

Deep-learning Assisted Detection and Quantification of (oo)cysts of *Giardia* and *Cryptosporidium* on Smartphone Microscopy Images

Suprim Nakarmi <https://orcid.org/0009-0003-7281-2729> suprimnakarmi@gmail.com
Nepal Applied Mathematics and Informatics Institute for research (NAAMII), Kathmandu, Nepal
Center for Analytical Sciences, Kathmandu Institute of Applied Sciences (KIAS), Lalitpur, Nepal

Sanam Pudasaini <https://orcid.org/0000-0001-9782-3369> sanam001@e.ntu.edu.sg
Center for Analytical Sciences, Kathmandu Institute of Applied Sciences (KIAS), Lalitpur, Nepal

Safal Thapaliya <https://orcid.org/0000-0002-4463-6700> safal.thapaliya@naamii.org.np
Nepal Applied Mathematics and Informatics Institute for research (NAAMII), Kathmandu, Nepal

Pratima Upretee upretee@gmail.com
Nepal Applied Mathematics and Informatics Institute for research (NAAMII), Kathmandu, Nepal

Retina Shrestha shresth.retina@gmail.com
Center for Analytical Sciences, Kathmandu Institute of Applied Sciences (KIAS), Lalitpur, Nepal

Basant Giri <https://orcid.org/0000-0003-4798-3414> chembasant@gmail.com
Center for Analytical Sciences, Kathmandu Institute of Applied Sciences (KIAS), Lalitpur, Nepal

Bhanu Bhakta Neupane <https://orcid.org/0000-0003-0731-2552> newbhanu@gmail.com
Central Department of Chemistry, Tribhuvan University, Kathmandu, Nepal

Bishesh Khanal <https://orcid.org/0000-0002-2775-4748> bishesh.khanal@naamii.org.np
Nepal Applied Mathematics and Informatics Institute for research (NAAMII), Kathmandu, Nepal

Abstract

The consumption of microbial-contaminated food and water is responsible for the deaths of millions of people annually. Smartphone-based microscopy systems are portable, low-cost, and more accessible alternatives for the detection of *Giardia* and *Cryptosporidium* than traditional brightfield microscopes. However, the images from smartphone microscopes are noisier and require manual cyst identification by trained technicians, usually unavailable in resource-limited settings. Automatic detection of (oo)cysts using deep-learning-based object detection could offer a solution for this limitation. We evaluate the performance of four state-of-the-art object detectors to detect (oo)cysts of *Giardia* and *Cryptosporidium* on a custom dataset that includes both smartphone and brightfield microscopic images from vegetable samples. Faster RCNN, RetinaNet, You Only Look Once (YOLOv8s), and Deformable Detection Transformer (Deformable DETR) deep-learning models were employed to explore their efficacy and limitations. Our results show that while the deep-learning models perform better with the brightfield microscopy image dataset than the smartphone microscopy image dataset, the smartphone microscopy predictions are still comparable to the prediction performance of non-experts. Also, we publicly release brightfield and smartphone microscopy datasets with the benchmark results for the detection of *Giardia* and *Cryptosporidium*, independently captured on reference (or standard lab setting) and vegetable samples. Our code and dataset are available at <https://zenodo.org/records/12587799> and <https://zenodo.org/records/7813183>, respectively.

Keywords: Automated Parasite Detection, Deep Learning, Giardia, Cryptosporidium, Smartphone Microscopy, Brightfield Microscopy

1. Introduction

Pathogen contamination of food and water is a serious global challenge. About 1.7 billion people do not have access to feces-free drinking water in the world¹. Such contaminated food and water can often lead to diarrhea. According to the World Health Organization (WHO), diarrheal diseases account for the loss of around half a million human lives annually. Similarly, nearly 600 million people fall ill due to the consumption of contaminated food, resulting in more than four hundred thousand deaths every year². *Giardia* and *Cryptosporidium* are two of the major causes of protozoan-induced diarrheal diseases and are the most frequently identified protozoan parasites causing outbreaks (Baldursson and Karanis, 2011)². In low- and middle-income countries, children under three years of age experience three episodes of diarrhea on average every year³. *Giardia* and *Cryptosporidium* cause intestinal illness called giardiasis and cryptosporidiosis, respectively. Infections due to these parasites are more common in low- and middle-income countries because of unhygienic lifestyles and poor sanitation (Fricker et al., 2002; Gupta et al., 2020; Tandukar et al., 2013; Sherchand et al., 2004). *Giardia* is more prominent in size with an elliptical shape having a major axis 8-12 μm and minor axis 7-15 μm . In contrast, *Cryptosporidium* is spherically shaped, having a diameter of 3-6 μm (Dixon et al., 2011). Accurate detection of these microorganisms could enable early diagnosis, saving millions of lives. Moreover, regularly screening these pathogens in real food and water samples could help prevent infections and disease outbreaks.

Polymerase Chain Reaction (PCR), immunological assays, cell culture methods, fluorescence in situ hybridization, and microscopic analysis (Van den Bossche et al., 2015; Adeyemo et al., 2018) are the primary methods for the detection of (oo)cysts of *Giardia* and *Cryptosporidium*. Even though these methods are reliable and accurate, they are laborious, time-consuming, costly, and require significant expertise, resulting in a lack of tests in many resource-limited regions. For instance, cell culture requires more than 24 hours, and immunological assays and PCR reactions require costly reagents and equipment (Guerrant et al., 1985, 2001). In addition, some viable bacterial pathogens are difficult to grow or are even non-culturable (Oliver, 2005). Fluorescence tagging of cells requires expertise, and the cost of reagents is high. Microscopic methods are the most widely used diagnostic methods for detecting parasites, especially in low-resource countries. The microscopic method examines a glass slide containing a sample under a microscope. However, the traditional microscopes are still costly, less portable, and require expertise to handle and accurately identify microorganisms on the slides (Chavan and Sutkar, 2014). Recently, smartphone-based microscopic methods have been developed to potentially replace or supplement more expensive and less portable microscopic methods, including the brightfield and fluorescence microscopy (Koydemir et al., 2015; Kobori et al., 2016; Kim et al., 2015; Saeed and Jabbar,

1. <https://www.who.int/news-room/fact-sheets/detail/drinking-water>

2. <https://www.who.int/news-room/fact-sheets/detail/food-safety>

3. <https://www.who.int/news-room/fact-sheets/detail/diarrhoeal-disease>

2018; Shrestha et al., 2020; Feng et al., 2016). These microscopes allow magnification of microorganisms enabling the user to observe them on the phone screen immediately and capture the images and videos using the smartphone. However, smartphone microscopy also requires a well-trained person to identify target organisms accurately, analyze the result, and report it. In the least developed countries, the lack of skilled technicians limits the use of such a new microscopic system for rapid field testing and clinical applications. Therefore, robust and automated detection of microorganisms using smartphone microscopic images could enable more widespread use of smartphone microscopes in large-scale screening of parasites.

In recent years, several deep learning-based algorithms have been developed for various biological and clinical applications such as automatic detection, segmentation, and classification of human cells and fungi species (Xue and Ray, 2017; Zieliński et al., 2020), bacteria (Wang et al., 2020), malaria detection (Vijayalakshmi et al., 2020), image segmentation of two-dimensional materials in microscopic images (Masubuchi et al., 2020). In addition, detecting a pollen grain from microscopy images using a deep neural network has been reported (Gallardo-Caballero et al., 2019). de Haan et al. (2020) automated the screening of sickle cells using deep learning on a smartphone-based microscope. Similarly, Xu et al. (2020) proposed ParasNet - a deep-learning-based network - to detect *Giardia* and *Cryptosporidium* in brightfield microscopic images. Luo et al. (2021) created MCellNet to classify *Giardia*, *Cryptosporidium*, microbeads, and natural pollutants from the images captured from imaging flow cytometry. Machine learning techniques have also been developed to classify *Giardia* from other parasites using features such as area, equivalent diameter, and intensity in fluorescent smartphone-based microscopic images (Koydemir et al., 2015). Several other studies have proposed deep learning-based algorithms to automate microorganism detection in smartphone microscopes (de Haan et al., 2020; Fuhad et al., 2020; Yang et al., 2019). However, it is not clear how well state-of-the-art deep learning models perform in automated detection of the (oo)cysts of *Giardia* and *Cryptosporidium* from smartphone microscopic images in comparison to the traditional microscope and non-experts. We also introduce labeled smartphone and brightfield microscopy datasets for detection of the (oo) cysts of *Giardia* and *Cryptosporidium*, making it publicly available for the scientific community.

Here, we explore the potential of deep learning algorithms to detect cysts of two different kinds of enteric parasites without human experts' involvement in images taken using a custom-built smartphone microscopy system. We created a custom dataset by capturing images of sample slides from a smartphone microscope and a traditional brightfield microscope. The sample slides were obtained for (oo)cysts on both reference and vegetable sample extracts. We trained four popular object detection models - Faster RCNN (Ren et al., 2016), RetinaNet (Lin et al., 2017), YOLOv8s (Jocher et al., 2023), and Deformable DETR (Zhu et al., 2020) - using our dataset and evaluated the performance of the models. Finally, we report the performance of these models compared with expert humans and non-expert humans and inter-operator variability. The performance of the models with smartphone microscopic images was also compared with brightfield microscopic images. The dataset and the source code are made publicly available.

2. Methods

2.1 Dataset

2.1.1 TRAINING-VALIDATION SET

We used two microscopes to capture images: i) a traditional brightfield microscope and ii) a sapphire ball lens-based smartphone microscope developed by Shrestha et al. (2020). The brightfield microscope images were captured with a rectangular Field Of View (FOV) of 190 μm X 350 μm and a magnification of 400X. In contrast, the smartphone microscope images were captured with a circular FOV of diameter 200 μm and magnification of 200X - using Samsung Galaxy J7 Prime.

We captured images by making microscope slides from two different types of samples: i) standard or reference (oo)cyst suspension samples (Waterborne Inc, PC101 G/C positive control), and ii) actual vegetable samples obtained from local markets in Nepal. Figure 1 shows a few examples of reference and actual vegetable samples along with their bounding box annotations. Likewise, Table 1 summarizes the number and types of images captured in the dataset.

Reference Samples: We prepared 25 slides each from 5 μL standard (oo)cyst suspension (or standard samples) mixed with Lugol’s iodine in equal proportion. Slides refer to microscope slides on which samples were mounted, generally used for microscopic examinations. From these slides, an expert with more than two years of experience in imaging and annotating *Giardia* and *Cryptosporidium* from microscopy images captured many images from both microscopes. To maintain consistency, we selected 830 best images, each based on the clarity of parasites. Images with one of the following criteria were removed: (i) images consisting of only broken parasites - either completely damaged or partially captured, and (ii) Noisy or blurry images where no single object was visible. Since there was no prior knowledge of parasite counts in individual slides, identification and counting of (oo)cysts by the expert was considered as a benchmark.

Vegetable Samples: We prepared 200 slides from 151 vegetable samples collected from local markets in Nepal. The expert captured and selected images using the same protocol as the one used for the reference samples. From these slides, 1005 images for each type of microscope were included for further steps. The images were used for training the object detection models.

2.1.2 INDEPENDENT TEST SET

To assess the generalization of deep learning models on smartphone microscope images, an independent test set was prepared as follows: One hundred ninety-three images were captured on a particular day using the same smartphone microscope. The expert captured images at random locations of the microscope slides, regardless of the presence of parasites. The expert annotated these images with bounding boxes and ellipses for all *Giardia* and *Cryptosporidium* using VGG annotator (Dutta and Zisserman, 2019).

2.1.3 NON-EXPERT ANNOTATED SET

The same expert trained three non-expert humans on smartphone vegetable sample images for three hours in a single-day training session. As the shape and size of (oo)cysts of

Giardia and *Cryptosporidium* played a crucial role in their identification, the non-experts were instructed to use these features to locate the parasites on images. After the training session, the three non-experts annotated the cysts in 193 independent test set images. Note that non-experts used MS-paint for annotations by encircling the cysts of the two parasites with ellipses of two different colors (yellow: *Giardia*, and blue: *Cryptosporidium*), and the time non-experts took to annotate the cysts in each of the images were recorded using a stopwatch.

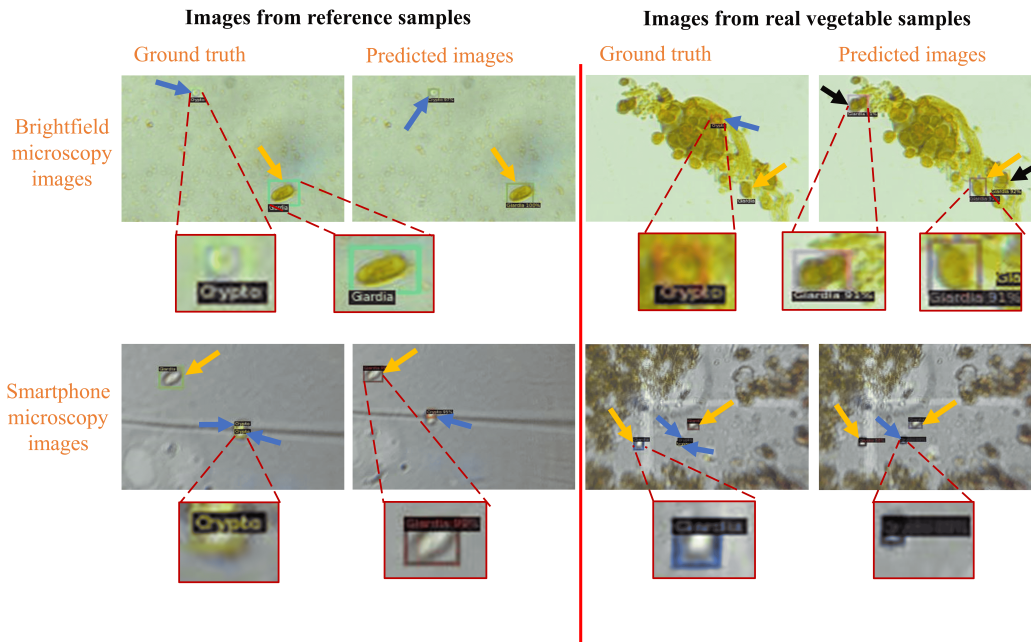


Figure 1: Representative images of reference and real vegetable samples captured using brightfield and smartphone microscopes. Example ground truth annotations of the parasites by the expert and predictions from the object detector models. Yellow arrow points *Giardia*, blue arrow points *Cryptosporidium*, and black arrow points debris. For this illustration, Faster RCNN was used to detect parasites.

Table 1: — Description of two types of the dataset (reference sample and vegetable sample) used in this study with the total number of parasites annotated by the expert using VGG software (Dutta and Zisserman, 2019).

Dataset	No. of images	Microscopes	Giardia’s annotation	Cryptosporidium’s annotation
Reference sample	830	Smartphone	839	534
		Brightfield	907	502
Vegetable sample	1005	Smartphone	439	796
		Brightfield	344	740
Test (vegetable sample)	193	Smartphone	165	137

2.2 Deep Learning-based Object Detection Models

Four state-of-the-art object detection models were selected for this study: Faster RCNN (Ren et al., 2016), RetinaNet (Lin et al., 2017), YOLOv8s (Jocher et al., 2023), and Deformable DETR (Zhu et al., 2020). Faster RCNN is one of the most popular networks from the Region-Based Convolutional Neural Networks (RCNN) family, which is an improved version based on two previous methods, RCNN (Girshick et al., 2014) and Fast RCNN (Girshick, 2015). RetinaNet is a popular single-stage detector that uses a *Focal Loss* to address the foreground-background class imbalance problem that gets more severe when detecting smaller objects in the images. YOLOv8s is a popular single-stage detector with relatively low computational costs, enabling it to be run on smartphones. To assess the quality of the real-time detection model that can be run on smartphones, we chose YOLOv8s - the smallest model with 11.2 million parameters - from the various available models for YOLOv8, as it is lightweight and has the lowest prediction time (Jocher et al., 2023). Deformable DETR is a state-of-the-art transformer-based object detector that incorporates deformable attention mechanisms, replacing the fixed grids used in traditional self-attention with deformable grids. This allows the model to capture spatial relationships more effectively by dynamically adjusting its attention (Zhu et al., 2020), which is the improved version of DETR.

2.3 Experimental Setup

Figure 2 illustrates the overall pipeline of the training and evaluation of the four object detectors using expert and non-expert annotated images captured from brightfield and smartphone images. During training, the images were first pre-processed (see subsection A.1) and then fed as an input to one of the four object detection models for classifying the object type (*Giardia* or *Cryptosporidium*) and localizing the objects with a bounding box.

2.3.1 EVALUATION APPROACH

Since the target application is to be able to assess the contamination in vegetable samples by identifying and counting the number of (oo)cysts of the two parasites in the microscopic images, we evaluate the four models using classification performance metrics: precision, recall, and F1-score. The model-predicted cysts can belong to one of the three categories: True Positive (TP) when the model's object prediction (of either *Giardia* or *Cryptosporidium* cysts) correctly matches with the Ground Truth (GT) annotation, False Positive (FP) when the model's object prediction is different than the GT, and finally False Negative (FN) when the model does not detect cysts annotated by the experts in the images. Precision, Recall, and F1-scores are calculated based on these three values to evaluate the models' performances.

$$Precision (P) = \frac{TP}{TP + FP} \quad (1)$$

$$Recall (R) = \frac{TP}{TP + FN} \quad (2)$$

$$F1 \text{ score} = \frac{2 \times P \times R}{P + R} \quad (3)$$

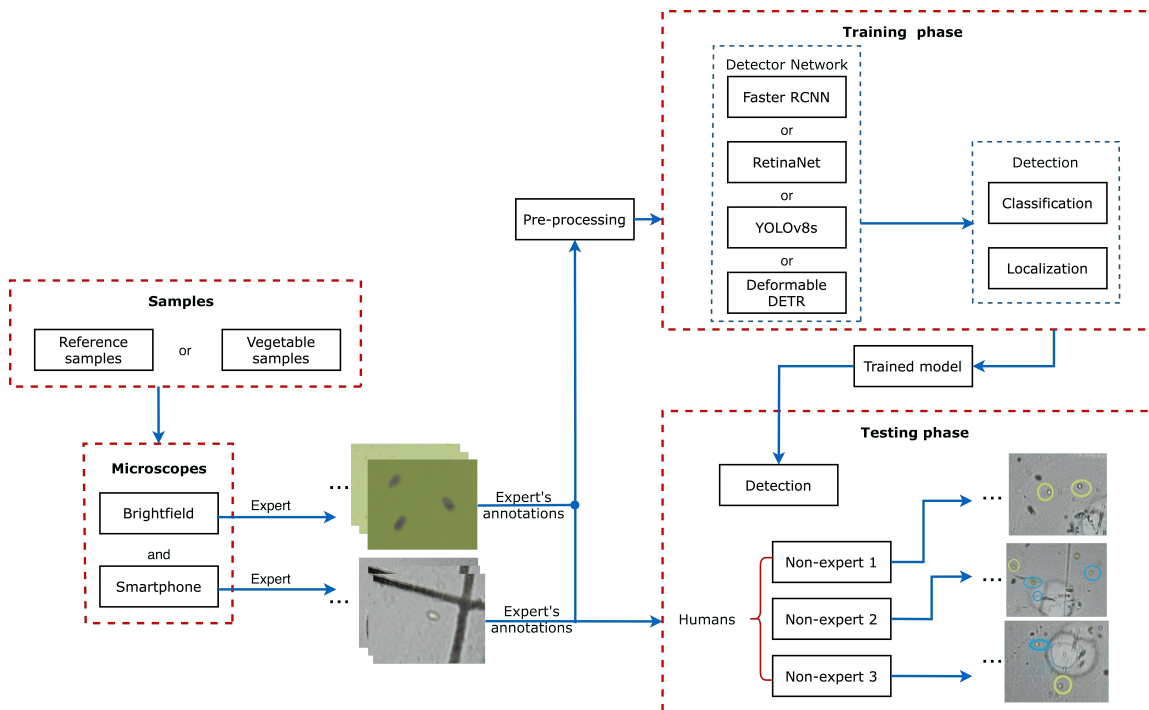


Figure 2: Schema showing the overview of the method implemented in this study. Note that the detector networks were trained separately using images captured from brightfield and smartphone microscopes on reference and vegetable samples. The testing phase consists of predicting (oo)cysts of the parasites only on images captured on vegetable samples using the smartphone microscope.

In addition to assessing the object detection model’s performance against the expert GT annotations, we also evaluate how the model compares with non-expert humans. This helps assess the utility of deploying the automated models in places where the experts are not available.

2.3.2 COMPARING THE FOUR MODELS AND NON-EXPERT HUMANS

We evaluated the object detection models - Faster RCNN, RetinaNet, YOLOv8s, and Deformable DETR - in the following settings:

- 5-fold cross-validation of the brightfield and smartphone microscope images using the training-validation dataset with expert annotations: 830×2 images for the two microscopes with reference and 1005×2 images with vegetable samples.
- Comparison of the detection models against non-expert humans in identifying the (oo)cysts of the two parasites in separate independent smartphone microscope test images ($n = 193$).
- Comparison of time taken by non-experts vs. detection model to identify the (oo)cysts in the smartphone microscope test images.

To test whether the differences in the performance of the non-expert humans and the detection models in the independent test set are statistically significant, we use paired Wilcoxon signed-rank test (Woolson, 2007). The details are provided in subsection A.2.

3. Results

3.1 5-fold Cross-validation in Training-validation Set

Table 2 presents the performance of the four object detection models for the brightfield microscope images of reference and vegetable samples. The results are reported for confidence scores (c) and iou-thresholds (i) optimized for each model separately. The details about the choice of these thresholds are provided in subsection A.3. We see that the models perform better for reference samples than vegetable samples. It is expected because the reference sample does not have debris, and hence, there are very few objects confounding with the cysts in the clean brightfield images (example images in Figure 1). Similarly, we observe that all the object detectors detect *Giardia* cysts better than *Cryptosporidium* ones, except for Faster RCNN, which has negligibly better performance in detecting *Cryptosporidium*.

Table 2: Brightfield microscope training-validation set: 5-fold cross-validation results on reference and vegetable samples for Faster RCNN, RetinaNet, YOLOv8s and Deformable DETR.

Dataset	Models	Thresholds	Precision	Recall	F1-score
<i>Giardia</i>					
Reference Sample	Faster RCNN	$c = 0.7, i = 0.4$	0.957 \pm 0.014	0.968 \pm 0.018	0.962 \pm 0.007
	RetinaNet	$c = 0.5, i = 0.5$	0.917 \pm 0.017	0.946 \pm 0.013	0.931 \pm 0.011
	YOLOv8s	$c = 0.4, i = 0.3$	0.954 \pm 0.016	0.975 \pm 0.010	0.965 \pm 0.010
	Deformable DETR	$c = 0.4, i = 0.2$	0.864 \pm 0.037	0.870 \pm 0.014	0.867 \pm 0.021
Vegetable Sample	Faster RCNN	$c = 0.8, i = 0.5$	0.783 \pm 0.024	0.901 \pm 0.026	0.837 \pm 0.005
	RetinaNet	$c = 0.4, i = 0.4$	0.753 \pm 0.029	0.927 \pm 0.035	0.830 \pm 0.024
	YOLOv8s	$c = 0.4, i = 0.3$	0.855 \pm 0.037	0.811 \pm 0.018	0.831 \pm 0.018
	Deformable DETR	$c = 0.4, i = 0.2$	0.636 \pm 0.042	0.782 \pm 0.063	0.700 \pm 0.042
<i>Cryptosporidium</i>					
Reference Sample	Faster RCNN	$c = 0.7, i = 0.4$	0.880 \pm 0.024	0.915 \pm 0.030	0.897 \pm 0.018
	RetinaNet	$c = 0.5, i = 0.5$	0.917 \pm 0.025	0.870 \pm 0.031	0.893 \pm 0.022
	YOLOv8s	$c = 0.4, i = 0.3$	0.890 \pm 0.034	0.887 \pm 0.031	0.888 \pm 0.028
	Deformable DETR	$c = 0.4, i = 0.2$	0.768 \pm 0.035	0.808 \pm 0.012	0.787 \pm 0.015
Vegetable Sample	Faster RCNN	$c = 0.8, i = 0.5$	0.845 \pm 0.029	0.835 \pm 0.045	0.839 \pm 0.024
	RetinaNet	$c = 0.4, i = 0.4$	0.801 \pm 0.025	0.851 \pm 0.022	0.826 \pm 0.023
	YOLOv8s	$c = 0.4, i = 0.3$	0.879 \pm 0.052	0.716 \pm 0.065	0.788 \pm 0.051
	Deformable DETR	$c = 0.4, i = 0.2$	0.642 \pm 0.029	0.720 \pm 0.054	0.678 \pm 0.033

Similarly, Table 3 shows 5-fold cross-validation results for smartphone microscope training-validation set images. As expected, the object detectors' performance is lower than brightfield microscope images because smartphone microscope images have more textured noise and lower magnification than traditional brightfield microscopes. However, Deformable DETR performs slightly better when predicting *Cryptosporidium* on vegetable samples. We see that most of the trends observed in brightfield microscope images can also be seen for smartphone microscope: *Giardia* cysts are better detected than *Cryptosporidium* cysts,

YOLOv8s and Faster RCNN provide better results than RetinaNet, except for vegetable samples, and the performance of the detectors is better for reference samples compared to the vegetable sample in general. However, one notable exception is that the F1-score for the vegetable sample is better than the reference sample for *Cryptosporidium* when using YOLOv8s and Deformable DETR. In the case of YOLOv8s, while the recall reduced slightly in the vegetable sample, the precision increased substantially, increasing the overall F1-score. Similarly, a remarkable increase in precision and recall is observed in Deformable DETR. RetinaNet provides better results in detecting *Giardia* and *Cryptosporidium* in vegetable samples, but Faster RCNN is better for reference samples.

Table 3: Smartphone microscope training-validation set: 5-fold cross-validation results on reference and vegetable samples for Faster RCNN, RetinaNet, YOLOv8s, and Deformable DETR.

Dataset	Models	Thresholds	Precision	Recall	F1-score
<i>Giardia</i>					
Reference Sample	Faster RCNN	$c = 0.5, i = 0.3$	0.872 \pm 0.038	0.919 \pm 0.019	0.895 \pm 0.024
	RetinaNet	$c = 0.3, i = 0.3$	0.803 \pm 0.024	0.912 \pm 0.022	0.854 \pm 0.018
	YOLOv8s	$c = 0.3, i = 0.4$	0.835 \pm 0.032	0.891 \pm 0.032	0.862 \pm 0.030
	Deformable DETR	$c = 0.4, i = 0.2$	0.768 \pm 0.038	0.745 \pm 0.022	0.756 \pm 0.027
Vegetable Sample	Faster RCNN	$c = 0.5, i = 0.4$	0.675 \pm 0.061	0.789 \pm 0.047	0.726 \pm 0.045
	RetinaNet	$c = 0.4, i = 0.3$	0.729 \pm 0.040	0.751 \pm 0.061	0.739 \pm 0.042
	YOLOv8s	$c = 0.2, i = 0.3$	0.761 \pm 0.053	0.729 \pm 0.093	0.740 \pm 0.038
	Deformable DETR	$c = 0.4, i = 0.2$	0.700 \pm 0.025	0.690 \pm 0.039	0.694 \pm 0.023
<i>Cryptosporidium</i>					
Reference Sample	Faster RCNN	$c = 0.5, i = 0.3$	0.761 \pm 0.030	0.673 \pm 0.051	0.714 \pm 0.037
	RetinaNet	$c = 0.3, i = 0.3$	0.698 \pm 0.027	0.688 \pm 0.045	0.692 \pm 0.025
	YOLOv8s	$c = 0.3, i = 0.4$	0.580 \pm 0.059	0.678 \pm 0.117	0.620 \pm 0.066
	Deformable DETR	$c = 0.4, i = 0.2$	0.580 \pm 0.038	0.598 \pm 0.058	0.588 \pm 0.042
Vegetable Sample	Faster RCNN	$c = 0.5, i = 0.4$	0.638 \pm 0.031	0.650 \pm 0.045	0.644 \pm 0.030
	RetinaNet	$c = 0.4, i = 0.3$	0.700 \pm 0.046	0.675 \pm 0.024	0.686 \pm 0.017
	YOLOv8s	$c = 0.2, i = 0.3$	0.658 \pm 0.021	0.676 \pm 0.103	0.663 \pm 0.046
	Deformable DETR	$c = 0.4, i = 0.2$	0.675 \pm 0.034	0.695 \pm 0.038	0.684 \pm 0.029

3.2 Independent Test Set with Smartphone Microscope Images

Table 4 presents the performance of the four models and non-expert humans on the independent test set images. This set consists of smartphone microscopic images of vegetable samples where the expert’s annotation is considered ground truth. The performance of all the models on the test set images is lower than that of cross-validation, which is typical for machine learning models, known as the generalization problem. YOLOv8s, which performed well in cross-validation, has the worst overall score in the test set, suggesting that the model was less robust to new data. In contrast, the RetinaNet model is more robust to the new data. Moreover, RetinaNet seems to perform better for *Cryptosporidium*, which could be due to the *Focal Loss* targeted for smaller objects, as *Cryptosporidium* cysts have a smaller size than *Giardia* cysts. Nevertheless, the test set results reinforce the observation that detecting *Cryptosporidium* is more difficult compared to *Giardia* cysts.

Table 4: Smartphone microscope independent test set: Performance of the four object detectors and non-expert humans. The standard deviation describes how diverse the scores are from the mean when predicting the test set on each cross-validation fold.

Detector	<i>Giardia</i>			<i>Cryptosporidium</i>		
	Precision	Recall	F1-score	Precision	Recall	F1-score
Faster RCNN	0.679 \pm 0.024	0.358 \pm 0.015	0.468 \pm 0.015	0.423 \pm 0.030	0.270 \pm 0.033	0.328 \pm 0.026
RetinaNet	0.657 \pm 0.034	0.333 \pm 0.025	0.441 \pm 0.025	0.473 \pm 0.013	0.286 \pm 0.049	0.355 \pm 0.039
YOLOv8s	0.656 \pm 0.020	0.240 \pm 0.037	0.350 \pm 0.040	0.337 \pm 0.019	0.260 \pm 0.037	0.293 \pm 0.031
Deformable DETR	0.578 \pm 0.018	0.272 \pm 0.036	0.368 \pm 0.034	0.350 \pm 0.019	0.340 \pm 0.012	0.344 \pm 0.009
Non-expert1	0.634	0.430	0.512	0.233	0.124	0.162
Non-expert2	0.399	0.539	0.459	0.191	0.606	0.290
Non-expert3	0.297	0.497	0.372	0.168	0.255	0.203
Expert (Benchmark)	1.000	1.000	1.000	1.000	1.000	1.000

Wilcoxon signed-rank test showed that all the models and non-expert humans mostly had significantly different results, except for Non-expert 1, who had non-significant statistical results with Faster RCNN and RetinaNet, for both *Giardia* and *Cryptosporidium*. (Details provided in subsection A.2.)

Since the ability to detect the cysts in real-time using only the smartphone can be valuable for rapid field testing scenarios, we computed the time required to detect the cysts for the different models and the human experts and non-experts. YOLOv8s was the fastest, predicting the objects in an average of 0.032 seconds per image, whereas Faster RCNN, RetinaNet, and Deformable DETR needed 1.4 seconds, 1.3 seconds, 0.25 seconds, respectively. Note that all the models were trained and predicted on Google Colab having Central Processing Unit (CPU) and Graphics Processing Unit (GPU) specifications of Intel (R) Xeon (R) 2vCPU @ 2.2 GHz and Tesla T4 (16GB, 2560 CUDA cores), respectively. Among the human annotators, the expert identified cysts in an average of 8.4 seconds per image, but the non-experts took as long as 24.818 seconds.

4. Discussion and Conclusion

In this study, we explored the possibility of using automatic parasite detection in brightfield and smartphone microscopes for the (oo)cysts of *Giardia* and *Cryptosporidium* in scenarios where experts are not available. Two different datasets were prepared by separately capturing the images of reference and actual vegetable samples using the smartphone and the brightfield microscopes. Four object detection models were explored, and their performance was compared against human non-experts while taking an expert annotation as ground truth. Precision, recall, and F1-scores were used as they are useful evaluation metrics when a target application requires counting objects (Xue and Ray, 2017; Brhane Hagos et al., 2019).

The results show that for the same range of training samples, models perform better on reference samples than vegetable samples, brightfield microscopes than smartphone microscopes, *Giardia* than *Cryptosporidium* cysts. Vegetable samples have debris similar to *Giardia* and *Cryptosporidium* (Figure 1), making the task more difficult. Smartphone images are more textured and noisy. Additionally, due to the curvature effect of the ball lens in the smartphone microscope, the objects get stretched toward the peripheral regions and appear bigger than those at the center of the image. In such cases, the models could falsely

predict *Giardia* as *Cryptosporidium* and vice versa, as shown in Figure 3. Some parasites on the smartphone images seem to be blurry, which causes false negatives as shown in Figure 3a. Figure 3c illustrates a scenario of multiple predictions where both classes were predicted for a single object. However, it was only observed when using RetinaNet. In other models, the problem was eliminated by increasing the confidence threshold level for the prediction. *Giardia* was better predicted by all four models used in the study, possibly due to its larger size and ellipse shape. At the time this article was published, to the best of our knowledge, no other work employed automation in sapphire ball lens-based smartphone microscopy to detect *Giardia* and *Cryptosporidium*. Some similar works are presented as follows: Luo et al. (2021) reported a combined average sensitivity (or recall) of 0.974 for both the parasites using deep learning-enabled imaging flow cytometry; however, they stated the ML model did not work well in very different unseen data. Koydemir et al. (2015) presented a fluorescent-based smartphone microscopy system integrated with a machine learning algorithm to detect *Giardia* that achieved a sensitivity of 0.840 on water samples, reporting a prediction time of 2 minutes. Ligda et al. (2020) implemented Linear Discriminant Function Analysis (LDFA) to quantify the *Giardia* and *Cryptosporidium* with an accuracy of 0.690 and 0.750, respectively, on water samples using Olympus fluorescence microscope. Note that, among all these works, the smartphone-based microscopy system we have used is the cheapest, costing only \$15 - excluding the price of the smartphone (Shrestha et al., 2020).

This work shows the feasibility and promise of integrating deep learning-based automated models into brightfield and smartphone microscopes, especially in resource-constrained areas where experts are not readily available. Although the models performed better than non-experts in *Cryptosporidium* cysts, the F1-scores for the models are still relatively low in test sets. Future work requires collecting a much larger dataset, which will improve the scores. Similarly, human experts can commit errors during annotation; therefore, it would be interesting to assess the inter- and intra-operator variability among the experts by annotating a certain subset and comparing this variability against AI models. A larger dataset can be used for self-supervised pretraining on unlabeled samples, followed by supervised fine-tuning on the annotated datasets to get better object detection performance. Moreover, in this study, we have not combined reference and vegetable sample images to train the models. The mixed dataset could be used in future works hoping for better performance and domain adaptation. Also, since the images obtained from smartphone microscopy were typically noisy, it would be beneficial to explore various image enhancement methods such as adaptive filters, Gaussian filters, deblurring techniques like Weiner deconvolution, Super-Resolution Convolutional Neural Network (SRCNN) (Dong et al., 2015), or deep-learning-based method (Zhao et al., 2020) before training. Future works can also focus on developing robust models in detecting tiny *Cryptosporidium* cysts. Additionally, defining the region of interest on the images to discard any stretched portions and applying domain adaptation techniques as suggested by (Becker et al., 2014) and (Farahani et al., 2021) to transform smartphone microscopy images into images resembling brightfield microscopy images could further improve the performance of the models.

In our experiments, YOLOv8s performed decently in the real vegetable samples. Considering its lightweight architecture (22 MB) and good performance in detecting the cysts,

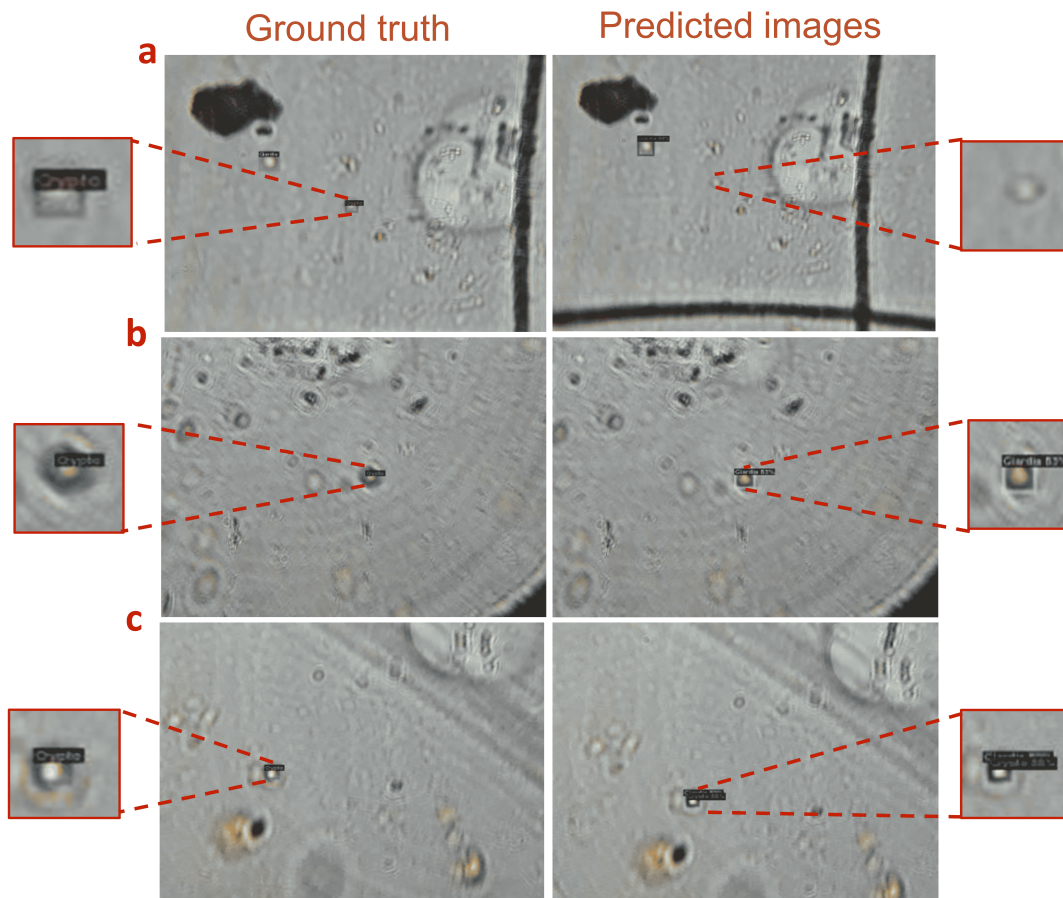


Figure 3: Three pairs of ground truth and predicted images (i.e., a, b, and c) showing error in prediction for vegetable sample test set taken from smartphone microscopy. (a) Prediction missed due to blurry parasite, (b) False prediction due to similar size, and (c) Multiple predictions for a single object. For this illustration, Faster RCNN was used to detect parasites.

YOLOv8s shows potential to be deployed on mobile devices without the need for a server (Kuznetsova et al., 2021; Chen et al., 2021).

This work provides the first step and shows the feasibility of a low-cost smartphone-based automated detection of (oo)cysts or other microorganisms in vegetables, water, stool, or other food products without needing an expert. More than specific deep learning models to choose from, future work should focus on larger datasets or semi-supervised approaches and designing experiments in prospective settings to compare against non-experts and experts for diagnostic end-points.

Acknowledgments

We thank Rabin Adhikari for helping us with formatting. The images used in this research were taken with support from NAS and USAID (to BG and BBN) through Partnerships for Enhanced Engagement in Research (PEER) (AIDOAA-A-11-00012). The opinions, findings, conclusions, or recommendations expressed in this article are those of the authors alone and do not necessarily reflect the views of USAID or NAS. Suprim Nakarmi is currently affiliated with the University of South Dakota.

Ethical Standards

The work follows appropriate ethical standards in conducting research and writing the manuscript, following all applicable laws and regulations regarding the treatment of animals or human subjects.

Conflicts of Interest

The authors declare no conflicts of interest.

References

- Folasade Esther Adeyemo, Gulshan Singh, Poovendhree Reddy, and Thor Axel Stenström. Methods for the detection of cryptosporidium and giardia: from microscopy to nucleic acid based tools in clinical and environmental regimes. *Acta tropica*, 184:15–28, 2018.
- Selma Baldursson and Panagiotis Karanis. Waterborne transmission of protozoan parasites: review of worldwide outbreaks—an update 2004–2010. *Water research*, 45(20):6603–6614, 2011.
- Carlos Becker, C Mario Christoudias, and Pascal Fua. Domain adaptation for microscopy imaging. *IEEE Transactions on Medical Imaging*, 34(5):1125–1139, 2014.
- Yeman Brhane Hagos, Priya Lakshmi Narayanan, Ayse U Akarca, Teresa Marafioti, and Yinyin Yuan. Concorde-net: Cell count regularized convolutional neural network for cell detection in multiplex immunohistochemistry images. *arXiv e-prints*, pages arXiv–1908, 2019.
- Sneha Narayan Chavan and Ashok Manikchand Sutkar. Malaria disease identification and analysis using image processing. *Int. J. Latest Trends Eng. Technol*, 3(3):218–223, 2014.
- Kai Chen, Jiaqi Wang, Jiangmiao Pang, Yuhang Cao, Yu Xiong, Xiaoxiao Li, Shuyang Sun, Wansen Feng, Ziwei Liu, Jiarui Xu, et al. Mmdetection: Open mmlab detection toolbox and benchmark. *arXiv preprint arXiv:1906.07155*, 2019.
- Yuwen Chen, Chao Zhang, Tengfei Qiao, Jianlin Xiong, and Bin Liu. Ship detection in optical sensing images based on yolov5. In *Twelfth International Conference on Graphics and Image Processing (ICGIP 2020)*, volume 11720, page 117200E. International Society for Optics and Photonics, 2021.

- Kevin de Haan, Hatice Ceylan Koydemir, Yair Rivenson, Derek Tseng, Elizabeth Van Dyne, Lissette Bakic, Doruk Karınca, Kyle Liang, Megha Ilango, Esin Gumustekin, et al. Automated screening of sickle cells using a smartphone-based microscope and deep learning. *NPJ Digital Medicine*, 3(1):1–9, 2020.
- BR Dixon, Ronald Fayer, Mónica Santín, DE Hill, and JP Dubey. Protozoan parasites: Cryptosporidium, giardia, cyclospora, and toxoplasma. *Rapid detection, characterization, and enumeration of foodborne pathogens*, pages 349–370, 2011.
- Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2015.
- Abhishek Dutta and Andrew Zisserman. The VIA annotation software for images, audio and video. In *Proceedings of the 27th ACM International Conference on Multimedia*, MM '19, New York, NY, USA, 2019. ACM. ISBN 978-1-4503-6889-6/19/10. . URL <https://doi.org/10.1145/3343031.3350535>.
- Abolfazl Farahani, Sahar Voghoei, Khaled Rasheed, and Hamid R Arabnia. A brief review of domain adaptation. *Advances in data science and information engineering: proceedings from ICDATA 2020 and IKE 2020*, pages 877–894, 2021.
- Steve Feng, Derek Tseng, Dino Di Carlo, Omai B Garner, and Aydogan Ozcan. High-throughput and automated diagnosis of antimicrobial resistance using a cost-effective cellphone-based micro-plate reader. *Scientific reports*, 6(1):1–9, 2016.
- CR Fricker, GD Medema, and HV Smith. Protozoan parasites (cryptosporidium, giardia, cyclospora). *Guidelines for drinking-water quality*, 2:70–118, 2002.
- KM Fuhad, Jannat Ferdousey Tuba, Md Sarker, Rabiul Ali, Sifat Momen, Nabeel Mohammed, and Tanzilur Rahman. Deep learning based automatic malaria parasite detection from blood smear and its smartphone based application. *Diagnostics*, 10(5):329, 2020.
- Ramón Gallardo-Caballero, Carlos J García-Orellana, Antonio García-Manso, Horacio M González-Velasco, Rafael Tormo-Molina, and Miguel Macías-Macías. Precise pollen grain detection in bright field microscopy using deep learning techniques. *Sensors*, 19(16):3583, 2019.
- Ross Girshick. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 1440–1448, 2015.
- Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587, 2014.
- Richard L Guerrant, David S Shields, Stephen M Thorson, John B Schorling, and Dieter HM Groschel. Evaluation and diagnosis of acute infectious diarrhea. *The American journal of medicine*, 78(6):91–98, 1985.

- Richard L Guerrant, Thomas Van Gilder, Ted S Steiner, Nathan M Thielman, Laurence Slutsker, Robert V Tauxe, Thomas Hennessy, Patricia M Griffin, Herbert DuPont, R Bradley Sack, et al. Practice guidelines for the management of infectious diarrhea. *Clinical infectious diseases*, 32(3):331–351, 2001.
- Ranjit Gupta, Binod Rayamajhee, Samendra P Sherchan, Ganesh Rai, Reena Kiran Mukhiya, Binod Khanal, and Shiba Kumar Rai. Prevalence of intestinal parasitosis and associated risk factors among school children of saptari district, nepal: a cross-sectional study. *Tropical medicine and health*, 48(1):1–9, 2020.
- Glenn Jocher, Ayush Chaurasia, and Jing Qiu. YOLO by Ultralytics, January 2023. URL <https://github.com/ultralytics/ultralytics>.
- Jung-Hyun Kim, Hong-Gu Joo, Tae-Hoon Kim, and Young-Gu Ju. A smartphone-based fluorescence microscope utilizing an external phone camera lens module. *BioChip Journal*, 9(4):285–292, 2015.
- Yoshitomo Kobori, Peter Pfanner, Gail S Prins, and Craig Niederberger. Novel device for male infertility screening with single-ball lens microscope and smartphone. *Fertility and sterility*, 106(3):574–578, 2016.
- Hatice Ceylan Koydemir, Zoltan Gorocs, Derek Tseng, Bingen Cortazar, Steve Feng, Raymond Yan Lok Chan, Jordi Burbano, Euan McLeod, and Aydogan Ozcan. Rapid imaging, detection and quantification of giardia lamblia cysts using mobile-phone based fluorescent microscopy and machine learning. *Lab on a Chip*, 15(5):1284–1293, 2015.
- Anna Kuznetsova, Tatiana Maleva, and Vladimir Soloviev. Yolov5 versus yolov3 for apple detection. *Cyber-Physical Systems: Modelling and Intelligent Control*, pages 349–358, 2021.
- Panagiota Ligda, Edwin Claerebout, Despoina Kostopoulou, Antonios Zdragas, Stijn Casaert, Lucy J Robertson, and Smaragda Sotiraki. Cryptosporidium and giardia in surface water and drinking water: Animal sources and towards the use of a machine-learning approach as a tool for predicting contamination. *Environmental Pollution*, 264: 114766, 2020.
- Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017.
- Shaobo Luo, Kim Truc Nguyen, Binh TT Nguyen, Shilun Feng, Yuzhi Shi, Ahmed Elsayed, Yi Zhang, Xiaohong Zhou, Bihan Wen, Giovanni Chierchia, et al. Deep learning-enabled imaging flow cytometry for high-speed cryptosporidium and giardia detection. *Cytometry Part A*, 2021.
- Satoru Masubuchi, Eisuke Watanabe, Yuta Seo, Shota Okazaki, Takao Sasagawa, Kenji Watanabe, Takashi Taniguchi, and Tomoki Machida. Deep-learning-based image segmentation integrated with optical microscopy for automatically searching for two-dimensional materials. *npj 2D Materials and Applications*, 4(1):1–9, 2020.

- James D Oliver. The viable but nonculturable state in bacteria. *Journal of microbiology*, 43(spc1):93–100, 2005.
- Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE transactions on pattern analysis and machine intelligence*, 39(6):1137–1149, 2016.
- Muhammad A Saeed and Abdul Jabbar. “smart diagnosis” of parasitic diseases by use of smartphones. *Journal of clinical microbiology*, 56(1), 2018.
- JB Sherchand, PN Misra, DN Dhakal, et al. Cryptosporidium parvum: An observational study in kanti children hospital, kathmandu, nepal. *Journal of Nepal Health Research Council*, 2004.
- Retina Shrestha, Rojina Duwal, Sajeev Wagle, Samiksha Pokhrel, Basant Giri, and Bhanu Bhakta Neupane. A smartphone microscopic method for simultaneous detection of (oo) cysts of cryptosporidium and giardia. *PLoS Neglected Tropical Diseases*, 14(9), 2020.
- Sarmila Tandukar, Shamshul Ansari, Nabaraj Adhikari, Anisha Shrestha, Jyotshana Gautam, Binita Sharma, Deepak Rajbhandari, Shikshya Gautam, Hari Prasad Nepal, and Jeevan B Sherchand. Intestinal parasitosis in school children of lalitpur district of nepal. *BMC research notes*, 6(1):449, 2013.
- Dorien Van den Bossche, Lieselotte Cnops, Jacob Verschueren, and Marjan Van Esbroeck. Comparison of four rapid diagnostic tests, elisa, microscopy and pcr for the detection of giardia lamblia, cryptosporidium spp. and entamoeba histolytica in feces. *Journal of microbiological methods*, 110:78–84, 2015.
- A Vijayalakshmi et al. Deep learning approach to detect malaria from microscopic images. *Multimedia Tools and Applications*, 79(21):15297–15317, 2020.
- Hongda Wang, Hatice Ceylan Koydemir, Yunzhe Qiu, Bijie Bai, Yibo Zhang, Yiyin Jin, Sabiha Tok, Enis Cagatay Yilmaz, Esin Gumustekin, Yair Rivenson, et al. Early-detection and classification of live bacteria using time-lapse coherent imaging and deep learning. *arXiv preprint arXiv:2001.10695*, 2020.
- RF Woolson. Wilcoxon signed-rank test. *Wiley encyclopedia of clinical trials*, pages 1–3, 2007.
- XF Xu, S Talbot, and T Selvaraja. Deep learning based cell parasites detection. *arXiv preprint arXiv:2002.11327*, 2020.
- Yao Xue and Nilanjan Ray. Cell detection in microscopy images with deep convolutional neural network and compressed sensing. *arXiv preprint arXiv:1708.03307*, 2017.
- Feng Yang, Mahdieh Poostchi, Hang Yu, Zhou Zhou, Kamolrat Silamut, Jian Yu, Richard J Maude, Stefan Jaeger, and Sameer Antani. Deep learning for smartphone-based malaria parasite detection in thick blood smears. *IEEE Journal of Biomedical and Health Informatics*, 24(5):1427–1438, 2019.

Huangxuan Zhao, Ziwen Ke, Ningbo Chen, Songjian Wang, Ke Li, Lidai Wang, Xiaojing Gong, Wei Zheng, Liang Song, Zhicheng Liu, et al. A new deep learning method for image deblurring in optical microscopic systems. *Journal of biophotonics*, 13(3):e201960147, 2020.

Xizhou Zhu, Weijie Su, Lewei Lu, Bin Li, Xiaogang Wang, and Jifeng Dai. Deformable detr: Deformable transformers for end-to-end object detection. *arXiv preprint arXiv:2010.04159*, 2020.

Bartosz Zieliński, Agnieszka Sroka-Oleksiak, Dawid Rymarczyk, Adam Piekarczyk, and Monika Brzychczy-Włoch. Deep learning approach to describe and classify fungi microscopic images. *PloS one*, 15(6):e0234806, 2020.

Appendix A. Supplementary Information

A.1 Implementation Details

The networks were implemented in Python 3.10.6. The open-source object detection library, mmdetection (Chen et al., 2019), was used to execute Faster RCNN, RetinaNet, and Deformable DETR. Note that the backbone network of ResNeXt101 accompanied by Feature Pyramid Network (FPN), ResNet101, and ResNet50 were used for Faster RCNN, RetinaNet, and Deformable DETR, respectively. Similarly, YOLOv8s was implemented by cloning the repository of Ultralytics (Jocher et al., 2023). For Faster RCNN, RetinaNet, and Deformable DETR the shortest edge was resized with sharp edge lengths of 640, 472, 704, 736, 768, and 800, and a random horizontal flip with a probability of 0.5 was used during data augmentation. For YOLOv8s, random horizontal flips were applied along with mosaic augmentation with probability 0.5 and 1.0, respectively, for reference and vegetable sample images. We adapted the learning rate, iterations, and number of classes for all three models during training using empirical experiments to achieve optimal performance. Additionally, for RetinaNet, the *Focal Loss's* alpha and gamma parameters were adjusted to improve the results from the default settings. The hyperparameters used for these three models are summarized in Table 5.

Table 5: Eight detection models and hyperparameter details. *W/N*: Warm-up/Maximum Iteration; *LR*: Learning Rate; *FPN*: Feature Pyramid Network; *PPI*: Proposals per image; all other hyperparameters were left as default in mmdetection implementation.

Microscope	Detector	Backbone	W/N	LR	Other
Brightfield	Faster RCNN	ResNeXt101	1200/1500	0.001	FPN, PPI = 64
	RetinaNet	ResNet101	800/1200	0.001	$\alpha = 0.93, \gamma = 1$
	YOLOv8s	CSPDarknet	3/100	0.01	batch size=16
	Deformable DETR	ResNet50	-	0.001	batch size =16, epoch=100
Smartphone	Faster RCNN	ResNeXt101	1500/2000	0.01	FPN, PPI = 64
	RetinaNet	ResNet101	1200/1500	0.001	$\alpha = 0.99, \gamma = 1.7$
	YOLOv8s	CSPDarknet	3/200	0.001	batch size = 16
	Deformable DETR	ResNet50	-	0.01	batch size=16, epoch=110

A.2 Statistical Analysis

The Quantile-Quantile (Q-Q) plot was used to check if the data were normally distributed. Since the data were not normally distributed, we selected paired Wilcoxon signed-rank test (Woolson, 2007) to test the significance between the predictions. A p-value of less than 0.05 (i.e., 5 %) was considered significant. We have assumed total images as the sample size (i.e., sample size = 193) and the count of parasites on each image as the scores. Since the sample size was large (i.e., $n > 50$), we used normal approximation on the Wilcoxon signed-rank test. Here, normal approximation does not mean the data distribution is normal, but the

Wilcoxon signed rank test statistic is assumed to be approximately normal. We used `scipy`⁴ to perform the Wilcoxon signed-rank test.

The null hypothesis is that there is no difference between the cysts-count predictions between the non-experts and models. The p-values are provided in Table 6 for *Giardia* and Table 7 for *Cryptosporidium*. We observed significantly different results among all the non-experts and between Non-expert 2 and models. All three models significantly outperformed the non-experts for the detection of *Cryptosporidium*. However, for *Giardia*, only Faster RCNN had a statistically significant difference in performance compared to Non-expert 1 and Non-expert 2. In contrast, RetinaNet had a statistically significant difference in performance compared to Non-expert 3. Faster RCNN and RetinaNet were significantly better than non-expert human 3 in predicting *Giardia* and *Cryptosporidium*.

Table 6: Table showing the p-values for *Giardia* - using Wilcoxon signed rank test - among expert humans, non-expert humans, and AI. (p-value threshold of 0.05).

	Expert	Non-expert 1	Non-expert 2	Non-expert 3	Faster RCNN	RetinaNet	YOLOv8s
Non-expert 1	1.64E-04						
Non-expert 2	4.72E-04	1.29E-08					
Non-expert 3	1.98E-04	4.04E-08	7.62E-01				
Faster RCNN	2.19E-09	2.60E-02	5.71E-13	7.88E-12			
RetinaNet	1.08E-09	2.59E-02	1.02E-13	2.60E-12	7.79E-01		
YOLOv8s	1.10E-13	1.57E-05	5.18E-17	6.42E-15	4.50E-05	7.74E-05	
Deformable DETR	4.09E-10	1.16E-3	1.16E-14	6.04E-21	3.56E-10	1.40E-24	6.69E-3

Table 7: Table showing the p-values for *Cryptosporidium* - using Wilcoxon signed rank test - among expert humans, non-expert humans, and AI. (p-value threshold of 0.05).

	Expert	Non-expert 1	Non-expert 2	Non-expert 3	Faster RCNN	RetinaNet	YOLOv8s
Non-expert 1	1.14E-04						
Non-expert 2	3.03E-21	5.52E-25					
Non-expert 3	9.37E-03	8.51E-10	8.88E-21				
Faster RCNN	5.60E-04	2.62E-02	1.86E-24	4.33E-09			
RetinaNet	1.37E-04	1.56E-01	3.52E-26	6.40E-10	3.23E-01		
YOLOv8s	6.67E-02	8.97E-04	2.01E-23	1.82E-06	5.43E-03	6.33E-04	
Deformable DETR	7.58E-2	3.34E-08	2.16E-22	5.74E-08	8.00E-08	7.04E-2	3.63E-10

A.3 Choice of Thresholds

To calculate the precision, recall, and F1-score of the object detection models, we plotted the precision-recall of the respective detectors for different values of iou thresholds and confidence scores, ranging from 0.1 to 1 with a step size of 0.1, as shown in Figure 5 and Figure 4. Then, we selected the thresholds by observing the respective plots to get the best precision and recall.

4. <https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.wilcoxon.html>

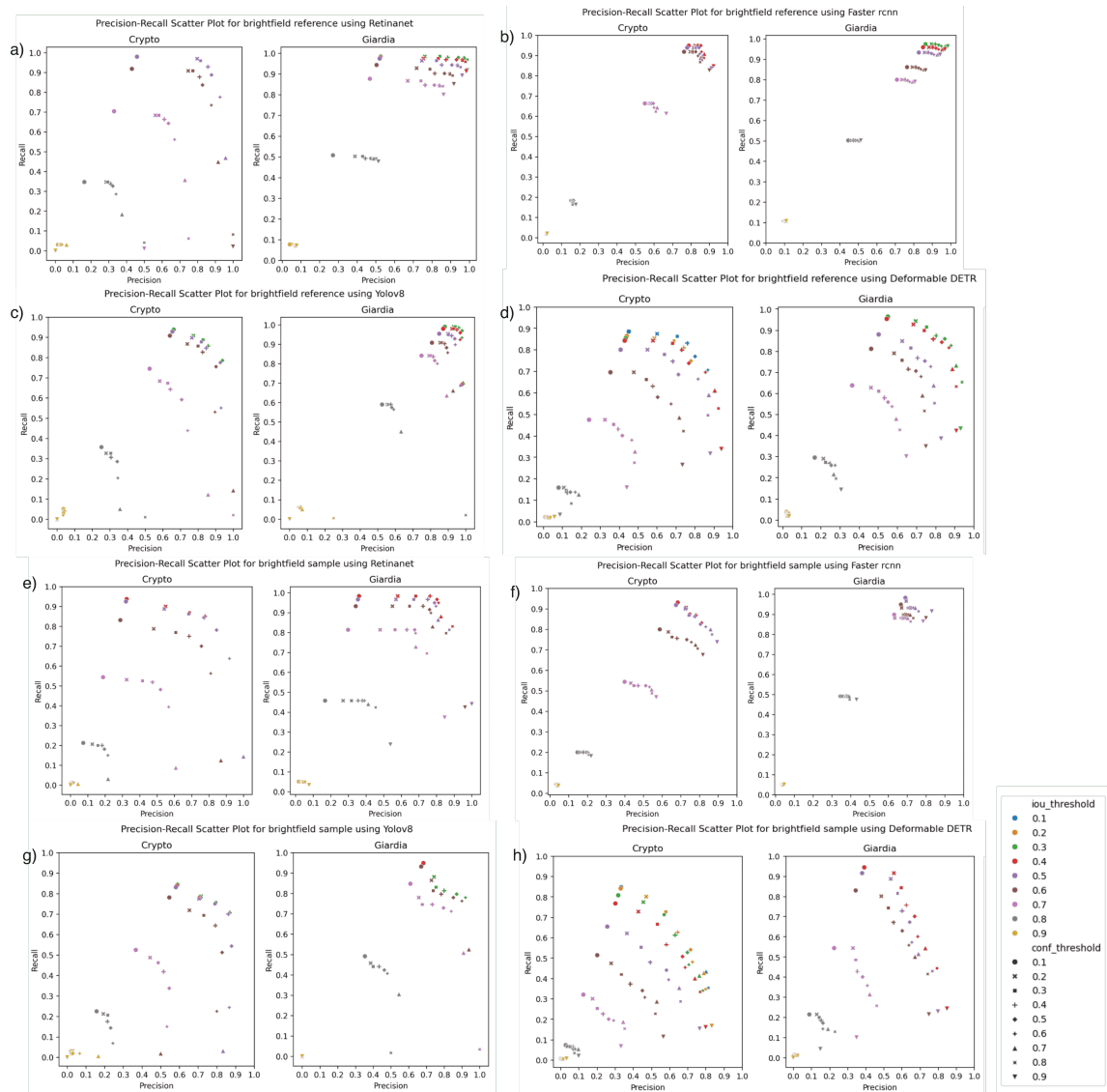


Figure 4: Precision-Recall scatterplot for different values of iou threshold and confidence score for brightfield reference samples and vegetable samples on four detection models. Each subfigure displays scatterplots for (a) brightfield reference sample using RetinaNet, (b) brightfield reference sample using Faster RCNN, (c) brightfield reference sample using YOLOv8s, (d) brightfield reference sample using Deformable DETR, (e) brightfield vegetable sample using RetinaNet, (f) brightfield vegetable sample using Faster RCNN, (g) brightfield vegetable sample using YOLOv8s, and (h) brightfield vegetable sample using Deformable DETR.

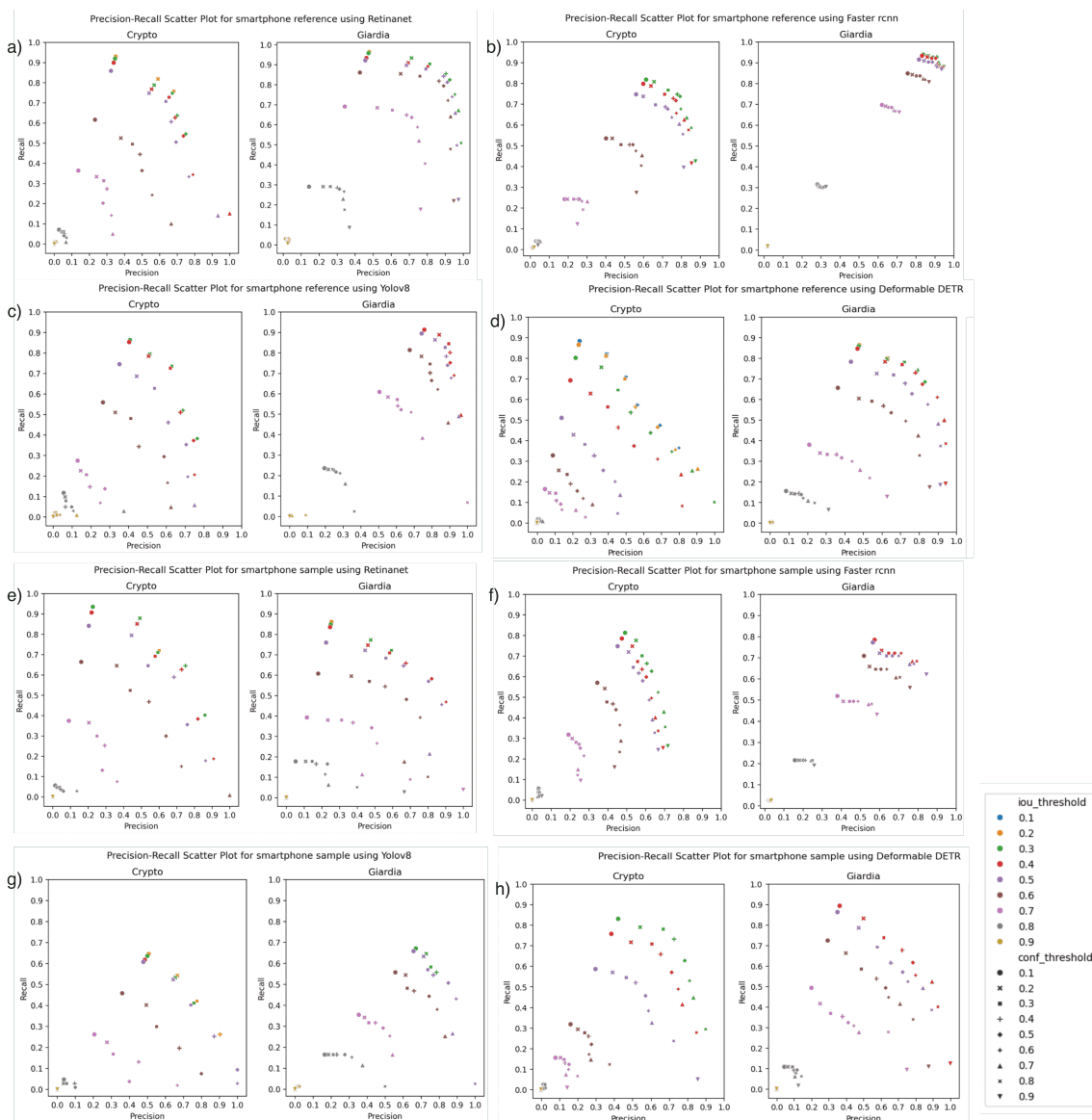


Figure 5: Precision-Recall scatterplot for different values of iou threshold and confidence score for smartphone reference samples and vegetable samples on four detection models. Each subfigure displays scatterplots for (a) smartphone reference sample using RetinaNet, (b) smartphone reference sample using Faster RCNN, (c) smartphone reference sample using YOLOv8s, (d) smartphone reference sample using Deformable DETR, (e) smartphone vegetable sample using RetinaNet, (f) smartphone vegetable sample using Faster RCNN, (g) smartphone vegetable sample using YOLOv8s, and (h) smartphone vegetable sample using Deformable DETR.