A Review on Feature Extraction Techniques for Plant Disease Classification

M.M.Gunarathna
Department of Computing and Information Systems
Sabaragamuwa University of Sri Lanka
Belihuloya, Sri Lanka
mmgunarathna95@gmail.com

R.M.K.T. Rathmayaa
Department of Physical Sciences and Technology
Sabaragamuwa University of Sri Lanka
Belihuloya, Sri Lanka
kapila.tr@gmail.com

***Abstract* — There are many advances in computer vision that helps to identify and classify plant diseases automatically. A classifier will detect the plant diseases as healthy and diseased with the given features as input in automatic detection. So, feature extraction plays a vital role in the identification of a disease. The main objective of this study is to review commonly used feature extractors that are more suitable for plant disease classification. This review has been carried out by dividing into three main sub-parts, namely planning, conducting and reporting. In the planning phase, a search strategy was defined with search terms, inclusion and exclusion criteria to identify relevant publications. In the next stage, the defined strategy was followed to identify the relevant published papers and features engineering techniques used in each study were identified. Hand-crafted feature engineering and deep learning feature extraction are the two main types of feature extraction methods reviewed and analyzed. In the last stage, all the findings were summarized and documented. With regards to the results obtained from reviewed studies, usually hand-crafted feature engineering techniques require human experts, although they display good results in identifying plant disease. However, by using this method, unwanted features can be skipped, or redundant features can be removed. Deep learning methods consist of neural networks that do not require any human expert intervention to extract the features automatically. From this review, it is identified that Convolutional Neural Network (CNN) in the deep learning approach is the commonly used feature extractor in plant disease classification that provides a high classification rate with an average of 98%. In contrast, other hand-crafted feature engineering techniques rely on the features that they choose**

Keywords — Feature extraction, Deep Learning, Hand-crafted Features

I. INTRODUCTION

Plant Disease identification plays a vital role in every farming and gardening system. In general, a farmer recognizes symptoms of the diseases in a plant using naked eye observation, which requires continuous monitoring that is tedious and expensive. There are many advances in computer vision that helps to identify and classify plant diseases automatically[1]. A classifier will detect the plant diseases as healthy and diseased with the given features as input in automatic detection. Accuracy is the main parameter that every researcher used to calculate the performance of the model. The classifier's accuracy depends primarily on the features which are extracted[2]. So, feature extraction plays a vital role in the identification of a disease. Proper selection of the correct features results in high diagnostic accuracy. Hand-crafted feature engineering and deep learning feature extraction are the two main types of feature extraction methods reviewed and analyzed in this paper.

Hand-crafted features of the plant leaf image can be mainly divided into three types, shape, texture, and colour. Hence Bera et al. have selected color and shape of the diseased part as features to identify rice diseases. After feature extraction and selection, they have used different classification techniques for the disease's identification [3]. There are a specific set of machine learning algorithms to extract each feature. In [4] they mentioned that color, shape and texture features can be extracted with Grey Level Co-occurrence Matrix (GLCM) , Blend vision and machine intelligence. Monika Jhuria, Ashwani Kumar and Rushikesh Borse have used colour, morphology and feature vectors to extract features from the existing dataset. They divided the colour image processing into three main distinct areas. They extracted the morphology feature using the erosion concept. For the texture extraction, they have used Daubecies 2-D wavelet packet decomposition. They concluded that they had achieved better results for colour and morphology when compared to texture, as the diseases are defined better by these features from the dataset [5]. Namrata Ghatol1 and Dr G.P. Dhok have suggested an image processing technique to detect and classify crop diseases with a texture-based feature approach. They have used Gray Level Co-occurrence Matrix to extract texture-based features like energy, contrast, correlation, mean and Homogeneity [6]. Sandeep et al. have mentioned there are various methods of feature extraction techniques that were used to identify species according to specific characteristics of the leaves. The study depicts that most studies have used shape, colour and texture features in leaves. They concluded that although the hand-engineered feature extraction method is successful in many studies, the performance of those approaches mainly relies on the features that they choose. With different leaf data and feature extraction techniques, these hand-crafted features are subject to change, which confuses the search for an active subset of features to represent leaf samples in species recognition studies.

Deep learning feature extraction includes various types of neural networks that automatically identify features by learning and adjusting the appropriate weights. Lee et al. have done a review on how deep learning extracts and learns leaf features for plant classification. They have learned essential leaf features using Convolutional Neural Networks (CNN) directly from the raw representations of input leaves. CNN itself automatically detect the features that are important to classify the plants according to the diseases with the help of CNN filters. They have concluded that the results show that the extracting features based on CNN can provide better feature extraction compared to the use of hand-crafted features[7]. Xiaolong Zhu, Meng Zhu and Honge Ren have proposed an improved deep CNN for plant leaf recognition. They have used inception V2 as the feature extractor consists of a layer-by-layer structure rather than using a conventional

CNN. Feature maps generated in the pooling layers were used to identify the features. They have concluded that this approach gave better results when compared to traditional CNN[8].

II. Objectives

In this study a Systematic Literature Review (SLR) has been conducted to determine the relevant related publications, and those papers were later analyzed deeper to identify the commonly used feature extraction techniques. The main objective behind this paper is to present the best feature extractors that can be used in plant disease identification by analyzing different recent methods of extraction of features. The broad range of applications on the field of plant disease classification makes it difficult for anyone to follow all possible useful ideas present in the literature, leading to missing potential solutions to problems. In this study, an effort has been made to highlight the progress made so far in the feature extraction phase of plant disease detection. This is a modest contribution to help new researchers to obtain an overall picture of recent best feature extraction methods

III. Methodology

The review was carried out by dividing the whole process into three main sub-parts, namely planning, conducting and reporting. In the planning phase, a search strategy was created. According to the strategy, search terms, inclusion and exclusion criteria were identified and defined. Search terms were used to filter only the related studies from large databases

Table 1. Search Terms

No	Search Terms
S1	"Feature Extraction" OR "Feature Engineering"
S2	("Feature Extraction" OR "Feature Engineering") AND "Image Processing"
S3	"Hand-Craft" AND ("Feature Extraction" OR "Feature Engineering")
S4	"Deep Learning" AND ("Feature Extraction" OR "Feature Engineering")
S5	("Feature Extraction" OR "Feature Engineering") AND ("Plant Diseases" OR Leaf Diseases OR "Crop Diseases") AND ("Recognition" OR "Identification" OR "Detection" OR "Classification")

In the conducting phase, a hybrid forwards and backward approach was followed to classify primary studies. The collected research papers were divided into two main categories as papers that have followed deep learning feature extraction methods and hand-crafted feature engineering techniques. They were stored in a local database. Then the necessary techniques for each purpose were extracted by reviewing and comparing. Finally, as the last phase, a table which returns all the crucial aspects of the features, methods and the comments was created

A. *Hand-craft feature Engineering method*

The computation of hand-crafted features is a two-step process wherein the first step characteristics of the images are located, and then each key point is distinguished with the help of a classifier. Reviewed studies have used several classifiers such as Support Vector Machine () and K-Nearest Neighbor (KNN)

B. *Deep Learning feature Engineering method*

Deep Learning has the capacity to construct and extrapolate new features from raw representations of input data without specifically telling which features to use and how to extract.

Widely used Deep learning feature extraction method in the plant classification area is CNN [3]. It learns by extracting the basic features in the first layers and evolving to learn complex features of the image in the deeper layers, resulting in more accurate image classification[9].

Table 2. Summary of Techniques

Feature	Feature Extraction Method	Comments and Accuracy
Texture Features: Contrast, Homogeneity, Energy, Entropy, Variance, Cluster Shade, Cluster, Prominence	Grey-Level Co-Occurrence Matrix (GLCM)	* Capture properties of texture, but they are not directly useful for further analysis, such as comparing two textures. * Easy to implement *High dimensionality
	Gabor Filter	* Sensitive to a different orientation and scale
	Local binary pattern (LBP)	* Very Simple * Low computational cost * Sensitive to image rotation
Color Features: Mean, Standard Deviation, Color moments	Color co-occurrence Method	* Competitive computational cost * It avoids the use of weights to combine individual color and texture features
	Color Histogram	* Represents the frequency distribution of color bins in an image * Easy to compute * Insensitive to small variations Not robust
Shape Features: Area, Euler Number, Orientation, Extent, Perimeter, Convex area, Filled area, Eccentricity, Major axis length, equidiameter, and Minaxislength	Elliptic Fourier and discriminant analyses	* Highly dependent on the segmented result of leaf images.
	Geometrical calculation + Moment invariants	* Instabilities and noise sensitivity

IV. Results and Discussion

The history of plant identification methods extracted from the studies shows that existing plant disease classification solutions heavily depends on the experts' ability to encode domain knowledge. Many researchers have used hand-engineering methods for their characterization for many morphological features pre-defined by pathologists. Colour and texture features are extracted widely by the people rather than shape feature because it does not have any specific details to reflect the plant disease's image. Some researchers have also used deep learning models to extract features for plant disease classification. Deep learning models can be used for both classification and feature extraction. Convolutional Neural networks (CNN), Probabilistic Neural Network (PNN), Back Propagation Neural Network (BPNN), and Radial Basis Function Neural Network (RBFNN) are examples of those deep learning models. The performance of these models highly depends on the amount of training data that is used. So, these are computationally expensive compared to other hand-engineered feature extraction techniques. Commonly used feature extractor out of the other deep learning methods is CNN. It extracts features from images and manages the entire feature engineering portion. Beginning layers in the usual CNN architecture extract the low-level characteristics, and end-level layers extract high-level characteristics from the image. From the studies, we can identify that all the issues with normal classifiers are addressed by CNN, and the accuracy of the CNN is also higher in the plant disease classification study area. Average accuracy achieved from the CNN considering all the reviewed papers is 98%. Below table shows the average of accuracy obtained by reviewing the papers with same dataset.

Table 3. Summary of Findings

Feature Extraction	Classifier	Average Accuracy
GLCM	KNN	91.37%
	SVM	95.8%
LBP	KNN	90.1%
	SVM	92.3%
CNN		98%

V. CONCLUSION

This study presents the main feature extraction methods that may help engineers, students, and researchers in selecting the appropriate algorithms for their plant disease classification. The usage of the feature extraction methods usually depends on the requirement of the classifier. Each method has its drawbacks and advantages. Before the deep learning approach, people need to spend time selecting the necessary characteristics when classifying the images. Too many hand-crafted features are available (local feature, global feature), so it will take too much time to select the correct features for a solution (image classification) and select the correct classification model. With regard to the studies, we can conclude that the most suitable approach for plant disease classification is deep learning models because the manual extraction is a complex task, and the identification is sensitive. Deep learning models can be used for feature extraction as well as for classification. So, without much human expert intervention with higher accuracy, we can classify plant diseases efficiently and effectively. This review makes a significant contribution by providing an essential and up-to-date analysis of the previous attempts made in the domain of feature extraction.

REFERENCES

[1] R. Meena Prakash, G. P. Saraswathy, G. Ramalakshmi, K. H. Mangaleswari, and T. Kaviya, "Detection of leaf diseases and classification using digital image processing," *Proc. 2017 Int. Conf. Innov. Information, Embed. Commun. Syst. ICIIECS 2017*, vol. 2018-Janua, pp. 1–4, 2018, doi: 10.1109/ICIIECS.2017.8275915.

[2] G. Kambale and N. Bilgi, "A Survey Paper on Crop Disease Identification and Classification Using Pattern Recognition and Digital Image Processing Techniques," *IOSR J. Comput. Eng.*, no. Acbcda, pp. 14–17, 2017.

[3] T. Bera, A. Das, J. Sil, and A. K. Das, *A survey on rice plant disease identification using image processing and data mining techniques*, vol. 814. Springer Singapore, 2019.

[4] U. Shruthi, V. Nagaveni, and B. K. Raghavendra, "A Review on Machine Learning Classification Techniques for Plant Disease Detection," *2019 5th Int. Conf. Adv. Comput. Commun. Syst. ICACCS 2019*, pp. 281–284, 2019, doi: 10.1109/ICACCS.2019.8728415.

[5] M. Jhuria, A. Kumar, and R. Borse, "Image processing for smart farming: Detection of disease and fruit grading," *2013 IEEE 2nd Int. Conf. Image Inf. Process. IEEE ICIIP 2013*, pp. 521–526, 2013, doi: 10.1109/ICIIP.2013.6707647.

[6] N. Ghatol, "TEXTURE BASED FEATURES APPROACH FOR CROP DISEASES CLASSIFICATION AND DIAGNOSIS-A RESEARCH," no. May, pp. 7469–7474, 2019.

[7] S. H. Lee, C. S. Chan, S. J. Mayo, and P. Remagnino, "How deep learning extracts and learns leaf features for plant classification," *Pattern Recognit.*, vol. 71, pp. 1–13, 2017, doi: 10.1016/j.patcog.2017.05.015.

[8] X. Zhu, M. Zhu, and H. Ren, "Method of plant leaf recognition based on improved deep convolutional neural network," *Cogn. Syst. Res.*, vol. 52, pp. 223–233, 2018, doi: 10.1016/j.cogsys.2018.06.008.

[9] M. Jogin, Mohana, M. S. Madhulika, G. D. Divya, R. K. Meghana, and S. Apoorva, "Feature extraction using convolution neural networks (CNN) and deep learning," *2018 3rd IEEE Int. Conf. Recent Trends Electron. Inf. Commun. Technol. RTEICT 2018 - Proc.*, pp. 2319–2323, 2018, doi: 10.1109/RTEICT42901.2018.9012507