ABSTRACT

This study examines the implementation of Vision Transformer (ViT) architecture for disease classification in cassava leaves with five classes: Cassava Bacterial Blight, Cassava Brown Streak Disease, Cassava Green Mottle, Cassava Mosaic Disease, and healthy leaves (Healthy). The secondary dataset consists of 21,397 images from the Kaggle platform, which were collected and processed through balancing stages (oversampling and undersampling) so that each class has 3,000 samples. The ViT model with token embedding and self-attention is used as the main approach in visual detection. Experiments were conducted across 18 scenario configurations varying the number of epochs (50 and 70), optimizer (SGD, Adam, AdamW), batch size (8, 16, and 32), and data splitting ratio (60:20:20, 70:15:15, and 80:10:10). Performance evaluation was determined using precision, recall, and F1-score metrics. The experimental results showed that the best configuration was achieved in Scenario 18 (70 epochs, AdamW optimizer, batch size 32, ratio 80:10:10), with the highest F1-score of 86%, which balanced precision and recall. The AdamW and Adam optimizers outperform SGD, with F1-score performance ranging from 79% to 86%, while SGD ranges from 55% to 73%. A larger batch size (32) and a higher training data ratio (80%) can boost model performance, especially at larger epochs (70). These findings indicate that the right combination of hyperparameters is crucial in optimizing Vision Transformer models for cassava leaf disease classification. Specifically, the use of adaptive optimizers such as AdamW, more training data, and adequate training duration have been shown to significantly contribute to generalization and accuracy. The contributions of this research are expected to serve as a reference for the development of computer vision technology in the agricultural field, particularly ViT-based plant disease detection.

Keywords: *Cassava Plants, Classification, Vision Transformer*