Classifying Zooplankton: A Collaborative Deep Learning Approach

STA2453 Final Project

Adele Lauzon

Supported by Dr. Vianey Leos Barajas and Dr. Sofia Ruiz Suarez

1 Introduction

1.1 Importance of Zooplankton

The Great Lakes and their surrounding ecosystem is the third largest regional economy in the world (Ontario Environment and Energy). In Ontario, the lakes are hugely important to the provincial economic landscape. The Great Lakes contribute to 91% of GDP generated in Ontario, with $80 million generated from aquaculture, $5.1 billion from recreational fishing, and $5.3 billion from farm receipts (Great Lakes Commission; Brouwer & Garcia Hernandez, 2020). The lakes are also immensely important to First Nations in the region, who have stewarded the land and water for generations. For the Saugeen Ojibway Nation, lake whitefish hold particular significance, playing a part in the creation story for all Anishinaabe and being an important food source for the community since time immemorial (Bardwell, 2024; Parks Canada). Unfortunately, due to reasons not entirely known, Lake Whitefish populations have significantly declined over the last twenty years (Cunningham & Dunlop, 2023), with some scientists believing the fish could disappear from some lakes entirely within the next five years (Katz, 2025).

In the Great Lakes system, zooplankton are the foundation of the food web, with all fish feeding only on zooplankton at some point during their life cycles (US EPA, 2016). Given the sequential nature of the food chain, understanding downstream effects on larger organisms (like whitefish) begins with evaluating the basal organisms. In freshwater ecosystems, this involves understanding the variety and density of zooplankton populations. Given the economic importance of the Great Lakes ecosystem and the responsibility to protect whitefish, the provincial government requires updated tools to automatically classify and enumerate zooplankton organisms.

1.2 Existing Research

To study zooplankton, water samples are first microscopically analyzed using tools like the FlowCam instrument, which uses flow imaging microscopy to create images of particles floating in a liquid. These images are then manually analyzed by scientists to classify and count species. However, manual classification is both time consuming and requires highly-skilled scientists who are able to identify subtle species differences. As such, development of methods to automatically classify plankton images has become an important area of ecological research (Kerr et al., 2020).

FlowCam produces two data sources for each water sample. First, a "mosaic" is created, which is a gridded image file that contains individual images for all particles detected in a sample. Each of these samples has a corresponding plain-text file that contains the manually assigned class for each particle, as well as FlowCam-generated geometric data related to the particle's image.

Most existing research tends to analyze the plain-text and image data separately. Class predictions for plain-text data can be found using multi-layer perceptrons (MLPs) or logistic regression, whereas image data is often modeled using convolutional neural networks (CNNs). While overall performance for each method independently can be impressive, it has been found that each model tends to underperform on minority classes. Most plankton datasets suffer from class imbalance, meaning some classes are far more prevalent than others. While methods like oversampling and image augmentation can improve minority class prediction, this has remained a problem in the field.

Kerr et al. (2020) propose an ensemble prediction method, drawing inspiration from other scientific fields. Ensemble methods involve multiple predictive models that are combined and the outputs are either averaged or some majority voting system is implemented. The authors propose that training different model architectures on identical training sets allows the ensemble to learn different methods and patterns, ultimately yielding higher accuracy measures. The authors demonstrate that this ensemble approach does in fact yield better results for minority classes compared to each model individually.

2 Methodology

This project aims to replicate the work of Kerr and colleagues to create a collaborative deep learning model to classify zooplankton data gathered from Lake Huron. The data for this project is provided by the Ontario Ministry of Natural Resources.

2.1 Data sources

There are three sources of data available for this project, and the objective was to utilize all available information. First, there is the Master Table, where each row represents a water sample, and contains environmental information common to the entire water sample -- for example, the depth at which the same was taken, surface water temperature, and the geographic coordinates of the trawl. Each water sample has both a corresponding image file and another plain-text file. The image file is a mosaic, or collage, of all individual particles detected in the sample. The plain-text file is a table of class assignments for each particle, as well as geometric information specific to that particle -- such as the diameter of the particle, the intensity, or the length. Some particles are labeled with a known zooplankton species, but some are too small for scientists to classify or bubbles that the machine mistakenly identified as an organism.

2.2 Data preprocessing

Due to computational limitations, a random sample of 100 water samples from Lake Huron was selected from the Master Table. Corresponding image mosaics and plain-text files were selected. Only particles from the 7 "important classes" were retained: Calanoid, Cylopoid, Bosmina, Herpacticoida, Chironomid, Chydoridae, and Daphnia. Unique particle IDs were generated to ensure the same particles existed in the image and text datasets.

The image mosaics were split into images of each individual particle, or vignettes, using the coordinates and dimensions provided in the geometric data. A corresponding table was generated to contain class, particle name, and particle image file location for each vignette. Each image was resized to be 300 by 300 pixels. Just before modelling, image data is converted to a 300 x 300 x 3 matrix, where each pixel is represented by RGB channels. The Huron vignettes are black and white, so the single grayscale pixel value was replicated across all three channels.

The two sources of text data (geometric and environmental) were merged to create one table of all available text information. Each row in this new dataset represented a particle, the geometric information for that particle as well as the environmental information related to that particle's water sample. Text data was also standardized and one-hot encoded before modelling.

This processing resulted in one image dataset of particles (supplemented by the table described above) and one text dataset of geometric and environmental information for each particle. To adjust for class imbalance, underrepresented classes (Chydoridae, Daphnia, Cylopoid) were to be oversampled, with targets of 100 observations. Further, data augmentation was performed on image data, wherein images are stretched, flipped, and shifted to increase data availability for minority classes (Figure 1). These augmented vignettes were only added to the image data, and were clearly labeled as being augmented.

Finally, the data was split into training (60%), validation (20%) and testing sets (20%). Particle IDs were matched to ensure the same particles appeared in each of the datasets.

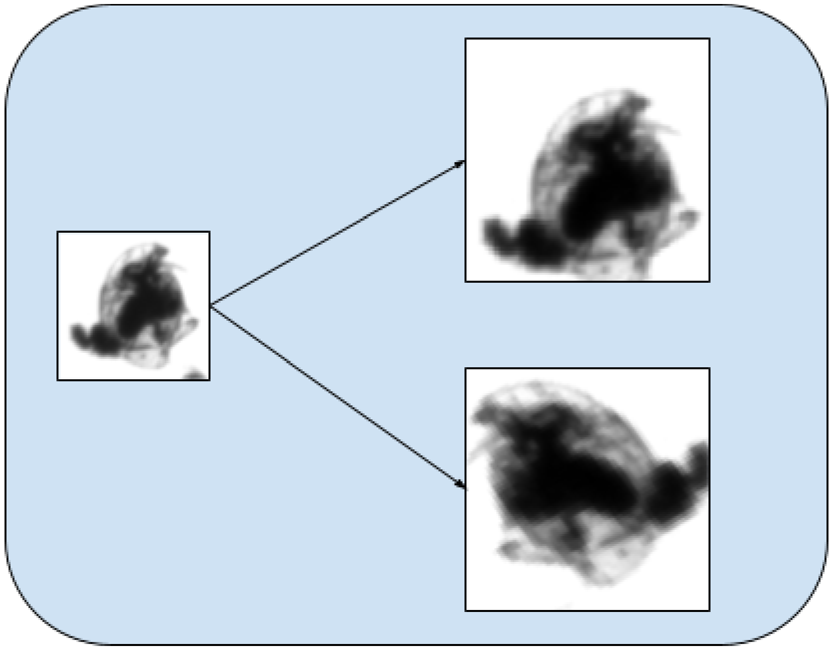

Figure 1: Augmented images for a Chydoridae particle

2.4 Plain-text classification: MLP

Multilayer perceptrons (MLPs) can be thought of as the most foundational form of a neural network. In general, the neural network is designed to resemble a human brain in which groups of neurons are organized hierarchically in layers (Kerr et al., 2020). Data is input to the network (analogous to a "signal" in a brain), and certain neurons are activated that pass on the data to all connected neurons in the following layers.

In an MLP, a one-dimensional array of data (in this case, the merged plain-text data) is passed through the network. In general, a single particle's text data is about 440 times smaller than its vignette (Kerr et al., 2020), resulting in much shorter training times. As such, the MLP was able to be trained from scratch. Due to the rapid training speed of this model, a grid search for model configurations was performed. Following the work of Kerr and colleagues, all combinations of hidden layers (1-5) and neurons per layer (256, 512, 1024, and 2048) were explored. Further, a dropout layer was included to prevent overfitting, which randomly deactivates a ratio of neurons (0.5) from the previous layer for each training step.

2.3 Image classification: CNN

If a neural network is conceptualized as a brain, convolutional neural networks can be thought of as adding a visual cortex to this model (Kerr et al., 2020). Unlike MLPs or more simple neural networks, CNNs involve convolutional filters that slide across an image, learning localized patterns, and creating new feature spaces or maps. These new feature maps are also images, which themselves can be processed by additional convolutional layers, allowing the network to have any number of layers which lead to increasingly abstract representations. As new data is fed through the network, weights and biases of neurons are learned and updated. At a high level, the goal of a CNN is to learn the filter values necessary for each convolutional function (layer) such that the network learns to extract important features that allow accurate class prediction.

Building a convolutional neural network from scratch requires significant computational resources and an abundance of training data. Given the limitations of this project, transfer learning was employed. Transfer learning involves initializing, re-training, and reconfiguring pre-trained models, where you begin with learned weights and biases and tune based on your particular dataset. Replicating the work of Bonin-Font et al. (2024), the EfficientNetv2 B3 model from TensorFlow was used, as the authors found that it performed well on plankton classification problems. Even with transfer learning, computational limitations did not allow for the same sort of grid search performed for the MLP.

2.5 Collaborative Model

To construct the collaborative model, both the CNN and MLP are loaded with trained weights. Weights are frozen when loaded into collaboration, and the output softmax (final) layer for each model is removed. In its place, the output just before the softmax layer for each model is connected to a new concatenated layer. Then, this layer is connected to a new fully connected layer with 512 neurons. This new layer acts as a new function for the collaborative model to learn how the MLP and CNN should each contribute to the final prediction. Finally, a new softmax layer was added, and the same training and validation data is used to learn the weights between the concatenated and fully connected layer .

3 Results

3.1 MLP

The results of grid search indicated that the best performing model had 2 hidden layers and 256 neurons per layer. Trained over 25 epochs, the final MLP had a training accuracy of 95% and test accuracy and F1 score of 97%. Training and validation accuracy across epochs is displayed in Figure 2. Note that the jaggedness in the graph is due to the scale of the y-axis.

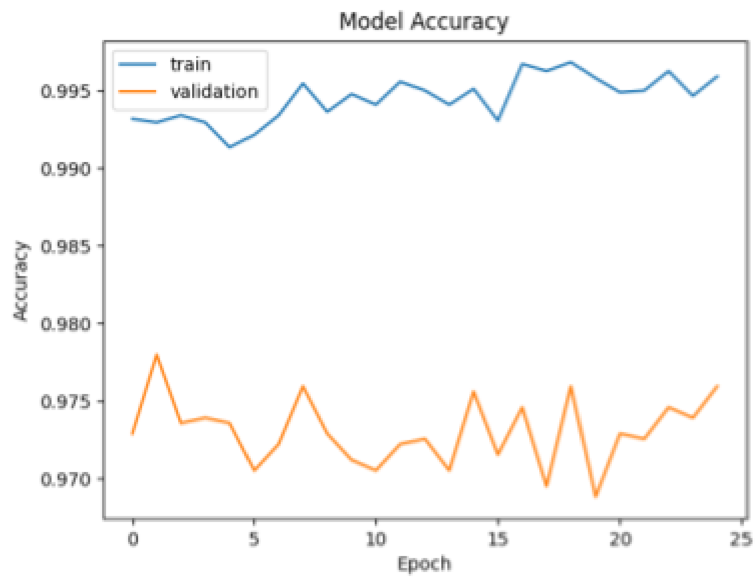

Figure 2: MLP training and validation accuracy across 25 epochs

3.2 CNN

The CNN model reached a training accuracy of 98% and testing accuracy and F1 of 97% as well. Unlike the MLP, the CNN required about 5 epochs to reach an accuracy above 90%. See Figure 3 for a detailed look at the model accuracy across epochs.

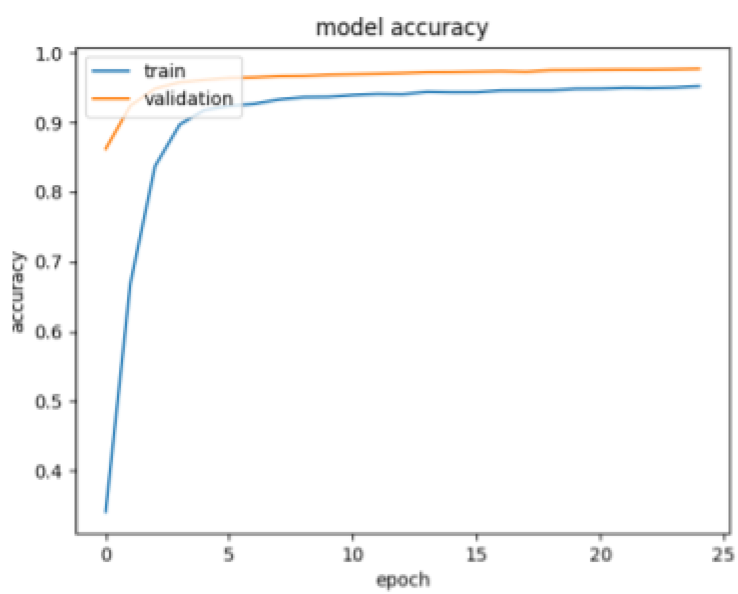

Figure 4: CNN training and validation accuracy across epochs

3.3 Collaborative Model

The final collaborative model reached a training accuracy of 99% and a test accuracy of 98%. As this collaborative approach was chosen in hopes of improved performance on classifying minority classes, the accuracy, precision, recall, and F1 score was disaggregated by class. See Table 1 for these results and Figure 5 for the confusion matrix of classifications by class.

	Precision	Recall	F1 Score	n
Bosmina_1	99%	89%	94%	224
Calanoid_1	99%	100%	99%	2462
Chironomid	99%	97%	98%	121
Chydoridae	62%	100%	77%	20
Daphnia	100%	100%	100%	3
Herpacticoida	96%	76%	85%	92

Table 1: Performance metrics for collaborative model, by class

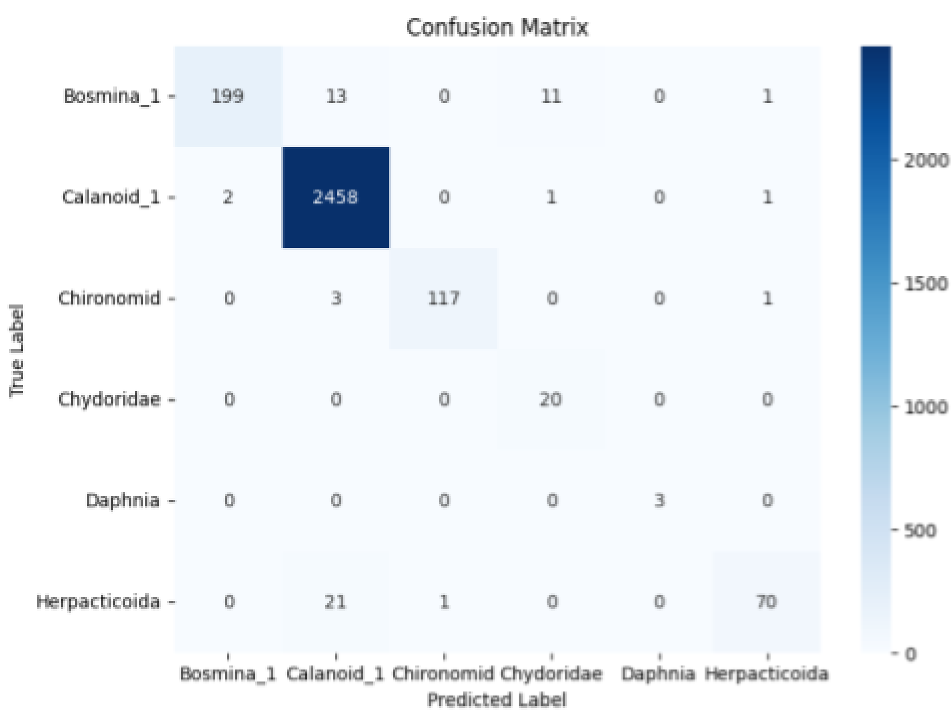

Figure 5: Confusion matrix of test predictions for collaborative model, with off-diagonal elements indicating misclassifications

4 Discussion

Each of the CNN and MLP individually had strong performance, and the collaborative model had even higher test accuracy. While the collaborative model had F1 scores of at least 85% for most classes, the performance for Chydoridae was the worst at 77%. This weaker performance is likely due to the limited particles of these classes that were included in the original subsample. Due to the sampling method chosen (selecting water samples from the master table, and then only pulling images and geometric data associated with these samples) limited the use of more robust sampling methods, such as stratified sampling or resampling (from the entire dataset) of underrepresented classes. Additionally, the random sample did not capture any observations from the class Cylopoid, therefore we are not able to evaluate model performance on this class.

Further limitations include procedural choices that were made without support from the literature. For example, it was not clear in the literature whether image augmentation was performed before or after image resizing. In this project, as part of the augmentation process, augmented images were resized to the dimensions required for the chosen network architecture (EfficientNet). Further, the architecture was chosen based on existing plankton classification literature, but these data differed in that they were greyscale, meaning it did not naturally suit the 3-channel dimensions required by the network. Future work could explore using architectures that have been shown to perform well for black and white images, such as ResNet50 (Hughes, 2022). Finally, there was some data loss, possibly due to the manual download of the geometric data associated with each water sample -- some of the geometric data for the selected subset of water samples were not found on the local machine.

These preliminary results, despite computational limitations that prevented the inclusion of the entire dataset, indicate that a collaborative framework may be a strong choice when it comes to both overall performance and performance on minority classes. If resources permit, future work should explore using the entire dataset and observing performance on underrepresented classes.

Citations

Bardwell, N. (2024). From Waters to Table: The Story of the Great Lakes Whitefish. *Native News Online*, https://nativenewsonline.net/sovereignty/from-waters-to-table-the-story-of-the-great-lakes-whitefish

Bonin-Font, F., Buenvaron, G., Kane, M., & Tuval, I. (2024). Microplankton Discrimination in FlowCAM Images Using Deep Learning: *Proceedings of the 19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, 606–613. https://doi.org/10.5220/0012460200003660

Cunningham, K. E., & Dunlop, E. S. (2023). Declines in lake whitefish larval densities after dreissenid mussel establishment in Lake Huron. *Journal of Great Lakes Research, 49(2),* 491-505. https://doi.org/10.1016/j.jglr.2022.12.015.

Garcia-Hernandez, J. A., & Brouwer, R. (2020). A multiregional input–output optimization model to assess impacts of water supply disruptions under climate change on the Great Lakes economy. *Economic Systems Research*. DOI: 10.1080/09535314.2020.1805414

Hughes, C. (2022). Transfer Learning on Greyscale Images: How to Fine-Tune Pretrained Models on Black-and-White Datasets. *Medium.* https://medium.com/data-science/transfer-learning-on-greyscale-images-how-to-fine-tune-pretrained-models-on-black-and-white-9a5150755c7a

Katz, E. (2025). ‘A crisis’: Lake whitefish survey paints an even more dire picture. *Detroit Public Radio.* https://wdet.org/2025/01/20/a-crisis-lake-whitefish-survey-paints-an-even-more-dire-picture

Kerr, T., Clark, J. R., Fileman, E. S., Widdicombe, C. E., & Pugeault, N. (2020). Collaborative Deep Learning Models to Handle Class Imbalance in FlowCam Plankton Imagery. *IEEE Access*, *8*, 170013–170032. IEEE Access. https://doi.org/10.1109/ACCESS.2020.3022242

Mirasbekov, Y., Zhumakhanova, A., Zhantuyakova, A., Sarkytbayev, K., Malashenkov, D. V., Baishulakova, A., Dashkova, V., Davidson, T. A., Vorobjev, I. A., Jeppesen, E., & Barteneva, N. S. (2021). Semi-automated classification of colonial Microcystis by FlowCAM imaging flow cytometry in mesocosm experiment reveals high heterogeneity during seasonal bloom. *Scientific Reports*, *11*(1), 9377. https://doi.org/10.1038/s41598-021-88661-2

Ontario Environment and Energy. (n.d.). Protecting the Great Lakes. ontario.ca. https://www.ontario.ca/page/protecting-great-lakes

Parks Canada Agency, Government of Canada (2024, October 18). Together with giigoonyag. Fathom Five National Marine Park. https://parks.canada.ca/amnc-nmca/on/fathomfive/nature/twg

US EPA, R. 05. (2016, January 26). *Great Lakes Zooplankton Monitoring* (Great Lakes) [Collections and Lists]. https://www.epa.gov/great-lakes-monitoring/great-lakes-zooplankton-monitoring https://medium.com/data-science/transfer-learning-on-greyscale-images-how-to-fine-tune-pretrained-models-on-black-and-white-9a5150755c7a