ABSTRACT

*Plankton are vital components of aquatic ecosystems, playing an essential role in the global food chain and oxygen production. Manual plankton classification is time-consuming and prone to errors due to the high visual similarity among species. This study compares the performance of three Transformer-based models: Swin-T, DeiT, and ConvNeXt, in the task of classifying 24 species of crustacean plankton. The dataset was obtained from the EcoTaxa platform and processed through several preprocessing stages, including scenarios with and without oversampling techniques. The models were optimized using pre-trained weights from ImageNet, parameter fine-tuning, and the AdamW optimizer. The results show that ConvNeXt achieved the best performance with a macro F1-score of 0.9231 using oversampled data and a learning rate of 0.00005. Oversampling improved the performance of Swin-T and ConvNeXt, but decreased the performance of DeiT, indicating that global self-attention architectures are more vulnerable to duplicated data. In addition, the quantity and quality of images significantly affected classification performance. Classes with abun*dant *samples (Ostracoda) or high-resolution images (Mysida) achieved consistently high accuracy, while classes with high visual similarity (Eucalanidae) were more likely to be misclassified.*

Keywords: *plankton classification, Transformer architecture, oversampling, Swin Transformer, DeiT, ConvNeXt.*