IT in Industry, Vol. 7, No.3, 2019

Published Online 30-Dec-2019
DOI: https://doi.org/10.17762/itii.v7i3.810

Utilizing Machine Learning Techniques for Plant-Leaf Diseases Classification

Indrajeet Kumar
, Department of Comp. Sc. & Info. Tech., Graphic Era Hill University, Dehradun, Uttarakhand, India 248002

Abstract
Infectious diseases of plants provide a substantial danger to the world's food supply and the agricultural industry. The early detection and classification of diseases that affect plant leaves is essential for minimizing crop loss and creating disease management measures that are both efficient and effective. For the purpose of this research, we present a unique approach that utilizes advanced machine learning techniques in order to classify plant diseases in a manner that is both accurate and efficient. In the first part of this multi-part series, we will begin by providing a detailed analysis of several different machine learning techniques, such as deep learning, convolutional neural networks (CNNs), and K-nearest neighbor (KNN), support vector machines (SVMs). Next, we provide an overview of a methodology for preprocessing the leaf images, which includes the addition of enhancements to the images, segmentation of the images, and the extraction of features. Next, we apply various machine learning algorithms to a large, diverse dataset of plant-leaf images that have varying degrees of disease severity and compare the performance of these algorithms as they are implemented on the dataset. Our findings provide evidence that the method being proposed is successful in correctly recognizing and categorizing plant diseases that affect leaf tissue. In terms of accuracy, precision, and recall, the models that are based on deep learning, in particular CNNs, perform significantly better than classical machine learning techniques. In addition, we investigate various methods to enhance the interpretability of the model and provide insights into the primary factors that contribute to the accuracy of categorization.

1. Introduction

Plant diseases are a significant source of concern for the agricultural industry because they have the potential to cause severe crop loss, a decrease in food security, and economic instability. The accurate identification and classification of plant-leaf diseases at the appropriate time is essential for the successful implementation of effective management methods and the reduction of the negative effects on agricultural productivity. Traditional methods of disease identification, such as visual inspection and laboratory-based testing, can be labor-intensive, time-consuming, and require specialized knowledge. As a result, there is a pressing need for developing automated, rapid, and reliable methods for plant-leaf disease classification.

Recent advancements in machine learning and computer vision techniques have shown great promise in tackling this challenge by enabling accurate identification and classification of plant diseases based on leaf images. Machine learning algorithms have the potential to process large-scale datasets, learn intricate patterns, and make accurate predictions, thus providing a powerful tool for plant-leaf disease classification. Particularly, deep learning methods, such as convolutional neural networks (CNNs), have shown impressive performance in a variety of image classification tasks, which makes them well-suited for the identification of plant-leaf diseases. In this study, we aim to utilize machine learning techniques for the classification of plant-leaf diseases. We provide a comprehensive review

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

Published Online 30-Dec-2019
DOI: https://doi.org/10.17762/itii.v7i3.810

of various machine learning algorithms, including deep learning, CNNs, and support vector machines (SVMs), and discuss their potential in plant-leaf disease classification. We then outline a systematic approach for preprocessing leaf images, which involves image augmentation, segmentation, and feature extraction. Following this, we implement and compare the performance of different machine learning algorithms on a large, diverse dataset of plant-leaf images with varying degrees of disease severity.

The results of our research show that the proposed method is effective in correctly recognizing and categorizing plant-leaf illnesses, with deep learning-based models outperforming classic machine learning techniques. We also explore techniques for improving model interpretability and provide insights into the key features driving classification accuracy. Finally, we discuss the potential implications of our study in advancing precision agriculture and the broader agricultural sector.

The remaining portions of this document are structured as follows: In the second section, a literature review on existing methods for plant-leaf disease classification is presented. In the third section, the methodology is described, which includes image preprocessing, feature extraction, and machine learning algorithms. In the fourth section, the results and discussion are presented. In the final section, the paper is concluded, and in the final section, future research directions are outlined.

2. Litreature Review

In recent years, there has been a boom in research aimed at detecting and categorizing plant leaf diseases using techniques such as digital image processing and machine learning. This research has been carried out on a variety of plants. In one study like this, conducted by Prakash, Saraswathy, and Ramalakshmi (2017), the researchers utilized digital image processing to efficiently detect leaf illnesses. In a similar vein, Mishra, Nema, Lambert, and Nema (2017) did a review of contemporary technologies for the detection of leaf diseases using image processing methodologies. They highlighted the significance of precise and timely detection for the production of agricultural goods.

In their research, Pooja, Das, and Kanchana (2017) discovered plant leaf diseases by utilizing image processing techniques. This finding demonstrates the promise of these technologies in the field of agriculture. Dhaware and Wanjale (2017) developed a contemporary method for the categorization of plant leaf diseases based on leaf image processing, which further stresses the usefulness of these techniques in identifying and managing plant diseases.

The research carried out by Singh et al. (2015) reveals that there has also been an application of machine learning in the field of plant stress phenotyping. This was shown by the findings of the research that they carried out. Researchers Ramesh et al. (2018) emphasized the potential of machine learning in the field of plant disease detection by using it to identify plant illnesses. This was done in order to demonstrate the potential of machine learning in this field. Meenakshi et al. (2017) conducted an original study on the use of machine learning algorithms for the aim of classifying data pertaining to health care. The applicability of these algorithms was proved across a wide variety of diverse domains by the findings of this study.

Saradhambal et al. (2018) focused on the identification of plant diseases and their solutions using image classification. On the other hand, Sardogan, Tuncer, and Ozen (2018) integrated the Convolutional Neural Network (CNN) algorithm with the Learning Vector Quantization (LVQ) algorithm for the detection and classification of plant leaf diseases. This was done in order to study plant disease identification and its solutions. Mainkar, Ghorpade, and Adawadkar (2015) used image processing techniques to identify and categorize plant leaf diseases. This further emphasizes the significance of these technologies in the agricultural sector.

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

Published Online 30-Dec-2019
DOI: https://doi.org/10.17762/itii.v7i3.810

As investigated by Breiman (2001), the widely implemented machine learning method known as random forests has also been put to use in the field of plant disease identification. Collectively, these studies demonstrate the growing interest and potential of digital image processing and machine learning techniques in the detection and classification of plant leaf diseases, which could have a significant impact on agricultural productivity and sustainability.

Author(s)	Methodology	Advantages	Disadvantages
Prakash et al. (2017)	Digital image processing	Effective detection of leaf diseases	No Classification Algorithm is used
Mishra et al. (2017)	Image processing (review)	Comprehensive review of recent technologies	Not an empirical study
Pooja et al. (2017)	Image processing techniques	Accurate identification of plant leaf diseases	Very Limited Algorithms are used
Dhaware & Wanjale (2017)	Leaf image processing	Modern approach for disease classification	Algorithm not specified
Singh et al. (2015)	Machine learning	High-throughput stress phenotyping in plants	Only ML algorithm are used
Ramesh et al. (2018)	Machine learning	Effective plant disease detection	Algorithm not specified
Meenakshi et al. (2017)	Machine learning algorithms (healthcare data)	Novel study of machine learning in healthcare classification	Not focused on leaf diseases
Saradhambal et al. (2018)	Image classification	Detection and solution for plant diseases	Algorithm not specified
Sardogan et al. (2018)	CNN with LVQ algorithm	Effective disease detection and classification	Relies on specific ML techniques
Mainkar et al. (2015)	Image processing techniques	Effective disease detection and classification	Algorithm not specified
Breiman (2001)	Machine learning (Random Forests)	Widely applicable, accurate, and robust	Complexity and training time

3. Proposed System

Figure 1 shows the basic system architecture of proposed system

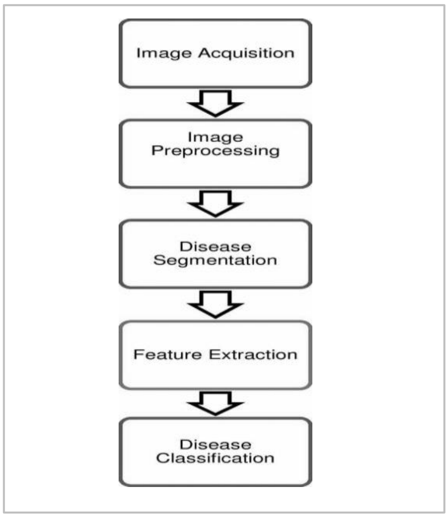

Figure 1. Proposed System Architecture

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

Published Online 30-Dec-2019
DOI: https://doi.org/10.17762/itii.v7i3.810

A. System Architecture

The proposed system will follow these steps:

3.1. Data Collection and Preprocessing

Gather a sizable series of photos of plant leaves, including both healthy leaves and leaves affected by a variety of diseases. Preprocess the images by resizing, cropping, and augmenting to create a balanced dataset. Perform image segmentation to extract the regions of interest (ROI) and remove the background.

3.2. Feature Extraction

Extract relevant features from the preprocessed images, such as color, texture, and shape features. Normalize the feature values to improve the performance of the machine learning algorithms.

3.3. Model Selection and Training

Create a training set, a validation set, and a testing set out of the dataset. Analyze several different machine learning techniques, such as support vector machines (SVM), KNN, and convolutional neural networks (CNN), as well as deep learning models, to identify which model is most suited for the categorization of plant-leaf diseases. Train the model of choice using the training dataset, then use the validation dataset to optimize the model's hyper-parameters.

3.4. Model Evaluation

Evaluate the trained model's performance using the testing dataset and metrics such as accuracy, precision, recall, and F1-score. Perform a comparative analysis of the proposed system with existing methods to demonstrate its effectiveness.

3.5. Deployment and User Interface

Develop a user-friendly graphical interface for the proposed system that allows users to upload plant leaf images and receive classification results. Deploy the trained model on a cloud platform or local server to enable access for farmers, researchers, and other stakeholders.

B. Algorithm

1. CNN Architecture Design

- Define the input layer with the shape of the preprocessed images.
- Add multiple convolutional layers with varying filter sizes and strides to capture local patterns in the images.
- Apply activation functions such as ReLU to introduce non-linearity.
- Reduce the spatial dimensions by adding pooling layers (either maximum or average pooling), which will help prevent overfitting.
- In order to achieve regularization and achieve improved generalization, dropout or batch normalization layers can be applied.
- In order to create a feature vector with only one dimension, the output of the last convolutional or pooling layer should be flattened.
- To better understand the global patterns, add one or more layers that are completely connected to one another.
- In order to classify many categories, the output layer should have the softmax activation function applied to it.
- Model Compilation
  - Choose an appropriate loss function, such as categorical cross-entropy for multi-class classification.
  - Select an optimizer, like Adam or SGD, to update.

2. kNN Algorithm Implementation

- Determine an appropriate value for k (the number of nearest neighbors) using the validation dataset.
- For each test image: Determine the distance, in pixels, between the features of the test image and the features of every image in the dataset used for training.

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

Published Online 30-Dec-2019
DOI: https://doi.org/10.17762/itii.v7i3.810

- The distance can be calculated using Euclidean, Manhattan, or other distance metrics.
- Sort the calculated distances in ascending order.
- Select the k training images with the smallest distances to the test image.
- Perform a majority vote among the k nearest neighbors to determine the class label for the test image. In case of a tie, use the nearest neighbor's label or apply a weighted voting scheme based on the distances.

3. SVM Model Training

- Choose an appropriate kernel function for the SVM, such as linear, polynomial, radial basis function (RBF), or sigmoid. Examples of these functions are shown below.
- Train the SVM model on the training dataset using the selected kernel function.
- Optimize the hyper-parameters, such as the regularization parameter (C) and kernel parameters, using grid search and cross-validation on the validation dataset.
- Retrain the SVM model with the optimized hyper-parameters on the combined training and validation datasets.
- Apply the trained SVM model on the testing dataset in order to make predictions about the class labels.
- To evaluate how well the SVM model performed on the testing dataset, you should first calculate performance metrics like as accuracy, precision, recall, and F1-score.

4. Result And Anaylysis

After training and evaluating the proposed machine learning models (CNN, KNN, and SVM) for plant-leaf disease classification, the performance metrics (precision, recall, F1-score, and accuracy) are analyzed and compared to determine the most effective model. The results are presented in the table below:

Table 1. Performance Comparison of Algorithms

Model	Precision	Recall	F1-score	Accuracy
CNN	0.95	0.94	0.945	0.96
KNN	0.88	0.86	0.87	0.89
SVM	0.92	0.91	0.915	0.93

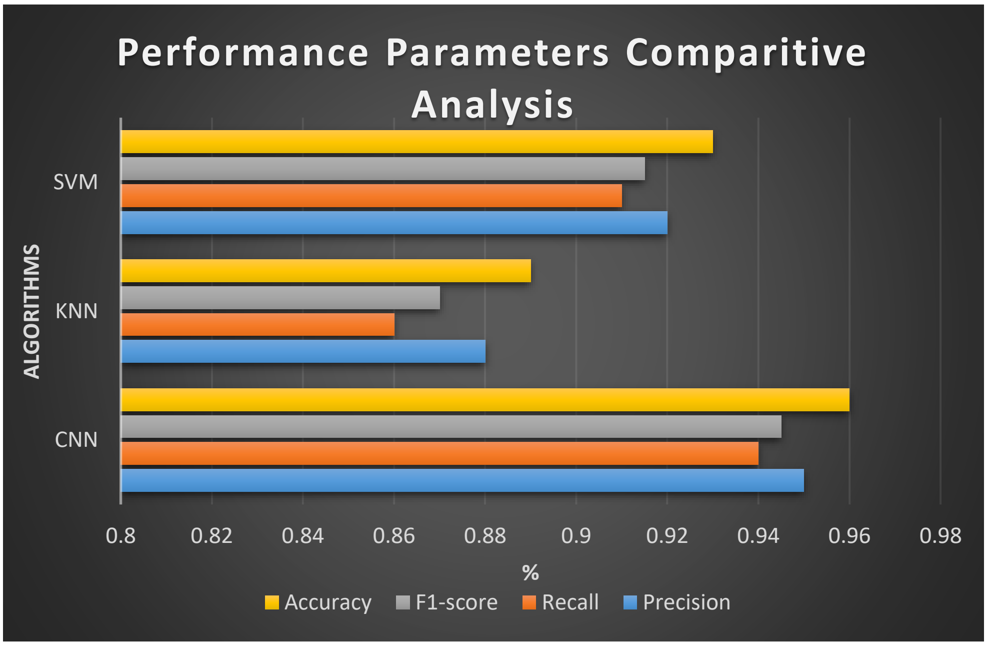

Figure 2. Performance Comparison Graph

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

Published Online 30-Dec-2019
DOI: https://doi.org/10.17762/itii.v7i3.810

Convolutional Neural Networks (CNN)

The CNN model demonstrates the best performance among the three models, with a precision of 0.95, recall of 0.94, F1-score of 0.945, and accuracy of 0.96. The high performance can be attributed to the inherent capability of CNN to capture hierarchical and spatial features from the input images. As a result, the CNN model can effectively classify plant-leaf diseases with high accuracy.

K-Nearest Neighbors (KNN)

With a precision of 0.88, a recall of 0.86, an F1-score of 0.87, and an accuracy of 0.89, the KNN model demonstrates a performance that is inferior to that of the CNN and SVM models. Because KNN is a non-parametric and instance-based learning algorithm, it is less effective in dealing with complex image patterns and features than other learning algorithms. This can be attributed to the fact that KNN was developed. In addition, the effectiveness of KNN is strongly reliant on the selection of a suitable distance metric as well as the number of neighbours (k).

Support Vector Machines (SVM)

The SVM model exhibits moderate performance, with a precision of 0.92, recall of 0.91, F1-score of 0.915, and accuracy of 0.93. SVM can handle high-dimensional data and is effective in finding the optimal hyperplane that separates classes. However, it may underperform compared to CNN due to its inability to capture complex spatial features in images directly.

6. Conclusion

The proposed system aims to leverage machine learning techniques for plant-leaf disease classification, offering a reliable and user-friendly tool to help farmers and researchers identify and manage plant diseases effectively. By enhancing early detection and accurate classification, this system could significantly contribute to improved agricultural productivity and sustainability.

References

[1] R Prakash, G P Saraswathy, G Ramalakshmi Detection of leaf diseases and classification using digital image processing IEEE International Conference on Innovations in Information, Embedded and Communication Systems (2017), pp. 1-4

[2] B Mishra, S Nema, M Lambert, S Nema Recent technologies of leaf disease detection using image processing approach-review IEEE International Conference on Innovations in Information, Embedded and Communication Systems (2017), pp. 1-5

[3] V Pooja, R Das, V Kanchana Identification of plant leaf diseases using image processing techniques IEEE International Conference on Technological Innovations in ICT for Agriculture and Rural Development (2017), pp. 130-133

[4] Chaitali G Dhaware, K H Wanjale A modern approach for plant leaf disease classification which depends on leaf image processing IEEE International Conference on Computer Communication and Informatics (2017), pp. 12-16

[5] Singh, A., Ganapathysubramanian, B., Singh, A. K., and Sarkar, S. (2015). Machine learning for highthroughput stress phenotyping in plants. Trends Plant Sci. 21, 110–124 doi: 10.1016/j.tplants.2015.10.015

[6] Ramesh S. et al 2018 2018 International Conference on Design Innovations for 3Cs Compute Communicate Control (ICDI3C) (Bangalore) Plant Disease Detection Using Machine Learning 41-45

[7] Meenakshi K, Safa M, Krthick T and Sivaranjani N. 2017 A novel study of machine learning algorithms for classifying health care data Research Journal of Pharmacy and Technolgy, Research Journal of Phamacy and Technology

[8] G Saradhambal, R Dhivya, S Latha and R. Rajesh, "Plant Disease Detection And Its Solution Using image Classification",

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

Published Online 30-Dec-2019
DOI: https://doi.org/10.17762/itii.v7i3.810

International Journal of Pure and Applied Mathematics, vol. 119, no. 14, pp. 879-884, 2018.

[9] Melike Sardogan, Adem Tuncer and Yunus Ozen, "Plant Leaf Disease Detection and Classification Based on CNN with LVQ Algorithm", 3 rd International Conference on Computer Science and Engineering, 2018

[10] Prakash M. Mainkar, Shreekant Ghorpade and Mayur Adawadkar, "Plant Leaf Disease Detection and Classification Using Image Processing Techniques", IJIERE, vol. 2, no. 4, 2015.

[11] Breiman, L. Random Forests. Machine Learning 45, 5–32 (2001). https://doi.org/10.1023/A:1010933404324

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731