

---

# Clasificación y detección de células infectadas *Plasmodium Vivax* en imágenes de frotis de sangre

---

Erick W. Díaz <sup>\*1</sup> Jorge A. Balsells <sup>\*1</sup>

## Abstract

En Guatemala, las técnicas actuales utilizadas para el diagnóstico de *P. Vivax* parásito de la malaria no son muy eficientes provocando un retraso en el diagnóstico, estas técnicas son el examen microscópico del protozoo a través de dos tipos de frotis de sangre: frotis de sangre espesa y frotis periférico. La investigación realizada consistió en la creación de un sistema de detección automática en imágenes que contienen cuatro de las *P. Vivax* parásito, utilizando dos arquitecturas CNN diferentes (YOLO y U-Net). Los resultados fueron satisfactorios y la arquitectura convolucional YOLO logró clasificar las fases trofozoíto, esquizonte y gameto de la *P. Vivax* parásito presente en imágenes de frotis de sangre. El código está disponible: [https://github.com/ErickDiaz/bioinformatic\\_thesis\\_project/tree/dev](https://github.com/ErickDiaz/bioinformatic_thesis_project/tree/dev)

## 1. Actividad Grupal

### Link de la conferencia:

<https://icml.cc/Conferences/2021>

Para esta actividad seleccionamos la conferencia: International Conference on Machine Learning (ICML 2021). Decidimos llenar la plantilla para la conferencia con un trabajo de investigación que realizamos en 2019. El tema es de aprendizaje profundo aplicado en imágenes de frotis de sangre ya que aplicaría para esta conferencia tan importante en el área de aprendizaje automático. El contenido de las secciones son extractos del documento de investigación

Las plantillas para esta conferencia están disponibles en:

---

<sup>\*</sup>Equal contribution <sup>1</sup>Department de Ingeniería, Universidad Internacional de la Rioja, España. Correspondence to: Erick Diaz <ewdiazsa@gmail.com>.

<http://icml.cc/>

La plantilla por defecto tiene desactivado mostrar el nombre de los autores, esto es mientras esta en proceso de revisión y aceptación, para poder desplegar los nombres seguimos las instrucciones que están como comentarios en el documento, por lo que hay que cambiar la forma en que se llama al paquete donde está el formato enviado la una variable con el valor 'accepted'. Ejemplo:

`usepackage[accepted]icml2021`

## 2. Introducción

Según el informe de la Organización Mundial de la Salud (OMS), la malaria causó 435.000 muertes en todo el mundo en 2017 (WHO, 2018). Actualmente, el método estándar para detectar y diagnosticar la malaria es mediante el uso de un microscopio bajo observación a través de un frotis extendido de sangre colocada en un portaobjetos (WHO, 2015). En Guatemala, este tipo de diagnóstico (*P. Vivax malaria*) tiene varias desventajas, se necesita personal microscopista capacitado y experimentado para identificar y cuantificar el parásito en la gota gruesa (estándar de oro) y frotis de sangre. También existe una gran carga de trabajo para el personal en el área en la que se concentra la enfermedad; tiempo estimado alto en la observación de frotis de sangre, en promedio una persona entrenada observa un frotis de sangre espeso durante 10 a 20 minutos, en áreas endémicas un centro de salud puede recibir entre 200 y 300 frotis de sangre semanales, produciendo fatiga ocular, que suele ser el factor más influyente en un diagnóstico incorrecto. Prueba de automatización a través de redes CNN que podría acelerar los diagnósticos, aumentar la eficiencia y el desempeño de los especialistas.

## 3. Conjunto de Datos

Usamos el conjunto de imágenes *P. Frotis de sangre humana infectada con Vivax* (malaria), número de acceso BBBC041 versión 1, disponible en Broad Bioimage Benchmark Collection (Ljosa et al., 2012). Las imágenes en el conjunto de datos contienen 1364 imágenes, con un total de 80,113 células etiquetadas, diferentes investigaciones contribuyeron al

Table 1. Comparación de métricas de evaluación en los diferentes modelos de redes neuronales.

MODELO	IoU	PRECISIÓN	F1 SCORE
U-NET	71.1%	27.81%	34.16%
YOLO	77.17%	45.41%	6.96%
YOLO 1ST DATASET	77.19%	4.08%	7.44%
YOLO 2ND DATASET	84.85%	5.13%	9.42%
YOLO 3RD DATASET	83.66%	7.16%	12.42%

etiquetar cada celda en el conjunto de datos. Estas imágenes fueron aportadas por Jane Hung del MIT y el Broad Institute en Cambridge, MA. Cada célula en las imágenes biológicas tiene una etiqueta de clase y las coordenadas del cuadro delimitador, para la célula infectada tiene cuatro clases de gametocitos, anillos, trofozoítos y esquizontes.

#### 4. Metodología

Estamos ejecutando múltiples experimentos usando dos arquitecturas CNN diferentes y diferentes conjuntos de configuraciones para cada modelo. Para la segmentación de imágenes, usamos U-Net (?) y para la detección y clasificación de objetos, decidimos experimentar con YOLO (Redmon et al., 2015). Para U-Net, desarrollamos la arquitectura usando TensorFlow, para la función de pérdida del modelo usamos disimilitud de Bray-Curtis e IoU para la función de costo. Para el entrenamiento del modelo YOLO, usamos el repositorio original, después de que el modelo fue entrenado con el conjunto de datos original, usamos el aprendizaje por transferencia para continuar el entrenamiento con los conjuntos de datos aumentados.

##### 4.1. Aumento de datos

Para aumentar los datos de entrenamiento y eliminar el desequilibrio del conjunto de datos para una de las clases y mejorar la capacidad de generalización de los modelos, primero recortamos todas las imágenes de forma de parásitos utilizando la ubicación del cuadro delimitador y aumentamos estas imágenes con rotaciones, luego generamos un método que agrega al azar parásitos en las imágenes de entrenamiento que tenían menos de cuatro parásitos, el método siempre agrega la clase de parásitos que tiene menos ocurrencias, esto se ejecuta hasta que las diferentes clases estén cerca de estar equilibradas.

#### 5. Resultados

En esta sección se presentan los resultados obtenidos.

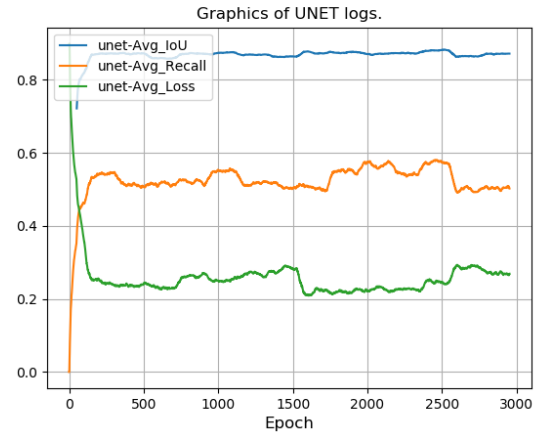


Figure 1. Métricas de evaluación durante el entrenamiento de la CNN U-Net.

#### References

- Ljosa, V., Sokolnicki, K. L., and Carpenter, A. E. Annotated high-throughput microscopy image sets for validation. *Nature Methods*, 9:637, 2012.
- Organization, W. H. *Guidelines for the treatment of malaria*. World Health Organization, 2015.
- Organization, W. H. *World malaria report 2018*. World Health Organization, 2018.
- Redmon, J., Divvala, S. K., Girshick, R. B., and Farhadi, A. You only look once: Unified, real-time object detection. *CoRR*, abs/1506.02640, 2015. URL <http://arxiv.org/abs/1506.02640>.

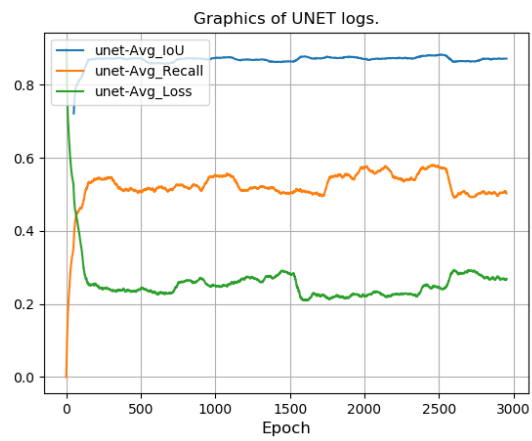


Figure 2. Métricas de evaluación durante el entrenamiento de la CNN YOLOv2.