Chapter 4

Combining collective and artificial intelligence for global health diseases diagnosis using crowdsourced annotated medical images

4.1 Introduction

Achieving Universal Health Coverage by 2030 is one of the Sustainable Development Goals and World Health Organization priorities (WHO) (UN, 2015). Half of the world's population lacks access to essential health services and diagnosis is a key step to achieve universal healthcare. Many of these diseases are diagnosed by visual inspection, which requires experts in front of the microscope and other medical devices at a certain time, a resource that is not always available. Artificial intelligence (AI) presents an opportunity to support these diagnostic processes.

The number of AI-based medical devices for diagnosis is increasing. From 2015 to 2020, 222 AI devices were approved in the USA and 240 in Europe (Muehlematter et al., 2021). Most of them were developed for radiology and cardiovascular diseases, and none of them is for microscopy applied to microbiology. Recently a few algorithms were trained to detect Malaria's parasites (Vijayalakshmi A & Rajesh Kanna B, 2020; F. Yang et al., 2020) and helminth's eggs in fecal samples (Holmström et al., 2017; Mathison et al., 2020; A. Yang et al., 2019), showing the potential of AI algorithms for microscopic images. Notwithstanding the above, more studies are needed to create algorithms approved by regulatory institutions.

Soil-transmitted helminthiasis (STH) is a neglected tropical disease that affects 1.5 billion people worldwide in the most deprived communities. To reduce the health burden, the WHO has outlined a roadmap for STH elimination, emphasizing the need for innovation and technologies like AI (WHO, 2020). The recommended diagnostic method for STH is the Kato-Katz technique, which prepares stool samples for microscopic detection and quantification of helminth eggs(Katz et al., 1972).

Image annotation to train AI models is a time-consuming labour that poses an important burden into experts. However, in recent years, the use of crowdsourcing has been proposed to overcome this problem by delegating this task on a large group of untrained annotators. Several studies have already demonstrated the validity of the use of crowdsourcing for annotating medical images. In 2012, Luengo-Oroz et. al demonstrated that the combination of the annotations collected using a video game of 22 players achieved a malaria parasite counting accuracy higher than 99% in thick malaria smears (Luengo-Oroz et al., 2012). In 2019, Linares et. al demonstrated that combined annotations form 25 players were able to distinguish most Malaria species with an accuracy of 99% (Linares et al., 2019). Furthermore, Keshavan et. al combined crowdsourcing and Deep Learning (DL) to predict the quality of Magnetic Resonance Imaging (Keshavan et al., 2019).

Within this context, this work proposes a methodology to train DL algorithms for quantifying parasitic infection in microscopy images with the following objectives: 1) to assess the feasibility of training DL algorithms for the differentiation of helminths eggs based on microscopy images with annotations obtained from crowdsourcing using a custom video game, 2) to identify the relationship between the amount of training data and the deep learning model performance using incremental training and 3) to compare the performance of AI models trained with data annotated by both untrained school-age children and adults (general population).

This study addresses **objective 2** of this PhD thesis. The primary contributions of this study are the introduction of crowdsourcing as a method for annotating medical images to address the shortage of expert annotators, along with evidence of its effectiveness in training deep learning algorithms using supervised methods.

4.2 Methodology

4.2.1 Crowdsourcing image annotation

We developed SpotWarriors (SW), a publicly available[1] set of mini-games that contribute to the diagnosis of diseases while playing, by generating crowdsourced annotated medical images. For the purpose of this project, we focused on a mini-game for the classification of small image patches from digitized stool samples for the identification of different helminths eggs, including *Ascaris* spp., *Trichuris* spp., Hookworms, and images without eggs. Figure 4.1 shows a screenshot of the game used.

All data used in this study for training, validation and testing of the AI algorithm came from 41 digitized stool samples from 6 different infected patients who were part of a follow-up study. Digitization of samples were made at 10x magnification. Ethical approval was obtained from the Kenya Medical Research Institute (KEMRI) Ethics Review Committee (SERU 3873).

From all digitized samples, we generated a total of 10319 cropped image patches (256x256 pixels) without overlap. Our interpretation is that all image patches can be considered as independent although they come from a limited number of subjects. We introduced 700 randomly selected image patches in the video game which were annotated by at least 20

[1] https://spotwarriors.org/en/, also available at Google Play Store and Apple App Store.

adults and 20 school-age children (from 11 to 18 years old) players. These annotated images were used for training DL algorithms. For comparative purposes, and to assess the quality of crowdsourced annotations, these training images were also analyzed by experts microscopists.

Annotations from school-age children were obtained by organizing workshops in different schools. The workshops, presented in collaboration with the teachers, included an explanation of the project and concepts related to global health, artificial intelligence and collective intelligence in addition to playing the game. Data from adults was collected anonymously from online players.

Additionally, the remaining images were annotated by experts and were used as validation and test sets (2932 and 6678 images respectively, randomly separated). For crowdsourced annotated images, the ground truth (GT) was generated using the majority voting rule, where the most common response among players was chosen. The final distributions for the training set (those images introduced in the game), validation set, and test set are presented in Table 4.1.

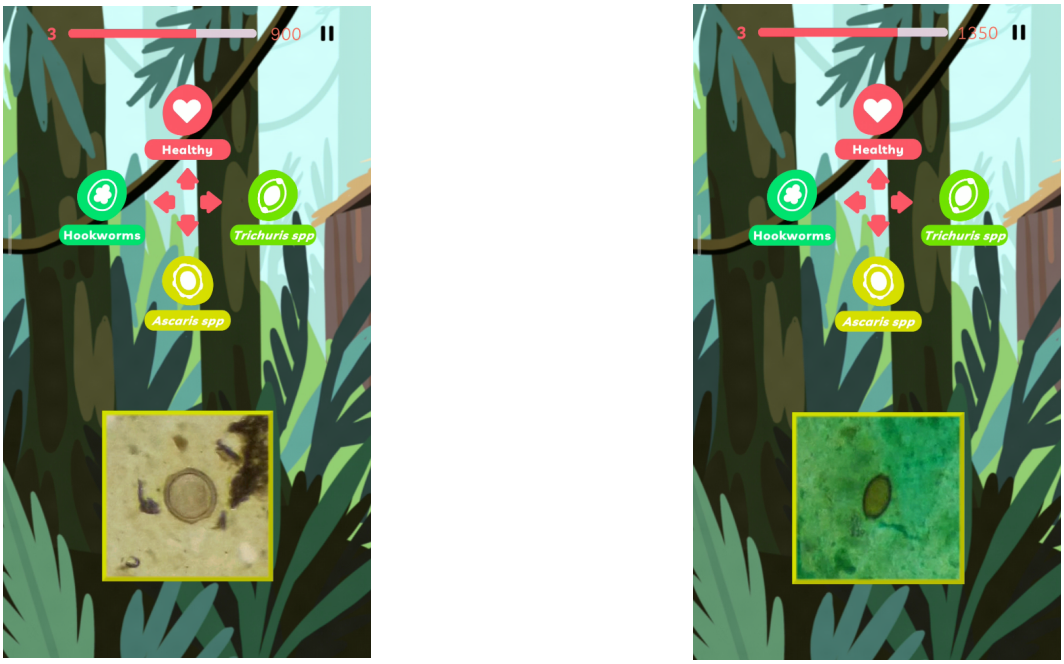


Figure 4.1: Screenshot of the mini-game used to collect data with *Ascaris* spp. (left) and *Trichuris* spp. (right)

	Train			**Validation**	**Test**
	E	SAC	A	E	E
Ascaris spp.	179	182	176	1258	2819
Trichuris spp.	212	198	211	241	570
Healthy	309	320	313	1433	3298
Total	700	700	700	2932	6687

Table 4.1: Distribution of the training, validation and test sets. Training images are annotated by three groups: experts (E), school-age children (SAC) and adults (A).

4.2.2 AI architecture: Deep learning model

In the present work, we used a Convolutional Neural Network (CNN)-based algorithm to solve the classification task for differentiating helminths eggs along with non infected sample images. The algorithm, given an image, returns an output probability distribution along the different classes under study, and the final predicted label is then computed as the one that has the highest probability. Particularly, we used the MobileNet V2 model (Sandler et al., 2018), a light-weighted architecture designed to run on mobile phones in an efficient manner. This particular architecture has three main components including depthwise convolutions that significantly reduce the number of parameters, inverted residual connection blocks which modify residual blocks for efficiency purposes, and linear bottleneck layers without any non-linear activation function in order to preserve information in the low dimensional space. MobileNet V2 architecture is composed by 157 layers and involves only 3.5 million of parameters, compared to the 138,4 million of parameters that are involved in the well known VGG-16 architecture along its 23 layers (Simonyan & Zisserman, 2014).

To overcome the limitation of having a small training dataset, we used a transfer learning technique by pretraining the MobileNet V2 model on a large dataset (ImageNet (Russakovsky, Deng, et al., 2015)) and fine-tuning it in our dataset for the classification of helminths eggs. Fine-tuning was performed by freezing the earlier layers which learn generic features, and retraining later layers, responsible for extracting specific features of the problem under study. Using this technique we can reduce the computational cost and result in better performance than training from scratch, specially when little training data is available.

Because crowd-sourced annotations usually contain some incorrect labels, we used soft bootstrapping cross entropy loss function, which minimize the damage of incorrect labels by dynamically updating the targets of the prediction based on the actual state of the model (Reed et al., 2015). The loss function is defined as equation 4.1, where $\mathbf{q}$ is the prediction, $\mathbf{t}$ is the target, β is the scaling factor between predictions and targets and L is the number of classes under study.

$$\mathcal{L}_{\text{soft}}\left(\mathbf{q}, \mathbf{t}\right) = \sum_{k=1}^{L} \left[\beta t_k + (1-\beta) q_k\right] \log\left(q_k\right) \tag{4.1}$$

Furthermore, we used data augmentation including rotation, shift, flip, zoom, and shear transformations to generate more training data to further improve the model performance. Additionally, we also used early stopping technique during the training process with a patience of 10 iterations to avoid overfitting on the training set.

DL models were designed and trained using Keras with Tensorflow, and using a GPU NVIDIA Tesla T4 16GB.

4.3 Experiments and Results

In this section, we first evaluated the quality of crowdsourced annotations using a majority voting mechanism. We continued with the study of the effect of the training sample size on the performance of the model. And finally we compared the DL model trained with school-age

children and adults annotations.

To evaluate the quality of crowdsourced annotations, we used the accuracy metric, which is defined as $ACC = TP + TN/N$ where TP, TN and N stand for true positives, true negatives and total number of samples respectively.

We used the area under the receiver operating characteristic curve (AUC) for evaluating the performance of DL algorithms. AUC measures the performance of the model across all possible probability thresholds. Macro-average AUC along classes is computed in order to avoid bias due to imbalanced class distribution.

In order to obtain a robust metric not affected by training instability, we repeated the training process 5 times, and calculated the mean and standard deviation of the performance metrics. The training of the models was carried out using the training set while the validation set was used for hyperparameter tuning. The incremental training experiment and the final performance evaluation were assessed on the test set.

It should be noted that we did not include Hookworm class in the analysis due to the lack of representativity in our database. No preprocessing was made on the images.

4.3.1 Quality of crowdsourced annotations

In order to select the optimal size of the quorum that best performs in comparison with the expert annotations, we used a bootstrap sampling method. For each image, we generated the final annotation using the majority voting rule considering only N randomly selected annotations. To measure the stability of each quorum size (N) we repeated this process 10 times. Table 4.2 shows the difference on the quality of crowdsourced annotations using different quorum sizes, by comparing the generated annotations by players and annotations made by experts. As derived from the table, we can observe that annotations based on 20 different player responses obtained the best performance.

Even though adult annotations obtained better accuracy with respect to expert annotations, annotations from 20 school-age children were found to be of enough quality (accuracy $> 94\%$).

Quorum size (N)	School-age children	Adults
5	0.842 (0.007)	0.966 (0.004)
10	0.910 (0.006)	0.988 (0.002)
15	0.931 (0.005)	0.989 (0.003)
20	0.946 (0.004)	0.991 (0.002)

Table 4.2: Mean accuracy and the standard deviation of crowdsourced annotations using different quorum sizes compared to the ones made by experts.

4.3.2 DL model: hyperparameter tuning

Mobilenet V2 is built on different blocks. To determine the optimal number of layers to be fine-tuned during transfer process, we froze a determined number of blocks and fine-tuned the rest. For this experiment, models were trained with experts annotations.

Furthermore, with the aim of evaluating the effectiveness of soft bootstrapping loss for noisy labels, and to select the best loss function for this particular case study, we trained the DL model using both conventional cross entropy and soft bootstrapping cross entropy ($\beta = 0.95$) loss functions. Models were trained with school-age children and adults annotations (expert annotations do not contain noisy labels). As derived from Table 4.3, the best performance was obtained when the first 46 layers were not fine-tuned and when bootstraping categorical cross entropy was used as the loss function.

Hyperparameter		**AUC**
Number of frozen layers		
	19 (block 2)	0.873 (0.033)
	46 (block 5)	**0.919 (0.019)**
	73 (block 8)	0.914 (0.014)
	99 (block 11)	0.917 (0.011)
Loss function		
Children	Cross entropy	0.927 (0.007)
	Bootstrapping cross entropy	**0.932 (0.006)**
Adults	Cross entropy	0.911 (0.014)
	Bootstrapping cross entropy	**0.925 (0.016)**

Table 4.3: Hyperparameter selection. Mean AUC and standard deviation are shown.

4.3.3 Incremental training

To study the effect of training sample size on the models performance we trained the model incrementally using different sample sizes. We performed this incremental training by steps of 100 training images from 100 to 700.

Figure 4.2 shows the results of the incremental training experiment, revealing the importance of the number of samples used for training and its impact on the model performance.

4.3.4 Differences between annotator groups

We compared the performance of the AI model trained with school-age children, adults and expert annotations. Table 4.4 summarizes the results of the three models using all available training samples (N=700). The results show that all models perform similarly, highlighting the power of the use of crowdsourced annotations which obtained similar results when compared to the ones obtained by the model trained on expert-based annotations. It should be noted that although the quality of school-age children annotations were lower than the ones collected by adults (difference of 4.5% in the accuracy, see Table 4.2), the AI algorithm is robust to noisy labels and decreases the difference in terms of the model performance (0.9%).

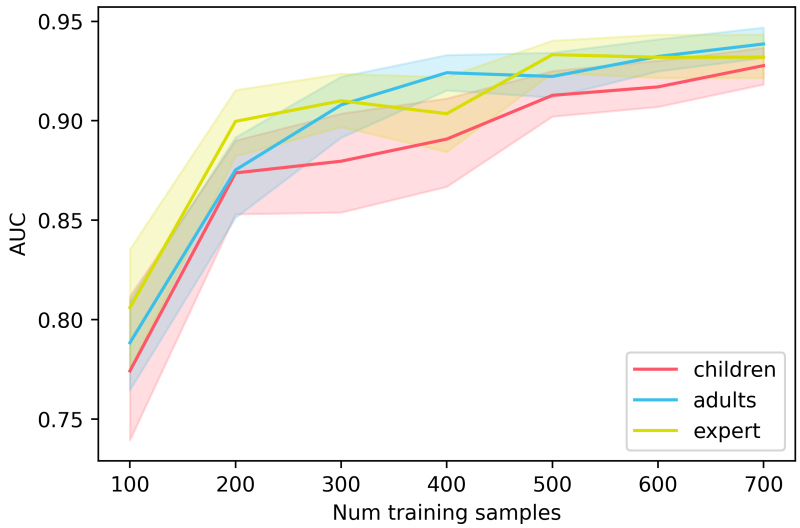


Figure 4.2: Evolution of the model performance (AUC) as the training sample size increases. Results from data annotated by children, adults and experts are presented independently.

AUC	Experts	Adults	Children
Ascaris spp.	0.960 (0.003)	0.958 (0.007)	0.948 (0.002)
Trichuris spp.	0.912 (0.002)	0.926 (0.015)	0.912 (0.026)
Healthy	0.924 (0.019)	0.932 (0.010)	0.923 (0.008)
Mean	0.932 (0.015)	0.939 (0.01)	0.928 (0.012)

Table 4.4: Detailed performance of the DL algorithm trained with different annotations. Mean AUC and standard deviation is reported.

Figure 4.3 shows the prediction result of the model on different images from the test set, including the three classes under study (*Ascaris* spp., *Trichuris* spp. and healthy). In addition, we computed the gradient-weighted class activation mapping (Grad-CAM) to visualize the the regions in the image that is important for the model to make the decision (Selvaraju et al., 2016).

4.4 Conclusions

This work shows promising results on the use of crowdsourced annotation for the development of AI-based diagnosis systems, and validates its use in the medical image field, where manual annotations from experts is a time-consuming labour, and requires high specialization.

In this work, we collected crowdsourced annotations by using a customized video game to classify different species of Helminths eggs, and used these annotations to train a CNN architecture (MobileNet V2). Particularly, we obtained crowdsourced annotations from both untrained school-age children and adults, and the results showed that DL models trained on those annotations performed in a similar manner compared to the ones trained with expert

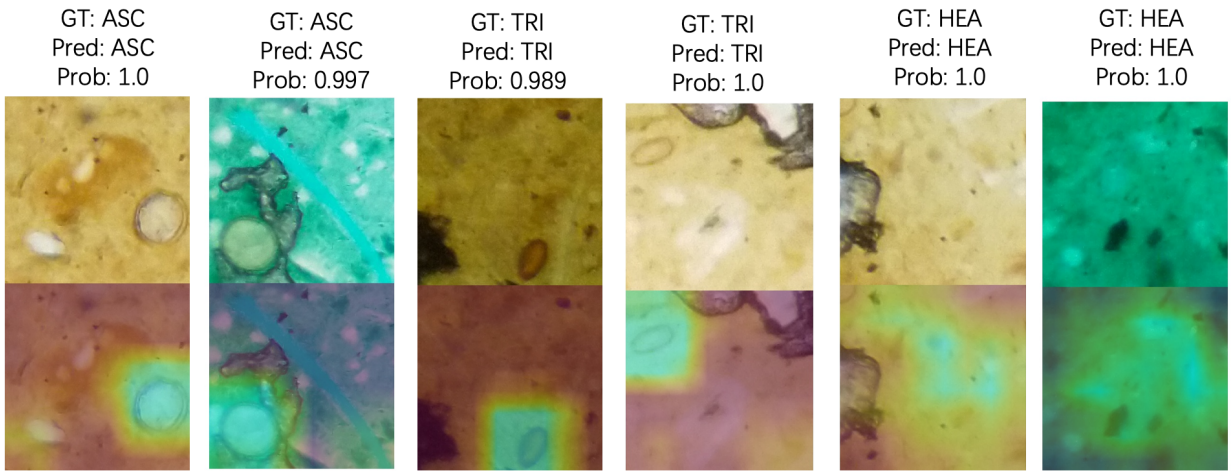


Figure 4.3: Examples of image samples representing all classes under study (upper row) along whit its activation maps generated by the DL algorithm (bottom row). Correct label (GT) as well as predictions (Pred) appear above each image. Activation maps for healthy samples (HEA) focused on the entire image with no significant activation, while activation maps for *Ascaris* spp. (ASC) and *Trichuris* spp. (TRI) focused exactly on the egg location.

annotations (AUC of 0.928, 0.939 and 0.932 for children, adults and expert annotations respectively). No significant differences in the model performance were found when used children, adults or experts annotations.

On the other hand, we showed the impact of the training sample size on the performance of the AI models. We showed that we obtained better performance as the training sample size increases. In particular, model performances increased approximately by a factor of 20% when trained on all available samples (700) compared to the result obtained when only 100 images were used for training. As derived from this experiment, we can conclude that we could train our DL model in an iterative manner as we obtain more images annotated by players, and thus obtaining a more robust algorithm with a better predictive capacity.

Additionally, this work lays the foundation for the use of video games as data enrichment platforms to automate and scale the medical image labeling process using human collective intelligence, enhancing human relevance in the process of developing AI algorithms.

4.5 Environmental Impact

In this study a cumulative of 43 hours of computation was performed on GPU (Tesla T4), which 20 hours contributed to obtain the final results. The total emissions, estimated by MachineLearning Impact calculator presented in (Lacoste et al., 2019), was 0.9 kg of CO_2. Virtual machines that host our SpotWarriors game are estimated to emit 0.39 kg CO2eq per month.

Chapter 5

Deployment, adaptation and evaluation in real time and real world settings of mobile AI assistants for soil-transmitted helminths quantification

5.1 Introduction

Soil-transmitted helminthiasis (STH), a neglected tropical disease, affects approximately 1.5 billion people, primarily in tropical and subtropical regions, with children being the most vulnerable (WHO, 2022). This disease leads to significant health issues such as malnutrition and impaired growth. In low- and middle-income countries (LMICs), the diagnosis of STH largely relies on conventional microscopy, with the WHO-recommended Kato-Katz technique as the standard for quantifying STH eggs in stool samples (Katz et al., 1972). While newer diagnostic methods, such as FECPAKG2, mini-FLOTAC, and DNA-based tests like PCR, offer higher sensitivity and advanced capabilities, they require well-equipped laboratories and skilled technicians, making them costlier and less accessible in resource-limited settings.

The ideal diagnosis tool is a point-of sampling test that allows the quantitative detection of STH in all groups, could be performed and interpreted by health personeel and community health workers with a single day of training, and be portable. The test should be specific ($\geq$ 94%) for *Ascaris*, *Trichuris* and hookworm and have a sensitivity of at least 60%. Moreover, the test should take less than nine minutes (at least seven samples per hour) and cost less than US $3 (WHO et al., 2021a).

Several studies have aimed to automate the analysis of STH samples using artificial intelligence (AI), but these efforts still face notable limitations. Some AI systems were designed to run on computers, which can be a constraint for point-of-care usage (Holmström et al., 2017; Li et al., 2020; P. Ward et al., 2022). Others have been deployed on smartphones, offering more

portability, but these studies are often limited by small sample sizes, which can impact the generalizability and robustness of the AI models (A. Yang et al., 2019).

A pioneering work was performed by Yang et. al in 2001, which used 82 microscopic images to develop an AI system for parasite egg detection. The system employed a median filter to remove small objects and then applied binary thresholding to binarize the image. From the binary image, objects larger than 600 pixels were segmented using a connected component operator, and features such as size, shape, and eggshell characteristics were extracted. These features were then processed through a two-stage artificial neural network: the first stage distinguished parasite eggs from artifacts, and the second stage identified the species of the detected eggs (Y. S. Yang et al., 2001)

In 2017 Holmström et. al introduced a study incorporating deep learning. They used a customized scanner called MoMic to acquire images, which were then uploaded to the WebMicroscope platform for image analysis using deep learning algorithms. The proposed algorithm consisted of two stages: the first stage analyzed the entire image to identify egg candidates, and the second stage classified the cropped candidate regions into four categories. Despite using a relatively small dataset of 410 positive images, the study demonstrated the potential of deep learning for the automatic detection and classification of helminth eggs, achieving a precision of 97.3% on a test set of 205 images (Holmström et al., 2017).

In 2019, Yang et al. developed KankaNet, the first object detection model that detects and classifies helminth eggs in one step. The model was trained using 516 samples, with images acquired using both mobile phones and USB video class endoscopes. KankaNet achieved a mean sensitivity of 61.1%, 97.8%, and 100% for *Ascaris*, *Trichuris*, and hookworm, respectively, on a validation set of 186 images. The model was also deployed as an Android app, making it accessible for field use (A. Yang et al., 2019).

In 2020, Jiménez et al. presented the Helminth Egg Automatic Detector (HEAD), a machine learning algorithm using a k-nearest neighbor classifier to classify images based on features extracted through traditional methods such as filtering and watershed segmentation. The algorithm is capable of classifying nine species of helminths, achieving a sensitivity of 96.82% and a specificity of 97.96%. The model is hosted on a web service, facilitating remote access and use (Jiménez et al., 2020). Li et al. introduced FecalNet, a deep learning-based algorithm developed with a large dataset of 22,440 images from 1,122 patients. FecalNet used a ResNet-152 architecture to detect six classes: erythrocytes, leukocytes, intestinal mucosal epithelial cells, hookworm eggs, *Ascaris* eggs, and whipworm eggs. The algorithm achieved an average precision of 92.16% and an average recall of 93.56%, highlighting its efficacy in handling a diverse set of image classes. However, ResNet-152 is a huge neural network that has 115 million parameters, requiring large computational resource for inference (Li et al., 2020).

In 2022, Naing et al. explored the application of YOLO object detection algorithms for stool sample examination. They trained the algorithm with 1,594 images across 34 classes and tested it with 176 images. YOLO v4 achieved a precision of 96.25% and a recall of 95.08%, demonstrating the potential of YOLO models for high-accuracy detection in complex datasets (Naing et al., 2022). In the same year, Ward et. al presented proof of concept for the detection of soil-transmitted helminths and *Schistosoma mansoni* eggs in Kato-Katz samples.

The model, Region-Based Fully Connected Network (R-FCN) with ResNet101 as feature extractor was trained and validated on samples from 4 countries, with a total of 7780 field of view images. The performance in the test set with 752 field-of-view images is 94.9% average precision (P. Ward et al., 2022).

While significant progress has been made in automating the analysis of STH samples using AI, several limitations persist. Many of the current AI systems require robust computational resources, making them impractical for deployment in point-of-care settings. Additionally, some models are trained on relatively small datasets, limiting their generalizability and effectiveness across diverse populations and environments. The need for specialized equipment, such as custom scanners or high-resolution digital microscopes, also poses challenges in regions where such resources are scarce. Furthermore, the existing AI solutions often rely on cloud-based processing, which can be hampered by limited internet connectivity in remote areas. These constraints underscore the need for more accessible, scalable, and user-friendly AI solutions tailored to the specific challenges of global health settings.

In response to these limitations, our work introduces a comprehensive AI development and validation system designed to overcome these barriers. This system includes AI-assisted image acquisition using a mobile phone coupled to the microscope's ocular, making it both cost-effective and easy to deploy in resource-limited settings. Additionally, we incorporate an AI-assisted, human-in-the-loop labeling process on a telemedicine platform, ensuring that the training data is both accurate and reflective of real-world conditions. Our fast deployment pipeline enables the trained AI models to be quickly adapted and rolled out in the field, facilitating rapid implementation. Importantly, our AI system is designed to be used at the point of sampling by community health workers, who can be trained in just one day, thereby reducing the reliance on specialized technicians and making advanced diagnostic tools more accessible in LMICs.

This study addresses several key objectives of this PhD thesis: the creation of a large-scale microscopic dataset (Objective 1), the design and implementation of an AI development and deployment workflow (Objective 2), and the evaluation of AI models and their integration in resource-limited settings (Objective 4). The primary contributions include the creation of large scale microscopic dataset for helminthiasis, the development and validation of an AI algorithm to assist in the diagnosis of STH, the creation of a pipeline for deploying AI models on smartphones, and the adaptation of this technology for use in resource-limited environments. Additionally, a human-in-the-loop labeling process was integrated into the telemedicine platform to expedite data labeling. These outcomes enable rapid expansion into new applications, simplifying the AI development process and enhancing the overall efficiency of diagnostics.

5.2 Material and methods

The study is structured into three distinct phases: proof of concept with limited data, massive data acquisition and AI model creation during a pilot study, and model refinement through human-in-the-loop labeling.

In the initial phase, the focus is on assessing the feasibility of developing an AI algorithm for the detection of STH eggs and its deployment on smartphone. This preliminary experiment is conducted in a controlled environment with participants who are well-acquainted with the technology. An AI model is deployed as a baseline, providing a foundation for subsequent analyses and serving as a proof of concept for the approach.

The second phase involves a pilot study conducted in Kenya during a deworming campaign. The primary objective is to establish a substantial image database of STH samples while evaluating the rapid development and deployment of AI models in a real-world setting. In this phase, a group of eight analysts, who are new to the system, received a one-day training session to use the AI-assisted technology. With this assistance, they digitize and analyze samples collected during the campaign. Some of the digitized images are labeled by experts on-site, allowing for immediate feedback and iterative model improvements. Throughout this phase, four iterations of model enhancements were released, each refining the AI system's performance and expanding its capabilities.

In the final phase, due to the large number of digitized images that remain unlabeled, the AI model is deployed on the cloud to establish a human-in-the-loop labeling system. This system accelerates the image labeling process by integrating human expertise with AI capabilities. Analysts interact with the AI to review and confirm labels, ensuring high accuracy and quality of the data. The final model developed in this phase reflects a collaboration between human expertise and AI, resulting in a robust and reliable diagnostic tool tailored to real-world conditions.

5.2.1 Phase 1: Baseline model

The primary objective of this phase is to conduct a proof of concept, assessing the feasibility of transitioning from a conventional, manual analysis workflow to a fully digital workflow. This involves developing an AI algorithm specifically for the detection of STH and integrating this technology into a smartphone. Additionally, this phase aims to establish a baseline model, which serves as a pre-trained foundation for further development and refinement in the subsequent phases of the study.

A total of 51 Kato-Katz slides were digitally captured using a smartphone-based digitization system. The system consisted of a custom Android app, a 3D printed adapter, and a conventional microscope. By aligning the smartphone camera with the eyepiece of the microscope, images of the entire field of view were captured (Figure 5.1 left). To ensure diversity in our database, two different smartphones, namely the Xiaomi Pocophone F1 and Bq Aquaris X2, were utilized during the digitization process. It is important to note that the database exclusively contains images of *Ascaris* and *Trichuris* eggs, as hookworm eggs have a tendency to disappear within 30 minutes of sample preparation. These samples were obtained from a total of 12 subjects, consisting of 6 positive and 6 negative cases. In total, 1508 image fields were digitized, contributing to the comprehensive dataset for analysis.

The uploaded images were stored in cloud storage and accessed through a telemedicine platform for the purpose of labeling. The identified parasites were marked by placing bounding boxes around them (Figure 5.1 right).

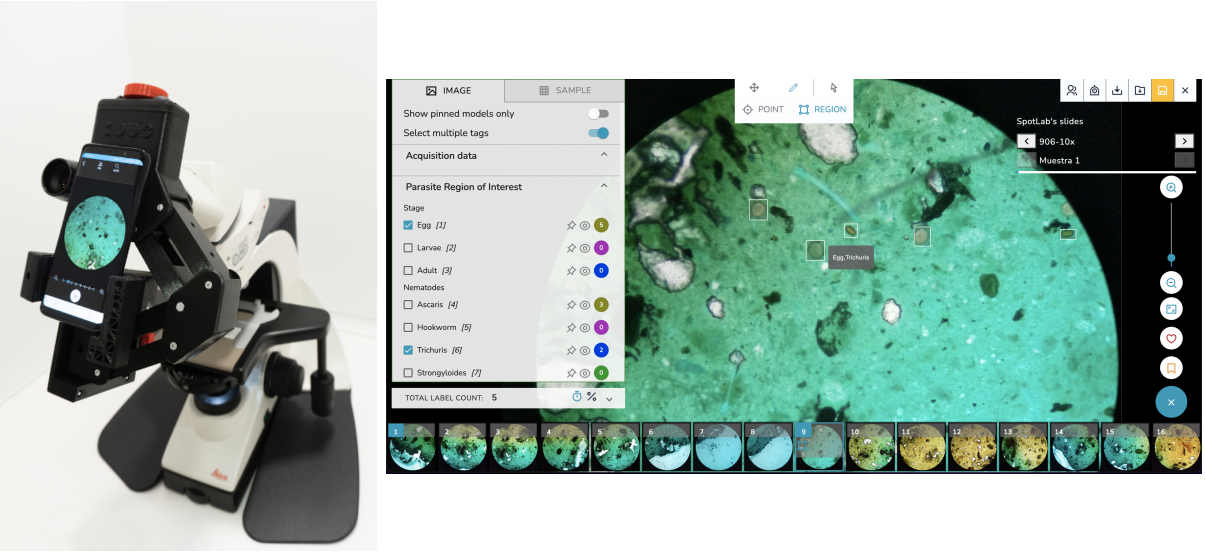


Figure 5.1: Digitisation and labeling system. (left) digitisation system, By aligning the smartphone camera with the microscope's ocular with the smartphone, the conventional microscope is converted to a digital microscope (right) Telemedicine platform for image visualisation and labeling.

To achieve real-time detection capabilities on a smartphone, the Single Shot Multibox Detector (SSD) architecture was selected, paired with a modified version of MobileNet V2 that incorporates a Feature Pyramid Network (FPN) as a feature extractor. This combination creates a mobile-friendly version of RetinaNet(Lin, Goyal, et al., 2017; Liu et al., 2016; Sandler et al., 2018). Compared to the original SSD, which uses the VGG network, replacing it with MobileNet V2 significantly enhances speed and efficiency, making it better suited for mobile deployment. The inclusion of the Feature Pyramid Network improves the model's ability to detect small objects by making predictions at multiple scales, which is crucial for accurately identifying tiny parasitic eggs. Additionally, the use of focal loss helps to address class imbalance by assigning greater weights to difficult or easily misclassified instances, improving the overall performance of the model.

The training dataset was created by generating 640x640 patches from larger 4032x3024 images, specifically focusing on regions where parasites were present. Multiple patches were generated around a single parasite to increase the likelihood of object presence in the training data, effectively augmenting the dataset. The model was trained using TensorFlow's Object Detection API and then exported to TensorFlow Lite (tflite) for mobile inference. An Android application was developed to run this model in real-time, enabling rapid and accurate detection of STH eggs directly on the smartphone.

5.2.2 Phase 2: Pilot study

The pilot study is structured around three main objectives:

1. **Evaluation of the AI-Assisted STH Analysis System**: The first objective is to assess the performance and usability of the AI-assisted system in real-world scenarios,

Figure 5.2: Real-time AI-assisted digitization and analysis of Kato-Katz samples by analysts

particularly with new users who are unfamiliar with the technology. This helps to understand how well the system functions in a practical setting, identifying any areas that may require further refinement or support.

2. **Acquisition of a Large Dataset of STH Images**: The second objective focuses on collecting an extensive dataset of STH images, with a particular emphasis on images of hookworm. This dataset is crucial for training and validating AI models, enhancing their accuracy and reliability in detecting STH infections. The acquisition of diverse and representative data is essential for improving the generalizability of the AI models.
3. **Assessment of AI Model Development and Implementation Capabilities**: The final objective is to evaluate the efficiency and effectiveness of the AI model development and deployment process. This includes assessing how quickly and accurately AI models can be trained, updated, and integrated into the analysis system based on real-world data and feedback from users.

The pilot encompasses four iterations of a methodology consisting of the following steps: data acquisition, labeling, AI model training and AI model deployment. Over two weeks, the iterative process that was carried out in 24-48h periods consisted of collecting samples in schools, preparing, digitizing, analyzing the Kato-Katz slides with an AI copilot before the hookworm eggs disappeared, and uploading the data to the telemicroscopy web platform where the images were labeled by experts to create an AI model for the following iteration. For each slide, the laboratory staff scans the sample at 10x magnification using a smartphone that is coupled to an optical microscope with a 3D printed adapter aligning the smartphone

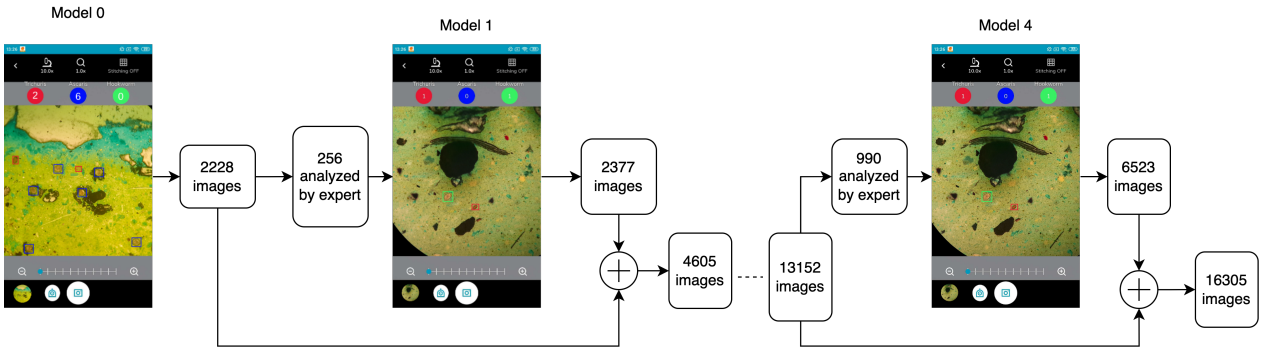


Figure 5.3: Workflow of the AI model creation and deployment during the pilot study

camera with the microscope ocular lens, as depicted on Figure 5.2. As the microscopist examinates the sample, an AI algorithm co-pilot analyzes the sample through the smartphone camera detecting and counting parasites in real time. All the fields in which the algorithm detected a parasite and those in which there was a parasite but the algorithm did not were saved by the analysts - a process triggered by the use of a foot pedal. These images were uploaded whenever there is connectivity to a telemedicine platform and part of them are labeled remotely by experts on the platform. While an AI model was trained first starting from a baseline model that lacked any hookworm examples but included images from other helminths (*Ascaris* spp. and *Trichuris* spp. eggs), the subsequent iterations allowed the AI model to be re-trained. It was then deployed in the acquisition app for use on the following day of the field campaign, assisting the slide analysis and digitization for the next model.

For each iteration (from model 1 to 4), the baseline model served as the pre-trained model, which was then fine-tuned using the newly annotated data specified in Table 5.1. The training process followed a similar approach to the baseline model. All images were resized to ensure that the minimum side was 640 pixels, and then patches of size 640x640 were generated. For minority classes, such as *Trichuris* spp. and hookworm, mosaic augmentation techniques were applied. This involved cropping the image into patches of size 320x320, ensuring that each patch contained at least one parasite, and then creating a new 640x640 image by combining four of these patches. Additionally, image augmentation techniques were employed during training, including random flipping, rotation, random adjustments of brightness, hue, and saturation, random cropping, and adjustments to JPEG quality. These augmentations were implemented to address the observed large variation in sample characteristics.

5.2.3 Phase 3: Post pilot study

After the two-week study, a total of 1,343 stool samples were collected, but only a small portion (71 samples with 2,235 labels) was initially analyzed by experts. To expedite the analysis process and manage the large volume of data, a human-in-the-loop annotation system was developed. This system involved deploying an AI algorithm on the cloud, which automatically analyzed each image uploaded to the telemedicine platform. Analysts could then review the AI's predictions, modify any incorrect labels, and save the revised analysis.

Model name	Training	Validation		N. Labels training	N. labels validation	N. labels total
Model 0 (baseline)	samples: 39 images: 613	samples: 14 images: 132	**Total**	**5039**	**1024**	**6063**
			Ascaris	3960	841	4801
			Trichuris	1079	183	1262
			Hookworm	0	0	0
Model 1	samples: 35 images: 198	samples: 13 images: 54	**Total**	**281**	**63**	**344**
			Ascaris	148	45	193
			Trichuris	109	11	120
			Hookworm	24	7	31
Model 2	samples: 44 images: 350	samples: 15 images: 117	**Total**	**680**	**195**	**875**
			Ascaris	218	86	304
			Trichuris	257	15	272
			Hookworm	205	94	299
Model 3	samples: 43 images: 445	samples: 17 images: 100	**Total**	**796**	**206**	**1002**
			Ascaris	235	69	304
			Trichuris	363	36	399
			Hookworm	198	101	299
Model 4	samples: 47 images: 679	samples: 25 images: 311	**Total**	**1685**	**550**	**2235**
			Ascaris	657	179	836
			Trichuris	494	211	705
			Hookworm	534	160	694

Table 5.1: Training and validation data distribution for each iteration, where model 0 is the baseline for model 1-4. Model 1-4 were deployed during the pilot study

During this phase, a significantly larger number of images from the pilot study were analyzed with AI assistance on the telemedicine platform, resulting in a total of 20,276 labels. These labels consisted of 13,775 *Ascaris* spp., 1,822 hookworm, and 4,679 *Trichuris* spp. The dataset was divided into three groups: training, validation, and test sets, according to the distribution outlined in Table 5.2. With this expanded dataset, we trained a new AI model, leveraging the increased volume and diversity of labeled images to enhance the model's accuracy and robustness.

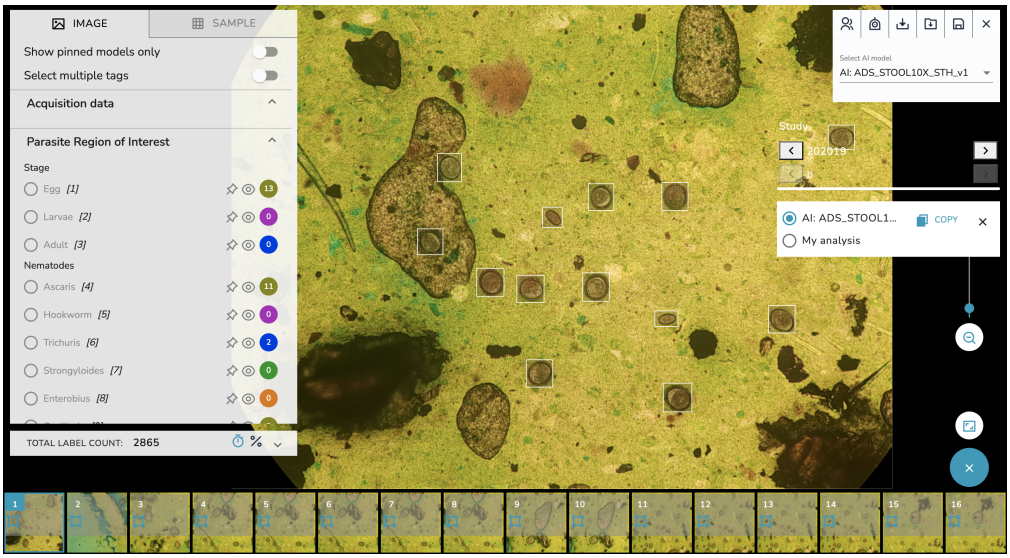


Figure 5.4: Human in the loop review. Instead of manually replacing the bounding boxes, analyst only has to review the predictions

Label	Train	Valid	Test	Total
Ascaris	11541	1350	884	13775
Trichuris	3163	703	813	4679
Hookworm	1276	331	215	1822
Total	15980	2384	1912	20276

Table 5.2: Data distribution of the latest AI model

5.3 Experiments and results

5.3.1 Phase 1: Baseline model

The algorithm was first focused on *Trichuris* spp. detection, and then extended to the detection of both *Ascaris* spp. and *Trichuris* spp.. We divided our dataset into 2 sets, training (85%) with 808 *Trichuris* spp. and 3649 *Ascaris* spp. eggs, and validation (15%) with 136 *Trichuris* spp. and 647 *Ascaris* spp. eggs. Table 5.3 summarizes the results and proves that the proposed model can be extended for the detection of different helminth eggs obtaining promising results including a mean precision of 94.36% and mean recall of 93.08%.

5.3.2 Pilot study

During the pilot study, a total of 1343 stool samples from children aged between 5-15 years old from 13 school were collected and analyzed by eight technicians. Together with the sample collection, an epidemiological questionnaire was carried out. All informed consent signs were obtained. The collected samples were analyzed by both conventional microscopy and using AI assisted analysis. Table 5.4 shows the result obtained with the conventional microscopy

Label	Precision	Recall	F-Score
Trichuris	95.31	89.71	92.43
Ascaris	93.41	96.45	94.91
Mean	**94.36**	**93.08**	**93.97**

Table 5.3: Overview of the results obtained for the detection of both *Trichuris* spp. and *Ascaris* spp. eggs. with the baseline model

for each school, including sample collection date, number of total collected samples, the prevalence, and the epg of each species. The mean prevalence in this zone is 33.63%, where Mzizima has the lowest prevalence (2.91%) and Mwkiro has the highest prevalence (82.41%).

The performance analysis of the AI model on both the validation set and real-time scenarios is detailed in Table 5.5. Each model was rigorously evaluated on the validation set before deployment to ensure that its performance surpassed that of previous iterations. As shown in Table 5.5, there was a consistent improvement in model performance across iterations. Specifically, the average precision increased from 44.46% in Model 1 to 87.27% in Model 4, while recall improved from 64.24% to 84.72%, respectively.

Despite these gains on the validation set, the real-time performance exhibited considerable variability. For instance, Model 3 achieved a precision of 72.78% and a recall of 79.85%. In contrast, the average performance of Model 4 dropped to 68.91% precision and 56.26% recall. This discrepancy in real-time performance can be attributed to differences in the datasets used for evaluation, highlighting the challenge of maintaining consistent performance across varied real-world conditions.

Additionally, we assessed the correlation between egg counts obtained using conventional methods and those obtained with AdaptaSpot (AI-assisted), as illustrated in Figure 5.5. This figure shows the relationship between egg counts from both methods along with the correlation coefficient R.

Overall, the correlation for *Ascaris* and *Trichuris* spp. is notably high, as these eggs remain visible throughout the analysis, providing consistent results. In contrast, the correlation for hookworm eggs exhibits more variability. This fluctuation is attributed to disappearance of hookworm eggs during the analysis, which can affect the accuracy of the counts. Consequently, the order in which samples are analyzed can impact the results.

Usability feedback

After the pilot study concluded, we gathered feedback from healthcare workers regarding the usability of the AI algorithm. Overall, the healthcare workers expressed satisfaction with the AI's functionality. They observed noticeable improvements in performance following several updates during the study, although they noted that further refinements are still needed. The AI was particularly appreciated for its ability to alleviate eye strain and was seen as a valuable tool for diagnosing various conditions such as malaria, filariasis, schistosomiasis, and tuberculosis. However, the healthcare workers emphasized that the presence of an expert

School	date	Total subjects	positive	mean AS. epg	mean TR. epg	mean HW. epg
Kidimu	2021-11-23	92	51.09%	241	73	823
Mwangwei	2021-11-24	97	43.30%	83	463	211
Majoreni	2021-11-25	100	13.00%	24	104	350
Jomo Kenyatta	2021-11-26	102	44.12%	96	525	249
Kanana Genesis	2021-11-29	106	46.23%		398	348
Mkono wa Ndugu	2021-11-30	112	19.64%		112	420
Mwazaro	2021-12-01	105	25.71%		101	248
Kichaka Mkwaju	2021-12-02	102	37.25%	18493	1143	70
Tswaka	2021-12-03	104	11.54%	21164	552	319
Mzizima	2021-12-06	103	2.91%			600
Shimoni	2021-12-07	108	14.81%		81	36
Wasini	2021-12-08	104	45.19%	11760	1305	67
Mwkiro	2021-12-09	108	82.41%	21780	1186	271
Mean		1343	33.63%	9205.1	503.6	308.6

Table 5.4: STH prevalence in Kwale in 2021. Result obtained with conventional microscopy. AS: *Ascaris lumbricoides*, TR: *Trichuris trichura*, HW: hookworm

Model	Label	Validation set			Real time		
		Labels	Precision	Recall	Labels	Precision	Recall
Model 0	Ascaris				73	41.75	58.9
	Trichuris				163	82.65	49.69
	Hookworm				104	0	0
	mean					62.2	36.2
Model 1	Ascaris	45	65.38	81.93	2	3.23	50
	Trichuris	11	64.29	56.25	419	90.45	76.85
	Hookworm	7	23.53	54.55	143	79.78	49.65
	mean		44.46	64.24		57.82	58.83
Model 2	Ascaris	86	87.64	93.98	1	0	0
	Trichuris	36	54.55	80	140	92.31	68.57
	Hookworm	101	57.98	63.37	190	93.84	72.11
	mean		66.72	79.12		62.05	46.89
Model 3	Ascaris	69	76.92	96.77	2549	95.21	86.58
	Trichuris	36	63.22	85.94	352	61.85	81.53
	Hookworm	101	80.5	72.73	392	61.27	71.43
	mean		73.55	85.15		72.78	79.85
Model 4	Ascaris	179	94.04	89.87	1417	93.04	80.24
	Trichuris	211	88.52	89.01	600	94.76	66.33
	Hookworm	160	79.25	75.27	63	18.92	22.22
	mean		87.27	84.72		68.91	56.26

Table 5.5: Model performance on validation set and on real time during the pilot study. Note that the performance between models are not comparable, as they are evaluated on different dataset.

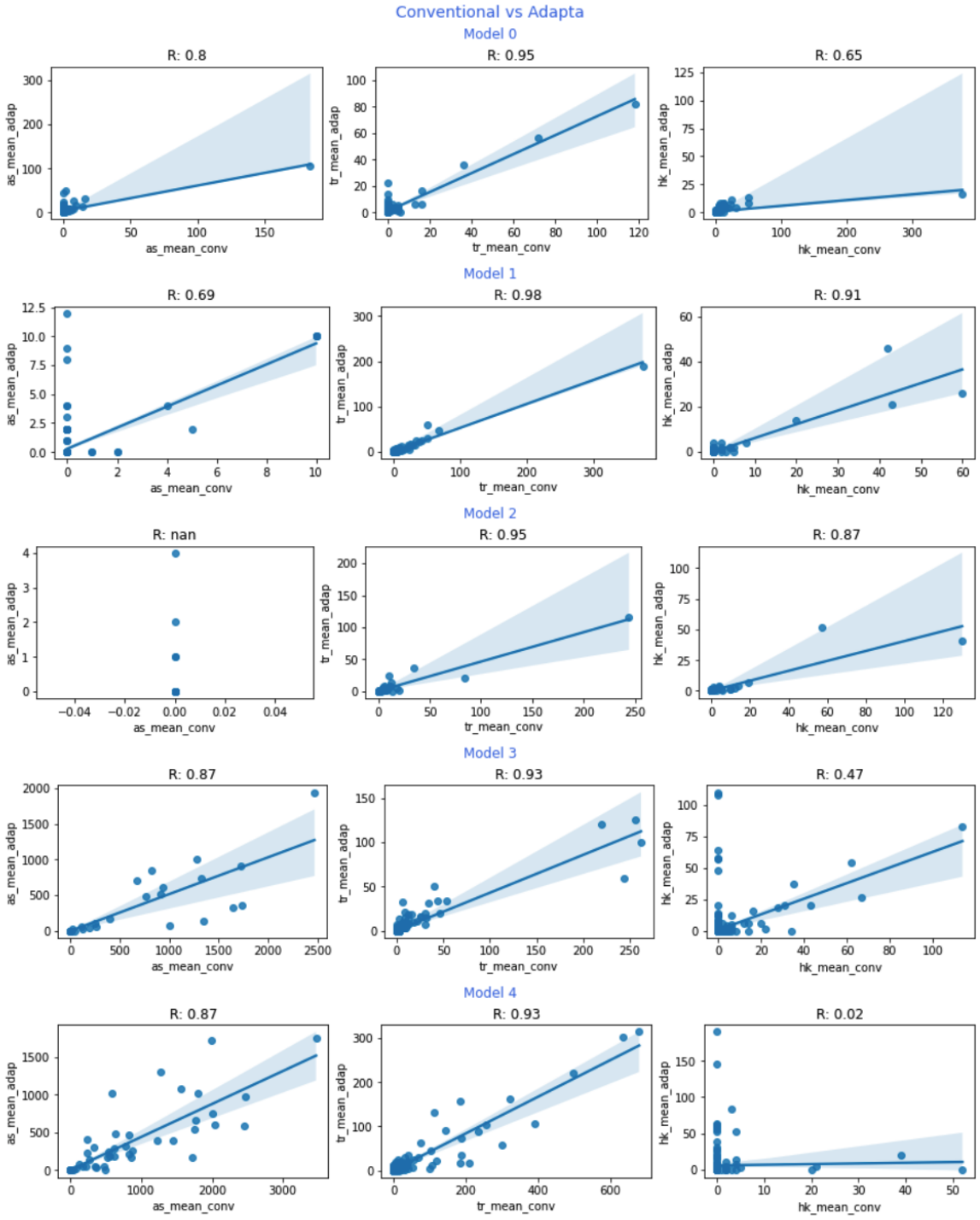


Figure 5.5: Correlation of egg counting with conventional microscopy and AI assistance, where R is the correlation coefficient.

remains crucial, as the AI algorithm occasionally produced false positives.

5.3.3 Post pilot study

With the integration of human in the loop labeling, we were able to significantly expand our dataset to include a total of 20,276 labeled instances. This expanded dataset comprised 13,775 instances of *Ascaris* spp., 4,697 instances of *Trichuris* spp., and 1,822 instances of hookworm.

Furthermore, the efficiency of the human-in-the-loop annotation system was evaluated by comparing the time required to label 200 objects from images of the same sample. Analysts using the human-in-the-loop system completed the task in approximately 9 minutes, compared to around 20 minutes without the system. This substantial reduction in labeling time highlights the significant efficiency gains achieved through the integration of this annotation approach.

Using the expanded dataset, we trained a new model, designated as Model 5, maintaining the same architecture and parameters as the previous iterations. The performance of Model 5 was first assessed on the validation set, which included 1,350 instances of *Ascaris* spp., 703 instances of *Trichuris* spp., and 331 instances of hookworm.

Model 5 demonstrated exceptional performance, achieving an average precision of 93.58%, recall of 91.74%, and an F1 score of 92.65%. Specifically, for *Ascaris* spp., the precision was 96.51% and the recall was 94.15%. For *Trichuris* spp., the precision was 89.68% with a recall of 92.75%, and for hookworm, the precision was 90.10% with a recall of 79.76%. These results indicate that Model 5 excels in detecting and classifying *Ascaris* spp., *Trichuris* spp., and hookworm, demonstrating high precision and recall values.

Following this, Model 5 was evaluated on the test set to assess the generalizability, which included 884 instances of *Ascaris* spp., 813 instances of *Trichuris* spp., and 215 instances of hookworm. On this test set, Model 5 achieved an average precision of 89.39%, a recall of 84.21%, and an F1 score of 86.71%.

To compare the performance, Models 0 through 4 were also evaluated on the same test set, with results summarized in Table 5.7. Consistent with the performance observed in real-time evaluations, the performance of the models varied significantly from Model 1 to Model 4. There was a notable improvement from Model 0 to Model 1, as Model 1 was trained with data from the study, including hookworm samples. However, despite being trained with more data, Model 4 did not achieve the best performance, potentially due to the similarity between the test data and the training data of Model 2. As expected, Model 5 outperformed all previous models, with the F1 score increasing dramatically from 35.33% in Model 0 to 86.72% in Model 5. This underscores the effectiveness of the new dataset and the improvements made in subsequent iterations.

5.4 Conclusions

In this chapter, we presented the entire workflow to create AI algorithm for soil-transmitted helminths detection in smartphone and we introduced it in a clinical workflow. The study

Ground truth		Artefact	*Ascaris*	*Trichuris*	*Hookworm*
	Artefact	0	53	55	24
	Ascaris	141	735	0	8
	Trichuris	77	0	735	1
	Hookworm	29	9	7	170
		Prediction			

Table 5.6: Confusion matrix of the model 5 on the unique test set

Model	Class	Precision	Recall	F1-score
Model 0	Ascaris	50.34	41.63	45.57
	Trichuris	93.93	28.54	43.78
	Hookworm		0	
	mean	72.13	23.39	35.33
Model 1	Ascaris	65.71	75.23	70.15
	Trichuris	88.5	59.66	71.27
	Hookworm	72.5	40.47	51.94
	mean	75.57	58.45	65.92
Model 2	Ascaris	85.9	59.95	70.62
	Trichuris	91.73	69.62	79.16
	Hookworm	65.36	54.42	59.39
	mean	81	61.33	69.72
Model 3	Ascaris	74.39	54.86	63.15
	Trichuris	90.87	74.66	81.97
	Hookworm	52.77	57.67	55.11
	mean	72.68	62.4	67.15
Model 4	Ascaris	92.07	57.81	71.02
	Trichuris	93.84	46.86	62.51
	Hookworm	44.72	82.79	58.07
	mean	76.88	62.49	68.94
Model 5	Ascaris	92.22	83.14	87.44
	Trichuris	92.22	90.41	91.31
	Hookworm	83.74	79.07	81.34
	mean	**89.39**	**84.21**	**86.72**

Table 5.7: Performance of the model on the unique test set

includes 3 phases: proof of concept to assess the viability to train and deploy AI model on the mobile phone; a pilot study with the aim to collect a large dataset, specially for hookworm, a specie that disappears within 30 minutes after the sample preparation, and test the usability of the AI algorithm on the real world setting; a post-pilot study that incorporated human in the loop labeling to fasten the image labeling process and to improve the model performance.

Microscopy continues to be the predominant method for diagnosing many NTDs, including STH. Nevertheless, this technique typically requires expert personnel, whose availability may be limited, and is time consuming. Given these constraints, there is a pressing need for digital microscopy and AI algorithms that enable remote and automated diagnosis of STH.

Several studies have proposed the use of AI for the diagnosis of STH, yet these approaches often require specialized hardware, such as scanners (Holmström et al., 2017; Lundin et al., 2024). Other studies have developed large models that demand significant computational resources for inference, rendering them impractical for point-of-care use (Li et al., 2020; P. Ward et al., 2022). While these methods have demonstrated promising results, their dependence on advanced hardware and high computational power limits their applicability in resource-constrained settings.

Compared to previous works, our approach offers several key advantages. It does not require dedicated hardware, relying only on a 3D-printed adapter and smartphones. Our system integrates seamlessly into the existing laboratory workflow, utilizing a conventional light microscope. Additionally, the AI algorithm operates in real-time and does not require an internet connection, making it highly practical for use in low-resource settings.

During this study, we collected a substantial dataset comprising 1,343 samples and 19,889 images, creating a comprehensive STH database. This dataset is valuable not only for AI training but also for educational purposes. Our pilot study demonstrated the feasibility of digitizing the STH analysis workflow with AI assistance at the point of care, showing that such an approach can be effectively integrated into real-world settings.

Several models were released and used by technicians throughout the study. When comparing AI-assisted analysis on the mobile phone with conventional microscopy, the main difference observed was in the quantification of hookworm. This discrepancy is primarily due to the fact that hookworms tend to deteriorate or disappear over time. Consequently, if a sample is first analyzed using conventional microscopy and then later with the AI-assisted digital method, the latter may detect fewer hookworms or none at all, as they may have already degraded or vanished.

Even though we presented the correlation between AI-assisted analysis and conventional microscopy, it is important to note that the performance of each model should not be directly compared, as they were evaluated on different datasets. To provide a fair comparison, we evaluated all models on the same test set. From model 0 to model 5, precision increased from 72.13% to 89.39%, and recall saw a sharp increase from 23.39% to 84.21%.

Model 5, which was trained with 15,980 labeled instances (11,541 *Ascaris*, 3,163 *Trichuris*, and 1,276 hookworms), achieved a precision of 92.22% for *Ascaris* spp., 92.22% for *Trichuris* spp., and 83.74% for hookworm, with corresponding recall values of 83.14%, 90.41%, and 84.21%.

These results demonstrate a significant improvement in the model's ability to accurately detect and classify STH eggs.

Despite the promising results of this study, several limitations should be acknowledged. Firstly, the generalizability of the AI model is a concern. The dataset was collected from a specific region in Kenya, and the model's performance might vary when applied to samples from different regions where STH prevalence, egg morphology, or sample preparation techniques could differ. Additionally, the quantification of hookworms presented challenges due to their tendency to deteriorate or disappear over time. This temporal factor can lead to discrepancies, particularly when comparing AI-assisted digital analysis with conventional microscopy, as the timing of the analysis can significantly affect the detection rates.

In terms of real-time application, although the model was designed to operate on a smartphone, practical deployment in field conditions may encounter challenges such as variability in lighting, smartphone camera quality, and user expertise, all of which could affect the quality of the digitized images and the model's performance. Additionally, the dataset contained fewer instances of hookworm compared to *Ascaris* and *Trichuris*, which may have impacted the model's ability to generalize well for this class. The relatively lower recall for hookworm detection suggests that more diverse data and further refinement are needed to enhance the model's performance for this species.

Lastly, the study primarily focused on short-term deployment and validation, and there is a need for longitudinal studies to assess the AI model's performance over time, especially under varying field conditions and across different population groups. Addressing these limitations in future work will be essential for improving the robustness and scalability of the AI model for broader applications in STH diagnosis.

Building on the findings and limitations of this study, several directions for future work are identified. Firstly, expanding the dataset to include samples from diverse geographic regions will be crucial to improve the generalizability of the AI model. This will involve collaboration with health centers in different countries, ensuring that the model can effectively handle variations in egg morphology and sample preparation techniques. Additionally, enhancing the model's ability to detect and quantify hookworm eggs, is another critical area for improvement.

Furthermore, refining the human-in-the-loop annotation system to further reduce reliance on expert input could enhance the system's scalability in resource-limited settings. This might include incorporating semi-supervised learning or self supervised learning techniques, where the model can continue to improve with minimal human intervention.

Finally, integrating the AI system with existing public health infrastructure and electronic medical records could facilitate better tracking of STH infections and improve overall disease management strategies.

This study demonstrates the potential of integrating AI into the diagnostic workflow for STH at the point of care. By leveraging a mobile phone-based system, we successfully digitized the microscopy process and utilized AI to assist in the detection and classification of STH eggs. The system is easily scalable and updatable, making it a practical solution for widespread deployment in resource-limited settings. Our approach can significantly contribute to the

diagnosis of STH by reducing the dependency on highly specialized microscopists. Moreover, this system has the potential to be expanded to other NTDs. By collecting more samples, a universal parasite detection system can be developed, aligning with the goals established in the WHO 2030 roadmap (WHO, 2020).

Chapter 6

Edge Artificial Intelligence (AI) for real-time automatic quantification of filariasis in mobile microscopy

6.1 Introduction

Filariasis is a tropical infectious disease caused by roundworms (Phylum Nematoda). There are at least eight filarial worms that are hosted in humans. These are the causative agents of four types of diseases: lymphatic filariasis, which is caused by *Wuchereria bancrofti*, *Brugia malayi*, and *Brugia timori*; Onchocerciasis, caused by *Onchocerca volvulus*; loiasis, caused by *Loa loa*; and mansonellosis, caused by *Mansonella perstans*, *Mansonella ozzardi*, and *Mansonella streptocerca*. Among these, lymphatic filariasis and onchocerciasis have significant clinical and public health implications and are included in the World Health Organization (WHO) list of Neglected Tropical Diseases, while loiasis and mansonellosis have historically received much less attention (Lima et al., 2016; Metzger & Mordmüller, 2014; Raccurt et al., 2014).

In 2000, the WHO launched the Global Programme for the Elimination of Lymphatic Filariasis (GPELF), which set the goal of eliminating lymphatic filariasis as a public health problem in 58 countries by 2030 (WHO, 2020). The program achieved a considerable reduction, but there are still 863 million people in 50 countries who require preventive chemotherapy (PC) (Cromwell et al., 2020). Similarly, onchocerciasis affects over 20.9 million people, with at least 220 million in need of PC (James et al., 2018). However, *L. loa* infection is hindering the elimination of lymphatic filariasis and onchocerciasis, as these diseases use ivermectin in massive drug administration (MDA), but ivermectin causes severe adverse effects when the individual has elevated levels of *L. loa* in the blood (Beng et al., 2020; Gardon et al., 1997).

Studies have reported that *M. perstans* is the most prevalent filariasis in Africa, with more than 100 million people estimated to be infected and 600 million living in 33 high-risk countries (Simonsen et al., 2011), yet it is one of the most neglected filariasis (Lima et al., 2016; Raccurt et al., 2014; Ta-Tang et al., 2018), and there are no control programs for it.

The correct diagnosis and appropriate treatment are paramount for the effective control and elimination of parasites and their approach depends on the filarias type. In addition to the ongoing elimination programme for lymphatic filariasis and onchocerciasis, there have been increasing calls for the treatment and control programme for mansonellosis and loasis in recent years (Ferreira et al., 2023; Jacobsen et al., 2022; Ta-Tang et al., 2021). WHO recommends utilizing the Alere Filariasis Test Strip (FTS) for all areas endemic for *W. bancrofti* and Brugia Rapid Test for all areas endemic for *Brugia* spp. However, these tests are species-specific and do not account for co-infections (WHO et al., 2021b). Molecular diagnosis methods have also been applied in surveillance studies with good results but without possibilities to perform on site (Moya et al., 2016). Microscopy remains the most widely used technique for all filarial species, enabling the detection of microfilariae through blood smears or skin snips. The routine examination is the screening at low magnification (10x magnification) and then uses higher magnification to identify the species (e.g., 40x). The sample should be scanned completely at 10x magnification to report the sample as negative (Mathison et al., 2019). Nonetheless, the diagnosis by microscopy is time-consuming and requires experienced microscopists, whose availability is not always assured (Petti et al., 2006; WHO, 2016). In that sense, different studies revealed the importance of mobile health (mHealth) to bring diagnostics to the point of care and scale access in low and middle-income countries (LMICs) (Feroz et al., 2020; McCool et al., 2022; Saeed & Jabbar, 2018). Notably, several investigations reported the use of mobile microscopy for parasite detection, such as LoaScope, which is a point-of-care microscope that detects *L. loa* microfilaria in blood smears automatically in video (D'Ambrosio et al., 2015; Pion et al., 2020) or SchistoScope, a mobile phone microscope for the screening of *Schistosoma haematobium* (Armstrong et al., 2022).

A possible tool to address the lack of trained specialists is the detection of parasites in microscopic images using Artificial Intelligence (AI). AI is revolutionizing the medical field and can be applied in different medical subfields (Cai et al., 2020; Topol, 2019). The development of AI algorithms for microscopy depends on the digitization samples, which can be done using digital microscopes that have embedded cameras, converting a conventional optical microscope to a digital microscope using mobile phones or other image acquisition modules.

In a recent review by Fan et al. focusing on AI applications for peripheral blood films, 95 studies addressed malaria, 81 leukaemia, 72 leukocytes, 25 mixed cell types, 15 erythrocytes, and 1 Myelodysplastic syndrome. Beyond the scope of peripheral blood films, limited attention was given to babesiosis, leishmania, trypanosomiasis, etc.. However, no work specifically addressing filariasis was identified in this review (Fan et al., 2023). Beside that, our research found numerous studies reporting the detection of parasites in microscopical images, revealing the potential of AI in this task. Quinn et al. created one of the first deep learning algorithms that consists of a four layer convolutional neural network (CNN) for malaria image classification from scratch. For that, they used a 3D printed adapter that aligns the mobile phone camera to the microscope eyepiece (Quinn et al., 2016). Davidson et al. presented a 3 phases analysis to detect and count malaria parasites and its life cycle stage. The first phase detects red blood cells using the Faster RCNN object detection algorithm, then crops the detected cell and feeds it to a ResNet50 to classify if the detected cell is infected, and finally classify the life cycle stage of the infected cell using ResNet-34. They achieved 98.5% average precision

in detecting RBCs and 99.8% in classifying the detected cells into infected or uninfected, and mean square error of 0.23 in the stage classification. Images in this study were acquired by manually aligning the mobile phone camera to the microscope eyepiece (Davidson et al., 2021). Similarly, Holmström et al. presented a deep learning algorithm for the detection of soil-transmitted helminths (STH) and *Schistosoma haematobium* with a custom microscope scanner and the commercially available image analysis software platform WebMicroscope (Holmström et al., 2017). Dacal et al. presented an object detection algorithm using Single-Shot multibox Detection (SSD) for STH that runs on a smartphone (Dacal et al., 2021). Oyibo et al. presented an automated microscope with an image segmentation algorithm using a u-net architecture for *Schistosoma haematobium* (Oyibo et al., 2022). Dedhiya et al. introduced the first study that uses machine learning over thermal imaging to predict the viability of onchocerca worms. In this work, they used five separated random forest classifiers and the final classification was obtained using a voting mechanism (Dedhiya et al., 2022). D'Ambrosio et al. presented an algorithm which detects *L. loa* microfilaria in video by subtracting subsequent frames of the video, then generating a single difference image and using a local peak-finding routine to find microfilarias. They correlated the automated counts with manual counts and achieved 94% specificity and 100% sensitivity (D'Ambrosio et al., 2015). Elvana et al. presented a lymphatic filariasis detection system using CNN, achieving an accuracy of 70% (Elvana & Suryanto, 2022). As far as we know, there have been very few attempts to deploy edge-AI systems using deep learning which are able to support in real-time and without connectivity the analysis of optical microscopy images for filaria detection with species differentiation and more broadly NTDs diagnostics.

The objective of this study is to propose, develop, and pilot a system for real-time, automatic detection and quantification of filariasis using an edge AI model. The proposed system aims to assist the screening and species differentiation of four worm species (*L. loa*, *M. perstans*, *W. bancrofti* and *B. malayi*) in blood smears for filariasis. For that, we proposed a pipeline with the following modules: the digitization of smear samples with smartphones coupled to a microscope through a 3D-printed device; sample analysis and data labeling in a telemedicine platform for training of an AI algorithm; integration of the trained algorithm on the smartphone to assist the diagnosis and validation of the model in a clinical environment.

This study addresses objectives 1, 3, and 4 of this PhD thesis. Following the workflow established in Chapter 5, this chapter develops and pilots a real-time, automated detection and quantification system for filariasis using an edge AI model. A key aspect of this work is the establishment of a clinical validation protocol, along with prospective validation in a clinical setting, ensuring the system's effectiveness and reliability for diagnosing filariasis in real-world applications.

6.2 Material and methods

6.2.1 Ethics statement

Ethical approval was obtained from the Research Ethics Committee (REC) Instituto de Salud Carlos III, Spain (CEI PI 74 2020).

6.2.2 Overview of the methodology

The study was conducted in two distinct phases. The initial phase involved digitizing blood smear samples to construct the database for the development of AI algorithms with 115 samples. In the subsequent phase, the AI model was integrated on the smartphone and a pilot study was conducted to evaluate the AI's performance on real world settings with a new dataset of 18 samples. The study design schema is presented in Figure 6.1.

All preparations included in the study were appropriately stained, positive and with well-preserved parasite morphology. Samples with varying levels of parasitemia and species were chosen based on results obtained from polymerase chain reaction (PCR) and/or conventional microscopy, ensuring the collection of both positive and negative fields of view. Additionally, the staining type and sample preparation details were systematically compiled.

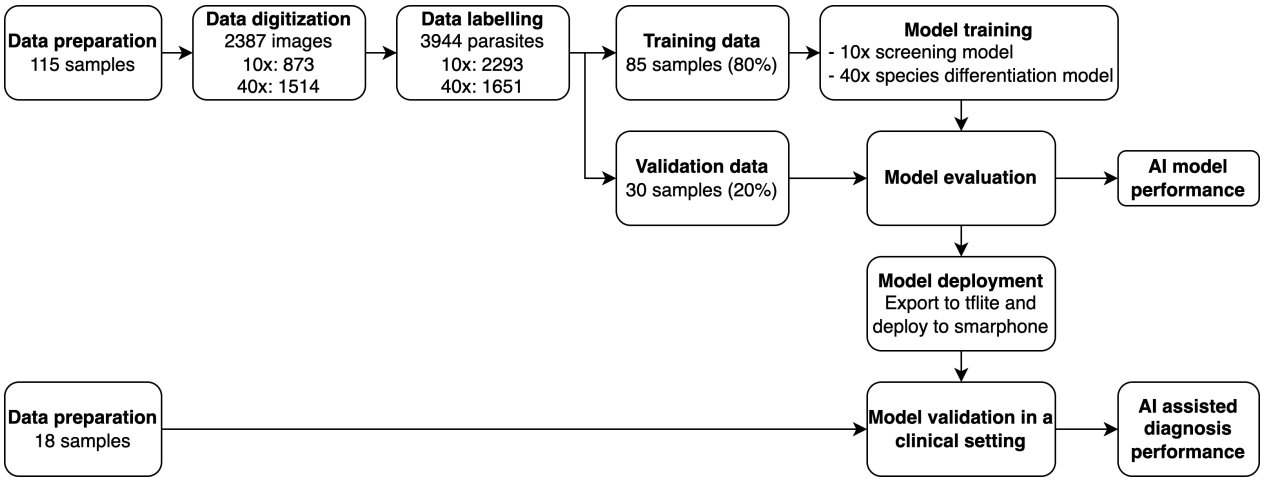


Figure 6.1: Schematic representation of the study design.

6.2.3 Creating a Filariasis differentiation AI model

Digitalize samples

In the initial phase, a total of 115 sample smears from 115 different subjects were collected from the sample collection of the Malaria and Emerging Protozoa Unit of the Instituto de Salud Carlos III (Spain). In addition, all preparations have been previously anonymized without the possibility of reverse coding. 112 of them were stained with Giemsa and 3 of them with Panopticon. Case distributions were presented in Table 6.1.

Images were digitized simulating the real diagnostic workflow, with a system previously described in Dacal et al. (Dacal et al., 2021). Briefly, this system uses a 3D printed device that allows coupling a mobile phone with a conventional optical microscope by aligning the smartphone camera with the objective of the microscope to acquire images, and that converts any conventional microscope into a digital microscope. Following the conventional workflow, the analyst scanned the samples at 10x magnification and captured photos of fields containing structures compatible with filarial parasites. Subsequently, the objective was switched to 40x

magnification, and photos of each detected parasite were taken. Slides were digitized using four different smartphone models (Huawei Ascend G7; (n=95 cases), Redmi Note7 (n=13 cases), Samsung Galaxy A32; (n= 5 cases), LG X Power K220; (n=1 case), Huawei Nova 5T (n=1 case)). In total, 873 FoV (images) of 10x and 1514 FoV (images) of 40x were captured.

To evaluate the AI model's capacity to generalize and address the issue of overfitting while ensuring accurate performance report, a case-level split was employed. This approach ensures that all images from the same case belong to the same dataset, whether used for training the AI model or validating its performance. The split is carried out after labelling the images to guarantee that all species are presented in both the training and validation sets. The cases are distributed randomly, striving to achieve an 80%-20% split between the two sets.

Parasite	**Number of cases**	**Training set**	**Validation set**
L. loa	19	11	8
M. perstans	43	37	6
W. bancrofti	9	5	4
B. malayi	6	4	2
L. loa + *M. perstans*	20	18	2
Negative	18	10	8
Total	115	85	30

Table 6.1: Cases included in the first phase and the training-validation split.

Labeling data

All acquired images were transferred from the smartphone to a telemedicine platform via mobile network, so that the images are stored and presented in an easy-to-use dashboard that allows their visualization, management, and labeling (Figure 6.2). In this web platform, standard clinical and analysis protocols were translated into digital tasks that were adapted to the clinical case and disease under study.

The annotation protocol was based on the placement of bounding boxes around the identified parasites. All visible parasites in the image were labeled by two analysts and reviewed by an expert. At 10x magnification, as the species can't be identified, all detected parasites belong to the microfilariae class. A total of 2293 parasites were located from 873 images. At 40x magnification, the parasite species were annotated with their corresponding class. A total of 1651 parasites were tagged from 1514 images. In addition, some artifacts that have a similar appearance to the parasite were labeled, which serves as a hard negative for the algorithm training.

The labeled data was divided into a training set for model development and a validation set for selecting the best model, as shown in Table 6.2. The training set for 10x images consists of 1965 microfilarias from 700 images, while the validation set for 10x images contains 328 microfilarias from 173 images (FoVs). In the training set for 40x images, there are 906 *L. loa*,

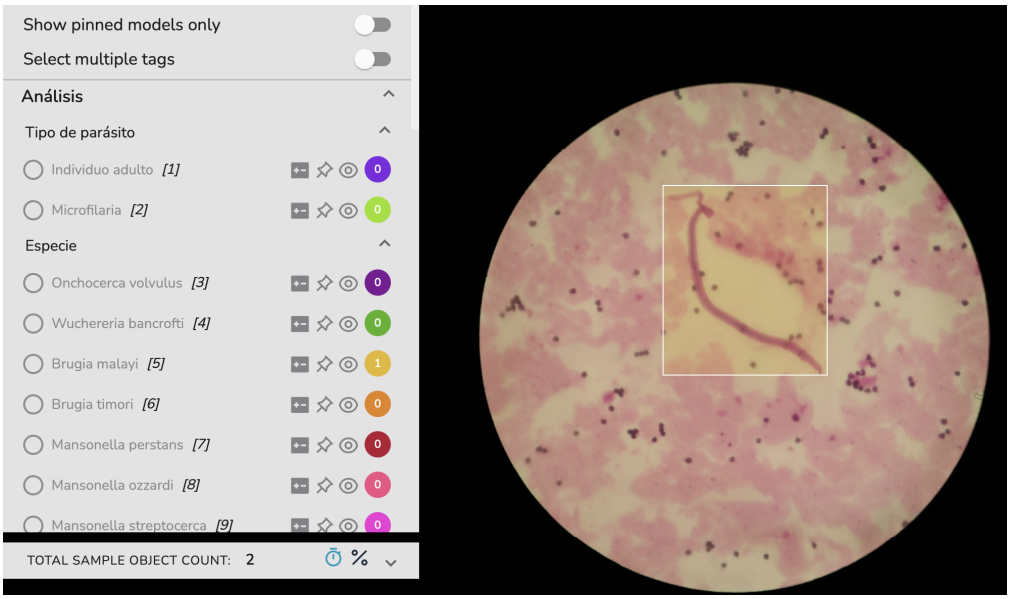


Figure 6.2: The telemedicine platform facilitates image visualization, management, and labeling. When an AI algorithm is deployed, analysts have the option to review the predictions rather than starting the labeling process from scratch.

Labels	total	training	validation
Microfilaria (10x)	2293	1965	328
L. loa (40x)	1044	906	138
M. perstans (40x)	480	378	102
W. bancrofti (40x)	64	35	29
B. malayi (40x)	63	58	5

Table 6.2: Label distribution of microfilaria species in the training and validation sets.

378 *M. perstans*, 35 *W. bancrofti*, and 58 *B. malayi* parasites from 1203 images, while the validation set includes 138 *L. loa*, 102 *M. perstans*, 29 *W. bancrofti*, and 5 *B. malayi* parasites from 311 images belonging to 30 cases.

Creating the AI model

A requirement for our AI model is that it can work offline or in limited bandwidth settings. To fulfill this requirement, we selected a lightweight model that can be run on a smartphone in real-time without internet connection. Given the multifaceted nature of the task, encompassing object localization, classification and counting, an object detection algorithm would be an appropriate solution. Specifically, we employed the Single-Shot Detection (SSD) MobileNet V2 detection model with a feature pyramid network as feature extractor, shared box prediction and focal loss (Lin, Dollár, et al., 2017; Liu et al., 2016; Sandler et al., 2018).

The SSD is a real-time object detection and localization algorithm comprising two fundamental

components: feature map extraction and the application of convolutional filters to detect objects. A simplified representation of SSD is illustrated in Figure 6.3. The feature extraction process leverages MobileNet V2, which encompasses a total of 52 convolutional layers. MobileNet V2 is structured around 16 bottleneck residual blocks, with each block having two 2D convolutional layers and one depthwise convolution layer. The output of MobileNet V2 then undergoes refinement through five additional feature blocks. These supplementary layers are designed to combine features from earlier layers, characterized by a low level of semantic information but a high spatial resolution, with later layers that possess high semantic information but a reduced spatial resolution. This fusion of features is facilitated through lateral connections, ultimately enhancing object detection accuracy. Finally, the processed data is directed through the convolutional box predictor to generate both bounding box predictions and class predictions.

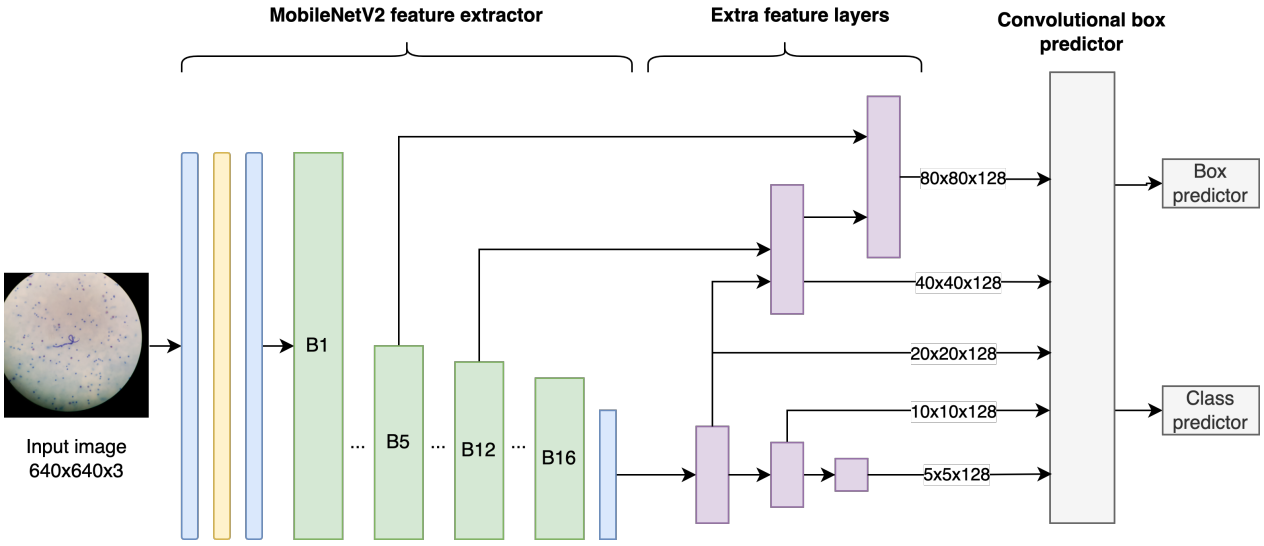


Figure 6.3: Simplified SSD MobileNet v2 detection architecture. MobileNet V2 initiates with one convolutional layer, succeeded by a depthwise convolutional layer and another convolutional layer. It is subsequently followed by 16 bottleneck residual blocks (green), each comprising two 2D convolutional layers and one depthwise convolutional layer, concluding with an additional convolutional layer. The SSD feature extractor is further enhanced by integrating five additional feature extractor blocks (purple). The resultant features are passed through a convolutional box predictor block (gray), which is responsible for predicting both the location and class of each detection.

In the context of object detection, CNN often generates thousands of candidate regions. However, only a few regions actually contain objects of interest, while the majority represent background elements. This class imbalance presents a significant challenge, as it can lead to training inefficiency. Notably, easy negatives, which correspond to background regions, constitute a substantial proportion of the total candidate regions, potentially overwhelming the loss function used during training. To mitigate this class imbalance issue, focal loss emerges as an enhanced alternative to the conventional cross-entropy loss. Focal loss addresses

class imbalance by assigning higher weights to hard-to-classify examples and down-weighting easier examples. This strategic adjustment helps focus the learning process on challenging cases, thereby improving the efficiency and effectiveness of object detection models.

Tensorflow object detection application programming interface (API) was used for model training because tensorflow has natively optimized the model to be executed in mobile phones and edge devices, whose code is publicly available on github (Abadi et al., 2016; tensorflow, n.d.). Given the relatively small size of our dataset, we used a pre-trained model that was trained with the COCO image database (Lin et al., 2014) and fine-tuned for this use case. The models were trained on Amazon Sagemaker, using the NVIDIA T4 GPU with 16 GB memory.

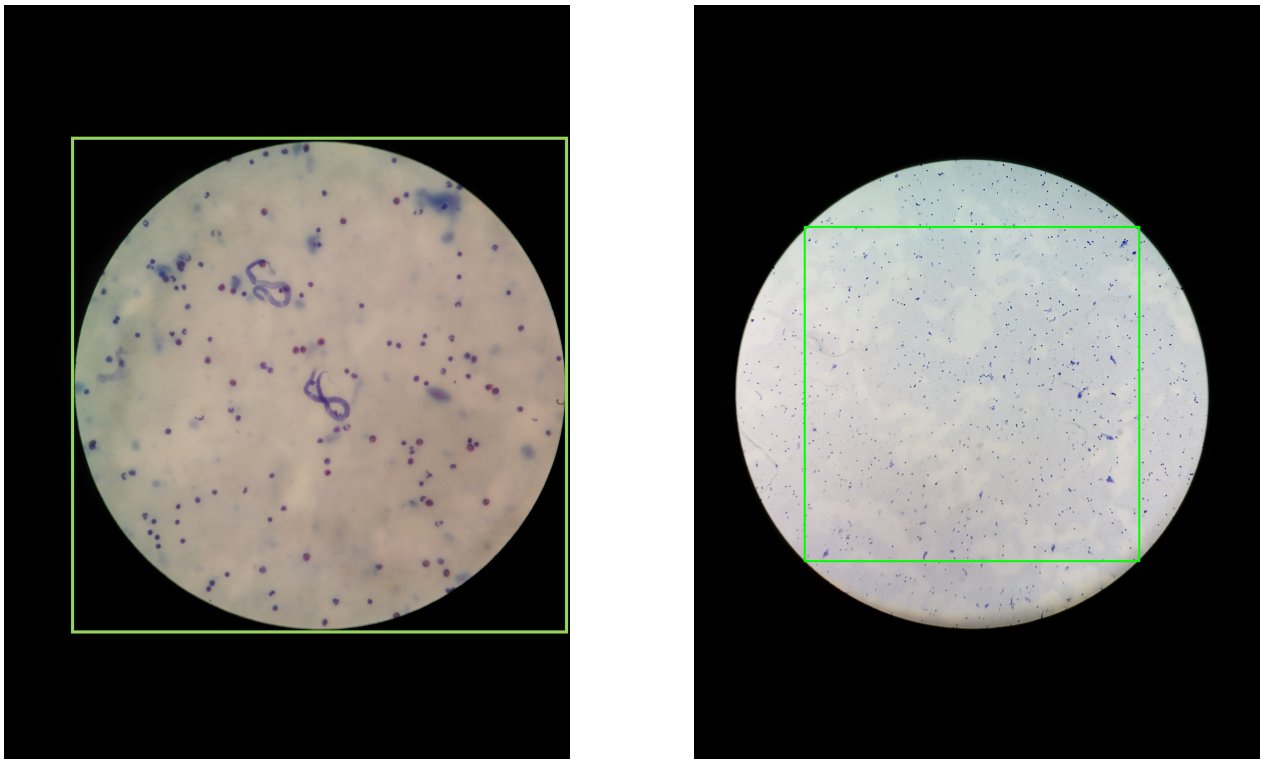
Two distinct algorithms were developed. The first algorithm, designed for screening at 10x magnification, focuses solely on detecting the presence of microfilariae. The second algorithm, developed for microfilaria species differentiation at 40x magnification, aims to classify the detected microfilariae into four species: *L. loa*, *M. perstans*, *W. bancrofti*, and *B. malayi*.

Given the alignment of the smartphone with the microscope eyepiece, the area visualized by the mobile phone is limited to a circular area, as depicted in Figure 2. In order to exclude non-informative regions (e.g., black areas), and to present other relevant information on the mobile phone screen (e.g., label count, AI activation, etc), we decided to use square images instead of rectangular images.

For the species differentiation algorithm that works with 40x magnification, we initially identify the circular region within the image and extract a square image encompassing the entire field of view, as illustrated on Figure 6.4a left. Subsequently, the cropped region was resized to 640x640 pixels. The reviewed data was split into two sets at case level as described above.

As Table 6.2 reflects, this dataset contains 1044 *L. loa*, 480 *M. perstans*, but only 64 *W. bancrofti* and 63 *B. malayi* in total, to address the imbalance nature of the dataset, oversampling of the minority classes (W. bancrofti and *B. malay*, and some *L. Loa*) was employed by generating mosaic images, which consists of cropping a 320x320 pixel patches that contains at least one parasite, and blending 4 images to create a new image of 640x640 pixels (see Figure 6.5). After augmentation, the training set contains 1116 *L. loa,* 378 *M. perstans*, 480 *W. bancrofti* and 533 *B. malayi*. Additional image augmentation, including random horizontal and vertical flip, 90 degree rotation with 50% of probability, random crop ensuring that the cropped image has a minimum area of the 80% of the original image, random brightness adjustment with a maximum range of 30%, random hue adjustment with a maximum range of 10%, and random saturation adjustment with a saturation factor between 0.8 and 1.25, was applied during training to enhance the model's robustness. The model was trained with a batch size of 2. Training employed the momentum optimizer, initialized with a learning rate of 0.01. To enhance training dynamics, a cosine decay strategy was employed, spanning a total of 50,000 training steps. This training required approximately 5 hours to complete.

In the screening algorithm that works with 10x magnification, a distinct strategy was implemented for image cropping in comparison to the species differentiation algorithm. Given the relatively small size of the parasite in 10x magnification, its visualization and detection



(a) **(b)**

Figure 6.4: a): example input of the species differentiation algorithm. b): green rectangle represents the sliding windows size

pose challenges for both human analysts and AI systems, necessitating the use of zoom. To optimize the visibility of the parasite and maximize the size of the image, we decided to crop the original image to the square inscribed within the circle as depicted in Figure 6.4b right. As it can be appreciated, a single crop of the inner square leaves some valuable information out. To overcome this limitation, we employed the sliding window technique, where 4 patches were generated for each image, ensuring that all the information within the field of view is represented. Then, patches were resized to 640x640 to fit the requirements set by the network used. The same data augmentation was applied as in the species differentiation algorithm. After data augmentation, the number of microfilarias in the training set increased from 328 to 10847, whereas the validation set was unmodified. The model was trained with a batch size of 2. Training employed the momentum optimizer, initialized with a learning rate of 0.01. To enhance training dynamics, a cosine decay strategy was employed, spanning a total of 20,000 training steps. This training required approximately 2 hours to complete.

6.2.4 Validation of the AI model in a lab setting

To assess the usability and performance of the proposed system within the clinical workflow, a lab validation was conducted. For that, we first deployed the model on the smartphone, and then piloted the AI assisted diagnosis workflow.

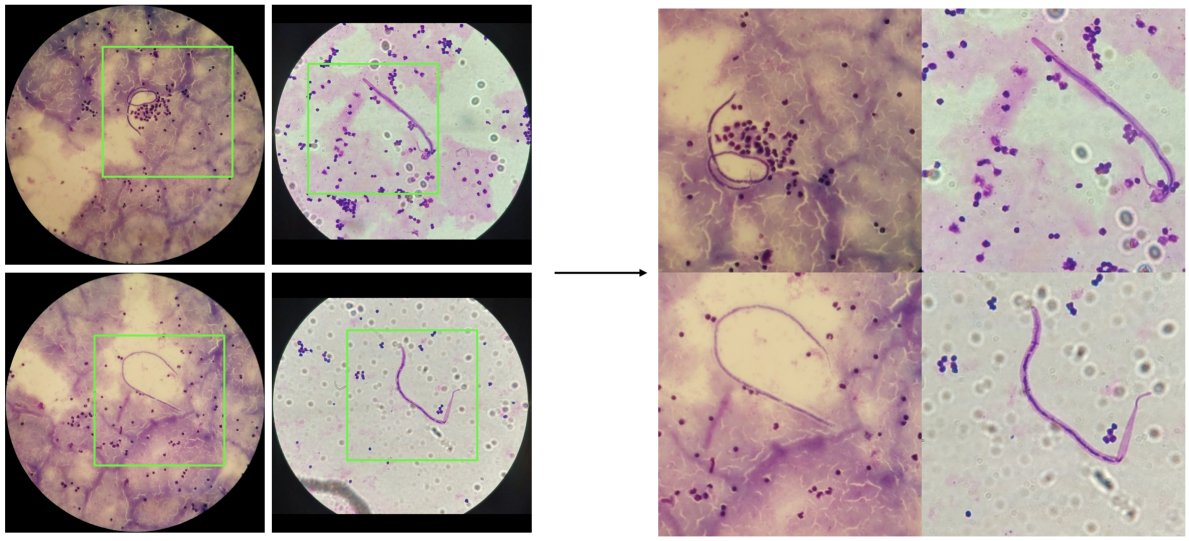


Figure 6.5: Mosaic augmentation. The original image is 640x640 pixels, for each image, we randomly select an area of 320x320 pixels that contains at least one parasite (green rectangle), using 4 cropped areas we compose a new mosaic image.

Deployment and integration of technology

The AI mobile model was optimized using post-training quantization, which is a conversion technique that reduces model size while also improving CPU and hardware accelerator latency, with little degradation on model accuracy. The model is exported as a tflite format, with a size of approximately 13 megabytes, which can be run on a middle range smartphone in real time. The execution time on BQ Aquaris X2 on CPU is 1400 milliseconds for a single image, while the Samsung S9, utilizing the GPU, accomplished the task in 610 milliseconds.

To facilitate the process of digitization and AI-assisted analysis, a customized Android application was developed. The application is not specifically for this research, but a proprietary platform that can be downloaded in Google play store. This application records both clinical data and images. While the user visualizes the image on the mobile phone screen, the selected AI algorithm, screening or species differentiation is running in real time depending on the magnification used (10x or 40x), generating predictions for the corresponding frame, and outlining the detected parasites within bounding boxes. When the user takes a photo, both the images and the prediction are saved, and the parasite detection counter is incremented no matter if the prediction is correct or not. In the case that the user finds parasites not detected by the AI algorithm, they can tap on the button of the corresponding label to increase the count of this parasite. Once the analysis is finished, this information is uploaded to the telemedicine platform, allowing users to review and correct the prediction and share information. Figure 6.6 explains how the smartphone is attached to the conventional microscope and the screening and species differentiation algorithm running on the smartphone. With AI running in real time, the analyst moves the sample and analyzes it with AI assistance (Spotlab, n.d.-a, n.d.-b).

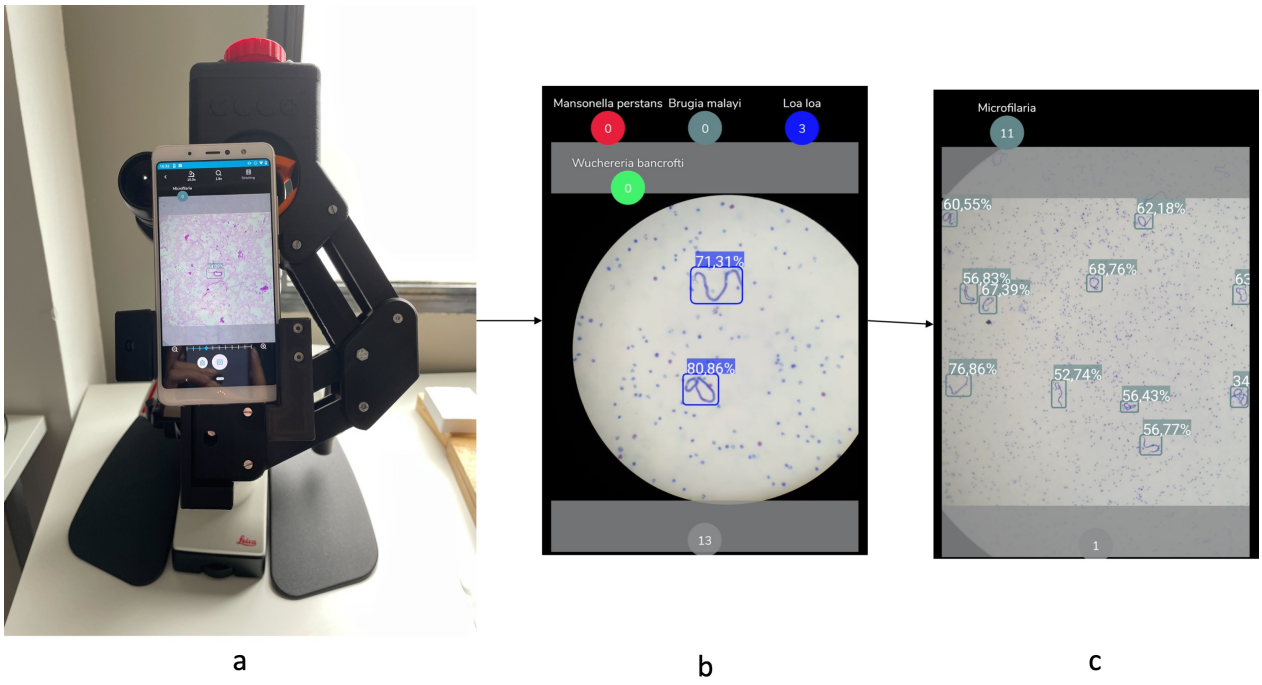


Figure 6.6: (a) smartphone attached to the conventional microscope with a 3D printed adapter. (b) screening algorithm working on the smartphone. (c) species differentiation working on the smartphone.

Experiment- Pilot replicating diagnostic workflow

To assess the performance of the AI models, a real-time pilot study was conducted to evaluate the effectiveness of the edge AI system in assisting parasite detection through the mobile application.

To pilot the proposed system, the algorithm developed in the first phase was integrated on the mobile phone and the telemedicine platform to be validated. For all selected samples (N=18), Figure 6.7 represents the ideal workflow: with the AI algorithm operating in real time, the analyst examines the complete sample using a 10x magnification objective. Depending on whether the algorithm detects a parasite, different actions are taken. When a parasite is identified by the AI algorithm, the analyst captures a photo and the detected parasites are automatically counted by the app, and switches to a 40x magnification objective. At this point, the species differentiation algorithm is activated to discern the specific species of the detected parasites, a photo is taken to count the parasite. In cases where parasites are present on the screen but not detected by the screening AI, the analyst manually adds the count by tapping on the corresponding label. Since the mobile application did not allow modification of the incorrect prediction, both images and the mobile prediction were uploaded to the telemedicine platform for further correction and validation. The results were independently reviewed by two analysts: analyst A, a junior researcher in parasitology, who analyzed images on real time using the mobile application; and analyst B, an expert in microscopy of infectious diseases and who only reviewed the digitized image on the telemedicine platform.

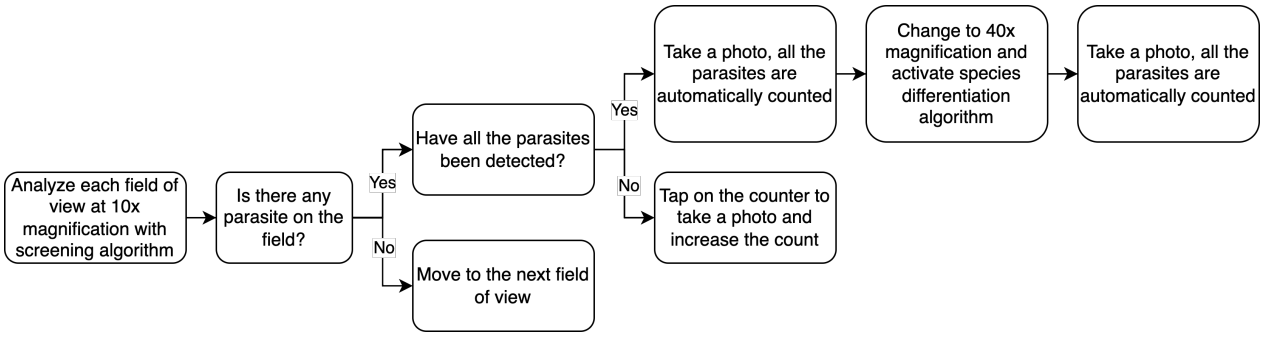


Figure 6.7: Schema represents the validation workflow of AI assisted filariae detection. At least 3 images of negative fields were acquired for each sample.

The evaluation of the algorithm's performance was based on precision (P), which measures the proportion of correctly identified objects among all the objects predicted by the model; recall (R), which measures the proportion of the correctly identified objects among all the ground truth objects; and F1 score, a combined metric that takes into account both precision and recall to provide a single value that represents the overall performance. Object detection algorithms have capabilities that go beyond classification algorithms, being able to detect multiple objects as well as their location and size within the image, in the form of bounding boxes. Therefore, to compute those metrics, additional considerations must be put in place. For each proposed bounding box with confidence score greater than 50%, it is considered as a true positive (TP) if the intersection over union with the ground truth is greater than 0.5 and the class is correct. Conversely, if the predicted area corresponds to artifacts or other parasites class then it is considered as false positive (FP). Furthermore, ground truth boxes that were not proposed by the algorithm were categorized as false negatives (FN). True negative (TN) were not computed as all areas without predictions are considered TN.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{6.1}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{6.2}$$

$$\text{F1 score} = 2 * \frac{P * R}{P + R} \tag{6.3}$$

6.3 Results

6.3.1 Evaluation of the AI model performance

The performance assessment of the model was conducted on the validation set with 30 cases as described in Table 6.1. The screening algorithm, designed to work with 10x magnification, achieved a precision of 88.17%, recall of 91.62%, and an F1 score of 89.85%. On the other hand, the species differentiation algorithm achieved a weighted precision of 84.08%, recall of 95.33%, and an F1 score of 94.70%. Breaking down the results per class, the precision rates

		AI Predictions				
		Artifact	*L. loa*	*M. perstans*	*W. bancrofti*	*B. malayi*
Ground truth	Artifact	-	7	3	2	3
	L. loa	6	129	0	1	1
	M. perstans	4	0	98	0	0
	W. bancrofti	0	0	0	47	1
	B. malayi	1	0	0	0	12

Table 6.3: Confusion matrix of the species differentiation algorithm (40x) on the validation set, each row represents the ground truth, and each column represents the prediction. The model may predict the artifacts as a parasite (false positive), but the analysts did not label all artifacts.

were 94.85% for *L. loa*, 97.03% for *M. perstans*, 94.00% for *W. bancrofti*, and 66.67% for *B. malayi*. The corresponding recall rates were 93.48%, 96.08%, 97.92%, and 92.31% respectively.

The resulting confusion matrix of the species differentiation algorithm that works with 40x magnification is presented in Table 6.3. It should be noted that the AI algorithm was not specifically trained with artifact labels. To avoid the confusion with artifacts, areas on the image that may look like a parasite (e.g., mycelium, fibers) were included in the training dataset as negative examples without specifying the class. However, in order to increase performance, it could be possible to create additional classes for the different structures that might lead to false positives, like hair or mycelium.

6.3.2 Validation of the AI-assisted mobile app

For the pilot study, a total of independent 18 samples from different subjects with respect to the ones used for training and validation were analyzed with AI assistance on the mobile phone by analyst A. 452 field of views of 10x magnification and 624 field of views of 40x magnification were analyzed on the mobile phone, uploaded to the telemedicine platform, and reviewed by another analyst, generating the ground truth to evaluate the model performance in real time.

To assess the potential benefits of reviewing AI analysis and its impact on inter-observer variability and time, we shuffled and split the uploaded images into four groups, 10x magnification with AI and without AI assistance (232 images with AI and 220 images without AI), and 40x magnification with and without AI assistance (320 images with AI and 304 images without AI). Both analysts analyze the two groups without AI assistance -10x without AI and 40x without AI - from scratch and analyze the two groups with AI assistance -10x with AI and 40x with AI - by reviewing the prediction generated by the AI on the smartphone. This allowed us to investigate the potential time-saving benefits and the potential reduction in inter-observer variability provided by AI assistance.

		AI predictions						
		Artifact	Microfilaria	Artifact	*L. loa*	M. perstans	*W. bancrofti*	*B. malayi*
		(10x)	(10x)	(40x)	(40x)	(40x)	(40x)	(40x)
Ground truth	Artifact (10x)	-	53					
	Microfilaria	75	851					
	Artifact (40x)			-	8	10	0	13
	L. loa			10	658	0	0	0
	M. perstans			0	0	15	0	0
	W. bancrofti			3	0	0	19	3
	B. malayi			0	0	0	0	23

Table 6.4: Performance of the AI algorithm on pilot study using mobile phone. Each row represents the ground truth and each column represents the AI prediction.

Real-time AI-Performance

The analysis conducted by analyst B, who has greater expertise compared to analyst A, was considered as the ground truth for our evaluation. According to analyst B, at 10x magnification, out of the 452 digitized images examined, 280 were identified as positive, indicating the presence of at least one parasite, while 172 were determined to be negative. The parasite count reported by analyst B and the AI for each image is significantly correlated, with a pearson correlation coefficient of 0.984. At parasite level, the screening algorithm achieved an overall performance of 94.14% precision, 91.90% recall, and 93.01% F1-score. In the context of differentiating between parasite species, according to analyst B, out of the 624 images assessed, 511 were classified as positive and 113 as negative. The parasite count reported by analyst B and the AI for each image is also significantly correlated for species differentiation algorithm, with a pearson correlation coefficient of 0.953. The AI algorithm demonstrated an overall precision of 95.46%, recall of 97.81%, and F1-score of 96.62% in this regard. The per-class precision values were determined as 98.80% for *L. loa*, 60.00% for *M. perstans*, 100.00% for *W. bancrofti*, and 58.97% for *B. malayi*. The corresponding recall rates were calculated as 98.50%, 100.00%, 76.00%, and 100.00%, respectively. Table 6.4 presents the confusion matrix of the AI model in relation to analyst B's analysis.

Inter-observer variability

To assess inter-observer variability, we compared the total number of parasites detected in each sample by analyst B and analyst A, both with and without AI assistance, and the results are presented in Table 5. Two-tailed t-test is used to analyze statistical significance. Additionally, we compared the parasite count of each image generated by the AI system with the count provided by analyst B (considered as the ground truth). This comparison allowed us to analyze the performance and agreement between the analysts and the AI system in identifying and quantifying parasites.

	With AI	Without AI
10x	0.994*	0.990*
40x	0.997*	0.992*

Table 6.5: Inter-observer agreement of detected parasites when analyzing with and without AI assistance of 2 experts and of the AI. Two-tailed t-test indicates that the analysis of analist A and B are significantly correlated.* p-value<0.05.

	without AI assistance	with AI assistance
Analyst A	23.5	3.5
Analyst B	12	12

Table 6.6: Average analysis time in seconds for both analysts with and without AI assistance for each image.

The results reported by both analysts are strongly correlated. The Pearson correlation coefficient between analyst A and analyst B when analyzing without AI is 0.990 for the screening algorithm and 0.992 for species differentiation algorithm. Similarly, when analyzing with AI assistance, the correlation coefficients are 0.994 and 0.997 respectively. Notably, the correlation coefficients are slightly higher when analysts utilize AI assistance during their assessments.

Furthermore, it is noteworthy that there is a high correlation between the parasite counts reported by the AI model and analyst B. The minimum Pearson correlation coefficient observed in this comparison was 0.928, further indicating a strong correlation between their reported counts.

6.3.3 Analysis time

In addition to evaluating the performance of the edge AI system, we also analyzed the potential time-saving effect of AI assistance on the telemedicine platform. We compared the time required for analysts to review AI predictions versus the time needed for labeling from scratch. Both analysts were asked to review 524 images without AI assistance, and then 524 different images with AI assistance. As shown in Table 6.6, for Analyst A, the analysis time significantly decreased from an average of 23.5 seconds per image to just 3.5 seconds per image when utilizing AI assistance. However, it should be noted that the analysis time for Analyst B remained unchanged.

6.4 Discussion and Conclusion

This study introduces the first real-time edge AI deployment on smartphones to assist in the screening and species differentiation for filarial samples in mobile microscopy and validated it in a clinical setting. To create and validate the AI powered mobile application, we proposed

a methodology that encompasses an image digitization system, a telemedicine platform to visualize and annotate images, a training and deployment pipeline, and an Android application to deploy AI models.

Diagnosis is an essential part of the monitorization of the effect of MDA, which is a recommended strategy to control or eliminate several neglected tropical diseases, including filariasis. Microscopy is a widely used technique for filariasis diagnosis, as it can distinguish parasite species, but it requires expert microscopists, and is time-consuming. Numerous studies have incorporated AI to aid in the diagnosis of microscopic images, targeting mostly malaria image analysis (Horning et al., 2021; Yu et al., 2023) - a recent review has identified 95 publications for malaria (Fan et al., 2023) -, and more recently also appeared works that deals with STH and schistosomiasis (Armstrong et al., 2022; Li et al., 2020; Meulah et al., 2023; Oyibo et al., 2022; P. Ward et al., 2022; P. K. Ward et al., 2023), leishmaniasis (Gonçalves et al., 2023), Chagas diseases (Morais et al., 2022; Pereira et al., 2020), etc. The number of studies in this topic is very limited, but have yielded promising outcomes, aiming to facilitate the diagnosis of these diseases in LMICs. For example, Yu et al. evaluated a malaria screener employing a custom CNN to detect Plasmodium falciparum using a smartphone on both thin and thick blood smears. Developed with 150 patients and 50 healthy subjects, the model achieved an accuracy of 74.1% compared to expert microscopy on a test set of 190 patients, meeting the WHO level 3 requirement for parasite detection (Rajaraman et al., 2019; F. Yang et al., 2020; Yu et al., 2023). Armstrong et al. proposed an object detection algorithm using ResNet101 to detect Schistosoma eggs on Google Pixel 4. Developed with 205 patients, the model achieved a sensitivity of 91%, a specificity of 85%, and an inference time of 6s (Armstrong et al., 2022). Li et al. presented a study with 1122 patients with a total of 22,444 images. The model trained with 15,700 images from 785 patients, detects visible components on human feces using an object detection algorithm with ResNet 152 as backbone, achieving a mean average precision of 92.16% and mean average recall of 93.56% on a test set with 6740 images from 337 patients (Li et al., 2020). Gonçalves et al. proposed the use of a u-net to segment human visceral leishmaniasis parasites on bone marrow samples. Developed with 150 images with 559 parasites (70% for training, 10% for validation and 20% for testing), the model achieved a Dice coefficient of 80.4% (Gonçalves et al., 2023). Morais et al. proposed the detection of chagas parasites using graph based segmentation algorithm and random forest. Developed with 33 mice samples with 1314 parasites (80% for training and 20% for testing), achieving a precision of 87.6% and a recall of 90.5% (Morais et al., 2022). Very few studies attempted to automate filarias parasite detection detecting microfilarias, without distinguishing species (D'Ambrosio et al., 2015; Elvana & Suryanto, 2022). The only preliminary proof of concept study that uses AI to detect microfilarias is proposed by Elvana et. al, who used a small database with 210 images and a custom CNN with 8 convolutional layers, achieving 70% accuracy (Elvana & Suryanto, 2022).

With respect to these prior works, our proposal allows us to replicate the full diagnostic workflow including 10x and 40x examinations, successfully distinguishing between different microfilariae species, making it particularly valuable in co-endemic areas where multiple species are prevalent. Our system also operates in real-time (610 milliseconds on Samsung S9) without the internet connection, enabling its deployment at the point of care and not relying on expensive or hard to find hardware as it can be utilized with any conventional microscope

and low- to middle-end mobile phones, making it accessible and affordable. The system is easily scalable, as it is deployed on smartphones.

The AI system that we propose follows the conventional workflow, screening the sample at 10x magnification and differentiating species at 40x magnification. Hence, two algorithms were deployed for each use case using 85 samples, which were first validated on 30 samples to assess the model performance and then deployed to the clinical environment to evaluate the whole system usability. The validation in the clinical environment was conducted by analyzing 18 samples with the AI model running on mobile phone in real time, achieving an overall precision of 94.14%, recall of 91.90% and F1 score of 93.01% for the screening algorithm and 95.46%, 97.81% and 96.62% for the species differentiation algorithm respectively.

In the inter-observer variability and analysis time comparison, we found that with AI assistance the correlation between two analysts increased slightly, and the analysis time reduced for the junior researcher in parasitology while it remained unchanged for the expert in infectious diseases microscopy.

It is important to highlight that our AI algorithm didn't incorporate all filarias species detectable in blood, B. timori and *M. ozzardi* were not available. The former is important to be monitored in order to decide when to stop the existing mass drug administration program for lymphatic filariasis (WHO, 2021). Regarding the latter, it shares a geographical distribution overlap with *M. perstans* (WHO, 1997). Since our model didn't include *M. ozzardi*, it can not differentiate between *M. ozzardi* from *M. perstans*, but the screening algorithm should detect the microfilaria even if it is *M ozzardi*. With the increasing calls for the mansonellosis treatment and control program (Ferreira et al., 2023; Ta-Tang et al., 2018), the inclusion of those species will further improve the utility of our AI model. It should be also noted that our study has a limited sample size in general, especially for *W. bancrofti*, and *B. malayi* (35, 58 labels for training respectively). Despite that, our algorithm achieved high precision and recall, even though the performance fluctuates a lot for minority classes. Additionally, the fact that all samples come from one research center may introduce bias and reduce generalizability of our algorithm, performing worse in samples from other centers, due to the sample preparation, etc. To address these limitations, future research will approach a multi-centric study, including training and validating samples from different research centers, involving more analysts, and including *B. timori* and *M. ozzardi* species. Such an extensive validation process would help to assess the robustness and generalizability of the AI system across various real-world settings and conditions to guarantee readiness for deployment to the local health centres.

In conclusion, the presented system can assist the diagnosis of filariasis in resource-constrained settings, particularly when healthcare workers are scarce, by transforming any optical microscope into an intelligent point-of-care device. The system is easily scalable, as it is deployed on smartphones. This approach could reduce the dependency of highly specialized personnel as we can empower community health workers to contribute to filariasis control. Additionally, the system's telemedicine platform provides the opportunity for seeking second opinions and quality control in cases of diagnostic uncertainty, enhancing overall accuracy. The platform also can be used as an epidemiological surveillance platform, contributing to the tracking and the monitoring of the prevalence and distribution of filariasis. Furthermore, our system

can be expanded to other neglected tropical diseases by collecting samples of other diseases, with the vision of creating a universal AI model for parasite detection. We also believe that future AI supporting systems will be multi-modal, incorporating a wide range of clinical inputs from diverse data sources beyond imaging, such as medical text or speech, enhancing the accuracy and generating comprehensible diagnostic interfaces reports (Moor et al., 2023; Rajpurkar & Lungren, 2023). The current AI revolution in medicine should also be viewed as an opportunity for NTDs.

Chapter 7

Towards a foundation model for fecal sample analysis

7.1 Introduction

Medical image analysis is essential for diagnosing a wide range of conditions across multiple fields, including radiology, histopathology, cytology, and parasitology. Recent advances in artificial intelligence (AI) have demonstrated exceptional performance, with over 900 AI systems approved by regulatory bodies for clinical use (FDA, 2024). These AI algorithms are typically based on supervised learning techniques, which require large amounts of labeled data for training—a resource that is often difficult to obtain. Moreover, the challenge of generalizability remains a significant hurdle; AI models that perform well in controlled, in-distribution settings often experience a considerable drop in performance when deployed in out-of-distribution environments. This issue is particularly pronounced in medical imaging, where differences in sample preparation protocols, imaging equipment, and population demographics across hospitals and regions can introduce variability that AI models struggle to handle effectively (Azizi et al., 2022; Y. Yang et al., 2024).

To address the challenges posed by the need for large labeled datasets and the generalizability issues in AI models, self-supervised learning (SSL) has emerged as a promising alternative in the field of medical imaging (Huang et al., 2023). Unlike supervised learning, SSL leverages the vast amounts of unlabeled data available in medical settings to learn useful representations. In recent years, several SSL techniques have been developed, such as SimCLR, SimSiam, BYOL, MoCo, and DINO and its successor DINOv2 (Caron et al., 2021; T. Chen, Kornblith, Norouzi, et al., 2020; T. Chen, Kornblith, Swersky, et al., 2020; X. Chen & He, 2021; Grill et al., 2020; Oquab et al., 2023). While these methods have demonstrated strong performance on the ImageNet benchmark, there are still relatively few studies focused on applying SSL specifically to medical imaging.

Building on these advances, foundation models represent a new paradigm in AI that significantly enhances the capabilities of SSL in medical imaging. Foundation models are large-scale, pre-trained models designed to serve as versatile base models for a wide range of downstream

tasks. Trained on extensive datasets, these models capture rich, generalizable representations that can be fine-tuned for specific applications with minimal labeled data. Their ability to leverage extensive pre-training allows them to excel in tasks where traditional models may struggle due to limited annotated data.

Several studies have showcased the potential of foundation models in medical imaging. For instance, Vorontsov et al. introduced Virchow, a foundation model for computational pathology and rare cancer detection, which utilized 1.5 million images and employed DINOv2. This model achieved a specimen-level area under the curve of 0.95 across nine common and seven rare cancers, demonstrating its robustness and versatility (Vorontsov et al., 2024). Similarly, Koch et al. developed Dinobloom, a foundation model trained with 380,000 blood cell images from 13 datasets, aimed at cell classification in peripheral blood and bone marrow smear samples (Koch et al., 2024). These advancements highlight the significant benefits of pre-training on domain-specific datasets as opposed to more general datasets like ImageNet.

The application of foundation models to stool sample analysis represents a promising frontier in medical diagnostics. For example, Soil-transmitted helminths (STH), including hookworm, *Ascaris lumbricoides*, and *Trichuris trichiura*, are prevalent in low- and middle-income tropical and subtropical regions, affecting over 1.5 billion people globally. These infections can lead to serious health issues such as anemia, gastrointestinal distress, and chronic fatigue, significantly impacting quality of life. Despite advancements in diagnostic methods, including DNA-based techniques like PCR and LAMP, as well as serological tests, microscopy remains the most widely used method for diagnosing diseases such as STH and malaria. The Kato-Katz technique, a microscopy-based method, is commonly used to diagnose STH infections. This technique involves preparing stool samples on microscope slides for visual inspection (Katz et al., 1972).

In the field of STH, multiple artificial intelligence algorithms have been proposed to assist the detection and quantification of STH parasites, all of them using supervised technique, (Armstrong et al., 2022; Holmström et al., 2017; Jiménez et al., 2020; Kitvimonrat et al., 2020; Lee & Lee, 2020; Li et al., 2020; Lundin et al., 2024; Meulah et al., 2023; Naing et al., 2022; Oyibo et al., 2024; P. Ward et al., 2022; P. K. Ward et al., 2023; A. Yang et al., 2019). Notably, no studies to date have applied SSL techniques to the analysis of STH, indicating an area ripe for further exploration and potential breakthroughs.

The objective of this study is to propose and develop the first foundation model for stool sample analysis. The proposed system consists of several key steps: the collection of stool samples, the labeling of a subset of images, the creation of an object detection algorithm using YOLOv8, and the development of a parasite classification algorithm using DINOv2, which learns representations from unlabeled images. By combining these approaches, this study aims to advance the field of parasitology with an AI-based system that is both scalable and effective, ultimately contributing to more efficient and accurate diagnostic workflows in clinical settings.

This study addresses and fulfills objective 2 of this PhD thesis by mitigating the scarcity of labels in medical imaging. It shifts the focus from supervised learning to unsupervised learning. The primary contribution of this chapter is the establishment of a pipeline for

generating foundation models for stool samples, which will serve as the building blocks for a system capable of performing various tasks and diagnosing multiple parasites in fecal samples.

7.2 Material and methods

7.2.1 Dataset

We collected an extensive dataset composed of 1,380 stool samples from children aged between 5-15 years in Kwale, Kenya. Each stool sample was prepared using the Kato-Katz thick smear method and visually analyzed by conventional microscopy. In parallel, each stool sample was digitized by taking pictures of the field of view and the images were transferred to a cloud telemedicine platform. Both processes were made at 100x magnification (0.08 µm/pixel).

The proposed digitization system is based on a 3D-printed adapter that allows coupling a smartphone to a conventional microscope by aligning the smartphone camera with the objective of the microscope to acquire the images. This adapter, which has been designed to be compact and portable (approximately 0.5kg, 30x15x10 cm) is able to convert any conventional microscope into a digital one in a low-cost manner, and enables the digitization of microscope samples without the need for expensive scanners. Additionally, the system has been designed to be universal, working with any microscope model and any smartphone model.

From this dataset, a total of 163 stool samples (comprising a total of 3,075 field-of-view images) were further digitally analyzed by labeling and tagging all parasites that can be seen in the images, distinguishing among different species, including *Ascaris lumbricoides*, *Trichuris trichiura* and hookworm. This annotated dataset was split at patient level into training (65%), validating (20%) and testing (15%). This split was used for both the parasite detection algorithm and the species classification stage. All images from the remaining unannotated stool samples (comprising a total of 1,217 stool samples with 14,083 field-of-view images) were used for training the self-supervised phase of the pipeline. For the unsupervised dataset, instead of using randomly cropped patches, we utilized an object detector that we had previously trained. We only cropped zones where the detector identified parasites, significantly increasing the amount of relevant data. Table 1 illustrates the dataset distribution, which is maintained consistently for both the detection and classification tasks. In the detection task, the entire field of view image is utilized, whereas for classification, patches cropped from the field of view images where parasites are identified are used.

7.2.2 Overview of the proposed foundational method

The proposed foundational model for parasite detection and classification operates as follows: First, an object detection algorithm identifies all parasites present in an image, regardless of their species. Then, a classification algorithm categorizes each para-site into its specific species. This classification algorithm is based on a foundational model for species differentiation. Initially, it is trained on a large unannotated dataset of parasite images to learn visual representations of stool parasites using a self-supervised approach. After learning these domain specific image-based features, the algorithm is further fine-tuned on a labeled dataset to accurately discern the species of each parasite. Fig. 1 illustrates the proposed approach.

Data set	#Patients	#Images	*Ascaris*	*Trichuris*	Hookworm	Artifact	Total
SSL	1,217	14,083	-	-	-	-	14,680
Train	84	1,637	6,294	1,283	586	2,253	10,416
Validation	43	817	1,351	698	325	937	3,311
Test	36	621	899	811	201	932	2,843

Table 7.1: Data Distribution: Breakdown of unannotated (SSL) and annotated (Train; Validation; Test) data sets. Expert labels include Ascaris, Trichuris, and hookworm classes, along with artifacts (false positives generated by the detection algorithm).

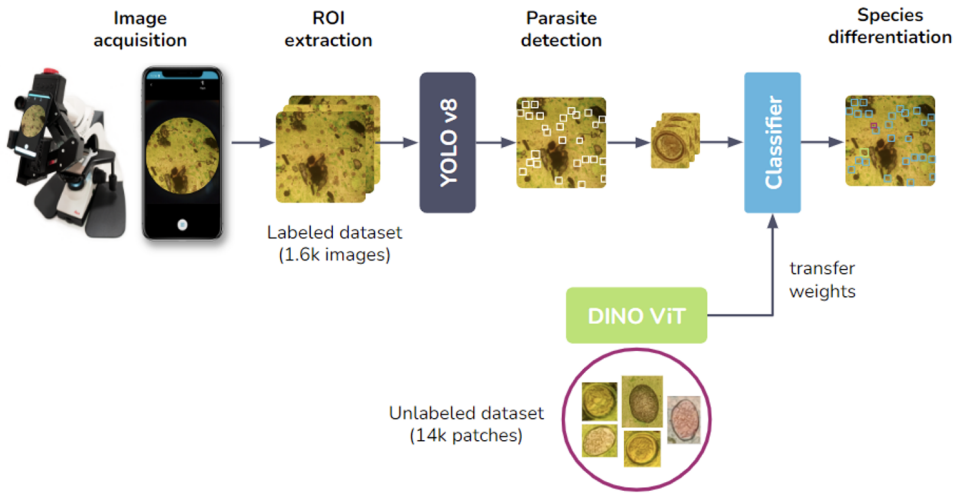


Figure 7.1: Overview of the proposed pipeline comprising: (1) image acquisition using a 3D-printed adapter; (2) field-of-view ROI extraction; (3) parasite detection using YOLOv8; (4) self-supervised pre-training with DINO; (5) species differentiation with our DINO pretrained ViT classifier.

7.2.3 Slide-level parasite detection

Each field of view image is processed through an object detection algorithm to detect all possible parasites regardless of the species. The proposed algorithm for parasite detection in stool sample images was based on a YOLO architecture (YOLOv8) (Jocher et al., 2023), which is a single-stage object detection algorithm that uses a convolutional neural network (CNN) as backbone. Unlike two-stage algorithms, single-stage detection models like YOLO offer enhanced processing speed, making them highly suitable for mobile deployment. This algorithm detects all parasites in a given image without regard to the species. For each detected parasite, an image patch is then extracted and further processed by the patch-level parasite classification to determine the specie.

7.2.4 Patch-level parasite classification

In this study, we aimed to enhance stool parasite differentiation by utilizing self-supervised learning (SSL) to train a feature extractor (backbone) for better data generalization using unlabeled datasets. Specifically, we propose the use of a vision transformer (ViT) (Dosovitskiy et al., 2020) (ViT-S/16) as the backbone, which has achieved significant success in various domains and whose application in the medical field is rapidly expanding. During the pre-training (self-supervised phase), we adopted the DINO technique (Caron et al., 2021), a knowledge distillation method that does not require labeled data. ViT operates by dividing an image of fixed-size patches (NxN), each patch is passed to a linear operator to obtain patch embedding. To preserve spatial information of each patch, their position is encoded and is added later to each patch embedding. The resulting sequence is fed to the transformer encoder. In order to perform classifi-cation, an extra learnable classification token (CLS) is added to the patch embedding, and the result is passed to a linear classification head to classify the image. This classification token and head is not required for SSL. The general concept of DINO is summarized in Fig. 2. It has two models that follow the same architecture, teacher and student, parameterized by t and s. For an input image I, patches of different sizes were generated: large patches (global crops) and small patches (local crops). All crops are followed by extensive augmentation. For a crop x, pair of view (x1, x2) were generated with random augmentation, both models produce output probabilities Pt and Ps, obtained by normalizing the output of the model with a softmax function. The student model learns to match distribution by minimizing the cross-entropy loss mins CE(Pt(x), Ps(x). The parameters of the stu-dent s are updated by stochastic gradient descent, and the parameters of the teacher t are updated using the exponential moving average (EMA) of the students.

After pre-training, we conducted supervised training. In this phase, we add a multi-layer perceptron (MLP) layer on top of the ViT encoder to perform classification. To ensure the need for SSL pretraining we tried two different approaches: first, by freezing all weights of the architecture except those from the classification layer (linear probing), and second, by enabling fine-tuning of all weights in the architecture, including the ViT encoder.

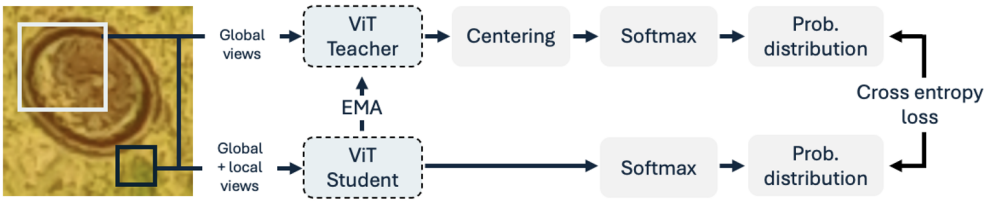


Figure 7.2: Overview of the proposed pipeline for training a ViT architecture with DINO SSL strategy.

7.2.5 Experimental setup

For comparative purposes, we compared different versions of the YOLOv8 architecture for parasite detection. This comparison aimed primarily to identify the optimal architecture for deployment on edge devices, such as smartphones, enabling real-time detections. Testing multiple YOLOv8 variants allowed us to determine their performance and suitability for this specific application.

Additionally, we trained the ViT model without pre-training to assess the improvement conferred by self-supervised learning (SSL). The comparison of these models with and without pre-training provided insights into the benefits of SSL in improving model performance.

All comparisons, including those for parasite detection and classification, were conducted on a validation set to ensure consistency and reliability in the performance assessments. The optimal configurations, determined through these validation tests, were then evaluated on an independent test set to confirm their efficacy and generalizability. This comprehensive approach allowed us to identify the best-performing models for both tasks and ensure their readiness for practical deployment.

The metrics used for evaluating the object detection algorithm included mean aver-age precision (mAP), the precision and recall. The performance of the classification algorithm was assessed by measuring the balanced accuracy (BACC) and F1-score to account for data imbalance and to provide a more comprehensive evaluation of the classifier's effectiveness across all classes

7.3 Experiments and results

We conducted a set of experiments to evaluate the performance of various YOLOv8 models: YOLOv8-n (nano), YOLOv8-s (small), YOLOv8-m (medium), and YOLOv8-l (large). Instead of training the model from scratch, we used the weights pretrained with COCO image dataset and fine tuned on our dataset. Our experiments utilized an input size of 640x640 pixels, a learning rate set to 0.01, and we trained each model for 100 epochs with early stopping implemented after 20 epochs without improvement by monitoring the loss in the validation set. We evaluated the mAP, precision, recall, and inference time of each one. Table 2 illustrates the performance of each YOLOv8 model. Our results indicate that all four models achieved a comparable mAP of 97%. However, YOLOv8-n demonstrated significantly faster inference

Architecture	mAP	Precision	Recall	Inference time/image (ms)
YOLOv8-n	97.07	91.27	87.19	372
YOLOv8-s	97.79	90.72	91.36	465
YOLOv8-m	97.47	89.18	92.33	696
YOLOv8-l	97.04	87.02	94.36	1060

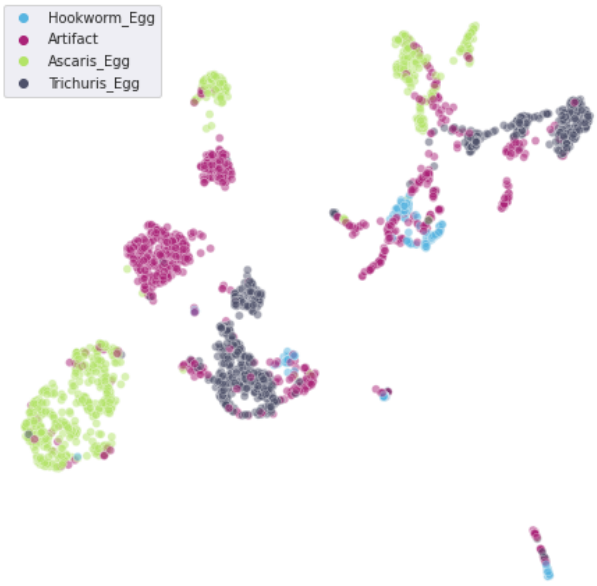


Figure 7.3: Uniform manifold approximation and projection (UMAP) visualization of the features extracted from our pre-trained ViT architecture.

times, being approximately three times quicker than YOLOv8-l. This makes YOLOv8-n particularly well-suited for deployment on edge devices where computational resources are limited.

7.3.1 Patch-level parasite classification

During the self-supervised training phase, we employed the following setup: a ViT-S/16 model with a patch size of 16. The model underwent training for 200 epochs with a batch size of 96 and an input size of 224x224, utilizing a cosine learning rate scheduler (initial LR: 5e-4, minimum LR: 1e-6). The backbone was trained using patch-es generated by the object detection algorithms (N=14,680). Fig. 3 illustrates the efficacy of the pretrained model in our domain-specific dataset, in extracting relevant features from stool parasites, demonstrating that features belonging to the same class are clustered together while those from different classes are distinctly separated.

To assess the benefits of using self-supervised learning (SSL) on a domain-specific dataset, we conducted an experiment comparing three different pre-training approaches for the ViT architecture: SSL on a domain-specific dataset (DINO-STH), SSL on a domain-agnostic

Pre-training	Fine tuning		Linear probing	
	BACC	**F1-score**	**BACC**	**F1-score**
None	57.48	55.13	47.90	46.61
DINO-ImageNet	55.45	53.48	85.10	85.15
DINO-STH	**91.48**	**91.70**	90.51	90.86

Table 7.2: Parasite classification performance on the validation set, obtained from full fine-tuning and linear probing

dataset (DINO-ImageNet), and no pre-training. In all three cases, we utilized the same ViT-S/16 architecture to ensure a fair comparison.

All experiments were conducted under the same hyperparameters: a batch size of 128, learning rate of 0.001, training for 100 epochs with early stopping patience of 10 epochs, and images resized to 224x224 pixels. During the supervised training phase, we executed two types of training strategies: fine-tuning, where both the backbone and the linear classifier (MLP) were trained together, and linear probing, where only the linear classifier was trained while keeping the backbone fixed. All these experiments used all available data of the training set. Table 3 presents the performance of all models on the validation set.

In addition, and for comparison purposes, we performed supervised training with only 200 images per parasite class, instead of using all available dataset, to assess the models' capacity when only a limited labeled dataset is available. When evaluated on the validation set, the model pre-trained on our domain-specific dataset (DINO-STH) achieved a BACC of 86.57% and a F1-Score of 86.24%, whereas when it was pre-trained on the domain-agnostic dataset (DINO-ImageNet) achieved 74.76% and 74.05% of BACC and F1-Score respectively.

7.3.2 Evaluation of the proposed system on an independent test set

To evaluate the generalization capabilities of our proposed system, we set aside a subset of labeled images, independent from those used for training and validation. After determining the best configurations from the validation set, we applied these configurations to the test set. The evaluation revealed that the optimal configuration for parasite detection was YOLOv8-n, which achieved a mean Average Precision (mAP) of 94.93%, precision of 89.63%, and recall of 88.50% on the test set. For parasite species classification, the best performance was achieved using the ViT-S/16 architecture pre-trained on our domain-specific unlabeled dataset through self-supervised learning (SSL). All detected boxes (those with a probability greater than 0.05) were then processed by the classification algorithm, and the predictions were compared to the annotations made by experts. The performance of the parasite classification algorithm was 80.32% balanced accuracy (BACC) and 80.17% F1-score. The whole system achieved a mAP of 82.5%, precision of 89.63% and recall of 82.14%. This assessment on an independent test set underscores the robustness and generalizability of our approach, demonstrating its potential for accurate parasite detection and classification in practical applications.

7.4 Conclusions

In this work, we presented a comprehensive approach for the detection and classification of soil-transmitted helminth (STH) parasites using YOLOv8 and Vision Transformers (ViT). For the classification task, we created a foundation model based on self-supervised learning (SSL) techniques to leverage a large amount of unannotated data, enabling the model to learn meaningful features. The proposed classification system trained on our domain-specific data achieved better results compared to the ViT backbone pre-trained on domain-agnostic dataset (ImageNet). This improvement is particularly noticeable when the available annotated training set is small. With only 200 training images per parasite class, the performance improved by 12% when com-paring our SSL-pretrained model to the one trained on a domain-agnostic dataset. This work is highly relevant because, in the field of medical imaging, the availability of labeled data is often limited. By incorporating SSL and leveraging unannotated data, we have shown that it is possible to enhance model performance, especially in data-scarce environments. This approach, which also leverages 3D-printing technologies and smartphones to enable data digitization without the need for expensive hard-ware, holds great promise for improving diagnostic accuracy and accessibility in low-resource settings, ultimately aiding in the fight against parasitic diseases. This marks a step toward meeting the World Health Organization's performance benchmarks for in-vitro diagnostic devices, leveraging AI to combat parasitic diseases.

Chapter 8

Concluding remarks

Artificial intelligence has experienced tremendous progress in recent years and is transitioning from research to deployment. In healthcare, AI is transforming service delivery in many high-income countries, with over 900 AI algorithms approved for use, more than 600 of which were approved after 2020, indicating a strong trend toward AI integration. Among these applications, AI in radiology, cardiology, and neurology is the most extensively researched. However, substantial obstacles remain before these products can achieve widespread clinical success, particularly in low- and middle-income countries.

Microscopy is a fundamental tool in biological and medical research, enabling the visualization of structures and organisms that are too small to be seen with the naked eye. In clinical settings, microscopy remains a cornerstone for diagnosing various conditions, such as hematology or infections caused by bacteria, parasites, and fungi. Despite advancements in molecular diagnostics, microscopy continues to be indispensable, particularly in resource-limited settings where it remains the most accessible and cost-effective diagnostic method.

While microscopy is a powerful and essential tool in biological and medical sciences, it has several limitations. First, it often requires trained technicians to accurately identify and interpret specimens, a resource that is not always available in all settings. Second, proper sample preparation is crucial; poor preparation can lead to artifacts or degradation of the sample, which can compromise the quality of observations. Third, the analysis can be difficult and subjective, leading to potential inconsistencies in results. Additionally, microscopy is labor-intensive and slow, especially when large numbers of samples need to be analyzed. The reliance on human observation introduces the potential for error, particularly when fatigue or inexperience come into play, resulting in variability in outcomes. Certain samples, such as those used in the Kato-Katz technique for detecting soil-transmitted helminths, degrade quickly and must be analyzed within a short time frame, adding another layer of complexity. Moreover, the process is largely manual, limiting efficiency.

However, recent technological advancements are beginning to address these challenges. The development of automatic scanners, the incorporation of cameras into microscopes, and the advent of mobile microscopy are transforming the field. By digitizing images, these innovations help prevent sample degradation, facilitate telemedicine to address healthcare shortages, and

lay the groundwork for automated analysis, potentially increasing accuracy and efficiency in microscopy.

In this context this PhD thesis has advanced in several lines of research with relevant contributions presented in the following paragraphs.

Data labeling is a crucial step in the development of an AI algorithm, serving as the foundation for the model's ability to learn and make accurate predictions. However, this process is notoriously time-consuming and poses a significant burden on experts, especially in specialized fields like medical imaging. In Chapter 4, we explored the use of crowd-sourcing as an innovative image labeling technique to alleviate this challenge. By collecting annotations from both school-age children and adults, and subsequently using these labels for training of AI models, we focused on classifying three species of soil-transmitted helminth eggs. Remarkably, the models trained on these crowdsourced annotations demonstrated performance levels comparable to those trained on expert-labeled data, highlighting the potential of crowd-sourcing as a viable and effective alternative for data labeling in AI development.

The training of AI models heavily depends on large datasets, and validating AI deployment is crucial to assess the system's usability in real-world settings. In Chapter 5, we conducted a proof-of-concept study aimed at digitalizing the soil-transmitted helminths (STH) analysis workflow. Initially, we collected samples containing two STH egg species, *Ascaris lumbricoides* and *Trichuris trichiura*, to create a baseline AI model for subsequent use during a pilot study. In this pilot study, we gathered 1,343 stool samples from 13 schools in Kenya, where 8 analysts digitized the samples using smartphones coupled to microscope oculars with AI assistance. Throughout the study, we labeled some of the images, retrained the AI algorithm, and redeployed it to support the following day's analysis. This iterative process allowed us to build a large dataset, including digitized images of hookworm eggs—a species that typically degrades within 30 minutes of sample preparation. The study demonstrated the feasibility of integrating AI algorithms in point-of-care settings, providing a valuable tool for improving diagnostic workflows in resource-limited environments.

The performance of AI algorithms in medical settings is influenced by various factors, including the quality of input data, the operation of the system, and the clinician's interpretation. To address these challenges, Chapter 6 details the development and pilot testing of a system for real-time, automatic detection and quantification of filariasis using an edge AI model. The proposed system is designed to aid in the screening and species differentiation of four filarial worm species (*Loa loa*, *Mansonella perstans*, *Wuchereria bancrofti*, and *Brugia malayi*) in blood smears. The system's pipeline includes several key modules: digitization of smear samples using smartphones coupled with microscopes via a 3D-printed adapter, sample analysis, and data labeling through a telemedicine platform for AI training. The trained algorithm is then integrated into the smartphone to assist with diagnosis, with its performance validated in a clinical environment. This study marks the first deployment of a real-time edge AI model on a smartphone, a significant achievement that could transform any optical microscope into an intelligent point-of-care device. This innovation is particularly valuable in resource-limited settings where healthcare workers are scarce, offering a practical and scalable solution for improving diagnostic accuracy and efficiency.

Traditional supervised learning techniques require large amounts of labeled data to achieve strong performance, which can be challenging to obtain, particularly in medical imaging. To overcome these challenges, self-supervised learning has emerged as a promising alternative, capable of learning meaningful representations from vast amounts of unlabeled data. In Chapter 7, we explored the application of SSL to enhance the diagnosis of soil-transmitted helminths. During the pilot study described in Chapter 5, a large dataset was collected, but the majority of these data remained unlabeled. We leveraged these unlabeled data and applied a self-supervised learning approach to train a vision transformer backbone. This pretrained backbone was then fine-tuned using a smaller labeled dataset. The experiment demonstrated that SSL could effectively utilize unlabeled data, improving the model's performance and reducing the reliance on extensive labeled datasets. This approach shows great potential for advancing AI-based diagnostics, particularly in resource-constrained settings where labeled data are limited.

The methodology, workflows, and outcomes from this PhD thesis serve as a foundation for developing trustworthy and user-friendly AI algorithms for microscopic image analysis. These results enable rapid expansion to new applications. The techniques have been integrated into the Spotlab AI platform and product portfolio, and are already in use in clinical research settings.

8.1 Future works

This thesis has established a strong foundation for integrating AI in diagnosing and managing neglected tropical diseases, but there are several areas that require further exploration to fully harness AI's potential in this field. One key direction for future research is the expansion of AI applications to other NTDs, such as schistosomiasis, leishmaniasis, and Chagas disease, which also rely heavily on microscopy for diagnosis. By including samples from these diseases, it is possible to develop two parasite parasite detection algorithms, one for stool samples and another for blood samples, thereby broadening the impact of AI-driven solutions across multiple conditions.

In addition to microscopy, incorporating other data sources such as medical records and speech analysis could enhance diagnostic accuracy and enable the generation of comprehensive diagnostic reports. This multimodal approach could provide a more holistic understanding of patient conditions, further improving outcomes.

Improving the generalizability of AI models is another critical area of focus, particularly when deploying models across different geographic regions and populations. Since the data collected for this thesis was from a single center, there may be biases that reduce the algorithm's effectiveness when applied to samples from other centers, due to variations in sample preparation and other factors. To overcome these limitations, future research should pursue multi-centric studies, including training and validating models with data from diverse research centers and involving a wider range of analysts.

The successful deployment of edge AI models on smartphones, as demonstrated in this thesis, presents opportunities for further advancement. Future research should aim to enhance

the computational efficiency of these models, enabling real-time analysis of more complex diagnostic tasks and expanding their application beyond microscopy.

Longitudinal studies will be crucial to assess the long-term impact of AI integration on disease management and health outcomes in NTD-endemic regions. This includes evaluating the cost-effectiveness of AI systems, their acceptance by healthcare workers and patients, and their overall impact on disease prevalence and control efforts.

As AI continues to be integrated into healthcare, ongoing research into the regulatory and ethical implications, especially in low- and middle-income countries, is essential. Future work should focus on developing guidelines that ensure the safe, ethical, and equitable deployment of AI technologies in these settings. Collaboration among AI researchers, healthcare professionals, policymakers, and local communities will be vital to creating trustworthy and effective AI solutions.

By addressing these future research directions, we can move closer to realizing the full potential of AI in transforming NTD diagnostics and treatment, ultimately contributing to global efforts to control and eliminate these diseases.

Chapter 9

Publications

The following sections list the journal articles and conference papers/abstracts published during the development of the PhD thesis. The rank of the journal (D1- first decile, Q1- first quartile from JCR/WoS), the contribution of the PhD candidate, as well as the number of citations (NC) are highlighted for each publication.

9.1 Journal articles

1. **Lin, L**, Dacal, E., Díez, N., Carmona, C., Martin Ramirez, A., Barón Argos, L., Bermejo-Peláez, D., Caballero, C., Cuadrado, D., Darias-Plasencia, O., García-Villena, J., Bakarjiev, A., Postigo, M., Recalde-Jaramillo, E., Flores-Chavez, M., Santos, A., Ledesma-Carbayo, M. J., Rubio, J. M., & Luengo-Oroz, M. (2024). Edge Artificial Intelligence (AI) for real-time automatic quantification of filariasis in mobile microscopy. PLoS Neglected Tropical Diseases, 18(4), e0012117. https://doi.org/10.1371/JOURNAL.PNTD.0012117

    - Journal ranking: Q1 (Tropical Medicine)
    - Number of citations: 1
    - Main contribution: The PhD candidate has been in charge of data inspection and preprocessing, AI model training and deployment, validation protocol design, statistical analysis and manuscript drafting. An editorial titled "AI Sees an End to Filariasis" is dedicated to this article, underscoring the significant impact of this work (Gaunt & Crainey, 2024).

2. Dacal, E*., Bermejo-Peláez, D.*, **Lin, L***, Álamo, E., Cuadrado, D., Martínez, Á., Mousa, A., Postigo, M., Soto, A., Sukosd, E., Vladimirov, A., Mwandawiro, C., Gichuki, P., Williams, N. A., Muñoz, J., Kepha, S., & Luengo-Oroz, M. (2021). Mobile microscopy and telemedicine platform assisted by deep learning for quantification of Trichuris trichiura infection. PLOS Neglected Tropical Diseases, 15(9), e0009677. https://doi.org/10.1371/journal.pntd.0009677

    - Journal ranking: Q1 (Tropical Medicine)

- Number of citations: 17
- Main contribution: The PhD candidate, along with Elena Dacal and David Bermejo-Peláez, contributed equally to the project. The PhD participated in data inspection and preprocessing, AI model training, statistical analysis, and manuscript drafting.

3. García-Villena, J., Torres, J. E., Aguilar, C., **Lin, L.**, Bermejo-Peláez, D., Dacal, E., Mousa, A., Ortega, M. D. P., Martínez, A., Vladimirov, A., Cuadrado, D., Postigo, M., Ordi, J., Bassat, Q., Salamanca, J., Rodriguez-Peralto, J. L., Linares, M., Ortuño, J. E., Ledesma-Carbayo, M. J., ... Luengo-Oroz, M. (2021). 3D-Printed Portable Robotic Mobile Microscope for Remote Diagnosis of Global Health Diseases. Electronics 2021, Vol. 10, Page 2408, 10(19), 2408. https://doi.org/10.3390/ELECTRONICS10192408
   - Journal ranking: Q2(Engineering, Electrical & Electronics)
   - Number of citations: 8
   - Main contribution: The PhD candidate contributed to technology development, clinical validation, and manuscript writing.
4. Bermejo-Peláez, D., Marcos-Mencía, D., Álamo, E., Pérez-Panizo, N., Mousa, A., Dacal, E., **Lin, L.**, Vladimirov, A., Cuadrado, D., Mateos-Nozal, J., Galán, J. C., Romero-Hernandez, B., Cantón, R., Luengo-Oroz, M., Rodriguez-Dominguez, M. (2022). A Smartphone-Based Platform Assisted by Artificial Intelligence for Reading and Reporting Rapid Diagnostic Tests: Evaluation Study in SARS-CoV-2 Lateral Flow Immunoassays. JMIR Public Health and Surveillance, 8(12). https://doi.org/10.2196/38533
   - Journal ranking: Q1 (Public, Environmental & Occupational Health)
   - Number of citations: 8
   - Main contribution: The PhD candidate contributed in the data acquisition and technical support.
5. Bermejo-Peláez, D., Charro, S. R., Roa, M. G., Trelles-Martínez, R., Bobes-Fernández, A., Soto, M. H., García-Vicente, R., Morales, M. L., Rodríguez-García, A., Ortiz-Ruiz, A., Sánchez, A. B., Urbina, A. M., Álamo, E., **Lin, L**, Dacal, E., Cuadrado, D., Postigo, M., Vladimirov, A., García-Villena, J., ... Luengo-Oroz, M. (2024). Digital Microscopy Augmented by Artificial Intelligence to Interpret Bone Marrow Samples for Hematological Diseases. Microscopy and Microanalysis, 30(1), 151–159. https://doi.org/10.1093/MICMIC/OZAD143
   - Journal ranking: Q1 (Microscopy)
   - Number of citations: 0
   - Main contribution: The PhD candidate participated in the design and development of the software.

9.2 Conference articles and abstracts

1. Ortuño, J. E., **Lin, L**, Del Pilar Ortega, M., García-Villena, J., Cuadrado, D., Linares, M., Santos, A., Ledesma-Carbayo, M. J., & Luengo-Oroz, M. (2020). Stitching Methodology for Whole Slide Low-Cost Robotic Microscope Based on a Smartphone. IEEE International Symposium on Biomedical Imaging, 2020-April, 503–507. https://doi.org/10.1109/ISBI45749.2020.9098355

   The PhD candidate contributed to the clinical validation and deployment of the design method.

   Lin, L*, Bermejo-Peláez, D.*, Capellán-Martín, D., Cuadrado, D., Rodríguez, C., García, L., Díez, N., Tomé, R., Postigo, M., Ledesma-Carbayo, M. J., & Luengo-Oroz, M. (2021). Combining collective and artificial intelligence for global health diseases diagnosis using crowdsourced annotated medical images. 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 3344–3348. 10.1109/EMBC46164.2021.9630868

   The PhD candidate, along with David Bermejo-Peláez, contributed equally to the project. The PhD candidate participated in data preprocessing, experiment design, model training and validation, statistical analysis, and manuscript drafting.

2. **Lin, L**, Bernal Fernández, M. J., Carmona, C., Martín Ramírez, A., Díez, N., Bermejo-Peláez, D., Dacal Picazo, E., García Villena, J., Luengo-Oroz, M., & Rubio Muñoz, J. M. (2022). Una aplicación móvil asistida por inteligencia artificial para el diagnóstico de filariasis en tiempo real. Sociedad Española de Medicina Tropical y Salud Internacional. Libro de actas, pp.178.

   The PhD candidate contributed in the data preprocessing AI model development, and edge deployment.

3. **Lin, L**, Kepha, S., Bermejo-Peláez, D., Cuadrado, D., Mousa, A., Vladimirov, A., García-Villena, J., Postigo, M., Díez, N., Aguilar, C., Recalde-Jaramillo, E., Gichuki, P., Mwandawiro, C., Muñoz, J., Santos, A., Ledesma-Carbayo, M. J., & Luengo-Oroz, M. (2023). Artificial intelligence for NTDs: a system to support the diagnosis of helminthiasis from microscopy using real-time artificial intelligence working in smartphones with limited connectivity. European Society for Paediatric Infectious Diseases. Abstract book, pp 96.

   The PhD candidate contributed in the data preprocessing AI model training, evaluation and deployment.

4. Bermejo-Peláez, D., Tligui, H., Arias, S., **Lin, L.**, Dacal, E., Vallés-López, R., Vladimorov, A., Caballero, C., Cuadrado, D., García-Villena, J., Darias, O., Ftouh, S. el, Babi, R., Bounagua, K., Abbass, C. el, Amalik, N., Zizi, I., Ouardighi, I., Elyazigi, L., ... Luengo-Oroz, M. (2023). Artificial Intelligence-Based System for Automated Cell Counting in Neonatal Cerebrospinal Fluid Analysis: A Remote Quality Control and Digital Diagnostic Tool. European Society for Paediatric Infectious Diseases Conference. Abstract book, pp 305.

The PhD candidate contributed in the data preprocessing and model development and deployment.

Lin, L, Rubio, C. C., Flores-Chavez, M., Torrico, M. C., Ramírez, A. M., Agreda, F. G., Paucara, B. C., Torrico, A. S., Lanza, M., Dacal, E., Pelaez, D. B., Diez, N., Chus, A., Rubio, J. M., & Luengo-Oroz, M. (2023). AI-Based Microscopy: A Game Changer for Neglected Tropical Disease Diagnosis. European Congress on Tropical Medicine and International Health. Abstract book pp 175. https://doi.org/10.1111/tmi.13931

The PhD candidate contributed in the data preprocessing AI model training, evaluation and deployment.

Mancebo-Martín, R., **Lin, L**, Dacal, E., Luengo-Oroz, M., & Bermejo-Peláez, D. (2024). How many labels do I need? Self-supervised learning strategies for multiple blood parasites classification in microscopy images. IEEE International Symposium on Biomedical Imaging, May 2024. https://doi.org//10.1109/ISBI56570.2024.10635899

The PhD candidate contributed in the experiment design, data preprocessing, AI model development, and manuscript drafting.

5. **Lin, L.**, Cuadrado, D., Mancebo-Martín, R., Kepha, S., Gichuki, P., Mwandawiro, C., Ledesma-Carbayo, M.J., Luengo-Oroz, M. Dacal, E., David Bermejo-Peláez, D. (2024). Enhancing Soil-transmitted helminths Diagnosis through AI: A Self-Supervised Learning Approach with Smartphone-Based Digital Microscopy. International Conference on Medical Image Computing, October 2024. Accepted.

   The PhD candidate contributed in the experiment design, data preprocessing, AI model development, and manuscript drafting.

9.3 Posters and Oral Presentations at Medical Conferences

1. **Lin, L**, & Callisaya, B. (2022). Un sistema de microscopía digital asistido por inteligencia artificial para el diagnóstico de la enfermedad de Chagas. XXV Congreso Nacional de La Sociedad Española de Enfermedades Infecciosas y Microbiología Clínica.

   The PhD candidate contributed in the deployment an AI algorithm for the chagas parasite detection.

2. Rubio, J. M., Carmona Rubio, C., Flores-Chavez, M., Bernal, M. J., Torrico, M. C., Martín Ramirez, A., Gonzalez, F., Callisaya, B., Solano, A., Illanes, D., Lanza, M., **Lin, L.**, Dacal, E., Bermejo-Peláez, D., Diíez, N., Mousa Urbina, A., & Luengo-Oroz, M. (2022). Microscopy, telemedicine and Artificial intelligence for the improvement of the blood parasites diagnosis. 20th International Congress for Tropical Medicine and Malaria.

   The PhD candidate contributed in the deployment an AI algorithm for the blood parasite parasite detection.

3. Rubio, J. M., Carmona Rubio, C., Flores-Chavez, M., Torrico, M. C., Gonzales Agreda, F., Callisaya Paucara, B., Solano Torrico, A., Lanza, M., **Lin, L.**, Dacal, E., Bermejo-pelaez, D., Diez, N., Mousa Urbina, A., & Luengo-Oroz, M. (2023). Artificial intelligence and microscopy for the improvement of blood parasites diagnosis. 33rd European Congress of Clinical Microbiology and Infectious Diseases.

   The PhD candidate contributed in the deployment an AI algorithm for the blood parasite parasite detection.

4. Dacal, E., Bermejo-Peláez, D., **Lin, L.**, Rubio, J. M., Dada Chechet, G., el Ftouh, S., Arias, S., Moussa, A., Torrico, M. C., Gonzalez, F., Solano, A., Illanes, D., Rozo-Anaya, J. C., Flores-Chavez, M., García-Villena, J., Cuadrado, D., Darias, O., Postigo, M., Vallés-López, R., ... Luengo-Oroz, M. (2023). Collaborative Research Workshop on Generative AI: Exploring the Potential of Large Language Models in Global Health. American Society of Tropical Medicine & Hygiene.

   The PhD candidate contributed in organization of the workshop.

5. García-Villena, J., Recalde-Jaramillo, E., Cuadrado, D., Bravo, I., Degraft Mensah, S., Opoku, N., Akanburichaab, M., Afi Krampah, C., Edem Agbogah, M., Rashid Adams, A., **Lin, L.**, Bermejo-Peláez, D., Dacal, E., Mahmood Seidu, A., King, C. L., & Luengo-Oroz, M. (2023). Portable Mobile-Based Microscope Scanner: A Case Study for Onchocerciasis Whole Slide Digitization. American Society of Tropical Medicine & Hygiene.

   The PhD candidate contributed in the validation of the Scanner and the development of the algorithm for whole slide imaging generalization.

6. **Lin, L.**, Callisaya, B., Solano, A., Gonzalez, F., Torrico, M. C., Illanes, D., Díez, N., Bermejo-Peláez, D., Dacal, E., Caballero, C., Darias, O., Bakardjiev, A., Postigo, M., Ledesma-Carbayo, M. J., Luengo-Oroz, M., Rubio, J. M., & Flores-Chavez, M. (2023). Tackling Chagas Disease: Advancements in AI-Based Parasite Detection Algorithm. American Society of Tropical Medicine & Hygiene.

   The PhD candidate contributed in the deployment an AI algorithm for the chagas parasite detection.

7. **Lin, L.**, Dacal, E., Díez, N., Carmona, C., Martín Ramírez, A., Barón Argos, L., Bermejo-Peláez, D., Caballero, C., Cuadrado, D., Darias, O., García-Villena, J., Bakardjiev, A., Postigo, M., Recalde-Jaramillo, E., Flores-Chavez, M., Santos, A., Ledesma-Carbayo, M. J., Rubio, J. M., & Luengo-Oroz, M. (2023). Advancing filariasis diagnosis with edge artificial intelligence in mobile microscopy. American Society of Tropical Medicine & Hygiene.

   The PhD candidate contributed in the data preprocessing, AI model development, and edge deployment.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., ... Zheng, X. (2016). Tensorflow: A system for large-scale machine learning [Software available from tensorflow.org]. http://arxiv.org/abs/1605.08695

Abdelmula, A. M., Mirzaei, O., Güler, E., & Süer, K. (2023). Assessment of deep learning models for cutaneous leishmania parasite diagnosis using microscopic images. *Diagnostics*, *14*(1), 12. https://doi.org/10.3390/diagnostics14010012

Akintayo, A., Tylka, G. L., Singh, A. K., Ganapathysubramanian, B., Singh, A., & Sarkar, S. (2018). A deep learning framework to discern and count microscopic nematode eggs. *Scientific Reports*, *8*(1), 1–11. https://doi.org/10.1038/s41598-018-27272-w

Ao Wang, L. L., Hui Chen. (2024). Yolov10: Real-time end-to-end object detection. *arXiv preprint arXiv:2405.14458*.

Armstrong, M., Harris, A. R., D'Ambrosio, M. V., Coulibaly, J. T., Essien-Baidoo, S., Ephraim, R. K., Andrews, J. R., Bogoch, I. I., & Fletcher, D. A. (2022). Point-of-care sample preparation and automated quantitative detection of schistosoma haematobium using mobile phone microscopy. *The American Journal of Tropical Medicine and Hygiene*, *106*(5), 1442. https://doi.org/10.4269/ajtmh.21-1071

Arshad, Q. A., Ali, M., Hassan, S.-u., Chen, C., Imran, A., Rasul, G., & Sultani, W. (2022). A dataset and benchmark for malaria life-cycle classification in thin blood smear images. *Neural Computing and Applications*, *34*(6), 4473–4485. https://doi.org/10.1007/s00521-021-06602-6

Azizi, S., Culp, L., Freyberg, J., Mustafa, B., Baur, S., Kornblith, S., Chen, T., MacWilliams, P., Mahdavi, S. S., Wulczyn, E., et al. (2022). Robust and efficient medical imaging with self-supervision. *arXiv preprint arXiv:2205.09723*. https://doi.org/10.48550/arXiv.2205.09723

Babalola, T. K., & Moodley, I. (2020). Assessing the efficiency of health-care facilities in sub-saharan africa: A systematic review. *Health services research and managerial epidemiology*, *7*, 2333392820919604. https://doi.org/10.1177/2333392820919604

Beng, A. A., Esum, M. E., Deribe, K., Njouendou, A. J., Ndongmo, P. W., Abong, R. A., Fru, J., Fombad, F. F., Nchanji, G. T., Amambo, G., et al. (2020). Mapping lymphatic filariasis in loa loa endemic health districts naïve for ivermectin mass administration and situated in the forested zone of cameroon. *BMC infectious diseases*, *20*, 1–11. https://doi.org/10.1186/s12879-020-05009-3

Bogoch, I. I., Andrews, J. R., Speich, B., Ame, S. M., Ali, S. M., Stothard, J. R., Utzinger, J., & Keiser, J. (2014). Quantitative evaluation of a handheld light microscope for field diagnosis of soil-transmitted helminth infection. *The American journal of tropical medicine and hygiene*, *91*(6), 1138. https://doi.org/10.4269/ajtmh.14-0253

Bogoch, I. I., Andrews, J. R., Speich, B., Utzinger, J., Ame, S. M., Ali, S. M., & Keiser, J. (2013). Mobile phone microscopy for the diagnosis of soil-transmitted helminth infections: A proof-of-concept study. *The American journal of tropical medicine and hygiene*, *88*(4), 626. https://doi.org/10.4269/ajtmh.12-0742

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., et al. (2021). On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*. https://doi.org/10.48550/arXiv.2108.07258

Bortsova, G., González-Gonzalo, C., Wetstein, S. C., Dubost, F., Katramados, I., Hogeweg, L., Liefers, B., van Ginneken, B., Pluim, J. P., Veta, M., et al. (2021). Adversarial attack vulnerability of medical image analysis systems: Unexplored factors. *Medical Image Analysis*, *73*, 102141. https://doi.org/10.1016/j.media.2021.102141

Bosch, F., Palmeirim, M. S., Ali, S. M., Ame, S. M., Hattendorf, J., & Keiser, J. (2021). Diagnosis of soil-transmitted helminths using the kato-katz technique: What is the influence of stirring, storage time and storage temperature on stool sample egg counts? *PLoS neglected tropical diseases*, *15*(1), e0009032. https://doi.org/10.1371/journal.pntd.0009032

Budd, S., Robinson, E. C., & Kainz, B. (2021). A survey on active learning and human-in-the-loop deep learning for medical image analysis. *Medical image analysis*, *71*, 102062. https://doi.org/10.1016/j.media.2021.102062

Cai, L., Gao, J., & Zhao, D. (2020). A review of the application of deep learning in medical image classification and segmentation. *Annals of translational medicine*, *8*(11). https://doi.org/10.21037/atm.2020.02.44

Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., & Joulin, A. (2021). Emerging properties in self-supervised vision transformers. *Proceedings of the IEEE/CVF international conference on computer vision*, 9650–9660.

Caton, S., & Haas, C. (2024). Fairness in machine learning: A survey. *ACM Computing Surveys*, *56*(7), 1–38. https://doi.org/10.1145/3616865

CDC-DPDx. (n.d.-a). Lymphatic filariasis. Retrieved April 23, 2024, from https://www.cdc.gov/dpdx/lymphaticfilariasis/index.html

CDC-DPDx. (n.d.-b). Mansonellosis. Retrieved September 9, 2022, from https://www.cdc.gov/dpdx/mansonellosis/index.html

Celi, L. A., Cellini, J., Charpignon, M.-L., Dee, E. C., Dernoncourt, F., Eber, R., Mitchell, W. G., Moukheiber, L., Schirmer, J., Situ, J., et al. (2022). Sources of bias in artificial intelligence that perpetuate healthcare disparities—a global review. *PLOS Digital Health*, *1*(3), e0000022. https://doi.org/10.1371/journal.pdig.0000022

Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A simple framework for contrastive learning of visual representations. *International conference on machine learning*, 1597–1607. https://doi.org/10.48550/arXiv.2002.05709

Chen, T., Kornblith, S., Swersky, K., Norouzi, M., & Hinton, G. E. (2020). Big self-supervised models are strong semi-supervised learners. *Advances in neural information processing systems*, *33*, 22243–22255. https://doi.org/10.48550/arXiv.2006.10029

Chen, X., & He, K. (2021). Exploring simple siamese representation learning. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 15750–15758.

Collins, J. T., Knapper, J., Stirling, J., Mduda, J., Mkindi, C., Mayagaya, V., Mwakajinga, G. A., Nyakyi, P. T., Sanga, V. L., Carbery, D., et al. (2020). Robotic microscopy for everyone: The openflexure microscope. *Biomedical Optics Express*, *11*(5), 2447–2460. https://doi.org/10.1364/boe.385729

Coulibaly, J. T., Ouattara, M., D'Ambrosio, M. V., Fletcher, D. A., Keiser, J., Utzinger, J., N'Goran, E. K., Andrews, J. R., & Bogoch, I. I. (2016). Accuracy of mobile phone and handheld light microscopy for the diagnosis of schistosomiasis and intestinal protozoa infections in côte d'ivoire. *PLoS neglected tropical diseases*, *10*(6), e0004768. https://doi.org/10.1371/journal.pntd.0004768

Coulibaly, J. T., Silue, K. D., Armstrong, M., de León Derby, M. D., D'Ambrosio, M. V., Fletcher, D. A., Keiser, J., Fisher, K., Andrews, J. R., & Bogoch, I. I. (2023). High sensitivity of mobile phone microscopy screening for schistosoma haematobium in azaguié, côte d'ivoire. *The American journal of tropical medicine and hygiene*, *108*(1), 41. https://doi.org/10.4269/ajtmh.22-0527

Cromwell, E. A., Schmidt, C. A., Kwong, K. T., Pigott, D. M., Mupfasoni, D., Biswas, G., Shirude, S., Hill, E., Donkers, K. M., Abdoli, A., et al. (2020). The global distribution of lymphatic filariasis, 2000–18: A geospatial analysis. *The Lancet Global Health*, *8*(9), e1186–e1194. https://doi.org/10.1016/S2214-109X(20)30286-2

Cybulski, J. S., Clements, J., & Prakash, M. (2014). Foldscope: Origami-based paper microscope. *PLoS ONE*, *9*(6), e98781. https://doi.org/10.1371/journal.pone.0098781

Dacal, E., Bermejo-Peláez, D., Lin, L., Álamo, E., Cuadrado, D., Martínez, Á., Mousa, A., Postigo, M., Soto, A., Sukosd, E., Vladimirov, A., Mwandawiro, C., Gichuki, P., Williams, N. A., Muñoz, J., Kepha, S., & Luengo-Oroz, M. (2021). Mobile microscopy and telemedicine platform assisted by deep learning for quantification of trichuris trichiura infection. (P. Steinmann, Ed.). *PLOS Neglected Tropical Diseases*, *15*(9), e0009677. https://doi.org/10.1371/journal.pntd.0009677

D'Ambrosio, M. V., Bakalar, M., Bennuru, S., Reber, C., Skandarajah, A., Nilsson, L., Switz, N., Kamgno, J., Pion, S., Boussinesq, M., et al. (2015). Point-of-care quantification of blood-borne filarial parasites with a mobile phone microscope. *Science translational medicine*, *7*(286), 286re4–286re4. https://doi.org/10.1126/scitranslmed.aaa3480

Daneshjou, R., Vodrahalli, K., Liang, W., Novoa, R. A., Jenkins, M., Rotemberg, V., Ko, J., Swetter, S. M., Bailey, E. E., Gevaert, O., et al. (2021). Disparities in dermatology ai: Assessments using diverse clinical images. *arXiv preprint arXiv:2111.08006*.

Das, D., Vongpromek, R., Assawariyathipat, T., Srinamon, K., Kennon, K., Stepniewska, K., Ghose, A., Sayeed, A. A., Faiz, M. A., Netto, R. L. A., Siqueira, A., Yerbanga, S. R., Ouédraogo, J. B., Callery, J. J., Peto, T. J., Tripura, R., Koukouikila-Koussounda, F., Ntoumi, F., Ong'echa, J. M., ... Dhorda, M. (2022). Field evaluation of the diagnostic performance of easyscan go: A digital malaria microscopy device based on machine-learning. *Malaria Journal*, *21*(1), 1–12. https://doi.org/10.1186/s12936-022-04146-1

Das, D. K., Ghosh, M., Pal, M., Maiti, A. K., & Chakraborty, C. (2013). Machine learning approach for automated screening of malaria parasite using light microscopic images. *Micron*, *45*, 97–106. https://doi.org/10.1016/j.micron.2012.11.002

Davidson, M. S., Andradi-Brown, C., Yahiya, S., Chmielewski, J., O'Donnell, A. J., Gurung, P., Jeninga, M. D., Prommana, P., Andrew, D. W., Petter, M., et al. (2021). Automated detection and staging of malaria parasites from cytological smears using convolutional neural networks. *Biological imaging*, *1*, e2. https://doi.org/10.1017/S2633903X21000015

Dedhiya, R., Kakileti, S. T., Deepu, G., Gopinath, K., Opoku, N., King, C., & Manjunath, G. (2022). Evaluation of non-invasive thermal imaging for detection of viability of onchocerciasis worms. *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 3518–3521. https://doi.org/10.1109/EMBC48229.2022.9871140

de Korne, C. M., van Lieshout, L., van Leeuwen, F. W. B., & Roestenberg, M. (2023). Imaging as a (pre) clinical tool in parasitology. *Trends in Parasitology*, *39*(3), 212–226. https://doi.org/Imagingasa(pre)clinicaltoolinparasitology

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*. https://doi.org/10.48550/arXiv.2010.11929

El Alaoui, Y., Elomri, A., Qaraqe, M., Padmanabhan, R., Yasin Taha, R., El Omri, H., El Omri, A., & Aboumarzouk, O. (2022). A review of artificial intelligence applications in hematology management: Current practices and future prospects. *Journal of Medical Internet Research*, *24*(7), e36490. https://doi.org/10.2196/36490

Elvana, A., & Suryanto, E. D. (2022). Lymphatic filariasis detection using image analysis. *Proceedings of the 4th International Conference on Innovation in Education, Science and Culture, ICIESC 2022, 11 October 2022, Medan, Indonesia*.

Ephraim, R. K., Duah, E., Cybulski, J. S., Prakash, M., D'Ambrosio, M. V., Fletcher, D. A., Keiser, J., Andrews, J. R., & Bogoch, I. I. (2015). Diagnosis of schistosoma haematobium infection with a mobile phone-mounted foldscope and a reversed-lens cellscope in ghana. *The American journal of tropical medicine and hygiene*, *92*(6), 1253. https://doi.org/10.4269/ajtmh.14-0741

EU. (2022). Artificial intelligence in healthcare [Accessed: 2022-11-28]. https://www.europarl.europa.eu/RegData/etudes/STUD/2022/729512/EPRS_STU(2022)729512_EN.pdf

Fan, B. E., Yong, B. S. J., Ruiqi, L., Wang, S. S. Y., Natalie, A. M. Y., Chia, M. F., Chen, D. T. Y., Neo, Y. S., Occhipinti, B., Wong, W. Y. T., et al. (2023). From microscope to micropixels: A rapid review of artificial intelligence for the peripheral blood film. *Blood Reviews*, 101144. https://doi.org/10.1016/j.blre.2023.101144

FDA. (2021). Artificial intelligence and machine learning in software as a medical device [Accessed: 2024-08-28]. https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device

FDA. (2024). Artificial intelligence and machine learning (ai/ml)-enabled medical devices. https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices

FDAn. (2021). Good machine learning practice for medical device development: Guiding principles [Accessed: 2022-11-28]. https://www.fda.gov/medical-devices/software-

medical-device-samd/good-machine-learning-practice-medical-device-development-guiding-principles

Feasey, N., Wansbrough-Jones, M., Mabey, D. C., & Solomon, A. W. (2010). Neglected tropical diseases. *British medical bulletin*, *93*(1), 179–200. https://doi.org/10.1093/bmb/ldp046

Feroz, A., Jabeen, R., & Saleem, S. (2020). Using mobile phones to improve community health workers performance in low-and-middle-income countries. *BMC public health*, *20*, 1–6. https://doi.org/10.1186/s12889-020-8173-3

Ferreira, M. U., Crainey, J. L., & Gobbi, F. G. (2023). The search for better treatment strategies for mansonellosis: An expert perspective. *Expert Opinion on Pharmacotherapy*, *24*(15), 1685–1692. https://doi.org/10.1080/14656566.2023.2240235

Finlayson, S. G., Bowers, J. D., Ito, J., Zittrain, J. L., Beam, A. L., & Kohane, I. S. (2019). Adversarial attacks on medical machine learning. *Science*, *363*(6433), 1287–1289. https://doi.org/10.1126/science.aaw4399

García-Villena, J., Torres, J. E., Aguilar, C., Lin, L., Bermejo-Peláez, D., Dacal, E., Mousa, A., Ortega, M. d. P., Martínez, A., Vladimirov, A., et al. (2021). 3d-printed portable robotic mobile microscope for remote diagnosis of global health diseases. *Electronics*, *10*(19), 2408. https://doi.org/10.3390/electronics10192408

Gardon, J., Gardon-Wendel, N., Kamgno, J., Chippaux, J.-P., Boussinesq, M., et al. (1997). Serious reactions after mass treatment of onchocerciasis with ivermectin in an area endemic for loa loa infection. *The Lancet*, *350*(9070), 18–22. https://doi.org/10.1016/S0140-6736(96)11094-1

Gaunt, M. W., & Crainey, J. L. (2024). Ai sees an end to filariasis. https://doi.org/10.1371/journal.pntd.0012260

Girshick, R. (2015). Fast r-cnn. *Proceedings of the IEEE international conference on computer vision*, 1440–1448.

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 580–587.

Glied, S., & Sacarny, A. (2018). Is the us health care system wasteful and inefficient? a review of the evidence. *Journal of Health Politics, Policy and Law*, *43*(5), 739–765. https://doi.org/10.1215/03616878-6951103

Gonçalves, C., Borges, A., Dias, V., Marques, J., Aguiar, B., Costa, C., & Silva, R. (2023). Detection of human visceral leishmaniasis parasites in microscopy images from bone marrow parasitological examination. *Applied Sciences*, *13*(14), 8076. https://doi.org/10.3390/app13148076

Gordon, W. J., & Stern, A. D. (2019). Challenges and opportunities in software-driven medical devices. *Nature biomedical engineering*, *3*(7), 493–497. https://doi.org/10.1038/s41551-019-0426-z

Grill, J.-B., Strub, F., Altché, F., Tallec, C., Richemond, P., Buchatskaya, E., Doersch, C., Avila Pires, B., Guo, Z., Gheshlaghi Azar, M., et al. (2020). Bootstrap your own latent-a new approach to self-supervised learning. *Advances in neural information processing systems*, *33*, 21271–21284. https://doi.org/10.48550/arXiv.2006.07733

Guo, J., & Li, B. (2018). The application of medical artificial intelligence technology in rural areas of developing countries. *Health equity*, *2*(1), 174–181. https://doi.org/10.1089/heq.2018.0037

Gyapong, M., Immurana, M., Manyeh, A., Odopey, C. T., Dean, L., & Krentel, A. (2024). The social and economic impact of neglected tropical diseases in sub-saharan africa. *Neglected tropical diseases-sub-saharan africa* (pp. 479–503). Springer. https://doi.org/10.1007/978-3-319-25471-5_15

Holmström, O., Linder, N., Ngasala, B., Mårtensson, A., Linder, E., Lundin, M., Moilanen, H., Suutala, A., Diwan, V., & Lundin, J. (2017). Point-of-care mobile digital microscopy and deep learning for the detection of soil-transmitted helminths and schistosoma haematobium. *Global health action*, *10*(sup3), 1337325. https://doi.org/10.1080/16549716.2017.1337325

Horning, M. P., Delahunt, C. B., Bachman, C. M., Luchavez, J., Luna, C., Hu, L., Jaiswal, M. S., Thompson, C. M., Kulhare, S., Janko, S., et al. (2021). Performance of a fully-automated system on a who malaria microscopy evaluation slide set. *Malaria journal*, *20*, 1–11. https://doi.org/10.1186/s12936-021-03631-3

Hotez, P. J., Aksoy, S., Brindley, P. J., & Kamhawi, S. (2020). What constitutes a neglected tropical disease? *PLoS neglected tropical diseases*, *14*(1), e0008001. https://doi.org/10.1371/journal.pntd.0008001

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*. https://doi.org/10.48550/arXiv.1704.04861

Huang, S.-C., Pareek, A., Jensen, M., Lungren, M. P., Yeung, S., & Chaudhari, A. S. (2023). Self-supervised learning for medical image classification: A systematic review and implementation guidelines. *NPJ Digital Medicine*, *6*(1), 74. https://doi.org/10.1038/s41746-023-00811-0

IMDRF. (2013). Software as a medical device (samd): Key definitions [Accessed: 2022-11-28]. *International Medical Device Regulators Forum*. https://www.imdrf.org/sites/default/files/docs/imdrf/final/technical/imdrf-tech-131209-samd-key-definitions-140901.pdf

Inácio, S. V., Ferreira Gomes, J., Xavier Falcão, A., Nagase Suzuki, C. T., Bertequini Nagata, W., Nery Loiola, S. H., Martins dos Santos, B., Soares, F. A., Rosa, S. L., Baptista, C. B., Borges Alves, G., & Saraiva Bresciani, K. D. (2020). Automated diagnosis of canine gastrointestinal parasites using image analysis. *Pathogens*, *9*(2), 139. https://doi.org/10.3390/pathogens9020139

Isaza-Jaimes, A., Bermúdez, V., Bravo, A., Castrillo, J. S., Lalinde, J. D. H., Fossi, C. A., Flórez, A., & Rodrıǵuez, J. E. (2020). A computational approach for leishmania genus protozoa detection in bone marrow samples from patients with visceral leishmaniasis. *AVFT–Archivos Venezolanos de Farmacologıá y Terapéutica*, *39*(7).

Jacobsen, K. H., Andress, B. C., Bhagwat, E. A., Bryant, C. A., Chandrapu, V. R., Desmonts, C. G., Matthews, T. M., Ogunkoya, A., Wheeler, T. J., & A'kayla, S. W. (2022). A call for loiasis to be added to the who list of neglected tropical diseases. *The Lancet Infectious Diseases*, *22*(10), e299–e302. https://doi.org/10.1016/S1473-3099(22)00064-0

James, S. L., Abate, D., Abate, K. H., Abay, S. M., Abbafati, C., Abbasi, N., Abbastabar, H., Abd-Allah, F., Abdela, J., Abdelalim, A., et al. (2018). Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: A systematic analysis for the global burden of

disease study 2017. *The Lancet*, *392*(10159), 1789–1858. https://doi.org/10.1016/S0140-6736(18)32279-7

Jiménez, B., Maya, C., Velásquez, G., Torner, F., Arambula, F., Barrios, J. A., & Velasco, M. (2016). Identification and quantification of pathogenic helminth eggs using a digital image system. *Experimental Parasitology*, *166*, 164–172. https://doi.org/10.1016/j.exppara.2016.04.016

Jiménez, B., Maya, C., Velásquez, G., Barrios, J. A., Pérez, M., & Román, A. (2020). Helminth egg automatic detector (head): Improvements in development for digital identification and quantification of helminth eggs and its application online. *MethodsX*, *7*, 101158. https://doi.org/10.1016/j.exppara.2020.107959

Jocher, G., Chaurasia, A., & Qiu, J. (2023). *Ultralytics yolov8* (Version 8.0.0). https://github.com/ultralytics/ultralytics

Joshi, G., Jain, A., Araveeti, S. R., Adhikari, S., Garg, H., & Bhandari, M. (2024). Fda-approved artificial intelligence and machine learning (ai/ml)-enabled medical devices: An updated landscape. *Electronics*, *13*(3), 498. https://doi.org/10.3390/electronics13030498

Jung, T., Anzaku, E. T., Özbulak, U., Magez, S., Van Messem, A., & De Neve, W. (2021). Automatic detection of trypanosomosis in thick blood smears using image pre-processing and deep learning. *Intelligent Human Computer Interaction: 12th International Conference, IHCI 2020, Daegu, South Korea, November 24–26, 2020, Proceedings, Part II 12*, 254–266.

Kamgno, J., Pion, S. D., Chesnais, C. B., Bakalar, M. H., D'Ambrosio, M. V., Mackenzie, C. D., Nana-Djeunga, H. C., Gounoue-Kamkumo, R., Njitchouang, G.-R., Nwane, P., Tchatchueng-Mbouga, J. B., Wanji, S., Stolk, W. A., Fletcher, D. A., Klion, A. D., Nutman, T. B., & Boussinesq, M. (2017). A test-and-not-treat strategy for onchocerciasis in loa loa –endemic areas. *New England Journal of Medicine*, *377*(21), 2044–2052. https://doi.org/10.1056/nejmoa1705026

Karimi, D., Dou, H., Warfield, S. K., & Gholipour, A. (2020). Deep learning with noisy labels: Exploring techniques and remedies in medical image analysis. *Medical Image Analysis*, *65*, 101759. https://doi.org/10.1016/j.media.2020.101759

Katz, N., Chaves, A., & Pellegrino, J. (1972). A simple, device for quantita tive stool thick-smear technique in schistosomiasis mansoni. *Revista do Instituto de medicina tropical de Sao Paulo*, *14*(6), 397–400.

Keshavan, A., Yeatman, J. D., & Rokem, A. (2019). Combining citizen science and deep learning to amplify expertise in neuroimaging. *Frontiers in Neuroinformatics*, *13*, 29. https://doi.org/10.3389/fninf.2019.00029

Kim, Y.-G., Choi, G., Go, H., Cho, Y., Lee, H., Lee, A.-R., Park, B., & Kim, N. (2019). A fully automated system using a convolutional neural network to predict renal allograft rejection: Extra-validation with giga-pixel immunostained slides. *Scientific reports*, *9*(1), 5123. https://doi.org/10.1038/s41598-019-41479-5

Kisantal, M., Wojna, Z., Murawski, J., Naruniec, J., & Cho, K. (2019). Augmentation for small object detection. *arXiv preprint arXiv:1902.07296*.

Kitvimonrat, A., Hongcharoen, N., Marukatat, S., & Watcharabutsarakham, S. (2020). Automatic detection and characterization of parasite eggs using deep learning methods. *2020 17th International Conference on Electrical Engineering/Electronics, Computer,*

Telecommunications and Information Technology (ECTI-CON), 153–156. https://doi.org/10.1109/ECTI-CON49241.2020.9158084

Klarmann-Schulz, U., Kuehlwein, D. A., Heine, J., Kuehlwein, J. M., Dubben, B., Chorowski, J., Debrah, A. Y., & Hoerauf, A. (2020). A proof of concept for automated histological analysis of onchocerciasis with deep learning (ai).

Koch, V., Wagner, S. J., Kazeminia, S., Sancar, E., Hehr, M., Schnabel, J., Peng, T., & Marr, C. (2024). Dinobloom: A foundation model for generalizable cell embeddings in hematology. *arXiv preprint arXiv:2404.05022*.

Koplan, J. P., Bond, T. C., Merson, M. H., Reddy, K. S., Rodriguez, M. H., Sewankambo, N. K., & Wasserheit, J. N. (2009). Towards a common definition of global health. *The Lancet*, *373*(9679), 1993–1995. https://doi.org/10.1016/S0140-6736(09)60332-9

Kovashka, A., Russakovsky, O., Fei-Fei, L., & Grauman, K. (2016). Crowdsourcing in computer vision. *Foundations and Trends in Computer Graphics and Vision*, *10*(3), 177–243. https://doi.org/10.1561/0600000071

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In F. Pereira, C. Burges, L. Bottou, & K. Weinberger (Eds.), *Advances in neural information processing systems*. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf

Lacoste, A., Luccioni, A., Schmidt, V., & Dandres, T. (2019). Quantifying the carbon emissions of machine learning. *arXiv preprint arXiv:1910.09700*.

Larson, D. B., Harvey, H., Rubin, D. L., Irani, N., Tse, J. R., & Langlotz, C. P. (2021). Regulatory frameworks for development and evaluation of artificial intelligence–based diagnostic imaging algorithms: Summary and recommendations. *Journal of the American College of Radiology*, *18*(3), 413–424. https://doi.org/10.1016/j.jacr.2020.09.060

Lee, C. S., & Lee, A. Y. (2020). Clinical applications of continual learning machine learning. *The Lancet Digital Health*, *2*(6), e279–e281. https://doi.org/10.1016/S2589-7500(20)30102-3

Li, Q., Li, S., Liu, X., He, Z., Wang, T., Xu, Y., Guan, H., Chen, R., Qi, S., & Wang, F. (2020). Fecalnet: Automated detection of visible components in human feces using deep learning. *Medical Physics*, *47*(9), 4212–4222. https://doi.org/10.1002/mp.14352

Liang, W., Tadesse, G. A., Ho, D., Fei-Fei, L., Zaharia, M., Zhang, C., & Zou, J. (2022). Advances, challenges and opportunities in creating data for trustworthy ai. *Nature Machine Intelligence*, *4*(8), 669–677. https://doi.org/10.1038/s42256-022-00516-1

Lima, N. F., Veggiani Aybar, C. A., Dantur Juri, M. J., & Ferreira, M. U. (2016). Mansonella ozzardi: A neglected new world filarial nematode. *Pathogens and Global Health*, *110*(3), 97–107. https://doi.org/10.1080/20477724.2016.1190544

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2117–2125.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. *Proceedings of the IEEE international conference on computer vision*, 2980–2988.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft coco: Common objects in context. *Computer Vision–ECCV*

2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13, 740–755.

Linares, M., Postigo, M., Cuadrado, D., Ortiz-Ruiz, A., Gil-Casanova, S., Vladimirov, A., Garcıá-Villena, J., Nuñez-Escobedo, J. M., Martıńez-López, J., Rubio, J. M., et al. (2019). Collaborative intelligence and gamification for on-line malaria species differentiation. *Malaria Journal*, *18*, 1–9. https://doi.org/10.1186/s12936-019-2662-9

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., & Berg, A. C. (2016). Ssd: Single shot multibox detector. *European conference on computer vision*, 21–37.

Luengo-Oroz, M. A., Arranz, A., & Frean, J. (2012). Crowdsourcing malaria parasite quantification: An online game for analyzing images of infected thick blood smears. *Journal of Medical Internet Research*, *14*(6), 1–14. https://doi.org/10.2196/jmir.2338

Lundin, J., Suutala, A., Holmström, O., Henriksson, S., Valkamo, S., Kaingu, H., Kinyua, F., Muinde, M., Lundin, M., Diwan, V., et al. (2024). Diagnosis of soil-transmitted helminth infections with digital mobile microscopy and artificial intelligence in a resource-limited setting. *PLOS Neglected Tropical Diseases*, *18*(4), e0012041. https://doi.org/10.1371/journal.pntd.0012041

Ma, J., He, Y., Li, F., Han, L., You, C., & Wang, B. (2024). Segment anything in medical images. *Nature Communications*, *15*(1), 654. https://doi.org/10.1038/s41467-024-44824-z

Ma, X., Niu, Y., Gu, L., Wang, Y., Zhao, Y., Bailey, J., & Lu, F. (2021). Understanding adversarial attacks on deep learning based medical image analysis systems. *Pattern Recognition*, *110*, 107332. https://doi.org/10.3390/jcm12093266

Manku, R. R., Sharma, A., & Panchbhai, A. (2020). Malaria detection and classificaiton. *arXiv preprint arXiv:2011.14329*.

Masud, M., Alhumyani, H., Alshamrani, S. S., Cheikhrouhou, O., Ibrahim, S., Muhammad, G., Hossain, M. S., & Shorfuzzaman, M. (2020). Leveraging deep learning techniques for malaria parasite detection using mobile application. *Wireless Communications and Mobile Computing*, *2020*(1), 8895429. https://doi.org/10.1155/2020/8895429

Matek, C., Schwarz, S., Spiekermann, K., & Marr, C. (2019). Human-level recognition of blast cells in acute myeloid leukaemia with convolutional neural networks. *Nature Machine Intelligence*, *1*(11), 538–544. https://doi.org/10.1038/s42256-019-0101-9

Mathison, B. A., Couturier, M. R., & Pritt, B. S. (2019). Diagnostic identification and differentiation of microfilariae. *Journal of clinical microbiology*, *57*(10), 10–1128. https://doi.org/10.1128/JCM.00706-19

Mathison, B. A., Kohan, J. L., Walker, J. F., Smith, R. B., Ardon, O., Ardon, O., Couturier, M. R., & Couturier, M. R. (2020). Detection of intestinal protozoa in trichrome-stained stool specimens by use of a deep convolutional neural network. *Journal of Clinical Microbiology*, *58*(6), 1–13. https://doi.org/10.1128/JCM.02053-19

Maturana, C. R., De Oliveira, A. D., Nadal, S., Bilalli, B., Serrat, F. Z., Soley, M. E., Igual, E. S., Bosch, M., Lluch, A. V., Abelló, A., et al. (2022). Advances and challenges in automated malaria diagnosis using digital microscopy imaging with artificial intelligence tools: A review. *Frontiers in microbiology*, *13*, 1006659. https://doi.org/10.3389/fmicb.2022.1006659

Maturana, C. R., de Oliveira, A. D., Nadal, S., Serrat, F. Z., Sulleiro, E., Ruiz, E., Bilalli, B., Veiga, A., Espasa, M., Abelló, A., et al. (2023). Imaging: A novel automated system for malaria diagnosis by using artificial intelligence tools and a universal low-cost

robotized microscope. *Frontiers in microbiology*, *14*, 1240936. https://doi.org/10.3389/fmicb.2023.1240936

McCool, J., Dobson, R., Whittaker, R., & Paton, C. (2022). Mobile health (mhealth) in low-and middle-income countries. *Annual Review of Public Health*, *43*, 525–539. https://doi.org/10.1146/annurev-publhealth-052620-093850

Mediannikov, O., & Ranque, S. (2018). Mansonellosis, the most neglected human filariasis. *New microbes and new infections*, *26*, S19–S22. https://doi.org/10.1016/j.nmni.2018.08.016

Mehanian, C., Jaiswal, M., Delahunt, C., Thompson, C., Horning, M., Hu, L., McGuire, S., Ostbye, T., Mehanian, M., Wilson, B., Champlin, C., Long, E., Proux, S., Gamboa, D., Chiodini, P., Carter, J., Dhorda, M., Isaboke, D., Ogutu, B., ... Bell, D. (2018). Computer-automated malaria diagnosis and quantitation using convolutional neural networks. *Proceedings - 2017 IEEE International Conference on Computer Vision Workshops, ICCVW 2017*, *2018-Janua*, 116–125. https://doi.org/10.1109/ICCVW.2017.22

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM computing surveys (CSUR)*, *54*(6), 1–35.

Mesko, B., Benjamens, S., & Dhunnoo, P. (2020). The state of artificial intelligence-based fda-approved medical devices and algorithms: An online database. *Journal of Medical Internet Research*, *3*(1), 118. https://doi.org/10.1038/s41746-020-00324-0

Metzger, W. G., & Mordmüller, B. (2014). Loa loa—does it deserve to be neglected? *The Lancet Infectious Diseases*, *14*(4), 353–357. https://doi.org/10.1016/S1473-3099(13)70263-9

Meulah, B., Bengtson, M., Van Lieshout, L., Hokke, C. H., Kreidenweiss, A., Diehl, J.-C., Adegnika, A. A., & Agbana, T. E. (2023). A review on innovative optical devices for the diagnosis of human soil-transmitted helminthiasis and schistosomiasis: From research and development to commercialization. *Parasitology*, *150*(2), 137–149. https://doi.org/10.1017/S0031182022001664

Meulah, B., Oyibo, P., Bengtson, M., Agbana, T., Lontchi, R. A. L., Adegnika, A. A., Oyibo, W., Hokke, C. H., Diehl, J. C., & van Lieshout, L. (2022). Performance evaluation of the schistoscope 5.0 for (semi-) automated digital detection and quantification of schistosoma haematobium eggs in urine: A field-based study in nigeria. *The American Journal of Tropical Medicine and Hygiene*, *107*(5), 1047. https://doi.org/10.4269/ajtmh.22-0276

Meulah, B., Oyibo, P., Hoekstra, P. T., Moure, P. A. N., Maloum, M. N., Laclong-Lontchi, R. A., Honkpehedji, Y. J., Bengtson, M., Hokke, C., Corstjens, P. L., et al. (2024). Validation of artificial intelligence-based digital microscopy for automated detection of schistosoma haematobium eggs in urine in gabon. *PLoS neglected tropical diseases*, *18*(2), e0011967. https://doi.org/10.1371/journal.pntd.0011967

Moor, M., Banerjee, O., Abad, Z. S. H., Krumholz, H. M., Leskovec, J., Topol, E. J., & Rajpurkar, P. (2023). Foundation models for generalist medical artificial intelligence. *Nature*, *616*(7956), 259–265. https://doi.org/10.1038/s41586-023-05881-4

Morais, M. C. C., Silva, D., Milagre, M. M., de Oliveira, M. T., Pereira, T., Silva, J. S., Costa, L. d. F., Minoprio, P., Junior, R. M. C., Gazzinelli, R., et al. (2022). Automatic detection of the parasite trypanosoma cruzi in blood smears using a machine learning approach applied to mobile phone images. *PeerJ*, *10*, e13470. https://doi.org/10.7717/peerj.13470

Moya, L., Herrador, Z., Ta-Tang, T. H., Rubio, J. M., Perteguer, M. J., Hernandez-Gonzalez, A., Garcia, B., Nguema, R., Nguema, J., Ncogo, P., et al. (2016). Evidence for suppression of onchocerciasis transmission in bioko island, equatorial guinea. *PLoS Neglected Tropical Diseases*, *10*(7), e0004829. https://doi.org/10.1371/journal.pntd.0004829

Muehlematter, U. J., Daniore, P., & Vokinger, K. N. (2021). Approval of artificial intelligence and machine learning-based medical devices in the usa and europe (2015–20): A comparative analysis. *The Lancet Digital Health*, *3*(3), e195–e203. https://doi.org/10.1016/S2589-7500(20)30292-2

Naing, K. M., Boonsang, S., Chuwongin, S., Kittichai, V., Tongloy, T., Prommongkol, S., Dekumyoy, P., & Watthanakulpanich, D. (2022). Automatic recognition of parasitic products in stool examination using object detection approach. *PeerJ Computer Science*, *8*, e1065. https://doi.org/10.7717/peerj-cs.1065

Ngwese, M. M., Ngwese, M. M., Manouana, G. P., Manouana, G. P., Moure, P. A. N., Ramharter, M., Ramharter, M., Esen, M., Adégnika, A. A., Adégnika, A. A., & Adégnika, A. A. (2020). Diagnostic techniques of soil-transmitted helminths: Impact on control measures. https://doi.org/10.3390/tropicalmed5020093

Niazi, M. K. K., Parwani, A. V., & Gurcan, M. N. (2019). Digital pathology and artificial intelligence. *The lancet oncology*, *20*(5), e253–e261. https://doi.org/10.1016/S1470-2045(19)30154-8

Nikolay, B., Brooker, S. J., & Pullan, R. L. (2014). Sensitivity of diagnostic tests for human soil-transmitted helminth infections: A meta-analysis in the absence of a true gold standard. *International journal for parasitology*, *44*(11), 765–774. https://doi.org/10.1016/j.ijpara.2014.05.009

Nkengasong, J. N., Yao, K., & Onyebujoh, P. (2018). Laboratory medicine in low-income and middle-income countries: Progress and challenges. *The Lancet*, *391*(10133), 1873–1875. https://doi.org/10.1016/S0140-6736(18)30308-8

OECD. (2017). *Tackling wasteful spending on health*. OECD Publishing. https://doi.org/10.1787/9789264266414-en

OECD. (2020). *Health at a glance: Europe 2020*. OECD Publishing Paris. https://doi.org/10.1787/82129230-en

Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., et al. (2023). Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*. https://doi.org/10.48550/arXiv.2304.07193

Oyibo, P., Agbana, T., van Lieshout, L., Oyibo, W., Diehl, J.-C., & Vdovine, G. (2024). An automated slide scanning system for membrane filter imaging in diagnosis of urogenital schistosomiasis. *Journal of Microscopy*, *294*(1), 52–61. https://doi.org/10.1111/jmi.13269

Oyibo, P., Jujjavarapu, S., Meulah, B., Agbana, T., Braakman, I., van Diepen, A., Bengtson, M., van Lieshout, L., Oyibo, W., Vdovine, G., et al. (2022). Schistoscope: An automated microscope with artificial intelligence for detection of schistosoma haematobium eggs in resource-limited settings. *Micromachines*, *13*(5), 643. https://doi.org/10.3390/mi13050643

Pattanaik, P. A., Mittal, M., Khan, M. Z., & Panda, S. N. (2022). Malaria detection using deep residual networks with mobile microscopy. *Journal of King Saud University-Computer*

and Information Sciences, *34*(5), 1700–1705. https://doi.org/10.1016/j.jksuci.2020.07.003

Pereira, A., Pyrrho, A., Vanzan, D., Mazza, L., & Gomes, J. G. (2020). Deep convolutional neural network applied to chagas disease parasitemia assessment. https://doi.org/10.21528/cbic2019-119

Petti, C. A., Polage, C. R., Quinn, T. C., Ronald, A. R., & Sande, M. A. (2006). Laboratory medicine in africa: A barrier to effective health care. *Clinical Infectious Diseases*, *42*(3), 377–382. https://doi.org/10.1086/499363

Pfeil, J., Nechyporenko, A., Frohme, M., Hufert, F. T., & Schulze, K. (2022). Examination of blood samples using deep learning and mobile microscopy. *BMC bioinformatics*, *23*(1), 65. https://doi.org/10.1186/s12859-022-04602-4

Pion, S. D., Nana-Djeunga, H., Niamsi-Emalio, Y., Chesnais, C. B., Deléglise, H., Mackenzie, C., Stolk, W., Fletcher, D. A., Klion, A. D., Nutman, T. B., et al. (2020). Implications for annual retesting after a test-and-not-treat strategy for onchocerciasis elimination in areas co-endemic with loa loa infection: An observational cohort study. *The Lancet Infectious Diseases*, *20*(1), 102–109. https://doi.org/10.1016/S1473-3099(19)30554-7

Quinn, J. A., Nakasi, R., Mugagga, P. K., Byanyima, P., Lubega, W., & Andama, A. (2016). Deep convolutional neural networks for microscopy-based point of care diagnostics. *Machine learning for healthcare conference*, 271–281.

Raccurt, C. P., Brasseur, P., & Boncy, J. (2014). Mansonelliasis, a neglected parasitic disease in haiti. *Memórias do Instituto Oswaldo Cruz*, *109*, 709–711. https://doi.org/10.1590/0074-0276140107

Rajaraman, S., Antani, S. K., Poostchi, M., Silamut, K., Hossain, M. A., Maude, R. J., Jaeger, S., & Thoma, G. R. (2018). Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images. *PeerJ*, *2018*(4), e4568. https://doi.org/10.7717/peerj.4568

Rajaraman, S., Jaeger, S., & Antani, S. K. (2019). Performance evaluation of deep neural ensembles toward malaria parasite detection in thin-blood smear images. *PeerJ*, *7*, e6977. https://doi.org/10.7717/peerj.6977

Rajaraman, S., Silamut, K., Hossain, M. A., Ersoy, I., Maude, R. J., Jaeger, S., Thoma, G. R., & Antani, S. K. (2018). Understanding the learned behavior of customized convolutional neural networks toward malaria parasite detection in thin blood smear images. *Journal of Medical Imaging*, *5*(03), 1. https://doi.org/10.1117/1.jmi.5.3.034501

Rajpurkar, P., Chen, E., Banerjee, O., & Topol, E. J. (2022). Ai in health and medicine. *Nature medicine*, *28*(1), 31–38. https://doi.org/0.1038/s41591-021-01614-0

Rajpurkar, P., & Lungren, M. P. (2023). The current and future state of ai interpretation of medical images. *New England Journal of Medicine*, *388*(21), 1981–1990. https://doi.org/10.1056/NEJMra2301725

Ramharter, M., Butler, J., Mombo-Ngoma, G., Nordmann, T., Davi, S. D., & Manego, R. Z. (2024). The african eye worm: Current understanding of the epidemiology, clinical disease, and treatment of loiasis. *The Lancet Infectious Diseases*, *24*(3), e165–e178. https://doi.org/10.1016/S1473-3099(23)00438-3

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 779–788. https://doi.org/10.48550/arXiv.1506.02640

Redmon, J., & Farhadi, A. (2017). Yolo9000: Better, faster, stronger. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 7263–7271. https://doi.org/10.48550/arXiv.1612.08242

Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*. https://doi.org/10.48550/arXiv.1804.02767

Reed, S. E., Lee, H., Anguelov, D., Szegedy, C., Erhan, D., & Rabinovich, A. (2015). Training deep neural networks on noisy labels with bootstrapping. *3rd International Conference on Learning Representations, ICLR 2015 - Workshop Track Proceedings*, 1–11.

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, *28*.

Ritchie, L. S. et al. (1948). An ether sedimentation technique for routine stool examinations. *Bulletin of the United States Army medical department*, *8*(4).

Roser, M. et al. (2024). Ensure healthy lives and promote well-being for all at all ages. *Our World in Data*.

Rösler, W., Altenbuchinger, M., Baeßler, B., Beissbarth, T., Beutel, G., Bock, R., von Bubnoff, N., Eckardt, J.-N., Foersch, S., Loeffler, C. M., et al. (2023). An overview and a roadmap for artificial intelligence in hematology and oncology. *Journal of cancer research and clinical oncology*, *149*(10), 7997–8006. https://doi.org/10.1007/s00432-023-04667-5

Ross, A., & Doshi-Velez, F. (2018). Improving the adversarial robustness and interpretability of deep neural networks by regularizing their input gradients. *Proceedings of the AAAI Conference on Artificial Intelligence*, *32*.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al. (2015). Imagenet large scale visual recognition challenge. *International journal of computer vision*, *115*(3), 211–252.

Russakovsky, O., Li, L.-J., & Fei-Fei, L. (2015). Best of both worlds: Human-machine collaboration for object annotation. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2121–2131.

Sadeghi, A., Sadeghi, M., Fakhar, M., Zakariaei, Z., Sadeghi, M., & Bastani, R. (2024). A deep learning-based model for detecting leishmania amastigotes in microscopic slides: A new approach to telemedicine. *BMC Infectious Diseases*, *24*(1), 551. https://doi.org/10.1186/s12879-024-09428-4

Saeed, M. A., & Jabbar, A. (2018). "smart diagnosis" of parasitic diseases by use of smartphones. *Journal of clinical microbiology*, *56*(1), e01469–17. https://doi.org/10.1128/JCM.01469-17

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4510–4520. https://doi.org/10.48550/arXiv.1801.04381

Sastry, A. S., & Bhat, S. (2018). *Essentials of medical parasitology*. Jp Medical Ltd.

Schlosser-Brandenburg, J., Midha, A., Mugo, R. M., Ndombi, E. M., Gachara, G., Njomo, D., Rausch, S., & Hartmann, S. (2023). Infection with soil-transmitted helminths and their impact on coinfections. *Frontiers in Parasitology*, *2*, 1197956. https://doi.org/10.3389/fpara.2023.1197956

Schwalbe, N., & Wahl, B. (2020). Artificial intelligence and the future of global health. *The Lancet*, *395*(10236), 1579–1586. https://doi.org/10.1016/S0140-6736(20)30226-9

Sellergren, A. B., Chen, C., Nabulsi, Z., Li, Y., Maschinot, A., Sarna, A., Huang, J., Lau, C., Kalidindi, S. R., Etemadi, M., et al. (2022). Simplified transfer learning for chest radiography models using less data. *Radiology*, *305*(2), 454–465. https://doi.org/10.1148/radiol.212482

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2016). Grad-cam: Visual explanations from deep networks via gradient-based localization. *International Journal of Computer Vision*, *128*(2), 336–359. https://doi.org/10.1007/s11263-019-01228-7

Shrestha, R., Duwal, R., Wagle, S., Pokhrel, S., Giri, B., & Neupane, B. B. (2020). A smartphone microscopic method for simultaneous detection of (oo) cysts of cryptosporidium and giardia. *PLoS Neglected Tropical Diseases*, *14*(9), e0008560. https://doi.org/10.1371/journal.pntd.0008560

Simonsen, P. E., Onapa, A. W., & Asio, S. M. (2011). Mansonella perstans filariasis in africa. *Acta tropica*, *120*, S109–S120. https://doi.org/10.1016/j.actatropica.2010.01.014

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Skandarajah, A., Reber, C. D., Switz, N. A., & Fletcher, D. A. (2014). Quantitative imaging with a mobile phone microscope. *PLoS ONE*, *9*(5), e96906. https://doi.org/10.1371/journal.pone.0096906

Soberanis-Mukul, R., Uc-Cetina, V., Brito-Loeza, C., & Ruiz-Piña, H. (2013). An automatic algorithm for the detection of trypanosoma cruzi parasites in blood sample images. *Computer Methods and Programs in Biomedicine*, *112*(3), 633–639. https://doi.org/10.1016/j.cmpb.2013.07.013

Spotlab. (n.d.-a). Ai copilot for neglected tropical diseases microscopy diagnosis with limited connectivity: Video 1. https://youtu.be/dqEKL5HMK6s?si=aK0rj4YtPFIOV-3r

Spotlab. (n.d.-b). Ai copilot for neglected tropical diseases microscopy diagnosis with limited connectivity: Video 2. https://youtu.be/YjXL5FBacA8?si=4Xr7H0rEouTO9TyC

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1–9.

Ta-Tang, T.-H., Crainey, J. L., Post, R. J., Luz, S. L., & Rubio, J. M. (2018). Mansonellosis: Current perspectives. *Research and reports in tropical medicine*, *9*, 9. https://doi.org/10.2147/RRTM.S125750

Ta-Tang, T.-H., Luz, S. L., Crainey, J. L., & Rubio, J. M. (2021). An overview of the management of mansonellosis. *Research and reports in tropical medicine*, 93–105. https://doi.org/10.2147/RRTM.S274684

tensorflow. (n.d.). Tensorflow object detection api. Retrieved May 29, 2024, from https://github.com/tensorflow/models/tree/master/research/object_detection

Tong, K., Wu, Y., & Zhou, F. (2020). Recent advances in small object detection based on deep learning: A review. *Image and Vision Computing*, *97*, 103910. https://doi.org/10.1016/j.imavis.2020.103910

Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature medicine*, *25*(1), 44–56. https://doi.org/10.1038/s41591-018-0300-7

Turner, H. C., Bettis, A. A., Dunn, J. C., Whitton, J. M., Hollingsworth, T. D., Fleming, F. M., & Anderson, R. M. (2017). Economic considerations for moving beyond the

kato-katz technique for diagnosing intestinal parasites as we move towards elimination. *Trends in parasitology*, *33*(6), 435–443. https://doi.org/10.1016/j.pt.2017.01.007

Uc-Cetina, V., Brito-Loeza, C., & Ruiz-Piña, H. (2013). Chagas parasites detection through gaussian discriminant analysis.

UN. (2015). Transforming our world: The 2030 agenda for sustainable development. Retrieved April 6, 2021, from https://sdgs.un.org/2030agenda

USAID. (2019). *Artificial intelligence in global health. defining a collective path forward* (tech. rep.). USAID CII, The Rockefeller foundation. www.usaid.gov/cii

Utzinger, J., Rinaldi, L., Lohourignon, L. K., Rohner, F., Zimmermann, M. B., Tschannen, A. B., N'Goran, E. K., & Cringoli, G. (2008). Flotac: A new sensitive technique for the diagnosis of hookworm infections in humans. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *102*(1), 84–90. https://doi.org/10.1016/j.trstmh.2007.09.009

Vasiman, A., Stothard, J. R., & Bogoch, I. I. (2019). Mobile phone devices and handheld microscopes as diagnostic platforms for malaria and neglected tropical diseases (ntds) in low-resource settings: A systematic review, historical perspective and future outlook. *Advances in Parasitology*, *103*, 151–173. https://doi.org/10.1016/bs.apar.2018.09.001

Vijayalakshmi A, & Rajesh Kanna B. (2020). Deep learning approach to detect malaria from microscopic images. *Multimedia Tools and Applications*, *79*(21), 15297–15317. https://doi.org/10.1007/s11042-019-7162-y

Vinkeles Melchers, N. V., Coffeng, L. E., Boussinesq, M., Pedrique, B., Pion, S. D., Tekle, A. H., Zouré, H. G., Wanji, S., Remme, J. H., & Stolk, W. A. (2020). Projected number of people with onchocerciasis–loiasis coinfection in africa, 1995 to 2025. *Clinical infectious diseases*, *70*(11), 2281–2289. https://doi.org/10.1093/cid/ciz647

Vorontsov, E., Bozkurt, A., Casson, A., Shaikovski, G., Zelechowski, M., Liu, S., Mathieu, P., van Eck, A., Lee, D., Viret, J., et al. (2023). Virchow: A million-slide digital pathology foundation model. *arXiv preprint arXiv:2309.07778*.

Vorontsov, E., Bozkurt, A., Casson, A., Shaikovski, G., Zelechowski, M., Severson, K., Zimmermann, E., Hall, J., Tenenholtz, N., Fusi, N., et al. (2024). A foundation model for clinical-grade computational pathology and rare cancers detection. *Nature Medicine*, 1–12. https://doi.org/10.1038/s41591-024-03141-0

Wahl, B., Cossy-Gantner, A., Germann, S., & Schwalbe, N. R. (2018). Artificial intelligence (ai) and global health: How can ai contribute to health in resource-poor settings? *BMJ global health*, *3*(4), e000798. https://doi.org/10.1136/bmjgh-2018-000798

Wang, C.-Y., Yeh, I.-H., & Liao, H.-Y. M. (2024). Yolov9: Learning what you want to learn using programmable gradient information. *arXiv preprint arXiv:2402.13616*.

Wang, Y., Ma, X., Bailey, J., Yi, J., Zhou, B., & Gu, Q. (2021). On the convergence and robustness of adversarial training. *arXiv preprint arXiv:2112.08304*.

Ward, P., Dahlberg, P., Lagatie, O., Larsson, J., Tynong, A., Vlaminck, J., Zumpe, M., Ame, S., Ayana, M., Khieu, V., et al. (2022). Affordable artificial intelligence-based digital pathology for neglected tropical diseases: A proof-of-concept for the detection of soil-transmitted helminths and schistosoma mansoni eggs in kato-katz stool thick smears. *PLoS neglected tropical diseases*, *16*(6), e0010500. https://doi.org/10.1371/journal.pntd.0010500

Ward, P. K., Roose, S., Ayana, M., Broadfield, L. A., Dahlberg, P., Kabatereine, N., Kazienga, A., Mekonnen, Z., Nabatte, B., Stuyver, L., et al. (2023). A comprehensive evaluation of an artificial intelligence based digital pathology to monitor large-scale deworming programs against soil-transmitted helminths: A study protocol. *medRxiv*, 2023–09. https://doi.org/10.1101/2023.09.28.23296266

WHO. (1997). *Bench aids for the diagnosis of filarial infections*. World Health Organization.

WHO et al. (2010). *Working to overcome the global impact of neglected tropical diseases: First who report on neglected tropical diseases*. World Health Organization.

WHO. (2016). *Global strategy on human resources for health: Workforce 2030*. World Health Organization. https://www.who.int/hrh/resources/global_strategy_workforce2030_14_print.pdf?ua=1

WHO. (2019). *Bench aids for the diagnosis of intestinal parasites*. World Health Organization.

WHO. (2020). *Ending the neglect to attain the sustainable development goals – a road map for neglected tropical diseases 2021–2030*. World Health Organization.

WHO et al. (2021a). *Diagnostic target product profile for monitoring and evaluation of soil-transmitted helminth control programmes*. World Health Organization.

WHO. (2021). *Diagnostic test for lymphatic filariasis to support decisions for stopping triple-therapy mass drug administration: Target product profile*. World Health Organization.

WHO et al. (2021b). *Diagnostic test for surveillance of lymphatic filariasis: Target product profile*. World Health Organization.

WHO. (2022). Soil-transmitted helminth infections. Retrieved April 23, 2022, from https://www.who.int/news-room/fact-sheets/detail/soil-transmitted-helminth-infections

WHO. (n.d.-a). Control of neglected tropical diseases. Retrieved May 3, 2024, from https://www.who.int/teams/control-of-neglected-tropical-diseases/interventions/strategies

WHO. (n.d.-b). Neglected tropical disease. Retrieved May 29, 2024, from https://www.who.int/health-topics/neglected-tropical-diseases

WHO. (n.d.-c). Soil-transmitted helminthiases data. Retrieved May 3, 2024, from https://apps.who.int/neglected_diseases/ntddata/sth/sth.html

Wu, E., Wu, K., Daneshjou, R., Ouyang, D., Ho, D. E., & Zou, J. (2021). How medical ai devices are evaluated: Limitations and recommendations from an analysis of fda approvals. *Nature Medicine*, *27*(4), 582–584. https://doi.org/10.1038/s41591-021-01312-x

Wu, Y. Y., Huang, T. C., Ye, R. H., Fang, W. H., Lai, S. W., Chang, P. Y., Liu, W. N., Kuo, T. Y., Lee, C. H., Tsai, W. C., & Lin, C. (2020). A hematologist-level deep learning algorithm (bmsnet) for assessing the morphologies of single nuclear balls in bone marrow smears: Algorithm development. *JMIR Medical Informatics*, *8*(4), e15963. https://doi.org/10.2196/15963

Xu, H., Usuyama, N., Bagga, J., Zhang, S., Rao, R., Naumann, T., Wong, C., Gero, Z., González, J., Gu, Y., et al. (2024). A whole-slide foundation model for digital pathology from real-world data. *Nature*, 1–8. https://doi.org/10.1038/s41586-024-07441-w

Yang, A., Bakhtari, N., Langdon-Embry, L., Redwood, E., Grandjean Lapierre, S., Rakotomanga, P., Rafalimanantsoa, A., de Dios Santos, J., Vigan-Womas, I., Knoblauch, A. M., et al. (2019). Kankanet: An artificial neural network-based object detection smartphone application and mobile microscope as a point-of-care diagnostic aid for soil-transmitted helminthiases. *PLoS neglected tropical diseases*, *13*(8), e0007577. https://doi.org/10.1371/journal.pntd.0007577

Yang, F., Poostchi, M., Yu, H., Zhou, Z., Silamut, K., Yu, J., Maude, R. J., Jaeger, S., & Antani, S. (2020). Deep learning for smartphone-based malaria parasite detection in thick blood smears. *IEEE Journal of Biomedical and Health Informatics*, *24*(5), 1427–1438. https://doi.org/10.1109/JBHI.2019.2939121

Yang, Y. S., Park, D. K., Kim, H. C., Choi, M.-H., & Chai, J.-Y. (2001). Automatic identification of human helminth eggs on microscopic fecal specimens using digital image processing and an artificial neural network. *IEEE Transactions on Biomedical Engineering*, *48*(6), 718–730. https://doi.org/10.1109/10.923789

Yang, Y., Zhang, H., Gichoya, J. W., Katabi, D., & Ghassemi, M. (2024). The limits of fair medical imaging ai in real-world generalization. *Nature Medicine*, 1–11. https://doi.org/10.1038/s41591-024-03113-4

Yu, H., Mohammed, F. O., Abdel Hamid, M., Yang, F., Kassim, Y. M., Mohamed, A. O., Maude, R. J., Ding, X. C., Owusu, E. D., Yerlikaya, S., et al. (2023). Patient-level performance evaluation of a smartphone-based malaria diagnostic application. *Malaria Journal*, *22*(1), 33. https://doi.org/10.1186/s12936-023-04446-0

Yu, H., Yang, F., Rajaraman, S., Ersoy, I., Moallem, G., Poostchi, M., Palaniappan, K., Antani, S., Maude, R. J., & Jaeger, S. (2020). Malaria screener: A smartphone application for automated malaria screening. *BMC Infectious Diseases*, *20*, 1–8. https://doi.org/10.1186/s12879-020-05453-1

Yuan, X., He, P., Zhu, Q., & Li, X. (2019). Adversarial examples: Attacks and defenses for deep learning. *IEEE transactions on neural networks and learning systems*, *30*(9), 2805–2824. https://doi.org/10.1109/TNNLS.2018.2886017.

Zare, M., Akbarialiabad, H., Parsaei, H., Asgari, Q., Alinejad, A., Bahreini, M. S., Hosseini, S. H., Ghofrani-Jahromi, M., Shahriarirad, R., Amirmoezzi, Y., et al. (2022). A machine learning-based system for detecting leishmaniasis in microscopic images. *BMC infectious diseases*, *22*(1), 48. https://doi.org/10.1186/s12879-022-07029-7

Zouré, H. G. M., Wanji, S., Noma, M., Amazigo, U. V., Diggle, P. J., Tekle, A. H., & Remme, J. H. (2011). The geographic distribution of loa loa in africa: Results of large-scale implementation of the rapid assessment procedure for loiasis (raploa). *PLoS neglected tropical diseases*, *5*(6), e1210. https://doi.org/10.1371/journal.pntd.0001210