ABSTRACT

MODELING DIABETES DISEASE CLASSIFICATION USING RANDOM FOREST WITH GREY WOLF OPTIMIZATION

By

Fiqi Nur Rohman

NIM: 24010122410001

Master's Program in Mathematics

Faculty of Science and Mathematics

Diponegoro University

Diabetes mellitus is one of the most alarming global chronic diseases, characterized by its ability to cause various serious complications such as blindness, kidney failure, heart attacks, and death. Although early detection is crucial for preventing or delaying complications, conventional diagnostic methods are often time-consuming and require clinical visits, making many patients reluctant to undergo them. The development of machine learning technology offers a promising solution to accelerate the detection and prediction of diabetes through intelligent and automated classification models. The Random Forest (RF) algorithm shows high potential for classification tasks, yet its performance is critically dependent on hyperparameter optimization. This study aims to enhance the classification performance of the RF model for diabetes classification prediction by optimizing its hyperparameters using the Grey Wolf Optimization (GWO) algorithm. This research utilizes the PIMA Indians Diabetes dataset (N=768). The data underwent a preprocessing stage that included handling illogical zero values, outlier management with Robust Scaling, data normalization, and class balancing on the training set using SMOTE (Synthetic Minority Over-sampling Technique). The performance of the baseline RF model was compared to the GWO-optimized RF model using K-Fold Cross-Validation (K=10). The evaluation metrics used were Accuracy, Precision, Recall, F1-Score, and AUC (Area Under the Curve). The results showed that the baseline RF model achieved an accuracy of 76.19% and an AUC of 0.88. After optimization with GWO, the RF model successfully achieved an accuracy of 82.53%, a precision of 80.0%, and an AUC of 0.93. These results demonstrate that hyperparameter optimization using GWO significantly enhances the accuracy and reliability of the Random Forest model for diabetes disease classification.

Keywords*: Diabetes Classification, Random Forest, Grey Wolf Optimization, Hyperparameter Optimization, Machine Learning*