Title: A comprehensive combined dataset on *Hibiscus and Tea plant leaf disease* images for classifications

Author Names: Md Masum Billah , Saifuddin Sagor , Mohammad Shorif Uddin

Abstract: In this study, we present a combined image dataset created from two distinct plant species: Hibiscus and Tea leaf. The dataset consists of high-resolution images of leaves from both species, captured using a SONY α7 II DSLR camera and a OnePlus 7T lubricant Tea Leaf dataset includes images categorized into five disease classes: Algal Leaf Spot, Brown Blight, Grey Blight, Red Leaf Spot, and Healthy, while the Hibiscus Leaf dataset includes images labeled across eight conditions, including citrus spot, fungal infection, mild edge damage, and healthy foliage. To ensure balanced representation and address class imbalances, extensive data augmentation techniques—such as flipping, rotation, zooming, shifting, noise addition, and brightness adjustment—were applied, resulting in a total of 1,413 combined original images and 13,000 augmented images. The ConvNextTiny deep learning model was fine-tuned on this combined dataset to classify the various leaf conditions, achieving an overall accuracy of 96%. This demonstrates the model's robust performance and high discriminatory power across the diverse set of leaf diseases and conditions. This experiment highlights the utility of combining multiple plant species into a single dataset and utilizing a lightweight yet effective model like ConvNextTiny for plant disease classification. The resulting dataset, along with the model and training scripts, is publicly available to facilitate further research in plant pathology, computer vision, and smart farming applications, enabling more accurate and efficient early-stage disease detection for both Hibiscus and Tea plants.

Keywords: Artificial intelligence ,Image processing ,Computer vision, Data science, Machine learning, Hibiscus plant leaves, Tea plant leaves

https://doi.org/10.1016/j.dib.2025.112357

https://www.sciencedirect.com/science/article/pii/S2352340925010716#:~:text=The%20dataset%20consists%20of%20high,while%20the%20Hibiscus%20Leaf%20dataset