# Working in a new era of automated classification in micro- (phyto)plankton monitoring: a FlowCAM case study

Lagaisse Rune[1], Dillen Nick[1,2], Deneudt Klaas[1] and Hablützel Pascal[1]

1 Flanders Marine Institute (VLIZ), Wandelaarkaai 7, 8400 Oostende, Belgium
E-mail: rune.lagaisse@vliz.be

2 Protistology and Aquatic Ecology Research Group (PAE), Department of Biology, Ghent University, Krijgslaan 281/S8, 9000 Gent, Belgium

Since 2017, VLIZ monitors micro- and phytoplankton in the Belgian Part of the North Sea within the LifeWatch framework. Monthly campaigns to onshore locations and seasonal campaigns to more offshore locations, lead up to around 140 samples per year. To deal with analyzing such large amounts of samples, VLIZ uses a FlowCAM device. This automated imaging device combines the principles of flow cytometry and microscopy to take traditional particle counting to the next level. As sample material is pipetted in the device, particles are brought in a constant fluid stream by a dosage pump and tubing system. Particles are then pulled through a photo chamber where each particle is captured in a unique image. In this way, the FlowCAM can create an image library of a sample in under 30 minutes. On a yearly basis, this yields several hundreds of thousands of images, all of which have to be identified by a taxonomist. Needless to say, this takes up a huge amount of time, not only hindering fast data releases to the public but also counteracting the fast and automated lab work with the FlowCAM device. To deal with this issue, we explored the use of deep learning approaches to analyze these image libraries.

In 2019, a prototype classifier was explored and implemented in the data-pipeline. Over the years, improved model versions were trained on yearly gathered FlowCAM data and all model predictions were verified, and corrected when needed, by a taxonomist. This fine-tuning has led to a current classifier with an accuracy around 88% and that can distinguish between 90 different microplankton classes, mainly belonging to diatoms, dinoflagellates and ciliates. So far, this automated approach to micro- (phyto)plankton monitoring has yielded over 2 million validated images, covering the full spatial scale of the Belgian Part of the North Sea at a monthly time interval[1].

This large set of validated data with model predictions allows us to take a next step in the automated image classification, where we explore the use of different measures indicating the quality of predictions and set thresholds for valuable observations. Currently, a first set of values explored are precision and recall. Precision is the proportion of model predictions that were correct for a model class. Recall is the proportion of true images of a class, according to a taxonomist, that the model also was able to retrieve, i.e. how many of the images of a class could the model recognize? Combining measures like these, and aiming for thresholds that indicate high quality predictions, can help us identify chunks of data that can be excluded from the manual validation check. This promising new approach would allow images with high quality predictions to be made available to the public within just a few days after laboratory processing, facilitating faster data releases while also providing a quality score for each validation to the user. Images that don't meet the required thresholds and still need a manual validation check can either be made available later, or immediately with a flag indicating the poor quality of the prediction. This new analysis allows us to evaluate different model versions used over the years and their performance, set standards for qualitative model predictions, and further automate the monitoring pipeline, which is crucial in the long-term context of the monitoring of micro- (phyto)plankton.

Reference

1 https://rshiny.lifewatch.be/flowcam-data/

Keywords: Phytoplankton; Microplankton; CNN; Modeling; FlowCAM