



# Resumen: Using transfer learning and dimensionality reduction techniques to improve generalisability of machine-learning predictions of mosquito ages from mid-infrared spectra

Universidad del Quindío  
Facultad de Ciencias Básicas y Tecnologías  
Programa de Física

Daniel Esteban Salazar Camacho

## Resumen

Este artículo parte de que los mosquitos viejos tienen más probabilidades de transmitir la malaria que los jóvenes. Por lo tanto, la predicción precisa de la edad de la población de mosquitos puede mejorar drásticamente la evaluación de las intervenciones dirigidas a los mosquitos. Sin embargo, los métodos estándar para determinar la edad de los mosquitos son laboriosos y costosos. Los autores investigaron si la aplicación de reducción de dimensionalidad y aprendizaje por transferencia a los datos MIRS puede mejorar la transferibilidad de las predicciones basadas en MIRS para las edades de los mosquitos.

Los investigadores criaron adultos del vector de la malaria *Anopheles arabiensis* en dos insectarios: uno de la Universidad de Glasgow, Reino Unido y otro del Instituto de Salud Ifakara, Tanzania. Las cabezas y tórax de las mosquitos fueron escaneados utilizando un espectrómetro infrarrojo de transformada de Fourier con reflexión total atenuada ( $4000 - 400 \text{ cm}^{-1}$ ), que se agruparon en dos clases de edad diferentes (jóvenes de 1 a 9 días y viejos de 10 a 17 días). La dimensionalidad de los datos espectrales se redujo utilizando análisis de componentes principales no supervisado o incrustación estocástica vecina distribuida en  $t$ , y luego se utilizó para entrenar clasificadores de aprendizaje profundo y aprendizaje automático estándar. También se evaluó el aprendizaje por transferencia para mejorar la transferibilidad de los modelos al predecir clases de edad de mosquitos de nuevas poblaciones. En particular, se entrenaron diferentes modelos de aprendizaje profundo; (1) Modelo de red neuronal convolucional (CNN) sin reducción de dimensionalidad, (2) Perceptrón multicapa (MLP) con PCA como reducción de dimensionalidad y (3) MLP con  $t$ -SNE como reducción de dimensionalidad. Para todos los modelos, se agregó una capa SoftMax para transformar las salidas no normalizadas de unidades  $K$  en una capa completamente conectada en una distribución de probabilidad de pertenecer a una de las dos clases de edad.

El transfer learning se desarrolló de la siguiente manera: El conjunto de datos de Ifakara se utilizó como dominio de origen para el entrenamiento previo de los modelos ML. El conjunto de datos de Ifakara se dividió en conjuntos de entrenamiento y prueba, y el rendimiento del estimador se evaluó mediante la validación cruzada. Para poner a prueba las opciones de transfer learning, se seleccionaron aleatoriamente 82 o 33 espectros de los 1635 de los datos de Glasgow, lo que representa el 5 % y el 2 % del conjunto de datos, respectivamente. Es decir, Los modelos de ML entrenados previamente con el conjunto de datos de Ifakara se ajustaron con precisión utilizando subconjuntos del 2 % o el 5 % del conjunto de datos de Glasgow.

Las precisiones del modelo para predecir la edad de los mosquitos de la misma población que las muestras de entrenamiento alcanzaron el 99 % para el deep learning. Sin embargo, estos modelos no se generalizaron a una población diferente, logrando solo un 46 % de precisión para el aprendizaje profundo, respectivamente. La reducción de la dimensionalidad no mejoró la generalización del modelo pero redujo el tiempo computacional. Transferir el aprendizaje mediante la actualización de modelos previamente entrenados con un 2 % de mosquitos de la población alternativa mejoró el rendimiento con una precisión de 98 % para predecir las clases de edad de los mosquitos en la población alternativa.