Herb-GANNet: Synthetic Data Generation through Conditional GANs for Improving Accuracy in Medicinal Leaf Classification

Smitha Reddy S[a] , Vaishnavi K[b] , Sapna R[c]*

[a]Department of Computer Science, School of Computer Science and Engineering and Information Science, Presidency University, Bengaluru, India

[b]Associate 1 software engineer, Capegemini, Bengaluru, India

[c]*Department of Information Technology, Manipal Institute of Technology Bengaluru, Manipal Academy of Higher Education, Manipal, India

Corresponding author: Sapna R

Email: sapna.r@manipal.edu

Abstract

Accurate classification of medicinal leaves is essential across various fields, including agriculture, Ayurveda, drug discovery, and biodiversity conservation. However, this task can be complex and time consuming for experts due to the complexity of plant morphology, limited public datasets, and inherent class imbalances among species. These issues not only hinder effective identification and utilization of medicinal plants but also impede research and development in related domains. This study explores the application of Conditional Generative Adversarial Networks (CGANs) to generate synthetic data aimed at improving medicinal leaf classification models. CGANs offer effective solution for augmenting datasets and addressing class imbalance issues. We employed a conditional Deep Convolution Generative Adversarial Network (cDCGAN) to produce 500 synthetic images for each of thirty different plant species. To evaluate the effectiveness of the generated data, we trained and evaluated three popular convolutional neural networks: ResNet-34, VGG-16, and EfficientNet-B1, on both the original and augmented datasets. Our results show that CGAN-generated data significantly improved the performance across all tested models. EfficientNet-B1 achieved the lowest test loss of 1.74% on the augmented dataset, while ResNet-34 exhibited the highest test accuracy of 98.26%. These findings indicate that cDCGANs can generate synthetic data that effectively mimics real images, leading to (1) larger training datasets, (2) reduced data collection cost, and (3) increased data diversity and model generalization by providing a broader range of training examples.

Keywords: medicinal leaf classification, data augmentation, generative adversarial networks (GAN), deep learning, image generation, classification, drug discovery