Nurdle Count – A machine learning approach to nurdle classification and quantification - Year 1 Quarter 4 Report

PI: Seneca Holland

May 1st, 2025

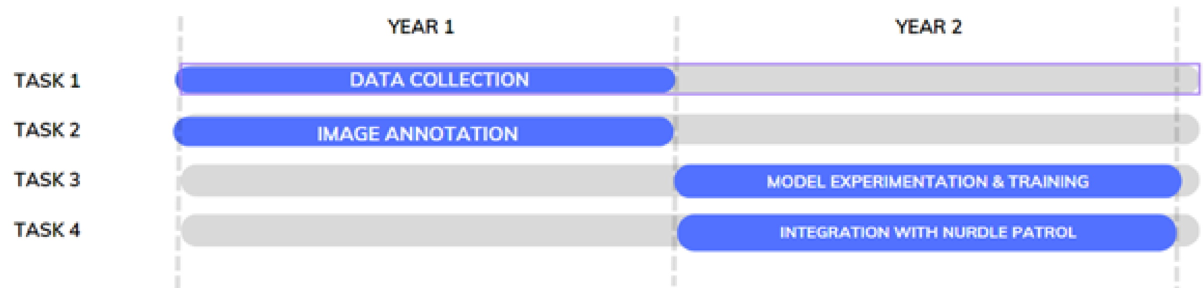
Administration:

The Nurdle Count – A machine learning approach to nurdle classification and quantification was approved for funding on January 8th, 2024, with a requested start date of May 1st, 2024.

Risks and Impacts:

None

Project Tasks:



1) Task 1 - Data collection:

a. **Collect training and test nurdle image data.**
b. **QA/QC collected nurdle image data.**
c. **Research and design AI training methods.**
d. **Develop a standard operating procedure (SOP) for capturing nurdle images.**

Task 1 – Subtasks 1a: Collect training and test nurdle image data

In Year 1, Quarter 1, the research team performed image capturing following the SOP developed for this purpose. Internally, using this SOP, 100 images were captured for Task 2 which is Image Annotation.

In Year 1, Quarter 2, this process was expanded with the help of middle school citizen scientists who are collecting images of nurdles in their classrooms and submitting them via the Nurdle Patrol Website using the QR code below.

In Year 1, Quarter 3, this process was expanded with the help of undergraduate students who collected images of nurdles in class and submitted them via the Nurdle Patrol Website using the QR code.

In Year 1, Quarter 4, this process was expanded with the help of several undergraduate students who added 700 images following strict collection parameters to the Nurdle Patrol Website using the QR codes (Figure 1).

Figure 1: Nurdle Count Image Submission QR Code

This task was completed in Year 1, Quarter 4.

Task 1 – Subtasks 1b: QA/QC collected nurdle image data

In Year 1, Quarter 4, Subtasks 1b (QA/QC of collected images) and 1d (development of the image capture SOP) became closely intertwined, forming an iterative and interdependent workflow. The QA/QC process required a finalized SOP to ensure consistent image quality and metadata, while the SOP's development relied on a fully functional Nurdle Swipe interface to validate and classify images. To support this integration, the Nurdle Swipe tool was upgraded to improve usability and streamline the review process. Notably, text-based buttons such as "swipe right" and "swipe left" were replaced with intuitive visual symbols to reduce user confusion and enhance accessibility (Figure 2).

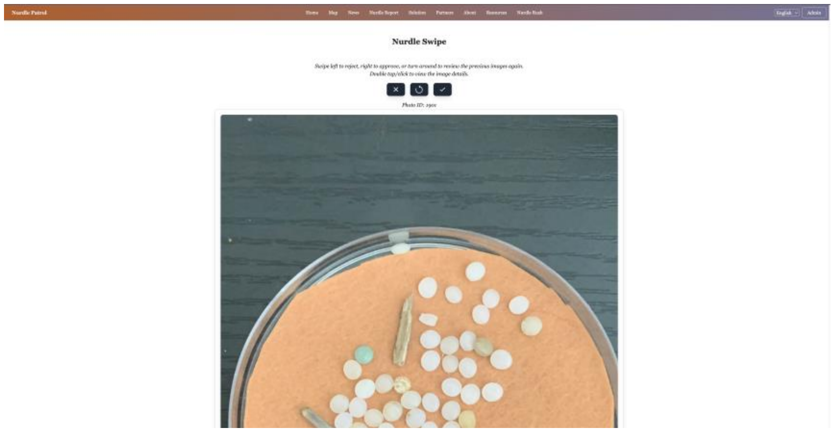


Figure 2: Nurdle Image

Additionally, a standardized list of disqualification reasons was created based on the most frequent issues identified in past image reviews. Validators can now select from this predefined list rather than entering reasons manually, streamlining the QA/QC process and promoting greater consistency (Figure 3).

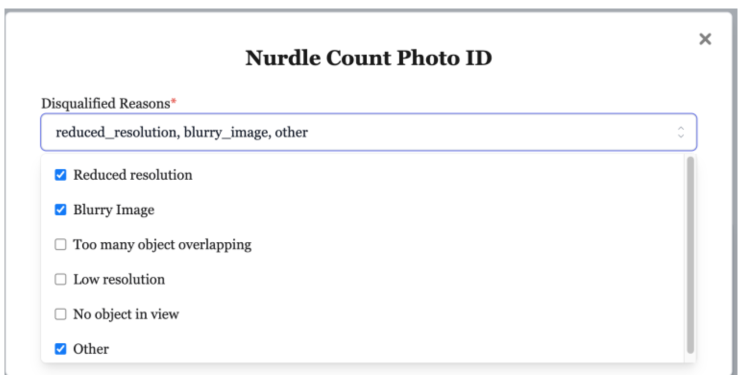


Figure 3: Nurdle Count Photo ID Disqualification Reasons

There is another option that can be used to point to disqualification reasons not yet included in the list, allowing validators to flag new issues. These entries will inform future updates by helping the research team expand and refine the standardized reason list (Figure 4).

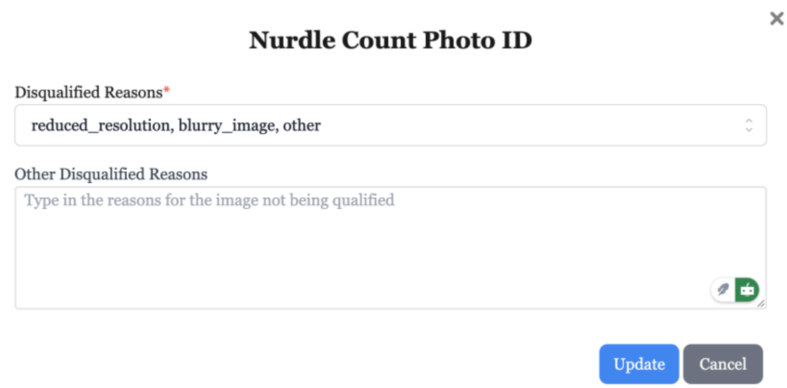


Figure 4: Nurdle Count Photo ID Disqualification Reasons Comment

Researchers assessed each image based on clarity, resolution, object visibility, and the absence of obstructions or excessive overlaps. Specifically, a total of 638 images were reviewed through this effort, resulting in 545 images being marked as qualified and 93 as disqualified. Disqualified images were excluded due to reasons such as "reduced resolution" (60 images), "blurry image" (8 images), "too many objects overlapping" (8 images), and "no visible object in view" (19 images), with some images falling into multiple disqualification categories. The qualified set of images will be used for training AI models for nurdle identification. Figure xxx shows the Nurdle Swipe webpage interface, where two images were classed as qualified and disqualified, respectively.

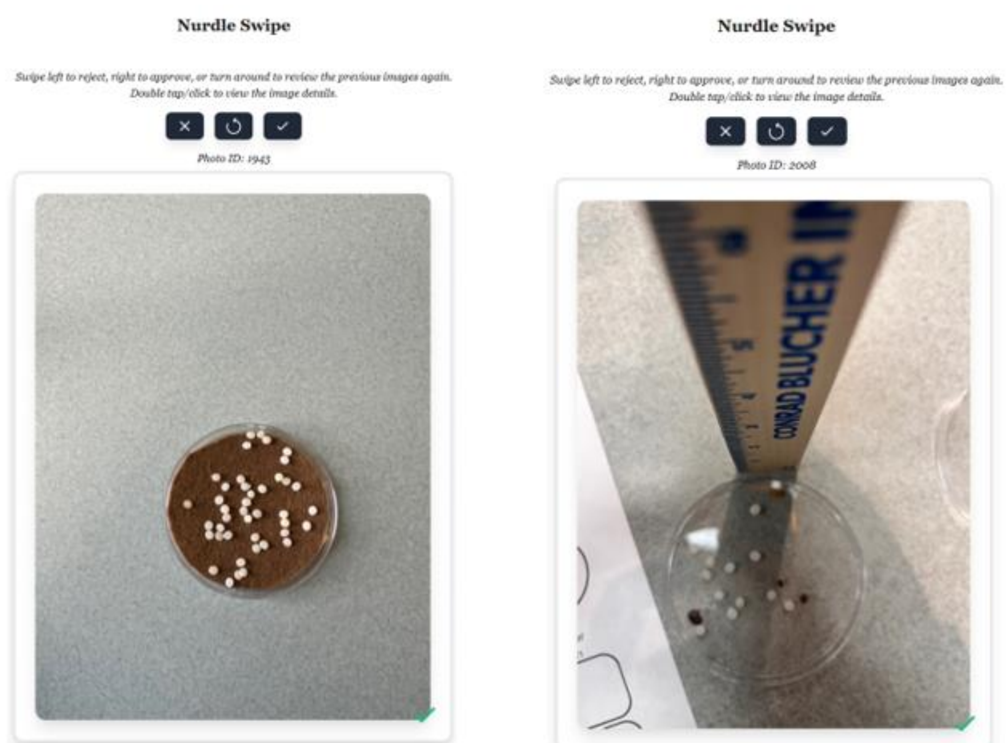


Figure 5: The Nurdle Swipe webpage interface. Left: a qualified nurdle image. Right: a disqualified nurdle image due to blurriness.

Task 1 – Subtasks 1c: Research and design AI training methods

This task was completed in Year 1, Quarter 1.

Task 1 – Subtask 1d: Develop a standard operating procedure (SOP) for capturing nurdle images.

In Year 1, Quarter 1, and in preparation for collecting training and testing the Nurdle image data, the Nurdle Count team first developed two Standard Operating Procedures (SOPs). After an extensive review, two Standard Operating Procedures (SOPs) were created, each tailored to different audiences: internal and external. The internal SOP is designed for use by the research team, while the external SOP is intended for 8th-grade students. Although both SOPs share similar content and workflow, the external SOP is written in language that is accessible and understandable at an 8th-grade reading level.

In Year 1, Quarter 2, project personnel developed a series of three videos detailing the nurdle capture process and made them available via YouTube for a wider audience. To ensure accessibility to a broader audience, YouTube settings enabled these videos to be viewed by kids, and closed captioning was enabled.

These videos are:

Part 1 – Setting up Nurdles in Nurdle Count: https://youtu.be/99pSZEfB37g
Part 2 – Capturing Pictures for Nurdle Count: https://youtu.be/rLRbYLwNVVg
Part 3 - Nurdle Count Image Submission: https://youtu.be/TyTd6OBw9HA

In Year 1, Quarter 3, these videos and materials were leveraged to collect nurdle image data, collect feedback, and improve the Nurdle Count application. This subtask was completed in Year 1, Quarter 3.

For additional information about Year 1, Quarter 4, please see the Subtask 1b section above.

Task 2 – Image Annotation

In Year 1, Quarter 3, to enhance the quality assurance (QA) and quality control (QC) of images collected for training images for Nurdle Count, the Nurdle Swipe feature was developed and successfully integrated into Nurdle Patrol. This process is detailed in Task 1 above.

In Year 1, Quarter 4, we worked to define the preliminary model for detecting support for fast and automatic annotation. Several ML/AI models have been experimented on for nurdle detection. In addition, these model results can be counted as the preliminary results in the early phases and are valued on the way to detect and count nurdles accurately.

Several YOLO-family models, including YOLOv5n, YOLOv8n, and the latest YOLO11n, have been experimented with the current annotated set of images. The models were evaluated based on several metrics, as described in Table 1 below.

Table 1: Model Metrics

Metric	**Description**
Precision	Of all the objects the model says it found, what fraction are real objects? Higher is better.
Recall	Of all the real objects in the image, what fraction did the model actually find? Higher is better.
mAP@0.5	A combined score (mean Average Precision) that rewards finding objects with at least 50% overlap accuracy. Think of it as an overall "accuracy" at a loose overlap threshold. Higher is better.
mAP@0.5-0.95	Similar to mAP@0.5 but averaged over a range of tighter overlap requirements (from 50% up to 95%). This penalizes sloppy bounding boxes more heavily. Higher is better.

Precision, which indicates the proportion of reported detections that are actual nurdles rather than false alarms, was found to be similar across all three models at around 83%, so false alarms are seldom raised. Recall, which measures the proportion of real nurdles in an image that are detected, was highest for YOLO11n at 78%, compared with 60% for YOLOv5n and 69% for YOLOv8n, indicating that substantially fewer pellets were missed by YOLO11n.

Mean Average Precision at a 50% overlap threshold (mAP@0.5), which combines precision and recall into a single accuracy score under a relatively loose matching requirement between predicted and true nurdle locations, was highest for YOLO11n at 0.823—over ten points above YOLOv5n's 0.732. When a tighter matching requirement was imposed (averaging overlap thresholds from 50% to 95%, known as mAP@0.5–0.95), the improvement offered by YOLO11n became even more pronounced: a score of 0.466 was achieved, compared with 0.336 for YOLOv5n and 0.360 for YOLOv8n. These results indicate that not only are more nurdles detected by YOLO11n, but bounding boxes are also drawn around them more precisely.

In practical applications, the use of YOLO11n can result in far fewer pellets are missed. This combination is critical when undetected nurdles can contribute to pollution or signal production defects, and when false alerts can lead to wasted time and resources. Overall, YOLO11n is demonstrated to provide the best balance of thoroughness and reliability for accurate nurdle detection (Table 2).

Table 2: YOLO Results

Model	Precision	Recall	mAP@0.5	mAP@0.5-0.95
YOLOv5n	**0.83**	0.596	0.732	0.336
YOLOv8n	0.815	0.685	0.777	0.36
YOLO11n	0.828	**0.784**	**0.823**	**0.466**

In Task 3 of the project, the research team will continue to work on the automatic annotation workflow, integrate the model for automatic annotation, and experiment with the workflow on the new batch of nurdle images.

Task 3 - Model experimentation and training: ***to be completed in year 2***

a. Train Nurdle Count AI.
b. Collect feedback and improve the AI model.

Task 4 - Integration with Nurdle Patrol: ***to be completed in year 2***

a. Implement the Nurdle Count feature on NurdlePatrol.org.
b. Implement the Nurdle Count feature in the Nurdle Patrol Apple iOS mobile application.
c. Implement the Nurdle Count feature in the Nurdle Patrol Android mobile application.
d. Publish AI model to the public.

Summary:

During Year 1, Quarter 4, the Nurdle Count project made significant strides in completing Tasks 1 and 2. We successfully finalized data collection efforts, adding over 700 rigorously captured nurdle images with the help of undergraduate students, and expanded the nurdle image dataset to strengthen model training. Substantial improvements were also made in image annotation processes, including the deployment of a self-hosted CVAT platform to support structured annotation and upgrades to the Nurdle Swipe tool for efficient quality control and validation. These enhancements streamlined workflows and supported the construction of a robust Nurdle Image Library.

In parallel, preliminary model experimentation began, with comparative testing of YOLO-family models (YOLOv5n, YOLOv8n, and YOLO11n). YOLO11n emerged as the top-performing model, achieving the highest scores across precision, recall, and mean average precision (mAP) metrics, laying a strong foundation for subsequent phases of model training.

Looking ahead to Year 2, project efforts will focus on advancing model experimentation and training. The research team will train the Nurdle Count AI using the expanded, high-quality image dataset and systematically collect feedback to refine and improve model performance. In

parallel, integration efforts will begin to deploy the Nurdle Count feature within the NurdlePatrol.org website and the Nurdle Patrol mobile applications for iOS and Android. Upon successful implementation, the trained AI model will be published and made available to the public, broadening the impact and accessibility of automated nurdle detection.

Synergistic Activities:

On April 3, Son Nguyen (Co-PI) and Khoi Nguyen presented in the morning session at the 7th Annual Texas Plastic Pollution Symposium at the Houston Zoo. The presentation covered how the project started and the project's progress until the end of March 2025. The intention of the presentation is to voice the idea of using artificial intelligence as an assisting tool to the citizen scientist community and to potentially spark more collaboration about the potential applications as well as future works of the Nurdle Count model among the researchers in the community. We also highlighted the educational outreach component in the project, in which Seneca and Jace helped the middle schoolers to learn more about nurdle pollution and artificial intelligence while giving the students an opportunity to contribute data to train the machine learning model.

Figure 6: Co-PIs Presenting at 7th Annual Texas Plastic Pollution Symposium

Obstacles: None