

MATHEMA 3(1) 71-81
REVISTA DE INVESTIGACIÓN
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
UNIVERSIDAD NACIONAL PEDRO RUIZ GALLO
LAMBAYEQUE-PERÚ

Algoritmo para detección de Trichuris trichiura en imágenes microscópicas de muestras coprológicas - Hospital Regional de Lambayeque - 2019

Algorithm for detection of Trichuris trichiura in microscopic images of coprological samples - Regional Hospital of Lambayeque - 2019

Ing. Eduar Anibal Vásquez Ortiz ¹
Ing. Oscar Uccelly Romero Cortéz ²
Dr. Blgo. Silva Díaz Heber³

RESUMEN

El objetivo de esta investigación fue elaborar un algoritmo de visión por computadora para detección de huevos de Trichuris trichiura en imágenes microscópicas de muestras coprológicas con alta sensibilidad y precisión; para lograr esto se usaron 1000 imágenes, 30 % para probar el funcionamiento del algoritmo y 70 % para su aprendizaje, de 65 x 65 píxeles extraídas de 30 imágenes microscópicas de 1280 x 960 píxeles que fueron recolectadas de una sola muestra coprológica positiva procesada en solución salina; se elaboraron programas en Python utilizando librerías OpenCV, Scikit-learn, imutils, argparse, os, cPickle, Numpy y Matplotlib para obtener las subimágenes, graficar histogramas, probar y guardar el clasificador con diferentes vectores de características. El vector con mejor rendimiento fue el histograma en espacio de color HSV con 3 intervalos de matiz, 4 de saturación y 4 de brillo usando el algoritmo del vecino más cercano con métrica Manhattan y un vecino para la clasificación, llegando la sensibilidad del algoritmo al 99,35% y la precisión al 96.1%.

Palabras clave: Algoritmo, Trichuris trichiura, sensibilidad y precisión

ABSTRACT

The objective of this research was to develop a computer vision algorithm for the detection of Trichuris trichiura eggs in microscopic images of coprological samples with high recall and precision; to achieve this, 1000 images were used, 30% to test the algorithm and 70% for its learning, of 65 x 65 pixels extracted from 30 microscopic images of 1280 x 960 pixels that were collected from a single positive coprological sample processed in saline solution; Python programs were developed using OpenCV, Scikit-learn, imutils, argparse, os, cPickle, Numpy and Matplotlib libraries to obtain sub-images, graph histograms, test and save the classifier with different feature vectors. The vector with the best performance was the HSV color space histogram with 3 hue intervals, 4 saturation and 4 brightness using the algorithm of the nearest neighbor with Manhattan metric and one neighbor for classification, reaching a recall of 99,35% and a precision of 96.1%.

Keywords: Algorithm, Trichuris trichiura, recall and precision

¹ Universidad Nacional Pedro Ruiz Gallo, Escuela Profesional de Ingeniería Electrónica. e28a12@gmail.com

² Universidad Nacional Pedro Ruiz Gallo, Departamento Académico de Ingeniería Electrónica. oromero@unprg.edu.pe

³ Hospital Regional de Lambayeque. h.silvadiaz185@gmail.com

INTRODUCCIÓN

Los médicos basan el diagnóstico de ciertas enfermedades en análisis que permiten la detección de agentes extraños en el organismo. Actualmente, el proceso de análisis de cualquier tipo de muestra a través de un microscopio en los hospitales se realiza mediante inspección humana, luego de haber preparado la muestra en laminillas de 22 x 22 mm (Fabián de Estrada et al, 2003).

Uno de los casos en el que se utiliza el análisis de muestras a través del microscopio es el de la detección de estructuras helmínticas parasitarias el cual se realiza mayormente con objetivo de 10x o 40x no siendo recomendable el de 100x (Fabián de Estrada et al, 2003). Una de dichas estructuras son los huevos de *Trichuris trichiura*, estos tienen un tamaño de 50 μm , forma alargada, con un tapón redondo y transparente en cada polo y una masa granulosa uniforme en el centro, su envoltura es anaranjada y su contenido amarillo (Zurita, 2013). De acuerdo a (Centers for Disease Control and Prevention, 2017) el *Trichuris trichiura* es el tercer gusano redondo más común en los humanos y se estima que 800 millones de personas están infectadas en todo el mundo; las infecciones graves causan problemas gastrointestinales y retraso del crecimiento en los niños.

La detección depende mucho del laboratorista, es por ello que se necesita un método rápido y cuyo resultado no dependa de su pericia y garantice confianza en los resultados.

La hipótesis fue: Si se elabora un algoritmo de visión por computadora, entonces se podrá detectar huevos de *Trichuris trichiura* en imágenes microscópicas de muestras coprológicas con alto grado de sensibilidad y precisión.

En un estudio en Sudán (Bashir et al, 2017) se detectó parásitos *Plasmodium* y eritrocitos utilizando características de intensidad y un clasificador de red neuronal artificial entrenada con el algoritmo de propagación hacia atrás, la arquitectura del sistema consistió en seis procesos principales: adquisición, pre-procesamiento, segmentación (extrayendo 1120 sub-imágenes de eritrocitos que se usaron en el entrenamiento y prueba del sistema), extracción de las características, comparación y clasificación; obteniéndose un 99,68% de precisión en la evaluación de la performance .

En otra investigación realizada en Malasia (Ghazali et al, 2013) hicieron uso de un sistema de filtración con determinaciones de umbrales para la clasificación y detección de huevos de *Ascaris lumbricoides* y *Trichuris trichiura*, obteniendo una tasa de éxito de casi 93% y 94 % respectivamente en simulaciones, realizaron además combinaciones diferentes de algoritmos de preprocesamiento de la imagen para así obtener el resultado más óptimo para la extracción de características concluyendo que el mejor detector de bordes es el método Canny y que usar un filtro de mediana dos veces da un mejor rendimiento.

Según Fadul, A. O. (2004), un algoritmo es una serie de pasos organizados, que describe el proceso que se debe seguir, para dar solución a un problema específico.

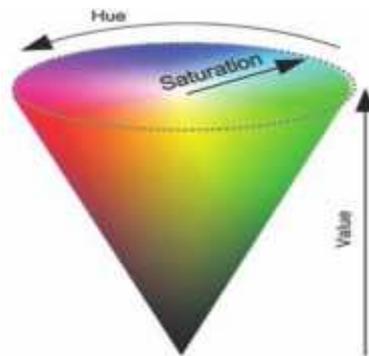
Espacio de color perceptual o HSV

En (Solomon & Breckon, 2011) se define el espacio de color perceptual como una forma alternativa de representar imágenes de color de una manera que sea más natural para la percepción humana y la comprensión del color que la representación RGB.

Sus parámetros son:

- . H (matiz / hue): Longitud de onda dominante del color.
- . S (saturación / saturation): Cantidad de luz blanca mezclada con el color.
- . V (valor / value): Brillo del color. (ver Fig. 1)

Figura 1. Espacio de color HSV (Solomon & Breckon, 2011)



Distribuciones de píxeles: histogramas

De acuerdo a (Solomon & Breckon, 2011) un histograma de imagen es un gráfico de la frecuencia relativa de ocurrencia de cada uno de los permitidos valores de píxeles en la imagen contra los valores mismos.

Vector de características

Los objetos pueden ser comparados por la similitud basada en sus representaciones como un vector de medidas. Suponga que cada objeto es representado por exactamente d medidas. La coordenada i -ésima de tal vector de características tiene el mismo significado para cada objeto. (Shapiro & Stockman, 2001)

El algoritmo del vecino más cercano

El principio del algoritmo del vecino más cercano es el de comparar patrones de imágenes de entrada con un número de paradigmas y luego clasificarlos de acuerdo a la clase del paradigma que dé la coincidencia más cercana. (Davies, 2012)

El concepto de “distancia” en un espacio de características es flexible y admite otras definiciones, en la Tabla 1, se aprecian fórmulas de dos métricas entre dos vectores N-dimensionales x y y . (Solomon & Breckon, 2011)

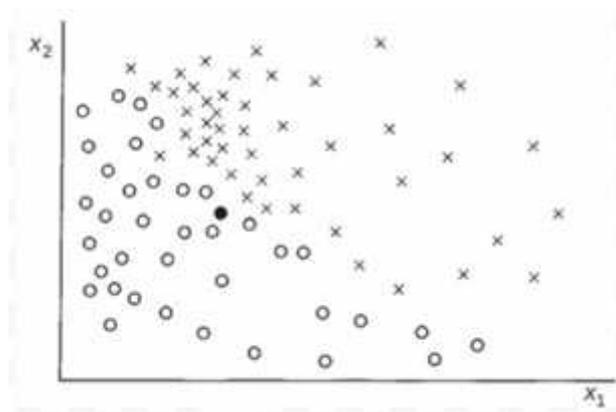
Tabla 1
Métricas para el algoritmo del vecino más cercano

Métrica	Fórmula
Distancia Euclidiana	$d(x, y) = \left[\sum_{i=1}^N (x_i - y_i)^2 \right]^{\frac{1}{2}}$
Distancia Manhattan	$d(x, y) = \sum_{i=1}^N x_i - y_i $

Nota: Tomada de (Solomon & Breckon, 2011)

En la Fig. 2, se muestra un ejemplo del uso del algoritmo para un problema de dos clases en un espacio vectorial de dos dimensiones; si $k=1$, el patrón de prueba será de clase 1, si se aumenta el valor de k la clasificación será más compleja debido a que ambas clases están cerca una de la otra en el espacio de características.

Figura 2. Principio del algoritmo del vecino más cercano para un problema de dos clases. \circ , conjunto de patrones de la clase 1; \times , conjunto de patrones de la clase 2; y \bullet patrón de prueba. (Davies, 2012)



Diseño de sistemas de clasificación

Los pasos a seguir para el diseño de un sistema de clasificación según (Solomon & Breckon, 2011) son: definición de clase, Exploración de datos, selección y extracción de características, construcción del clasificador usando los datos de entrenamiento y prueba del clasificador.

Precisión y sensibilidad

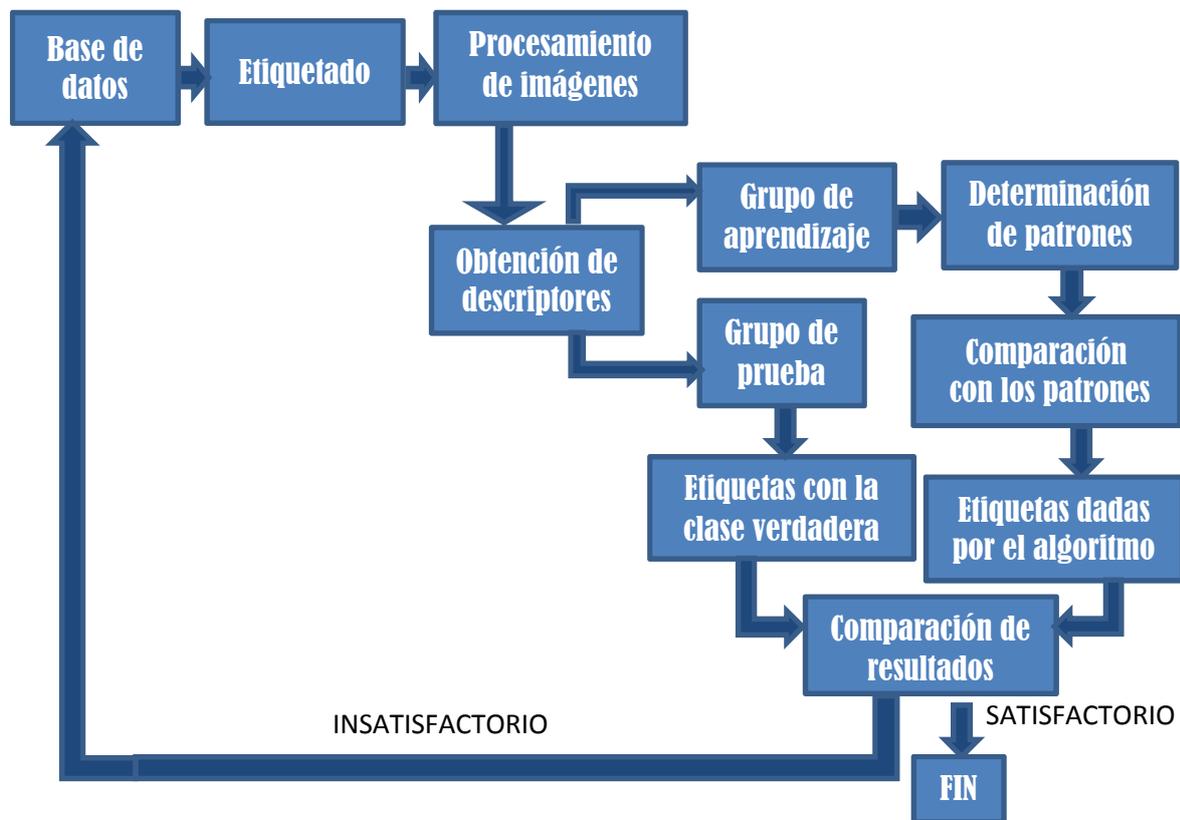
De acuerdo a (Shapiro & Stockman, 2001), la precisión es el número de documentos relevantes recuperados dividido entre el número total de documentos recuperados y la sensibilidad es el número

de documentos relevantes recuperados dividido entre el número total de documentos relevantes en la base de datos.

MÉTODO Y MATERIALES

Diseño de Contrastación de Hipótesis

Figura 3. Esquema con la lógica a seguir para la contrastación de la hipótesis



Usando los descriptores obtenidos del procesamiento de la base de datos, se realizó el reconocimiento de patrones en el grupo de aprendizaje para poder detectar los huevos de Trichuris trichiura. Los resultados obtenidos por el algoritmo al clasificar el grupo de prueba fueron comparados con sus respectivas clases verdaderas, para probar así su sensibilidad y precisión. En la Fig. 3 se presenta la lógica a seguir para la contrastación de la hipótesis.

La población estuvo constituida por las imágenes microscópicas de las muestras coprológicas de los pacientes que se atienden en el Hospital Regional de Lambayeque las cuales fueron procesadas para detectar la presencia de huevos de Trichuris trichiura.

El tamaño de muestra de imágenes fue no probabilístico por conveniencia del investigador, siendo un total de 1000 imágenes de 65 x 65 píxeles (500 positivas y 500 negativas), 30 % para probar el funcionamiento del algoritmo y 70 % para su aprendizaje, obtenidas de la segmentación de 30 imágenes microscópicas de 1280 x 960 píxeles que fueron recolectadas de una sola muestra coprológica positiva procesada en solución salina.

Las técnicas usadas fueron de adquisición y segmentación de imágenes, extracción de histogramas del espacio de color HSV, y el algoritmo del vecino más cercano con las métricas Euclidiana y Manhattan. Se utilizaron programas escritos en lenguaje Python con distintas librerías especializadas (OpenCV, Scikit-learn (Pedregosa et al, 2011), imutils, argparse, cPickle, Numpy y Matplotlib). Para analizar la muestra coprológica se usó un microscopio Biológico Binocular Olympus CX31 con el objetivo de 10x que combinado con el ocular de 10x se logró un aumento de 100 veces.

RESULTADOS Y DISCUSIÓN

Para la obtención de subimágenes, se realizó un programa en Python que permitió extraer regiones de 65 x 65 píxeles de las imágenes microscópicas obtenidas en la Fig. 4 se observa el programa en funcionamiento.

Fig 4 Interfaz del programa para la segmentación



Histogramas en espacio HSV a vector

Con el programa para segmentación se obtuvieron subimágenes como la Fig 5, en donde se puede apreciar un huevo; a esta imagen se le extrajeron sus histogramas (Fig. 6, 7 y 8)

Fig 5 Subimagen con huevo



Fig 6 Histograma de H

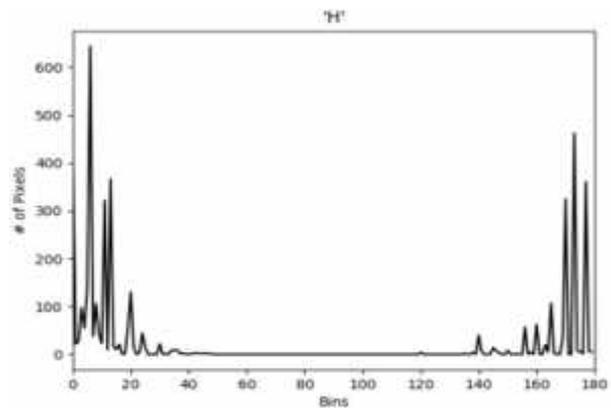


Fig 7 Histograma de S

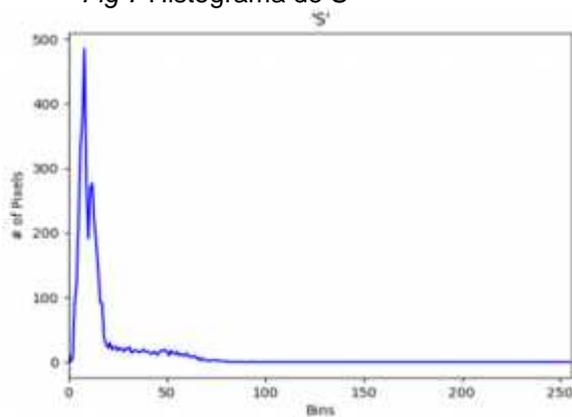
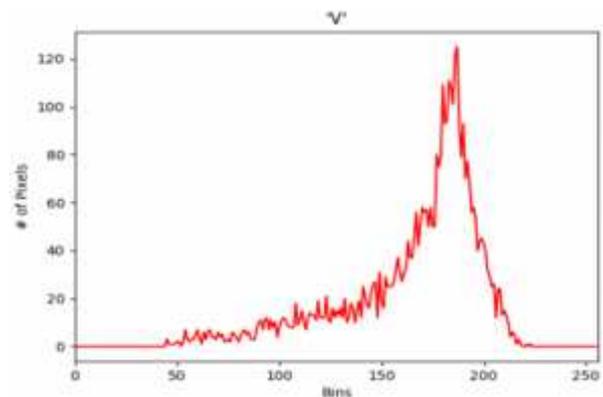


Fig 8 Histograma de V



Los histogramas fueron obtenidos con 180 'bins' (intervalos) para 'H' y 256 para 'S' y 'V'. Se observó que el histograma de H tiene picos justo entre los matices rojo y amarillo, colores propios del huevo, el histograma de S nos indicó una baja saturación y el histograma de V presentó una densidad alta de valores altos de brillo. Luego de realizar una comparación con imágenes sin huevos, se concluyó que sí era un buen descriptor de la imagen. Luego de esto, se realizó la conversión de los histogramas a un vector, sin embargo no se recomienda colocar todos los datos completos ya que el vector resultante tendría 11'796,480 elementos, obtenido de una matriz de 180*256*256; por lo tanto se debe disminuir el número de intervalos. Se decidió usar 8 intervalos para matiz, saturación y brillo, obteniéndose un vector de 512 elementos, que es un valor adecuado para iniciar las pruebas.

Descriptor usando rango de colores

Se elaboró un programa para determinar en qué rangos del espacio de colores HSV se encuentran los huevos, eliminando la mayor cantidad de colores sin perjudicar la forma básica de los mismos, los rangos obtenidos fueron: $0 < H < 17$, $22 < S < 80$ y $85 < V < 170$

De esto, se obtuvieron dos vectores de características: uno de 4,225 elementos que es la nueva imagen completa en blanco y negro, y otro de 512 elementos que es el histograma de la nueva imagen sin los colores fuera de rango.

Sensibilidad y precisión del clasificador

Para la clasificación de las imágenes se utilizó el algoritmo del vecino más cercano con los parámetros (número de vecinos y métrica) elegidos por la función GridSearchCv de la librería Scikit-learn, que busca los más adecuados según el vector de características usado.

En la tabla 1 se aprecia que los valores más altos de sensibilidad y precisión se consiguieron con el histograma de la imagen completa como vector, un vecino y métrica Manhattan. La imagen en blanco y negro como vector, luego de eliminar colores fuera de rango, da como resultado valores bajos de sensibilidad y precisión, esto probablemente se deba a que el color de los huevos es variable por lo que existe el riesgo que colores fuera del rango también se encuentren dentro del huevo, dando una forma que no es buen descriptor de lo que se desea clasificar; sin embargo, el histograma de los colores dentro del rango sí da un alto valor de sensibilidad, pero la baja precisión hace que se descarte; por último, la imagen en blanco y negro como vector concatenada con el histograma de la imagen completa da valores aceptables de sensibilidad y precisión, pero no lo suficientemente buenos para la clasificación que se realiza, pues esto se traduce en posibles errores en el diagnóstico médico.

Tabla 1

Sensibilidad y precisión del clasificador de acuerdo al vector de características

Vector de características	V	M	S (%)	P (%)
Imagen en Blanco y negro como vector, luego de eliminar colores fuera del rango	3	E	57.05	69.67
Histograma de los colores dentro del rango	3	Ma	97.32	68.72
Histograma de la imagen completa	1	Ma	97.99	94.19
Imagen en B/N como vector concatenada con el histograma de la imagen completa	1	Ma	76.51	85.07

Nota: V= vecinos, M=métrica, S=sensibilidad, P=Precisión, E=Euclidiana, Ma=Manhattan

Habiendo elegido el histograma de la imagen completa como vector de características a usar, se procedió a probar con diferentes intervalos de matiz (H), saturación (S) y brillo (V); como se observa en la tabla 2.

Tabla 2
Sensibilidad y precisión del clasificador de acuerdo a los intervalos del histograma

Intervalos	V	M	S (%)	P (%)
(8,8,8)	1	Ma	97.99	94.19
(10,10,10)	1	Ma	97.99	94.81
(12,12,12)	1	Ma	97.99	96.05
(14,14,14)	1	Ma	98.66	93.04
(6,6,6)	1	Ma	98.66	93.63
(4,4,4)	1	Ma	98.66	94.23
(2,2,2)	3	Ma	94.63	85.98
(3,3,3)	1	Ma	93.96	89.74
(3,3,4)	1	Ma	97.99	94.81
(3,4,4)	1	Ma	98.66	95.45

Nota: V= vecinos, M=métrica, S=sensibilidad, P=Precisión, Ma=Manhattan

Se obtuvo el mejor rendimiento en dos combinaciones de intervalos (12,12,12) con 97.99% de sensibilidad y 96.05% de precisión, y (3,4,4) con 98.66% de sensibilidad y 95.45% de precisión. Se le dio prioridad al de mayor sensibilidad, puesto que en el diagnóstico médico conviene un bajo número de falsos negativos, además en este caso conviene tener menos valores para que el procesamiento de los datos sea más rápido.

Por último, se realizaron diferentes pruebas agrupando de diferentes maneras las imágenes, llegándose a obtener un 99.35% de sensibilidad en una prueba y 96.1% de precisión en otra. Además en la tabla 3 se puede observar que la sensibilidad siempre se mantiene mayor de 97%, mientras que la precisión fluctúa entre 92% y 96%.

Tabla 3
Sensibilidad y precisión del clasificador con distintos datos de aprendizaje, con intervalo de (3,4,4), métrica Manhattan y un vecino más cercano.

Prueba	Sensibilidad (%)	Precisión (%)
1	98.66	95.45
2	97.37	96.1
3	99.35	93.87
4	98.72	92.22
5	98.76	93.53
6	97.44	95.00

La cantidad de subimágenes extraídas se decidió basándose en (Bashir et al, 2017), y lo obtenido en esta investigación llega a superar en resultados para *Trichuris trichiura* a la investigación realizada en Malasia (Ghazali et al, 2013).

CONCLUSIONES

1. El algoritmo de visión por computadora elaborado para detección de huevos de *Trichuris trichiura* contiene las siguientes fases: obtención de imágenes - extracción de histogramas en HSV - clasificación con el algoritmo del vecino más cercano.
2. Se procesaron imágenes microscópicas de 65 x 65 píxeles obteniéndose cuatro descriptores:
 -) Imagen en B/N, luego de eliminar colores fuera del rango.
 -) Histograma de la imagen luego de eliminar los colores fuera del rango.
 -) Histograma de la imagen completa.
 -) Imagen en B/N como vector concatenada con el histograma de la imagen completa.
3. En un espacio de características N-dimensional se ubicaron los descriptores del grupo de aprendizaje, obteniéndose patrones de clasificación.
4. La clasificación del grupo de prueba se obtuvo utilizando el algoritmo del vecino más cercano con distintos parámetros.
5. Con el histograma con 3 intervalos de matiz (H), 4 de saturación (S) y 4 de brillo (V) y el algoritmo del vecino más cercano con métrica Manhattan y un solo vecino, se obtuvo 99.35% de sensibilidad y 96.1% de precisión para el algoritmo elaborado

AGRADECIMIENTO

Al Hospital Regional de Lambayeque por permitirme usar sus equipos y acceder a la muestra coprológica necesitada para el desarrollo del proyecto.

REFERENCIAS BIBLIOGRÁFICAS

Bashir, A., Mustafa, Z. A., Abdelhameid, I., & Ibrahem, R. (2017, January). Detection of malaria parasites using digital image processing. In 2017 International Conference on Communication, Control, Computing and Electronics Engineering (ICCCCEE) (pp. 1-5). IEEE.

Centers for Disease Control and Prevention (2017). Trichuriasis. Recuperado de <https://www.cdc.gov/dpdx/trichuriasis/index.html>

Davies, E. R. (2012). Computer and machine vision: theory, algorithms, practicalities (4th ed). Amsterdam ; Boston: Elsevier.

Fabián de Estrada, M., Tello Casanova, R., & Náquira Velarde, C. (2003). Manual de Procedimientos de Laboratorio para el Diagnóstico de los Parásitos Intestinales del Hombre.

Fadul, A. O. (2004). Diseño Estructurado de Algoritmos. Colombia: Sincelejo.

Shapiro, L. G., & Stockman, G. C. (2001). Computer Vision. Upper Saddle River, NJ: Pearson.

Ghazali, K. H., Hadi, R. S., & Mohamed, Z. (2013). Automated system for diagnosis intestinal parasites by computerized image analysis. Modern Applied Science, 7(5), 98.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. Journal of machine learning research, 12(Oct), 2825-2830.

Solomon, C., & Breckon, T. (2011). Fundamentals of Digital Image Processing: A Practical Approach with Examples in Matlab (1st edition). Chichester, West Sussex ; Hoboken, NJ: Wiley.

Zurita Macalupú, S. (2013). Procedimientos de laboratorio: Laboratorios locales I - Laboratorios locales II (2nd ed.)