

Project Final Report: Near real-time detection and monitoring of invasive mussel species in Texas waterways

Greg Hamerly, Ryan McManamay, Shaif Chowdhury, and John Brubacher

October 31, 2023

PI Contact Information

Greg Hamerly - Greg_Hamerly@baylor.edu

TPWD Contract Number

CA-0002652

We acknowledge funding from the TPWD Statewide Aquatic Vegetation and Invasive Species Management funding, Rider 28.

Executive Summary

This report summarizes the findings of the TPWD-funded project “Near real-time detection and monitoring of invasive mussel species in Texas waterways”, funded by Texas Parks and Wildlife Department, September 2021-August 2023.

The purpose of this project was to explore the suitability of new imaging technology and analysis software for detecting the larvae of invasive zebra mussels in water samples taken from Texas waterways.

In this document we describe the methodologies we developed and executed for water sample collection, sample processing, and prediction using machine learning. We made significant progress in each of these areas. In particular:

- In water sample collection, we tried several methods for sampling from lentic sources using a deployed plankton net, and lotic sources using a pump, kayak, tubing, and an on-shore plankton net.
- In sample processing, we developed and improved our process that produces more consistent and representative sub-samples for imaging, compared to our previous simpler sampling processes that did not involve sub-sampling.
- In prediction, we developed three successively more effective models, ending with a model that is able to use sequential and contextual information for high sensitivity

and accuracy. Our top-performing model has an accuracy of 99.5% in discriminating between dreissenid veligers and other objects, which is a dramatic increase from the 89.4% accuracy of our initial model.

- We collected 68 water samples at a number of sites, focused primarily in the Highland Lakes region of the Lower Colorado River, for use in training and testing our system. We added to this a set of samples we collected prior to the project as well as samples provided by other agencies such as USGS (Lake Minnetonka) and LCRA (also in the Highland Lakes).
- We analyzed the water samples and looked for patterns that correlate between veliger presence with dam operations, method of collection, and time of collection. We did find some correlations indicating that the method of collection and dam operations are associated with different veliger prevalence rates, in particular that collection with kayak, in shallow water, during dam open/close is most effective for observing veligers.

Introduction

Zebra mussels (*Dreissena polymorpha*) and Quagga mussels (*Dreissena bugensis*) are not native to North American waters and likely arrived as stowaways in commercial vessels from Europe in the 1980s (Wong and Gerstenberger 2015). These species can spread rapidly, cause ecological disruption, and clog water pipes and other machinery (Wong and Gerstenberger 2015, Connelly et al. 2007). Due to economic and environmental damage it is important to detect and prevent the spread of these invasive species. Dreissenid mussels reproduce rapidly, laying millions of eggs per female per season, and can become established rapidly. By the time these invasive species have established themselves in a waterway, eradicating or mitigating their presence becomes very difficult and costly. Thus it is important to detect and monitor zebra mussels at the larval (aka veliger) stage (Counihan and Bollens 2017, Sepulveda et al. 2023).

Detection of veligers is usually done by collecting water samples and using microscopy with cross-polarized light for identification (Wong and Gerstenberger 2015), or using DNA-based methods (Sepulveda et al. 2023, Peñarrubia et al. 2016). Microscopy is expensive, time-consuming, and requires expert manual analysis. DNA-based methods are time-consuming, expensive, and are able to detect veliger presence but not prevalence. It's important to have an automated process that can monitor both veliger presence and *prevalence* (Wong and Gerstenberger 2015; Michigan. Zebra Mussel Task Force 1991; Fischer et al. 2021). There is other prior work with video-based detection of veligers using the FlowCam device (Johansson et al. 2020). Previous results show that FlowCam image analysis was correlated with DNA-based techniques. However, the narrow “flow cell” imaging unit tends to become clogged. Our project, which is not based on FlowCam, developed a new imaging chamber that reduces the potential for clogging, commodity imaging devices, and custom software with advanced machine learning for video analysis and discrimination between veligers and other objects.

We have been working since 2018 to develop automated methods of enumerating dreissenid veliger larvae using water samples, video capture, and machine learning. Prior to 2021, we developed early-stage technology for this task, including a microscopic video imaging device, tracking software, and a simple classification model based on single images. From August 2021 until July 2023, we worked on improving these technologies. In particular, we have worked on improving imaging (via improvements in the 3D-printed imaging chamber geometrical optimizations, and using a newer industry-standard imaging format with greater dynamic range capability), sample collection (from lentic sites using drift nets and lotic sites using shore-based pumps), sample processing (using centrifuges, sub-sampling, and sample agitation prior to imaging), video processing (video conversion and object tracking), and finally object prediction (using multiple machine learning models). We have also gathered water samples from the Highland Lakes chain on the Lower Colorado River in Central Texas, as well as obtained water samples from other infested sites elsewhere in the USA. We have analyzed these samples for prevalence of veligers as a function of location, time of collection, and related dam operations. Our results showed that our best prediction model has an accuracy of 99.5% at discriminating between veligers and non-veliger objects.

Methods

Sample collection

Two different sampling approaches were used to obtain water samples for analysis from a water body depending on the habitat and degree of flowing water (i.e., lentic versus lotic sites). Sample collection was also timed in some instances to coincide with dam releases to evaluate the effects of release on veliger presence and abundance.

Lotic Sites

In flowing water (water flowing sufficiently to suspend a drift net), a 1-m long, 30-cm diameter plankton net with 64- μm mesh was suspended at roughly 60% of the depth (Figure 1). In most cases, the net was attached downstream of a t-post, hammered in the substrate, but in some cases of individual samples, the net was held by a person. Keeping the net close to the post or person anchoring the net ensured that it remained at the correct depth. Nets were typically deployed for 15-minute durations. Samples were collected and concentrated in a cod-end filter and transferred to sample jars. An Ott-flow meter and wading rod was used to measure stream velocity at 0.6 depth and to calculate volume sampled. Volume sampled using the plankton net ranged from 13 to 45 m^3 .



Figure 1: Left image: below Ladybird Lake, 2021-12-03, setting up drift net staked to a post for capturing a sample in flowing water. Right image: below Mansfield Dam, 2022-05-26, setting up drift net in the dark.

Lentic Sites

In relatively still water, two methods were employed to examine differences in capture efficiency. For most samples in lentic water, a 300-gallon-per-minute rated gasoline-powered semi-trash pump was used with 3" diameter PVC hose or corrugated plastic pipe, with the end submersed to settle (0.3-0.5m above the bottom of the river, or 1-2m below the reservoir surface) at approximately 20-30 meters from shore (Figure 2). Sampling depth was limited in our collection by pipe length. A kayak was used to suspend the pipe from the bottom. During sampling, the pump was operated for 10-20 minutes, depending on depth (change in head affects pump velocity). The outlet of the pump was oriented into the plankton net mentioned above to collect samples. Pump volumetric discharge was measured by calculating the time required to fill a 5-gallon bucket. Time to fill a 5-gallon bucket generally ranged from 9s to 12s, which resulted in pump rates of 25 to 33 gallons per minute. As a comparison of sampling efficiencies in relatively still water during the 24-hr sampling event, a plankton net was attached 5m downstream from a kayak, suspended 0.5 to 1m below the surface, and was towed a straight line distance (approximately 100m) in the main channel. The distance of the tow was measured to calculate volume filtered, which ranged from 11 to 14 m³.



Figure 2: These images are from below Mansfield Dam, 2021-09-18. In the left image, the kayak, water pump (foreground), and tubing are visible. On the right, the mesh net for sample collection (used for pump water) is visible.

Water Sampling Locations

Water samples for analysis were collected from sources in the Highland Lakes chain: Ladybird Lake, Lake Austin, Lake Travis, and Lake LBJ. For each of these locations, we collected samples both in the lake and just below the reservoir dam. We also collected 4 samples from below the dam at Lake Waco. In total, we collected 64 water samples from the Highland Lakes.

Outside the work of this project, we collected additional water samples and were able to use them for analysis. Prior to this project start, we collected approximately 25 water samples from below Davis Dam on Lake Mohave. Approximately 20 water samples were collected in Lake Minnetonka and provided to us by US Geological Survey. These USGS samples were of dreissenid veligers sorted by life stage (D-hinge, umbonal, and pediveliger).

Sample processing

Sample Preservation

Samples were preserved immediately upon collection by adding 95% reagent-grade ethanol to the sample soon enough after acquisition to preserve the organic material without degradation.

Sub-sampling of Preserved Samples

Samples were agitated and thoroughly mixed, and 5- to 20-ml sub-samples were taken for centrifugation. A centrifuge was utilized to create some separation of the higher-density veligers from the other planktonic organisms. A micropipette was then used to place 1 ml or less of concentrated sample into a 5 ml centrifuge tube containing a preservative (isopropyl alcohol or similar). The sample was centrifuged as necessary to increase the concentration of veligers near the bottom of the well in the tube. A 1 ml micropipette was then used to collect a sample from the “well” at the bottom of the 5 ml tube and place it into a ≤ 2 ml volume tube, filling it with isopropyl alcohol as necessary.

Imaging of Sub-samples

We used a 10 μ l micropipette to place 0.5 μ l of sample into a 1.5 ml tube, and we performed this step 6 times. The result was 6 sub-samples in 6 tubes. Then we added 1.0 ml of distilled water into each tube, mixing the sample into the overall tube and let this stabilize for approximately an hour. Once the imaging system was configured to accept and process the injection of sub-samples, we used a 1.0 ml micropipette to inject the sub-sample into the stream of water flowing through the system. Depending on the density of this sub-sample, we injected the sub-sample for a duration of approximately 30 to 60 seconds.

As there are 6 sub-samples per sample, these imaging procedures generated 6 imaging events to be processed per sub-sample.

Software analysis

After obtaining the video imagery of a water sample, the video was processed using multiple software steps in a pipeline, with six imaging events for each processed sub-sample. Processing time during the object tracking, extraction, and grouping stage was correlated to the spatial density of the sub-sample as it flows through the imaging system. The approximate processing time for each imaging event, by stages, was as follows:

- Video processing: ~ 5 minutes
- Object tracking, extraction, and grouping: ~ 0.5 to 1.5 hours
- Training (when needed)
 - Object labeling: ~ 1 to 2 days
 - Machine learning model training ~ 1 hour on a NVIDIA 3080Ti GPU
- Machine learning object prediction (when testing or during deployment) ~15 mins for each sub-sample.

Video Processing

The video was acquired in a ProRes RAW format to maximize the dynamic range of luminance and color spectrum data. A video clip was then created to include the “dead-space” before any particles are present in the image to normalize the process, the entire imaging event when the microscopic particles are flowing through the system is captured, and then the clip ends once the particle flow is depleted.

Object Tracking, Extraction, and Grouping

Object tracking was accomplished by using a Kalman-like filter (Saho 2017). This allowed us to automatically follow objects through the video viewport across different frames. We also used a simple size rejection filter at this stage to automatically ignore objects that were either too large ($>200\ \mu\text{m}$) or too small ($<70\ \mu\text{m}$) according to the expected sizes of dreissenid veligers (USGS, n.d.). Typically an object appeared in view for 20-60 frames, depending on the speed of the water flow in the chamber. Once we had tracked them, we extracted a cropped image from each video frame of that object, typically resulting in 20-60 cropped images per object. Since we tracked the object, we could then group images from the same objects. We placed these grouped images into a directory structure containing one directory of images per tracked object.

Model Training

When needed, images extracted from a video source are used to train our machine learning models. Training only needs to be done once, but can be updated later. Updates are only needed when the distribution of the data significantly changes, such as sampling from water sources that have significantly different populations. However, a robust training set with a representative sample of dreissenid veligers should make it unnecessary to retrain often.

The first step in training is to assign a label to each object (group of images) of dreissenid or non-dreissenid. These labels came from human experts observing the video and extracted images. For samples which were provided to us, the samples were already classified, and we used those labels. We labeled the samples we collected, and sought confirmation from outside experts for our process. We can then use these labeled images to train or improve a machine learning model (Kandhare 2013).

Machine Learning Object Prediction

The next step is to apply a trained machine learning model to predict whether each object is dreissenid or non-dreissenid. This is useful for evaluating the predictive performance of the system (on known samples), or in deployment for detecting dreissenid mussels in samples taken from the field.

Machine Learning Methods

The central prediction problem in this project was discriminating between images of objects that appear to be dreissenid larvae (the positive class) and everything else that appears in the video (the negative class) (Kandhare 2013; Dixon 2018). Because we have a video-based imaging system, most objects that are tracked end up having many (30-50 or more) images per object. Due to tracking through the video frame, the video processing pipeline can automatically group images of the same object (Kandhare 2013; Dixon 2018) even if their class is not known.

With this in mind, we developed and tested three very different machine learning-based prediction models:

1. Voting model. This model uses a voting scheme to combine individual votes on separate images for a single object.
2. Feature set averaging model. This model makes a single prediction on a single feature vector for a single object, where the vector comes from averaging the different image representations in a latent feature space.
3. Video model. This model makes a single prediction on a sequence of images taken from the video.

All of these models use some similar components, including multiple convolutional layers for extracting features from images. The best-performing model was the video model. Below we give more details on each of these three models. Model parameters are shown in Table 1.

Voting model

This model makes a single prediction (dreissenid/non-dreissenid) **per image** that was extracted from the video. The basic prediction model is a convolutional neural network (CNN) with a fully-connected layer, similar to VGGNet (Brownlee 2019). So, if we have N images of an object, then we make N separate (independent) predictions on those images. We then combine them using a voting scheme. We can set a threshold on the vote to make it more or less sensitive (and less or more likely to produce false positive predictions). Figure 3 shows a diagram of the voting model.

The advantage of this model is its simplicity: we can train one neural network model on a collection of single images, and make predictions in the same way. The disadvantage of this model is that it does not take full advantage of the fact that we have multiple images of the same object; each image is separately classified without regard to the information in other images of the same object. These different images can have different perspectives (as the objects tumble through the field of view) and have related context that we want to use. To make use of this additional context, we turned to the feature set average model.

Feature set averaging model

This model makes a single prediction **per object**; each object is represented in the prediction process by multiple images of that object extracted from different frames of the video. This model is really a pipeline of two models: one deep CNN-based autoencoder network for feature creation, and one fully-connected network for prediction (Brownlee,

2019). For one object with N images, the feature set average model converts those N images into N latent feature vectors using a CNN-based image autoencoder that has been pre-trained for this task. The autoencoder has two parts: an encoder followed by a decoder. It is trained to reduce the dimension of the image to a latent feature space (using the encoder), and then the decoder takes that representation and generates an output image that is as similar as possible to the input image. After training, the second half of the network (the decoder) is discarded, and we use only the encoder to produce a compressed, latent feature representation of each image. Figure 3 shows a diagram of the feature set averaging model.

Given N images, passing each of them through the encoder gives N feature vectors in a lower-dimension latent feature space (the original image space is typically $20 \times 20 \times 3 = 1200$ dimensions, while the latent space is typically 64 dimensions). We then combine those N latent feature vectors into one vector in the same space using averaging. Then we make one prediction on the averaged feature vector using another feedforward neural network. The advantage of this model is that it makes one prediction rather than N predictions that must be combined (Yue et al. 2023).

This model and its results were published in ISVC 2022 (Chowdhury and Hamerly, 2022). Some disadvantages of this model are that it discards sequence information (which image comes from which video frame over time) and it requires two stages (encoder, then prediction model). These led us to the next model, which was a video-based model.

Video model

This model makes a single prediction **per object**. The video model takes a fixed-length ordered sequence of K images (which is typically fewer than the entire set available) for a single object. It then makes a prediction on the sequence. These K images are passed into a model that is useful for video sequences due to two key features: the attention-based transformer units (Peng et al. 2023) and the LSTM (long short-term memory) units (Sabry 2023). The advantage of this model is that it allows us to make predictions on a sequence of images while keeping their order as context. A disadvantage of this model is that it is limited to a fixed length set of images, thus discarding some information on objects that have many images. It also requires significant training time. An advantage of this model is that it performs very well and effectively uses the context of multiple images (Hsu, Liao, and Huang 2023). This model was the best-performing model, and the one we believe is best suited for dreissenid detection. Figure 4 shows a diagram of the video model. The use of this model for dreissenid detection in this study was published by Chowdhury et al. (2023).

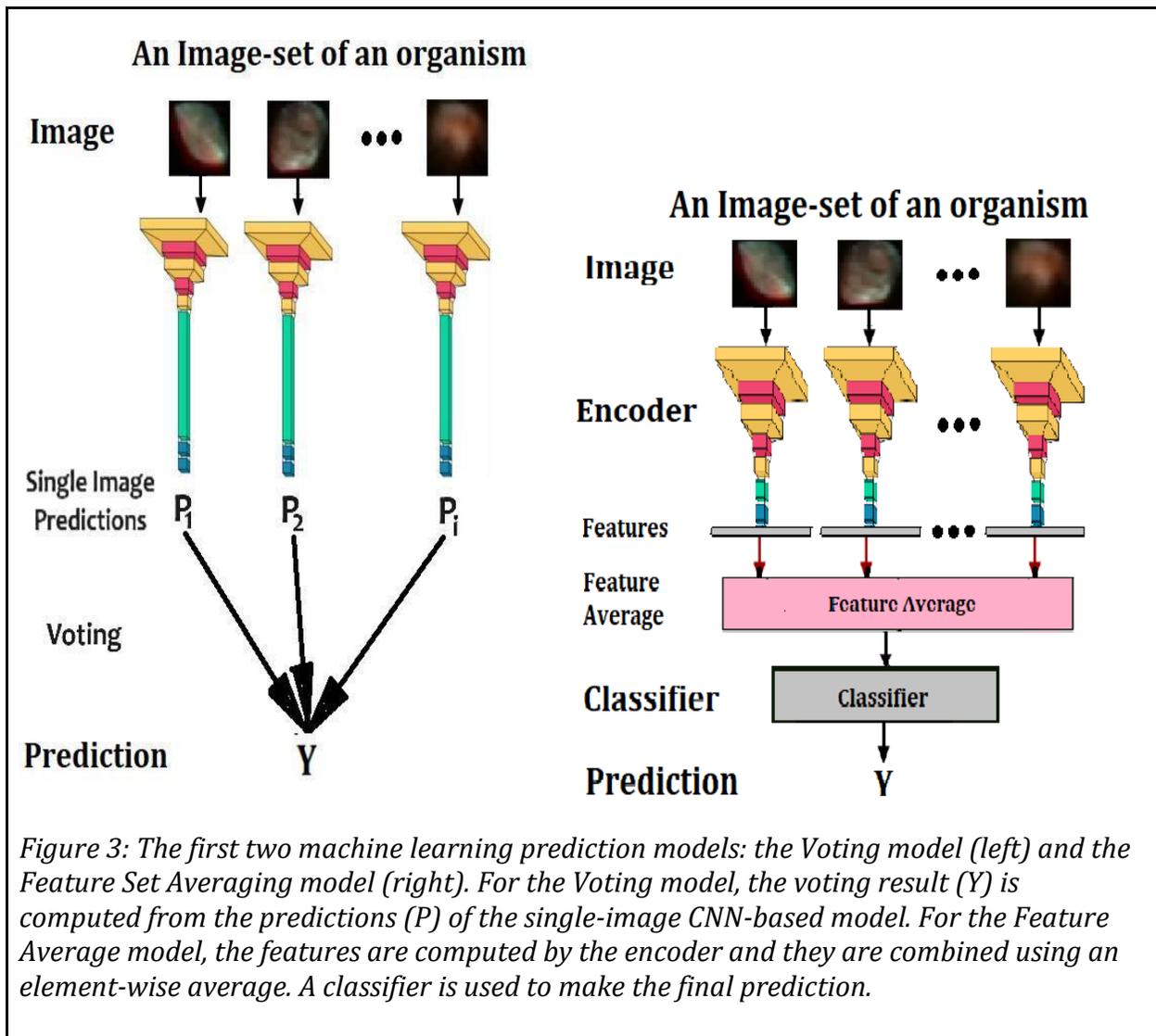


Figure 3: The first two machine learning prediction models: the Voting model (left) and the Feature Set Averaging model (right). For the Voting model, the voting result (Y) is computed from the predictions (P) of the single-image CNN-based model. For the Feature Average model, the features are computed by the encoder and they are combined using an element-wise average. A classifier is used to make the final prediction.

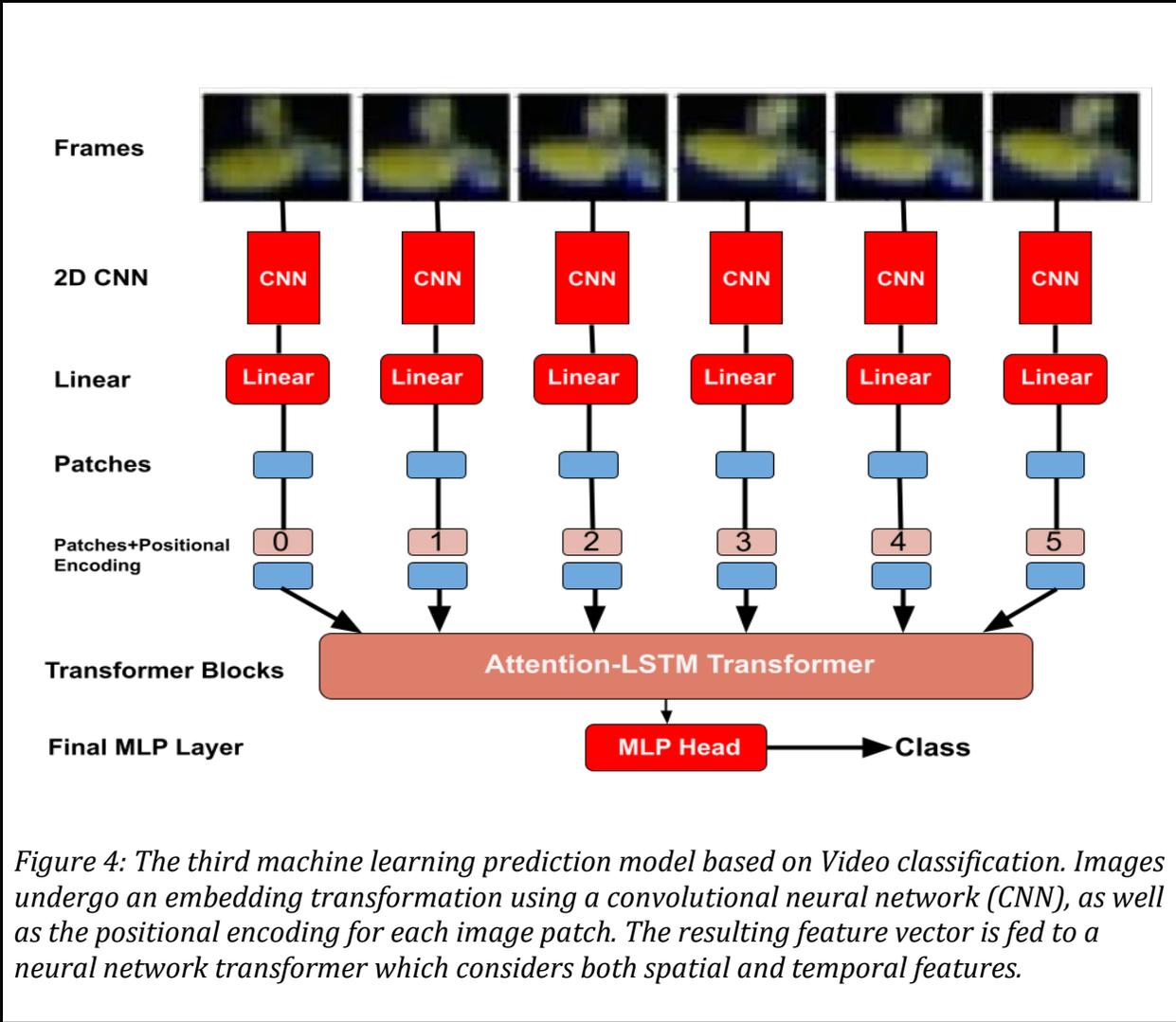


Figure 4: The third machine learning prediction model based on Video classification. Images undergo an embedding transformation using a convolutional neural network (CNN), as well as the positional encoding for each image patch. The resulting feature vector is fed to a neural network transformer which considers both spatial and temporal features.

Table 1: Configurations of three types of neural network models (voting, feature average and video). The Feature Average models are composed of two different parts – the autoencoder and the classifier. So for those models, we list the sizes of the two parts separated by plus (+). The Feature Average model also lists the number of latent features output by the autoencoder. We attempted many more models than what are shown here; we show only a small subset for brevity. For more complete details, please see Chowdhury and Hamerly (2022) and Chowdhury et al. (2023).

| | ID | Model | # Parameters | # Convolution Layers | # Dense Layers |
|------------------------|-----|---|----------------|----------------------|----------------|
| Voting model | V | Single-image CNN voting model | 204,512 | 3 | 3 |
| Feature Average models | F64 | Autoencoder + classifier (64 features) | 38,297 + 2,362 | 7+0 | 3+3 |
| | F48 | Autoencoder + classifier (48 features) | 34,281 + 1,850 | 7+0 | 3+3 |
| | F16 | Autoencoder + classifier (16 features) | 26,249 + 276 | 7+0 | 3+3 |
| Video Models | A1 | Attention-LSTM-S 2D CNN (32 patch size) | 47,066 | 1 | 4 |
| | A2 | Attention-LSTM-B 2D CNN (64 patch size) | 844,154 | 1 | 8 |

Results

Procedural Refinements for Sample Processing

During this study, we achieved significant refinements in the sample processing pipeline to improve ease of processing, image quality, and image processing. These are all precursors to the machine-learning based prediction stage.

Imaging

We upgraded the imaging camera to a Lumix S5 and Sigma 70mm 1:2.8 DG Macro Lens and made design parameter changes to the 3D-printed imaging chamber and related components to support compatibility with a 2021-era camera. During the supply chain issues of 2021 - 2022 we had significant challenges acquiring clear resin for our SLA-based 3D printer that wasn't too porous and/or brittle, causing this component to leak or fracture. In order to increase the homogeneous-spatial-distribution of the particles flowing through the chamber, we reinstalled a sample agitator (helix, see Sample Manipulation for Imaging below), which we had done for previous projects when we were acquiring live, and not preserved samples.

Sub-sample Preparation for Imaging

We introduced a centrifuge to our sample process to provide some increased separation of higher-density veligers from other planktonic material to minimize loss of veligers into a subsample. The results of this step were that we did increase separation of material types, but it also tended to introduce a matrix of organic material into sub-samples, which could sometimes cause imaging periods with poor spatial separation of organisms. This was ameliorated by agitating the sample before imaging using the helix (discussed below).

While our imaging system can work with preserved samples, it is intended and tends to operate more easily with fresh samples directly at the source. Hence, for preserved samples we added a dilution step prior to imaging. We diluted each preserved sub-sample to roughly 95% H₂O before injection into the imaging system to increase the homogeneity of the fluid sample flowing through the system.

Sample Manipulation for Imaging

We reinstalled a 3D-printed helix ("corkscrew") component that the sample flows through prior to entering the imaging chamber. This increased agitation of each sample for better orientation-invariant, or "pseudo-3D" imaging. It also helped mix and distribute each sample more evenly within the imaging chamber.

Software

To enhance video processing, we modified video and image processing parameters to better resolve distinctions between similar microscopic organisms such as ostracods, and different veliger life-stages. To enhance object tracking, extraction, grouping, we adjusted size rejection parameters to better optimize tracking results around veligers. We do not have metrics to demonstrate this improvement, but it became visually apparent that the clarity of our imaging improved with these modifications.

Neural Network Results

The results showed large improvements in the modeling capabilities from the original voting model (“V”) to the feature averaging model (“F16” through “F64”) to the video models (“A1” and “A2”; we use “A” because these are “attention-based” models). These models make progressively better use of the sequential information that our imaging system produces. The resulting performance improved from a per-organism accuracy of approximately 88.1% (for voting model “V”) to over 99% (for the video model “A1”).

The Feature Average models performed significantly better than the Voting model on all the metrics we measured: F1 Score, Balanced Accuracy (BAC), and Recall (Table 2). The video model then improved on the feature average model in both F1 Score and Accuracy (Table 3).

| Table 2: Neural network results comparing the Voting model (V) with the Feature Average model(s) (ID beginning with F) on a dataset with 4,374 organisms and 112,788 images. There were 674 (15%) organisms that had the class of “dreissenid” (positive class), and 3,700 (85%) with the class of “non-dreissenid” (negative class). BAC is balanced accuracy. We used 20% of the held-out test data for the results shown here. The statistics shown in this table are per organism, rather than per image. | | | |
|---|-----------------------|--------------------|---------------------|
| ID | F1 Score (Dreissenid) | BAC | Recall (Dreissenid) |
| V | 88.1 ± 0.7% | 89.4 ± 0.3% | 82.5 ± 0.6% |
| F64 | 97.1 ± 0.9% | 98.2 ± 0.7% | 96.3 ± 0.5% |
| F48 | 97.1 ± 0.3% | 98.2 ± 0.3% | 96.3 ± 0.4% |
| F16 | 90.5 ± 0.3% | 95.2 ± 1.2% | 88.8 ± 1.5% |

Table 3: Neural network results for two of the video models on a dataset with 6,905 organisms and 221,702 images. There were 1,220 (17%) organisms that had the class of “dreissenid” (positive class), and 5,685 (83%) with the class of “non-dreissenid” (negative class). We used 30% of the held-out test data for the results shown here. The predictions shown here are per organism, rather than per image.

| ID | F1 Score (Dreissenid) | Accuracy |
|-----|-----------------------|----------------------|
| A1 | 99.15 ± 1.5% | 99.51 ± 0.87% |
| A2 | 98.56 ± 2.71% | 99.51 ± 0.87% |
| F48 | 97.1 ± 0.3% | 98.2 ± 0.3% |
| F16 | 90.5 ± 0.3% | 95.2 ± 1.2% |

More comprehensive technical details of these neural network models and model comparison analysis results from this study can be found in Chowdhury and Hamerly (2022) and Chowdhury et al. (2023). In those papers we also compare with a number of other network variants and other machine learning classifiers such as SVMs, PCA, and k-Nearest Neighbor.

False Positive Analysis

In the second year of this project, we conducted experiments to analyze the tendency of our prediction techniques to incorrectly identify organisms visually similar to dreissenid larvae. In particular, we looked at larvae of Asian clams (*Corbicula fluminea*) and ostracods. We obtained ostracod samples from TPWD as well as purchased and grew our own samples of these other organisms in the McManamay laboratory. While we obtained Asian clam specimens, we were not successful in obtaining or cultivating larvae of this species. Therefore, evaluation of potential for false positives for visually similar organisms focused on juvenile ostracods.

This study showed strong evidence that our prediction model can discriminate between dreissenid larvae, ostracods, and other objects in the water column (Table 4). For this small study, the accuracy was 98.5%, with none of the true ostracods being predicted as dreissenid.

Table 4. This is the confusion matrix for our experiment involving ostracod larvae, dreissenid veligers, and other objects. This model does not have much difficulty discriminating between the three classes, and most importantly that none of the true ostracods were falsely predicted as dreissenid.

| | Predicted other | Predicted ostracod | Predicted dreissenid |
|-----------------|-----------------|--------------------|----------------------|
| True other | 1270 | 2 | 13 |
| True ostracod | 0 | 18 | 0 |
| True dreissenid | 6 | 2 | 267 |

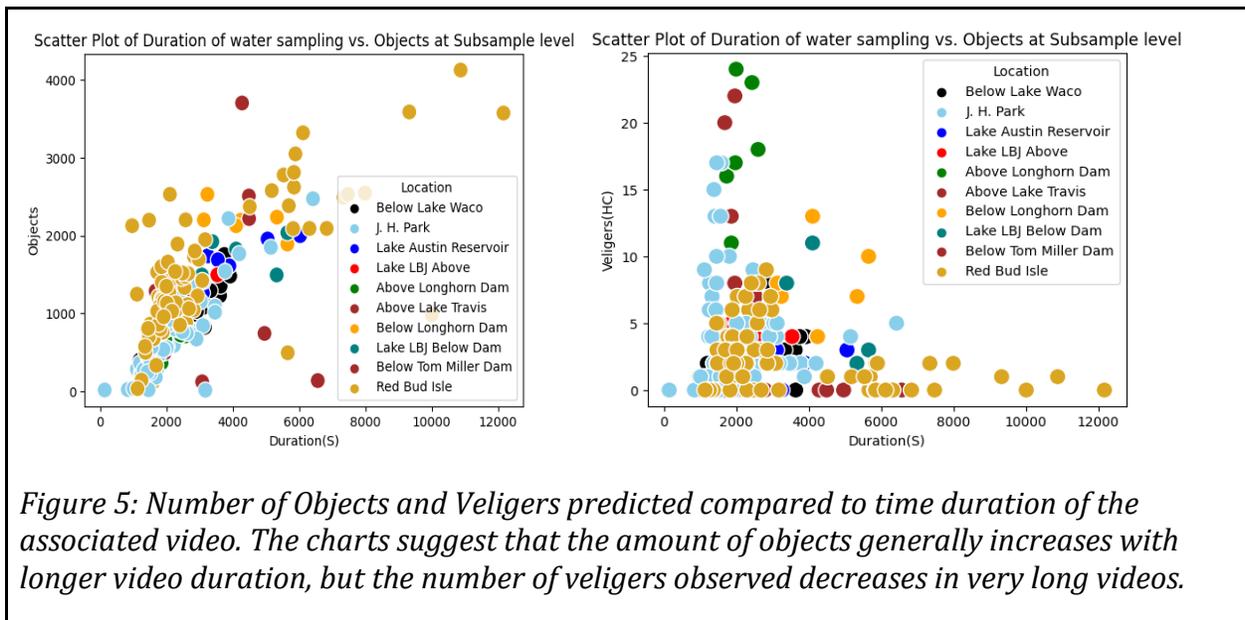
Water Sample Analysis

This technology (i.e., video model) was applied to analyze water samples we collected and processed for dreissenid early detection.

We trained each prediction model in two steps with all images taken from every organism. At first we train from images taken from older dreissenid species data from Davis Dam on the Colorado River downstream from Lake Mead (quagga mussel infested) collected by us in 2019, as well as data we collected in the Highland Lakes from March 2022 and Jan 2023. Then we fine-tune the model on a smaller set of labeled data obtained from the U.S. Geological Survey (USGS) from Lake Minnetonka serving as ground truth. We train for 50 epochs (i.e. dataset presentations) with three different classes: dreissenid, ostracods, and other (Bertin 2017; Sarkar, Bali, and Ghosh 2018). While ostracods are also non-invasive, we treat them as a separate class so that we can observe false positive rates related to the larvae of this similar-appearing organism.

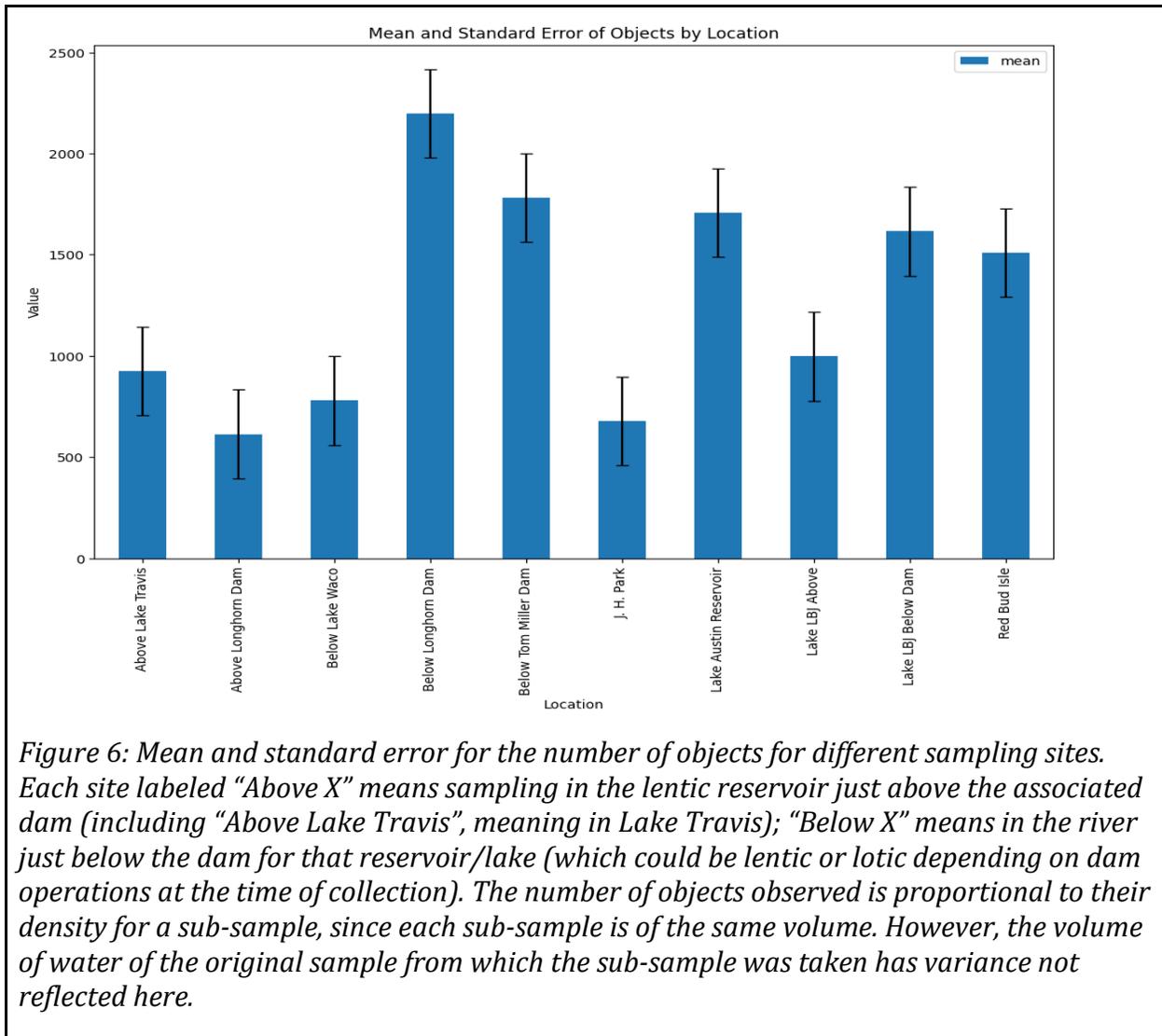
We have analyzed the data generated by water samples taken from Red Bud Isle, Lake Waco, Longhorn Dam, Lake Travis, Tom Miller Dam, Austin Reservoir and Jessica Hollis Park in Austin, Texas. The data we have comes from five sub-samples of each original water sample. (In Methods, we mentioned using six sub-samples, which is our current practice; the data we have to analyze was from a time when we were using five.) Each sub-sample was processed as described in Methods to create a video, and then multiple images of each object appearing in the video. We have measurements for the volume of water sampled (based on duration and flow rate for both lentic and lotic sources), length of the video (i.e., seconds), number of objects appearing in the video, and the predicted class for each object.

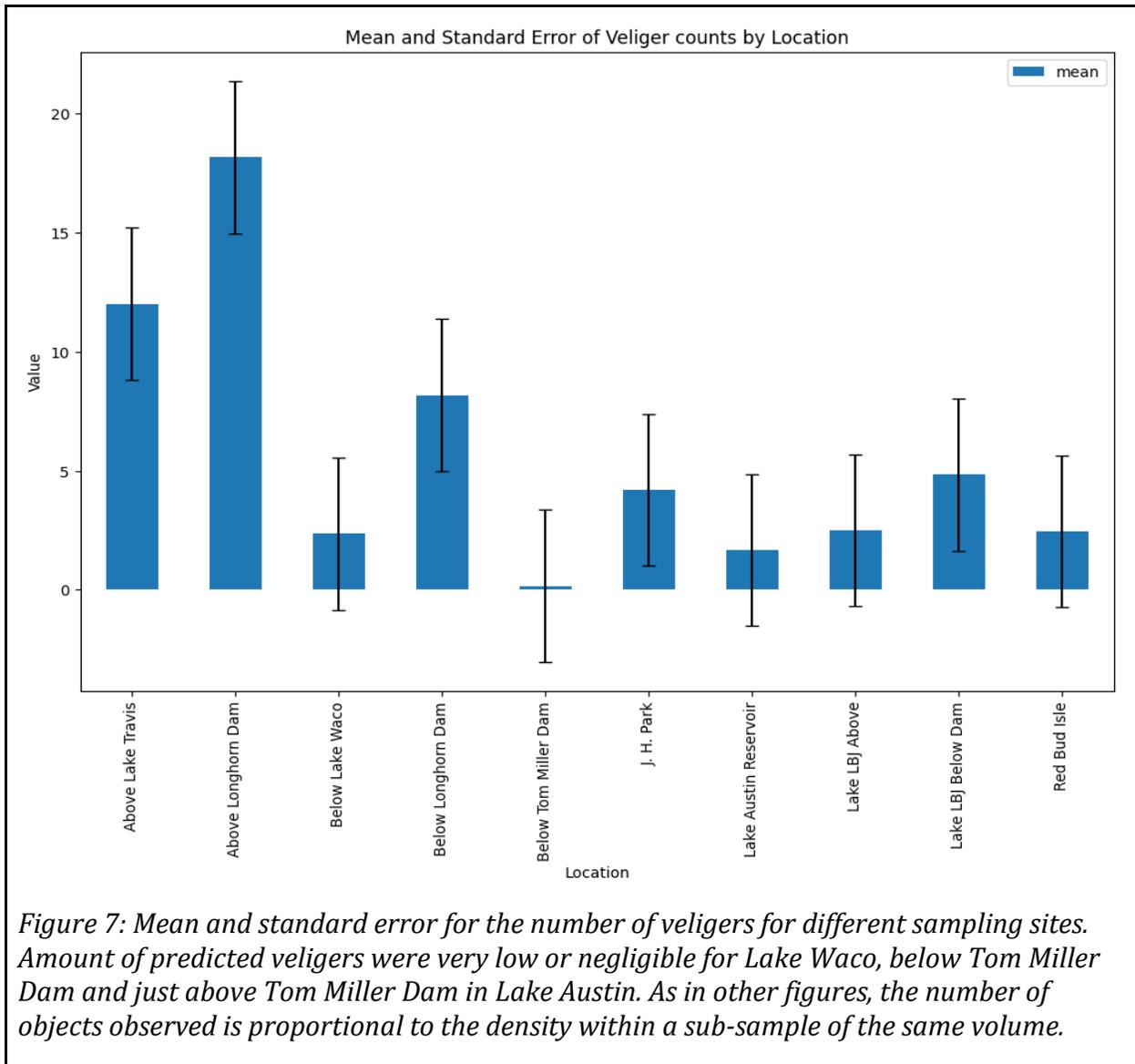
First we can look at the number of objects, veligers, and ostracods based on the duration of the video. As indicated in Figure 5 the number of objects increases linearly with the video duration ($R^2 = 0.93$). However, the number of veligers increases sharply and then decreases as the video duration increases ($R^2 = 0.126$). We believe therefore that the number of veligers observed is more influenced by other factors.



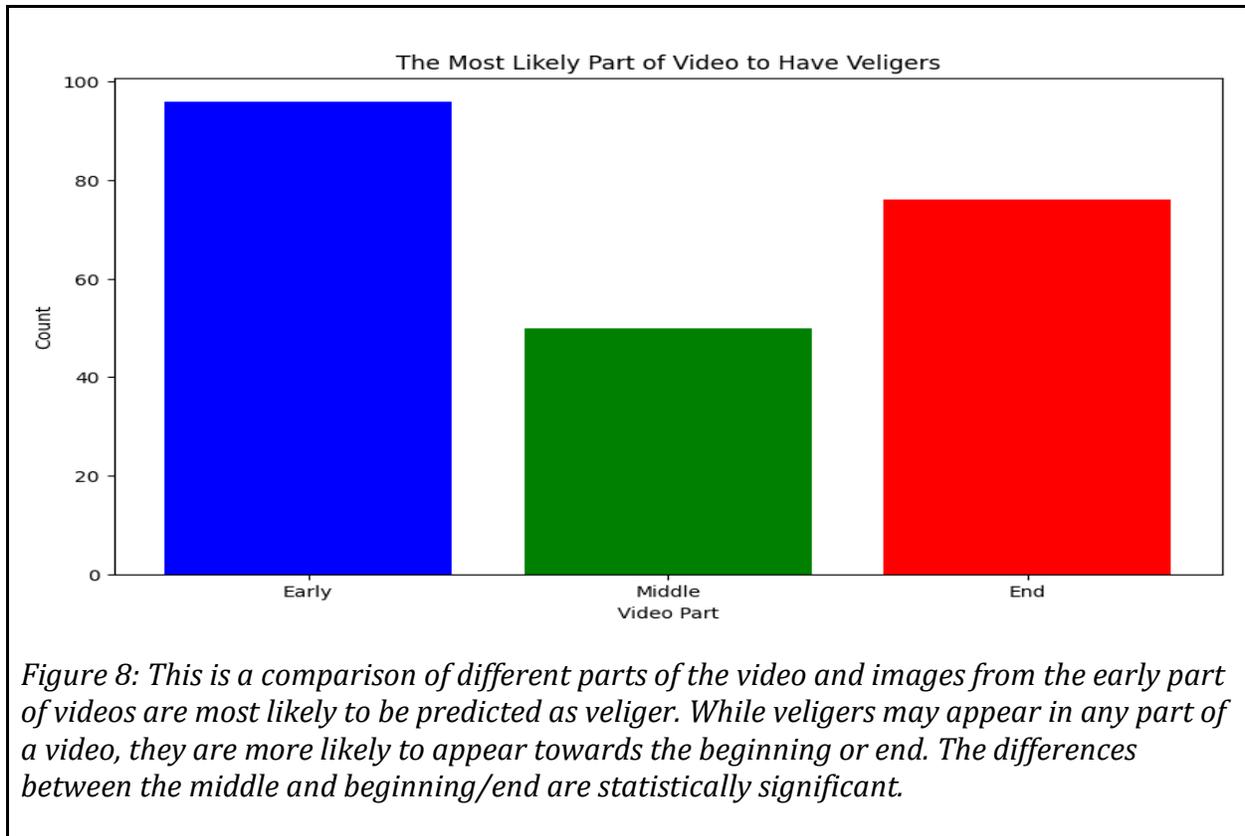
Throughout this analysis, we enumerated veliger presence (i.e. counts). This enumeration is correlated to, but not the same as veliger density in the original water sample. Each subsample is of the same volume, so a higher count is representative of a higher density within that fixed volume. But the sub-samples come from original samples which can have different volumes.

We further evaluated data for objects (Figure 6) and predicted veligers (Figure 7). The number of veligers was high for Lake Travis and LongHorn Dam. The number of veligers was very low or negligible for Lake Waco (downstream from the dam), below Tom Miller Dam, and just above Tom Miller Dam in Lake Austin. Given efforts to eradicate zebra mussels in Lake Waco and recent confirmed lack of establishment post-eradication efforts, we presume that detection of veligers downstream of Lake Waco is likely a false positive; however, we cannot be certain without DNA-testing. We recommend further investigation and routine monitoring of the site be conducted.

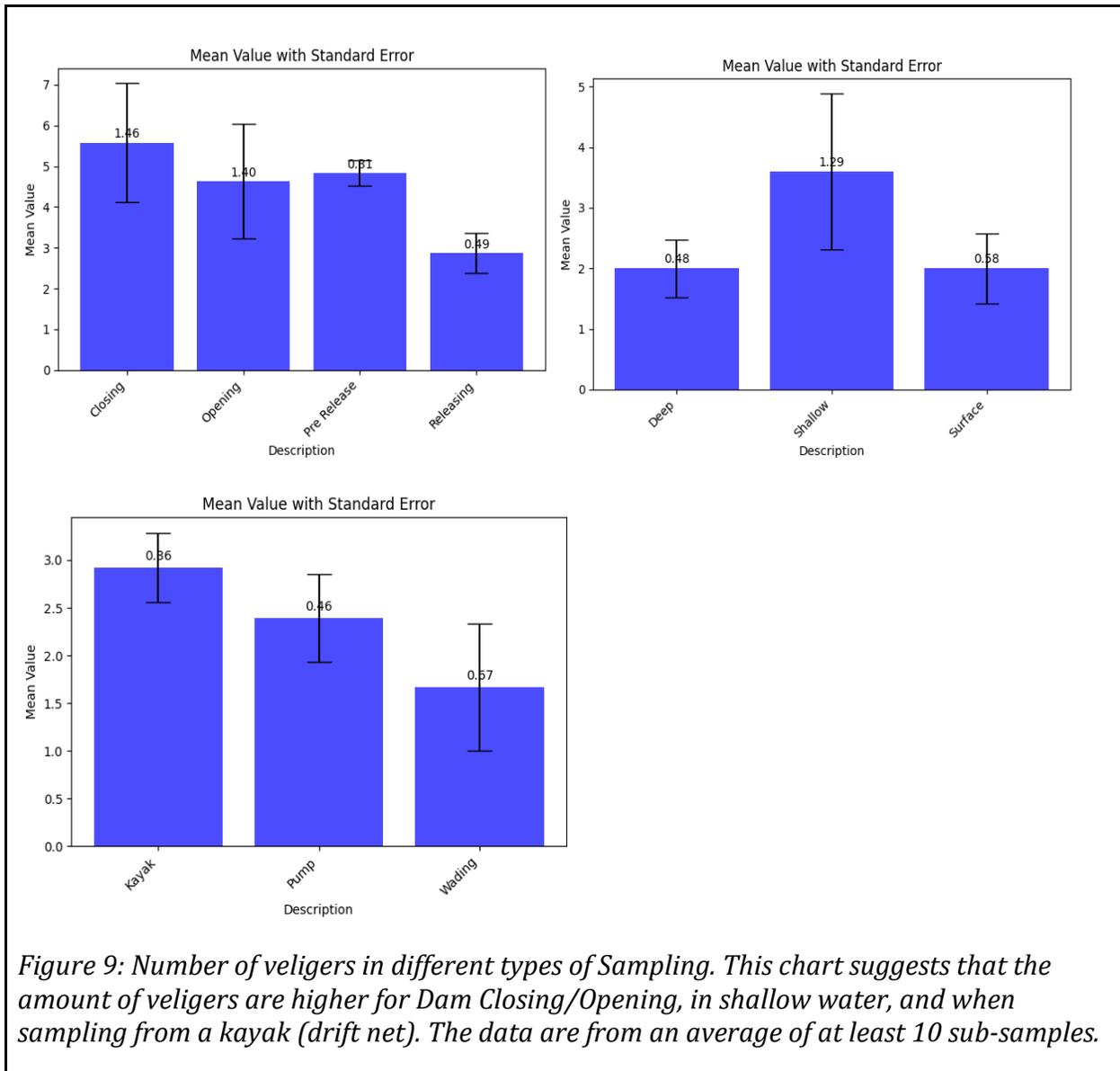




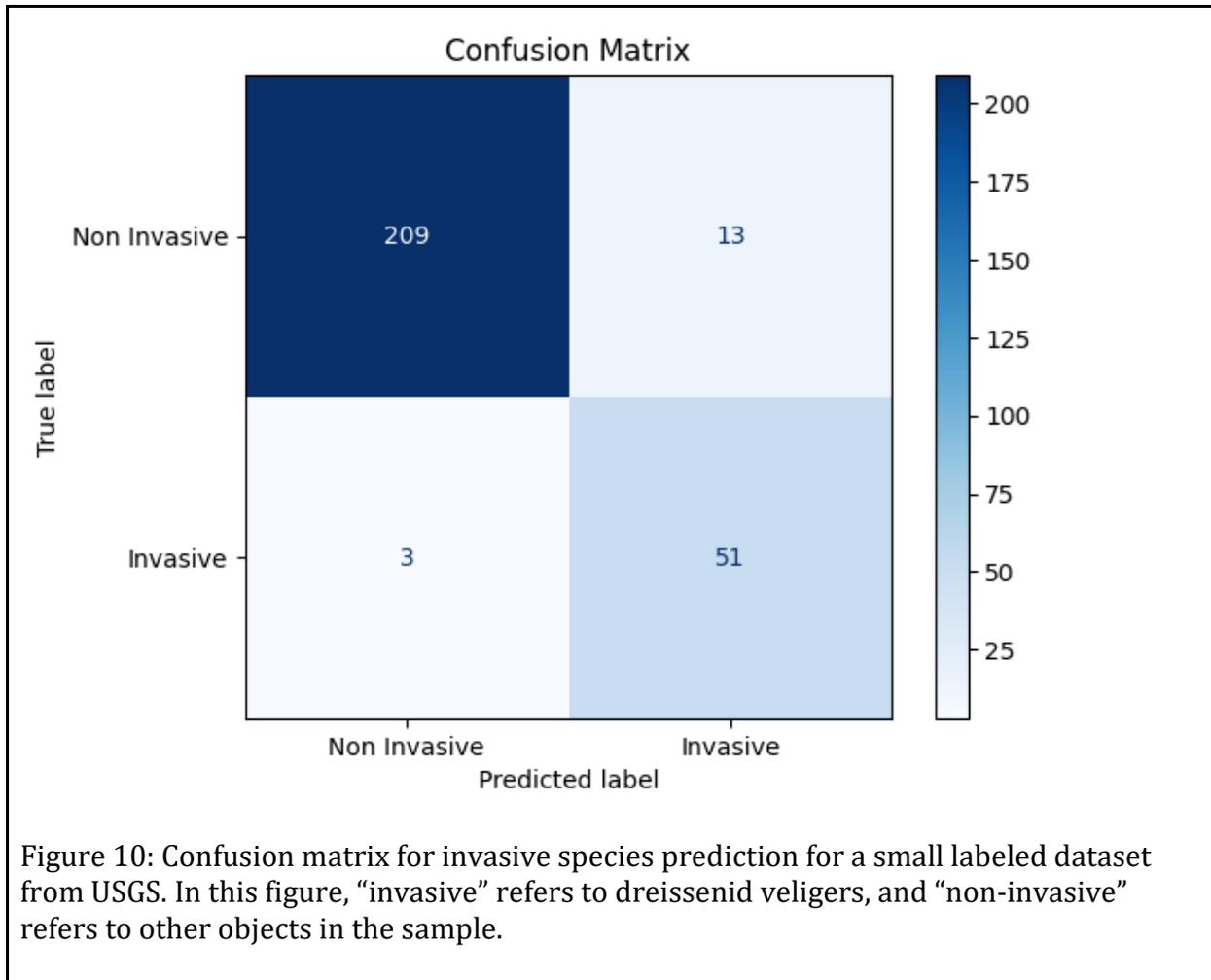
The predictions are made from the set of images for every object. It is interesting to see which part of the video is most likely to include images that are predicted to be veligers. A comparison of the part of the video that is most likely to be predicted as veliger (i.e., early, middle, or end; see Figure 8). Our data suggest that video samples tend to display veligers more prominently at the beginning and ends rather than in the middle of the video, and this trend is statistically significant.



We also evaluated whether veliger presence changes based on dam operation (i.e., recently opened or closed), or whether the sample was originally taken from deep or shallow water, and sample collection method. We plot the veliger percentage based on position of sampling (Figure 9). Our data suggest that there was a correlation between dam operation and location and veliger concentration, that we observe more veligers for dam opening (5.50 +/- 1.46) and closing (4.50 +/- 1.40) and pre-release (4.60 +/- 0.81) than during release (2.70 +/- 0.49). We observe more veligers when sampling from shallower water (3.50 +/- 1.29) than deeper water (2.00 +/- 0.48) and when using a kayak (2.59 +/- 0.36) versus pumping (2.50 +/- 0.46) versus wading (1.55 +/- 0.57). We do not have sufficient data to show statistically significant differences in all comparisons, though the lowest and highest sampling types – kayak versus wading – are significantly different. Shallower water and the use of a kayak with a stationary drift net likely provide conditions more conducive for detecting veliger presence, due to factors like water flow and habitat suitability (Delaney 1997; Wong and Gerstenberger 2015). Indeed, it is easier to obtain larger volume water samples with a drift net in flowing water than using a pump of the size we employed.



Finally, we conduct a validation experiment on a small amount of labeled data from USGS processed on 08/30/2023. This is a small data set with 222 non-dreissenid objects and 54 dreissenid veligers. Our model trained on the aforementioned data set has accuracy of 94% and an F1 score of 90% with dreissenid as positive class. The confusion matrix for prediction is given in Figure 10. Note that these results are for a particular prediction threshold (i.e. if prediction > threshold, predict dreissenid).



Discussion

Dreissenid larvae detection is possible at high accuracy rates (over 99% on a per-object rate for our most sophisticated prediction model, the video model) using automated visual detection, for a very small cost in false positives. The FPR in the confusion matrix above is 5.5%, but this is for a very small test set of 276 images. On a much larger training set with 122,788 which was used in pre-training the model, our FPR is only 0.053%. No prediction algorithm can ever completely eliminate the risk of false positives, but adjusting the threshold of prediction can reduce this risk at the cost of true positives.

This study showed that water sampling by drift net (either kayak tow or in flowing water) was more effective than water sampling via pumping, as it is easier to collect larger water volumes, which leads to greater numbers of observed veligers.

The time required to apply our methods depends on the type of sample. Fresh samples are much faster, on the order of 2 hours to process beginning to end. With the additional work required for processing preserved samples, the process takes a longer time – roughly day of work to process. As the technology and our processes improve, we expect these times to drop. Traditional microscopy requires 18-24 hours to settle the sample, 5 minutes to concentrate the sample, and 15-60 minutes to do cross-polarized light microscopy analysis (according to email communication with Greg Southard at TPWD).

One key and unique quality of this approach is obtaining multiple views of the same object at different times and from different perspectives. We developed models that are able to integrate these different views to improve accuracy and prediction confidence. Not only can this technology efficiently detect the presence of invasive organisms, but also their prevalence on a large scale, and gives evidence that can be verified visually. While this shares these visual aspects with microscopy analysis, what sets it apart are the scale, automation (prediction, enumeration), and ability to review past recordings. In addition, samples processed by our system are not harmed and can be collected for further analysis (i.e. with our system in a treatment/response loop, or using microscopy or DNA analysis for confirmation analysis).

We have identified methodology improvements for preparing water samples for imaging. In particular we used sub-sample processing (with centrifuges and micropipettes) that allow greater consistency in sub-sample composition.

There are many possibilities for further work. These prediction models require thresholds, and tuning of these thresholds is important for adapting things like false positive rates, balanced against accuracy. In a multiclass setting (e.g. dreissenid/ostracod/other), there are multiple thresholds that need to be balanced. We also observed in this study that veligers tend to be more prevalent on the ends of captured videos, and this merits further investigation as to whether this pattern is consistently occurring and why. The addition of the centrifuge helped separate veligers from other material, but appears to increase the structure of an "organic matrix" in the sub-sub samples which then generates some imaging periods with poor spatial separation, causing challenges for the tracking algorithm. The helix helps breakup the "matrix", but it appears more agitation is necessary. As the technology and methodology both mature it enables higher-frequency water sampling, and finding the right sampling frequency is essential for monitoring and detecting veliger prevalence. We also would like to study the variation in detection rates for different life stages of veligers.

Conclusion

This analysis represents a significant step forward in detection of dreissenid larvae. By leveraging deep learning techniques and conducting large scale experiments, we have demonstrated the potential for accurate and timely identification of these invasive species. In particular, we have measurably improved the per-organism neural network prediction accuracy from around 88% (for the "voting model") to over 99% (for the "video model") by

improving models so they make better use of the sequential and contextual information provided by our imaging platform, as well as the different perspectives. We also showed how data from different water samples sheds light on the optimal time and type of water sampling for larval dreissenid detection.

Furthermore, our analysis of temporal variations in veliger presence offers valuable insights for targeted sampling in response to dam operations. Hopefully, this research can help with preservation of ecological balance and protection of vital water resources in Texas waterways.

While this is the project final report, we are continuing with this work, as there are many open questions we intend to address. Some of the further ideas we are intending to look at include: prediction threshold analysis to fine-tune false positive/true positive rates, life stage analysis, analysis of veliger viability, and automating the imaging pipeline process.

Management Implications

Our hope is that this technology becomes more widely accepted in detecting and managing early dreissenid invasions, or detecting when existing infestations are spawning. It would allow processing water samples much more quickly and thoroughly than current practice. Thus, with reduced sample analysis time, additional time and effort could be spent on additional and more frequent water sampling, enhancing the possibility of detecting new dreissenid invasions early, as well as understanding the growth and distribution of known infestations. Using this system in the field, with fresh samples collected by drift net, seems to be the most effective and efficient method. Collection during dam opening/closing seems to be more effective than during dam release based on our study. Many more details on this work, particularly on the machine learning approaches, can be found in the two papers published during this study (Chowdhury and Hamerly 2022; Chowdhury et al. 2023).

References

- Arterburn, Heather Mae. 2020. *Population and Reproductive Dynamics of Zebra Mussels (Dreissena Polymorpha) in Texas Reservoirs: Predicting the Sustainability of Zebra Mussel Populations in Warmer Waters*.
- Bertin, Marvin. 2017. *Training, Evaluating, and Tuning Deep Neural Network Models with TensorFlow-Slim: Advanced Topics in Training, Evaluating, and Tuning Deep Neural Network Models*.
- Brownlee, Jason. 2019. *Deep Learning for Computer Vision: Image Classification, Object Detection, and Face Recognition in Python*. Machine Learning Mastery.
- Chowdhury, S. and Hamerly, G., 2022, October. Recognition of Aquatic Invasive Species Larvae Using Autoencoder-Based Feature Averaging. In the International Symposium on Visual Computing (pp. 145-161). Cham: Springer International Publishing.
- Chowdhury, S., Tisha, S.N., McGarrity, M. and Hamerly, G., 2023. Video-Based Recognition of Aquatic Invasive Species Larvae Using Attention-LSTM Transformer. In the International Symposium on Visual Computing.
- Connelly, Nancy A., Charles R. O'Neill Jr, Barbara A. Knuth, and Tommy L. Brown. 2007. "Economic Impacts of Zebra Mussels on Drinking Water Treatment and Electric Power Generation Facilities." *Environmental Management* 40 (1): 105–12.

- Counihan, Timothy D., and Stephen M. Bollens. 2017. "Early Detection Monitoring for Larval Dreissenid Mussels: How Much Plankton Sampling Is Enough?" *Environmental Monitoring and Assessment* 189 (3): 98.
- Delaney, Mary Anne. 1997. *Potential Impact of Zebra Mussels on Warmwater Aquaculture and Aquaculture's Potential to Disseminate Zebra Mussels Into Public and Private Waters*.
- Dixon, Dewayne A. 2018. *Neural Network and Its Application to an Image Binary Classification Problem*.
- Fischer, Samuel M., Martina Beck, Leif-Matthias Herborg, and Mark A. Lewis. 2021. "Managing Aquatic Invasions: Optimal Locations and Operating Times for Watercraft Inspection Stations." *Journal of Environmental Management* 283 (April): 111923.
- Hsu, Tzu-Chun, Yi-Sheng Liao, and Chun-Rong Huang. 2023. "Video Summarization With Spatiotemporal Vision Transformer." *IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society* 32 (May): 3013–26.
- Johansson, M.L., Lavigne, S.Y., Ramcharan, C.W., Heath, D.D. and MacIsaac, H.J., 2020. Detecting a spreading non-indigenous species using multiple methodologies. *Lake and Reservoir Management*, 36(4), pp.432-443.g
- Kandhare, Pravinkumar G. 2013. *Object Tracking in Video Sequence Using Modified Kalman Filter with a Shrinking Active Contour as a Measuring Tool*.
- Michigan. Zebra Mussel Task Force. 1991. *The Zebra Mussel, (Dreissena Polymorpha): A Strategy to Control Its Spread in Michigan : A Report to the Michigan Legislature*.
- Peñarrubia, Luis, Carles Alcaraz, Abraham Bij de Vaate, Nuria Sanz, Carles Pla, Oriol Vidal, and Jordi Viñas. 2016. "Validated Methodology for Quantifying Infestation Levels of Dreissenid Mussels in Environmental DNA (eDNA) Samples." *Scientific Reports* 6 (December): 39067.
- Peng, Xinggan, Hanhui Li, Yuxuan Lin, Yongming Chen, Peng Fan, and Zhiping Lin. 2023. "TCF-Trans: Temporal Context Fusion Transformer for Anomaly Detection in Time Series." *Sensors* 23 (20). <https://doi.org/10.3390/s23208508>.
- Sabry, Fouad. 2023. *Long Short Term Memory: Fundamentals and Applications for Sequence Prediction*. One Billion Knowledgeable.
- Saho, K., 2017. Kalman filter for moving object tracking: Performance analysis and filter design. *Kalman Filters-Theory for Advanced Applications*, pp.233-252.
- Sarkar, Dipanjan, Raghav Bali, and Tamoghna Ghosh. 2018. *Hands-On Transfer Learning with Python: Implement Advanced Deep Learning and Neural Network Models Using TensorFlow and Keras*. Packt Publishing Ltd.
- Sepulveda, Adam J., Christine E. Dumoulin, Denise L. Blanchette, John McPhedran, Colin Holme, Nathan Whalen, Margaret E. Hunter, et al. 2023. "When Are Environmental DNA Early Detections of Invasive Species Actionable?" *Journal of Environmental Management* 343 (October): 118216.
- USGS. "NAS - Nonindigenous Aquatic Species - Dreissena polymorpha". USGS website, <https://nas.er.usgs.gov/queries/FactSheet.aspx?speciesID=5>. Accessed 13 January 2024.
- Wong, Wai Hing, and Shawn L. Gerstenberger. 2015. *Biology and Management of Invasive Quagga and Zebra Mussels in the Western United States*. CRC Press.
- Yue, Guanghui, Peishan Wei, Tianwei Zhou, Youyi Song, Cheng Zhao, Tianfu Wang, and Baiying Lei. 2023. "Specificity-Aware Federated Learning with Dynamic Feature Fusion Network for Imbalanced Medical Image Classification." *IEEE Journal of Biomedical and Health Informatics* PP (September). <https://doi.org/10.1109/JBHI.2023.3319516>.