Nurdle Count – A machine learning approach to nurdle classification and quantification - Year 2 Quarter 2 Report

PI: Seneca Holland

November 21st, 2025

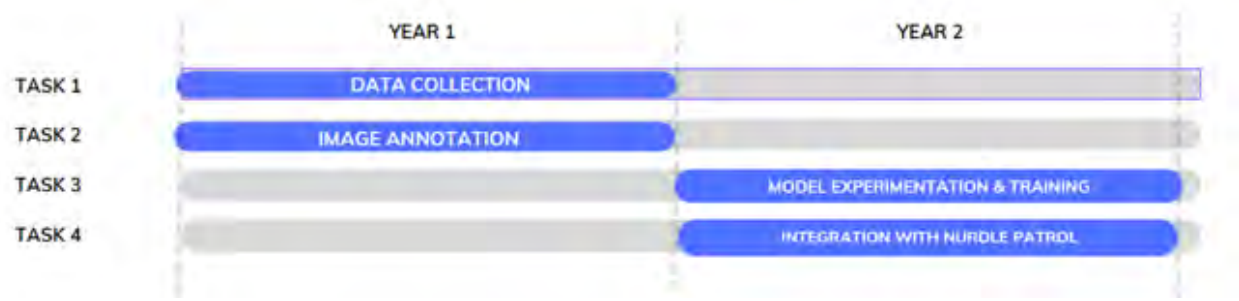
Administration:

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

Task 1 – Subtask 1a 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).

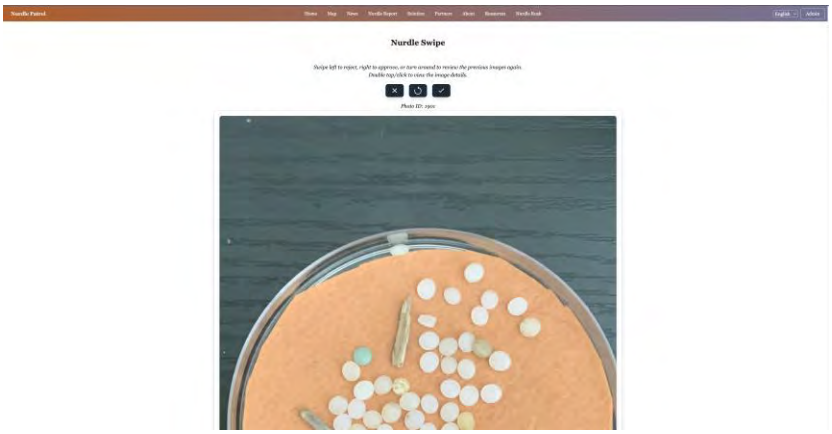


Figure 2: Nurdle Image

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).

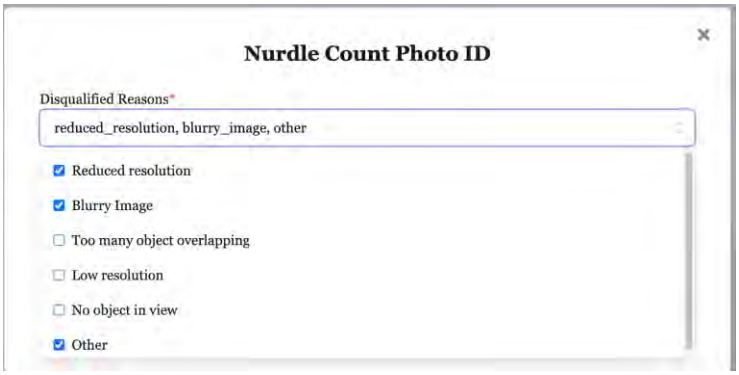


Figure 3: Nurdle Count Photo ID Disqualification Reasons

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. Task 1 – Subtask 1b was completed in Year 1.

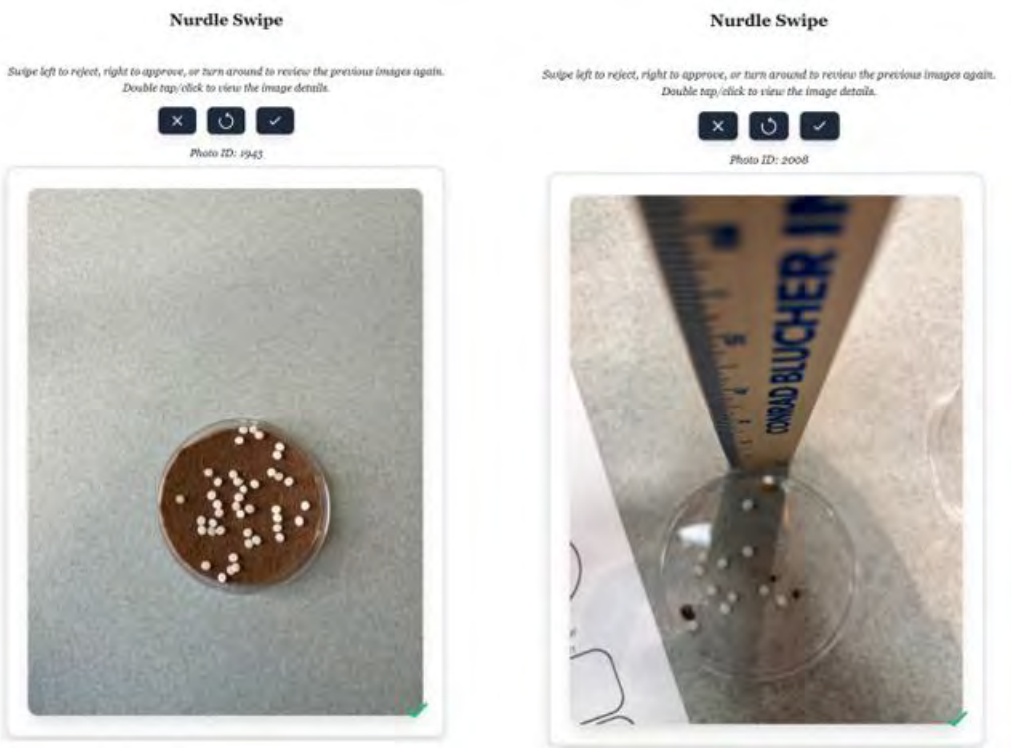


Figure 5: The Nurdle Swipe webpage interface. Left: a qualified nurdle image. Right: a disqualified nurdle image due to blurriness.

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***

In preparation for Task 3, the research team advanced efforts to curate a high-quality image dataset through the Nurdle Swipe tool, a web-based platform hosted on the Nurdle Patrol website (https://nurdlepatrol.org/app/nurdle-swipe). The tool was developed to support AI model training by systematically reviewing nurdle images collected in accordance with the established image collection SOP. Using a swipe interface, researchers approved images that met quality standards (swipe right) or disqualified those that did not (swipe left).

Each image was evaluated for clarity, resolution, object visibility, and freedom from obstructions or excessive overlap. Through this process, 638 images were reviewed, of which 545 were classified as qualified and 93 were disqualified. Disqualifications were attributed to reduced resolution (60 images), blurry capture (8 images), excessive object overlap (8 images), or absence of visible objects (19 images), with some images falling into multiple categories. The resulting set of 545 qualified images will serve as a training dataset for AI-based nurdle identification models, ensuring that only rigorously vetted imagery is used to improve detection accuracy. Figure 5 illustrates the Nurdle Swipe interface, displaying examples of qualified and disqualified images.

Looking ahead, Task 3 will focus on model experimentation and training using this expanded, high-quality dataset. The research team will conduct comparative testing across multiple computer vision architectures to evaluate precision, recall, and mean average precision (mAP) metrics. The top-performing model will then be selected for iterative training and refinement. Feedback loops from annotation and quality control processes will be integrated to further improve performance. By the conclusion of Year 2, the trained Nurdle Count AI will be ready for deployment into the Nurdle Patrol website and mobile applications, providing a scalable solution for automated nurdle detection.

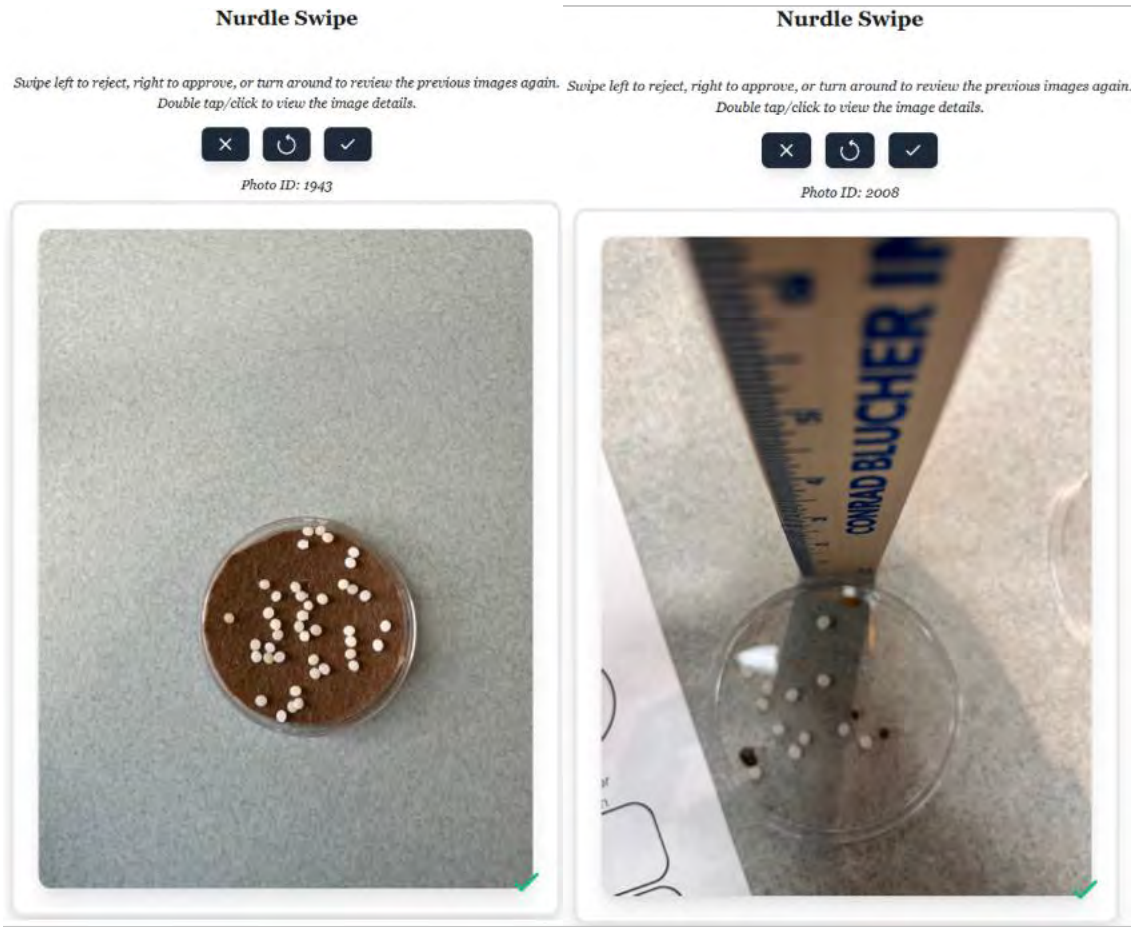


Figure 6: Nurdle Swipe Images

In Year 2, Quarter 2, work continued on the YOLO segmentation model experiment.

In the last quarter, the research team experimented with YOLO segmentation model. This approach required a different way of annotating in which the boxes fit more to the shape of the nurdles. A deployment of Facebook Segmentation Anything Model has been done to support the annotation process.

Figure 7: Example of an annotated image with SAM for YOLO segmentation model

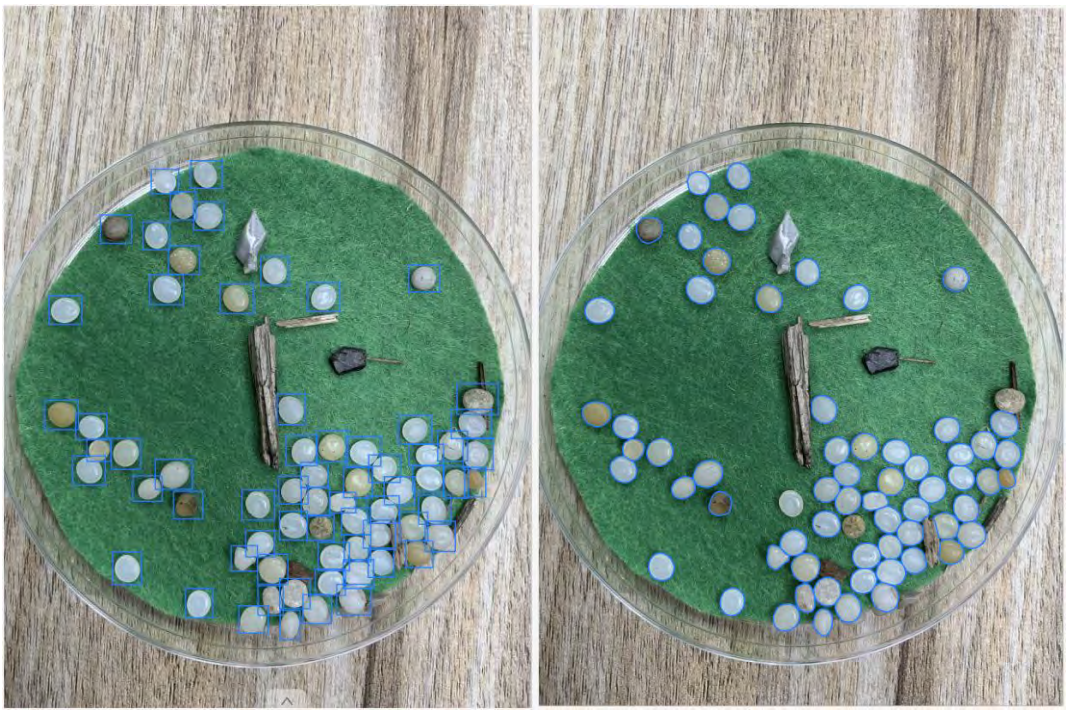


Figure 8: Original annotation for YOLO-based model (left) and annotation for YOLO segmentation model (right)

To assess the performance of the new YOLOv8n segmentation, it will be compared to the YOLO11n, the best model in the previous reports.

Table 3: Performance Metrics Comparison

Metric	YOLOv8n-seg	YOLOv11n	Difference
Overall Accuracy (mAP@0.5)	99.3%	82.3%	-17.0%
Strict Accuracy (mAP@0.5-0.95)	~90%	~40%	-50%
Correct Detection Rate (Precision)	~100%	~85-95%	-5 to -15%
Finding All Nurdles (Recall)	~100%	~87%	-13%
Balanced Score (F1)	0.98	0.81	-0.17
Confidence Needed for Reliability	57.3%	95.5%	+38.2%

The performance comparison reveals substantial differences between the two models. YOLOv8n with segmentation achieved an exceptional accuracy of 99.3%, meaning it correctly identifies nurdles 99 times out of 100. YOLOv11n reached only 82.3%, representing a significant 17 percentage point gap. This difference becomes even more pronounced when examining strict accuracy measures that require very precise box placement, where YOLOv8n achieves about 90% compared to YOLOv11n's 40%. When examining how correctly the models identify nurdles versus falsely detecting non-nurdles, YOLOv8n maintains nearly perfect accuracy at approximately 100%, while YOLOv11n ranges between 85-95%. The ability to find all nurdles similarly favors YOLOv8n at approximately 100% versus YOLOv11n's 87%.

Perhaps most critically for practical deployment, YOLOv8n achieves 100% accurate identifications when it's at least 57% confident, while YOLOv11n requires an extremely strict 95.5% confidence threshold. This means in real-world use, YOLOv8n can reliably detect nurdles with moderate confidence, while YOLOv11n must be almost completely certain before its detections can be trusted. This makes YOLOv11n far less practical, as most legitimate nurdle detections fall below this very high threshold, causing the model to miss many real nurdles just to avoid making errors.

Table 3: YOLOv8n Segmentation Results:

What's Actually There	Model Says "Nurdle"	Model Says "Background"
Nurdle	211 (96.8%)	7 (3.2%)
Background	2 (0.9%)	Perfect

Table 4: YOLOv11n Detection Results:

What's Actually There	Model Says "Nurdle"	Model Says "Background"
Nurdle	816 (80.9%)	193 (19.1%)
Background	0 (0%)	Perfect

The confusion matrix shows exactly how each model performs in different situations. For YOLOv8n segmentation, the model correctly identified 211 nurdles out of 218 presents, missing only 7 nurdles. This means it has a 96.8% success rate at finding nurdles that are there. The model also produced only 2 false alarms where it thought it saw a nurdle when there wasn't one, demonstrating excellent ability to distinguish between nurdles and background objects.

YOLOv11n detection tells a different story. While it correctly identified 816 nurdles, it missed 193 nurdles that were present. This means it has a 19.1% failure rate - nearly one in five nurdles goes undetected. Interestingly, YOLOv11n never produces false alarms, achieving a perfect 0% false positive rate. However, this "perfection" comes at a severe cost: the model is so conservative that it refuses to make a detection unless it's extremely certain, which causes it to miss many real nurdles. Missing one in five nurdles is 27 times worse than YOLOv8n's miss rate.

Detection Quality Analysis

When looking at the actual images with detection boxes drawn on them, the differences become visually apparent. YOLOv8n segmentation produces tight, well-fitted boxes that precisely outline individual nurdles. Even when multiple nurdles are clustered close together, the model successfully identifies each individual pellet separately. The detections remain accurate across various challenging conditions: different lighting (bright, shadowy, or dim), different angles and rotations of the objects, and different background materials (green surfaces, brown media, white plates, or wooden surfaces). The confidence scores the model assigns are generally high for correct detections, indicating it "knows" when it has found a real nurdle.

YOLOv11n detection exhibits several quality problems in its predictions. In many images, multiple overlapping boxes are detected on the same nurdle, showing that the model is detecting the same object several times instead of recognizing it as a single item. The boxes are generally less precisely fitted around the actual nurdle boundaries, often being too large or poorly positioned. The model particularly struggles when nurdles are packed closely together - it either groups multiple nurdles into one detection or misses individual nurdles within crowded scenes. The confidence scores vary widely from 30% to 90%, and because the model needs 95.5% confidence to be truly reliable, it ends up missing many valid nurdles that fall below this very strict threshold.

Precision-Recall Characteristics

The precision-recall relationship tells us how well a model can balance between being accurate and being thorough. YOLOv8n segmentation maintains over 95% accuracy in its identifications even while finding 95% of all nurdles present. Only when trying to find nearly every single nurdle (above 95% recall) does its accuracy begin to drop. This creates a nearly ideal performance curve where the model can be both accurate and thorough simultaneously.

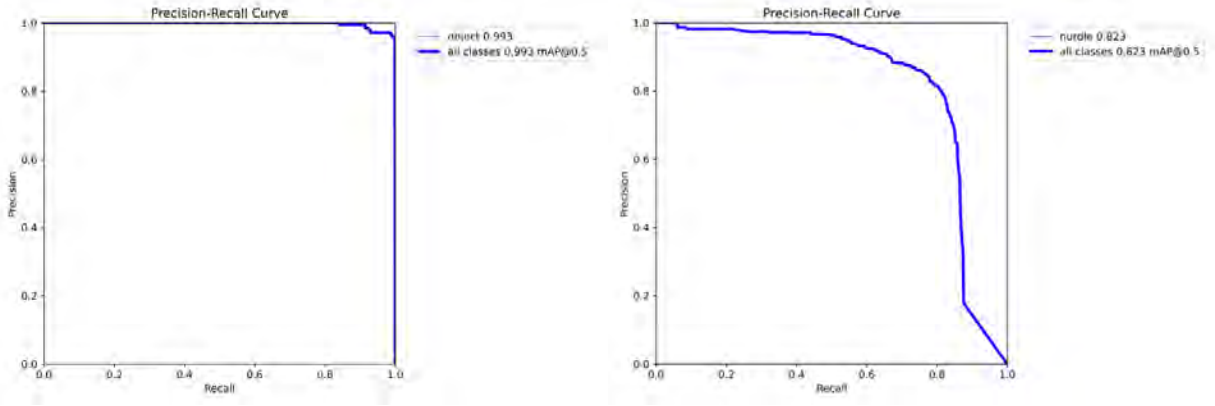


Figure 9*: YOLOv8n segmentation (left) achieves 0.993 mAP@0.5 with a nearly rectangular curve, maintaining over 95% precision while finding 95% of nurdles. YOLOv11n detection (right) achieves 0.823 mAP@0.5 with earlier precision degradation, dropping to 85% precision at 80% recall. The area under the curve difference (0.993 vs 0.823) represents a 17-point performance gap.*

YOLOv11n detection shows a much less favorable balance. It can maintain 98-100% identification accuracy only when it's being very selective and finding less than 40% of the nurdles present. As the research team lower the selectivity to find more nurdles, the accuracy degrades much earlier and more severely than YOLOv8n. By the time it finds 80% of nurdles, its identification accuracy has dropped to about 85%. This curved relationship indicates a harsh tradeoff where attempting to detect more nurdles rapidly compromises how accurately it identifies them. This makes it very difficult to find a setting that gives both reasonable coverage and reliable identifications.

Confidence Calibration Analysis

Confidence calibration determines how much we can trust a model's stated confidence in its predictions. Think of it like a weather forecast - if the forecast says 70% chance of rain, it should rain about 70% of the time for the forecast to be well-calibrated. YOLOv8n segmentation demonstrates excellent calibration. When the research team use a confidence threshold of 57.3% (meaning we only trust detections where the model is at least 57.3% certain), we get 100% accurate identifications. The model works well even at very low confidence thresholds, and its confidence scores align well with actual accuracy, making it straightforward to choose appropriate settings for different needs.

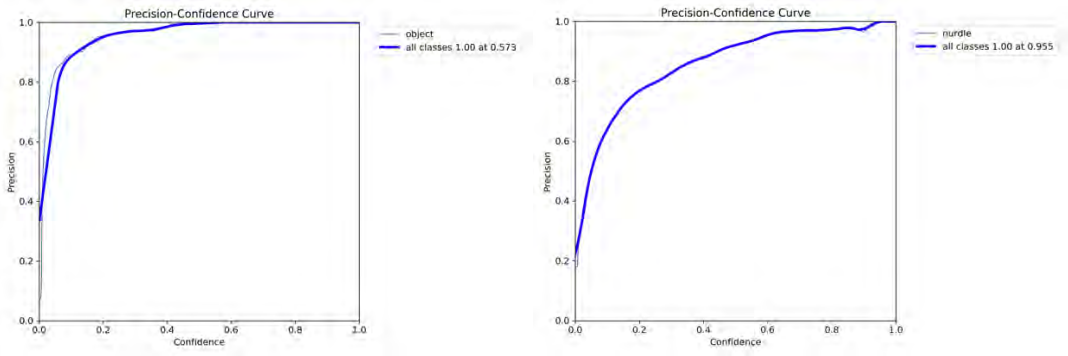


Figure 10: YOLOv8n segmentation (left) achieves 100% precision (perfect accuracy) at 57.3% confidence, with smooth calibration showing reliability at moderate thresholds. YOLOv11n detection (right) requires 95.5% confidence for 100% precision, making it impractical as most detections fall below this strict threshold. The 38-point difference in required confidence (95.5% vs 57.3%) represents a critical disadvantage for real-world deployment.

YOLOv11n detection has severe calibration problems. To get 100% accurate identifications, we must use an extremely strict confidence threshold of 95.5%, meaning we can only trust detections where the model is 95.5% certain. This is impractical because most real nurdle detections fall below this threshold, so we'd miss most nurdles just to avoid errors. Even when accepting all predictions regardless of confidence, the model still misses 13% of nurdles, showing that even its most confident predictions aren't finding everything. This poor calibration means users face a difficult choice: either use a very high confidence threshold and miss most nurdles or use a lower threshold and accept many unreliable detections.

Architecture Comparison

The architectural differences between these models are fundamental to understanding their performance. YOLOv8n with instance segmentation doesn't just draw boxes around objects - it also traces the exact outline of each nurdle at the pixel level, like carefully coloring within the lines. This segmentation capability provides several important advantages. It forces the model to understand object boundaries very precisely, not just approximately. The additional task of drawing outlines helps the model learn more detailed features during training. For small objects like nurdles, this detailed attention is especially valuable. The outlines also naturally help

separate overlapping objects since the model must trace each one individually. The main downsides are that it runs about 10-20% slower and uses slightly more computer memory.

YOLOv11n represents newer technology designed to be faster and more efficient. In theory, it should perform better through various technical improvements and optimizations. The simpler task of just drawing boxes (without outlines) should be easier to learn. However, in practice, these expected advantages haven't materialized for nurdle detection. The model performs significantly worse across all measures.

This suggests the improvements in YOLOv11 were focused on different types of detection tasks, possibly sacrificing the ability to detect small objects for gains in speed or efficiency. The segmentation component in YOLOv8n forces the model to learn precise shape and texture features at the pixel level - you can observe this in the detection images where bounding boxes fit tightly and strictly around each nurdle's actual boundaries. This detailed shape learning helps the model distinguish individual nurdles even in dense clusters and understand the subtle texture differences between nurdles and background. Without the segmentation head, YOLOv11n misses this extra learning signal and detailed feature understanding. The model learns only approximate locations rather than exact object shapes and textures, which proves particularly problematic for small, round objects like nurdles where shape precision is critical. The architectural changes that make YOLOv11 work well for some tasks apparently make it worse for detecting small, clustered objects like nurdles.

Potential Overfitting Concerns with YOLOv8n Segmentation

While YOLOv8n shows impressive 99.3% accuracy, this near-perfect performance raises an important concern: overfitting. Think of overfitting like a student who memorizes specific test questions instead of truly understanding the subject. The student might score perfectly on practice tests but fail when questions are worded differently. Similarly, YOLOv8n might have "memorized" the 1,300 training images rather than learning what nurdles generally look like, which could cause problems when encountering new situations.

Several warning signs suggest overfitting might be occurring. The small training set of only 1,300 images is concerning. The validation images likely came from the same collection effort using similar cameras, lighting, and locations as the training images, so they may not represent truly different real-world conditions. The segmentation approach, while accurate, requires learning very precise, detailed patterns that might make the model too sensitive to any changes in image quality or nurdle appearance.

There are practical situations where YOLOv8n's overfitting could make it perform worse than YOLOv11n. If deployed with a different camera or smartphone than used in training, YOLOv8n might struggle with different color profiles or image quality. When lighting conditions change, the model might fail to recognize nurdles it hasn't seen under those specific conditions. Different types of nurdles (different colors, sizes, or materials), dirty or damaged nurdles, wet versus dry surfaces, unusual camera angles, or new background types could all cause unexpected failures. YOLOv11n's more cautious, conservative approach might handle these surprises better because it hasn't learned overly specific patterns.

An overfitted model tends to be overconfident, making predictions even in uncertain situations. YOLOv11n's extreme caution, while causing it to miss nurdles, might be more appropriate when facing completely new scenarios. In truly different environments, YOLOv8n might produce many more false alarms than the 2 seen in testing.

To verify whether overfitting is a real problem, the model needs testing on completely independent images from different locations, times, cameras, and conditions that it has never encountered before. Currently, there's no evidence such comprehensive testing has occurred. Until then, the 99.3% accuracy should be viewed cautiously - it might represent genuine capability, or it might collapse when facing real-world diversity. The model should be expanded to at least 5,000-10,000 diverse training images and tested across multiple equipment types and environmental conditions before full confidence in deployment.

YOLOv8n appears significantly better than YOLOv11n based on current testing, but there's a risk this advantage might not hold up in all practical situations. The model might work perfectly in scenarios like training conditions but struggle when things look different. YOLOv11n's poorer performance but more conservative approach might be more reliable when encountering unexpected situations.

Next Steps

In Year 2 Quarter 3,the research team will re-annotate the existing image dataset with segmentation masks and retrain YOLOv8n with enhanced data augmentation (Mosaic, MixUp, color jittering, geometric transformations) and systematic exploration of training parameters including learning rates, dropout rates, weight decay, and batch sizes. Cross-validation across multiple data splits and temporal test set separation will rigorously assess whether the model learns generalizable nurdle characteristics or simply memorizes training patterns. This expanded approach using the same underlying dataset as YOLOv11n will determine whether segmentation's advantages persist with equal data volumes and proper validation protocols, or whether the current 99.3% accuracy reflects overfitting to a fortunate train/validation split.

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

In Year 2 Quarter 1, initial design work began on the integration of Nurdle Count AI into the Nurdle Patrol Website (Design).

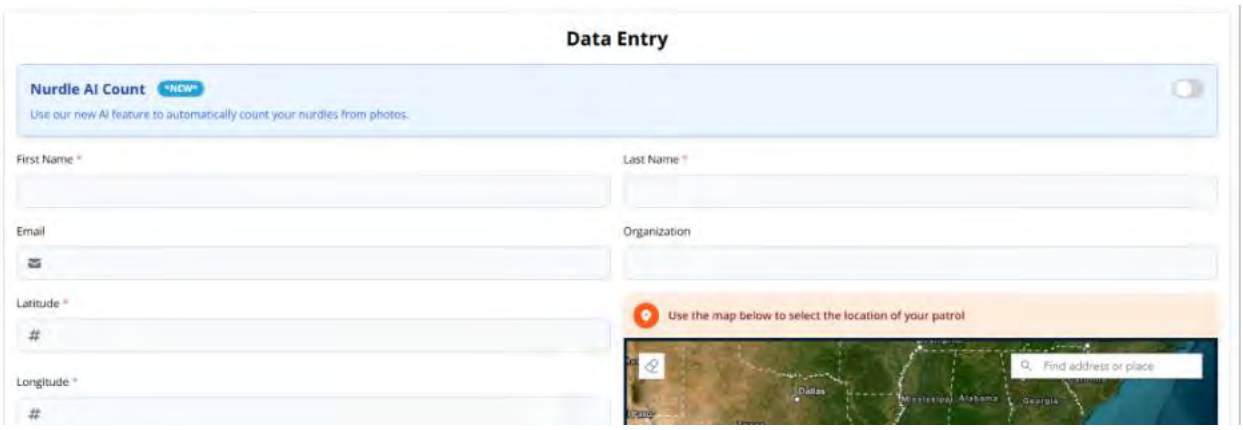


Figure 11: UI Data Entry

The Data Entry form on the Nurdle Patrol Website will include a toggle to enable the Nurdle Count AI feature. If it is enabled, the actions below will follow:

Figure 12: AI Detection & Confirmation

First, a pop-up will appear that uses Nurdle Count AI to automatically detect and estimate the total number of nurdles in the submitted image. The user will then be prompted to either confirm or reject the AI's detected count.

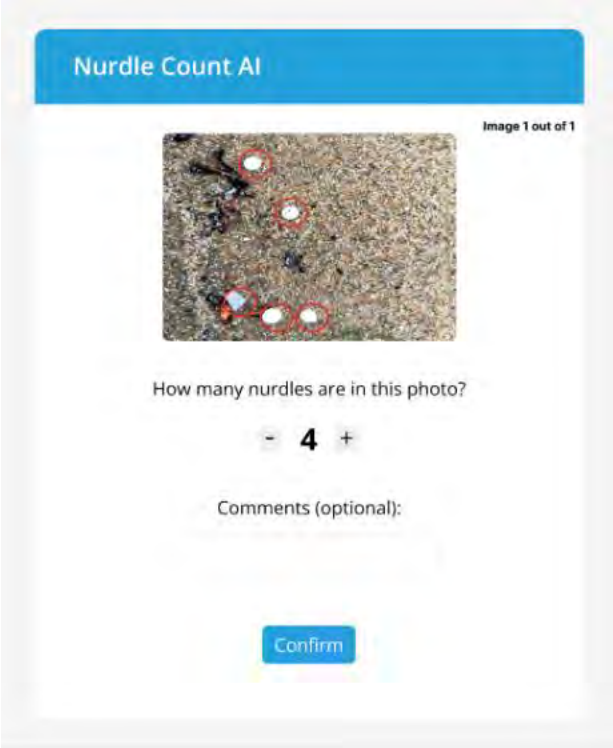


Figure 13: User Rejection & Manual Input

If the user rejects the AI's estimate, they will be asked to manually input the correct number of nurdles. An optional comment field will also be provided for additional notes or clarifications.

Figure 4: Form Auto-Population

Finally, once the user confirms, the form will be auto-populated.

In Year 2 Quarter 2, work continued to further the integration of Nurdle Count AI into the Nurdle Patrol website by developing workflows that support up to 2 photos (Figure x) (Figure x). The workflow allows users to submit their photo(s) and includes options to confirm, reject, or adjust the count as needed. Based on whether the user confirms or rejects the count, a different set of actions are presented to accurately account for all nurdles. Users can also leave comments within

the interface, and finally, the count is auto-populated into the form, making the experience useful and easy.

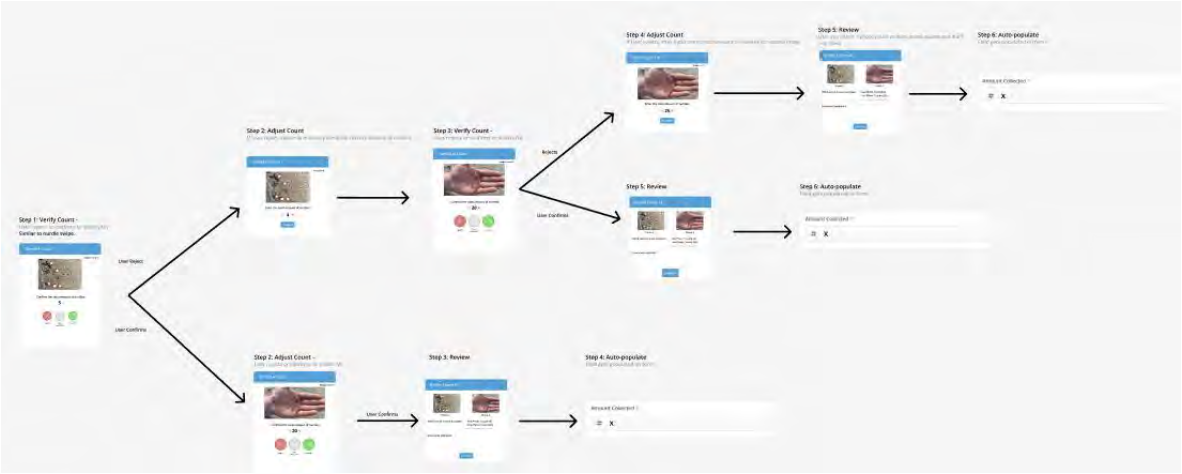


Figure 14: Workflow for 2 photos

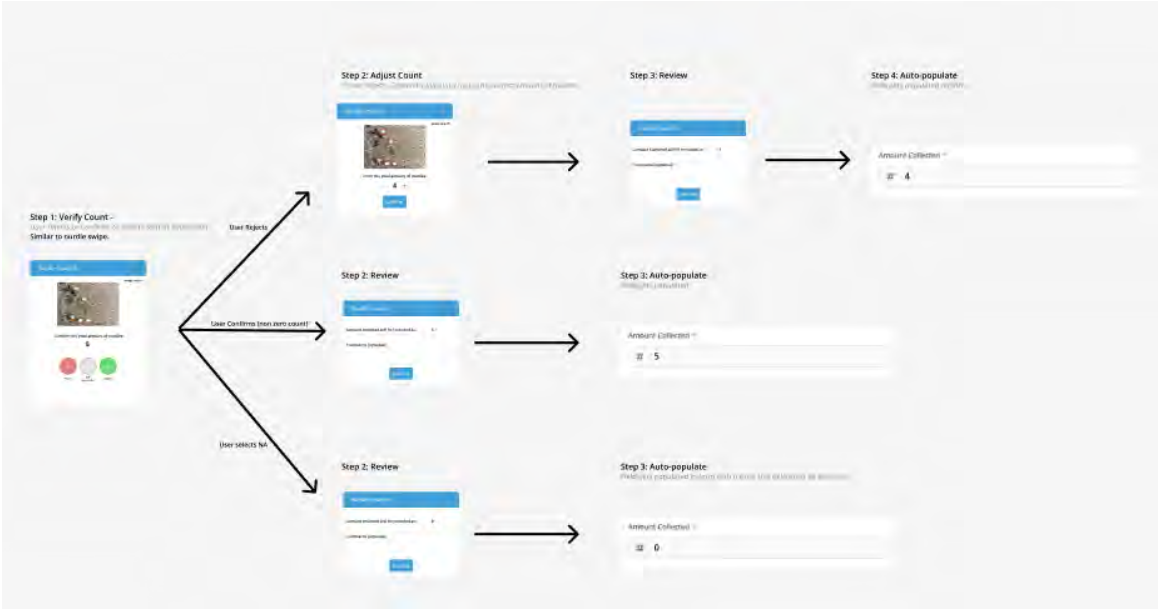


Figure 15: Workflow for 1 photo

Several popups were designed for ease of use and to guide users through every step of the process while using Nurdle Count AI to count their nurdles.

Integration Popup: Allows users to add counts for up to 2 photos. Users also have the option to select one count or the other if preferred. This is useful when users submit 2 photos of 2 different sets of nurdles that pertain to one location (Figure X).

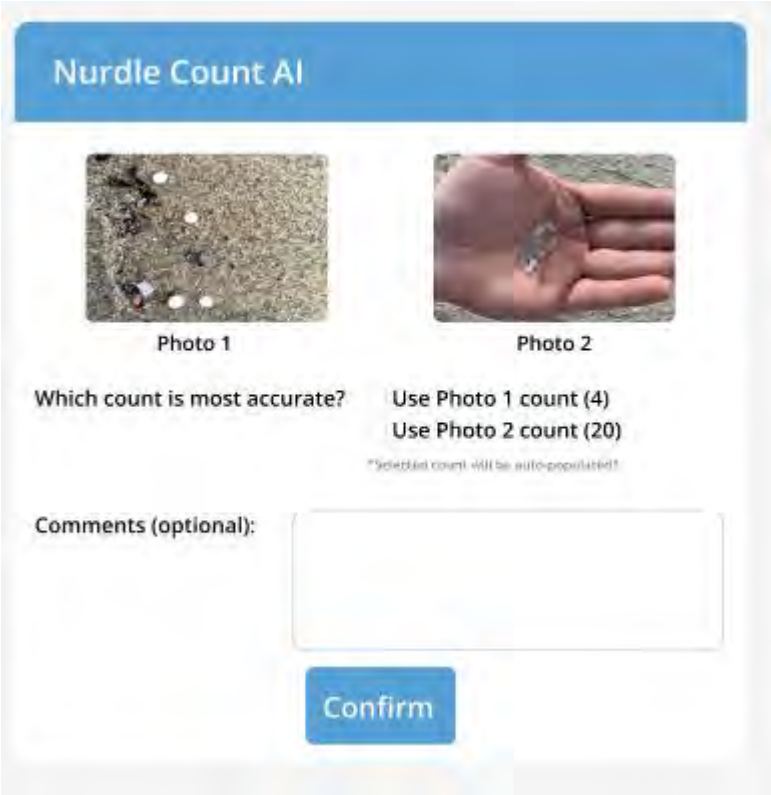


Figure 16: Workflow for 2 photos

Loading Popup: Displays while the Nurdle Count AI is detecting nurdles (Figure X).

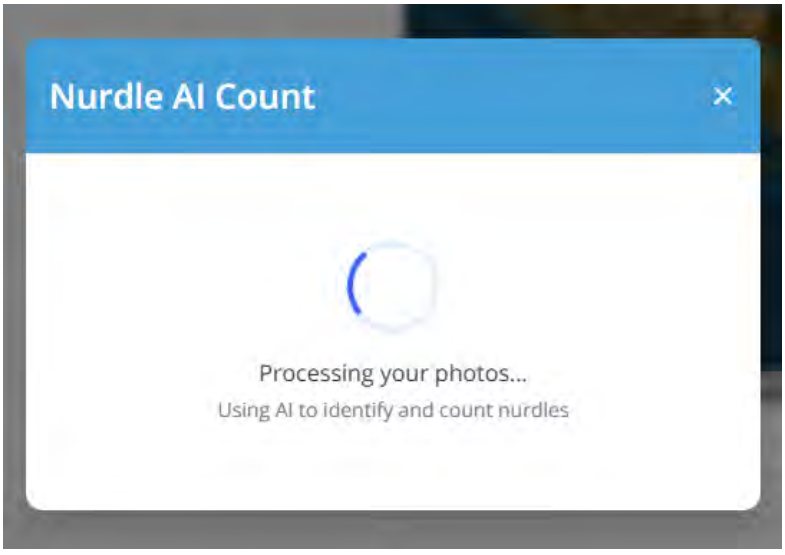


Figure 17: Pop-ups

Confirmation Popup: Displays once the user has finished confirming/adjusting the count (Figure X).

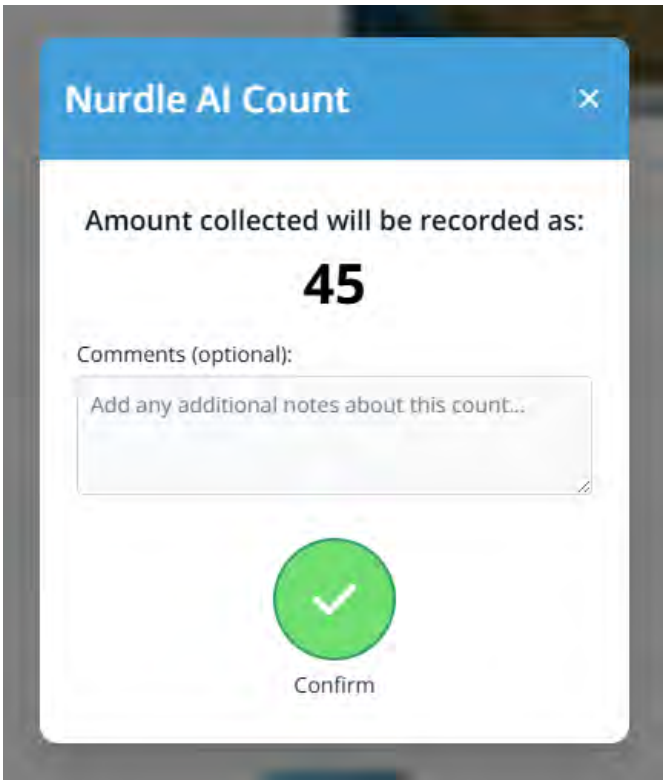


Figure 18: Confirmation Pop-up

Future work:

The next steps are to finalize all logic and implement the design changes in the code. We will be integrating the modal, testing Nurdle Count AI for ease of use, and ensuring the interface is intuitive for users.

Summary:

During Year 2 Quarter 2, Tasks 3 and 4 advanced significantly, building on earlier development to refine model experimentation and strengthen integration workflows for the Nurdle Count AI.

For Task 3**,** work centered on continued experimentation with the new YOLO segmentation approach. Building on the previously curated dataset of 545 qualified images, the research team evaluated the YOLOv8n segmentation model and compared its performance directly with the YOLO11n detection model. This quarter's analyses demonstrated substantial gains in segmentation accuracy, with YOLOv8n achieving approximately 99.3 percent mAP at a 0.5 threshold and near perfect precision and recall. Confusion matrix results further highlighted

YOLOv8n's ability to detect nearly all nurdles with minimal false positives. In contrast, YOLO11n continued to exhibit missed detections, weaker bounding box performance, and poor confidence calibration requiring unusually high confidence thresholds. This quarter also documented potential overfitting risks associated with YOLOv8n's near perfect results, noting the need for expanded datasets and more rigorous cross validation in upcoming quarters. Collectively, the work in Y2Q2 provides compelling evidence in favor of the segmentation based modeling approach and lays the groundwork for the re annotation and retraining efforts planned for Year 2 Quarter 3.

For Task 4**,** the team advanced the integration of the Nurdle Count AI into the Nurdle Patrol website by developing and refining workflows that support submissions of one or two photos. Multiple user interface popups were designed and tested to guide users through uploading images, confirming or adjusting AI generated counts, selecting between two images when needed, and automatically populating the data entry form with the verified count. The new logic accommodates flexible user paths, including confirming the AI estimate, rejecting it and entering a manual count, or toggling between counts from two different photos within a single submission. These refinements move the project closer to a fully functional front end integration that is intuitive for citizen scientists and consistent with the existing Nurdle Patrol user experience.

Also during Year 2 Quarter 2, Nurdle Patrol outreach and education efforts continued at a strong pace, led by Jace Tunnell across Texas and through virtual platforms. These activities focused on promoting environmental literacy, engaging citizen scientists, and increasing public awareness of plastic pellet (nurdle) pollution.

Across the reporting period, 16 events were delivered between August 1 and November 30, 2025, reaching more than 1,260 participants. Audiences included K–12 students, teachers, community organizations, university groups, coastal stakeholders, and international partners. Events were conducted in both in-person and virtual formats, broadening geographic and demographic reach.

Programming emphasized hands-on learning in beachcombing, nurdle identification, microplastics, marine debris, oysters and water quality, and the role of citizen science in coastal stewardship. The outreach extended well beyond the Texas Coastal Bend, including participation at Dallas College and a virtual international session with the Universidad Autónoma de Baja California in La Paz, Mexico**.**

Several high impact engagements took place during this quarter. These included a large STEM night event with 300 students at Baker Middle School, a Friends of Padre community booth with 200 participants**,** and outreach to multiple middle and high schools across the region. Additional visibility came from three podcast interviews with Texas-based hosts, expanding the program's reach to broader public audiences.

Collectively, these activities demonstrate the strong demand for Nurdle Patrol's educational programming and the continued importance of citizen science based environmental education. This quarter's outreach strengthened community partnerships, supported stewardship across

coastal environments, and expanded the program's footprint at local, regional, national, and international levels.

Synergistic Activities:

In the Fall season of 2025, Jace Tunnell conducted a robust series of Nurdle Patrol outreach and education events across Texas and beyond, reaching a wide range of audiences from high school students to international organizations. These activities emphasized hands-on environmental education, citizen science engagement, and science communication.

Highlights include:

- 16 events delivered between August 1st and November 30, 2025, with a mix of in-person and virtual formats.
- Total reach of more than 1,260 participants**,** spanning K–12 students, teachers, community groups, scientists, and the general public.
- Local, regional, and national impact, with events hosted in the Texas Coastal Bend, at national and international venues such as Dallas College in person in Dallas, Texas, and the Universidad Autónoma de Baja California in La Paz, Mexicothrough virtual sessions that connected with broader audiences.
- Educational themes focused on beachcombing, plastic pellet (nurdle) pollution, science communication, oysters and water quality, and the role of citizen science in coastal stewardship.

Notable engagements:

- A large-scale event with 300 students at Baker Middle School to STEM night.
- An outdoor booth event with 200 community members at the Briscoe Pavillion on North Padre Island and hosted by the Friends of Padre.
- Three podcast interviews with Texas based groups with large audience reach.
- A community partnership with the Neighbor League of Corpus Christi (40 attendees).
- Virtual session with the Universidad Autónoma de Baja California in La Paz, Mexico (50 participants) extending the program's reach internationally.
- An appearance at the Dallas College in Dallas, Texas (50 attendees)**,** introducing Nurdle Patrol to a broader geographic audience.

These events illustrate the strong demand for Nurdle Patrol's educational programming and the value of integrating science communication with citizen science opportunities. Collectively, this summer's outreach fostered environmental awareness, built community connections, and encouraged active participation in monitoring plastic pollution across coastal environments.

Here is a full list of events conducted during this reporting period:

Date	Organization	Type	Subject	Title	Location	Number of Attendees
8/1/2025	Texas Surf Camp	In-person	Beachcombing/Nurdle Patrol	Beachcombing and Nurdle Patrol	Bob Hall Pier	25
8/13/2025	Logan OnAir Podcast	In-person	Beachcombing and Nurdle Patrol	Podcast	Hometown Seafood on North Padre Island	
8/18/2025	Shellphone Podcast	In-person	Nurdle Patrol	Podcast	Zoom	
8/25/2025	Surfrider Foundation	In-person	Beachcombing/Nurdle Patrol	Beachcombing and Nurdle Patrol	TAMUCC NRC building	40
8/26/2025	Breakaway Tackle	In-person	Beachcombing/Nurdle Patrol	Podcast	Nick Meyer Studio on North Padre	
9/11/2025	Neighbor League of Corpus Christi	In-person	Beachcombing/Nurdle Patrol	Beachcombing and Nurdle Patrol	Omni Hotel	40
9/20/2025	Surfrider Foundation	In-person	Marine Debris	Marine Debris and Adopt a Beach Cleanup	Bob Hall Pier	50
9/26/2025	Dallas College	In-person	Nurdle Patrol	Microplastics and Nurdle Patrol	Dallas, TX	50
10/3/2025	University of Texas Marine Science Institute	In-person	Nurdle Patrol	Nurdle Patrol	Port Aransas, TX	20
10/20/2025	Islander CCA	In-person	Beachcombing/Nurdle Patrol	Beachcombing and Nurdle Patrol	TAMUCC	20
10/23/2025	TAMUCC President's Circle	In-person	Beachcombing/Nurdle Patrol	Booth	HRI-127	50
10/24/2025	Kaffie Middle School	In-person	Beachcombing/Nurdle Patrol	Beachcombing Texas Beaches	Kaffie Middle School	120
10/26/2025	Port Aransas Farmers Market	In-person	Beachcombing/Nurdle Patrol	Beachcombing Port A	Port Aransas, TX	100

10/28/2025	Dallas College presentation to Universidad Autónoma de La Paz, Mexico	Virtual	Nurdle Patrol	Nurdle Patrol	Zoom	50
10/30/2025	Baker Middle School	In-person	Beachcombing/Nurdle Patrol	Booth	Baker Middle School, C.C.	300
11/1/2025	Friends of Padre	In-person	Beachcombing/Nurdle Patrol	Booth	Briscoe Pavillion on North Padre	200
11/11/2025	Veteran's Memorial High School	In-person	Beachcombing/Nurdle Patrol	Beachcombing the Coastal Bend	Veteran's Memorial High School	120
11/14/2025	Austin school	Virtual	Nurdle Patrol	Nurdle Patrol Citizen Science Project	Zoom	35
11/21/2025	Martin Middle School	In-person	Nurdle Patrol	Nurdle Patrol	CBI	44

Figure 19: Tracy Weatherall at Friends of Padre Event November 1st 2025

Figure 20: Friends of Padre Event November 1st 2025