**Intestinal Worms Final Technical Report**

*December 6, 2022*

Global Solutions VIP

Team Members:

Anna Hart

Nathaly Jose-Maria

Siri Reddi Tanguturi

Raghav Raahul Manoharan Jayanthi

Esther Yael Shalom

ADVISORS:

Dr. James Stubbs, PhD

Dr. Kelsey Kubelick, PhD

Kaveen Vijay Kumar, TA

**Executive Summary**

A major issue posed to global health efforts is soil transmitted helminth (STH), particularly among children, and diagnosis and treating the communities in need. To do this, we aim to develop an automated diagnosis system to enable faster, more reliable diagnosis of STH to enable public health workers to screen and treat communities (Childrens Without Worms, 2021a; Children Without Worms, 2021b). Our goals this semester were two-fold: from the computer science side, we aimed to analyze the code from Ward's paper, propose modifications, and prototype our new model (Ward et al., 2022a, Ward et al., 2022b). From the hardware side, we aimed to increase the compactness of the microscope scanner base and automate the device's rotation. We succeeded in accessing the code which classifies intestinal worm eggs and modified the notebook to make it more accessible for future team members (Ward et. al, 2021b). Furthermore, we translated our use-case into specifications for our new prototype, ordered by priority. From the hardware side, we prototyped a new base design and researched potential solutions for automatic rotation. This work lays the base for future semesters to complete prototyping of both the hardware and software.

**Introduction**

Soil transmitted helminths (STH) are a significant blight within the realm of neglected tropical diseases that pose immense risks worldwide particularly to children and pregnant mothers. Endemic STH infections are considered widespread community issues by the World Health Organization once more than 2% of a population is affected to a moderate or severe degree. However, there does not currently exist an efficient, effective way to determine the most at-risk communities requiring more urgent resource allocation. The major STH species of concern at the moment are roundworms, whipworms, threadworms, and hookworms (Childrens Without Worms, 2021a, Childrens Without Worms, 2021b). The Kato Katz technique is a manual process that counts the number of worm eggs present in stool samples and is regarded as the goal standard (Mbong et al., 2020). However, this process exposes trained health professionals to noxious materials for long periods of time, therefore hindering their quality of work and data accuracy as longer work hours in this type of environment increases human error (Children Without Worms, 2021b; Akintayo et al., 2018). We seek to minimize the data collection and processing time given its role as the most time-consuming step of the process (Stuyver & Levecke, 2021). Automating this process will increase the number of samples analyzed at a higher accuracy, therefore enhancing the care patients can receive with a quicker, more convenient diagnosis (Children Without Worms, 2021a; Children Without Worms, 2021b). Currently, a machine learning algorithm exists to classify four species of intestinal worm eggs, but it does not robustly compute the diagnosis and severity of infection to clinical standards (Ward et. al, 2021a). Moreover, we need to develop a 3D-printed device capable of imaging slides with basic microscopy. Uganda, Bangladesh, and Kenya are the main countries of operation for Children Without Worms, and these areas often lack reliable electricity, transportation and laboratories (Children Without Worms, 2021a; Children Without Worms, 2021b; Jain & Keith, 2021). This device and automated STH infection quantification method in unison are necessary to allocate deworming medication in these areas and ultimately mitigate the impact of these infections (Children Without Worms, 2021a; Children Without Worms, 2021b; Butploy et al., 2021).

**Project Goals and Scope**

***CS Subteam*** The overall project goal is to detect, classify, and quantify intestinal worm eggs in stool samples (Children Without Worms, 2021a). In a study conducted, a proof of concept model was developed and trained such that it can detect and identify a single egg and classify it as one of four different species (Ward et al., 2022a). We plan to expand and enhance this proof of concept in two ways: adapt this model to the hardware component of the project by modifying the parameters when training the model and move away from Kaggle so our modified algorithm can be applied in mobile devices and beyond. Our goals this semester were to understand and document the workings of Ward's code, develop a list of specifications for our new model, and begin making modifications to the current code. Our expected outcomes were adapted documentation for the current code, design considerations and specifications for our new model, and potentially the change in model accuracy due to our modifications.

The scope of our project is limited in a few ways. First, there exists little to no annotated sets of data that we can use to test our model excluding the set in the aforementioned paper. Thus, our model will be limited to detecting and classifying only these 4 species of intestinal worms until more datasets are acquired: soil-transmitted helminths (STHs: *Ascaris lumbricoides*, *Trichuris trichiura* and hookworms) and *Schistosoma mansoni* (SCH) (Ward et al., 2022a). Secondly, compatibility between the software and hardware components is difficult to iterate on since we are unable to acquire images of samples taken from the hardware.

***Hardware Subteam***

The project goals on the hardware side of the project were to make improvements to the Fusion360 design for the base to make the base more compact using lesser material. The specifications for the base include a decrease in the amount of material used while maintaining or increasing the strength and durability of the base as measured by stress simulation and buckling tests (The Engineering ToolBox, n.d, SkyCiv 2016). We also aimed to design a rotating motor that would enable automatic rotation of the microscope as it rotates to view each slide with a stool sample. The implementation of this motor will speed up the process of inspecting the stool samples and reduce the human effort involved in the process. Some limitations to the scope of this project involve the lack of electrical and mechanical engineering experience of the team, leading to a need to research the materials needed for circuitry and motor control and dedicated time to learn how to use them.

**Current Project Status**

***CS Subteam***

As of this previous summer, novel code in the field of ACH egg identification was released by a Johnson and Johnson (J&J) research group (Ward 2022a; Ward 2022b). The team sought to spend the majority of this semester fully understanding the J&J group’s code for its own research purposes. With that came additional research into specific aspects of machine learning as it pertained to the code to enhance comprehension.

The J&J code was stored in a Kaggle notebook, a crowd-source platform for data science and machine learning. As this was a new environment to all group members, it took some time to understand how specific aspects of the website worked, how the prewritten code could be executed, and how the code could be manipulated in a different medium independent yet connected to the original source. This led to the development of a notebook duplicated from the original copy with modified wording to ease comprehension among the group and future members (Jose-Maria et al., 2022; Ward et al., 2022b). Re-translating the original text seems menial and unnecessary, but it took a great amount of time to digest all the inner workings of the notebook and thus required adequate documentation to share that understanding with the rest of the team. Moreover, once the team was able to better understand what exactly was occurring, then the team could work better to propose changes that would distinguish our work from the original J&J team. The main goal of the semester was to get to a point of manipulating the original code and proposing changes with subsequent, valuable effects; however, this ended up being a future goal as comprehension of this notebook took much longer than expected, and the notebook could not run independently due to its storage limitations, which is detailed further.

Before discussing the workings of the notebook, it is important to develop context on two key concepts: confusion matrices, and precision versus recall analysis. A confusion matrix works to identify information out of several classifications and errors (Saxena, 2018). Figure 1 is a formally annotated confusion matrix output of the faster\_rcnn\_resnet101\_sthsch\_3 model pipeline configuration pre-trained offline. The purple cell is of note as it highlights the number of eggs that were not counted as part of the ground truth total number of eggs. Other pretrained models are roughly annotated in Figures 3-7. To define precision and recall, the output matrix and precision versus recall results of Figure 2 will be referenced. The precision statistic highlights the number of relevant results (Saxen, 2018). A relevant result is a correct egg tally. For example, for the 833 eggs classified as ascaris, the 809 that were correctly classified would be considered the relevant results as opposed to the 24 that were unclassified but legitimately of the ascaris-type in the ground truth. The recall statistic highlights the accuracy of the relevant results (Saxena, 2018). Continuing with the aforementioned pretrained model, an accuracy percentage can then be derived by looking at the 809 correctly identified ascaris eggs compared to the total 833 eggs identified as ascaris across all egg types presented. For these reasons, we end with a 97.1188% precision and recall rate for ascaris classification with this model (Ward et al., 2022b). All of the pre-trained models’ precision versus recall statistics can be found in Figure 8.

Now as for how the notebook works, it trains and tests a machine learning model that identifies and counts specific types of STH eggs from an input image dataset. The egg types are detailed in an input egg\_label\_map.pbtxt file and can be manipulated as needed (Ward et al, 2022b). The notebook follows a traditional machine learning process of ingesting organized data into a pipeline that is inputted to train, test and run a model (Pachyderm, 2022). As such, the input image dataset is organized by using a base pretrained model and applying transfer learning to set up the STH egg detection model. The particular model chosen in the notebook is an EfficientDet-Do, the smallest model of the EfficientDet family storage-wise, but this can also be manipulated to lessen model training time. From there a model pipeline configuration is imputed in the model to train it, the model undergoes a primary evaluation, the model is exported, and the model undergoes a final secondary evaluation. The secondary evaluation outputs a confusion matrix that details the number of correct egg type predictions and the number of confused egg type predictions. These values are then used to generate a precision versus recall analysis to determine the effectiveness of the model compared to the ground truth that highlights the correct number of each egg type. Since the notebook cannot be run independently, Ward et al. go on to include models pre-trained offline to demonstrate how the code should run and what the output should be, which are the pretrained models referenced during earlier context building of confusion matrices and precision versus recall analysis (Ward et al., 2022b; Figures 1-8).

Given this comprehensive understanding of the code’s individual parts, the team can now go about recommending specific changes to the code, inputs, and parameters, to produce different outputs and demonstrate different results. Some of these changes include batch size, training steps, class label mapping, base pre-trained model, and model pipeline configuration. Once these changes are implemented and optimized, the team can work to incorporate other datasets and ensure the model’s validity. However, this may prove to be an issue in the short-term future as credible STH datasets are currently very difficult to come across. Additionally, these changes require a more powerful computer with more storage to actually see any changes without any errors as seen in the Kaggle notebook (Ward et al., 2022b).

Another piece of our work this semester was identifying properties and functions of the code that we will need to write, and analyzing the gaps between the current code and the needed code. We analyzed the needs in the field compared to the current code from the J&J group, and developed high, medium, and low priority specifications for our new design (Ward et al., 2022b; Children Without Worms, 2021a). High priority specifications include the ability to take in a slide and count the number and species of egg per sample. We also need a user interface which can upload slide images to the cloud, evaluate the slides, and return the information in a format easily understandable to health workers and compliant with patient privacy (Children Without Worms, 2021a, Children Without Worms, 2021b). Our medium priority specifications include defining and evaluating the clinical specificity and sensitivity of the design - which will likely necessitate software to detect overlap between slides (Ward et al, 2022a). Furthermore, we need to detect ambiguous samples for follow up with a pathologist, decrease the number of slides necessary, and develop a user interface that aggregates individual slide data (Children Without Worms, 2021a, Children Without Works, 2021b). Additionally, we need to develop a way to iterate on our design to improve the product as new data evolves. Finally, we have a few low priority specifications, to be worked on after completing the high and medium priority specifications. We could improve the efficacy of the model in order to enable researchers to use an even lower-power computer in the field. J&J’s Tynong's group was able to use the Jetson AGX developer kit, which they claim is accessible for field researchers (Ward et al., 2022a). Finally, we must test our design with multiple microscopes, cameras, and operating systems to give flexibility allowing workers to use the devices that they have access to in the field without the need to purchase new ones (World Health Organization, 2021).

***Hardware Subteam***

On the hardware side of the project, the team sought to ideate automatic slide rotation to be viewed under the microscope and to improve the previous capstone team's SolidWorks model for the container/base holding the lid in place. The model needs to reduce the amount of material used, utilize a more compact shape, firmly attach the lid to the base, and implement a more even bottom surface. We were able to successfully design a new prototype for the base, improving upon the previous Capstone team's model, and conducted analysis of this prototype using Fusion 360. We also researched next steps for designing a rotating motor to automate slide rotation (AUTODESK Instructables, 2020a; AUTODESK Instructables, 2020b). The proposed final design is an automated slide reader using a direct current (DC) motor rotation control. This system will rely on an electronic circuit, mechanical components and Arduino software tools, where a microcontroller controls a motor based on light detection. This motor should be encased inside the paradigm, and its rotational speed should be determined by the time it takes to identify parasitic eggs. The supplies needed are as follows: a L298N motor controller, a breadboard, an ATMega 2560 microcontroller, a 12 V DC motor, an LED emitter and receiver, a 12 volts power supply for the motor, a 5 volts power supply for the lights, two 220 ohms resistors, a 1k ohms resistor, a 10k ohms resistor, a 2N222A transistor, and a BC547 transistor (AUTODESK Instructables, 2020a; AUTODESK Instructables, 2020b). Moreover, the shape and dimensions of the lid were used to create a new base that ensured a more snug, compact fit (Figures 9-12).

We analyzed the new design’s efficiency with Fusion 360 static stress simulations, buckling tests, and physical comparisons. The static stress simulation calculated the safety factor of various materials with a 100N load (Table 2; Figure 13).The recommended safety factor for most materials under normal conditions must be above 2, and the design demonstrated safety factors well over the requirement at 15, thus indicating the structure’s stability (The Engineering ToolBox, n.d.). The physical comparison demonstrated that the new design used considerably less material and had a more compact shape with a mass, volume, area, and Y value of the center of mass being lower by 11.69%, 11.69%, 16,27%, and 30.72% respectively (Table 1). The lower center of mass of the Y value confirmed structural stability as suggested by our background research (Access Physiotherapy, n.d.). Finally, a structural buckling test measured deformation due to stress from loads (SkyCiv, 2016). Even at a force 540.9 times the applied load of 100 N, the majority of the structure stayed intact with some portions deforming between the range of 0 to 1 mm, further validating the stability of the structure (Figure 14).

**Conclusions**

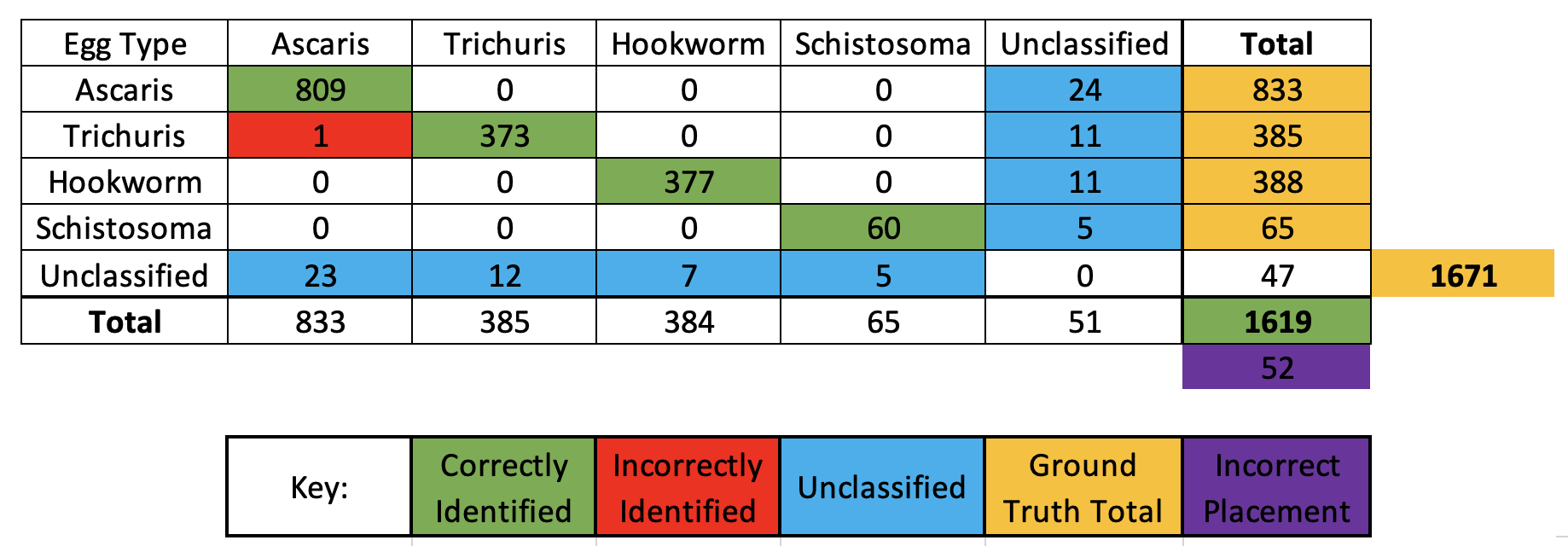
Overall, we aim to contribute to the worldwide mitigation of STH infection through more effective diagnosis and easing health professional work conditions. Utilizing an artificial intelligence solution will provide more accurate diagnoses and subsequent patient treatment lacking trained pathologists (Children Without Worms, 2021a; Children Without Worms, 2021b). Much of our work this semester involved groundwork for future teams. Our computer science subteam adapted the Kaggle notebook by working through the code, understanding and discussing the code, and editing its documentation for readability of future teams (Ward et al., 2022b). We researched the computing requirements and applied for resources which future teams can use to prototype changes to the machine learning model. We also developed a list of high, medium, and low priority specifications for our code that will guide future teams in setting goals. The hardware team improved the compactness of the base, with documentation and prototype testing which will inform future teams as to the specifications, safety, and strength of the model. Furthermore, we researched design possibilities for the automatic rotation of the device which will allow subsequent teams to begin prototyping.

One of the main issues we found with the software team was the lack of computing resources to run the model, as well as less than user-friendly documentation for the existing notebook. We spent considerable time going through the notebook to understand what was occurring and why, and adapted much of the documentation for future teams (Ward et al., 2022b). Additionally, we troubleshooted issues in the code and discovered that it was from a lack of resources. We then researched resources available to us, compared to the resources used in the paper, and discovered that we will need access to Georgia Tech's PACE-ICE cluster (Ward et. al 2022a). From the hardware side, one of the main roadblocks was a lack of experience in electrical engineering. Our team spent considerable time researching circuits and motor control before being able to make progress on the prototyping.

In future semesters, we aim to fully prototype the machine learning model. From our adapted notebook, future teams will be able to easily onboard, and once given access to the PACE-ICE cluster, can begin following our recommendations for modifications to the code. After optimizing the machine learning model, we will begin working on the high-priority specifications by creating a wrap-around algorithm to group slides from patients, detect overlap, and calculate the severity of infection for each of the four egg species. From the hardware side, we will prototype our design research for the automatic rotation of the microscope. Ultimately, we aim to combine the hardware and software aspects to have a software system which links to our hardware and assists researchers in automatically determining severity and type of STH infection across patients and communities.

**Addendum**

**Figure 1: Formally Annotated Confusion Matrix of faster\_rcnn\_resnet101\_sthsch\_3 Model**



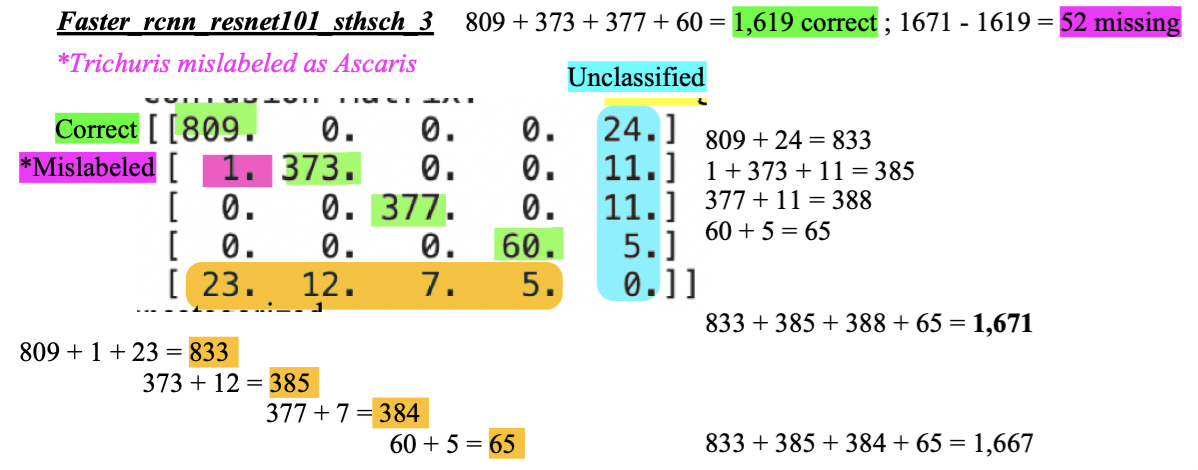
Constructed from raw output of the J&J group’s original Kaggle notebook (Ward et al., 2022b). Figure 2 demonstrates the original raw matrix output from the notebook correlating to this figure. Figure 3 demonstrates the roughly annotated matrix correlating to this figure.

**Figure 2: Raw Output Confusion Matrix and Precision vs Recall of faster\_rcnn\_resnet101\_sthsch\_3**

Confusion Matrix: 
5.] 
[ [809. 
[ 23. 
o. 
373. 
o. 
o. 
12. 
0. 
0. 
377. 
0. 
o. 
o. 
o. 
60. 
5. 
24.] 
11.] 
11.] 
1 
2 
3 
category 
Ascaris 
Trichuris 
Hookwo rm 
Schistosoma 
preclsxon 
0.5101.1 
i. 971188 
o. 968831 
o. 981771 
o. 923077 
o. 971188 
o. 968831 
o. 971649 
o. 923077 

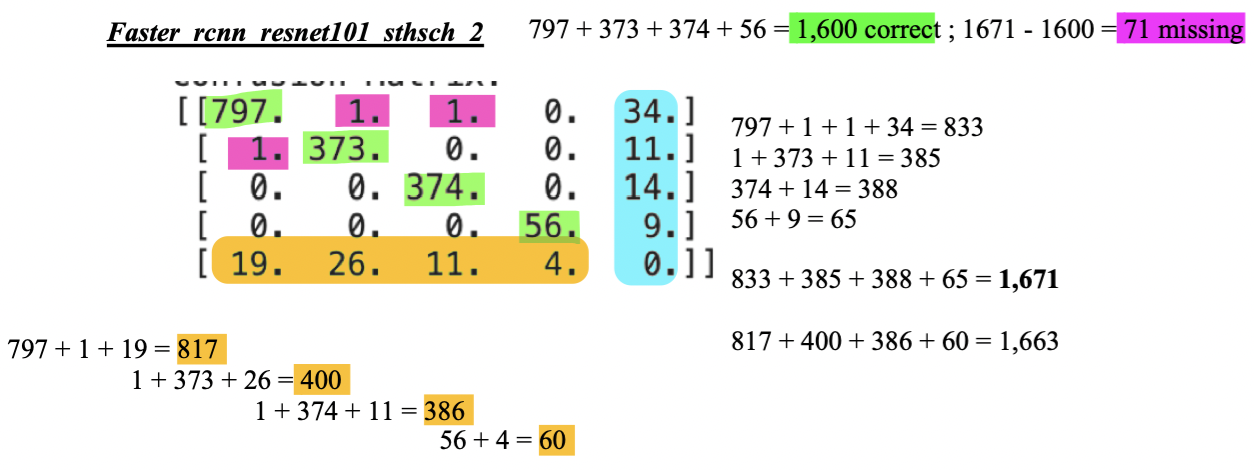
This is the direct output faster\_rcnn\_resnet101\_sthsch\_3 model of the J&J group’s original Kaggle notebook (Ward et al., 2022b). The rows and columns follow the same order as that illustrated in Figure 1 with the exception of the “Total” row and column. The matrix demonstrated the number of one egg type identified as another across both axes. For example, following an xy coordinate system increasing in the right and down directions starting at value 809, position (0,0) represents 809 eggs correctly identified as ascaris. Position (0,1) represents 1 egg incorrectly identified as ascaris despite being trichuris.

**Figure 3: Roughly Annotated Confusion Matrix of faster\_rcnn\_resnet101\_sthsch\_3**



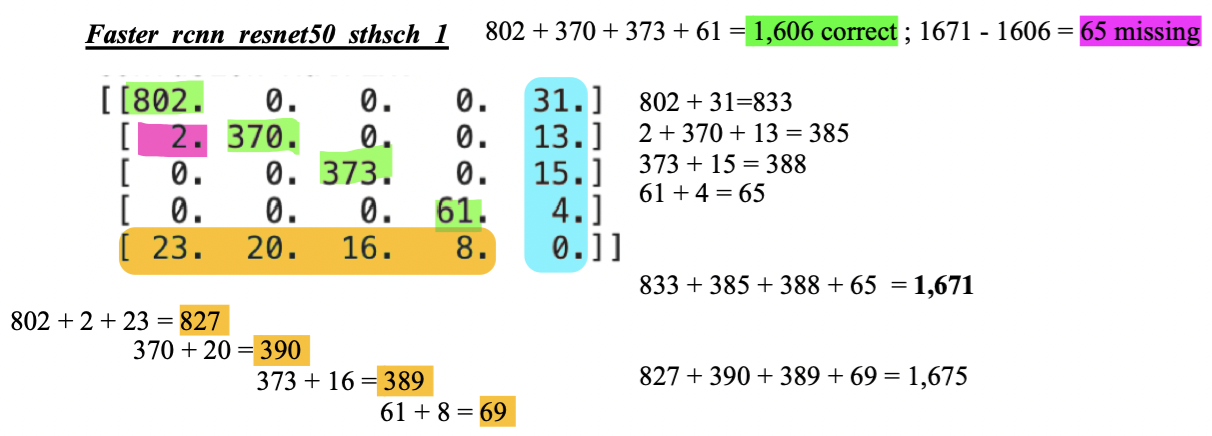
These are annotations on the direct matrix output of the faster\_rcnn\_resnet101\_sthsch\_3 model of the J&J group’s original Kaggle notebook (Ward et al., 2022b). The rows and columns follow the same order as that illustrated in Figure 1. The annotations worked to help identify what exactly the confusion matrices were outputting as it was not explicitly stated anywhere in the original notebook. The key is similar to Figure 1 with green highlight representing correctly identified eggs, pink highlight representing both misidentified eggs and eggs missing from the final ground truth egg count, and both aqua and orange highlight representing unclassified eggs. The figure also shows calculations for the totals of each row and column and the final sum of these totals to identify the number of eggs missing from the final correct count.

**Figure 4: Roughly Annotated Confusion Matrix of faster\_rcnn\_resnet101\_sthsch\_2**



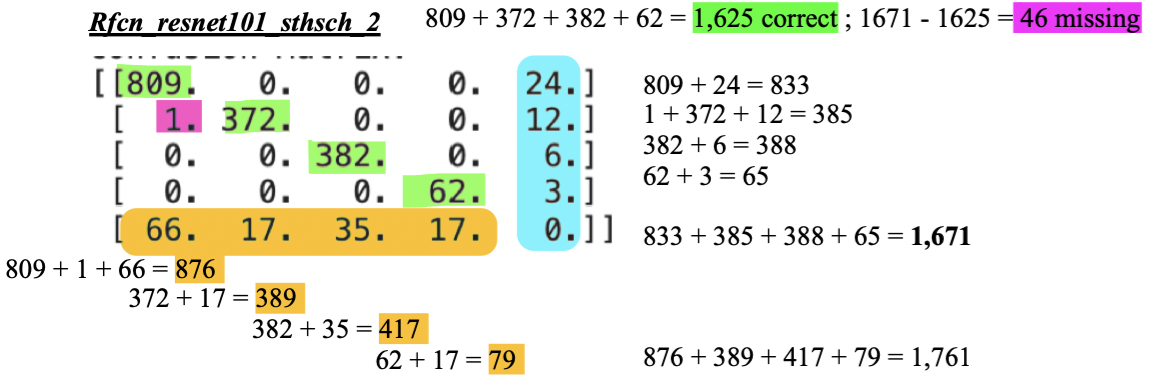
These are annotations on the direct matrix output of the faster\_rcnn\_resnet101\_sthsch\_2 model of the J&J group’s original Kaggle notebook (Ward et al., 2022b). The annotations - color key and calculations - follow the same format as specified for Figure 3: green highlight represents correctly identified eggs, pink highlight represents both misidentified eggs and eggs missing from the final ground truth egg count, and both aqua and orange highlight represent unclassified eggs.

**Figure 5: Roughly Annotated Confusion Matrix of faster\_rcnn\_resnet101\_sthsch\_1**



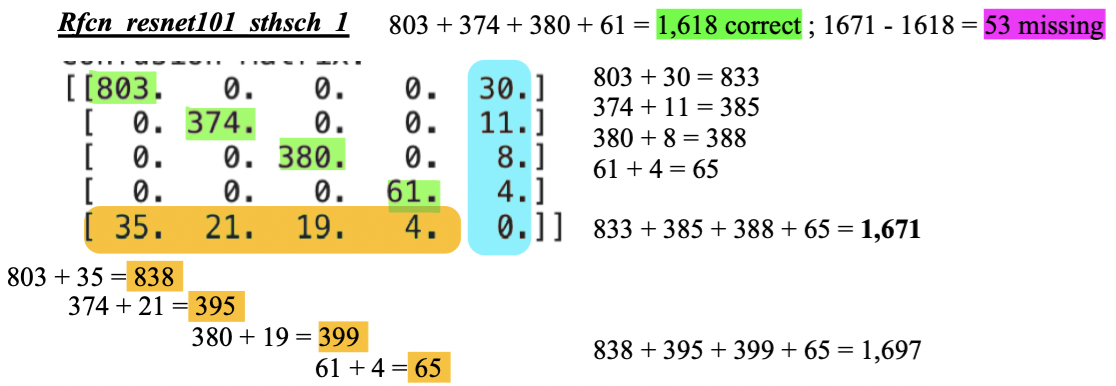
These are annotations on the direct matrix output of the faster\_rcnn\_resnet101\_sthsch\_1 model of the J&J group’s original Kaggle notebook (Ward et al., 2022b). The annotations - color key and calculations - follow the same format as specified for Figure 3: green highlight represents correctly identified eggs, pink highlight represents both misidentified eggs and eggs missing from the final ground truth egg count, and both aqua and orange highlight represent unclassified eggs.

**Figure 6: Roughly Annotated Confusion Matrix of rfcn\_resnet101\_sthsch\_2**



These are annotations on the direct matrix output of the rfcn\_resnet101\_sthsch\_2 model of the J&J group’s original Kaggle notebook (Ward et al., 2022b). The annotations - color key and calculations - follow the same format as specified for Figure 3: green highlight represents correctly identified eggs, pink highlight represents both misidentified eggs and eggs missing from the final ground truth egg count, and both aqua and orange highlight represent unclassified eggs.

**Figure 7: Roughly Annotated Confusion Matrix of rfcn\_resnet101\_sthsch\_1**



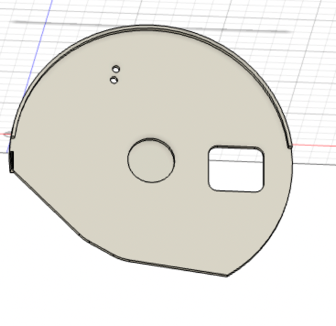
These are annotations on the direct matrix output of the rfcn\_resnet101\_sthsch\_1 model of the J&J group’s original Kaggle notebook (Ward et al., 2022b). The annotations - color key and calculations - follow the same format as specified for Figure 3: green highlight represents correctly identified eggs, pink highlight represents both misidentified eggs and eggs missing from the final ground truth egg count, and both aqua and orange highlight represent unclassified eggs.

**Figure 8: Pretrain Model Precision versus Recall Statistics and Graph**

/kagg le/wo rking/confusion_matrix 
rfcn resnet1Ø1 sthsch 1 
rfcn resnet1Ø1 sthsch 2 
faster rcnn resnet5Ø sthsch 1 
faster rcnn resnet101 sthsch 2 
faster rcnn resnet1Ø1 sthsch 3 
0.970 
• rfcn resnetIOI sthsch 2 
precision 
o. 948978 
0.895173 
o. 940354 
o. 952566 
o. 961217 
recall 
o. 963314 
o. 968951 
o. 955906 
o. 937767 
o. 958687 
0.965 
0.960 
= 0.955 
0.950 
0.945 
0.940 
0.90 
• rfcn resnet 
I sthsch I 
ster rcnn resnetIOI sthsch 3 
• faster rcnn resnet5 sthsch I 
0.91 
• faster rc n resnetIOI sthsch 2 
0.95 
0.96 
0.92 
0.93 
Precision 
0.94 

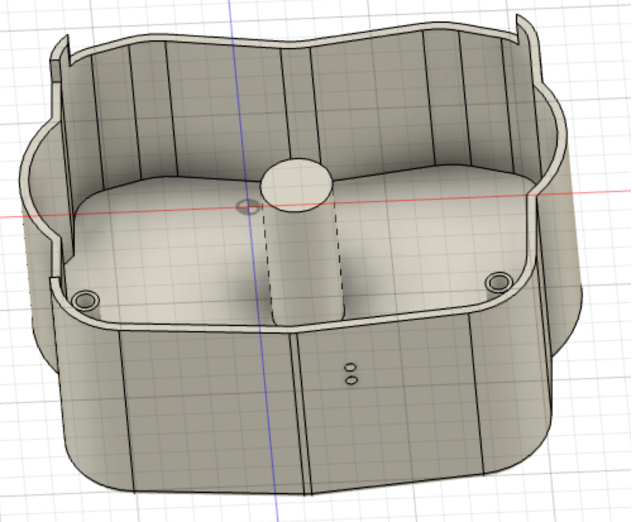
This is the direct aggregate output of all pre-trained models of the J&J group’s original Kaggle notebook (Ward et al., 2022b). The figure demonstrates the precision versus recall analysis of each pre-trained model after going through the already trained model as test models. This is why there is a discrepancy with Figure 2’s precision versus recall statistics as Figure 2 demonstrated the training results as opposed to the testing results. The figure shows that rfcn\_resnet101\_sthsch\_1, faster\_rcnn\_resnet101\_sthsch\_1, and faster\_rcnn\_resnet101\_sthsch\_3 performed within close effectiveness of each other while rfcn\_resnet101\_sthsch\_2 demonstrated immensely high recall but low precision and faster\_rcnn\_resnet101\_sthsch\_2 demonstrating immensely high precision but low recall.

**Figure 9: Reference Lid Design**



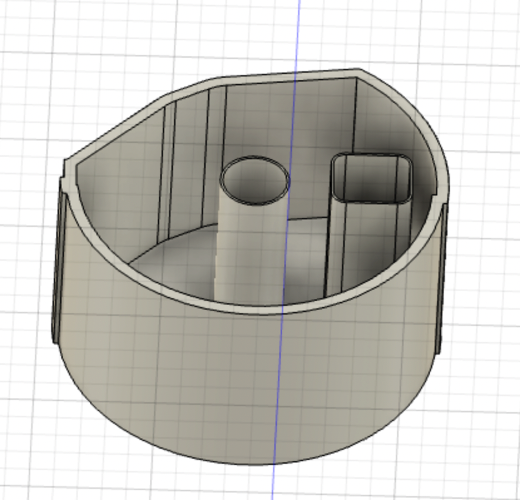
This is the lid design that was used as reference. This is the lead that encases the base of the already existing physical prototype with the rectangular hole because the slot of the microscopic slide to be placed. Modeling done with Fusion 360.

**Figure 10: Old Base Design (Top View)**

****

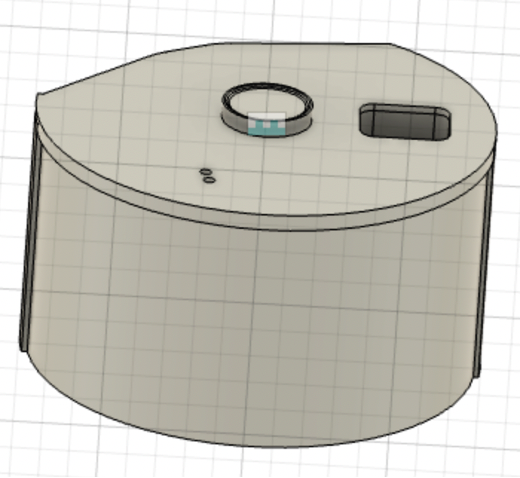
This is the old design of the base (top view). This is the base of the already existing physical on top where the lead of the prototype is placed. This structure helps hold the lid and therefore the microscopic slides in place but has a lot of additional material through the bends which can be removed. Modeling done with Fusion 360.

**Figure 11: New Base Design (Top View)**

****

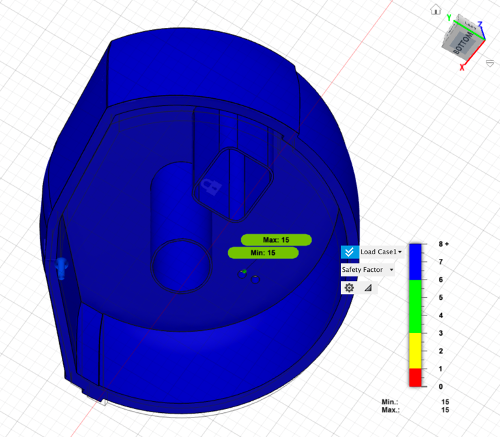
This is the new design of the base (top view). This is an adaptation of the old design of the base and fits in perfectly with the lid (as shown in Figure 12) and uses significantly less material (~40 grams) and is a more compact shape without any bends. Modeling done with Fusion 360.

**Figure 12: Lid Placement Atop New Base Design**

****

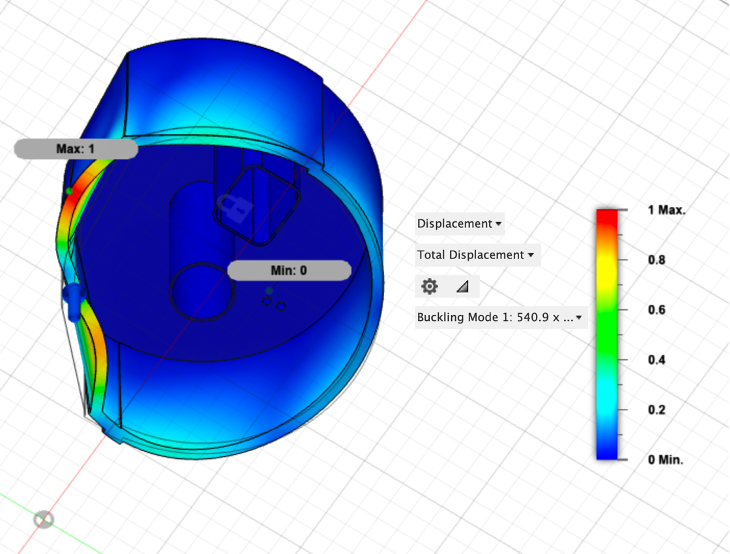
This is the lid placed on the new base design. This figure demonstrates how the lid fits perfectly onto the newly designed base. Modeling done with Fusion 360.

**Figure 13: New Base Design Static Stress Simulation**



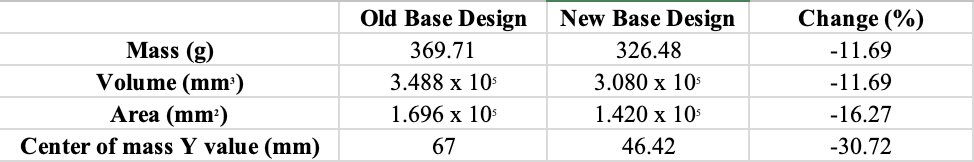
This is a static stress simulation of the new base design. Modeling done with Fusion 360.

**Figure 14: New Design Structural Buckling Simulation**



This is a structural buckling simulation of the new base design. Modeling done with Fusion 360.

**Table 1: Difference in Base Physical Properties**



This table demonstrates the difference in physical properties between the old and new base. The old and new base can be viewed in the Addendum section. The calculations were performed on Fusion 360.. Modeling done with Fusion 360.

**Table 2. Static Stress Simulation (on Fusion 360) - Resulting Safety Factors Based on Material Type**



This table demonstrates the safety factors of various material types in the static stress simulation from Figure 13. Safety factor was calculated as part of the static stress simulation performed on Fusion 360. Modeling done with Fusion 360.

**References**

Access Physiotherapy. (n.d.). *Chapter 14. The Center of Gravity and Stability* . Retrieved from <https://accessphysiotherapy.mhmedical.com/content.aspx?bookid=446&sectionid=41564591#:~:text=Lowering%20the%20center%20of%20gravity%20will%20increase%20the%20stability%20of,of%20support%20(Figure%2014.8)>.

Akintayo, A., Tylka, G.L., Singh, A.K., Ganapathysubramian, B., Singh, A., & Sarkar, S. (2018). A deep learning framework to discern and count microscopic nematode eggs. *Sci Rep* 8, 9145. <https://doi.org/10.1038/s41598-018-27272-w>

AUTODESK Instructables. (2020, March 15). *DC Motor Rotation Control With Start-Stop-Start and a Latch* . Retrieved from Instructables: <https://www.instructables.com/DC-Motor-Rotation-Control-With-Start-Stop-Start-an/>

AUTODESK Instructables. (2020, September 8). *How to control DC Motors Rotation?* Retrieved from Instructables: <https://www.instructables.com/How-to-Control-DC-Motors-Rotation-/>

Butploy, N., Kanarkard, W., & Intapan, P.W. (2021). "Deep Learning Approach for Ascaris lumbricoides Parasite Egg Classification", *Journal of Parasitology Research*, vol. 2021, Article ID 6648038, 8 pages.<https://doi.org/10.1155/2021/6648038>

Children Without Worms (27 May 2021). "Innovative Technology to Address the Elimination of Morbidity of Soil Transmitted Hemelinth."<https://gatech.instructure.com/courses/214292/files/24538503?module_item_id=1970154>

Children Without Worms (2021). "Welcome Georgia Tech BME Students!" <https://gatech.instructure.com/courses/214292/files/24797641?module_item_id=1982086>

Jain, Manav, Keith, Emily, et, al. (28 September 2021) “Design Inputs Report, Paradigm” Georgia Institute of Technology.

Jose-Maria, N., Hart, A., & Tanguturi, S. (2022). GTVIPFall22 STH ML Notebook. (Version 7). Kaggle. <https://www.kaggle.com/code/nathalyjosemaria/gtvipfall22-sth-ml-notebook?scriptVersionId=111027532> Adapted from Ward et. al, 2021b

Mbong Ngwese, M., Prince Manouana, G., Nguema Moure, P. A., Ramharter, M., Esen, M., & Adégnika, A. A. (2020). Diagnostic techniques of soil-transmitted helminths: Impact on control measures. *Tropical Medicine and Infectious Disease*, *5*(2), 93.

Pachyderm. (2022). "What is a Data Pipeline for Machine Learning?" *Pachyderm*. Retrieved November 7th from <https://www.pachyderm.com/blog/what-is-a-data-pipeline-for-machine-learning/>.

Saxena, S. (2018). Precision vs Recall. Medium.<https://medium.com/@shrutisaxena0617/precision-vs-recall-386cf9f89488>

SkyCiv. (2016, April 29). *The Importance of Buckling Analysis*. Retrieved from SkyCiv: <https://skyciv.com/education/the-importance-of-a-buckling-analysis/>

Stuyver LJ, Levecke B (2021) The role of diagnostic technologies to measure progress toward WHO 2030 targets for soil-transmitted helminth control programs. PLOS Neglected Tropical Diseases 15(6): e0009422.<https://doi.org/10.1371/journal.pntd.0009422>

The Engineering ToolBox. (n.d.). *Engineering ToolBox - Resources, Tools and Basic Information for Engineering and Design of Technical Applications!* Retrieved from The Engineering ToolBox: <https://www.engineeringtoolbox.com/factors-safety-fos-d_1624.html>

Ward P, Dahlberg P, Lagatie O, Larsson J, Tynong A, Vlaminck J, 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 Negl Trop Dis 16(6): e0010500.<https://doi.org/10.1371/journal.pntd.0010500>

Ward P, Dahlberg P, Lagatie O, Larsson J, Tynong A, Vlaminck J, et al. (2022). AI4NTD KK2.0 P1.5 STH & SCHm. (Version 6). Kaggle.<https://www.kaggle.com/code/peterkward/ai4ntd-kk2-0-p1-5-sth-schm#Explore-the-data-sets>

​​World Health Organization. (2021). *Diagnostic Target Product Profile*. <https://apps.who.int/iris/rest/bitstreams/1355684/retrieve)>