

ABSTRAK

Klasifikasi spesies plankton secara otomatis merupakan tantangan dalam bidang oseanografi dan ekologi laut, terutama karena keberagaman bentuk morfologi serta keterbatasan jumlah data yang tersedia. Masalah utama dalam penelitian ini adalah bagaimana memperoleh representasi numerik yang akurat dari citra plankton serta bagaimana meningkatkan akurasi dan stabilitas klasifikasi dengan menggunakan pendekatan *ensemble learning*. Tujuan dari penelitian ini adalah mengoptimalkan proses segmentasi dan ekstraksi fitur citra plankton, menerapkan metode ensemble learning yang menggabungkan *Support Vector Machine* (SVM) dan *Random Forest* (RF), serta mengevaluasi performa klasifikasi menggunakan berbagai skema validasi yaitu: *train-test split*, *cross-validation*, dan *Grid Search CV*.

Penelitian ini menggunakan model U-Net untuk segmentasi citra plankton dan mengekstraksi fitur numerik berdasarkan karakteristik morfologi dan tekstur (GLCM). Selanjutnya, model klasifikasi dibangun dengan SVM, RF, dan *Voting Classifier*. Evaluasi dilakukan menggunakan dua skenario *train-test split* (80%-20% dan 95%-5%), serta validasi tambahan berupa 5-Fold dan 10-Fold *Cross-Validation*. Untuk optimasi *hyperparameter*, *Grid Search CV* diterapkan pada model SVM dan RF.

Hasil pengujian menunjukkan bahwa model *Voting Classifier* mencapai akurasi tertinggi sebesar 98,40% pada skenario 95%-5%, sementara *Random Forest* memperoleh akurasi tertinggi sebesar 99,20% namun menunjukkan penurunan stabilitas saat proporsi data berubah. Model SVM menunjukkan performa yang konsisten dengan akurasi 97,84% setelah dilakukan optimasi. Hasil penelitian ini menunjukkan pendekatan *ensemble learning* terbukti efektif dalam meningkatkan akurasi dan kestabilan klasifikasi plankton dibandingkan model individual.

Kata Kunci: Klasifikasi Plankton, *Ensemble Learning*, SVM, *Random Forest*, *Voting Classifier*, Segmentasi Citra, *Grid Search CV*.

ABSTRACT

Automatic classification of plankton species poses a significant challenge in the fields of oceanography and marine ecology, primarily due to the diverse morphological structures of plankton and the limited availability of labeled datasets. The main problem addressed in this research is how to obtain accurate numerical representations from plankton images and how to improve classification accuracy and stability using ensemble learning approaches. The objectives of this study are to optimize the segmentation and feature extraction process of plankton images, to implement an ensemble learning method that combines Support Vector Machine (SVM) and Random Forest (RF), and to evaluate the classification performance using various validation schemes such as train-test split, cross-validation, and Grid Search CV.

This study employs the U-Net model for image segmentation and extracts numerical features based on morphological and texture characteristics (GLCM). Classification models are built using SVM, RF, and a Voting Classifier. Evaluation is performed using two train-test split scenarios (80%-20% and 95%-5%) and additional validation through 5-Fold and 10-Fold Cross-Validation. Hyperparameter tuning is applied using Grid Search CV for both SVM and RF. Experimental results show that the Voting Classifier achieves the highest accuracy of 98.40% in the 95%-5% train-test split scenario, while Random Forest reaches a peak accuracy of 99.20% but suffers from reduced stability when the test data proportion increases. SVM demonstrates consistent performance with an improved accuracy of 97.84% after hyperparameter tuning. The results of this study show that the ensemble learning approach has proven effective in increasing the accuracy and stability of plankton classification compared to individual models.

Keywords: *Plankton Classification, Ensemble Learning, SVM, Random Forest, Voting Classifier, Image Segmentation, Grid Search CV.*