AI Algae Classification using Neural Networks

Avishmita Hemesh Sarabesh Yashaswini
Indiana University - Bloomington
{avmandal,heraaj,sneelam,ysampath}@iu.edu

Abstract

*The **identification** and monitoring of algal populations are crucial for assessing water quality, understanding ecosystem dynamics, and mitigating risks associated with harmful algal blooms (HABs). This project proposes the use of cutting-edge artificial intelligence (AI) image recognition algorithms to revolutionize the process of identifying algae and monitoring water quality.*

*We propose a novel approach to address the challenges of training deeper **neural networks** in the context of algae identification and monitoring. Drawing inspiration from the residual learning framework and Transformers, we aim to ease the training process of networks substantially deeper than those used previously. Through comprehensive empirical evaluations of datasets obtained from FlowCam imagery and human classifications, our goal is to demonstrate the effectiveness of our approach in developing a robust neural network model capable of accurately identifying and categorizing phytoplankton species present in water samples. To achieve this, we employed a range of models including **ViT, CNN, AlexNet, ResNet**, and **FNN**, with **ViT** emerging as the most successful model, achieving remarkable test accuracies of 98.45% and 97.27% for the 5-Classes and 10-Classes classification tasks, respectively. These results underscore the efficacy of our approach in enhancing the accuracy of algae classification, facilitating proactive management of water quality, and safeguarding public health.*

1. Introduction

Rapid urbanization, industrialization, and agricultural activities have significantly impacted the quality of freshwater ecosystems worldwide, leading to the proliferation of harmful algal blooms (HABs). These blooms pose serious threats to aquatic life, human health, and ecosystem balance, highlighting the urgent need for robust monitoring and management strategies. Traditional methods of algal identification and monitoring are often labour-intensive, time-consuming, and prone to human error.

Our project addresses this pressing need by proposing a multifaceted approach that integrates state-of-the-art computer vision techniques. Our primary aim is to improve the accuracy and efficiency of identifying and quantifying algae species, facilitating researchers' ability to monitor changes in water quality over time and detect harmful algal blooms (HABs) promptly. Through the use of powered by AI image identification instead of traditional statistical analysis, we aim to greatly improve the accuracy of algal species classification from FlowCam images.

1.1 Dataset

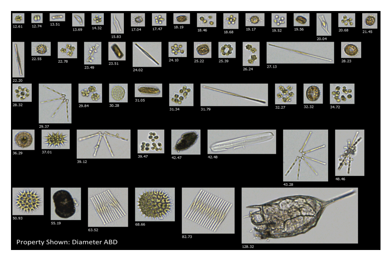

Fig 1. Algal microscopic images from FlowCam

Our dataset comes from an extensive database dump produced by FlowCam, a tool designed to take high-resolution images of microscopic algae. These photos have been meticulously categorized into discrete groupings, creating an extensive database of years of data-gathering efforts by the City of Bloomington researchers (Fig 1).

Through this collaboration, we have been able to gather a wide range of images of algae, which has given us important information on the many species found in the area's water bodies and how they have distributed over time. Through the collaborative efforts of researchers and city authorities, our dataset emerges as a valuable resource for studying algae populations, tracking changes in water quality, and identifying potential environmental risks such as harmful algal blooms (HABs).

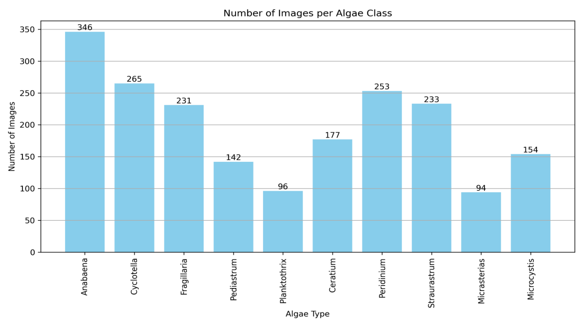

Fig 2: Dataset distribution across different algae classes

The dataset's distribution among different types of algae is shown in Fig. 2. The visual representation of the proportional distribution of various algal kinds within the dataset is facilitated by the classes, each of which represents a unique category of algae species.

2. Background and Related Work

At the City of Bloomington, the current approach to algae species classification relies heavily on statistical methods. Unfortunately, this method often falls short in terms of accuracy, leading to a need for additional manual effort to ensure proper classification.

In response to these challenges, recent advancements in Deep Learning have provided promising alternatives. Researchers have explored various deep-learning models [1] to distinguish between different types of microalgae. Impressively, these models have achieved accuracies of up to 95.27%. Such results highlight the potential of Deep Learning techniques to significantly improve the accuracy and efficiency of algae species classification.

Moreover, to enhance the accuracy of algae detection even further, researchers have integrated various techniques into their models [7]. These include adaptive thresholding, Sobel edge detection, and Canny edge detection. By incorporating these methods, researchers aim to refine the model's ability to accurately identify and classify algae species, thereby reducing the need for manual intervention.

Additionally, Convolutional Neural Networks (CNNs) have emerged as a popular choice for algae classification tasks.[3] The inherent capabilities of CNNs to capture complex patterns and features make them particularly well-suited for achieving high accuracy in this domain. As a result, many researchers have advocated for the use of CNNs as a means to further enhance the accuracy of algae classification models.

3. Method and Implementation

There is a total of 3059 images, with 147 classes, where each class has images ranging from 252 images to 5 images per class. This huge variance in several images is problematic for training neural networks, as the model will fail to generalize. So to tackle this problem, the solution is to do Data Augmentation using methods like transformations of images or GAN. So top classes are chosen based on the counts, and all other classes' images are put into a class called "other", and similarly for top-10 class are chosen for the second set of experimentation along with the "other" class.

Top - 5 included classes such as Peridinium, Straustrum, Ceratrium, Cyclotella, Microcystis and other classes. The top - 10 included all classes from the Top and also an additional 5 classes namely Pedastrium, Anabaena-coiled, Fragillaria, Anabaena, and Planktothrix

When using GAN it mixes two or three classes of distinct features, which is not the correct approach, and also produces low-resolution images. So we choose the other approach which is data augmentation to increase the number of images and give all classes the same ratio of images, so the model will generalize better.

3.1. Edge Detection

The augmentation and preprocessing steps remain consistent for both the top-5 and top-10 classes, ensuring comparability across experiments. This uniformity allows for a meaningful comparison of results.

Given the significant role edges play in algae identification, a critical preprocessing step involves edge detection using Sobel's edge detection technique. This process enhances the distinction between edges and the background, facilitating easier identification of key features such as shape, colour, and size. By amplifying the visibility of edges, the aim is to improve the model's ability to accurately discern between different types of algae. These steps lay the foundation for subsequent data augmentation and preprocessing procedures.

3.2. Data Augmentation and Preprocessing

Once the images are split into the train, validation, and test split using an 80:10:10 ratio of total images. Each train, test, and validation folder, have a subfolder named after the class.

To increase the training size data augmentation techniques such as rotation, width shift, height shift, horizontal flip, vertical flip, and sheer range. Each subfolder goes through an augmentation function and applies one of the augmentations mentioned above randomly, and this loop continues until a set number of images are reached in each folder. To keep experimentation comparable, each folder had 300 images for both top-5 and top-10 classification.

Also since there is an excess number of “other” class images, which might negatively affect the model’s ability to generalize, only 300 images from this sub-folder are chosen. Now all the folders have the same number of images and ratio of images.

3.3. Training & Fine-tuning

Several key hyperparameters were carefully fine-tuned during our model training process to optimize the classifier performance. First, we carefully selected the regularization factor, denoted as 'l2_lambda', which is critical in controlling overfitting by penalizing the model's weights. We chose a low value of 0.001 to strike an appropriate balance between preventing overfitting and ensuring that the model could capture key patterns in the data without undue constraint.

Second, the initial learning rate ('initial_learning_rate') was a critical determinant of the optimization process's effectiveness. Setting it to 0.001 allowed for a gradual initiation of parameter updates, allowing for effective solution space exploration without the risk of overshooting optimal parameter values too early during training.

Apart from the above two we have fine-tuned a few parameters like the ExponentialDecay function to create a dynamic learning rate schedule, with both the decay steps ('decay_steps') and the decay rate ('decay_rate') carefully selected to ensure that the learning rate gradually decreased over time, adapting to the changing dynamics of the optimization process and promoting stable convergence. We used Adam Optimizer to refine the model's parameters during the training. This helped in optimizing our classification task.

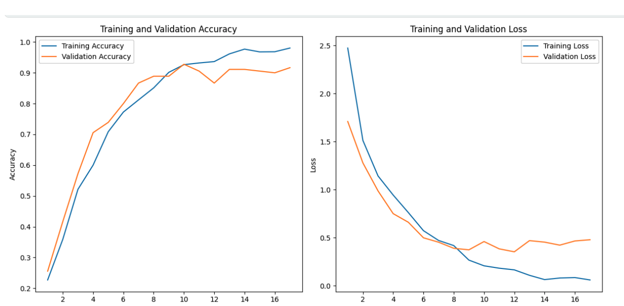

Fig 3.1: Training and Validation Accuracy and Loss for AlexNet for Top 5 classes

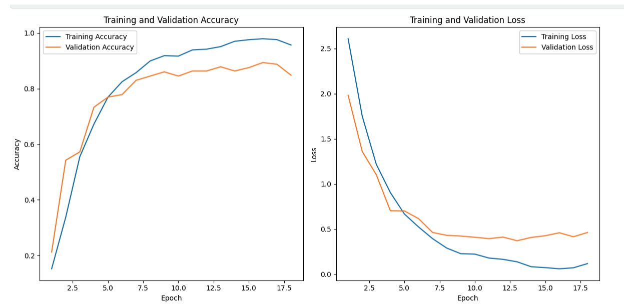

Fig 3.2: Training and Validation Accuracy and Loss for AlexNet for Top 10 classes

Fig 3.1 and 3.2 show the training and validation accuracy of the AlexNet model for the top 5 and top 10 images demonstrating its efficacy in learning and generalizing complex patterns.

3.4. Model Architecture

In our model architecture as shown in Fig 4, we developed a systematic approach to dealing with a large corpus of images obtained from a SQL dump. We use SQL queries to extract images class-wise, ensuring a structured data retrieval process. These images are then subjected to a rigorous data augmentation pipeline, as described in Section 3.2 of our research. We standardize image size and appearance across classes by using techniques such as edge detection, global transformations, and rotation. This preprocessing step is critical for ensuring dataset uniformity, which improves the robustness and generalizability of our model.

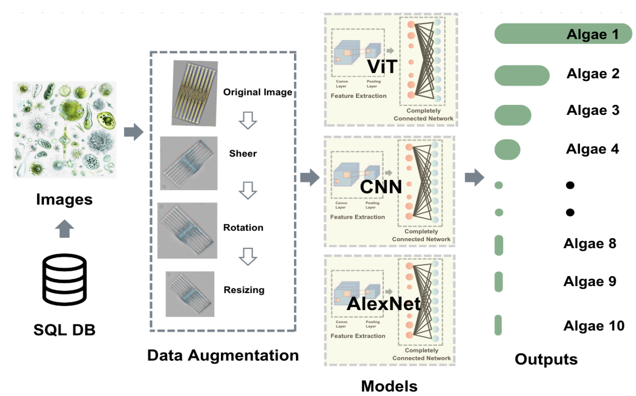

Fig 4: Model Workflow

After data augmentation, we look into a wide range of neural network architectures, including AlexNet, ResNet, AlexNet and various CNN variants. Each architecture is extensively tested, fine-tuned, and validated to ensure that it is effective in handling the complexities of our classification task. Section 3.3 goes into detail about the hyperparameters that were fine-tuned for each model, encapsulating our meticulous approach to optimizing performance.

Finally, our models generate softmax classification scores for the processed images, resulting in probabilistic predictions for each class. This classification framework not only facilitates accurate categorization, but it also provides information about the relative confidence levels associated with each prediction.

4. Experiments

After the images had been preprocessed, various models were tried for both the top -5 and top-10 classes, the intuition and results and their interpretation would be discussed under each model.

All models used a learning rate scheduler early stopping, l2 regularization, and decay rate, to make the model more generalizing and get the best weights of the model.

a) FNN(Baseline model):

There are other statistical numerical values accompanying each image, which give more details about the images. So simple FNN was used on these statistical values to predict the class, which may be combined with CNN models but since we have used ensembling to increase the number of images, ensembling wouldn't be possible, so this model was left out to be a standalone.

b) CNN:

Initially Simple CNN has 2 convolution layers each followed by a max-pooling layer, and 2 two dense layers. The goal is to see how much a simple model would perform, and how well, the model can extract the details from the images.

The model consists of two convolution layers, followed by max-pooling; this layer helps to extract hierarchical features from the images. Both CNNs use RELU activation and L2 regularization to alleviate the effects of overfitting. Once these features are extracted, they are flattened and passed through a 256-neuron fully connected layer that also uses L2 regularization, followed by a layer with a respective number of output neurons—either 6 or 11—using softmax.

c) AlexNet:

Next AlexNet model was used, the reasoning being, that we tried to build a more complex model of our simple CNN model, such as normalization, dropout, and deeper network, and settled for Alexnet as it was a standard model after a couple of experiments of building a custom model from scratch.

AlexNet has 5 CNN layers, where for the first two CNN layers max-pooling is followed after it. Then it has three fully connected layers followed by a final layer that uses softmax which outputs probabilities of various classes (it could be 6 or 11 according to the experiment). In this architecture, the ReLU activation function is used to bring non-linearity into the picture and also uses dropout to mitigate the effects of overfitting in deeper models.

d) ResNet:

Alexnet was slightly overfitting, so we tried to build a model using a skip connection, and as the size of the model increased, started using blocks to build the model. Then decided to use a standardized model, resnet50 which was trained from scratch. Since the images are low resolution, ResNet, which utilizes a deep residual framework, is suitable as it can effectively learn from less detailed content.

ResNet-50 is a type of residual network that was used to alleviate the vanishing gradient problem in deeper models. As the name suggests, it has a total of 50 layers and uses residual connections/skip connections, which allow the gradient to take shortcuts through some layers. This helps in reducing the value of the gradient descent.

The architecture of ResNet includes a CNN layer followed by a block called the bottleneck block, which has three layers of 1x1, 3x3, and 1x1 convolutions. Skip connections serve as highways, connecting some layers and allowing the gradient to pass through. This block is repeated several times to learn about the residual function. Finally, it has an average pooling of all these layers, followed by a fully connected layer with softmax activation, with the number of classes depending on the task (6 or 11).

e) ViT

Due to its superior global understanding of images, the Vision Transformer (ViT) addresses issues our previous models faced with classes that closely resemble each other. Traditional models often struggled with these distinctions, leading to confusion and inaccurate classifications. Although techniques, like skip connections, dropout, and regularization, were implemented to mitigate overfitting, they did not fully resolve the issue satisfactorily.

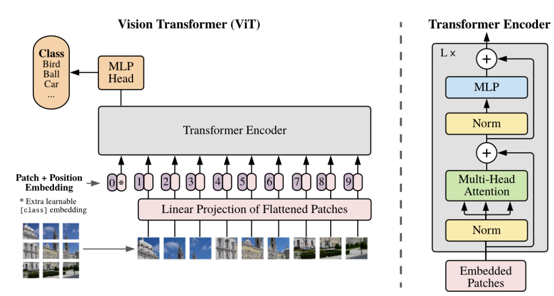

Fig 5: ViT Model (Referenced image from paper [9])

The ViT model's ability to integrate broader context and relationships between different parts of the image help in more accurately differentiating similar classes and reduce the tendency to overfit compared to conventional approaches

ViT breaks each image into fixed-size patches and flattens them. This is then embedded along with each patch's positional encodings. Embedding is then fed into transfer blocks to calculate self-attention for each patch concerning other patches. Doing so helps ViT understand the global dependencies between various patches for an image, which is better than CNN because CNN captures only local respective fields. This array of patches which has to be embedded in it is passed through various transformer blocks, and for output last layer is passed to a classification head which has softmax activation to calculate the class probability.

.

5. Results

Models	5-Classes		10-Classes	
	Test Accuracy	Test Loss	Test Accuracy	Test Loss
ViT	**96.90%***	0.7949	**96.97%***	0.8910
CNN	85.16%	0.6361	71.48%	1.2449
AlexNet	92.22%	0.2085	87.88%	0.4113
ResNet	88.89%	0.4623	87.27%	0.4556
FNN	62.13%	0.7910	58.89%	0.8303

Table 1: Test Accuracy and Loses on Various Models for 5 and 10 Classes

** Best Performing Model*

Using our enhanced data, we have trained and evaluated five distinct models for both top-5 and top-10 class classifications. Table 1 displays the accuracy and test loss for each model. Without using any image-related data for classification, we first constructed a fully connected neural net to obtain a baseline of achievable performance. We first obtained the numerical statistics from flow cam. Compared to this baseline model, we achieved significantly superior outcomes even with our original model that used CNN with few layers.

We kept seeing a significant improvement in performance over the basic CNN when we explored the deeper model further, according to the AlexNet paradigm. However, it proved difficult to outperform the baseline Alexnet model mostly due to insufficient data to build deeper models even after data augmentations. With 92% for top-5 class and 87.88% for top-10 class classification, it is evident that AlexNet is already performing admirably. After that, we attempted to employ ResNet—which has more layers than AlexNet—to induce learning at the deeper levels, but this did not improve accuracy.

Eventually, we gave the ViT model a shot. We utilized a pre-trained ViT model and adjusted it for our data, as was mentioned in the Experiments section. When compared to the other models, the ViT model fared the best. After five average runs, we achieved a test accuracy of above 96%. With the dataset, pretrained ViT that has been fine-tuned performs best. All the attention is on it.

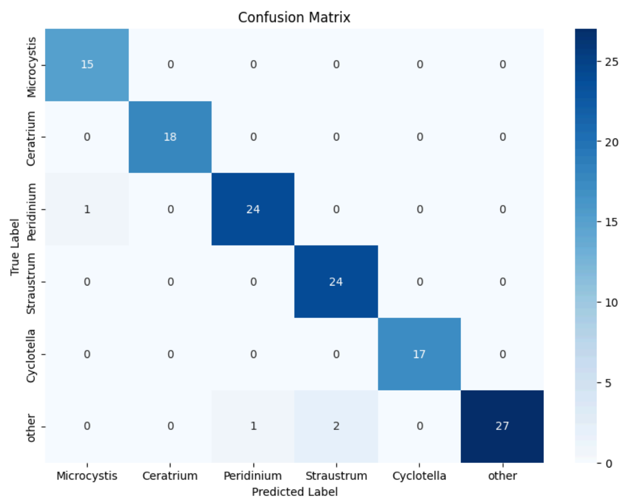

Fig 6: Confusion Matrix from VIT model results for top-5 class classification.

```
Classification Report:
              precision    recall  f1-score   support

          12       0.94      1.00      0.97        15
          43       1.00      1.00      1.00        18
          46       0.96      0.96      0.96        25
          53       0.92      1.00      0.96        24
           9       1.00      1.00      1.00        17
         999       1.00      0.90      0.95        30

    accuracy                           0.97       129
   macro avg       0.97      0.98      0.97       129
weighted avg       0.97      0.97      0.97       129
```

Fig 7: Classification Report from VIT model results for top-5 class classification.

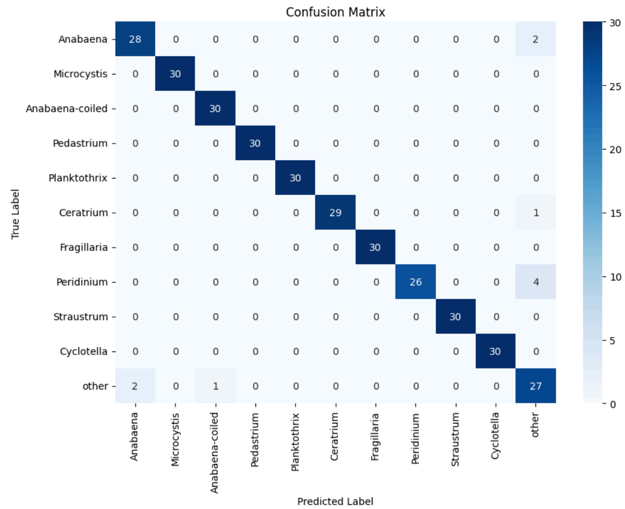

Fig 8: Confusion Matrix from VIT model results for top-10 class classification

```
Classification Report:
              precision    recall  f1-score   support

           1       0.93      0.93      0.93        30
          12       1.00      1.00      1.00        30
           2       0.97      1.00      0.98        30
          25       1.00      1.00      1.00        30
          27       1.00      1.00      1.00        30
          43       1.00      0.97      0.98        30
          44       1.00      1.00      1.00        30
          46       1.00      0.87      0.93        30
...
    accuracy                           0.97       330
   macro avg       0.97      0.97      0.97       330
weighted avg       0.97      0.97      0.97       330
```

Fig 9: Classification Report from VIT model results for top-10 class classification.

The Confusion Matrix derived from the VIT model results for both top-5 (Fig 6) and top-10 (Fig 8) class classifications shows impressive accuracy, with the vast majority of predictions perfectly aligned with the ground truth labels. Just seven of the ten classes considered have discrepancies, indicating that the model's performance could be improved further. Despite these obtained misclassifications, overall precision and recall scores (Fig 7 for the top 5) and (Fig 9 for the top 10 classes) remain significantly high, demonstrating the VIT model's robustness and efficacy in accurately distinguishing complex visual patterns across multiple classes.

5.1. Classification Output

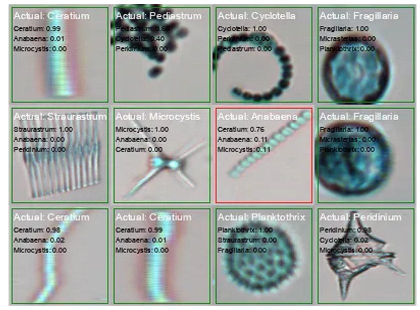

Fig 10: Softmax Outputs of top-3 Most Probable Algae

Figure 10 represents the softmax outputs of the top three most likely algae species, as predicted by our best-performing model. These softmax outputs provide insight into the model's confidence levels for each class prediction.

6. Conclusions

In conclusion, our experiments showed that the Vision Transformer (ViT) architecture outperformed traditional convolutional neural networks (CNNs) and well-known models such as ResNet and AlexNet. We believe several factors contribute to ViT's superior performance on our classification task. ViT's self-attention mechanism allows it to effectively capture global dependencies within images, allowing for a more holistic understanding of complex visual patterns. Also, due to its ability to grasp long-range dependencies.

Visualizing the probabilities assigned to each algae species provides researchers with valuable insights into the model's decision-making process. This information not only helps us understand which algae species are most likely to be present in the images, but it also provides a level of certainty for each prediction. Such insights are useful in ecological studies, biodiversity assessments, and environmental monitoring efforts, allowing scientists to make informed decisions.

Researchers can use these advanced architectures to classify images with precision, significantly reducing labour hours previously spent on manual classification tasks. This increased efficiency not only frees up valuable resources but also allows researchers to focus their efforts on more complex analyses and innovative research projects. The goal is not to replace human researchers but rather to make their job easier. Once the model provides an output for an image, researchers can give feedback to the model, which helps it improve using a feedback loop. Additionally, a confidence score for the output could also be provided along with the output, which is the softmax probability score.

7. Limitations

In classification tasks, only a select few classes are actively considered, typically the top 5 or top 10 most prevalent ones. The remaining classes are often grouped under a single category termed "others." This approach simplifies the classification process by prioritizing the most significant classes while relegating less common ones to a secondary status. Hence in this case we haven't considered all types/species of Algae classes.

More preprocessing steps could be explored, which complement the model’s performance. Since the images are low resolution, many models did not perform well for the top-5 classes.

8. Future Work

1. To tackle the limited number of images, GAN could generate more images, but this GAN should have controllable parameters, so it doesn’t mix features of one class with another, and after generating such images, it should be verified with a subject human expert in this field to check for the integrity of the produced images.
2. Based on these generated images, other statistical values can also be calculated for these new images. Two separate models could be trained: one for image classification using images, and the other using statistical values to predict the class. These could be used as components in an ensembling model or to build a student-teacher model.
3. To test the model on the live dataset, a streamlined pipeline can be built to deploy the model in the real world. And the model could be an online model so that it can adapt to changes in the real world.
4. Could build a self-supervised model, which classifies the wrong class, subject matter experts could correct it.
5. More images could be added with better resolution.
6. Try models that could handle low resolution such as U-Net, Efficient Net, Super-Resolution Convolutional Neural Networks (SRCNN), Feature Pyramid Networks (FPN) and other models that handle low resolution well.

9. Acknowledgement

We extend our sincere gratitude to Professor David Crandall, and Associate Instructors Vibas Vats and Sahaj Singh Maini, for their invaluable guidance throughout the Computer Vision course, which laid the foundation for our project. Additionally, we express our appreciation to Jill Minor, data analyst from the City of Bloomington, for her insights into the data and for generously sharing the dataset with us.

References

[1] Otálora Berenguel, Pablo & Guzmán, José & Acien, Gabriel & Berenguel, M. & Reul, Andreas. (2023). “*An artificial intelligence approach for identification of microalgae cultures. New biotechnology.”* 77. 10.1016/j.nbt.2023.07.003.

[2] Santhi N, Pradeepa C, Subashini P, Kalaiselvi S. “*Automatic identification of algal community from microscopic images. Bioinform Biol Insights.”* 2013 Oct 10;7:327-34. doi: 10.4137/BBI.S12844. PMID: 24151424; PMCID: PMC3798295.

[3] Salmi, Pauliina, Marco Calderini, Salli Pääkkönen, Sami Taipale, and Ilkka Pölönen. "*Assessment of microalgae species, biomass, and distribution from spectral images using a convolution neural network.*" Journal of Applied Phycology 34, no. 3 (2022): 1565-1575.

[4] Samantaray, Arabinda, Baijian Yang, J. Eric Dietz, and Byung-Cheol Min. "*Algae detection using computer vision and deep learning.*" arXiv preprint arXiv:1811.10847 (2018).

[5] Deglint, Jason L., Chao Jin, Angela Chao, and Alexander Wong. "*The feasibility of automated identification of six algae types using feed-forward neural networks and fluorescence-based spectral-morphological features*." Ieee Access 7 (2018): 7041-7053.

[6] Zhang, H., S. Yang, and X. Zhang. "*Automatic recognition and classification of microalgae using an inception-v3 convolution neural network model.*" International Journal of Environmental Science and Technology 21, no. 4 (2024): 4625-4634.

[7] Chong, Jun Wei Roy, Kuan Shiong Khoo, Kit Wayne Chew, Huong-Yong Tingmax-poolingKoji, Roger Ruan, Zengling Ma, and Pau Loke Show. "*Artificial intelligence-driven microalgae autotrophic batch cultivation: A comparative study of machine and deep learning-based image classification models.*" Algal Research (2024): 103400.

[8] Nash, R. (2015). An Introduction to Convolutional Neural Networks. *ArXiv*. /abs/1511.08458

[9] Alexey Dosovitskiy, Lucas Beyer, “An Image is worth 16X16 words: Transformers for image Recognition at Scale”,ICLR 2021