AI Algae Classification using Neural Networks

Avishmita, Hemesh, Sarabesh, Yashaswini

Motivation

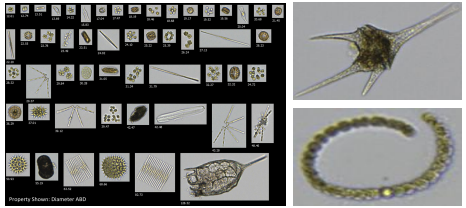

Fig 1. Algal microscopic images from FlowCam

1. To help **accurate identification** and quantification of algae species, so researchers can track changes in water quality over time, detect **harmful algal blooms (HABs)**.
2. Transitioning from regular statistical analysis to **AI image recognition**, we aim to substantially enhance the **accuracy** of classifying algae obtained from FlowCam imagery.
3. **Automation** of classification tasks frees up human resources for other critical aspects of monitoring and conservation.

Dataset

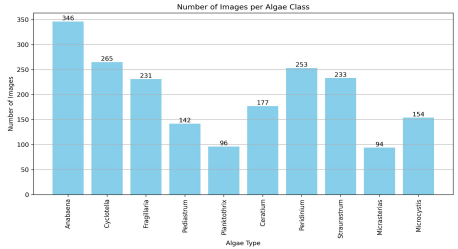

Fig 2. Dataset distribution across different algae classes

- Dataset is obtained from a **DB dump** taken by FlowCam capturing microscopic algae images.
- Images are **meticulously classified** into their groups, contributing to a comprehensive dataset accumulated over several years from the **City of Bloomington**.

Method & Model Architecture

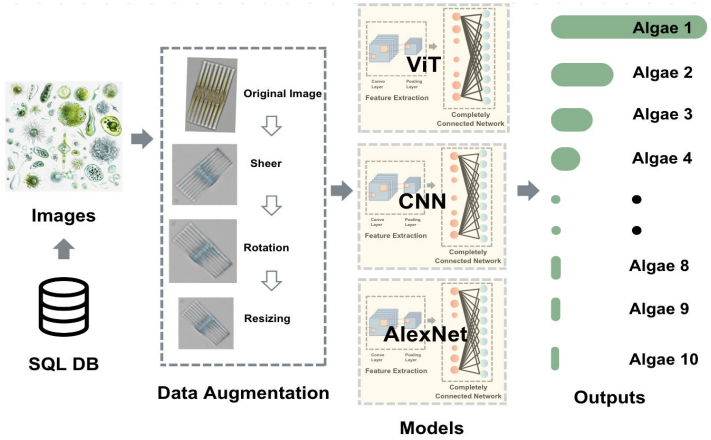

Fig 3. Model Workflow

Models	Reason and Advantages	Drawbacks
ViT	**Pretrained Vision Transformers (ViTs) as it can be trained with just few more layers used for training.**	**Though this provided us the best result, we require data more representative of the classes.**
CNN	CNNs were preferred as it works well for image classification due to their capacity to automatically extract hierarchical features from raw pixel data.	We got maximum of 94% test accuracy, but were unable to go higher with basic changes in model. .
AlexNet	Tried AlexNet for pioneering deep CNN capabilities.	It performed worse than our CNN model, due to less data.
ResNet	Considered ResNet for its deep architecture with residual connections, enabling training of deeper networks.	It got lowest accuracy than all other models with CNNs, we are assuming again it's due to lack of data.
FNN (Statistical)	To compare model based only on numerical features provided from FlowCam.	It performed poorly, than even a simple CNN improved the accuracy by large margins.

Table 1. Models explored

Results

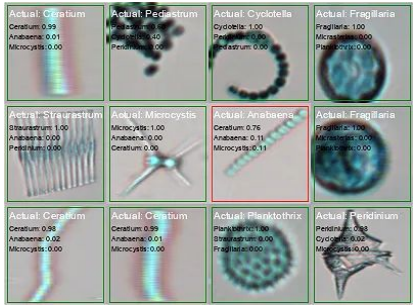

Fig 4 : Softmax Outputs of top-3 most probable algae

Models	5-Classes		10-Classes	
	Test Accuracy	Test Loss	Test Accuracy	Test Loss
ViT	**98.45%***	0.7949	**97.27%***	0.6010
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

** Best Performing Model*

Table 2: Test Accuracy and Loses on Various models for 5 and 10 classes

Future Work

- Plan to utilize advanced machine learning techniques such as **RNNs** and **GNNs** alongside CNNs for more nuanced algae classification.
- Gather more data from **diverse sources** to enhance the dataset, enabling comprehensive training and improving classification accuracy.
- Carry out classification of **more number of classes** of algae.