Figura 24. Ejemplo de visualización de detección y clasificación de leucocitos	33
Figura 25. Resultado de la segmentación por K-means	34
Figura 26. Interfaz inicial presentada al especialista	44
Figura 27. Cuadro de diálogo para abrir una imagen	45
Figura 28. Imagen cargada en el espacio de trabajo	45
Figura 29. Imagen en la cual se ha detectado los leucocitos	46
Figura 30. Imagen en la cual se ha clasificado los leucocitos	47
Figura 31. Reporte generado en formato PDF	47
Figura 32. Ayuda	48
Figura 33. Figura 8. Acerca de ...	48

## LISTA DE ANEXOS

Anexo A: Códigos de Python	Pág. 42
Anexo B: Guía de usuario	44

## GLOSARIO

**Leucocito:** componente importante de la sangre y una pieza clave en el sistema inmunológico del cuerpo.

**Conteo Diferencial de Sangre:** valorar y reconocer la cantidad de los distintos tipos de leucocitos que se observa en una placa.

**Python:** Lenguaje de programación.

## **RESUMEN**

Los glóbulos blancos también llamados leucocitos son las células del sistema inmunitario que están involucradas en la protección del cuerpo tanto contra enfermedades infecciosas como contra los invasores extraños. El análisis de los leucocitos proporciona información valiosa a los especialistas médicos, ayudándolos a diagnosticar diferentes enfermedades hemáticas importantes, como el SIDA y el cáncer de sangre (leucemia). En la actualidad, los patólogos anotan manualmente los leucocitos, lo que lleva mucho tiempo y es susceptible de error, debido a la naturaleza tediosa del proceso. Este proyecto tiene como objetivo la automatización del proceso del Conteo Diferencial de Sangre, a fin de aumentar la productividad y eliminar los errores humanos. El sistema propuesto toma las imágenes microscópicas de frotis de sangre como entrada para la detección y clasificación de leucocitos. En el marco propuesto, primero se segmenta el núcleo, seguido de la extracción de las características de textura y geométricas. Finalmente, los leucocitos detectados se clasifican en sus respectivas clases. Se usó la crosvalidación en los experimentos en los cuales Support Vector Machine (SVM), Análisis discriminante lineal (LDA) y Análisis discriminante cuadrático (QDA) fueron aplicados como clasificadores. Los resultados indican un 92.3% de exactitud del sistema para clasificar las células.

### **PALABRAS CLAVE:**

Leucocito, Conteo Diferencial de Sangre, Máquinas de Soporte Vectorial, Análisis discriminante lineal y Análisis discriminante cuadrático.

## INTRODUCCIÓN

Las imágenes microscópicas del frotis de sangre son investigadas rutinariamente por hematólogos para diagnosticar la mayoría de las enfermedades de la sangre. En la sangre circulan tres tipos básicos de células: eritrocitos (glóbulos rojos), leucocitos (glóbulos blancos) y plaquetas. Los leucocitos son, en realidad, un grupo de diferentes células, con diferentes funciones en el sistema inmunológico. Algunos leucocitos atacan directamente al invasor, otros producen anticuerpos, otros apenas hacen la identificación y así sucesivamente. Bajo el microscopio, el valor normal de los leucocitos varía entre 4000 y 11000 células por ml. Existen cinco tipos, neutrófilos (40-60%), linfocitos (20-40%), monocitos (2-8%), eosinófilos (1-4%) y basófilos (0.5-1%), cada uno con sus particularidades<sup>1</sup>.

La observación de frotis de sangre entera bajo un microscopio proporciona importante información cualitativa y cuantitativa sobre el diagnóstico de diversas enfermedades, incluida la leucemia. La identificación y clasificación de leucocitos en estos tipos es con frecuencia realizada manualmente por técnicos o patólogos experimentados e involucra dos elementos principales. El primero es el estudio cualitativo de la morfología de los leucocitos que proporciona información sobre la adecuación del frotis entre otros parámetros, incluidas las características morfológicas de las células. El segundo es cuantitativo y consiste en un recuento diferencial de al menos 100 glóbulos blancos. La precisión de la clasificación y el conteo de las células se ve notablemente afectada por las habilidades y la experiencia del operador individual. Además, la identificación y el recuento diferencial de las células sanguíneas es una tarea repetitiva que consume mucho tiempo. Normalmente, el experto toma una gota de sangre de la muestra y la extiende sobre un portaobjetos de vidrio. La muestra se tiñe con un tinte especial, el cual ayuda a establecer la diferencia entre diversos tipos de leucocitos. Finalmente se realiza el conteo en el microscopio<sup>2</sup>. Sustituir tal proceso por una detección y clasificación automática de las diferentes clases de células es un desafío importante en el dominio de los laboratorios de diagnóstico clínico.

---

<sup>1</sup> NARANJO A. Carmen Beatriz. "Atlas de hematología células sanguíneas". En: (2008), pp. 60-64.

<sup>2</sup> NARANJO A. Carmen Beatriz. "Atlas de hematología células sanguíneas". En: (2008), pp. 20-22.

## **1. PLANTEAMIENTO DEL PROBLEMA**

La presencia de diferentes enfermedades hematológicas, han llevado al ser humano a investigar métodos más sencillos, eficaces y confiables para la detección y posterior tratamiento de ellas. Excluyendo la técnica manual, otras técnicas representan una mayor inversión de dinero para los laboratorios clínicos, ya que es necesario el uso de equipos costosos y la compra de reactivos químicos necesarios para el funcionamiento de estos. La técnica manual, en cambio, no representa mayor inversión de dinero, pero si requiere una gran dedicación por parte del profesional, ya que en esta técnica se deben contar manualmente las células sanguíneas observadas en el microscopio a partir de un frotis de sangre periférica, lo que conlleva errores en el resultado del hemograma, debido al esfuerzo que esto requiere.

Este problema se agrava principalmente en centros de salud del departamento del Huila, donde la atención diaria se da a cientos de personas que muchas veces tienen que esperar un tiempo considerable para la obtención de sus resultados médicos.

El proyecto en cuestión aparece, entonces, como una alternativa confiable, eficiente y económica, pues tiene como objetivo otorgar una herramienta destinada a la detección y clasificación de cinco tipos de leucocitos (basófilo, eosinófilo, neutrófilo, linfocito y monocito) empleando los rasgos morfológicos del núcleo a partir de imágenes de frotis de sangre, mediante procesamiento digital de imágenes, siendo una herramienta que servirá de apoyo para el diagnóstico de una infección, anemia o leucemia.

## 2. JUSTIFICACIÓN

El sistema de clasificación se enfoca en el análisis de imágenes microscópicas de frotis de sangre periférica humana para identificar características morfológicas de los leucocitos contenidos en la imagen, de manera que se puedan clasificar dentro de los cinco tipos de leucocitos más comunes. Dado que dichos tipos de células están relacionadas con la presencia de diversas enfermedades, se crea una herramienta de clasificación que ofrece al médico elementos de juicio y criterios objetivos para reducir la naturaleza subjetiva del diagnóstico.

De acuerdo con lo mencionado antes, la importancia de este trabajo radica precisamente en la combinación de herramientas de procesamiento de imágenes y técnicas de inteligencia computacional, para determinar métricas discriminantes y usarlas en la clasificación de algunos tipos de células que pueden ser indicadores de diferentes enfermedades.

### **3. OBJETIVOS**

#### **3.1 OBJETIVO GENERAL**

Desarrollar un software que permita la detección y clasificación de leucocitos empleando solo información del núcleo.

#### **3.2 OBJETIVOS ESPECÍFICOS**

- Adquirir las imágenes digitales del frotis de sangre periférica humana.
- Implementar el algoritmo más adecuado, para la detección de los leucocitos comprendidos en las imágenes digitales.
- Extraer las características morfológicas del núcleo de los leucocitos contenidos en las imágenes en cuanto a su forma y textura.
- Implementar el método de clasificación óptimo empleando los rasgos seleccionados.
- Comparar los resultados alcanzados con resultados obtenidos por un médico especialista.
- Desarrollar el software que permita la detección y clasificación de leucocitos.

## 4. LEUCOCITOS

Figura 1. Tipos de leucocitos



De izquierda a derecha: monocito, linfocito, basófilo, eosinófilo, y neutrófilo. Las imágenes han sido etiquetadas a mano por la bacterióloga Sandra Rojas.

Existen varios tipos y subtipos diferentes de leucocitos, que, en función de su aspecto nuclear, se pueden dividir en dos categorías: células **mononucleares** y células **polimorfonucleares**. En la Figura 1, se puede observar los cinco tipos de leucocitos de interés.

Los leucocitos mononucleares comprenden tanto **linfocitos** como **monocitos**, ambas células tienen núcleos redondos y grandes de color violeta-azul, que ocupa casi toda la célula<sup>3</sup>.

Por otro lado, los polimorfonucleares, tienen citoplasma granuloso y núcleos alargados o lobulados, con 2 a 5 lóbulos conectados por delgados puentes. Este tipo de células incluyen **neutrófilos**, **eosinófilos** y **basófilos**<sup>4</sup>.

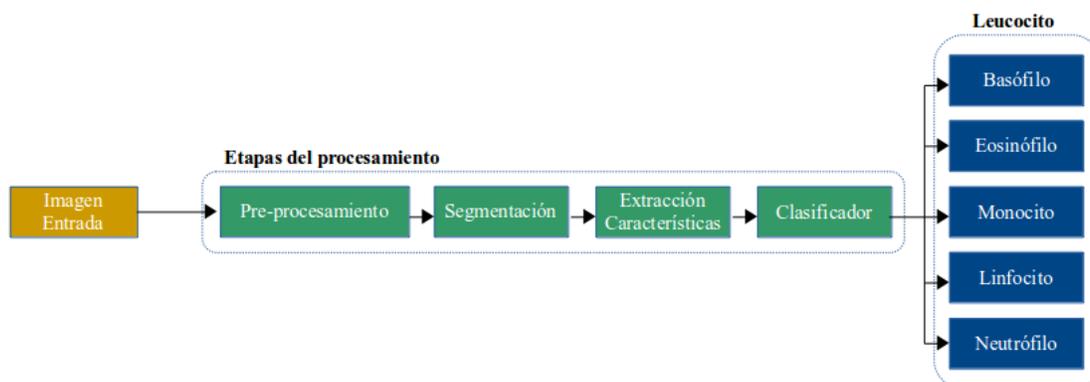
---

<sup>3</sup> NARANJO A. Carmen Beatriz. "Atlas de hematología células sanguíneas". *En*: (2008), pp. 76-79.

<sup>4</sup> Rafael Amadeo Mateo Capilla. "Análisis clínico de rutina". *En*: El recuento leucocitario.

## 5. SISTEMA DE CLASIFICACIÓN AUTOMÁTICA DE LEUCOCITOS

Figura 2. Sistema base del proyecto



El sistema base en el cual se desarrolla el trabajo, visto en la Figura 2, consta de cuatro pasos: 1) adquisición y pre-procesamiento de imágenes de frotis de sangre microscópicas, 2) segmentación de núcleos de los leucocitos, 3) extracción de características de los núcleos segmentados, y 4) clasificación de leucocitos en sus respectivas categorías.

El lenguaje de programación elegido es **Python**, el cual cuenta con bibliotecas y **librerías** de acceso libre, que tienen bajo consumo de recursos. También cuenta con una amplia lista de funciones para la construcción de aplicaciones de procesamiento de imágenes. Las librerías utilizadas durante el proyecto principalmente son **Scipy**, **Scikit-Image** y **Scikit-Learn**.

### 5.1 ADQUISICIÓN DE DATOS Y PRE-PROCESAMIENTO

Las imágenes se adquirieron a partir de diferentes frotis de sangre periférica humana teñidas previamente con tinte Wright, dichas muestras fueron tomadas del centro de salud E.S.E. María Auxiliadora de Garzón, Huila.

La captura se realiza en el Laboratorio Clínico de dicho centro de salud, se usa un Microscopio Primo Star, visto en la Figura 3, adaptado a una cámara fotográfica de un smartphone Samsung J5. El aumento de captura fue de **100x** para permitir mejor observación y separación entre las células. Se tomaron **147** muestras de leucocitos y se guardaron en formato **JPG** de dimensión **2560 × 1440 píxeles**. El detalle del conjunto de datos se proporciona en la Tabla 1. Las imágenes digitales fueron clasificadas previamente por la bacterióloga en las 5 clases de interés. Los basófilos representan el menor porcentaje adquirido en el laboratorio, en efecto se eliminan ya que son muy pocos para ser generalizados.

Figura 3. Microscopio Primo Star



Microscopio Binocular con iluminación LED. Marca Zeiss, modelo Primo Star Pack 1.  
Fuente: Carl Zeiss MicroImaging Gamba. “Microscopía de Carl Zeiss”.

Tabla 1. Conjunto de datos de imágenes

Tipo	Basófilo	Eosinófilo	Monocito	Linfocito	Neutrófilo	Total
No de Imágenes	1	14	15	51	66	147

### Pre-procesamiento

En esta sección se enfoca en mejorar la calidad de la imagen adquirida. Esta etapa es necesaria debido a que en las imágenes microscópicas es frecuente encontrar defectos como iluminación no uniforme y variación en la intensidad causadas por las condiciones de tinción, como lo indica Silva E.<sup>5</sup>. En primera instancia el tamaño de las imágenes se ajusta a **720 × 720 píxeles** durante el procesamiento en **Python**. Las matrices de las imágenes se dividen entre **255** para llevar el rango de valores entre **0 y 1**, Figura 4, y se aplica la ecualización de histograma para redistribuir la intensidad de la imagen para cubrir todo el rango de intensidad<sup>6</sup>, visto en la Figura 5.

<sup>5</sup> SILVA, Edwin. “Detección de cáncer de cuello uterino usando técnicas de procesamiento y análisis de imágenes microscópicas”. En: Tesis de Maestría (Ingeniería Electrónica). Universidad Industrial de Santander. Facultad de Ingenierías Fisicomecánicas. Escuela de Ingeniería Eléctrica, Electrónica y Telecomunicaciones. p. 172 (2006).

<sup>6</sup> Rafael C. Gonzalez and Richard E. Woods. “Digital Image Processing”. pp. 91-93.

Figura 4. Histograma de la imagen original

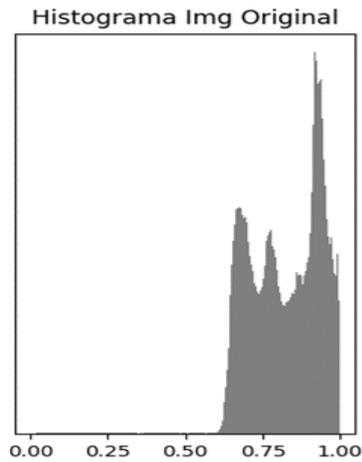
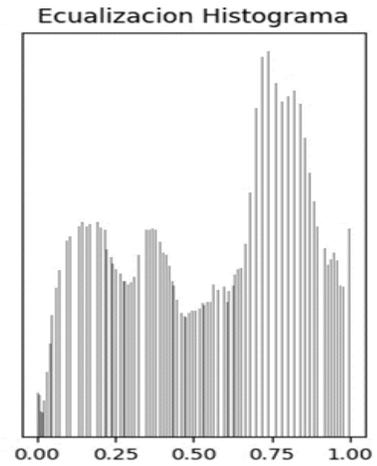


Figura 5. Histograma Ecuilizado



Para eliminar la influencia de la intensidad de la iluminación, la imagen se transforma del espacio **RGB** al espacio **cromático rg**<sup>7</sup>, visto en la Figura 6 y Figura 7. La conversión del espacio de color del espacio RGB al espacio cromático rg, se define como,

$$r = \frac{R}{R+G+B} \quad (1)$$

$$g = \frac{G}{R+G+B} \quad (2)$$

$$b = 1 - r - g \quad (3)$$

Donde R, G, B son los tres componentes de canal de la imagen RGB original; r, g, b son los tres componentes de canal en el espacio cromático rg.

Figura 6. Imagen RGB

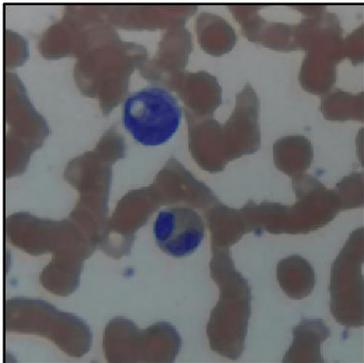
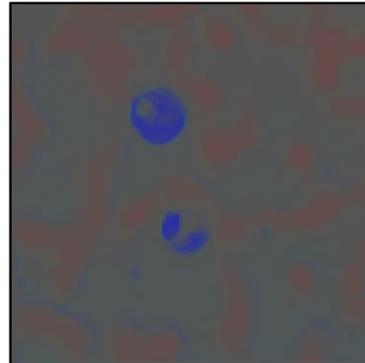


Figura 7. Imagen cromática rg

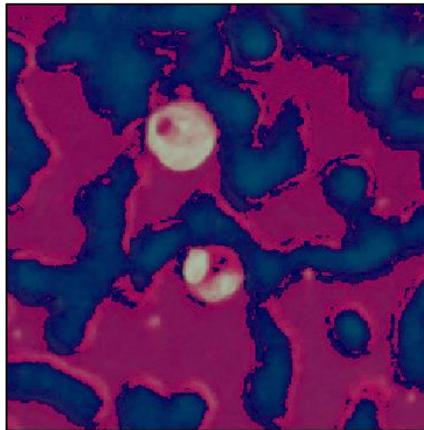


<sup>7</sup> Zhi Liu, Jing Liu, Xiaoyan Xiao, Hui Yuan, Xiaomei Li, Jun Chang and Chengyun Zheng." Segmentation of White Blood Cells through Nucleus Mark Watershed Operations and Mean Shift Clustering". En: (2015).

## 5.2 SEGMENTACIÓN

Primero la imagen cromática rg se convierte en imagen **HSI**, como se observa en la Figura 8. Para segmentar la imagen se implementa la técnica de segmentación no supervisada, es decir, el **algoritmo de agrupamiento k-means**, el cual se implementa con las librerías **Numpy** y **Scikit-Learn**<sup>8</sup> en **Python**, tomado como referencia del trabajo original<sup>9</sup>, el cual especifica que, *es el mejor candidato para las tareas subyacentes, ya que es simple y comparativamente más preciso para la segmentación basada en el color*<sup>10</sup>.

Figura 8. Modelo HSI.



### 5.2.1 Selección del número óptimo de Clústers

El **algoritmo de agrupamiento K-means** es un proceso iterativo de mover los centros de los **clústers**, o los **centroides**, a la posición media de sus puntos constituyentes, y reasignar instancias a sus clústers más cercanos<sup>11</sup>. El **K** especifica la cantidad de clústers que se deben crear; K-means asigna automáticamente observaciones a los clústers, pero no puede determinar el número apropiado de clústers<sup>12</sup>.

El **método del codo (Elbow Method)**<sup>13</sup>, se implementa para hallar el número apropiado de clústers. Este método utiliza los valores de la inercia obtenidos tras aplicar el K-means a

---

<sup>8</sup> Gavin Hackling. "Mastering Machine Learning with Scikit-learn". En: (2014).

<sup>9</sup> Omid Sarrafzadeh and Alireza Mehri Dehnavi. "Nucleus and cytoplasm segmentation in microscopic images using K-means clustering and region growing". En: Advanced Biomedical Research (2015).

<sup>10</sup> Muhammad Sajjad, Siraj Khan, Zahoor Jan, Khan Muhammad, Hyeonjoon Moon, Jin Tae Kwak, Seungmin Rho, Sung Wook Baik and Irfan Mehmood. "Leukocytes Classification and Segmentation in Microscopic Blood Smear: A Resource-Aware Healthcare Service in Smart Cities". En: (2016).

<sup>11</sup> Gavin Hackling. "Mastering Machine Learning with Scikit-learn". En:(2014), pp. 115-116.

<sup>12</sup> Moya Ricardo. "Machine Learning (en Python)". En: (2016).

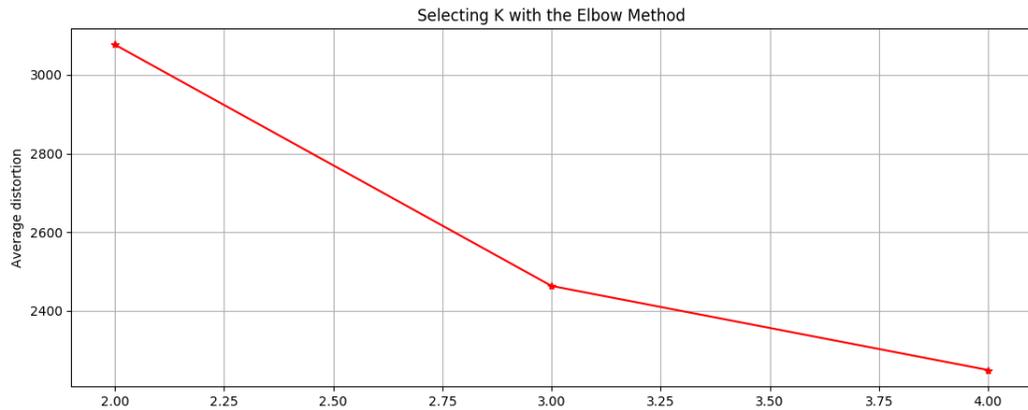
<sup>13</sup> Purnima Bhalowalia and Arvind Kumar. "EBK-Means: A Clustering Technique based on Elbow Method and K-Means in WSN". En: International Journal of Computer Application (2014).

diferentes clústers, desde **1 a N** clústers, siendo la inercia la suma de las distancias al cuadrado de cada objeto del clúster a su centroide<sup>14</sup>.

$$Inercia = \sum_{i=0}^N \|x_i - \mu\|^2 \quad (4)$$

La Figura 9 muestra un cambio brusco en la evolución de la inercia, las líneas representan una forma similar a la de un brazo y su codo, y el punto que representa al codo del brazo es el número óptimo de Clusters, el cual es igual a **3**.

Figura 9. Elbow Curve



Se observa un cambio en K igual a 3, por lo que, este es el número apropiado de clústers para llevar a cabo el algoritmo de agrupamiento de K-means.

### 5.2.2 Algoritmo de agrupamiento de K-means

El algoritmo de agrupamiento K-means, es tomado de **Scikit Learn**<sup>15</sup>. Específicamente, consiste en seleccionar un número **K** de centroides de forma aleatoria, donde **K** es el número grupos (clústers) deseado. Cada clúster generado está asociado a un centroide (punto central) y los puntos se asignan al clúster más cercano al centroide, Tabla 2. La Figura 10 ilustra el resultado después de aplicar el algoritmo.

Tabla 2. Algoritmo de agrupamiento K-means

---

#### Algoritmo K-means

---

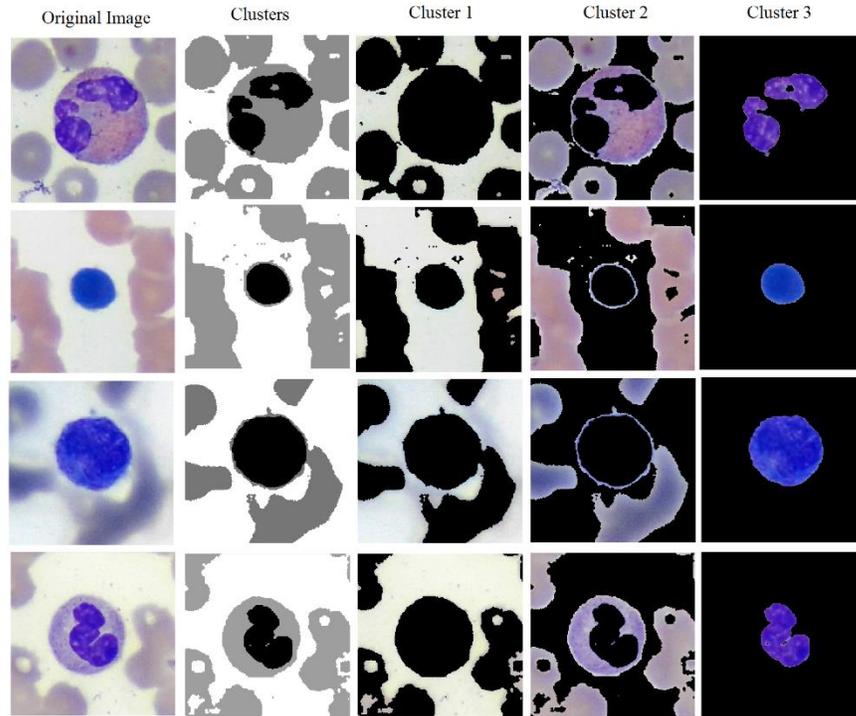
1. Seleccionar K puntos como centroide central
  2. Repetir
  3. Formar K *clústers* asignando cada punto a su centroide más cercano
  4. Calcular de nuevo el centroide de cada *clúster*
  5. No hay cambio en los centroides
- 

Fuente: Amparo Olaya Benavides, Álvaro Enrique Ortiz Dávila and Wilson Soto. "Learning Styles Selection for Geomatic Teaching using Intelligent Algorithms". En: (2012).

<sup>14</sup> Mashiat Fatma and Jaya Sharma. "Leukemia Image Segmentation using K-Means Clustering and HSI Color Image Segmentation". En: (2014).

<sup>15</sup> Scikit-learn user guide. (Consultado 11/10/2017).

Figura 10. Aplicando el algoritmo de agrupación K-means



### 5.2.3 Filtrado espacial

Después de aplicar el algoritmo de agrupamiento K-means, la imagen segmentada presenta algunas regiones o ruidos no deseados. Para esta etapa, como lo indican Dhanachandra, N., Manglem k. y Chanu Y.<sup>16</sup>, se utilizan **filtros de mediana**, para obtener una imagen libre de ruido.

Para fortalecer el débil enlace entre objetos, aumentar la definición de formas y eliminar los pequeños objetos oscuros no deseados, se utilizan los filtros morfológicos **closing** y **opening**<sup>17</sup>, en efecto, se obtuvo una imagen con los objetos de interés más definidos (núcleos).

## 5.3 EXTRACCIÓN DE CARACTERÍSTICAS

La extracción de características se realiza tomando como entrada la imagen obtenida en la sección anterior. Aquí, se describen los núcleos para su reconocimiento, que es un paso fundamental hacia la precisión de la clasificación.

<sup>16</sup> Nameirakpam Dhanachandra, Khumanthem Manglem and Yambem Jina Chanu. "Image Segmentation using K-means Clustering Algorithm and Subtractive Clustering Algorithm". *En: Eleventh International Multi-Conference on Information Processin* (2015).

<sup>17</sup> Rafael C. Gonzalez and Richard E. Woods. "Digital Image Processing". pp. 528-531.

El módulo **Measure** de la librería **Scikit-Image**<sup>18</sup>, ofrece las funciones de extracción necesarias para describir los diferentes núcleos de los leucocitos. En la Tabla 3 se observan las respectivas características con su descripción.

Tabla 3. Características extraídas

<b>Característica</b>	<b>Descripción</b>	<b>Función</b>
<b>Circularidad</b>	Relación entre el área de la célula y el cuadrado de su perímetro normalizada con respecto a $\frac{1}{4}\pi$ .	Para diferenciar el tipo el núcleo, mononuclear o polimorfonuclear.
<b>Excentricidad</b>	Desviación de la forma circular ideal del núcleo.	
<b>Área</b>	Área total de píxeles que conforman el núcleo contorneado.	Para diferencia entre células con núcleos mononuclear, linfocito o monocito.
<b>Entropía</b>	Medida de entropía de los píxeles.	
<b>Extensión</b>	Proporción de los píxeles en el <i>bounding box</i> que también están en la región.	Para diferenciar entre células de núcleos polimorfonuclear, neutrófilos o eosinófilos.
<b>Intensidad Mínima</b>	Valor mínimo del canal de luminancia.	
<b>Intensidad Máxima</b>	Valor máximo del canal de luminancia.	
<b>Número de Lóbulos</b>	Número de regiones divididas.	

Fuente: Rafael C. Gonzalez and Richard E. Woods. "Digital Image Processing". Pp. 644-674

### 5.3.1 Circularidad

La circularidad del núcleo se calcula utilizando la Fórmula 5, si la circularidad es **uno**, indica un círculo perfecto y si es **0**, indica un polígono cada vez más alargado. Entonces, basado en la circularidad, los leucocitos se pueden dividir en **mononuclear** y **polimorfonuclear**, ya que los primeros tienen un núcleo casi circular, mientras que los últimos tienen un núcleo de forma irregular<sup>19</sup>.

$$Circularidad = 4 * \pi * (A/P^2) \quad (5)$$

Donde A es el tamaño del núcleo y P es el perímetro del núcleo en la célula sanguínea.

<sup>18</sup> Scikit-learn user guide. (Consultado 08/11/2017).

<sup>19</sup> Chhaya S. Hinge, Aarti G. Ambekar and Samidha S. Kulkarni. "Classification of RBC And WBC in Peripheral Blood Smear Using KNN". En: (2013).

### 5.3.2 Número de Lóbulos

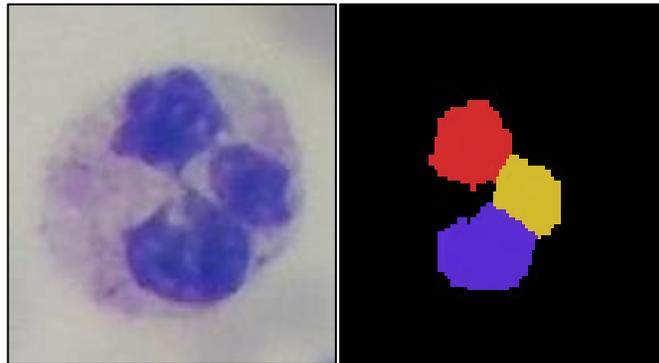
El número de lóbulos en los **linfocitos** y **monocitos** tiene un valor menor; la mayoría de ellos son **lobulados**. Por otro lado, los **eosinófilos** y **neutrófilos** tienen un mayor número de lóbulos. Por lo tanto, la cantidad de lóbulos es una característica distintiva importante.

Para estimar el número de lóbulos en el núcleo de un leucocito, se toma como referencia el método propuesto por Atin Mathur, Ardhendu S. y Manohar Kuse<sup>20</sup>, donde se ha calculado dividiendo el núcleo en regiones de  $N_i$ , donde  $N_i \in (2, 3, 4, 5)$ , utilizando el algoritmo de división de regiones propuesto por Costas et al.<sup>21</sup>. Si la relación entre el área del núcleo y la de su caja delimitadora, es decir, 'Extensión' es mayor que **0,7**, el número de lóbulos es igual a uno, de lo contrario, el número de lóbulos es dada por la Fórmula 6.

$$N_i | C(N_i) = \text{Media}(R_i, e_i, E_i) \quad (6)$$

Donde  $(R_i, e_i, E_i)$ , son la media de factores de circularidad, extensiones y excentricidades de las regiones divididas de  $N_i$ , respectivamente. La Figura 11 ilustra el método de recuento de lóbulos utilizado.

Figura 11. Cálculo de Lóbulos



Aplicando la Fórmula 6 en una célula neutrófila, se observa un total de 3 lóbulos.

## 5.4 CLASIFICADORES

Los tipos de clasificadores a utilizar son el **Support Vector Machine (SVM)**, **Linear Discriminant Analysis (LDA)** y **Quadratic Discriminant Analysis (QDA)**, los cuales son ampliamente utilizados en el campo de la visión por computador<sup>22</sup>.

<sup>20</sup> Atin Mathur, Ardhendu S. Tripathi, and Manohar Kuse. "Scalable system for classification of white blood cells from Leishman stained blood stain images". En: Journal of Pathology Informatics (2013).

<sup>21</sup> Panagiotakis C, Ramasso E, Tziritas G. "Lymphocyte segmentation using the transferable belief model". En: Lecture notes in Computer Science (LNCS). Recognizing Patterns in Signals, Speech, Images and Videos. Proceedings of the ICPR 2010 Contests, Springer. 2010:253–62 (2010).

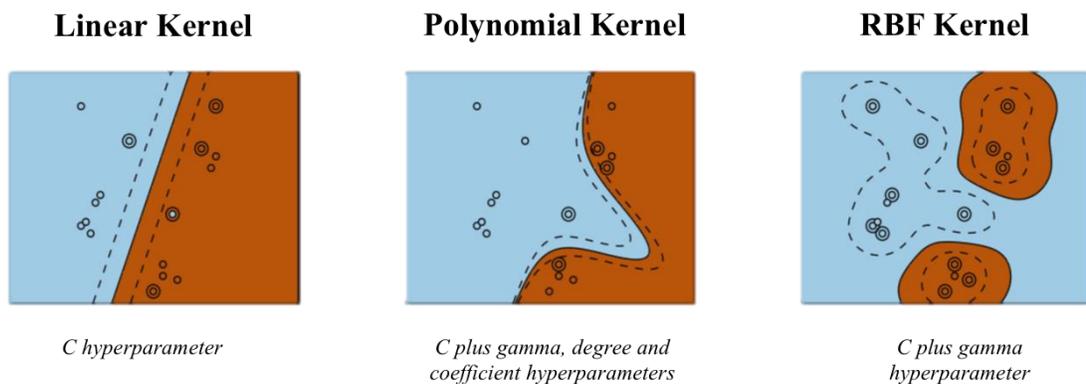
<sup>22</sup> Anisas Bouzalmat, Jamal Kharroubi and Arsalane Zarghili. "Comparative Study of PCA, ICA, LDA using SVM Classifier". En: Journal of Emerging Technologies in Web Intelligence, Vol.6, No 1 (2014).

### 5.4.1 Support Vector Machine (SVM)

Este tipo de clasificador funciona correlacionando datos a un espacio de características de grandes dimensiones de forma que los puntos de datos se puedan categorizar, incluso si los datos no se puedan separar linealmente de otro modo<sup>23</sup>. Se detecta un separador entre las categorías y los datos se transforman de forma que el separador se puede extraer como un **hiperplano**. Tras ello, las características de los nuevos datos se pueden utilizar para predecir el grupo al que pertenece el nuevo registro<sup>24</sup>.

SVM permite utilizar las llamadas *funciones Kernel (no lineales)*<sup>25</sup>. Estas funciones resuelven el problema de clasificación trasladando los datos a un espacio donde el hiperplano solución es **lineal** y, por tanto, más sencillo de obtener, como se observa en la Figura 12. Incluyen parámetros iniciales para su entrenamiento, siendo la constante de margen suave **C** y el parámetro de kernel **gamma** ( $\gamma$ ), las cuales ajustan la importancia de maximizar, o minimizar la cantidad de muestras a tomar en cuenta y que tanta influencia puede tener una muestra sobre el clasificador respectivamente<sup>26</sup>.

Figura 12. Ejemplo de funciones Kernel para SVM



Ejemplo de los diferentes tipos de Kernel utilizados para el clasificador SVM.  
Fuente: “Scikit-learn user guide”. Consultado (10/01/2018).

### Selección de los parámetros SVM

Una buena clasificación depende de la correcta elección de parámetros. El tipo de **kernel** a utilizar es el **Lineal**, depende de **C**, el cual se somete a variaciones para encontrar un balance aceptable de separación de las clases y sus respectivas muestras.

La Figura 13 ilustra las muestras azules que son leucocitos mononucleares y las rojas los polimorfonucleares. Las sombras rojas y azules representan la totalidad de la zona que el

<sup>23</sup> “Scikit-learn user guide”. (Consultado 10/01/2018).

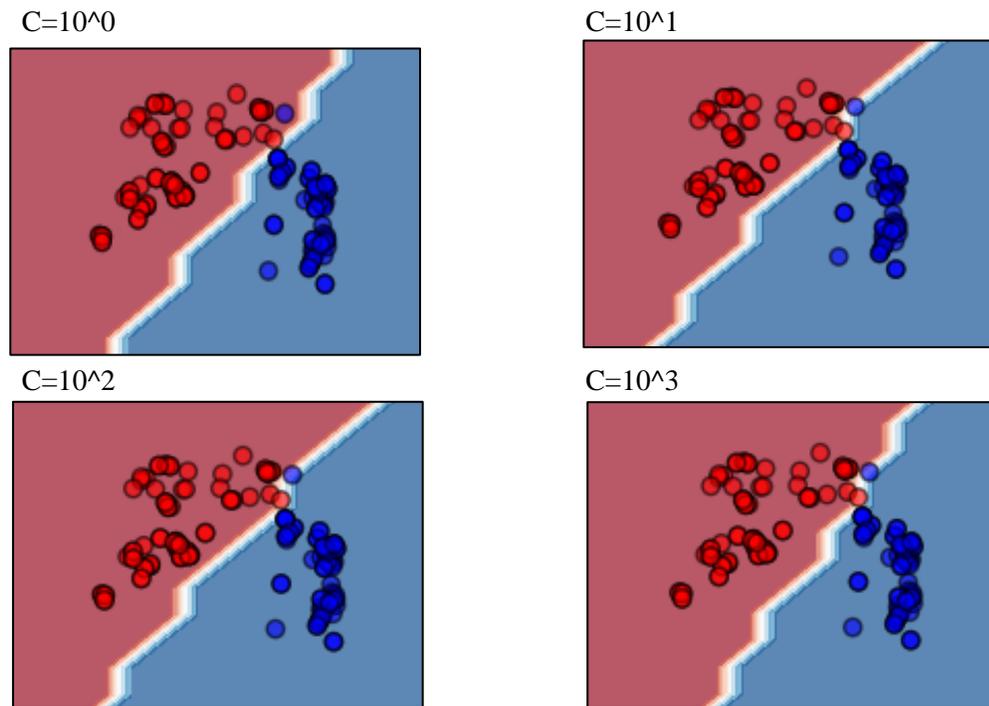
<sup>24</sup> Enrique J. Carmona Suárez. “Tutorial sobre Máquinas de Vectores Soporte (SVM)”. *En*: Dpto. de Inteligencia Artificial, ETS de Ingeniería Informática, Universidad Nacional de Educación a Distancia (UNED), C/Juan del Rosal, 16, 28040-Madrid (Spain) (2014).

<sup>25</sup> Geoff Gordon. “Support Vector Machines and Kernel Methods”. (Consultado 10/01/2018).

<sup>26</sup> Ben-hur, Asa; Weston, “A User’s Guide to Support Vector Machines”.

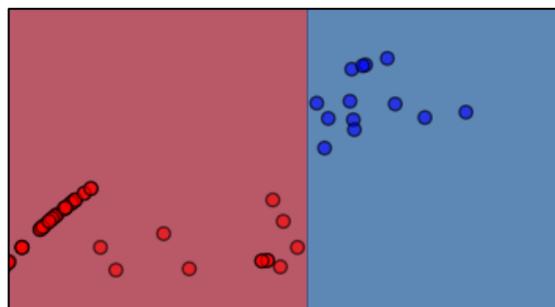
valor de  $C$  otorga para que el clasificador encierre las muestras que estén dentro de dicha margen. La zona blanca, es la margen divisoria que separa las muestras y se toman como no clasificadas. Para los valores de  $C$  igual **100**, se mantiene acorde una separación aceptable de las muestras, dejando pocas muestras sin clasificar.

Figura 13. Plot de parámetro Lineal para el tipo de núcleo



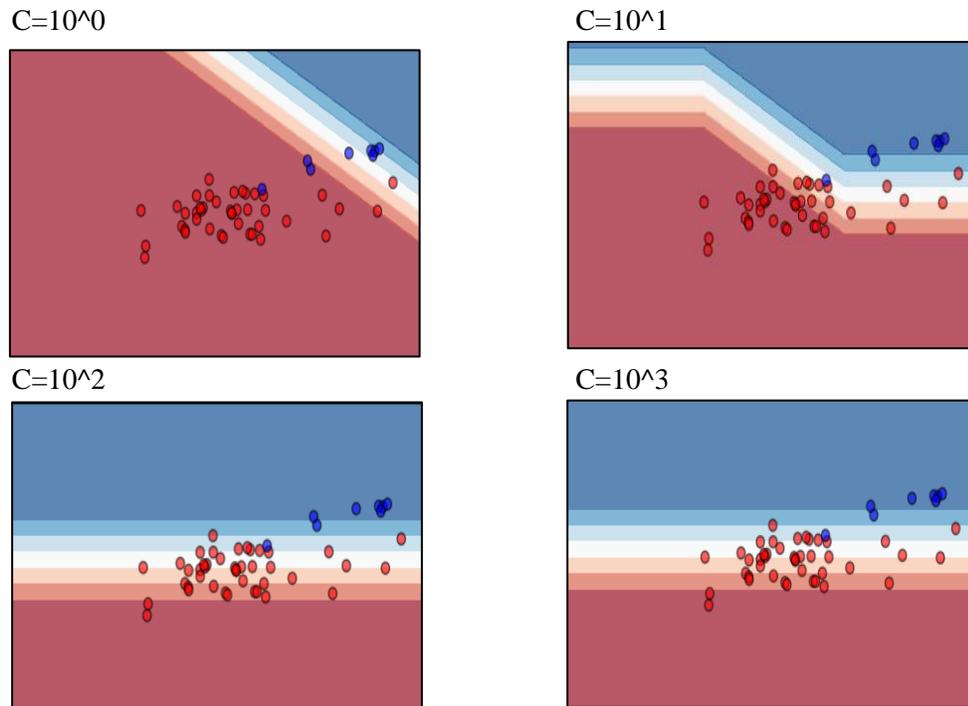
Con respecto a las células mononuclear, Figura 14,  $C$  puede tomar distintos valores para alcanzar una separación aceptable y clasificación óptima de las muestras azules (monocitos) frente a las rojas (linfocitos), se selecciona el valor de **10**.

Figura 14. Plot de parámetro Lineal para las células mononuclear



Para el caso de las células polimorfonucleares, se puede apreciar las muestras de eosinófilos (azules) y neutrófilos (rojas), para valores de  $C$  igual a  $10$ , presenta una mejor separación de las muestras, no obstante, ambas clases están cerca de la zona central, dejando en la frontera gran cantidad de muestras a no clasificar, otorgando un nivel mayor de dificultad al clasificar, Figura 15.

Figura 15. Plot de parámetro Lineal para las células polimorfonuclear



#### 5.4.2 Linear Discriminant Analysis (LDA) y Quadratic Discriminant Analysis (QDA)

Estos tipos de clasificadores, como su nombre lo indica, una superficie de decisión **lineal** y **cuadrática**, respectivamente, son atractivos porque tienen soluciones cerradas que se pueden calcular fácilmente, han demostrado funcionar bien en la práctica<sup>27</sup>, y no tienen parámetros para sintonizar<sup>28</sup>.

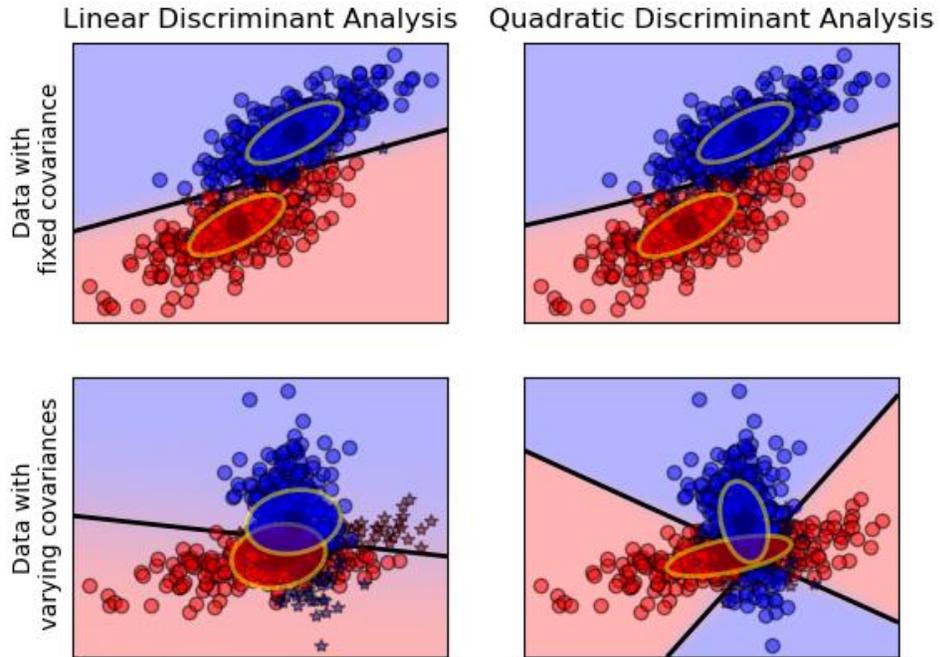
Estos dos métodos suponen que cada clase es de distribución gaussiana multivariante y usan propiedades estadísticas de los datos, como la varianza y la media para establecer el clasificador<sup>29</sup>. En la Figura 16 se evidencia un ejemplo sobre estos, donde la fila inferior demuestra que **LDA** solo puede aprender los límites lineales, mientras que **QDA** puede aprender los límites cuadráticos y, por lo tanto, es más flexible.

<sup>27</sup> Scikit-Learn guide. “Linear and Quadratic Discriminant Analysis”. (Consultado 12/01/2017).

<sup>28</sup> Sofus A. Macskassy. “Machine Learning”.

<sup>29</sup> Miguel Cárdenas-Montes. “Linear Discriminant Analysis”.

Figura 16. Ejemplo LDA vs QDA



Fuente: Scikit-Learn, Linear and Quadratic Discriminant Analysis.

La Figura 17 muestra el Plot **LDA** versus **QDA** para el tipo de núcleo de la célula, las muestras mononucleares (azules) y las polimorfonucleares (rojas) mantienen una mejor separación entre sí para ambos clasificadores, donde las sombras respectivas ayudan a encerrar estas fácilmente. El clasificador **QDA** deja mayor cantidad de muestras en la zona fronteriza respecto a **LDA**. Para las células mononuclear, Figura 18, ambos clasificadores separan de buena manera las muestras, sin dejar muestras en la zona fronteriza, por lo cual se pueden clasificar de forma oportuna.

Figura 17. Plot LDA vs QDA para el tipo de núcleo

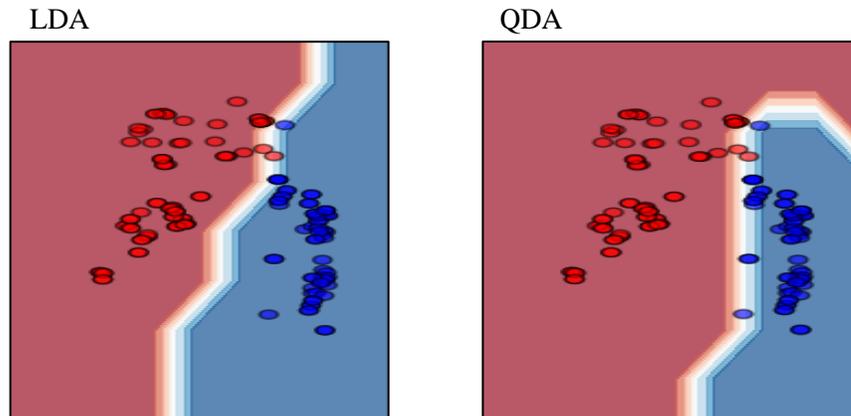
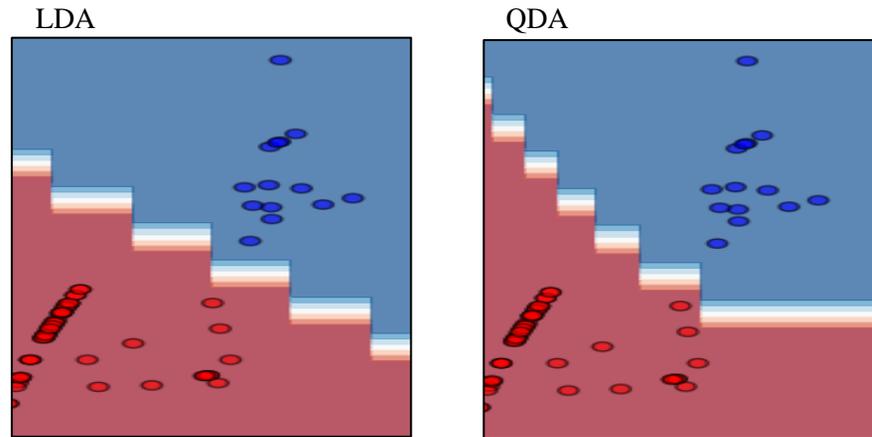
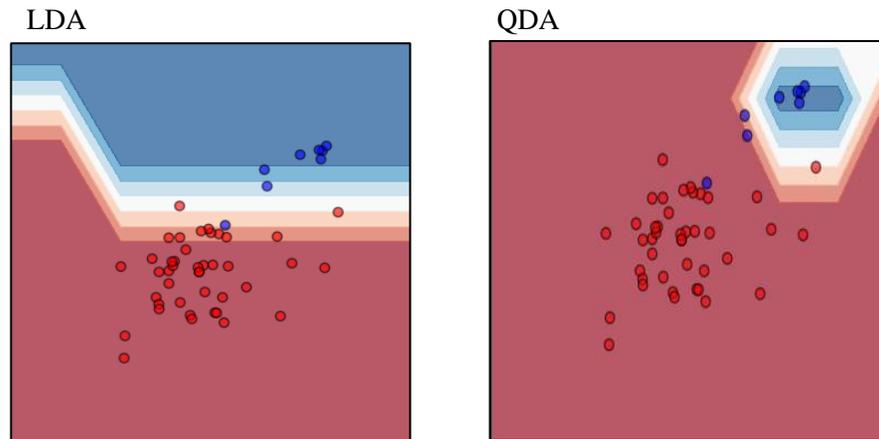


Figura 18. Plot LDA vs QDA para las células mononuclear



Respecto a las células polimorfonuclear, Figura 19, **LDA** separa mejor las muestras respecto al otro clasificador, no obstante, varias muestras quedan ubicadas en la frontera y se toman como no clasificadas.

Figura 19. Plot LDA vs QDA para las células polimorfonuclear

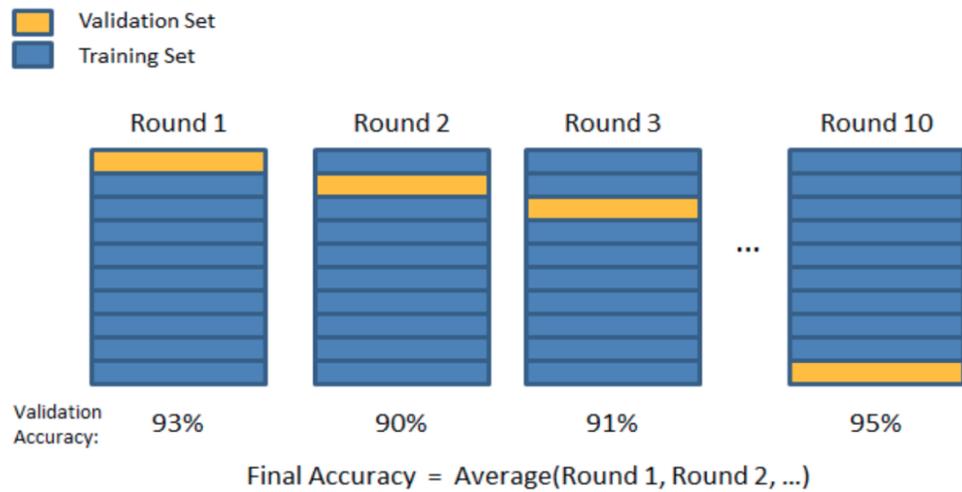


### 5.4.3 Selección del clasificador y Cross Validation

Cuando los datos de entrenamiento son limitados, se usa **cross validation**, que comienza dividiendo el conjunto de datos en **k** subconjuntos, y entrenamos en **k-1** uno de esos subconjuntos, manteniendo el último subconjunto para la prueba. Esto se hace para cada uno de los subconjuntos y luego se promedian los resultados<sup>30</sup>. Una representación visual de este selector de clasificador se puede observar en la Figura 20.

<sup>30</sup> “Scikit-Multilearn”. Consultado (10/02/18).

Figura 20. Cross Validation



Fuente: Joseph Nelson. "Visual Representation of Train/Test Split and Cross Validation".

La Tabla 4 ilustra los porcentajes de exactitud de los clasificadores, tomando en cuenta el hecho de que, a cada prueba de validación, las muestras se toman aleatorias para el **data** y **test**, lo único constante son los parámetros escogidos para el clasificador **SVM**. La función **KFold** de la librería **Scikit-learn** se ha utilizado para validar los datos.

Tabla 4. 10 Fold Cross Validation expresado en porcentaje

	<b>SVM</b> (%)	<b>LDA</b> (%)	<b>QDA</b> (%)
<b>Tipo de célula</b>	97.8	98.62	98.62
<b>Mononuclear</b>	98.0	98.7	98.7
<b>Polimorfonuclear</b>	84.0	88.0	86.0

Los clasificadores **LDA** y **QDA** no presentan mucha diferencia entre sí, únicamente en la clasificación para células polimorfonuclear con **88%** y **86%** respectivamente, pero presentan mayor porcentaje de exactitud frente a **SVM**. Se optó por **LDA** para desempeñar la prueba final.

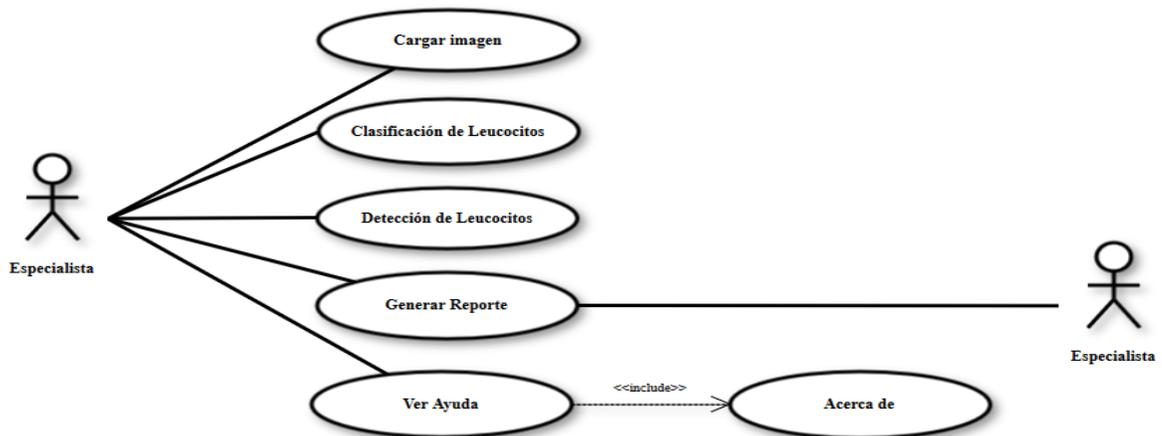
## 6. IMPLEMENTACIÓN

El desarrollo aplicativo del sistema de clasificación automática de leucocitos se realizó en el lenguaje de programación de **Python** junto a las bibliotecas **Scikit-learn** y **Tkinter**<sup>31</sup>, esta última orientada a diseñar la interfaz gráfica para aplicaciones de escritorio. Disponible para plataforma Windows, dado que es el sistema operativo de uso frecuente en los centros de salud.

### 6.1 DIAGRAMAS DE CASOS DE USO

En la Figura 21 se presenta la descripción de los diagramas de casos de uso que se crearon para orientar el desarrollo y la planificación de las características del ambiente computacional.

Figura 21. Diagrama de casos de uso



### 6.2 INTERFAZ GRÁFICA

La interfaz gráfica consta de una ventana principal que se mantiene durante la sesión del usuario, como se muestra en la Figura 22, donde se realiza la detección celular, clasificación y visualización de resultados, Figura 23 y la Figura 24.

Con la opción Cargar Imagen se accede al directorio de la imagen que se desea analizar. La herramienta soporta los formatos **JPG**, **PNG** o **BMP** ya que estos son los formatos que normalmente arrojan las cámaras de adquisición.

<sup>31</sup> Alvarez Alejandro. "Guia Tkinter Documentation". En: (2016).

Figura 22. Interfaz gráfica del sistema

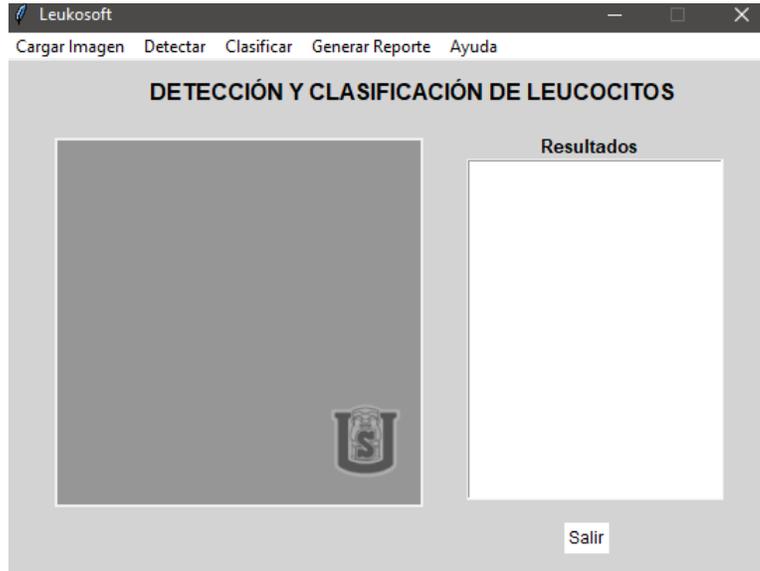
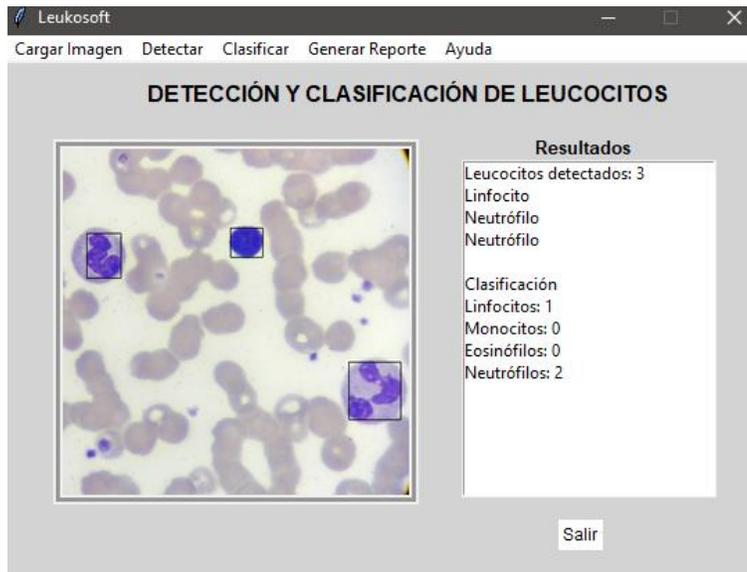


Figura 23. Ejemplo de visualización de leucocitos



Una vez seleccionada la imagen, se realiza la visualización y las operaciones de detección y clasificación. Se determina un cuadrado de borde de color negro para identificar los núcleos de los leucocitos, como se muestra en la Figura 23.

Figura 24. Ejemplo de visualización de detección y clasificación de leucocitos



### 6.3 REQUISITOS

La Tabla 5 muestra los requisitos de **memoria** y **procesador** para la instalación y uso de la herramienta computacional.

Tabla 5. Requisitos del sistema

<b>Componente</b>	<b>Requisito</b>
Sistema Operativo	Windows 10, Windows 8.1, Windows 8, Windows 7, Windows Server 2008 R2 o Windows Server 2012; Linux.
Velocidad del procesador	Mínimo: 1 GHz (32 y 64 bits) Recomendado: 2 GHz o más rápido (32 y 64 bits)
Memoria RAM	Mínimo: 1 GB (32 y 64 bits) Recomendado: 2GB (32 y 64 bits)
Espacio libre requerido	1GB o superior

## 7. PRUEBAS Y RESULTADOS

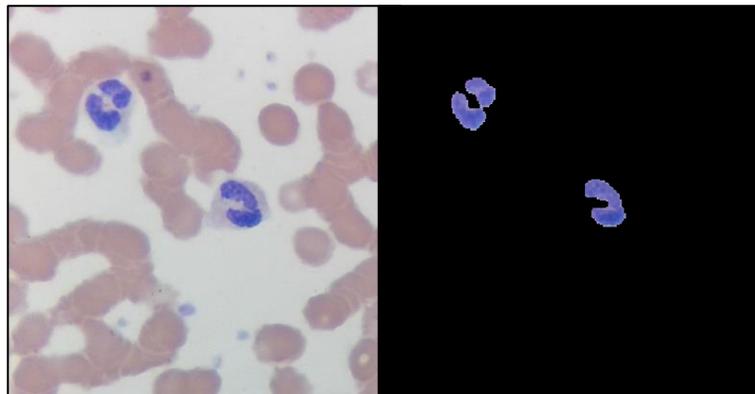
### 7.1 SEGMENTACIÓN

En esta subsección, se evalúa el rendimiento del algoritmo de agrupamiento de k-means. En el proceso de evaluación, se aplica la técnica propuesta individualmente a cada una de **147** imágenes de frotis de sangre; **139** de ellos fueron segmentados con precisión, alcanzando una exactitud de **94.6%**. La Fórmula 7, muestra la proporción de núcleos detectada correctamente, donde  $A_1 = 1$  indica que todos los núcleos se detectan correctamente en las imágenes microscópicas de frotis de sangre<sup>32</sup>.

$$A_1 = \frac{\text{Lacantidad de núcleos actualmente segmentados}}{\text{Número total de núcleos segmentados}} \times 100 \quad (7)$$

La Figura 25 ilustra los resultados obtenidos después de aplicar la segmentación de color utilizando el espacio de color HSI y la agrupación de K-means.

Figura 25. Resultado de la segmentación por K-means



a. Imagen original, donde se puede distinguir dos leucocitos. b. Segmentación de los núcleos basada en el color y el algoritmo de agrupación K-means.

### 7.2 CLASIFICACIÓN

La prueba consiste en clasificar **52** leucocitos presentes en imágenes obtenidos en el centro de salud de María Auxiliadora de Garzón, Huila, durante dos sesiones, las cuales contiene **14** linfocitos, **12** monocitos, **8** eosinófilos y **18** neutrófilos previamente clasificados por la bacterióloga. La Tabla 6 muestra el número de leucocitos de cada tipo presente en las imágenes anteriormente clasificados por la especialista frente al número de objetos

---

<sup>32</sup> Muhammad Sajjad, Siraj Khan, Zahoor Jan, Khan Muhammad, Hyeonjoon Moon, Jin Tae Kwak, Seungmin Rho, Sung Wook Baik, Irfan Mehmood. "Leukocytes Classification and Segmentation in Microscopic Blood Smear: A Resource-Aware Healthcare Service in Smart Cities". En: (2016).

detectados y clasificados por el sistema propuesto. Estos resultados se validaron en dicho centro de salud por la bacterióloga Sandra Rojas.

Tabla 6. Leucocitos detectados versus datos reales clasificados por un especialista

<b>Tipo</b>	<b>Datos Reales</b>	<b>Sistema propuesto</b>
<b>Linfocito</b>	14	15
<b>Monocito</b>	12	15
<b>Eosinófilo</b>	8	8
<b>Neutrófilo</b>	18	14

La eficacia de la aplicación se calcula en función del número de detecciones incluidas en la Tabla 6. La Tabla 7 muestra los resultados obtenidos para la clasificación de los tipos de leucocitos. Mientras que, la Tabla 8 muestra los valores logrados con las fórmulas relacionadas con una matriz de confusión, Fórmulas 8, 9 y 10<sup>33</sup>.

$$Sensibilidad = \frac{TP}{(TP+FN)} \quad (8)$$

$$Especificidad = \frac{TN}{(FP+TN)} \quad (9)$$

$$Exactitud = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (10)$$

Para los linfocitos, monocitos y eosinófilos se tiene una sensibilidad del **100%**, lo que indica la capacidad del sistema para dar como casos positivos los casos realmente positivos. Además, para los linfocitos, monocitos, eosinófilos y neutrófilos se alcanza una exactitud de **98.1%**, **90.4%**, **92.3%** y **88.5%**, respectivamente, y la exactitud total del sistema es de **92.3%**.

Tabla 7. Clasificación de los resultados obtenidos para cada tipo de leucocito

	<b>TP</b>	<b>TN</b>	<b>FP</b>	<b>FN</b>
<b>Linfocitos</b>	14	37	1	0
<b>Monocitos</b>	10	37	5	0
<b>Eosinófilos</b>	4	44	4	0
<b>Neutrófilos</b>	12	34	2	4

TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative

<sup>33</sup> Jaroonrut Prinyakupt and Charnchai Pluempitiwiriwaj. "Segmentation of white blood cells and comparison of cell morphology by linear and naïve Bayes classifiers". *En*: Prinyakupt and Pluempitiwiriwaj. *BioMed Eng. Online* (2015).

Tabla 8. Valores calculados de sensibilidad, especificidad y exactitud

<b>Tipo</b>	<b>Sensibilidad</b>	<b>Especificidad</b>	<b>Exactitud</b>
<b>Linfocitos</b>	1	0.974	0.981
<b>Monocitos</b>	1	0.881	0.904
<b>Eosinófilos</b>	1	0.917	0.923
<b>Neutrófilos</b>	0.75	0.944	0.885
<b>Total</b>	0.938	0.929	0.923

## 8. CONCLUSIONES

Fue posible efectuar la detección y clasificación de las células considerando únicamente características que describen el núcleo, de acuerdo con su morfología. La herramienta de software hace uso de técnicas de procesamiento de imágenes y de clasificación mediante LDA. La clasificación automática coincide en más del 92% con la realizada por la bacterióloga experta.

Por otro lado, durante la fase de adquisición de la imagen existieron condiciones que no se lograron controlar como la iluminación ambiental y de la fuente, por lo cual se optó por el uso de técnicas que atenuaran dichas variaciones, la ecualización del histograma y la conversión de las imágenes del espacio de color RGB al espacio cromático rg.

Se optó por eliminar los basófilos, puesto que el número es limitado según las imágenes adquiridas en el laboratorio y no se pueden generalizar, por lo cual se presenta un sistema de análisis de imágenes para reconocer únicamente cuatro grupos de glóbulos blancos en la sangre periférica.

Los resultados muestran que la segmentación no supervisada a través del algoritmo de agrupamiento de K-means se implementaron con éxito, con una exactitud del 94.6%, para imágenes más complejas, con un fondo complicado y que contienen varios eritrocitos (glóbulos rojos) cerca de los leucocitos, asimismo hay células con múltiples núcleos, donde se obtienen buenos resultados de segmentación.

La circularidad y excentricidad son características claves para diferenciar las células mononucleares de las polimorfonuclear, dado que sus núcleos son redondeados o ligeramente redondeados, excéntricos y no presentan lobulaciones. Con respecto las características número de lóbulos e intensidad mínima y máxima, son discriminantes para los eosinófilos y neutrófilos por la cantidad de lóbulos que presentan sus núcleos, además, el color en los eosinófilos es menos violeta que los neutrófilos, siendo características esenciales para distinguirlas del resto.

En términos de clasificación, LDA y QDA encuentra el hiperplano que mejor separa todos los puntos de datos mientras que SVM de kernel lineal busca el hiperplano que mejor separa solo los puntos en la frontera entre las dos clases, debido a esto se escogieron para realizar la clasificación de las células.

Tanto Python como sus bibliotecas y librerías son gratuitas, con un consumo de recursos bastante bajo, por lo que la hace esencial para la construcción de aplicaciones de procesamiento de imágenes. Con respecto al tiempo de procesamiento, el programa requiere 3s para analizar un leucocito en una PC Intel Core i5, corriendo a 2.3 GHz, con 4 GB de RAM. Por lo tanto, el conteo diferencial de 100 leucocitos dura aproximadamente 5 minutos.

En comparación, un experto requiere aproximadamente 15 minutos para llevar a cabo este proceso.

Los resultados son alentadores con un promedio de 93.8% de sensibilidad, 92.9% promedio de especificación y exactitud promedio del 92.3%, de manera que el sistema puede ser tomado como una alternativa confiable, eficiente y económica, esto último en comparación al precio de los sistemas electrónicos modernos que cumplen la misma función.

## 9. TRABAJOS FUTUROS

En primer lugar, sería deseable disponer de un mayor número de imágenes en las que se encuentren muestras de cada uno de los tipos de leucocitos, en cantidades suficientes para obtener resultados estadísticamente significativos. En efecto, con las imágenes disponibles no fue posible trabajar con los basófilos. La hipótesis por verificar sería que la metodología propuesta conduciría a resultados de clasificación igualmente confiables, pero con un mayor número de tipos de células clasificadas.

De otro lado, midiendo otras características sobre estos otros tipos de células con una base de datos más representativa, podría llegarse a encontrar nuevos parámetros simples y discriminantes que hagan más objetivos los procesos de diagnóstico.

Con un mayor número de muestras clasificadas y con más características discriminantes, sería de interés estudiar otras técnicas de clasificación basadas en inteligencia computacional o en otros métodos estadísticos para evaluar, comparar y seleccionar la más adecuada para este tipo de propósitos.

Dado que cada célula identificada tiene más subtipos, con una base de datos amplia se puede tener una descripción más completa de la célula y esto puede ampliar el número de clases a reconocer.

Lo ideal sería desarrollar un sistema automático de reconocimiento de células en tiempo real, adaptando una cámara fotográfica al microscopio. Donde se separe de una muestra de sangre en sus células constituyentes y extraer información útil al patólogo como cantidad porcentual, tamaño promedio y características morfológicas de los eritrocitos, leucocitos y plaquetas. La utilidad de las técnicas de procesamiento digital de imágenes en el desarrollo de herramientas para ayuda diagnóstica, en particular para el caso de la detección de leucocitos, es de gran importancia clínica en la identificación de diferentes patologías.

## BIBLIOGRAFÍA

ABDUL, Aimi Salehi, MASOR, Mohd Yusuff and MOHAMED, Zeehaida. “Colour Image Segmentation Approach for Detection of Malaria Parasiter using Various Colour Models and k-Means Clustering”, En: WSEAS Transaction on Biology and Biomedecine., vol. 10, (2013).

ALVAREZ, Alejandro. “Guia Tkinter Documentation”. Disponible en Internet : <https://media.readthedocs.org/pdf/guia-tkinter/master/guia-tkinter.pdf>. Consultado (02/02/2018).

ATIN, Mathur, ARDHENDU, S. Tripathi, and MANOHAR, Kuse. “Scalable system for classification of white blood cells from Leishman stained blood stain images”. En: Journal of Pathology Informatics (2013).

BEN-HUR, Asa; WESTON, “A User’s Guide to Support Vector Machines”. Disponible en Internet: <http://pymml.sourceforge.net/doc/howto.pdf>.

BHALOWALIA, Purnima and KUMAR, Arvind. “EBK-Means: A Clustering Technique based on Elbow Method and K-Means in WSN”. En : International Journal of Computer Application (2014).

CÁRDENAS, Montes Miguel. “Linear Discriminant Analysis”. Disponible en Internet : <http://wwwae.ciemat.es/~cardenas/docs/lessons/LDA.pdf>. Consultado (23/01/2018).

CARMONA, Suárez Enrique J. “Tutorial sobre Máquinas de Vectores Soporte (SVM)”. En: Dpto. de Inteligencia Artificial, ETS de Ingeniería Informática, Universidad Nacional de Educación a Distancia (UNED), C/Juan del Rosal, 16, 28040-Madrid (Spain) (2014).

CHHAYA, S. Hinge, AARTI, G. Ambekar and SAMIDHA, S. Kulkarni. “Classification of RBC And WBC in Peripheral Blood Smear Using KNN”. En: (2013).

FATMA, Mashiat and SHARMA, Jaya. “Leukemia Image Segmentation using K-Means Clustering and HSI Color Image Segmentation”. En: (2014).

GAVIN, Hackeling. “Mastering Machine Learning with Scikit-learn”. En:(2014).

GONZALEZ, Rafael C.and WOODS, Richard E. “Digital Image Processing”. pp. 91-93.  
ZHI, Liu, JING, Liu, XIAOYAN, Xiao, HUI, Yuan, XIAOMEI, Li, JUN, Chang and CHENGYN, Zheng.” Segmentation of White Blood Cells through Nucleus Mark Watershed Operations and Mean Shift Clustering”. En: (2015).

JOSEPH, Nelson. “Visual Representation of Train/Test Split and Cross Validation”. Disponible en Internet: <https://towardsdatascience.com/train-test-split-and-cross-validation-in-python-80b61beca4b6>. Consultado (25/01/2018).

KANADE, C. M., T. P. D., and ROWELY, "Anomaly Detection through registration, Pattern Recognition," ed, pp. 113-128 (2013).

MACSKASSY, Sofus A. “Machine Learning”. Consultado (21/01/2018).

MOYA, Ricardo. “Machine Learning (en Python)”. Disponible en Internet: <https://github.com/RicardoMoya?tab=overview&from=2015-09-01&to=2015-09-30>.

MUHAMMAD, Sajjad, SIRAJ, Khan, ZAHOOR, JAN, KHAN, Muhammad, HYEONJOON, Moon, JIN, Tae K., SEUNGMIN, Rho, SUNG, Wook Baik and IRFAN, Mehmood. “Leukocytes Classification and Segmentation in Microscopic Blood Smear: A Resource-Aware Healthcare Service in Smart Cities”. En: (2016).

NAMEIRAKPAM, Dhanachandra, KHUMANTHEM, Manglem and YAMBEM, Jina Chanu. “Image Segmentation using K-means Clustering Algorithm and Subtractive Clustering Algorithm”. En: Eleventh International Multi-Conference on Information Processin (2015).

NARANJO A. Carmen Beatriz. “Atlas de hematología células sanguíneas”. En: (2008), pp. 60-64.

OLAYA, B. Amparo, ORTIZ, D. Álvaro Enrique and SOTO, Wilson. “Learning Styles Selection for Geomatic Teaching using Intelligent Algorithms”. En: (2012).

PANAGIOTAKIS, C., RAMASSO, E., TZIRITAS, G. “Lymphocyte segmentation using the transferable belief model”. Lecture notes in Computer Science (LNCS). Recognizing Patterns in Signals, Speech, Images and Videos. Proceedings of the ICPR 2010 Contests, Springer. 2010:253–62. (2010).

“Scikit-learn user guide”. Disponible en Internet : <http://scikit-learn.org/stable/>.

SILVA, Edwin. 2006, Detección de cáncer de cuello uterino usando técnicas de procesamiento y análisis de imágenes microscópicas. Bucaramanga, 2006, 172 p. Tesis de Maestría (Ingeniería Electrónica). Universidad Industrial de Santander. Facultad de Ingenierías Fisicomecánicas. Escuela de Ingeniería Eléctrica, Electrónica y Telecomunicaciones.

## ANEXO A: CÓDIGOS DE PYTHON

- Parte de Código de Pre-procesamiento

---

```
import math
import numpy as np
from skimage import exposure
image1 = exposure.equalize_hist(image,2)
image1 = median_filter(image1, 3)
R = image1[:, :, 0]
G = image1[:, :, 1]
B = image1[:, :, 2]
HSV = rgb2hsv(image1)
S = HSV[:, :, 1]
r = R / (R + G + B)
g = G / (R + G + B)
b = 1 - r - g
croma_rg = np.dstack((r, g, b))
```

---

- Parte de Código de Segmentación

---

```
import numpy as np
from sklearn.cluster import KMeans
# croma_rg Imagen en el espacio croma rg"
rows = croma_rg.shape[0]
cols = croma_rg.shape[1]
image = croma_rg.reshape(croma_rg.shape[0] *
croma_rg.shape[1], 3)
kmeans = KMeans(n_clusters=3, n_init=10, max_iter=3)
kmeans.fit(image)

clusters = np.asarray(kmeans.cluster_centers_,
dtype=np.uint8)
labels = np.asarray(kmeans.labels_, dtype=np.uint8)
labels = labels.reshape(rows, cols)
```

---

- Parte de Código Extracción de Características

---

```
import math
```

```

import numpy as np
import pandas as pd
import os
from skimage.measure import regionprops, label,
shannon_entropy

img_yiq = rgb2yiq(img)
intensity_img = img_yiq[:, :, 0]
label_img = label(img_binary)
props = regionprops(label_img, intensity_image=intensity_img)
ffactor = (4 * math.pi * props[i].area) /
(math.pow(props[i].perimeter, 2))
b_box = props2[i].intensity_image
entropia = shannon_entropy(b_box, base=2)
No_Lobulos= (props[i].extent + props[i].eccentricity +
ffactor) / 3
fdata.at[i, 'Area'] = props[i].area
fdata.at[i, 'Extent'] = props[i].extent
fdata.at[i, 'FormFactor'] = ffactor
fdata.at[i, 'Eccs'] = props[i].eccentricity
fdata1.at[i, 'Min_Intensity'] = props[i].min_intensity
fdata1.at[i, 'Max_Intensity'] = props[i].max_intensity
fdata1.at[i, 'Entropy'] = entropia
fdata2.at[i, 'Eccs'] = props[i].eccentricity
fdata2.at[i, 'No_Lobulos'] = No_Lobulos

```

---

- Parte de Código Clasificación

---

```

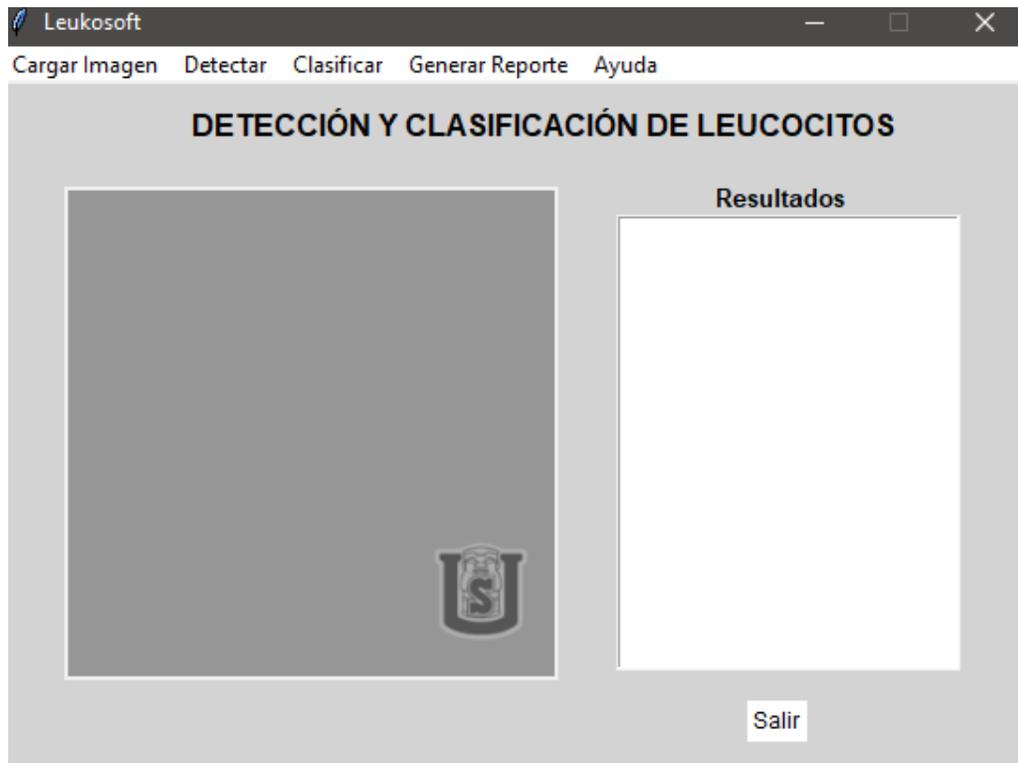
import numpy as np
from sklearn.discriminant_analysis import
LinearDiscriminantAnalysis
X = np.array(pd.read_csv('data.csv', header=0))
y = np.array(pd.read_csv('target.csv', header=0))
qda = QuadraticDiscriminantAnalysis().fit(X, y)
leu = np.array(fdata)
leu = leu[:, -3:]
leu = leu.reshape(1, -1)
prediction = qda.predict(leu)

```

## ANEXO B: GUÍA DE USUARIO

**Ejecución:** Para iniciar el software basta con ubicarse en el directorio donde se encuentra la aplicación.

Figura 26. Interfaz inicial presentada al especialista



**Cargar una imagen:** Para cargar una imagen, se debe dar clic al botón “Cargar Imagen” e inmediatamente se muestra un cuadro de dialogo, que se abre por defecto en el directorio que contiene el banco de imágenes tomadas. A continuación, se puede elegir la imagen que se desea analizar. Si no abre una imagen, los botones “Detectar”, “Clasificar” y “Generar Reporte”, no se pueden utilizar. Figura 27 y Figura 28.

Figura 27. Cuadro de diálogo para abrir una imagen

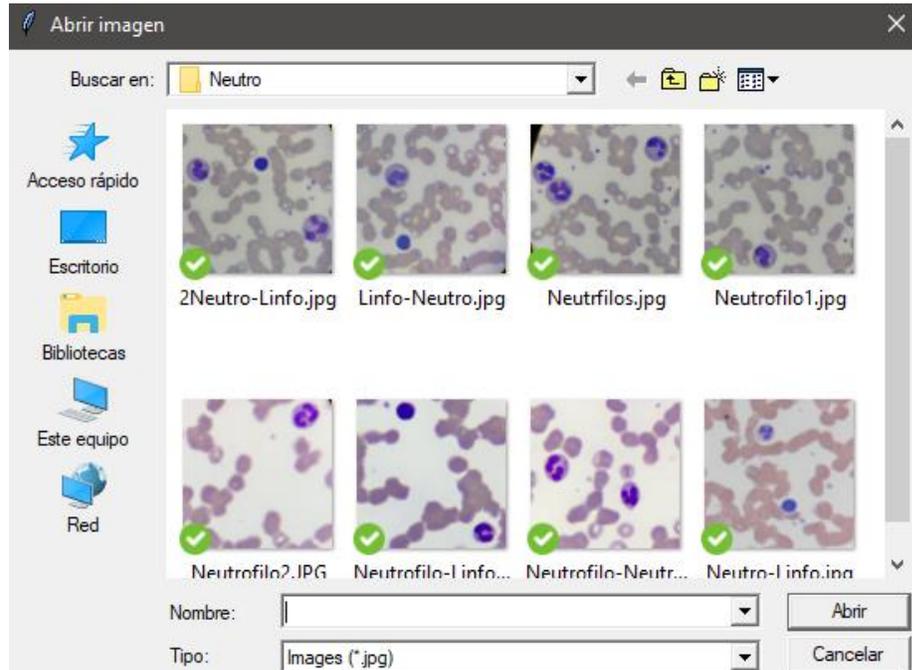
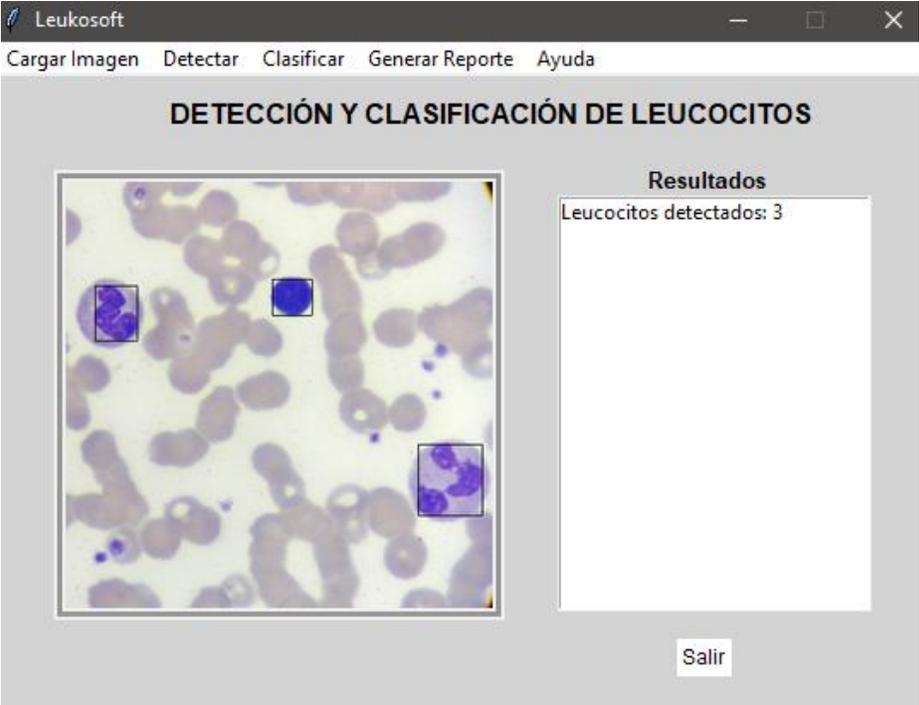


Figura 28. Imagen cargada en el espacio de trabajo



**Detectar leucocitos:** después de haber cargado la imagen deseada del banco de imágenes se procede a detectar el núcleo de los leucocitos, dando clic en el botón “Detectar”, Figura 29.

Figura 29. Imagen en la cual se ha detectado los leucocitos



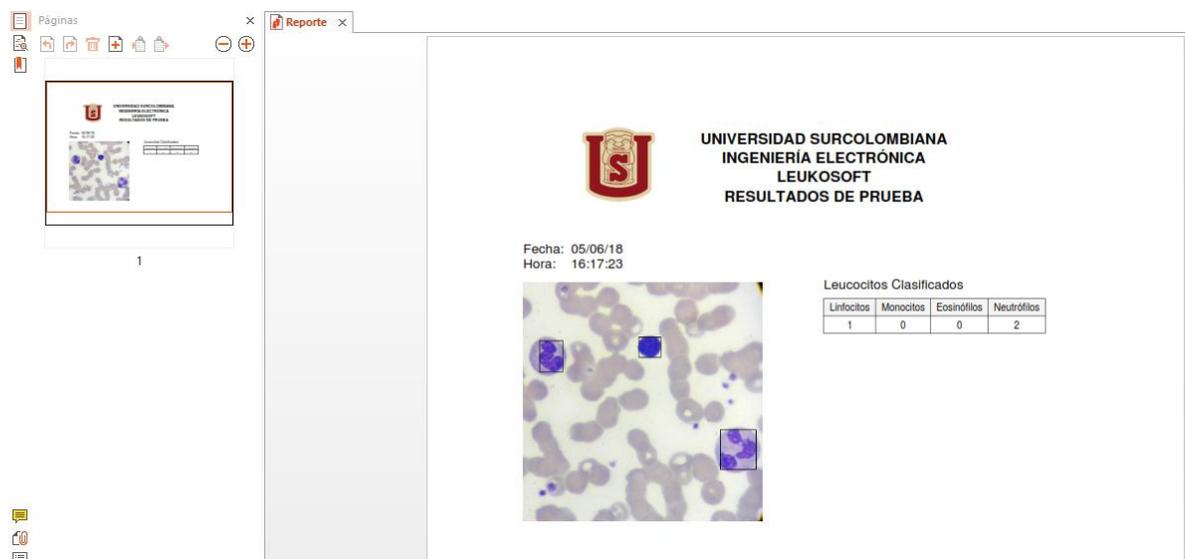
**Clasificar leucocitos:** Una vez detectado los leucocitos se puede realizar la clasificación de estos, dando clic en el botón “Clasificar”, Figura 30.

Figura 30. Imagen en la cual se ha clasificado los leucocitos



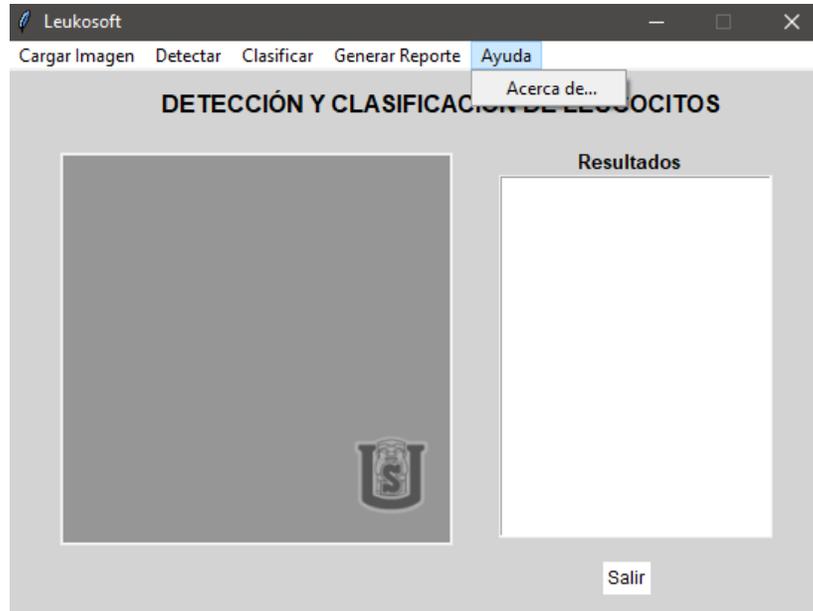
**Generar Reporte:** Ya realizada la clasificación de los leucocitos se puede generar un archivo PDF con el reporte de los resultados obtenidos, Figura 31.

Figura 31. Reporte generado en formato PDF



**Ayuda:** Al dar clic en el botón “Ayuda” se despliega un nuevo menú donde aparecen dos nuevos botones “Guía de Usuario” y “Acerca de...”, Figura 32.

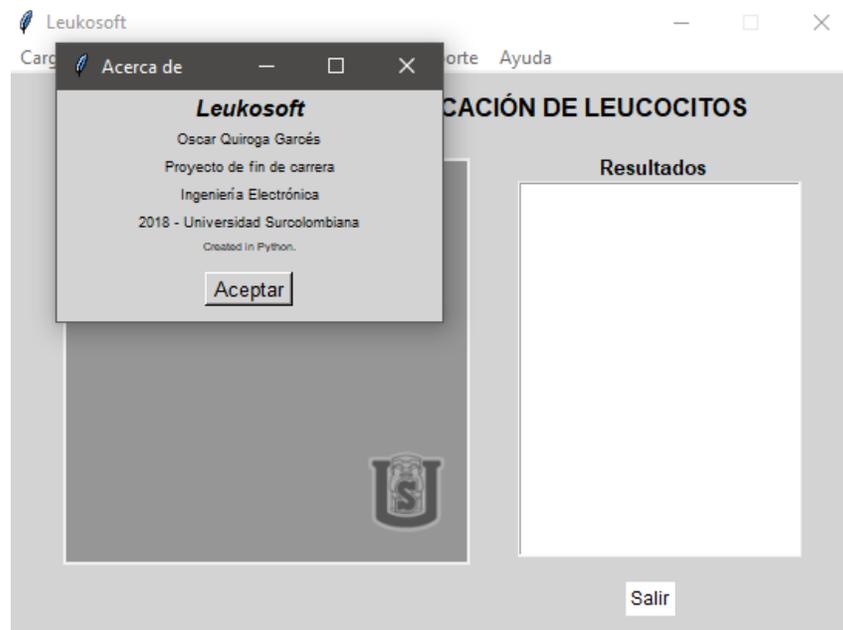
Figura 32. Ayuda



**Guía de usuario:** Proporciona la guía de usuario.

**Acerca de:** Genera una pequeña descripción sobre la aplicación, Figura 33.

Figura 33. Acerca de ...



**Salir:** Si se desea cerrar la aplicación basta con dar un clic en el botón “Salir”.