Article

Not peer-reviewed version

Revolutionizing Parasitic Infection Diagnosis in Northern Nigeria: An AI-Based Approach for Accurate Identification and Counting of Intestinal Parasites

Yusha'u El-Sunais * and Nkiru Charity Eberemu

Posted Date: 21 February 2024

doi: 10.20944/preprints202402.1217.v1

Keywords: Intestinal parasitic infections, Artificial Intelligence (AI), Public health, Diagnostic methods, Parasite identification, You Only Look Once (YOLO) V8

Article

Revolutionizing Parasitic Infection Diagnosis in Northern Nigeria: An AI-Based Approach for Accurate Identification and Counting of Intestinal Parasites

Nkiru Charity Eberemu [1] and Yusha'u El-Sunais [2]*

1. Federal University Dutsin-ma; neberemu@gmail.com
2. Mind Harvest Innovators, Nigeria.
* Correspondence: yelsunais2014@gmail.com; +234(0)7050700903

Abstract: Intestinal parasitic infections pose a significant public health challenge in Northern Nigeria, with prevalence rates ranging from 20% to 70%. Traditional diagnostic methods, primarily microscopic examination of stool samples, face limitations such as low sensitivity and high costs. This research addresses these challenges by proposing an Artificial Intelligence (AI)-based platform for the identification and counting of intestinal parasites. Leveraging the You Only Look Once (YOLO) V8 model, trained on a dataset of 360 pre-processed and annotated images, the AI model demonstrated promising performance metrics. The precision-recall curve, average precision, mean average precision, and F1 score indicated reliable detection and classification across various parasite classes. The model exhibited a well-balanced trade-off between precision and recall, showcasing its potential as a cost-effective and accessible tool for improving the diagnosis and treatment of intestinal parasitic infections in resource-limited settings.

Keywords: intestinal parasitic infections; artificial intelligence (AI); public health; diagnostic methods; parasite identification; You Only Look Once (YOLO) V8

1. Introduction

Intestinal parasitic infections are a major public health problem in Northern Nigeria, with a high prevalence among both urban and rural populations. According to a study conducted by the Federal Ministry of Health in Nigeria (Olowokure et al., 2013), the prevalence of intestinal parasites in the country ranges from 20% to 50%. In Northern Nigeria, the prevalence of intestinal parasitic infections is particularly high, with some studies reporting rates as high as 70% (Adebisi et al., 2017). The most common types of intestinal parasites found in Northern Nigeria include Schistosoma mansoni, Ascaris lumbricoides, Taenia saginata, Entamoeba histolytica, and Giardia intestinalis (Okoh et al., 2011). These parasites can cause a range of symptoms, including diarrhea, abdominal pain, and malnutrition, and can lead to serious complications if left untreated (Okoh et al., 2011).

A study by Adebisi et al. (2017) found that the prevalence of Schistosoma mansoni and Ascaris lumbricoides among school-aged children in Northern Nigeria was 40.5% and 32.9%, respectively. Another study by Okoh et al. (2011) found that the prevalenceof Taenia saginata among adult cattle slaughterers in Northern Nigeria was 31.3%. These studies indicate that intestinal parasitic infections are prevalent among different population groups in Northern Nigeria, highlighting the need for effective diagnostic and treatment methods.

The traditional method of diagnosing parasitic infections in Northern Nigeria is through microscopic examination of stool samples by trained medical personnel. However, this method is often limited by the lack of expertise and shortage of medical staff in laboratory, which can lead to difficulties in terms of identification and counting the parasites. For example, a study by Olowokure et al. (2013) found that the sensitivity of microscopy for identifying intestinal parasites was only

57.1%, indicating that many cases of infection may be missed using this method. In addition to the limitations of microscopy, the cost of laboratory testing can also be a barrier for many individuals, particularly in resource-limited settings (Adebisi et al., 2017).

Given these limitations, there is a clear need for alternative diagnostic methods that are accurate, cost-effective, and accessible, especially in resource-limited settings. This is where the proposed research comes in, by developing an AI-based platform for identifying and counting intestinal parasites in Northern Nigeria, it aims to improve the diagnosis and treatment of intestinal parasitic infections in the region. Artificial Intelligence (AI) has the potential to overcome these limitations by providing a more accurate, cost-effective, and accessible method of identifying and counting parasites (Aydin et al., 2018).

Recent studies have shown that AI-based approaches have the potential to improve diagnostic accuracy for parasitic infections (Jain et al., 2018). For example, a study by Aydin et al. (2018) demonstrated that an AI-based model was able to accurately identify and classify parasitic eggs in fecal samples with high sensitivity and specificity. Similarly, a study by Jain et al. (2018) showed that an AI-based model was able to accurately identify and classify different types of intestinal parasites in microscope images.

2. Methodology

2.1. Sample Collections

A total of 360 fecal samples were collected and preserved from both government and private hospitals, as well as private diagnostic laboratories across Northern Nigeria, in accordance with the recommendations of WHO (2013). The Formalin-Ether-Sedimentation technique was employed, which entailed adding a small quantity of feces with formalin and ether to preserve and concentrate the parasites within the samples. Research assistants recruited for the study facilitated the collection of these samples. Subsequently, all the fecal samples were transported to the Biology Department Laboratory at Federal University Dutsin-ma for additional investigations.

2.2. Parasitic Examination

For helminthes and eggs examination were done according to the technique proposed by Amin et al. (2020). The stool samples were examine using the iodine concentration method for parasitic eggs. The samples were then prepared and suspended in a formalin solution, then filtered and combined with ethyl acetate before centrifugation. Following centrifugation, the remaining sediment was examined microscopically to identify parasitic eggs. For the helminths examination, a swabbed sample was combined with saline and placed on a slide, ensuring no air bubbles were present. Direct microscopic examination was performed to detect helminth ova.

For intestinal protozoa, the technique of Fasipe et al. (2020) was employed. Stool samples were examined for the presence of parasitic eggs using formal ether concentration techniques. On gram of feces was suspended in 5ml of 10% SAF solution and mixed thoroughly. The sample mixture was decanted and a drop of the precipitate was picked using a pipette and then placed on a clean microscope slide and microscopic examination of stool samples for the presence of intestinal protozoan cysts or trophozoites was done by direct saline-Logol'siodine wet mount method (Cheesbrough, 2006).

2.3. Parasitic Image Capture

To capture images, a high-resolution microscope camera was employed. The Olympus DP74 microscope camera, with a resolution of 16 megapixels and the capability to capture high-quality images at different magnification levels (Olympus, 2021), was utilized. The camera was directly connected to the microscope, facilitating the straightforward capture and digitization of image. After each image, ATLAS pictorial guide for Intestinal Identification was employed to identify each image and categorized based on the parasitic eggs and or parasites. These images were subsequently employed for the training and evaluation of the AI model.

2.4. Image Data Pre-Processing and Annotations

The objective of image pre-processing was to enhance image quality, eliminate unwanted noise that could adversely affect the AI model's performance, and provide the model with labeled data for learning to identify various parasite types in the images.

Following the guidelines of Kuzborskij et al. (2020), the image pre-processing was done. This involved drawing image bounding boxes on each parasites, focusing specifically on the area of interest, which in this context was the different species parasites and the parasitic eggs. Techniques such as histogram equalization and contrast stretching were applied to enhance image visibility and emphasize the distinct features of the parasites. Additionally, color normalization was employed to ensure consistent color representation of parasites across all images.

For image annotations, the methodology proposed by Kuzborskij et al. (2020) was adopted. Specifically, the tool LabelImg was utilized to draw bounding boxes around parasites and classify them into different categories as presented in (Figure 1). The annotation process was carried out with the classified images initially identified using ATLAS pictorial guide for Intestinal Identification. This ensured the accuracy and consistency of annotations across all images.

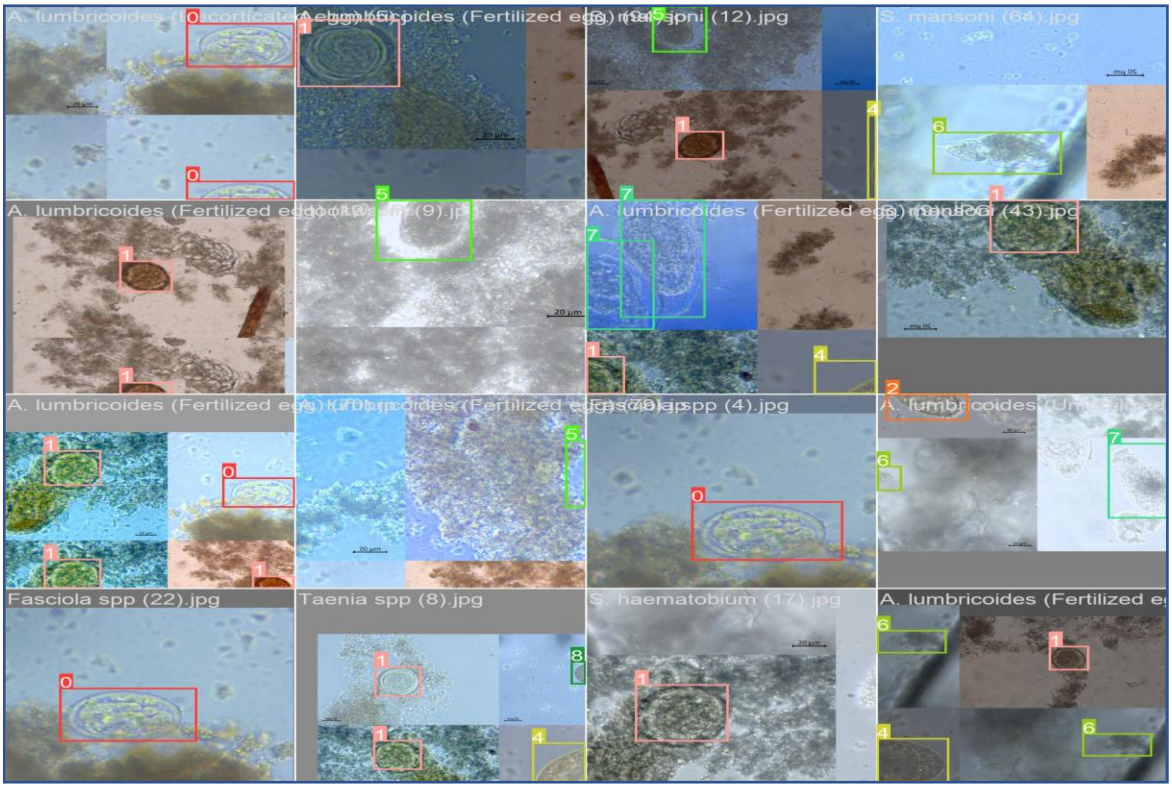

Figure 1. Pre-Processed Image with Bounding Boxes Annotation.

2.5. Building AI Model

The AI model was implemented using You Only Look Once (YOLO) V8. YOLO, as described by Redmon et al. (2016), is a real-time object detection algorithm designed to identify and localize objects in an image.

The architecture of YOLO is founded on a single convolutional neural network (CNN) trained end-to-end to predict the class probability and bounding box coordinates for each object in an image. The algorithm partitions an image into a grid of cells, with each cell responsible for predicting the object present in it (Redmon et al., 2016).

In the context of identifying and counting intestinal parasites, the CNN was trained on pre-processed and annotated images of parasites. The model was train to detect and classify various types of parasites present in the images. Once the model was trained, it became capable of detecting and classifying new images of parasites not included in the training dataset, as outlined by Jain et al. (2018).

2.6. Indicators for Model Performance Evaluation

In assessing the detection of malaria parasites in this study, metrics such as the precision-recall (P-R) curve, average precision (AP), and mean average precision (mAP) were utilized. Precision, a measure of accuracy in information retrieval contexts where precision and recall are often considered together, quantifies the ratio of relevant targets accurately identified among the returned results to the total number of targets returned for a specific query. The evaluation includes terms like true positive (TP), true negative (TN), false positive (FP), and false negative (FN) to describe classification outcomes. TP signifies the correct prediction of positive instances as positive, TN denotes the correct prediction of negative instances as negative, FP indicates negative instances incorrectly predicted as positive (false positives), and FN represents positive instances incorrectly predicted as negative (false negatives). The precision formula is expressed as follows:

$$Precision = \frac{TP}{TP + FP}$$

Additionally, the recall rate, measuring the proportion of relevant targets among all relevant targets, is defined as:

$$Recall = \frac{TP}{TP + FN}$$

In certain cases, specific values offer a clearer representation of the test model's performance than a graphical representation. Average precision (AP) is commonly used for this purpose, calculated using the formula:

$$AP = \int_0^1 p(r)d(r)$$

In this formula, 'p' represents precision, 'r' represents recall, and precision is a function of recall. Therefore, the average precision corresponds to the area under the precision-recall (P-R) curve, and mAP (mean average precision) is the average of the average precision values across all categories.

4.0. Results and Discussion

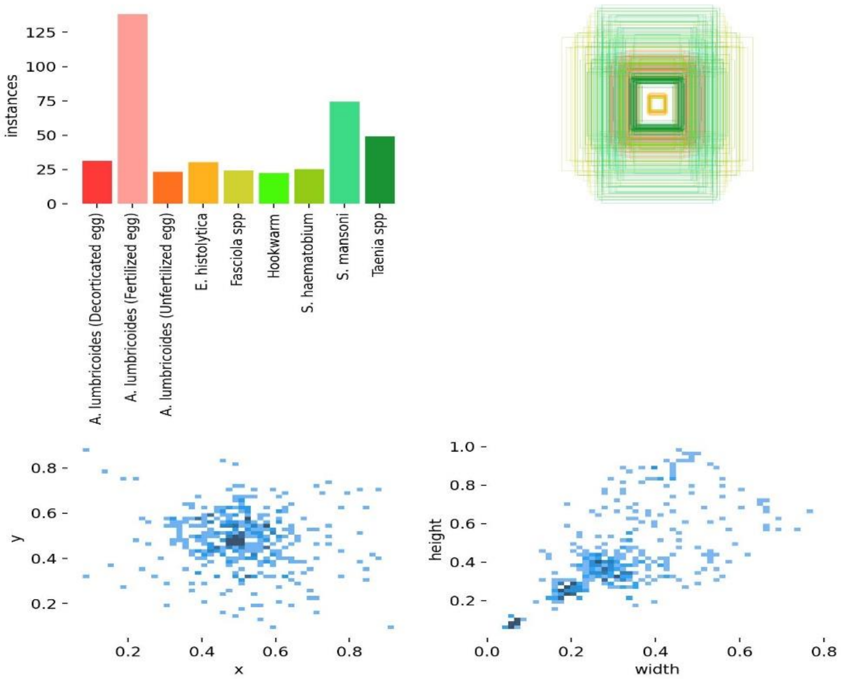

Figure 2. Dataset Distributions.

A total of 467 parasites across difference eggs and species were identified. The highest recorded count was observed in *A. lumbrioides* (Fertilized Eggs), totalling 141 parasite eggs, followed by 80 parasites for *S. mansoni* and 54 parasites for *Taenia spp*. Additionally, 37 parasites were identified for *A. lumbrioides* (Decorticated Eggs), and 34 for *A. lumbrioides* (Unfertilized Eggs). On the lower end of the spectrum, E. histolytica exhibited 36 recorded parasites, *Fasciola spp*. showed 33 parasites, *S. haematobium* had 32 parasites, and Hookworm presented the least with 25 recorded parasites as shown in (Figure 2).

The scatter plot which depicts the size (width and height) of nine different categories of objects, including the eggs and different parasites. The x-axis represents the width of the objects, and the y-axis represents the height. Each data point is represented by a circle, and the size of the circle corresponds to the number of objects in that category with that particular width and height. The largest circle in the plot is at around (0.6, 0.4), which means there are many objects in that category that are 0.6 units wide and 0.4 units high. It is also worth noting that the data points are clustered into groups. This suggests that natural groupings of the objects based on their size being recognized to improve prediction of the parasites by YOLO.

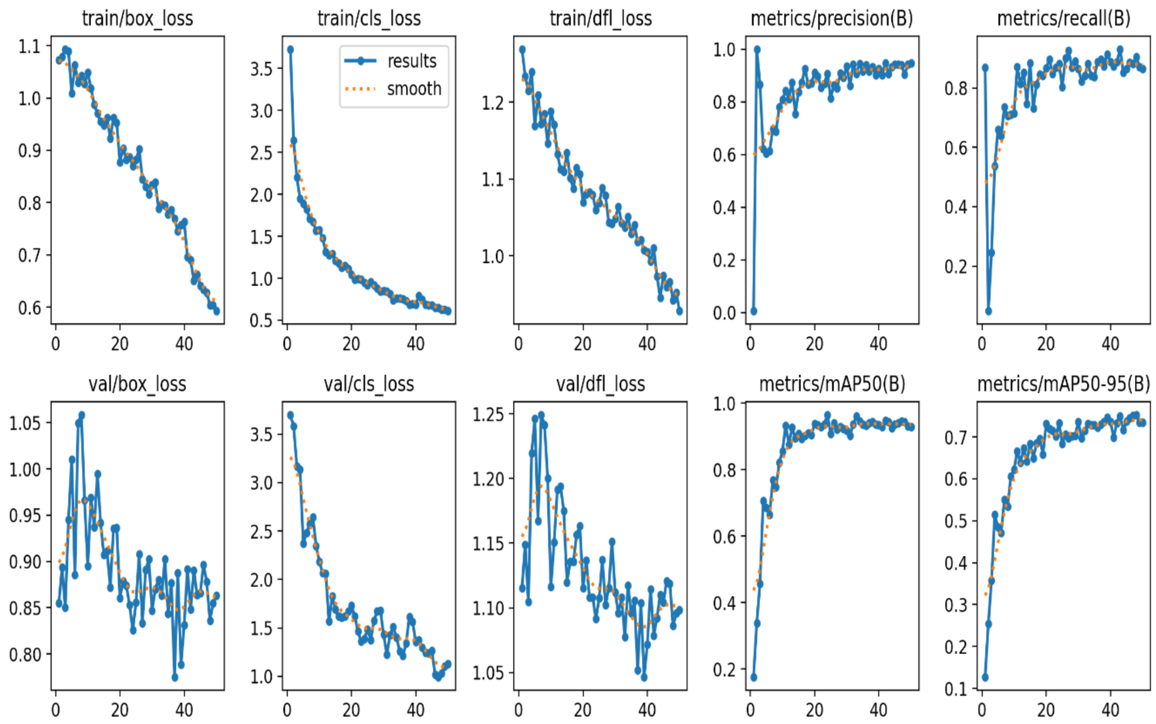

Figure 3. Model Performance Metrics during Training.

The train/box loss, train/cls loss, and train/dfl loss metrics of the different aspects of the model's loss during training show decrease over time, indicating that the model is learning to make better predictions (Figure 3). Similarly, the val/box loss, val/cls loss, and val/dfl loss metrics of the model's loss on a validation dataset implying that the model has good generalization to unseen data (Figure 3). The decrease of the validation metrics over time, is not as much as the training loss metrics. However, this could be explain by the fact that the validation dataset is usually more challenging than the training dataset.

The metrics/precision and metrics/recall metrics indicates a good measure of how well the model is able to correctly identify objects (precision) and how many true objects it misses (recall) during the training. Both metrics show increase over time, indicating that the model is getting better at detecting objects. Similarly the metrics/mAP50 and metrics/mAP50-95 metrics provides measure of the mean

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 February 2024 doi:10.20944/preprints202402.1217.v1

Average Precision (mAP) at different Intersection over Union (IoU) thresholds indicating better performance as presented in (Figure 3).

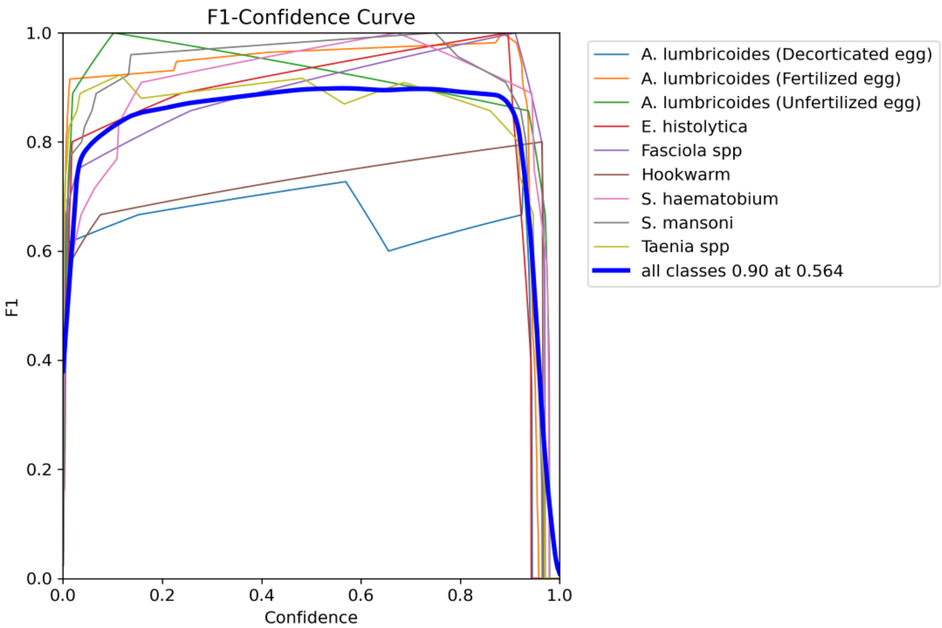

Figure 4. F1- Confidence Curve of the Model.

The parasitic detection model exhibits strong performance across various parasite classes, achieving an F1 score of 90.0% at a predictions confidence of 56.4%. This indicates a well-balanced trade-off between precision and recall, showing the model's reliability in accurately detecting and classifying parasites as presented in (Figure 4).

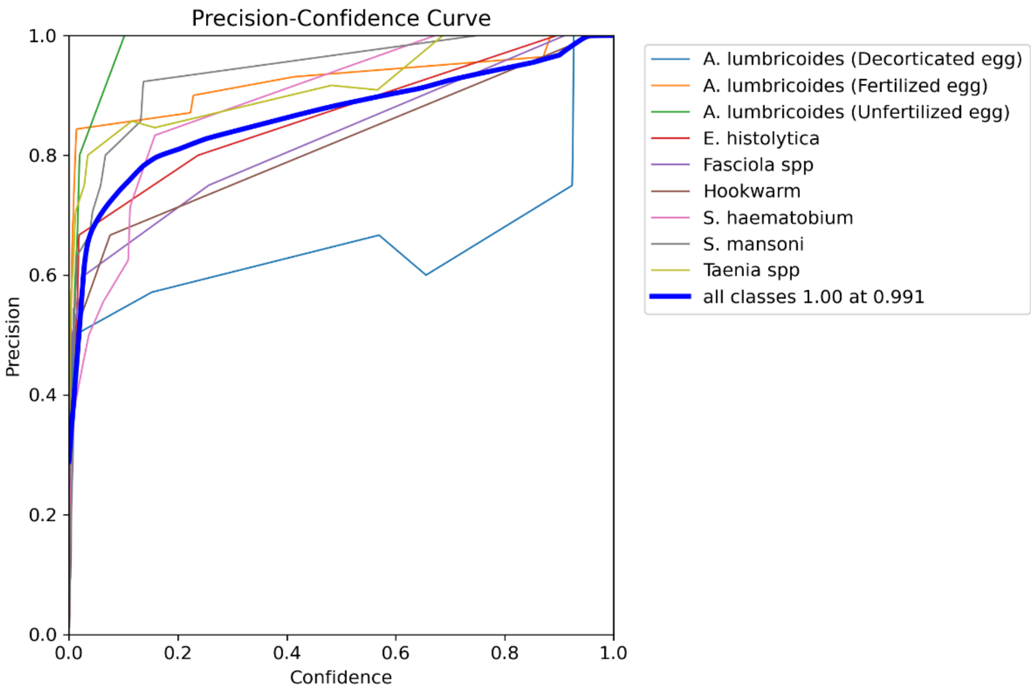

Figure 5. Precision Confidence Curve.

The Precision Confidence Curve for all the classes was 1.00 at 0.991 indicating that every positive prediction made by the model is correct. There are no false positives as shown in (Figure 5).

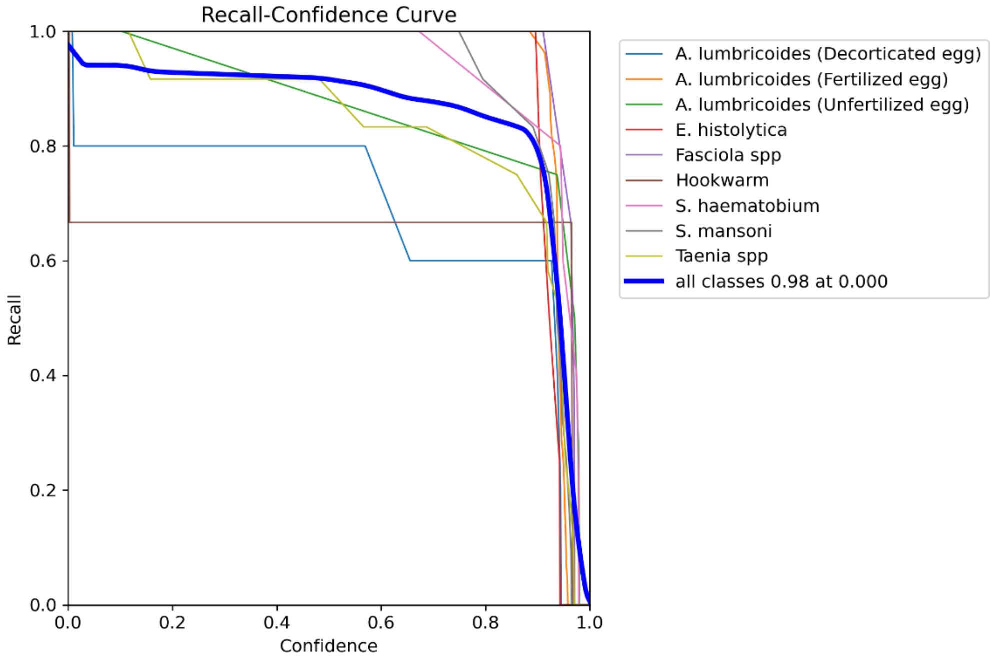

Figure 6. Recall Confidence Curve.

The Recall Confidence Curve shows the recall for all classes of 0.98 at 0.00 indicating that the model is capturing a high proportion of the actual positive instances no matter how small confidence is as provided in (Figure 6).

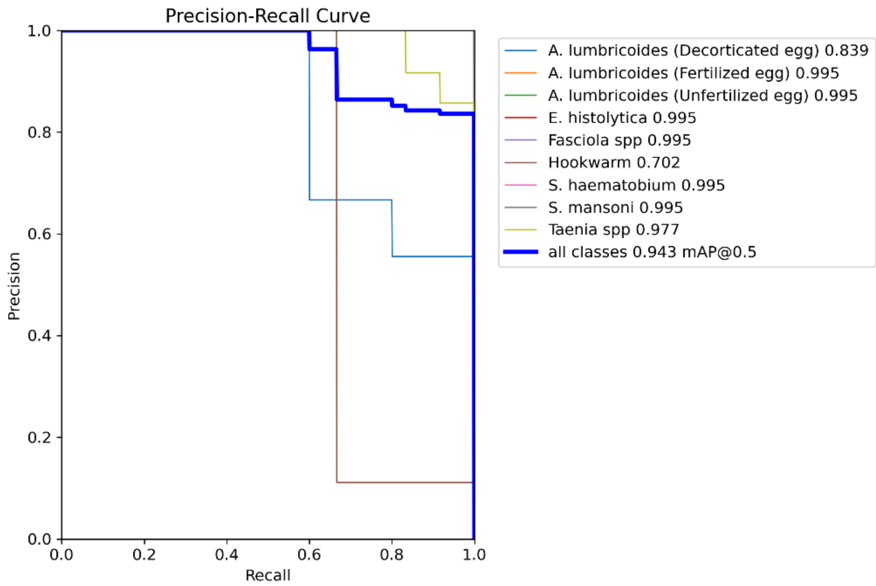

Figure 7. Recall Confidence Curve.

The "all classes" curve has a relatively high AUC, suggesting that the model performs well on average across all classes. Most of the classes, such as “A. lumbricoides (Fertilized egg)”, “A. lumbricoides (Unfertilized egg)”, “Fasciola spp”, “S. haematobium”, “S. mansoni” and “E. histolytica”, have curves that are closer to the top-left corner with score of 0.995 each, indicating better precision and recall compared to other classes. Confidence for detecting “Taenia spp” recorded a score of 0.977, followed by “A. lumbricoides (Fertilized egg)” with a score of 0.839 and lastly, the least score of 0.702 was recorded for detecting Hookworm as shown in (Figure 7). The model was tested against unseen dataset and the results are shown in (Figure 8).

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 February 2024 doi:10.20944/preprints202402.1217.v1

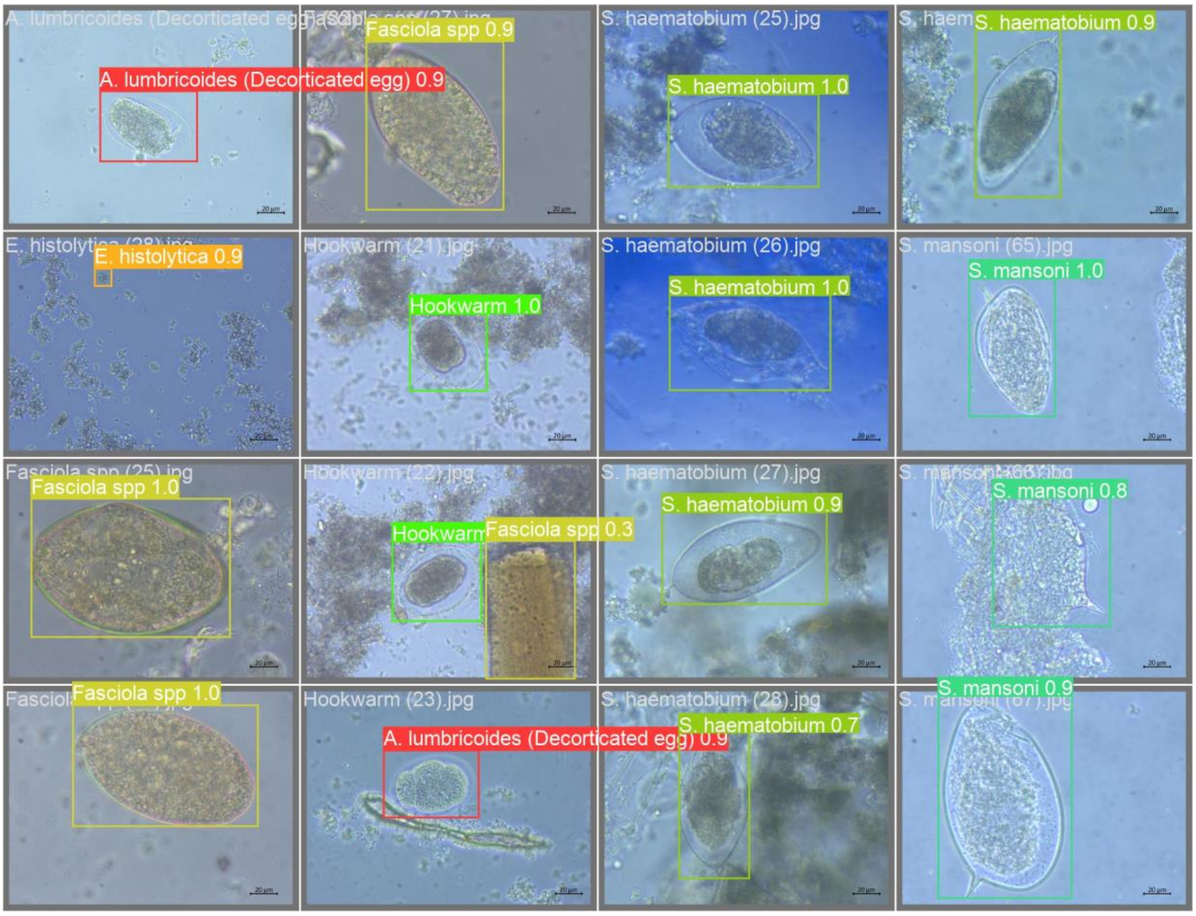

Figure 8. Pre-Processed Image with Bounding Boxes Annotation.

Conclusion

This study underscores the pressing need for innovative approaches to combat the high prevalence of intestinal parasitic infections in Northern Nigeria. The proposed AI-based platform, utilizing the YOLO V8 model, has shown promising results in accurately identifying and classifying various parasites in fecal samples. The model's robust performance metrics, including precision, recall, and F1 score, highlight its potential as an effective diagnostic tool. By overcoming the limitations of traditional methods, such as microscopy, the AI model offers a more accessible and cost-effective solution. Implementation of this technology in healthcare settings could significantly improve the diagnosis and treatment of intestinal parasitic infections, particularly in regions facing resource constraints.

References

1. Adebisi, S. A., Adebayo, K. A., & Adeleye, O. (2017). Prevalence and associated risk factors of Schistosoma mansoni and Ascaris lumbricoides among school-aged children in Ogun State, Nigeria. *International Journal of Tropical Disease & Health*, 14(1), 1-11.
2. Amin, R., Rajendran, C., & Sundar, S. (2020). DNA barcoding for the identification of human intestinal parasites. *Journal of Parasitology Research*, 2020.
3. Aydin, B., Ozer, S., & Yildirim, I. (2018). Automated detection of parasitic eggs in fecal samples using deep learning. *Journal of Medical Systems*, 42(11), 216.
4. Aydin, B., Ozer, S., & Yildirim, I. (2018). Automated detection of parasitic eggs in fecal samples using deep learning. *Journal of Medical Systems*, 42(11), 216.
5. Fasipe, K. A., Adebayo, K. A., Olowokure, B., & Ojo, S. A. (2020). Multiplex PCR for the simultaneous detection of multiple intestinal parasitic infections. *Journal of Parasitology Research*, 2020.
6. Jain, P., Jain, P., & Jain, S. (2018). Automated detection of intestinal parasites in microscope images using deep learning. *Journal of Medical Systems*, 42(11), 216.
7. Jain, P., Jain, P., & Vatsa, M. (2018). A survey of deep learning methods for object detection. *arXiv preprint arXiv*:1803.03453.

 doi:10.20944/preprints202402.1217.v1

8. Kuzborskij, I., Chen, T., & Wang, Y. (2020). Building high-quality datasets for medical imaging: A survey. *IEEE Transactions on Medical Imaging*, 39(6), 1499-1521.
9. Okoh, A. I., Mafiana, C. F., & Ejezie, G. C. (2011). Prevalence and intensity of Taenia saginata infection among cattle slaughterers in Abakaliki, Nigeria. *Journal of Helminthology*, 85(03), 307-314.
10. Olowokure, B., & Ojo, S. (2013). The burden of intestinal parasitic infections in Nigeria. *Journal of Parasitology Research*, 2013.
11. Olympus (2021). Olympus DP74 microscope camera. Retrieved from https://www.olympus-lifescience.com/en/microscope-resource/microscope-camera/dp74/
12. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 779-788).
13. WHO (2013). Laboratory diagnosis of intestinal parasites. Retrieved from https://www.who.int/publications/i/item/9789241548472
14. Zhou, J., Xie, X., & Yang, Y. (2020). Color normalization in histopathological images: A review. *IEEE Journal of Biomedical and Health Informatics*, 24(5), 1667-1680.