

# HAB detection within Aquaculture Industry: A Case Study in the Atlantic Area\*

Bruna Guterres<sup>\*†</sup>, Kauê Sbrissa<sup>†</sup>, Amanda Mendes<sup>†</sup>, Lucas Meireles<sup>†</sup>, Lucie Novoveska<sup>‡</sup>  
Francisca Vermeulen<sup>‡</sup>, Javier Martinez<sup>§</sup>, Aitor Garcia<sup>§</sup>, Lisl Lain<sup>¶</sup>, Marié Smith<sup>¶</sup>  
Paulo Drews<sup>†</sup>, Nelson Duarte<sup>†</sup>, Vinicius Oliveira<sup>†</sup>, Marcelo Pias<sup>†</sup>, Silvia Botelho<sup>†</sup> and Rafaela Machado<sup>†</sup>

<sup>\*†</sup>Federal University of Rio Grande - FURG - Brazil

<sup>‡</sup>The Scottish Association for Marine Science - SAMS - United Kingdom

<sup>§</sup>Acondicionamiento Tarrasense Asociación - LEITAT - Spain

<sup>¶</sup>The Council for Scientific and Industrial Research - CSIR - South Africa

Email: \*guterres.bruna@furg.br

**Abstract**—Fisheries and aquaculture industries notably contribute to animal-source protein production worldwide. Climate change is creating environmental conditions suitable for harmful algal blooms (HAB) on a global scale. Some phytoplankton species can also release toxins, which may cause large-scale marine mortality with knock-on effects on coastal economies. Reliable phytoplankton monitoring and early HAB detection are also essential in climate-resilient solutions for aquaculture applications. Currently, phytoplankton monitoring is primarily based on traditional microscopy. However, it is time-consuming and requires an experienced taxonomist. There is a need to expedite and automate phytoplankton monitoring to support aquaculture industries. Analytical instruments based on microscopy coupled with artificial intelligence (AI) models may be vital to monitoring applications. Digital plankton data sets are usually imbalanced and reflect natural environmental differences. The lack of data to represent minority species/genera prevents AI models from understanding some taxa completely. It compromises system reliability for HAB monitoring applications. The present study investigates state-of-the-art models for class imbalance problems tailored for HAB monitoring within multi-trophic aquaculture farms from Brazil, South Africa, and Scotland. A unified benchmark database covering publicly available microscopic image-based datasets supported phytoplankton modelling. AI deep collaborative models and threshold moving techniques provided the best results compared to standard architectures. It prevailed, especially for low-abundant yet toxic organisms.

**Index Terms**—Harmful Algal Blooms, Class Imbalance, Climate Change, Aquaculture, Deep Learning

## I. INTRODUCTION

Aquaculture is a significant and expanding industry that provides a sustainable source of seafood for people worldwide. In 2022, it produced 76.9 tons of animal-based protein for human consumption, making up 49% of total seafood production and creating a USD 265 billion industry [1]. This highlights the increasing importance of aquaculture in meeting the growing demand for fish as a source of food [?]. However, Aquaculture industry faces a number of challenges. Harmful Algal Blooms (HAB) have expressively compromised aquaculture sector worldwide. It has caused expressive economic losses and impacted food production worldwide (e.g. loss of

thousands of salmon within Scottish (£10 million loss) and Chilean (US\$ 50 million) aquaculture industries). In this sense, providing AI solutions for phytoplankton monitoring and early HAB detection tailored for aquaculture industry has paramount importance.

HAB are characterized by the fast growth of phytoplankton biomass. It may exert many adverse effects such as the proliferation of toxin-producing species [2] and large-scale marine mortality leading to economic impacts in coastal regions and serious consequences for aquaculture industries [2]. Climate change has increased the frequency and severity of HAB worldwide [3], making reliable HAB monitoring within aquaculture applications an imperative need. Thus, phytoplankton monitoring and early HAB detection are essential for safeguarding marine life, economic activities, and human health. AI models coupled with microscopy image-based analytical instruments may best support early HAB detection within aquaculture applications.

Plankton databases are usually skewed and reflect natural distribution differences within the environment [4]. Plankton images from the same taxonomic species have inherent cell orientation, colour and size variability. There may not be enough data to represent minority HAB species properly. As a result, it prevents AI models from gaining a complete understanding of low-abundant yet potentially toxic taxa [4].

Recent technological advances in machine learning have enabled in situ plankton image capture in real-time at low cost [5]. State-of-the-art discussions present deep learning as a crucial solution for early HAB detection [6], [7]. However, they still struggle with phytoplankton image classification within aquatic monitoring scenarios. For instance, high intra-class variability and inter-class similarity prevent the practical identification of morphologically similar species [7], and low-abundant organisms due to imbalanced databases [4].

Publicly available microscopic image databases (e.g. WHOI [8]) have been the basis for building most AI solutions for phytoplankton monitoring. However, phytoplankton diversity makes communities highly heterogeneous in size, shape, and morphology within different areas. State-of-the-art models built upon representative image databases may best support

aquaculture industry applications. There is an imperative to consider target phytoplankton species, end-user needs and expectations for building deployable and reliable deep-learning solutions.

This work investigates early HAB detection and phytoplankton monitoring within Integrated Multi-Trophic Aquaculture (IMTA) farms based on state-of-the-art deep learning techniques tailored for class imbalance classification problems. A data integration pipeline builds a representative database from publicly available datasets considering end-user needs and constraints. The work aims to provide automatic phytoplankton monitoring with high throughput for rapid HAB detection. Open challenges for reliable phytoplankton monitoring and technology deployment within aquaculture applications are also discussed.

## II. RELATED LITERATURE

Biased plankton databases reflect natural imbalances within the aquatic environment [4]. Many approaches can address class imbalance and phytoplankton classification challenges. Convolutional Neural Networks (CNNs) have extracted relevant features from images [9]. However, they still face misclassification because of similarity in shape, size and texture among phytoplankton species. CNNs with repeated layers (e.g. ResNet, MobileNet), optimization functions, and ensemble techniques present promising research areas [10] to support HAB monitoring applications.

A heterogeneous ensemble of CNN models harnesses the limited understanding of individual models to provide a collective and more accurate classification of minority classes [4]. Two-phase learning allows the minority classes to contribute more to the gradient descent during a pre-training stage [11]. New loss functions also address class imbalance problems.





















Other techniques include Cost Sensitive (CS) learning and threshold moving [12]. CS assigns higher weights to minority classes and minimizes misclassification cost [11]. Unlike other methods, threshold moving may be quickly implemented on already trained models to improve classification results. This approach has outperformed baseline CNNs for different levels of class imbalance [12].

Several studies provide comparisons among classic CNN architectures. Densenets [13], Nasnets [13], Resnets [14], and VGGNets [14] are commonly described as prominent models to boost the classification of minority classes individually. The best performance varies within pre-processing strategies, datasets, training parameters, and other aspects. In this sense, testing different approaches for classifying phytoplankton is an essential primary task for reliable monitoring.

## III. MATERIALS AND METHODS

### A. Dataset

AI models must be trained on phytoplankton images that represent usually encountered species within aquaculture industry applications. The present work employs the data integration pipeline proposed by [15]. It considers publicly

Genus	Aquaculture farm	Genus	Aquaculture farm
Alexandrium		Nodularia	
Anabaena		Prorocentrum	
Chaetoceros		Protoцератium	
Dinophysis	 	Pseudo-nitzschia	 
Gonyaulax	 	Skeletonema	
Karenia	 	Tetraselmis	
Katodinium		Thalassiosira	
		Lingulodinium	 




Brazil (  ), South Africa (  ) and UK (  )

Fig. 1. Target phytoplankton genera within ASTRAL IMTA labs from Brazil, South Africa and Scotland.

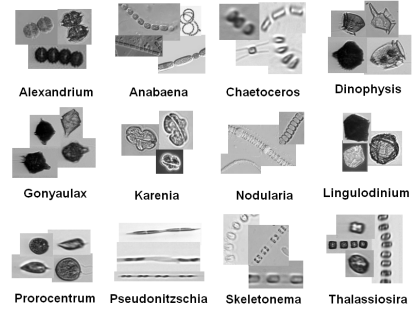


Fig. 2. Image samples within target phytoplankton genera. The database provides classification challenges due to intra-class variability and interclass similarity.

available image-based datasets and targets phytoplankton genera (Figure 1) within ASTRAL IMTA labs from Brazil, South Africa, and Scotland. The resulting unified benchmark database is tailored for IMTA application needs. It aims to support ASTRAL technology development and validation in industrially relevant environments.

The data integration pipeline proposed by [15] uses the most comprehensive public dataset (WHOI - Woods Hole Oceanographic Institution [8]) as the basis for data processing. Output data are grey-scaled, fixed-size images, considering expected size ranges within target phytoplankton genera. Grey-scale information is replicated into three image channels to address AI input requirements. The pipeline succeeded towards a more representative database for the target IMTA applications. Unfortunately, the data integration yielded a severely imbalanced database. Only 73% of target phytoplankton genera have at least 20 images (Table I) and were used for AI modelling.

Some phytoplankton organisms release toxins providing additional threats to aquatic ecosystems. They can be harmful even at very low cell abundance. Table I depicts target phytoplankton genera within the ASTRAL IMTA labs. It also provides information for toxin producers. Figure 2 illustrates image samples within target phytoplankton genera. Intra-class variability and interclass similarity may provide challenges for accurate classification.

The database is randomly split into training and testing (80% and 20% of images, respectively). Validation employed

TABLE I  
NUMBER OF PHYTOPLANKTON IMAGES USED IN THE MODEL. KNOWN TOXINS PRODUCED BY THE TARGET GENERA ARE ALSO LISTED (PSP IS PARALYTIC SHELLFISH POISONING).

Class	Total	Toxins	Size ( $\mu m$ )
<i>Alexandrium</i>	20	PSP toxins (e.g. saxitoxin)	20 - 50
<i>Anabaena</i>	61	Anatoxin and microcystin	Filaments, 4 $\mu m$ wide
<i>Chaetoceros</i>	48249	-	-
<i>Dinophysis</i>	838	Diarrhoetic toxins	20-100
<i>Gonyaulax</i>	592	PSP yessotoxin	25-50
<i>Lingulodinium</i>	27	PSP yessotoxin	40-55
<i>Nodularia</i>	45	Nodularin	Filaments, 8 $\mu m$ wide
<i>Prorocentrum</i>	2622	Diarrhoetic toxins	10-75
<i>Pseudonitzschia</i>	3542	Domoic acid	25-160
<i>Skeletonema</i>	13979	-	-
<i>Thalassiosira</i>	11416	-	-

twenty per cent of the training images.

### B. Class imbalance approaches

Several deep learning architectures have shown promise in boosting the classification of minority classes in phytoplankton datasets, including DenseNets [14], NasNets [14], ResNets [13], [14], and VGGNets [13]. However, the best-performing model architecture can vary depending on factors such as pre-processing strategies, training parameters, and the specific dataset being used. Therefore, testing different approaches for phytoplankton classification is a crucial step in supporting early HAB monitoring in aquaculture industries. In this study, we investigate several CNN architectures tailored for phytoplankton monitoring within industrial IMTA applications. The most effective CNN architecture will serve as a baseline for implementing state-of-the-art approaches and addressing the class imbalance classification problem. Table II summarizes hyper-parameters used to train and select the baseline model. Specifically, we investigate the following architectures:

- **VGG16** introduces the use of smaller receptive fields (3x3) compared to conventional convolutional networks. As a result, this network achieves high performance by having more activation layers and fewer weight parameters than 5x5 and 7x7 models. [16].
- **InceptionV3** employs factorized convolutions and dimension reduction in a 48-layer deep learning model. This architecture improves computational cost and may be over three times faster than similar networks [17].
- **NASNetMobile** architecture is also optimized for mobile and embedded vision tasks. NASNet focuses on searching for an optimal CNN architecture using reinforcement learning. NAS (Neural Architecture Search) proposes to search for a good architecture on a small dataset (CIFAR-10) and then transfer the learned architecture to a more extensive dataset (ImageNet).
- **MobileNetV2** targets mobile and resource-constrained platforms. They encompass depth-wise separable convolutions. It is a form of factorized convolution that significantly reduces computational cost and model size.

MobileNetV2 also introduces the inverted residual with a linear bottleneck layer.

TABLE II  
HYPERPARAMETERS FOR TRAINING CNN ARCHITECTURES.

Architecture	Loss Function	Number of Epochs	Optimizer
VGG16	Dice Loss	100	Adam
InceptionV3	Dice Loss	100	Adam
NASNetMobile	Dice Loss	100	Adam
MobileNetV2	Dice Loss	30	SGD

1) *Focal loss*: For instance, Focal Loss (FL) reshapes the cross entropy loss to reduce the impact caused by more easily classified samples during the training process [11].

FL comprises changing the Cross-Entropy (CE) loss to prevent large numbers of easily classified samples from the primary classes from overwhelming the training process [18]. In the Focal Loss equation (Eq. (1)),  $\alpha_t$  is a class-wise factor. It increases the relevance of minority classes. The hyperparameter  $\gamma$  defines the rate that down weights easy examples [18].

$$FL(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t) \quad (1)$$

It is adapted for a multi-class problem (Eq. (2)) by summing the individual loss for each of the  $n$  classes [19].  $y_t$  and  $p_t$  represent the expected and predicted probabilities for the class  $t$ , respectively.

$$FL = \sum_{t=1}^n -\alpha_t(1 - p_t)^\gamma \cdot y_t \cdot \log(p_t) \quad (2)$$

2) *Cost-sensitive*: Cost-sensitive learning assigns penalties to each class through a cost matrix. For example, increasing the cost of the minority group is equivalent to increasing its importance, decreasing the likelihood that the learner will incorrectly classify instances from specific classes.

The cost-sensitive approach applies different penalties to the learner, depending on the class of a misclassified sample [11]. Each instance contributes to the loss proportionally to its class weight. Therefore, the cost of a class is directly proportional to its importance in updating weights. The present work empirically defines the cost of each class. Keras feeds it into the network through the *class\_weight* parameter in Keras' fit method. The cost for minority classes ( $n < 100$ ) is set to ten (10x) times higher than for abundant classes.

3) *Two-phase learning*: Two-phase learning usually combines RUS with transfer learning. The pre-training phase adjusts the model based on a balanced dataset. [20] experimentally defined a balanced database with 5000 images per class to support the pre-training stage. The present study employs a hybrid approach with RUS and ROS to build a balanced dataset ( $N = 5000$ ) [11]. A final training phase employs the original imbalanced data for model fine-tuning.

RUS randomly selects  $N$  images from classes with over 5000 images. Then, ROS employs data augmentation techniques for the remaining classes. It artificially generates additional training images, considering the data augmentation parameters adapted from [4]. The augmentation applies to the original training images randomly. It aims to allow the model to understand low-abundant phytoplankton genera better. The pre-training phase runs until the metric cannot improve for over five epochs. The second training stage comprises model fine-tuning with the original class distribution.

4) *Dynamic Sampling*: The Dynamic Sampling [21] aims to boost the classification of minority classes. It changes the class distribution of the training samples dynamically. The model iteratively focuses on classes with poor performance within the training process.

Dynamic Sampling splits the database into training, reference, and testing sets. Initially, the number of samples for each class is  $N^*$ , which is the average number of samples. By the end of each training iteration, F1-Score assesses performance for the reference set. It is the basis for defining the number of samples from each class during the next training iteration. For example, equation (3) defines the number of samples  $N$  of a class  $c_k$  in the iteration  $i$ .

$$N_{i,c_k} = \frac{1 - f1_{i,c_k}}{\sum_{c_k \in C} 1 - f1_{i,c_k}} \quad (3)$$

The present work uses 20% of the training dataset as a reference set, leaving 80% for the training itself. Considering this split, the average number of samples in the training set of all eleven classes is 4735. However, the number of samples from eight of the eleven classes is lower than this value. Therefore, two dynamic sampling approaches are implemented.

The first approach uses the ROS method to ensure all classes have at least 4735 training samples. Alternatively, the other approach uses no over-sampling. Instead, the number of images of each class is defined by  $\min(N_{i,c_k}, n_{c_k})$  where  $N_{i,c_k}$  is calculated as in Eq. (3) and  $n_{c_k}$  is the number of samples of class  $c_k$  in the training set. The F1-score assesses class performances for both approaches according to the reference set. The next training iteration employs images randomly sampled from the training database (Eq. 3).

5) *Ensemble methods*: Deep collaborative models have provided outstanding performance compared to individual CNNs [4], [13]. The work assembles different models tailored for class-imbalance applications to boost low-abundant phytoplankton genera classification. The deep learning collaborative model includes the two models with the highest performance compared to the baseline.

6) *Threshold moving*: The threshold moving adjusts the decision threshold of a classifier during the test phase by changing the output class probabilities. The most basic version compensates for prior class probabilities.

Considering neural networks estimate Bayesian *a posteriori* probabilities, the output  $y$  for class  $i$  implicitly corresponds to  $y_i(x) = p(i|x) = p(i) * p(x|i)/p(x)$  for a data point  $x$  [22]. Thus, dividing the network output for each class

by its estimated prior probability provides the correct class probabilities (Eq. (4)).

$$p(i) = \frac{|i|}{\sum_k |k|} \quad (4)$$

where  $|i|$  denotes the number of unique examples in class  $i$ . The present work applies the threshold moving technique to the resulting model with the highest performance within the target phytoplankton genera.

#### IV. EVALUATING MODEL PERFORMANCE

Accuracy is the ratio between correctly classified samples against the total number of tested data. Although intuitive, overall accuracy is a misleading performance metric for imbalanced scenarios [12]. AI models may provide high accuracy levels and still achieve poor performance for low-abundant taxa. F-score (Eq. 7) is sensitive to the performance within minority classes [23]. It provides the harmonic mean between precision (Eq. 5) and recall (Eq. 6) metrics.

$$precision = \frac{TP}{TP + FP} \quad (5)$$

$$recall = \frac{TP}{TP + FN} \quad (6)$$

$$F - score = \frac{2.Precision.Recall}{Precision + Recall} \quad (7)$$

Where  $TP$ ,  $FP$  and  $FN$  are the numbers of true positives, false positives and false negatives in a classification process.

Since precision provides a more accurate representation of a model performance in a skewed distribution [24], the evaluation metrics also include the area under the precision-recall curve (AUC-PR). The present work also assesses model size. It supports further model integration into embedded platforms and resource-constrained environments.

#### V. RESULTS AND DISCUSSIONS

Table III summarizes CNN performance. MobileNetV2 provided the best individual results and was selected as the baseline model. It achieved the best performance and comprised a smaller model size which may be helpful for embedded and resource-constrained applications. Table IV depicts MobileNetV2 performance within target phytoplankton genera. However, it still struggles with phytoplankton genus classification, especially for low-abundant classes.

TABLE III  
CLASSIFICATION PERFORMANCE AND MODEL SIZE REGARDING INVESTIGATED CNN MODELS. PERFORMANCE IS DESCRIBED THROUGH MACRO F-SCORE AND AUC-PR METRICS.

Architecture	F-Score	AUC-PR	Model Size
VGG16	0.57	0.972	59.6MB
InceptionV3	0.63	0.987	97.1MB
NASNetMobile	0.75	0.992	43MB
MobileNetV2	0.75	0.996	29.0MB

TABLE IV  
CLASSIFICATION REPORT OF MOBILENETV2 REGARDING TARGET  
PHYTOPLANKTON GENERA WITHIN ASTRAL IMTA LABS

Genus	Precision	Recall	F-Score	Support
<i>Alexandrium</i>	0.00	0.00	0.00	4
<i>Anabaena</i>	0.33	0.23	0.27	13
<i>Chaetoceros</i>	0.98	0.98	0.98	9650
<i>Dinophysis</i>	0.93	0.92	0.93	168
<i>Gonyaulax</i>	0.96	0.99	0.98	119
<i>Lingulodinium</i>	1.00	0.50	0.67	6
<i>Nodularia</i>	0.47	0.78	0.58	9
<i>Prorocentrum</i>	0.97	0.99	0.98	525
<i>Pseudonitzschia</i>	0.97	0.98	0.98	709
<i>Skeletonema</i>	0.95	0.95	0.95	2796
<i>Thalassiosira</i>	0.98	0.96	0.97	2284
<b>Accuracy</b>			0.97	16283
<b>Macro avg</b>	0.78	0.75	0.75	16283
<b>Weighted avg</b>	0.97	0.97	0.97	16283

*Alexandrium*, *Anabaena*, *Lingulodinium* and *Nodularia* genera ( $n_i > 100$ ) include toxin-producing species also related to HAB occurrences. Therefore, improving model performance for low-abundant classes is imperative for reliable phytoplankton monitoring within aquaculture applications.

Focal loss, threshold moving, dynamic sampling and deep collaborative methods are some approaches investigated to address the class imbalance problem. Tables V show the resulting performances within target phytoplankton genera. Table VI summarizes overall performance and model sizes.

The focal loss provided minor performance gains compared to the baseline model. Cost-sensitive learning and dynamic sampling yielded comparable performance gains without increasing model size and employability within embedded systems. A dynamic sampling approach performed better than focal loss and cost-sensitive learning. It also provided a smaller model size, which may better support integration to embedded resource-constrained systems. Although compromising model compactness, two-phase learning allowed for valuable performance gains compared to baseline, focal loss, and dynamic sampling techniques.

The deep collaborative approach ensembles dynamic sampling and two-phase learning models. They achieved better performance when applied individually. The combined approach provided outstanding performance compared to baseline and individual class imbalance methodologies. The resulting model size was 90% greater than the baseline CNN. However, it provided a smaller model size compared to two-phase learning. In addition, the moving threshold technique was applied to the deep collaborative model. It allowed further performance improvement without affecting model size (Table VI). The outcomes indicate deep collaborative models and threshold moving as possible solutions towards reliable HAB monitoring through embedded system solutions.

The baseline model provided poor performance in the identification of crucial phytoplankton organisms. The proposed method provided promising performance gains to address phytoplankton monitoring within aquaculture applications. It allowed for building deep collaborative models tailored to

aquaculture needs and requirements. The main results include outstanding classification improvement of low-abundant and toxin-producing genera. For instance, the F-score metric for identifying *Alexandrium*, *Anabaena*, *Lingulodinium* and *Nodularia* genera increased from 0.00, 0.27, 0.67 to 0.86, 0.67, 0.80 and 0.84, respectively. Two-phase learning has enabled performance gains for the low-abundant general classification. However, it increased model size by 90% compared to baseline MobileNetV2 which may affect AI usability within resource-constrained prototypes.

Deep collaborative modelling, dynamic sampling and threshold moving techniques have also allowed model optimization. Threshold moving may be employed upon the latest deep learning architectures without compromising model employability in embedded and resource-constrained applications.

There is a trade-off between model performance and size which may play an important role in reliable embedded solutions for early HAB detection within aquaculture applications. It prevails especially for low-cost resource-constrained embedded systems. Assessment of model integration feasibility and consequent effects on prototype throughput, power consumption and autonomy is part of ongoing work.

## VI. CONCLUSIONS

This work presents the development of AI deep learning models tailored for phytoplankton monitoring and early HAB detection within multi-trophic aquaculture industries. It provided a broad discussion about the main challenges to support early HAB detection within industrially relevant scenarios. The investigated techniques addressed class imbalance issues within a benchmark database built upon publicly available phytoplankton images.

Results showed that deep collaborative models, dynamic sampling and threshold moving techniques are potential solutions to better support phytoplankton monitoring within aquaculture applications. Significant improvements were observed when comparing the performance of these approaches and the baseline model, enabling the identification of toxin-producing species in exchange for increased model sizes. The feasibility of model integration to the embedded platform and consequent impacts on prototype throughput and autonomy are part of ongoing work.

## VII. ACKNOWLEDGEMENTS

This work was developed as part of the ASTRAL (All Atlantic Ocean Sustainable, Profitable and Resilient Aquaculture) project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement 863034. The authors would like to acknowledge Kati Michalek from the Scottish Association for Marine Science (SAMS) for phytoplankton sample provision and draft revision.

TABLE V  
CLASSIFICATION REPORT FROM THE THRESHOLD MOVING TECHNIQUE APPLIED AS PART OF THE DEEP COLLABORATIVE (ENSEMBLE) MODEL.

	Baseline	Focal Loss	CS (10x)	DS	2PL	Ensemble (DS + 2PL)	TM (Ensemble)
<i>Alexandrium</i>	0.00	0.40	0.57	0.40	0.57	0.86	0.86
<i>Anabaena</i>	0.27	0.19	0.20	0.52	0.61	0.67	0.67
<i>Chaetoceros</i>	0.98	0.98	0.98	0.97	0.98	0.98	0.99
<i>Dinophysis</i>	0.93	0.89	0.95	0.93	0.95	0.93	0.94
<i>Gonyaulax</i>	0.98	0.91	0.95	0.96	0.97	0.85	0.98
<i>Lingulodinium</i>	0.67	0.60	0.62	0.67	0.80	0.80	0.80
<i>Nodularia</i>	0.58	0.52	0.63	0.75	0.78	0.84	0.84
<i>Prorocentrum</i>	0.98	0.94	0.98	0.97	0.99	0.97	0.98
<i>Pseudonitzschia</i>	0.98	0.97	0.98	0.96	0.98	0.95	0.97
<i>Skeletonema</i>	0.95	0.95	0.96	0.94	0.96	0.97	0.96
<i>Thalassiosira</i>	0.97	0.96	0.97	0.95	0.98	0.97	0.98
<b>macro avg</b>	<b>0.75</b>	<b>0.75</b>	<b>0.80</b>	<b>0.82</b>	<b>0.87</b>	<b>0.89</b>	<b>0.91</b>

TABLE VI

RESULTS FROM THE DIFFERENT CLASS IMBALANCE APPROACHES. IT INCLUDES FOCAL LOSS, COST SENSITIVE (CS), DYNAMIC SAMPLING (DS), TWO-PHASE LEARNING (2PL), DEEP COLLABORATIVE MODEL (ENSEMBLE) AND THRESHOLD MOVING. THE DEEP COLLABORATIVE MODEL COMBINED DS AND 2PL MODELS TOWARDS ACCURATE PHYTOPLANKTON CLASSIFICATION. THRESHOLD MOVING WAS APPLIED TO ENSEMBLE OUTPUT. PERFORMANCE METRICS INCLUDE PRECISION, RECALL AND F-SCORE. MODEL SIZE IS ALSO INCLUDED.

Method	Recall	Precision	F-Score	Model Size
<b>None (Baseline)</b>	0.75	0.78	0.75	29.1MB
<b>Focal Loss</b>	0.79	0.79	0.75	29.0MB
<b>CS (10x)</b>	0.82	0.79	0.80	29.0MB
<b>DS</b>	0.88	0.77	0.82	23.0MB
<b>2PL</b>	0.84	0.91	0.87	54.8MB
<b>Ensemble (DS + 2PL)</b>	0.87	0.94	0.89	50.0MB
<b>TM (Ensemble)</b>	0.88	0.94	0.91	50.0MB

## REFERENCES

- [1] Food and A. O. of the United Nations FAO, *The State of World Fisheries and Aquaculture - Towards Blue Transformation*, 2022.
- [2] E. Zohdi and M. Abbaspour, "Harmful algal blooms (red tide): a review of causes, impacts and approaches to monitoring and prediction," *International Journal of Environmental Science and Technology*, vol. 16, no. 3, pp. 1789–1806, 2019.
- [3] M. L. Wells, V. L. Trainer, T. J. Smayda, B. S. Karlson, C. G. Trick, R. M. Kudela, A. Ishikawa, S. Bernard, A. Wulff, D. M. Anderson *et al.*, "Harmful algal blooms and climate change: Learning from the past and present to forecast the future," *Harmful algae*, vol. 49, pp. 68–93, 2015.
- [4] T. Kerr, J. R. Clark, E. S. Fileman, C. E. Widdicombe, and N. Pugeault, "Collaborative deep learning models to handle class imbalance in flowcam plankton imagery," *IEEE Access*, vol. 8, pp. 170 013–170 032, 2020.
- [5] T. Pollina, A. G. Larson, F. Lombard, H. Li, S. Colin, C. de Vargas, and M. Prakash, "Planktonscope: affordable modular imaging platform for citizen oceanography," *BioRxiv*, 2020.
- [6] L. Vaughan, A. Zamyadi, S. Ajampur, H. Almutaram, and S. Freguia, "A review of microscopic cell imaging and neural network recognition for synergistic cyanobacteria identification and enumeration," *Analytical Sciences*, pp. 1–19, 2022.
- [7] R.-M. Plonus, J. Conradt, A. Harmer, S. Janßen, and J. Floeter, "Automatic plankton image classification—can capsules and filters help cope with data set shift?" *Limnology and Oceanography: Methods*, vol. 19, no. 3, pp. 176–195, 2021.
- [8] E. C. Orenstein, O. Beijbom, E. E. Peacock, and H. M. Sosik, "Whoiplankton—a large scale fine grained visual recognition benchmark dataset for plankton classification," *arXiv preprint arXiv:1510.00745*, 2015.
- [9] V. P. Pastore, T. G. Zimmerman, S. K. Biswas, and S. Bianco, "Annotation-free learning of plankton for classification and anomaly detection," *Scientific reports*, vol. 10, no. 1, pp. 1–15, 2020.
- [10] J. S. Ellen, C. A. Graff, and M. D. Ohman, "Improving plankton image classification using context metadata," *Limnology and Oceanography: Methods*, vol. 17, no. 8, pp. 439–461, 2019.
- [11] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance. j. big data 6 (1), 1–54 (2019)."
- [12] M. Buda, A. Maki, and M. A. Mazurowski, "A systematic study of the class imbalance problem in convolutional neural networks," *Neural Networks*, vol. 106, pp. 249–259, 2018.
- [13] A. Lumini, L. Nanni, and G. Maguolo, "Deep learning for plankton and coral classification," *Applied Computing and Informatics*, 2020.
- [14] K. Cheng, X. Cheng, Y. Wang, H. Bi, and M. C. Benfield, "Enhanced convolutional neural network for plankton identification and enumeration," *PLoS One*, vol. 14, no. 7, p. e0219570, 2019.
- [15] B. Guterres, S. Khalid, M. Pias, and S. Botelho, "A data integration pipeline towards reliable monitoring of phytoplankton and early detection of harmful algal blooms," in *NeurIPS 2021 Workshop Tackling Climate Change with Machine Learning*. NeurIPS, 2021.
- [16] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv 1409.1556*, 09 2014.
- [17] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," *arXiv 1409.4842*, 2014.
- [18] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.
- [19] W. Liu, L. Chen, and Y. Chen, "Age classification using convolutional neural networks with the multi-class focal loss," in *IOP conference series: materials science and engineering*, vol. 428, no. 1. IOP Publishing, 2018, p. 012043.
- [20] H. Lee, M. Park, and J. Kim, "Plankton classification on imbalanced large scale database via convolutional neural networks with transfer learning," pp. 3713–3717, 2016.
- [21] S. Pouyanfar, Y. Tao, A. Mohan, H. Tian, A. S. Kaseb, K. Gauen, R. Dailey, S. Aghajanzadeh, Y.-H. Lu, S.-C. Chen *et al.*, "Dynamic sampling in convolutional neural networks for imbalanced data classification," in *2018 IEEE conference on multimedia information processing and retrieval (MIPR)*. IEEE, 2018, pp. 112–117.
- [22] M. D. Richard and R. P. Lippmann, "Neural network classifiers estimate bayesian a posteriori probabilities," *Neural computation*, vol. 3, no. 4, pp. 461–483, 1991.
- [23] Y. Marrakchi, O. Makansi, and T. Brox, "Fighting class imbalance with contrastive learning," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2021, pp. 466–476.
- [24] S. A. Khan and Z. A. Rana, "Evaluating performance of software defect prediction models using area under precision-recall curve (auc-pr)," in *2019 2nd International Conference on Advancements in Computational Sciences (ICACS)*. IEEE, 2019, pp. 1–6.