Plant Species Classification Based on Hyperspectral Imaging via a Lightweight Convolutional Neural Network Model

Keng-Hao Liu¹, Meng-Hsien Yang¹, Sheng-Ting Huang¹ and Chinsu Lin²*¹

¹ Department of Mechanical and Electro-Mechanical Engineering, National Sun Yat-sen University, Kaohsiung, Taiwan
¹ Department of Forestry and Natural Resources, National Chiayi University, Chiayi, Taiwan

In recent years, many image-based approaches have been proposed to classify plant species. Most methods utilized red green blue (RGB) imaging materials and designed custom features to classify the plant images using machine learning algorithms. Those works primarily focused on analyzing single-leaf images instead of live-crown images. Without considering the additional features of the leaves’ color and spatial pattern, they failed to handle cases that contained leaves similar in appearance due to the limited spectral information of RGB imaging. To tackle this dilemma, this study proposes a novel framework that combines hyperspectral imaging (HSI) and deep learning techniques for plant image classification. We built a plant image dataset containing 1,500 images of 30 different plant species taken by a 470–900 nm hyperspectral camera and designed a lightweight conventional neural network (CNN) model (LtCNN) to perform image classification. Several state-of-art CNN classifiers are chosen for comparison. The impact of using different band combinations as the network input is also investigated. Results show that using simulated RGB images achieves a kappa coefficient of nearly 0.90 while using the combination of 3-band RGB and 3-band near-infrared images can improve to 0.95. It is also found that the proposed LtCNN can obtain a satisfactory performance of plant classification (kappa = 0.95) using critical spectral features of the green edge (591 nm), red-edge (682 nm), and near-infrared (762 nm) bands. This study also demonstrates the excellent adaptability of the LtCNN model in recognizing leaf features of plant live-crown images while using a relatively smaller number of training samples than complex CNN models such as AlexNet, GoogLeNet, and VGGNet.

Keywords: plant species classification, live-crown features, leaf feature recognition, plant stress detection, dimensionality reduction, convolutional neural network, hyperspectral imaging, deep learning

INTRODUCTION

Species composition provides basic individual biological features of a landscape and a forest ecosystem. The ability to identify species of individual plants or trees over an inventory plot as well as a forest stand is essential for the automatic mapping of plant distribution, biological diversity, stand structure, and even for diagnosing the dynamics of a forest stand (Lin et al., 2016; Lin, 2019; Santos et al., 2019). The development of plant mapping techniques has the benefit of
identifying signals of climate change based on plant phenology (Lin C. et al., 2018) and advanced tree segmentation (Lin C.Y. et al., 2018; Jaskierniak et al., 2021) via remote sensing images. Remote sensing images have recently been used to map species distribution mainly according to spectral information with classification techniques. The gaps in/between tree crowns, which tend to be caused by lower crown density, greenness, and background materials, create a challenge for species classification using high-resolution satellite images (Lin et al., 2015a). Image fusion that integrates very high spatial resolution images with atmospherically corrected high spectral resolution can benefit tree crown delineation and improve the mapping (Lin et al., 2015b; Lin C.Y. et al., 2018). However, pixels of the inter- and intra-canopy gaps in a fused image became more significant and increased impact on species crown reflectance (Lin et al., 2015a). Consequently, plant species recognition with remote sensing images becomes a more complicated task involving not merely pixel-based but also object-based approaches. Recently, advanced sensor technology can acquire very high spatial resolution (VHRS) images from various platforms such as in-situ, drone, airborne, and spaceborne for environmental studies. With regard to plant studies, VHRS images are capable of sensing every subtle difference of reflectance in a scale from sub-centimetric to decimetric size allowing better opportunity to reveal detailed features of materials. This is particularly evident in in-situ hyperspectral imaging systems. Moreover, to address the impact of climate change on a vulnerable vegetation community or ecosystem, the dynamics of the community must be derived from the perspective of plant species composition. Therefore, more effort is needed to investigate the problem of developing suitable remote sensing algorithms for classifying a large number of plant species.

During the last decade, most research on plant classification with red green blue (RGB) images was primarily based on extracting leaf features and performing classification with machine learning (ML; Zhang, 2020) classifiers on single-leaf images. The ML classifiers used included support vector machine, K-nearest neighbor, probability neural network, and so on. The methods of leaf feature extraction include polar Fourier transform (Kadir et al., 2013), Canny edge detection (Salman et al., 2017), Fourier transforms (Hossain and Amin, 2010; Khmag et al., 2017), and wavelet decomposition method (Zhang H. et al., 2012), in which the most frequently used features were the color, shape, contour, and texture of leaves. Additional features such as leaf width factor and leaf edge were also used to develop multiscale-distance feature matrices to improve classification by Beghin et al. (2010) and Hu et al. (2012). As noted, the issues raised in these image-based plant recognition methods are highly dependent on feature engineering and the lack of leaf composition information. In other words, much more effort should be made to achieve noise removal, leaf feature measurements, and texture divergence calculations. The classification seems very dependent on leaf preprocessing.

With the breakthrough of hardware technology, deep learning (DL; Bengio et al., 2017) became the mainstream data processing method in recent years. Among many DL approaches, the convolutional neural network (CNN) is the most popular and representative one in computer vision and imaging processing communities (Ioffe and Szegedy, 2015; Simonyan and Zisserman, 2015; Szegedy et al., 2015; He et al., 2016; Krizhevsky et al., 2017; Wang et al., 2021; Yang et al., 2021). Different from ML methods, CNN can integrate feature derivation, feature learning, and classifier into a single architecture. Many studies have reported that using CNN approaches can produce significantly higher accuracy than using conventional ML ones, as long as with sufficient training data. The reason is that CNN can automatically learn objective, multi-scale, and most discriminative features from raw data without human subjectivity. Following this trend, a few CNN-based plant recognition methods were proposed (Lee et al., 2015; Grinblat et al., 2016; Carranza-Rojas et al., 2017; Lee et al., 2017; Chen et al., 2018; Zhu et al., 2019; Chen et al., 2021). A two-dimensional (2D)-CNN model is adopted in each work to learn the discriminative features from the entire RGB plant images. The spatial relationship of leaf arrangement (phyllotaxy) and overlapping patterns can also be discovered. In other words, various spatial features of interest objects revealed in a VHSR image can be processed by suppressing background materials’ signals and therefore recognized based on the spatial pattern in spectra. This allows us to identify plant species in a way very similar to phytologists with plant morphological features such as leaf color and size, contour, surface, venation, and even phyllotaxy.

Although the current DL approaches demonstrate a certain level of reliability, they still may fail to handle the cases that contain plant species that are similar in appearance, even with enough training data, due to the limited spectral information provided by RGB imaging. If two or more plants have similar outer appearance characteristics, the CNN-based methods may misclassify them. Under such circumstances, it is necessary to use the imaging system providing more delicate spectral information to improve the recognition performance. With the advancement of remote sensing imaging technology, hyperspectral imaging (HSI; Chang, 2013) was developed and widely applied to many topics such as agriculture (Nicolai et al., 2006; Baiano et al., 2012; Teena et al., 2014; Jung et al., 2015; Marshall et al., 2015; Rapaport et al., 2015; Adão et al., 2017; Gao et al., 2018; Mirzaei et al., 2019; Sun et al., 2019; Sinha et al., 2020; Feng et al., 2021), military defense (Briottet et al., 2006), environment (Zhang B. et al., 2012; Schmitter et al., 2017; Harrison et al., 2018; Abbas et al., 2021), plant phenotyping (Ubbens and Stavness, 2017; Nasiri et al., 2021), and medical imaging (Liu et al., 2007; Fei, 2020). The familiar HSI image contains hundreds of spectral bands ranging from the visible spectrum to the near-infrared (NIR) spectrum so that it can capture the complete spectral characteristics of target objects. Due to its superior spectral resolution, many substances indistinguishable to the naked eye can be recognized. In recent years, hyperspectral cameras have been gradually commercialized. The use of micro hyperspectral cameras for research has become more and more popular. Therefore, using HSI technology to classify plant species has great potential.
Since the use of both HSI and DL techniques for plant species recognition has not been fully explored, in this paper, we conducted a study that adopts hyperspectral plant images as the sample materials and designed a lightweight CNN model to achieve accurate image classification. Firstly, we collect the hyperspectral images from 30 plants of different species with a hyperspectral camera and build a dedicated HSI plant dataset containing 1,500 images. Since the existing CNN-based classification models were designed for the datasets composed of tens of thousands of RGB images of specific objects with a large number of categories, they may not be suitable for the training of our HSI plant dataset. Therefore, this study proposed an improved lightweight convolutional neural network based on the architecture of GoogLeNet (Szegedy et al., 2015) to disclose the issues of species classification through hyperspectral images by deep learning technique. Hyperspectral images have hundreds of bands that are highly correlated, and spectral information of the bands is excessively redundant for vegetation application such as water content modeling (Lin et al., 2012), hyperspectral signal restoration (Lin, 2017; Lin, 2018), and chlorophyll concentration estimation (Lin et al., 2015c; Lin and Lin, 2019). Appropriate feature selection strategies in deriving critical bands for accurate species classification were also explored. Plant classification will be beneficial in diagnosing the stress of individual trees and, therefore, the forest. The objectives of this study are:

(1) Applying hyperspectral imaging technique to build a plant species hypercube dataset consisting of 1500 images of plant species to support developing ML models for the plant species classification,

(2) Investigating the feasibility of applying published deep learning architectures to the species classification based on spectral-textural information of plant live-crown images,

(3) Proposing a lightweight CNN model to catch plant live-crown features in the hyperspectral images to achieve optimistic classification performance, and

(4) Exploring the appropriateness of feature selection for hyperspectral images in species classification and the influence of using a limited number of spectral bands on classification performance.

HYPERSPECTRAL DATA COLLECTION

Plant Preparation

As mentioned above, this study aims to recognize plant species based on plant morphology via features of leaf color, size, contour, surface, venation, and phyllotaxy. Thirty species of foliage plants from 27 genera and 18 families were collected to produce plant images for analysis. To increase the leaf features and plant geometry diversity in the images, at least 2 or 3 plant individuals were gathered for replications. As shown in Figure 1, leaf features of the plants appeared similar or dissimilar in color, size, venation, and leaf edge. Detailed taxonomy information of the species is shown in Table 1.

Plant Image Acquisition

The IMEC Snapscan VNIR B150 imaging system1 was used to capture the hyperspectral images of the species. This system composes of the spectral image sensor, HSI camera, optics, and some other components that can acquire hypercube datasets up to a full-image size of 3,600 × 2,048 pixels covering a spectrum range from 458 nm to 913 nm. The system’s spectral and radiometric resolutions are 2.8 nm (equivalent to 161 bands) and 10 bits. In the image acquisition, the camera is mounted on a tripod facing downward to the plant at a distance of 40 and 60 cm. Two 50 w/12 V halogen lamps were deployed, one on each side of the plant at a 45-degree elevation angle from the horizontal plane. A black material was used to minimize the background/neighboring material reflectance effects on the target reflectance. The aperture of the camera was set to f5.6 for every single snapshot. Due to the vertical and horizontal variations of the leaves locations, changing the orientation of the plant led to changes in light intensity over the crown area and therefore helped to increase the diversity of the sample images. With the fixed positions of the two light sources, the plant was set to rotate 90 degrees to generate diverse hyperspectral images of the same plant. The image-acquisition scheme is shown in Figure 2. Accordingly, the snapshot acquisition produced a hypercube raw image with a dimension of 1,200 rows × 1,200 columns × 161 bands, and a dynamic range of 10 bits. With the combination of two camera-target distances and four plant orientations, eight HSI raw images of every individual plant of the 30 species were acquired. Due to the significant noise in the wavelengths at both ends of the sensor, the raw image was spectrally subset to 147 bands with a spectrum range of 468–898 nm for the analysis.

Data Calibration

To eliminate the impact of inconsistent image quality caused by the environmental factors, such as different illuminations or sensor response, each acquired HSI raw image \( R_0 \) was calibrated with the formula to derive the HSI reflectance image \( R_f \):

\[
R_f = \frac{R_0 - I_B}{I_w - I_B} \times 100\% ,
\]

where \( I_B \) denotes the dark reference image with 0% reflectance recorded with the lens closed, and \( I_w \) presents the white reference image with more than 95% reflectance recorded with white a Teflon panel.

Hypercube Dataset Preparation

To increase the total number of images for DL and reduce the computational complexity of training a CNN model, we adopted the following steps to segment a large image into multiple smaller sub-images. First, each 1,200 × 1,200 HSI reflectance image is evenly segmented into nine non-overlapping 400 × 400 sub-images. Then, those sub-images with a noticeable shadow or insufficient leaves, e.g., the leaf/background ratio does not exceed 60%, were removed. As a result, 50 sub-images were inspected and retained for each species, and a total of 1,500 HSI reflectance images (hereafter hypercube images) were generated.

---

1https://www.imec-int.com/en/
FIGURE 1 | The RGB sample images of the 30 plant species.

TABLE 1 | The taxonomy information and image samples list of the 30 plant species.

| ID | Family            | Scientific name     | Abbreviation | Number of individuals | Number of full-size images | Number of sub-images |
|----|-------------------|---------------------|--------------|-----------------------|---------------------------|----------------------|
| 0  | Acanthaceae       | Fittonia albivenis  | F.a          | 2                     | 16                        | 50                   |
| 1  | Apocynaceae       | Hoya carnosa        | H.c          | 2                     | 16                        | 50                   |
| 2  | Apocynaceae       | Hoya keni           | H.k          | 2                     | 16                        | 50                   |
| 3  | Apocynaceae       | Ammocallis rosea    | A.r          | 2                     | 16                        | 50                   |
| 4  | Araceae           | Spathiphyllum kochii| S.k          | 2                     | 16                        | 50                   |
| 5  | Araceae           | Zamiaechites zamifolia | Z.z      | 2                     | 16                        | 50                   |
| 6  | Araceae           | Aglaonema anyamanee | A.an         | 2                     | 16                        | 50                   |
| 7  | Araceae           | Aglaonema commutatum| A.c         | 2                     | 16                        | 50                   |
| 8  | Araceae           | Alocasia amazonica | A.am         | 2                     | 16                        | 50                   |
| 9  | Araceae           | Hydrocotyle ventricolata | H.v | 2                     | 16                        | 50                   |
| 10 | Araliaceae        | Polyscias guilloyle | P.g          | 2                     | 16                        | 50                   |
| 11 | Araliaceae        | Schefflera arboricola | S.a       | 2                     | 16                        | 50                   |
| 12 | Asparagaceae      | Sansevieria trifasciata | S.t     | 3                     | 24                        | 50                   |
| 13 | Asparagaceae      | Dracaena marginata  | D.m          | 2                     | 16                        | 50                   |
| 14 | Asparagaceae      | Cryptanthus plicatus| C.p         | 2                     | 16                        | 50                   |
| 15 | Begoniaceae       | Begonia cathayana   | B.c          | 3                     | 24                        | 50                   |
| 16 | Bromeliaceae      | Cryptanthus bivittatus | C.b      | 3                     | 24                        | 50                   |
| 17 | Clusiaceae        | Clusia rosea       | C.r          | 2                     | 16                        | 50                   |
| 18 | Davalliaceae      | Davalla griffithiana| D.g         | 2                     | 16                        | 50                   |
| 19 | Haloragaceae      | Myriophyllum aquaticum | M.a        | 2                     | 16                        | 50                   |
| 20 | Lamiaceae         | Glechoma hederacea  | G.h          | 2                     | 16                        | 50                   |
| 21 | Lamiaceae         | Plectranthus amboinicus | P.am       | 2                     | 16                        | 50                   |
| 22 | Lamiaceae         | Plectranthus amboinicus cv. | P.a.cv | 2                     | 16                        | 50                   |
| 23 | Malvaceae         | Pachira aquatica   | P.aq         | 2                     | 16                        | 50                   |
| 24 | Marantaceae       | Calthea lancifolia | C.l          | 3                     | 24                        | 50                   |
| 25 | Nephrolepidaceae  | Nephrolepis exaltata| N.e       | 2                     | 16                        | 50                   |
| 26 | Orchidaceae       | Spathoglottis plicata | S.p      | 2                     | 16                        | 50                   |
| 27 | Piperaceae        | Peperomia puteolata | P.p        | 2                     | 16                        | 50                   |
| 28 | Podocarpaceae     | Podocarpus macrophyllus | P.m      | 3                     | 24                        | 50                   |
| 29 | Urticaceae        | Pilea cadierei     | P.c          | 2                     | 16                        | 50                   |
for the study. It is worth noting that each sub-image still retains sufficient spatial information of which type of plant it belongs to. The overall process is illustrated in Figure 3. After that, the images were randomly divided into training and test datasets. The former contains 1,200 hypercube images, and the latter has 300 ones. Figure 4 shows some example images of the species at some selected wavelengths in the visible and NIR regions. The reflectance and the features of leaves are retained in each band of the hypercube image.

**Deriving Representative Spectra of the Plant Species**

In remote sensing, a reflectance curve is typically representative of the spectral behavior of an object. In addition to the deep learning approach, this study also investigates the spectral reflectance of the plant species. The representative reflectance at each wavelength of the visible-NIR region is determined as the average of the plant leaves. To do this, it is necessary to eliminate background components to extract the region-of-interest (ROI), referred to as leaf regions. The steps to extract the representative reflectance curve of the species are described as follows.

First, two particular bands with the largest globally average reflectance, abbreviated maxBand, and the smallest globally average reflectance, abbreviated minBand, among all bands of a hypercube image were identified. Second, a different image of the two specific bands is determined as maxBand – minBand. Third, a thresholding method (Ma et al., 2015) is used to differentiate the different images into two parts, i.e., the region-of-interest vs. the background. The threshold value was set to be 0.3, determined based on the experience. Fourth, a regional-averaging model is adopted to the hypercube image to calculate the average reflectance of every pixel in the ROI. Fifth, the representative reflectance of a particular wavelength of a Plant Species is generalized as the mean of all the corresponding average values of the 50 hypercube images of the species. Finally, the full-wavelength reflectance Spectra of the species are restored by assembling the representative reflectance at every wavelength. Figure 5 illustrates the overall procedures for deriving the generalized reflectance curve.

**METHODS**

The Lightweight Convolution Neural Network Architecture

Considering the limitation of gathering a large number of species and images, this study followed the concept of "compact" in ML to design the lightweight CNN (LtCNN) for better modeling fitting. The LtCNN model is developed by referring to GoogLeNet (Szegedy et al., 2015) and other networks (Ioffe and Szegedy, 2015; Simonyan and Zisserman, 2015; Szegedy et al., 2015). As shown in Figure 6, the architecture of LtCNN is only composed of three parts. The first two are responsible for feature extraction, and the last one is for prediction. The details of those parts are explained below, and the setting of network parameters of the LtCNN model is summarized in Table 2.

**Part I:** Part I aims to convert the input image into low-level (or shallow) features as the input of Part II. It comprises two convolutional layers (5 × 5 and 3 × 3) and one pooling layer.

**Part II:** The objective of Part II is to learn the high-level features in a multi-scale manner as the input of Part III. It adopts three “Inception modules” originating from GoogLeNet. Our first two inception modules adopt a 3-path structure and replace the 5 × 5 convolution in the original version with two 3 × 3 ones to reduce the number of parameters while maintaining the same receptive field. The third inception module only adopts a 2-path structure since the size of the feature map has been reduced.

**Part III:** Aims to perform classification via the features received from Part II. It uses global average pooling (GAP) to integrate all the features and then applies two fully-connected (FC) layers, one dropout layer, and a Softmax classifier to predict the species of the input image.

**Loss Function**

The cross-entropy is selected as the loss function to measure the difference between two probability distributions of the target ground truth and the model’s prediction. It is defined by

\[
\text{loss}_{CE} = \sum_{c=1}^{C} \sum_{i=1}^{S} -y_{c,i} \log(p_{c,i}),
\]

where \(C\) stands for the number of classes, \(S\) denotes the batch size, \(y_{c,i}\) is a binary indicator, and \(p_{c,i}\) is the predicted probability. In our experiment, we set \(S = 12\) and \(C = 30\).

**Experimental Setting**

The experiments were implemented on the hardware environment with an Intel i7-7700k CPU, 32 GB RAM, and NVIDIA GTX-1080Ti GPU. Three well-known CNN models such as AlexNet (Krizhevsky et al., 2017), VGGNet (Simonyan and Zisserman, 2015), and GoogLeNet (Szegedy et al., 2015) were applied as a referring method for Performance comparison of the Species recognition/Classification. Since AlexNet was designed for the classification of a large number of categories with very deep neural networks, the number of neurons of FC layers was reduced. Specifically, the number of output classes was set to 30 in this study. This model is, therefore, named AlexNetr. Similarly, it is difficult to reach convergence when training the original VGGNet (16 layers) on our plant dataset. The original architecture of VGGNet (16 layers) is therefore simplified by preserving the first eight convolution layers and three FC layers and reducing the number of neurons in the FC layers. It is named VGGNetr in this study. Similar to the LtCNN, a ReLu activation function is applied to improve the nonlinearity.

All the models are trained from scratch without pre-trained parameters or transfer learning techniques. They are implemented on Tensorflow 1.8.0. The size of the input is set to \(200 \times 200 \times L\) for our lightweight model and \(224 \times 224 \times L\) for other CNN models, where \(L\) denotes the number of selected bands. If we set \(L = 6\), the number of parameters of AlexNetr, GoogLeNet, VGGNetr, and the proposed lightweight model are 10865310, 15901982, 10404938, and 1388950, respectively. For data augmentation, we used random crop, random flip in
horizontal or vertical to expand the size of the training dataset. For parameter settings, the batch size was set to 12. The learning rate was set to be exponential decay with an initial rate of $10^{-3}$ and a decay rate of 0.9 for every 5 epochs. The training epoch was assigned as 200, and the optimizer was ADAM. To evaluate model performance, four quantitative metrics were used: overall accuracy (OA), precision, macro F1-Score, and kappa coefficient.

**Feature Selection Methods**
The hyperspectral image bands are mostly correlated, particularly those in a similar spectral region. Using full spectral bands for data analysis may lead to the curse of Dimensionality (Hughes, 1968) and increase the computational burden. To achieve a better calculation efficiency while retaining classification accuracy, data dimensionality reduction (Chang, 2013) is required to
select critical bands or discriminative spectral features for the species classification/recognition with DL techniques. The band selection was made via two approaches: manual inspection of the reflectance curves and automatic selection based on the spectral heterogeneity of bands. The methods are summarized in Table 3, and the suggested bands with the corresponding wavelength are listed in Table 4.

RESULTS

Use of the Visible-Infrared Spectra Variation of Plant Species for Classification

A generalized reflectance curve of species leaves (average spectra) is essential for differentiating and labeling pixels in a pixel-based classification. Figure 7 shows the averaged spectra of the ROI regions of the 30 plant species, where the x-axis denotes wavelength and the y-axis indicates reflectance values. Each curve was drawn by averaging the spectral reflectance vectors of all the leaf pixels of that particular hypercube image of a species. As can be seen, the reflectance spectra of all species vary at each of the wavelengths. At the same time, the particular features of green peaks, blue and red valleys, and near-infrared plateau remain evident and visually differentiated. Due to the complicated light environment in leaf pixels and even a natural variety of leaf colors for the same species, the reflectance of the species changed dramatically and consequently showed a wide SD band along the visible-infrared regions. The high variation of reflectance of the same materials will lead to difficulty of species classification using pixel-based methods. For example, in Figure 8, the leaf of species D.m (#13) has a white line feature distributed from the bottom to the top of the leaf rib, but species C.c (#14) has two white stripes on the leaf edges. In contrast to the all-green-leaf image of species Z.z (#5), the red spots randomly distributed over the leaf mesophylls of species A.an (#6) make it more challenging for pixel-based species classification.

As noted in the subfigures on the right column of Figure 8, four hypercube images of species A.an highlight the difference
between the global and local mean reflectance curves. The former is derived from every pixel of the whole image (the blue curve), and the latter is derived from those red leaf pixels (the red curve). In the visible region, the red-curve spectra spread far from the blue curve and locate almost close to the border of the standard deviation band of the global mean spectra. In contrast, the red curve in the infrared region distributes very close to that of the blue curve. These subfigures reveal that infrared reflectance of leaves is not correlated to the leaf color but to the chlorophyll contents and water contents in mesophyll tissues. In other words, the infrared reflectance of leaves behaves very similarly for the same species as the standard curve. According to Lin et al. (2012, 2015c), leaf reflectance over the VNIR-SWIR region may be affected by leaf water and concentration contents. Their study first highlighted the effect of water stress on chlorophyll concentration estimation and further proposed effective chlorophyll indices to account for the influence of water content to achieve accurate estimation of leaf chlorophyll concentration. Therefore, the dissimilarity in the infrared reflectance among hypercube images indicates the possibility of species difference or physiological stress such as water content shortage. Including near-infrared spectra with visible spectra is beneficial to species classification because leaf pattern features and mesophyll structure are considered simultaneously.

An Overall Assessment of Species Classification Accuracy for the Four Deep Learning Models

Table 5 shows the accuracy measures of the CNN models performing on 6 different spectral features combinations of the

| TABLE 2 | Detailed network parameters of the proposed lightweight CNN model. |
|----------|-----------------|-----------------|-----------------|-----------------|
| Parameter | Kernel size | Depth | Strides | Output |
| Input | 200 x 200 x L | 1 | 2 | 100 x 100 x 64 |
| Convolution | 5 x 5 | 1 | 2 | 100 x 100 x 64 |
| Max pooling | 3 x 3 | 0 | 2 | 50 x 50 x 64 |
| Convolution | 3 x 3 | 1 | 1 | 50 x 50 x 96 |
| Max pooling | 3 x 3 | 0 | 2 | 25 x 25 x 96 |
| Inception module1 | 2 | 25 x 25 x 288 |
| Inception module2 | 2 | 13 x 13 x 296 |
| Inception module3 | 2 | 7 x 7 x 480 |
| Global average pooling | 7 x 7 | 0 | 1 | 1 x 1 x 480 |
| FC layer | 1 | 1 x 1 x 196 |
| Dropout (40%) | 0 | 0 |
| Output | 1 | 1 x 1 x 30 |
| Softmax | 0 | 1 x 1 x 30 |

The number of parameters (Input dimension = 6): 1388950
In the classification with the natural-color RGB bands, the AlexNetr, GoogLeNet, VGGNetr achieved an OA of 76.3, 71.7, and 79.9%, a macro F1-score of 0.758, 0.699, and 0.782, and a kappa value of 0.755, 0.707, and 0.790, respectively. Each of the models had a precision of 0.832, 0.705, and 0.806, which indicates that the AlexNetr model and VGGNetr model showed better adaptability in retrieving information of species leaf features and leaf structure in RGB reflectance and therefore achieved a prediction with lower commission error or false positive than the GoogLeNet model. In contrast, the LtCNN model performed at the best accuracy with a value of OA = 89.7%, macro F1-score = 0.896, and kappa = 0.893 which is correspondingly higher than the previous models by 10–18%, 0.11–0.19, and 0.10–0.19. Using only RGB spectral features, the LtCNN model performed species classification with a commission error around 0.1.

The accuracy measures decreased significantly when checking with the classification results using three NIR bands or the simulated false-color RGB bands, as mentioned in Table 5. For example, the decrease of OA was 15, 5, 10, and 21% for the AlexNetr model, GoogLeNet model, VGGNetr model, and LtCNN model, respectively. These four DL architectures performed the species classification at an OA between 60 and 70% using only NIR-based false-color images. Obviously, the natural-color RGB images provide more diverse spectral information and inherently spatial information of the species than the false-color NIR images. Such cases are due to some species having a similar leaf mesophyll structure (Hopkins and Hüner, 2004; Lin et al., 2015c) and behaving similarly in the near-infrared bands. As shown in Figure 7, the reflectance in the visible-NIR region varied dramatically and overlapped significantly. This leads to a higher degree of omission and commission error in it. In contrast, the natural-color RGB is supposed to catch leaf color, shape, and surface texture changes and consequently contribute species classification accuracy.

In general, the reflectance of RGB bands is low correlated to the NIR bands. The leaf features derived simultaneously from the RGB natural-color bands (471.44, 535.06, and 602.94 nm) and the NIR false-color bands (750.75, 799.99, and 851.04 nm) are assumed to be of benefit to species classification. However, as noted in Table 5, the kappa coefficient achieved by the GoogLeNet model was 0.693, which is even slightly smaller than 0.707, the performance baseline achieved in the classification using only the RGB natural-color bands. In contrast, the AlexNetr, VGGNetr, and LtCNN models revealed a lively performance as the OA, F1-score, and kappa significantly increased by nearly 5%, 0.05, and 0.05, respectively. This verifies that additional NIR bands in respect to the RGB basic spectral information are beneficial to species classification.

Although the principal component analysis (PCA) method can transform the spectral information of bands in the hypercube image into several components, the classification using most informational details through the four CNN models did not perform better than the RGB bands’ baseline. For example, the F1-score of PCA and RGB for the AlexNetr, GoogLeNet, VGGNetr, and LtCNN was 0.703/0.758, 0.700/0.699, 0.778/0.782, and 0.881/0.896, respectively. The result implies that the PCA is most likely inappropriate for use in the reduction of the dimensionality of hypercube images in the view of species classification via DL. Considering the performance improvement of RGB+NIR classification, the linear transformation of hyperspectral bands most likely destroyed the physical properties of the materials in each band, thereby weakening the spatial relationship between the features or different tissues on the leaf which decreases the classification ability of a CNN.

A Comprehensive Examination of the Species Confusion in the Models

To illustrate the prediction results of the four CNN models more comprehensively, a confusion matrix is used to examine...
the confusion among the species for the classification scenario using the RGB+NIR dataset. In Figure 9, the matrix entries contain the numbers of prediction rates of the four models, which are marked with different colored squares. The value of one in diagonal entries specifies the classification of 100% from the view of ground truth. The perfect true-positive rate indicates the superior excellence of a CNN model in describing the leaf features of the plant species. As can be seen, there were 10, 4, 13, and 22 species being classified with 100% of true positive rate for the AlexNet, GoogLeNet, VGGNet, and LtCNN models, respectively. As noted in Figure 9, the species Ammocallis rosea (A.r, #3) was completely misclassified by the GoogLeNet model. It is mostly recognized as Sansevieria trifasciata (S.t, #12) with a false-negative rate of 0.6 while as Alocasia amazonica (A.am, #8), Clusia rosea (C.r, #17), Glechoma hederacea (G.h, #20), Plectranthus amboinicus cv. (P.am.cv, #22), Spathoglottis plicata (S.p, #26), and Podocarpus macrophyllus (P.m, #28) with a true-positive rate of 0.9, 0.5, 0.9, 0.7, 0.9, 0.8, 0.9, and 0.8, respectively. Poorer confidence of classification occurred in species #5 and #17. The false-negative in species #17 is mainly due to the lack of the full leaf shape in the sub-images randomly generated during the convolution, which resulted in a partial leaf and therefore increased the feature similarity of species #17, #5, and #2. Looking into the false-negative classification of species #5, whose images were misclassified as the species #4, #11, and #10 with a rate of 20, 20, and 10%, respectively. These species are
visibly differentiated based on leaf margin, surface leathery, and petiole features, but the LtCNN model misclassified 50% of the species. The misclassification is also evident in the other models. This is highly probably due to the inability of retaining leaf margin (serrate and entire), surface leathery, and petiole features during the convolution and pooling processes as the features are too small to detect with respect to the leaf area. Figure 10 illustrates some examples of the confusion in species #3, #5, #13, and #14, and the excellent recognition in species #6 and #19 for the AlexNetr, GoogLeNet, VGGNetr, and LtCNN models. It is also noted that the images of species #13 and #14 were partially misclassified by AlexNetr, GoogLeNet, and VGGNetr models mainly due to the leaf shape similarity; meanwhile, the models missed their heterogeneous features. In contrast, the LtCNN model showed excellence in successfully learning the key features, and therefore the images were classified as the species.

**DISCUSSION**

**Band Selection Contributes to Improving the Performance of Species Classification**

The accuracy figures for the species classification using the dataset with the predetermined bands of RGB, NIR, or RGB+NIR in Table 5 show the proposed LtCNN is more appropriate than the AlexNetr, GoogLeNet, and VGGNetr models. The LtCNN model misclassified species by AlexNetr, GoogLeNet, VGGNetr, and LtCNN models. And, the LtCNN is the most significant of the four models to achieve reliable and stable classification performance with a minimum number of bands and the most informative spectra. Specifically, the most appropriate spectral features for species classification are the green-edge (591.46 nm), red-edge (681.80 nm), and near-infrared (761.18 nm).

**Appropriate Dimensionality of Hyperspectral Imaging Images in Recognizing and Classifying Plant Species**

As noted in Figure 11, the four CNN models revealed diverse sensitivity of spectral bands in species classification. An
interesting question arises: Can using more bands help improve accuracy? Or what is the best accuracy achievable by the four models? To address this question this study conducted extended experiments by adding the number of bands progressively up to 60 with an interval of 3 as the input image of species classification for the CNN models. All the models are trained with the parameters mentioned in section “Experimental Setting”, except for the VGGNet model, because it was unable to handle higher dimensional data under our hardware environment. The adaptability of the AlexNetr, GoogLeNet, and LtCNN models to high-dimensional data is shown in Figure 12.

The x-axis presents the number of bands (M) in each subfigure, and the y-axis shows the corresponding accuracy measure. The yellow, blue, and red curves denote the accuracy trends of the three CNN models, respectively. From the point of view of the species classification, the main observation is that increasing M cannot help to improve accuracy and may even cause worse results. This phenomenon mainly occurred when using AlexNetr and GoogLeNet. The impacts of M on the CNN models are summarized below.

(1). For the cases of UBS, as shown in Figure 12A, the kappa coefficient of AlexNetr starts at 0.624 and increases to 0.779–0.800 when M = 6 to 18. As M increases, the accuracy is no longer improved but becomes unstable. The kappa of GoogLeNet starts at 0.797 and gradually decreases
when M increases. For the proposed LtCNN model, the increase of M does not cause a significant change in the classification accuracy. The best value appears at $M = 15$, which is 0.976. After that, the accuracy curve maintains between 0.941 and 0.969. It implies that the amount of spectral information is saturated. On average, the kappa accuracy for the AlexNetr, GoogLeNet, and LtCNN models was 0.717 ± 0.079, 0.666 ± 0.084, and 0.951 ± 0.021, respectively.

(2). Similarly, as shown in Figure 12B for the cases of FNGBS, the trends generated by AlexNetr and GoogLeNet are gradually declining when M increases or fluctuates between 0.503 and 0.866 and 0.538 and 0.807, respectively. On the contrary, the LtCNN model can maintain accuracy between 0.92 and 0.973 and is not sensitive to M. Each of the three models was averaged 0.696 ± 0.096, 0.674 ± 0.081, and 0.940 ± 0.018.

(3). From the point of view of the amount of spectral information, it is evident that using a sufficient number of bands can achieve the highest accuracy. For example, the LtCNN model obtained 0.976 with $M = 15$ selected by UBS in Figure 12A and 0.972 with $M = 18$ recommended by FNGBS in Figure 12B. The other two CNNs models also follow the same fashion. This proves that using hyperspectral imaging for species classification can obtain good results without too many bands.

**Comparison of Conventional Neural Network Models**

The AlexNetr and GoogLeNet models produced lower, unstable, and downward accuracy in the plant classification is most likely due to two reasons. Firstly, they were designed and specialized for handling large databases with a large number of categories. ...
FIGURE 11 | A comparison of the sensitivity of spectral bands for species classification in the four CNN models. Charts (A,B) show the kappa coefficients with 3-band and 6-band classifications recommended by the band selection methods. A considerable difference of kappa values in any classifications with different datasets indicates higher sensitivity of the CNN models.

FIGURE 12 | The relationship between the number of bands (M) and produced accuracy when using a subset of hypercube images for classification. Line charts (A,B) present the kappa variation for the uniform band selection (UBS) and Fast Neighborhood Grouping Band Selection (FNGBS) method. The numbers below the M list the kappa values of the CNN models.

and training images. Since the tested plant database has only 30 classes and 1,500 images, the training data is relatively insufficient for them. Besides, the nature of the plant image is distinct from the objects’ colors, shapes, and patterns for which the models were originated. This may explain why AlexNet and GoogLeNet produce lower accuracy performance. Secondly, when M increases, the inter-band correlation of data increases. The input data with excessive redundant information may further interfere with the training process of the more extensive network under insufficient training data. This additionally imposes the difficulty of getting convergence in network training. This may explain why AlexNet and GoogLeNet produce an unstable performance at different M values. In contrast, the proposed model LtCNN adopts a simplified architecture that is optimized for smaller datasets and significantly performs better than the other two in both accuracy and stability. This emphasizes the importance of designing a dedicated network for processing a particular dataset. And it suggests that as long as the network design is correct, it will not be too sensitive to data redundancy. Meanwhile, it can also be efficient with minimum bands to
achieve satisfactory accuracy. Such a conclusion is significant for dealing with hyperspectral images.

**Limitation and Opportunity**

The main strength of the proposed plant image classification method is that it uses the abundant spectral information provided by HSI, and uses the “deep features” learned by CNNs for plant species classification. However, this approach suffers from some drawbacks and limitations. Firstly, the cost of hyperspectral cameras is high so it is difficult for it to become widely adopted. Its long-shooting time also limits data acquisition and the possibilities for in-field investigations. Secondly, limited by the existing network architecture of CNN and memory size of GPU, it is hard to use high-resolution HSI images as the material to learn the more comprehensive features. Thirdly, our framework relies on band selection to reduce the data dimensionality. Ideally, we can feed all the bands into CNN and let it learn the discriminative bands automatically. However, limited by the network architecture, memory size, and the amount of training data, it is temporarily impossible to achieve. Finally, if we want to increase the species, except for adding image data to expand the database, it is also necessary to expand the network architecture and retrain the model. This is one of the crucial shortcomings of the current CNN approaches.

Even with these limitations, we are confident that this study will contribute to educational use as well as to the development of plant identification and forest remote sensing. One of the critical findings of this study is that applying only green-edge, red-edge, and near-infrared bands can substantially improve the species classification via the proposed model LtCNN. This finding provides an additional opportunity for sensor design specifically for plant applications and therefore benefits the imaging technology development for plant science research and education at a lower investment cost.

**CONCLUSION**

From the point of view of individual tree recognition and mapping, this study applied hyperspectral imaging to build a plant image dataset via a VNIR imaging system with a spectral resolution of 2.8 nm and a radiometric resolution of 10 bits. The plant dataset contains 1500 images accounting for the crown and leaf features of 30 species. The plant images show dramatic reflectance values over the spectral range from the visible to the near-infrared region and therefore reveal the dilemma of pixel-based plant classification via remote sensing images. Although a pixel-based inspection of plant images reveals that diverse leaf colors increase the difficulty of plant classification using merely visible spectra, the near-infrared reflectance of colorful leaves of the same species remains very similar and behaves homogeneously and stably. In contrast to the variation of visible spectra, the species consistency of near-infrared spectral features provides an optimistic opportunity for plant classification.

According to the results, the complex deep learning architecture of AlexNetr, GoogLeNet, and VGGNetr models are not suitable for plant classification using a limited number of training samples and therefore failed to obtain satisfactory performance when integrating the features in 3-band RGB and 3-band NIR bands. Correspondingly, the best kappa accuracy for these models was 0.817, 0.693, and 0.834. The proposed lightweight conventional neural network, the LtCNN model, however, achieved an optimistic kappa accuracy of 0.945. Interestingly, this novel model has demonstrated its excellence in retrieving critical features from limited training samples through three bands suggested by the fast neighborhood grouping band selection method. The classification using the bands of green-edge (591.46 nm), red-edge (681.80 nm), and near-infrared (761.18 nm), the LtCNN model can achieve a kappa accuracy of 0.945, a value equal to the accuracy of a classification using 6 bands of RGB and NIR. Because the accuracy is very close to the maximum accuracy of 0.976, the best performance with 15 spectral bands of the hyperspectral images, the LtCNN model is concluded to be very efficient and reliable in classifying plant images. It is also concluded that a feature selection should be implemented before applying hyperspectral images to plant classification to reduce training cost and hardware loading significantly.

Many studies developed deep learning techniques for plant classification based on single-leaf images. This study is devoted to exploring an appropriate method for recognizing and classifying plant species according to live-crown and leaf features. Although the hyperspectral imaging technique can provide a hyperspectral dataset with critical spectral features for the application, some false-positive and false-negative errors still occurred in some species by the AlexNetr, GoogLeNet, VGGNetr, and LtCNN models simultaneously. These species are visual recognizably based on the features of leaf margin, surface leathery, and petiole. Developing a new network model with a 3D-CNN module should enhance feature learning in the spectral domain. The ability to retrieve tiny leaf features would also be an essential task for the future.

**DATA AVAILABILITY STATEMENT**

The raw data supporting the conclusions of this article will be available at the following link: https://github.com/asufdhkjlj456/Plant_Classification_with_HSI_and_DL.

**AUTHOR CONTRIBUTIONS**

CL and K-HL: conceptualization, project administration, methodology, and writing – original draft. K-HL, M-HY, and S-TH: data curation and formal analysis. CL: writing – review and editing. All authors contributed to the article and approved the submitted version.

**FUNDING**

This work was supported by the Ministry of Science and Technology, Taiwan, in Grant No.: MOST 109-2221-E-110-047-MY2 and 109-2221-E-415-003, and in part by the Council of Agriculture of the Executive Yuan, Taiwan, under Grant 110AS-8.3.2-ST-a7.
Lin, C., Tsogt, K., and Chang, C. I. (2012). An empirical model-based method for signal restoration of SWIR in ASD field spectroradiometry. *Photogramm. Eng. Remote Sens.* 78, 119–127. doi: 10.14358/PERSS.78.2.119

Lin, C., Tsogt, K., and Zandraabal, T. (2016). A decompositional stand structure analysis for exploring stand dynamics of multiple attributes of a mixed-species forest. *For. Ecol. Manage.* 378, 111–121. doi: 10.1016/j.foreco.2016.07.022

Liu, Z., Yan, Q., Zhang, D., and Li, Q. L. (2007). Automated tongue segmentation in hyperspectral images for medicine. *Appl. Opt.* 46, 8328–8334. doi: 10.1364/ao.46.08328

Ma, J., Pu, H., Sun, D.-W., Gao, W., Qu, J.-H., and Ma, K.-Y. (2015). Application of Vis–NIR hyperspectral imaging in classification between fresh and frozen-thawed pork Longissimus Dorsi muscles. *Int. J. Refrig.* 50, 10–18. doi: 10.1016/j.ijrefrig.2014.10.024

Marshall, S., Kelman, T., Qiao, T., Murray, P., and Zabalza, J. (2015). “Hyperspectral imaging for food applications,” in *Proceedings of the 2015 23rd European Signal Processing Conference (EUSIPCO)*, (Piscataway, NJ: IEEE), 2854–2858. doi: 10.1109/EUSIPCO.2015.7362906

Mirrazi, M., Marofi, S., Abbasi, M., Solgi, E., Karimi, R., and Verrelst, I. (2019). Scenario-based discrimination of common grapevine varieties using in-field hyperspectral data in the western of Iran. *Int. J. Appl. Earth Obs. Geoinf.* 80, 26–37. doi: 10.1016/j.jaeg.2019.04.002

Nasiri, A., Taheri-Garavand, A., Fanourakis, D., Zhang, Y. D., and Nikoloudakis, N. (2021). Automated grapevine cultivar identification via leaf imaging and deep convolutional neural networks: a proof-of-concept study employing primary iranian varieties. *Plants* 10:1628. doi: 10.3390/plants10081628

Nicolai, B. M., Lötze, E., Peirs, A., Scheerlinck, N., and Theron, K. I. (2006). Non-destructive measurement of bitter pit in apple fruit using NIR hyperspectral imaging. *Postharvest Biol. Technol.* 40, 1–6. doi: 10.1016/j.postharvbio.2005.12.006

Rapaport, T., Hochberg, U., Shoshany, M., Karniel, A., and Rachmilevitch, S. (2015). Combining leaf physiology, hyperspectral imaging and partial least squares regression (PLS-R) for grapevine water status assessment. *ISPRS J. Photogramm. Remote Sens.* 109, 88–97. doi: 10.1016/j.isprsjprs.2015.09.003

Salman, A., Semwal, A., Bhatt, U., and Thakkar, V. M. (2017). “Leaf classification and identification using Canny Edge Detector and SVM classifier,” in *Proceedings of the 2017 International Conference on Inventive Systems and Control (ICISC)*, (Piscataway, NJ: IEEE), 1–4. doi: 10.1109/ICISC.2017.8068597

Santos, F., Meneses, P., and Hostert, P. (2019). Monitoring long-term forest dynamics with scarce data: a multi-class data classification implementation in the Amazonian Ecuador. *Eur. J. Remote Sens.* 52(Suppl. 1), 62–78. doi: 10.1080/22797254.2018.1533793

Schmitter, P., Steinruecken, J., Roemer, C., Ballvora, A., Leon, J., Rascher, U., et al. (2017). Unsupervised domain adaptation for early detection of drought stress in hyperspectral images. *ISPRS J. Photogramm. Remote Sens.* 131, 65–76. doi: 10.1016/j.isprsjprs.2017.07.003

Simonyan, K., and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *arXiv* [preprint]. *arXiv:1409.1556*, doi: 10.3390/s21082852

Sinha, P., Robson, A., Schneider, D., Kilic, T., Mugera, H. K., Ilukor, J., et al. (2020). The potential of in-situ hyperspectral remote sensing for differentiating 12 banana genotypes grown in Uganda. *ISPRS J. Photogramm. Remote Sens.* 167, 85–103. doi: 10.1016/j.isprsjprs.2020.06.023

Sun, J., Zhou, X., Hu, Y., Wu, X., Zhang, X., and Wang, P. (2019). Visualizing distribution of moisture content in tea leaves using optimization algorithms and NIR hyperspectral imaging. *Comput. Electron. Agric.* 160, 153–159. doi: 10.1016/j.compag.2019.03.004

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al. (2015). “Going deeper with convolutions,” in *Proceedings of the IEEE conference on CVPR* , (Piscataway, NJ: IEEE), 1–9. doi: 10.1109/CVPR.2015.7298594

Teena, M. A., Manickavasagan, A., Ravikanth, L., and Jayas, D. S. (2014). Near infrared (NIR) hyperspectral imaging to classify fungal infected date fruits. *J. Stored Prod. Res.* 59, 306–313. doi: 10.1016/j.jspr.2014.09.005

Ubbens, J. R., and Stavness, I. (2017). Deep plant phenomenics: a deep learning platform for complex plant phenotyping tasks. *Front. Plant Sci.* 8:1190. doi: 10.3389/fpls.2017.01190

Wang, Q., Li, Q., and Li, X. (2020). A fast neighborhood grouping method for hyperspectral band selection. *IEEE Trans. Geosci. Remote Sens.* 59, 5028–5039. doi: 10.1109/TGRS.2020.3011002

Wang, Y., Qin, Y., and Cui, J. (2021). Occlusion robust wheat ear counting algorithm based on deep learning. *Front. Plant Sci.* 12:65899. doi: 10.3389/fpls.2021.65899

Yang, G., Chen, G., Li, C., Fu, J., Guo, Y., and Liang, H. (2021). Convolutional rebalancing network for the classification of large imbalanced rice pest and disease datasets in the field. *Front. Plant Sci.* 12:671134. doi: 10.3389/fpls.2021.671134

Zhang, B., Wu, D., Zhang, L., Jiao, Q., and Li, Q. (2012). Application of hyperspectral remote sensing for environment monitoring in mining areas. *Environ. Earth Sci.* 65, 649–658. doi: 10.1007/s12665-011-112-y

Zhang, H., Yanne, P., and Liang, S. (2012). “Plant species classification using leaf shape and texture,” in *Proceedings of the 2012 International Conference on Industrial Control and Electronics Engineering*, (Piscataway, NJ: IEEE), 2025–2028. doi: 10.1109/ICICEE.2012.538

Zhang, X. D. (2020). A Matrix Algebra Approach to Artificial Intelligence, (Singapore: Springer), 223–440. doi: 10.1007/978-981-15-2770-8_6

Zhu, Y., Sun, W., Cao, X., Wang, C., Wu, D., Yang, Y., et al. (2019). TA-CNN: two-way attention models in deep convolutional neural network for plant recognition. *Neurocomputing* 365, 191–200. doi: 10.1016/j.neucom.2019.07.016

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Liu, Yang, Huang and Lin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.