



UNIVERSITAT DE
BARCELONA

Clasificación automática de malaria e inclusiones eritrocitarias anómalas en sangre periférica utilizando modelos de *machine learning*

Angel Molina Borrás

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (www.tdx.cat) i a través del Dipòsit Digital de la UB (diposit.ub.edu) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (www.tdx.cat) y a través del Repositorio Digital de la UB (diposit.ub.edu) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (www.tdx.cat) service and by the UB Digital Repository (diposit.ub.edu) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.



UNIVERSITAT DE
BARCELONA

FACULTAD DE MEDICINA
DEPARTAMENTO DE BIOINGENIERIA
PROGRAMA DE DOCTORADO DE BIOMEDICINA

CLASIFICACIÓN AUTOMÁTICA DE MALARIA E INCLUSIONES ERITROCITARIAS ANÓMALAS EN SANGRE PERIFÉRICA UTILIZANDO MODELOS DE MACHINE LEARNING

Memoria presentada por **Angel Molina Borrás** para optar al título de doctor por la
Universidad de Barcelona

Directores:

Anna Merino González
Consultora Senior
Laboratorio Core
Hospital Clinic de Barcelona

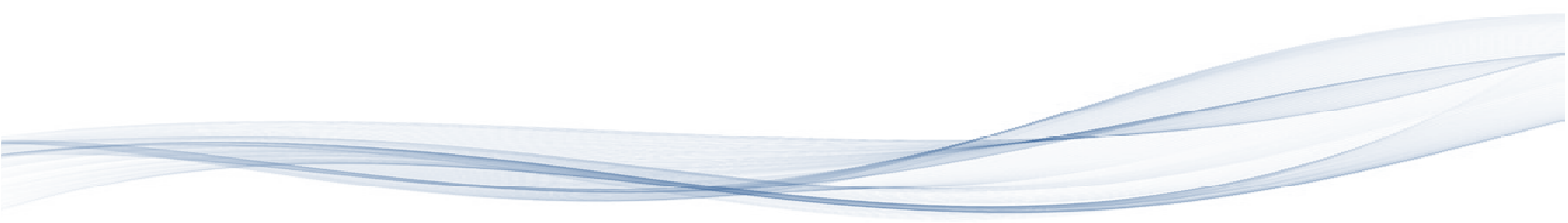
José Rodellar Benedé
Catedrático
Dpto. de Matemáticas
Universitat Politècnica de Catalunya

Tutora:

Roser Sala Llonch
Profesora asociada
Dpto. de Biomedicina
Universitat de Barcelona



Índice



Listado de Figuras y Tablas.....	I
1. Listado de Figuras.....	III
2. Listado de Tablas.....	V
3. Abreviaturas.....	VII
Resumen.....	IX
 CAPÍTULO 1	
Introducción	1
1.1 MOTIVACIÓN.....	3
1.2 ORGANIZACIÓN DE LA TESIS	4
 CAPÍTULO 2	
Alteraciones de la morfología eritrocitaria.....	3
2.1 Hemograma y frotis sanguíneo	3
2.2 Morfología eritrocitaria normal y anormal	4
2.3 La malaria	11
 CAPÍTULO 3	
Estado del arte:	15
Morfología eritrocitaria e inteligencia artificial.....	15
3.1 Segmentación de hematíes.....	18
3.2 Extracción de características.....	21
3.3 Métodos de clasificación.....	23
3.4 Redes neuronales y aprendizaje profundo	26
3.5 Analizadores para el reconocimiento automático de hematíes	29
 CAPÍTULO 4	
Objetivos.....	35
4.1 Objetivo general.....	37
4.2 Objetivos específicos.....	37

CAPÍTULO 5

Reconocimiento automático de malaria mediante modelos de *machine learning*.....41

5.1	PROCESAMIENTO DIGITAL DE IMÁGENES.....	43
5.1.1	Obtención de muestras	43
5.1.2	Obtención de imágenes	45
5.1.3	Etiquetado	46
5.1.4	Preprocesamiento	47
5.1.5	Binarización	50
5.1.6	Postprocesamiento y segmentación	51
5.2	EXTRACCIÓN DE CARACTERÍSTICAS.....	55
5.3	CLASIFICACIÓN AUTOMÁTICA.....	62
5.3.1	Clasificador binario.....	62
5.3.1.1	Entrenamiento del clasificador binario	63
5.3.1.2	Evaluación del clasificador binario	65
5.3.2	Sistema de clasificación secuencial.....	65
5.3.2.1	Entrenamiento del sistema de clasificación secuencial	66
5.3.2.2	Evaluación del sistema de clasificación secuencial	68
5.4	RESULTADOS	69
5.4.1	Segmentación.....	69
5.4.2	Clasificación.....	73
5.4.2.1	Clasificador binario para hematíes normales <i>versus</i> malaria.....	73
5.4.2.2	Sistema de Clasificación secuencial.....	77
5.4.2.2.1	Evaluación del sistema secuencial utilizando el conjunto de evaluación	86
5.4.2.2.2	Evaluación del sistema utilizando el conjunto de diagnóstico.....	87
5.5	PUBLICACIONES.....	90

CAPÍTULO 6

Reconocimiento automático de malaria y otras inclusiones eritrocitarias mediante modelos de *deep learning*.....91

6.1	REDES NEURONALES CONVOLUCIONALES.....	94
6.1.1	Estructura general.....	94
6.1.2	Entrenamiento del modelo	97
6.1.3	Evaluación del modelo	99
6.2	RESULTADOS	102
6.2.1	Entrenamiento del modelo	102
6.2.2	Evaluación del modelo	105
6.3	PUBLICACIONES.....	111

CAPÍTULO 7

Reconocimiento automático de malaria y otras inclusiones eritrocitarias mediante *deep learning* utilizando imágenes obtenidas con CellaVision 113

7.1	MATERIAL Y MÉTODOS	115
7.1.1	Adquisición de imágenes mediante el sistema CellaVision	115
7.1.2	Entrenamiento del modelo	117
7.1.3	Evaluación del modelo	118
7.2	RESULTADOS	119
7.2.1	Entrenamiento del modelo	119
7.2.2	Evaluación del modelo	121

CAPÍTULO 8

Discusión125

CAPÍTULO 9

Conclusiones137

9.1	CONCLUSIONES	139
-----	--------------------	-----

9.1.1 Conclusiones derivadas del primer objetivo específico basado en el reconocimiento de malaria y otras inclusiones eritrocitarias mediante modelos de <i>machine learning</i>	139
9.1.2 Conclusiones derivadas del segundo objetivo específico basado en el reconocimiento de malaria y otras inclusiones eritrocitarias mediante modelos de <i>deep learning</i>	141
9.1.3 Conclusiones derivadas del tercer objetivo específico basado en el reconocimiento de malaria y otras inclusiones eritrocitarias mediante modelos de <i>deep learning</i> utilizando imágenes obtenidas con el analizador CellaVision	142
9.2 CONTRIBUCIONES	142
9.3 PERSPECTIVAS FUTURAS	143

CAPÍTULO 10

Bibliografía.....	145
--------------------------	------------

CAPÍTULO 11

Producción científica.....	155
11.1 Artículos publicados	171
11.2 Artículos en colaboración con el grupo de investigación.....	171
11.3 Publicaciones en congresos.....	172
11.4 Conferencias y cursos impartidos en relación a la tesis doctoral	174

CAPÍTULO 12

ANEXO I	171
----------------------	------------



Listado de Figuras y Tablas

Abreviaturas

1. Listado de Figuras

Figura 1. Alteraciones del tamaño de los hematíes.

Figura 2. Alteraciones de la coloración de los hematíes.

Figura 3. Alteraciones de forma de los hematíes.

Figura 4. Presencia de inclusiones en el interior de los hematíes.

Figura 5. Alteraciones en la distribución de los hematíes.

Figura 6. Ciclo de vida de la malaria.

Figura 7. Aspecto macro y microscópico de una extensión y una gota gruesa de sangre periférica.

Figura 8. Imagen de campo de un frotis de sangre periférica.

Figura 9. Analizadores para la revisión automatizada del frotis de sangre periférica.

Figura 10. Procedimiento de la extensión de sangre periférica.

Figura 11. Efecto de la tinción sobre una muestra de sangre periférica.

Figura 12. Gradación de la escala de grises en imágenes de 12 y 8 bits de profundidad.

Figura 13. Esquema de la obtención de imágenes desde la toma de muestra.

Figura 14. Proceso de etiquetado de los hematíes en una imagen de campo mediante la aplicación *Image Labeller*.

Figura 15. Representación del histograma y de las imágenes en los canales rojo, verde y azul del espacio de color RGB.

Figura 16. Representación del histograma y de las imágenes al aplicar una ecualización del histograma.

Figura 17. Comparación de las imágenes al aplicar un filtro Gaussiano.

Figura 18. Representación de la binarización mediante el método de Otsu.

Figura 19. Representación de la imagen en el canal b del espacio de color Lab.

Figura 20. Eliminación de los leucocitos y plaquetas de la imagen binarizada.

Figura 21. Representación del efecto de la función “rellenar huecos”.

Figura 22. Resultados de la segmentación mediante el método de Watershed.

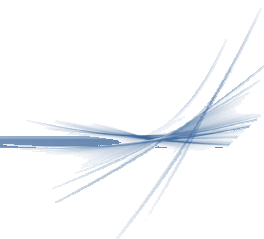


Figura 23. Resultados de la imagen final antes de aplicar el proceso de recorte de los hematíes individuales.

Figura 24. Resumen de los algoritmos aplicados a la imagen original de campo con la finalidad de obtener las imágenes de hematíes individuales.

Figura 25. La imagen enmascarada nos permite analizar únicamente los píxeles de la imagen que realmente son relevantes.

Figura 26. Ejemplo de cálculo de la matriz de coocurrencia (GLCM).

Figura 27. Ejemplo de cálculo de la matriz de largo recorrido (GLRLM).

Figura 28. Ejemplo de cálculo de la matriz por tamaño de zona (GLSZM).

Figura 29. Ejemplo de cálculo de la matriz por diferencia de tonalidad en la vecindad (NGTDM).

Figura 30. Ejemplo de cálculo de la matriz de dependencia (GLDM).

Figura 31. Esquema de entrenamiento mediante la validación cruzada 5x.

Figura 32. Diseño del clasificador basado en herramientas clásicas de *machine learning*. Estructura y funcionamiento de sus módulos.

Figura 33. Resultado del algoritmo de Watershed sobre hematíes de una imagen de campo con diferente grado de solapamiento.

Figura 34. Ejemplos de la metodología de segmentación

Figura 35. Resultados del entrenamiento de los diferentes algoritmos de clasificación.

Figura 36. Descriptores más relevantes en el clasificador SVM de hematíes normales y hematíes con malaria.

Figura 37. Resultados del entrenamiento de cada uno de los módulos que componen el sistema final de reconocimiento de inclusiones eritrocitarias.

Figura 38. Descriptores más relevantes en el módulo encargado de reconocer entre hematíes normales y hematíes con cualquier tipo de inclusión.

Figura 39. Representación de diferentes tipos de hematíes en el canal U del espacio de color YUV y en el canal Cb del espacio de color YCbCr.

Figura 40. Descriptores más relevantes en el módulo encargado de reconocer entre hematíes con malaria y hematíes con cualquier otro tipo de inclusión.

Figura 41. Representación de diferentes tipos de hematíes en el canal H del espacio de color LCH y en el canal B del espacio de color LAB.

Figura 42. Descriptores más relevantes en el módulo encargado de reconocer entre hematíes con cuerpos de Howell-Jolly (HJ), con cuerpos de Pappenheimer (PP), con punteado basófilo (PB), y con plaquetas alojadas en su superficie (PLT).

Figura 43. Representación de diferentes tipos de hematíes en el canal R del espacio de color RGB y en el canal E del espacio de color HED.

Figura 44. Esquema de la estructura de una red neuronal convolucional.

Figura 45. Representación del desarrollo de un modelo basado en redes neuronales convolucionales.

Figura 46. Representación de las imágenes generadas por generación artificial.

Figura 47. Monitorización de la función de pérdidas en función de la tasa de aprendizaje (A) y del número de ciclos de entrenamiento (B).

Figura 48. Mapas de características obtenidos del modelo VGG-16 entrenado y sin entrenar.

Figura 49. Distribución de valores de correlación en los mapas de características obtenidos de la quinta capa convolucional de los modelos entrenados y sin entrenar.

Figura 50. Mapas de activación obtenidos con el modelo VGG-16 entrenado.

Figura 51. Esquema de obtención de imágenes de hematíes individuales a partir de la imagen de campo obtenida por el sistema CellaVision.

Figura 52. Visión general de la clasificación realizada por CellaVision.

2. Listado de Tablas

Tabla 1. Número de frotis e imágenes individuales que constituyen los diferentes conjuntos de entrenamiento y evaluación.

Tabla 2. Resultados de la evaluación del modelo SVM que diferencia entre hematíes normales y hematíes infectados con malaria.

Tabla 3. Matriz de confusión en el conjunto de evaluación por imágenes individuales.

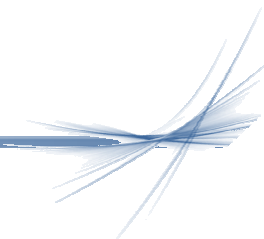


Tabla 4.1. Matriz de confusión en el conjunto de evaluación por frotis completos en el grupo de muestras de pacientes infectados con malaria.

Tabla 4.2. Matriz de confusión en el conjunto de evaluación por frotis completos en el grupo de muestras de pacientes no infectados.

Tabla 5. Distribución de imágenes en el conjunto de evaluación por frotis.

Tabla 6. Distribución de imágenes en los diferentes conjuntos de entrenamiento, validación y evaluación.

Tabla 7. Resultados de la clasificación sobre el conjunto de validación de diferentes arquitecturas de redes neuronales convolucionales.

Tabla 8. Matriz de confusión que refleja los resultados de la clasificación a nivel de imágenes individuales del conjunto de evaluación.

Tabla 9. Matriz de confusión que refleja los resultados de la clasificación a nivel de frotis completos del conjunto de evaluación.

Tabla 10. Número total de imágenes y frotis utilizados en los conjuntos de entrenamiento y evaluación.

Tabla 11. Resultados en porcentaje de la clasificación de las diferentes arquitecturas de CNN utilizadas sobre el conjunto de imágenes de validación.

Tabla 12. Matriz de confusión del sistema CellaVision sobre un conjunto de 48 frotis de sangre periférica.

Tabla 13. Matriz de confusión del modelo seleccionado y entrenado (Resnet-34) sobre un conjunto de 48 frotis de sangre periférica.

Tabla 14. Resultados de la clasificación sobre las mismas muestras a nivel de frotis completos tanto por el sistema Cellavision (izquierda) como por el modelo desarrollado Resnet-34.

Tabla 15. Comparación entre los indicadores de rendimiento en la clasificación de las principales publicaciones que utilizan modelos basados en *deep learning*.

3. Abreviaturas

PB: Punteado basófilo

HJ: Cuerpos de Howell-Jolly

PP: Cuerpos de Pappenheimer

ICHS: comité internacional de estandarización en hematología

ROI: Regiones de interés

RGB: Espacio de color rojo, verde y azul

HSV: Espacio de color matiz, saturación y valor

VPP: Valore predictivo positivo

SVM: Máquinas de soporte vectorial

KNN: K vecinos más cercanos

LDA: Análisis discriminante lineal

YCbCr: espaciodecolor luma, crominancia de azul y crominancia de rojo.

LR: Regresión logística

LBP: patrones binarios locales

SIFT: Descriptores de transformación de características invariantes de escala

VAR: Varianza de la textura local de la imagen

DAPI : 4',6-Diamidino-2-fenilindol diclorhidrato

MLP: Perceptrones multicapa

ANN: Redes neuronales artificiales

CNN: Redes neuronales convolucionales

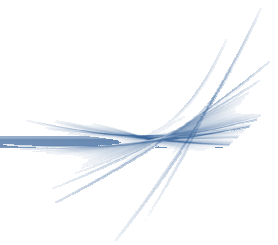
ARBCA: *Advance Red Blood Cell Application*

PLT: Plaquetas alojadas en la superficie de los hematíes.

EDTA: Ácido etieldiamínico tetraacético

MGG: May Grunwald-Giemsa

LAB: luminosidad de negro a blanco, de rojo a verde, y de azul a amarillo (espacio de color).



GLCM: Matriz de coocurrencia de los niveles de grises.

GLRLM: Matriz de largo recorrido de los niveles de grises.

GLSZM: Matriz por tamaños de zona de los niveles de grises.

NGTDM: Matriz por diferencia de tonalidad de grises en la vecindad.

GLDM: Matriz de dependencia de los niveles de grises.

XYZ: Espacio de color de los valores triestímulos.

HED: Espacio de color hematoxilina, eosina, DAB.

YIQ: Espacio de color luminancia, en fase y cuadratura.

LCH: Espacio de color brillo, saturación, tono.

YUV: Espacio de color luma, crominancia de proyección azul, crominancia de proyección roja.

CMYK: espacio de color cyan. magenta, amarillo, negro.

RF: Bosques aleatorios.

GNB: Clasificador de Bayes Gaussiano.

CMIM: Maximización de la información condicional mutua.

ReLU: función de activación lineal rectificada

TA: Tasa de aprendizaje.



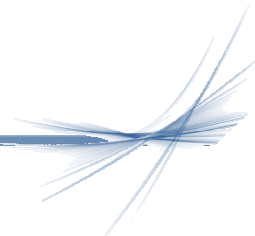
Resumen

Los especialistas de laboratorio juegan un papel esencial en el diagnóstico del paludismo mediante la identificación de características morfológicas cualitativas de los hematíes en la revisión del frotis de sangre periférica. Este análisis visual conlleva un consumo de tiempo, requiere un personal bien entrenado, y es sensible a la variabilidad entre observadores. .

El objetivo general de esta tesis doctoral es el reconocimiento automático de hematíes parasitados con malaria, así como su diferenciación no solo de hematíes normales sino también de hematíes con otros tipos de inclusiones, mediante herramientas de inteligencia artificial. La primera parte de este trabajo comienza con el desarrollo de un algoritmo de segmentación de hematíes a partir de imágenes de sangre periférica obtenidas con microscopio y el desarrollo de un modelo de reconocimiento automático basado en aprendizaje automático (*machine learning*) a partir de la extracción de descriptores cuantitativos de las imágenes segmentadas. En una segunda parte de esta tesis, el reconocimiento automático de la malaria se aborda a través de modelos basados en redes neuronales convolucionales. En último lugar, se desarrolla un modelo para la identificación de malaria basado en redes neuronales convolucionales, pero en este caso, mediante el uso de imágenes de sangre periférica obtenidas con el analizador CellaVision.

En comparación con el actual estado del arte, esta tesis doctoral supone un avance significativo en la implementación de modelos de reconocimiento automático de malaria en un entorno clínico realista. Esto es posible principalmente al haber incluido en los modelos de clasificación otras inclusiones eritrocitarias que tienen gran similitud morfológica con los parásitos de la malaria. Con esta estrategia ha podido evitarse el alto número de falsos positivos en la detección de malaria observado en modelos previos de reconocimiento automático.

Palabras clave: *Malaria, Inclusiones eritrocitarias, Análisis morfológico, Reconocimiento automático, Análisis digital de imágenes, Inteligencia artificial, Aprendizaje profundo, Redes neuronales convolucionales.*





CAPÍTULO 1

Introducción

1.1 MOTIVACIÓN

Esta tesis doctoral se enmarca dentro de una línea conocida recientemente como “hematología computacional”. El contexto general engloba el desarrollo y la implementación de modelos y algoritmos que ayuden al patólogo clínico en el diagnóstico y la toma de decisiones a partir de variables medidas en el laboratorio. Más en concreto, este trabajo se enfoca en el análisis de la morfología eritrocitaria, en la revisión del frotis de sangre periférica para la detección de malaria y de las alteraciones eritrocitarias asociadas a esta enfermedad.

El análisis morfológico visual del frotis es esencial para la detección de anomalías. Sin embargo, una revisión adecuada requiere tiempo y la disponibilidad de profesionales con un alto grado de experiencia. Aun así, la interpretación de ciertas alteraciones tiene un componente subjetivo asociado al observador, lo que implica que la revisión del frotis posea un grado adicional de variabilidad en comparación con otras técnicas de laboratorio automatizadas. Un ejemplo de esta variabilidad puede encontrarse en la malaria, una enfermedad potencialmente mortal cuyo diagnóstico se basa en el análisis visual de los hematíes en un frotis de sangre periférica. Los organizadores de programas de evaluación externa de la calidad muestran que frecuentemente los laboratorios no detectan la presencia de malaria en un frotis de sangre periférica, lo que demuestra la variabilidad en su detección y diferenciación.

Las desventajas asociadas con el análisis morfológico ya mencionadas, han contribuido en los últimos años al desarrollo de métodos y técnicas basadas en el análisis de imágenes digitales y en métodos de inteligencia artificial, como el aprendizaje automático (*machine learning*) y el aprendizaje profundo (*deep learning*). Esta combinación puede contribuir a llevar a la microscopía a una forma más rápida, más eficiente y estandarizada de realizar el análisis morfológico.

Existe una amplia y reciente bibliografía a nivel de investigación y desarrollo de métodos de inteligencia artificial para el reconocimiento automatizado de hematíes infectados con malaria. De su análisis, que será presentado más adelante en esta memoria, se observa que la mayoría de modelos presentan un notable número de falsos positivos debido a la presencia de otros tipos de inclusiones eritrocitarias que son clasificadas erróneamente como parásitos de malaria. Este hecho es la principal motivación para la realización de esta tesis doctoral, que pretende contribuir con nuevos modelos de *machine learning* y *deep learning* capaces de



reconocer hematíes con malaria y también hematíes con otras inclusiones eritrocitarias, tales como los cuerpos de Howell-Jolly, los cuerpos de Pappenheimer, el punteado basófilo, y los hematíes con plaquetas alojadas en su superficie.

1.2 ORGANIZACIÓN DE LA TESIS

Esta tesis sigue el formato de tesis clásico, según lo establecido en la normativa de la Universidad de Barcelona y se organiza en 12 capítulos:

El presente Capítulo 1 explica brevemente el marco global que ha servido de motivación para la realización de esta tesis doctoral, así como su estructura.

El Capítulo 2 servirá de introducción para definir la importancia del análisis de la morfología eritrocitaria en la revisión del frotis de sangre periférica para la detección de diferentes patologías. Se introducirá también la relevancia clínica de la malaria, así como las alteraciones eritrocitarias asociadas a esta enfermedad. Estos dos apartados constituyen el contexto clínico sobre la que se asienta la tesis doctoral.

El Capítulo 3 resume el estado del arte de las principales herramientas de inteligencia artificial actualmente desarrolladas para la detección de malaria.

El Capítulo 4 detalla los objetivos que se han definido para esta tesis doctoral.

El Capítulo 5 muestra el desarrollo del primer objetivo específico de la tesis doctoral, basado en el desarrollo de un modelo de reconocimiento automático de malaria y otras inclusiones eritrocitarias mediante herramientas de *machine learning* a partir de imágenes obtenidas mediante microscopía. Se detallan los métodos desarrollados y los resultados obtenidos asociados a este primer objetivo.

El Capítulo 6 muestra el desarrollo del segundo objetivo específico de la tesis doctoral, basado en el desarrollo de un modelo de reconocimiento automático de malaria y otras inclusiones eritrocitarias mediante herramientas de *deep learning* a partir de imágenes obtenidas mediante microscopía. Se detallan los métodos desarrollados y los resultados obtenidos asociados a este segundo objetivo.

El Capítulo 7 muestra el desarrollo del tercer objetivo específico de la tesis doctoral, basado en el desarrollo de un modelo de reconocimiento automático de malaria y otras inclusiones eritrocitarias mediante herramientas de *deep learning* a partir de imágenes obtenidas mediante el autoanalizador CellaVision. Se detallan los métodos desarrollados y los resultados obtenidos asociados a este tercer objetivo.

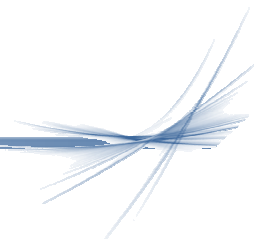
El Capítulo 8 realiza una discusión de los resultados obtenidos a lo largo de la tesis doctoral.

El Capítulo 9 detalla las conclusiones y contribuciones más significativas derivadas de la tesis doctoral.

El Capítulo 10 enumera la bibliografía utilizada en esta memoria.

El Capítulo 11 resume la actividad científica producida asociada a este trabajo de investigación.

Por último, el Capítulo 12 constituye un anexo donde se muestran los artículos publicados a lo largo de la elaboración de la tesis doctoral.





CAPÍTULO 2

Alteraciones de la morfología eritrocitaria

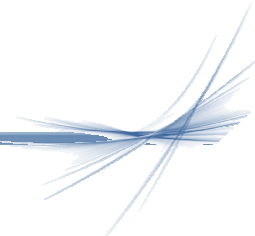
En este Capítulo se describe el contexto clínico de la presente tesis doctoral, cuyo eje central gira en torno a la revisión del frotis de sangre periférica. A continuación, se analizarán los siguientes apartados:

- La importancia del análisis de la morfología eritrocitaria en la revisión del frotis de sangre periférica para la detección de diferentes patologías (Apartados 2.1 y 2.2).
- La relevancia clínica de la malaria, así como las alteraciones eritrocitarias asociadas a esta enfermedad (Apartado 2.3).

2.1 Hemograma y frotis sanguíneo

Los laboratorios de Análisis Clínicos se fundamentan en aquellas actuaciones que, a través de métodos diagnósticos analíticos, ayudan al diagnóstico, pronóstico, terapéutica médica y prevención de las enfermedades. Una de las pruebas más solicitadas en el laboratorio de análisis clínicos es el hemograma (1). El hemograma es un perfil analítico solicitado como parte del examen médico rutinario para monitorizar el estado de salud general y para cribar una gran variedad de enfermedades como los cuadros de anemia, las infecciones, y la leucemia entre otras patologías.

El hemograma constituye la base del diagnóstico en el laboratorio de hematología y proporciona información importante acerca de las diferentes poblaciones celulares sanguíneas, caracterizando la serie leucocitaria, la serie roja y la serie plaquetar a través de los diferentes parámetros hematológicos. Cuando el hemograma es realizado por los analizadores de hematología, la presencia de resultados anormales desencadena alarmas cuantitativas y/o cualitativas que pueden indicar la necesidad de realizar un frotis de sangre periférica (2). La revisión manual del frotis de sangre periférica implica la observación al microscopio de una gota de sangre adecuadamente extendida sobre un portaobjetos y teñida con un colorante apropiado. Esta revisión microscópica proporciona información adicional y es parte esencial de la orientación diagnóstica. La evaluación conjunta de los resultados del hemograma y del frotis de sangre periférica constituye la esencia en el diagnóstico y seguimiento de diferentes patologías, adquiriendo una especial relevancia en las enfermedades hematológicas.



2.2 Morfología eritrocitaria normal y anormal

Una correcta evaluación morfológica de los hematíes constituye un aspecto clave en el diagnóstico diferencial ante la aparición de un cuadro de anemia no filiada. La mayoría de los glóbulos rojos o eritrocitos normales tienen forma de disco bicóncavo. En una extensión de sangre periférica teñida los hematíes son aproximadamente de contorno circular y en condiciones normales muestran solo pequeñas variaciones en la forma y variaciones moderadas de tamaño. En promedio, el diámetro de un eritrocito es de aproximadamente 7,5 μm y en el área del frotis de sangre periférica donde las células forman una monocapa se observa un área central más pálida que ocupa aproximadamente el tercio medio de la célula. Para realizar una evaluación de la morfología eritrocitaria es necesario considerar los siguientes aspectos de los hematíes (3):

Alteraciones en el tamaño (ver Figura 1)

Microcitosis: Es una disminución en el tamaño de los eritrocitos. Los microcitos se detectan en un frotis de sangre por una reducción del diámetro de los glóbulos rojos a menos de 7 a 7,2 μm . La microcitosis se refleja en el hemograma con una reducción en el volumen corpuscular medio por debajo de los valores de referencia. alguna de las causas más comunes de la microcitosis puede ser la talasemia o la anemia por déficit de hierro.

Macrocitosis: Supone un aumento en el tamaño de los eritrocitos que en algunas ocasiones puede ser fisiológica o consecuencia de ciertas patologías. La macrocitosis se refleja en el hemograma con un aumento en el volumen corpuscular medio por encima de los valores de referencia. Algunas de las causas más comunes de la macrocitosis son la anemia por déficit de ácido fólico o vitamina B₁₂, las hepatopatías o los síndromes mielodisplásicos.

Anisocitosis: Se trata de un aumento en la variabilidad del tamaño de los eritrocitos. La anisocitosis es una anomalía común e inespecífica en los trastornos hematológicos. La anisocitosis se refleja en el hemograma con un aumento en el ancho de distribución eritrocitaria (ADE o RDW).

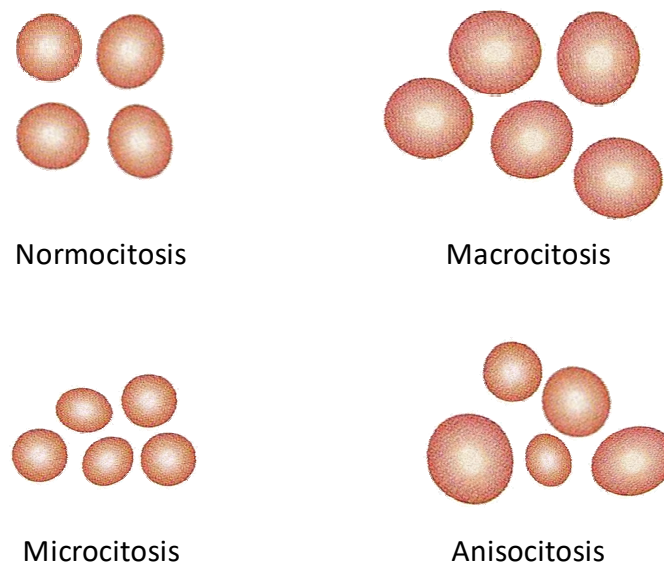


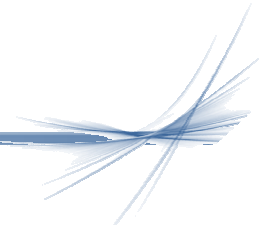
Figura 1. Alteraciones del tamaño de los hematíes.

Alteraciones en el color (ver Figura 2)

Hipocromía: Refleja una disminución en la concentración de hemoglobina en el interior del hematíe. La hipocromía se refleja en el frotis de sangre periférica como un aumento en la palidez central, ocupando más de un tercio del diámetro del hematíe. Generalmente la hipocromía refleja una deficiencia en la disponibilidad de hierro que se necesita para una eritropoyesis normal, como sucede en las anemias ferropénicas.

Hipercromía: Es un aumento en la concentración de hemoglobina en el interior del hematíe. La hipercromía se refleja en el frotis de sangre periférica como una disminución o ausencia en la palidez central del hematíe. La hipercromía es un hallazgo poco frecuente, normalmente relacionado con la presencia de esferocitos o de hematíes irregularmente contraídos.

Anisocromía: Supone un aumento de la variabilidad en el grado de hemoglobinización de los hematíes. En la práctica se utiliza para expresar la presencia de hematíes que presentan desde hipocromía a normocromía, diferenciándose claramente dos poblaciones eritrocitarias (dimorfismo eritrocitario). La anisocromía es característica de situaciones cambiantes en las que todavía coexisten poblaciones antiguas de hematíes que fueron originadas en un ambiente diferente a la población actual. La anisocromía puede encontrarse en pacientes con



anemia ferropénica que inician tratamiento con hierro, tras la transfusión, o en casos de inicio o resolución de la anemia asociada a enfermedades crónicas.

Policromasia: Se refiere a un aumento de hematíes inmaduros que muestran una tonalidad azulada en el frotis de sangre periférica. Estos hematíes corresponden a reticulocitos, cuyo ARN ribosomal es responsable de la tonalidad azulada. La policromasia indica una situación regenerativa, en un intento de la médula ósea por compensar una pérdida de la población eritrocitaria, como sucede en las anemias hemolíticas.

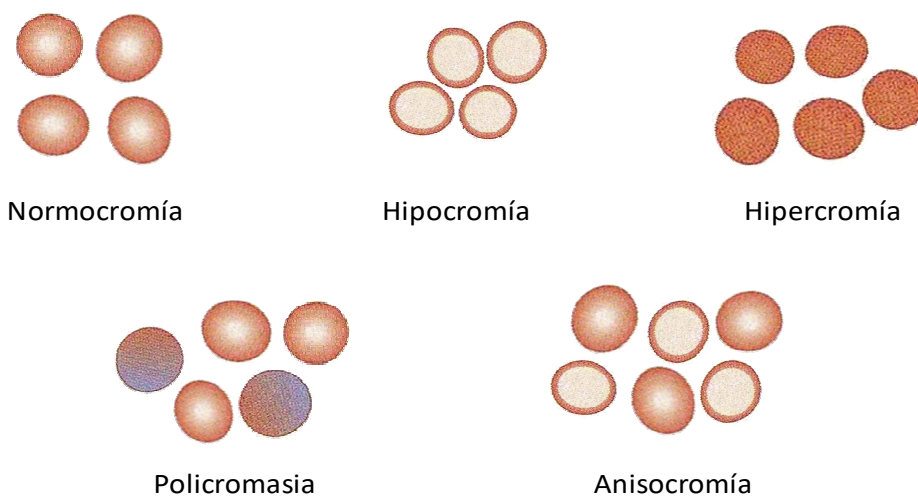


Figura 2. Alteraciones de la coloración de los hematíes.

Alteraciones de forma (ver Figura 3)

Poiquilocitosis: Consiste en una variabilidad en la forma de los hematíes. Se trata de un hallazgo común e inespecífico en el contexto de muchas enfermedades hematológicas. Es muy marcada en la mielofibrosis y en la piropoiquilocitosis hereditaria.

Esferocitos: Son hematíes que han adquirido una forma esférica o casi esférica, por lo que carecen de la palidez central. Aun con el mismo volumen, el diámetro de una esfera es menor que la forma normal de disco bicóncavo, por lo que los esferocitos pueden aparentar ser más pequeños que los hematíes normales. Los esferocitos pueden aparecer en la esferocitosis hereditaria, anemias hemolíticas autoinmunes y anemias hemolíticas microangiopáticas, entre otras.

Eliptocitos y ovalocitos: Se trata de hematíes con forma alargada. Se diferencian en que los eliptocitos tienen un eje mayor con una longitud de más del doble del de su eje menor, mientras que los ovalocitos tienen un eje mayor con una longitud de menos del doble la longitud del eje menor. Aparecen en gran proporción en la eliptocitosis y ovalocitosis hereditaria, pero también pueden presentarse en la anemia ferropénica, talasemias, síndromes mielodisplásicos y anemia megaloblástica.

Equinocitos: Son hematíes que presentan numerosas espículas cortas distribuidas regularmente por toda la superficie. En muchos casos se debe a un artefacto en muestras de sangre almacenada durante un tiempo prolongado con el anticoagulante EDTA. También se pueden ver equinocitos en el déficit de piruvatoquinasa y en pacientes con enfermedad renal.

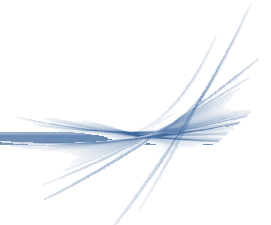
Acantocitos: Son hematíes que presentan escasas espículas más alargadas distribuidas irregularmente por la superficie. Se observan en las neuroacantocitosis, tras la esplenectomía y en la *spur cell anemia* (anemia hemolítica con acantocitos asociada a hepatopatía grave).

Queratocitos: Hematíes que presentan dos proyecciones laterales que confieren al hematíe forma de casco. Este defecto puede ser la consecuencia de un daño oxidativo (déficit de glucosa 6 fosfato deshidrogenasa) o mecánico (anemias hemolíticas microangiopáticas), o tras la eliminación de un cuerpo de Heinz por el bazo.

Excentrocitos: Son hematíes en los que la hemoglobina queda polarizada en una parte de la célula. Este defecto puede ser la consecuencia de un daño oxidativo (déficit de glucosa 6 fosfato deshidrogenasa) o un daño en la membrana por precipitación de hemoglobinas inestables.

Esquistocitos: Son fragmentos de hematíes, que se forman habitualmente por fragmentación mecánica. Son frecuentes en las anemias hemolíticas microangiopáticas, como la púrpura trombótica trombocitopénica. También se observan en la piropoiquilocitosis hereditaria y en pacientes con prótesis valvulares.

Dianocitos: Se forman por un exceso de membrana en relación al volumen de citoplasma que le confiere forma de diana en el frotis de sangre periférica. Los dianocitos se observan en ictericia obstructiva, hepatopatías graves, en talasemias y hemoglobinopatías.



Dacriocitos: Son hematíes que presentan una prolongación distal que les confiere forma de lágrima. Son frecuentes en la anemia megaloblástica, en la mielofibrosis o en la infiltración medular metastásica.

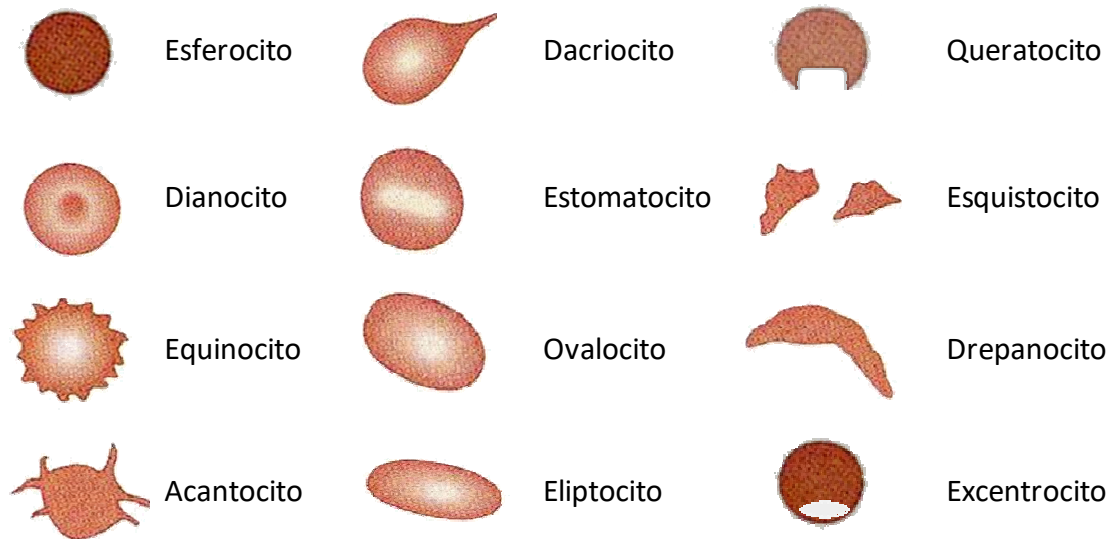


Figura 3. Alteraciones de forma de los hematíes.

Estomatocitos: Hematíes en los que la claridad central tiene forma alargada o de boca, aunque *in vivo* tienen forma de cuenco. Los estomatocitos son frecuentes en personas con hábito enólico y también en la estomatocitosis hereditaria.

Drepanocitos: Son hematíes muy largos y estrechos y con forma de hoz. También son denominados hematíes falciformes. Se deben a la formación de polímeros de hemoglobina S cuando esta se desoxigena. Característico de la enfermedad falciforme o drepanocitosis.

Presencia de inclusiones (ver Figura 4)

Punteado basófilo (PB): Son hematíes que presentan gránulos basófilos distribuidos regularmente por toda la célula. Constituyen precipitados de ribosomas y ARN ribosómico. Puede observarse un punteado basófilo tenue en los reticulocitos. Es grueso en el déficit de la enzima eritrocitaria pirimidina 5' nucleotidasa. La inhibición de esta enzima podría ser responsable también del punteado basófilo grueso que se observa en el saturnismo

(intoxicación por plomo). En condiciones normales aparece asociado a la presencia de reticulocitosis y muestra un estado regenerativo de la eritropoyesis. Su asociación con una distribución anómala de la hemoglobina es sugestiva de anemia sideroblástica.

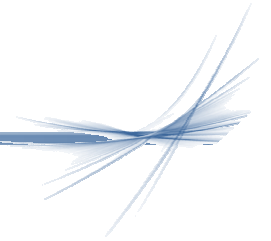
Cuerpos de Howell-Jolly (HJ): Son inclusiones redondeadas, habitualmente únicas por cada hematíe, que se localizan en la periferia, y se tiñen de color violeta intenso. Se trata de restos nucleares que en condiciones normales son eliminados de los hematíes por el bazo. Se observan tras la esplenectomía y en estados de hipoesplenismo.

Cuerpos de Pappenheimer (PP): Inclusiones basófilas, redondeadas, de tamaño intermedio entre los HJ y el PB, que habitualmente se presentan en pequeños grupos en la periferia de los hematíes. Se trata de agregados de hemosiderina que contienen hierro. Pueden observarse tras la esplenectomía, en situaciones de gran sobrecarga férrica y en las anemias sideroblásticas y diseritropoyéticas congénitas.

Parásitos: En la revisión del frotis de sangre periférica puede observarse la presencia de parásitos intraeritrocitarios, como en el caso de la malaria o la babesiosis. Estos parásitos invaden los hematíes, alojándose en su interior y apareciendo como estructuras intracelulares. En el interior de los hematíes, los trofozoitos jóvenes adoptan inicialmente una forma anular, con un citoplasma azulado en la que puede observarse uno o dos núcleos de color violáceo. A medida que la enfermedad avanza es posible observar una evolución morfológica del parásito hacia formas más irregulares o ameboides. La detección de parásitos intraeritrocitarios es crucial para el tratamiento y pronóstico de estas enfermedades.



Figura 4. Presencia de inclusiones en el interior de los hematíes



Alteraciones en la distribución (ver Figura 5)

Rouleaux: Es el fenómeno de apilamiento de eritrocitos de forma ordenada que recuerda a una disposición como en pilas de monedas. Cuando está presente, se correlaciona con un aumento de la velocidad de sedimentación globular. Existe una relación entre la presencia de pilas de monedas y un aumento en la concentración de proteínas en sangre que hacen que los hematíes se adhieran unos a otros. Puede observarse, por tanto, en las gammapatías monoclonales o en los procesos inflamatorios crónicos.

Crioaglutininas: Es un fenómeno de agrupación de hematíes de forma irregular que recuerda a una disposición en racimos de uvas. Esta agrupación característica de los hematíes está causada por anticuerpos fríos o crioaglutininas, que son inmunoglobulinas de tipo IgM que se unen a los hematíes a temperaturas más bajas que la temperatura corporal. Estas inmunoglobulinas son capaces de fijar el complemento y producir cuadros de anemia hemolítica.

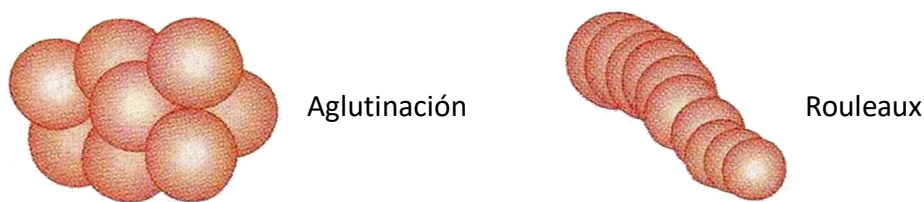


Figura 5. Alteraciones en la distribución de los hematíes.

Las alteraciones morfológicas de la serie roja que se detectan en la revisión manual del frotis de sangre periférica pueden suponer una importante ayuda al diagnóstico clínico, ya que no solo permiten confirmar los resultados anómalos que aparecen en el hemograma, además permiten detectar alteraciones que los analizadores de hematología actuales no son capaces de identificar y que tienen un importantísimo valor para establecer una orientación diagnóstica presuntiva. En muchos casos también permite orientar acerca de las pruebas adicionales que serían necesarias realizar para poder confirmar una sospecha diagnóstica. Esto adquiere una gran relevancia en aquellos casos en los que la orientación diagnóstica pone de manifiesto una patología en la que la toma de decisiones clínicas es crítica para el pronóstico

del paciente. En el ámbito del laboratorio clínico, los resultados que requieren una urgencia en su comunicación son los conocidos como resultados o valores críticos, y en muchos casos esta notificación puede ser determinante para el tratamiento inmediato y la evolución del paciente.

Con respecto a la morfología eritrocitaria, actualmente solo hay dos tipos de alteraciones que son consideradas como valores críticos. Según recomendaciones del comité internacional de estandarización en hematología (ICSH), los hallazgos morfológicos que se consideran como valores de urgente notificación son la presencia de esquistocitos en un frotis sugestivo de microangiopatía trombótica y la presencia de parásitos intraeritrocitarios como los de la malaria (4). A diferencia de otras alteraciones morfológicas, actualmente la malaria constituye junto a la leucemia aguda los dos hallazgos morfológicos considerados como valor crítico en la mayoría de los laboratorios clínicos.

2.3 La malaria

El paludismo es una enfermedad febril aguda causada por parásitos del género *Plasmodium*, que se transmiten a las personas por la picadura de hembras de mosquito infectadas del género *Anopheles*. Hay cinco especies de parásitos que causan paludismo en el ser humano, de las que las más peligrosas son dos: *P. falciparum* y *P. vivax*. El primero es el parásito palúdico más mortífero y el más prevalente en el continente africano, siendo el causante del 90% de las infecciones por malaria (5). Si bien la malaria es poco común en climas templados, sigue siendo común en países tropicales y subtropicales. Según el último Informe mundial sobre el paludismo, en 2020 hubo 241 millones de casos de paludismo, y fue el causante de 627.000 muertes. Algunos grupos de población corren un riesgo considerablemente mayor que otros de contraer la enfermedad y presentar un cuadro clínico grave: los lactantes, los menores de 5 años, las embarazadas y los pacientes con VIH/sida, así como las personas con baja inmunidad que se desplazan a zonas de intensa transmisión palúdica, como puedan ser trabajadores migrantes, viajeros y poblaciones itinerantes. La malaria continúa siendo un grave problema de salud pública en muchas regiones, especialmente en los países de África subsahariana.

Los parásitos de malaria presentan un ciclo biológico (ilustrado en la Figura 6) que involucra a dos hospedadores. El ciclo del parásito se inicia en el mosquito hembra del género *Anopheles*, donde tiene lugar la reproducción sexual del parásito, produciendo formas infectivas o esporozoitos que se quedan almacenadas en las glándulas salivales del mosquito. A través de

la picadura, los mosquitos inoculan los esporozoitos al hospedador humano. Los esporozoitos infectan las células del parénquima hepático, donde se dividen y multiplican dando lugar a merozoitos. Cuando la célula hepática se rompe, los parásitos pasan a la sangre e infectan los eritrocitos produciendo ciclos de multiplicación asexual. Algunos parásitos se pueden diferenciar en el estadio sexual (gametocitos) que son las formas infectivas de nuevos mosquitos, cerrando el ciclo de vida del parásito. *P. vivax* y *P. ovale* pueden persistir en el hígado en un estado de latencia (hipnozoitos) y causar recaídas al invadir el torrente sanguíneo por semanas e incluso años después de la infección (6).

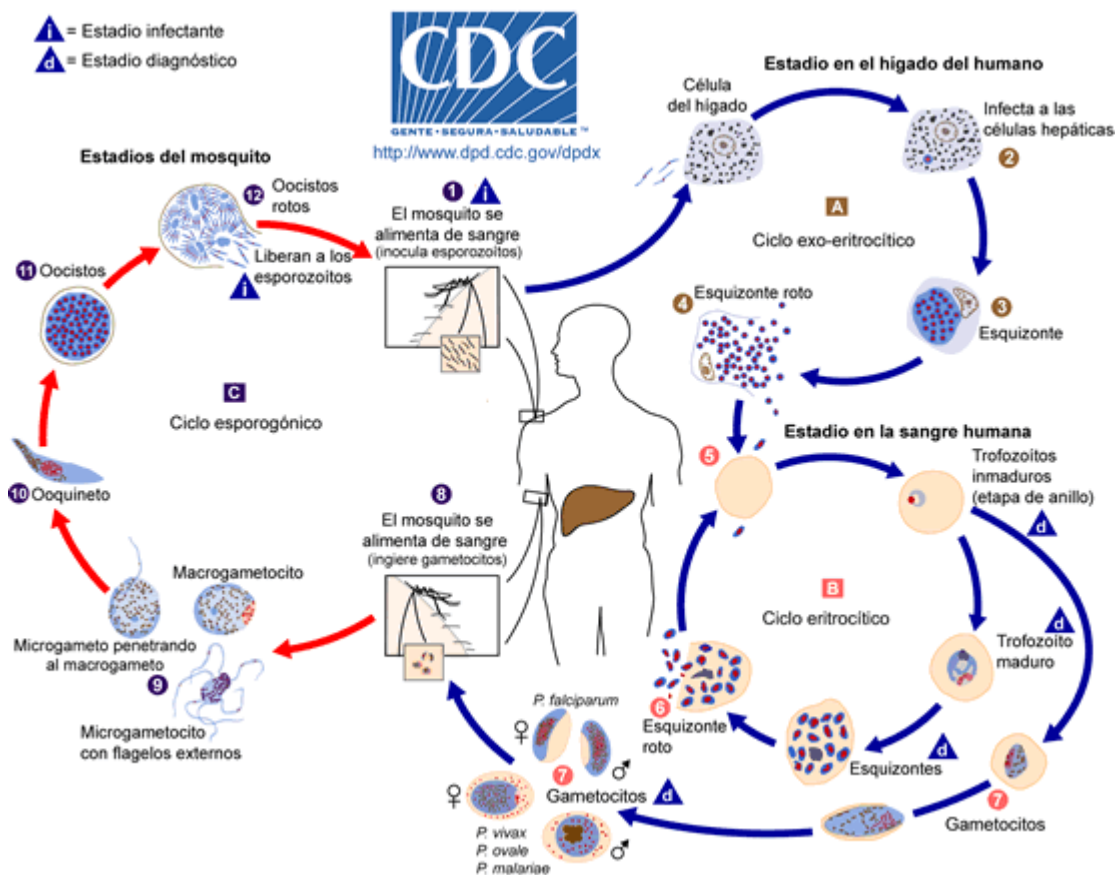
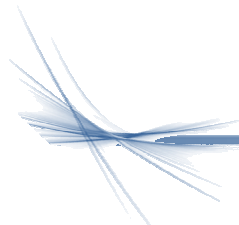


Figura 6. Ciclo de vida de la malaria (fuente: www.cdc.gov/dpdx/malaria/index.html).

Los estadios parasitarios sanguíneos y los ciclos de replicación son los responsables de las manifestaciones clínicas de la enfermedad. Desde la picadura del mosquito y la adquisición del parásito hasta el comienzo de los síntomas existe un periodo de incubación, que dura entre 9 y 30 días. La manifestación más característica de las distintas formas de paludismo es la aparición de episodios de fiebre, precedidos de escalofríos intensos, que ceden con sudoración

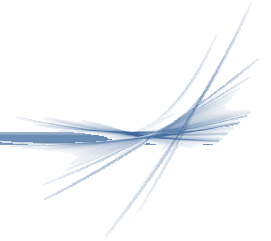


muy abundante, dando paso a una fase de relajación y baja temperatura. Estos ciclos se repiten con una cadencia distinta dependiendo del ciclo vital de la especie infectante, produciendo la llamada fiebre terciana (cada tres días) en las infecciones por *P. falciparum*, *P. viva* o *P. ovale*; o produciendo la fiebre cuartana (cada cuatro días) en las infecciones por *P. malariae*. Estos síntomas son más o menos intensos según la especie de plasmodio que cause la infección, y puede adoptar formas crónicas en los adultos de regiones endémicas, casi sin fiebre, que constituyen un síndrome de esplenomegalia tropical.

La malaria es una enfermedad potencialmente mortal, y diagnosticar a tiempo una malaria puede ser vital para el enfermo, ya que la aparición de complicaciones está muy relacionada con la demora en la instauración del tratamiento. Por lo que en el diagnóstico de la malaria es altamente recomendable disponer del resultado de forma urgente, menos de 3h tras la extracción (7). El retraso en el diagnóstico no debe demorar el inicio de tratamiento antimalárico empírico si la probabilidad de malaria es alta y el paciente está grave. Las pruebas de diagnóstico de malaria se han de solicitar a cualquier paciente con fiebre procedente de un área endémica.

Desde que se describiera por primera vez en 1880, el diagnóstico de esta enfermedad se ha realizado mediante la observación de las distintas formas del parásito en el examen microscópico de extensiones de sangre periférica teñidas con diversos colorantes. Hoy día, 120 años después, esta técnica sigue siendo el método de referencia. Para la observación microscópica del parásito generalmente se utilizan dos tipos de frotis (gota gruesa y extensión) según el objetivo del análisis, según se ilustra en la Figura 7.

La gota gruesa consiste en una gota de sangre en un portaobjetos de vidrio en la que se lisan los hematíes previa a la tinción, mientras que una extensión es el resultado de esparcir la gota por un portaobjetos, realizándose una tinción mientras se mantiene la integridad de la morfología eritrocitaria. Las gotas gruesas son más sensibles y teóricamente permiten la visualización de parásitos incluso en casos con niveles bajos de parasitemia. Por lo tanto, son más útiles para detectar la presencia de parásitos. Sin embargo, solo se realiza en los casos en los que existe una sospecha previa de infección por malaria. Las extensiones permiten la visualización de detalles morfológicos que son importantes para el diagnóstico y la instauración del tratamiento. Además, las extensiones de sangre periférica suelen complementar los resultados aportados por el hemograma completo, una de las pruebas de



laboratorio más solicitadas. No es infrecuente la observación de parásitos de la malaria en una extensión como hallazgo casual en los laboratorios clínicos de urgencias.

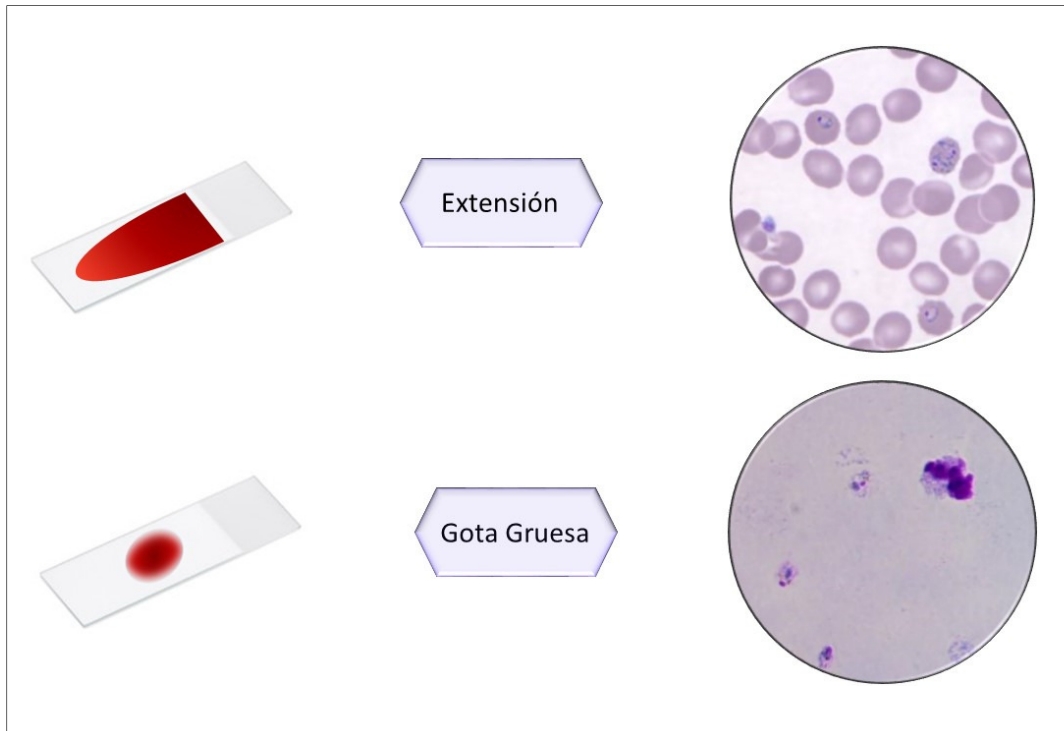


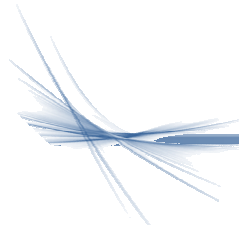
Figura 7. Aspecto macro y microscópico de una extensión y una gota gruesa de sangre periférica.



CAPÍTULO 3

Estado del arte:

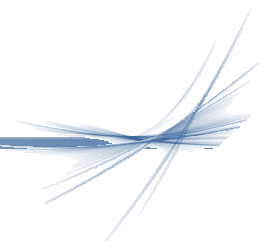
Morfología eritrocitaria e inteligencia artificial



El análisis morfológico del frotis de sangre periférica es esencial para detectar anomalías en las células sanguíneas. Sin embargo, el análisis visual adolece de ciertos inconvenientes: requiere mucho tiempo para valorar correctamente las características cuantitativas y cualitativas del frotis, también requiere un conocimiento profundo y necesita profesionales expertos para revisar los frotis de una manera objetiva y fiable. Aun disponiendo de una amplia experiencia en este campo, la interpretación de un frotis de sangre periférica es propensa a la variabilidad entre diferentes observadores (8). Los resultados publicados por los organizadores de programas externos de evaluación de la calidad muestran que la presencia de malaria en un frotis de sangre periférica constituye un hallazgo que frecuentemente no es detectado por algunos laboratorios (9), aun incluso cuando el examen del frotis es una herramienta crucial en el diagnóstico de infecciones por *Plasmodium*. Las desventajas asociadas con el análisis morfológico ya mencionadas han contribuido al desarrollo de técnicas basadas en el procesamiento de imágenes digitales que sirven de apoyo al laboratorio clínico, con la finalidad de preclasificar células sanguíneas, como leucocitos y glóbulos rojos. Estas herramientas pueden suponer la oportunidad de llevar la microscopía a una forma más rápida, más eficiente y estandarizada de realizar un análisis morfológico (10,11).

El contenido del presente capítulo analiza las principales aportaciones bibliográficas respecto a las metodologías con mayor repercusión en el análisis digital de imágenes, como la detección automática de hematíes dentro de una imagen de sangre periférica (ver Apartado 3.1), la extracción de datos cuantitativos que describan estas imágenes de hematíes (ver Apartado 3.2), y el análisis de estos descriptores con la finalidad de realizar un reconocimiento automático de las imágenes mediante algoritmos de clasificación basados en modelos de *machine learning* (ver Apartado 3.3). Se realizará también una revisión de los principales modelos publicados que utilizan herramientas de aprendizaje profundo (*deep learning*) en el reconocimiento de la malaria (ver Apartado 3.4). Por último, se revisará el papel de los principales analizadores automatizados a nivel comercial desde el punto de vista del reconocimiento automático de la malaria (ver Apartado 3.5).

Como se podrá observar posteriormente, estos modelos de clasificación automática mediante modelos de *machine learning* actualmente publicados presentan un notable número de falsos positivos debido a la presencia de otros tipos de inclusiones eritrocitarias que son erróneamente clasificados como parásitos de malaria. Ello ha motivado el desarrollo de esta tesis doctoral de modelos basados en *machine learning* y *deep learning* para el reconocimiento



automático de hematíes parasitados por malaria y capaces de discriminarlos de hematíes con otro tipo de inclusiones.

3.1 Segmentación de hematíes

Uno de los primeros problemas que aparece a la hora de trabajar con imágenes digitales de hematíes está relacionado con la obtención de las imágenes. Cualquier sistema de microscopía analógico o digital a partir del cual se desee obtener un elemento de interés, dará lugar a una imagen en la que se observará la presencia de diferentes y numerosos elementos, generalmente con una mayoritaria presencia de hematíes, pero donde también pueden aparecer diferentes elementos leucocitarios y plaquetas. A estas imágenes se las denomina “imágenes de campo” (ver Figura 8).

Por lo tanto, si se desea trabajar con alguno de los elementos que aparecen en un frotis de sangre periférica es necesario aplicar algoritmos de análisis digital de imágenes para obtener de una manera automatizada imágenes individuales de cada elemento. De esta manera se podrá estudiar y extraer información de las regiones de interés (ROI) sin que el análisis se vea interferido por la presencia de otros elementos no deseados.

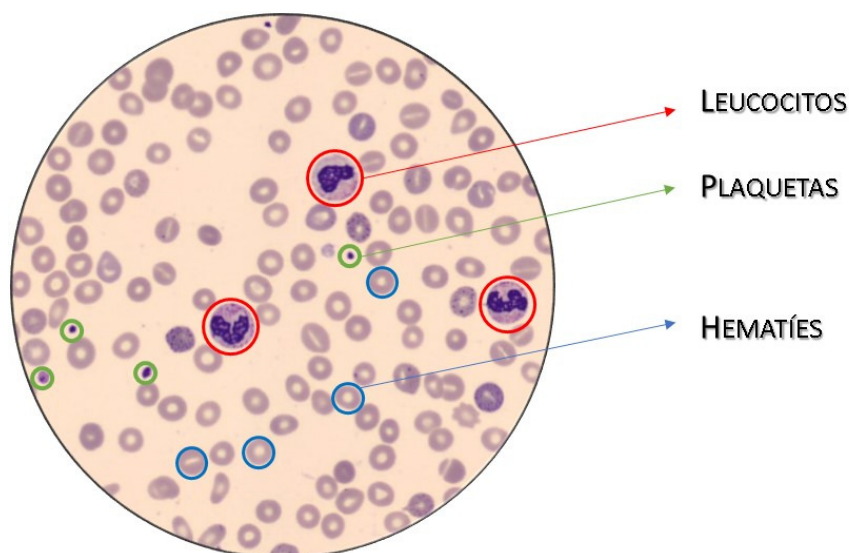
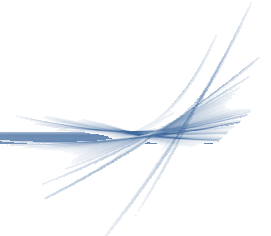


Figura 8. Imagen de campo de un frotis de sangre periférica.

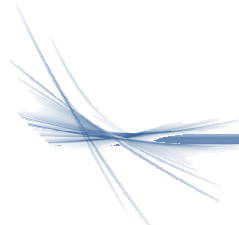
La literatura muestra una multitud de métodos diferentes aplicados a la segmentación de los hematíes a partir de imágenes de sangre periférica. Generalmente, los algoritmos de segmentación de los hematíes suelen ser más sencillos en comparación con los de otras células de la sangre como los leucocitos, en los que para hacer una correcta valoración a través de análisis de imagen es necesario tener en cuenta los diferentes componentes celulares. La segmentación por separado del núcleo, el citoplasma, e incluso la región colindante del leucocito será de gran utilidad para un correcto reconocimiento de las diferentes subpoblaciones leucocitarias (12). Sin embargo, la morfología de los hematíes es más sencilla, lo que facilita el desarrollo de algoritmos para su detección. Halim *et al* (13) propuso una estrategia de emparejamiento con plantillas (*template matching*) para la detección de eritrocitos. En este trabajo construyeron plantillas binarias predefinidas asumiendo modelos elípticos con una excentricidad variable como el modelo que mejor se adapta a la morfología del hematíe. Las imágenes reales en escala de grises eran comparadas con las plantillas diseñadas por medio de correlaciones cruzadas que describen el grado de emparejamiento. Los hematíes detectados también fueron utilizados para la creación de nuevas plantillas en las sucesivas tareas de segmentación. Este método mostró una precisión en el reconocimiento de hematíes del 85-89%. Tek *et al* utilizaron el análisis por granulometría en procesos locales para proporcionar independencia de escala de las diferentes imágenes (14). Se basaron en la utilización de un umbral para el valor del pico de la función de granulometría en un área determinada como la mejor estimación para el área del hematíe promedio, basándose en el hecho de que los hematíes normales tienen límites de tamaño esperados. Frejlichowski *et al* usaron un enfoque en el que realizaron dos transformadas para una imagen de la célula (15). En primer lugar, utilizaron la transformación de coordenadas cartesianas a polares, para posteriormente usar la transformada en dos dimensiones de Fourier a partir la imagen polar obtenida. Finalmente, el subespectro obtenido se extrajo con la finalidad de representar el hematíe y emparejarlo con plantillas, usando la medida de la disimilitud. Veluchami *et al* utilizaron algoritmos de detección de contornos para la segmentación de hematíes del fondo de la imagen (16). Los algoritmos de detección de bordes usan información de gradiente seguida de operaciones morfológicas de cierre de los contornos detectados. Otros autores argumentaron que el uso de algoritmos de detección de bordes no siempre facilita el reconocimiento de objetos, especialmente en presencia de una cantidad sustancial de ruido de fondo o de suciedad en las imágenes de frotis de sangre periférica. Con la finalidad de solventar este inconveniente, propusieron utilizar la minimización de energía de la imagen



para una detección de hematíes más precisa dentro de una imagen de campo (17). Gual Arnau *et al* utilizaron modelos de contornos activos para la detección de hematíes, enfoque que garantiza además cierta flexibilidad en la topología de los contornos (18). Otros autores usaron métodos de detección de hematíes a través del uso de una versión modificada de la transformada circular de Hough para detectar círculos de un tamaño determinado (19,20). El propósito de esta técnica es encontrar instancias imperfectas de objetos con cierta flexibilidad dentro de una clase de formas mediante un procedimiento de votación.

Uno de los procedimientos más utilizados con la finalidad de segmentar imágenes de hematíes a partir de una imagen de campo se basa en la umbralización. Esta metodología busca seleccionar un grupo de píxeles dentro de una imagen en función de su escala de valores de intensidad en un espacio de color determinado. Dentro del espacio de color RGB (rojo, verde y azul), las imágenes de hematíes muestran un mayor contraste respecto al fondo de la imagen en la representación del componente verde (21). Esto permite conseguir un buen rendimiento en la segmentación utilizando métodos de umbralización de sencilla implementación, siendo el método de Otsu uno de los más utilizados en la literatura con diferencia (21-24).

Uno de los principales problemas que aparece en la segmentación de los hematíes es que en numerosas ocasiones aparecen dos o más hematíes superpuestos. Este hecho es especialmente llamativo si se obtienen imágenes de una zona interior del frotis de sangre periférica, donde las células tienden a aglomerarse. Sin embargo, incluso cuando las imágenes son obtenidas en una zona óptima del frotis, es bastante común que aparezcan hematíes superpuestos. Esto puede plantear problemas para la correcta detección de los hematíes, ya que los métodos de segmentación basados en umbrales de los histogramas de color, reconocerán los hematíes superpuestos como un único elemento, mientras que los métodos basados en la presunción de tamaño y o circularidad, podrán pasar estos elementos desapercibidos. Algunas estrategias han intentado lidiar con este problema. Chen *et al* propusieron un método de separación de hematíes superpuestos basado en la técnica de códigos de cadena de ocho conexiones (22). Este algoritmo analiza la dirección del contorno de los hematíes para calcular los puntos de alta concavidad en dos hematíes superpuestos, y utilizar esta información para separar los hematíes adyacentes. Díaz *et al* utilizaron un método de separación de hematíes superpuestos a partir del emparejamiento con plantillas de hematíes modelo (25). La idea básica es que una plantilla segmente los hematíes superpuestos encontrando la mejor correlación entre una forma agrupada y la plantilla, que es un eritrocito

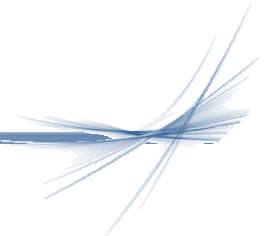


"ideal" construido a partir de la imagen de un conjunto de eritrocitos aislados provenientes del mismo frotis.

Sin embargo, el método más utilizado para lidiar con la separación de hematíes superpuestos es la utilización de la transformada de Watershed. Este algoritmo se fundamenta en que cualquier imagen en escala de grises se puede ver como una superficie topográfica donde una intensidad alta indica picos y colinas, mientras que intensidades bajas indican valles. En este algoritmo se empieza por llenar cada valle aislado (que corresponde a los mínimos locales) con agua de diferentes colores que posteriormente originarán las etiquetas. A medida que el agua sube, dependiendo de los picos (pendientes) cercanos, el agua de diferentes valles, obviamente con diferentes colores, comenzará a fusionarse. Para evitar esto, se construyen barreras en los lugares donde se une el agua. Luego, se continúa el trabajo de rellenar con agua y construir barreras hasta que todos los picos estén bajo el agua. Así, las barreras creadas en este proceso, no son más que la segmentación de la imagen. Aunque el método de Watershed tiene el inconveniente de que puede producir una sobresegmentación debido al ruido y a irregularidades de la imagen, el uso de marcadores dentro de los objetos de interés ayuda a obtener un buen rendimiento en la segmentación. Este hecho ha dado lugar a que en la literatura sea uno de los métodos más comúnmente empleados en la segmentación y separación de los hematíes (26-30).

3.2 Extracción de características

Para desarrollar un algoritmo de clasificación en las herramientas clásicas de *machine learning*, es necesario alimentar un modelo con datos cuantitativos. Por lo tanto, una de las fases principales del desarrollo consiste en extraer características medibles partiendo de una imagen. Estas características son conocidas como descriptores, de los que existen una variada representación. Los descriptores más sencillos son los de forma o geométricos, que describen información acerca del tamaño y la forma de las ROI. Estos tipos de descriptores han sido utilizados preferentemente para el desarrollo de algoritmos de clasificación de las principales alteraciones eritrocitarias (16, 24,26,31,32,33), ya que las alteraciones de forma suponen una de las características más importantes en el diagnóstico diferencial de una gran variedad de causas de anemia. Algunas de estas características morfológicas también han sido utilizadas como entrada a algoritmos publicados para la discriminación entre hematíes infectados con



malaria y hematíes no infectados (31). Sin embargo, en la mayoría de los trabajos se opta por escoger los descriptores de color y textura como los más representativos en la detección de la malaria. Este tipo de descriptores han sido propuestos como aquellos que tienen mayor poder discriminante y mayor relevancia en la diferenciación entre hematíes infectados y no infectados. Los principales descriptores de color utilizados han sido los estadísticos de primer orden calculados a partir de histogramas de espacios de color concretos (25,34,35). Los descriptores de textura más utilizados se basan en el análisis de la entropía, patrones locales binarios (LBP), descriptores de Tamura y descriptores de Haralick, basado en el cálculo de matrices de coocurrencia (35-37).

Un enfoque muy interesante utilizado por algunos autores es el uso de descriptores que pretenden representar características morfológicas que se utilizan visualmente para el reconocimiento de parásitos de malaria en la revisión manual del frotis de sangre periférica. La idea consiste en intentar parametrizar en forma de datos cuantitativos los aspectos visuales que un observador normalmente utiliza para reconocer un hematíe normal de uno infectado, e incluso entre las diferentes especies de *Plasmodium* que son capaces de infectar hematíes humanos. En un estudio llevado a cabo por Ross *et al* se abordó esta estrategia para calcular descriptores morfológicos como el tamaño relativo del hematíe infectado, su excentricidad, la suavidad del contorno celular, la textura del citoplasma de los hematíes infectados, el número de parásitos por hematíe o el número de puntos de cromatina de cada parásito (38). La intención es intentar diferenciar entre las especies de *P. falciparum*, *P. vivax*, *P. ovale* y *P. malariae*. El tamaño del hematíe infectado sirve para diferenciar *P. falciparum* y *P. malariae* de *P. vivax* o *P. ovale*, ya que estos dos últimos tienen más afinidad por infectar hematíes jóvenes que suelen tener un mayor tamaño que los hematíes maduros. La textura del citoplasma de los hematíes infectados también puede servir para identificar *P. vivax* o *P. ovale*, ya que los hematíes infectados por ambas especies suelen mostrar una granulación muy similar morfológicamente al PB conocida como granulación de Schüffner. El número de parásitos por hematíe o el número de puntos de cromatina de cada parásito son descriptores muy útiles para la diferenciación de *P. falciparum* del resto de especies. En un estudio similar llevado a cabo por Sheikhosseini *et al*, también se estudió el uso de este tipo de descriptores morfológicos del propio parásito, como su área, excentricidad, elongación o características geométricas del núcleo y del citoplasma del parásito y como se relacionan entre sí (39). Este análisis de las características morfológicas del parásito lo utilizaron con la finalidad de distinguir entre los diferentes estadios parasitarios: trofozoito, esquizonte o gametocito. En

resumen, el análisis de todas estas características permite el desarrollo de modelos de reconocimiento automático de malaria que tienen en cuenta características citológicas de las imágenes, lo que muestra un intento de transmitir la forma en la que observamos el mundo a los algoritmos desarrollados.

3.3 Métodos de clasificación

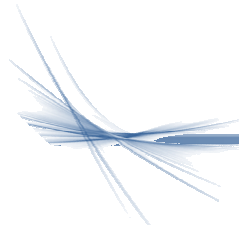
Los datos obtenidos en la extracción de descriptores nos permiten describir cuantitativamente una imagen. Los datos que representan la imagen sirven como elementos de entrada a los algoritmos de clasificación. El desarrollo de estos algoritmos puede realizarse mediante procedimientos de aprendizaje supervisado o no supervisado. La diferencia entre estas dos metodologías radica, desde un punto de vista general, en la existencia de etiquetas en el conjunto de datos de entrenamiento o *training* (40). La mayoría de estudios aplicados al reconocimiento automático de malaria utilizan técnicas de aprendizaje supervisado, donde los algoritmos intentan predecir y clasificar la etiqueta predeterminada, y la exactitud en el reconocimiento conjuntamente con otras medidas de desempeño depende de la cantidad de etiquetas predeterminadas correctamente clasificadas o no.

Un enfoque muy utilizado en el entrenamiento supervisado es el uso de técnicas de umbralización como métodos de clasificación. Ma *et al* utilizaron un enfoque de umbralización a partir del promedio de los valores de píxeles locales que están contenidos dentro de los hematíes (20). Desarrollaron un software que realiza un recuento de la parasitemia (porcentaje de hematíes infectados por malaria en una muestra de sangre periférica), con un ajuste respecto al conteo manual que mostró un coeficiente de determinación (R^2) del 0,995 en muestras con un rango de parasitemias del 0,1% al 58%. Halim *et al* utilizaron un umbral definido a partir de una matriz de coocurrencia de color en la que se analiza la intensidad de los píxeles locales (13). Desarrollaron un modelo que predecía la presencia de gametocitos y esquizontes de malaria con una sensibilidad del 89% y un valor predictivo positivo (VPP) del 94%. Prasad *et al* publicaron un algoritmo mediante técnicas de umbralización para el reconocimiento de malaria a partir de los siguientes descriptores: el porcentaje de píxeles teñidos como puntos de cromatina, el porcentaje de píxeles teñidos de la parte del citoplasma y la desviación estándar del canal V del espacio de color HSV (matiz, saturación y valor) de cada ROI (41). Su modelo mostró un valor de R de *Spearman* del 0,888 respecto al conteo manual.



Otros trabajos también se basaron en el uso de umbrales a partir de los valores de intensidad de color de los píxeles con una buena tasa de reconocimiento de malaria (42-45). Das *et al* realizaron un estudio en el que compararon el rendimiento en reconocimiento automático de malaria en dos clasificadores diferentes, uno basado en un clasificador de Bayes y otro basado en máquinas de soporte vectorial (SVM) (35). Obtuvieron una exactitud en la clasificación del 84% con un clasificador de Bayes que utilizaba 19 descriptores y del 83,5% con un clasificador de SVM con solo 9 descriptores. Ghosh *et al* realizaron un trabajo similar en el que compararon el rendimiento de un clasificador de Bayes y un SVM con 4 descriptores de textura en el que obtuvieron unas exactitudes del 98% y 95%, respectivamente, en el reconocimiento de los diferentes estadios parasitarios de muestras infectadas con *Plasmodium vivax* (46). Tek *et al* emplearon un clasificador basado en los K vecinos más cercanos (KNN) en el reconocimiento automático de malaria a partir de 4 descriptores de color, forma y textura (47), obteniendo una sensibilidad y especificidad del 74% y 98%, respectivamente. Los mismos autores en un trabajo posterior compararon un clasificador basado en KNN y otro en análisis discriminante lineal (LDA) a partir de 83 descriptores (14). El reconocimiento de malaria proporcionado por ambos clasificadores se midió con exactitudes del 93% y 90% para KNN y LDA, respectivamente. Diaz *et al* también realizaron una comparación entre clasificadores basados en KNN, SVM (lineal y radial) y un clasificador de Bayes con descriptores extraídos en cuatro espacios de color diferentes: RGB, RGB normalizado, HSV y YCbCr (luma, crominancia de azul y crominancia de rojo) (48). En este estudio se obtuvieron los mejores rendimientos de clasificación con los clasificadores KNN y SVM lineal, con valores de F score para el reconocimiento de parásitos de 0,83 para el clasificador KNN que utiliza los descriptores en el espacio de color RGB normalizado, y un valor de F score de 0,81 para el clasificador SVM lineal que utiliza los descriptores extraídos en el espacio de color YCbCr. Park *et al* aplicaron diferentes técnicas de aprendizaje automático, que incluían LDA, la regresión logística (LR) y KNN, para formular algoritmos que combinan 23 descriptores diferentes (49). Los resultados mostraron que LDA proporcionaba la mayor exactitud de hasta un 99,7% en la detección de células infectadas en etapa de esquizonte en comparación con hematíes no infectados. KNN mostró una exactitud ligeramente mejor (99,5%) que LDA (99,0%) o LR (99,1%) para discriminar los trofozoítos tardíos de los hematíes no infectados. Sin embargo, para los trofozoítos jóvenes, LDA mostró también los mejores valores de exactitud (98%).

Uno de los algoritmos más utilizados en el desarrollo de modelos de reconocimiento automático son las SVM. Savkare *et al* desarrollaron un modelo basado en SVM lineal



mediante un entrenamiento con descriptores morfológicos y de color (50). Consiguieron una sensibilidad y una especificidad del 93% en el reconocimiento de dos clases: hematíes infectados y hematíes no infectados. En un trabajo posterior, el mismo grupo a través del desarrollo de un modelo de similares características, con el mismo tipo de descriptores y un algoritmo basado en SVM de función radial, consiguieron un 96% de exactitud en el reconocimiento de diferentes estadios parasitarios (51). Linder *et al* elaboraron una interfaz de clasificación de malaria basado en SVM mediante el uso de descriptores como LBP, la varianza de la textura local de la imagen (VAR) y descriptores de transformación de características invariantes de escala (SIFT) a partir de imágenes digitalizadas de frotis de sangre periférica escaneados y (37). Su modelo mostró una sensibilidad y especificidad en la clasificación del 85% y 100%, así como un coeficiente de correlación de 0,97 respecto a un recuento de parasitemia realizado mediante observación manual. Eshel *et al* evaluaron un sistema comercial automatizado de reconocimiento de malaria a partir del examen del frotis de sangre periférica (52). Este dispositivo realiza una tinción de ADN y ARN presente en la muestra. Las imágenes de estas muestras teñidas son reconocidas por un modelo de SVM para la detección y recuento de parásitos de malaria con una sensibilidad y especificidad del 99%. Yang *et al* desarrollaron un sistema portable de reconocimiento de malaria con una tasa de error inferior al 0,25% a partir de modelos de SVM que utilizaba imágenes de frotis obtenidos a partir de cultivos de malaria (53).

Otro de los modelos más frecuentemente utilizados en el reconocimiento automatizado de malaria son los algoritmos basados en árboles de decisión. Sheikshoeini *et al* desarrollaron un modelo a partir de reglas de decisión aplicadas sobre descriptores morfológicos (39). Estos descriptores resultaron útiles cuando la segmentación consigue distinguir el núcleo y el citoplasma del parásito por separado. Su modelo demostró una sensibilidad y una especificidad del 82% y 98 % respectivamente. Pamungkas *et al* diseñaron un sistema capaz de identificar la fase de desarrollo de *P. falciparum* en los hematíes infectados con una clasificación basada en árboles de decisión a partir de dos descriptores extraídos en el espacio de color HSV: la ratio entre el número de píxeles de la región de interés y el tamaño de la imagen, y la excentricidad de la región de interés (54). Para ello tuvieron que realizar una segmentación específica de los parásitos sin tener en cuenta la presencia de hematíes. Su modelo obtuvo una exactitud en la clasificación de los diferentes estadios parasitarios del 83%. Soni *et al* publicaron un sistema de clasificación basado en arboles de decisión en el que a través de descriptores de color y textura se diferenciaba entre hematíes infectados y no



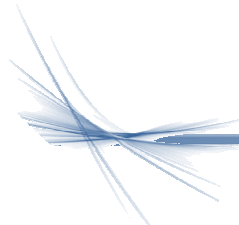
infectados (55). En un segundo paso, a través de descriptores morfológicos del parásito, el clasificador discrimina entre las diferentes especies causantes de malaria con una sensibilidad y un VPP del 98% y 96%, respectivamente. Moon *et al* elaboraron un algoritmo de análisis de imágenes para la detección, cuantificación y clasificación de las diferentes etapas del ciclo de vida de *Plasmodium* obtenido de muestras en cultivo, así como discriminación entre parásitos viables y no viables (muertos) en muestras tratadas con fármacos (56). Este algoritmo identifica, a través de árboles de decisión, eritrocitos parasitados sobre muestras teñidas con 4',6-Diamidino-2-fenilindol diclorhidrato (DAPI) para la detección de núcleos de parásitos, y teñidas con Mitotracker Red CMXRos para la detección de mitocondrias en parásitos viables. Su sistema mostró un error respecto al conteo manual inferior al 5%.

Otro tipo de algoritmos utilizados en el reconocimiento automático de la malaria incluyen trabajos como los de Vink *et al*, en el que desarrollaron modelos basados en el *boosting* adaptativo acoplado a un sistema de escáner fácil de usar en muestras teñidas con naranja de acridina para la detección de malaria en muestras con parasitemias bajas (57). Su sistema mostró una sensibilidad del 75%. Suryawanshi *et al* realizaron trabajos similares a los mencionados mediante el uso de un clasificador euclidiano lineal, aunque no mostraron resultados en el rendimiento de la clasificación de su modelo (58).

Purwar *et al* son los únicos que desarrollaron un modelo a través de un aprendizaje no supervisado (17). Utilizaron una versión probabilística y ponderada de *K-means clustering* (agrupamiento de clusters) a partir de los valores de intensidad de los píxeles de la imagen, para identificar los píxeles que corresponden a los parásitos de malaria. Su modelo mostró un gran nivel de sensibilidad (100%) para la detección de casos de paludismo, mientras que obtuvieron una especificidad que oscilaba entre el 50 y el 88 % para el reconocimiento de las diferentes especies de parásitos.

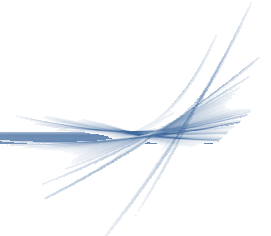
3.4 Redes neuronales y aprendizaje profundo

Las redes neuronales artificiales (ANN) forman un subconjunto del aprendizaje automático inspirado en la estructura y función del cerebro humano compuesto por unidades de cálculo multicapa. Los primeros trabajos que aplicaron las ANN en el reconocimiento de la malaria utilizaron perceptrones multicapa (MLP) con un rendimiento equivalente a los modelos



clásicos de aprendizaje automático. Diaz *et al* diseñaron un modelo basado en MLP, y de entre varias configuraciones posibles, tuvieron los mejores resultados de clasificación con una arquitectura de 7 neuronas en la capa oculta y una tasa de aprendizaje de 0,1 (25). Sin embargo, al comparar el rendimiento en el reconocimiento, el modelo basado en MLP obtuvo peores resultados en la clasificación cuando era comparado con un modelo basado en SVM lineal, obteniendo valores de la métrica *F score* de 0,94 para el MLP y 0,96 para SVM. Ross *et al* también usaron ANN para el reconocimiento automático de malaria a partir de imágenes de sangre periférica. Desarrollaron un modelo secuencial con dos MLP de arquitectura similar, dos clasificadores que contenían dos capas ocultas, pero con diferente configuración de neuronas (38). La tarea del primer clasificador se centraba en distinguir si el hematíe de entrada estaba infectado o no, y en un segundo paso, los hematíes reconocidos como infectados entraban en un segundo clasificador cuya finalidad consistía en distinguir la especie causante de malaria. Este modelo mostró una buena sensibilidad (92%), pero a costa de una gran tasa de falsos positivos, ya que el 60% de los hematíes reconocidos como infectados en realidad correspondían a hematíes normales. La exactitud en el reconocimiento a nivel de la especie resultó ser del 73%. Devi *et al* diseñaron una ANN con una capa de entrada que aceptaba 71 descriptores calculados y una capa oculta con una configuración de 20 neuronas (36). Este diseño de ANN ofrecía mejores resultados de exactitud (94,15%) en el reconocimiento de malaria al compararlo con un clasificador de Bayes, SVM o KNN. Sin embargo, esta exactitud conseguía mejorar hasta el 96,5% cuando se construía un modelo combinado que unificaba las predicciones de ANN con KNN y SVM por medio de una estrategia de mayoría de voto.

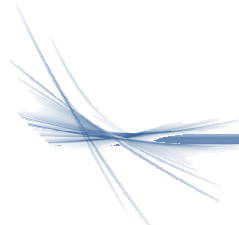
Las técnicas de *deep learning* han sufrido una gran evolución durante los últimos años. Estudios recientes han demostrado que las redes neuronales convolucionales (CNN) muestran una gran superioridad en el reconocimiento y clasificación automática de imágenes (59) en comparación con las ANN convencionales. Las CNN no requieren el uso de descriptores calculados y extraídos manualmente por el usuario, lo que permite ahorrar tiempo y recursos tanto en el desarrollo como en requerimientos computacionales. Sin embargo, se requieren conjuntos de datos más grandes y recursos de hardware más sofisticados para entrenar un modelo. A diferencia de otros modelos de ANN que trabajan con datos cuantitativos, las CNN están espacialmente diseñadas para el reconocimiento de imágenes, ya que analizan píxeles con una distribución local en lugar de analizar píxeles individuales. Esto favorece el reconocimiento de patrones en las imágenes, y debido a esto en los últimos años ha habido



una explosión de publicaciones en las que se explora el uso de las CNN para el reconocimiento y diagnóstico de pruebas clínicas basadas en imágenes (60), con grandes aplicaciones en el campo de la hematología, tanto para el reconocimiento de leucocitos displásicos (61), como para la identificación automática de los diferentes tipos de leucemias agudas (62), e incluso para el reconocimiento automático de linfocitos reactivos sugestivos de infección por SARS-COV-2 (63).

La arquitectura de las CNN permite un gran abanico de posibilidades. Su estructura puede ser definida por el propio usuario. Gopakumar *et al* entrenaron una CNN creada desde cero con una arquitectura lineal basada en cuatro capas convolucionales (64). Su modelo se entrenó para detectar hematíes infectados a partir de imágenes obtenidas de frotis sanguíneos. Consiguieron una sensibilidad y una especificidad del 97,1% y 98,5%, respectivamente. Liang *et al* también utilizaron una CNN creada desde cero con 6 capas convolucionales para el reconocimiento de malaria (65). Su modelo demostró una exactitud del 97,4%.

Existen arquitecturas CNN publicadas han demostrado una gran eficacia en la clasificación de imágenes y actualmente se encuentran disponibles en librerías de código abierto. Esto ha hecho que se haya explotado su uso en el reconocimiento de imágenes de una multitud de objetivos. Pan *et al* entrenaron una de las arquitecturas CNN más sencillas y pequeñas publicadas, conocida como Lenet (66). Este modelo consiste en dos capas convolucionales y 2 capas *fully connected*, con el objetivo de diferenciar hematíes infectados respecto a hematíes normales. Este modelo consiguió un buen rendimiento en la clasificación de la malaria, con un 95% de exactitud, pero mostraba unos resultados similares al compararlo con un modelo basado en SVM (95%). Dong *et al* compararon el rendimiento de diferentes arquitecturas de CNN (67). GoogleNet obtuvo los mejores resultados en la clasificación respecto a Lenet, Alexnet y respecto a SVM, con una exactitud del 98%. Rajaraman *et al* realizaron varios trabajos en los que estudiaban el rendimiento de diferentes CNN del estado del arte preentrenadas en la clasificación de hematíes normales respecto a hematíes infectados con malaria (68,69). En todas sus publicaciones se muestra un mejor rendimiento en la clasificación mediante el uso de VGG-16, VGG-19 y Resnet-50 respecto a otras arquitecturas, con resultados de exactitud que alcanzaron valores de 99,3%. Este rendimiento incluso parecía mejorar al combinar la capacidad predictiva de varios modelos, obteniendo una exactitud del 99,5% promediando las predicciones de una VGG-19 y una Squeezenet entrenadas por separado con imágenes de hematíes individuales.



Sin embargo, los modelos preentrenados de CNN también admiten cierta versatilidad en su configuración. Algunos autores han logrado combinar la potencia de una CNN en la lectura de imágenes y la conversión de estas imágenes en vectores de datos, con la potencia de algoritmos convencionales de clasificación. Vijayalakshmi *et al* obtuvieron un 93% de exactitud en el reconocimiento automático de malaria, modificando la estructura de una VGG-19 en la que sus capas convolucionales eran sustituidas por un SVM con un kernel de función radial (70).

Todos estos trabajos presentados, que tienen como finalidad el reconocimiento automático de parásitos de malaria a partir de muestras de sangre periférica, indican que existe una multitud de enfoques diferentes que pueden aplicarse para abordar este problema. Aunque la mayoría de trabajos publicados constituyen un valioso aporte científico en técnicas de análisis de imagen y clasificación, en ninguno de ellos se explora la posibilidad de que los modelos desarrollados tengan que enfrentarse a elementos para los que no han sido entrenados pero que pueden tener un impacto significativo en el rendimiento de la clasificación.

3.5 Analizadores para el reconocimiento automático de hematíes

Tradicionalmente, el análisis morfológico del frotis de sangre periférica se ha realizado mediante microscopía óptica manual. Aunque este método constituye el *gold standard*, tiene las desventajas de ser bastante laborioso, requiere un entrenamiento continuo del personal y está sujeto a una elevada variabilidad interobservador. Por lo tanto, existe una demanda creciente para el desarrollo de sistemas de microscopía digital, capaces de realizar un análisis morfológico automatizado de los frotis de sangre periférica. Las empresas de diagnóstico de laboratorio *in vitro* han sido conscientes de esta necesidad y es posible encontrar en el mercado del laboratorio clínico analizadores que son capaces de procesar automáticamente un frotis de sangre periférica a través de un análisis digital de imágenes. Estos sistemas disponen de un sistema automatizado de toma de imágenes o de escaneo para posteriormente analizar sus elementos mediante un modelo de reconocimiento automático. Ejemplos de estos analizadores lo constituye el CellaVision (71), Vision Hema (72), HemaCam (73), Cobas m511 (74), o DiffMaster Octavia (75) (ver Figura 9).

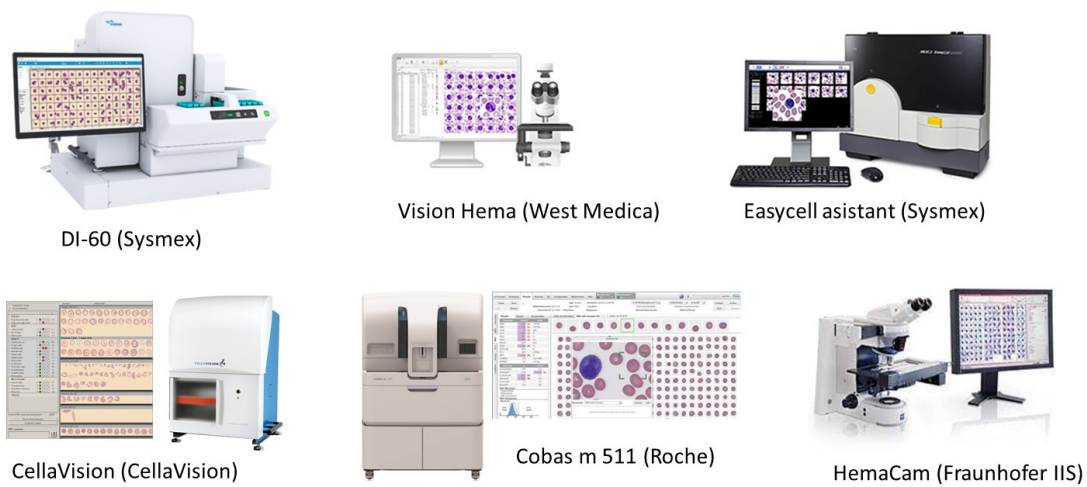


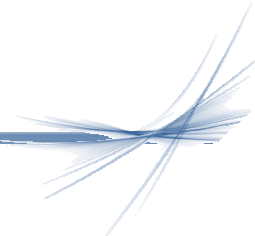
Figura 9. Analizadores para la revisión automatizada del frotis de sangre periférica.

De entre todos los proveedores de analizadores automatizados para la revisión del frotis de sangre periférica, CellaVision ha sido de las primeras compañías en proporcionar un analizador digital de imágenes en el laboratorio de hematología, lo que ha resultado en que se conviertan en los líderes del mercado. Prueba de ello es que la práctica totalidad de las publicaciones en las que se evalúa el rendimiento de este tipo de analizadores pertenecen a estudios del rendimiento del CellaVision (76). Este tipo de analizadores automatizados trabajan con imágenes digitales en las cuales localizan las células que se encuentran en el frotis de sangre periférica, para ello tienen un algoritmo de localización de la sección de la monocapa óptima para evaluar el frotis, y capturan imágenes de bajo y gran aumento como soporte para la evaluación de la serie roja y de los leucocitos, respectivamente. Inicialmente en la versión DM96, utiliza un software de análisis de imágenes basados en redes neuronales artificiales para preclasificar las diferentes subpoblaciones leucocitarias, para luego mostrar estas imágenes en una aplicación de ordenador en la que el usuario puede modificar y/o validar este recuento. Kratz *et al* documentaron que la preclasificación de los leucocitos era bastante precisa para los elementos normalmente mayoritarios de la sangre periférica, con exactitudes superiores al 90% para los neutrófilos y linfocitos, y superiores al 80% en el caso de los monocitos (77). Además, el análisis del frotis se realiza con mayor rapidez en comparación con la revisión manual al microscopio, lo que unido a una buena tasa de reconocimiento global, hace que el flujo de trabajo diario sea bastante fluido. Sin embargo, tiene un peor rendimiento en el reconocimiento de granulocitos inmaduros y de blastos, con exactitudes que se

encuentran entre el 50-60% (78), por lo que las muestras patológicas procesadas por el CellaVision requieren una atención más minuciosa.

En condiciones normales, el software del CellaVision no realiza una caracterización de las principales alteraciones eritrocitarias, aunque permite visualizar una imagen de campo desde bajo aumento con una definición aceptable para poder ser valorada visualmente. Cada una de estas imágenes ofrece un campo de visión en el que se pueden apreciar entre 2.000 y 3.000 hematíes, por lo que ofrece una buena representación de la población de la serie roja. Horn *et al* valoraron la capacidad de poder valorar la imagen global de la serie roja que ofrece el CellaVision en la detección de alteraciones eritrocitarias (79), definiendo como presencia de alteración las recomendaciones definidas por el ICSH (80). Según los resultados de este estudio el CellaVision tiene una buena especificidad, pero posee un rango muy amplio de sensibilidades en la detección de alteraciones eritrocitarias, con valores de sensibilidad que oscilan entre el 60%, como el caso de los esquistocitos, y el 100% en el caso de los drepanocitos. Esta imagen de campo que ofrece el CellaVision también ha sido evaluada para valorar su utilidad en la detección de parásitos de malaria por Racsa *et al* (81), mostrando unos niveles de sensibilidad similares a los obtenidos mediante microscopía manual, lo que pone de manifiesto la utilidad de las imágenes que ofrece este analizador en la detección de parásitos intraeritrocitarios.

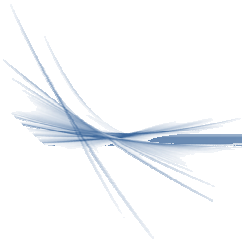
Sin embargo, las nuevas versiones de los analizadores CellaVision incorporan un módulo de software avanzado de análisis de la serie roja, el *Advance Red Blood Cell Application* (ARBCA) (76). Este módulo de análisis de la serie roja agrupa e identifica los hematíes en diferentes alteraciones morfológicas basadas en alteraciones de color, tamaño, forma e inclusiones mediante el uso de redes neuronales artificiales. Las diferentes categorías morfológicas que se consideran son: anisocitosis, microcitosis, macrocitosis, policromasia e hipocromía, poiquilocitosis, dianocitos, esquistocitos, queratocitos, drepanocitos, esferocitos, eliptocitos, ovalocitos, dacriocitos, estomatocitos, acantocitos y equinocitos. ARBCA también incluye una categoría de reconocimiento de inclusiones eritrocitarias que incluyen HJ, PP, PB y la presencia de parásitos intraeritrocitarios. Egelé *et al* evaluaron el rendimiento de ARBCA para la detección y clasificación de esquistocitos y dacriocitos, utilizando los umbrales (*cut-offs*) predefinidos por el propio CellaVision (82,83). Los resultados de la precaracterización ofrecidos por ARBCA mostraron un 20-30% de sobreestimación respecto a los recuentos obtenidos en la postclasificación, aunque con una buena correlación (superior a 0,95) entre ambos recuentos.



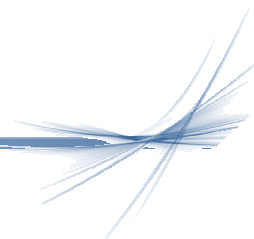
En un trabajo posterior en el que evaluaban la detección de diferentes alteraciones de tamaño y forma también encontraron un alto grado de acuerdo en todas las categorías, superior al 90%, al compararlo con el recuento manual (84). El grupo de Criel *et al* realizaron un trabajo similar en el que también se evaluó el rendimiento de ARBCA en la detección de la presencia de las diferentes alteraciones eritrocitarias teniendo en consideración diferentes *umbrales* (85). Dentro de todas las alteraciones eritrocitarias, encontraron que el ARBCA tenía muy baja sensibilidad en la detección de esferocitos, parásitos de malaria y presencia de PB, argumentando que para la detección de estas alteraciones siempre es necesaria la evaluación de la imagen de campo por parte del observador. Este módulo de ARBCA ha sido específicamente evaluado para la detección de malaria en algunos trabajos. Florin *et al* realizaron un análisis de regresión entre el recuento de parásitos ofrecido por ARBCA y el recuento manual en 40 muestras de frotis de sangre periférica (86). Este análisis estadístico mostró una pendiente de 0,75, reflejando una menor capacidad de detección de parásitos por el módulo ARBCA en comparación con la microscopía manual. En concreto, en aquellas muestras que tenían unos niveles de parasitemia inferior a 0,14%, ARBCA no era capaz de detectar la presencia de parásitos.

En todos los trabajos presentados donde se evalúa el rendimiento del módulo ARBCA se muestran resultados en términos de sensibilidad y especificidad de la capacidad de detección de una determinada alteración, lo cual tiene un sentido de aplicación clínica, pero por otra parte, son resultados que no muestran en detalle el comportamiento del clasificador ARBCA. Park *et al* fue el único grupo que estudió la exactitud en el reconocimiento automático de parásitos de malaria por el módulo ARBCA (87). Aunque solo evaluaron el rendimiento en el reconocimiento en un solo frotis con presencia de malaria analizado varias veces por el ARBCA, observaron que en la mayoría de los casos los parásitos eran confundidos con PP y con HJ. Yoon *et al*, en un trabajo de más alcance en el que evaluaron 98 frotis con presencia de parásitos de malaria por ARBCA, también encontraron que en la mayoría de los casos los hematíes con parásitos de malaria eran incorrectamente identificados como hematíes con HJ, con PP y con PB en un 57%, 85% y 76% de los casos, respectivamente (88).

Todos estos resultados demuestran que el analizador CellaVision supone una gran herramienta de soporte en el laboratorio clínico, especialmente en la clasificación de los leucocitos. Sin embargo, sus algoritmos de detección y clasificación automática de las inclusiones



eritrocitarias todavía necesitan ser mejorados para que tengan una aplicación práctica, tal como se ha comentado al comienzo de este capítulo.





CAPÍTULO 4

Objetivos

En este capítulo se describe el objetivo general y los objetivos específicos de la presente tesis doctoral. Este trabajo pretende contribuir a la detección de parásitos intraeritrocitarios en pacientes infectados por malaria, utilizando métodos de inteligencia artificial, tales como *machine learning* y *deep learning*, así como su diferenciación con respecto a otras inclusiones eritrocitarias.

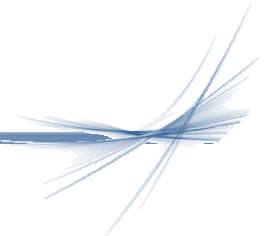
4.1 Objetivo general

El objetivo general de esta tesis doctoral consiste en diseñar e implementar un método de reconocimiento automático de malaria, que no solamente sea sensible a la presencia de parásitos sino también altamente específico en el reconocimiento de hematíes normales. Para evitar que otras inclusiones eritrocitarias sean falsamente reconocidas como malaria (sistema de baja especificidad), este método debe ser capaz de discriminar los hematíes parasitados no solo de los hematíes normales, sino también de los cuerpos de Howell-Jolly (HJ), cuerpos de Pappenheimer (PP), punteado basófilo (PB), así como de plaquetas que se encuentran alojadas sobre la superficie de los hematíes (PLT).

4.2 Objetivos específicos

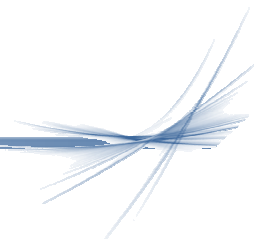
1. Desarrollar un método de reconocimiento automático de malaria y otras inclusiones eritrocitarias mediante herramientas de *machine learning* a partir de imágenes de sangre periférica obtenidas con microscopio. Este objetivo se plantea mediante las siguientes etapas:

1. Diseñar un método de preprocesamiento y segmentación de hematíes individuales a partir de las imágenes obtenidas de frotis de sangre periférica.
2. Implementar el cálculo y la extracción de descriptores que definan las características eritrocitarias.
3. Evaluar el número y el tipo de descriptores óptimos con mayor poder discriminatorio entre las diferentes alteraciones eritrocitarias.
4. Evaluar diferentes modelos de clasificación automática basados en el entrenamiento con los descriptores extraídos e implementar el clasificador con mejor rendimiento.



5. Evaluar el desempeño del clasificador en el reconocimiento de imágenes individuales a través de una prueba de concepto.
 6. Evaluar la capacidad del clasificador en el diagnóstico de pacientes infectados con malaria mediante una prueba de concepto.
2. Desarrollar un método de reconocimiento automático de malaria y otras inclusiones eritrocitarias mediante herramientas de *deep learning* a partir de imágenes de sangre periférica obtenidas con microscopio. Este objetivo se plantea mediante las siguientes etapas:
1. Desarrollar un nuevo método de obtención de imágenes de hematíes individuales enfocado al entrenamiento de algoritmos de *deep learning*.
 2. Evaluar diferentes modelos de *deep learning* para el reconocimiento automático de malaria y otras inclusiones eritrocitarias.
 3. Evaluar el desempeño del clasificador en el reconocimiento de imágenes individuales a través de una prueba de concepto.
 4. Evaluar la capacidad del clasificador en el diagnóstico de muestras de pacientes infectadas con malaria mediante una prueba de concepto.
 5. Estudiar los mapas de características asociados a las capas convolucionales con la finalidad de interpretar el aprendizaje de las redes neuronales en el reconocimiento de las inclusiones eritrocitarias.
3. Desarrollar un método de reconocimiento automático de malaria y otras inclusiones eritrocitarias mediante herramientas de *deep learning* a partir de imágenes de sangre periférica obtenidas con el autoanalizador CellaVision. Este objetivo se plantea mediante las siguientes etapas:
1. Desarrollar un nuevo método de segmentación para la obtención de imágenes de hematíes individuales a partir de imágenes obtenidas con el CellaVision.

2. Evaluar diferentes modelos de *deep learning* para el reconocimiento automático de malaria y otras inclusiones eritrocitarias.
3. Evaluar el desempeño del clasificador en el reconocimiento de imágenes individuales a través de una prueba de concepto.
4. Evaluar la capacidad del clasificador en el diagnóstico de pacientes infectados con malaria mediante una prueba de concepto.
5. Comparar el rendimiento en el reconocimiento automático entre el modelo de *deep learning* obtenido y el software *Advance RBC Application* del autoanalizador CellaVision.





CAPÍTULO 5

Reconocimiento automático de malaria mediante modelos de *machine learning*

Este capítulo presenta la metodología desarrollada y los resultados obtenidos para llevar a cabo y evaluar el primer objetivo específico de la presente tesis doctoral: “desarrollar un método de reconocimiento automático de malaria y otras inclusiones eritrocitarias mediante herramientas de *machine learning* a partir de imágenes de sangre periférica obtenidas con microscopio”.

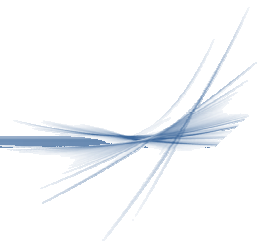
Se describe en primer lugar todos los procesos relacionados con las imágenes, desde su obtención a partir del frotis de sangre periférica (Apartado 5.1), la extracción de características (Apartado 5.2), hasta la clasificación de las imágenes en el Apartado 5.3. Los resultados se presentan en el Apartado 5.4.

5.1 PROCESAMIENTO DIGITAL DE IMÁGENES

5.1.1 Obtención de muestras

El flujo de trabajo en el laboratorio de hematología empieza por la extracción de la sangre periférica en tubos que generalmente contienen un anticoagulante basado en ácido etildiamínico tetraacético (EDTA). Esta muestra es procesada por los autoanalizadores hematológicos, consistentes en citómetros de flujo que son capaces de analizar las diferentes poblaciones de la sangre a través del análisis de la dispersión de luz, láser y/o fluorescencia de sus elementos. Aquellas muestras en las que se observan alteraciones tanto cualitativas como cuantitativas son seleccionadas para su revisión manual por microscopía. La revisión del frotis de sangre periférica es una técnica sencilla y barata que supone la base del diagnóstico en enfermedades hematológicas y no hematológicas. Para llevar a cabo el objetivo descrito al inicio de este capítulo, se han seleccionado 87 muestras de pacientes que incluyen pacientes infectados con malaria. A partir de muestras de sangre extraídas en tubos con K₂EDTA, se han realizado frotis de sangre periférica teñidos con May Grunwald-Giemsa (MGG), siguiendo procedimientos de tinción publicados en recomendaciones internacionales (89).

El frotis de sangre periférica es una técnica manual que requiere cierta destreza. Se deposita una gota de sangre sobre un portaobjetos, y se coloca el extremo de un segundo portaobjetos sobre dicha gota formando un ángulo de 45º respecto al primero. El ángulo de inclinación del portaobjetos que se desliza sobre la gota de sangre condiciona el grosor de la extensión. Por



último, se arrastra el portaobjetos sobre la gota de sangre con suavidad y decisión, siguiendo la dirección del primer portaobjetos (ver Figura 10).

Una vez extendida la sangre y secada se procede a la tinción. La tinción de MGG consiste en sumergir las extensiones en May Grünwald puro durante minuto y medio aproximadamente, lavar dichas extensiones con agua del grifo, sumergir posteriormente las extensiones en una segunda cubeta que contiene Giemsa al 10 % durante 10 minutos, sumergir las extensiones en agua del grifo durante unos segundos y finalmente, dejar secar. Las muestras de sangre teñidas con MGG observadas al microscopio muestran resaltadas las estructuras celulares, lo que facilita su identificación (ver Figura 11).

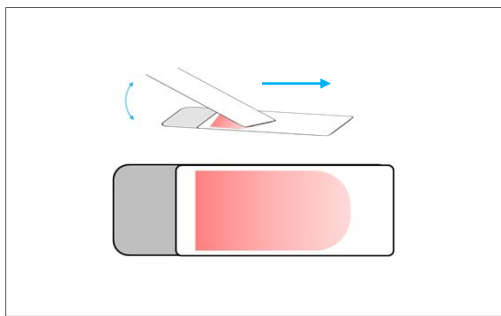


Figura 10. Procedimiento de la extensión de sangre periférica, deslizamiento de la gota de sangre sobre el portaobjetos y aspecto final de una extensión bien realizada.

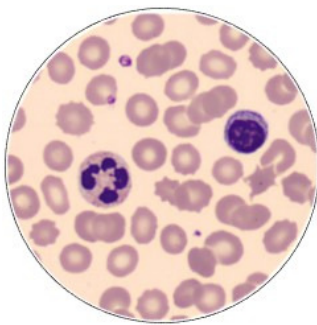


Figura 11. Efecto de la tinción sobre una muestra de sangre periférica.

El procedimiento de extensión-tinción ha sido realizado mediante el analizador automatizado Sysmex SP-1000 (Kobe, Japan), lo que confiere una mayor estandarización a este proceso.

5.1.2 Obtención de imágenes

Las muestras de frotis han sido analizadas mediante un microscopio óptico Olympus BX43 con una cámara Olympus DP73 acoplada. Se han obtenido un total de 4.150 imágenes de campo a partir de las extensiones. La captura de las imágenes se ha realizado mediante el uso del software CellSens v3.2 de Olympus, almacenando las imágenes en formato TIFF con una profundidad de 12 bits y 2400 x 1800 píxeles de resolución. Las mismas imágenes han sido guardadas también en formato comprimido JPG con una profundidad de 8 bits.

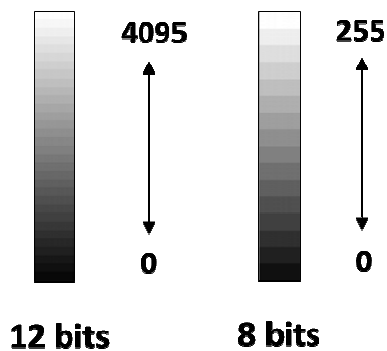


Figura 12. Gradación de la escala de grises en imágenes de 12 y 8 bits de profundidad.

En una imagen de 12 bits, los valores de los píxeles que lo componen pueden adquirir un rango de valores que oscilan entre 0 y 4.095, mientras que en las imágenes de 8 bits se observan valores entre 0 y 255 (ver Figura 12). Esto le confiere a la imagen una mayor profundidad, lo que se traduce en una mayor gama de colores totales que se podrán reproducir. Si pensamos que las imágenes digitales son representaciones cuantitativas (números), la imagen de 12 bits “almacenan mayor información”, y debido a que la extracción de descriptores que se realizará posteriormente tendrá un gran impacto los valores de los píxeles, se ha decidido almacenar las imágenes en un formato TIFF de alta calidad (12 bits) y un formato convencional JPG de menor calidad (8 bits).

Todo el proceso de trabajo desde la obtención de la muestra hasta la captura de las imágenes de campo puede observarse en el esquema de la Figura 13:

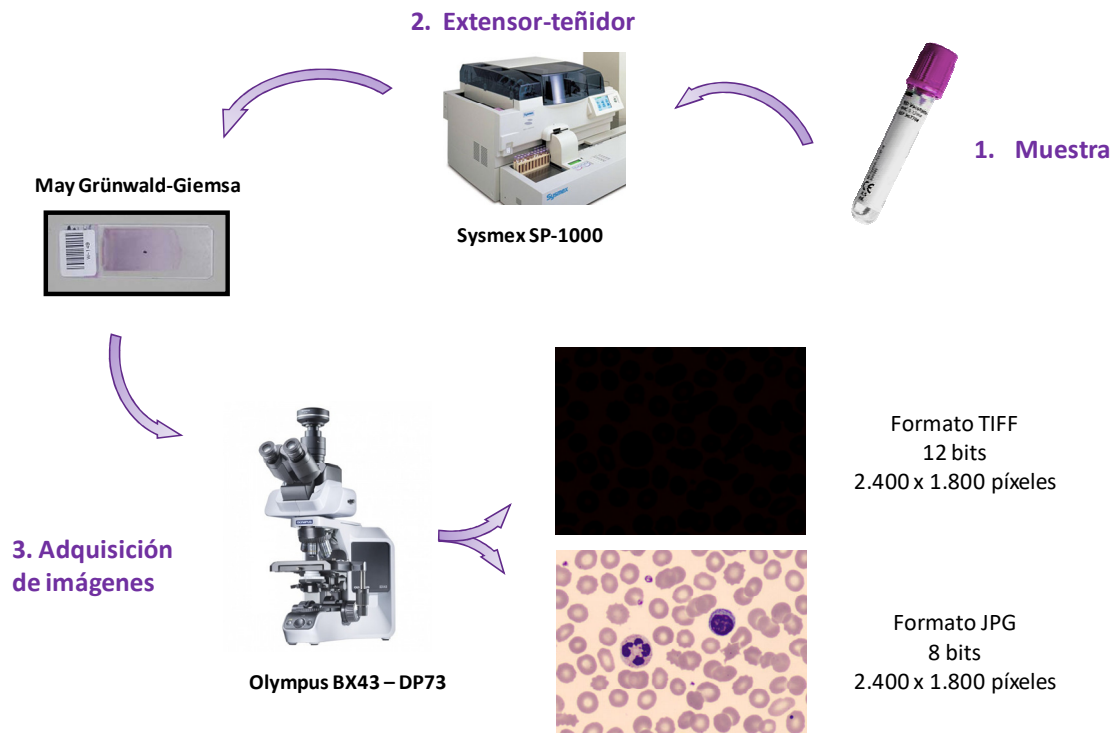


Figura 13. Esquema de la obtención de imágenes desde la toma de muestra.

5.1.3 Etiquetado

Una vez obtenidas las imágenes, es importante identificar los elementos que queremos detectar. El etiquetaje permite establecer la clase a la que pertenece cada ROI, esto determinará el entrenamiento, las diferentes clases que será capaz de reconocer el clasificador, y nos permitirá conocer también si un elemento determinado es correctamente reconocido o no. El etiquetaje de los hematíes se realizó mediante la aplicación *Image Labeller* de la plataforma de programación y cálculo matemático MATLAB® v. R2017b.

Las imágenes de 12 bits no son visualizables por la mayoría de los programas reproductores de imagen. Por lo que el etiquetaje se realizó a través de las imágenes obtenidas en JPG, donde cada hematíe individual de una imagen de campo queda etiquetado con el tipo de alteración definida previamente por especialistas en citología de referencia. Esta etiqueta, además de asociar el tipo de alteración, delimita el hematíe por un rectángulo que también almacena sus coordenadas dentro de la imagen. A través de un sencillo algoritmo en MATLAB® es posible exportar a un archivo en formato CSV los hematíes etiquetados de todas las imágenes

correspondientes a un mismo frotis junto con sus coordenadas (ver Figura 14), las cuales son igualmente extrapolables a las imágenes TIFF de 12 bits. Este archivo CSV servirá posteriormente para recortar cada hematíe tanto de la imagen original como de la imagen binarizada y obtener así imágenes de células individuales.

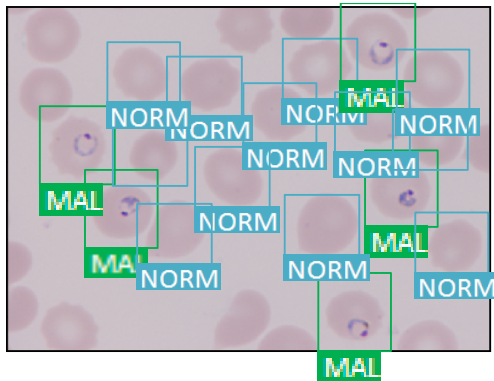


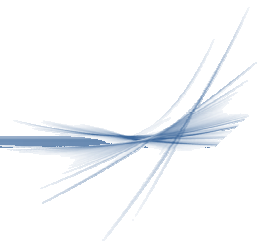
Figura 14. Proceso de etiquetado de los hematíes en una imagen de campo mediante la aplicación Image Labeller.

5.1.4 Preprocesamiento

Las técnicas de preprocesamiento pretenden mejorar o realzar las propiedades de las imágenes con la finalidad de facilitar las operaciones sobre las ROI que se realizarán en etapas posteriores, como la segmentación, la extracción de características y la preparación del dataset para el entrenamiento del modelo. Todo el procesamiento y el tratamiento de las imágenes se ha realizado bajo el entorno de programación *Python* v3.6, utilizando la librería de código abierto *scikit-image* (90), que se basa en una colección de algoritmos para el procesamiento de imágenes y la visión artificial.

Para trabajar con las imágenes y así poder tratarlas es necesario transformarlas a un formato reproducible que nos permita su visualización. Las imágenes adquiridas a través del microscopio óptico convencional tienen una profundidad de 12 bits. Por tanto, con una operación sencilla se transforman las imágenes a 16 bits, con la finalidad de poder visualizarlas y procesarlas.

Dentro del espacio de color RGB en el que normalmente representamos una imagen, los hematíes muestran un mayor contraste respecto al fondo de la imagen en la representación del componente verde (21). Esto permite diferenciar mejor los píxeles que pertenecen a los hematíes respecto a los píxeles que forman el fondo de la imagen. Un elemento muy importante en el procesamiento de imágenes es el histograma. Consiste en una función



discreta que cuenta el número de ocurrencias de cada nivel de gris que una imagen presenta. Se representa en un diagrama donde en el eje de abscisas está el nivel de gris y el de ordenadas la frecuencia de cada nivel de gris. Un histograma no da información sobre la disposición espacial de las diferentes intensidades, por ejemplo, dos imágenes distintas pueden tener el mismo histograma. Si observamos el histograma de los valores de los píxeles en el canal rojo, verde y azul observamos que en todos los canales se aprecian 2 picos (ver Figura 15): uno de intensidades más bajas (más oscuras), de curva más aplanada que corresponden a los píxeles de los hematíes, y un segundo pico más acentuado, de intensidades más elevadas (de color más claro), que corresponden a los píxeles del fondo de la imagen. El canal donde se observa una mayor distancia entre ambos picos es el verde, por lo que se escoge este canal como imagen de base para una separación de ambos objetos por binarización.

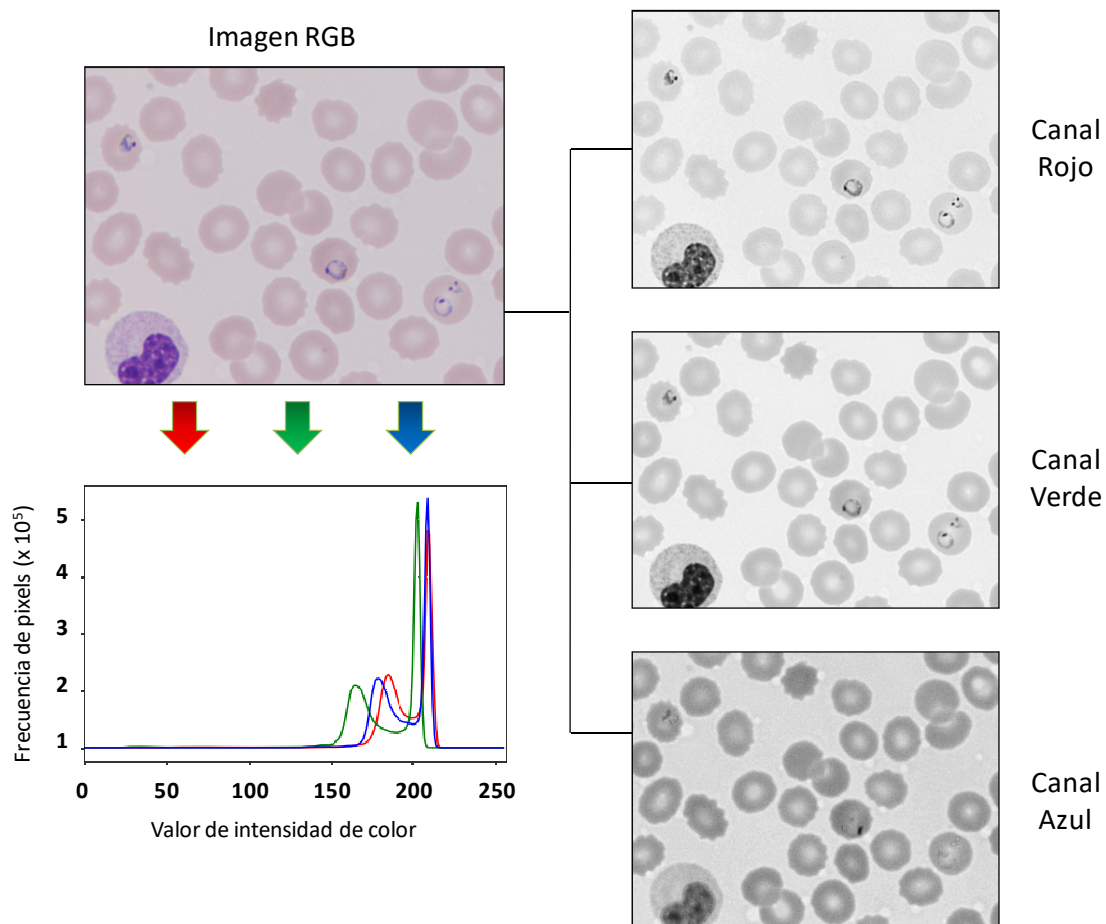


Figura 15. Representación del histograma y de las imágenes en los canales rojo, verde y azul del espacio de color RGB.

La imagen puede ser ajustada previamente para conseguir un mejor rendimiento en la binarización. Se ajusta el contraste, parámetro asociado al histograma. Consiste en el mapeo de valores de intensidad en la escala de grises a nuevos valores de intensidad para incrementar el contraste. La ecualización de histograma es una transformación muy útil que proporciona la distribución de cada valor de escala de grises de la imagen transformada sea la misma. Como consecuencia, el primer pico del histograma de la imagen se desplaza hacia valores de intensidades más oscuras, mientras que el pico del fondo se desplaza hacia intensidades más claras. La consecuencia es que en la imagen se observa un mayor contraste entre los hematíes y el fondo de la imagen, tal como se ilustra en la Figura 16.

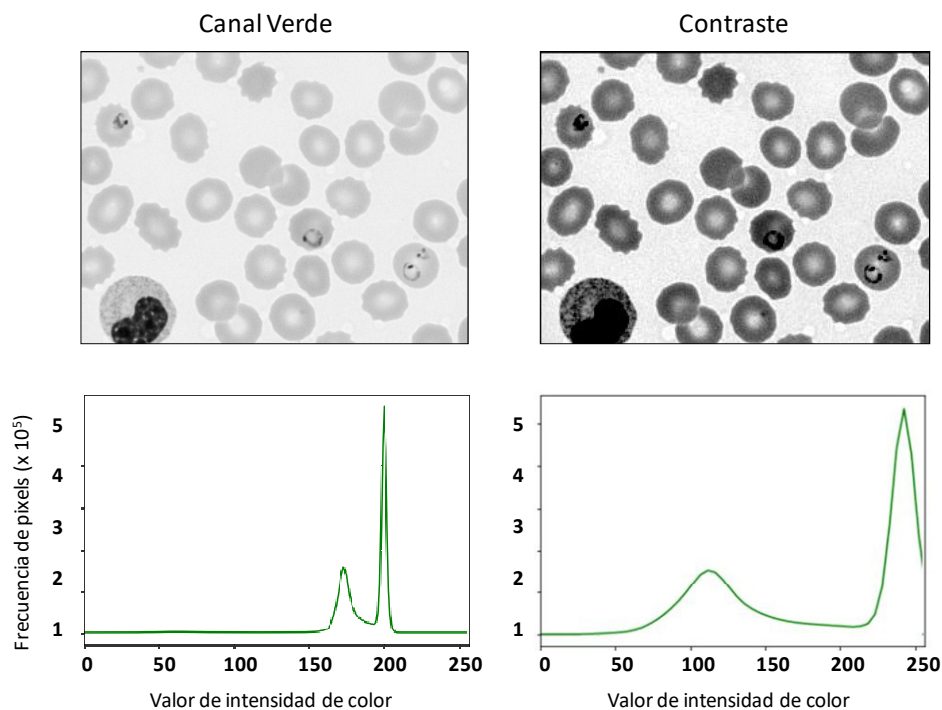
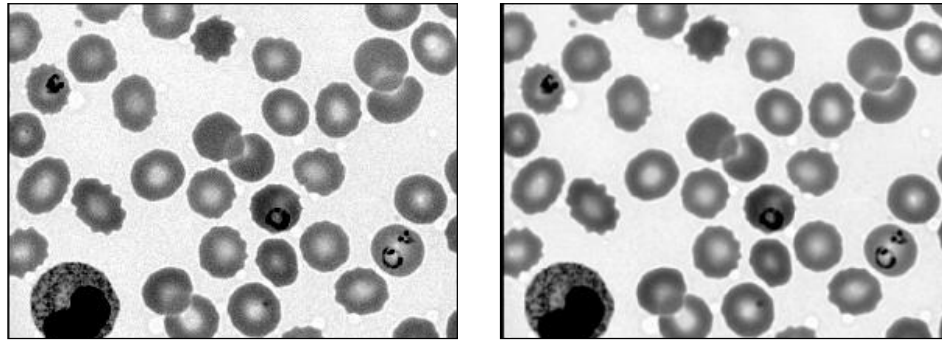
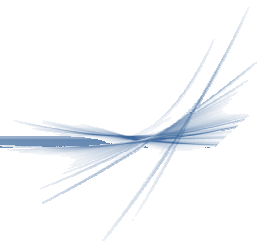


Figura 16. Representación del histograma y de las imágenes al aplicar una ecualización del histograma.

Los filtros no lineales tienen muchas aplicaciones, especialmente en la eliminación de ciertos tipos de ruido que no son aditivos. Por ejemplo, el filtro mediana actúa reemplazando el valor de gris original de cada pixel por la mediana del valor de los píxeles de su entorno. De esta forma se homogenizan los píxeles de intensidad muy distinta respecto a los vecinos, la imagen se suaviza y esto también ayuda a eliminar el ruido de fondo.



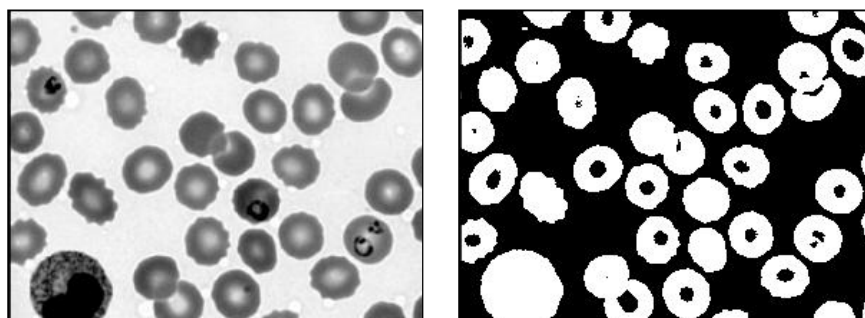
Ecualización del histograma

Filtro mediana

Figura 17. Comparación de las imágenes al aplicar un filtro mediana.

5.1.5 Binarización

Seguidamente se binariza la imagen mediante el método de Otsu para obtener una imagen binaria de las células sin el fondo de la imagen. Este método de binarización calcula un valor umbral que minimiza la varianza de los píxeles blancos y negros, devolviendo una imagen binaria. Este método de umbralización busca que la dispersión dentro de cada segmento sea lo más pequeña posible, pero a la vez que la dispersión entre segmentos diferentes sea lo más elevada posible. El resultado es que los píxeles que forman los hematíes ahora tienen un valor de 1, mientras que el resto de la imagen adquiere un valor 0 (ver Figura 18). Esto hace que sea sencillo identificar los píxeles que nos interesan dentro de la imagen. La umbralización por el método de Otsu se ha convertido en un procedimiento muy utilizado por su sencilla implementación a la par que ofrece muy buenos resultados de segmentación.



Filtrado + Contraste

Umbralización de Otsu

Figura 18. Representación de la binarización mediante el método de Otsu.

5.1.6 Postprocesamiento y segmentación

La metodología de preprocesamiento realizada sobre las imágenes ha servido de ayuda para facilitar la detección de los hematíes dentro de la imagen a través de la umbralización. Sin embargo, esto no evita que este método de binarización también detecte otro tipo de objetos que realmente no corresponden con regiones de interés para el estudio, como los leucocitos y las plaquetas que no están alojadas en la superficie de los hematíes. Por lo tanto, es necesario desarrollar un método para eliminar estos objetos no deseados de las imágenes binarizadas. Una solución a este problema podría consistir en la selección de elementos en base a su tamaño, ya que los leucocitos tienen un tamaño mayor al de un hematíe y las plaquetas muestran un tamaño inferior. Pero este método podría suponer un problema en la segmentación de las imágenes ante ciertas condiciones: presencia de plaquetas gigantes o protoplaquetas, poblaciones de hematíes con una marcada microcitosis como en los síndromes talasémicos o los fragmentos de hematíes propios de microangiopatías trombóticas, en poblaciones de hematíes con macrocitosis donde el tamaño de los hematíes puede solaparse con el de los linfocitos, etc.

Una solución más apropiada es desarrollar un algoritmo similar al que se ha utilizado hasta ahora para la detección exclusiva de plaquetas y leucocitos. El contenido de las plaquetas y del núcleo de los leucocitos suelen mostrar un color rosado-violáceo. Esto hace que estas estructuras aparezcan muy resaltadas en el canal B del espacio de color LAB (luminosidad de negro a blanco, de rojo a verde, y de azul a amarillo). Mediante un método de umbralización en este espacio de color es posible seleccionar solamente las plaquetas y los leucocitos ver Figura 19).

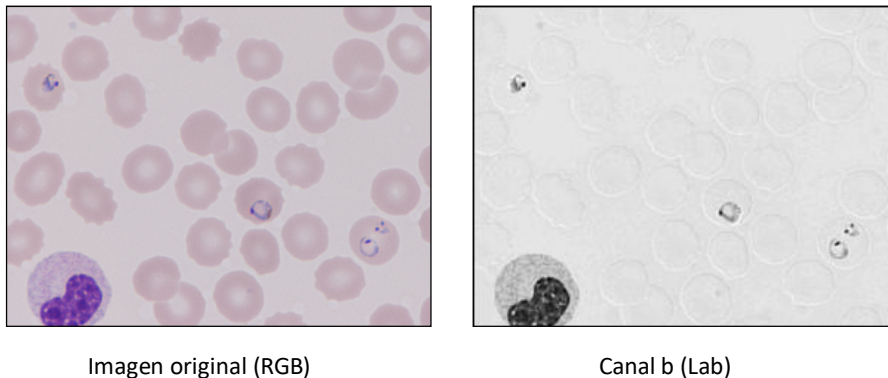
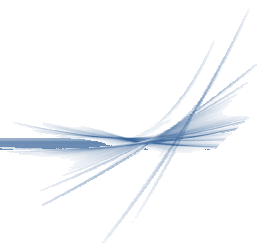


Figura 19. Representación de la imagen en el canal b del espacio de color Lab.



La imagen obtenida puede servir de plantilla para que, mediante una sencilla operación lógica, eliminar estos elementos de la imagen binarizada que contiene a los hematíes (ver Figura 20). Este método es un algoritmo más robusto para eliminar las regiones que no nos interesan de las imágenes, ya que puede ser igualmente efectivo en diferentes situaciones. La binarización y los filtros aplicados en ambos procedimientos (para la detección de los hematíes y para la detección de leucocitos/plaquetas) pueden originar pequeñas diferencias en el tamaño de los elementos que hace que queden pequeñas agrupaciones de píxeles con valor de 1. Por lo que se aplica una función para eliminar todos aquellos píxeles conectados que forman objetos con un área inferior a 1.000 píxeles.

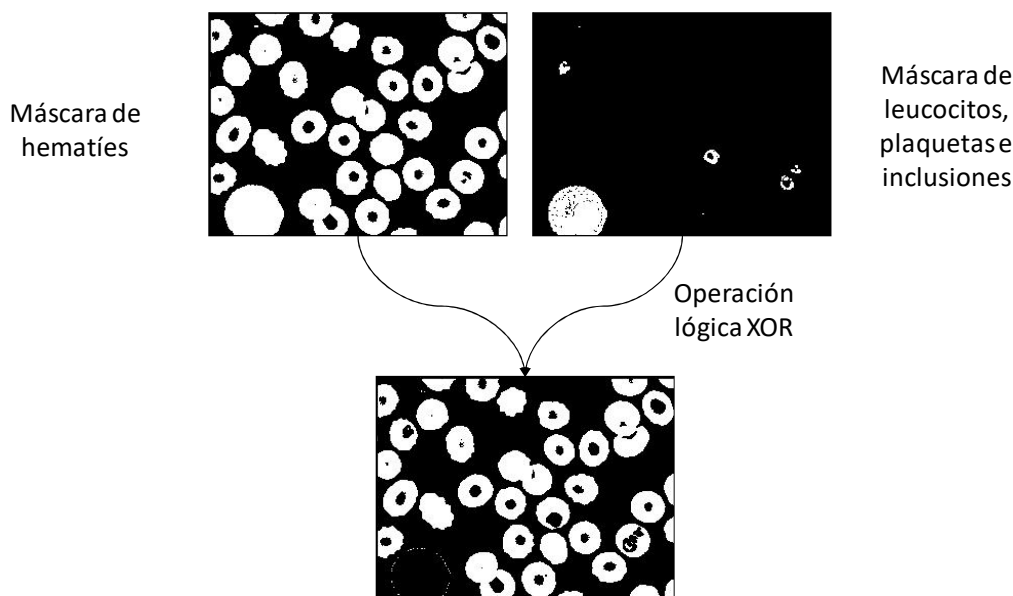


Figura 20. Eliminación de los leucocitos y plaquetas de la imagen binarizada.

La mayoría de los hematíes se caracterizan por tener una zona central pálida. Esto deja un hueco que en la binarización es reconocido como píxeles que corresponden al fondo de la imagen y no propiamente al eritrocito. La función *"fill_holes"* utiliza morfología matemática a través de procesos de dilación que conlleva a asociar como una ROI aquellos píxeles con valores de 0 que se encuentran dentro de una zona con valores de 1 cerrada y delimitada. Como consecuencia las ROIs ahora engloban a todos los hematíes enteros, incluyendo la zona de palidez central (ver Figura 21).

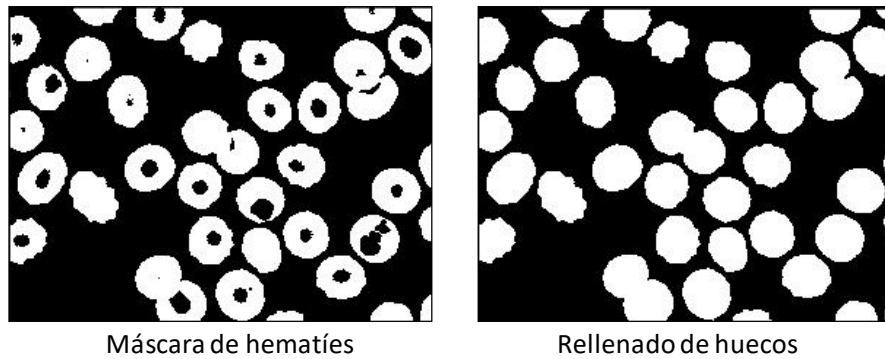
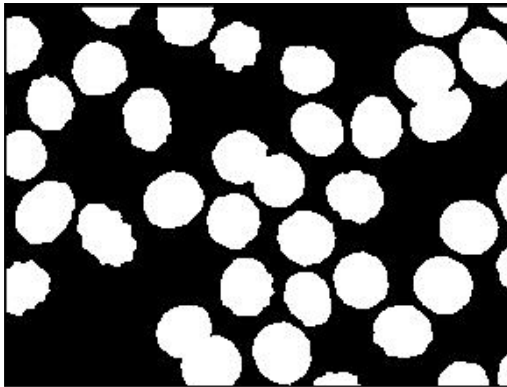
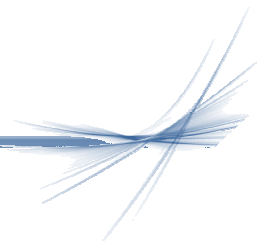
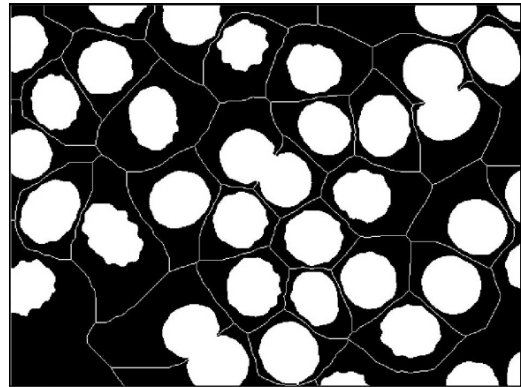


Figura 21. Representación del efecto de la función “rellenar huecos”.

Ahora que se dispone de una imagen nítida en la que solamente aparecen resaltados los hematíes, es necesario por una parte, separar los hematíes que adyacentes y que en la imagen binarizada aparecen como un único elemento, y por otra parte, identificar cada hematíe como un elemento individual y diferente a los demás. Un método para separar objetos es la transformada de Watershed. Esta técnica de segmentación interpreta la imagen como si fuese un relieve topográfico e inicia una inundación desde varios puntos. En las fronteras donde se unen varias inundaciones, se crea una pared que representa el borde entre varios segmentos de la imagen. Este proceso de “llenado” empieza por los mínimos locales de la imagen, y la obtención de estos mínimos es un aspecto de gran importancia para obtener una correcta segmentación. Para el cálculo de los mínimos locales se ha utilizado la transformada de la distancia mediante la métrica euclídea a partir de la imagen binarizada. La imagen resultante consiste en una imagen en escala de grises donde el valor de cada píxel representa la distancia entre ese píxel y el píxel distinto de cero más cercano de la imagen de partida binarizada. Sobre esta imagen se calcula su inversa de modo que los píxeles claros representen elevaciones altas y los píxeles más oscuros representen los valles, que corresponden a las zonas centrales de los hematíes. Esta imagen obtenida de la transformada de la distancia servirá como plantilla a partir de la cual se realice el procedimiento de Watershed. Como resultado, se crean fronteras entre los diferentes hematíes que componen la imagen, lo que permite identificar cada uno de ellos como un elemento individual (ver Figura 22).



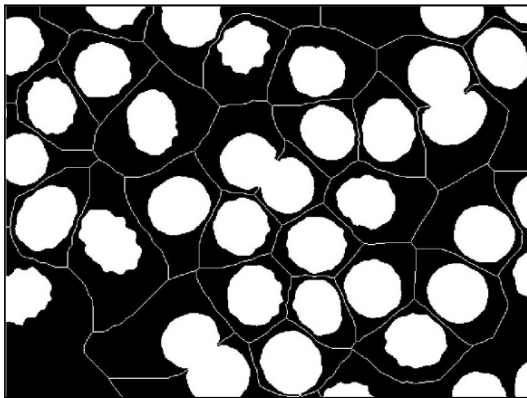
Máscara de hematíes



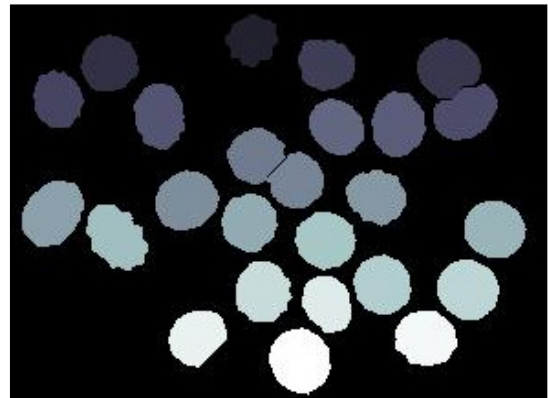
Segmentación Watershed

Figura 22. Resultados de la segmentación mediante el método de Watershed.

Otro aspecto a tener en cuenta son los hematíes que quedan en los márgenes de la imagen. Estos hematíes no corresponden a hematíes enteros y por lo tanto son fragmentos de hematíes en los que es posible no encontrar una inclusión en caso de que los hubiera. Al no tratarse de elementos que puedan ser representativos para su correcta evaluación, todos los hematíes que se encuentran colindando con los márgenes de la imagen son descartados a través de la función “clear_borders” (ver Figura 23).



Segmentación Watershed



Eliminar hematíes de los bordes

Figura 23. Resultados de la imagen final antes de aplicar el proceso de recorte de los hematíes individuales.

El algoritmo completo que comprende desde la captura de las imágenes hasta la obtención de las imágenes de hematíes individuales puede resumirse en el diagrama de la Figura 24.

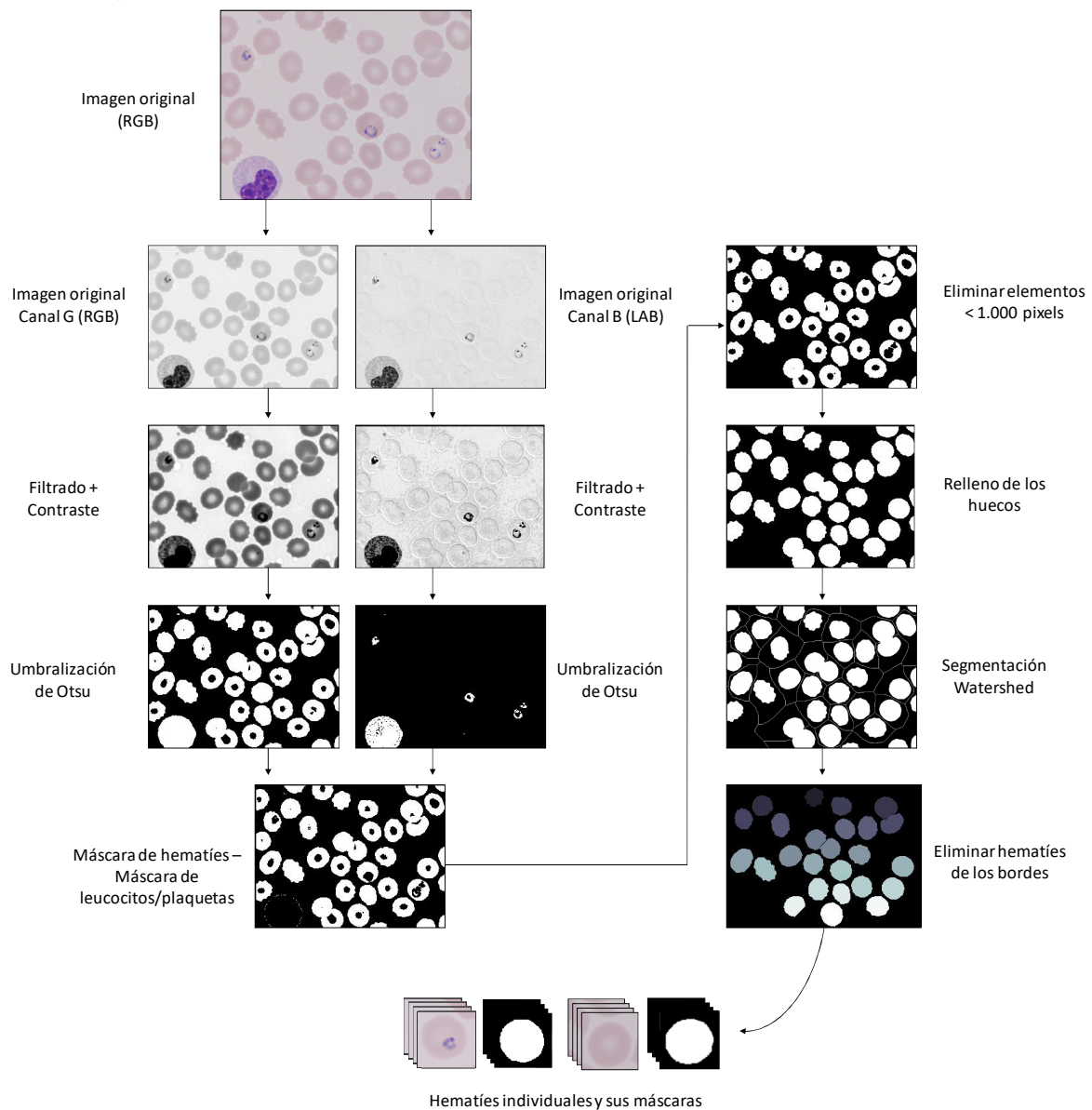
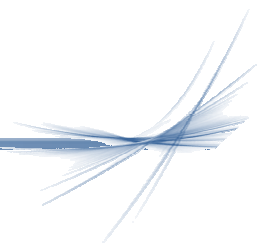


Figura 24. Esquema de bloques de los algoritmos aplicados a la imagen original de campo con la finalidad de obtener las imágenes de hematíes individuales.

5.2 EXTRACCIÓN DE CARACTERÍSTICAS

Antes de proceder al entrenamiento de los algoritmos de clasificación en las metodologías clásicas de *machine learning*, es necesario alimentar un modelo con datos cuantitativos. Para ello hay que extraer de cada imagen de hematíe individual una serie de descriptores que ayuden a caracterizar cada imagen y a distinguirla de las imágenes que pertenecen a otras clases. En este sentido, es importante analizar solamente los píxeles que forman la ROI.



Aunque las imágenes obtenidas por el método de segmentación pertenecen a hematíes individuales, también encontramos píxeles que pertenecen al fondo de la preparación y a fragmentos de otros hematíes cercanos. Pero la segmentación también nos ha permitido obtener la imagen de la máscara de cada hematíe, por lo que combinando ambas imágenes (original y máscara), es posible obtener los descriptores únicamente de los píxeles que forman el hematíe (ver Figura 25).

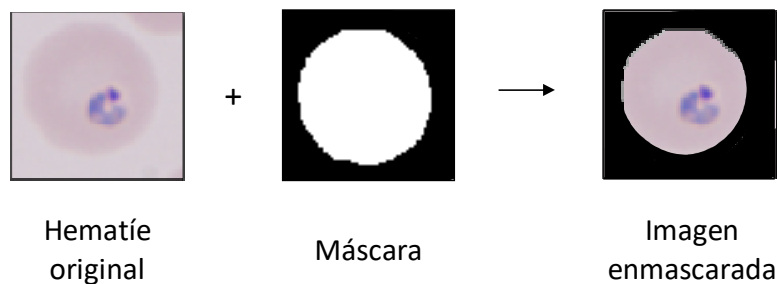


Figura 25. La imagen enmascarada utilizada para analizar únicamente los píxeles de la imagen que realmente son relevantes.

Por otra parte, dado que la presencia de inclusiones eritrocitarias puede aparecer en cualquier tipo de hematíe independientemente de la forma de éstos, el análisis de los descriptores se ha realizado únicamente a través de descriptores de color y textura, dado que con toda seguridad serán los más relevantes en la diferenciación de las distintitas inclusiones eritrocitarias objeto de estudio. Se han excluido del estudio, por lo tanto, aquellos descriptores que analizan cuantitativamente la forma y la geometría del hematíe. La extracción de los descriptores de color y textura se ha realizado a través de la librería de código abierto *Pyradiomics* (91). Los descriptores analizados se agrupan en las siguientes clases:

- Estadísticos de primer orden (*First Order Statistics*): Las estadísticas de primer orden describen la distribución de las intensidades de los píxeles dentro de la ROI de la imagen definida por la máscara. Proporcionan valores cuantitativos a través de métricas básicas y de uso común. Los descriptores de primer orden calculados fueron los siguientes:
 - *Energy*
 - *Total Energy*
 - *Entropy*
 - *Minimum*
 - *Maximum*
 - *10th percentile*
 - *90th percentile*
 - *Mean*
 - *Median*
 - *Interquartile Range*

- *Range*
- *Mean Absolute Deviation*
- *Robust Mean Absolute Deviation*
- *Root Mean Squared*
- *Standard Deviation*
- *Skewness*
- *Kurtosis*
- *Variance*
- *Uniformity*

- Matriz de coocurrencia (*Gray Level Cooccurrence Matrix*, GLCM): Una matriz de coocurrencia de nivel de grises de tamaño $N \times N$ describe la función de probabilidad conjunta de segundo orden de una región de la imagen que se define como $P(i,j|\delta,\theta)$. El par (i,j) de esta matriz representa el número de veces que la combinación de los niveles de intensidad i y j ocurren en dos píxeles de la imagen, que están separados por una distancia δ a lo largo del ángulo θ . La Figura 26 muestra un ejemplo del cálculo de una matriz de coocurrencia.

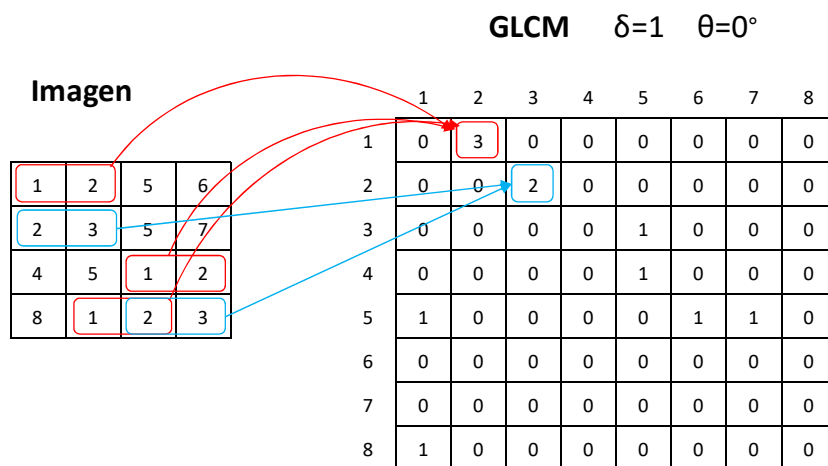
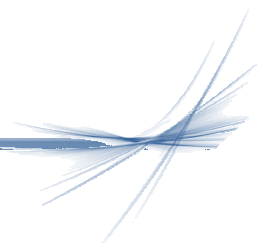


Figura 26. Ejemplo de cálculo de la matriz de coocurrencia (GLCM).

Los descriptores calculados a partir de la GLCM fueron los siguientes:

- *Autocorrelation*
- *Joint Average*
- *Cluster Prominence*
- *Cluster Shade*
- *Cluster Tendency*
- *Contrast*
- *Correlation*
- *Difference Average*
- *Difference Entropy*
- *Difference Variance*
- *Joint Energy*
- *Joint Entropy*



- Informational Measure of Correlation 1
- Informational Measure of Correlation 2
- Inverse Difference Moment
- Maximal Correlation Coefficient
- Inverse Difference Moment Normalized
- Inverse Difference
- Inverse Difference Normalized
- Inverse Variance
- Maximum Probability
- Sum Average
- Sum Entropy
- Sum of Squares

- Matriz de largo recorrido (Gray Level Run Length Matrix, GLRLM): este tipo de matriz cuantifica el recorrido de nivel de gris, que se define como la longitud de píxeles consecutivos que tienen el mismo valor de nivel de gris. En una matriz de largo recorrido de nivel de gris $P(i,j|\theta)$, el elemento (i,j) describe el número de repeticiones con nivel de gris i y longitud j que aparece en la imagen a lo largo del ángulo θ . Ver ejemplo en la Figura 27.

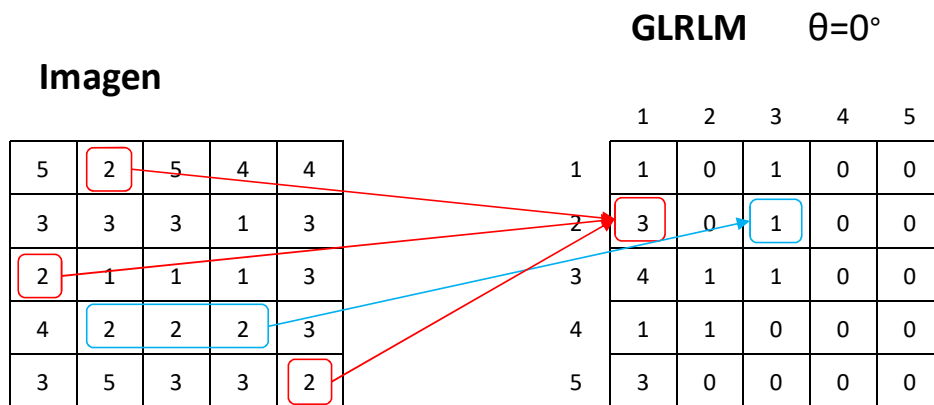


Figura 27. Ejemplo de cálculo de la matriz de largo recorrido (GLRLM).

Los descriptores calculados a partir de la GLRLM fueron los siguientes:

- Short Run Emphasis
- Long Run Emphasis
- Gray Level Non-Uniformity (GLNU)
- GLNU Normalized
- Run Length Non-Uniformity (RLNU)
- RLNU Normalized
- Run Percentage
- Gray Level Variance
- Run Variance
- Run Entropy
- Low Gray Level Run Emphasis
- High Gray Level Run Emphasis
- Short Run Low Gray Level Emphasis
- Short Run High Gray Level Emphasis

- *Long Run Low Gray Level Emphasis*
- *Long Run High Gray Level Emphasis*
- Matriz por tamaños de zona (*Gray Level Size Zone Matrix*, GLSZM): se trata de una matriz que cuantifica las zonas de niveles de gris similares en una imagen. Una zona de nivel de gris se define como el número de píxeles conectados que comparten la misma intensidad de nivel de gris. En una matriz de zona de tamaño de nivel de gris $P(i,j)$, el elemento (i,j) es igual al número de zonas con nivel de gris i y tamaño j que aparecen en la imagen. A diferencia de GLCM y GLRLM, GLSZM es independiente de la rotación, con solo una matriz calculada para todas las direcciones en la misma imagen. Se muestra un ejemplo del cálculo de este tipo de matrices en la Figura 28.

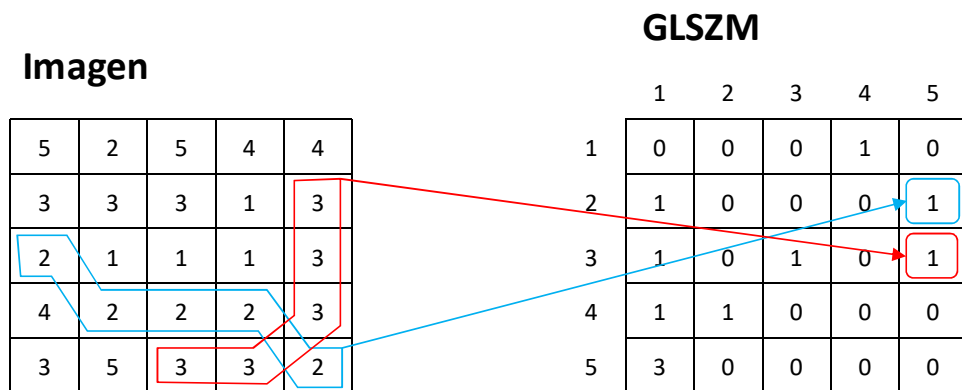
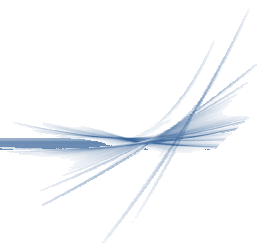


Figura 28. Ejemplo de cálculo de la matriz por tamaño de zona (GLSZM).

Los descriptores calculados a partir de la GLSZM fueron los siguientes:

- *Small Area Emphasis*
- *Large Area Emphasis*
- *Gray Level Non-Uniformity*
- *Gray Level Non-Uniformity Normalized*
- *Size-Zone Non-Uniformity*
- *Size-Zone Non-Uniformity Normalized*
- *Zone Percentage*
- *Gray Level Variance*
- *Zone Variance*
- *Zone Entropy*
- *Low Gray Level Zone Emphasis*
- *High Gray Level Zone Emphasis*
- *Small Area Low Gray Level Emphasis*
- *Small Area High Gray Level Emphasis*
- *Large Area Low Gray Level Emphasis*
- *Large Area High Gray Level Emphasis*
- Matriz por diferencia de tonalidad en la vecindad (*Neighbouring Gray Tone Difference Matrix*, NGTDM): cuantifica la diferencia entre un valor de gris y el valor de gris promedio de sus vecinos dentro de la distancia δ . La matriz NGTDM viene determinada por el valor



de los píxeles, su frecuencia absoluta y relativa, y por la suma de las diferencias absolutas para el nivel de gris i y el valor promedio de sus vecinos. La Figura 29 muestra un ejemplo ilustrado.

Imagen				NGTDM			
1	2	5	2	i	n_i	p_i	s_i
3	5	1	3	1	6	0,375	13,35
1	3	5	5	2	2	0,125	2
3	1	1	1	3	4	0,25	2,63
				4	0	0	0
				5	4	0,25	10,08

Valor del píxel (i) = 1

Número de ocurrencias (n_i) = 6

Frecuencia relativa (p_i) = $6/16 = 0,375$

$s_1 = |1-10/3| + |1-30/8| + |1-15/5| + |1-13/5| + |1-15/5| + |1-11/3| = 13,35$

Figura 29. Ejemplo de cálculo de la matriz por diferencia de tonalidad en la vecindad (NGTDM).

Los descriptores calculados a partir de la NGTDM fueron los siguientes:

- *Coarseness*
 - *Contrast*
 - *Busyness*
 - *Complexity*
 - *Strength*
- Matriz de dependencia (*Gray Level Dependence Matrix*, GLDM): Este tipo de matriz cuantifica las dependencias del nivel de gris en una imagen. Una dependencia de nivel de gris se define como el número de píxeles conectados dentro de la distancia δ que dependen de un píxel central. Un píxel vecino con nivel de gris j se considera dependiente del píxel central con nivel de gris i si $|i-j| \leq \alpha$. En una matriz de dependencia de nivel de gris $P(i,j)$, el elemento (i,j) describe el número de veces que aparece en la imagen un píxel con nivel de gris i con j píxeles dependientes en su vecindad. Véase el ejemplo de la Figura 30.

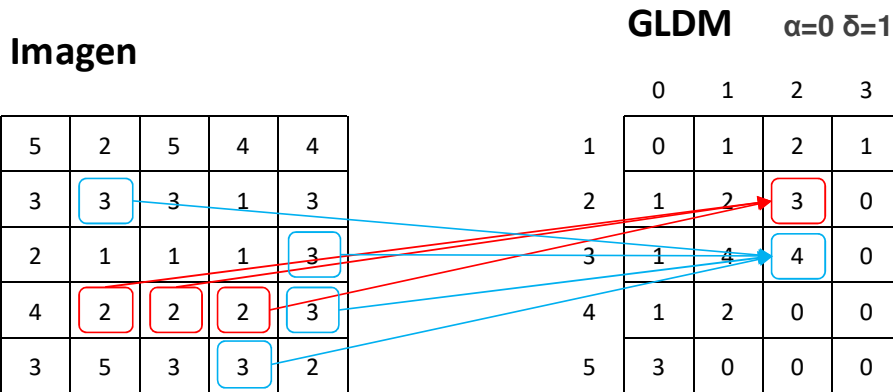
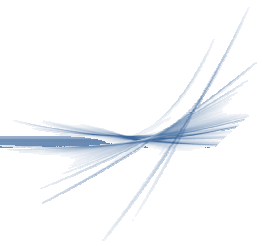


Figura 30. Ejemplo de cálculo de la matriz de dependencia (GLDM).

Los descriptores calculados a partir de la GLDM fueron los siguientes:

- *Small Dependence Emphasis*
- *Large Dependence Emphasis*
- *Gray Level Non-Uniformity (GLN)*
- *Dependence Non-Uniformity*
- *DNU Normalized (DNN)*
- *Gray Level Variance*
- *Dependence Variance*
- *Dependence Entropy*
- *Low Gray Level Emphasis*
- *High Gray Level Emphasis*
- *Small Dependence Low GL Emphasis*
- *Small Dependence High GL Emphasis*
- *Large Dependence Low GL Emphasis*
- *Large Dependence High GL Emphasis*

Todos los descriptores analizados y mencionados previamente hacen un total de 92 datos cuantitativos. Además, todas estas características se han extraído de cada componente que forman los siguientes espacios de color: RGB, HSV, XYZ (valores triestímulos), LAB, HED (hematoxilina, eosina, DAB), YIQ (luminancia, en fase y cuadratura), LCH (brillo, saturación, tono), YUV (luma, cromaticidad de proyección azul, cromaticidad de proyección roja), YCbCr y CMYK (cyan, magenta, amarillo, negro). Esto nos ha proporcionado un total de 2.852 valores cuantitativos para cada una de las imágenes de hemáties individuales que se utilizarán en el sistema de clasificación automática.



5.3 CLASIFICACIÓN AUTOMÁTICA

Los descriptores de cada tipo de hematíe obtenidos en la sección anterior servirán como datos de entrada a los módulos de reconocimiento automático a través del entrenamiento de algoritmos matemáticos de clasificación. En este apartado se presenta el desarrollo de dos modelos de clasificación de hematíes:

- 1) Un clasificador binario entrenado para diferenciar exclusivamente entre hematíes normales y hematíes infectados con malaria.
- 2) Un sistema secuencial de mayor complejidad de reconocimiento automático para diferenciar entre seis clases diferentes de hematíes:
 - Hematíes normales que no presentan ningún tipo de inclusión.
 - Hematíes con parásitos de malaria.
 - Hematíes con cuerpos de Howell-Jolly (HJ).
 - Hematíes con cuerpos de Pappenheimer (PP).
 - Hematíes con punteado basófilo (PB).
 - Hematíes con plaquetas alojadas en su superficie (PLT).

5.3.1 Clasificador binario

Este clasificador se entrena para diferenciar solamente entre hematíes normales y hematíes con parásitos de malaria. El objetivo de este clasificador es valorar su rendimiento cuando se enfrenta a la clasificación de hematíes con otros tipos de inclusiones, que constituyen hematíes para los que el clasificador no ha sido entrenado. Este tipo de hematíes son frecuentes en las muestras de sangre periférica que se analizan en los laboratorios clínicos. Por otro lado, los modelos de clasificación binaria descritos mayoritariamente en la literatura se enfocan exclusivamente en diferenciar entre hematíes normales y hematíes con malaria (25,35,36,47,48,64,67-69). Por lo tanto, es de esperar que este tipo de clasificadores generen una gran cantidad de falsos positivos al confundir hematíes con otros tipos de inclusiones como hematíes infectados con malaria, lo que los hace de dudosa utilidad en un entorno clínico. El clasificador desarrollado en este apartado pretende reproducir los clasificadores publicados en la literatura y demostrar su poca especificidad en muestras reales de pacientes, evidenciando la necesidad de desarrollar modelos con un mejor rendimiento en la

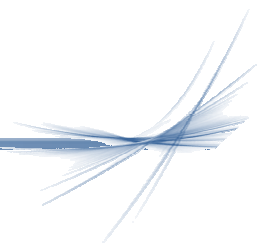
clasificación. Este apartado constituye la base para el desarrollo de los modelos propuestos en los objetivos específicos.

5.3.1.1 Entrenamiento del clasificador binario

Esencialmente, un clasificador usa un conjunto de características seleccionadas para predecir la clase a la que pertenece la imagen. Para ello, se evaluaron los siguientes algoritmos de aprendizaje automático:

- Máquinas de soporte vectorial: este tipo de clasificador busca el hiperplano que tenga la máxima distancia con los puntos que estén más cerca de él mismo, cuando los elementos son representados en un espacio p-dimensional.
- K vecinos más cercanos: es un modelo que clasifica un nuevo elemento en función de la mayoría de clases a las que pertenezcan sus k vecinos más cercanos, siendo k un parámetro del algoritmo.
- Análisis discriminante lineal: realiza una separación de las clases a través de funciones lineales de los descriptores que definen las clases. Puede considerarse como una generalización de la regresión logística para la predicción de más de dos clases.
- Bosques aleatorios (RF): implica la construcción al azar de una gran cantidad de árboles de decisión sobre un mismo conjunto de datos, y la decisión final de la clasificación es tomada a partir del cálculo de las predicciones ofrecidas por cada uno de los árboles que conforman el bosque según un criterio de mayoría de voto.
- Clasificador de Bayes Gaussiano (GNB): algoritmo que predice la probabilidad de pertenencia a una clase en base al Teorema de Bayes, también conocido como teorema de la probabilidad condicionada

Cada tipo de algoritmo de clasificación ha sido entrenado a través de sucesivos procedimientos de validación cruzada 5x en los que se han incrementado secuencialmente el número de descriptores ordenados por orden de relevancia. La técnica *Conditional Mutual Information Maximization* (CMIM) es un criterio que realiza un ordenamiento de las características de acuerdo a la capacidad que tienen éstas para predecir la clase a la que pertenecen, utilizando la información mutua condicional. CMIM no selecciona una característica similar a las ya recogidas, aunque de forma individual sea de gran alcance, puesto que no aporta información



adicional acerca de la clase a predecir. Por lo que garantiza un buen equilibrio entre independencia y discriminación. La diferencia principal de CMIM con respecto a otros criterios consiste en que prioriza la minimización de la redundancia entre las características seleccionadas sobre la información que éstas tienen de la clase. El objetivo detrás de esta metodología es obtener aquellas características que proporcionan información única de la clase (92). La aplicación de este método se realizó a través de la librería PyFeast v1.1 para Python (93).

El ajuste del modelo se ha realizado a través de un entrenamiento con validación cruzada 5x. Esta técnica divide el conjunto de entrenamiento en cinco subconjuntos, de los cuales cuatro son utilizados para realizar el entrenamiento propiamente dicho, y el quinto subconjunto es utilizado para su validación (ver Figura 31). Esta operación se repite cinco veces, en las que los subconjuntos reservados para la validación y para el entrenamiento van rotando por todo el conjunto de datos. Dado que el subconjunto de validación no se utiliza para el entrenamiento, se obtiene una visión no sesgada del comportamiento del modelo cuando se enfrenta con datos no conocidos.

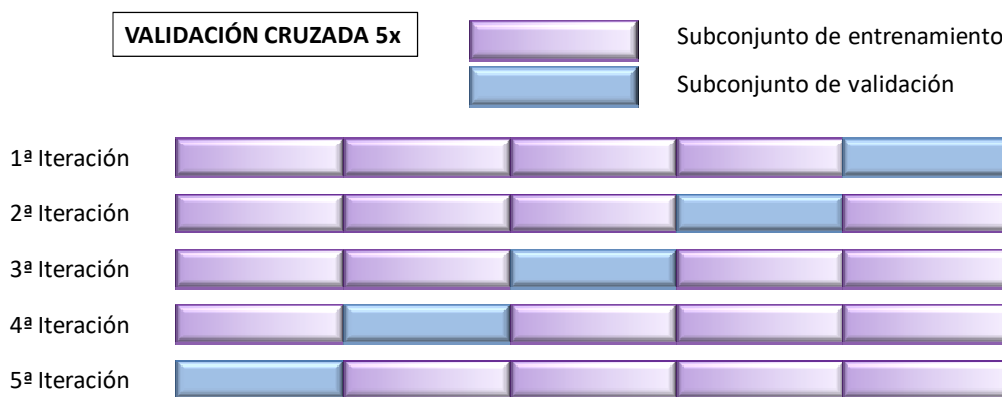


Figura 31. Esquema de entrenamiento mediante la validación cruzada 5x.

Este esquema de entrenamiento se realiza para todos los algoritmos matemáticos de clasificación (SVM, LDA, KNN, RF, GNB) y para diferentes subconjuntos de descriptores que se van incrementando secuencialmente mediante la técnica CMIM. El objetivo de este procedimiento de búsqueda iterativa es encontrar la mejor combinación de algoritmo de clasificación y número de características que proporcione los mejores resultados de exactitud global, definido como el porcentaje de imágenes bien clasificadas sobre el número total de imágenes. Una vez obtenido la mejor combinación, se procede al entrenamiento del modelo con todo el conjunto de entrenamiento.

5.3.1.2 Evaluación del clasificador binario

El modelo seleccionado se evalúa con un conjunto de imágenes nuevas para el modelo. Este conjunto de imágenes procede de muestras de frotis sangre periférica que no han sido utilizadas para el entrenamiento. El hecho de utilizar imágenes y frotis específicamente destinadas a la evaluación proporciona una idea realista de cómo se comportará el modelo en condiciones reales de trabajo.

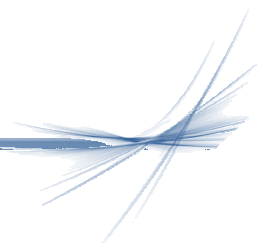
Se analizará el rendimiento en la clasificación del clasificador desarrollado a través de dos conjuntos de imágenes:

- Un conjunto de evaluación constituido por imágenes de hematíes normales y hematíes infectados con malaria que proceden de muestras de pacientes que no han sido previamente utilizados durante el entrenamiento. Esta valoración nos proporcionará un reflejo del rendimiento del clasificador cuando se enfrenta a muestras nuevas de pacientes.
- Un conjunto de evaluación constituido por imágenes de hematíes con otros tipos de inclusiones (HJ, PP, PB y PLT), lo que nos mostrará una aproximación del comportamiento de los analizadores publicados en la literatura cuando el clasificador desarrollado no ha sido entrenado específicamente para esas clases.

5.3.2 Sistema de clasificación secuencial

Esta es la propuesta original desarrollada en la tesis para la clasificación de hematíes normales, hematíes con malaria, así como hematíes con otros tipos de inclusiones. Esta propuesta responde a dos finalidades: elaborar un sistema de clasificación automática de las diferentes clases de hematíes, pero por otro lado, desarrollar un sistema que sea capaz de identificar hematíes con parásitos de malaria con una alta sensibilidad y especificidad. Para ello se consideran las siguientes clases de hematíes:

- Hematíes normales que no presentan ningún tipo de inclusión.
- Hematíes con parásitos de malaria.
- Hematíes con cuerpos de Howell-Jolly (HJ).
- Hematíes con cuerpos de Pappenheimer (PP).



- Hematíes con punteado basófilo (PB).
Hematíes con plaquetas alojadas en su superficie (PLT).

5.3.2.1 Entrenamiento del sistema de clasificación secuencial

Las diferentes clases de hematíes objeto de estudio tienen diferente peso clínico-morfológico. Por ejemplo, no tendrá la misma repercusión clínica el reconocimiento de un hematíe con presencia de malaria (patognomónico de paludismo, lo cual constituye un valor crítico de urgente notificación y unas medidas terapéuticas urgentes), que el reconocimiento de una plaqueta alojada sobre la superficie del hematíe, que no tiene valor ni repercusión alguna. Por este motivo, y con la finalidad de conseguir un reconocimiento más específico y efectivo, planteamos el diseño de un sistema constituido por tres módulos independientes que trabajarán de manera secuencial, cada uno de ellos destinado al reconocimiento de elementos específicos, evitando que el sistema confunda las diferentes inclusiones eritrocitarias con parásitos. La Figura 32 presenta la estructura del sistema y de los módulos que la componen, que se describen a continuación:

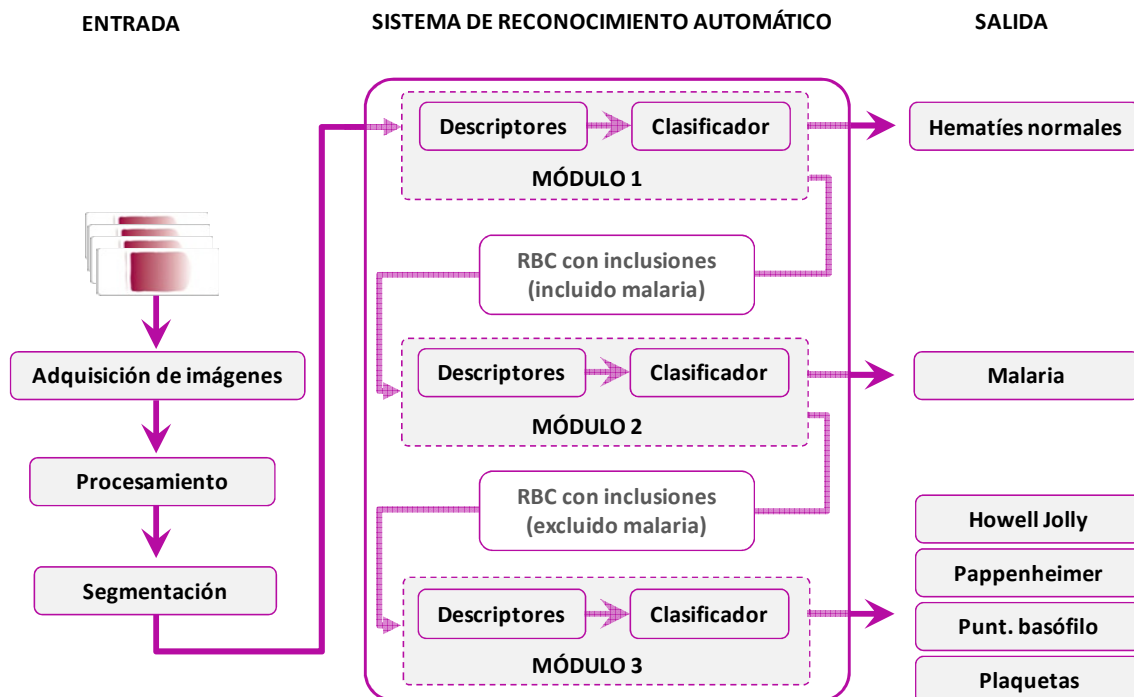


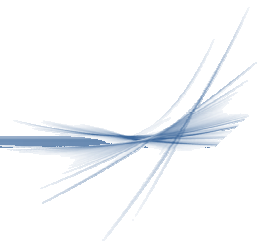
Figura 32. Diseño del sistema de clasificación basado en métodos de machine learning mediante el uso de descriptores obtenidos a partir de la segmentación de las imágenes. Estructura y funcionamiento de sus módulos.

- **Módulo 1:** Las imágenes de entrada serán reconocidas por un **primer clasificador** encargado de la diferenciación entre hematíes normales y hematíes con cualquier tipo de inclusión. Por una parte, la malaria, HJ, PP, PB y PLT, son elementos que comparten una característica común: todos son inclusiones y pertenecen a clases de hematíes que tienen un contenido en su interior. Por otra parte, es importante hacer un buen reconocimiento de los hematíes normales, no solo con la finalidad de obtener pocos falsos positivos, también por el hecho de que en la mayoría de muestras existe una frecuencia relativa de hematíes normales muy superior a la de hematíes con inclusiones. Por lo tanto, es importante enfocar como primer paso la diferenciación de los hematíes normales respecto al resto.

- **Módulo 2:** Las imágenes que en el primer clasificador hayan sido reconocidas como hematíes con inclusiones, constituirán los elementos de entrada a un **segundo clasificador** encargado de la diferenciación entre hematíes infectados con malaria respecto a hematíes con cualquier otro tipo de inclusión. Dada la importancia clínica de la malaria, es importante desarrollar un clasificador específico para el reconocimiento de hematíes infectados, con el objetivo de encontrar las características con mayor relevancia en la diferenciación de este tipo de hematíes. Dado que la presencia de un solo hematíe infectado con parásitos de malaria es patognomónico de paludismo, es importante que el clasificador desarrollado posea una elevada sensibilidad.

- **Módulo 3:** Aquellas imágenes que en el segundo módulo han sido identificadas como hematíes con otros tipos de inclusiones (excluyendo la malaria), entrarán en un **tercer clasificador** de reconocimiento que identificara HJ, PP, PB y PLT.

Para cada uno de los tres módulos seguimos el mismo procedimiento de entrenamiento que el descrito en el apartado anterior. El propósito consiste en entrenar por separado cada clasificador (ver Figura 32) para ser altamente específico para cada tarea particular de reconocimiento. Cada uno de los tres clasificadores que componen el sistema final ha sido entrenado con diferentes algoritmos de clasificación (SVM, LDA, KNN, RF, y GNB) a través de procedimiento iterativo de validación cruzada 5x. Del mismo modo que en el clasificador binario, en cada iteración se han incrementado secuencialmente el número de descriptores ordenados por orden de relevancia según el criterio CMIM (ver Apartado 5.3.1.1). El objetivo de este procedimiento es encontrar para cada módulo la mejor combinación de algoritmo de



clasificación y número de características que proporcione los mejores resultados de exactitud global, definido como el porcentaje de imágenes bien clasificadas sobre el número total de imágenes. Una vez encontrada la mejor combinación, se procede al entrenamiento de cada módulo con todo el conjunto de imágenes de entrenamiento.

5.3.2.2 Evaluación del sistema de clasificación secuencial

El modelo seleccionado se evalúa con un conjunto de imágenes nuevas para el modelo, que no han sido utilizadas para el entrenamiento y que proceden de frotis completamente nuevos para el clasificador. Este conjunto de imágenes de evaluación ha sido dividido, a su vez, en dos grupos de imágenes: un conjunto de *test* y un conjunto de diagnóstico, que realizarán una evaluación del sistema de reconocimiento automático final mediante dos enfoques diferentes:

- El conjunto de *test* se usará para la evaluación del rendimiento en la clasificación de imágenes individuales de hematíes. Este conjunto de imágenes procede de 18 frotis de sangre periférica que no han sido utilizadas para el entrenamiento.
- El conjunto de diagnóstico se usará para comprobar la habilidad del sistema de reconocimiento en identificar pacientes infectados con malaria a través del análisis de todo un frotis en conjunto. Estas imágenes proceden de 18 frotis de sangre periférica que no fueron utilizados en el entrenamiento y que proceden de muestras diferentes a las del conjunto de *test*. Del total, 8 muestras proceden de pacientes infectados con malaria, mientras que las otras 10 proceden de pacientes no infectados, en los que aparecen también otro tipo de inclusiones eritrocitarias. A diferencia del conjunto de *test*, el objetivo de esta evaluación consiste en analizar todas las imágenes que proceden de un mismo frotis. Ante la presencia de un único hematíe que queda reconocido por el sistema como un hematíe con parásitos de malaria, el frotis queda automáticamente etiquetado como “infectado”.

Los indicadores utilizados para la evaluación del modelo se basan en las siguientes métricas:

Exactitud: Porcentaje de imágenes correctamente clasificadas respecto al total de imágenes.

$$Exactitud = \frac{VP + VN}{VP + FP + FN + VN}$$

Valor Predictivo Positivo: Porcentaje correctamente identificadas como malaria respecto al total de imágenes reconocidas como malaria.

$$VPP = VP/(VP + FP)$$

Sensibilidad: Porcentaje de hematíes con malaria que son correctamente identificados.

$$Sensibilidad = VP/(VP + FN)$$

Especificidad: Porcentaje de hematíes no infectados que son correctamente identificados.

$$Especificidad = VN/(VN + FP)$$

F score: Promedio entre VPP y sensibilidad. Ofrece una visión global de los hematíes con malaria que son correctamente reconocidos, pero teniendo en cuenta tanto los FN como los FP.

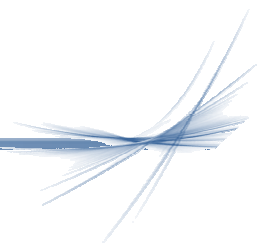
$$F\ Score = 2 * (Sensibilidad * VPP) / (Sensibilidad + VPP)$$

VP: Verdadero Positivo; VN: Verdadero Negativo; FP: Falso Positivo; FN: Falso Negativo.

5.4 RESULTADOS

5.4.1 Segmentación

Las etapas de procesamiento y segmentación permitieron obtener imágenes individuales de hematíes a partir de las imágenes de campo obtenidas manualmente con el microscopio. Además, el algoritmo de Watershed aplicado permitió no solo diferenciar cada uno de los hematíes como elementos individualmente diferentes, también nos proporcionó una metodología para diferenciar hematíes solapados. Es importante establecer un delicado equilibrio en la sensibilidad de segmentación del algoritmo de Watershed. En nuestro caso, los mejores resultados se obtuvieron a través del cálculo de la transformada de la distancia euclídea. Mientras este solapamiento no exceda de un determinado nivel en el que todavía pueden distinguirse los halos centrales de ambos hematíes, el algoritmo de Watershed será capaz de identificar por separado dos ROI diferentes. Sin embargo, un mayor grado de acoplamiento entre hematíes solapados tienden a identificarse como un único elemento. Estos elementos pueden ser reconocidos en la mayoría de los casos mediante un filtrado del área de los píxeles, ya que los hematíes que pertenecen una agrupación pueden ser difíciles de valorar.



La Figura 33 muestra el efecto de estos ajustes ante la presencia de hematíes con diferente grado de solapamiento. Un grupo de dos hematíes parcialmente solapados, cuyos puntos de intersección muestran mínimo grado de concavidad permiten ser separados mediante los parámetros aplicados en el algoritmo (Figura 33.B). Pero no es efectivo en grupos de hematíes con un grado mayor de solapamiento. En la Figura 33.C puede apreciarse a partir de qué punto los hematíes son reconocidos como una unidad. De los tres hematíes que forman el grupo, los dos inferiores han permanecido unidos con la metodología utilizada. Esta metodología demuestra ser útil también en solapamientos de hematíes que se encuentran en los límites de la imagen, ya que estos son eliminados al tratarse de imágenes parciales que no permiten la valoración correcta de su estructura. Sin embargo, un hematíe que no linda con los bordes pero que se encuentra acoplado a un hematíe que sí limita con los cantos, es segmentado correctamente y tenido en cuenta para su análisis (Figura 33.A).

La metodología utilizada en este trabajo mostró un rendimiento de segmentación de 97,4%, considerando una correcta segmentación cuando el resultado de la máscara conservó la forma original de su RBC.

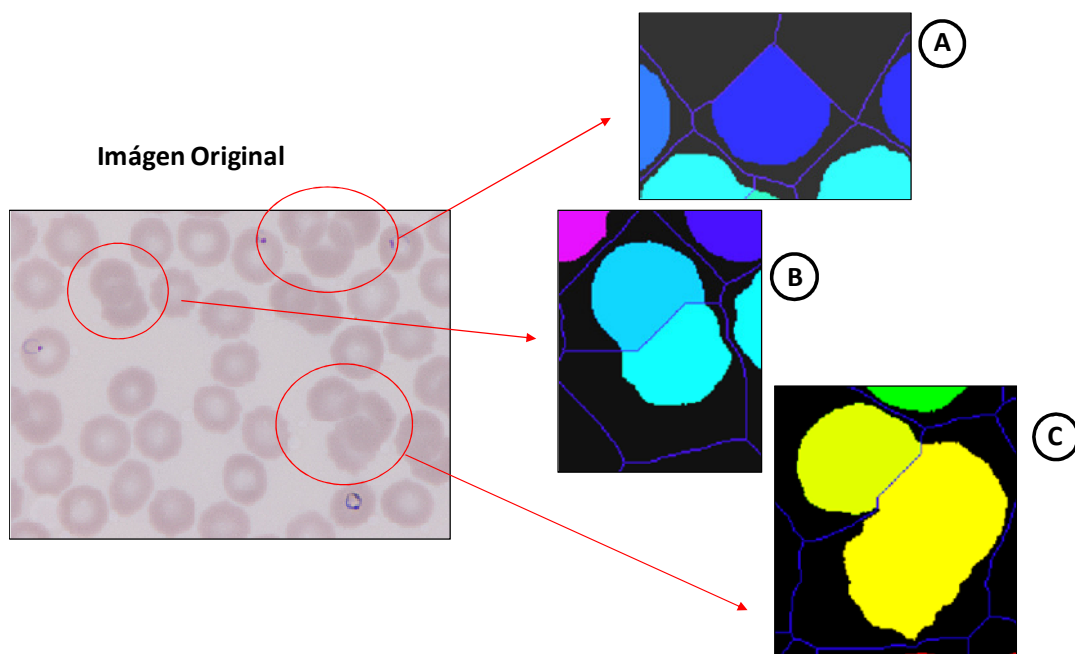


Figura 33. Resultado del algoritmo de Watershed sobre hematíes de una imagen de campo con diferente grado de solapamiento.

Uno de los inconvenientes de la segmentación mediante la metodología de Watershed es que en ocasiones proporciona una sobresegmentación de algunos elementos. Esto hace que algunos hematíes queden “divididos” artificialmente en diferentes porciones y reconocidos por lo tanto como elementos diferentes. Esta sobresegmentación ha aparecido en un 2,5% de las imágenes obtenidas. Para evitar este error se ha optado por el uso de la función “*convex hull*”, la cual busca “rellenar” los píxeles de la imagen binaria que contiene el polígono convexo más pequeño que rodea todos los píxeles blancos de la imagen. Un ejemplo de la aplicación de esta función puede observarse en la Figura 34.G. A partir de la imagen original puede apreciarse como el algoritmo de Watershed ha dividido al hematíe por la mitad en dos secciones diferentes, fruto de una sobresegmentación. Sin embargo, al aplicar la función *convex hull*, puede obtenerse la máscara completa del hematíe como un único elemento sin que se haya perdido la morfología de la imagen original. Por otro lado, como puede verse en la Figura 34.A-F, la implantación de esta función no ha tenido ningún efecto perceptible en aquellos hematíes que han sido segmentados correctamente desde un inicio.

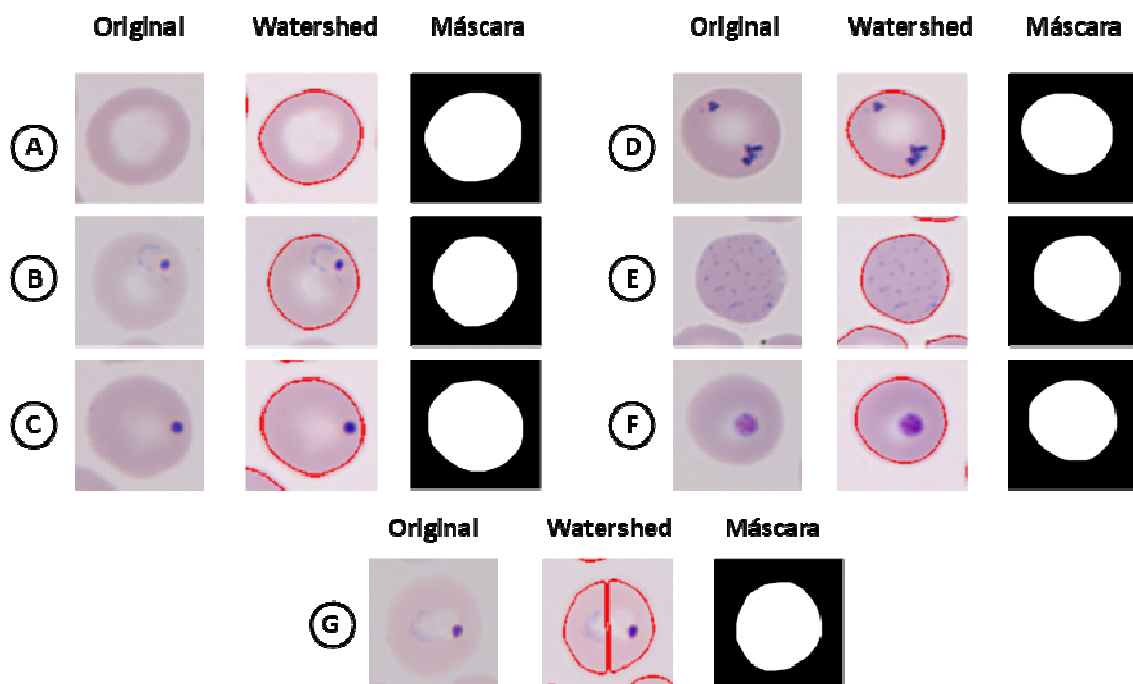
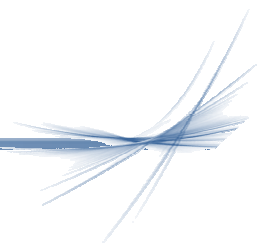


Figura 34. Ejemplos de la metodología de segmentación. A-F: Imagen original, contorno detectado mediante el algoritmo de Watershed, y máscara resultante en ejemplos de cada una de las inclusiones eritrocitarias. G: Obtención de la máscara binaria en un hematíe que ha sido sobresegmentado.



Añadir a la metodología de segmentación un último paso con la adición de la función *convex hull* ha permitido aumentar el rendimiento de la segmentación a un 99,9%. Como resultado, se obtuvieron un total de 15.660 imágenes de hematíes individuales provenientes de 87 frotis de sangre periférica. Estas muestras de frotis, con sus respectivas imágenes, fueron divididas en los siguientes subconjuntos: Las imágenes de 51 frotis para el entrenamiento, mientras que se reservaron 18 frotis para la evaluación, tanto para evaluar la efectividad del modelo para la clasificación de hematíes individuales, como para comprobar el rendimiento diagnóstico sobre muestras de frotis completos infectadas con malaria. El número total de imágenes para cada tipo de RBC se muestra en la Tabla 1.

		CONJUNTO DE IMÁGENES			
		Entrenamiento	Evaluación		Total
			Imagen individual	Por frotis	
Número de frotis		51	18	18	87
Número de imágenes	nRBC	1.276	733	8.443	10.452
	MAL	933	187	71	1.191
	HJ	578	539	76	1.193
	PP	698	256	72	1.026
	PB	701	359	65	1.125
	PLT	447	162	64	673
Total		4.633	2.236	8.791	15.660

Tabla 1. Número de frotis e imágenes individuales que constituyen los diferentes conjuntos de entrenamiento y evaluación. nRBC: Hematíes normales; MAL: hematíes con parásitos de malaria; HJ: Hematíes con cuerpos de Howell-Jolly; PP: hematíes con cuerpos de Pappenheimer; PB: hematíes con punteado basófilo; PLT: hematíes con plaquetas alojadas en su superficie.

5.4.2 Clasificación

Tal como se ha descrito en el Apartado 5.3, en este capítulo se han desarrollado dos clasificadores que responden a dos finalidades diferentes:

- Un clasificador que diferencia exclusivamente entre hematíes normales y hematíes infectados con malaria, con la finalidad de valorar su rendimiento cuando se encuentra con hematíes con otros tipos de inclusiones. Los resultados de este clasificador se describen en el Apartado 5.4.2.1.
- El objetivo principal de este capítulo, que consiste en un sistema de reconocimiento automático de hematíes normales, hematíes con malaria y hematíes con otros tipos de inclusiones, con la finalidad de obtener un sistema fiable de apoyo al laboratorio clínico. Los resultados de este modelo se describen en el Apartado 5.4.2.2.

5.4.2.1 Clasificador binario para hematíes normales *versus* malaria

En este caso se realizó un entrenamiento de los principales algoritmos matemáticos de clasificación a través de un entrenamiento con validación cruzada 5x, utilizando solamente el conjunto de imágenes de hematíes normales y de hematíes infectados con malaria.

En la Figura 35 se muestran los resultados de clasificación en términos de exactitud global del promedio de las cinco validaciones realizadas en las sucesivas iteraciones en las que se ha ido incrementando el número de descriptores con mayor relevancia. Los resultados muestran que la tarea de diferenciar hematíes normales de hematíes infectados con malaria parece ser de fácil consecución. Con muy pocos descriptores, todos los algoritmos de clasificación muestran una exactitud en el conjunto de validación superior al 99%, por lo que cualquier algoritmo podría ser escogido para realizar esta tarea. Sin embargo, se observa una exactitud global ligeramente superior (99,84%) mediante el uso de una SVM con los siete descriptores con mayor relevancia según el criterio CMIM. Cuando este modelo es enfrentado al conjunto de evaluación, en la matriz de confusión puede apreciarse que todos los hematíes infectados con malaria son reconocidos correctamente, mientras que solo un 0,3 % de los hematíes normales son reconocidos erróneamente como hematíes infectados con malaria (ver Tabla 2).

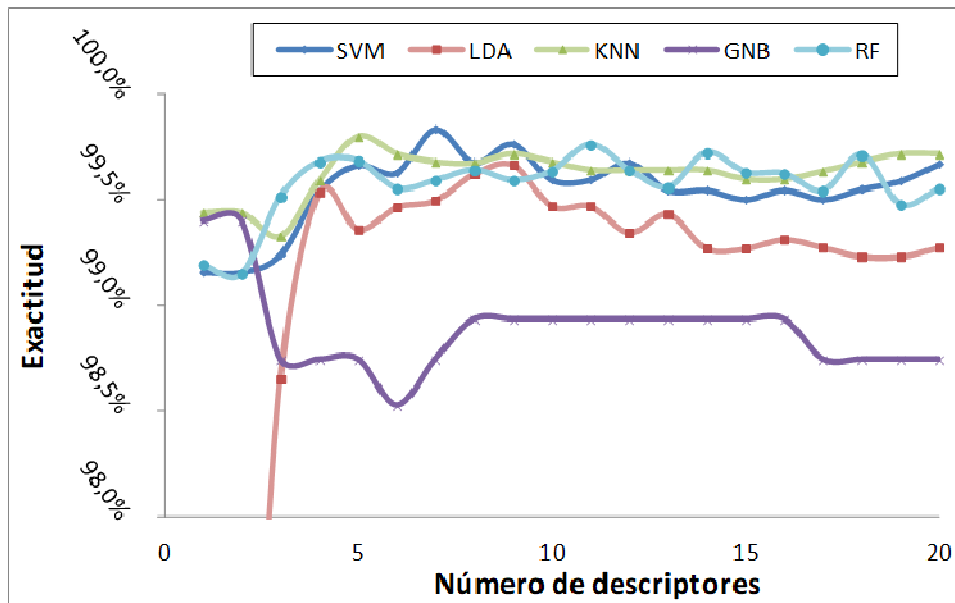


Figura 35. Resultados del entrenamiento de los diferentes algoritmos de clasificación. Los valores de exactitud corresponden a la clasificación del conjunto de validación realizado sobre el entrenamiento. La gráfica muestra los valores de exactitud obtenidos para cada conjunto de descriptores utilizados en el entrenamiento. SVM: máquinas de soporte vectorial; LDA: análisis discriminante lineal; KNN: K vecinos más cercanos; GNB: clasificador de Bayes Gaussiano; RF: bosques aleatorios.

	NORM	MAL
NORM	99,7	0,3
MAL	0	100

	NORM	MAL
HJ	0	100
PP	0,7	99,3
PB	24,4	75,6
PLT	16,1	83,9

Tabla 2. Resultados (en porcentaje) de la evaluación del modelo SVM que diferencia entre hematíes normales y hematíes infectados con malaria. Izquierda: conjunto de evaluación de hematíes normales y malaria. Derecha: evaluación del clasificador con el resto de hematíes que contienen cualquier otro tipo de inclusión diferente a la malaria (HJ: hematíes con cuerpos de Howell-Jolly, PP: hematíes con cuerpos de Pappenheimer, PB: hematíes con punteado basófilo, PLT: hematíes con plaquetas alojadas en su superficie).

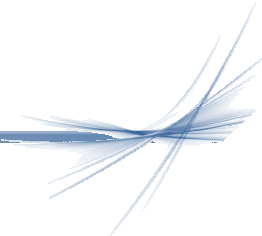
En una etapa posterior, este modelo ha sido sometido al total de imágenes de hematíes que contienen otros tipos de inclusiones. Dado que el modelo ha sido específicamente entrenado

para el reconocimiento de hematíes normales y hematíes infectados con malaria, otro tipo de inclusiones son elementos completamente desconocidos para el modelo. Los resultados de esta clasificación muestran que todas las inclusiones son reconocidas incorrectamente como hematíes con parásitos de malaria en un porcentaje superior al 75% (ver Tabla 2). Visto de otra manera, solo el 24,4% en el caso de PB y solo el 16% en el caso de PLT son reconocidos como hematíes normales. Sin embargo, el resto de inclusiones han sido reconocidas como malaria, lo que sugiere que HJ, PP, PB y PLT son más similares morfológicamente a la malaria que a los hematíes normales, y lo que también puede explicar que los modelos publicados en la literatura adolezcan de un comportamiento similar.

Se ha realizado un estudio de los descriptores más relevantes a la hora de diferenciar hematíes normales de hematíes con parásitos de malaria. De todos los descriptores analizados en el Apartado 5.2, los que presentan mayor relevancia en la clasificación entre hematíes normales y hematíes con malaria son los siguientes:

- La asimetría de los valores los píxeles del canal u, que procede del espacio de color YUV.
- El rango intercuartílico del canal K del espacio de color CMYK.
- El énfasis de corto recorrido del canal R del espacio de color RGB.
- La suma de la entropía del canal R del espacio de color RGB.

De todos ellos, la asimetría de los píxeles del canal U ha mostrado ser el descriptor con mayor relevancia. En la representación mediante el diagrama de cajas de este descriptor se observa que el 50% de los valores centrales procedentes de hematíes normales tienen una asimetría entre -0,19 y - 0,66, mientras que el 50% de los hematíes infectados con malaria tienen una asimetría del 2,69 y 4,58. Al tratarse de valores que no siguen una distribución normal, se ha realizado un test de U de Mann Whitney en el que se observa una diferencia altamente significativa entre ambas distribuciones de valores. Aunque este clasificador solo reconoce entre hematíes normales y hematíes con parásitos de malaria, en la Figura 36 se muestran también los valores de este descriptor en las imágenes de hematíes que contienen otros tipos de inclusiones (este grupo se muestra en gris en la figura). Las imágenes que contienen otras inclusiones eritrocitarias muestran unos valores de asimetría que se solapan claramente con la de las imágenes de los hematíes con malaria, con un valor de rango intercuartílico entre 1,86 - 4,80. En el resto de los descriptores más relevantes encontramos situaciones completamente equiparables a las encontradas en el caso de la asimetría. La distribución de valores de los hematíes normales difiere estadísticamente de la de los hematíes infectados con malaria,



mientras que los hematíes con cualquier tipo de inclusión muestran valores muy similares a los de la malaria en la mayoría de descriptores, aunque en algún caso como en la suma de la entropía también se solapan con la población de hematíes normales, lo que puede significar un problema para su correcta identificación.

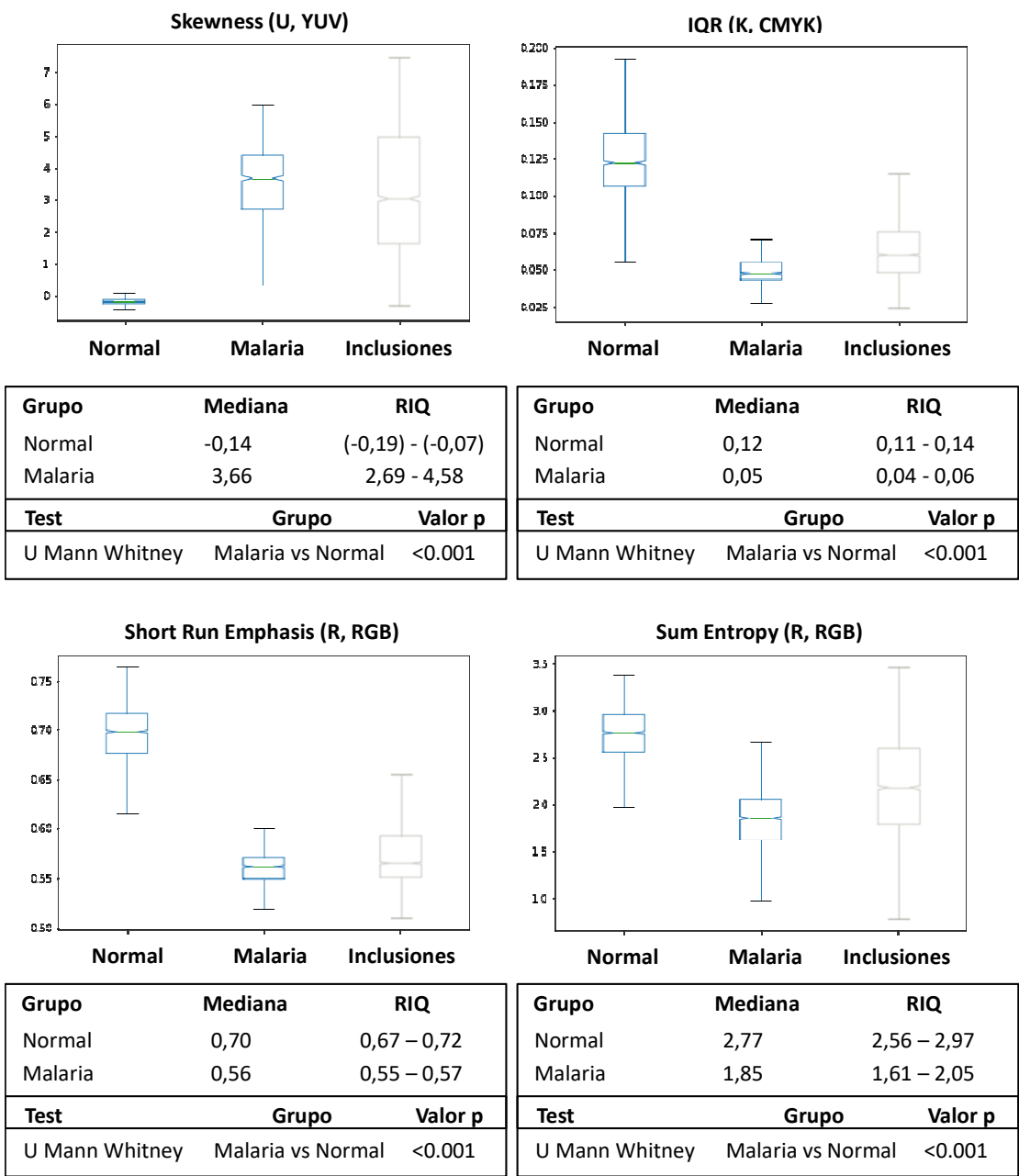


Figura 36. Descriptores más relevantes en el clasificador SVM de hematíes normales y hematíes con malaria.

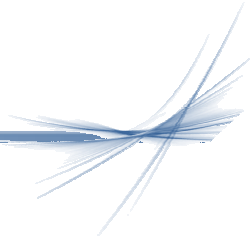
5.4.2.2 Sistema de Clasificación secuencial

Una vez demostrado que los clasificadores basados en la discriminación entre hematíes normales y hematíes infectados con malaria no son eficaces cuando aparecen otros tipos de inclusiones, se procedió al desarrollo de un modelo efectivo capaz de diferenciar todos los tipos de inclusiones. Este sistema final combina tres módulos individuales que funcionan secuencialmente y cada uno de los cuales ha sido entrenado por separado para una tarea concreta mediante la búsqueda de una combinación adecuada del número de descriptores con mayor relevancia con un algoritmo de clasificación adecuado. Los resultados principales del entrenamiento se presentan en la Figura 37.

El primer módulo del sistema final es el encargado de diferenciar entre hematíes normales y hematíes con cualquier tipo de inclusión (incluida la malaria). Durante el entrenamiento de este módulo se observaron unos resultados muy similares a los de la sección anterior, prácticamente todos los algoritmos de clasificación mostraron muy buenos resultados de clasificación en términos de exactitud global (superiores al 97% en todos los casos) con el uso de muy pocos descriptores. Sin embargo, SVM junto con los siete descriptores más relevantes mostraron los mejores resultados de exactitud (99,12%), teniendo en cuenta las imágenes del conjunto de validación en el proceso de la validación cruzada.

El segundo módulo del sistema final tiene como entrada los hematíes que han sido reconocidos como hematíes con cualquier tipo de inclusión en el primer módulo. Su función se basa en reconocer si las inclusiones corresponden a parásitos de la malaria o no. Los resultados de este entrenamiento muestran que este segundo módulo se enfrenta a una tarea mucho más compleja que en el primer módulo, ya que para obtener los mejores resultados de la clasificación es necesario un gran número de descriptores. Tampoco todos los algoritmos de clasificación parecen igualmente adecuados para esta tarea, siendo LDA y SVM los métodos que mostraron una mayor exactitud con diferencia. La combinación de LDA con las 610 características más relevantes mostró mejor rendimiento en el reconocimiento del conjunto de validación, con un valor de exactitud global del 96,52%.

Los hematíes que en el segundo módulo son reconocidos como hematíes con otro tipo de inclusiones diferente a los de la malaria constituyen la entrada a un tercer módulo que será el encargado de diferenciar si se trata de HJ, PP, PB o PLT. Los resultados obtenidos en el entrenamiento de este tercer módulo muestran valores muy similares a los obtenidos durante



el entrenamiento del segundo módulo: se mantiene la superioridad de LDA y SVM, mientras que KNN, RF y GNB muestran los peores resultados de clasificación. La combinación de LDA junto con los 700 descriptores más relevantes muestra los mejores resultados de exactitud (96,91%), como se observa en la Figura 37.

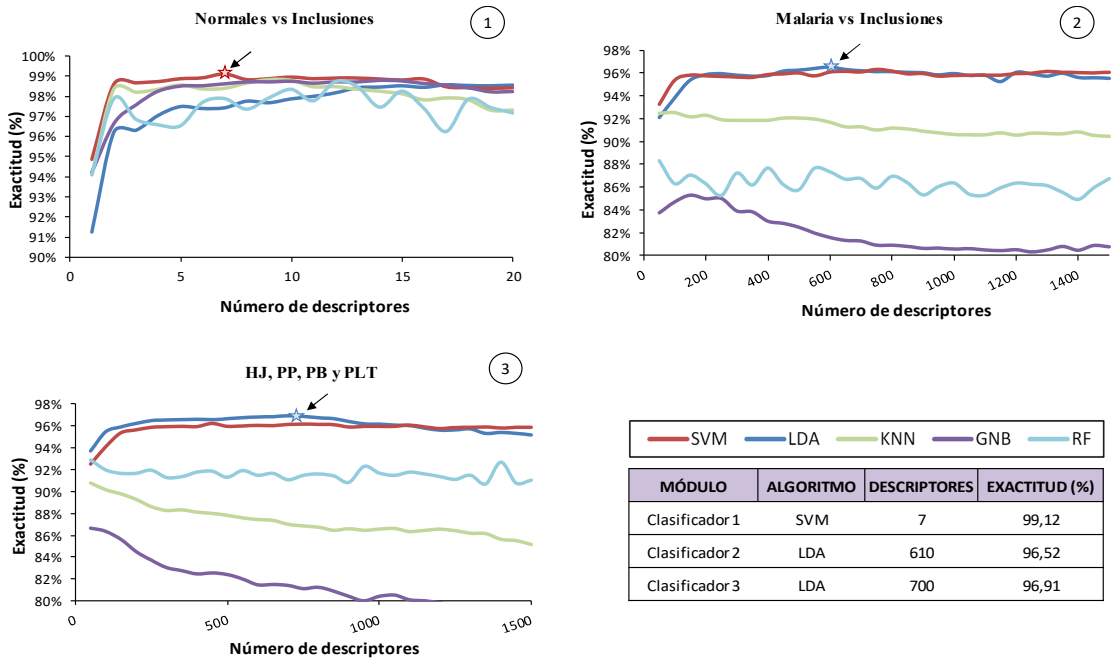


Figura 37. Resultados del entrenamiento de cada uno de los módulos que componen el sistema final de reconocimiento de inclusiones eritrocitarias. Los valores de exactitud corresponden a la clasificación del conjunto de validación realizado sobre el entrenamiento. Las gráficas muestran los valores de exactitud obtenidos para cada conjunto de descriptores utilizados en el entrenamiento. SVM: máquinas de soporte vectorial; LDA: análisis discriminante lineal; KNN: k vecinos más cercanos; GNB: clasificador de Bayes Gaussiano; RF: bosques aleatorios.

Cuando analizamos los descriptores con mayor relevancia en cada uno de los módulos del sistema final, encontramos que en el primer módulo (encargado de la diferenciación entre hematíes normales y hematíes con cualquier tipo de inclusión, incluido la malaria) los descriptores más relevantes han sido los siguientes:

- La asimetría del canal U del espacio de color YUV.
- El énfasis de dependencia de corta distancia del canal Y del espacio de color YCbCr.
- La correlación del canal Cb del espacio de color YCbCr.
- El coeficiente de correlación máxima del canal Cb del espacio de color YCbCr.

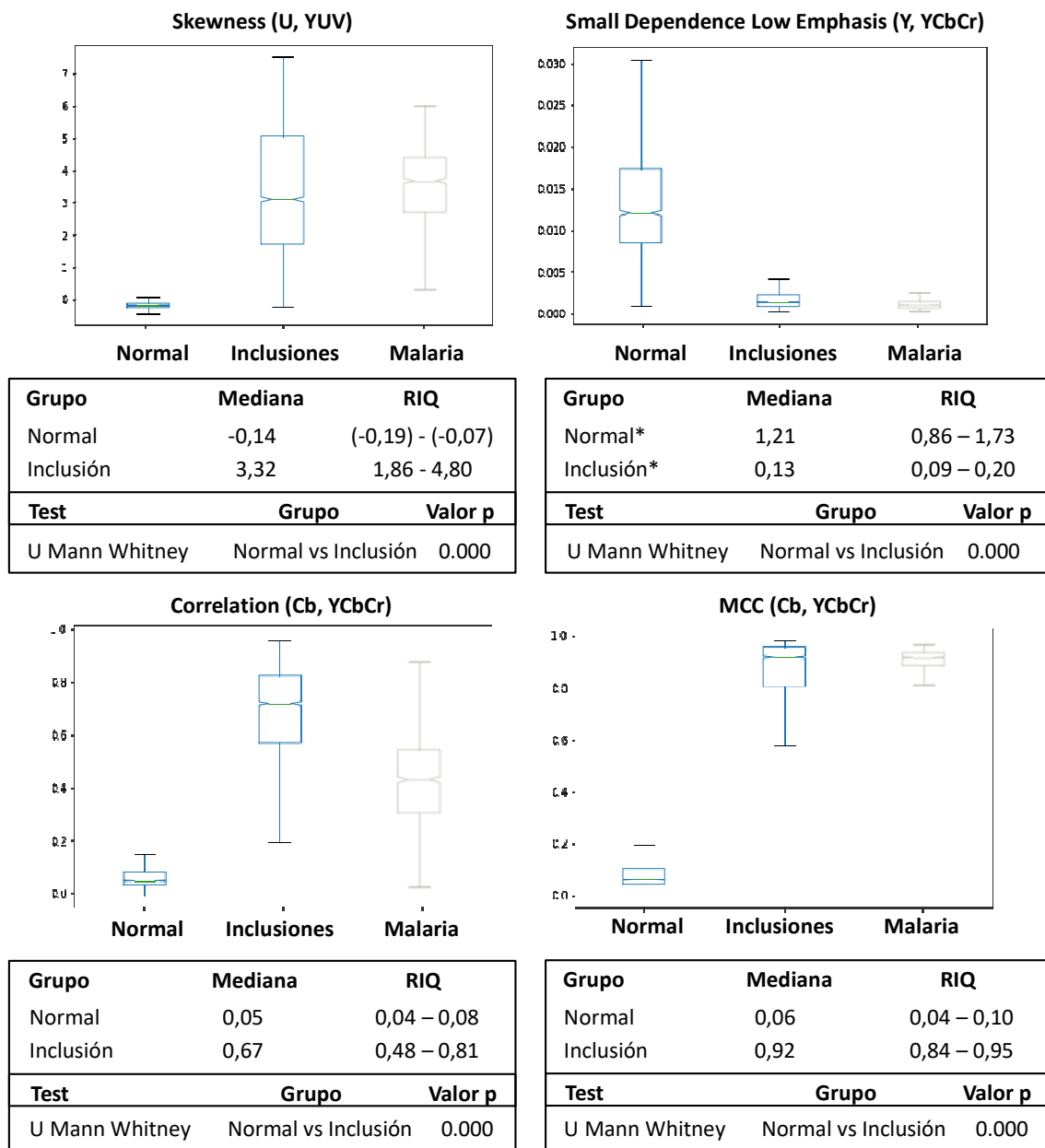
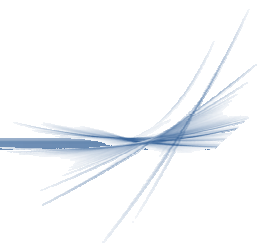


Figura 38. Descriptores más relevantes en el Módulo 1 encargado de reconocer entre hematíes normales y hematíes con cualquier tipo de inclusión.

El descriptor más relevante de este módulo coincide con el descriptor del clasificador binario desarrollado para la diferenciación entre hematíes normales y hematíes infectados con malaria (ver Apartado 5.4.2.1). Y esto en cierta manera tiene sentido, ya que los valores de la asimetría del canal U muestran una distribución muy similar en las imágenes de hematíes infectados con malaria o con cualquier otro tipo de inclusión, y con una diferencia muy notable respecto de



los hematíes normales. En la Figura 38 puede observarse con más detalle los valores de las medianas y rango intercuartílico entre las poblaciones de hematíes normales y hematíes con inclusiones, que muestran una diferencia estadísticamente significativa. Aunque no forma parte de los grupos que intervienen en la clasificación del primer módulo, se muestra como información adicional la distribución de valores de los hematíes infectados con malaria (gris en la figura). En prácticamente la mayoría de los descriptores que intervienen en el primer módulo, la malaria muestra valores muy similares a los del total de inclusiones. En todos los casos, se observan diferencias muy significativas entre los hematíes normales y los hematíes con cualquier tipo de inclusión. Las diferencias entre ambas poblaciones son tan notables que permiten crear un clasificador de muy buen rendimiento con muy pocos descriptores.

Los espacios de color que mayor relevancia han tenido en el desarrollo de este primer módulo han sido YUV y YCbCr. A modo de ejemplo, en la Figura 39 se muestra la representación de imágenes de diferentes hematíes en el canal U del espacio YUV y en el canal Cb del espacio YCbCr. Tanto en el canal U como en el canal Cb se obtienen imágenes en las que quedan resaltadas las inclusiones eritrocitarias, aunque con ligeras diferencias en ambos casos. No es de extrañar que la asimetría del canal U sea el descriptor más relevante para la diferenciación entre hematíes normales y hematíes con cualquier tipo de inclusión, ya que en este canal todas las inclusiones quedan muy resaltadas.

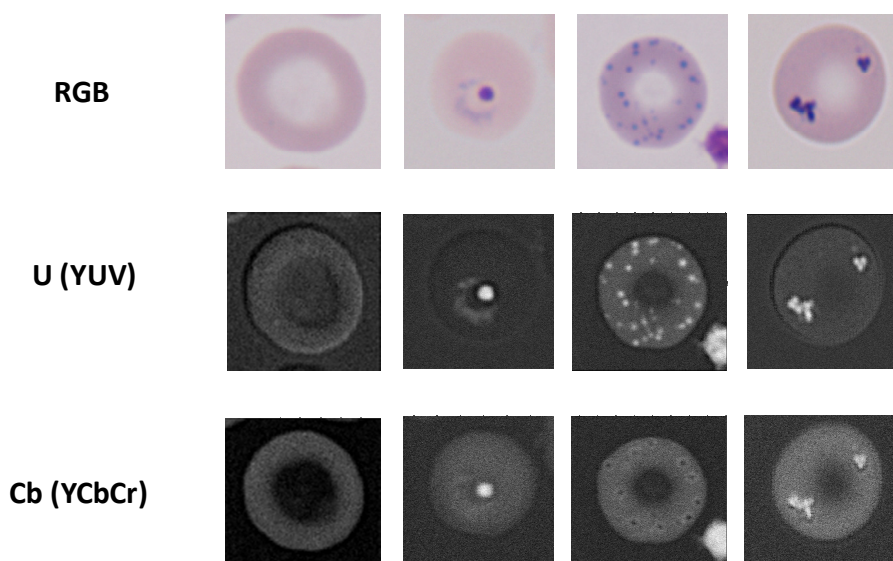


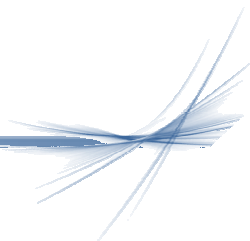
Figura 39. Representación de diferentes tipos de hematíes en el canal U del espacio de color YUV y en el canal Cb del espacio de color YCbCr.

La asimetría en este canal mide la forma de la distribución de los valores de los píxeles de la imagen. En un hematíe normal apenas se observan píxeles blancos (valores elevados), los valores tienen un aspecto muy homogéneo, y esto hace que su representación en un histograma tenga un aspecto mucho más simétrico, lo que explicaría que los valores de este descriptor en los hematíes normales tengan valores cercanos a cero. Sin embargo, en el resto de hematíes que presentan inclusiones se observan grupos de píxeles con valores más elevados (de aspecto muy claro) sobre un fondo de hematíes que muestran una mayoría de píxeles oscuros. La representación del histograma en estos hematíes se desplaza por lo tanto hacia la derecha, lo que le confiere una asimetría positiva a la curva de distribución. Por esa razón los hematíes con inclusiones muestran valores de asimetría variables pero muy superiores a la de los hematíes normales. Por otra parte, aunque el canal Cb también resalta la presencia de inclusiones, no lo realiza en el mismo sentido que el canal U. En la Figura 39 se observa como el citoplasma del trofozoito de la malaria y las pequeñas inclusiones del punteado basófilo quedan más atenuadas que lo que se observa en el canal U, quedando solo resaltado aquellas inclusiones que muestran un mayor grado de solidez. El citoplasma de los hematíes tampoco se representa de la misma manera, quedando ligeramente más resaltado en el canal Cb que en el canal U. Como consecuencia, la diferencia entre los valores de los píxeles de la inclusión respecto del hematíe donde se encuentra alojada es menor en el canal Cb que en el U. Probablemente por eso el canal U ha quedado en el descriptor con mayor relevancia para la diferenciación que lleva a cabo el primer módulo.

En el segundo módulo del sistema final (encargado de la diferenciación entre hematíes con parásitos de malaria y hematíes con cualquier otro tipo de inclusión) los descriptores más relevantes han sido los siguientes:

- La varianza del canal H del espacio de color LCH.
- La entropía de zona del canal U del espacio de color YUV.
- El énfasis de área pequeña del canal B del espacio de color LAB.
- El énfasis de corta dependencia de valores bajos de intensidad del canal Cb del espacio de color YCbCr.

En los descriptores más relevantes que intervienen en el segundo módulo (ver Figura 40) se puede observar que, aunque existen diferencias estadísticamente significativas entre los valores de los descriptores de los hematíes infectados con malaria respecto a los hematíes que contienen otros tipos de inclusiones, el grado de solapamiento entre las dos poblaciones es



mayor que el observado en los descriptores más relevantes del primer módulo (Figura 38). Esto no es de extrañar si tenemos en cuenta que para el desarrollo de un segundo módulo, que tenga un rendimiento óptimo en la clasificación, es necesario utilizar alrededor de 600 descriptores.

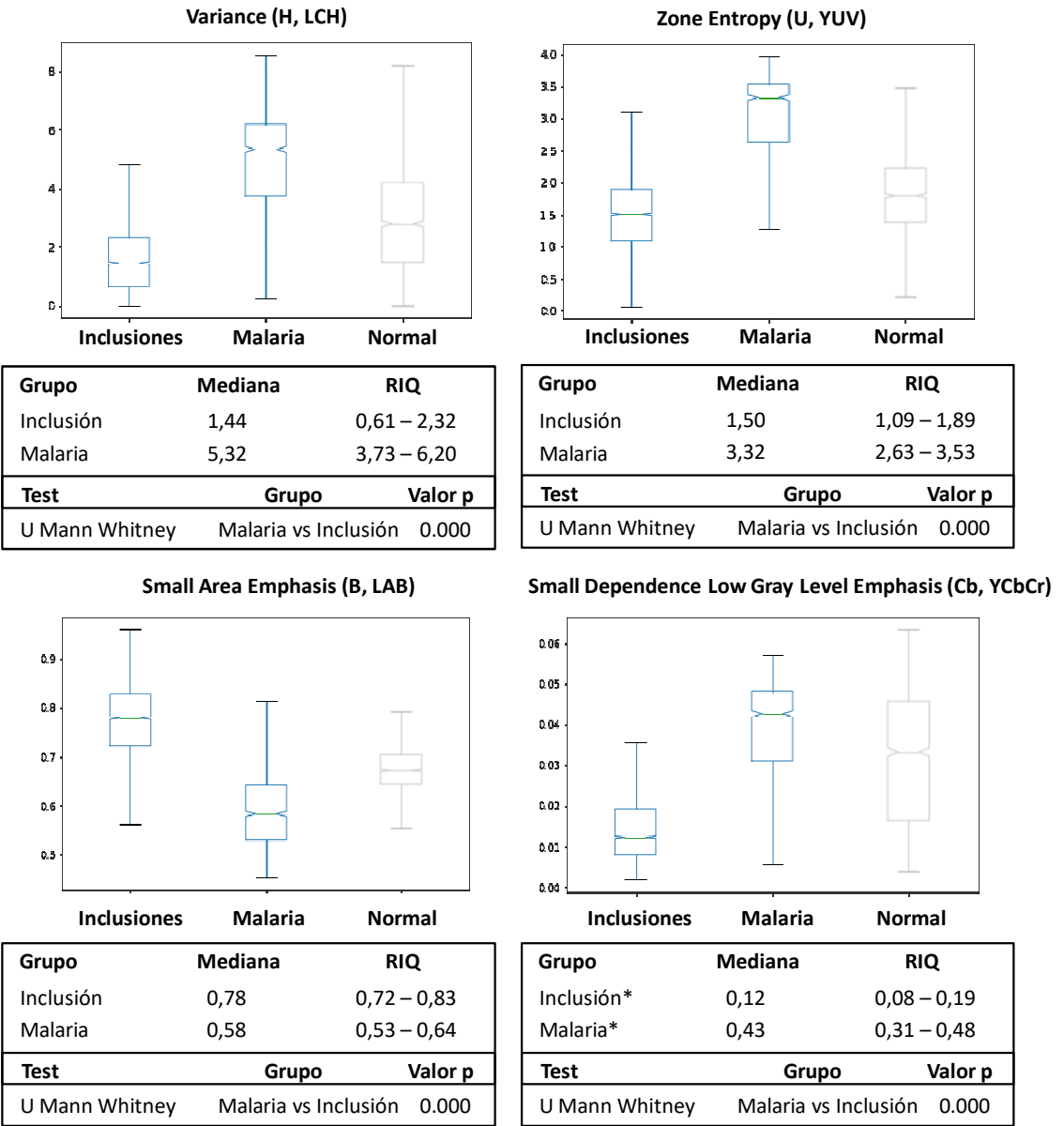


Figura 40. Descriptores más relevantes en el Módulo 2 encargado de reconocer entre hematíes con malaria y hematíes con cualquier otro tipo de inclusión.

En la Figura 40 puede apreciarse con más detalles los estadísticos de ambas poblaciones de hematíes en cada uno de los descriptores, el valor de significación al compararlas, y la representación de los valores de la población de hematíes normales (representada en gris) como información adicional. Es de destacar que la población de hematíes normales muestra unos valores en estos descriptores que se solapan con los de la malaria y con los de otras inclusiones. Esto demuestra que intentar abordar directamente el problema de la clasificación entre estos tres grupos es difícil, ya que estos descriptores no sirven para diferenciar hematíes normales. En el análisis de los descriptores del segundo módulo aparecen dos canales de color que habíamos visto previamente, como el U procedente de YUV y el Cb procedente de YCbCr. También aparecen dos canales de color que no habíamos visto en el primer módulo, el canal H del espacio de color LCH y el canal B del espacio de color LAB. Teniendo en cuenta que el uso de máscaras reduce la ROI al contenido del hematíe, la varianza del canal H nos informa de la diferencia entre los valores de cada pixel y el valor promedio de todo el conjunto. Para los hematíes con cualquier tipo de inclusión, el contenido de los hematíes muestra un aspecto más homogéneo, por lo que la varianza es superior en el caso de la malaria que en el resto de inclusiones. Respecto al énfasis de área pequeña, este descriptor indica que un valor más alto, como en el caso de las inclusiones, es debido a un mayor número de zonas de pequeño tamaño con valores similares de píxeles. Esto implica un mayor grado de textura fina, probablemente debido a que, en el caso de las inclusiones, los píxeles que representan al propio hematíe están más resaltados respecto a las imágenes de hematíes que contienen malaria (ver Figura 41).

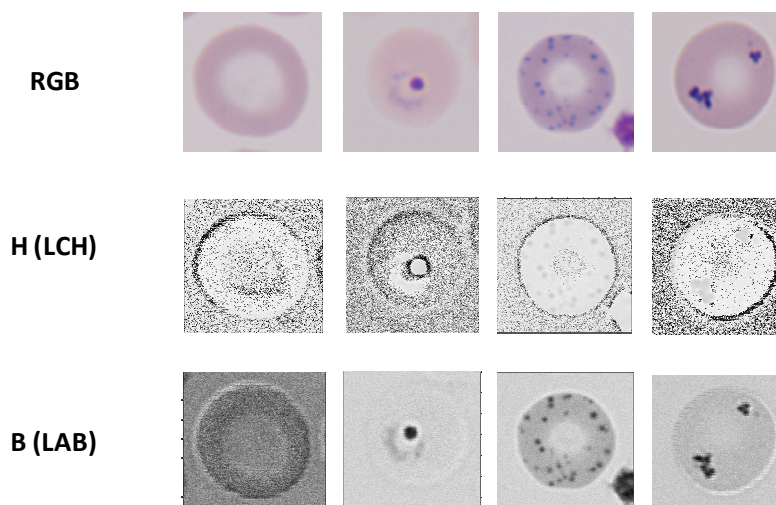


Figura 41. Representación de diferentes tipos de hematíes en el canal H del espacio de color LCH y en el canal B del espacio de color LAB.

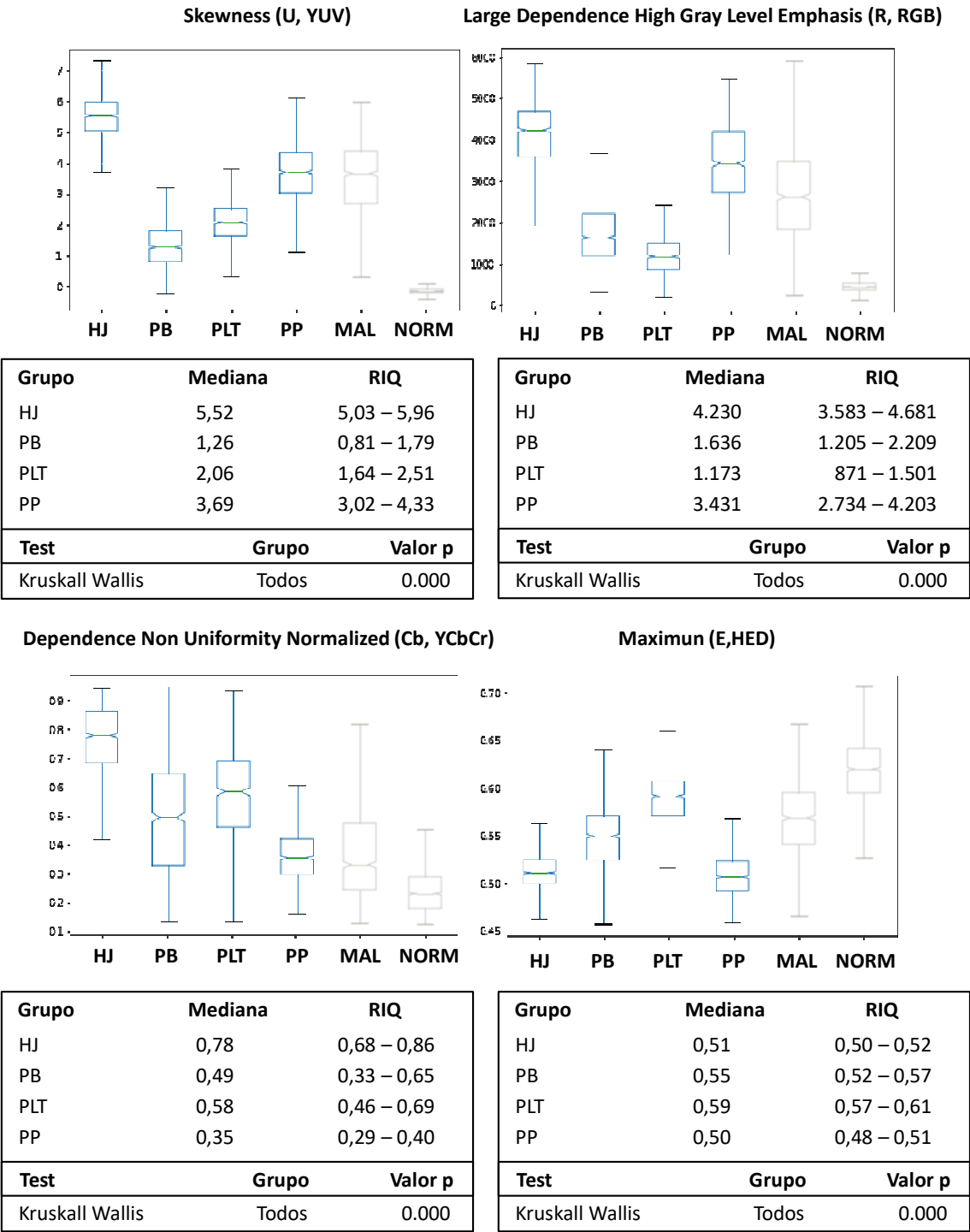
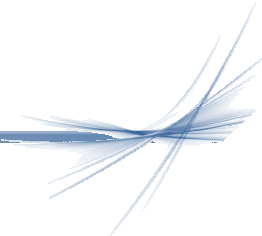


Figura 42. Descriptores más relevantes en el Módulo 3 encargado de reconocer entre hematíes con cuerpos de Howell-Jolly (HJ), con cuerpos de Pappenheimer (PP), con punteado basófilo (PB), y con plaquetas alojadas en su superficie (PLT).

En el tercer módulo del sistema final (encargado de la diferenciación entre HJ, PP, PB y PLT), los descriptores más relevantes han sido los siguientes:

- La asimetría del canal U del espacio de color YUV.
- El énfasis de larga dependencia de valores altos del canal R del espacio de color RGB.
- La dependencia no uniforme normalizada del canal Cb del espacio de color YCbCr.
- El valor máximo del canal E del espacio de color HED.

Lo primero que llama la atención en este tercer módulo es que el descriptor más relevante para diferenciar entre los hematíes con cualquier inclusión diferente a la malaria, es precisamente el mismo que para diferenciar hematíes normales de hematíes con cualquier inclusión (incluyendo la malaria) observado en el primer módulo del sistema final.

Sin embargo, la asimetría del canal U no nos serviría para la identificación específica de hematíes con malaria, ya que muestran unos valores muy similares a los de la población de PP (ver Figura 42). Pero en lo que concierne al resto de clases, es el descriptor que presenta las mayores diferencias. Al igual de lo que sucedía en el módulo anterior, los valores de los descriptores a lo largo de las diferentes clases a identificar se encuentran más solapados que en el primer módulo, y por esa razón también hace falta un gran número de descriptores (700) para conseguir las mejores tasas de rendimiento en la clasificación.

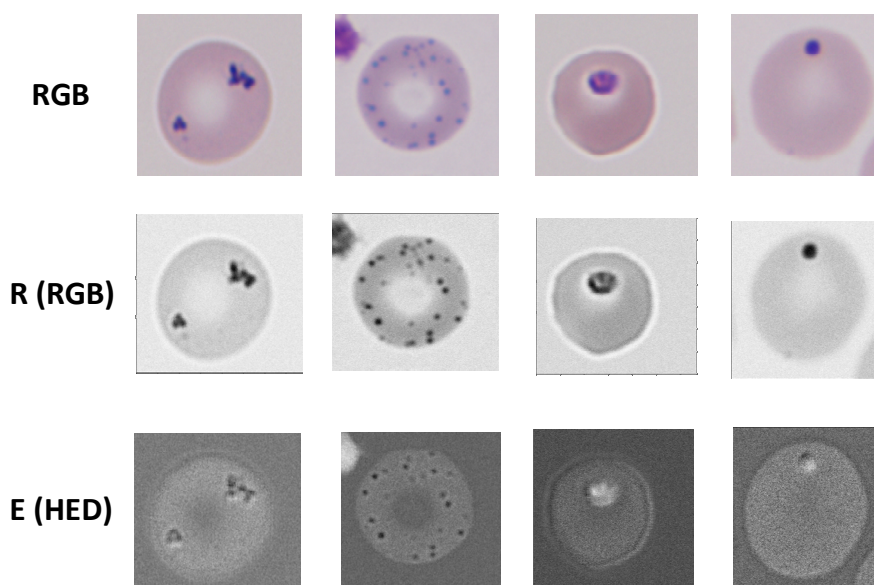
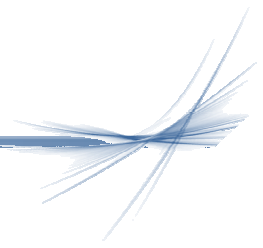


Figura 43. Representación de diferentes tipos de hematíes en el canal R del espacio de color RGB y en el canal E del espacio de color HED.



De los canales que no han aparecido anteriormente, el canal R del espacio RGB resalta mucho cualquier tipo de inclusión eritrocitaria, y con un poco menos de contraste respecto al fondo de la imagen, también resalta los píxeles contenidos en los hematíes. El énfasis de larga dependencia de valores altos del canal R habla de textura gruesa en colores de la imagen que tienden a ser claros. En la Figura 43 se observa que la intensidad de los píxeles que proceden de los hematíes son más similares entre las inclusiones HJ y PP. De aquí que la distribución de valores de este descriptor entre estas poblaciones sea más similar que entre PB y PLT. De hecho, en la mayoría de descriptores se encuentra una distribución de valores más similar entre los HJ y los PP, lo cual también tiene un sentido morfológico ya que ambos comparten una similitud morfológica en comparación con los otros tipos de inclusiones. El canal E del espacio de color HED resalta sobretodo estructuras de color rosada-violacea con una mayor intensidad que sobre estructuras azuladas. Probablemente por esa razón el valor máximo de las plaquetas en este canal queda por encima del resto de inclusiones eritrocitarias. En prácticamente todos los descriptores analizados, los parásitos de malaria adquieren unos valores intermedios completamente solapados con el resto de inclusiones, mientras que solo en alguno de ellos los hematíes normales también se solapan con los valores de las otras inclusiones. La distribución de los valores en cada descriptor de las clases pertenecientes a la malaria y a hematíes normales puede observarse de color sombreado en la Figura 42.

5.4.2.2.1 Evaluación del sistema secuencial utilizando el conjunto de evaluación

El rendimiento de la clasificación utilizando el conjunto de evaluación se resume en la matriz de confusión que se muestra en la Tabla 3, que ofrece la intersección entre las clases a las que pertenecen realmente las imágenes enfrentadas a las clases predichas por el clasificador, tanto en porcentaje como en número absoluto de imágenes. El valor correspondiente a la diagonal principal muestra la sensibilidad o la tasa de verdaderos positivos para cada clase de hematíes (porcentaje de imágenes pronosticadas como una clase sobre el total de imágenes que verdaderamente pertenecen a esa clase): 98,9% para los hematíes normales; 100% para hematíes infectados con parásitos de la malaria; 97,4% para HJ; 97,7% para PP; 94,4% para PB; y 98,1% para PLT. El promedio de estos valores (97,7%) corresponde a la exactitud ponderada, que es el porcentaje promedio de imágenes individuales que se han reconocido correctamente.

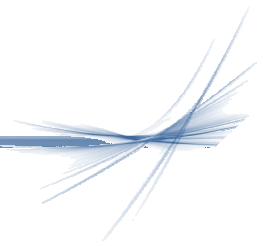
		Predicción					
		NORM	MAL	HJ	PP	PB	PLT
Clase Verdadera	NORM	98,9 (725)	0,0 (0)	0,0 (0)	0,0 (0)	0,3 (2)	0,8 (6)
	MAL	0,0 (0)	100,0 (187)	0,0 (0)	0,0 (0)	0,0 (0)	0,0 (0)
	HJ	0,0 (0)	0,4 (2)	97,4 (525)	2,2 (12)	0,0 (0)	0,0 (0)
	PP	0,0 (0)	0,4 (1)	2,0 (5)	97,7 (250)	0,0 (0)	0,0 (0)
	PB	0,3 (1)	5,3 (19)	0,0 (0)	0,0 (0)	94,4 (339)	0,0 (0)
	PLT	0,6 (1)	1,2 (2)	0,0 (0)	0,0 (0)	0,0 (0)	98,1 (159)

Tabla 3. Matriz de confusión en el conjunto de evaluación por imágenes individuales. Los resultados se muestran en porcentaje (en paréntesis se muestra el número absoluto de imágenes)

Es destacable el 100% de sensibilidad alcanzado para el reconocimiento de hematíes infectados con malaria respecto al resto de clases correspondientes a hematíes no infectados. Para este caso particular, podemos calcular la especificidad como la proporción de imágenes reconocidas como hematíes no infectados (la suma de todas las columnas en la Tabla 3 excluyendo malaria: 2.025) respecto al número de imágenes que verdaderamente pertenecen a cualquiera de las clases que no presentan malaria (la suma de todas las filas de la Tabla 3 excepto malaria: 2.049). Este indicador obtuvo un valor de especificidad del 98,8%.

5.4.2.2.2 Evaluación del sistema utilizando el conjunto de diagnóstico

Esta evaluación de frotis completos se centró en el reconocimiento automático de malaria utilizando 18 muestras correspondientes a nuevos pacientes cuyas imágenes no fueron utilizadas previamente en el entrenamiento del modelo. La Tabla 4 (separada en dos secciones, las muestras de pacientes con malaria, y las muestras de pacientes no infectados) presenta los resultados del reconocimiento en las muestras de 18 frotis pertenecientes a 8 pacientes con paludismo, con diferentes porcentajes de parasitemia (del 0,10% al 14,16%); y 10 pacientes no infectados con malaria, pero con presencia de inclusiones en algunos de sus hematíes. Para cada frotis, el sistema clasifica todas las imágenes en tres clases: hematíes normales, hematíes con malaria o hematíes que contiene otras inclusiones.



			Predicción			Diagnóstico Verdadero			Predicción	
			NORM	MAL	INC					
Frotis 1	Clase Verdadera	NORM	200	0	0	Frotis 1	Parasitemia:	Infectado	Infectado	
		MAL	0	33	0		14,20%			
		INC	0	0	0					
Frotis 2		NORM	200	0	0	Frotis 2	Parasitemia:	Infectado	Infectado	
		MAL	0	19	0		8,70%			
		INC	0	0	0					
Frotis 3		NORM	200	0	0	Frotis 3	Parasitemia:	Infectado	Infectado	
		MAL	0	7	0		3,40%			
		INC	0	0	0					
Frotis 4		NORM	198	0	0	Frotis 4	Parasitemia:	Infectado	Infectado	
		MAL	0	4	0		2,00%			
		INC	0	0	0					
Frotis 5		NORM	499	0	0	Frotis 5	Parasitemia:	Infectado	Infectado	
		MAL	0	3	0		0,60%			
		INC	0	0	0					
Frotis 6		NORM	500	0	0	Frotis 6	Parasitemia:	Infectado	Infectado	
		MAL	0	2	0		0,40%			
		INC	0	0	0					
Frotis 7		NORM	999	0	0	Frotis 7	Parasitemia:	Infectado	Infectado	
		MAL	0	2	0		0,20%			
		INC	0	0	0					
Frotis 8		NORM	1000	0	0	Frotis 8	Parasitemia:	Infectado	Infectado	
		MAL	0	1	0		0,10%			
		INC	0	0	0					
TOTAL		NORM	3796	0	0					
		MAL	0	71	0					
		INC	0	0	0					

Tabla 4.1. Matrices de confusión en el conjunto de evaluación por frotis completos en el grupo de muestras de pacientes infectados con malaria.

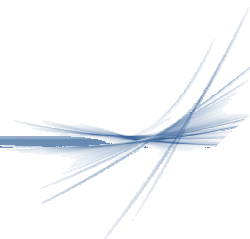
Las matrices de confusión de las muestras de pacientes con malaria se muestran en la Tabla 4.1. Se detectaron parásitos de malaria en los ocho frotis (71 imágenes de hematíes con parásitos), incluidas aquellas muestras con un bajo porcentaje de parasitemia. Un paciente ha sido reconocido como infectado cuando al menos un solo hematíe con presencia de parásitos ha sido identificado en el frotis correspondiente. La evaluación de las muestras infectadas con malaria mostró un valor de sensibilidad del modelo del 100%.

Respecto a las muestras no infectadas (ver Tabla 4.2), solo encontramos un frotis en el que una sola imagen de RBC (con presencia de una inclusión tipo HJ) se identificó erróneamente como malaria. De acuerdo a al criterio diagnóstico establecido, este frotis fue el único (uno entre más de 10 muestras) erróneamente diagnosticado como infectado, dando un valor de especificidad del 90%. Pero considerando las imágenes de hematíes individuales, solo una imagen de hematíe no infectado sobre un total de 8.720 se identificó erróneamente como infectada por malaria. El rendimiento de la clasificación considerando imágenes de hematíes

individuales del conjunto de diagnóstico resultó en una sensibilidad y una especificidad del 100% y 99,98%, respectivamente.

			Predicción			Diagnóstico Verdadero		Predicción
			nRBC	MAL	INC			
Frotis 9	Clase Verdadera	NORM	500	0	0	Frotis 9 Pappenheimer	No Infectado	No Infectado
		MAL	0	0	0			
		INC	0	0	54			
Frotis 10		NORM	500	0	0	Frotis 10 Punteado Basófilo	No Infectado	No Infectado
		MAL	0	0	0			
		INC	0	0	38			
Frotis 11		NORM	498	0	1	Frotis 11 Plaquetas sobre hematías	No Infectado	No Infectado
		MAL	0	0	0			
		INC	0	0	52			
Frotis 12		NORM	442	0	2	Frotis 12 Howell Jolly	No Infectado	Infectado
		MAL	0	0	0			
		INC	0	1	39			
Frotis 13		NORM	312	0	0	Frotis 13 Plaquetas sobre hematías	No Infectado	No Infectado
		MAL	0	0	0			
		INC	0	0	12			
Frotis 14		NORM	533	0	0	Frotis 14 Pappenheimer	No Infectado	No Infectado
		MAL	0	0	0			
		INC	0	0	18			
Frotis 15	NORM	601	0	1	Frotis 15 Punteado Basófilo	No Infectado	No Infectado	
	MAL	0	0	0				
	INC	0	0	27				
Frotis 16	NORM	448	0	0	Frotis 16 Howell Jolly	No Infectado	No Infectado	
	MAL	0	0	0				
	INC	0	0	36				
Frotis 17	NORM	393	0	0	Frotis 17 Frotis Normal	No Infectado	No Infectado	
	MAL	0	0	0				
	INC	0	0	0				
Frotis 18	NORM	416	0	0	Frotis 18 Frotis Normal	No Infectado	No Infectado	
	MAL	0	0	0				
	INC	0	0	0				
TOTAL	NORM	4643	0	4				
	MAL	0	0	0				
	INC	0	1	276				

Tabla 4.2. Matrices de confusión en el conjunto de evaluación por frotis completos en el grupo de muestras de pacientes no infectados.



5.5 PUBLICACIONES

El trabajo presentado en este capítulo ha sido publicado como artículo original en la revista *Journal of Clinical Pathology* con la siguiente referencia:

- **Molina A**, Alférez S, Boldú L, Acevedo A, Rodellar J, Merino A. Sequential classification system for recognition of malaria infection using peripheral blood cell images. *J Clin Pathol* 2020.

DOI: 10.1136/jclinpath-2019-206419.

Factor de impacto: 3,411

Indexación de la revista mediante Journal Impact Factor: percentil 30/77 (tercil 2).



CAPÍTULO 6

Reconocimiento automático de malaria y otras inclusiones eritrocitarias mediante modelos de *deep learning*

Este capítulo presenta la metodología desarrollada y los resultados obtenidos para llevar a cabo y evaluar el segundo objetivo específico de la presente tesis doctoral: “Desarrollar un método de reconocimiento automático de malaria y otras inclusiones eritrocitarias mediante herramientas de *deep learning* a partir de imágenes de sangre periférica obtenidas con microscopio”.

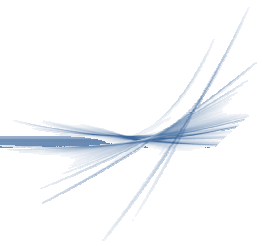
El desarrollo metodológico para la obtención de imágenes de hematíes individuales ha seguido los mismos pasos que los detallados en el Capítulo anterior:

- Obtención de muestras (Apartado 5.1.1).
- Obtención de imágenes de campo (Apartado 5.1.2).
- Etiquetado (Apartado 5.1.3).
- Preprocesamiento (Apartado 5.1.4).
- Binarización (Apartado 5.1.5).
- Postprocesamiento y segmentación (Apartado 5.1.6).

Las únicas diferencias en las metodologías realizadas en este Capítulo se basan principalmente en dos aspectos:

- Tal como se describe en el apartado siguiente, la arquitectura de las redes neuronales utilizadas en el desarrollo de este capítulo permite trabajar directamente con imágenes, a diferencia de los modelos utilizados en el capítulo anterior que utilizaban datos cuantitativos. Como consecuencia, no es necesario aplicar algoritmos de extracción de descriptores sobre las imágenes de hematíes individuales.
- Las imágenes de campo utilizadas para la obtención de hematíes individuales fueron las mismas imágenes almacenadas en formato comprimido JPG. Esta elección tiene una doble justificación. Por una parte, dado que no es necesario la extracción de características, tampoco existe la necesidad de utilizar las imágenes que contienen una mayor profundidad de bits. Por otra parte, es interesante valorar el funcionamiento del modelo sobre imágenes de menor tamaño ya que puede suponer una mayor agilidad computacionalmente.

En este capítulo se aborda en detalle los aspectos relacionados con el entrenamiento de la red neuronal convolucional. Además, también se muestran los resultados obtenidos en el entrenamiento de la red, el clasificador desarrollado, su interpretación, y se analiza el rendimiento que ofrece en la clasificación de las inclusiones eritrocitarias.



6.1 REDES NEURONALES CONVOLUCIONALES

6.1.1 Estructura general

Las CNNs han demostrado ser muy útiles en el reconocimiento de imágenes. Se trata de redes neuronales multicapa que intentan imitar la conectividad neuronal que se encuentra en la corteza visual del cerebro biológico y están específicamente diseñadas para reconocer patrones visuales. Las CNNs se componen de dos bloques principales, tal como se muestra en la Figura 44. La primera parte está compuesta por diferentes capas convolucionales de número variable en función de la arquitectura de la red. Es en estas capas donde comienza el “procesado distintivo” o “convoluciones”, donde se analizan grupos de píxeles locales de la imagen y se transforman matemáticamente a través de un producto escalar contra otra pequeña matriz que se llama *kernel*. Este *kernel* recorre toda la imagen y genera una nueva matriz de salida que se conoce como mapa de características. En cada capa convolucional puede haber configurado varios *kernels* que darán lugar a diferentes mapas de características, que a su vez serán elementos de entrada a las siguientes capas convolucionales. Al tratarse de operaciones lineales consecutivas, todas las operaciones realizadas en las diferentes capas convolucionales podrían ser equivalentes a una única capa, por lo que después de cada convolución se aplica una función de activación no lineal. La más comúnmente aplicada por sus buenos resultados y sencillez de aplicación es la función de activación lineal rectificada (ReLU), que transforma los valores negativos de la salida de una capa convolucional en 0, mientras que mantiene intactos los valores positivos. Si se hiciera una nueva convolución a partir de estos resultados el número de neuronas crecería exponencialmente. Por lo que para reducir el número de neuronas se realiza un proceso de agrupamiento a través de una función *Max-Pooling*, en el que se preserva el valor más alto de entre un grupo de neuronas, lo que reduce considerablemente el tamaño de las matrices resultantes, manteniendo la información más relevante de la imagen. Este proceso se puede repetir varias veces hasta que los mapas de características obtenidos de la última capa convolucional se transformen en un vector unidimensional. El resultado final es un conjunto de características cuantitativas que representan la imagen de entrada.

Este vector es la entrada del segundo bloque de la CNN, compuesto por capas totalmente conectadas (*fully connected*) que tiene la estructura de una red neuronal tradicional cuya última capa tendrá la cantidad de neuronas correspondientes con las clases que estamos

clasificando a través de una función conocida como *Softmax*. Esta función asigna una probabilidad de pertenencia de la imagen a las diferentes clases, de forma que el modelo predice la clase más probable.

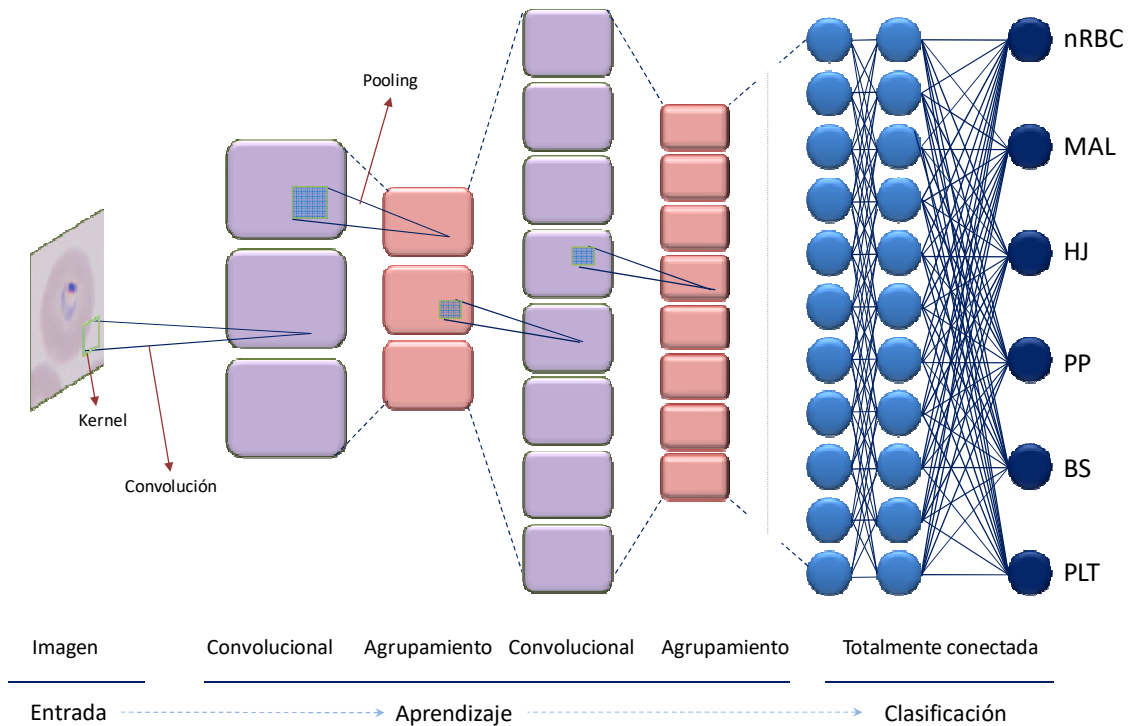
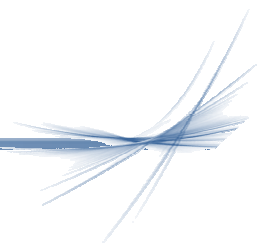


Figura 44. Esquema de la estructura de una red neuronal convolucional.

Tal como se ha descrito al comienzo de este capítulo, el esquema de trabajo para la obtención de las imágenes es similar al descrito en el capítulo anterior. Aunque debido al gran potencial de reconocimiento que tienen las CNN, se ha optado por el uso de las imágenes de campo en formato comprimido JPG con profundidad de 8 bits. A partir de estas imágenes se ha realizado un algoritmo de procesamiento y segmentación utilizando las mismas etapas que en el capítulo anterior (ver Figura 24, Apartado 5.1.6) con la finalidad de obtener las imágenes de hematíes individuales objetos de estudio. La gran diferencia ahora respecto al capítulo anterior está en que no se realiza una extracción de características como se ha descrito en el Apartado 5.2. Como se ilustra en la Figura 45, las arquitecturas basadas en CNN utilizan directamente las imágenes individuales y extraen características cuantitativas mediante las diferentes capas de convolución, las cuales se usan directamente para la clasificación.



Todos los frotis recolectados para el estudio se dividieron en entrenamiento, validación, y conjunto de evaluación del modelo. Por un lado, para el desarrollo del modelo se usaron los conjuntos de entrenamiento y validación. El conjunto de entrenamiento se utilizó para el ajuste del modelo, mientras que el conjunto de validación se utilizó para proporcionar una evaluación puntual en la finalización de cada ciclo, lo que ayudó a establecer los hiperparámetros del modelo.

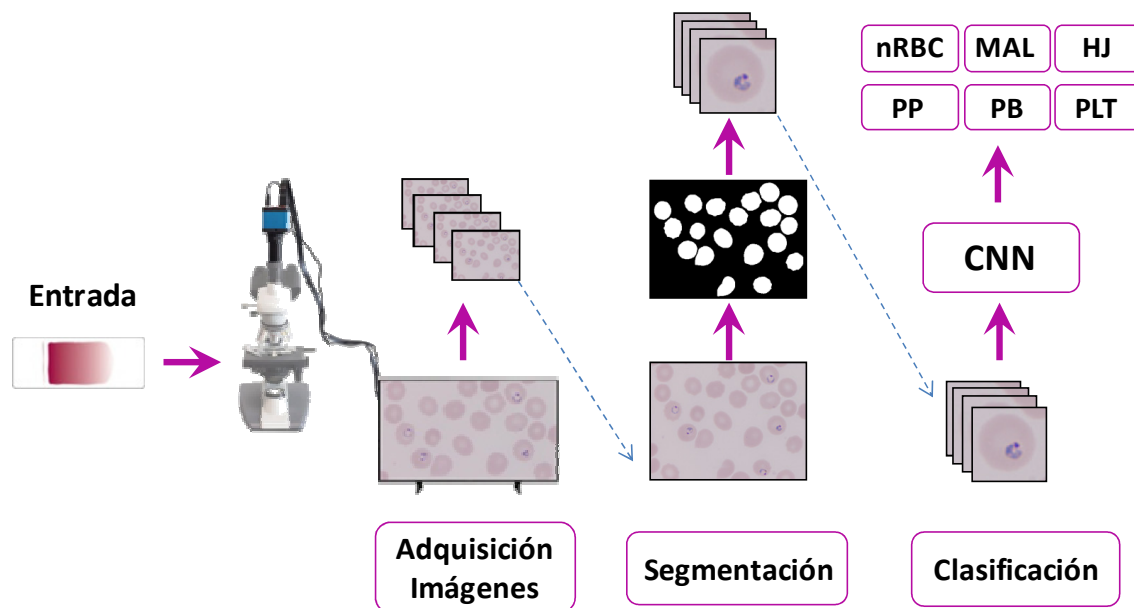


Figura 45. Representación del desarrollo de un modelo basado en redes neuronales convolucionales.

Las imágenes de validación siempre son diferentes a las imágenes de entrenamiento, aunque pueden proceder de los mismos frotis. Sin embargo, el conjunto de evaluación es importante que esté compuesto por imágenes de un conjunto de frotis que no se haya utilizado previamente en el modelo desarrollo para asegurar una evaluación fiable y representativa. El número de imágenes en el conjunto de validación se definió a partir del número total de imágenes destinadas al desarrollo del modelo. El porcentaje de imágenes destinadas a la validación se definió en un 20% para todas las clases de hematíes con presencia de inclusiones. Sin embargo, con la finalidad de mantener un delicado equilibrio entre la proporcionalidad del conjunto de validación y la superioridad en la prevalencia de hematíes normales, esta clase se definió con una división del 15%.

Para conseguir un modelo entrenado que proporcione un buen rendimiento de reconocimiento es necesario utilizar un elevado número de imágenes y una proporción equilibrada entre las diferentes clases del conjunto de entrenamiento. Un procedimiento muy utilizado en este desarrollo de modelos de *deep learning* es someter las imágenes del entrenamiento a un proceso de generación artificial (*data augmentation*). Este procedimiento aplica ciertas modificaciones al azar a una imagen original, modificando el nivel de rotación de imagen, el nivel de zoom y/o produciendo variaciones de luminosidad, lo que genera fácilmente imágenes nuevas a partir de otras ya preexistentes. Como resultado se obtiene un aumento en el conjunto de datos sin usar imágenes repetidas, equilibra el conjunto de datos de entrenamiento, mejora la generalización del modelo y evita el sobreajuste (94). El procedimiento de generación artificial se realizó hasta alcanzar 2.000 imágenes de cada clase del conjunto de entrenamiento y después de dividir el conjunto de datos de entrenamiento y validación, para garantizar que las imágenes “generadas” no se utilizaran en la validación del entrenamiento. La Figura 46 presenta un ejemplo de imágenes creadas por generación artificial.

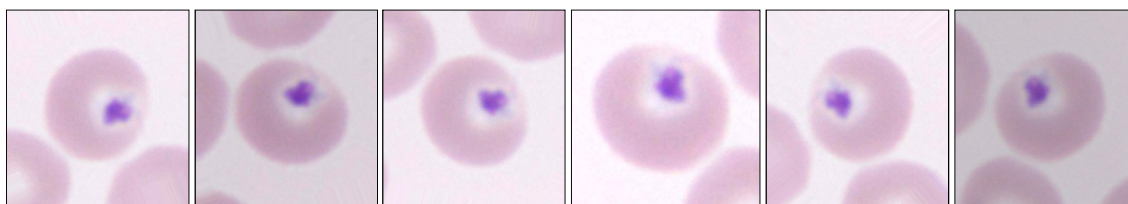
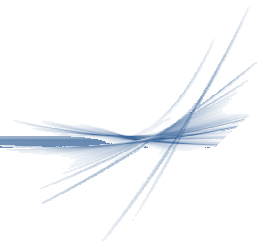


Figura 46. Representación de las imágenes generadas por generación artificial. Puede observarse como a partir de una imagen original (izquierda) se generan imágenes similares en las que la posición u orientación de la plaqueta y de los hematíes colindantes varían, así como el tamaño de la plaqueta y la luminosidad de la imagen.

6.1.2 Entrenamiento del modelo

En este trabajo se investigaron diferentes arquitecturas CNN para el desarrollo del modelo. Las arquitecturas utilizadas han sido escogidas debido a su sencilla disponibilidad en las librerías de código abierto, lo que ha desembocado en que sean las más utilizadas en la literatura científica (ver Apartado 3.4): AlexNet, ResNet, SqueezeNet, DenseNet, VGG, Xception e Inception (95-99). Se ha mantenido la estructura principal del primer bloque de todas las arquitecturas utilizadas. Para entrenar el modelo, primero necesitábamos adaptar las capas



totalmente conectadas y sus pesos a nuestro conjunto de datos/clases. Las primeras capas se definieron con 512 nodos en el caso de AlexNet; 1024 nodos para VGG, ResNet18, ResNet34 y SqueezeNet; 2048 nodos para DenseNet; 1456 para Xception; 3072 para Inicio; y 4096 nodos para ResNet50 y ResNet101. Cada nodo completamente conectado procesa una suma ponderada de sus características de entrada. Para ayudar a la generalización del modelo, esos nodos se eliminaron aleatoriamente (*dropout*) en cada iteración con una probabilidad de 0,5. Después de la primera capa totalmente conectada, las funciones de activación introducen la no linealidad en el modelo mediante la función ReLU. La segunda capa se configuró con 512 nodos, mientras que la tercera capa totalmente conectada se configuró con 6 nodos, un nodo para cada clase de hematíe que se desea detectar. Esta última capa predice la clase de las imágenes de entrada como valores de probabilidad a través de la función *Softmax*.

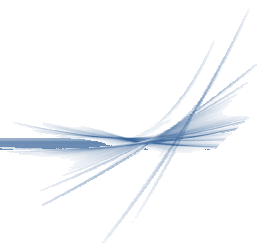
El proceso de entrenamiento de las CNN se basó en técnicas de aprendizaje por transferencia (*transfer learning*). Este método de aprendizaje sirve como una inicialización de pesos eficaz y ha demostrado que proporciona excelentes resultados de clasificación. Si lo comparamos con el entrenamiento de una arquitectura cuyos pesos se han inicializado aleatoriamente, la gran ventaja del aprendizaje por transferencia radica en que es especialmente útil cuando el conjunto de imágenes destinadas al entrenamiento es relativamente pequeño, simplificando el proceso de aprendizaje y dando lugar a modelos con un mejor rendimiento (100). Un modelo que ya ha sido entrenado con una gran cantidad de datos (generalmente *ImageNet*) podrá manejar una tarea nueva y relativamente similar con muchos menos datos. En primer lugar, se entrenaron las capas totalmente conectadas, ya que las capas posteriores tienen el mayor impacto en el reconocimiento del rendimiento. Durante el proceso de entrenamiento de estas últimas capas, se mantienen congeladas las capas convolucionales, lo que evita que se modifiquen sus pesos. En el entrenamiento existen dos hiperparámetros importantes a determinar, la tasa de aprendizaje y el número de ciclos (*epochs*) a entrenar. La tasa de aprendizaje determina la magnitud del gradiente de actualización de los pesos que se propagará hacia atrás durante el entrenamiento, y cuánto nos movemos hacia el punto que minimiza la función de pérdidas. Una tasa de aprendizaje muy pequeña hace que el modelo converja lentamente, mientras que una tasa demasiado grande hace que el modelo diverja. La determinación de la tasa de aprendizaje se basó en el enfoque propuesto por Leslie Smith conocido como *one cycle policy* (101). Esta estrategia recomienda hacer un ciclo de tasa de aprendizaje en 2 pasos de igual tamaño, pasando de una tasa muy baja a una elevada en el paso 1, y de vuelta a una tasa baja en el paso 2. Durante el proceso intermedio del aprendizaje,

cuando la tasa de aprendizaje es más alta, esta tasa funciona como método de regularización y evita que la red sea sobreentrenada. Para determinar la tasa de aprendizaje de partida, previamente realizamos un entrenamiento simulado con pocos ciclos para encontrar la tasa de aprendizaje máxima adecuada que se utilizará en el procedimiento de entrenamiento de un ciclo real. Un ciclo constituye un solo paso de todo el conjunto de imágenes de entrenamiento a través de la red neuronal, y en general, entrenar un modelo puede tomar una gran cantidad de ciclos para obtener un nivel aceptable de precisión. El número de ciclos óptimo para el entrenamiento del modelo se determinó a través de la monitorización de las pérdidas durante el proceso de aprendizaje. Los pesos del modelo se han ajustado empleando la función de entropía cruzada como función de pérdida y la función de optimización ADAM (estimación de momento adaptativo) (102). El entrenamiento se realizó utilizando un tamaño de lote de 64 a lo largo de 750 iteraciones. Una vez que se entrenaron las últimas capas totalmente conectadas, descongelamos todo el modelo y lo volvimos a entrenar para afinar el ajuste de los pesos de todas las capas, de modo que el modelo completo fuera específico para el problema que queríamos resolver.

El desarrollo del modelo de *deep learning* se realizó a través de FastAi. FastAi es una nueva biblioteca de código abierto que está construida sobre la librería de *machine learning* PyTorch y proporciona una interfaz práctica para las aplicaciones de *deep learning*, simplificando el entrenamiento y la implantación de modelos basados en CNN, que suelen tener un rendimiento excepcional en clasificación de imágenes.

6.1.3 Evaluación del modelo

De las diferentes arquitecturas CNN probadas en el Apartado anterior, se seleccionó aquella que mostró mejores resultados en la identificación de las imágenes del conjunto de validación para el modelo final. La evaluación del modelo seleccionado se realizó con un conjunto de imágenes perteneciente a 23 frotis nuevos que no se usaron previamente en el entrenamiento para realizar una evaluación imparcial y no sesgada. De cada uno de los frotis de evaluación se obtuvieron mil imágenes de hematíes individuales, distribuidas como se detalla en la Tabla 5. Los primeros 11 frotis pertenecían a pacientes infectados con malaria, con muestras que presentaban diferentes niveles de parasitemia. Los frotis 12–14 contenían únicamente hematíes normales. Los frotis restantes mostraban la presencia de eritrocitos con algunas de



las inclusiones consideradas en este trabajo: HJ (frotis 15, 16 y 21), PP (frotis 21–23), PB (frotis 17 y 18) y PLT (frotis 19 y 20).

TEST	Parasitemia	Imágenes					
		NORM	MAL	HJ	PP	PB	PLT
Frotis 1	3,2%	968	32	-	-	-	-
Frotis 2	14,0%	858	140	-	-	-	2
Frotis 3	4,7%	947	47	-	-	-	6
Frotis 4	2,4%	976	24	-	-	-	-
Frotis 5	1,9%	977	19	-	-	-	4
Frotis 6	2,1%	979	21	-	-	-	-
Frotis 7	7,6%	921	76	-	-	-	3
Frotis 8	1,1%	989	11	-	-	-	-
Frotis 9	0,5%	995	5	-	-	-	-
Frotis 10	0,2%	998	2	-	-	-	-
Frotis 11	0,1%	999	1	-	-	-	-
Frotis 12	-	1.000	-	-	-	-	-
Frotis 13	-	1.000	-	-	-	-	-
Frotis 14	-	1.000	-	-	-	-	-
Frotis 15	-	952	-	48	-	-	-
Frotis 16	-	809	-	191	-	-	-
Frotis 17	-	916	-	-	-	84	-
Frotis 18	-	831	-	-	-	169	-
Frotis 19	-	950	-	-	-	-	50
Frotis 20	-	917	-	-	-	-	83
Frotis 21	-	932	-	15	53	-	-
Frotis 22	-	878	-	-	122	-	-
Frotis 23	-	920	-	-	80	-	-
TOTAL	-	21.712	378	254	255	253	148

Tabla 5. Distribución de imágenes en el conjunto de evaluación por frotis.

La evaluación del modelo se realizó considerando dos escenarios de clasificación diferentes: imágenes individuales y frotis completos.

- Clasificación de imágenes individuales:

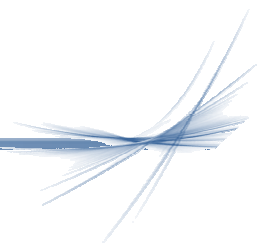
Se valoró el rendimiento en la clasificación a nivel de las imágenes de hematíes individuales provenientes de los frotis reservados para la evaluación del modelo. El objetivo de esta

evaluación consiste en confirmar la capacidad del modelo para clasificar correctamente cada una de las imágenes individuales de entrada como hematíes normales, hematíes con malaria, HJ, PP, BS o PLT. La visión global de la clasificación se representará a través de una matriz de confusión, mientras que el rendimiento específico se describirá a través de los indicadores: exactitud global, exactitud ponderada, sensibilidad, especificidad, VPP y F score.

- Detección de malaria en frotis completos:

Este objetivo consistió en evaluar el modelo desarrollado en un marco de detección de malaria de cada frotis. En este caso, la entrada es un conjunto de imágenes de un frotis y la salida es una clasificación binaria, en la que se determina si el paciente al que pertenece el frotis sufre una infección por malaria o no. Dado que muchos casos de malaria se asocian con niveles bajos de parasitemia y la visualización microscópica de un solo parásito es patognomónica de una infección por malaria, una muestra se considera infectada si el modelo propuesto reconoce al menos un hematíe con presencia de malaria dentro de todas las imágenes que componen el frotis de sangre periférica. En este sentido, la parasitemia es un concepto clínicamente muy relevante. Se puede definir como el porcentaje de hematíes que contienen parásitos de malaria en la población global de eritrocitos de un paciente. El nivel de parasitemia refleja por lo tanto la gravedad de la infección, pero indirectamente también determinará la dificultad en el diagnóstico de malaria por parte del laboratorio mediante la revisión de un frotis. Del mismo modo, la parasitemia constituye un factor determinante para el reconocimiento de malaria por parte de los sistemas automatizados, ya que es más probable que un frotis se reconozca como infectado en presencia de un elevado número de parásitos en comparación con muestras con un nivel bajo de parasitemia.

Los indicadores utilizados del rendimiento en la clasificación del modelo desarrollado son los mismos utilizados en el Capítulo anterior (ver Apartado 5.3.2.2).



6.2 RESULTADOS

6.2.1 Entrenamiento del modelo

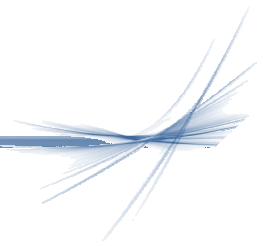
Aplicando el algoritmo de segmentación utilizado previamente en el Apartado 5.1 sobre el conjunto de imágenes en formato JPG con una profundidad de 8 bits, se han construido los conjuntos de entrenamiento, validación y evaluación tal como se detalla en la Tabla 6.

	Frotis	Imágenes					
		nRBC	MAL	HJ	PP	BS	PLT
Entrenamiento	53	2000	1083	626	648	550	350
Validación		348	286	138	145	126	92
Evaluación	Parasitemia	Imágenes					
		nRBC	MAL	HJ	PP	BS	PLT
Frotis 1	3,2%	968	32	-	-	-	-
Frotis 2	14%	858	140	-	-	-	2
Frotis 3	4,7%	947	47	-	-	-	6
Frotis 4	2,4%	976	24	-	-	-	-
Frotis 5	1,9%	977	19	-	-	-	4
Frotis 6	2,1%	979	21	-	-	-	-
Frotis 7	7,6%	921	76	-	-	-	3
Frotis 8	1,1%	989	11	-	-	-	-
Frotis 9	0,5%	995	5	-	-	-	-
Frotis 10	0,2%	998	2	-	-	-	-
Frotis 11	0,1%	999	1	-	-	-	-
Frotis 12	-	1.000	-	-	-	-	-
Frotis 13	-	1.000	-	-	-	-	-
Frotis 14	-	1.000	-	-	-	-	-
Frotis 15	-	952	-	48	-	-	-
Frotis 16	-	809	-	191	-	-	-
Frotis 17	-	916	-	-	-	84	-
Frotis 18	-	831	-	-	-	169	-
Frotis 19	-	950	-	-	-	-	50
Frotis 20	-	917	-	-	-	-	83
Frotis 21	-	932	-	15	53	-	-
Frotis 22	-	878	-	-	122	-	-
Frotis 23	-	920	-	-	80	-	-
TOTAL	-	21.712	378	254	255	253	148

Tabla 6. Distribución de imágenes en los diferentes conjuntos de entrenamiento, validación y evaluación.

El primer paso antes de proceder al entrenamiento del modelo basado en CNN ha sido buscar los valores óptimos de los principales hiperparámetros asociados a este proceso: la tasa de aprendizaje y el número de ciclos de entrenamiento. La Figura 47 muestra como varía la función de pérdidas en relación a la tasa de aprendizaje y al número de ciclos seleccionado. La función de pérdidas evalúa la desviación entre las predicciones realizadas por la red neuronal y los valores reales etiquetados por los patólogos. Cuanto menor es el resultado de esta función, más eficiente es la red neuronal. Su minimización es el objetivo principal para conseguir un modelo con un buen rendimiento de reconocimiento. La búsqueda de la tasa de aprendizaje óptima se realiza a través de un entrenamiento simulado en el que se evalúa una amplia gama de tasas de aprendizaje y qué impacto tienen éstos sobre la función de pérdida. La Figura 47.A muestra una representación de este entrenamiento simulado, en el que se dan las pérdidas obtenidas en función de diferentes valores de tasa de aprendizaje. El valor ideal se encuentra ligeramente antes del punto mínimo de la función, donde la pérdida todavía tiene margen suficiente de mejora sin sacrificar la velocidad de aprendizaje. Según los resultados obtenidos, este valor se estableció en 10^{-2} . Por otra parte, cada ciclo de aprendizaje supone un solo paso de las imágenes de todo el conjunto de entrenamiento a través de la red, y en general, entrenar un modelo puede tomar una gran cantidad de ciclos para obtener un nivel aceptable de exactitud. Sin embargo, la estrategia de *“one cycle policy”* para la aplicación de la tasa de aprendizaje optimiza este procedimiento de entrenamiento. Como se ilustra en la Figura 47.B, los valores de las pérdidas llegan a un punto de máximo rendimiento aparentemente estacionario solo con tres o cuatro ciclos. Según los resultados obtenidos, las capas totalmente conectadas fueron entrenadas con una tasa de aprendizaje máxima de 0,01 y con cuatro ciclos de entrenamiento, manteniendo los pesos del resto de la arquitectura constantes durante el entrenamiento. Después de entrenar estas capas totalmente conectadas se procedió a descongelar todo el modelo y realizar un nuevo entrenamiento con solo dos ciclos para afinar los pesos de toda la red neuronal.

La arquitectura adecuada para el modelo final fue escogida en función del rendimiento en la clasificación del conjunto de validación, cuyos resultados se muestran en la Tabla 7. De todas las CNN probadas en la fase de validación, se seleccionó la red neuronal basada en la arquitectura VGG-16 en base a los resultados de su clasificación, teniendo en cuenta los siguientes factores:



- Exactitud global: el porcentaje del total de imágenes que han sido correctamente identificadas. Este indicador proporciona una visión general del rendimiento de todas las arquitecturas. Todos los modelos probados mostraron buenos resultados de exactitud global, con valores superiores al 99%, incluso para las arquitecturas más sencillas como AlexNet. Sin embargo, VGG-16 fue el modelo que obtuvo los mejores resultados, con un 99,75% de exactitud. VGG-19 y ResNet-101 mostraron el segundo mejor resultado con una precisión general del 99,6 %.
- Valor predictivo positivo: es crucial que los eritrocitos no infectados no se identifiquen como infectados, ya que, debido a la superioridad numérica de los hematíes normales, incluso una pequeña proporción de falsos positivos pueden representar una proporción significativa con respecto a frecuencia de hematíes con parásitos de malaria. VGG-16 también tuvo el mejor valor predictivo positivo (97%) entre todas las CNN probadas.
- Sensibilidad: Es clínicamente imperativo poder detectar todos los parásitos analizados. Muchos casos de malaria están asociados con niveles bajos de parasitemia, y un falso negativo puede significar identificar una muestra como normal. Por lo tanto, los modelos deben ser muy sensibles en su identificación. Con la excepción de Inception y ResNet-101, todos los modelos tenían valores de sensibilidad cercanos al 100%.

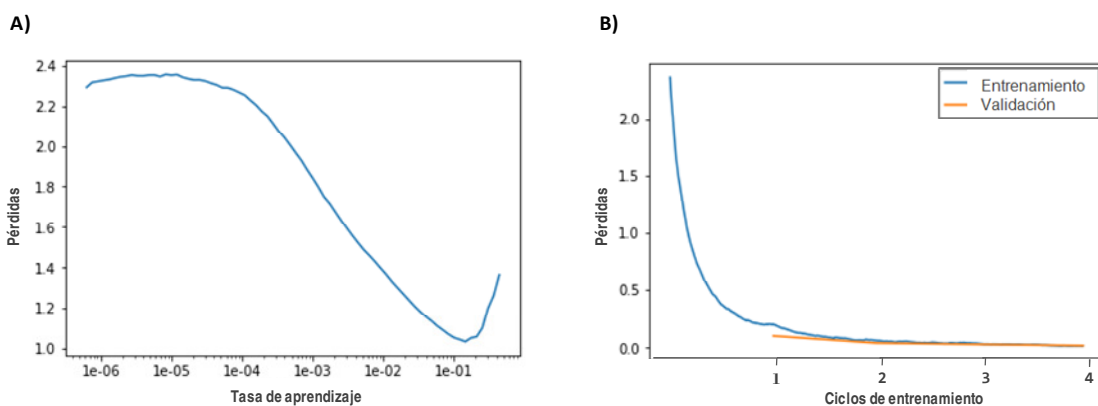


Figura 47. Monitorización de la función de pérdidas en función de la tasa de aprendizaje (A) y del número de ciclos de entrenamiento (B).

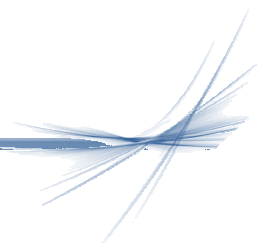
VGG-16 fue seleccionado como el modelo de clasificación final debido a que presentó los mejores resultados en todos los factores considerados. Esta CNN se basa en una arquitectura secuencial que mantiene una estructura relativamente simple. VGG-16 se compone de cinco bloques convolucionales secuenciales. Los bloques 1 y 2 tienen dos capas convolucionales, mientras que los bloques 3, 4 y 5 disponen de una capa adicional. Las capas convolucionales hacen uso de diferentes filtros de tamaño 3×3 para generar mapas de características. Al final de cada bloque, se utiliza una capa de agrupación para reducir el tamaño de la imagen tanto en dimensiones de alto como de ancho. La entrada para la primera capa totalmente conectada es una representación aplanada del mapa de características de salida del quinto bloque. Posteriormente, dos capas totalmente conectadas terminan en una función de activación *Softmax*, utilizada para representar una función categórica sobre las seis clases incluidas en el estudio, proporcionando una probabilidad para cada una de ellas.

	NORM	MAL	HJ	PB	PLT	PP	PROMEDIO	FP	FN
AlexNet	99,4	100	99,3	100	100	96,6	99,2	1,0	0
DenseNet121	99,4	100	100	100	98,9	96,6	99,2	2,3	0
Inceptionv4	99,4	99,7	99,3	100	98,9	97,2	99,1	1,1	0,3
ResNet18	99,4	100	99,3	100	100	97,9	99,4	1,0	0
ResNet34	99,4	100	100	100	100	95,9	99,2	2,4	0
ResNet50	99,4	100	100	100	98,9	96,6	99,2	2,4	0
ResNet101	99,7	99,7	100	100	98,6	97,9	99,3	1,7	0
SqueezeNet	99,1	100	100	100	98,9	98,6	99,4	1,4	0
VGG16	99,7	100	100	100	100	98,6	99,7	0,3	0
VGG19	99,7	100	100	100	97,8	97,9	99,2	3,2	0
Xception	99,1	100	100	100	98,9	97,2	99,2	2,1	0

Tabla 7. Resultados de la clasificación sobre el conjunto de validación de diferentes arquitecturas de redes neuronales convolucionales.

6.2.2 Evaluación del modelo

La valoración del modelo seleccionado se realizó mediante el conjunto de evaluación consistente en 23 frotis nuevos que no se han usado previamente en el entrenamiento. Esta evaluación de la clasificación se enfocó tanto desde el punto de vista del reconocimiento de



imágenes individuales como de frotis completos con la finalidad de establecer un diagnóstico a partir del hallazgo de muestras infectadas.

El rendimiento de la clasificación de las imágenes de hematíes individuales se resume en la Tabla 8 en forma de matriz de confusión. Para cada imagen se compara la clase real a la que pertenece frente a la clase a la que el modelo predice que pertenece. Los elementos de la diagonal representan las tasas de verdaderos positivos (sensibilidad) para cada clase de hematíe: 99,6 % para hematíes normales; 99,2 % para hematíes infectados con malaria; 98,4% para HJ; 97,6% para PP; 98,0% para PB; y 99,3% para PLT. La exactitud global del modelo fue del 99,5%. Es especialmente relevante que la alta sensibilidad del modelo permite que más del 99% de las imágenes de los hematíes infectados con malaria son identificados correctamente. Pero también es importante destacar que más de 22.000 imágenes de hematíes no infectados, entre las que también se incluyen otros tipos de inclusiones eritrocitarias, son también identificadas correctamente por el modelo. Dado que el modelo en realidad realiza una clasificación multiclase, la especificidad fue calculada considerando todas las clases que componen los hematíes que no contienen malaria, dando un valor de especificidad del 99,9%

		Predicción					
		NORM	MAL	HJ	PP	PB	PLT
Clase Verdadera	NORM	99,6 (21.604)	0,1 (16)	0,1 (13)	0,2 (33)	0,1 (13)	0,1 (15)
	MAL	0 (0)	99,2 (375)	0 (0)	0,5 (2)	0 (0)	0,3 (1)
	HJ	0 (0)	0 (0)	98,4 (250)	1,6 (4)	0 (0)	0 (0)
	PP	0 (0)	0 (0)	0 (1)	97,6 (249)	0,2 (5)	0 (0)
	PB	0 (0)	0 (0)	0 (0)	2,0 (5)	98,0 (248)	0 (0)
	PLT	0 (0)	0,7 (1)	0 (0)	0 (0)	0 (0)	99,3 (147)

Tabla 8. Matriz de confusión que refleja los resultados de la clasificación a nivel de imágenes individuales del conjunto de evaluación.

Los resultados de la clasificación por frotis se resumen en la matriz de confusión binaria de la Tabla 9, teniendo en cuenta que un solo hematíe que haya sido identificado por el modelo como malaria asignará automáticamente el diagnóstico de ese frotis de sangre periférica como infectado. Los resultados muestran que los 11 frotis de pacientes infectados con malaria (100%) fueron reconocidos correctamente. Por otra parte, sólo uno de los 12 frotis de los pacientes sin malaria fue reconocido incorrectamente como infectado. Esto significa que los valores de sensibilidad y especificidad del modelo fueron de 100% y 91,7%, respectivamente.

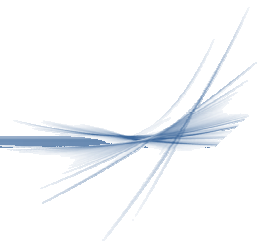
Lo que indica un alto rendimiento no solo en el reconocimiento de la infección por malaria, sino también en el reconocimiento de frotis pertenecientes a pacientes no infectados. En este sentido, es importante señalar que algunos de estos frotis contenían exclusivamente hematíes normales, mientras que otros frotis contenían hematíes con inclusiones.

		Predicción	
		Infectado	No Infectado
Clase Verdadera	Infectado	100 (11)	0
	No Infectado	8,3 (1)	91,7 (11)

Tabla 9. Matriz de confusión que refleja los resultados de la clasificación a nivel de frotis completos del conjunto de evaluación.

Para poder analizar el comportamiento de la red es necesario entender el funcionamiento de sus procesos. Cuando una imagen entra en una arquitectura de red neuronal convolucional, en su primera capa se encuentra con un conjunto de matrices que actúan como filtros. Estas matrices son agrupaciones de píxeles que tienen un tamaño de 3x3, y cada una de ellas procesa la imagen de entrada, aplicando un producto escalar de la matriz proveniente del filtro con un pequeño sector de la imagen de entrada de igual tamaño. Esta matriz recorre toda la imagen de izquierda-derecha y de arriba-abajo. Y así, con cada uno de los productos que realiza logra generar una nueva matriz de salida, que en definitiva será la entrada de una nueva capa convolucional. Por lo tanto, de cada capa convolucional se generan numerosas matrices de datos que pueden representarse de una manera similar a las de una imagen. Las primeras capas convolucionales son muy generales y son capaces de detectar características primitivas como líneas o curvas. Sin embargo, a medida que avanza a través del modelo, esta imagen se va simplificando, va reduciendo su tamaño y esto obliga a la red a transformar estos detalles en conceptos más generales. A estas matrices de salida de las diferentes capas convolucionales se les denomina mapas de características que pueden exportarse de un modelo concreto para visualizar como realiza una clasificación.

En la Figura 48 se representa un ejemplo de la salida de diferentes capas del modelo propuesto. Se ha escogido un punto inicial de la red: la salida de la segunda capa convolucional de primer bloque; un punto intermedio: la salida de la tercera capa convolucional del tercer bloque; y un punto avanzado de la red: la salida de la tercera capa convolucional del quinto bloque. A modo ilustrativo, para facilitar la interpretación del resultado, se muestran los



misimos mapas de características obtenidos tanto en el modelo ya entrenado y evaluado, como en el mismo modelo sin entrenar. Los mapas de características obtenidos en el primer bloque convolucional muestran con gran detalle la presencia de inclusiones y del propio hematíe. Esto puede observarse tanto en el modelo entrenado como sin entrenar. Y en cierta manera no es de extrañar, ya que las primeras capas son muy generales y pueden detectar cualquier tipo de estructura. Sin embargo, a medida que avanzamos por la red encontramos dos particularidades. Primero, los mapas de los bloques 3 y 5 muestran que la imagen se va volviendo más borrosa. Esto es lógico dado que la imagen va reduciendo su tamaño, y por lo tanto adquiere una menor resolución. Mientras que la salida del primer bloque produce una imagen de 224x224, la del tercer bloque es de 56x56, y la del quinto de 14x14 píxeles. Esto adquiere una importancia conceptual, y es que al simplificar la imagen la red es capaz de abstraer un concepto y aprender una propiedad de la imagen más general.

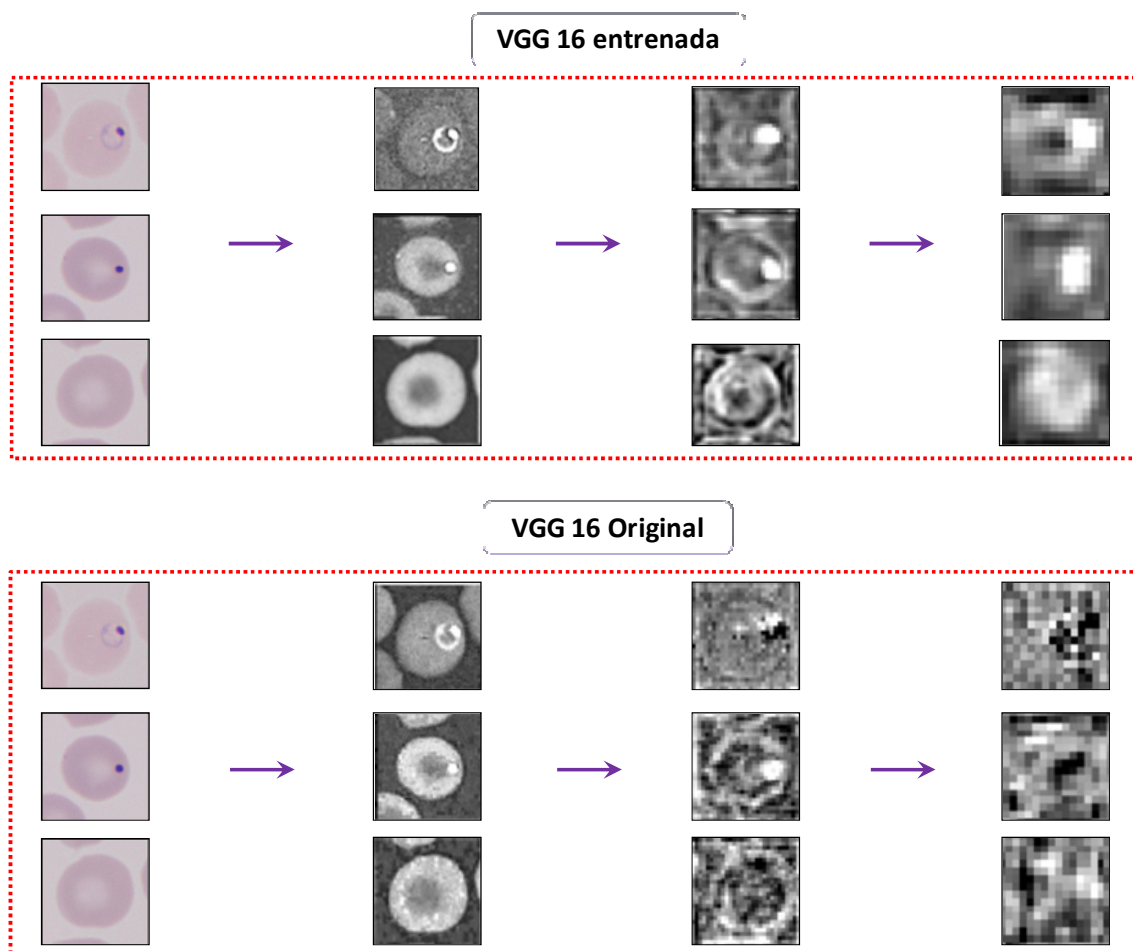


Figura 48. Mapas de características obtenidos del modelo VGG-16 entrenado y sin entrenar.

En segundo lugar, si nos fijamos en lo que sucede en una red no entrenada puede observarse que este proceso de “conceptualización” falla (ver Figura 48). La red no sabe en qué características tiene que fijarse para realizar una buena predicción, y por lo tanto la distribución de los píxeles en el mapa de características del quinto bloque parece más errática.

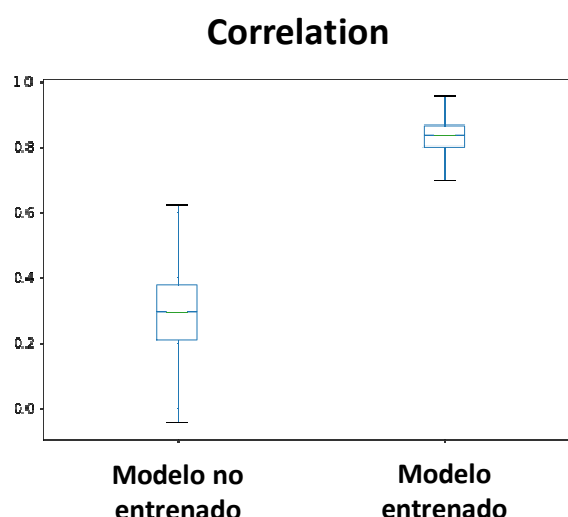
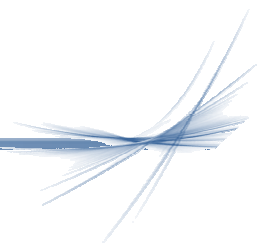


Figura 49. Distribución de valores de correlación en los mapas de características obtenidos de la quinta capa convolucional de los modelos entrenados y sin entrenar.

Estas diferencias que observamos en la generalización conceptual pueden cuantificarse. En la red entrenada, la imagen del quinto bloque se muestra más ordenada, y es que en el fondo existe una dependencia lineal entre el valor de un píxel y el valor de sus píxeles más cercanos. Esta dependencia lineal no se aprecia en la red no entrenada, y los valores de los píxeles aparecen con una disposición más errática. La disposición de los píxeles en los dos tipos de mapas de características (tanto de la red entrenada como la no entrenada) pueden medirse a través de descriptores de textura, que no es más que una medida de cómo se distribuyen los valores de los píxeles en la imagen. Se ha analizado el valor del descriptor de correlación de la matriz de coocurrencia en un conjunto de los mapas de características escogido al azar de todas las imágenes del estudio, que se representa en la Figura 49. Este descriptor refleja la tendencia lineal que puede encontrarse en la relación existente entre los diferentes valores de píxeles. Mientras que los mapas de características de la red entrenada muestran una correlación cercana a la unidad, en las imágenes que provienen de la red no entrenada se aprecia una baja correlación en los valores de los píxeles. No solo se aprecian diferencias significativas entre las dos poblaciones (prueba de U de Mann Whitney: $p < 0,001$), sino que



además este descriptor puede suponer un marcador de la calidad del entrenamiento de un modelo.

Otra manera de visualizar el comportamiento de una red consiste en calcular una combinación ponderada de las activaciones (valores de los mapas de características) de los canales de la última capa de convolución. Esto genera mapas de calor con las que puede visualizarse de modo muy intuitivo qué partes de la imagen han llevado a la red neuronal a tomar su decisión en la clasificación, lo que nos permite conocer si la red está “mirando” hacia una zona correcta de la imagen que tiene que identificar. En la Figura 50 se representan los mapas de activación obtenidos sobre un ejemplo de imágenes de cada clase objeto de estudio. Las imágenes resultantes nos muestran como en aquellas inclusiones que tienen una localización concreta dentro del hematíe, como los PP, los parásitos de malaria, los HJ y las PLT, el modelo desarrollado enfoca precisamente las inclusiones como los píxeles más relevantes de la imagen a la hora de establecer una clasificación. Sin embargo, tanto en los hematíes normales como en las inclusiones que tienen una distribución esparcida por toda la superficie del hematíe como el caso de PB, el foco de atención de la red se centra en todo el hematíe. Estos resultados nos muestran que la red es capaz de aprender por si sola a reconocer los detalles más relevantes que diferencian a todos estos tipos de hematíes.

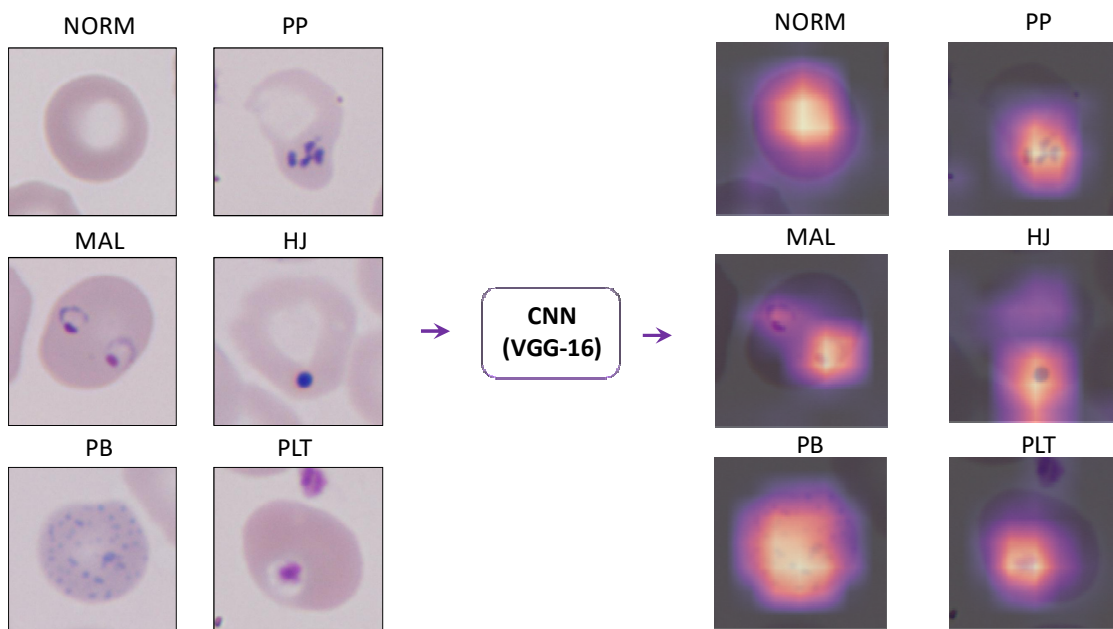


Figura 50. Mapas de activación obtenidos con el modelo VGG-16 entrenado. En la imagen de la derecha puede observarse que posteriormente al entrenamiento, los píxeles más relevantes para la clasificación de las imágenes corresponden a la localización de las inclusiones

6.3 PUBLICACIONES

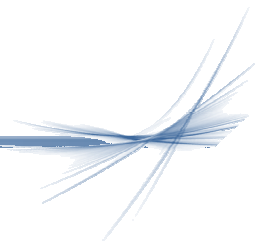
El trabajo presentado en este capítulo ha sido publicado como artículo original en la revista *Computer Methods and Programs in Biomedicine* con la siguiente referencia:

- **Molina A**, Rodellar J, Boldú L, Acevedo A, Alférez S, Merino A. Automatic identification of malaria and other red blood cell inclusions using convolutional neural networks. *Comput Biol Med* 2021.

DOI: 10.1016/j.compbiomed.2021.104680

Factor de impacto: 7,027

Indexación de la revista mediante Journal Impact Factor: percentil 20/113 (Cuartil 1).





CAPÍTULO 7

Reconocimiento automático de malaria y otras inclusiones eritrocitarias mediante *deep learning* utilizando imágenes obtenidas con CellaVision

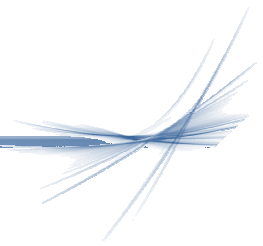
En este capítulo se describen los métodos utilizados para llevar a cabo el tercer objetivo específico de la presente tesis doctoral: “Desarrollar un método de reconocimiento automático de malaria y otras inclusiones eritrocitarias mediante herramientas de *deep learning* a partir de imágenes de sangre periférica obtenidas con el autoanalizador CellaVision”.

La obtención de imágenes de campo y su procesamiento hasta la obtención de hematíes individuales ha seguido el mismo proceso que el descrito en los Capítulos 5 y 6. En este capítulo, al Apartado 7.1.1 describe algunas particularidades del sistema CellaVision. El resto del capítulo aborda en detalle los aspectos relacionados con el entrenamiento del modelo. Además, también se muestran los resultados obtenidos al aplicar la metodología de segmentación, el entrenamiento del clasificador desarrollado y se analiza el rendimiento que ofrece en la clasificación de las inclusiones eritrocitarias, tanto individualmente como en comparación con la clasificación aportada por el propio sistema CellaVision.

7.1 MATERIAL Y MÉTODOS

7.1.1 Adquisición de imágenes mediante el sistema CellaVision

El sistema Cellavision es un analizador automatizado de imágenes digitales de frotis de sangre periférica (71). Este analizador contiene un sistema de introducción de muestras de frotis en racks de carga, escanea una porción de un portaobjetos mediante microscopía y automáticamente identifica una zona apropiada (monocapa) para el análisis, en la que realiza una captura a bajo aumento (50x) de una imagen de campo de gran alcance (ver Figura 51), y capturas a gran aumento (100x) de imágenes de leucocitos individuales. Esta imagen de campo a menor aumento cubre una extensión considerable en la que pueden detectarse alrededor de 2.000 hematíes por imagen, por lo que constituye una fuente automatizada de obtención de imágenes de frotis de sangre periférica. A diferencia de las imágenes obtenidas con el microscopio, las imágenes de la sección de hematíes tienen una dimensión de 2.920 x 2.940 pixels, con una resolución de 96 ppp, se almacenan en un formato comprimido JPG y contienen un peso digital de tan solo 1.200 Kb para toda la zona de extensión que cubren, lo que por una parte hace que sean imágenes de baja cantidad de información, pero por otra las convierte en elementos muy ligeros con los que trabajar.



Este sistema automatizado, que se encuentra disponible en un gran número de laboratorios clínicos, constituye un gran avance en la estandarización de la adquisición de imágenes. Esto supone un gran beneficio para aquellos laboratorios que no dispongan de un sistema manual de captura de imágenes. Incluso para aquellos laboratorios que posean una cámara acoplada a microscopio, el sistema CellaVision ofrece una homogenización en las características de las imágenes captadas, solventando las variabilidades que se pueden encontrar entre los diferentes softwares de capturas. Todo esto nos ha llevado a desarrollar un modelo basado en *deep learning* a partir de imágenes adquiridas por el sistema CellaVision.

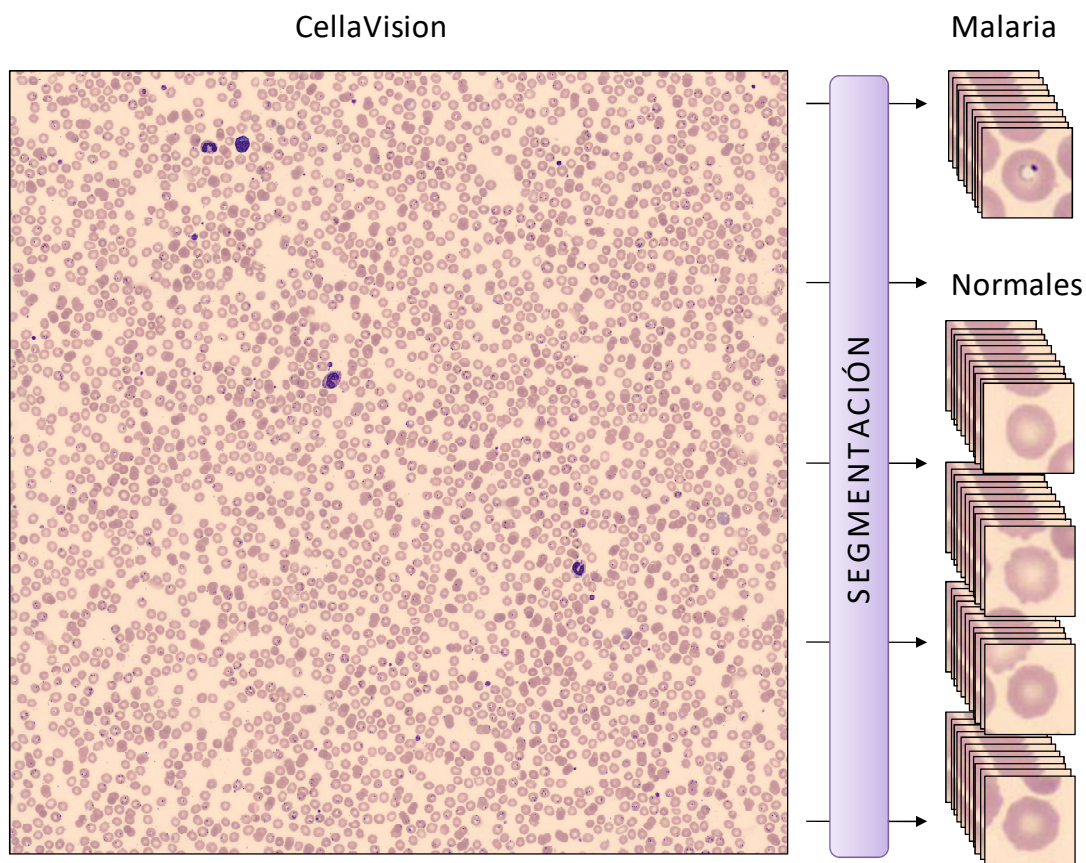


Figura 51. Esquema de obtención de imágenes de hematíes individuales a partir de la imagen de campo obtenida por el sistema CellaVision.

El algoritmo de segmentación para la obtención de imágenes de hematíes individuales ha sido modificado para adaptarse a las imágenes obtenidas con el sistema CellaVision y para facilitar

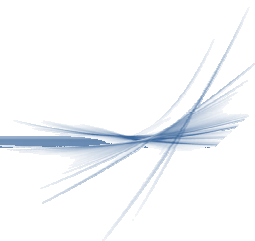
su utilización en modelos de *deep learning*. El proceso de segmentación se ha basado en los pasos que se han descrito previamente en el Apartado 5.1:

- Selección del canal verde del espacio de color RGB.
- Aumento del contraste mediante ecualización del histograma.
- Suavizado de la imagen mediante el filtrado mediana.
- Umbralización de Otsu.
- Morfología matemática: eliminar objetos pequeños, rellenado de los huecos de los hematíes y eliminar hematíes que quedan tocando los bordes de la imagen que no corresponden a hematíes enteros.
- Algoritmo de Watershed.
- Recorte de cada hematíe individual.

En este caso, las imágenes de hematíes van a ser clasificadas por modelos basados en CNN. Por lo tanto, no es necesario la obtención de las máscaras de cada imagen individual ni tampoco se realizó cálculo de descriptores.

7.1.2 Entrenamiento del modelo

A partir de las imágenes de hematíes individuales obtenidas mediante la segmentación de las imágenes proporcionadas mediante el sistema Cellavision se han construido dos conjuntos de imágenes: el conjunto destinado al entrenamiento y el conjunto para la evaluación. El conjunto de entrenamiento está constituido por 150 frotis de sangre periférica, de muestras que contenían hematíes normales, hematíes con parásitos de malaria, hematíes HJ, con PP, con PB y con PLT. Tal como se ha explicado en apartados anteriores, el conjunto de entrenamiento ha sido utilizado para el desarrollo del modelo, para lo que a su vez ha sido dividido en los subconjuntos de entrenamiento y validación mediante un procedimiento de validación cruzada 5x. El desarrollo del modelo se ha realizado mediante el uso de herramientas de *deep learning*, en la que se han probado las siguientes arquitecturas de CNN: AlexNet, VGG16, el modelo VGG16 desarrollado en el segundo objetivo específico de esta tesis doctoral (ver Capítulo 6, Apartado 6.1.2), VGG19, ResNet18, ResNet34, ResNet50, ResNet101 y SqueezeNet. El proceso de entrenamiento de las CNN se basó en técnicas de aprendizaje por transferencia, utilizando como inicialización de los pesos de las redes la información aprendida previamente en el



conjunto de imágenes de *ImageNet*, y en su caso, de la VGG desarrollada con las imágenes obtenidas con el microscopio. Primero se entrenaron las capas totalmente conectadas de cada arquitectura mientras que posteriormente se procedió a descongelar toda la arquitectura para afinar los pesos toda la red. Los valores de tasa de aprendizaje y ciclos utilizados en el entrenamiento se basaron en los valores obtenidos en el modelo anterior. El rendimiento en la clasificación de las diferentes arquitecturas se valoró mediante la clasificación sobre el conjunto de validación. La arquitectura que mostró un mejor rendimiento global en la clasificación se escogió para el modelo final y se evaluó mediante el conjunto de imágenes de evaluación, con el objetivo de analizar su rendimiento en la clasificación con un conjunto de imágenes procedentes de nuevos frotis que no fueron utilizados durante el entrenamiento. El desarrollo del modelo se realizó a través de FastAi.

7.1.3 Evaluación del modelo

Por una parte, se valoró el rendimiento en la clasificación a nivel de las imágenes de hematíes individuales provenientes de los frotis reservados para la evaluación del modelo. Se utilizaron 48 frotis de sangre periférica: 17 frotis de pacientes infectados con malaria y 31 frotis que incluyen hematíes con diferentes tipos de inclusiones. La visión global de la clasificación de las imágenes individuales se representó a través de una matriz de confusión, mientras que el rendimiento específico se describió a través de los indicadores: exactitud global, exactitud ponderada, sensibilidad, especificidad, VPP y F score.

También se realizó una evaluación por frotis de sangre periférica, para valorar la capacidad del modelo para predecir muestras de pacientes como infectados o no infectados. En esta ocasión, tanto en la evaluación de imágenes individuales como en la evaluación por frotis se realizó además mediante una comparación entre el rendimiento del modelo finalmente seleccionado y el mencionado software ARBCA (ver Apartado 3.5) en los mismos 48 frotis seleccionados para el conjunto de evaluación. ARBCA es un software de clasificación automática de alteraciones eritrocitarias de reciente aparición para su uso en el laboratorio clínico. Dentro de las alteraciones eritrocitarias, incluye las siguientes inclusiones en el interior de los hematíes: hematíes parasitados con malaria, HJ, PP y PB. Por lo tanto, es posible comparar el modelo propuesto en esta tesis doctoral con el software ARBCA en las mismas muestras de pacientes. Además de los indicadores de rendimiento en la clasificación, también se comparó la eficacia

en el procedimiento de segmentación de nuestro trabajo respecto a la obtenida por el CellaVision.

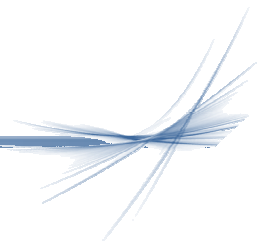
7.2 RESULTADOS

7.2.1 Entrenamiento del modelo

Se presentan aquí los resultados de la comparación entre el rendimiento en la clasificación de las inclusiones eritrocitarias realizadas por el propio Cellavision y por nuestro modelo. Para ello, se han utilizado las propias imágenes captadas por el Cellavision. Estas imágenes de campo proporcionan una extensa visión del área de un frotis, por lo que en primer lugar ha sido necesario desarrollar un algoritmo de segmentación de los hematíes contenidos en la imagen, presentado en el Apartado 7.1.1. Con ello se obtuvo un conjunto total de imágenes de 393.026 hematíes individuales. De todo el conjunto de imágenes, en primer lugar, se construyó un conjunto de evaluación con las imágenes de hematíes correspondientes a 48 frotis de sangre periférica (25 % de los frotis originales). Las imágenes restantes fueron divididas en un 80 % para el entrenamiento y 20 % para la validación, aunque a diferencia de la división realizada para la evaluación, esta vez no se tuvo en cuenta si las imágenes procedían de frotis diferentes. Las imágenes pertenecientes al conjunto de entrenamiento fueron sometidas a un proceso de generación artificial a través de modificaciones al azar como rotaciones, cambios de zoom y luminancia (excepto en los hematíes normales que han sido sometidos a un proceso de reducción mediante una selección al azar) hasta conseguir un total aproximado de 5.000 imágenes de hematíes de cada clase. El conjunto de imágenes final utilizado para el entrenamiento/evaluación de los modelos se muestra en la Tabla 10.

	Frotis	Imágenes					
		nRBC	MAL	HJ	PP	PB	PLT
TRAINING^a	150	281.626	4.990	1.969	1.584	1.611	980
TRAINING^b		5.000	4.990	4.907	4.752	4.833	4.900
TESTING	48	95.716	3.154	355	551	284	206

Tabla 10. Número total de imágenes y frotis utilizados en los conjuntos de entrenamiento y evaluación. nRBC: hematíes sin inclusiones. MAL: hematíes con parasites de malaria; HJ: hematíes con cuerpos de Howell-Jolly; PP: hematíes con cuerpos de Pappenheimer; PB:



hematíes con punteado basófilo; PLT: hematíes con plaquetas en su superficie. Training^a muestra la distribución original de imágenes. Training^b muestra el conjunto de imágenes que se utilizará para el entrenamiento después de aplicar un proceso de generación artificial y reducción, dependiendo del número de imágenes de partida de cada clase.

Los resultados del entrenamiento de las diferentes arquitecturas de CNN muestran muy buenos resultados para el reconocimiento de inclusiones eritrocitarias sobre el conjunto de validación (ver Tabla 11). En todos los casos se observan unos promedios de exactitud global en la clasificación superiores al 97 %, con muy buenos valores de sensibilidad, especificidad y valor predictivo positivo en el reconocimiento de la malaria. De todas las arquitecturas probadas (en las que también se ha incluido el modelo de *deep learning* del segundo objetivo específico desarrollado en el Capítulo 6 de esta tesis doctoral), se ha encontrado un rendimiento ligeramente superior en la clasificación por ResNet-34, con una exactitud global en el conjunto de validación del 98,58 %. Además, esta arquitectura también es la que muestra un mayor VPP. Por lo que se decidió escoger esta red como modelo final para realizar una prueba de concepto con el conjunto de imágenes de evaluación, formado por 48 frotis que no fueron utilizados para el entrenamiento del modelo.

	Sensibilidad	Especificidad	VPP	Exactitud Global
AlexNet	96,99	99,63	98,08	97,43
ResNet-18	97,99	99,96	99,80	98,30
ResNet-34	99,19	99,92	99,60	98,58
ResNet-50	98,99	99,94	99,50	98,55
ResNet-101	98,90	99,84	99,20	98,42
SqueezeNet	96,79	99,82	99,07	98,02
VGG-16	97,99	99,84	99,19	98,23
VGG-19	98,39	99,84	99,49	98,50
VGG-16 *	97,79	99,90	99,49	98,39

Tabla 11. Resultados en porcentaje de la clasificación de las diferentes arquitecturas de CNN utilizadas sobre el conjunto de imágenes de validación. TPR: Tasa de verdaderos positivos; VPP: Valor predictivo positivo. VGG-16* corresponde al modelo de *deep learning* entrenado con las imágenes del microscopio (ver Capítulo 5) y reentrenado para las imágenes de CellaVision.

7.2.2 Evaluación del modelo

El analizador CellaVision contiene una aplicación para la identificación automática de alteraciones eritrocitarias llamada *Advance RBC Application (ARBCA)*. Estas alteraciones se agrupan en cuatro categorías: color, tamaño, forma e inclusiones. Dentro del grupo de inclusiones eritrocitarias, CellaVision está diseñado para reconocer los siguientes tipos de alteraciones eritrocitarias: HJ, PP, PB y parásitos de malaria. Este analizador ofrece una visión del resultado de la clasificación de cada uno de los hematíes que consigue detectar en una imagen de campo (ver Figura 52). Esto nos llevó a realizar una evaluación de nuestro modelo a dos niveles: por una parte, evaluar el rendimiento de la clasificación de nuestro modelo en el reconocimiento de los hematíes individuales; por otra parte, analizar el rendimiento de la clasificación del sistema CellaVision con la finalidad de poder compararlo con nuestro modelo sobre las mismas muestras.

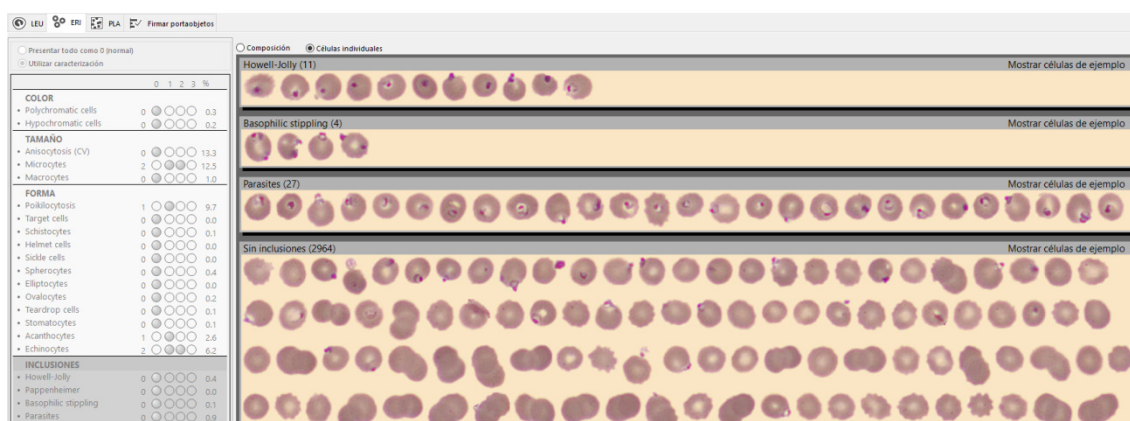
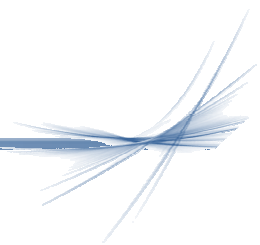


Figura 52. Visión general de la clasificación de las inclusiones eritrocitarias realizada por CellaVision.

Con respecto a la metodología de segmentación de los hematíes, al utilizar los mismos frotis de sangre periférica, el módulo de la serie roja del CellaVision obtuvo un total de 87.394 imágenes de hematíes individuales para clasificar, mientras que el modelo propuesto en nuestro trabajo obtuvo un total de 100.266 imágenes. Estos resultados muestran que a nivel global nuestro algoritmo de segmentación ofrece un mejor rendimiento en la obtención de imágenes de hematíes individuales que el propio algoritmo de CellaVision. Los resultados de la



clasificación sobre los 48 frotis destinados a la evaluación, tanto por el sistema CellaVision como por el modelo propuesto en este trabajo, se muestran en las Tablas 12 y 13, respectivamente.

		Predicción					
		nRBC	MAL	HJ	PP	PB	PLT
Clase real	nRBC	98,77	0,70	0,46	0,01	0,05	0
	MAL	20,56	47,35	30,31	0,05	1,73	0
	HJ	8,79	12,30	78,91	0	0	0
	PP	27,71	44,34	16,86	0,46	10,62	0
	PB	30,65	3,63	1,21	0,40	64,11	0
	PLT	42,82	47,78	7,31	0,26	1,83	0

Tabla 12. Matriz de confusión del sistema CellaVision sobre un conjunto de evaluación constituido por 48 frotis de sangre periférica. Los resultados se muestran en porcentaje. nRBC: hematíes sin inclusiones. MAL: hematíes con parasites de malaria; HJ: hematíes con cuerpos de Howell-Jolly; PP: hematíes con cuerpos de Pappenheimer; PB: hematíes con punteado basófilo; PLT: hematíes con plaquetas en su superficie.

Los resultados de la clasificación sobre el conjunto de evaluación por el sistema CellaVision muestran una exactitud global de 96,5%, aunque en gran parte es debido a la alta cantidad de imágenes de hematíes normales que son identificados correctamente. Si realizamos una ponderación para cada una de las clases, los valores de exactitud disminuyen hasta el 48%, lo cual refleja de una manera más realista el rendimiento en la clasificación. Mientras que los hematíes normales muestran una sensibilidad del 98,8 %, en el caso de los hematíes con inclusiones este porcentaje disminuye de un modo variable. Por ejemplo, solo un 47,4% de los hematíes infectados con malaria son identificados correctamente. Dentro de las inclusiones eritrocitarias, los HJ y el PB muestran las mejores tasas de reconocimiento, con una sensibilidad del 78,9% y 64,11%, respectivamente. Muchos de estos hematíes con reconocidos como hematíes con malaria o como hematíes normales. Por otro lado, las peores tasas de reconocimiento se dan en PP y en PLT. En el caso de PP, estos son identificados normalmente como cualquiera de las otras clases de hematíes. El caso de PLT es muy particular, ya que CellaVision no tiene contemplada esta categoría entre los elementos que tiene que reconocer.

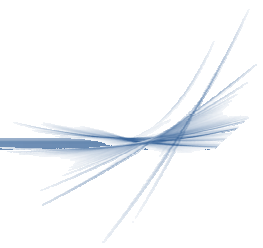
Sin embargo, sí que segmenta hematíes con plaquetas alojadas en su superficie, incluso pueden observarse plaquetas aisladas segmentadas, y estos elementos son reconocidos principalmente como malaria o como hematíes normales.

		Predicción					
		nRBC	MAL	HJ	PP	PB	PLT
Clase real	nRBC	99,58	0,03	0,12	0,05	0,12	0,11
	MAL	0,22	99,12	0	0,44	0	0,22
	HJ	0	0,67	97,99	1,34	0	0
	PP	1,56	1,56	0	94,53	2,34	0
	PB	0	0	0	0	100	0
	PLT	0	0	0	0	0	100

Tabla 13. Matriz de confusión del modelo seleccionado y entrenado (Resnet-34) sobre un conjunto de evaluación constituido por 48 frotis de sangre periférica. Los resultados se muestran en porcentaje. nRBC: hematíes sin inclusiones. MAL: hematíes con parasites de malaria; HJ: hematíes con cuerpos de Howell-Jolly; PP: hematíes con cuerpos de Pappenheimer; PB: hematíes con punteado basófilo; PLT: hematíes con plaquetas en su superficie.

Utilizando las mismas extensiones, los resultados de la clasificación obtenidos con nuestro modelo basado en la arquitectura ResNet-34 mostraron valores de sensibilidad, para cada una de las clases, igual o superiores al 98%, excepto para PP en los que se obtuvo una sensibilidad del 94,5%. Estos resultados en la clasificación mostraron un valor de exactitud global de 99,6%, e incluso realizando un promedio en los valores de exactitud de cada clase, el valor solamente disminuyó hasta el 98,54%. Los valores de sensibilidad y especificidad para el reconocimiento de hematíes parasitados con malaria fueron del 99,12 y 99,67 %, respectivamente. Además, el VPP fue del 95,15 %, por lo que no solo la mayoría de hematíes con malaria fueron correctamente identificados, sino que también la mayoría de elementos que el clasificador reconoció como malaria realmente lo eran.

La baja especificidad del CellaVision tiene un gran impacto en la clasificación a nivel de frotis completos. Si tenemos en cuenta que una muestra de frotis es identificada como infectada si contiene un mínimo de dos imágenes que han sido identificadas como hematíes con parásitos



de malaria, observamos que el sistema CellaVision mostró una especificidad muy baja ya que prácticamente todos los frotis evaluados tanto normales como patológicos fueron reconocidos como infectados (ver Tabla 14). Nuestro modelo sin embargo mantuvo un buen equilibrio entre sensibilidad y especificidad, lo que permitió la detección de muestras infectadas sin que quedaran identificadas como falsos negativos. Además, mantuvo una baja proporción de falsos positivos. Los valores de sensibilidad y especificidad a nivel de frotis por el CellaVision fueron de 100% y 9,7% respectivamente, mientras que en el caso del modelo propuesto se observaron unos valores de 100% y 96,7%.

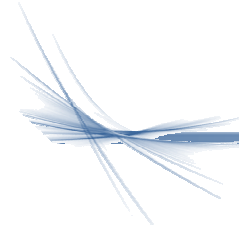
<u>CELLAVISION</u>				<u>RESNET-34</u>			
		Predicción				Predicción	
Clase Real		NORMAL	INFECTADO	Clase Real		NORMAL	INFECTADO
		NORMAL	17	0	NORMAL	17	0
	INFECTADO	28	3		INFECTADO	1	30
Sensibilidad		100%		Sensibilidad		100%	
Especificidad		9,7%		Especificidad		96,7%	

Tabla 14. Resultados de la clasificación sobre las mismas muestras a nivel de frotis completos tanto por el sistema CellaVision (izquierda) como por el modelo desarrollado (Resnet-34) en este capítulo.



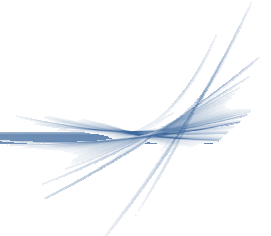
CAPÍTULO 8

Discusión



Aunque existen muchos trabajos publicados acerca del reconocimiento automático de malaria, hasta la fecha no existen publicaciones en los cuales se aborde simultáneamente la diferenciación de la malaria respecto de otras inclusiones. Es muy importante considerar esta cuestión, ya que en los clasificadores propuestos por otros autores (en los que solamente se entrena un modelo para discernir entre hematíes normales y hematíes infectados con malaria) un hematíe con otro tipo de inclusiones puede dar lugar a un falso positivo, disminuyendo enormemente la especificidad de los trabajos publicados. La metodología propuesta en esta tesis asegura que los hematíes reconocidos como infectados por malaria realmente lo son y no son confundidos con otros elementos. En este estudio, desarrollamos un sistema de reconocimiento que es capaz de diferenciar no sólo los hematíes infectados con malaria, sino también hematíes que contiene otras inclusiones y que son importantes en el contexto de la revisión citológica del frotis. Hasta la fecha actual, es la primera vez en la literatura que se ha considerado este enfoque en el reconocimiento de la malaria.

En este trabajo se ha desarrollado un algoritmo de segmentación basado en umbralización y en la transformada de Watershed. Este enfoque ha sido utilizado por la mayoría de los procedimientos de segmentación publicados ya que mantiene un buen equilibrio entre la sencillez de aplicación y la obtención de buenos resultados. Este algoritmo permite reconocer como diferentes hematíes que tienen cierto grado de solapamiento. Sin embargo, este procedimiento también ha producido una sobresegmentación en el 2,5% de los hematíes que se ha solventado mediante la aplicación de la función *convex hull*. Esta función no altera la morfología de la máscara del hematíe, o al menos no de manera apreciable siempre que no nos encontremos con alteraciones significativas de forma. Eso significa que el algoritmo de segmentación propuesto puede ser válido para el estudio de inclusiones eritrocitarias, pero puede dar lugar a máscaras poco precisas si se utilizan para el estudio de otras alteraciones eritrocitarias de forma. En la literatura son escasos los trabajos que realizan una valoración del rendimiento en el proceso de segmentación. Somasekar *et al* mostraron unos valores de exactitud del 97,2% en su algoritmo de segmentación basado en la detección de contornos (103). Diaz *et al* solamente valoraron su precisión sobre la segmentación de hematíes solapados a través de un proceso de emparejamiento mediante plantillas. Su algoritmo mostró un valor de exactitud del 95% (25). Aunque otros autores no muestran una valoración de su proceso de segmentación, la literatura muestra que existen multitud de enfoques diferentes y completamente válidos en la metodología de segmentación de hematíes. El desarrollo de un algoritmo de segmentación preciso es imprescindible si se desea implementar el modelo en un



entorno real de trabajo, ya que con toda seguridad la fuente de obtención de imágenes provenientes de frotis de sangre periférica dará lugar a imágenes de campo.

Todo nuestro trabajo se ha desarrollado bajo la hipótesis de que los modelos que se han publicado en la bibliografía para la diferenciación entre hematíes normales y hematíes con malaria, no son válidos en la práctica real porque no se han tenido en cuenta la presencia de hematíes con otro tipo de inclusiones. En este caso, si las inclusiones eritrocitarias las reconociera como hematíes normales no tendría mayor repercusión porque clínicamente tienen una relevancia concreta. Pero la malaria supone un hallazgo con enormes repercusiones, por lo que es importante que el reconocimiento automático de estos parásitos sea muy sensible y muy específico. Una alta sensibilidad porque es importante que en muestras con baja parasitemia, cualquier hematíe con presencia de malaria sea correctamente identificado, y una alta especificidad para evitar que la gran cantidad de muestras normales a los que se enfrentará el sistema no den la alarma de presencia errónea de malaria, ya que acabaría generando un aumento de la carga asistencial del especialista de laboratorio al tener que revisar innecesariamente un mayor número de frotis de sangre periférica. Esta hipótesis se ha demostrado en primer lugar mediante el desarrollo de un clasificador para diferenciar hematíes normales de hematíes con malaria con una SVM que utiliza solamente siete descriptores. Este modelo, que simula el entrenamiento de los clasificadores publicados en la bibliografía, ha sido puesto a prueba con el conjunto de imágenes de hematíes que contienen otras inclusiones y en la mayoría de los casos se confirma que las inclusiones serían erróneamente reconocidas como malaria. Esto tiene una cierta relación a nivel morfológico, ya que un hematíe que contenga un trofozoito de malaria parece más similar a otros tipos de inclusiones que respecto a los hematíes normales. De todas las inclusiones eritrocitarias, los cuerpos de Howell Jolly son los que muestran mayor similitud con los núcleos de los parásitos, y puede observarse que en este modelo el 100% son identificados como malaria, por lo que nuestra percepción morfológica de los elementos parece tener una correlación en los algoritmos de clasificación. Prueba de esta similitud se refleja en la distribución de valores de los descriptores más relevantes para diferenciar hematíes normales de hematíes con malaria. El hecho de que los hematíes con inclusiones tengan una distribución de valores muy similar a la de los hematíes infectados con malaria hace que los modelos desarrollados con esta finalidad reconozcan mayoritariamente las inclusiones como hematíes infectados con malaria.

Purwar *et al* (17) son los únicos autores que estudiaron el reconocimiento automático de parásitos de malaria e intentó distinguirlos de otros artefactos. Utilizó una versión modificada de *k means clustering* para identificar agrupaciones de píxeles de una determinada intensidad de color. Este algoritmo reconocía tanto los parásitos de malaria como otro tipo de artefactos. Los artefactos que consideraron fueron leucocitos, plaquetas, suciedad y restos de colorante. La aproximación que utilizó para separar los artefactos de los parásitos fue suponer que los artefactos consistían en estructuras que quedaban fuera de los hematíes, mientras que los parásitos de malaria se alojan en el interior. Este enfoque, aunque intenta solventar la posible presencia de elementos que ocasionen falsos positivos, puede evitarse mediante el desarrollo de un adecuado algoritmo de segmentación que seleccione específicamente los hematíes y no otras estructuras externas. Aun así, aunque en este trabajo se buscaron elementos adicionales que pudieran confundir su modelo, no se valoró el problema de discriminar la malaria respecto a otras inclusiones eritrocitarias. A la hora de desarrollar un modelo de inteligencia artificial para la identificación automática de cualquier elemento, es muy importante tener en cuenta y entrenar el modelo con estructuras que puedan tener algún tipo de similitud con los elementos diana. El éxito en la implementación de modelos de aprendizaje automático está estrechamente relacionado con los detalles particulares del conjunto de datos con los que se entrenó el modelo. Por lo tanto, si se produce una deriva entre el conjunto de imágenes de entrenamiento y las imágenes a las que el modelo se enfrentará en condiciones reales, con toda seguridad se producirá una pérdida del rendimiento en la clasificación que en otros entornos diagnósticos incluso podría tener consecuencias clínicas (104). En este sentido, para el desarrollo de modelos de inteligencia artificial con aplicación en el laboratorio clínico es fundamental la colaboración de los propios especialistas de laboratorio ya que éstos son los que en la práctica diaria tienen la experiencia de los posibles elementos que pueden ser una fuente de confusión para las clases que se desea identificar automáticamente.

La estructura principal de esta tesis doctoral se ha basado en el desarrollo de tres objetivos secundarios: un modelo basado en herramientas clásicas de *machine learning* desarrollado con imágenes de microscopio, un modelo basado en herramientas de *deep learning* desarrollado con imágenes de microscopio, y un último modelo de *deep learning* que utiliza imágenes procedentes de CellaVision. En el modelo basado en herramientas clásicas de *machine learning* se han evaluado diferentes algoritmos estadísticos de clasificación para cada uno de los módulos que componen el sistema final. Respecto al módulo encargado de la discriminación entre hematíes normales y hematíes infectados con malaria todos los



algoritmos mostraron muy buenos valores de exactitud global. Sin embargo, SVM presentó los mejores resultados de clasificación utilizando solamente los siete descriptores más relevantes. Para los módulos dos y tres (encargados de la identificación de los hematíes con malaria o con otro tipo de inclusiones), SVM y LDA mostraron mejores resultados respecto a los demás por medio del uso de más de 600 descriptores, con una ligera superioridad de LDA respecto a SVM. No es de extrañar que en la literatura SVM constituye uno de los algoritmos más utilizados para abordar el reconocimiento automático de los parásitos de la malaria (1,35,37,46-53). Según los resultados de nuestro trabajo, SVM constituye una buena opción tanto para el desarrollo de tareas sencillas como complejas, ya que en todos los casos es capaz de proporcionar excelentes resultados de clasificación. Sin embargo, LDA mostró el mejor rendimiento en los módulos dos y tres del sistema final. Nuestro grupo también ha demostrado que un sistema de aprendizaje automático basado en LDA tiene un alto nivel de rendimiento en la clasificación de diferentes tipos de leucemias agudas (105), por lo que parece un algoritmo que adapta bien a tareas complejas. Respecto a la identificación de hematíes parasitados con malaria solamente se han encontrado dos trabajos que utilicen LDA, donde además se compara el rendimiento proporcionado por KNN. Tek *et al* encontraron que KNN era mejor que LDA en la discriminación entre hematíes normales y hematíes con malaria, aunque los propios autores argumentaron que la selección de descriptores se realizó en función de los resultados obtenidos con el clasificador KNN, lo que podría conferir una ventaja a este modelo sobre LDA (14). Park *et al* también compararon KNN respecto a LDA, y encontraron que KNN proporcionaba mejores resultados para diferenciar entre trofozoitos de malaria y hematíes normales, mientras que LDA mostró mejores resultados para la diferenciación del estadio parasitario. Aunque son pocas las publicaciones que utilizan LDA, los estudios de Tek y Park también parecen sugerir que LDA tiene un mejor rendimiento que otros algoritmos en tareas complejas, por lo que constituye una buena opción a tener en cuenta en el desarrollo de modelos de aprendizaje automático.

El descriptor más relevante del primer módulo (hematíes normales vs hematíes con cualquier tipo de inclusión) coincide con el del clasificador de que simula los modelos publicados en la literatura (hematíes normales vs hematíes con malaria). La asimetría del canal U muestra unos valores similares entre los hematíes con malaria y los hematíes con inclusiones, por lo que este descriptor representa la mejor opción para diferenciar los hematíes normales del resto de clases. Estos resultados refuerzan la idea de que los clasificadores publicados en la literatura reconocerían las inclusiones eritrocitarias como parásitos de malaria con una alta probabilidad,

y que para el reconocimiento automático de la malaria es fundamental diferenciarlas del resto de inclusiones. Además, el hecho de que no sea un descriptor importante para el segundo módulo indica que para diferenciar entre malaria y el resto de inclusiones es imprescindible tener en cuenta otros descriptores específicos que no son útiles en la diferenciación de los hematíes normales. Curiosamente, el tercer módulo que diferencia entre HJ, PP, PB y PLT vuelve a obtenerse como el descriptor más importante la asimetría del canal U, lo que indica que es un buen descriptor para la diferenciación de cualquier tipo de hematíe excepto los que están infectados por malaria. Estos resultados demuestran intentar diferenciar todos estos tipos de hematíes de una manera directa puede resultar difícil de conseguir, y que la estrategia de desarrollar un modelo secuencial en el que cada clasificador se enfoque en tareas concretas parece una aproximación que produce mejores resultados.

En el segundo objetivo secundario es esta tesis doctoral se ha desarrollado un modelo basado de reconocimiento basado en *deep learning*. Durante el entrenamiento, todas las arquitecturas de CNNs probadas mostraron un alto rendimiento en la identificación de la malaria y otras inclusiones eritrocitarias. Decidimos que la arquitectura VGG-16 era la más adecuada por dos razones: 1) mostraba una ligera superioridad en el rendimiento y 2) está formada por una arquitectura simple y uniforme con un número relativamente bajo de capas. Al igual que en nuestro trabajo, otros autores han evaluado diferentes CNNs para el reconocimiento de malaria, obteniendo las mejores puntuaciones de clasificación con arquitecturas VGG (68-70). VGG-16 contiene una gran cantidad (138 millones) de parámetros entrenables; este número es mayor que el de otras arquitecturas más complejas, como ResNet-101, que contiene solo 44 millones de parámetros entrenables con una densidad de capas muchísimo mayor. Nuestros resultados sugieren que el alto nivel de rendimiento en el reconocimiento de inclusiones de hematíes podría ser relativamente independiente de la complejidad de la CNN y podría estar relacionado con la cantidad de parámetros entrenables. Esto explicaría que incluso arquitecturas sencillas como AlexNet también mostrara buenos resultados de reconocimiento. Las primeras publicaciones sobre el reconocimiento automático de malaria mediante herramientas de *deep learning* utilizaron perceptrones multicapa (25,36,38), aunque no se obtuvieron mejores resultados al compararlo con algoritmos convencionales de clasificación como el SVM. Sin embargo, las CNNs suponen un cambio radical, ya que su manera de trabajar con imágenes supone una gran mejora en términos de rendimiento. La Tabla 15 muestra los resultados comparativos de nuestro modelo de clasificación con respecto a otros trabajos publicados anteriormente en los que también se han valorado diferentes CNNs. Nuestro



sistema de reconocimiento automático mostró unos valores de exactitud, sensibilidad y especificidad superior al 99%, teniendo en cuenta que es el único que fue entrenado con seis clases diferentes de Imágenes de hematíes.

	CNN	Exactitud	VPP	Valor F1	Sensibilidad	Especificidad
Molina	VGG16	99,5	95,7	97,4	99,2	99,9
Bibin ¹⁰⁶	DBN	96,4	82,9	89,7	97,6	95,9
Gopakumar ⁶⁴	custom CNN	98,5	55,9	70,9	97,1	98,5
Rajaraman ⁶⁸	VGG16	99,3	99,7	99,3	-	-
Rajaraman ⁶⁹	VGG16	95,1	-	95,2	94,6	95,7
Vijayalakshmi ⁷⁰	VGG19-SVM	93,1	89,9	91,7	93,4	92,9

Tabla 15. Comparación entre los indicadores de rendimiento en la clasificación de las principales publicaciones que utilizan modelos basados en deep learning.

Normalmente, los frotis de sangre periférica procedente de pacientes infectados con malaria poseen una proporción de hematíes desbalanceada, con una frecuencia de hematíes normales mucho mayor respecto a los parasitados. Este hecho puede tener un impacto en los indicadores del rendimiento en la clasificación y por lo tanto, es importante usar conjuntos de datos no balanceados en el proceso de evaluación que reproduzcan los casos reales. Gopakumar *et al* utilizaron un conjunto de imágenes desbalanceada (en proporciones reales) en la evaluación de un modelo para la detección de parásitos de la malaria basado en una CNN personalizada (64). Su modelo mostró unos valores de sensibilidad y especificidad del 97,1% y 98,5%, respectivamente. Estos valores indican una buena tasa de reconocimiento tanto de hematíes normales como de hematíes parasitados. Sin embargo, cuando el número de hematíes normales es mucho mayor en comparación con los hematíes parasitados, incluso una pequeña tasa de falsos positivos puede ser muy significativo en relación a los verdaderos positivos. Por este motivo se observa en su trabajo un VPP del 56%. Este resultado indica que prácticamente la mitad de las imágenes que son reconocidas como malaria en realidad no lo son. Por lo tanto, a la hora de evaluar modelos de reconocimiento automático de malaria es muy importante, por una parte, utilizar proporciones de hematíes desbalanceadas que reflejen la proporción de hematíes normales/parasitados que se encuentran en casos reales; y por otra

parte, en este tipo de proporciones uno de los indicadores de rendimiento en la clasificación más apropiado es el VPP. En nuestro trabajo, incluso utilizando un conjunto de datos desbalanceados en la evaluación, se obtuvo un VPP del 95,7%. Otros autores como Vijayalakshmi, Rajaraman y Bibin evaluaron sus modelos usando conjuntos de datos balanceados (68-70,106) por lo que se podría esperar VPP más bajos en el procesamiento de muestras de pacientes reales. Los sistemas desarrollados también se diseñaron para analizar un frotis completo como un único elemento para valorar la capacidad de los modelos de identificar muestras infectadas con malaria. Este es un aspecto importante cuando se desarrollan nuevas herramientas de apoyo al diagnóstico para la práctica diaria del laboratorio. Por ejemplo, CellaVision es un sistema de análisis de frotis de sangre periférica que sigue un proceso de escaneado e identificación. Las identificaciones que realiza de cada frotis se mantienen almacenadas como una lista de trabajo para su revisión. Nuestro sistema podría aplicarse a un analizador tipo CellaVision que "etiquetara" o mostrara una alarma en aquellos frotis que presentan algún hematíe con parásitos de malaria. Esta aplicación permitiría llamar la atención del especialista de laboratorio para facilitar su rápida examinación, sobre todo en aquellos laboratorios con una gran carga de trabajo asistencial. Nuestro sistema supondría un complemento ideal ya que una altísima sensibilidad (cercana al 100%) y un alto nivel de especificidad (alrededor del 90%) en el diagnóstico de frotis completos. Actualmente el software del propio CellaVision proporciona una gran cantidad de falsos positivos que hace inviable su uso en el diagnóstico de frotis completos.

En el desarrollo de esta tesis doctoral, los modelos propuestos correspondientes a los dos primeros objetivos específicos (ver capítulos 4 y 5) utilizan imágenes que han sido obtenidas manualmente a través de un microscopio con cámara acoplada. El hecho de que cada laboratorio necesite un sistema de captura de imágenes a nivel local puede suponer un inconveniente para su implementación. Sin embargo, en el tercer objetivo secundario de esta tesis doctoral se ha desarrollado un modelo basado en CNNs que utiliza imágenes de campo tomadas por el CellaVision. Esto supone un avance muy significativo en la automatización y estandarización de la toma de imágenes, lo que facilita la aplicación práctica de nuestro modelo en un entorno real de trabajo, ya que el CellaVision es un sistema automatizado muy extendido en los laboratorios clínicos. Pero además, los resultados conseguidos mediante los diferentes objetivos muestran que la diferenciación específica de la malaria respecto a otras inclusiones eritrocitas puede lograrse mediante diferentes enfoques, y por lo tanto, puede adaptarse a las necesidades y recursos de diferentes tipos de laboratorios.



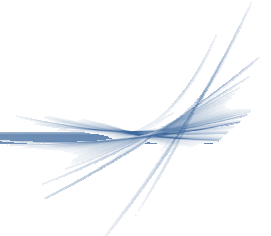
Este trabajo es el primero en el que se utilizan las imágenes de campo proporcionadas por el CellaVision en el desarrollo de modelos de clasificación automática de la serie roja. Para poder trabajar con este tipo de imágenes es necesario desarrollar un algoritmo de segmentación. En nuestro caso, de cada imagen de campo se ha obtenido un promedio de 2.000 hematíes individuales por lo que nuestro modelo sería capaz de reconocer muestras infectadas con malaria con parasitemias iguales o superiores al 0,05%. Además, el algoritmo de segmentación propuesto ha demostrado un rendimiento 1,15 veces superior al mostrado por el propio autoanizador CellaVision. Nuestro modelo no solo ha mostrado un mejor rendimiento en la segmentación de hematíes, sino también en la clasificación automática de malaria e inclusiones eritrocitarias. La clasificación con el sistema CellaVision ha mostrado que solo el 47 % de los hematíes infectados con malaria son reconocidos como tales, siendo el resto clasificados como hematíes normales o como HJ. Por otra parte, un 44 % de los PP y un 48 % de las PLT son erróneamente reconocidos como parásitos de malaria por el CellaVision. Estos datos de clasificación muestran un VPP del 46,6%, lo que significa que más de la mitad de los hematíes reconocidos como infectados en realidad no lo son. Aunque el sistema CellaVision ha demostrado ser una gran herramienta de soporte al especialista del laboratorio clínico para la clasificación de leucocitos de sangre periférica, su rendimiento en la clasificación de inclusiones eritrocitarias es muy mejorable y ha resultado ser de escasa aplicación práctica a nivel real. Si el sistema CellaVision alertara al usuario de posibles muestras infectadas con malaria, prácticamente en casi todas las muestras habría algún hematíe que sería reconocido como infectado. El modelo propuesto en este objetivo basado en la clasificación utilizando ResNet-34 ha demostrado un rendimiento mucho más eficiente comparado con el sistema CellaVision. Utilizando las mismas muestras con las que se ha evaluado CellaVision, nuestro modelo ha mostrado una exactitud global superior al 99,5 % y un VPP de 98,31%.

A la hora de realizar una evaluación de un modelo de reconocimiento automático de imágenes de sangre periférica es importante tener en cuenta los frotis sobre los que proceden las imágenes al realizar la división para los conjuntos de entrenamiento y evaluación. La metodología de esta división puede influir en el proceso de evaluación. Aunque se usen imágenes diferentes en el entrenamiento y evaluación, si estas proceden del mismo frotis compartirán características similares de morfología, tinción y luminosidad, lo que puede dar lugar a una sobreestimación en el rendimiento real del modelo. Rajaraman demostró que no tener en cuenta la entidad del frotis entero en la evaluación puede sobreestimar la exactitud global del modelo en un 5% (69). Para asegurar que el rendimiento de la evaluación no se

encuentra sobreestimado, una buena práctica para la evaluación de cualquier modelo de clasificación propuesto consiste en dividir las imágenes en conjuntos de entrenamiento y evaluación a partir de diferentes frotis y pacientes. En todos nuestros modelos, incluso siguiendo esta metodología, se obtuvieron valores de sensibilidad y especificidad cercanos al 100%, más elevados que los publicados en la literatura.

El diagnóstico de la malaria se basa en la visualización microscópica de los parásitos en gota gruesa o en extensión de sangre periférica. Se recomiendan ambos tipos de muestras en la búsqueda de parásitos de malaria. Yang *et al* desarrollaron una aplicación para *smartphones* que usaba una CNN personalizada para detectar parásitos de malaria en gotas gruesas con una exactitud del 93,46 % (107), un valor elevado considerando la gran cantidad de artefactos que suelen encontrarse en las gotas gruesas. Nosotros hemos centrado nuestros modelos de clasificación en el uso de extensiones de sangre periférica ya que permiten la visualización de detalles morfológicos que son importantes para el diagnóstico, como el tamaño y la forma, la presencia de pigmentos y manchas, y la morfología del parásito (108). La extensión de sangre periférica es, además, una herramienta de cribado y de diagnóstico muy frecuentemente utilizada para la confirmación de los resultados aportados por el hemograma. Este tipo de frotis es la muestra ideal que permite abordar la diferenciación de parásitos de la malaria de otras inclusiones eritrocitarias a través de modelo de *machine learning*.

En los objetivos de la tesis doctoral desarrollados se han utilizado diferentes modelos que han utilizado una fuente de imágenes de características diferentes. En el primer objetivo basado en *machine learning* clásico se utilizaron imágenes en formato TIFF de muy alta resolución obtenidas con microscopio. Las imágenes de los hematíes individuales obtenidos de la segmentación tienen un tamaño aproximado de 600 Kb. El modelo de *deep learning* desarrollado durante el segundo año utilizó imágenes de alta resolución obtenidas con microscopio en formato comprimido JPG dieron lugar a imágenes de hematíes individuales de un tamaño aproximado de 8 Kb. El modelo de *deep learning* que utiliza imágenes de campo tomadas con el autoanalizador CellaVision proporciona imágenes que están en formato comprimido jpg de baja resolución. En este caso, las imágenes de hematíes individuales tienen un tamaño aproximado de 1 Kb. Los resultados aportados durante los diferentes años de tesis doctoral muestran que, independientemente de la resolución de las imágenes, se pueden obtener muy buenos resultados de clasificación automática de las inclusiones eritrocitarias. Esto supone un factor importante a tener en cuenta, ya que trabajar con imágenes livianas



hace que los costes computacionales y los tiempos de ejecución sean mucho menores, lo que agiliza la implementación práctica de los modelos.

Las herramientas de aprendizaje automático son cada vez más importantes en el campo de la medicina de laboratorio. Todos los trabajos publicados en relación a la detección automática de la malaria muestra que existen multitud de enfoques diferentes tanto en el procesamiento de imágenes, segmentación, extracción de descriptores, entrenamiento de modelos, algoritmos de clasificación, cuyas variables pueden dar la ventaja a algunos modelos sobre otros. En este trabajo se han desarrollado diferentes modelos para el reconocimiento automático de la malaria a partir de imágenes digitales provenientes de frotis de sangre periférica. Estos modelos mejoran significativamente el rendimiento de los modelos publicados en la literatura, siendo capaz de reconocer no solo los parásitos de malaria, sino también otras inclusiones eritrocitarias que no están relacionadas a la infección por paludismo. La alta sensibilidad y especificidad de nuestro modelo reduce el número de frotis que debe revisar el patólogo clínico, mejorando la eficiencia del laboratorio flujo de trabajo en el laboratorio. Este trabajo supone un avance prometedor que puede ayudar a reducir los tiempos de respuesta y ayudar al patólogo clínico a realizar análisis morfológico más objetivo, eficiente y rápido de las muestras.



CAPÍTULO 9

Conclusiones

En la presente tesis se han desarrollado nuevos sistemas para la detección de malaria basados en métodos de *machine learning* y *deep learning* a partir de imágenes individuales de hematíes de sangre periférica. Estos sistemas pretenden contribuir al soporte diagnóstico de la malaria en el laboratorio clínico, dada su alta sensibilidad y especificidad.

Estos sistemas son capaces, a diferencia de los publicados anteriormente en la literatura, de discriminar diferentes tipos de inclusiones eritrocitarias, presentes muy frecuentemente en la sangre periférica, de los parásitos intraeritrocitarios del tipo *Plasmodium* responsables de la malaria.

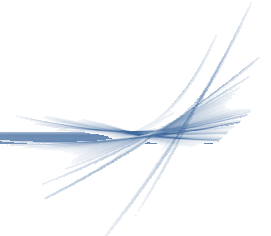
Las conclusiones derivadas de la investigación realizada se resumen en el Apartado 9.1. Las contribuciones principales de la presente tesis se destacan en el Apartado 9.2. Finalmente, las perspectivas de futuro surgidas de este trabajo se describen en el Apartado 9.3.

9.1 CONCLUSIONES

La conclusión general de la tesis está relacionada con el objetivo principal de la misma y es que se ha desarrollado un sistema de elevada sensibilidad y especificidad para la clasificación automática de malaria y que es capaz de identificar también las diferentes inclusiones eritrocitarias como tales y, por tanto, resuelve el problema de los falsos positivos.

9.1.1 Conclusiones derivadas del primer objetivo específico basado en el reconocimiento de malaria y otras inclusiones eritrocitarias mediante modelos de *machine learning*

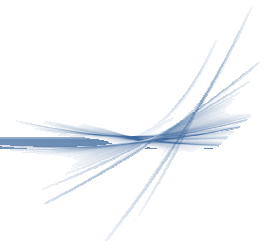
- 1 Se ha diseñado un método de procesamiento y segmentación de eritrocitos basado en umbralización y en la transformada de *Watershed*, que ha mostrado un alto rendimiento para la detección de hematíes en una imagen de campo obtenida mediante microscopio.
- 2 Se ha implantado con éxito un algoritmo de extracción de descriptores basados en variables de color y principalmente de textura.



- 3 El estudio de los descriptores mostró que los hematíes infectados con malaria tienen características morfológicas cuantitativas distinguibles con respecto a los que contienen otras inclusiones.
- 4 Entre los descriptores que presentan mayor relevancia en la clasificación entre hematíes normales y hematíes con parásitos de malaria, así como para diferenciar entre los diferentes tipos de inclusiones, el más importante fue la *asimetría de los píxeles del canal U*.
- 5 Las inclusiones eritrocitarias fueron clasificadas erróneamente como parásitos de malaria por los modelos de clasificación automática entrenados exclusivamente para distinguir entre hematíes normales y hematíes infectados.
- 6 Los mejores resultados de la validación del primer módulo de clasificación para diferenciar entre hematíes normales o con cualquier tipo de inclusión (incluida la malaria) se obtuvieron utilizando el método de clasificación de máquinas de soporte vectorial (SVM) y siete descriptores relevantes, con una exactitud del 99,12%.
- 7 En el segundo módulo, para la discriminación entre hematíes con parásitos de malaria o hematíes con otras inclusiones, la combinación del método de clasificación análisis discriminante lineal (LDA) con las 610 características más relevantes mostró una exactitud global del 96,52%.
- 8 En el caso del tercer módulo, la combinación de LDA con los 700 descriptores más relevantes mostró los mejores resultados en la clasificación de hematíes con las otras inclusiones eritrocitarias diferentes a los parásitos de malaria. La exactitud fue del 96,91%.
- 9 El diagnóstico de malaria considerando frotis individuales de sangre periférica de los pacientes utilizando nuestro modelo de *machine learning* mostró una sensibilidad del 100% y especificidad del 90%.
- 10 El sistema de clasificación desarrollado fue capaz de detectar infección por malaria con valores de parasitemia iguales o superiores al 0,1%.

9.1.2 Conclusiones derivadas del segundo objetivo específico basado en el reconocimiento de malaria y otras inclusiones eritrocitarias mediante modelos de *deep learning*

- 1 Las herramientas de *deep learning* mostraron una gran superioridad en términos de tiempos de computación y de rendimiento en la clasificación con respecto a las herramientas clásicas de *machine learning*. Como ventaja adicional, permiten trabajar con imágenes más livianas.
- 2 Entre los modelos con redes neuronales convolucionales (CNN) empleados en esta tesis para la clasificación automática de hematíes con malaria y otras inclusiones, las mejores métricas de rendimiento se obtuvieron utilizando la estructura VGG-16: promedio del 99% de exactitud, 97% de valor predictivo positivo (precisión) y 99,9% tanto de sensibilidad como de especificidad.
- 3 En la detección de parásitos intraeritrocitarios a partir de frotis individuales de pacientes con malaria, la sensibilidad y especificidad de nuestro modelo de *deep learning* fue del 100 y 91,7%, respectivamente. Estos resultados demuestran que nuestro sistema es eficiente en la práctica para la identificación de pacientes con malaria.
- 4 Las últimas capas convolucionales del modelo de clasificación fueron las más relevantes para la detección de malaria y otras inclusiones eritrocitarias.
- 5 Se ha propuesto la correlación de la matriz de coocurrencia como un índice (ICMC) para cuantificar el grado de ordenación en los píxeles de los mapas de características. Este índice ha mostrado valores que aumentan significativamente a medida que la red se entrena con las imágenes de los hematíes normales, con parásitos intraeritrocitarios y conteniendo las otras inclusiones objeto de estudio.



9.1.3 Conclusiones derivadas del tercer objetivo específico basado en el reconocimiento de malaria y otras inclusiones eritrocitarias mediante modelos de *deep learning* utilizando imágenes obtenidas con el analizador CellaVision

- 1 Se ha diseñado un método de segmentación de eritrocitos a partir de imágenes de campo obtenidas con el autoanalizador CellaVision. Su rendimiento es superior al del propio sistema CellaVision.
- 2 Se ha diseñado un método de clasificación de eritrocitos con presencia de parásitos de malaria y otras inclusiones mediante *deep learning* (basado en ResNet-34) a partir de imágenes obtenidas en el CellaVision, con un promedio de exactitud en la clasificación del 99,6 %.
- 3 A partir de imágenes obtenidas con el CellaVision nuestro modelo de clasificación secuencial basado en CNNs mostró una mayor sensibilidad y especificidad para la detección de malaria en comparación con el *software* propio del sistema CellaVision.

9.2 CONTRIBUCIONES

Las contribuciones tangibles derivadas del trabajo realizado en esta tesis se detallan a continuación:

1. Un método de segmentación de hematíes válido para diferentes tipos de imágenes de campo que muestra un alto rendimiento en la obtención de hematíes individuales.
2. Un sistema secuencial basado en herramientas de *machine learning* para la detección de malaria y desarrollado a partir de imágenes obtenidas manualmente con microscopio.
3. Un sistema basado en herramientas de *deep learning* para la detección de malaria y desarrollado a partir de imágenes obtenidas manualmente con microscopio.

4. Un sistema basado en herramientas de *deep learning* y desarrollado a partir de imágenes obtenidas con el sistema CellaVision.

Los sistemas desarrollados en la presente tesis doctoral son capaces de reconocer automáticamente hematíes con parásitos de malaria frente a hematíes normales y, además, frente a hematíes con otras inclusiones. Estos sistemas de clasificación suponen un avance prometedor para la identificación automática de pacientes con malaria.

Como consecuencia del trabajo de investigación llevado a cabo en esta tesis doctoral, se han publicado dos artículos científicos en revistas de alto impacto, una de ellas de ámbito clínico y otra en el ámbito de la ingeniería biomédica:

- **Molina A**, Rodellar J, Boldú L, Acevedo A, Alférez S, Merino A. Automatic identification of malaria and other red blood cell inclusions using convolutional neural networks. *Comput Biol Med* 2021.

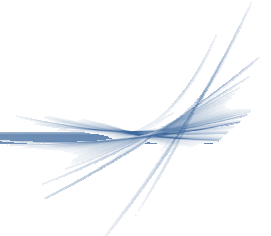
DOI: 10.1016/j.combiomed.2021.104680.

- **Molina A**, Alférez S, Boldú L, Acevedo A, Rodellar J, Merino A. Sequential classification system for recognition of malaria infection using peripheral blood cell images. *J Clin Pathol* 2020.

DOI: 10.1136/jclinpath-2019-206419.

9.3 PERSPECTIVAS FUTURAS

Trabajar en modelos de reconocimiento automático que permitan utilizar imágenes procedentes de frotis de sangre periférica que han sido sometidos a diferentes procesos de tinción. Esto permitiría “universalizar” los modelos de reconocimiento de malaria a los diferentes entornos de trabajo de los laboratorios.



Desarrollar una interfaz a través de un aplicativo web que permita el análisis de imágenes de sangre periférica para el reconocimiento de malaria en tiempo real. Esto permitirá valorar el rendimiento de los modelos desarrollados en condiciones reales de trabajo.

Extender los algoritmos de segmentación y los modelos de reconocimiento automático a otros tipos de alteraciones eritrocitarias con la finalidad de automatizar y estandarizar la revisión de la serie roja.



CAPÍTULO 10

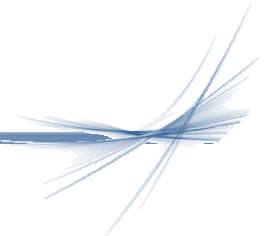
Bibliografía

1. Osei-Bimpong A, Mclean R, Bhonda E, Lewis SM. The use of the white cell count and haemoglobin in combination as an effective screen to predict the normality of the full blood count. *Int J Lab Hematol* 2012;34: 91–97.
2. Barnes PW, McFadden SL, Machin SJ, Simson E. International consensus group for hematology. The international consensus group for hematology review: suggested criteria for action following automated CBC and WBC differential analysis. *Lab Hematol Off Publ Int Soc Lab Hematol* 2005;11:83–90.
3. J Ford. Red blood cell morphology. *Int J Lab Hematol* 2013;35:351-7
4. Keng TB, De La Salle B, Bourner G, Merino A, Han J, Kawai Y, Peng MT, McCafferty R, International Council for Standardization in Haematology (ICSH). Standardization of haematology critical results management in adults: an International Council for Standardization in Haematology, ICSH, survey and recommendations. *Int J Lab Hematol* 2016;38:457-71.
5. World Health Organization. World malaria report 2020. ISBN: 978-92-4-001579-1.
6. Merino A. Manual de Citología de Sangre Periférica y Líquidos Biológicos. Ed. Médica Panamericana, 2020. ISBN: 978-84-9-110262-5.
7. Muñoz J, Rojo-Marcos G, Ramírez-Olivencia G, Salas-Coronas J, Treviño B, Perez Arellano JL, Torrús D, Muñoz Vilches MJ, Ramos JM, Alegría I, López-Vélez R, Aldasoro E, Perez-Molina JA, Rubio JM, Bassat Q. Diagnóstico y tratamiento de la malaria importada en España: recomendaciones del Grupo de Trabajo de Malaria de la Sociedad Española de Medicina Tropical y Salud Internacional (SEMTSI). *Enferm Infecc Microbiol Clin* 2015;33:e1-e13.
8. Constantino BT. Reporting and grading of abnormal red blood cell morphology. *Int J Lab Hematol* 2015;37:1-7.
9. Gutiérrez G, Merino A, Domingo A, Jou JM, Reverter JC. EQAS for peripheral blood morphology in Spain: a 6-year experience. *Int J Lab Hematol* 2008;30:460-6.
10. Díaz G, Manzanera A. Automatic analysis of microscopic images in hematological cytology applications. In *Clinical Technologies: Concepts, Methodologies, Tools and Applications*. 2011:325-52. IGI Global.
11. Shouval R, Fein JA, Savani B, Mohty M, Nagler A. Machine learning and artificial intelligence in haematology. *Br J Haematol* 2021;192:239-50.
12. Alférez S, Merino A, Acevedo A, et al. Color clustering segmentation framework for image analysis of malignant lymphoid cells in peripheral blood. *Med Biol Eng Comput* 2019;57:1265–83.
13. Halim S, Bretschneider TR, Li Y, Preiser PR, Kuss C. Estimating malaria parasitaemia from blood smear images. In: 9th international conference on control, automation, robotics and vision. IEEE; 2006:1–6.
14. Tek FB, Dempster AG, Kale I. Parasite detection and identification for automated thin blood film malaria diagnosis. *Comput Vis Image Underst* 2010;114:21–32.



15. Frejlichowski D. Identification of erythrocyte types in greyscale MGG images for computer-assisted diagnosis. In: Iberian Conference on Pattern Recognition and Image Analysis 2011:636-643.
16. Veluchamy M, Perumal K, Ponuchamy T. Feature extraction and classification of blood cells using artificial neural network. *Am J App Sci* 2012;9:615.
17. Purwar Y, Shah SL, Clarke G, Almugairi A, Muehlenbachs A. Automated and unsupervised detection of malarial parasites in microscopic images. *Malar J* 2011;10:364.
18. Gual-Arnau X, Herold-García S, Simó A. Erythrocyte shape classification using integral-geometry-based methods. *Med Bio Eng Comp* 2015;53:623-633.
19. Maitra M, Gupta RK, Mukherjee M. Detection and counting of red blood cells in blood cell images using Hough transform. *Int J Comput Appl* 2012;53:18–22.
20. Ma C, Harrison P, Wang L, Coppel RL. Automated estimation of parasitaemia of Plasmodium yoelii-infected mice by digital image analysis of Giemsa-stained thin blood smears. *Malar J* 2010;9:348.
21. Tomari R, Zakaria WN, Jamil MMA, Nor FM, Fuad NFN. Computer aided system for red blood cell classification in blood smear image. *Proc Comp Sci* 2014;42:206-213.
22. Chen HM, Tsao YT, Tsai SN. Automatic image segmentation and classification based on direction texton technique for hemolytic anemia in thin blood smears. *Mach Vis Appl* 2014;25:501-510.
23. Lotfi M, Nazari B, Sadri S, Sichani NK. The detection of dacrocyte, schistocyte and elliptocyte cells in iron deficiency anemia. In: 2nd International Conference on Pattern Recognition and Image Analysis 2015:pp. 1-5. IEEE.
24. Lee H, Chen YPP. Cell morphology based classification for red cells in blood smear images. *Pattern Recognit Lett* 2014;49:155-161.
25. Díaz G, González FA, Romero E. A semi-automatic method for quantification and classification of erythrocytes infected with malaria parasites in microscopic images. *J Biomed Inform* 2009;42:296-307.
26. Maity M, Mungle T, Dhane D, Maiti AK, Chakraborty C. An Ensemble Rule Learning Approach for Automated Morphological Classification of Erythrocytes. *J Med Syst* 2017;41:56.
27. Tek FB, Dempster AG, Kale I. Blood cell segmentation using minimum area watershed and circle radon transformations. In: Mathematical morphology. 40 years on:441-454. Springer, Dordrecht. 2005.
28. Jiang P, Zhang X, Wang F. Red blood cell detection by the improved two-layer watershed segmentation method with a full convolutional neural network. *J Med Imaging Health Inform* 2018;8:50-54.
29. Ajala FA, Fenwa OD, Aku MA. A comparative analysis of watershed and edge based segmentation of red blood cells. *Int J Med Biomed Res* 2015;4:1-7.

30. Biswas S, Ghoshal D. Blood cell detection using thresholding estimation based watershed transformation with Sobel filter in frequency domain. *Procedia Comput Sci* 2016;89:651-657.
31. Das D, Ghosh M, Chakraborty C, Pal M, Maity AK. Invariant moment based feature analysis for abnormal erythrocyte recognition. In *2010 International Conference on Systems in Medicine and Biology*: 242-247. IEEE.
32. Khot ST, Prasad RK. Image analysis system for detection of red blood cell disorders using artificial neural network. *Int J Eng Res Technol* 2012;1:1-14.
33. Wheelless LL, Robinson RD, Lapets OP, Cox C, Rubio A, Weintraub M, Benjamin LJ. Classification of red blood cells as normal, sickle, or other abnormal, using a single image analysis feature. *Cytometry* 1994;17:159-166.
34. Abbas N, Saba T, Mohamad D, Rehman A, Almazayad AS, Al-Ghamdi JS. Machine aided malaria parasitemia detection in Giemsa-stained thin blood smears. *Neural Comput Appl* 2017;29:803-18.
35. Das DK, Ghosh M, Pal M, et al. Machine learning approach for automated screening of malaria parasite using light microscopic images. *Micron* 2013;45:97-106.
36. Devi SS, Laskar RH, Sheikh SA. Hybrid classifier based life cycle stages analysis for malaria-infected erythrocyte using thin blood smear images. *Neural Comput Appl* 2018;29:217-35.
37. Linder N, Turkki R, Walliander M, Mårtensson A, Diwan V, Rahtu E, Lundin J. A malaria diagnostic tool based on computer vision screening and visualization of Plasmodium falciparum candidate areas in digitized blood smears. *PLoS One* 2014;9:e104855.
38. Ross NE, Pritchard CJ, Rubin DM, et al. Automated image processing method for the diagnosis and classification of malaria on thin blood smears. *Med Biol Eng Comput* 2006;44:427-36.
39. Sheikhhosseini M, Rabbani H, Zekri M, Talebi A. Automatic diagnosis of malaria based on complete circle-ellipse fitting search algorithm. *Journal of microscopy* 2013;252:189-203.
40. Berry MW, Mohamed A, Yap BW. Supervised and unsupervised learning for data science. Springer Nature 2019. ISBN 978-3-030-22474-5
41. Prasad K, Winter J, Bhat UM, Acharya RV, Prabhu GK. Image analysis approach for development of a decision support system for detection of malaria parasites in thin blood smear images. *J Digit Imaging* 2012;25:542-9.
42. Suradkar PT. Detection of malarial parasite in blood using image processing. *Int J Eng Innov Technol* 2013;2:124-6.
43. Le MT, Bretschneider TR, Kuss C, Preiser PR. A novel semiautomatic image processing approach to determine Plasmodium falciparum parasitemia in Giemsa-stained thin blood smears. *BMC Cell Biol* 2008;9:15-27.
44. Di Ruberto C, Dempster A, Khan S, Jarra B. Analysis of infected blood cell images using morphological operators. *Image Vision Comput* 2002;20:133-46.



45. Sio SW, Sun W, Kumar S, et al. MalariaCount: an image analysis-based program for the accurate determination of parasitemia. *J Microbiol Methods* 2007;68:11–8.
46. Ghosh M, Das D, Chakraborty C, Ray AK. Quantitative characterisation of Plasmodium vivax in infected erythrocytes: a textural approach. *Int J Artificial Intell Soft Comput* 2013;3:203–21.
47. Tek FB, Dempster AG, Kale I. Malaria parasite detection in peripheral blood images. In: BMVC. 2006:347–56.
48. Diaz G, Gonzalez F, Romero E. Infected cell identification in thin blood images based on color pixel classification: comparison and analysis. *Progr Pattern Recogn Image Anal Appl* 2007;812–21.
49. Park HS, Rinehart MT, Walzer KA, Chi JTA, Wax A. Automated detection of P. falciparum using machine learning algorithms with quantitative phase images of unstained cells. *PLoS One* 2016;11:e0163045.
50. Savkare S, Narote S. Automatic detection of malaria parasites for estimating parasitemia. *Int J Comput Sci Sec* 2011;5:310–5.
51. Savkare S, Narote S. Automatic system for classification of erythrocytes infected with malaria and identification of parasite's life stage. *Proc Technol* 2012;6:405–10.
52. Eshel Y, Houry-Yafin A, Benkuzari H, et al. Evaluation of the Parasight platform for malaria diagnosis. *J Clin Microbiol* 2017;55:768–75.
53. Yang D, Subramanian G, Duan J, et al. A portable image-based cytometer for rapid malaria detection and quantification. *PLoS ONE* 2017;12:e0179161.
54. Pamungkas A, Adi K, Gernowo R. Identification of Plasmodium falciparum development phase in malaria infected red blood cells using adaptive color segmentation and decision tree based classification. *Int J Appl Eng Res* 2015;10:4043–55.
55. Soni J, Mishra N, Kamargaonkar N. Automatic difference between RBC and malaria parasites based on morphology with first order features using image processing. *Int J Adv Eng Technol* 2011;1:290–7.
56. Moon S, Lee S, Kim H, et al. An image analysis algorithm for malaria parasite stage classification and viability quantification. *PLoS ONE* 2013;8:e61812.
57. Vink J, Laubscher M, Vlutters R, et al. An automatic vision-based malaria diagnosis system. *J Microsc* 2013;250:166–78.
58. Suryawanshi MS, Dixit V. Improved technique for detection of malaria parasites within the blood cell images. *Int J Sci Eng Res* 2013;4:373–6.
59. W. Rawat, Z. Wang, Deep convolutional neural networks for image classification: a comprehensive review. *Neural Comput* 2017;29:2352–2449.
60. Bakator M, Radosav D. Deep learning and medical diagnosis: A review of literature. *Multimodal Technologies and Interaction* 2018;2:47.

61. Acevedo A, Merino A, Boldú L, Molina A, Alférez S, Rodellar J. A new convolutional neural network predictive model for the automatic recognition of hypogranulated neutrophils in myelodysplastic syndromes. *Comput Biol Med* 2021; 134:104479.
62. Boldú L, Merino A, Acevedo A, Molina A, Rodellar J. A deep learning model (ALNet) for the diagnosis of acute leukaemia lineage using peripheral blood cell images. *Comput Methods Programs Biomed* 2021;202:105999.
63. Merino A, Vlaga A, Molina A, Egri N, Laguna J, Barrera K, Boldú L, Acevedo A, Díaz-Pavón M, Sibina F, Bascón F, Sibila O, Juan M, Rodellar J. Atypical lymphoid cells circulating in blood in covid-19 infection: morphology, immunophenotype and prognosis value. *J Clin Pathol* 2020;75:104-111.
64. Gopakumar GP, Swetha M, SaiSiva G, et al. Convolutional neural network-based malaria diagnosis from focus stack of blood smear images acquired using custom built slide scanner, *J Biophot* 2018;11:e201700003.
65. Liang Z, Powell A, Ersoy I, Poostchi M, Silamut K, Palaniappan K, Guo P, Hossain MA, Sameer A, Maude RJ, Huang JX, Jaeger S, Thoma G. 2017. CNN-based image analysis for malaria diagnosis. In: Proceedings 2016 IEEE international conference on bioinformatics and biomedicine: 493-496.
66. Pan WD, Dong Y, Wu D, Classification of malaria infected cells using deep convolutional neural networks, in: Machine Learning: Advanced Techniques and Emerging Applications, vol. 159, Intech Open, 2018.
67. Dong Y, Jiang Z, Shen H, Pan WD, Williams LA, Reddy VV, Bryan AW. Evaluations of deep convolutional neural networks for automatic identification of malaria infected cells. In 2017 IEEE EMBS international conference on biomedical & health informatics 2017:101-104.
68. Rajaraman S, Jaeger S, Antani SK. Performance evaluation of deep neural ensembles toward malaria parasite detection in thin-blood smear images. *PeerJ* 2019;7:e6977.
69. Rajaraman S, Silamut K, Hossain MA, Ersoy I, Maude RJ, Jaeger S, Antani SK. Understanding the learned behavior of customized convolutional neural networks toward malaria parasite detection in thin blood smear images. *J Med Imaging* 2018;5:034501.
70. Vijayalakshmi A. Deep learning approach to detect malaria from microscopic images. *Multimed Tool Appl* 2020;79:15297–15317.
71. Lee LH, Mansoor A, Wood B, Nelson H, Higa D, Naugler C. Performance of CellaVision DM96 in leukocyte classification. *J Pathol Inform* 2013;4:14
72. Sosnin DY, Onyanova LS, Kubarev OG, Kozonogova EV. Evaluation of efficacy of white blood cell identification in peripheral blood by automated scanning of stained blood smear images with variable magnification. *Biomed Eng* 2018;52:31-36.
73. Münzenmayer C, Schlarb T, Steckhan D, Haßlmeyer E, Bergen T, Aschenbrenner S, Zerfaß, T. HemaCAM—A computer assisted microscopy system for hematology. In *Microelectronic Systems*:233-242. Springer, Berlin, 2011.



74. Bruegel M, George TI, Feng B, Allen TR, Bracco D, Zahniser DJ, Russcher H. Multicenter evaluation of the cobas m 511 integrated hematology analyzer. *Int J Lab Hematol* 2018;40:672-682.
75. Swolin B, Simonsson P, Backman S, Löfqvist I, Bredin I, Johnsson M. Differential counting of blood leukocytes using automated microscopy and a decision support system based on artificial neural networks—evaluation of DiffMaster™ Octavia. *Clin Lab Haematol* 2003;25:139-147.
76. Kratz A, Lee SH, Zini G, Riedl JA, Hur M, Machin S, International Council for Standardization in Haematology. Digital morphology analyzers in hematology: ICSH review and recommendations. *Int J Lab Hematol* 2019;41:437-447.
77. Kratz A, Bengtsson HI, Casey JE, Keefe JM, Beatrice GH, Grzybek DY, Van Cott EM. Performance evaluation of the CellaVision DM96 system: WBC differentials by automated digital image analysis supported by an artificial neural network. *Am J Clin Pathol* 2005;124:770-781.
78. Briggs C, Longair I, Slavik M, Thwaite K, Mills R, Thavaraja V, Machin SJ. Can automated blood film analysis replace the manual differential? An evaluation of the CellaVision DM96 automated image analysis system. *Int J Lab Hematol* 2009;31:48-60.
79. Horn CL, Mansoor A, Wood B, Nelson H, Higa D, Lee LH, Naugler C. Performance of the CellaVision® DM96 system for detecting red blood cell morphologic abnormalities. *J Pathol Inform* 2015;6:11.
80. Palmer L, Briggs C, McFadden S et al. ICSH recommendations for the standardization of nomenclature and grading of peripheral blood cell morphological features. *Int J Lab Hematol* 2015;37:287-303.
81. Racsa LD, Gander RM, Southern PM, McElvania TeKippe E, Doern C, Luu HS. Detection of intracellular parasites by use of the CellaVision DM96 analyzer during routine screening of peripheral blood smears. *J Clin Microbiol* 2015;53:167–171.
82. Egelé A, van Gelder W, Riedl J. Automated detection and classification of schistocytes by a novel RBC module using digital imaging/microscopy. *Int J Lab Hematol* 2015;4:184–6.
83. Egelé A, van Gelder W, Riedl J. Automated detection and classification of teardrop cells by a novel RBC module using digital imaging/ microscopy. *Int J Lab Hematol* 2015;37:153–6.
84. Egelé A, Stouten K, van der Heul-Nieuwenhuijsen L, de Bruin L, Teuns R, van Gelder W, Riedl J. Classification of several morphological red blood cell abnormalities by DM 96 digital imaging. *Int J Lab Hematol* 2016;38:98-101.
85. Criel M, Godefroid M, Deckers B, Devos H, Cauwelier B, Emmerechts J. Evaluation of the Red Blood Cell Advanced Software Application on the CellaVision DM 96. *Int J Lab Hematol* 2016;38:366-374.
86. Florin L, Maelegheer K, Muyldermans A, Van Esbroeck M, Nulens E, Emmerechts J. Evaluation of the CellaVision DM96 advanced RBC application for screening and follow-up of malaria infection. *Diagn Microbiol Infect Dis* 2018;90:253-256.

87. Park M, Hur M, Kim H, Kim HN, Kim SW, Moon HW, Cheong HS. Detection of Plasmodium falciparum using automated digital cell morphology analyzer Sysmex DI-60. *Clin Chem Lab Med* 2018;56:284-287.
88. Yoon J, Kwon JA, Yoon SY, Jang WS, Yang DJ, Nam J, Lim CS. Diagnostic performance of CellaVision DM96 for Plasmodium vivax and Plasmodium falciparum screening in peripheral blood smears. *Acta Trop* 2019;193:7-11.
89. Vives Corrons JL, Albarède S, Flandrin G, et al. Guidelines for blood smear preparation and staining procedure for setting up an external quality assessment scheme for blood smear interpretation. Part I: control material. *Clin Chem Lab Med* 2004;42:922-6.
90. Stéfan van der Walt, Johannes L. Schönberger, Juan Nunez-Iglesias, François Boulogne, Joshua D. Warner, Neil Yager, Emmanuelle Gouillart, Tony Yu and the scikit-image contributors. scikit-image: Image processing in Python. *PeerJ* 2014;2:e453.
91. van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res* 2017;77:e104-7.
92. Fleuret F. Fast binary feature selection with conditional mutual information. *J Mach Learn Res* 2004;5:1531-55.
93. Brown G, Pocock A, Lujan M, Zhao MJ. Conditional Likelihood Maximisation: A unifying framework for information theoretic feature selection. *J Mach Learn Res* 2012;13:27-66.
94. Acevedo A, Alférez S, Merino A, Puigví L, Rodellar J. Recognition of peripheral blood cell images using convolutional neural networks. *Comput Methods Progr Biomed* 2019;180:105020
95. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems*, 2012: 1097-1105.
96. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *ArXiv preprint* 2014:1409-1556.
97. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 2016:770-778.
98. Chollet F. Xception: deep learning with depthwise separable convolutions. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 2017:1251-1258.
99. Szegedy C, Ioffe S, Vanhoucke V, Alemi AA. Inception-v4, inception-resnet and the impact of residual connections on learning. In: *Thirty-first AAAI Conference on Artificial Intelligence* 2017.
100. Huh M, Agrawal P, Efros AA. What makes ImageNet good for transfer learning? *ArXiv preprint* 2016:1608.08614.
101. D.P. Kingma, J.L. Ba, Adam: a method for stochastic optimization, in: *3rd Int Conf Learn Represent ICLR 2015 - Conf Track Proc*, 2015, pp. 1-15.
102. Smith LN. Cyclical learning rates for training neural networks. En *2017 IEEE winter conference on applications of computer vision (WACV)*, IEEE (2017) 464-472.
103. Somasekar J, Reddy B, Reddy E, Lai C. Computer vision for malaria parasite classification in erythrocytes. *Int J Comp Sci and Eng* 2011;3:2251-2256.



104. Subbaswamy A, Saria, S. From development to deployment: dataset shift, causality, and shift-stable models in health AI. *Biostatistics* 2020;21:345-52.
105. Boldú L, Merino A, Alférez S, Molina A, Acevedo A, Rodellar J. Automatic recognition of different types of acute leukaemia in peripheral blood by image analysis. *J Clin Pathol* 2019;72:755-761.
106. Bibin D, Nair MS, Punitha P. Malaria parasite detection from peripheral blood smear images using deep belief networks. *IEEE Access* 2017;5:9099–9108.
107. Yang F, Poostchi M, Yu H, Zhou Z, Silamut K, Yu J, Antani S. Deep learning for smartphone-based malaria parasite detection in thick blood smears. *IEEE J Biomed Health Inform* 2019;24:1427–1438
108. Wongsrichanalai C, Barcus MJ, Muth S, Sutamihardja A, Wernsdorfer WH. A review of malaria diagnostic tools: microscopy and rapid diagnostic test (RDT). *Am J Trop Med Hyg* 2007;77:119-127.



CAPÍTULO 11

Producción científica

11.1 Artículos publicados

Esta tesis doctoral está basada en el compendio de los siguientes artículos:

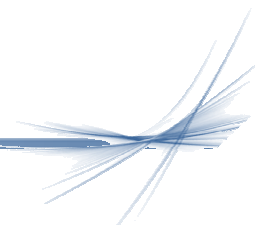
- **Molina A**, Rodellar J, Boldú L, Acevedo A, Alférez S, Merino A. Automatic identification of malaria and other red blood cell inclusions using convolutional neural networks. *Comput Biol Med* 2021. DOI: 10.1016/j.combiomed.2021.104680.
- **Molina A**, Alférez S, Boldú L, Acevedo A, Rodellar J, Merino A. Sequential classification system for recognition of malaria infection using peripheral blood cell images. *J Clin Pathol* 2020. DOI: 10.1136/jclinpath-2019-206419.

Ambos artículos se muestran completamente en el Capítulo 12 (Anexo I).

11.2 Artículos en colaboración con el grupo de investigación

Colaboraciones en publicaciones del grupo de investigación relacionadas con la aplicación de la inteligencia artificial en el reconocimiento automático de imágenes de sangre periférica:

- Acevedo A, Merino A, Boldú L, **Molina A**, Alférez S, Rodellar J. A new convolutional neural network predictive model for the automatic recognition of hypogranulated neutrophils in myelodysplastic syndromes. *Comput Biol Med* 2021.
- Merino A, Laguna J, **Molina A**, Vlagea A, Sibila O. SARS-CoV-2 pneumonia and atypical lymphocyte morphology in pleural fluid. *Int J Lab Hematol* 2021.
- Boldú L, Merino A, Acevedo A, **Molina A**, Rodellar J. A deep learning model (ALNet) for the diagnosis of acute leukaemia lineage using peripheral blood cell images. *Comput Methods Programs Biomed* 2021.
- Merino A, Vlagea A, **Molina A**, Egri N, Laguna J, Barrera K, Boldú L, Acevedo A, Díaz-Pavón M, Sibila F, Bascón F, Sibila O, Juan M, Rodellar J. Atypical lymphoid cells circulating in blood in covid-19 infection: morphology, immunophenotype and prognosis value. *J Clin Pathol* 2020.
- Merino A, **Molina A**, Rodríguez M, Vicente-Steijn R, Tomé A, Maduell F, Xipell M, Castro P. Detection and significance of green inclusions in peripheral blood neutrophils and monocytes. *Int J Lab Hematol* 2020.
- Alcaraz J, **Molina A**, Laguna J, Rodríguez M, Gutiérrez G, Bedini JL, Merino A. Peripheral blood morphology review and diagnostic proficiency evaluation by a new Spanish EQAS during the period 2011-2019. *Int J Lab Hematol* 2020.



- Boldú L, Merino A, Alférez S, **Molina A**, Acevedo A, Rodellar J. Automatic recognition of different types of acute leukaemia in peripheral blood by image analysis. *J Clin Pathol* 2019.
- Rodellar J, Alférez S, Acevedo A, **Molina A**, Merino A. Image processing and machine learning in the morphological analysis of blood cells. *Int J Lab Hematol* 2018. DOI: 10.1111/ijlh.12818.
- Delgado-Ortet M, **Molina A**, Alférez S, Rodellar J, Merino A. A deep learning approach for segmentation of red blood cell images and malaria detection. *Entropy* 2020;22:657.
- Acevedo A, Merino A, Alférez S, **Molina A**, Boldú L, Rodellar J. A dataset of microscopic peripheral blood cell images for development of automatic recognition systems. *Data Brief* 2020.
- Boldú L, **Molina A**. Importancia de la detección de formas eritrocitarias anómalas en el diagnóstico desde el laboratorio clínico. *Ed. Cont. Lab. Clin* 2019;47: 98 – 118.

11.3 Publicaciones en congresos

- Molina A**, Rodellar J, Boldú L, Acevedo A, Alférez S, Merino A. Deep learning approach for malaria recognition using peripheral blood images acquired with cellavision DM96 (XXXIII International Symposium on Technological Innovations in Laboratory Hematology. 2021). **Comunicación premiada por el congreso con el premio “trainee award”.**
- **Molina A**, Rodellar J, Boldú L, Acevedo A, Alférez S, Merino A. Automatic classification of red blood cell inclusions using a deep learning model. (XXXIII International Symposium on Technological Innovations in Laboratory Hematology. 2020). **Comunicación premiada por el congreso con el premio “Travel award”.**
- Boldú L, **Molina A**, García S, Carballo C, Rodríguez M, Laguna J, Bascón F, Merino A. Performance evaluation of the Mindray BC-6800 haematology analyser in comparison with the Advia 2120® and manual review using the Cellavision®DM96. (XXXIII International Symposium on Technological Innovations in Laboratory Hematology. 2020).
- Acevedo A, Merino A, Boldú L, **Molina A**, Alférez S, Rodellar J. A new model for the automatic detection of dysplastic cells in peripheral blood: dysplasianet. (XXXIII International Symposium on Technological Innovations in Laboratory Hematology. 2020).
- Boldú L, Acevedo A, **Molina A**, Alférez S, Rodellar J, Merino A. A deep learning model for the diagnosis of acute leukaemia subtypes using peripheral blood cell images. (XXXIII International Symposium on Technological Innovations in Laboratory Hematology. 2020).
- Merino A, Boldú L, Acevedo A, Alférez S, **Molina A**, Serrando M, Rodellar J. Automatic differentiation of acute leukemia, lymphoma and reactive lymphocytes in peripheral blood

(PB) using a novel convolutional neural network (CNN). (XXXIII International Symposium on Technological Innovations in Laboratory Hematology. 2020).

- Barrera K, Pérez A, Acevedo A, Alférez S, Boldú L, **Molina A**, Rodellar J, Merino A. Automatic generation of artificial images of peripheral blood cells using generative artificial networks (GAN'S). (XXXIII International Symposium on Technological Innovations in Laboratory Hematology. 2020).

- **Molina A**, Alférez S, Boldú L, Acevedo A, Rodellar J, Merino A. Automatic classification of red blood cells infected with malaria by means of digital image analysis. (XXXII International Symposium on Technological Innovations in Laboratory Hematology. 2019). Comunicación seleccionada por el congreso para presentación oral.

- Acevedo A, Merino A, Alférez S, Boldú L, **Molina A**, Rodellar J. Automatic detection of dysplastic cells using deep learning. (XXXII International Symposium on Technological Innovations in Laboratory Hematology. 2019).

- Alférez S, Merino A, Boldú L, Acevedo A, **Molina A**, Rodellar J. A deep learning approach to automatically classify pathological cell images in peripheral blood. (XXXII International Symposium on Technological Innovations in Laboratory Hematology. 2019).

- Boldú L, Alférez S, **Molina A**, Acevedo A, Rodellar J, Merino A. Automatic recognition of different types of acute leukaemias in peripheral blood by image analysis. (XXXII International Symposium on Technological Innovations in Laboratory Hematology. 2019).

- Alférez S, Merino A, Boldú L, **Molina A**, Puigví L, Acevedo A, Rodellar J. A new image-based machine-learning system (cellsimatic) for the automatic recognition of hematologic neoplasia versus infections in peripheral blood. (XXXII International Symposium on Technological Innovations in Laboratory Hematology. 2019).

- **Molina A**, Pladevall C, Alferez S, Rodellar J, Merino A. Automatic classification of sickle cells in peripheral blood through digital image analysis. (XXXI International Symposium on Technological Innovations in Laboratory Hematology. 2018).

- A Merino, J Laguna, L Boldú, **A Molina**, K Barrera, A Acevedo, J Rodellar. Morphology and prognosis value of atypical lymphocytes in peripheral blood incovid-19 patients.IFCC 2021

- Merino A, Vlaga A, **Molina A**, Egri N, Laguna J, Barrera K, Boldú L, Rodellar J. Células linfoides atípicas circulantes en la sangre periférica de pacientes con infección por covid-19: morfología, inmunofenotipo y valor pronóstico. (XIV Congreso Nacional del Laboratorio Clínico. 2020).

- **Molina A**, Torrens N, Alférez S, Rodellar J, Merino A. Clasificación automática de esferocitos y drepanocitos por medio de análisis digital de imágenes. (XII Congreso Nacional del Laboratorio Clínico. 2018).

11.4 Conferencias y cursos impartidos en relación a la tesis doctoral

- Ponente invitado en el curso precongreso: “Nuevas tecnologías en el laboratorio de hematología”.

Título de la ponencia: “Estado actual de la detección automática de inclusiones eritrocitarias y su diagnóstico diferencial”.

Organizado por el Congreso Nacional de Laboratorio Clínico. 2021.

- Ponente invitado en el curso: “Machine Learning y Ciencia de Datos en el Laboratorio Clínico”.

Título de la ponencia: “Introducción a Python” y “Análisis Digital de Imágenes”.

Organizado por la Sociedad Española de Medicina de Laboraotrio (SEQC^{ML}). Septiembre 2022.



CAPÍTULO 12

ANEXO I

Sequential classification system for recognition of malaria infection using peripheral blood cell images

Angel Molina ¹, Santiago Alf  rez,² Laura Bold   ¹, Andrea Acevedo,^{1,3} Jos   Rodellar,³ Anna Merino ^{1,4}

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/jclinpath-2019-206419>).

¹Biochemistry and Molecular Genetics, Biomedical Diagnostic Center, Hospital Cl  nic de Barcelona, Barcelona, Spain

²School of Engineering, Science and Technology, Universidad del Rosario Facultad de Ciencias Naturales y Matem  ticas, Bogot  , Cundinamarca, Colombia

³Matem  ticas CoDALab, Universitat Polit  cnica de Catalunya, Barcelona, Catalunya, Spain

⁴Hospital Cl  nic de Barcelona, Barcelona, Catalunya, Spain

Correspondence to

Dr Angel Molina, Biochemistry and Molecular Genetics, Biomedical Diagnostic Center, Hospital Cl  nic de Barcelona, Barcelona 08036, Spain; molinaborras@gmail.com

Received 28 December 2019

Revised 21 February 2020

Accepted 25 February 2020

Published Online First

16 March 2020

ABSTRACT

Aims Morphological recognition of red blood cells infected with malaria parasites is an important task in the laboratory practice. Nowadays, there is a lack of specific automated systems able to differentiate malaria with respect to other red blood cell inclusions. This study aims to develop a machine learning approach able to discriminate parasitised erythrocytes not only from normal, but also from other erythrocyte inclusions, such as Howell-Jolly and Pappenheimer bodies, basophilic stippling as well as platelets overlying red blood cells.

Methods A total of 15 660 erythrocyte images from 87 smears were segmented using histogram thresholding and watershed techniques, which allowed the extraction of 2852 colour and texture features. Dataset was split into a training and assessment sets. Training set was used to develop the whole system, in which several classification approaches were compared with obtain the most accurate recognition. Afterwards, the recognition system was evaluated with the assessment set, performing two steps: (1) classifying each individual cell image to assess the system's recognition ability and (2) analysing whole smears to obtain a malaria infection diagnosis.

Results The selection of the best classification approach resulted in a final sequential system with an accuracy of 97.7% for the six groups of red blood cell inclusions. The ability of the system to detect patients infected with malaria showed a sensitivity and specificity of 100% and 90%, respectively.

Conclusions The proposed method achieves a high diagnostic performance in the recognition of red blood cell infected with malaria, along with other frequent erythrocyte inclusions.

INTRODUCTION

Malaria is among the most important parasitic diseases in the world caused by different species of *Plasmodium* parasites. WHO reported 219 million cases of malaria in 2017, and over 435 000 deaths.¹ An early diagnosis is essential for effective disease management.² Thin peripheral blood smear (PBS) microscopic examination is a low-expensive and easily accessible tool for malaria diagnosis being the gold standard for species identification,³ for parasite quantification and it is essential to detect cell abnormalities.^{4,5} However, PBS examination is time consuming, needs expert professionals and is prone to interobserver variability.⁶ Moreover, external quality programmes show that the presence of malaria in PBS samples is overlooked by some laboratories.⁷ These drawbacks have contributed to

develop image processing techniques that help to preclassify blood cells^{8,9} contributing to more standardised morphological analysis.^{10–12}

Machine learning methods have been proposed for malaria recognition.¹³ Segmentation is the first step in machine learning approaches. Usually, it is done from the green component of RGB images, using histogram thresholding techniques^{14–16} and applying watershed algorithm.^{17–19} In other works, segmentation is based on granulometry analysis²⁰ or in the use of thresholding from the HSV colour space.²¹ Segmentation is followed by feature extraction, which uses colour, textural and geometric morphological information.²² Features are used to train classifiers, which in most cases are based on support vector machine (SVM),^{19,23–25} k-nearest neighbours (KNN)^{16,17,21,26} or decision trees algorithms.²⁷ Latest trends in classification of red blood cells (RBC) infected with malaria propose the use of deep learning approaches.^{14,17,18,25,28} Literature reveals that the automatic recognition of malaria has been addressed: (1) to discern between infected versus non-infected RBC, (2) to differentiate between *Plasmodium* species or (3) to discriminate among different parasite stages. It is striking that none of the studies have considered other RBC inclusions such as Howell-Jolly (HJ) and Pappenheimer (PP) bodies, basophilic stippling (BS) or platelets (PLT) overlying RBC, which can confuse machine learning tools.

Evaluations of the Advanced RBC Application, a CellaVision software that performs RBC classification and recognises inclusions, reported sensitivity and specificity values of 23.5% and 81.1%, respectively.^{29,30} Furthermore, malaria parasites are commonly misinterpreted as HJ bodies, PP bodies or BS. ARBC algorithms need further improvements, and require postclassification by clinical pathologist.

The objective of this work is to design a new system for the automatic recognition of malaria using digital image analysis and machine learning tools, which must be highly sensitive and specific for clinical application. This methodology aims to discriminate parasitised RBC not only from normal RBC, but also from other RBC inclusions, such as HJ bodies, PP bodies, BS as well as PLT overlying RBC.

MATERIALS AND METHODS

The proposed system is illustrated in figure 1. Inputs are digital RBC images acquired from PBS. After preprocessing, RBC are segmented and



   Author(s) (or their employer(s)) 2020. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Molina A, Alf  rez S, Bold   L, et al. *J Clin Pathol* 2020;**73**:665–670.

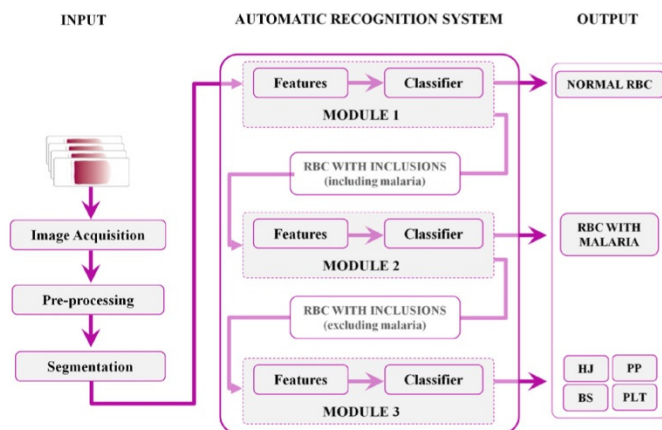


Figure 1 Illustration of the proposed system for the automatic classification of malaria and inclusions in red blood cells (RBC). The left track starts with the acquisition of images and ends up with the segmentation of the regions of interest, which are the inputs to the recognition system. The automatic recognition system includes three sequential modules. The first module recognises normal RBC with respect to those containing inclusions (including malaria). The second identifies whether the inclusions correspond to malaria parasites or not. In a third step, non-malaria inclusions are recognised individually as Howell-Jolly bodies (HJ), pappenheimer bodies (PP), basophilic stippling (BS) or platelets (PLT) overlying RBC.

represented by individual masks. Images and masks are the inputs to the automatic recognition system, which has a hierarchical architecture with three modules working sequentially: (1) the first recognises normal RBC with respect to RBC containing inclusions (including malaria); (2) the second identifies whether the inclusion corresponds to malaria or not and (3) the third discriminates among HJ, PP, BS or PLT. The output of the overall system is the automatic classification of the RBC images in one of the six classes under study.

The core of each module is a set of quantitative features along with a machine-learning classification algorithm. The study is designed in three steps: (1) datasets and image processing; (2) development, where classifiers are trained with selected features and (3) assessment of the overall system performance. Algorithms were developed in the programming environment Python 3.6, with the exception of RBC labelling that was performed in the scientific software MATLAB R2017b.

Datasets and image processing

Image acquisition and labelling

Blood samples, collected in Ethylenediaminetetraacetic acid (EDTA) anticoagulant, from 87 patients were stained with May Grünwald-Giemsa using the slide maker-stainer SP1000i (Sysmex, Kobe, Japan).^{31 32} Field images were acquired using the Olympus BX43 microscope and Olympus DP73 camera. Images

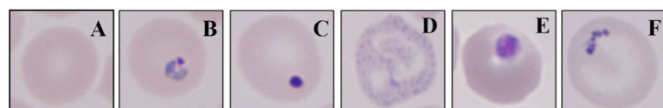


Figure 2 Example of different types of red blood cells (RBC) considered in the study. (A) Normal RBC. (B) RBC infected with malaria parasites. (C) RBC containing Howell-Jolly bodies. (D) RBC with basophilic stippling. (E) RBC with platelets allocated on its surface. (F) RBC containing Pappenheimer bodies.

were taken at 1000 magnifications, 12 bits of colour depth and 2400×1800 pixels. Images were split into two sets of different smears: a training set, to develop the whole system and an assessment set to evaluate the system. This split was made ensuring that both datasets contained images from different patients.

Image labelling was performed by clinical pathologists, using Image Labeller app from MATLAB. It allows to select erythrocytes from field images, providing a marked area that associates the ground truth (true class) and location of every RBC. We considered the following types of images: (1) normal RBC; (2) RBC showing malaria parasites and (3) RBC containing other inclusions (HJ, PP, BS or PLT overlying them). Examples are shown in figure 2.

Image preprocessing and segmentation

Preprocessing aims to enhance image properties to facilitate further operations over the regions of interest (ROIs). The green component was selected from the original RGB image. Contrast between RBC and background was improved using contrast local adaptive histogram equalisation. Gaussian and median filter were used to smooth the image and decrease background noise.

Segmentation was the next step, which gives a separation of the ROIs. Since nucleus is absent in RBC, segmentation is simpler compared with leukocytes.³³ Image was binarised by thresholding using Otsu's method. Post-processing techniques were applied, such as removing noise and PLT, filling holes to deal with the RBC central pallor and clearing borders to select the complete RBC. Watershed method was used to separate adjacent cells, allowing to identify each RBC as an individual element.

Using coordinates stored in the previous labelling step, it was possible to crop individual RBC images and its mask from the original and segmented images.

Feature extraction

Feature extraction from the segmented images of the training set was done using PyRadiomics 2.2.0,³⁴ an open source platform which calculates different types of features: shape, colour (first order statistics) and texture (grey-level co-occurrence matrix, grey-level run-length matrix, grey-level size zone matrix, neighbouring grey tone difference matrix and grey-level dependence matrix features). Features give quantitative information to characterise each ROI, facilitating the subsequent learning and classification steps. Since inclusions are found in RBC with different shapes, extraction was based on the calculation of colour and texture features. They were extracted from each component of different colour spaces included in the python library scikit-image: RGB (red, green, blue), HSV (hue, saturation, value), XYZ (tristimulus values), LAB (lightness from black to white, from green to red, and from blue to yellow), HED (haematoxylin, eosin, dab), YIQ (luminance, in-phase quadrature), LCH (luminance, chroma, hue), YUV (luminance, blue chroma projection, red chroma projection), YCbCr (luminance, blue-difference chroma, red-difference chroma) and CMYK (cyan, magenta, yellow, black), resulting in 2852 features calculated from each RBC.

System development

For each one of the three modules we followed the same procedure. The purpose was to train separately each classifier (see figure 1) to be highly specific for its particular recognition task. Essentially, a classifier uses a set of selected features to predict the class to which the image belongs. Several well-known machine learning algorithms were evaluated: SVM, KNN, linear

discriminant analysis (LDA), random forest and Gaussian Naive Bayes. For each classifier, all images of the training set were used to perform successive tests of fivefold cross-validation for different increasing numbers of features. Features obtained were ordered by relevance according to the conditional mutual information maximisation algorithm.³⁵ Selection of the best classification algorithm and the appropriate number of features was made according to highest overall accuracy, which is defined as the percentage of images well classified over the total number of images.

System assessment

The **assessment set** was divided into two sets (test and diagnostic) to perform an evaluation of the final trained recognition system through two approaches:

1. The **test set** was used to evaluate the performance to classify individual images. This set included images from 18 smears that were not used to develop the model.
2. The **diagnostic set** was used to assess the ability of the system to identify patients infected with malaria from the analysis of a whole PBS. A number of 18 new smears of different patients (not previously used) were included, eight of them corresponding to malaria patients showing different parasitaemia percentages. The remaining 10 were obtained from patients free of malaria, in which other RBC inclusions were observed.

RESULTS

Image processing and segmentation

Preprocessing and segmentation stages allowed to obtain images of individual RBC from microscopic field images. As an example, the different stages are illustrated in **figure 3**. The methodology used in this work showed a segmentation performance of 97.4%, considering a correct segmentation when the resulting mask preserved the original shape of its RBC. As a result, a total of 15 660 RBC images were obtained: 4633 RBC images for the

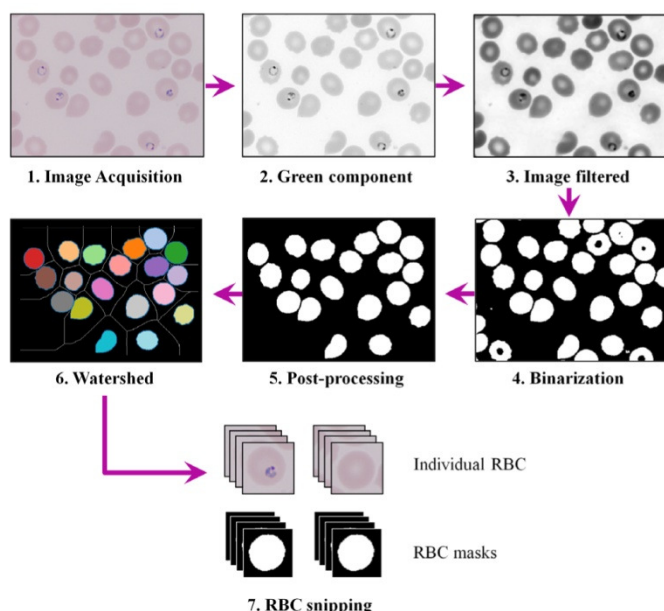


Figure 3 Illustration of the processing methodology used for the automatic classification of red blood cells (RBC), starting from the acquisition of images and ending up with the obtention of individual RBC images and their binary masks.

Table 1 Total number of red blood cell (RBC) images and smears used in this work

		Image dataset			Total
		Training	Test	Diagnostic	
No of smears		51	18	18	87
No of images	nRBC	1276	733	8443	10 452
	MAL	933	187	71	1191
	HJ	578	539	76	1193
	PP	698	256	72	1026
	BS	701	359	65	1125
	PLT	447	162	64	673
Total		4633	2236	8791	15 660

Images are grouped by class for each dataset (training set, test set, and diagnostic set).

BS, basophilic stippling; HJ, Howell-Jolly bodies; MAL, malaria parasites; nRBC, RBCs without inclusions; PLT, platelets; PP, Pappenheimer bodies.

training set and 11 027 RBC images for the assessment set (test and diagnostic sets). The total number of images for each RBC type is shown in **table 1**.

System training

The final system combines three individual modules that work sequentially. Considering the discrimination among normal RBC and RBC containing any type of inclusion, we observed high accuracies with all methods along with low number of features. SVM, together with the seven most relevant features, exhibited the best accuracy (99.12%) using the training set. Regarding the second module, which recognises whether the RBC inclusions correspond to malaria parasites or not, LDA and SVM were the methods with highest accuracies. The combination of LDA with the 610 most relevant features showed accuracy values of 96.52%. LDA and SVM were the methods showing the highest accuracies for the third module to discriminate among HJ or PP bodies, BS and PLT overlying them. The highest accuracy (96.91%) was obtained for LDA along with the 700 most relevant features. Online supplementary figure 1 plots the accuracies with the five machine learning methods mentioned above as a function of different increasing numbers of features.

System assessment using the test set

Classification performance using the test set is summarised by the confusion matrix shown in **table 2**, which gives the sight of

Table 2 Confusion matrix of the classification results (in percentage and absolute values in brackets) using red blood cell (RBC) images of the test set

		Predicted class					
		nRBC	MAL	HJ	PP	BS	PLT
True class	nRBC	98.9 (725)	0	0	0	0.3 (2)	0.8 (6)
	MAL	0	100 (187)	0	0	0	0
	HJ	0	0.4 (2)	97.4 (525)	2.2 (12)	0	0
	PP	0	0.4 (1)	2 (5)	97.7 (250)	0	0
	BS	0.3 (1)	5.3 (19)	0	0	94.4 (339)	0
	PLT	0.6 (1)	1.2 (2)	0	0	0	98.1 (159)

Rows indicate the true class and columns represent the predicted class provided by the system. Diagonal values are the true positive rates.

BS, basophilic stippling; HJ, Howell-Jolly bodies; MAL, malaria parasites; nRBC, RBCs without inclusions; PLT, platelets; PP, Pappenheimer bodies.

Table 3 System evaluation of the diagnostic assessment set using individual smears from eight malaria-infected patients and 10 non-infected patients

	A	Predicted class				B	Predicted class	
		nRBC	MAL	INC			Infected	Non-infected
True class	nRBC	8439	0	4	True class	Infected	8	0
	MAL	0	71	0		Non-infected	1	9
	INC	0	1	276				

A: confusion matrix of the classification results (in absolute number of images) for the diagnostic assessment of the recognition system. Rows indicate the true class and columns represent the predicted class provided by the system. Diagonal values are the true positive rates. B: confusion matrix for true and predicted diagnosis of individual smears tested. Smears are diagnosed as infected if the model recognises at least one single RBC image containing parasites. INC, inclusions; MAL, malaria parasites; nRBC, normal red blood cells.

predicted vs true classes, in percentage and in absolute number of images. The main diagonal shows the sensitivity or true positive rate for each cell class (percentage of images predicted as positive over the total number of images that truly belong to the class): 98.9% for normal RBC, 100% for RBC infected with malaria parasites, 97.4% for HJ bodies, 97.7% for PP bodies, 94.4% for BS and 98.1% for PLT overlying RBC. The mean of these values (97.7%) is the overall accuracy, which is the percentage of single images that have been correctly recognised.

It is remarkable the 100% sensitivity achieved for the recognition of RBC infected with malaria versus all the non-infected RBC types. For this case, we may calculate the specificity as the ratio of images predicted as non-malaria (the sum of all columns in table 2 excluding malaria: 2025) over the number of images that truly belong to any of the non-malaria classes (the sum of all the rows in table 2 except for malaria: 2049). This gave a remarkable specificity of 98.8%.

System assessment for malaria infection using the diagnostic set

This evaluation was focused on the automatic recognition of malaria diagnosis using 18 smears corresponding to new patients whose images were not previously used. Table 3 presents the results for the 18 smears belonging to 8 patients with malaria, with different parasitaemia percentages (from 0.10% to 14.16%); and 10 non-malaria infected patients, but containing inclusions in some of their RBC. For each smear, the system classifies all the images in three classes: normal RBC, malaria or RBC containing other inclusions. The confusion matrices are shown in table 3A. Parasites were detected in all eight smears (71 RBC images with parasites), including those with low parasitaemia percentages. A patient was considered infected (predicted diagnosis) when at least one single RBC containing parasites was detected in the corresponding smear. Table 3B shows that sensitivity of the classification by smear was 100%.

We only found in one smear a single RBC image (with presence of an HJ inclusion) that was misidentified as malaria. According to the established diagnostic criterion, this smear was the only one (over 10) misdiagnosed as infected (90% of specificity). But considering individual RBC images, only one non-infected RBC image over a total of 8720 was misidentified as infected by malaria. Classification rates considering individual RBC images of the diagnostic set resulted into a sensitivity and a specificity of 100% and 99.98%, respectively. Online supplementary tables 1 and 2 show confusion matrices and diagnostic prediction for each specific smear.

DISCUSSION

Previous publications have addressed the malaria recognition problem.¹³ Some of these works described machine learning

approaches to recognise among different species of malaria or among different parasite stages.²² However, these systems have only been designed to address the recognition between infected and non-infected RBC. In real practice, there are other inclusions that should be considered to obtain a sensitive and specific support tool. This remark is important because the recognition of infected samples is a critical result of urgent notification, since it can condition the patient's immediate treatment and prognosis.³⁶

In this study, we developed a recognition system which is able to differentiate not only RBC infected with malaria, but also RBC containing other inclusions. To the authors' knowledge, it is the first time in the literature that this approach has been considered.

With respect to the segmentation of RBC and parasites, other works have used the histogram thresholding, euclidean distance transform,³⁷ edge detection algorithms³⁸ or fuzzy C means.³⁹ Those works that assess the segmentation performance show accuracies over 95%. We have obtained an overall segmentation accuracy of 97.4%. Our approach is based on Otsu's thresholding, which shows a good balance between implementation simplicity and performance.¹³

We evaluated different machine learning approaches for the design of three sequential classification modules, which showed high accuracy rates. Results showed that only a small number of features was required for the first classifier to discriminate between normal and RBC containing inclusions (including malaria). However, a higher number of features was necessary when addressing the recognition between malaria parasites and other inclusions (module two), or among different types of inclusions (module three). This is understandable because of the morphological similarities between the RBC images showing any type of inclusion.

SVM^{19 23–25} and KNN^{16 17 21 26} are among the most frequently used approaches to address the recognition of malaria parasites. In our work, SVM was also the best classifier for the first module. For modules two and three, LDA showed the best performance, while KNN showed lower accuracies. The methodological differences in the segmentation approach with respect to other publications, as well as the different techniques used for the extraction and analysis of features, may give the advantage of some models over the others. LDA has also been reported by our group to have a high accuracy in a machine learning system for the automatic recognition of different types of acute leukaemia.⁴⁰ Previous papers have reported accuracies ranging from 86% to 99%. However, they have not considered other RBC inclusions different from malaria. In our work, an overall accuracy of 97.7% was achieved in the assessment of the system by individual images, but addressing the recognition of malaria along with several other inclusions. This more complete differentiation is important in daily practice to ensure higher specificity in the automatic recognition of malaria.

It is important to consider individual patients when organising datasets for training and assessing the classifier. Using different images from the same smear in both training and assessment could result in an accuracy overestimation. To avoid this issue, we split the initial dataset into two sets of different smears. We only found two previous publications that followed this procedure when performing training/assessment splitting.^{14 24} The classification of the test set images showed a very high performance: none of the normal RBC was classified as containing malaria and none of the infected RBC was classified as normal (sensitivity 100%). This ensures that patients with malaria do not go unnoticed. Besides, it is important to remark that the specificity of the proposed system is 98.8% in spite of the similarity of most of the RBC inclusions under study with respect to malaria parasites, which can easily confuse the classifier.

We also evaluated the ability of the system for the diagnosis of malaria infection on a different set of smears from new patients. The results showed that all the smears from the eight new patients containing RBC parasites were identified as infected. Although a single smear showing other RBC inclusions (due to a HJ body) was identified as infected, in practice, there was only one RBC image misclassified out of 8720 images well recognised as non-infected. These results showed that the approach proposed in this work could be suitable for clinical laboratory practice. The diagnosis by smear showed sensitivity of 100% and specificity of 90%, but considering the classification of individual RBC images, the specificity raised up to 99.98%. This work was performed acquiring images using a camera attached to a microscope. The proposed approach could be also integrated in an automated image acquisition system, which would require an algorithm to find a suitable area of the smear for the proper assessment of RBC morphology. Furthermore, since some cases of malaria show very low levels of parasitaemia, the identification of infected RBC would be conditioned to the number of field images taken. Our system proved to be able to recognise malaria infection with parasitaemia values of 0.1% or higher.

Machine learning tools are becoming increasingly important in the field of laboratory medicine. This work developed a framework for the recognition of malaria in thin PBS digital images that improved specificity of previous approaches, being able to recognise other inclusions in RBC which are not related to malaria infection. The high sensitivity and specificity of our system reduces the number of smears to be reviewed by the clinical pathologist, improving the efficiency of the laboratory workflow.

Take home messages

- Thin blood smear microscopic examination is a low-expensive and easily accessible tool very valuable for malaria diagnosis. Image processing techniques may help to preclassify blood cells contributing to more standardised morphological analysis.
- The contribution of this paper is to provide an automatic recognition system able to discriminate parasitised RBC not only from normal RBC, but also from other erythrocyte inclusions, such as Howell-Jolly bodies, Pappenheimer bodies, basophilic stippling as well as platelets overlying RBC.
- The proposed system showed high sensitivity and specificity and it would be a practical tool to assist the pathologist in the diagnosis of malaria during the morphological peripheral blood examination.

Handling editor Mary Frances McMullin.

Contributors All authors have contributed to the development of the present work.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as online supplementary information.

ORCID iDs

Angel Molina <http://orcid.org/0000-0002-9584-3646>

Laura Boldú <http://orcid.org/0000-0002-5162-3182>

Anna Merino <http://orcid.org/0000-0002-1889-8889>

REFERENCES

- 1 World Health Organization. *World malaria report 2018*. ISBN: 978-92-4-156565-3.
- 2 Muñoz J, Rojo G, Ramírez G, et al. Diagnosis and treatment of imported malaria in Spain: recommendations from the malaria Working group of the Spanish Society of tropical medicine and international health (SEMTRI). *Enferm Infecc Microbiol Clin* 2015;33:e1–13.
- 3 World Health Organization. *Malaria microscopy quality assurance manual, version 2* (ISBN 978-92-4-154939-4), 2016.
- 4 Merino A. *Manual de Citología de Sangre Periférica*. Madrid: Acción Médica, 2005.
- 5 Bain BJ. Diagnosis from the blood smear. *N Engl J Med* 2005;353:498–507.
- 6 Constantino BT. Reporting and grading of abnormal red blood cell morphology. *Int J Lab Hematol* 2015;37:1–7.
- 7 Gutiérrez G, Merino A, Domingo A, et al. EQAS for peripheral blood morphology in Spain: a 6-year experience. *Int J Lab Hematol* 2008;30:460–6.
- 8 Diaz G, Manzanera A. Automatic analysis of microscopic images in hematological cytology applications. In *Clinical Technologies: Concepts, Methodologies, Tools and Applications* 2011:325–52.
- 9 Acevedo A, Alferez S, Merino A, et al. Recognition of peripheral blood cell images using convolutional neural networks. *Comput Methods Programs Biomed* 2019;180:105020.
- 10 Kratz A, Lee S-H, Zini G, et al. Digital morphology analyzers in hematology: ICSH review and recommendations. *Int J Lab Hematol* 2019;41:437–47.
- 11 Merino A, Puigvi L, Boldú L, et al. Optimizing morphology through blood cell image analysis. *Int J Lab Hematol* 2018;40:54–61.
- 12 Rodellar J, Alferez S, Acevedo A, et al. Image processing and machine learning in the morphological analysis of blood cells. *Int J Lab Hematol* 2018;40:46–53.
- 13 Poostchi M, Silamut K, Maude RJ, et al. Image analysis and machine learning for detecting malaria. *Transl Res* 2018;194:36–55.
- 14 Ross NE, Pritchard CJ, Rubin DM, et al. Automated image processing method for the diagnosis and classification of malaria on thin blood smears. *Med Biol Eng Comput* 2006;44:427–36.
- 15 Moon S, Lee S, Kim H, et al. An image analysis algorithm for malaria parasite stage classification and viability quantification. *PLoS One* 2013;8:e61812.
- 16 Prasad K, Winter J, Bhat UM, et al. Image analysis approach for development of a decision support system for detection of malaria parasites in thin blood smear images. *J Digit Imaging* 2012;25:542–9.
- 17 Devi SS, Laskar RH, Sheikh SA. Hybrid classifier based life cycle stages analysis for malaria-infected erythrocyte using thin blood smear images. *Neural Comput Appl* 2018;29:217–35.
- 18 Gopakumar GP, Swetha M, Sai Siva G, et al. Convolutional neural network-based malaria diagnosis from focus stack of blood smear images acquired using custom-built slide scanner. *J Biophotonics* 2018;11:e201700003.
- 19 Das DK, Ghosh M, Pal M, et al. Machine learning approach for automated screening of malaria parasite using light microscopic images. *Micron* 2013;45:97–106.
- 20 Di Ruberto C, Dempster A, Khan S, et al. Analysis of infected blood cell images using morphological operators. *Image Vis Comput* 2002;20:133–46.
- 21 Diaz G, Gonzalez FA, Romero E. Infected cell identification in thin blood images based on color pixel classification: comparison and analysis. *Progress in Pattern Recognition, Image Analysis and Applications* 2007:812–21.
- 22 Das DK, Mukherjee R, Chakraborty C. Computational microscopic imaging for malaria parasite detection: a systematic review. *J Microsc* 2015;260:1–19.
- 23 Yang D, Subramanian G, Duan J, et al. A portable image-based cytometer for rapid malaria detection and quantification. *PLoS One* 2017;12:e0179161.
- 24 Linder N, Turkki R, Walliander M, et al. A malaria diagnostic tool based on computer vision screening and visualization of Plasmodium falciparum candidate areas in digitized blood smears. *PLoS One* 2014;9:e104855.

- 25 Diaz G, González FA, Romero E. A semi-automatic method for quantification and classification of erythrocytes infected with malaria parasites in microscopic images. *J Biomed Inform* 2009;42:296–307.
- 26 Tek FB, Dempster AG, Kale İzzet. Parasite detection and identification for automated thin blood film malaria diagnosis. *Comput Vis Image Underst* 2010;114:21–32.
- 27 Sheikhsosseini M, Rabbani H, Zekri M, et al. Automatic diagnosis of malaria based on complete circle-ellipse fitting search algorithm. *J Microsc* 2013;252:189–203.
- 28 Pan WD, Dong Y, Wu D. Classification of malaria-infected cells using deep convolutional neural networks, in machine learning: advanced techniques and emerging applications. *Intech Open* 2018;159.
- 29 Florin L, Maelegheer K, Muyldermans A, et al. Evaluation of the CellaVision DM96 advanced RBC application for screening and follow-up of malaria infection. *Diagn Microbiol Infect Dis* 2018;90:253–6.
- 30 Racsa LD, Gander RM, Southern PM, et al. Detection of intracellular parasites by use of the CellaVision DM96 analyzer during routine screening of peripheral blood smears. *J Clin Microbiol* 2015;53:167–71.
- 31 Piaton E, Fabre M, Goubin-Versini I, et al. [Technical recommendations and best practice guidelines for May-Grünwald-Giemsa staining: literature review and insights from the quality assurance]. *Ann Pathol* 2015;35:294–305.
- 32 Vives Corrons J-L, Albarède S, Flandrin G, et al. Guidelines for blood smear preparation and staining procedure for setting up an external quality assessment scheme for blood smear interpretation. Part I: control material. *Clin Chem Lab Med* 2004;42:922–6.
- 33 Alférez S, Merino A, Acevedo A, et al. Color clustering segmentation framework for image analysis of malignant lymphoid cells in peripheral blood. *Med Biol Eng Comput* 2019;57:1265–83.
- 34 van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res* 2017;77:e104–7.
- 35 Brown G, Pocock A, Zhao M-J, et al. Conditional likelihood maximisation: a unifying framework for information theoretic feature selection. *J Mach Learn Res* 2012;13:27–66.
- 36 Keng TB, De La Salle B, Bourner G, et al. Standardization of haematology critical results management in adults: an international Council for standardization in haematology, ICSH, survey and recommendations. *Int J Lab Hematol* 2016;38:457–71.
- 37 Le M-T, Bretschneider TR, Kuss C, et al. A novel semi-automatic image processing approach to determine Plasmodium falciparum parasitemia in Giemsa-stained thin blood smears. *BMC Cell Biol* 2008;9:15.
- 38 Purwar Y, Shah SL, Clarke G, et al. Automated and unsupervised detection of malarial parasites in microscopic images. *Malar J* 2011;10:364.
- 39 Somasekar J, Eswara Reddy B, Reddy BE. Segmentation of erythrocytes infected with malaria parasites for the diagnosis using microscopy imaging. *Comput Electr Eng* 2015;45:336–51.
- 40 Boldú L, Merino A, Alférez S, et al. Automatic recognition of different types of acute leukaemia in peripheral blood by image analysis. *J Clin Pathol* 2019;72:755–61.

A		Predicted class			
		nRBC	MAL	INC	
Smear 1	True class	nRBC	200	0	0
		MAL	0	33	0
		INC	0	0	0
Smear 2		nRBC	200	0	0
		MAL	0	19	0
		INC	0	0	0
Smear 3		nRBC	200	0	0
		MAL	0	7	0
		INC	0	0	0
Smear 4		nRBC	198	0	0
		MAL	0	4	0
		INC	0	0	0
Smear 5		nRBC	499	0	0
		MAL	0	3	0
		INC	0	0	0
Smear 6		nRBC	500	0	0
		MAL	0	2	0
		INC	0	0	0
Smear 7		nRBC	999	0	0
		MAL	0	2	0
		INC	0	0	0
Smear 8		nRBC	1,000	0	0
		MAL	0	1	0
		INC	0	0	0
TOTAL		nRBC	3,796	0	0
		MAL	0	71	0
		INC	0	0	0

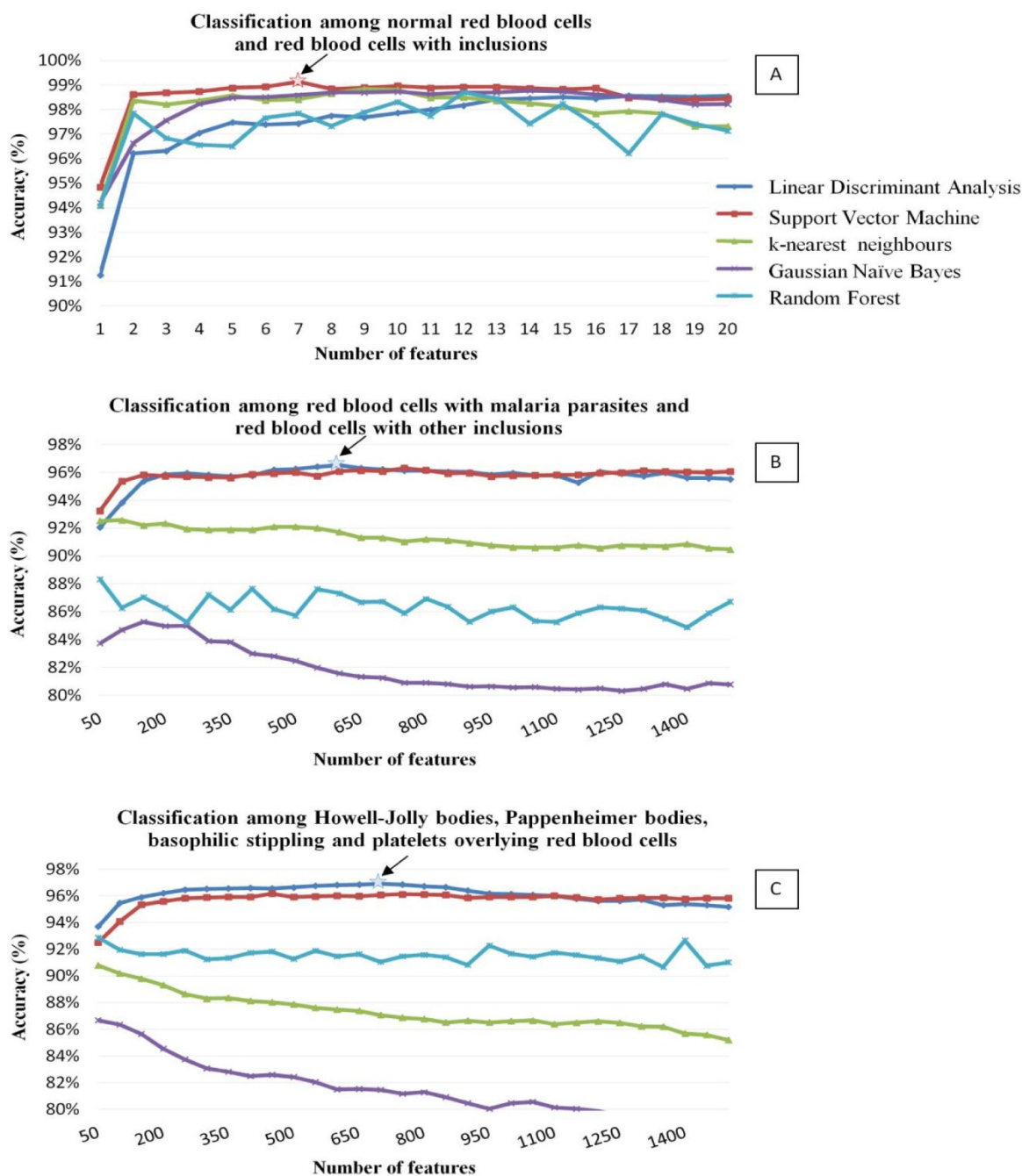
B		TRUE DIAGNOSIS	PREDICTED DIAGNOSIS
Smear 1			
Parasitaemia: 14.20%		INFECTED	INFECTED
Smear 2			
Parasitaemia: 8.70%		INFECTED	INFECTED
Smear 3			
Parasitaemia: 3.40%		INFECTED	INFECTED
Smear 4			
Parasitaemia: 2.00%		INFECTED	INFECTED
Smear 5			
Parasitaemia: 0.60%		INFECTED	INFECTED
Smear 6			
Parasitaemia: 0.40%		INFECTED	INFECTED
Smear 7			
Parasitaemia: 0.20%		INFECTED	INFECTED
Smear 8			
Parasitaemia: 0.10%		INFECTED	INFECTED

Supplementary Table 1. System evaluation using the infected samples of the diagnostic assessment set. **A:** Confusion matrices of the classification results (in absolute number of images) for the diagnostic assessment of the recognition system. Rows indicate the true class and columns represent the class predicted by the system for each smear analyzed. Diagonal values are the true positive rates. nRBC: normal red blood cells (RBC). MAL: RBC containing malaria parasites; INC: RBC containing other inclusions, excluding malaria. **B:** True and predicted diagnosis for individual smears tested. Smears are diagnosed as infected if the model recognizes at least one single RBC image containing parasites.

A		Predicted class			
		nRBC	MAL	INC	
Smear 9	True class	nRBC	500	0	0
		MAL	0	0	0
		INC	0	0	54
Smear 10		nRBC	500	0	0
		MAL	0	0	0
		INC	0	0	38
Smear 11		nRBC	498	0	1
		MAL	0	0	0
		INC	0	0	52
Smear 12		nRBC	442	0	2
		MAL	0	0	0
		INC	0	1	39
Smear 13		nRBC	312	0	0
		MAL	0	0	0
		INC	0	0	12
Smear 14		nRBC	533	0	0
		MAL	0	0	0
		INC	0	0	18
Smear 15	nRBC	601	0	1	
	MAL	0	0	0	
	INC	0	0	27	
Smear 16	nRBC	448	0	0	
	MAL	0	0	0	
	INC	0	0	36	
Smear 17	nRBC	393	0	0	
	MAL	0	0	0	
	INC	0	0	0	
Smear 18	nRBC	416	0	0	
	MAL	0	0	0	
	INC	0	0	0	
TOTAL	nRBC	4,643	0	4	
	MAL	0	0	0	
	INC	0	1	276	

B		TRUE DIAGNOSIS	PREDICTED DIAGNOSIS
Smear 9	Pappenheimer	NON INFECTED	NON INFECTED
Smear 10	Basophilic stippling	NON INFECTED	NON INFECTED
Smear 11	Platelets overlying RBC	NON INFECTED	NON INFECTED
Smear 12	Howell Jolly	NON INFECTED	INFECTED
Smear 13	Platelets overlying RBC	NON INFECTED	NON INFECTED
Smear 14	Pappenheimer bodies	NON INFECTED	NON INFECTED
Smear 15	Basophilic stippling	NON INFECTED	NON INFECTED
Smear 16	Howell Jolly bodies	NON INFECTED	NON INFECTED
Smear 17	Normal RBCs	NON INFECTED	NON INFECTED
Smear 18	Normal RBCs	NON INFECTED	NON INFECTED

Supplementary Table 2. System evaluation of the diagnostic assessment set using individual smears for non-infected patients. **A:** Confusion matrices of the classification results (in absolute number of images) for the diagnostic assessment of the recognition system. Rows indicate the true class and columns represent the predicted class provided by the system. Diagonal values are the true positive rates. nRBC: normal red blood cells (RBC). MAL: RBC containing malaria parasites; INC: RBC containing other inclusions, excluding malaria. **B:** True and predicted diagnosis for individual smears tested. Smears are diagnosed as infected if the model recognizes at least one single RBC image containing parasites.



Supplementary Figure 1. Graphs showing the overall accuracies (Y-axis) obtained with each classification algorithm for every combination of the most relevant features (X-axis). The highest classification accuracy is marked with an arrow. A) Model that recognizes normal red blood cells (RBC) with respect to those containing malaria parasites or other inclusions. B) Model that differentiates between malaria parasites with respect to other inclusions. C) Model that differentiates among RBC inclusions, excluding malaria parasites.



Automatic identification of malaria and other red blood cell inclusions using convolutional neural networks

Angel Molina^{a,*}, José Rodellar^b, Laura Boldú^a, Andrea Acevedo^{a,b}, Santiago Alférez^c, Anna Merino^a

^a Biochemistry and Molecular Genetics Department, Biomedical Diagnostic Center, Hospital Clínic of Barcelona, Barcelona, Catalonia, Spain

^b Department of Mathematics, Barcelona East Engineering School, Technical University of Catalonia, Barcelona, Catalonia, Spain

^c Universidad Del Rosario. Faculty of Natural Science and Mathematics, Colombia

ARTICLE INFO

Keywords:

Malaria
Erythrocyte
Peripheral blood smear
Digital image processing
Deep learning
Convolutional neural networks

ABSTRACT

Malaria is a serious disease responsible for thousands of deaths each year. Many efforts have been made to aid in the diagnosis of malaria using machine learning techniques, but to date, the presence of other elements that may interfere with the recognition of malaria has not been considered. We have developed the first deep learning model using convolutional neural networks capable of differentiating malaria-infected red blood cells from not only normal erythrocytes but also erythrocytes with other types of inclusions. 6415 images of red blood cells were segmented from digital images of 53 peripheral blood smears using thresholding and watershed transformation techniques. These images were used to train a VGG-16 architecture using transfer learning. Using an independent test set of 23 smears, this model was 99.5% accurate in classifying malaria parasites and other red blood cell inclusions. This model also exhibited sensitivity and specificity values of 100% and 91.7%, respectively, classifying a complete smear as infected or not infected. Our model represents a promising advance for automation in the identification of malaria-infected patients. The differentiation between malaria parasites and other red blood cell inclusions demonstrates the potential utility of our model in a real work environment.

1. Introduction

Malaria is a life-threatening disease caused by the protozoan parasite *Plasmodium*. It is an acute febrile illness with an incubation period of 7 days or more, and its severe forms can lead to organ failure, circulatory collapse, coma, and death. Severe malaria is often associated with increased mortality. According to the World Health Organization, an estimated 228 million cases of malaria occurred worldwide in 2018, and children under the age of 5 accounted for 67% of all malaria deaths [1].

The laboratory gold standard in the diagnosis of malaria is based on microscopic visualization of the parasite in blood smears [2]. However, the examination of peripheral blood smear (PBS) under the microscope is time consuming and requires a great deal of experience to avoid inconsistent results. In recent years, significant advances have been reported in digital image processing associated with clinical laboratory diagnosis, the product of collaboration between pathologists, mathematicians, and engineers [3].

Generally, two types of smears (thick and thin) are used depending

on the purpose of the analysis. A thick smear consists of a drop of blood on a glass slide, whereas a thin smear is the result of spreading the drop across the slide. Thick smears are more sensitive, and they theoretically allow visualization of parasites even in cases with low levels of parasitemia. Thus, they are more useful for detecting the presence of parasites; however, they are only performed in cases in which there is a previous suspicion of malaria infection. A recent work [4] developed a deep learning algorithm implemented in smartphones for the automatic detection of malaria parasites in thick smears with 93.46% accuracy. The work also included a review of other machine learning approaches to thick smear detection.

Thin PBS allows the visualization of morphological details that are important for diagnosis and for the initiation of treatment [5]. Furthermore, thin PBS often complements the complete blood count, one of the most requested laboratory tests, and it is not uncommon to observe malaria parasites in PBS as a casual finding in clinical laboratories.

The methodology proposed in this work is based on thin smears,

* Corresponding author. Biochemistry and Molecular Genetics Department. Biomedical Diagnostic Center, Hospital Clínic of Barcelona. Villarroel 170, Barcelona, 08036, Spain.

E-mail address: amolinas@clinic.cat (A. Molina).

<https://doi.org/10.1016/j.combiomed.2021.104680>

Received 26 March 2021; Received in revised form 19 July 2021; Accepted 20 July 2021

Available online 22 July 2021

0010-4825/© 2021 Elsevier Ltd. All rights reserved.

because our objective is to develop a deep learning tool to help in the review of PBS, which helps to detect the presence of malaria parasites, especially in those cases in which malaria infection is not suspected. Our system considers the presence of other elements that can interfere with the automatic recognition of malaria, such as other red blood cell (RBC) inclusions.

1.1. Related works

Recently, different research groups have addressed the automatic recognition of malaria-infected RBCs using digital images of blood smears. The classic approach includes the following steps: image segmentation, feature extraction/selection, and classification. The goal of image segmentation is to automatically detect all RBCs from a field-of-view image. One of the most commonly used techniques for background subtraction is the Otsu algorithm of the green histogram [6–8]. Once the RBCs have been identified from the rest of the image, the watershed algorithm is also used to separate the overlapping RBCs as individual elements [9–11]. These methods have shown good performance in detecting RBCs. However, other works have demonstrated that alternative segmentation techniques can be used, such as the Euclidean distance transform [12], edge detection algorithms [13], and fuzzy C means [14]. Deep learning techniques have also been proposed as an accurate method for RBC segmentation [15].

RBC images can be described quantitatively by means of “hand-crafted” features. Typically, three types of features are used: geometric, color, and texture [16]. Since infected RBCs come in different shapes, the geometric features only appear to be useful for analyzing segmented parasites. By contrast, the color and texture features are preferable choices in most works, because these descriptors seem to be better candidates to highlight the specific staining properties of the parasites, especially the texture characteristics, which describe spatial patterns of color intensities. Additionally, these features can be used in different color spaces, highlighting complementary information provided by alternative colors. As a last step, a mathematical discrimination method uses features to classify the segmented objects into the different classes of interest.

Prasad et al. [8] developed a support decision system using three features of color and texture that detects 96% of parasites but with a 20% false-positive rate. Tek et al. [17] compared the classification performance of the K-nearest neighbor classifier (KNN) and Fisher’s linear discriminant algorithms. They developed a KNN support system with 93% accuracy, which is higher than the accuracy obtained for linear systems. However, the selection of the feature parameter values was based solely on the KNN results. Devi et al. [9] found that artificial neural networks or KNNs outperformed the classification rate of support vector machine (SVM) or Bayes classifiers. However, they found that the combination of SVM, KNN, and artificial neural networks with the majority voting technique provided higher performance, with an accuracy of 96.5%, than to that due to using only one of them. Yang et al. [18] used a three-dimensional feature to train an SVM with less than 5% error detection. Linder et al. [19] used digitized smears to extract local binary patterns and SIFT features from segmented RBCs. The images were used to train an SVM to detect infected RBCs versus normal RBCs. The researchers obtained 97% accuracy with the proportion of RBCs that reflects a number of normal and parasitized RBCs from real cases. However, a 74.2% positive predictive value was obtained. Díaz et al. [20] used characteristics of different color spaces to compare different algorithms. They obtained the best results with the KNN algorithm, with F1 score values of 0.99 for the detection of erythrocytes and 0.76 for the detection of parasites. In another work, Díaz et al. also obtained F1 score of 0.96 using SVM on segmented images using a template matching strategy [21]. Sheikhhosseini et al. [22] used decision rules to detect the presence of malaria parasites inside RBCs, with 97.3% accuracy and a 68% positive predictive value. In previous work, our group also addressed the problem of malaria identification by comparing different

classical machine learning algorithms [23]. We found that a large number of features can be used in linear discriminant analysis (LDA) or SVM to obtain 97.7% accuracy and 98.8% specificity to discriminate infected RBCs between RBCs with other types of inclusions, as well as normal RBCs.

In the last decade, there has been a significant increase in the use of deep learning, a subset of machine learning inspired by the structure and function of the human brain comprised of multilayer neural networks. Deep learning does not rely on handcrafted functions, which saves most of the computational steps. However, more sophisticated hardware resources and larger data sets are required to train a model. Initially, the first works that applied deep learning in the recognition of malaria used multilayer perceptron with a performance equivalent to the classic machine learning models [6,19]. However, recent studies have reported that convolutional neural networks (CNN) show higher performances in the recognition and automatic classification of images [24]. Pan et al. used one of the simplest and most direct CNN architectures, called LeNet-5, to differentiate infected RBCs from uninfected ones with 95% accuracy [25]. Gopakumar et al. trained a customized four-layer CNN to detect infected RBCs from images obtained from a slide scanner. They reached sensitivity and specificity values of 97.1% and 98.5%, respectively [10]. Vijayalakshmi et al. obtained 93% accuracy in the automatic recognition of malaria by modifying the structure of a VGG-19 model, where the fully connected layers were replaced by SVM with a radial basis function kernel [26]. Rajaraman et al. compared several state-of-the-art CNNs and found that the combination of a VGG-19 and SqueezeNet outperformed the recognition of infected versus uninfected RBCs, using segmented images that removed background details from the source [27].

1.2. Contribution of this work

The aim of previous publications in the literature on automatic recognition of malaria has been exclusively the detection of malaria parasites. However, RBCs may also contain other types of inclusions associated with red cell abnormalities [28], which have not been considered in the differential diagnosis with respect to malaria parasites. These RBC inclusions show a morphology similar to that of parasites; therefore, classifiers trained in the automatic recognition of normal and malaria-infected RBCs could misidentify other inclusions as parasites. As a result, a higher false-positive rate will increase the number of smears that pathologists must review, reducing laboratory efficiency and turn-around times.

In this work, we propose a CNN model for the automatic identification of RBCs infected with malaria parasites, which is able to differentiate them from not only normal RBCs but also RBCs that contain other types of inclusions, such as Howell-Jolly bodies (HJ), Pappenheimer bodies (PP), basophilic stippling (BS), and platelets distributed on the surface of RBCs (PLT). All RBC inclusions can be differentiated by their morphological characteristics [29]. Malaria parasites are microorganisms recognizable by their blue ring-shaped cytoplasm containing a purplish nucleus. HJs are fragments of nuclear material (DNA) and are typically single, small, dense, and perfectly round basophilic inclusions. PPs are aggregates of ferritin in red cells, which appear as multiple basophilic inclusions of variable size and shape and are normally distributed in a limited cytoplasmic area. BSs are aggregated ribosomes that appear as fine blue granules uniformly distributed throughout the RBC. PLTs are fragments of pale blue cytoplasm that contain numerous fine, purple granules. An example of all these types of inclusions can be found in Fig. 1.

2. Materials and methods

2.1. Overview

Fig. 1 illustrates the general workflow of the proposed system. The

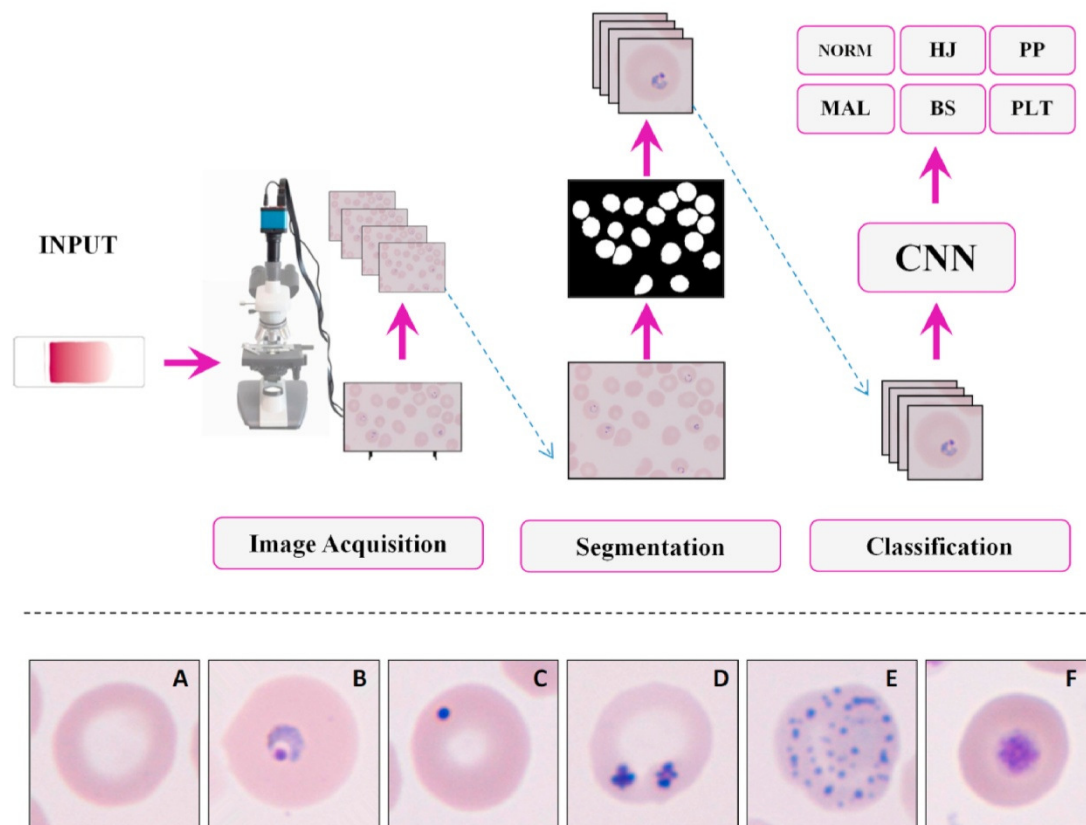


Fig. 1. Illustration of the proposed system for the automatic classification of malaria and inclusions in red blood cells (RBCs) (above). Example of the different target RBCs (below): **A)** Normal RBC (NORM). **B)** RBC infected with malaria parasites (MAL). **C)** RBC containing Howell–Jolly bodies (HJs). **D)** RBC containing Pappenheimer bodies (PPs). **E)** RBC with basophilic stippling (BS). **F)** RBC with platelet allocated on its surface (PLT).

workflow begins with the acquisition of thin PBS images. The microscope allows you to see a small region of the smear (field-of-view), where various blood cells (leukocytes, RBCs, and platelets) can be found. Therefore, it is necessary to segment the RBCs from each field-of-view image. Finally, the images of individual RBCs are the inputs to the CNN-based recognition model that classifies them into one of the following six classes: 1) normal RBCs, 2) RBCs containing malaria parasites, 3) RBCs showing HJ, 4) RBCs showing PP, 5) RBC showing BS, and 6) RBC containing PLT on their surface.

The entire pipeline was done using the Python 3.6 open source software. The Scikit image library was used for image processing [30]. The development of the model was carried out using the FastAI library [31] and an Nvidia Titan XP GPU.

The remainder of this section describes the procedures and image data sets used in this work. Regarding the classification model, we separate the details and the data sets into two subsections: one dedicated to development (see details of training the model in Section 2.5) and one that describes two different approaches designed to evaluate model performance (see details of the model test in Section 2.6).

2.2. Image acquisition

Peripheral blood samples were collected from 76 different patients in EDTA tubes. The samples were used to prepare thin PBS stained with May Grünwald-Giemsa. This process was fully automated with the SP1000i slide maker-stainer (Sysmex, Kobe, Japan). Smear staining was performed following international recommendations [32]. The slides were reviewed using an Olympus BX43 microscope at 1,000 $\times$ magnification. Images were captured with an attached camera (Olympus DP73) and acquired in JPEG format with a resolution of 2400 $\times$ 1800 pixels and a color depth of 12 bits. Images of individual RBCs were labeled by

clinical pathologists using the MATLAB R2017b Image Labeller application. Each label stored the inclusion type along with its field image location coordinates. The coordinates were then used to associate the segmented RBCs with their corresponding inclusion type.

2.3. Segmentation

Pre-processing is required to remove unwanted noise and adjust images for subsequent steps. The green channel of the RGB color space has been reported to be the best component for differentiating RBCs from a blood smear field image, as it offers the best contrast between the RBCs and the background [33]. Then, a median filter was applied, which is a useful technique for reducing random noise, especially in periodic patterns. This filter replaces each pixel value with the mean values of its neighboring pixels.

Segmentation was performed to obtain images of individual RBCs from field-of-view images. Otsu's method exhaustively searches for the histogram threshold that minimizes intraclass variance while maximizing interclass variance. This segmentation technique strikes a good balance between simplicity of implementation and good performance. Through mathematical morphology, it is possible to remove unwanted objects (platelets and leukocytes), fill holes to treat the central area of RBCs, and remove fragmented RBCs from the edges of the image. Finally, overlapping erythrocytes were addressed using the watershed method, separating adjacent catchments surrounding regional minima. Highly overlapping RBCs that were not separated by the watershed algorithm were excluded, as they might hamper proper examination of morphological characteristics. Overlapping RBCs could be easily excluded depending on their area. For this study, only simple or slightly overlapping RBCs were considered. The center of mass of each ROI (centroid) was used to obtain individual 400 $\times$ 400 pixel RBC images.

2.4. Dataset arrangement

All smears collected for the study were divided into training, validation, and testing data sets. On the one hand, the model development used training and validation sets. The training set was used to fit the model, whereas the validation set was used to provide an evaluation at the end of each epoch, which helped establish the hyperparameters of the model. The validation images were always different from the training images, although they could come from the same smears. With regard to the test set, it is important that it be composed of images of a complete smear set that have not been previously used in the model development. This is why of the 76 smears, 23 were reserved for testing to ensure a reliable evaluation. The data set for the development of the model was organized with images obtained from the remaining 53 smears and divided into training and validation sets, as detailed in Table 1. The number of images in the validation set was defined from the number total of images destined to the development of the model. The imaging division of all types of RBC inclusions was defined at 20%. On the contrary, to maintain a delicate balance between the proportionality of the validation set and the superiority of the presence of normal RBCs, this class was defined with a division of 15%.

The images of the training data set were augmented until 2000 images were obtained for each class and after dividing the data set into training and validation data. This procedure easily generates more than 10 new images by applying modifications to an original image (random level of image rotation, zoom, and/or luminance variation). It results in an increase in the data set without using repeated images, balances the training data set, improves the generalizability of the model, and avoids overfitting [34]. Fig. 2 presents an example of images generated by increasing data. The magnification procedure was carried out entirely on the training data set and not on the validation data set, to ensure that the magnified images were not used in both the training and validation phases.

2.5. CNN model design and training

Convolutional Neural Networks (CNNs) have proven to be very useful in image recognition. They are multi-layered networks that attempt to mimic the neural connectivity found in the visual cortex of the brain and are designed to recognize visual patterns. In this work, different CNN architectures were investigated for the development of the model: AlexNet, ResNet, SqueezeNet, DenseNet, VGG, Xception, and Inception [35–39]. CNNs are made up of two main blocks. The first part filters images with multiple convolutional cores and returns feature maps. This process can be repeated several times until the feature maps obtained from the last convolutional layer are transformed into a one-dimensional vector. This vector is the input to the second CNN block, made of fully connected layers. The main structure of the first block of all the tested architectures was kept. In our work, the model training was based on transfer learning. This technique serves as efficient weight initialization and has been shown to provide excellent

Table 1

Training and validation data sets. Total number of red blood cell (RBC) images and smears used in the model development. The data set for model development was split into training and validation data sets. **nRBC**, RBC without inclusions; **MAL**, malaria parasites; **HJ**, Howell–Jolly bodies; **PP**, Pappenheimer bodies; **BS**, basophilic stippling; **PLT**, platelets. **TRAINING^a**, original number of images of the training data set; **TRAINING^b**, number of images of the training data set after data augmentation.

	Smears	Images					
		nRBC	MAL	HJ	PP	BS	PLT
TRAINING^a	53	2000	1083	626	648	550	350
TRAINING^b		2000	2000	2000	2000	2000	2000
VALIDATION		348	286	138	145	126	93

classification results. This procedure is especially useful when the target data set is relatively small, and it has been shown to provide better performance compared to starting a randomly initialized model [40]. A model that has already been trained with a large amount of data (typically ImageNet) will be able to handle a new but similar task with much less data. To train the model, we first needed to adapt the fully connected layers and their weights to our data set/classes. The first fully connected layers were defined with 512 nodes in the case of AlexNet; 1024 nodes for VGG, ResNet18, ResNet34, and SqueezeNet; 2048 nodes for DenseNet; 1456 for Xception; 3072 for Inception; and 4096 nodes for ResNet50 and ResNet101. Each fully connected node processes a weighted sum of its input characteristics. To aid generalizability of the model, these nodes were randomly dropped at each iteration with a probability of 0.5. After this fully connected layer, the ReLU activation function was used to introduce non-linearity into the model. The second fully connected layer was introduced with 512 nodes, whereas the third fully connected layer was configured with 6 nodes, one node for each class of cell under study for all CNNs evaluated. This last layer predicts the class of the input images as probability values by applying a *softmax* operation. Fully connected layers were trained with a maximum learning rate of 0.01 and four epochs. The model weights were adjusted using the cross entropy as the loss function and ADAM (Adaptive Moment Estimation) as the optimizer [41]. The training was carried out using a batch size of 64 over 750 iterations, implementing the cycling learning rate policy [42]. The top layers have the largest impact on performance recognition, so the models were trained in fully connected layers first. Convolutional layers' freezing prevents their weights from changing. Once the last fully connected layers were trained, we unfreeze the entire model and retrained it to adjust the weights of all the layers, so that the entire model was specific to the problem we wanted to solve.

2.6. Model testing

To make an unbiased evaluation of our model, a data set was organized with 23 new smears that were not previously used for training. A thousand images of each smear were used, distributed as detailed in Table 2. The first 11 smears belonged to patients infected with malaria, showing different levels of parasitemia. Smears 12–14 contained normal RBCs. The remaining smears had erythrocytes with some of the inclusions considered in this work: HJ (smears 15, 16, and 21), PP (smears 21–23), BS (smears 17 and 18), and PLT located on the surface of RBC (smears 19 and 20).

The evaluation of the model was performed considering two different scenarios.

Images classification.

Image level tests, the purpose of which was to confirm the ability of the model to correctly classify individual RBCs as nRBC, MAL, HJ, PP, BS, or PLT. Sorting performance was assessed using the following known metrics [43]:

Accuracy, percentage of correctly predicted images out of all images:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

Precision, percentage of images correctly recognized as malaria out of all images recognized as malaria:

$$Precision = TP / (TP + FP)$$

Sensitivity, proportion of actual malaria images correctly identified:

$$Sensitivity (Recall) = TP / (TP + FN)$$

Specificity, proportion of actual normal RBC images correctly identified:

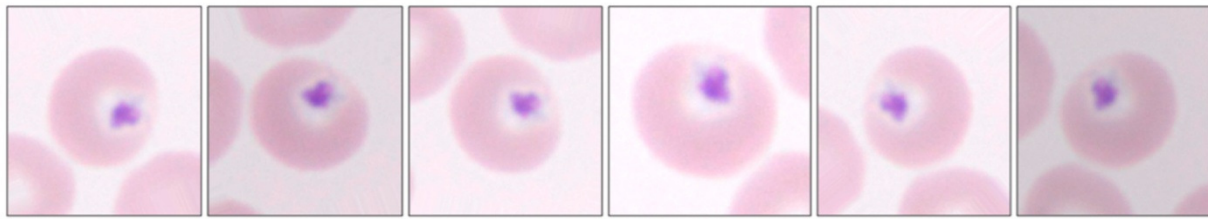


Fig. 2. Example of data augmentation over a PLT image. In this process, the original image has been transformed into new images applying different effects of rotation, zooming, and luminance variation. Although the new images are similar, it can be observed that the platelet localization and the surrounding RBCs are always different in the obtained images. There is also zooming and luminance variation in some of the new images.

Table 2

Testing data set. The total number of red blood cell (RBC) images of each smear used in the model assessment. Parasitemia indicates the percentage of RBCs that contain malaria parasites in each infected smear, and it reflects the severity of the disease. **nRBC**, RBC without inclusions; **MAL**, malaria parasites; **HJ**, Howell-Jolly bodies; **PP**, Pappenheimer bodies; **BS**, basophilic stippling; **PLT**, platelets.

TEST	Parasitemia	Images					
		nRBC	MAL	HJ	PP	BS	PLT
Smear 1	3.2%	968	32	–	–	–	–
Smear 2	14%	858	140	–	–	–	2
Smear 3	4.7%	947	47	–	–	–	6
Smear 4	2.4%	976	24	–	–	–	–
Smear 5	1.9%	977	19	–	–	–	4
Smear 6	2.1%	979	21	–	–	–	–
Smear 7	7.6%	921	76	–	–	–	3
Smear 8	1.1%	989	11	–	–	–	–
Smear 9	0.5%	995	5	–	–	–	–
Smear 10	0.2%	998	2	–	–	–	–
Smear 11	0.1%	999	1	–	–	–	–
Smear 12	–	1000	–	–	–	–	–
Smear 13	–	1000	–	–	–	–	–
Smear 14	–	1000	–	–	–	–	–
Smear 15	–	952	–	48	–	–	–
Smear 16	–	809	–	191	–	–	–
Smear 17	–	916	–	–	–	84	–
Smear 18	–	831	–	–	–	169	–
Smear 19	–	950	–	–	–	–	50
Smear 20	–	917	–	–	–	–	83
Smear 21	–	932	–	15	53	–	–
Smear 22	–	878	–	–	122	–	–
Smear 23	–	920	–	–	80	–	–
TOTAL	–	21,712	378	254	255	253	148

$$\text{Specificity} = \text{TN} / (\text{TN} + \text{FP})$$

F1 score: It is a weighted average between precision and recall that takes into account both false positives and false negatives. It is a useful metric, especially in an uneven class distribution.

$$\text{F1 Score} = 2 * (\text{Sensitivity} * \text{Precision}) / (\text{Sensitivity} + \text{Precision})$$

TP: True positives; **TN:** True negatives; **FP:** False positives; and **FN:** False negatives.

Smear diagnosis.

Smear diagnosis at the patient level, the objective of which was to evaluate the model developed in a malaria screening framework. In this case, the input is a set of images from a smear, and the output is a binary classification whether the smear is infected or not. Since many cases of malaria are associated with low levels of parasitemia and the microscopic visualization of a single parasite is pathognomonic for malaria infection, an individual smear was considered infected if our model recognized at least one RBC with the presence of malaria. Parasitemia is a clinically important concept. It can be defined as the percentage of erythrocytes that are actually infected. Parasitemia reflects the severity

of the infection, but it will also determine how difficult it is for the laboratory to diagnose malaria by reviewing a PBS. Similarly, parasitemia will determine the performance of automated systems in recognizing malaria, as a smear will be more likely to be recognized as infected in the presence of numerous parasites compared to when there is only one parasite.

3. Results

3.1. CNN model design and training

Fig. 3 illustrates the results when looking for the optimal values of the main hyperparameters: learning rate (LR) and the number of training epochs. We found the optimal LR through simulated training that was performed through a wide range of LRs and calculation of the corresponding values of the loss function. **Fig. 3-A** presents the loss values against the LRs. We chose a value of the LR a little before the minimum, where the loss even improved without sacrificing the speed of learning. An epoch is a single step through the entire training set, and generally, training a model can take a large number of epochs to obtain an acceptable level of accuracy. However, the one cycle policy for LR optimizes this training procedure. As illustrated in **Fig. 3-B**, the loss function stopped decreasing only with three to four epochs. The fully connected layers were trained with a maximum LR of 0.01 and four epochs. After that, we proceeded to unfreeze the entire model and perform a new training to adjust the weights of the entire neural network.

The selection of the best architecture of the deep learning model was carried out by classifying the validation set, the results of which are given in **Table 3**. Of all the CNNs tested in the validation phase, VGG-16 was selected based on the results of their classification. The selected model was chosen based on the following factors:

- **Overall Accuracy:** The percentage of images predicted correctly provides an overview of the performance of all architectures. The different models showed high global accuracies, with values higher than 99%, even for the simplest deep learning architectures like AlexNet. Among all of them, VGG-16 was the model that gave the best result in the classification with 99.75% accuracy. VGG-19 and ResNet-101 had second best results with 99.6% overall accuracy.
- **Positive predictive value:** it is crucial that uninfected erythrocytes are not recognized as infected. Due to the superiority of uninfected RBCs, even a small proportion of false positives can represent a significant proportion with respect to malaria parasites. VGG-16 also had the best positive predictive value (97%) among all CNNs tested.
- **Sensitivity:** It is clinically imperative to detect all analyzed parasites. Many cases of malaria are associated with low levels of parasitemia. Therefore, the models should be able to cope with the recognition of a single malaria image from a complete sample. With the exception of Inception and ResNet-101, all models had sensitivity values close to 100%.

VGG-16 presented the best results in all the factors considered.

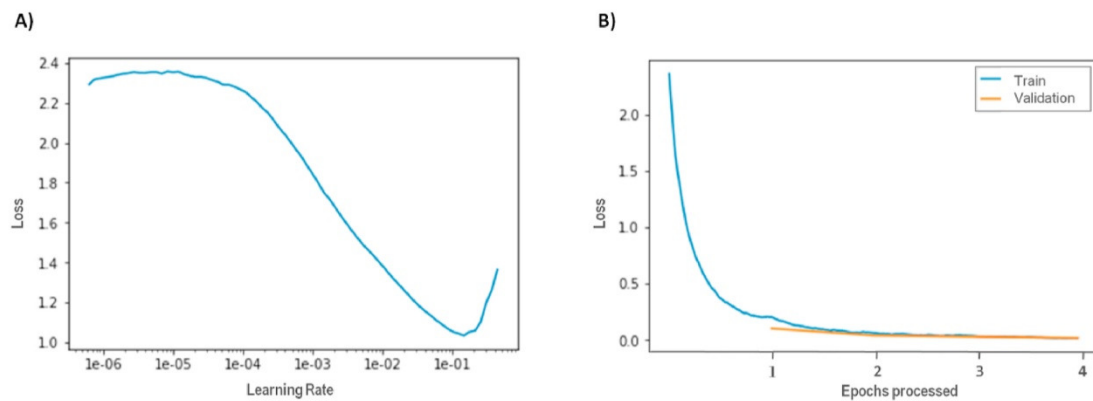


Fig. 3. A, Representation of losses against the learning rate obtained. B, Representation of losses against the number of training epochs.

Table 3

Classification results of the validation set. Sensitivities (in percentage) obtained for each RBC class with every CNN architecture. **nRBC**, RBC without inclusions; **MAL**, malaria parasites; **HJ**, Howell–Jolly bodies; **PP**, Pappenheimer bodies; **BS**, basophilic stippling; **PLT**, platelets. Results in bold are those obtained using the VGG-16 architecture.

	Sensitivity (%)						Overall Accuracy (%)
	nRBC	MAL	HJ	PB	PLT	PP	
AlexNet	99.4	100	99.3	100	100	96.6	99.43
DenseNet-121	99.4	100	100	100	98.9	96.6	99.41
Inception-v4	99.4	99.7	99.3	100	98.9	97.2	99.35
ResNet-18	99.4	100	99.3	100	100	97.9	99.48
ResNet-34	99.4	100	100	100	100	95.9	99.44
ResNet-50	99.4	100	100	100	98.9	96.6	99.41
ResNet-101	99.7	99.7	100	100	98.6	97.9	99.60
SqueezeNet	99.1	100	100	100	98.9	98.6	99.30
VGG-16	99.7	100	100	100	100	98.6	99.75
VGG-19	99.7	100	100	100	97.8	97.9	99.59
Xception	99.1	100	100	100	98.9	97.2	99.24

Therefore, VGG-16 was selected as the final classification model, ready for the testing stage.

VGG-16 is a sequential CNN with a relatively simple architecture. It has five convolutional blocks. Blocks 1 and 2 have two convolutional layers, whereas blocks 3, 4, and 5 have an additional one. Convolutional layers use 3×3 core filters to generate feature maps. At the end of each block, a max pooling layer is used to reduce the image size in width and

height. The input for the first fully connected layer is a flattened representation of the output feature map from block 5. Then, two fully connected layers are followed by a *softmax* classifier, which provides the probability for each class under study.

3.2. Model assessment

Once the model was selected, it was evaluated using the test set (see Table 2) for classification of individual images and complete smears for diagnostic purposes.

3.2.1. Classification of individual RBC images

The classification performance of individual RBC images is summarized in Table 4. Diagonal elements represent true-positive rates (sensitivity) for each individual cell type: 99.6% for normal RBCs, 99.2% for malaria-infected RBCs, 98.4% for HJ, 97.6% for PP, 98.0% for BS, and 99.3% for platelets distributed on RBCs. The overall accuracy was 99.5%.

In addition to the sensitivity, it is particularly relevant in this work to obtain the specificity for the case of malaria. As the model performs a multiclass classification, the confusion matrix in Table 4 was used to calculate the specificity for the recognition of RBCs containing malaria versus all other classes of uninfected RBCs. This gave 99.9% specificity for detecting malaria.

3.2.2. Diagnostic classification of patient smears

The assessment of the model's ability to detect malaria-infected smears corresponding to individual patients is summarized in the

Table 4

Classification of individual images of the testing set. Results are presented as percentages (absolute values in brackets). Rows indicate the true class, and columns represent the predicted class provided by the system. Diagonal shows the sensitivity values. **nRBC**, RBC without inclusions; **MAL**, malaria parasites; **HJ**, Howell–Jolly bodies; **PP**, Pappenheimer bodies; **BS**, basophilic stippling; **PLT**, platelets.

True class	Predicted class					
	nRBC	MAL	HJ	PP	BS	PLT
nRBC	99.6 (21,604)	0.1 (16)	0.1 (13)	0.2 (33)	0.1 (13)	0.1 (15)
MAL	0 (0)	99.2 (375)	0 (0)	0.5 (2)	0 (0)	0.3 (1)
HJ	0 (0)	0 (0)	98.4 (250)	1.6 (4)	0 (0)	0 (0)
PP	0 (0)	0 (0)	0 (1)	97.6 (249)	0.2 (5)	0 (0)
BS	0 (0)	0 (0)	0 (0)	2.0 (5)	98.0 (248)	0 (0)
PLT	0 (0)	0.7 (1)	0 (0)	0 (0)	0 (0)	99.3 (147)

binary confusion matrix in Table 5. We observed that the 11 smears (see details in Table 2) from malaria-infected patients were recognized correctly. Furthermore, only one of the 12 smears from patients without malaria (see details in Table 2) was incorrectly recognized as infected. This means that sensitivity and specificity values of 100% and 91.7% were obtained, respectively. These results indicate that our model has a high performance not only in the recognition of malaria infection (considering that only an erythrocyte recognized as infected will automatically predict the diagnosis of smears as infected) but also in the recognition of smears belonging to patients without malaria infection. It is important to note that some of these smears contained exclusively normal RBCs, whereas other smears contained normal RBCs with inclusions.

4. Discussion

RBC inclusions are morphologically similar structures to malaria parasites. When automatic systems are trained exclusively for the recognition of malaria parasites against normal RBCs, other inclusions are prone to being recognized as parasites. These misidentifications will lead to more false-positive cases, increasing workload and misusing negative sample review efforts. To the authors' knowledge, this work presents the first deep learning model that differentiates infected RBCs from not only normal RBCs but also from other types of RBC inclusions. Consideration of different types of RBC inclusions guarantees that our model is highly specific for the recognition of malaria, as shown in the confusion matrix (Table 4) corresponding to the classification of the test set. In fact, only 16 out of 21,694 normal RBCs (0.07%) were wrongly recognized as malaria parasites. Furthermore, the proposed model exhibits a high rate of parasite detection, because 375 out of 378 infected RBCs were automatically recognized.

From a technical point of view, all tested CNN architectures showed high performance in identifying malaria and other RBC inclusions. We decided that the VGG-16 architecture was the most appropriate for two reasons: 1) It showed a slight performance superiority, and 2) it was a simple and uniform architecture with a relatively low number of layers. As in our work, other authors have evaluated different CNNs for the recognition of malaria, obtaining the best classification scores with VGG architectures [26,27,44]. VGG-16 contains a large number (138 million) of trainable parameters; this number is larger than that of other more complex architectures, such as ResNet-101, which contains only 44 million trainable parameters. Our results suggest that the high accuracies obtained in the recognition of erythrocyte inclusions could be independent of the complexity of CNN and may be related to the number of trainable parameters. This would explain that even the simplest architecture such as AlexNet also showed good recognition results.

Some of the published papers have used public data sets made from segmented images. Typically, the segmentation procedure is missing in works that have used this type of data set. Therefore, it can be considered as a limitation for the implementation of these models in a real work environment, because it will not be possible to obtain similar images used for model training. Using our own data set can make our results difficult to compare with those of other works. However,

designing a complete pipeline, ranging from image acquisition and segmentation to automatic image classification, ensures the implementation of our model in a real work environment.

Table 6 shows the comparative results of our classification model with respect to those previously published. Our automatic recognition system was the only one that was trained with six different groups of RBC images, showing accuracy, sensitivity, and specificity values greater than 99%. Regular PBS obtained from malaria-infected patients in clinical laboratories are unbalanced, with a high predominance of normal versus parasitized RBCs. Therefore, it is important to use unbalanced data sets to test classification models that reproduce real samples, with precision being the most appropriate performance indicator. Gopakumar et al. [10] used an unbalanced image data set to test a customized CNN model for the detection of malaria parasites, showing specificity values of 98.5%. However, their model had a significant false-positive rate relative to the number of true positives, resulting in low precision values (56%). In our work, even when an unbalanced data set was used to test the model, precision reached a value of 95.67%. Other authors such as Vijayalakshmi, Rajaraman, and Bibin evaluated their models using balanced data sets [26,27,44,45], which would lead to decreased precision values if their models were tested with real patient samples (see Table 6).

Rajaraman et al. demonstrated that the performance of the model evaluation is influenced by the methodology applied in the train/test division [44]. They reported 95% accuracy when the classification model was tested using smears from patients whose data had not been used before in the training set. However, when the test was performed with different images that were not used for training but belonged to the same smear/patient, the overall accuracy increased to 99%. To ensure that the accuracy of the tests is not due to overestimation, good practice for any proposed classification model should be to divide the images into training and test sets of different smears and patients. In our work, different individual patients were selected for both sets to avoid overestimation (53 for training/validation and 23 for testing), and the proposed model was able to achieve a high accuracy value of 99.5%.

Rajaraman et al. also obtained better recognition results when modifying the structure of previously trained CNNs. They found that a customized VGG outperformed other known architectures, achieving 99.3% accuracy in recognizing malaria [27]. Modifying previously trained CNNs is also common practice in deep learning approaches. Vijayalakshmi et al. used the model VGG-19 as a feature extractor to feed an SVM classifier. The literature reveals that the problem of automatically identifying malaria-infected RBCs can be addressed in many different approaches, but our results indicate that previously trained CNNs produce satisfactory performance.

Our research group also addressed this problem using classical machine learning algorithms [23]. We showed that the differentiation between normal RBCs and RBCs that have any type of inclusion can be easily done with LDA, SVM, KNN, Gaussian Naive Bayes, or random forest classifiers. All algorithms exhibited overall accuracies greater than 98% with a small set of characteristics. This means that malaria

Table 5

Assessment of smears' classification. Results are presented as percentages (absolute values in brackets) for true and predicted diagnosis of individual smears tested. Smears are recognized as infected if the model classifies at least one single RBC image as containing parasites.

		Predicted diagnosis	
		INFECTED	NON-INFECTED
True diagnosis	INFECTED	100 (11)	0 (0)
	NON-INFECTED	8.3 (1)	91.7 (11)

Table 6

State-of-the-art metrics in malaria classification. Comparison of performance metrics between our proposed methodology and other state-of-the-art studies in malaria detection using deep learning.

	Accuracy	Precision	F1 Score	Sensitivity	Specificity
Proposed method	99.5	95.7	97.4	99.2	99.9
Bibin [45]	96.4	82.9	89.7	97.6	95.9
Gopakumar [10]	98.5	55.9	70.9	97.1	98.5
Rajaraman [44]	95.1	–	95.2	94.6	95.7
Rajaraman [27]	99.3	99.7	99.3	–	–
Vijayalakshmi [26]	93.1	89.9	91.7	93.4	92.9

parasites and other types of inclusions have very similar morphological characteristics compared to normal RBCs. These results suggested that other studies will predict HJ, PP, BS, or platelets as parasites of malaria, thus classifying the smears as pathological when they are not. However, differentiating malaria parasites from other types of inclusions is a more difficult problem to solve. More than 600 features need to be considered to obtain 96.5% accuracy. In this situation, LDA and SVM performed significantly better compared to other types of algorithms. Both LDA and SVM calculate optimal hyperplanes with respect to their individual targets; this seemed to be a more effective technique for tackling highly challenging problems than KNN, Gaussian Naive Bayes, or random forest classifiers. Comparing those results with the proposed CNN, we found that CNNs have three main advantages over LDA or SVM for automatically identifying malaria parasites with high specificity: 1) numerous “handcrafted” features have high computational and time-consuming costs; 2) it was necessary to design a three-step classifier to focus on each particular problem (the first module recognized normal RBCs with respect to RBCs containing inclusions, the second identified whether the inclusion corresponded to malaria or not, and the third module discriminated among HJ, PP, BS, and PLT); and 3) CNNs outperform other classical algorithms.

Our system was ultimately designed to analyze the entire smear as a single item; this is important when developing new diagnostic support tools for laboratory practice. In the evaluation of the model, all the smears from individual patients that contained malaria parasites in the RBCs were identified as infected. Furthermore, one smear out of 12 healthy controls was incorrectly recognized as infected. Malaria-infected patients can be at high risk of death, so it is extremely important not to misdiagnose them as uninfected patients. Therefore, high sensitivity to specificity is preferable when considering new screening tools for detecting malaria. By contrast, the sensitivity and specificity values obtained with our model are higher than those obtained with the RBC Advanced software in the CellaVision® image analyzer (CellaVision, Lund) [46,47]. Using the aforementioned software, sensitivity and specificity values of 23.5% and 81.1%, respectively, have been reported.

A limitation of our model is that it was developed using images obtained manually with a camera attached to a microscope. For a practical deployment of our model in clinical laboratories, use of a motorized microscope and the automatic acquisition of peripheral blood images in a simpler and more standardized way would be desirable.

The diagnosis of malaria is based on the microscopic visualization of the parasites in thick or thin smears. Both types of film are recommended when looking for malaria parasites in peripheral blood. Yang et al. developed a smartphone application that used a customized CNN to detect RBC parasites in thick films that exhibited 93.46% accuracy [4], which was high considering the amount of artifacts commonly seen in this type of smear. The reason we focused our classification model on thin smears was that they enable analysis of the morphological details of RBCs that are important for diagnosis, such as size and shape, the presence of pigments and spots, and the morphology of the parasite [5]. The thin smears allowed us to address the differentiation of malaria parasites from other inclusions through a deep learning model.

5. Conclusion

This work proposes a new deep learning system capable of recognizing malaria-infected RBCs from normal erythrocytes and from erythrocytes with other types of inclusions. By testing the classifier with a separate image data set, excellent performance measures were obtained in single-cell recognition. An evaluation using individual smears demonstrated that our system is efficient and practical for automatic identification of patients infected with malaria. Compared to other published studies, our approach not only offers a final model with remarkable results, but it is also based on a suitable design capable of discriminating malaria parasites from other RBC inclusions. This approach has not been considered before and ensures its actual

implementation for the first time. It is a promising advance that can help reduce response times and help the clinical pathologist perform efficient, objective, and rapid morphological analysis of samples.

Our future work is to standardize the acquisition of peripheral blood cell images and move toward implementing the system on portable devices and conducting a clinical proof of concept.

Declaration of competing interest

All authors declare there is no any financial or personal relationship with other people or organizations that could inappropriately influence the presented work.

Acknowledgment

This work is part of a research project funded by the Ministry of Science and Innovation of Spain, with reference PID2019-104087RB-I00.

References

- [1] World Health Organization, *World Malaria Report, 2019*. ISBN: 978-92-4-156572-1.
- [2] C.K. Murray, R.A. Gasser, A.J. Magill, R.S. Miller, Update on rapid diagnostic testing for malaria, *Clin. Microbiol. Rev.* 21 (2008) 97–110, <https://doi.org/10.1128/CMR.00035-07>.
- [3] R. Shouval, J.A. Fein, B. Savani, M. Mohty, A. Nagler, Machine learning and artificial intelligence in haematology, *Br. J. Haematol.* (2020), <https://doi.org/10.1111/bjh.16915>.
- [4] F. Yang, M. Poostchi, H. Yu, Z. Zhou, K. Silamut, J. Yu, S. Antani, Deep learning for smartphone-based malaria parasite detection in thick blood smears, *IEEE J Biomed Health Inform* 24 (2019) 1427–1438, <https://doi.org/10.1109/JBHI.2019.2939121>.
- [5] C. Wongsrichanalai, M.J. Barcus, S. Muth, A. Sutamihardja, W.H. Wernsdorfer, A review of malaria diagnostic tools: microscopy and rapid diagnostic test (RDT), *Am. J. Trop. Med. Hyg.* 77 (2007) 119–127, <https://doi.org/10.4269/ajtmh.2007.77.119>.
- [6] N.E. Ross, C.J. Pritchard, D.M. Rubin, et al., Automated image processing method for the diagnosis and classification of malaria on thin blood smears, *Med. Biol. Eng. Comput.* 44 (2006) 427–436, <https://doi.org/10.1007/s11517-006-0044-2>.
- [7] S. Moon, S. Lee, H. Kim, et al., An image analysis algorithm for malaria parasite stage classification and viability quantification, *PLoS One* 8 (2013), e61812, <https://doi.org/10.1371/journal.pone.0061812>.
- [8] K. Prasad, J. Winter, U.M. Bhat, et al., Image analysis approach for development of a decision support system for detection of malaria parasites in thin blood smear images, *J. Digit. Imag.* 25 (2012) 542–549, <https://doi.org/10.1007/s10278-011-9442-6>.
- [9] S.S. Devi, R.H. Laskar, S.A. Sheikh, Hybrid classifier based life cycle stages analysis for malaria infected erythrocyte using thin blood smear images, *Neural Comput. Appl.* 29 (2018) 217–235, <https://doi.org/10.1007/s00521-017-2937-4>.
- [10] G.P. Gopakumar, M. Swetha, G. SaiSiva, et al., Convolutional neural network-based malaria diagnosis from focus stack of blood smear images acquired using custom built slide scanner, *J. Biophot.* 11 (2018), e201700003, <https://doi.org/10.1002/jbio.201700003>.
- [11] D.K. Das, M. Ghosh, M. Pal, et al., Machine learning approach for automated screening of malaria parasite using light microscopic images, *Micron* 45 (2013) 97–106, <https://doi.org/10.1016/j.micron.2012.11.002>.
- [12] M.T. Le, T.R. Bretschneider, C. Kuss, et al., A novel semiautomatic image processing approach to determine Plasmodium falciparum parasitemia in Giemsa stained thin blood smears, *BMC Cell Biol.* 9 (2008) 15, <https://doi.org/10.1186/1471-2121-9-15>.
- [13] Y. Purwar, S.L. Shah, G. Clarke, et al., Automated and unsupervised detection of malarial parasites in microscopic images, *Malar. J.* 10 (2011) 364, <https://doi.org/10.1186/1475-2875-10-364>.
- [14] J. Somasekar, B. Eswara Reddy, B.E. Reddy, Segmentation of erythrocytes infected with malaria parasites for the diagnosis using microscopy imaging, *Comput. Electr. Eng.* 45 (2015) 336–351, <https://doi.org/10.1016/j.compeleceng.2015.04.009>.
- [15] M. Delgado-Ortíz, A. Molina, S. Alf  rez, J. Rod  llar, A. Merino, A deep learning approach for segmentation of red blood cell images and malaria detection, *Entropy* 22 (2020) 657, <https://doi.org/10.3390/e22060657>.
- [16] M. Poostchi, K. Silamut, R.J. Maude, et al., Image analysis and machine learning for detecting malaria, *Transl. Res.* 194 (2018) 36–55, <https://doi.org/10.1016/j.trsl.2017.12.004>.
- [17] F.B. Tek, A.G. Dempster, İzzet Kale, Parasite detection and identification for automated thin blood film malaria diagnosis, *Comput. Vis. Image Understand.* 114 (2010) 21–32, <https://doi.org/10.1016/j.cviu.2009.08.003>.
- [18] D. Yang, G. Subramanian, J. Duan, et al., A portable image-based cytometer for rapid malaria detection and quantification, *PLoS One* 12 (2017), e0179161, <https://doi.org/10.1371/journal.pone.0179161>.

- [19] N. Linder, R. Turkki, M. Walliander, et al., A malaria diagnostic tool based on computer vision screening and visualization of *Plasmodium falciparum* candidate areas in digitized blood smears, *PloS One* 9 (2014), e104855, <https://doi.org/10.1371/journal.pone.0104855>.
- [20] G. Díaz, F.A. González, E. Romero, A semi-automatic method for quantification and classification of erythrocytes infected with malaria parasites in microscopic images, *J. Biomed. Inf.* 42 (2009) 296–307, <https://doi.org/10.1016/j.jbi.2008.11.005>.
- [21] G. Díaz, F.A. González, E. Romero, Infected cell identification in thin blood images based on color pixel classification: comparison and analysis, *Progress in Pattern Recognition, Image Analysis and Applications* (2007), https://doi.org/10.1007/978-3-540-76725-1_84, 812–21.
- [22] M. Sheikhsosseini, H. Rabbani, M. Zekri, et al., Automatic diagnosis of malaria based on complete circle ellipse fitting search algorithm, *J. Microsc.* 252 (2013) 189–203, <https://doi.org/10.1111/jmi.12081>.
- [23] A. Molina, S. Alf  rez, L. Bold  , A. Acevedo, J. Rodellar, A. Merino, Sequential classification system for recognition of malaria infection using peripheral blood cell images, *J. Clin. Pathol.* (2020), <https://doi.org/10.1136/jclinpath-2019-206419>, <https://doi.org/10.1136/jclinpath-2019-206419>.
- [24] W. Rawat, Z. Wang, Deep convolutional neural networks for image classification: a comprehensive review, *Neural Comput.* 29 (2017) 2352–2449, https://doi.org/10.1162/neco_a_00990.
- [25] W.D. Pan, Y. Dong, D. Wu, Classification of malaria infected cells using deep convolutional neural networks, in: *Machine Learning: Advanced Techniques and Emerging Applications*, vol. 159, Intech Open, 2018, <https://doi.org/10.5772/intechopen.72426>.
- [26] A. Vijayalakshmi, Deep learning approach to detect malaria from microscopic images, *Multimed. Tool. Appl.* 79 (2020) 15297–15317, <https://doi.org/10.1007/s11042-019-7162-y>.
- [27] S. Rajaraman, S. Jaeger, S.K. Antani, Performance evaluation of deep neural ensembles toward malaria parasite detection in thin-blood smear images, *PeerJ* 7 (2019) e6977, <https://doi.org/10.7717/peerj.6977>.
- [28] A. Merino, *Manual de Citolog  a de Sangre Perif  rica y L  quidos Biol  gicos*, M  dica Panamericana, 2020. ISBN: 978-84-9-110262-5.
- [29] L. Palmer, C. Briggs, S. McFadden, et al., ICSH recommendations for the standardization of nomenclature and grading of peripheral blood cell morphological features, *Int J Lab Hematol* 37 (2015) 287–303.
- [30] S. Van der Walt, J.L. Sch  nberger, J. Nunez-Iglesias, F. Boulogne, J.D. Warner, Yager, et al., scikit-image: image processing in Python, *PeerJ* 2 (2014) e453, <https://doi.org/10.7717/peerj.453>.
- [31] J. Howard, Gugger S. Fastai, A layered API for deep learning, *Information* 11 (2014) 108, <https://doi.org/10.3390/info11020108>.
- [32] J.-L. Vives Corrons, S. Albar  de, G. Flandrin, et al., Guidelines for blood smear preparation and staining procedure for setting up an external quality assessment scheme for blood smear interpretation. Part I: control material, *Clin. Chem. Lab. Med.* 42 (2004) 922–926, <https://doi.org/10.1515/CCLM.2004.149>.
- [33] R. Tomari, W.N.W. Zakaria, M.M.A. Jamil, F.M. Nor, N.F.N. Fuad, Computer aided system for red blood cell classification in blood smear image, *Procedia Comput Sci* 42 (2014) 206–213, <https://doi.org/10.1016/j.procs.2014.11.053>.
- [34] A. Acevedo, S. Alf  rez, A. Merino, L. Puigv  , J. Rodellar, Recognition of peripheral blood cell images using convolutional neural networks, *Comput. Methods Progr. Biomed.* 180 (2019) 105020, <https://doi.org/10.1016/j.cmpb.2019.105020>.
- [35] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105, <https://doi.org/10.1145/3065386>.
- [36] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, *arXiv preprint* (2014) 1409–1556.
- [37] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
- [38] F. Chollet, Xception: deep learning with depthwise separable convolutions, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1251–1258.
- [39] C. Szegedy, S. Ioffe, V. Vanhoucke, A.A. Alemi, Inception-v4, inception-resnet and the impact of residual connections on learning, in: *Thirty-first AAAI Conference on Artificial Intelligence*, 2017.
- [40] Huh M, Agrawal P, Efros AA. What makes ImageNet good for transfer learning? *arXiv preprint arXiv* 2016:1608.08614.
- [41] D.P. Kingma, J.L. Ba, Adam: a method for stochastic optimization, in: *3rd Int Conf Learn Represent ICLR 2015 - Conf Track Proc*, 2015, pp. 1–15.
- [42] L.N. Smith, Cyclical learning rates for training neural networks. En 2017 IEEE winter conference on applications of computer vision (WACV), IEEE (2017) 464–472, <https://doi.org/10.1109/WACV.2017.58>.
- [43] M. Kubat, *An Introduction to Machine Learning*, Springer International Publishing AG, 2017. ISBN: 978-3-319-63913-0.
- [44] S. Rajaraman, K. Silamut, M.A. Hossain, I. Ersoy, R.J. Maude, S. Jaeger, S. K. Antani, Understanding the learned behavior of customized convolutional neural networks toward malaria parasite detection in thin blood smear images, *J. Med. Imaging* 5 (2018), 034501, <https://doi.org/10.1117/1.JMI.5.3.034501>.
- [45] D. Bibin, M.S. Nair, P. Punitha, Malaria parasite detection from peripheral blood smear images using deep belief networks, *IEEE Access* 5 (2017) 9099–9108, <https://doi.org/10.1109/ACCESS.2017.2705642>.
- [46] L. Florin, K. Maelegheer, A. Muyldermans, et al., Evaluation of the CellaVision DM96 advanced RBC application for screening and follow-up of malaria infection, *Diagn. Microbiol. Infect. Dis.* 90 (2018) 253–256, <https://doi.org/10.1016/j.diagmicrobio.2017.12.002>.
- [47] L.D. Racsa, R.M. Gander, P.M. Southern, et al., Detection of intracellular parasites by use of the CellaVision DM96 analyzer during routine screening of peripheral blood smears, *J. Clin. Microbiol.* 53 (2015) 167–171, <https://doi.org/10.1128/JCM.01783-14>.