Deep Learning Models for Phytoplankton Detection for Water Quality Assessment

Rawit Phanphoowong[1*], Jirawat Amornophatsathein[1], Kanthorn Khaimuk[1]
Tanyaratana Dumkua[1] & Orawan Chaowalit[2]
[1]*Mahidol Wittayanusorn School, Nakhon Pathom, Thailand*
[2]*Department of Computing, Faculty of Science, Silpakorn University, Nakhon Pathom, Thailand*
**Email: rawit.pha_g31@mwit.ac.th*

Abstract

In Thailand, prevalent euglenoid genera like *Euglena, Phacus, Trachelomonas, Lepocinclis,* and *Strombomonas* are typically found in organic-rich water, indicating hypereutrophic and eutrophic conditions. These euglenoids have diverse shapes, making accurate identification challenging. To address this, an object detection application was developed using deep-learning neural network models to reduce identification errors. Photographic datasets collected between June and October 2022, using both microscope and phone cameras in Khlong Mahasawat, Nakhon Pathom, Thailand, covered five Euglenoid genera. These datasets were manually labeled and used to train four deep-learning neural network models: Detectron2, YOLOv5, YOLOv7, and YOLOv8. Precision and recall of the models were improved through image augmentation, mimicking variations in image quality. The model which proved best for phytoplankton identification was YOLOv5l, which yielded precision and recall of 0.839 and 0.873, respectively. This model exhibited high performance in terms of the accuracy, with a low rate of misclassification.

Keywords: Water quality assessment, Euglenoid, Object detection, Deep learning

I. INTRODUCTION

Assessing water quality is important in environmental science research, and is quantified using physical properties such as color and turbidity, chemical properties such as dissolved oxygen (DO), pH, and salinity, and biological properties such as biological oxygen demand (BOD) and biodiversity. In Thailand, rapid urbanization and industrialization have depleted the water quality of the Chao Phraya and Tha Chin rivers, as evidenced through decreased DO levels, elevated ammonia-nitrogen content, and increased BOD (Simachaya, W., 2003). These changes significantly impact water biodiversity. Different water quality impacts the varieties and population density of phytoplankton; thus, the presence of certain species of phytoplankton can be used as an indicator of water quality.

This research utilizes the AARL-PP (Assessment of Water Quality in Standing Water by Using Dominant Phytoplankton) developed by Y. Peerapormpisal (2007) as a method for assessing water quality. AARL-PP offers a chemical-free approach, allowing for long-term water quality assessment. It was developed from a two-part scoring system. First, an index of water quality based on a standard range of water quality tests was defined in six levels, based on research, using a score of 1-10, with lower scores indicating clean water and higher scores indicating polluted water, as shown in Table 1.

Score	Water quality by trophic level	General water quality
1.0 - 2.0	Oligotrophic	Clean
2.1 - 3.5	Oligo-mesotrophic	Clean-moderate
3.6 - 5.5	Mesotrophic	Moderate
5.6 - 7.5	Meso-eutrophic	Moderate-polluted
7.6 - 9.0	Eutrophic	Polluted
9.1 - 10.0	Hypereutrophic	Very polluted

Table 1 Water quality scores defined by trophic level and general water quality

Euglenoid genus	Sources of database	Taxonomic references
Euglena Ehrenberg, 1830	Images of water sample under microscope	M.D. Guiry *in* Guiry, M.D. & Guiry, G.M. (2017)
Phacus Dujardin, 1841		E.A. Molinari Novoa *in* Guiry, M.D. & Guiry, G.M. (2021)
Trachelomonas Ehrenberg, 1834		M.D. Guiry *in* Guiry, M.D. & Guiry, G.M. (2017)
Lepocinclis Perty, 1849	Images of water sample under microscope and from Solito de Solis (2020)	G.M. Guiry *in* Guiry, M.D. & Guiry, G.M. (2022)
Strombomonas Deflandre, 1930		M.D. Guiry *in* Guiry, M.D. & Guiry, G.M. (2018)

Table 2 Sources of database and taxonomic references for each genus of euglenoids

For the second part, the dominant genera of phytoplankton found in the different water quality levels were identified and matched to the rating score. Three to five dominant genera of the phytoplankton were selected in accordance with their respective populations. During the assessment of water quality, phytoplankton was collected, identified and the quantity of each genus was determined. The score of each genus following the water quality in the second part was averaged and compared with the standard score of water quality from the first part. The AARL-PP Score has been tested in 50 water systems in northern Thailand and 20 water systems each in northeastern and southern Thailand, with phytoplankton identification being more than 95% in agreement with physical and chemical water quality.

Five genera of euglenoid were studied: *Euglena, Phacus, Trachelomonas* and *Strombomonas,* with scores in the AARL-PP score, 10, 8, 8, and 8, respectively, and *Lepocinclis*, though it should be noted that its score is yet to be determined within the AARL-PP system. The high scores assigned to these euglenoids suggest their prevalence in hypereutrophic waters, indicative of significant pollution levels.

Euglenoids tend to have a highly variable morphology, which increases the chances of genus misidentification. As each genus of euglenoids represents a different AARL-PP score, genus identification affects the accuracy of score. One proposed solution is to detect the “eye spot”, a plastid-like organelle found in euglenoids. The eye spot is easily identified under a microscope, so accurately identifying the eye spot can be used to aid in the identification of euglenoids. We propose the development of an object detection program, using a deep-learning neural network model, to reduce the identification error and to further develop an accurate AARL-PP score assessment program.

Deep learning models are artificial intelligence algorithms, mimicking human brain neurons, consisting of multiple layers of neural network. The models are trained in object detection and classification using datasets of images. In one training cycle, the deep learning model divides the image into grids, processes each grid with the neuron-like network of the mathematical model, and the result is shown. During each round of training, called an epoch, some parameters in the neural network change and the accuracy of the model improves. What makes each neural network model different is its architecture and size of the model. Models can be evaluated and compared with evaluation metrics such as precision, recall, and F1-score (IBM, 2024)

To create an object detection program, the dataset of images is divided into three parts: a training set, a validation set, and a test set, at an appropriate ratio. In the training set images, the objects of interest are labeled to locate and specify types of objects for the model to learn. The program is trained with the labeled training images, then validated and tested with unlabeled images.

II. METHODS

Data Collection and Image Preparation

Water samples were collected using a 21-micron plankton net at coordinates 13°48'25"N, 100°17'3"E in Khlong Mahasawat canal, Phutthamonthon district, Nakhon Pathom, Thailand, from June to October 2022. Images of euglenoids were captured

under a bright field microscope at 400x magnification, showcasing distinct morphologies of *Euglena* spp., *Phacus* spp., *Trachelomonas* spp., *Lepocinclis* spp., and *Strombomonas* spp. The classification of each genus was based on taxonomic references, including Wongrat *et al.* (2017) and AlgeaBase, a comprehensive global database curated by M.D. & G.M. Guiry, as shown in Table 2. Images from videos by Solito de Solis (2020) were also used. Selected images were required to meet specific criteria: (1) inclusion of at least one genus of euglenoids, (2) presence of other organisms (e.g. different types of algae) or contaminants (e.g. sediment), (3) variation in color, brightness, and aperture, and (4) image resolution was not controlled.

Images were labeled with the respective genera and the position of each euglenoid within the image. Two labeling methods were employed: bounding box and instance segmentation, according to the requirements of the object detection models.

The image dataset was then split into three sets: a 70% training set, a 20% validation set, and a 10% testing set. In total, 284 images of *Euglena* spp., 180 images of *Phacus* spp., 115 images of *Trachelomonas* spp., 187 images of *Lepocinclis* spp., and 102 images of *Strombomonas* spp. were obtained after completing the data collection process.

After completing the image labeling process, the next step involves augmenting the image dataset to simulate factors affecting genus identification, such as brightness and color of light from the microscope, focus and sharpness, position of euglenoid features, and obstructions within the image. Image augmentation techniques included cropping (with adjustments ranging from 0% to 30%), altering hue (ranging from -25% to +25%), applying cutout to portions of the image (with a cutout rate of 20%), inducing blur (ranging from 0 to 2 pixels), and adjusting brightness (ranging from -25% to +25%), as shown in Figure 1. Upon completion of image augmentation, the dataset consisted of a total of 2,055 images.

Model Training

Deep learning network models were developed based on four different neural network models: Detectron2 , YOLOv5 , YOLOv7 , and YOLOv8 , with five sub-models of YOLOv5 tested, including YOLOv5s, YOLOv5n, YOLOv5m, YOLOv5l, and YOLOv5x. The fundamental principle of these deep learning network models is to divide the images in the dataset into multiple cells, process each cell using a mathematical model, display the results, and then iteratively adjust the parameters in the model to improve the results.

The YOLO models are designed with different sizes appropriate for varying dataset sizes, with the size of the model indicated by the letter at the end of the name, such as 's' representing small size (Jocher, 2022) . Currently, development has progressed to version 8, known as YOLOv8 . YOLOv5 , the fifth version, remains the most popular. The Detectron model is a methodical model for high-speed object detection and flexible use (Wu, Y., 2019).

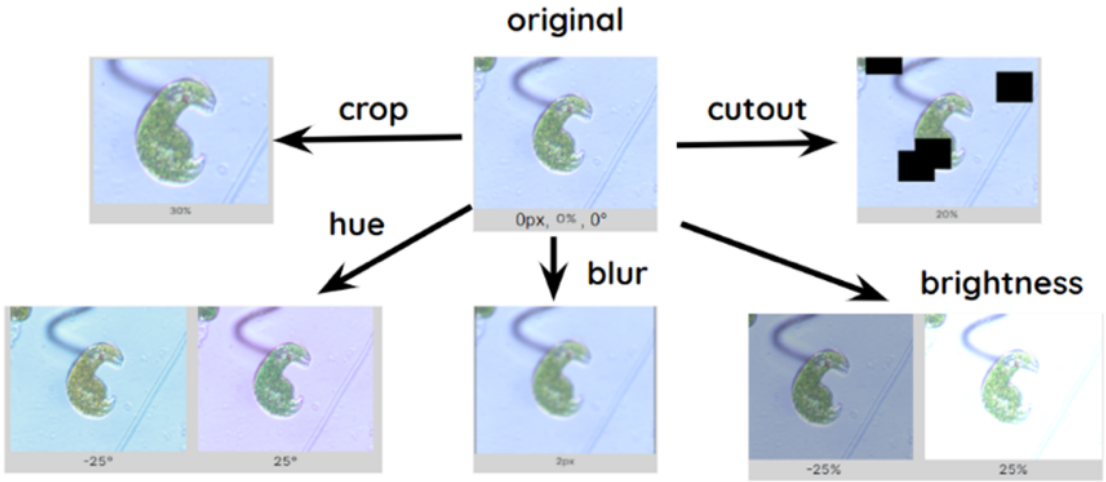

Figure 1 Datasets after image augmentation

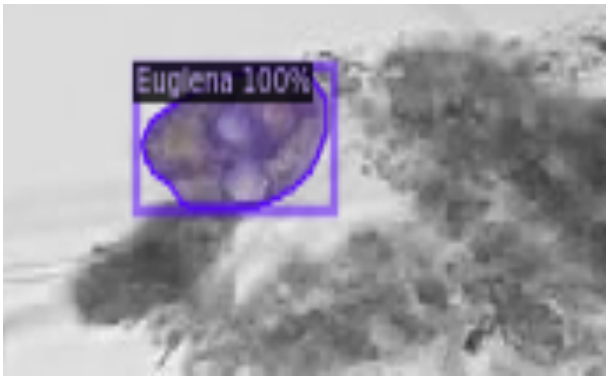

Figure 2 Example of the results from Detectron2

The model training involved 300 epochs, after which the following evaluation metrics were examined: classification accuracy, precision, recall/sensitivity, and F1-score. The F1-score is often used for comparing deep learning models, where models with higher F1-scores have better classification abilities. Each metric is calculated with a formula based on the amount of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).

The 'accuracy' in classification is calculated based on the proportion of (TP + TN) to (TP + TN + FP + FN), while 'precision' is calculated based on the proportion of (TP) to (TP + FP), recall is calculated based on the proportion of (TP) to (TP + FN), and the F1-score is the harmonic mean of precision and recall. Additionally, the mean Average Precision at k (mAP@k) is evaluated as an appropriate evaluation metric for ranking and sorting tasks, particularly suitable for assessing the quality of water using the AARL-PP score in object detection. Commonly used values for k are from 50 to 95.

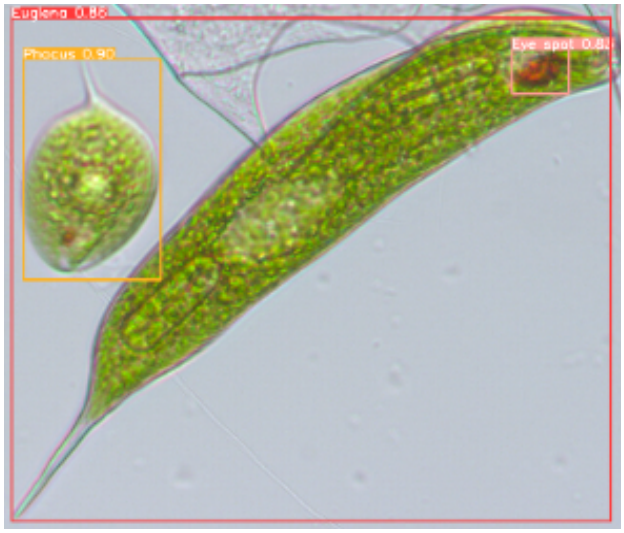

Figure 3 Example of the results from YOLOv5

III. RESULTS AND DISCUSSION

After building the model and conducting image input experiments, it was found that the model could detect images, as shown in Figures 2 and 3. The detected genus of euglenoids is identified along with the confidence level. For example, in Figure 2, *Euglena* sp. Was detected with 100% confidence, and in Figure 3, *Euglena* sp., *Phacus* sp., and an eye spot were detected with confidence levels of 88%, 90%, and 83%, respectively.

After training the models with the obtained images using all 8 deep learning models, the results are presented in Table 3. The best-performing model was found to be YOLOv5l, as it achieved the highest

Model	Precision	Recall	F1-score	mAP@50	mAP@50-95
Detectron2	0.833	0.847	0.840	0.915	0.731
YOLOv5					
YOLOv5s	0.871	0.804	0.836	0.844	0.527
YOLOv5n	0.877	0.750	0.809	0.831	0.532
YOLOv5m	0.841	0.858	0.849	0.868	0.687
YOLOv5l	0.839	0.873	0.856	0.894	0.715
YOLOv5x	0.833	0.847	0.840	0.915	0.731
YOLOv7	0.773	0.789	0.781	0.829	0.584
YOLOv8	0.854	0.833	0.803	0.866	0.567

Figure 3 Results from the training for all the models tested.

Genus	Precision	Recall	F1-score	mAP@50	mAP@50-95
Euglena	0.772	0.881	0.822	0.900	0.763
Phacus	0.972	0.863	0.914	0.938	0.834
Trachelomonas	0.732	0.952	0.827	0.921	0.841
Lepocinclis	0.888	0.855	0.871	0.933	0.804
Strombomonas	0.904	0.947	0.925	0.964	0.757

Table 4 Evaluation metrics of YOLOv5l for each genus of euglenoids

F1-score. The F1 -score represents the harmonic mean of precision and recall, which are two evaluation metrics for the learning performance of efficient models. Considering the harmonic mean of both variables provides a balanced assessment without bias towards any specific variable, ensuring an unbiased evaluation of the model's performance.

Based on the evaluation metrics of the YOLOv5l model, it was able to classify each genus as shown in Table 4. During the training process, there were trends in the changes of evaluation metrics for each epoch, as depicted in Figure 4.

From Figure 4, it can be observed that the values of box loss, objectness loss, and classification loss decrease in each training epoch for both the training dataset (as shown in Figures 4A, 4B, and 4C) and the validation dataset (as shown in Figures 4F, 4G, and 4H). The box loss indicates the model's ability to identify the center position of detected objects and predict bounding boxes around objects. Objectness represents the likelihood of objects being present in the area of interest in the image, while classification loss demonstrates the model's ability to predict object classifications.

Figures 4D, 4E, 4I, and 4J show evaluation metrics such as precision, recall, mAP@50, and mAP@50-95, respectively. The increase in values for these metrics across all four figures indicates an overall improvement in the model's performance with an increasing number of training epochs.

Figure 5 depicts the graph of the relationship between precision and recall. The area under the

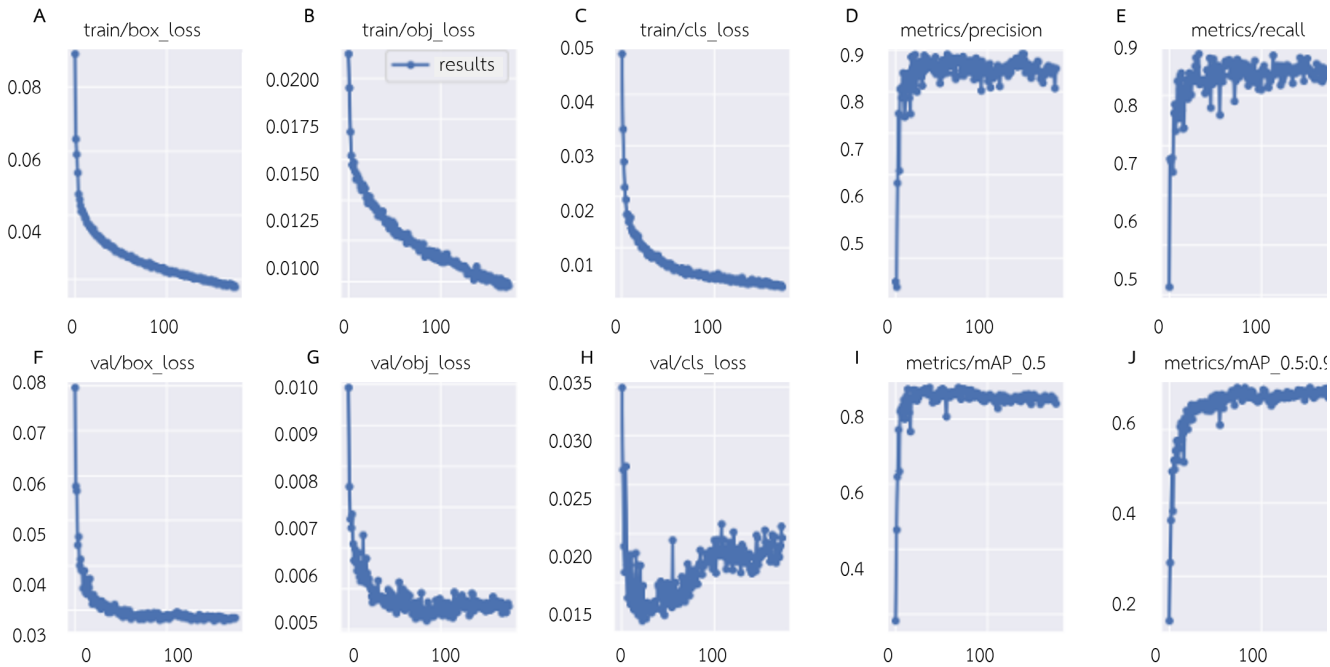

Figure 4 Evaluation metrics graph of YOLOv5l

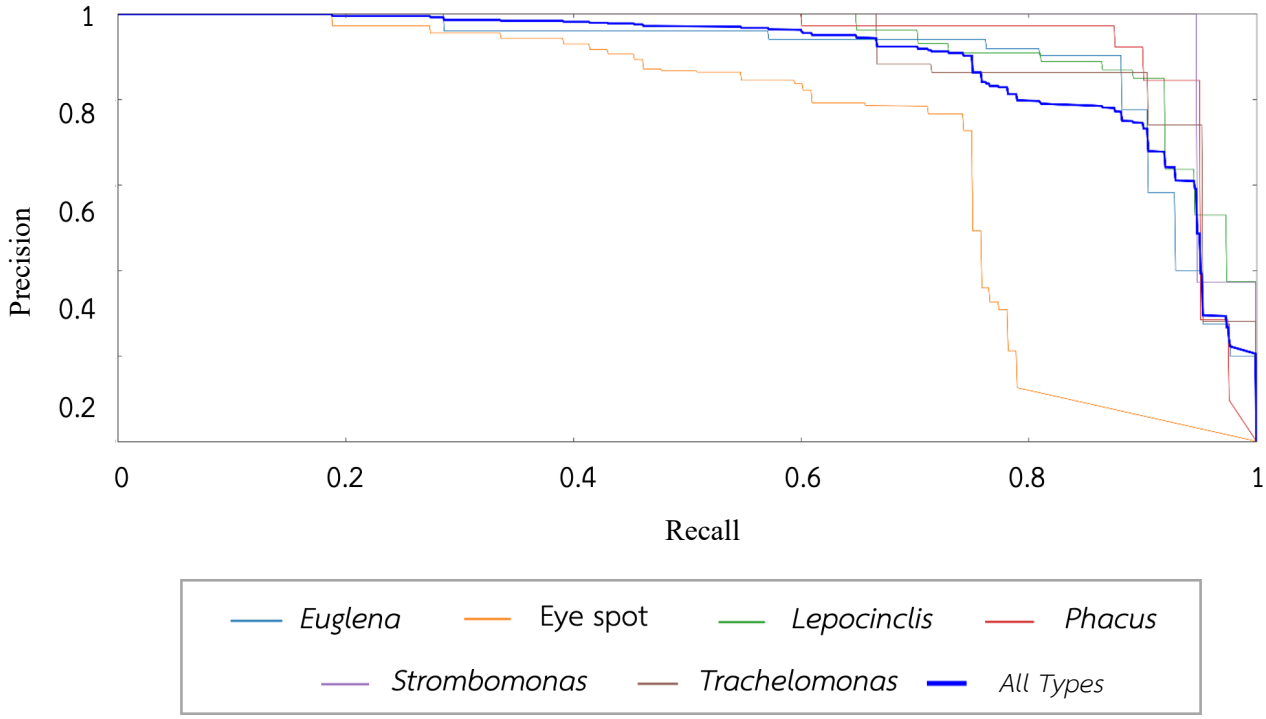

Figure 5 The Precision-Recall curve for the YOLOv5l model for *Euglena*, Eye spot, *Lepocinclis, Phacus, Strombomonas, Trachelomonas*, and *All Types*.

graph indicates that the model exhibits both high precision and high recall, meaning it has a low rate of false positives and false negatives. The interpretation of the area under the graph is similar to that of the F1-score mentioned earlier, as it is a single variable that can be used to compare both accuracy and recall without bias or distortion in assessing the model's performance.

From Figure 5, it can be observed that the model can detect eye spots with low efficiency, while the groups of euglenoids have higher efficiency. Therefore, it is not feasible to rely on an eye spot for identifying the type of euglenoid, as the model's ability to detect euglenoids is better than that of eye spot.

In the confusion matrix shown in Figure 6, it is found that the model correctly identified the species *Lepocinclis* and *Phacus* with accuracies of 84% and 82%, respectively. The most common misclassification was observed for *Lepocinclis*, which was misclassified as *Euglena*, *Phacus*, and *Trachelomonas*, while *Phacus* was misclassified as *Euglena* and *Trachelomonas*, as well as being labeled as the background in the image. The model correctly identified *Euglena* with an accuracy of 88% but misclassified it as *Lepocinclis* and *Trachelomonas*. Conversely, there were two instances of misclassification: one for *Strombomonas*, which was correctly identified 95% of the time but misclassified as *Trachelomonas*, and one for *Trachelomonas*, which was correctly identified 90% of the time but misclassified as *Euglena*.

Furthermore, the confusion matrix indicates that the model has difficulty distinguishing between eye spots and backgrounds (to be more specific, when there is no object in the image). Backgrounds are often misclassified as eye spots up to 75% of the time, suggesting that the model still struggles to differentiate between eye spots and backgrounds.

It can be observed that each genus had misclassifications, mainly because some species exhibit similar morphological characteristics. For instance, *Phacus* sp. resembles a green leaf-like structure, and sometimes, it may lack color due to not being exposed to sunlight. This led the model to misclassify some of them as background rather than detecting them as part of the sample. Additionally, eye spots exhibit various colors, such as dark green, red, transparent red, and white, all of which are

colors found in *Euglena* spp. Consequently, the model struggled to differentiate between the colors of *Euglena* spp. and the colors of the eye spots, resulting in low accuracy in eye spot identification. Furthermore, some genera of euglenoids share similar morphological features from certain angles, making it challenging to distinguish between them, such as *Phacus* sp. and *Trachelomonas* sp.

With current information, the model could infer that the water source under study is hypereutrophic. However, further work must be done to develop a water quality assessment program capable of calculating AARL-PP scores because the assessment requires identifying and counting all phytoplankton genera in Peerapormpisal (2007), which exceeds the number of genera considered in this study. Therefore, to develop an accurate AARL-PP score calculation program, it is necessary to create a model that can identify all phytoplankton genera in Peerapormpisal (2007) and validate the program's scores against other water quality indicators such as DO and BOD values over various time periods and sampling locations.

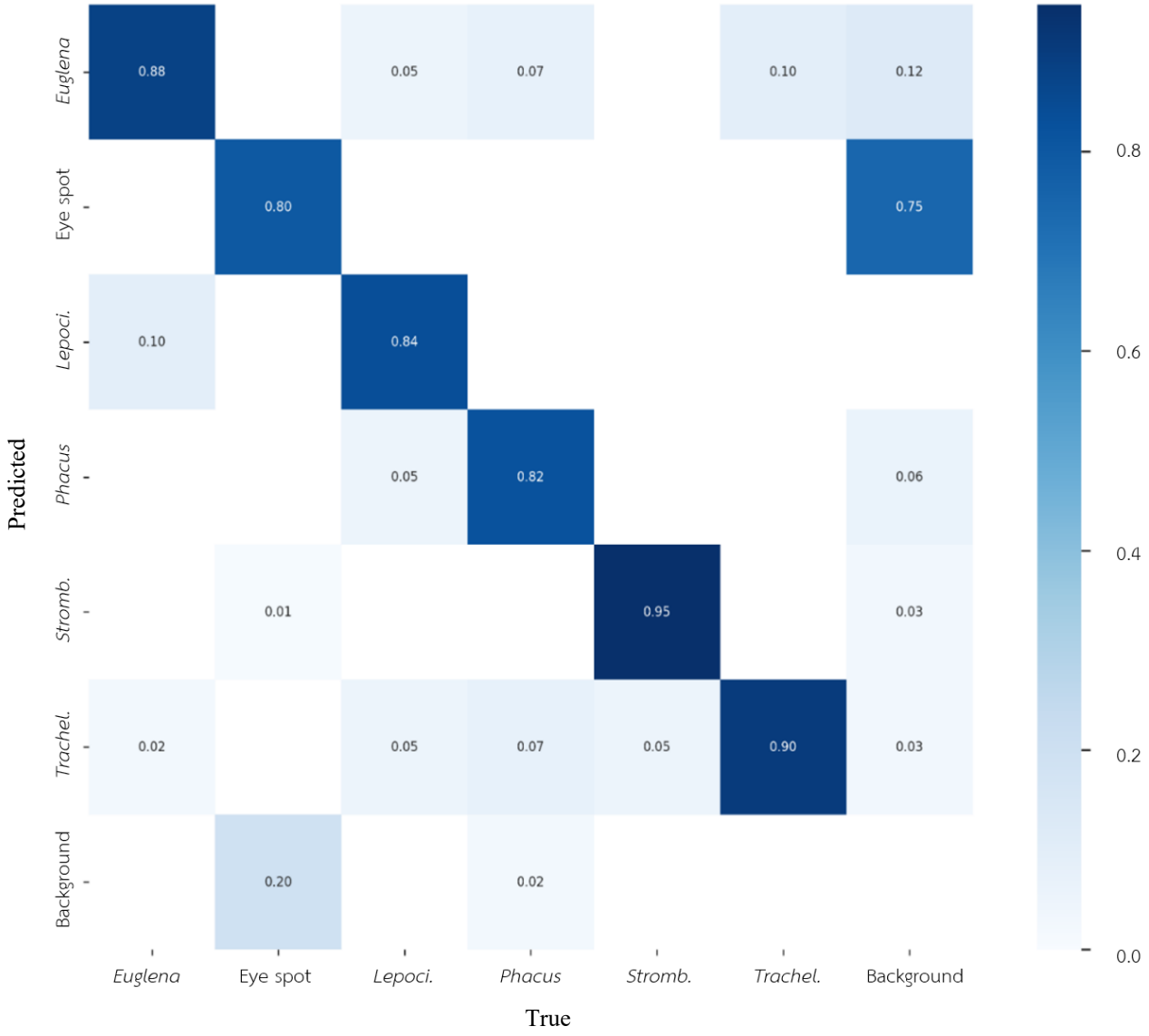

Figure 6 Confusion matrix of YOLOv5l, showing the truth and the predictions, dark blue represents a proportion of 1.0, while white represents 0.0. The intensity of the color corresponds to the proportion of data in each cell

IV. CONCLUSION

In the water quality assessment, the identification of various phytoplankton genera is crucial. However, the freeform morphology within euglenoids makes species identification challenging, leading to high error rates. To mitigate this issue, a deep learning model was developed to reduce errors in species identification and further extended into a water quality assessment program, the AARL-PP scoring system. The YOLOv5l model achieved an F1-score of 0.856, the highest among the eight tested models. It demonstrated efficient detection of *Euglena, Phacus, Trachelomonas, Lepocinclis,* and *Strombomonas* species with similar performance levels, while unable to identify eye spots. This model exhibited high accuracy and low misclassification rates. In the future, it is possible to enhance the program by developing a model capable of identifying all phytoplankton genera in Peerapormpisal, Y. *et al.* (2007). Subsequently, the program's scores can be validated against other water quality indicators.

V. REFERENCES

Guiry, M.D., *in* Guiry, M.D. & Guiry, G.M. (2017, December 30). *AlgaeBase.* World-wide electronic publication, National University of Ireland, Galway. https://www.algaebase.org

Guiry, M.D., *in* Guiry, M.D. & Guiry, G.M. (2018, January 1). *AlgaeBase.* World-wide electronic publication, National University of Ireland, Galway. https://www.algaebase.org

Guiry, G.M., *in* Guiry, M.D. & Guiry, G.M. (04 March 2022). *AlgaeBase*. World-wide electronic publication, National University of Ireland, Galway. https://www.algaebase.org

IBM. (2024). *What is deep learning?* https://www.ibm.com/topics/deep-learning

Jocher, G., Chaurasia, A., Stoken, A., Borovec, J. & Kwon, Y. (2022). ultralytics/yolov5: v7.0 - YOLOv5 SOTA Realtime Instance Segmentation (v7.0). *Zenodo.* https://doi.org/10.5281/zenodo.7347926

Molinari Novoa, E.A., *in* Guiry, M.D. & Guiry, G.M. (10 May 2021). *AlgaeBase*. World-wide electronic publication, National University of Ireland, Galway. https://www.algaebase.org

Peerapormpisal, Y., Pekkoh, J., Powangprasit, D., Tonkhamdee, T., Hongsirichat, A. & Kunpradid, T. (2007). Assessment of Water Quality in Standing Water by Using Dominant Phytoplankton (AARL- PP Score). J*ournal of Fisheries Technology Research, 1*(1): 71-81.

Simachaya, W. (2003). A decade of water quality monitoring in Thailand's four major rivers: the results and the implications for management. *6th International Conference on the Environmental Management of Enclosed Coastal Seas*, November 18-21, 2003, Bangkok, Thailand.

Solito de Solis. (2020, August 31). *Strombomonas gibberosa* [Video]. Youtube. https://youtu.be/0kJilYOjqF0

Solito de Solis. (2020, September 2). *Strombomonas eurystoma* [Video]. Youtube. https://youtu.be/lrmUDwzsef0

Solito de Solis. (2020, September 26). *Lepocinclis spirogyroides* [Video]. Youtube. https://youtu.be/CsxC2KcrnZs

Solito de Solis. (2022, March 27). *Lepocinclis fusca* [Video]. Youtube. https://youtu.be/87Tlb1FeQ-8

Wongrat, L., Wongrat, P., & Ruangsomboon, S. (2017). Freshwater Phytoplankton in Thailand I: Euglenophyceae. *Kasetsart University Fishery Research Bulletin, 27*: 1–34.

Wu, Y., Kirillov, A., Massa, F., Lo, W. & Girshick, R. (2019). *Detectron2.* Github. https://github.com/facebookresearch/detectron2

Acknowledgments

This research has been successfully completed thanks to the kindness and valuable assistance from Tanyaratana Dumkua and Associate Professor Dr. Orawan Chaowalit. They generously devoted their time to provide guidance throughout the research process, including reviewing and correcting deficiencies from the beginning until completion. Special thanks to the faculty members in the Department of Biology and Health sciences for their advice and support. Gratitude also goes to Dr. Jiroat Sangrattanaprasert for his assistance in equipment usage. Lastly, sincere appreciation to Mahidol Wittayanusorn School for providing material and equipment support as well as facilitating the dissemination of research findings. Ultimately, thanks to the Institute for the Promotion of Teaching Science and Technology (IPST) for organizing programs to develop environmental innovations, providing guidance, and offering financial support for future research endeavors.