Abstract

Banana production in Indonesia in 2022 reached 9.6 million tons of fruit. The conventional method used to determine the ripeness of bananas still relies on human visual perception by observing changes in the banana skin's color. However, this method of assessing banana ripeness has several drawbacks, such as being time-consuming and subjective, leading to varying judgments among individuals. Therefore, computer vision technology can offer an effective solution for automatically classifying the ripeness levels of bananas. This research employs the Vision Transformer (ViT) methodology to classify banana ripeness levels into four classes: unripe, semi-ripe, ripe, and overripe. The study utilizes five pre-trained ViT models: ViT-B-P16, ViT-B-P32, ViT-L-P16, ViT-L-P32, and ViT-H-P14, pre-trained on ImageNet-21k and ImageNet-1k. Subsequently, these ViT models are evaluated and compared with convolutional neural network (CNN) models. The evaluation is performed using a test dataset comprising 5,068 banana images aggregated from publicly available online datasets. The evaluation results reveal that the ViT-L-P16-in21k model achieves the highest accuracy at 91.61%. ViT models demonstrate superior generalization capabilities, while CNN models exhibit more efficient model sizes and training times.

Keywords: banana ripeness classification, computer vision, vision transformer, pre-trained model