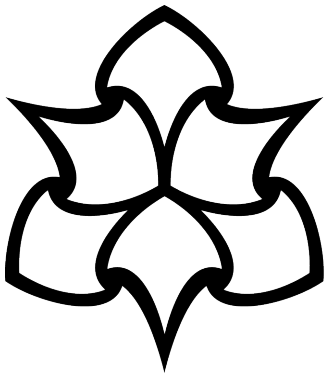

MDPI

Article

A Novel CropdocNet for Automated Potato Late Blight Disease Detection from the Unmanned Aerial Vehicle-based Hyperspectral Imagery

Yue Shi [1], Liangxiu Han [1,*], Anthony Kleerekoper[1], Sheng Chang [2] and Tongle Hu [3]

[1] Department of Computing and Mathematics, Faculty of Science and Engineering, Manchester Metropolitan University, Manchester M1 5GD, UK.;
[2] State Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101, China;
[3] College of Plant Protection, Hebei Agriculture University, Baoding 070001, China;
* Correspondence: L.Han@mmu.ac.uk

Abstract: Accurate and automated diagnosis of potato late blight disease, one of the most destructive potato diseases, is critical for precision agricultural control and management. Recent advances in remote sensing and deep learning offer the opportunity to address this challenge. This study has proposed a novel end-to-end deep learning model (CropdocNet) for accurate and automated late blight disease diagnosis from UAV-based hyperspectral imagery. The proposed method considers the potential disease specific reflectance radiation variance caused by the canopy structural diversity, introduces the multiple capsule layers to model the part-to-whole relationship between spectral-spatial features and the target classes to represent the rotation invariance of the target classes in the feature space. We have evaluated the proposed method with the real UAV-based HSI data under the controlled and natural field conditions. The effectiveness of the hierarchical features has been quantitatively assessed and compared with the existing representative machine learning/deep learning methods on both testing and independent datasets. The experiment results show that the proposed model significantly improves the accuracy performance when considering hierarchical-structure of spectral-spatial features, with the average accuracies of 98.09% for the testing dataset and 95.75% for an independent dataset, respectively.

Keywords: Potato late blight; Automated crop disease diagnosis; UAV-based hyperspectral imagery; deep learning; classification

Citation: Title. *Remote Sens.* **2021**, *1*, 0. https://doi.org/

Received:
Accepted:
Published:

1. Introduction

Potato late blight disease, caused by *Phytophthora infestans (Mont.)* de Bary, is one of the most destructive potato diseases, resulting in significant potato yield loss across the major potato growing areas worldwide [1,2]. The yield loss to the infestation of late blight disease is around 30% to 100% [3,4]. The current control measure mainly relies on application of fungicides [5], which is expensive and has negative impacts on the environment and human health due to excessive use of pesticides. Therefore, early accurate detection of potato late blight disease is vital for effective disease control and management with minimal application of fungicides.

Since the late blight disease affects the potato leaves, stems and tubers with visible symptoms (e.g. black lesions with granular regions and green halo) [6,7], the current detection of late blight disease in practice is mainly based on the visual observation [8,9]. However, this manual inspection method is time consuming and costly, and often causes a delay in the late blight disease management, especially at an early stage across large fields [10]. In addition, the field surveyors diagnose the diseases based on their domain knowledge, which may introduce inconsistency and bias due to individual subjectivity

[11]. An automated approach for fast and reliable potato late blight disease diagnose is important to ensure effective disease management and control.

With the advancements in low-cost sensor technology, computer vision and remote sensing, machine vision technology based on images (such as the red, green and blue (RGB) images, thermal images, multispectral and hyperspectral images) have been successfully used in agricultural and engineering fields [12–21]. For example, Wu *et al.* [20] developed a deep learning based model to detect the edge images of the flower buds and inflorescence axes, and successfully applied this algorithm to the banana bud-cutting robot for real-time operation. Cao *et al.* [21] developed a multi-objective particle swarm optimizer for a multi-objective trajectory model of the manipulator, which has improved the stability of the fruit picking manipulator and facilitated the nondestructive picking. Particularly, in the area of automated crop disease diagnosis [22,23], Unmanned Aerial Vehicles (UAVs) equipped with RGB camera and thermal sensors have been used for plant physiological monitoring (e.g. transpiration, leaf water, etc) [13]. Li *et al.*[24] acquired the potato biomass associated spatial and spectral features from the UAV based RGB and hyperspectral imagery, respectively, and then they fed them into a random forest (RF) model to predict the potato yield. Wan *et al.*[25] fused the spectral and structural information from multispectral imagery into a multi-temporal vegetation index model to predict the rice grain yield.

In addition, with the advancements in remote sensing technologies, the remote sensing-based vison technology has shown great potential for agricultural control and management, especially for automatic crop disease diagnosis [22,23]. The existing remote sensing-based computer vision model are developed based on the characteristics of the images (such as the red, green and blue (RGB) images, thermal images, multispectral and hyperspectral images) [12–16]. For instance, Unmanned Aerial Vehicles (UAVs) equipped with RGB camera and thermal sensors have been used for plant physiological monitoring (e.g. transpiration, leaf water, etc) [13]. Li *et al.*[24] acquired the potato biomass associated spatial and spectral features from the UAV based RGB and hyperspectral imagery, respectively, and then they fed them into a random forest (RF) model to predict the potato yield. Wan *et al.*[25] fused the spectral and structural information from multispectral imagery into a multi-temporal vegetation index model to predict the rice grain yield.

Benefiting from many more narrow spectral bands over a contiguous spectral range, hyperspectral imagery (HSI) provides spatial information in two dimensions and rich spectral information in the third one, which captures detailed spectral-spatial information of the disease infestation with the potential to provide better diagnostic accuracy [26,27]. However, how to extract the effective infestation features from the abundant spectral and spatial information from hyperspectral images is a key challenge for disease diagnosis. Currently, based on the features used in the HSI-based disease detection, the existing models can be divided into three categories: ***spectral feature-based approaches*** focusing on spectral signatures composed by the associated radiation signal of each pixel of image scene in various spectral ranges [28–30]; ***spatial feature-based approaches*** focusing on features such as shape, texture and geometrical structures [31–34], and ***the joint spectral-spatial feature-based approaches*** focusing on combination of spectral and spatial features [35–42]. The detailed discussion on these methods can be found in Section 2 of Related work.

Despite existing works are encouraging, the existing models do not consider the hierarchical structure of the spectral and spatial information of the crop diseases (for instance, canopy structural information and reflectance radiation variance of the ground objects hidden in HSI data), which are important indicators for crop disease diagnosis. In fact, changes on reflectance due to plant pathogens and plant diseases are highly disease-specific since the optical properties of plant diseases are related to a number of factors such as foliar pathogens, canopy structural information, pigment content, etc..

Therefore, to address the issue above, the hierarchical structure of the spectral-spatial features should be considered in the learning process. In this paper, we propose a novel CropdocNet for automated detection and discrimination of potato late blight disease. The contributions of the proposed work include:

1) Development of an end-to-end deep learning framework (CropdocNet) for potato disease detection;

2) Proposal of introducing multiple capsule layers to handle the hierarchical structure of the spectral-spatial features extracted from HSIs;

3) Combination of the spectral-spatial features to represent the part-to-whole relationship between the deep features and the target classes (i.e. healthy potato and the potato infested with late blight disease).

The remaining part of this paper is organized as follow: Section 2 describe the related work, Section 3 describes the study area, data collection, and the proposed model; Section 4 presents the experimental results; Section 5 provides discussions, and Section 6 summarizes this work and highlights the future works.

2. Related work in crop disease detection based on hyperspectral imagery

In this section, we mainly discuss related work in crop disease detection based on hyperspectral imagery (HSI). Based on features used for HSI-based crop disease detection, there are broadly three main categories including: ***spectral feature-based approaches***; ***spatial feature-based approaches*** and ***the joint spectral-spatial feature-based approaches***. Table. 1 summarizes the existing models on potato late blight disease detection based on different features used in the machine learning process, which provides a baseline for the hyperspectral imagery based late blight disease detection. The detailed review for each class are described below:

Table 1: The existing models comparison on potato late blight disease detection.

Approach type	Model name	Calssification accuracy	Observation scale	Reference
Spectral feature-based	Support vector machien (SVM)	84%	Leaf	[2]
	Partial least square discriminant analysis (PLSDA)	82.1%	Leaf	[6]
	Multiclass support vector machine (MSVM)	87.5%	Canopy	[11]
Spatial feature-based	Random forest (RF)	79%	Leaf	[2]
	Texture segmentation (TS)	86%	Leaf	[8]
	Simplex Volume Maximization (SiVM)	88.5%	Canopy	[10]
Spectral-spatial feature-based	Full convolutional network (FCN)	88.9%	Leaf	[6]
	3D convolutional network (3DCNN)	85.4%	Canopy	[22]

In the category of ***spectral feature-based approaches***, it exploits the spectral features associated with plant diseases, which represents the biophysical and biochemical status of the plant leaves from the spectral domain of HSI [28–30]. For example, Nagasubramanian *et al.* [43] found that the spectral bands associated to the depth of chlorophyll absorption is very sensitive to the occurrence of plant diseases, and they extracted the optimal spectral bands as the input of the Genetic Algorithm (GA) based SVM for early identification of charcoal rot disease in soybean, with a 97% classification accuracy. Huang *et al.* [44] extracted twelve sensitive spectral features for Fusarium head blight, which were then fed into a SVM model to diagnose the Fusarium head blight severity with good performance.

For the category of ***spatial feature-based approaches***, it exploits the spatial texture of the hyperspectral image, which represents the foliar contextual variances, such as the colour, density, and leaf angle, and is one of important factors for crop disease diagnosis [31–34]. For example, Mahlein *et al.* [45] summarized the spatial features of the RGB, multi-spectral, and hyperspectral images used in the automatic detection of disease detection. Their study showed that the spatial properties of the crop leaves were affected by leaf chemicals parameters (e.g., pigments, water, sugars, etc.) and light reflected from internal leaf structures. For instance, the spatial texture of the hyperspectral bands from 400 to 700 nm is mainly influenced by foliar content, and the spatial texture of the bands from 700 to 1,100 nm reflects the leaf structure and internal scattering processes. Yuan *et al.* [46] introduced the spatial texture of the satellite data into the spatial angle mapper (SAM) to monitor wheat powdery mildew at the regional level.

In the category of ***the joint spectral-spatial feature-based approaches***, there are two main strategies for extracting joint spectral-spatial features to represent the characteristics of the crop diseases in HSI data. The first strategy is to extract spatial and spectral features separately and then combine them together based on the 1D or 2D approaches (e.g. feature stacking, convolutional filters, etc) [40–42]. For example, Xie *et al.* [47] investigated the spectral and spatial features extracted from hyperspectral imagery for detecting early blight disease on eggplant leaves, and they then stacked these features as the input of an AdaBoost models to detect the healthy and infected samples. The second strategy is to jointly extract the correlated spectral-spatial information of the HSI cube through the 3D kernel based approaches [48–50]. For instance, Nguyen *et al.* [51] tested the performance of the 2D convolutional neural network (2D-CNN) and 3D convolutional neural network (3D-CNN) for early detection of grapevine viral diseases. Their findings demonstrated that the 3D convolutional filter was able to produce promising results over the 2D convolutional filter from hyperspectral cubes. Benefiting from the advanced self-learning performance of 3D convolutional kernel, the depth of the 3D convolutional kernel has also been investigated for crop disease diagnosis [35–39]. For instance, Suryawati *et al.* [52] compared the CNN baselines with the depth of 2, 5, and 13 3D-convolutional layers, their findings suggested that the deeper architecture achieved the higher accuracy for the plant disease detection tasks. Nagasubramanian *et al.* [53] developed a 3D deep convolutional neural network (DCNN), with 8 3D convolutional layers, to extract the deep spectral-spatial features for representing the inoculated stem images from the soybean crops. Kumar *et al.* [54] proposed a 3D convolutional neural network (CNN) with 6 3D convolutional layers to extract the spectral-spatial features for various crop diseases.

However, these existing methods fail to model the various reflectance radiation of the crop disease and the hierarchical structure of the disease specific features, which are affected by the particular combination of multiple factors, such as the foliar biophysical variations, the appearance of typical fungal structures, canopy structural information, from region to region [27]. A reason behind is that the convolutional kernels in the existing CNN methods are independent to each other, which is hard to model the part-to-whole relationship of the spatial-spatial feature and to characterize the complexity and diversity of the potato late blight disease on HSI data [36]. Therefore, this study proposes a novel end-to-end deep learning model to address the limitations with consideration of the hierarchical structure of spectral-spatial features associated with plant diseases.

3. Materials and methods

3.1. Data acquisition

3.1.1. Study site

The field experiments were conducted at three experimental sites (see Fig. 1), wherein, the first two sites were conducted under the controlled conditions to collect the high-quality labelled data for model training, and the third site was conducted under the natural conditions to obtain an independent dataset for model evaluation. All of the experiments were located in Guyuan county, Hebei province, China. The detailed information for each experimental site are descripted below:

The site 1 was located at (41°41′2.41″ N, 115°44′47.39″ E). The potato cultivars, 'Yizhangshu No.12' and 'Shishu No.1', were selected due to their different susceptibility to late blight infestation. There were two control groups and four infected groups of late blight were applied. Each field group occupied 410 m^2 of field campaigns. Seedlings of these cultivars were inoculated with late blight on 13 May, 2020. The spores concentration of 9 $mg100^{-1}mL^{-1}$ was used. A total of nine $1m \times 1m$ observation plots were set for the ground truth data investigation (see Fig. 1). There are two reasons for using $1m \times 1m$ observation plots: 1) it allows for collecting the canopy spectral-spatial variations of the potato leaves; 2) it enables easy identification of the same patches on hyperspectral

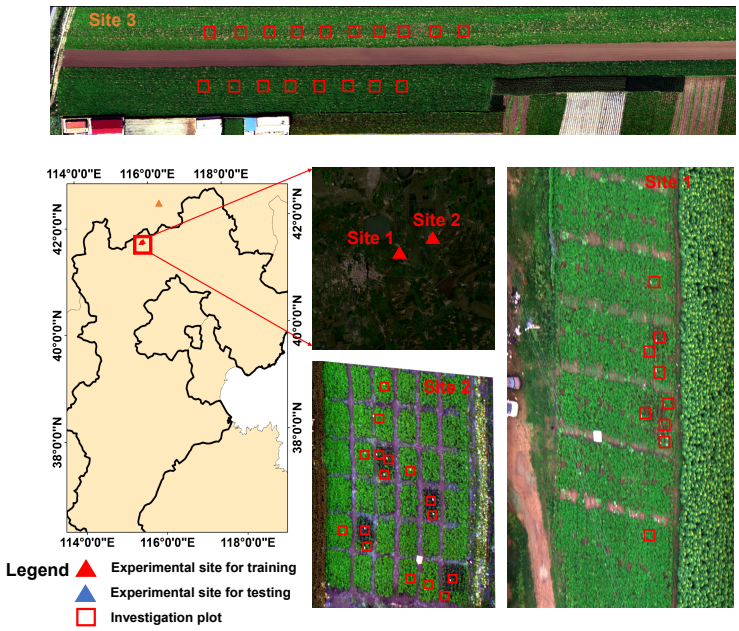

Figure 1. The experimental sites in Guyuan, Hebei province, China.

images to ensure the right match between the ground truth investigation patches and the pixel-level labels. The field observations were conducted on August 16, 2020.

The site 2 was located at (41°42′2.4″ N, 115°47′44.39″ E). The potato cultivars as same as site 1 were selected. There were six control groups and thirty infected groups of late blight were applied. Each field group occupied 81 m^2 of field campaigns. Seedlings of these cultivars were inoculated with late blight on 14 May, 2020. In the infected groups, the spores concentration of 9 $mg100^{-1}mL^{-1}$ was used. A total of eighteen $1m \times 1m$ observation plots were set for the ground truth data investigation. The field observations were conducted on August 18, 2020.

The site 3 was located at (42°34′1.12″ N, 115°46′52.39″ E). The potato cultivar,'Shishu No.1', was selected. The late blight disease naturally occurred in this experimental site under natural conditions. A total of eighteen $1m \times 1m$ in-situ observation plots were set for the ground truth data investigation. The field observations were conducted on August 20, 2020.

3.1.2. Ground truth disease investigation

Four types (classes) of ground truth data were investigated including: healthy potato, late blight disease, soil, and background (i.e. the roof, road, and other facilities). Among which, the classes of soil and background can be easily labelled based on visual investigation from the UAV HSI. For the classes of healthy potato and late blight disease, we firstly investigated the disease ratio (i.e. the diseased area / the total leaf area) of the experiment sites based on National Rules for Investigation and Forecast Technology of the Potato Late Blight ($NY/T1854-2010$). Then, we labelled diseased ratio in a sampling plot lower than 7% as a healthy potato class, otherwise it was labelled as a diseased class. The reason for choosing the threshold of 7% is mainly because the hyperspectral signal and the spatial texture of the potato leaves with the disease ratio lower than 7% are indistinguishable from the healthy leaves in our HSI data (with the spatial resolution of 2.5cm).

3.1.3. UAV-based HSIs collection

The UAV-based HSIs were collected by Dajiang (DJI) S1000 (ShenZhen (SZ) DJI Technology Co Ltd., Gungdong, China) equipped with UHD-185 Imaging spectrometer (Cubert GmbH, Ulm, Baden-Warttemberg, Germany). The collected HSI imagery cover-

ing the wavelength ranges from 450 nm to 950 nm with 125 bands. In the measurements, a total of 23 HSIs (the overlap rate is set as 30% to avoid the mosaicking errors [55]) were mosaicked to cover the experiment site 1 and the full size for experimental site 1 is 16382×8762 *pixels*. A total of 16 HSIs were mosaicked to cover the experiment site 2 and the full size for experimental site 2 is 8862×7625 *pixels*. A total of 14 HSIs were mosaicked to cover the experiment site 3 and the full size for experimental site 2 is 15822×6256 *pixels*. All of the UAV-based HSI data were collected between the 11:30 a.m. and 13:30 p.m. under a cloud-free condition. The spatial resolution of the HSI is 2.5 cm, with a height of 30 m. HSI data were manually labelled based on the ground truth investigations. The HSIs for experimental site 1 and site 2 were used as training dataset for model training and cross-validation, the HSI for experimental site 3 was used as an independent dataset for model evaluation.

3.2. The proposed CropdocNet model

Since the traditional convolutional neural networks extract the spectral-spatial features without considering the hierarchical structure representations among the features, it may lead to a suboptimal performance on characterizing the part-to-whole relationship between the features and the target classes. In this study, inspired by the dynamic routing mechanism of capsules [56], the proposed CropdocNet model introduces multiple capsule layers (see below) with the aim to model the effective hierarchical structure of spectral-spatial details and generate the encapsulated features to represent the various classes and the rotation invariance of the disease attributes in the feature space for accurate disease detection.

Essentially, **the design rationale behind our proposed approach** is that, unlike the traditional CNN methods extracting the abstract scalar features to predict the classes, the spectral-spatial information extracted by the convolutional filters in the form of scalars are encapsulated into a series of hierarchical class-capsules to generate the deep vector features, which represents the specific combination of the spectral-spatial features for the target classes. Based on this rationale, the length of the encapsulated vector features represent the membership degree of an input belonging to a class, and the direction of the encapsulated vector features represent the consistency of the spectral-spatial feature combination between the labelled classes and the predicted classes.

Fig. 2 shows the proposed framework, which consists of ***a spectral information encoder***, ***a spectral-spatial feature encoder***, ***a class-capsule encoder***, and ***a decoder***.

Specifically, the proposed CropdocNet firstly extracts the effective information from the spectral domain based on the 1-D convolutional blocks, and then, it encodes the spectral-spatial details around the central pixels by using the 3-D convolutional blocks. Subsequently, these spectral-spatial features are sent to the hierarchical structure of the class-capsule blocks in order to build the part-to-whole relationship, and to generate the hierarchical vector features for representing the specific classes. Finally, a decoder is employed to predict the classes based on the length and direction of the hierarchical vector features in the feature space. The detailed information for the model blocks are described below.

3.2.1. Spectral information encoder

The spectral information encoder, located at the beginning of the model, is set to extract the effective spectral information from the input HSI data patches. It is composed of a serial connection of two 1D convolutional layers, two batch normalization layers, and a ReLu layers.

Specifically, as shown in Fig. 2, the HSI data with H rows, W columns, and B bands, denoted as $X \in \mathbb{R}^{H \times W \times B}$, can be viewed as a sample set with $H \times W$ pixel vectors. Each of the pixels represents a class. And then, the 3-D patches with a size of $d \times d \times B$ around each pixel are extracted as the model input, where d is the patch size. In this study, d is

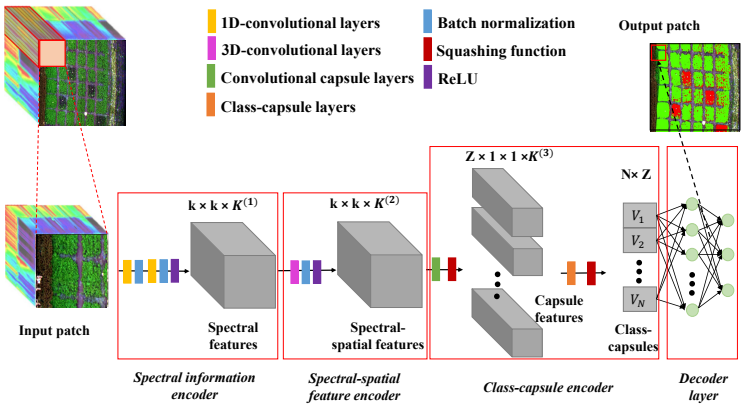

Figure 2. The workflow of the CropdepcNet framework for potato late blight disease diagnosis (k is the spatial size of the convolutional kernel, K is the number of the channel of the convolutional kernel, Z is the dimensionality of the class-capsule, N is the number of the class-capsule, and V representing a vector of the high-level features)

set as 13 so that the input patch is able to capture at least on intact potato leaf. These patches are labelled with the classes same as their central pixels.

Subsequently, the joint 1D convolution and batch normalization series, which receive the data patch from the input HSI cube, are introduced to extract the radiation magnitude of the central band and their neighbours bands. A total of $K^{(1)}$ convolutional kernels with a size of $1 \times 1 \times L_{rf}$ are employed by the 1D convolutional layer, where, L_{rf} is the length of the receptive field for the spectral domain. The 1D convolutional layer is calculated as follows:

$$C_p^j = \sum_{l=1}^{L_{rf}} W_l^j I_l^p \tag{1}$$

where C_p^j is the intermediate output of the p^{th} neuron with the j^{th} kernel, W_l^j is the weight for the l^{th} unit of the j^{th} kernel, and I_l^p is the feature value of the l^{th} unit corresponding to the p^{th} neuron.

The second 1D convolution and batch normalization series are used to extract the abstract spectral details from the low-level spectral features. Finally, a ReLu activation function is used to obtain a spectral feature output denoted as $\mathbf{X}_{out}^1 \in \mathbb{R}^{H \times W \times K^{(1)}}$.

3.2.2. Spectral-spatial feature encoder

The spectral-spatial feature encoder is located after the spectral information encoder, which aims to arrange the extracted spectral features in $\mathbf{X}_{out}^1$ into the joint spectral-spatial features that feed to the subsequent capsule encoder. Firstly, a total of $K^{(2)}$ global convolutional operations are used on the $\mathbf{X}_{out}^1$ with a kernel size of $c \times c \times K^{(1)}$, where c is the kernel size, which is set as 13 in order to match the size of input patch. Then, the a batch normalization step and a ReLu activation function is used to generate the output volume $\mathbf{X}_{out}^2 \in \mathbb{R}^{H \times W \times K^{(2)}}$.

3.2.3. Class-capsule encoder

The class-capsule encoder, the most important module of the proposed network, is introduced to generate the hierarchical features to represent the translational and rotational correlations between the low-level spectral-spatial and the target classes of healthy and diseased potato. It comprises two layers: a feature encapsulation layer and a class-capsule layer.

Specifically, The feature encapsulation layer consists of Z convolutional-based capsule units, and each of the capsule unit composed by K convolutional filters, and the size of each filter is $\mathrm{k} \times \mathrm{k} \times K^{(3)}$. In the training process, the $\mathbf{X}^2_{out}$ from the spectral-spatial feature encoder will input into a series of capsules units to learn the potential translational and rotational structure between the features in $\mathbf{X}^2_{out}$. An output vector $\mathbf{u}^{(\mathbf{m})} \in \mathbb{R}^K = [u_1^{(m)}, u_2^{(m)}, \cdots, u_K^{(m)}]$ would be generated by the K convolutional kernel of m^{th} capsule. The orientation of the output vector represents the class-specific hierarchical structure characteristics, while its length represents the degree a capsule is corresponding to a class (e.g. health or disease). To measure the length of the output vector as a probability value, a nonlinear squash function is used as follow:

$$\breve{u}_m = \frac{||u_m||^2}{1 + ||u_m||^2} \cdot \frac{u_m}{||u_m||} \tag{2}$$

wherein, $\breve{u}_m^{(l)}$ is the scaled vector of $\mathbf{X}^2_{out}$. This function compresses the short vector features to zero and enlarge the long vector features a value close to 1. The final output is denoted as $\mathbf{X}^3_{out} \in \mathbb{R}^{Z \times 1 \times 1 \times K}$.

Subsequently, the class-capsule layer is introduced to encode the encapsulated vector features in $\mathbf{X}^3_{out}$ in to the class-capsule vectors corresponding to the target classes. The length of the class-capsule vectors indicate the probability of belonging to corresponding classes. Here, a dynamic routing algorithm is introduced to iteratively update the parameters between the class-capsule vectors with the previous capsule vectors. The dynamic routing algorithm provides a well-designed learning mechanism between the feature vectors, which reinforces the connection coefficients between the layers, and highlights the part-to-whole correlation relationship between the generated capsule features. Mathematically, the class-capsule $\hat{u}_{n|m}^{(l)}$ is calculated as:

$$\hat{u}_{n|m}^{(l)} = W_{m,n}^{(l)} \cdot \breve{u}_m^{(l-1)} + B_n^{(l)} \tag{3}$$

where $W_{m,n}^{(l)}$ is a transformation matrix connecting the layer $l-1$ with layer l, $\breve{u}_m^{(l-1)}$ is the m^{th} feature of layer $l-1$, and $B_n^{(l)}$ is the biases. This function allows the vector features in low-level to make prediction for the rotation invariance of high-level features corresponding to the target classes. After that, the prediction agreement can be computed by a dynamic routing coefficient $c_{m,n}^{(l)}$:

$$s_n^{(l)} = \sum_m^{z^{(l-1)}} c_{m,n}^{(l)} \cdot \hat{u}_{n|m}^{(l)} \tag{4}$$

where $c_{m,n}^{(l)}$ is a dynamic routing coefficient measuring the weight of the m^{th} capsule feature of layer $l-1$ activating the n^{th} class-capsule of layer l, the sum of all the coefficients would be 1, and the dynamic routing coefficient can be calculated as:

$$c_{m,n}^{(l)} = \frac{e^{b_{m,n}}}{\sum_i^{z^{(l)}} e^{b_{m,i}}} \tag{5}$$

where $b_{m,n}$ is the *log* prior representing the correlation between layer $l-1$ and layer l, which is initialized as 0 and is iteratively updated as follow:

$$b_{m,n}^{l} = b_{m,n}^{l-1} + v_n^{l-1} \cdot \hat{u}_{n|m}^{(l-1)} \tag{6}$$

where v_n^l is the activated capsule of layer l, which can be calculated based on the function as follows:

$$v_n^l = \frac{||s_n^{(l)}||^2}{1+||s_n^{(l)}||^2} \cdot \frac{s_n^{(l)}}{||s_n^{(l)}||} \tag{7}$$

Updated by the dynamic routing algorithm, the capsule features with similar prediction will be clustered, and a robust prediction based on these capsule clusters is performed. Finally, the the loss function (L) is defined as follow:

$$\begin{aligned} L_{margin} = \sum_{i}^{n_{class}} T_i \max(0, edge^+ - ||v_n^l||^2) + \\ \mu(1-T_i)(\max(0, ||v_n^l|| - edge^-)^2) \end{aligned} \tag{8}$$

where T_i is set as 1 when class i is currently classified in the data, otherwise is 0. The $edge^+$, set as 0.9, and $edge^-$, set as 0.1, are defined to force the v_n^l into a series of small interval values to update the loss function. μ, defined as 0.5, is a regularization parameter to avoid over-fitting and reduce the effect of the negative activity vectors.

3.2.4. The decoder layer

The decoder layer, composed by two fully connected layers, is designed to reconstruct the classification map from the output vector features. The final output of this model is regarded as $\tilde{\mathbf{Y}} \in \mathbb{R}^{H \times W}$. For the model updating, the model loss aims to minimize the difference between the labelled map, $\bar{\mathbf{Y}}$, and the output map, $\tilde{\mathbf{Y}}$. The final loss function is defined as follow:

$$L_{end} = L_{margin} + \theta \cdot L_{reconstruction} \tag{9}$$

where, $L_{reconstruction} = \|\tilde{\mathbf{Y}} - \bar{\mathbf{Y}}\|$ is the mean square error (MSE) loss between the labelled map and the output map, and θ is the learning rate, in this study, θ is set to 0.0005 in order to trade-off the contribution of L_{margin} and $L_{reconstruction}$. And an Adam optimizer is used to optimize the learning process.

3.3. Model evaluation on the detection of potato late blight disease

3.3.1. Experimental design

In order to evaluate the performance of the proposed CropdocNet on the detection of potato late blight disease, three experiments have been conducted: 1) Model sensitivity to the network depth. 2) An accuracy comparison study between the CropdocNet and the existing machine/deep learning models for potato late blight disease detection 3) The accuracy performance evaluation at both pixel and patch scales. The detailed experimental settings are described as follow.

1) Experiment one: Model sensitivity to the depth of the network

The depth of the network is an important parameter that determines the model performance on spectral-spatial feature extraction. To investigate the effect of the depth of the network, we change the number of the 1D convolutional layers and the 3D convolutional layers in the proposed model to control the model depth. For each of the configuration, we compare the model performance on the potato late blight disease detection and show the best accuracy.

2) Experiment two: An accuracy comparison study between the CropdocNet and the existing machine/deep learning models

In order to evaluate the effectiveness of the hierarchical structure of the spectral-spatial information in our model on the detection of potato late blight disease, we

compare the proposed CropdocNet considering the hierarchical structure of the spectral-spatial information with the existing representative machine/deep learning approaches using a) spectral features only, b) the spatial features only, and c) joint spectral-spatial features only. Based on literature review, SVM, Random Forest (RF) and 3D-CNN are selected as existing representative machine learning/deep learning models for comparison study. For the spectral feature-based models, the works in [43,44,57] have reported the Support Vectors Machine (SVM) is an effective classifier for plant disease diagnosis based on spectral features. For the spatial feature based models, researches in [27,33,34] have demonstrated that the Random Forest (RF) is an effective classifier for analysis of the plant stress associated spatial information in disease diagnosis. For joint spectral-spatial feature based models, a number of deep learning models have been proposed for extracting the spectral-spatial features from the HSI data. Among which, 3D convolutional neural network (3DCNN)-based models [39,50,53] are the most commonly used in plant disease detection. All these existing methods didn't consider the hierarchical structure of the spectral-spatial information.

3) Experiment three: The accuracy performance evaluation at both pixel and patch scales

To evaluate the model performance on mapping of the potato blight disease occurrence situation under different observation scales, two evaluation methods are conducted: 1) pixel-scale evaluation, which focuses on the performance evaluation of the proposed model for detection of the detailed late blight disease occurrence at the pixel-level based on the pixel-wised ground truth data. In addition, to validate model robustness and generalisability, we have also compared the classification maps of all four models based on the independent dataset. 2) patch-scale evaluation, which focuses on performance evaluation at the patch level by aggregation of the pixel-wised classification into the patches with a given size. For instance, in our case, the field is divided into the $1m \times 1m$ patches/grids, the disease predictions at the pixel level are aggregated into the $1m \times 1m$ patches, which is compared against the corresponding real disease occurrence within that given patch area. In this study, the patch size of $1m \times 1m$ is used for two reasons: 1) to enable easy pixel-level data labelling; 2) to enable easy identification of the patches on HSIs to ensure the right match between the ground truth investigation patches and the pixel-level labels. This patch-scale evaluation further indicates the classification robustness of the disease detection at different observation scales.

3.3.2. Evaluation metrics

A set of widely used evaluation metrics are introduced to evaluate the accuracy of the detection of potato late blight disease including: Confusion Matrix, Sensitivity, Specificity, Overall Accuracy (OA), Average Accuracy (AA), and Kappa coefficient. These evaluation metrics are computed based on the statistics of condition positive (P), condition nagative (N), true positive (TP), false positive (FP), true nagative (TN), false negative (FN). Specifically, for a given class (e.g. the late blight disease), the real P indicates the samples labelled as late blight disease and the real N indicates the samples labelled as non-late blight disease. TP, TN, FP, and FN are obtained from the model output. The detailed definition of the metrics are set as follow:

Table 2: The definition of Confusion Matrix: P = Condition Positive; N = Condition Negative; TP = True Positive; FP = False Positive; TN = True Negative; FN = False Negative; UA = user's accuracy; PA = producer's accuracy. Wherein, the producer's accuracy refers to the probability that a certain class is classified correctly, and the user's accuracy refers to the reliability of a certain class.

	P	N	UA (%)
P	TP	FP	$TP/(TP+FP)\times 100\%$
N	FN	TN	$TN/(TN+FN)\times 100\%$
PA(%)	$TP/(TP+FN)\times 100\%$	$TN/(TN+FP)\times 100\%$	

$$Sensitivity = TP/TP + FN \tag{10}$$

$$Specificity = TN/TN + FP \tag{11}$$

$$OA = TP + TN/TP + TN + FP + FN \tag{12}$$

$$AA = 1/2 \times (\frac{TP}{TP+FN} + \frac{TN}{TN+FP}) \tag{13}$$

$$Observation = TP + TN \tag{14}$$

$$Expect = \frac{((TP+FP)*(TP+FN)+(TN+FP)*(TN+FN))}{(TP+TN+FP+FN)} \tag{15}$$

$$Kappa = \frac{Observation - Expect}{(TP+TN+FP+FN) - Expect} \tag{16}$$

3.3.3. Model training

In this study, a slide window approach is used to extract the input samples for model training. Here, the slide window size is set as 13×13. A total of $3,200$ (i.e. 800 for each class) HSI blocks with a size of $13 \times 13 \times 125$ are randomly extracted from the HSI data collected from the controlled field conditions (i.e. experimental site 1 and 2). In order to prevent the over-fitting in the training process, a 5-fold cross validation is used. For model optimization, an Adam optimizer, with a batch size of 64, is used to train the proposed model. The learning rate is initially set as 1×10^{-3}, and is iteratively increased with a step of $1 \times 10^{(-6)}$.

The hardware environment for model training consists of an Intel (R) Xeon (R) CPU E5-2650, NVIDIA TITAN X (Pascal) and 64 GB memory. The software environment is Tensorflow 2.2.0 framework and python 3.5.2 as the programming language.

4. Results

4.1. The CropdocNet model sensitivity to the depth of the convolutional filters

In the proposed method, we will need to set the parameters $K^{(1)}$, $K^{(2)}$, and $K^{(3)}$, which represent the depth of the 1D convolutional layers for the spectral feature extraction, the depth of the 3D convolutional layers for the spectral-spatial feature extraction, and the number of the capsules vector features respectively. Due to the fact that, in our model, the high-level capsule vector features are derived from the low-level spectral-spatial scalar features, the depth of the convolutional filters is the main factor that influences this process. Therefore, we firstly set the $K^{(3)}$ as a fixed value of 16 to evaluate the effect of using different depths of $K^{(1)}$ and $K^{(2)}$ for spectral-spatial scalar feature extraction. Fig. 3a shows the overall accuracy of the potato late blight disease classification using the the various $K^{(1)}$ and $K^{(2)}$ from 32 to 256 with a step of 16. It can

be seen that both $K^{(1)}$ and $K^{(2)}$ have the positive effects on the classification accuracy. The accuracy convergence is more sensitive to $K^{(2)}$than to $K^{(1)}$. This is because the $K^{(2)}$ controls the joint spectral-spatial features with more correlation with the plant stress, and affects the final disease recognition accuracy. Overall, the classification accuracy reaches convergence (approximately 85.05%) when $K^{(1)} = 128$ and $K^{(2)} = 64$. Thus, in the following experiments, we set $K^{(1)} = 128$ and $K^{(2)} = 64$ for optimal model performance and computing efficiency.

Subsequently, we test the effect of the parameter $K^{(3)}$ with the fixed $K^{(1)}$ and $K^{(2)}$ values of 128 and 64. Fig. 3b shows that the classification accuracy increases when $K^{(3)}$ increases from 8 to 32, and then converges to approximately 97.15% when $K^{(3)}$ is greater than 32. These findings suggest that the number of 32 capsule vector blocks is the minimum configuration for our model on detection of potato late blight disease. Therefore, in order to trade off between the model performance and computing performance, $K^{(3)}$ is set as 32 in the subsequent experiments.

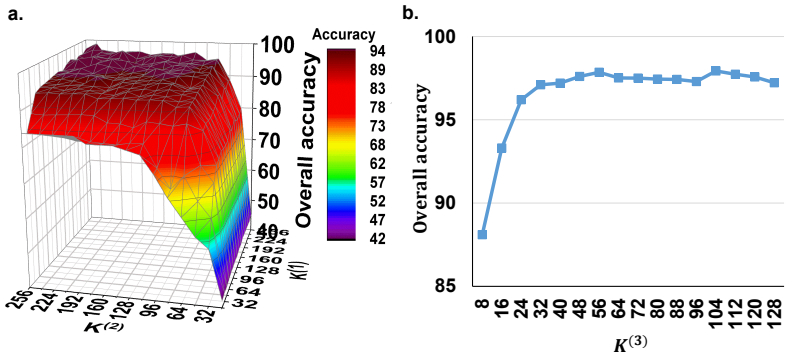

Figure 3. The model sensitivity to the depth of the convolutional filters. (a) the overall accuracy of using the different $K^{(1)}$ and $K^{(2)}$ with fixed $K^{(3)}$ of 16. (b) the overall accuracy of using different $K^{(3)}$ under the fixed $K^{(1)}$ and $K^{(2)}$ values of 128 and 64. Here, $K^{(1)}$ is the depth of the 1D convolutional layers for the spectral feature extraction, $K^{(2)}$ is the depth of the 3D convolutional layers for the spectral-spatial feature extraction, and $K^{(3)}$ is the number of the capsules vector features.

4.2. Accuracy comparison study between the CropdocNet and existing machine learning-based approaches for potato disease diagnosis

In this experiment, we quantitatively investigate the performance of the proposed model considering the hierarchical structure of the spectral-spatial information and the representative machine/deep learning approaches without considering it (i.e. SVM with the spectral feature only, RF with the spatial feature only, and 3D-CNN with the joint spectral-spatial feature only) on potato late blight disease detection with different feature extraction strategies. Wherein, for SVM, we used Radial Basis Function (RBF) kernel to learn the non-linear classifier, two kernel parameters C and γ were set to 1000 and 1, respectively [43,44]. For RF, a quantity of 500 decision trees were employed because this value has been proven to be effective in crop disease detection tasks [33,34]. For 3D-CNN, we employed the model architecture and configurations reported in Nagasubramanian *et al.* [53]'s study. All of the models were trained on the training dataset and validated on both of the testing and independent datasets.

Table. 3 shows the accuracy comparison between the proposed model and the competitors using the test dataset and the independent dataset. The results suggest that the proposed model using the hierarchical vector features consistently outperforms the representative machine/deep learning approaches with scalar features in all of the classes. The OA and AA of the proposed model are 97.33% and 98.09% respectively with a Kappa value of 0.82 on the test dataset, which is 7.8% in average higher than the

second best model (i.e. the 3D-CNN model with joint spectral-spatial scalar features). In addition, the classification accuracy of the proposed model achieves 96.14%, which is 11.8% higher than the second best model. For the independent test dataset, the OA and AA of the proposed model achieve 95.31% and 95.73% respectively with a Kappa value of 0.80, which is the best classifier. The classification accuracy achieves 93.36%, 9.88% which is higher than the second best model. These findings demonstrate that the proposed model with the hierarchical structure of the spectral-spatial information outperforms scalar spectral-spatial feature based models on the classification accuracy of the late blight disease detection.

To further explore the classification difference significance between the proposed method and the existing machine models, the McNemar's Chi-Squared (χ^2) test is conducted between two-paired models. The statistic significant is shown in Table. 4. Our results show that the overall accuracy improvement of the proposed model is statistically significant with $\chi^2 = 32.92(p \leq 0.01)$ for SVM, $\chi^2 = 31.52(p \leq 0.01)$ for RF, and $\chi^2 = 29.34(p \leq 0.01)$ for 3D-CNN.

Moreover, a sensitivity and specificity comparison of detailed class is shown in Fig. 4. Similar to the classification evaluation results, the proposed model achieves the best sensitivity and specificity on all of the ground classes, especially for the class of potato late blight disease.

Table 3: The accuracy comparison between the proposed model and existing representative machine/deep learning models on potato late blight disease detection.

Class	Models on test dataset				Models on independent test dataset			
	Proposed	SVM	RF	3D-CNN	Proposed	SVM	RF	3D-CNN
Healthy potato	**97.21**	86.82	90.64	94.24	**96.32**	82.25	88.92	85.21
Late blight disease	**96.14**	80.15	82.31	85.51	**93.36**	71.76	79.01	83.48
Soil	**99.85**	89.91	92.19	93.31	**98.44**	87.42	83.78	85.12
Background	**99.14**	90.31	93.52	91.16	**94.88**	89.85	86.35	83.85
OA(%)	**97.33**	84.89	87.77	90.32	**95.31**	79.45	83.97	90.32
AA(%)	**98.09**	86.8	89.67	91.06	**95.75**	82.82	84.52	91.06
Kappa	**0.822**	0.549	0.614	0.728	**0.801**	0.512	0.595	0.699

Table 4: The McNemar's Chi-Square Test of the proposed model and the existing representative machine/deep learning models on potato late blight disease detection (* means $p < 0.1$, ** means $p < 0.05$).

Class	Proposed vs. SVM	Proposed vs. RF	Proposed vs. 3D-CNN
Healthy potato	31.82**	30.25**	28.82**
Late blight disease	35.91**	33.24**	32.31**
Soil	33.25**	32.12**	30.33**
Background	32.15**	30.14**	27.42**
Overall	32.92**	31.52**	29.34**

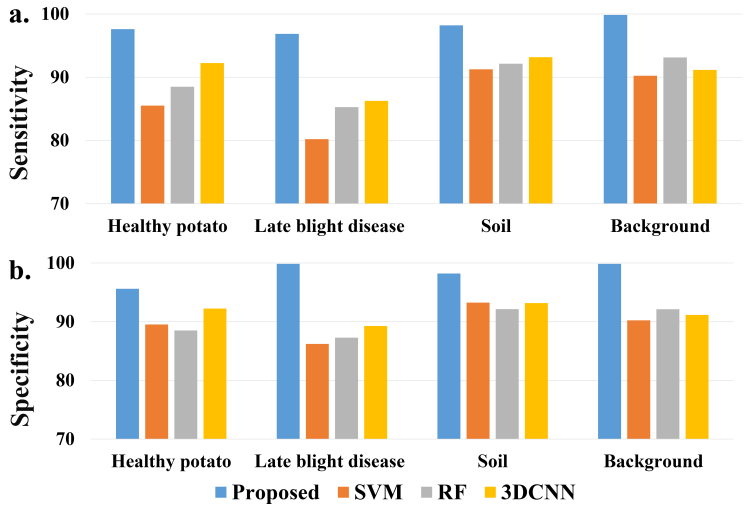

Figure 4. A comparison of (a) sensitivity and (b) specificity of each classes from different models.

4.3. The model performance on mapping the potato late blight disease from the UAV HSI data

In order to show the model performance and generalisability on the detection of potato late blight disease, Fig. 5 illustrates the classification maps of all four models for the independent testing dataset (collected under natural conditions). Here, for highlighting the display of healthy potato and late blight, we show the classes of soil and background in a same colour. We find that the potato late blight disease area produced by the proposed CropdocNet is located in a hot-spot area, which is consistent with our ground investigations. In comparison, there are noticeable "salt and pepper" noises found in the classification maps produced by SVM, RF, and 3DCNN. More importantly, the proposed CropdocNet outperforms the competitors in the classification of the mixed pixels located in the potato field edge and low density area, thus, a clear boundary between the plant (i.e. the class of healthy potato) and bare soil (i.e. the class of background) can be observed in the classification map of CropdocNet (see Fig. 5e), but the pixels in the potato field edge and low density area are misclassified as late blight disease in the maps of SVM, RF, and 3DCNN (see Fig. 5b-d).

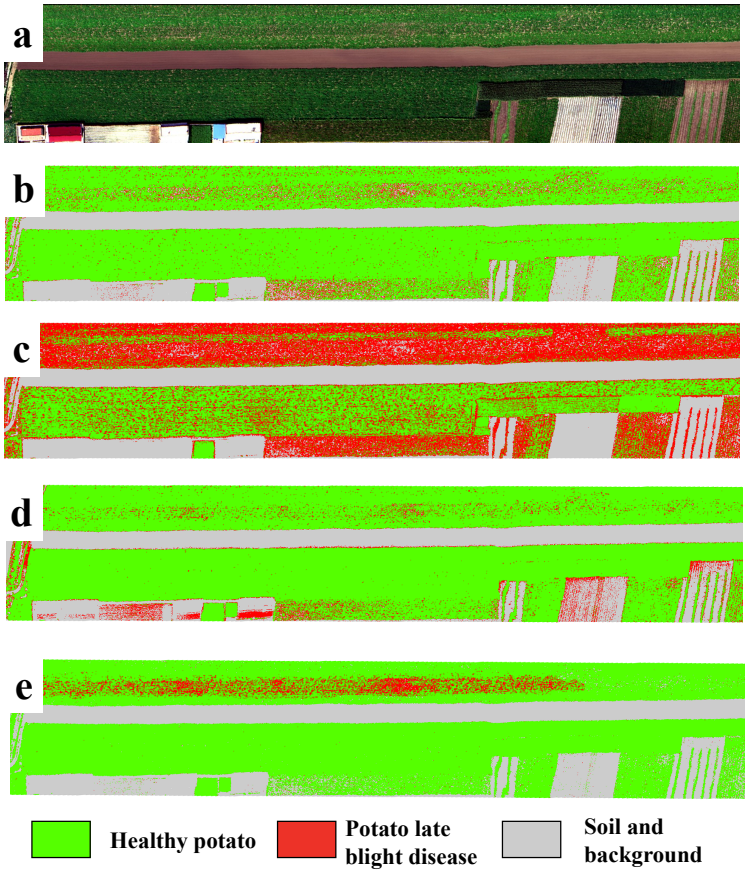

Figure 5. A comparison of the classification maps for the independent testing dataset from four model. (a) the RGB composition map of the raw data, (b-e) the classification maps of SVM, RF, 3DCNN, and the proposed CropdocNet.

Table 5 shows the confusion matrix of the proposed model and the existing model on the pixel-scale disease classification by using the independent testing dataset from the site 3. Our results demonstrate that, compared with the accuracies based on the test dataset mentioned in section 4.2, the proposed model performs a robust classification on the evaluation dataset with the overall accuracy of 98.2% and Kappa of 0.812. In comparison, the competitors that only considered spectral (i.e. SVM) or spatial information (i.e. RF) reveals a significant degradation in terms of classification accuracy and robustness. The execution time of the proposed model is 721 ms, which is faster than the 3DCNN, but lower than the ones of SVM and RF. This findings suggest that, the proposed model has a better performance in terms of both accuracy and computing efficiency, compared to 3DCNN.

In addition, a patch scale evaluation between the ground truth and classification result is significant for guiding the agricultural management and control in practise. Fig. 6 shows the patch-scale test for the classification maps of healthy potato and potato late blight disease overlaying on the UAV HSI in experimental site 1 and site 2, respectively. Wherein, the percentage rate revealed in each patches are the ratio of the late blight disease pixels and the total pixels of the patch. For the experimental site 1, 9 patches with $1m \times 1m$ size are ground truth data. Our results illustrate that, the average differences of disease ratio within the patches between the ground truth data and the classification map is 2.6%. The maximum difference occurring in the patch 8 is 5%. For the experimental site 2, there are 16 $1m \times 1m$ ground truth patches. Our findings suggest that the average differences of disease ratio within the patches between the ground truth patches and the

patches from the classification map is 1%, and the maximum difference occurring in the patch 1 is 3%.

Table 5: The confusion matrix of the proposed model and the existing models on the pixel scale detection of potato late blight disease. Here, UA is the User's accuracy, PA is the Producer's accuracy.

		Healthy potato	Late blight disease	Soil	Background	U(%)	OA(%)	Kappa	Computing time (ms)
CropdocNet	Healthy potato	81	1	0	0	98.8			
	Late blight disease	2	82	0	0	97.6	98.2	0.812	721
	Soil	0	2	89	0	97.8			
	Background	0	0	1	72	98.6			
	P(%)	97.6	96.5	98.9	100				
SVM	Healthy potato	69	11	2	0	84.1			
	Late blight disease	10	70	3	5	79.5	82.7	0.571	162
	Soil	3	5	75	8	82.4			
	Background	1	0	11	59	83.1			
	P(%)	83.1	81.4	82.4	81.9				
RF	Healthy potato	65	11	2	2	81.3			
	Late blight disease	12	66	4	4	76.7	78.8	0.615	117
	Soil	3	5	73	8	82			
	Background	3	3	11	56	76.7			
	P(%)	78.3	77.6	81.1	80				
3DCNN	Healthy potato	73	6	0	0	92.4			
	Late blight disease	5	75	2	3	88.2	88.8	0.771	956
	Soil	1	2	80	4	92			
	Background	1	1	8	65	86.7			
	P(%)	91.3	89.3	88.9	90.3				

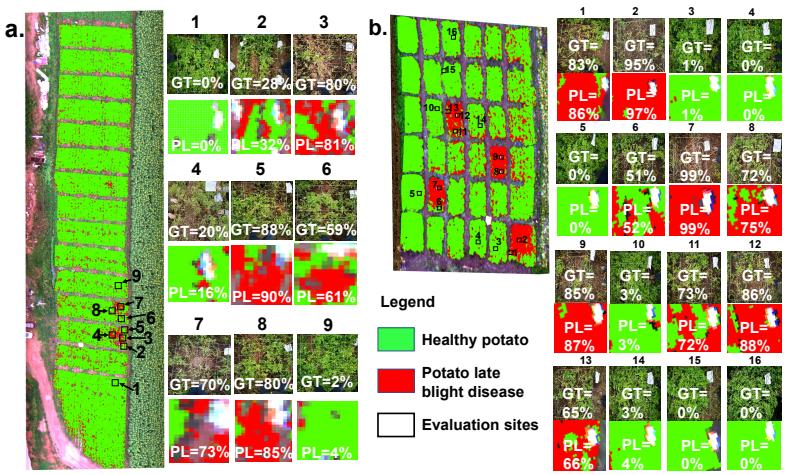

Figure 6. The patch scale test for the classification maps of the healthy potato and the potato late blight disease in a) experimental site 1 and 2) experimental site 2. Here, the example patches on the right side illustrate the accuracy comparison between the ground truth (GT) investigations and the predicted levels (PL) of the late blight disease. Each value inside the patch represents the disease ratio (the late blight disease pixels / the total pixels).

5. Discussion

The hierarchical structure of the spectral-spatial information extracted from HSI data have been proven effective for representing the invariance of the target entities on HSI [36]. In this paper, we propose a CropdocNet for learning the late blight disease associated hierarchical structure information from the UAV HSI data, providing more accurate crop disease diagnosis at the farm scale. Unlike the traditional scalar features used in the existing machine learning/deep learning approaches, our proposed method introduces the capsule layers to learn the hierarchical structure of the late blight disease-associated spectral-spatial characteristics, which allows for capturing the rotation invariance of the late blight disease under the complicated field conditions, leading to improvements in terms of the model accuracy, robustness, and generalisability.

To trade off between the accuracy and computing efficiency, the effects of the depth of the convolutional filters are investigated. Our findings suggest that there is no obvious improvement in accuracy when the depth of 1-D convolutional kernels $K^{(1)} = 128$ and the depth of 3-D convolutional kernels $K^{(2)} = 64$. We also find that, by using the multi-scale capsule units ($K^{(3)} = 32$), the model performance on HSI-based potato late blight disease detection could be improved.

To investigate the effectiveness of using the hierarchical vector features for accurate disease detection, we have compared the proposed model with three typical machine learning models considering only the spectral or spatial scalar features. The results illustrate that the proposed model outperforms the traditional models in terms of overall accuracy, average accuracy, sensitivity and specificity on both the training dataset (collected under controlled field conditions) and the independent testing dataset (collected under natural conditions). In addition, the classification differences between the proposed model and the existing models are statistically significance based on the McNemar's Chi-Square Test.

5.1. The assessment of the hierarchical vector feature

To further visually demonstrate the benefit of using hierarchical vector features in the proposed CropdocNet, we have compared the visualized feature space and the mapping results of the healthy (see the first row of Fig. 7) and diseased plots (see the second row of Fig. 7) from three models: SVM, 3DCNN, and the proposed CropdocNet. Our quantitative assessment reveals that the accuracy of the potato late blight disease plots is 76.8%, 83.2%, and 94.2% for SVM, 3DCNN, and CropdocNet, respectively. Specifically, for the SVM-based model which only maps the spectral information into the feature space, a total of 81% areas in the healthy plots are misclassified as the potato late blight disease (see the left subgraph of Fig. 7b). And the feature space of the samples in the yellow frame, as shown in the right subgraph of Fig. 7b, explains the reason for these misclassifications. Thus, there is no cluster characteristics can be observed between the spectral features in the SVM-based feature space, indicating that the inter-class spectral variances are not significant between the SVM decision hyperplane.

In contrast, the spectral-spatial information based on 3DCNN (Fig. 7c) performs better than the SVM-based model. However, looking at the edge of the plots, there are obvious misclassifications. The right subgraph of Fig. 7c, reveals the averages and the standard deviations of the activated high-level features of the samples within the yellow frame. It is worth noting that, for the healthy potato (the first row of Fig. 7c), the average values of the activated joint spectral-spatial features for different classes are quite close, and the standard deviations are relatively high, illustrating that the inter-class distance between the healthy potato and the potato late blight disease are not significant in the features space. The similar results can be found in the late blight disease (see second row of Fig. 7c). Thus, no significant inter-class separability can be represented in the joint spectral-spatial feature space owning to the mixed spectral-spatial signatures of plant and background.

In comparison, the hierarchical vector features-based CropdocNet model provides more accurate classification because the hierarchical structural capsule features can express the various spectral-spatial characteristics of the target entities. For example, the white panels in the diseased plot (see the second row of of Fig. 7d) are successfully classified as the background. The right subgraphs of Fig. 7d demonstrate the average, direction, and standard deviations of the activated hierarchical capsule features of the samples within the yellow frame. It's noteworthy that the average length and direction of the activated features for different classes are quite different, and the standard deviations (see the shadow under the arrows) do not overlap with each other. These results fully demonstrate the significant clustering of each class in the hierarchical capsule feature space, thus, the hierarchical vector features are capable of capturing most of the spectral-spatial variability found in practices.

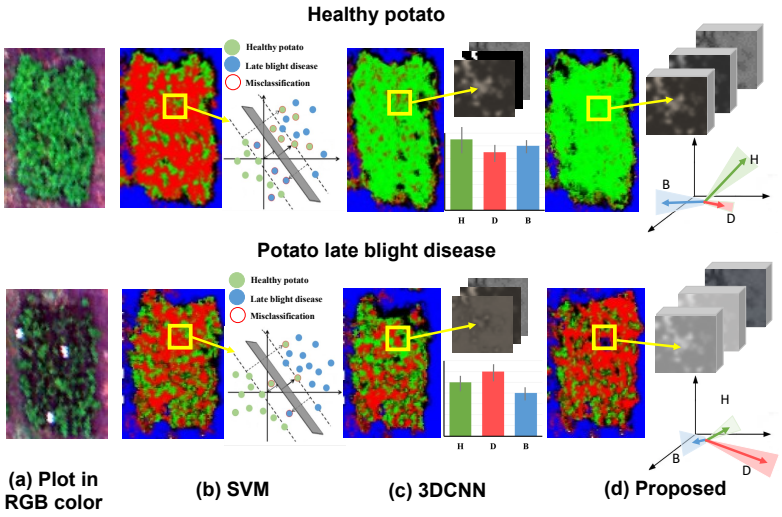

Figure 7. The visualized feature space and the mapping results of the healthy and diseased plots based on the different machine learning/deep learning methods: a) the original RGB image for the healthy potato (H) and diseased potato (D), and background (B). b) the classification results and the visualized spectral feature space of SVM, c) the classification results and the averages and the standard deviations of the activated high-level spectral-spatial features of 3DCNN, and d) the classification results and the visualized hierarchical capsule feature space of the proposed CropdocNet.

5.2. The general comparison of the CropdocNet and the existing models

For indirect comparison between the proposed CropdocNet and the existing case studies, we have drawn a Table. 6 and provided accuracy performance and computing efficiency. As shown in Table. 6, our proposed CropdocNet has the best accuracy performance (95.75%), compared to the existing works. For the computing efficiency, due to the deep-layered network architecture and large scale samples, the deep learning models (3DCNN and CropdocNet) require more computing time, compared to traditional machine learning methods (such as SVM, RF) where they only use fewer samples.

Table 6: The performance comparison of the proposed CropdocNet with the existing study cases. Note: the '-' means no record found in the relative literature.

Model name	Studied crop and disease	Classification accuracy	Number of training sample	Number of parameters	Model execution time	Reference
SVM	Potato late blight	84%	892	-	-	[2]
	Grape leaf disease	88.89%	137	-	182ms	[58]
	Tomato leaf disease	92%	708	-	-	[59]
RF	Tomato leaf disease	95.20%	882	-	-	[60]
	Rice leaf blight	69.44%	423	-	104ms	[61]
	Potato late blight	79%	892	-	-	[2]
3DCNN	Tea leaf blight	89.90%	13262	770k	-	[62]
	Tomato leaf disease	91.83%	3852	600k	687ms	[49]
	Tomato leaf disease	90.30%	7176	840k	871ms	[63]
	Potato late blight	85.40%	5142	560k	564ms	[22]
CropdocNet	Potato late blight	95.75%	3200	690k	721ms	This study

5.3. The limitations and future works of this study

Benefiting from the hierarchical capsule features, the proposed CropdocNet performs better on potato late blight disease detection than the existing spectral-based or spectral-spatial based deep/machine learning models, and the generalisability of the network architecture is better than the existing models. The previous experimental evaluation has demonstrated the robustness and generalisability of our proposed model. Our model can be adapted to other crop disease detection since our proposed method introduces the capsule layers to learn the hierarchical structure of the disease-associated spectral-spatial characteristics, which allows for capturing the rotation invariance of diseases under complicated conditions. However, it is worth mentioning our current

input data for model training is mainly based on the full bloom period of potato growth when the canopy closure reaches maximum and the field microclimate is mostly suitable for the occurrence of late blight disease, the direct use of the pre-trained model may lead to a limited performance. The reason is that the hyperspectral imagery is generally influenced by the mixed pixel effect, which depends on the crop growth and stress types. Therefore, in the future study, we will validate the proposed model on more UAV-based HSI data with various potato growth stages and various diseases. Specifically, we will further test the receptive field of the CropdocNet and fine-tune the model on the HSI data for performance enhancement under various field conditions.

6. Conclusions

In this study, a novel end-to-end deep learning model (CropdocNet) is proposed for extracting the spectral-spatial hierarchical structure of the late blight disease, and automatically detecting the disease from the UAV HSI data. The innovation of the CropdocNet is the deep-layered network architecture that integrates the spectral-spatial scalar features into the hierarchical vector features for representing the rotation invariance of the potato late blight disease in the complicated field conditions. The model has been tested and evaluated on the controlled and the natural field data, and compared with the existing machine/deep learning models. The average accuracies for the training dataset and independent testing dataset are 98.09% and 95.75%, respectively. The experimental findings demonstrate that the proposed model is able to significantly improve the accuracy of the potato late blight disease on the HSI data.

Since the proposed model is mainly based on data collected from the limited potato growth stage and one type of potato diseases. To further enhance the proposed model, future work will include two aspects: 1) we will validate the proposed model on more UAV-based HSI data with various potato growth stages and various diseases under various field conditions. This is important for the UAV-based crop disease detection and monitoring at the canopy and regional levels since the hyperspectral imaging is generally influenced by the mixed pixel effect, which is highly dependent on the canopy geometry associated with the crop growth and stresses. 2) We will also investigate whether the size of the receptive field of the CropdocNet is able to characterize the spectral-spatial hierarchical features for different crop diseases.

Author Contributions: Conceptualization, all authors; methodology, Y.S. and L.H..; software, Y.S.; data acquisition and validation, T. H. and C. S.; analysis: Y.S., L.H. and T. H.; writing—original draft preparation, Y.S. and L.H.; writing—review and editing, all authors; supervision, L.H.; funding acquisition, L.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by BBSRC (BB/S020969/1), BBSRC (BB/R019983/1), the Open Research Fund of Key Laboratory of Digital Earth Science, Chinese Academy of Sciences(No.2019LDE003), and National key R&D Program of China (2017YFE01227000)

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Demissie, Y.T. Integrated potato (Solanum tuberosum L.) late blight (Phytophthora infestans) disease management in Ethiopia. *American Journal of BioScience* **2019**, *7*, 123–130.
2. Patil, P.; Yaligar, N.; Meena, S. Comparision of performance of classifiers-svm, rf and ann in potato blight disease detection using leaf images. 2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC). IEEE, 2017, pp. 1–5.
3. Hirut, B.G.; Shimelis, H.A.; Melis, R.; Fentahun, M.; De Jong, W. Yield, Yield-related Traits and Response of Potato Clones to Late Blight Disease, in North-Western Highlands of Ethiopia. *Journal of Phytopathology* **2017**, *165*, 1–14.
4. Namugga, P.; Sibiya, J.; Melis, R.; Barekye, A. Yield Response of Potato (Solanum tuberosum L.) Genotypes to late blight caused by Phytophthora infestans in Uganda. *American Journal of Potato Research* **2018**, *95*, 423–434.
5. Zhang, X.; Li, X.; Zhang, Y.; Chen, Y.; Tan, X.; Su, P.; Zhang, D.; Liu, Y. Integrated control of potato late blight with a combination of the photosynthetic bacterium Rhodopseudomonas palustris strain GJ-22 and fungicides. *BioControl* **2020**, *65*, 635–645.

6. Gao, J.; Westergaard, J.C.; Sundmark, E.H.R.; Bagge, M.; Liljeroth, E.; Alexandersson, E. Automatic late blight lesion recognition and severity quantification based on field imagery of diverse potato genotypes by deep learning. *Knowledge-Based Systems* **2021**, *214*, 106723.
7. Lehsten, V.; Wiik, L.; Hannukkala, A.; Andreasson, E.; Chen, D.; Ou, T.; Liljeroth, E.; Lankinen, Å.; Grenville-Briggs, L. Earlier occurrence and increased explanatory power of climate for the first incidence of potato late blight caused by Phytophthora infestans in Fennoscandia. *PloS one* **2017**, *12*, e0177580.
8. Sharma, R.; Singh, A.; Dutta, M.K.; Riha, K.; Kriz, P.; others. Image processing based automated identification of late blight disease from leaf images of potato crops. 2017 40th International Conference on Telecommunications and Signal Processing (TSP). IEEE, 2017, pp. 758–762.
9. Tung, P.X.; Vander Zaag, P.; Li, C.; Tang, W. Combining Ability for Foliar Resistance to Late Blight [Phytophthora infestans (Mont.) de Bary] of Potato Cultivars with Different Levels of Resistance. *American Journal of Potato Research* **2018**, *95*, 670–678.
10. Franceschini, M.H.D.; Bartholomeus, H.; Van Apeldoorn, D.F.; Suomalainen, J.; Kooistra, L. Feasibility of unmanned aerial vehicle optical imagery for early detection and severity assessment of late blight in potato. *Remote Sensing* **2019**, *11*, 224.
11. Islam, M.; Dinh, A.; Wahid, K.; Bhowmik, P. Detection of potato diseases using image segmentation and multiclass support vector machine. 2017 IEEE 30th canadian conference on electrical and computer engineering (CCECE). IEEE, 2017, pp. 1–4.
12. Shi, Y.; Huang, W.; González-Moreno, P.; Luke, B.; Dong, Y.; Zheng, Q.; Ma, H.; Liu, L. Wavelet-based rust spectral feature set (WRSFs): a novel spectral feature set based on continuous wavelet transformation for tracking progressive host–pathogen interaction of yellow rust on wheat. *Remote sensing* **2018**, *10*, 525.
13. Parra-Boronat, L.; Parra-Boronat, M.; Torices, V.; Marín, J.; Mauri, P.V.; Lloret, J. Comparison of single image processing techniques and their combination for detection of weed in Lawns. *International Journal On Advances in Intelligent Systems* **2019**, *12*, 177–190.
14. Shin, J.; Chang, Y.K.; Heung, B.; Nguyen-Quang, T.; Price, G.W.; Al-Mallahi, A. A deep learning approach for RGB image-based powdery mildew disease detection on strawberry leaves. *Computers and Electronics in Agriculture* **2021**, *183*, 106042.
15. Shi, Y.; Huang, W.; Luo, J.; Huang, L.; Zhou, X. Detection and discrimination of pests and diseases in winter wheat based on spectral indices and kernel discriminant analysis. *Computers and Electronics in Agriculture* **2017**, *141*, 171–180.
16. Yang, N.; Yuan, M.; Wang, P.; Zhang, R.; Sun, J.; Mao, H. Tea diseases detection based on fast infrared thermal image processing technology. *Journal of the Science of Food and Agriculture* **2019**, *99*, 3459–3466.
17. Høye, T.T.; Ärje, J.; Bjerge, K.; Hansen, O.L.; Iosifidis, A.; Leese, F.; Mann, H.M.; Meissner, K.; Melvad, C.; Raitoharju, J. Deep learning and computer vision will transform entomology. *Proceedings of the National Academy of Sciences* **2021**, *118*.
18. Tang, Y.; Chen, M.; Wang, C.; Luo, L.; Li, J.; Lian, G.; Zou, X. Recognition and localization methods for vision-based fruit picking robots: A review. *Frontiers in Plant Science* **2020**, *11*, 510.
19. Iqbal, U.; Perez, P.; Li, W.; Barthelemy, J. How computer vision can facilitate flood management: A systematic review. *International Journal of Disaster Risk Reduction* **2021**, p. 102030.
20. Wu, F.; Duan, J.; Chen, S.; Ye, Y.; Ai, P.; Yang, Z. Multi-target recognition of bananas and automatic positioning for the inflorescence axis cutting point. *Frontiers in plant science* **2021**, *12*.
21. Cao, X.; Yan, H.; Huang, Z.; Ai, S.; Xu, Y.; Fu, R.; Zou, X. A Multi-Objective Particle Swarm Optimization for Trajectory Planning of Fruit Picking Manipulator. *Agronomy* **2021**, *11*, 2286.
22. Dhingra, G.; Kumar, V.; Joshi, H.D. Study of digital image processing techniques for leaf disease detection and classification. *Multimedia Tools and Applications* **2018**, *77*, 19951–20000.
23. Zhu, W.; Chen, H.; Ciechanowska, I.; Spaner, D. Application of infrared thermal imaging for the rapid diagnosis of crop disease. *IFAC-PapersOnLine* **2018**, *51*, 424–430.
24. Li, B.; Xu, X.; Zhang, L.; Han, J.; Bian, C.; Li, G.; Liu, J.; Jin, L. Above-ground biomass estimation and yield prediction in potato by using UAV-based RGB and hyperspectral imaging. *ISPRS Journal of Photogrammetry and Remote Sensing* **2020**, *162*, 161–172.
25. Wan, L.; Cen, H.; Zhu, J.; Zhang, J.; Zhu, Y.; Sun, D.; Du, X.; Zhai, L.; Weng, H.; Li, Y.; others. Grain yield prediction of rice using multi-temporal UAV-based RGB and multispectral images and model transfer–a case study of small farmlands in the South of China. *Agricultural and Forest Meteorology* **2020**, *291*, 108096.
26. Moghadam, P.; Ward, D.; Goan, E.; Jayawardena, S.; Sikka, P.; Hernandez, E. Plant disease detection using hyperspectral imaging. 2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA). IEEE, 2017, pp. 1–8.
27. Golhani, K.; Balasundram, S.K.; Vadamalai, G.; Pradhan, B. A review of neural networks in plant disease detection using hyperspectral data. *Information Processing in Agriculture* **2018**, *5*, 354–371.
28. Zhang, N.; Pan, Y.; Feng, H.; Zhao, X.; Yang, X.; Ding, C.; Yang, G. Development of Fusarium head blight classification index using hyperspectral microscopy images of winter wheat spikelets. *Biosystems Engineering* **2019**, *186*, 83–99.
29. Khan, I.H.; Liu, H.; Cheng, T.; Tian, Y.; Cao, Q.; Zhu, Y.; Cao, W.; Yao, X. Detection of wheat powdery mildew based on hyperspectral reflectance through SPA and PLS-LDA. *International Journal of Precision Agricultural Aviation* **2020**, *3*.
30. Abdulridha, J.; Batuman, O.; Ampatzidis, Y. UAV-based remote sensing technique to detect citrus canker disease utilizing hyperspectral imaging and machine learning. *Remote Sensing* **2019**, *11*, 1373.
31. Gogoi, N.; Deka, B.; Bora, L. Remote sensing and its use in detection and monitoring plant diseases: A review. *Agricultural Reviews* **2018**, *39*.
32. Arivazhagan, S.; Shebiah, R.N.; Ananthi, S.; Varthini, S.V. Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features. *Agricultural Engineering International: CIGR Journal* **2013**, *15*, 211–217.

33. Behmann, J.; Bohnenkamp, D.; Paulus, S.; Mahlein, A.K. Spatial referencing of hyperspectral images for tracing of plant disease symptoms. *Journal of Imaging* **2018**, *4*, 143.
34. De Castro, A.I.; Torres-Sánchez, J.; Peña, J.M.; Jiménez-Brenes, F.M.; Csillik, O.; López-Granados, F. An automatic random forest-OBIA algorithm for early weed mapping between and within crop rows using UAV imagery. *Remote Sensing* **2018**, *10*, 285.
35. Qiao, X.; Jiang, J.; Qi, X.; Guo, H.; Yuan, D. Utilization of spectral-spatial characteristics in shortwave infrared hyperspectral images to classify and identify fungi-contaminated peanuts. *Food chemistry* **2017**, *220*, 393–399.
36. Shi, Y.; Han, L.; Huang, W.; Chang, S.; Dong, Y.; Dancey, D.; Han, L. A Biologically Interpretable Two-Stage Deep Neural Network (BIT-DNN) for Vegetation Recognition From Hyperspectral Imagery. *IEEE Transactions on Geoscience and Remote Sensing* **2021**.
37. Behmann, J.; Mahlein, A.K.; Rumpf, T.; Römer, C.; Plümer, L. A review of advanced machine learning methods for the detection of biotic stress in precision crop protection. *Precision Agriculture* **2015**, *16*, 239–260.
38. Saleem, M.H.; Potgieter, J.; Arif, K.M. Plant disease detection and classification by deep learning. *Plants* **2019**, *8*, 468.
39. Zhang, X.; Han, L.; Dong, Y.; Shi, Y.; Huang, W.; Han, L.; González-Moreno, P.; Ma, H.; Ye, H.; Sobeih, T. A deep learning-based approach for automated yellow rust disease detection from high-resolution hyperspectral UAV images. *Remote Sensing* **2019**, *11*, 1554.
40. Rumpf, T.; Mahlein, A.K.; Steiner, U.; Oerke, E.C.; Dehne, H.W.; Plümer, L. Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance. *Computers and electronics in agriculture* **2010**, *74*, 91–99.
41. Li, J.; Tang, Y.; Zou, X.; Lin, G.; Wang, H. Detection of fruit-bearing branches and localization of litchi clusters for vision-based harvesting robots. *IEEE Access* **2020**, *8*, 117746–117758.
42. Chen, M.; Tang, Y.; Zou, X.; Huang, Z.; Zhou, H.; Chen, S. 3D global mapping of large-scale unstructured orchard integrating eye-in-hand stereo vision and SLAM. *Computers and Electronics in Agriculture* **2021**, *187*, 106237.
43. Nagasubramanian, K.; Jones, S.; Sarkar, S.; Singh, A.K.; Singh, A.; Ganapathysubramanian, B. Hyperspectral band selection using genetic algorithm and support vector machines for early identification of charcoal rot disease in soybean stems. *Plant methods* **2018**, *14*, 1–13.
44. Huang, L.; Li, T.; Ding, C.; Zhao, J.; Zhang, D.; Yang, G. Diagnosis of the severity of fusarium head blight of wheat ears on the basis of image and spectral feature fusion. *Sensors* **2020**, *20*, 2887.
45. Mahlein, A.K. Plant disease detection by imaging sensors–parallels and specific demands for precision agriculture and plant phenotyping. *Plant disease* **2016**, *100*, 241–251.
46. Yuan, L.; Pu, R.; Zhang, J.; Wang, J.; Yang, H. Using high spatial resolution satellite imagery for mapping powdery mildew at a regional scale. *Precision Agriculture* **2016**, *17*, 332–348.
47. Xie, C.; He, Y. Spectrum and image texture features analysis for early blight disease detection on eggplant leaves. *Sensors* **2016**, *16*, 676.
48. Zhang, N.; Wang, Y.; Zhang, X. Extraction of tree crowns damaged by Dendrolimus tabulaeformis Tsai et Liu via spectral-spatial classification using UAV-based hyperspectral images. *Plant Methods* **2020**, *16*, 1–19.
49. Karthik, R.; Hariharan, M.; Anand, S.; Mathikshara, P.; Johnson, A.; Menaka, R. Attention embedded residual CNN for disease detection in tomato leaves. *Applied Soft Computing* **2020**, *86*, 105933.
50. Francis, M.; Deisy, C. Disease detection and classification in agricultural plants using convolutional neural networks—a visual understanding. 2019 6th International Conference on Signal Processing and Integrated Networks (SPIN). IEEE, 2019, pp. 1063–1068.
51. Nguyen, C.; Sagan, V.; Maimaitiyiming, M.; Maimaitijiang, M.; Bhadra, S.; Kwasniewski, M.T. Early Detection of Plant Viral Disease Using Hyperspectral Imaging and Deep Learning. *Sensors* **2021**, *21*, 742.
52. Suryawati, E.; Sustika, R.; Yuwana, R.S.; Subekti, A.; Pardede, H.F. Deep structured convolutional neural network for tomato diseases detection. 2018 International Conference on Advanced Computer Science and Information Systems (ICACSIS). IEEE, 2018, pp. 385–390.
53. Nagasubramanian, K.; Jones, S.; Singh, A.K.; Sarkar, S.; Singh, A.; Ganapathysubramanian, B. Plant disease identification using explainable 3D deep learning on hyperspectral images. *Plant methods* **2019**, *15*, 1–10.
54. Kumar, K.V.; Jayasankar, T. An identification of crop disease using image segmentation. *Int. J. Pharm. Sci. Res* **2019**, *10*, 1054–1064.
55. La Rosa, L.E.C.; Sothe, C.; Feitosa, R.Q.; de Almeida, C.M.; Schimalski, M.B.; Oliveira, D.A.B. Multi-task fully convolutional network for tree species mapping in dense forests using small training hyperspectral data. *arXiv preprint arXiv:2106.00799* **2021**.
56. Cai, S.; Shu, Y.; Wang, W. Dynamic Routing Networks. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2021, pp. 3588–3597.
57. Zhao, J.; Fang, Y.; Chu, G.; Yan, H.; Hu, L.; Huang, L. Identification of leaf-scale wheat powdery mildew (Blumeria graminis f. sp. Tritici) combining hyperspectral imaging and an SVM classifier. *Plants* **2020**, *9*, 936.
58. Padol, P.B.; Yadav, A.A. SVM classifier based grape leaf disease detection. 2016 Conference on advances in signal processing (CASP). IEEE, 2016, pp. 175–179.
59. Semary, N.A.; Tharwat, A.; Elhariri, E.; Hassanien, A.E. Fruit-based tomato grading system using features fusion and support vector machine. In *Intelligent Systems' 2014*; Springer, 2015; pp. 401–410.
60. Govardhan, M.; Veena, M. Diagnosis of Tomato Plant Diseases using Random Forest. 2019 Global Conference for Advancement in Technology (GCAT). IEEE, 2019, pp. 1–5.

61. Saha, S.; Ahsan, S.M.M. Rice Disease Detection using Intensity Moments and Random Forest. 2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD). IEEE, 2021, pp. 166–170.
62. Hu, G.; Wu, H.; Zhang, Y.; Wan, M. A low shot learning method for tea leafs disease identification. *Computers and Electronics in Agriculture* **2019**, *163*, 104852.
63. Elhassouny, A.; Smarandache, F. Smart mobile application to recognize tomato leaf diseases using Convolutional Neural Networks. 2019 International Conference of Computer Science and Renewable Energies (ICCSRE). IEEE, 2019, pp. 1–4.