BENCH TOP VISUAL HAB MONITORING: SUPPORTING CLIMATE RESILIENT AQUACULTURE IN THE ATLANTIC AREA

B. Guterres, K. Sbrissa, A. Mendes, L. Pereira, F. Vermeulen, K. Michalek, L. Lain, M. Smith, J. Martinez, S. Botelho, P. Drews, N. Duarte, V. Menezes, L. Novoveska and M. Pias*

Federal University of Rio Grande, Av. Itália – KM 8, Rio Grande (Brazil)
E-mail: guterres.bruna@furg.br

Introduction

Harmful algae blooms (HABs) are characterized by a massive proliferation of phytoplankton organisms, which provide a myriad of adverse effects such as large-scale marine mortality, economic impacts in coastal regions and consequences to aquaculture industries. Climate change has affected HAB frequency and severity on a global scale. In this scenario, machine learning may be an invaluable tool in helping society adapt to the effects of climate change through reliable HAB monitoring and early detection.

Plankton data sets are usually imbalanced and reflect natural differences within the environment. For minority classes, there may not be enough data to properly represent this variability, preventing AI models from gaining a full understanding of these classes (Kerr et al., 2020). The present work employs state-of-the-art Deep Learning (DL) models to support HAB monitoring applications within the Atlantic area.

Materials and Methods

A unified benchmark database covering publicly available phytoplankton images has been built through a data integration pipeline (Guterres et al., 2021) considering target phytoplankton genera from Integrated Multi-Trophic Aquaculture (IMTA) farms from Brazil, South Africa and Scotland. Classic convolutional neural networks architectures are trained for phytoplankton classification. Best individual models serve as a baseline for investigating state-of-the-art methods for class imbalance classification of target phytoplankton organisms. The following approaches have been evaluated to best support climate resilient solutions on HAB monitoring:

- Two-phase Learning (2PL) combines the Random Under Sampling technique with transfer learning. The model is first pre-trained using threshold data and then fine-tuned using original unbalanced datasets (Buda et al., 2018).
- Dynamic Sampling (DS) dynamically changes the class distribution of the training samples. Initially, the number of samples of each class equals the average number of samples. For every other iteration, the number of samples of each class is calculated based on the F1-Score from the previous training round (Johnson et al., 2019).
- Threshold Moving may be implemented on already trained models to improve classification results. It adjusts the decision threshold of a classifier during the test phase. Considering neural networks estimate Bayesian *a posteriori probability*, the output *y* for class *i* implicitly corresponds to $□_{□}(□) = □(□|□) = \frac{□(□)\cdot □(□|□)}{□(□)}$ for a given datapoint *x*. The correct class probabilities can be obtained by dividing the network output for each class by its estimated prior probability (Buda et al., 2018).
- Deep collaborative models (ensemble) may harness the limited understanding of individual models to provide a collective and more accurate classification specially for minority classes. It is a heterogeneous ensemble of DL models which grants a substantial performance improvement regarding other state-of-the-art approaches (Buda et al., 2018).

Results and Discussions

MobileNetV2 was selected as baseline model for further DL modeling since they provided best results among other architectures (NasNet, Resnet and VGG16). They are also targeted towards embedded and resource constrained environments. Table 1 depicts performance results within state-of-the-art approaches for phytoplankton classification.

Table 1 - Performance results within state-of-the-art methods for classification of target phytoplankton genera in IMTA applications.

Method	Recall	Precision	F1-Score	Model Size
None (Baseline)	0.75	0.78	0.75	29.1MB
DS	0.88	0.77	0.82	23MB
2PL	0.84	0.91	0.87	54.8MB
Ensemble (DS + 2PL)	0.87	0.94	0.89	50MB
Threshold Moving (Ensemble)	0.88	0.94	0.91	50MB

All investigated methods have improved classification performance compared to the baseline architecture. Collaborative deep learning model showed promising results. It enabled the combination of other state-of-the-art approaches towards reliable phytoplankton and HAB monitoring. Threshold moving has provided outstanding performance compared to other investigated approaches.

Conclusions

The present work investigated state-of-the-art approaches to class imbalance classification, considering target phytoplankton organisms within the Atlantic area. Deep collaborative models and threshold moving may be key methods towards climate resilient solutions for HAB monitoring since they can be employed upon latest DL models and architectures.

Acknowledgements

This work is 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 Nº 863034.

References

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

Guterres, B., Khalid, S., Pias, M., Botelho, S., 2021. 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

Buda, M., Maki, A., & Mazurowski, M. A. (2018). A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, *106*, 249-259.

Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. *Journal of Big Data*, *6*(1), 1-54.