Comprehensive Analysis of Machine Learning Algorithms for Crop Disease Prediction

V. Praba* and K. Krishnaveni

Department of Computer Science,Sri S. Ramasamy Naidu Memorial College (Affiliated to Madurai Kamaraj University, Madurai), Sattur, Virudhunagar District, India

*Corresponding authors: praba@srnmcollege.ac.in, kkrishnaveni@srnmcollege.ac.in

ABSTRACT

The crop disease prediction plays a significant role for improving overall crop production in agriculture field. All prominent parts of cultivated plants get naturally caused by pests, insects, and pathogens, and if not promptly handled, they significantly reduce the yield. This research focuses on the application of five prominent machine learning algorithms: Support Vector Machines (SVM), Naïve Bayes (NB), Random Forest (RF), Logistic Regression (LR), and k-Nearest Neighbours (KNN) to the task of crop disease prediction. The process involves several critical steps in image pre-processing, feature extraction and classification. Initially, crop images are converted into grayscale images, and Gaussian Blur is applied to enhance image smoothness. Subsequently, the grayscale image is threshold using Otsu's algorithm to create a binary image. Morphological transformations are employed to close small holes in the foreground part of the binary image. After foreground detection, the original color image is reconstructed to obtain an RGB image of the segmented leaf. Shape features are extracted from the pre-processed image using contour analysis. Furthermore, the image is converted to the HSV color space to calculate the ratio of pixels with pixel intensity of the hue (H) channel between 30 and 70 to the total number of pixels in the channel. Mean and Standard Deviation values are computed for each channel in the RGB image. Additionally, Non-green regions are identified by subtracting the green color component from 1, and texture features are extracted using the Gray-Level Co-Occurrence Matrix (GLCM) method. This fusion of texture and shape features significantly enhances the algorithms' ability to discriminate between healthy and diseased crops. The performance of these machine learning models is rigorously evaluated using a battery of metrics, including accuracy, precision, recall and F1-score. After extensive evaluation process, Random Forest emerges as the optimal choice among the considered models.

Keywords: Crop Disease prediction, Gray-Level Co-Occurrence Matrix (GLCM), Random Forest

DOI: 10.21467/abstracts.158 ISBN: 978-81-965621-0-6 (eBook)