Generative Adversarial Network-based Data Augmentation for Papaya Disease Detection using Deep Learning

K. Kosala[1*], K. Banujan[1], B.T.G.S. Kumara[1], and T. Kovarthanan[2]

[1]Department of Computing and Information Systems, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka, Belihuloya, Sri Lanka.
[2]Edith Cowan University, Melbourne VIC 3000, Australia.

[*]kosalakumara372@gmail.com

Fruit diseases are a severe problem for all farmers, especially for fruit growers. A major challenge for large farms is the spread of diseases that make the fruits unfit for consumption, which also significantly impacts on farmer's income. Farmers must detect disease early in its life cycle in order to prevent it from spreading. Traditional fruit disease detection and identification rely on a person's ability to see the diseased fruit. Even though this approach sufficiently caters for small-scale farmers, it requires a high level of expertise to correctly identify the disease. Machine Learning and Image Processing techniques have been used in recent research to develop automated solutions for this problem. Papaya which is a popular fruit in Sri Lanka which is also having a high postharvest loss has been considered in this study. Among various papaya diseases, the most prevalent papaya diseases in Sri Lanka namely anthracnose, black spot, powdery mildew, phytophthora, and ringspot were selected. Data were collected from public image sources from the internet and from actual fields. VGG 16 as a Convolutional Neural Network technique was used to develop a computerized model for detecting papaya diseases. Literature reveals that many of the image-based disease recognition systems possess limitations due to insufficient data. Therefore, it is believed that novel data augmentation methods have promising advantages. In this approach, Deep Convolutional Generative Adversarial Network (DCGAN) was used to develop the data set. The VGG 16 model accuracy was found using the same pre-processed data set with and without being subjected to DCGAN. According to the results, the VGG 16 model showed a high accuracy for all diseases. Accuracy values for Anthracnose, black spot, powdery mildew, phytophthora, and ringspot were 90%, 85%, 70%, 65%, and 90% respectively. The results revealed that the proposed DCGAN model outperforms the basic data augmentation approaches.

Keywords: Artificial Neural Network, Source Code, Java Parser Library, Abstract Syntax Tree