Leaf Image based Plant Disease Identification using Transfer Learning and Feature Fusion

Authors and Affiliations

Xijian Fan[1,†,*]

Peng Luo[2,3,†]

Yuen Mu[1]

Rui Zhou[1],

Tardi Tjahjadi[4],

Yi Ren[5]

[1] College of Information Science and Technology, Nanjing Forestry University, Nanjing, Jiangsu, 210037, China; xijian.fan@njfu.edu.cn, muyuan@njfu.edu.cn, rui.zhou@njfu.edu.cn

[2] Institute of Forest Resource Information Techniques, Chinese Academy of Forestry, Beijing 100091, China; lozpeng@ifrit.ac.cn,

[3] Key Laboratory of Forestry Remote Sensing and Information System, National Forestry and Grassland Administration, Beijing 100091, China

[4] School of Engineering, University of Warwick, Gibbet Hill Road, Coventry, CV4 7AL, UK; t.tjahjadi@warwick.ac.uk

[5] Academy of Forestry Inventory and Planning, National Forestry and Grassland Administration, Beijing, 100714

* Correspondence: xijian.fan@njfu.edu.cn;

† These authors contribute equally to this work.

Abstract

With the continuing changes in the structure of plant and cultivation patterns, new diseases are constantly appearing on the leaves of plant, exacerbating the threat to food security and agricultural production in many areas of the world. Thus, a rapid and accurate recognition of various diseases in plant will not only significantly reduce unnecessary planting costs, but also alleviate the economic losses and environmental pollution caused by incorrect disease diagnosis. Recent advances in deep learning have improved the performance in recognizing plant leaf diseases. In this paper, we present a general framework for recognizing plant diseases. Firstly, we propose a deep feature descriptor based on transfer learning to obtain a high-level latent feature representation. Then, we integrate the deep features with traditional handcrafted features by feature fusion to capture the local texture information in plant leaf images. In addition, centre loss is incorporated to further enhance the discriminative ability of the fused feature. The centre loss simultaneously minimizes intra-class distance and maximizes inter-class distance to learn both compact and separate features. Extensive experiments have been conducted on three publicly available datasets (two Apple Leaf datasets and one Coffee Leaf dataset) to validate the effectiveness of proposed method. The propose method achieves 99.79%, 92.59% and 97.12% classification accuracies on the three datasets, respectively. The experiment results demonstrate that the proposed method effectively captures the discriminative feature representation for distinguishing plant leaf diseases.

Keywords： plant disease; transfer learning; feature fusion; convolutional neural network

1. Introduction

In recent years, there has been an increased number of plant diseases, that heavily influenced the agricultural production and food security. Early detection of plant diseases is thus of great importance for full-blown disease prevention and plant treatment at a later stage. It also plays a vital role in the management and decision-making of agricultural production **(Liu et al., 2018)**. Disease-infected plants tend to show obvious marks or lesions on leaves, flowers or fruits. Generally, each disease presents a unique visible pattern that can be used to diagnose plant abnormalities. The leaves of plants are the primary source for identifying plant diseases, and most of the symptoms of diseases appear on the leaves **(Ebrahimi et al., 2017)**.

There are two main traditional methods for identifying diseases in a plant via its leaves: visual inspection of plant tissues by trained experts; and machine detection based on image processing **(Dutot et al., 2013)**. The visual inspection by agricultural and forestry experts requires observing the morphology of the leaf surface and analyzing the condition of the lesion one at a time to identify the diseased leaves. Such manual process is time consuming which leads to high cost and low efficiency. In addition, there exists the risk of errors due to the subjective perception in the process **(Mahlein et al., 2013; Yuan et al., 2014)**.

The state-of-the-art machine detection methods generally use low-level image processing techniques, i.e., noise removal, morphological operations and image enhancement, to pre-process the images of the diseased leaves **(Qin et al. 2016, Rumpf et al. 2010, Padol et al. 2016**). This is followed by applying handcrafted feature extraction techniques to capture low-level information of the leaves, e.g., colour, shape and texture. **Patil et al. (2017)** proposed a content-based image retrieval system that uses colour, shape and texture features of leaf for identifying diseased leaves of soybean.

Sandika et al. (2016) developed a method for disease identification of grape leaves with complex background. They compared the identification performance using different features, e.g., local binary pattern (LBP) feature, and some statistical features in RGB planes, and different machine learning algorithms, e.g., Support Vector Machine (SVM) and Random Forest. **Sharif et al (2018)** proposed a texture-feature based approach for identifying citrus fruit plant diseases. They employed a hybrid feature selection technique based on principal component analysis and feature statistics. A statistical method based on scale-invariant feature transform (SIFT) for the classification of plant diseases **(Hlaing et al. 2017)** reduces the computation cost in using SIFT feature. However, such handcraft techniques merely extract shallow feature representations, which fails to mine the inner relation information within the same disease classes. Moreover, such techniques require researchers to possess the relevant domain knowledge for obtaining good features, that cannot be generalized well in different environments.

In recent years, deep learning-based methods have made a significant advance in the field of computer vision **(Krizhevsky et al., 2012; He et al., 2016; Szegedy et al., 2015; Ye et al., 2019; Chen et al., 2014; Badrinarayanan et al., 2017, Wang et al. 2021)**. Due to its ability to capture meaningful feature representation, deep learning methods have also been applied to plant disease recognition and detection **(Sladojevic et al. 2016; Nachtigall et al., 2016; Mohanty et al. 2016; Jalal et al. 2020**; **Bi et al. 2020, Shrivastava et al. 2021a, Shrivastava et al. 2021b)**. **Sladojevic et al. (2016)** proposed a plant disease classification model using deep convolutional networks. This is the first deep learning-based plant disease recognition method. **Mohanty et al. (2016)** presented a Convolution Neural Network (CNN) based model for detecting 26 diseases and 14 crop species, achieving promising results. **Yu et al. (2020)** proposed the

104 attention mechanism to highlight the leaf area, which helps to capture more
105 discriminative feature in diseased leaves. **Tetila et al. (2019)** utilized data augmentation
106 and fine-tuning to train a deep network model to automatically identify soybean leaf
107 disease. **Yu et al. (2020)** developed a region-of-interest based two-stream network for
108 recognizing the apple leaf disease, enhancing the diseased areas of leaf while separating
109 the background. **Jalal et al. (2020)** applied deep neural network (DNN) to detect apple
110 leaf diseases by using SURF for feature extraction and an evolutionary algorithm for
111 feature optimization to create a plant disease detection system. **Bi et al. (2020)** proposed
112 a leaf disease model based on MobileNet model and compared it with ResNet152 and
113 InceptionV3 models. However, the recognition of leaf diseases remains difficult as the
114 following two exclusive key properties need to be considered: 1) Coarse local spatial
115 properties: most spots of leaf diseases are very small in size and tend to vary; and 2)
116 Fine-grained properties: some factors such as complex background and uneven
117 illumination resulted in low inter-class and high intra-class variation of diseased
118 samples.

119 Although deep learning methods have been shown to be very capable in depicting
120 both high-level and low-level features, they are less reliable than handcrafted features
121 in representing local spatial characteristic (Cai et al. 2018). Thus, to better capture local
122 characteristics that exclusively exist in plant leaf image, we propose to fuse the
123 handcrafted and deep features, with the handcrafted features complementing the deep
124 features. We further impose discriminative constraint to uncover the fine-grained
125 properties in diseased leaf image.

126 The main objectives of this study are to propose a more powerful and robust
127 framework to address the problems using deep learning methods to improve the
128 recognition performance of different plant diseases. The proposed framework combines

the deep learning features and handcrafted features based on feature fusion and transfer learning, where the handcrafted features effectively capture the coarse local information, as complementary to deep features. The combined features are capable of extracting not only the distinctive texture and shape patterns existing in diseased leaves, but also the high-level semantics corresponding to the leaves with different types of diseases. Furthermore, the study aims to further enhance the representation ability for different diseased plant leaves. To this end, the study adopted the auxiliary discriminative constraint, i.e., centre loss during the deep feature training, where the discriminative features are simultaneously learned under the joint supervision of softmax loss and centre loss.

In summary, the main contributions of this study are as follows:

1) We proposed to learn a deep model based on transfer learning, where an adapted CNN network is adopted to accelerate model training and extract the semantic information of different categories of diseased leaf images.

2) We constructed an integrated model by fusing deep features and traditional handcrafted feature, which effectively captures the local spatial texture information widely existing in plant leaf images.

3) We introduced the centre loss constraint to simultaneously optimize the learning process of the fused features, which further enhances the discriminative power of the extracted deep features.

2. Materials and Methods

The proposed method combines deep discriminative features with handcrafted feature for plant disease recognition which includes three main modules: preprocessing; feature extraction including fusion; and classification. Specifically, the proposed framework first extracts the deep features of plant disease samples based on transfer

learning, where the pre-trained CNN is utilized to accelerate the training process. To improve the discriminative ability, centre loss is introduced together with softmax loss to jointly supervise the learning of the deep feature. The framework then further fuses the handcrafted feature in the form of oriented histogram of gradient (HOG) and discriminative deeply learned features to obtain the final characteristic feature of a disease sample for plant disease recognition. Figure 1 shows the proposed method.

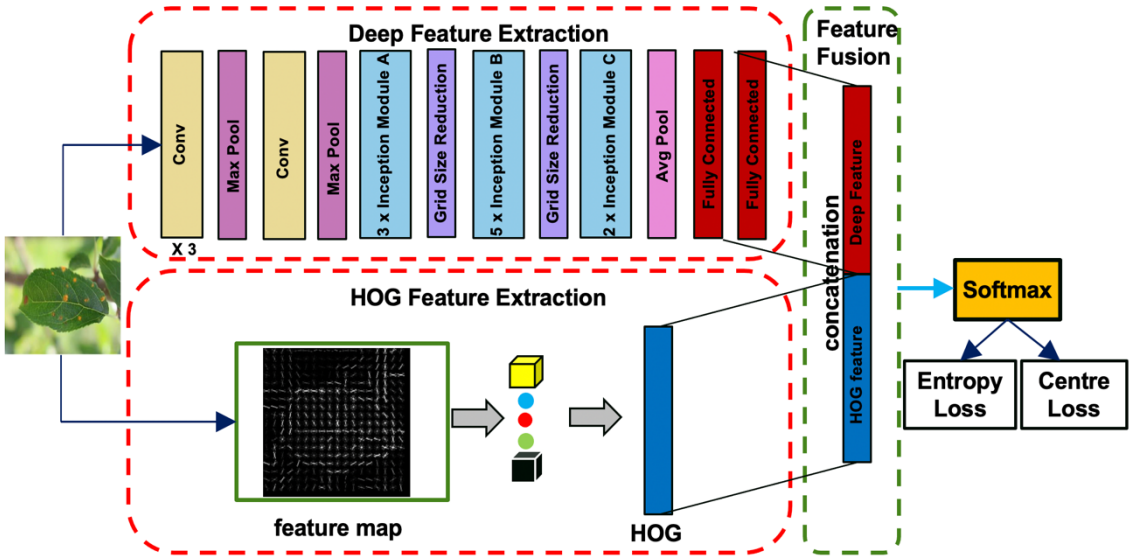

Figure 1. The proposed method for plant leaf disease recognition.

2.1. Deep Features Extraction

CNN is mainly composed of a convolutional layer, a pooling layer and a fully connection (FC) layer. An important research direction of CNN is to build a network model with deeper and more complex layers. However, with the increase in deep layer of a network model, the computational efficiency of the model reduces. In addition, there is limited availability in training samples for plant disease recognition. For these reasons, we employ transfer learning which uses a pre-trained CNN model to estimate the various semantics associated with different plant leaf diseases. InceptionV3 **(Wang et al 2012, Szegedy et al. 2016)** is selected as the backbone network of our transfer learning model due to its two advantages: 1) the use of convolution decomposition

reduces the amount of convolution computation; and 2) the use of the label smoothing regularization module avoids the occurrence of over-fitting.

The core idea of Inception based network model is factorizing convolutions by reducing the size of convolution kernel, thus increasing the computational efficiency without much or any loss in representational power. For example, let us consider the decomposition of a 5×5 convolution into two 3×3 convolutions (as illustrated in Figure 2). The number of parameters in the case of 5×5 convolution and 3×3 convolutions are respectively $5 \times 5 = 25$ and $3 \times 3 + 3 \times 3 = 18$. Thus, the number of parameters is reduced by 28%. Figure 3 shows the factorized Inception Block A.

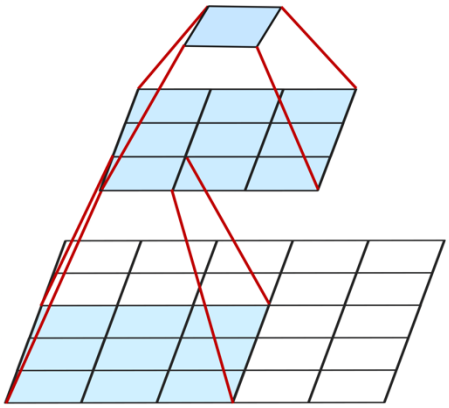

Figure 2. Use of two 3 × 3 convolutions to replace a 5 × 5 convolution.

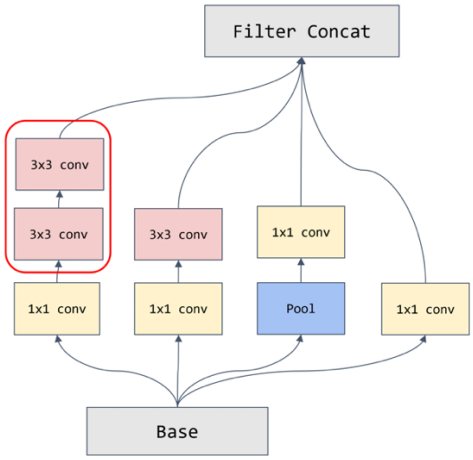

Figure 3. Transformation of a 5 × 5 convolution in Inception block into two 3 × 3 convolutions (enclosed within red rectangle).

In addition, a $n \times n$ convolution kernel in the InceptionV3 network can be decomposed into $1 \times n$ and $n \times 1$ convolution. For example, the 3×3 convolution of one is equivalent to the 1×3 convolution of one and then the 3×1 convolution of one. In the case of 3×3 convolution the number of parameters is $3 \times 3 = 9$, while in the case of convolution 1×3 and 3×1 the number of parameters is $1 \times 3 + 3 \times 1 = 6$. Thus, the number of parameters is 33% fewer than that of a single 3×3 convolution. The factorised Inception Block B is shown in Figure 4. However, this structure is not suitable for the earlier layers, but suitable only for middle-sized features of the middle layer (for features of size $m \times m$, the value of m is between $12 \sim 20$). Following another principle of CNN design that higher dimensional representations are easier to process locally within a network, the convolution can be decomposed into a sequence of expanded filter banks including 1×1, 1×3 and 3×3 convolutions. The factorized Inception Block C is shown in Figure 5. The overall architecture of InceptionV3 is shown in Figure 6.

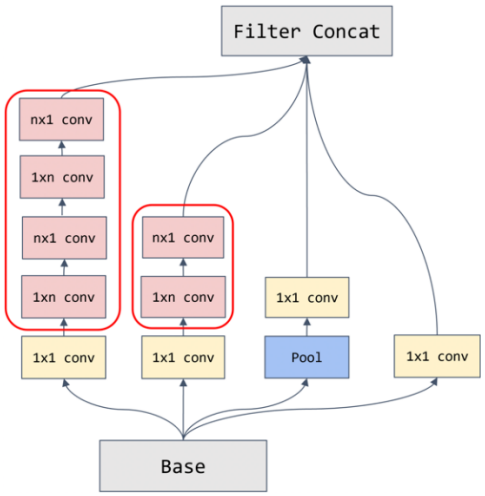

Figure 4. Transforming the $n \times n$ convolution in Module B into $1 \times n$ and $n \times 1$ convolutions.

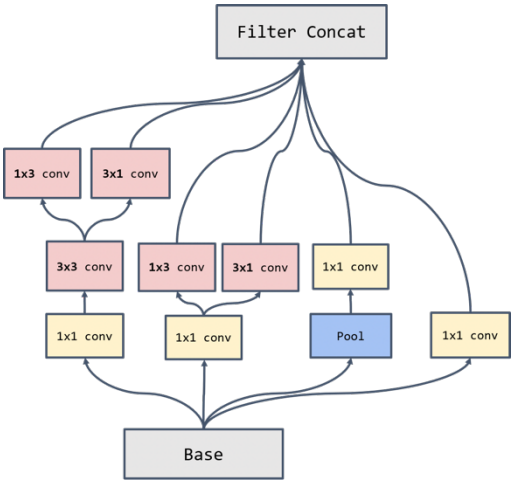

Figure 5. Transforming the $n \times n$ convolution in Module C into $1 \times n$ and $n \times 1$ convolutions.

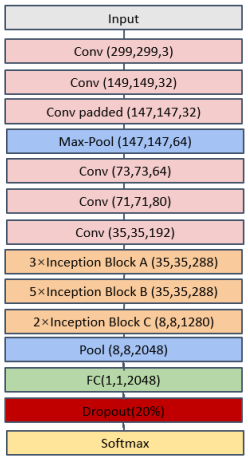

Figure 6. Structure of the learning model based on InceptionV3. Different colours denote different layer including convolution (Conv) layer, pooling layer, Inception block, fully connected layer and dropout layer.

The label smoothing mechanism is designed to regularise the classifier layer in the InceptionV3 model, which effectively avoids the occurrence of over-fitting to a certain extent. For example, given each training plant disease sample x, the probability of marking the corresponding label of each sample is

$$k \in \{1 \ldots K\}: p(k|x) = \frac{exp(Z_k)}{\sum_{i=1}^{K} exp(Z_i)}, \quad (1)$$

where Z_i denotes the logits probability.

The ground-truth distribution of the sample x is denoted by $q(k|x)$, where $\sum_k q(k|x) = 1$ for normalization. Thus, the cross entropy of the training sample is defined as

$$\ell = -\sum_{k=1}^{K} \log(\mathrm{p(k)})\mathrm{q(k)}, \quad (2)$$

where the minimization of cross entropy is approximately equivalent to the maximization of the logarithmic likelihood expectation of the label. Consider a distribution over labels $u(k)$, independent of the training example x, and a smoothing parameter ε, the label distribution $q(k)$ can be replaced with

$$q'(k \mid x) = (1-\varepsilon)qk + \varepsilon\, u(k), \quad (3)$$

which is a mixture of the original ground-truth distribution $q(k)$ and the fixed distribution $u(k)$, with weights 1-ε and ε, respectively. For easy implementation, a uniform distribution of $u(k)$, i.e., $u(k) = 1/K$, is used. Thus, Eq. 3 can be rewritten as

$$q'(k) = (1-\varepsilon)\delta_{k,y} + \frac{\varepsilon}{K}\ , \quad (4)$$

where this change in ground-truth label distribution is label smoothing regularization (LSR). Such smoothing operation avoids the largest logit from getting much larger than all the others. We use smoothed ground-truth label to estimate the cross-entropy, i.e.,

$$L_{en} = H(q', p') = -\sum_{k=1}^{K} log\, p\,(k) q'(k) = (1-\varepsilon)H(q,p) + \varepsilon H(u,p)\,. \quad (5)$$

It can be seen from Eq. 5 that the loss penalizes the deviation of the predicted label p from the prior u, which supervises the learning of deep feature.

Transfer learning is a method which applies the structure and weight of the pre-trained model trained on a large dataset to other training models **(Pan et al. 2009, Neyshabur et al. 2020)**. The simplest approach to transfer learning is to fine-tune the network using a pre-trained model. In this paper, transfer learning is specifically performed as follows. First, we download the pre-training weight of InceptionV3 on

ImageNet as the initialization of our network parameters. We then train the network model with the plant dataset used. The final deep feature is obtained from the output of the fully connected layer, which is for further fusion process. Figure 7 shows the block diagram of the transfer learning method.

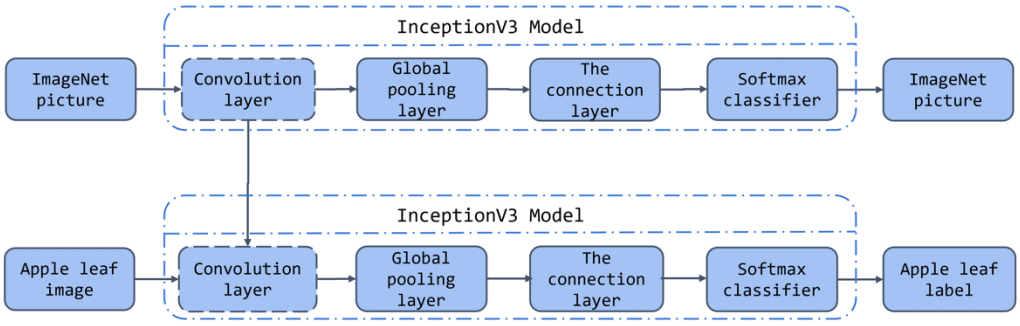

Figure 7. The transfer-learning based model.

2.2. Handcrafted Descriptors

HOG is one of the feature descriptors widely used in computer vision tasks, which is capable of capturing local texture and appearance information of an image by computing statistical histograms (**Dalal et al. 2005**). As there exist various local texture information in different types of diseased leave, HOG is effective for capturing such local leaf properties and thus improve the performance of apple leaf disease recognition. The specific process of HOG feature extraction is as follows:

1) The RGB image is converted into greyscale image $I(m,n)$. The image is normalized by Gamma correction to reduce the impact of illumination variation in the image using

$$I(m,n) = I(m,n)^{1/2}. \quad (6)$$

2) The gradient direction and amplitude of each pixel in the cell are calculated. Specifically, the gradient in the horizontal and vertical directions of the image $(G_m(m,n), G_n(m,n))$ are calculated. The gradient value $G(m,n)$ and gradient direction value $\theta(m,n)$ of each pixel position are computed using

$$G_m(m,n) = I(m+1,n) - I(m-1,n)$$
$$G_n(m,n) = I(m,n+1) - I(m,n-1), \quad (7)$$

$$G(m,n) = \sqrt{G_m^2(m,n) + G_n^2(m,n)}$$
$$\theta(m,n) = arctan\frac{G_{yn}(m,n)}{G_m(m,n)}. \quad (8)$$

3) The resulting image is divided into several cells, the size of which is set as 10×10 pixels, and the direction $0^0 \quad -180^0$ is divided into 9 bins as histogram channels to construct the cell histogram.

4) The feature vectors of all cells are concatenated to form the features of the block, and normalization is employed to obtain the final HOG features F_n i.e.,

$$F_n = \frac{v}{\sqrt{||v||_2^2 + \varepsilon'}}, \quad (9)$$

where v is the non-normalized vector containing all histograms in a given block and ε' is some small constant.

In order to make the feature dimension fit for classification, the size of the block and cell in our proposed method are set to 20×20 and 10×10, respectively. The 9-bins histogram is adopted, resulting in a 36-dimension feature vector for each image. The HOG feature extraction of a single image is illustrated in Figure 8.

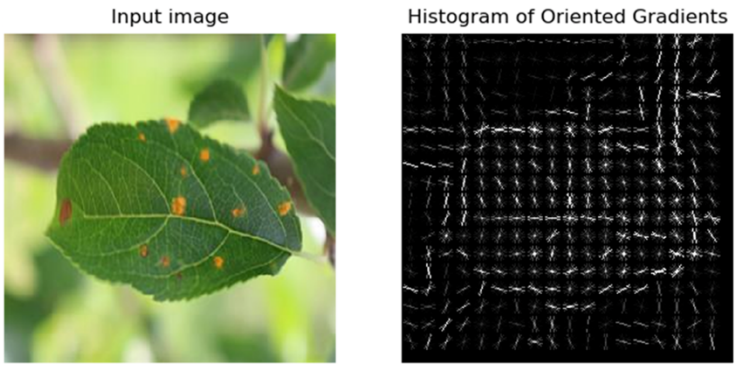

Figure 8. Illustration of HOG feature extraction: input image (left) and the extracted HOG (right).

2.3. Feature fusion with Centre Loss

A deep feature based on CNN captures the semantic information which is robust to environmental factors (e.g., variation in scene illumination), while HOG effectively extracts the local texture information. We fuse these two different types of feature to obtain the final feature of plant disease samples. In this paper, we directly integrate the extracted HOG with deep feature via the first fully connected layer in InceptionV3 model. The integrated feature is then fed to the second fully connected layer and softmax for classification.

To deal with the fine-grained characteristics in plant disease images, i.e., mitigating the intra-class variations while retaining the features of different classes separable, we introduce the auxiliary constraint, i.e., center loss, to increase the discriminative ability, i.e.,

$$L_C = \frac{1}{2}\sum_{i=1}^{m}\left\|x_i - C_{y_i}\right\|_2^2 , \quad (10)$$

where x_i denotes the fused features belonging to each image sample, and $C_{y_i} \in R^d$ denotes the y_ith class centre of fused features. In our experiments, instead of updating the centres of the entire training set, we perform the update based on mini batch. In each epoch, the centres are computed by averaging the features of the corresponding class. (In this case, some of the centres may not be updated). Eq. 10 effectively characterizes the intra-class variations. We adopt the joint supervision of softmax loss and centre loss to train the CNN based transfer learning model for discriminative feature learning, i.e.,

$$L = L_{en} + \lambda L_C = L_{en} + \frac{\lambda}{2}\sum_{i=1}^{m}\left\|x_i - C_{y_i}\right\|_2^2 , \quad (11)$$

where L_{en} is cross-entropy loss and a scalar λ is used for balancing the two loss functions. In each epoch, the centres are computed by averaging the features of the related classes. We use standard stochastic gradient decent (SGD) to optimize Eq. 11.

The gradient of L_C with respect to x_i and update equation of C_{y_i} are computed as

$$\frac{\partial L_C}{\partial x_i} = x_i - C_{y_i} \tag{12}$$

$$\Delta c_j = \frac{\sum_{i=1}^{m} \delta(y_i=j)\cdot(c_j - x_i)}{1+\sum_{i=1}^{m} \delta(y_i=j)} \ , \tag{13}$$

where $\delta(condition) = 1$ if the condition is satisfied, and $\delta(condition) = 0$ otherwise. It is clear that the CNN based model supervised by centre loss is trainable using the standard SGD.

Feature fusion is used to directly combine the smoothed vector of the convolutional layer and the traditional handcrafted feature vector. The fused feature is then fed to the fully connected layer as an input. During the classification, the more discriminative and meaningful features are learned under the joint supervision of cross-entropy and centre loss, that effectively capture the discriminative semantic information associated with plant leaves with various diseases.

2.4. Data Preparation

To evaluate the efficacy of the proposed method, we used the two public Apple Leaf dataset. The first is provided by the Plant Pathology Challenge (PPC) held at the CVPR FGVC (Fine-Grained Visual Classification) workshop in 2020 **(Thapa et al. 2020)**. This is referred as Apple Leaf Dataset 1 in our paper. The dataset consists of 1821 apple leaf images, and each image is labelled as heathy or three disease patterns (i.e., rust, scab, or with both diseases), with a ratio of approximately 6:6:6:1 for each of these four categories. The samples from each category are shown in Figure 9, and the number of images for each category are shown in Table 1. The second Apple Leaf dataset was presented by the Apple Research Institute in Korea **(Yu et al., 2020)**, which includes three classes: the diseased leaf with marssonia blotch; the diseased leaf with alternaria leaf spot; and the normal health leaf. The total numbers of samples belonging to three

classes are 120, 166 and 118, respectively. We referred this dataset as Apple Leaf Dataset 2 in our paper.

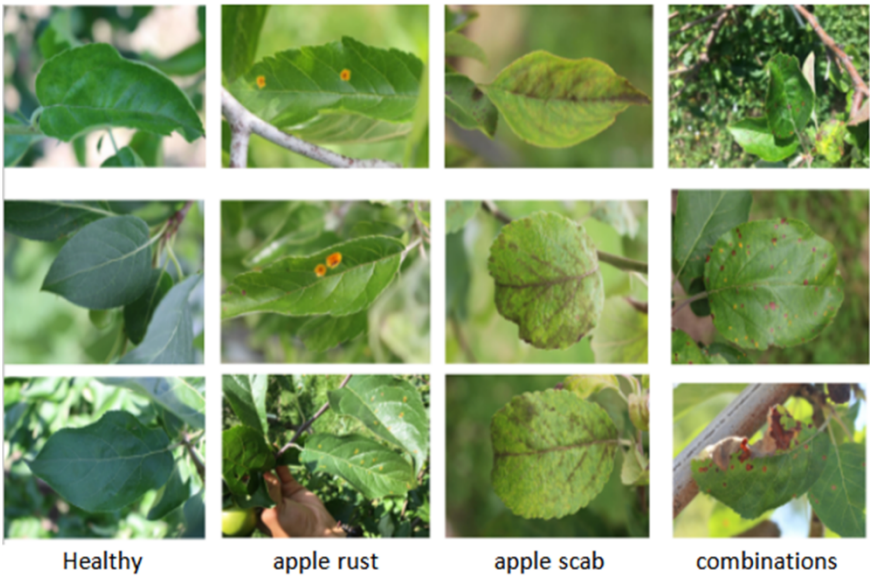

Figure 9. Sample images of the Apple Leaf dataset 1.

Table 1. The number of images for each category in two Apple Leaf datasets.

	Apple Leaf Dataset 1 (**Thapa et al. 2020**)				
Category	Health	Apple rust	Apple scab	Both	Total
Samples	516	622	592	91	1821
	Apple Leaf Dataset 2 **(Yu et al., 2020)**				
Category	Normal	Marssonia blotch	Alternaria leaf spot		Total
Samples	118	120	166		404

We also used Coffee Leaf dataset **(Esgario et al. 2020)**. The dataset contains 1747

images of arabica coffee leaves captured using different mobile phones. It includes healthy leaves and leaves affected by one or multiple diseases, e.g., leaf miner, rust, brown leaf spot and cercospora leaf spot. The samples from each category are shown in Figure 10, and the number of images for each category are presented in Table 2.

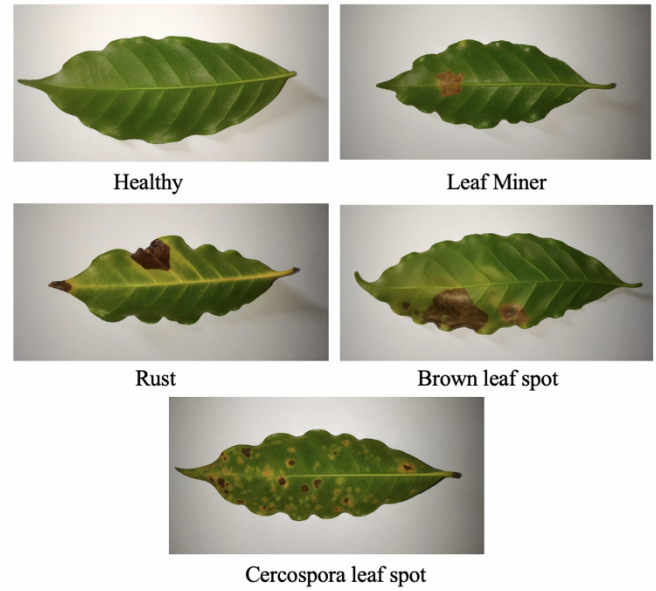

Figure 10. Sample images of the Coffee Leaf dataset.

Table 2. The number of images for each category in the Coffee Leaf dataset.

Category	Samples
healthy	272
Leaf Miner	387
Rust	531
Brown leaf spot	348
Cercospora leaf spot	147
Total	1685

To avoid the problem of over-fitting during the training of the deep learning

algorithm, data augmentation techniques have been used to increase the size of a training set by applying geometric and photometric transformations to the original data such that their distribution transformation remains unchanged. Such a technique helps in adding more varieties to the training set, and effectively addresses the over-fitting problem in the deep network training process for improved accuracy **(Zhong et al. 2020)**.

In this work, we apply data augmentation to the three datasets used in our experiments, increasing the size of the datasets without changing their biological characteristics. We focus only on the most common geometric transformation techniques, e.g., decentralization, standardization, rotation, translation, horizontal projections, flipping, scaling and vertical translation. The augmented results of a sample image are illustrated in Figure 11.

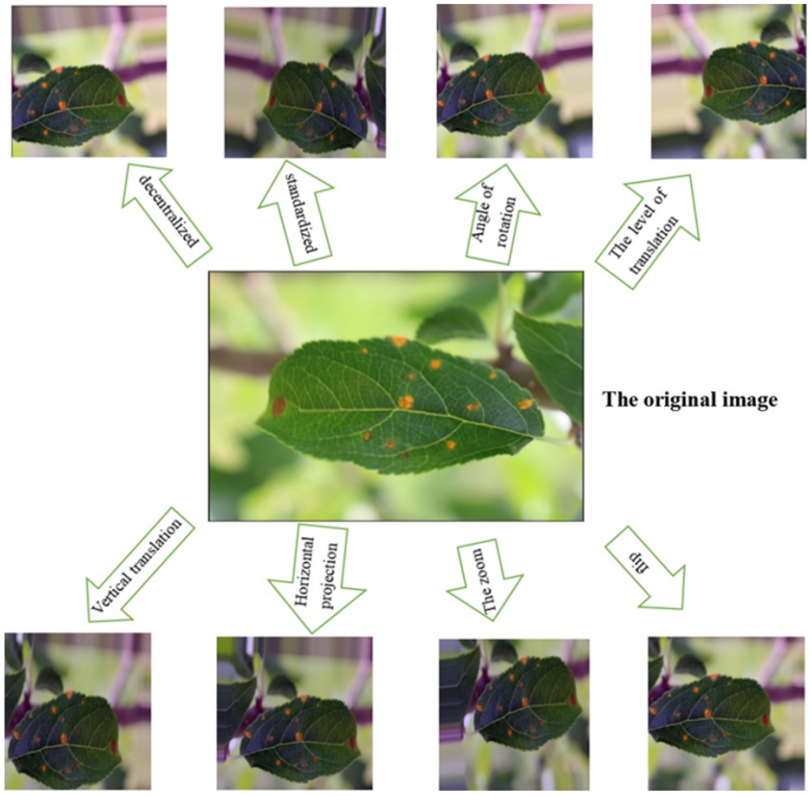

Figure 11. Illustration of different augmented results of an original sample image.

3. Results

3.1. Experimental setting

The experiments were conducted under the following hardware configuration: Intel(R) Core(TM) i7-9750h CPU, 16GB memory, and graphics card NVIDIA GeForce GTX 1080Ti. During the training of the networks, we used SGM to jointly optimize the cross-entropy and centre loss. The batch size, learning rate, momentum and weight decay were set to 32, 0.01, 0.9, 0.0005, respectively. For each dataset, we randomly allocated 80% samples for training, and the remaining 20% data are used for testing.

3.2. Experimental Results

The first experiment focused on investigating the effectiveness of transfer learning on different CNN backbone networks. DenseNet, VGG16, VGG19, ResNet50 and InceptionV3 were selected for comparison, where the final classification layer of each was modified according to the apple leaf disease dataset used. In this paper, we used the classification accuracy as the metric for evaluating the performance of the leaf disease recognition using transfer learning. The classification accuracy is measured by

$$Accuracy = \frac{T_{correct}}{T_{sum}} \times 100\%, \quad (8)$$

where $T_{correct}$ denotes the number of correctly classified samples for each type of diseased leaves, and T_{sum} denotes the total number of samples of this type of diseased leaves.

The results for leaf disease recognition are presented in Table 3. The table shows that the classification accuracy using InceptionV3 model performs better than the other networks, showing that InceptionV3 effectively extracts the useful information in plant disease images.

Table 3. The classification accuracy of transfer learning model using different networks.

Backbone network	Val_acc
DenseNet	78.52%
VGG16	73.79%
VGG19	81.49%
InceptionV3	91.28%

The curves of testing loss and testing accuracy are shown in Figure 12. The loss is consistently low, and the accuracy remains stably high, which shows the effectiveness of IncepetionV3 on the feature learning of plant disease images.

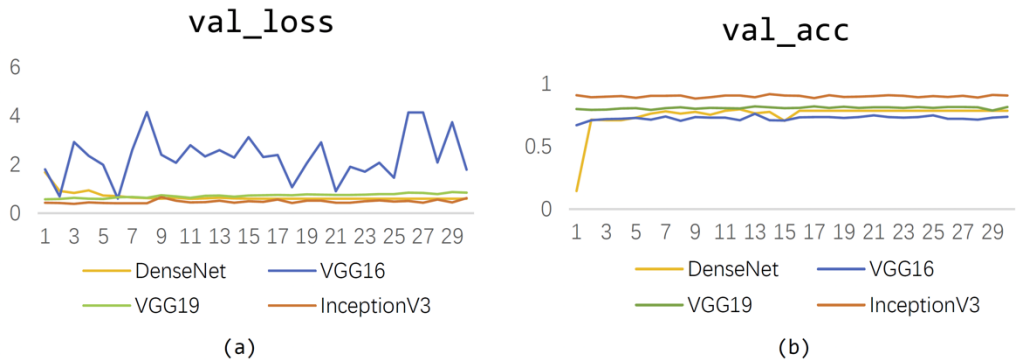

Figure 12. Line graph of classification loss (a) and accuracy (b) of different transfer networks.

To better demonstrate the efficacy of the proposed discriminative feature fusion methods, an ablation study is conducted on the two datasets (where the results are shown in Table 4). By combining the deep feature and HOG, the proposed method using feature fusion yields better performance than the individual features. Furthermore, the combination of deep feature and HOG is more effective than the combination of deep feature and other popular handcrafted features, i.e., SIFT and LBP. In addition, a significant increase in accuracy is obtained by incorporating the auxiliary discriminative constraint. As can be seen from Table 4, the proposed method provides the most discriminative for distinguishing two apple leaf and one coffee leaf diseases.

Table 4. Comparison of experimental results for individual initial diseased leaf images.

Task	Method	Accuracy(%)
Apple Leaf dataset 1 (**Thapa et al. 2020**)	HOG+SVM	82.53
	Deep feature only	91.27
	Deep feature + SIFT	92.30
	Deep feature + LBP	92.50
	Deep feature + HOG	93.19
	Deep feature + HOG + Discriminative constraint	94.02
Apple Leaf dataset 2 **(Yu et al., 2020)**	HOG+SVM	72.84
	Deep feature only	81.48
	Deep feature + SIFT	83.95
	Deep feature + LBP	82.72
	Deep feature + HOG	83.95
	Deep feature + HOG + Discriminative constraint	85.19
Coffee leaf dataset (**Esgario et al. 2020**)	HOG+SVM	78.68
	Deep feature only	90.00
	Deep feature + SIFT	92.89
	Deep feature + LBP	92.54
	Deep feature + HOG	93.39
	Deep feature + HOG + Discriminative constraint	94.07

The third experiment was conducted to demonstrate the effect of the proposed method using data augmentation, and the recognition results are shown in Table 5. The results show that the data augmentation increases the recognition accuracy of all methods, where the recognition rate of the proposed method obtain an accuracy 99.79%, 82.59% and 97.12% on the two apple leaf datasets and one coffee leaf dataset, respectively. These demonstrate that data augmentation effectively solved the problems of over-fitting and sample imbalance in plant leaf disease images. The confusion matrix of the proposed method on **the three datasets** are shown in Table 6, Table 7 and Table 8, where the leading diagonal entries represent the recognition accuracy for each disease category. Tables 6 to 8 show the proposed method perform consistently well in each disease category. However, confusion frequently occurs in recognizing the leaves affected by multiple diseases because the leaves of this category contain pattern similar to other categories, and there were limited samples for this category.

Table 5. Comparison of experimental results for individual initial diseased leaf using data augmentation

Task	Method	Accuracy (%)
Apple Leaf dataset 1	HOG+SVM	83.43
	Deep feature only	96.77
	Deep feature + SIFT	98.01
	Deep feature + LBP	98.49
	Deep feature + HOG	98.97
	Deep feature + HOG + Discriminative constraint	99.79
Apple Leaf dataset 2	HOG+SVM	76.54

	Deep feature only	88.90
	Deep feature + SIFT	87.65
	Deep feature + LBP	87.65
	Deep feature + HOG	91.36
	Deep feature + HOG + Discriminative constraint	92.59
Coffee leaf dataset	HOG+SVM	85.08
	Deep feature only	94.92
	Deep feature + SIFT	96.10
	Deep feature + LBP	96.61
	Deep feature + HOG	96.95
	Deep feature + HOG + Discriminative constraint	97.12

Table 6. Confusion matrix for the recognition results (%) for the proposed method using the Apple Leaf dataset 1.

	Heath	Rust	Scab	Multiple Diseases
Heath	**100**	0	0	0
Rust	0	**100**	0	0
Scab	0	0	**100**	0
Multiple Diseases	0	1.39	2.78	**95.83**

Table 7. Confusion matrix for the recognition results (%) for the proposed method using the Coffee Leaf dataset.

	Heath	Miner	Rust	Brown leaf spot	Cercospora leaf spot
Heath	**95.79**	1.58	1.05	1.58	o
Miner	0.74	**96.31**	1.11	1.11	0.74
Rust	o	0.54	**98.92**	0.27	0.27
Brown leaf spot	o	0.82	0.82	**98.36**	o
Cercospora leaf spot	0.97	0.97	2.91	2.91	**92.23**

Table 8. Confusion matrix for the recognition results (%) for the proposed method using the Apple Leaf dataset 2.

	Heath	Marssonia blotch	Alternaria Leaf spot
Normal	**100**	o	o
Marssonia blotch	o	**83.33**	12.5
Alternaria leaf spot	o	6.06	**93.94**

4. Discussion

As expected, the experimental results show that using fine-tuned Inceptionv3 achieves high accuracy and reduces the error rate (Table 3), which shows that InceptionV3 model is suitable for the identification of the plant disease. This is because InceptionV3 network employs multi size convolution to capture the disease spots with different size and shape, and performs the label smoothing to deal with the problem of

limited samples in plant datasets.

The results also show that fusing the deep learned feature and handcrafted feature increase the identification performance on all three datasets (Table 4). Among the handcrafted features, HOG feature is more suitable for supplementing the deep feature due to its ability in depicting the local texture information. We also visualized the feature maps of HOG feature and Deep features, which are shown in Figure 13. From the figure, we can see that the HOG feature and deep features capture different spatial properties of diseased leaf images. These spatial properties are beneficial and increase the recognition performance. Therefore, this motivates us to fuse two types of features for better performance in recognizing diseased leaves. Our experiment results validate the effectiveness of fusing two types of features.

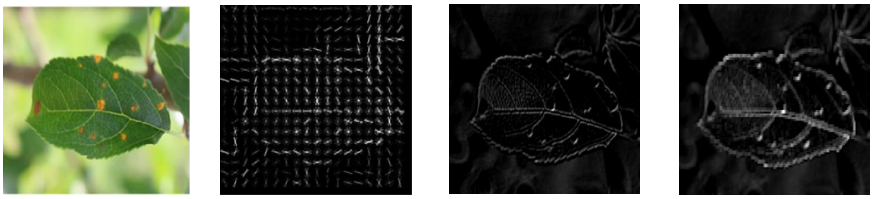

Figure 13. Visualization of HOG feature maps and Deep features low-level feature maps. From left to right are the original image, HOG feature map, feature map of the 2nd Convolutional layer, and feature map of the 3rd Convolutional layer.

In addition, the introduction of the centre loss is beneficial to the learning of the fused feature, where the accuracy increases from 93.19% to 94.02% and from 83.85% to 85.19% for Apple Leaf Dataset 1 and 2, respectively, and from 93.39% to 94.07% for Coffee Leaf dataset. It shows more discriminative feature are learned by the joint supervision of the center loss and cross-entropy loss.

To deal with the problem of insufficient training samples and image imbalance, we increase the size of the training set using image augmentation. Such technique helps

in adding more varieties to the training set and avoid the over-fitting. The experiment result shows the effectiveness of image augmentation (see Table 5). Note that we only use the simple strategy of image augmentation in this work. There are a few promising augmentation methods **(e.g., Cubuk et al. 2019, Shorten et al. 2019)** that have been proposed. An investigation of such methods is left for future research.

We use confusion matrix to explore the recognition performance for different leaf diseases (Tables 6, 7 and 8). For the Apple Leaf dataset 1, the leaves affected by multiple diseases are relatively hard to identify (with the accuracy of 95.83%) because the leaves of this category contain pattern similar to other categories. Moreover, there are limited samples for this category. For the Apple Leaf dataset 2, the recognition accuracy is lower than for the other apple leaf dataset. The main reason is the insufficient number of diseased samples despite the use of data augmentation. Thus, the representation that can distinguish the two diseases are not well learnt. Similarly, for the Coffee dataset, Cercospora leaf spot is hard to be accurately classified. This is because cercospora leaf spots are not restricted to the small local regions. Instead, the spots are located all over the whole leaf, which could confuse the classification of algorithms.

5. Conclusion

In this paper, a feature-fusion based method for identifying apple tree diseased leaves is proposed. The classical InceptionV3 network was improved and its corresponding features extracted by transfer learning. This enables such features to be fused with those extracted by traditional feature extraction method, e.g., HOG, speeding up the convergence speed and reducing the training parameters. The model was trained on 1821 images of apple leaves for identifying apple leaf diseases. The experiment results show that the accuracy of the model after integrating the traditional feature extraction method reached 93.19%, which is 1.91% higher than the model without the

fusion. After data augmentation in the training, the accuracy of the recognition on the data set reaches 99.83%.

The CNN model proposed in this paper is able to quickly and accurately identify apple leaf diseases, providing a feasible scheme for the identifying apple leaf diseases. It is widely acknowledged that deep learning methods tend to require more data for their training. Generally, increasing the size and variety of training samples could increase the capability to represent the images for the recognition methods. How to acquire more useful samples is also our future focus.

Literature Cited

[1] Badrinarayanan, V., Kendall, A., & Cipolla, R. 2017. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 39(12): 2481-2495.

[2] Bi C, Wang J, Duan Y, et al. 2020. MobileNet Based Apple Leaf Diseases Identification. **Mob. Netw. Appl**. (2020): 1-9.

[3] Cai, L., Zhu, J., Zeng, H., Chen, J., Cai, C., & Ma, K. K., 2018. HOG-assisted deep feature learning for pedestrian gender recognition. J. Franklin. Inst., 355(4): 1991-2008.

[4] Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. 2017. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Trans. Pattern Anal. Mach. Intell. 40(4): 834-848.

[5] Cubuk, E.D., Zoph, B., Mane, D., Vasudevan, V. and Le, Q.V., 2019. Autoaugment: Learning augmentation strategies from data. In Proceedings of the **IEEE Conference on Computer Vision and Pattern Recognition.** IEEE. Page 113-123. Salt Lake City, USA.

[6] Dalal, N., & Triggs, B. 2005. Histograms of oriented gradients for human detection. In 2005 **IEEE computer society conference on computer vision and pattern recognition (CVPR'05)**. 1: 886-893. San Diego, CA, USA.

[7] Dutot, M., Nelson, L. M., & Tyson, R. C. 2013. Predicting the spread of postharvest disease in stored fruit, with application to apples. **Postharvest Biol Tec**. 85: 45-56.

[8] Ebrahimi M. A., Khoshtaghaza M. H., Minaei S., et al. 2017. Vision-based pest detection based on SVM classification method. **Comput. Electron. Agric**. 137: 52-58.

[9] Esgario, J.G.; Krohling, R.A.; Ventura, J.A. 2020. Deep learning for classification and severity estimation of coffee leaf biotic stress. **Comput. Electron. Agric**. 169: 105162.

[10] He, K., Zhang, X., Ren, S., & Sun, J. 2016. Deep residual learning for image recognition. In Proceedings of the **IEEE conference on computer vision and pattern recognition**, Page 770-778. Las Vegas, Nevada, USA.

[11] Hlaing, C.S.; Zaw, S.M.M. 2017. Plant diseases recognition for smart farming using model-based statistical features. 2017 IEEE 6th **global conference on consumer electronics (GCCE)**. IEEE. Page 1–4.

[12] Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K. Q. 2017. Densely Connected Convolutional Networks. 2017 **IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. IEEE. Page 2261-2269. Honolulu, Hawaii, USA.

[13] Jalal, S., Burak, B. 2020 Evolutionary Feature Optimization for Plant Leaf Disease Detection by Deep Neural Networks for Apple Leaf. **Int. J. Comput. Intell. Syst**. 13(1): 12-23.

[14] Krizhevsky, A., Sutskever, I., & Hinton, G. E. 2012. Imagenet classification with deep convolutional neural networks. **Advances in neural information processing systems**, 25:1097-1105. Lake Tahoe Nevada, USA.

[15] Liu, B., Zhang, Y., He, D. and Li, Y. 2018. Identification of apple leaf diseases based on deep convolutional neural networks. **Symmetry**. 10(1)：11.

[16] Mahlein, A. K., Rumpf, T., Welke, P., Dehne, H. W., Plümer, L., Steiner, U., & Oerke, E. C. 2013. Development of spectral indices for detecting and identifying plant diseases. **Remote Sens Environ**. 128: 21-30.

[17] Mohanty, S. P., Hughes, D. P., & Salathé, M. 2016. Using deep learning for image-based plant disease detection. Front. Plant Sci. 7: 1419.

[18] Nachtigall, L. G., Araujo, R. M., & Nachtigall, G. R. 2016. Classification of apple tree disorders using convolutional neural networks. In 2016 IEEE 28th **International Conference on Tools with Artificial Intelligence (ICTAI)**. IEEE. Page 472-476)

[19] Neyshabur, B., Sedghi, H., & Zhang, C. 2020. What is being transferred in transfer learning?. arXiv preprint arXiv:2008.11687.

[20] Padol, P. B., & Yadav, A. A. 2016. SVM classifier based grape leaf disease detection. In 2016 Conference on advances in signal processing (**CASP**). Pages 175-179. IEEE.

[21] Pan, S. J., & Yang, Q. 2009. A survey on transfer learning. **IEEE Trans Knowl Data Eng.** 22(10): 1345-1359.

[22] Patil J K, Kumar R. 2017. Analysis of content based image retrieval for plant leaf diseases using color, shape and texture features. **Eng. Agric. Environ. Food**. 10(2): 69-78.

[23] Qin, F., Liu, D., Sun, B., Ruan, L., Ma, Z., & Wang, H. 2016. Identification of alfalfa leaf diseases using image recognition technology. **PLoS One**. 11(12): e0168274.

[24] Rumpf, T., Mahlein, A. K., Steiner, U., Oerke, E. C., Dehne, H. W., & Plümer, L. 2010. Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance. **Comput. Electron. Agric**. 74(1): 91-99.

[25] Sandika, B., Avil, S., Sanat, S., Srinivasu, P. 2016. Random forest based classification of diseases in grapes from images captured in uncontrolled environments. 2016 IEEE 13th **International Conference on Signal**

561 **Processing (ICSP)**. IEEE, Page 1775–1780.
562 [26] Sharif, M., Khan, M.A., Iqbal, Z., Azam, M.F., Lali, M.I.U., Javed, M.Y. 2018. Detection and classification of
563 citrus diseases in agriculture based on optimized weighted segmentation and feature selection. **Comput. Electron.**
564 **Agric**. 150: 220–234.
565 [27] Simonyan, K. and Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition.
566 arXiv preprint arXiv:1409.1556.
567 [28] Shorten, C. and Khoshgoftaar, T.M., 2019. A survey on image data augmentation for deep learning. **J. Big**
568 **Data**. 6(1): 1-48.
569 **[29] Shrivastava, V. K., and Pradhan, M. K., 2021. Rice plant disease classification using color features: a**
570 **machine learning paradigm. Journal of Plant Pathology, 103(1), 17-26.**
571 **[30] Shrivastava, V. K., Pradhan, M. K., & Thakur, M. P. 2021., Application of Pre-Trained Deep**
572 **Convolutional Neural Networks for Rice Plant Disease Classification. In 2021 International Conference on**
573 **Artificial Intelligence and Smart Systems (ICAIS), Page 1023-1030. IEEE.**
574 [31] Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., & Stefanovic, D. 2016. Deep neural networks based
575 recognition of plant diseases by leaf image classification. Comput. Intell. Neurosci.
576 [32] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., & Rabinovich, A. 2015. Going deeper with
577 convolutions. In Proceedings of the **IEEE conference on computer vision and pattern recognition**, Page 1-9.
578 Boston, MA, USA
579 [33] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. 2016. Rethinking the Inception Architecture for
580 Computer Vision. 2016 **IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. Las Vegas,
581 Nevada,
582 **[34] Tetila, E, Machado, B., Menezes, G., Oliveira, et al., 2019. Automatic recognition of soybean leaf diseases**
583 **using UAV images and deep convolutional neural networks. IEEE Geosci. Remote Sens. Lett. 17(5): 903-907.**
584 [35] Thapa, R.; Snavely, N.; Belongie, S., Khan, A. 2020. The Plant Pathology 2020 challenge dataset to classify
585 foliar disease of apples. arXiv preprint arXiv:2004.11958.
586 [36] Wang H, Li G, Ma Z, et al. 2012. Image recognition of plant diseases based on principal component analysis
587 and neural networks. 2012 8th **International Conference on Natural Computation**. IEEE. Page 246-251.
588 **[37] Wang, W.; Xie, E.; Li, X., et al. Pyramid vision transformer: A versatile backbone for dense prediction**
589 **without convolutions. arXiv 2021, arXiv:2102.12122**
590 [38] Ye, Q., Li, Z., Fu L., et al. 2019. Nonpeaked Discriminant Analysis, **IEEE Trans. Neural Netw. Learn. Syst**.
591 30(12): 3818-3832.
592 [39] Yu, H. J., Son, C. H. 2020. Leaf spot attention network for apple leaf disease identification, **IEEE Conference**
593 **on Computer Vision and Pattern Recognition Workshop**, Virtual, June 2020, Page. 229-237.
594 [40] Yu, H. J., Son, C. H., & Lee, D. H. 2020. Apple leaf disease identification through region-of-interest-aware deep

convolutional neural network. **Journal of Imaging Science and Technology**, 64(2), 20507: 1-10.

[41] Yuan, L., Huang, Y., Loraamm, R. W., Nie, C., Wang, J., & Zhang, J. 2014. Spectral analysis of winter wheat leaves for detection and differentiation of diseases and insects. **Field Crops Res.** 156: 199-207.

[42] Zhong, Z., Zheng, L., Kang, G., Li, S., & Yang, Y. 2020. Random erasing data augmentation. In Proceedings of the **AAAI Conference on Artificial Intelligence**. 34(07): 13001-13008.