Smart Agricultural Technology 11 (2025) 101023

Contents lists available at ScienceDirect

Smart Agricultural Technology

journal homepage: www.journals.elsevier.com/smart-agricultural-technology

Classifying early-stage soybean fungal diseases on hyperspectral images using convolutional neural networks

Chieh Fu Hsiao *, Georg Feyrer, Anthony Stein

Department of Artificial Intelligence in Agricultural Engineering, Institute of Agricultural Engineering, University of Hohenheim, Garbenstraße 9, 70599 Stuttgart, Germany

ARTICLE INFO

Keywords:
Hyperspectral images
Soybean
Fungal diseases
Early detection
Deep learnin
Pathogen detection

ABSTRACT

Using convolutional neural networks (CNNs) to detect plant diseases has proven to reach high accuracy in the classification of infected and non-infected plant images. However, most of the existing researches are based on RGB images due to the availability and the comparably low cost of image collection. The limited spectral information restricts the detectability of plant diseases, especially in the early stage where often symptoms of pathogen infection have not yet become visible. To this end, in this study, hyperspectral imaging (HSI) data are combined with deep learning models to test the classification ability of two soybean fungal diseases: Asian soybean rust (*Phakopsora pachyhizi*) and soybean stem rust (*Sclerotinia sclerotiorum*). Different CNNs employing 2D, 3D convolution, and hybrid approaches are compared. The influences of the depth of the convolutional layer and the regularization techniques are also discussed. Besides, image augmentation methods are investigated to overcome the problem of data scarcity. The results indicate the 6-convolutional-layer depth hybrid model to have the best capacity in classifying Asian soybean rust in the early-mid to mid-late stage when there are over 2 % visible symptoms but a limited detectability in the early stages when there are below 2 % visible symptoms on leaves. On the other hand, the optimized CNN model shows a limited capability to detect both diseases when there are no visible symptoms observable. Overall, this study suggests a hybrid 2D-3D convolutional model with augmentation and regularization methods has a high potential in the early detection of fungal diseases. This research is expected to contribute to a new cropping system that vastly reduces the chemical-synthesis plant protection products, where a continuous pathogen disease monitoring plays a key to manage the crop stands.

1. Introduction

Soybean cultivation, a cornerstone of global agriculture, faces a continuous threat from various diseases that impede plant physiology resulting in substantial economic losses. Among these pathogens, Asian soybean rust (ASR) disease caused by the fungi species *Phakopsora pachyhizi* (Pp) and the soybean Sclerotinia stem rust (SSR) caused by *Sclerotinia sclerotiorum* (Ss) stand out as notorious adversaries, wreaking havoc on crop yield and quality. ASR once reported leading 10 % to 60 % of the yield loss in the northern-central and south-east of the United States which are the second-largest producer in the world [1,2]. SSR, on the other hand, is the predominant disease relevant to stem and root which caused the yield loss of over 200 million bushels in the United States from 2015 to 2019 [3]. Therefore, developing a method and reduce the threat of fungal disease is a critical issue.

In conventional agriculture, traditionally, the detection of fungal diseases is mostly done by manual checking of farmers or plant protection experts. However, it highly depends on the expertise of people, which leads to the potential risk of bias, and the extra cost of the labor force [4]. Besides, the timing of the observation of symptoms influences the effectiveness of the disease control. Applying fungicides in early infection stages can usually reach a better control efficacy, thus it can attenuate the spread of the disease, reduce further fungicide spraying, and save input resources [5,6]. However, early visual detection of fungal diseases by humans is difficult because symptoms are not visible to eyes, which makes machine-supported detection a potential solution.

In recent years, artificial intelligence (AI) has emerged as a powerful ally in the realm of plant protection [7]. The Convolutional Neural Network (CNN), as a branch of deep learning, has demonstrated exceptional capabilities in classifying images, scenes, and objects, which are widely utilized in many industries and affect our daily lives [8]. In the past few years, its application in agriculture is booming. Multiple

* Corresponding author.
E-mail address: chiehfu.hsiao@uni-hohenheim.de (C.F. Hsiao).

https://doi.org/10.1016/j.atech.2025.101023
Received 19 November 2024; Received in revised form 9 March 2025; Accepted 18 May 2025
Available online 19 May 2025
2772-3755/© 2025 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

CNN-based models have been designed to detect plant diseases [9]. By leveraging deep learning, CNNs can analyze intricate plant images, identifying subtle indicators of infection with high accuracy [10]. The integration of AI into plant pathology practices has ushered in a new era of precision agriculture, enabling timely responses to potential threats and enhancing overall crop management strategies [11].

Most of the research regarding the utilization of CNNs in crop protection focuses on RGB images because of the availability and the cost of image collection compared to hyperspectral images. Some of them [12, 13] demonstrated high accuracy rates over 90 % with the transfer learning approach in open database resources such as ImageNet and PlantVillage.

Hyperspectral sensors constitute a more sophisticated alternative to RGB cameras. Instead of measuring 3 bands of the visible electromagnetic spectrum, a hyperspectral sensor primarily utilizes very narrow and adjacent spectral bands over a quasi-continuous spectral range. Each pixel in a hyperspectral image represents a more informative spectral information [14]. One of the previous research [15] demonstrated that *Fusarium* spp. can be detected 9 days after inoculation by using hyperspectral imaging (HSI) data of wheat spike, which is the early stage of the infection. By obtaining not only the human eye's visible but also invisible light reflectance, such as near-infrared light, HSI has proven its ability to recognize the physiological processes in plants. Thus, it has the potential to detect pathogen infection and other stresses through analyzing HSI data in generally assumed [16].

CNNs have been used to perform such analysis. Depending on the dimensionality of the convolutional kernels, the CNNs can be categorized as 1 dimensional (D), 2D, and 3D CNNs. The increase in dimensionality of the kernels helps to extract complex features [17]. Since the hyperspectral data cube is a 3D array matrix that includes one spectral dimension (z) and two spatial dimensions (x, y), it is more common to use 3D CNN models for analyzing HSI [18]. However, with the increase of the dimensionality and the depth of successive convolutional layers, the complexity of the model increases. This results in much more trainable parameters' weights in the neural network and thus increases the cost in terms of the training time and data required. Hybrid models therefore are a compromising solution that balances model capacity and the training time. They concatenate 3D with 2D convolutional layers to maintain the ability to learn spectral-spatial and abstract features with a comparably non-complex structure in contrast to full 3D convolutional models [19]. Hybrid models have proven to reach over 90 % accuracy in the classification of some satellite-based HSI data [20], while their applicability to the leaf-scale HSI data of plant diseases remains largely unclear.

Furthermore, although HSI data is much more informative compared to RGB images, the amount of research on HSI in detecting crop diseases is far less than RGB because of the difficulty and high cost of image collection. Available HSI datasets applications for crop leaf disease research are also very scarce. In recent research on detecting soybean rust disease, combining HSI data with a pure artificial neural network or random forest approach can reach up to 80 % accuracy for classifying infected leaves in 25 % severity [21]. Applying weighting or selecting important spectral bands such as principle component analysis (PCA), Shapley value, and attention mechanism help facilitate the model's performance [22,23]. Besides, by capturing complex spatial features by deformable and dilated convolutional layers, it is possible to distinguish regions of interest extracted from healthy leaves and asymptomatic leaves that represent early infection [23]. Although a high accuracy rate was achieved to classify asymptomatic leaves by randomly extracting regions of interest from samples, it is challenging to guarantee that extracted regions have already had physiological changes that can reflect on the spectrum. It is also unclear what the spectral pattern looks like for the asymptomatic leaves. In contrast, using whole leaf samples with a full spectrum retains more information and avoids missing any important information. Therefore, it is worth exploring whether the application of CNNs on HSI data can reliably detect certain soybean fungi diseases in the early stages on a leaf scale.

With the aim to bridge the knowledge gap of the application of HSI data with CNNs in the early-stage detection of soybean fungal diseases, a HSI dataset for ASR and SSR detection on soybean leaves is created. The images come from the NOcsPS project at the University Hohenheim. Multiple CNN architectures (2D, 3D, and hybrid models) are designed and tested for their ability of disease detection, especially focusing on models' performances in the early stages of the infestation. Furthermore, data pre-processing and image augmentation methods are investigated to tackle the challenge of data scarcity. The following hypothesis of this study was tested: (1) CNNs can classify fungal diseases on a leave scale HSI data of various infection stages, (2) 3D-CNN architectures show a better classification ability compared to 2D and hybrid CNN models, and (3) Augmentation techniques and image transformation methods increase classification performance. To summarize, this study contributes to the research of applying deep learning on HSI data for crop protection through the early detection of fungal diseases.

2. Materials and methods

2.1. Overall workflow

The overall workflow design for detecting soybean diseases is demonstrated in Fig. 1. The process contains the following steps: (1) soybean cultivation, (2) fungal inoculation, (3) HSI data acquisition, (4) image cutting, (5) image preprocessing, (6) data augmentation, (7) model training, and (8) detect diseases with the severity degree. The process (1), (2), and (3) were conducted in the context of the research project NOcsPS (cf. Acknowledgement) and is thus not an original part of this paper.

2.2. Image acquisition

For the purpose of data acquisition, soybean plants were grown under laboratory conditions in the greenhouses of the Institute of Phytomedicine at the University Hohenheim. Two kinds of fungal diseases (ASR and SSR) have been inoculated in the soybean plants in different batches of experiments. For SSR, five batches of experiments (Ss_1~Ss_5) were performed from March 19th to September 16th in 2021. For ASR, two batches of the experiment (Pp_1~Pp_2) were run from March 4th to June 22nd in 2021. Each batch contains 1 control group (Healthy) and 3 to 4 experimental groups (i.e., inoculated soybeans, denoted V_1~V_4). HSI data was collected by a hyperspectral line scanner in the range from 1 day before inoculation up to a maximum of 8 days after inoculation (DAI) to trace the development of infection (cf. Table S2). The used hyperspectral camera has the following specifications: 'Corning microHSITM410 Vis-NIR Hyperspectral Sensor' with 75 spectral bands ranging from 402 nm to 994 nm. See a more detailed list of the spectral band information in Table S3.

2.3. Image pre-processing

Raw data are canopy images including several leaves and branches with an average spatial image resolution of 1368 × (2700±100) pixels. To keep the required computational capacity for the conducted experiments (including CNN model architecture comparisons) at an acceptable level, the study focuses on the CNNs' capability to classify fungal diseases on single leaves. Therefore, isolated leaves are cut out and stored separately from the original hyperspectral images. The image processing software FluxTrainer (version: 4.17.05) has been utilized to process the HSI data. The full image preprocessing workflow, as depicted in Fig. 2, consists of three steps: (1) *Background masking* is applied by the GlobalAverageLesser function with a 0.0800 threshold to remove background noise. All channels of an input pixel are first averaged, and the resulting average is then compared to the provided threshold value. If a pixel's value is smaller than the threshold, it will be considered as

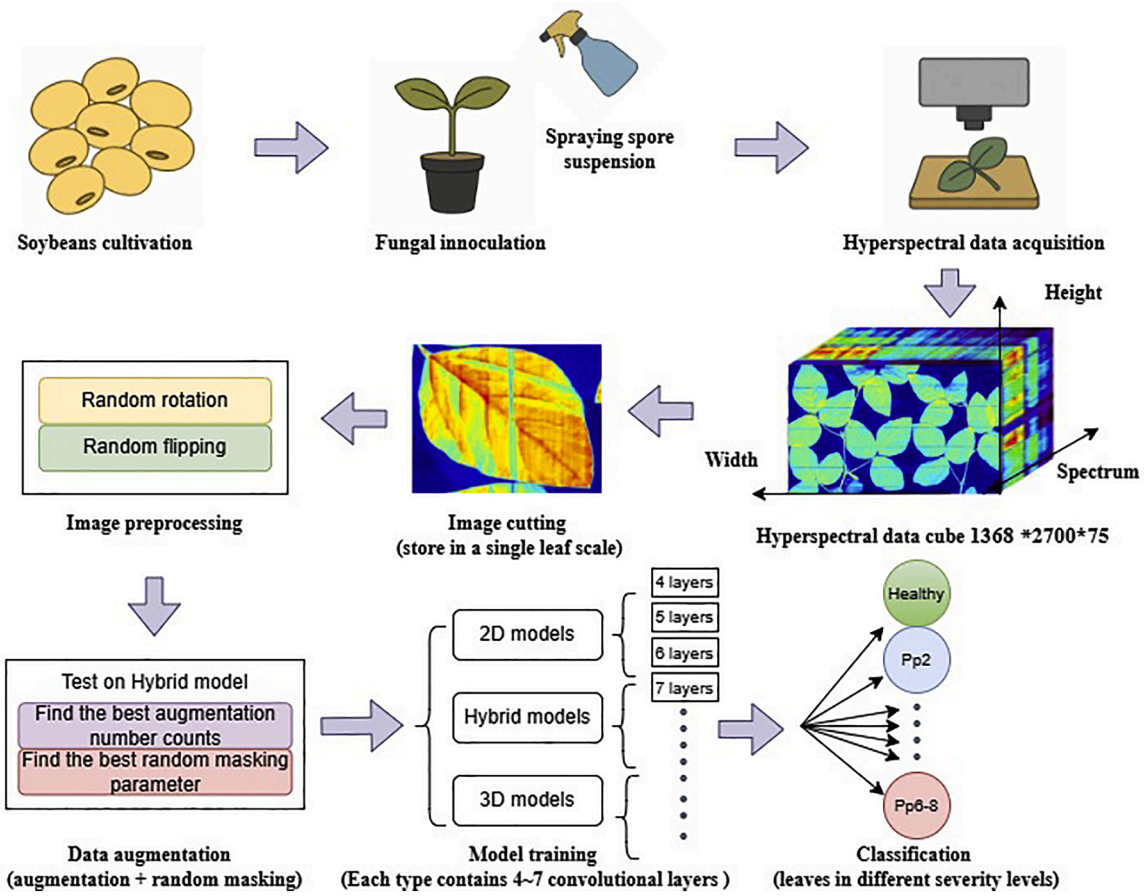


Fig. 1. The overall workflow of this study.

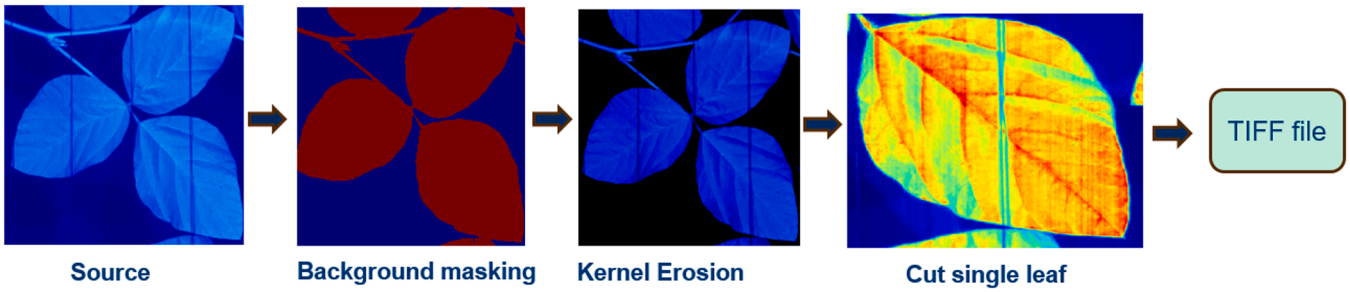


Fig. 2. The applied image processing workflow. The vertical lines inside the images are caused by fishing wires used to fix soybean samples on a glass plate for image collection by the HSI line scanner.

background, and vice versa. (2) After background masking, a morphological manipulation using an *Erosion* kernel is applied to smooth the edges of the leaves. The erosion kernel operates as a smoothing object in the image by replacing each pixel with the minimum pixel value within the kernel's range. (3) The final step of *single-leaf cutting* was performed manually in this work. We note that applying further, more advanced computer vision techniques could have been investigated to further automate the preprocessing, however, for this study's objective, we neglected this and concentrated on the actual task of disease classification on HSI data. Leaf images containing interference factors such as damage caused by insects, deformities, and curling are excluded, but overlapping leaves, i.e., leaves partially covered by others, are kept to maintain enough training samples and also imitate the canopy effect. See Fig. 3 for an example of pseudo-RGB images showing these disturbing factors. Before images are used in the model training, they are all resized to 605 × 555 with padding set as 0. This approach will make the biggest size input (605 × 555) be trained without any pixel loss and keep all inputs in the same size.

2.4. Image labeling

For the two different soybean diseases considered in this study, we utilized different classification schemes. For recognizing the SSR disease, because there are no visible symptoms that can be observed on leaves within the range of 8 DAI, the labeling method is based on the general infection dynamic as described in [23,25]. According to these studies, it takes 4 to 6 days after the infection for visible symptoms to appear on the plant's surface. Normally, during 5 to 7 DAI, white mycelium starts to grow in the stem and causes white bleach to appear on branches and pods. In this stage, the transportation of nutrients is inhibited, and the water content in the leaves starts reducing which leads to the beginning of withering. Besides, the content change of

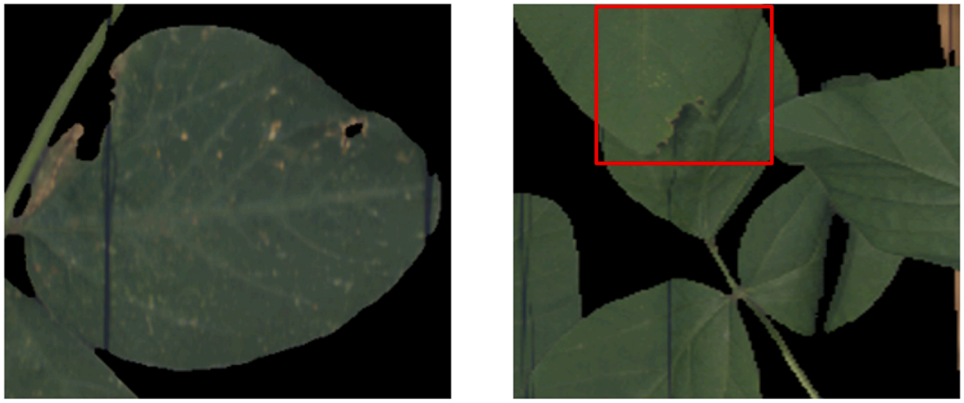


Fig. 3. (a) An example leaf showing non-pathogenic damage caused by insects. (b) An example of overlapping leaves (leaves that cover other leaves or are covered by other leaves). See the red frame highlighting the cover part.

secondary metabolisms caused by the infection such as the flavonoid and anthocyanin can reflect on hyperspectral signals [26,27]. Thus, SSR was separated into two classes: The first is denoted `Ss1–4′, which includes images falling in the range DAI1 to DAI4 from the experimental groups Ss1, Ss2, Ss3, and Ss4. This class thus represents the early stage of SSR infection which entails no (obvious) visible symptom on the whole plant. The second class is denoted `Ss5–8′ which contains images from DAI5 to DAI8, representing the mid-late stage of infection. In the stage Ss5–8, the symptoms were not visible on leaves but only visible on stems and branches as white lesions and mycelium appeared. However, in order to keep the consistency of the dataset, these infected stems and branches were excluded. The dataset only contains images in a single leave scale as shown in Fig. 1.

For the ASR disease, the labeling method is based on the severity scale of visible symptoms on the leaves. In our data, visible symptoms could only be observed without further zooming the images after DAI4. Grounded on [28,29], five classes with a particularly narrow range regarding the infection area in the early stage have been created: The class 'Pp2' contains most images from the DAI2 subset. In this class, there are no observable symptoms, even when amplifying images over 30 times. It thus represents the very early stage of infection. The second class, 'Pp3', represents the early infection stage with almost no symptoms to very light symptoms (i.e., 0.5 % to 2 % of leaf surface). In this class, there is still no obvious symptom that can be detected by the naked eye while some tiny black spots can be recognized when zooming in images >10 times. This class contains most of the images of the data subset DAI3. The third class, 'Pp4–5′, represents the early-mid stage of infection with visible symptoms occurring on 2 % to 10 % of the leaf surface. Images in this class were mainly selected from DAI4 and DAI5 but also included some images from the DAI6 subset when the symptom area met 2 % to 10 %. The fourth class is denoted 'Pp6–8′ and represents the mid-late infection stage with the lesion area on single leaves counting 10 % and more. Most DAI6 to DAI8 images are assigned to this class. Finally, a class named 'Healthy' has been created as a representation of healthy leaves for both schemes. Images from DAI-1 (before inoculation) and DAI0 in experimental groups as well as images in each control group are assigned to this class. In summary, seven classes have been created for the HSI training dataset. Table 1 depicts an example pseudo-RGB image calculated from the corresponding HSI datacube along with the number of images per class and furthermore illustrates the visible state of infection by zooming in salient spots. Fig. 4(a) demonstrates the spectral pattern of the corresponding classes in Table 1. Fig. 4(b) depicts the spectral pattern by zooming the region in Fig. 4(a) for the wavelength ranging from 520 nm to 580 nm. The spectral pattern for Ss1–4 and Ss5–8 are quite similar to Pp2. To avoid the confusion, they are not presented here.

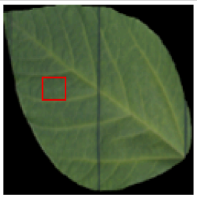
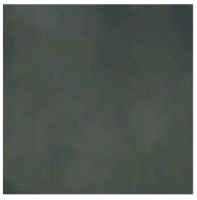
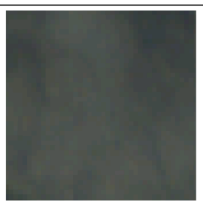
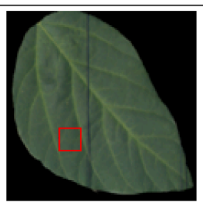
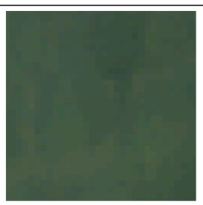
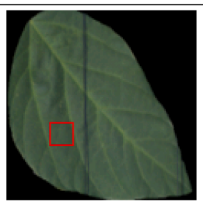
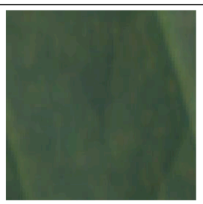
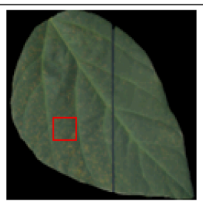
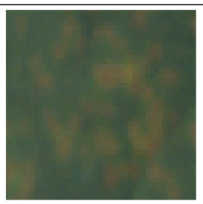
2.5. Image augmentation

Data augmentation has been applied to further enlarge the amount of available image data. By means of image transformation, the original HSI data is modified to make up new examples which in turn adds the resulting variation to the data pool. Applied image transformations comprise *random rotation* and *random flipping* which change the objects of interest geometrically inside images. However, these techniques do not change the spectral dimension in HSI data. To this end, *random masking* is introduced as a spectral transformation method. By randomly masking out a portion of pixels, it changes the average value of the spectral information and thus is intended to facilitate the model to learn incomplete data [30,31]. The masking ratio was tested in the range of 20 % to 40 % to find a suitable configuration. The transformation steps were applied to all images to increase the dataset before splitting into training, validation, and testing sets. The original dataset was augmented 2 to 12 times to test the optimal dataset size. The optimal performance was observed at 8 times augmentation, which counts in average 1056 samples per class (cf. Section 3.3). This augmentation size was applied for the rest of the models' training.

2.6. CNN architectures

Three different types of CNN models (2D CNNs, 3D CNNs, and 3D-2D CNN hybrids) are constructed to classify soybean leaf images. The depth of the networks, i.e., the number of convolutional layers, influences the predictive capacity of models [32]. According to [33], a model with 4 convolutional layers is recommended to do the physiological disorder diagnosis of plant leaves based on HSI with a small-size dataset (each class has <100 samples). However, the task in this study required the model to classify early stage symptoms which is very hard to observe. The difficulty of the task might lead to the need to increase the depth of convolutional layers. Thus, four-layer, five-layer, six-layer, and seven-layer architectures (in this study, the layer depth is the count of convolutional layers) have been constructed for each of the three types of CNN models to compare their performance. The used CNN backbone structure (i.e., the part of the deep neural network responsible for feature learning) is inspired by [10,34,35]. In the six-layer CNN model, the backbone consists of four 3D convolutional layers followed by two 2D layers. The reshaping method is applied to reduce the dimensionality by one to match the expected input of the 2D layers coming from the 3D layers. The kernel moves through the HSI datacubes to calculate the data information and transform it as the number on the output feature maps.

Table 1
Summary of the applied classification scheme with example pseudo-RGB images calculated from the preprocessed HSI datacubes and the amount of original samples in each class.

Class	Pseudo-RGB images	10 times zoomed images***	NC **
Healthy *DAI: -1			133
Ss1-4 *DAI: 2			143
Ss5-8 *DAI: 6			128
Pp2 *DAI: 2			121
Pp3 *DAI: 3			126
Pp4-5 *DAI: 5			133
Pp6-8 *DAI: 7			137

*For each depicted image, the DAI number, represents the data subset which was sampled from.
** In the rightmost column, number count (NC), which represents the total number of images existing in the class, is given.
***In the third column-'10 times zoomed images', salient spots (marked by the red frames) on the images in the second column –'Pseudo-RGB images' are zoomed in ten times to illustrate the visible state of the infection.

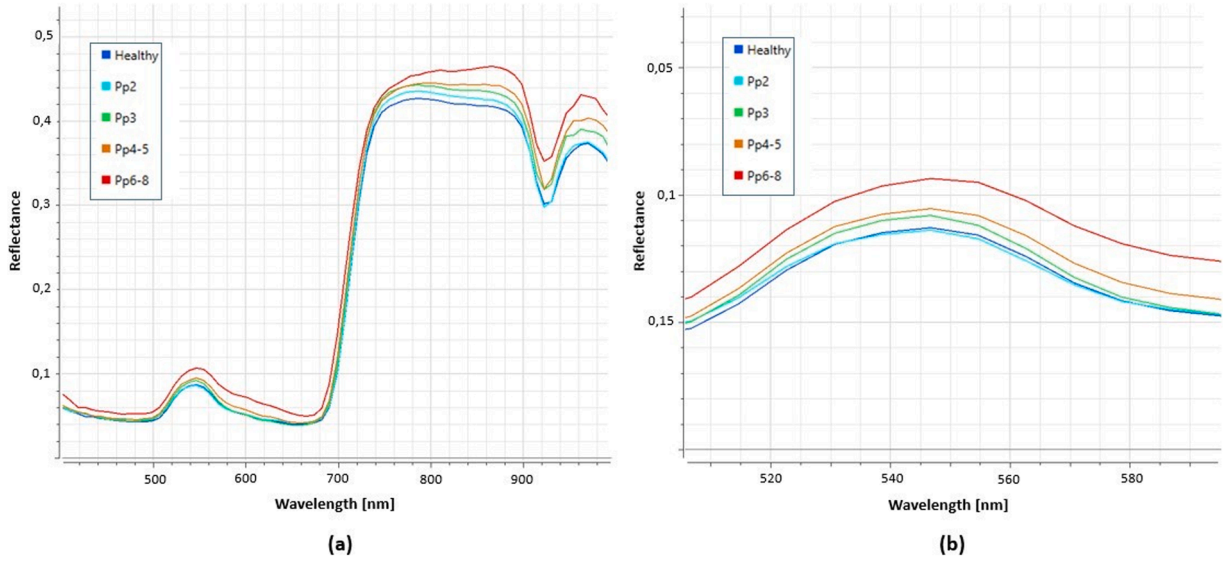


Fig. 4. (a) Spectrum pattern of classes Healthy, Pp2, Pp3, Pp4–5, and Pp6–8 depicted in Table 1. (b) Spectrum pattern by zooming the region of wavelength ranging 520 to 580 nm from (a).

The size of the kernels is presented in Table S5. The kernel filters contain parts of the learnable model parameters (i.e., the CNN's weights) which are optimized by backpropagation during training. Batch normalization is appended after every convolutional layer to standardize the output from one hidden layer before it is passed to the next hidden layer. It has been proposed to facilitate the convergence of deep neural network models [32,36].

Rectified linear unit (ReLU) served as an activation function between the convolutional layers and in the fully connected layers. It is defined as $f(x) = max(0,x)$. Unlike the sigmoid function which generally squeezes any input value into the range between 0 and 1, ReLU cuts off all negative values and maps them to 0 while the positive values are directly mapped by a linear function part. To this end, it has been found to allow for more complex tasks to be efficiently learned and is especially commonly utilized in multi-class classification [37,38].

The pooling layer, in general, is used for down-sampling the resulting feature maps resulting from the convolutional operations and activation functions. It reduces the size and parameters of preceding layers in the deep neural network. The feature map produced by the convolutional layer is based on the precise location of the input data, which leads to the problem of translation variance for different input images where the objects of interest occur in different regions. Pooling layers provide a solution to achieve translational invariance. In this study, max-pooling is concatenated after each ReLU activation layer. Max-pooling applies a kernel-operation that extracts the maximum number in the range of the kernel sliding over the feature maps. It thus, strives to retain only the most prominent information. The size of the pooling layer was adjusted along with different model types. They are set to reduce the size of output feature maps to a similar size before they go to the fully connected layer. See Table 4 and Appendix Tables S4, S5, and S6 for the details of model architectures and the output shape of feature maps before they go through the fully connected layer. The fully connected (FC) layer, also known as the dense layer, constitutes the standard type of neural network in which each neuron is connected to all neurons in the previous layer. Each neuron takes a weighted sum of all input values (usually after passing the activation function) from the previous layer and adds a bias term. In so-called dense layers, these operations are computed for all neurons in the current layer (usually in vectorized form). The FC layer consists of multiple dense layers followed by activation layers and dropout layers. It is placed at the end of the deep neural network, as a so-called head, to predict the target variables based on the extracted features from the preceding convolutional layers. The architecture of the FC for all models are same. It consists of a flatten layer, followed by 3 dense layer layers which firstly reduced the input feature dimension to 1024 and then 1024 to 512. In the final dense layer, the input feature dimension was reduced from 512 to the number of the classification category.

In this work, we constructed various CNN architectures which are specified in Appendix Tables S4, S5, and S6. Fig. 5 exemplarily depicts our hybrid 3D-2D deep neural network architecture.

In our experiments, 4-layer, 5-layer, 6-layer, and 7-layer 3D-2D CNN models have been designed with the number of 2D convolutional layers always maintained at two but the number of 3D convolutional layers varied in the range of two to five. Furthermore, fully 2D and fully 3D CNN models have been tested. The depth of the purely 2D and 3D CNNs was varied by including convolutional layers in the range of four to seven before the final FC layer. All CNN models' backbones are shown in Appendix Tables S4, S5, and S6. The main difference between the three considered types of architectures is the reshaping process that only exists in the hybrid 3D-2D model. The fully 2D and 3D models accordingly only apply 2D and 3D convolutional layers. Additionally, the kernel size and stride are adjusted in 3D and 2D CNNs to ensure proper output sizes. Another difference exists in the 7-layer 2D, 7-layer hybrid, and 7-layer 3D model. In order to maintain a proper output size, their number 5 convolutional layers are concatenated directly after the batch normalization of number 4 convolutional layers without going through a Max-pooling process. This might influence its comparability to other models but help to understand how the depth of the convolutional layer influences the performance of models.

2.7. Regularization methods

Overfitting is a phenomenon where a model performs very well in classifying data used for training while the performance tremendously declines when predicting previously unseen data. It results from the weights in the model which are trained to recognize features from the training dataset. If the number of training epochs on these training data is too high this can lead to the effect that any little change in the input data causes high prediction errors. This problem is particularly notable in small datasets – as we have in this study – because of the low data variance represented in the training data.

The first method for the regularization is called *dropout* [39]. Dropout randomly deactivates parts of the neurons in a layer at each update in the training step. This approach is expected to reduce the

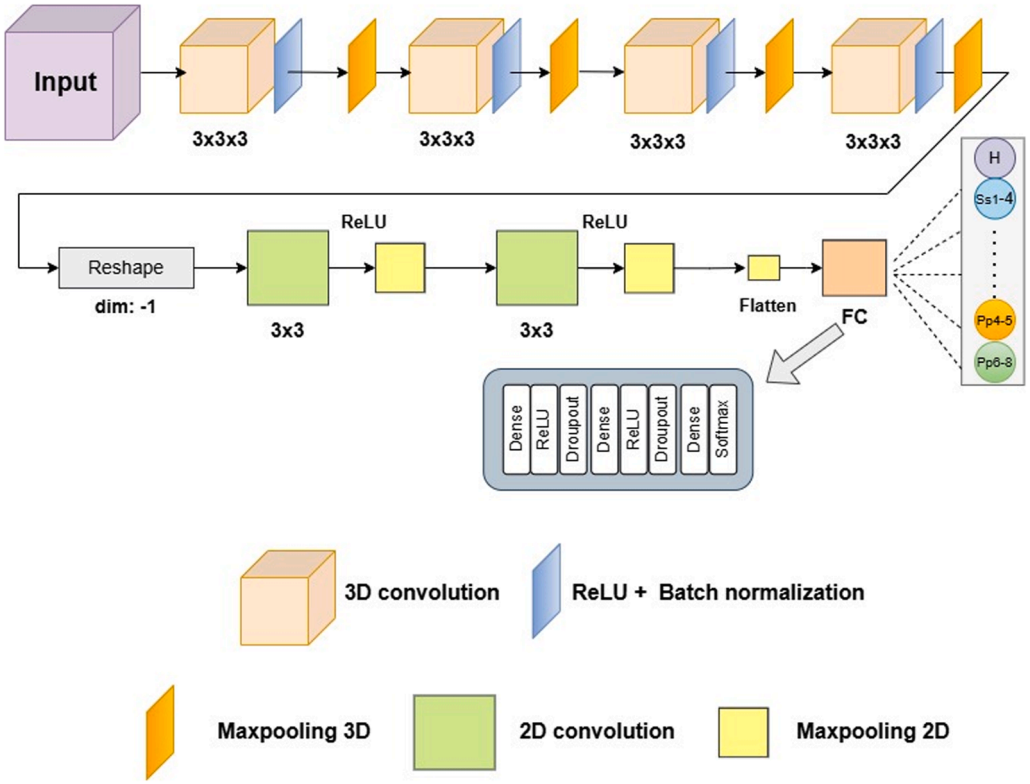


Fig. 5. Neural architecture of the 6-layer 3D-2D CNN model.

reliance on specific neurons and thus increase the robustness of the model. Such dropout layers are commonly placed in the FC layer. The choice of the hyperparameter 'dropout rate' is critical. A higher dropout rate introduces more regularization but may slow down training and impact the overall model capacity. In this study, through conducting a hyperparameter study, the optimal dropout rate has been found to be 0.25.

Weight decay is the method commonly used in the training step to optimize weights and avoid overfitting. The essential idea is to introduce a penalty for more complex models, effectively discouraging the learning algorithm from finding parameters (weights) that put too much emphasis on specific features [40]. In this study, we applied L2 regularization. It augments the standard loss function by adding a penalty term proportional to the square of the L2 norm of the model's weights. Formally, the term of L2 regularization $R(W)$ can be expressed as Eq. (1):

$$R(W) = \sum_{i}^{n} \frac{1}{2n} \lambda \cdot W_i^2 \tag{1}$$

Here, W represents the model's weight matrix, λ is the regularization strength, and W_i is the norm of the weight vector of the $i-$th layer. For the task of multi-class classification, the loss function (*Loss*) commonly employed is the cross-entropy loss, which we also applied during model training. Adding L2 regularization to this scenario involves combining the loss function with the L2 regularization term. In this study, the cross-entropy loss was applied from the *PyTorch* package – *CrossEntropyLoss*. It contains two steps to calculate the cross-entropy loss, which includes probability transformation and loss calculation with L2 regularization. The first step transforms the model's raw output logits, represented by X, into log probabilities for each class. The logits X are structured as a matrix with dimensions $(N,\ C)$, where N is the number of samples in a batch, and C is the total number of classes. For each sample i and each class j, the logit $X_{i,j}$ represents the unnormalized score for class j. To convert these logits into a stable probability distribution, *PyTorch Softmax* function is applied as Eq. (2):

$$p_{i,j} = \frac{\exp\left(X_{i,k}\right)}{\sum_{k=1}^{C} \exp\left(X_{i,k}\right)} \tag{2}$$

Where $\exp(X_{i,k})$ is the exponential of the logit for class j in sample i, ensuring non-negative values, The denominator, $\sum_{k=1}^{C} \exp\left(X_{i,k}\right)$, is the sum of the exponentiated logits for all classes C for sample i, enforcing that the resulting probabilities of each sample sum to 1. Thus, $p_{i,j}$ represents the model's estimated probability that sample i belongs to class j. To stabilize computation in the exponential function, *Pytorch* directly transforms the $p_{i,j}$ into log-probabilities as given by Eq. (3):

$$\log p_{i,j} = X_{i,j} - \log\left(\sum_{k=1}^{C} \exp\left(X_{i,k}\right)\right) \tag{3}$$

Once the log-probabilities are computed, the negative log-likelihood (NLL) is calculated to measure how closely the predicted probabilities match the true class labels Y, where Y_i is the correct class index for each sample i in a batch. For each sample i, the NLL for the true calss Y_i is computed as Eq. (4):

$$NLL\ Loss_i = -\log\left(p_i, Y_i\right) \tag{4}$$

The total cross-entropy loss (TCEL) is then expressed as the average over all N samples in the batch with the L2 regularization term $R(W)$ added to penalized large weights. It is given as Eq. (5):

$$TCEL = \frac{1}{N} \sum_{i=1}^{N} -\log(p_i, Y_i) + R(W) \tag{5}$$

The regularization term $R(W)$ penalizes large weights quadraticlly, and is thereby assumed to positively influence the optimization process during the error backpropagation. Overall, regularization methods in this study help to facilitate the robustness of the model and are intended to alleviate the potentially negative impact due to the small training dataset and the kind of replicated sampling naturally occurring through

image augmentation.

In the process of optimization, all weights are initially initialized randomly; they are then taught using the back propagation algorithm with the *Adam* optimizer and *Softmax* classification. We train the network for 100 epochs of minibatches 4 and both learning rate and weight decay parameters have been set to 10 E^−4 as found during hyperparameter optimization for our experiments.

2.8. Evaluation method

For evaluation purposes, we make use of the *confusion matrix* which constitutes a common method for the visualization of a model's performance in machine learning and which originates from the contingency table [41]. The evaluating indicators including *F1 score* (F1), *macro-averaged F1 score* (MAF1), *overall accuracy* (OA), *average accuracy* (OA), and *Kappa coefficient* (Kappa) are calculated by *precision* (P) and *recall* (R) based on the number of *true positives* (TP), *true negatives* (TN), *false positives* (FP) and *false negatives* (FN) which derived from the confusion matrix. The F1 score provides a comprehensive assessment of a model's ability to make accurate predictions across binary classes. The MAF1 is an extension of F1 to evaluate multiple classes [42]. The OA is a straightforward metric that reflects the proportion of correctly classified samples relative to the total number of samples. The AA, by contrast, calculates the mean accuracy across all classes, thus giving equal weight to each class regardless of its frequency. Kappa measures the agreement between predicted and true classifications while adjusting for the agreement that could occur by chance. These equations are given as Eqs. (6)–10:

$$F1 = \frac{2 \cdot P \cdot R}{R + P} \tag{6}$$

$$MAF1 = \frac{1}{n} \sum_{i=1}^{n} F1_i \tag{7}$$

Where $F1_i$ is denoted as the F1 score for each class, and n represents the number of all classes.

$$OA = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

$$AA = \frac{1}{2} \times \left(\frac{TP}{TP + FN} + \frac{TN}{FP + TN} \right) \tag{9}$$

$$P_e = \frac{(TP + FN) \times (TP + FP) + (FP + TN) \times (FN + TN)}{(TP + TN + FP + FN)^2} \tag{10}$$

$$Kappa = \frac{OA - P_e}{1 - P_e} \tag{11}$$

Each model was trained three times with a fixed ratio of train-test split as 0.7, 0.15, and 0.15 for the training, validation, and testing. Thus, the mean value of metrics calculated from three replications is presented in the next section reporting the results.

3. Results

3.1. Recognition of Asian soybean rust (ASR)

For the soybean disease detection of ASR, three types of model architectures (3D, hybrid 3D-2D, and 2D) with four different neural network depths (4 to 7 convolutional layers) are tested for their ability to recognize five classes of infection stages. All models used the image augmentation methods as described in Sections 2.5 and 3.3 for multiplying the training dataset's size by repeating the number of images in the training set 8 times. The means of the observed predictive performance metrics of the compared CNN architectures are reported in Table 2. The F1 scores for the two most different classes, Pp6–8 and Healthy, are particularly presented in the table to support the comparison of MAF1, OA, AA, and Kappa in different architectures. A model that has a high F1 score in the two most different classes represents a good performance in the classification because a good model should at least be able to recognize obvious symptoms and healthy leaves.

For the 2D CNN architectures, all evaluating indicators reach the highest value for a layer depth of five. Increasing the number of convolutional layers leads to decreasing values of evaluating indicators. The same observation can be made for the hybrid and the fully 3D architectures, for which the optimal performance appeared for a depth of six layers and five layers, respectively. 2D and 3D architectures both revealed the best performance when containing five convolutional layers, but the best 2D network model shows overall higher values of evaluating indicators compared to purely 3D models. Overall, the best performance is achieved by the hybrid model with six convolutional layers. This result indicates that for our disease detection task, combining an appropriate number of 3D convolutional layers with subsequent 2D layers can enhance the predictive capacity of CNNs as

Table 2
Evaluation of different CNN architectures for the task of early ASR disease detection on HSI data of soybean leaves.

Model	Layer depth**	MAF1±SD	OA±SD	AA±SD	Kappa±SD	F1 for Pp6–8*±SD	F1 for Heathy*±SD
2D	4	0.396±0.019	0.423±0.015	0.429±0.017	0.316±0.025	0.788±0.021	0.375±0.030
	5	**0.473±0.020**	**0.461±0.019**	**0.479±0.016**	**0.365±0.023**	**0.886±0.015**	**0.484±0.016**
	6	0.453±0.013	0.442±0.011	0.466±0.013	0.347±.017	0.876±0.007	0.413±0.019
	7	0.270±0.026	0.267±0.023	0.278±0.024	0.161±0.019	0.712±0.005	0.225±0.028
Hybrid	4 (2 × 3D + 2 × 2D)	0.315±0.014	0.329±0.009	0.345±0.013	0.241±0.013	0.740±0.018	0.214±0.042
	5 (3 × 3D + 2 × 2D)	0.469±0.021	0.457±0.017	0.463±0.023	0.365±0.018	0.861±0.023	0.183±0.011
	6 (4 × 3D + 2 × 2D)	**0.537±0.009**	**0.513±0.015**	**0.534±0.011**	**0.401±0.016**	**0.892±0.012**	**0.325±0.014**
	7 (5 × 3D + 2 × 2D)	0.384±0.037	0.395±0.028	0.386±0.025	0.307±0.024	0.777±0.007	0.296±0.027
3D	4	0.240±0.017	0.251±0.014	0.242±0.013	0.176±0.015	0.726±0.011	0.285±0.022
	5	**0.334±0.019**	**0.346±0.010**	**0.358±0.015**	**0.247±0.012**	**0.773±0.008**	**0.435±0.007**
	6	0.262±0.009	0.288±0.008	0.295±0.012	0.158±0.009	0.722±0.010	0.326±0.012
	7	0.226±0.025	0.238±0.022	0.255±0.018	0.163±0.024	0.651±0.016	0.267±0.015

* F1 scores for Pp6–8 and Healthy represent the prediction capacity for the two most different classes.
**The layer depth indicates the number of convolutional layers preceding the final FC layer.

Table 3
MAF1 scores and F1 scores of each class for the best performing 2D and hybrid CNN model for the task of early ASR disease detection on HSI data of soybean leaves.

Model	Layer depth	MAF1±SD	Pp6–8±SD	Pp4–5±SD	Pp3±SD	Pp2±SD	Healthy±SD
2D	5	0.473±0.020	0.886±0.015	0.380±0.020	0.276±0.021	0.340±0.038	0.484±0.016
Hybrid	6 (4 × 3D + 2 × 2D)	0.537±0.009	0.892±0.012	0.581±0.027	0.462±0.023	0.427±0.011	0.325±0.014

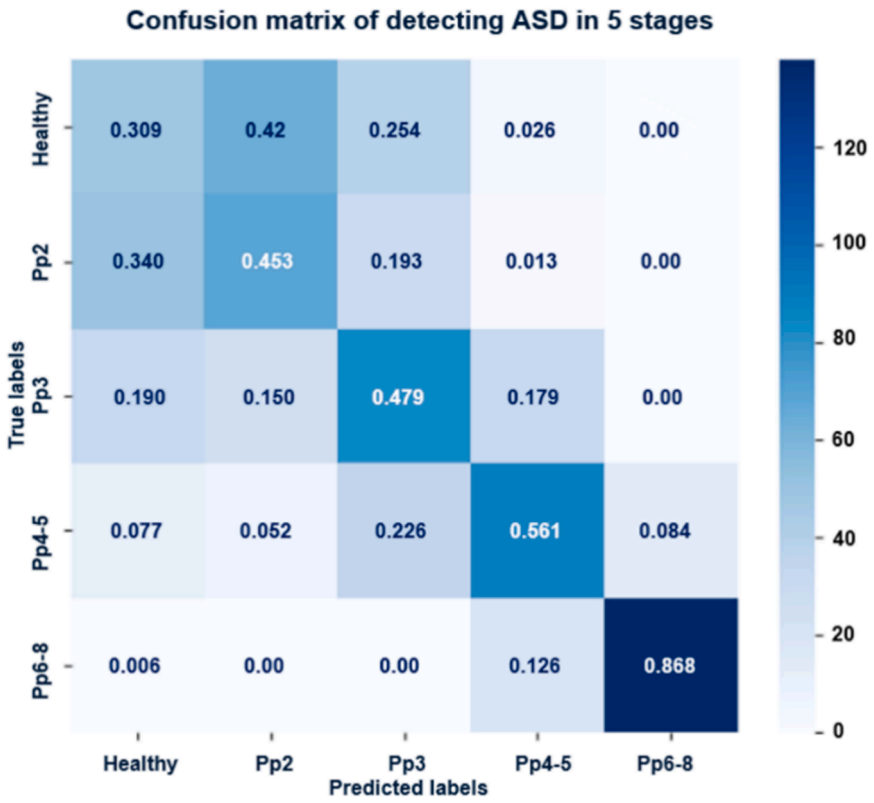


Fig. 6. Normalized confusion matrix obtained by classifying four stages of ASR and healthy plants by the 6-convolutional layer hybrid CNN model. The x-axis represents labels predicted by the model. The y-axis represents the true labels of samples. Figures in the matrix are normalized by calculating the percentage of correctly predicted samples in corresponding labels, which is also known as class-specific accuracy. The bar on the right presents the counts of samples in the matrix. The samples in each ASR class are selected and amplified by a factor of eight in the augmentation process.

compared to fully 3D or 2D architectures when the input consists of 3D HSI datacubes.

Table 3 shows the results of a further investigation of the F1 scores for each class for the best-performing models, i.e., the 5-convolutional layer 2D model and the 6-convolutional layer hybrid model. As can be seen, the hybrid model not only demonstrates high prediction performance in identifying obvious symptoms in the mid-late ASR infection stage (Pp6–8) but also presents an acceptable performance in detecting ASR in the early-mid stage (Pp4–5) in which only 2–10 % of the leaves area shows visible symptoms. In the early and very early stages of infection (Pp3 and Pp2), the hybrid model again shows substantially increased F1 scores when compared to a pure 2D CNN. However, 2D CNN performs better in the Healthy class. Overall, the hybrid CNN model revealed the best performance in recognizing the ASR disease on HSI data.

The normalized confusion matrix of the 6-layer hybrid model is depicted in Fig. 6. The model shows the highest accuracy in classifying ASR infection in stage Pp6–8, and it also demonstrates an acceptable capability to detect Pp4–5 and Pp3. However, the class-specific recall for Pp3 is 0.479 which indicates poor prediction ability in the range of early infection stage (only 0.5 %~2 % of leaf area with visible symptoms). The class-specific recall for Pp2 is similar to Pp3 which is under 50 %. Furthermore, it can be observed that the classes Pp2 and Healthy both have >30 % of images recognized wrongly as belonging to each other. This indicates that the model struggles to distinguish the difference in images between these two classes. In the class Pp-4–5 and Pp 6–8, images are mostly predicted correctly but some images are predicted as the classes nearby in the confusion matrix, which presents a dispersion of the prediction.

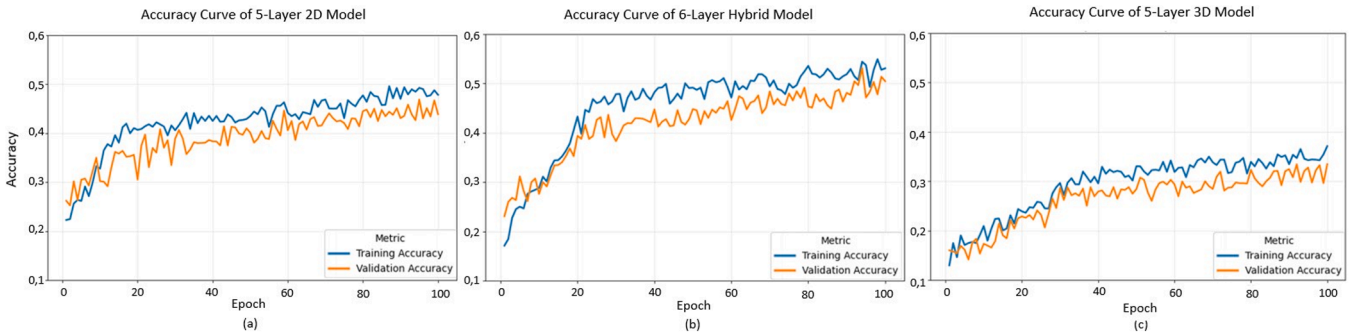


Fig. 7. The accuracy curves of training and validation were obtained after each epoch and presented as (a) the accuracy curve of the 5-layer 2D model, (b) the accuracy curve of the 6-layer hybrid model, and (c) the accuracy curve of the 5-layer 3D model.

3.2. Detecting Sclerotinia stem rust (SSR) and Asian soybean rust (ASR) in the mid-late infection stage

The six-layer depth hybrid model demonstrates the best capability in the previous experiment. For the next experiment, we thus applied this candidate to now detect SSR. Unlike the result of classifying ASR samples which has a good performance in prediction, the six-layer depth model predicted all samples as Ss1–4. This result might be because the model can not identify the difference between Healthy, Ss1–4, and Ss5–8. After all, all of them have no visible symptoms. As a consequence, the model predicts all samples to the class having the largest amount of training instances compared to other classes to reduce the cross-entropy loss. To verify this assumption, the classes Pp4–5 and Pp6–8 as the two most obvious ones for the ASR disease are selected with SSR samples to test the model. The corresponding normalized confusion matrix is depicted in Fig. 8.

In the confusion matrix, the prediction of the Pp6–8 class has greater than 92 % class-specific recall and only a marginal fraction is predicted as Pp4–5. The prediction of Pp4–5 disperses to Pp6–8 and Ss1–4, a similar phenomenon as observed in the previous experiment (cf. Fig. 6). However, the instances of Healthy, Ss1–4, and Ss5–8 are still accumulated as predictions of Ss1–4 by the model. This result demonstrates that the model is not able to distinguish these three classes. The reason that the distribution of the prediction accumulates at Ss1–4 is deemed to occur due to Ss1–4 having the largest amount of training instances (cf. Table 1).

3.3. Impact of image transformation and augmentation on the hybrid model

In this study, image augmentation has been applied in the experiment to tackle the data scarcity problem (cf. Section 2.5). Geometric transformation of the images including random rotation and random flipping have been used. The r*andom masking* which serves as spectral transformation, is particularly tested for its masking ratio ranging from 0.2 to 0.4. The experiment is conducted based on the six-layer depth hybrid model because it showed the best performance in the classification of ASR disease. Table 5 summarizes the results.

The evaluating indicators of MAF1, OA, AA, and Kappa increase with the samples in each class augmented from 264 to 1056. When the data set size including the augmented training samples increased to over 1056 in each class, the evaluating indicators however were found to decline. Accordingly, a saturation of indicators was observed when the original dataset was amplified by factor eight, and the amount of training samples reached a number of 1056.

Comparing the model's performance in different ratios of random masking, it has been found that for the ratio hyperparameter set to 0.3, on average higher evaluating indicators across different dataset sizes can be observed. For a smaller set masking ratio of 0.2, mostly the indicators have been found inferior except for the largest augmented dataset. However, the indicators decreased in every setup when the masking ratio was set to 0.4, which indicates that a higher ratio of area masking can harm the model's learning ability. Overall it can be stated that an appropriate amount of training samples by means of augmentation with a proper ratio of random masking in our case led to a substantially increased model performance in terms of MAF1, OA, AA, and Kappa.

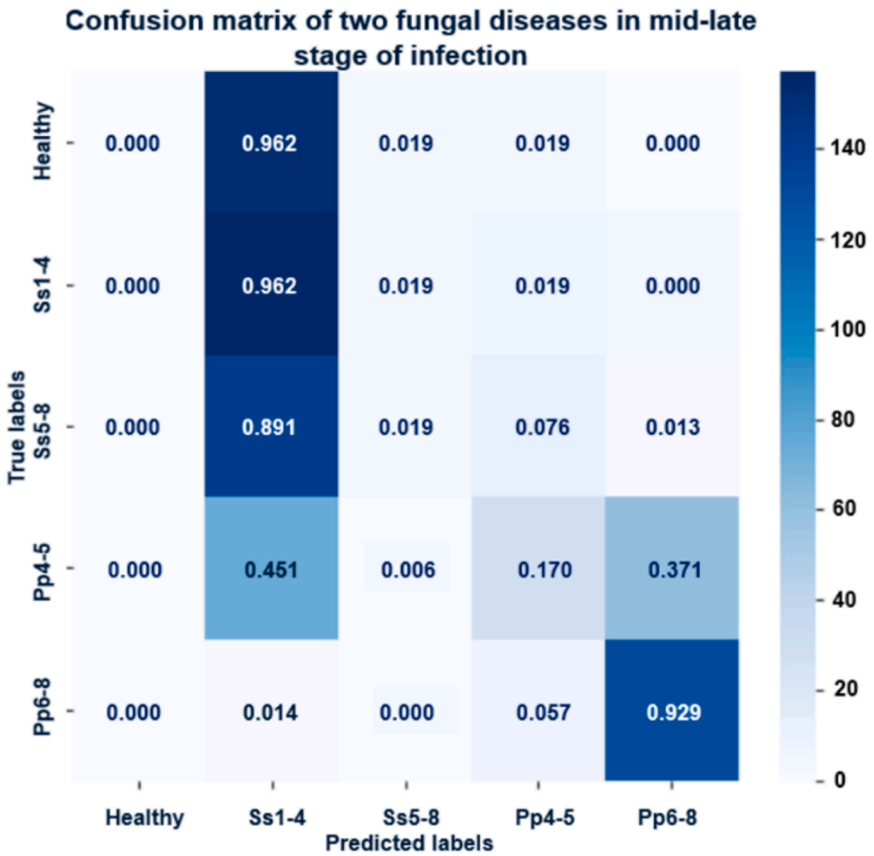


Fig. 8. The normalized confusion matrix of the detection of 2 fungal diseases in the early-mid and mid-late stages by 6-layers 3D-2D CNN. The x-axis represents labels predicted by the model. The y-axis represents the true labels of samples. Figures in the matrix are normalized by calculating the percentage of correctly predicted samples in corresponding labels, which is also known as class-specific recall. The bar on the right presents the counts of samples in the matrix. The samplers in each class are selected and multiplied 8 times in the augmentation process.

Table 4
The output dimension of the last Maxpooling layer and the total trainable parameters in each model.

Model	2D 4-layer	2D 5-layer	2D 6-layer	2D 7-layer	Hybrid 4-layer	Hybrid 5-layer	Hybrid 6-layer	Hybrid 7-layer	3D 4-layer	3D 5-layer	3D 6-layer	3D 7-layer
*Output shape of the last Maxpooling layer	6 × 5	6 × 5	7 × 6	7 × 6	6 × 5	6 × 6	7 × 6	7 × 7	2 × 6 × 5	2 × 6 × 6	2 × 7 × 6	2 × 7 × 6
Total trainable parameters	17,792,837	17,848,325	24,141,365	24,206,117	17,831,549	21,888,517	24,098,501	27,170,885	35,997,605	40,462,037	46,236,341	46,266,437

* Output shapes of the 2D and hybrid models follow the structure as height × width, and output shapes in the 3D models are depth × height × width. The filter number for all output shapes from their last Maxpooling layer is 512 which has been specified in Table S5, S6, and S7.

4. Discussion

4.1. Limited predictive capacity of the model for non-visible symptoms of SSR

In the process of manual labeling the leaf samples inoculated with SSR disease, we have confirmed that there are no visible symptoms on the leaves' surface for the infection stages Ss1–4 and Ss5–8. The criteria we used to separate these two classes is the existing white lesion on branches. According to [24,43], the metabolic response change of infected leaves starts from 5 days after the inoculation. However, the response of soybean plants to the same inoculation method can vary among different environmental conditions and the resistance of the plant. Literature investigating a CNN approach to detect SSR turned out to be scarce. To the authors' knowledge, there is no proper insight available for standardizing the labeling method for detecting SSR at leaf scale with different degrees of infection severity. This knowledge gap could be due to the morphology of this disease. Since SSR is a soil-based pathogen that infects the soybean from the root, symptoms that appear on the leaf surfaces are not apparent in the early to middle stages of infection. In today's practical disease-control assessments, symptom recognition on leaves is not dominant because the white lesions at the roots are visible in the early stage already [44]. Considering the samples available for this study, although some wither and defoliate phenomena have been lightly apparent, an exact confirmation of the cause of this symptom being SSR is difficult because some of the symptoms are similarly prevalent after insect attacks or other biotic and abiotic stressors [45,46].

Overall, the research regarding the identification of the SSR fungal disease on soybean leaves at a very early stage remains scarce. One related study on tomato bacteria disease detection indicates that the hyperspectral signal of non-symptomatic leaves is highly similar to healthy leaves and accordingly, the biological change that induces the transition of optical signals occurs with the appearance of visible symptoms or withering parts on the leaves [47]. This can be brought in line with the results of our study, i.e., that the model struggles to distinguish leaves being either healthy or infected but without visible symptoms on their surfaces (cf. Table 1).

4.2. Impact of CNN architecture and depth on disease detection performance

Our reported results for detecting ASR in different severity scales with different CNN architectures and depths reveal that in our experimental setting, a hybrid CNN model comprised of four 3D-convolutional layers and two subsequent 2D-convolutional layers has superior performance compared to pure 3D-CNNs. We deem the vast number of trainable parameters of a pure 3D model in combination with the small dataset available a potential explanation for this observation. The amount of trainable parameters for the 6-layer 3D model is 190 % more than the 6-layer hybrid model (cf. Table 4). It took 18, 25, and 37 epochs for the 5-layer 2D model, 6-layer hybrid model, and 5-layer 3D model to converge respectively, see Fig. 7. According to [48–50], excessive numbers of convolutional layers and kernel size can lead to decreased overall accuracy in classification. While on the one hand, deeper convolutional layers can extract more abstract and complex features, it has been found that this requires a proper dataset size during training to improve the classifying capability of a model [51]. If the dataset is insufficient for training, this usually causes overfitting of the model which leads to higher generalization errors [52–54].

In this study, we found that the evaluating indicators decrease when the models' number of successive convolutional layers is less than five. Improved disease detection performance is observed in the six-layer model of the hybrid architecture and the five-layer model of the fully 2D-architecture. This result is contradictory to the insights reported in [55], which found that two layers of 3D-CNNs can reach 95 % accuracy

Table 5
The influence of the augmentation in the model performance.

Model	Amplification factor	Training sample count per class (average)	Ratio of random masking	MAF1 ± SD	OA± SD	AA± SD	Kappa± SD
Hybrid (6-layer)	2	264	0.2	0.317 ± 0.019	0.325±0.017	0.331±0.011	0.241±0.013
			0.3	0.346 ± 0.012	0.338±0.010	0.350±0.014	0.252±0.015
			0.4	0.340 ± 0.011	0.335±0.016	0.345±0.008	0.248±0.011
	4	528	0.2	0.453 ± 0.009	0.437±0.014	0.462±0.007	0.364±0.012
			0.3	0.474 ± 0.014	0.461±0.013	0.479±0.015	0.376±0.015
			0.4	0.446 ± 0.007	0.431±0.020	0.455±0.017	0.358±0.018
	6	792	0.2	0.475 ± 0.006	0.456±0.012	0.470±0.009	0.395±0.014
			0.3	0.481 ± 0.027	0.476±0.015	0.492±0.016	0.391±0.019
			0.4	0.458 ± 0.013	0.452±0.013	0.464±0.011	0.377±0.008
	8	**1056**	0.2	0.526 ± 0.007	0.505±0.011	0.538±0.009	0.399±0.013
			0.3	**0.537 ± 0.009**	**0.513±0.015**	**0.534±0.011**	**0.401±0.016**
			0.4	0.518 ± 0.006	0.496±0.010	0.520±0.014	0.382±0.012
	10	1320	0.2	0.492 ± 0.025	0.484±0.018	0.513±0.015	0.385±0.016
			0.3	0.516 ± 0.009	0.503±0.009	0.524±0.007	0.391±0.010
			0.4	0.504 ± 0.012	0.489±0.014	0.511±0.011	0.388±0.008
	12	1584	0.2	0.513 ± 0.008	0.497±0.021	0.519±0.017	0.395±0.018
			0.3	0.510 ± 0.004	0.493±0.013	0.515±0.009	0.387±0.015
			0.4	0.479 ± 0.013	0.471±0.016	0.488±0.012	0.360±0.016

over all classes in recognizing Charcoal rot disease at the soybean stem. A possible reason for this discrepancy in the number of required convolutional layers to achieve viable performance metrics is the more complex task in our case. Instead of classifying images using a binary classification scheme (infected or healthy), we applied a multi-class scheme for detecting ASR at different stages of the disease. We hypothesize that this increase in the complexity of the underlying classification task results in the need for deeper CNN architectures. The results reported in [56], which used both six-layer 2D CNN and four-layer 3D CNN structures to classify the gray mold disease of strawberries in three classes at the leaf scale support our rationale.

Compared to [21,22] which weighted or reduced the spectral bands by PCA, random forest, and attention mechanism. Our approach by training whole spectrum data demonstrates lower performance in accuracy. However, these studies focus on the infection stage where over 25 % leaf area has the symptom. During this stage, the difference in the spectrum is obvious, while the difference in the early stage is unclear. The reduction of spectral channels may miss important information. One another study indicates the high accuracy in detecting early infection of asymptomatic leaves by weighting the contribution of spectral channels. Instead of using a whole leaf as a sample, it extracted regions of interest from leaves [23]. This approach did not consider the overall severity of leaves but it unveils the possibility of improving model performance by weighting spectral bands. In the future, reducing spectral channels or weighting channels' contribution will be applied for further research.

4.3. Impact of image augmentation on CNN-model training

In Table 5, the optimal amount of samples per class in our setting has been identified to be 1056, which constitutes an eight times augmentation of the original dataset. A further increase in the amount of samples has been found to reduce the values of evaluating indicators. This can be relevant for dealing with the often-faced limitation of small datasets [57]. Random rotation and random flipping only inject variance by changing the geometric information of the data [58]. However, the spectral variance is not affected. When overly augmenting the original data, the efficacy of changing geometric information declines because the variance that can be achieved by these two transformation methods is limited by the original data. Regarding the influence caused by the different masking ratios, the optimal ratio for our setting has been found to be 0.3. The observed decline in performance for higher masking rations (0.4) is most likely because it masks out too much important information vital for learning. The lower masking ratio (here set to 0.2) appears not sufficient for adding proper variance to the augmented dataset. It can thus be concluded that random masking can be a promising augmentation method when dealing with HSI datacubes, as long as a proper masking ratio is set.

4.4. Limiting factors influencing the accuracy of ASR detection

Although the proposed hybrid CNN model demonstrates viable recognition performance for the applied five-stage classification scheme, some samples have been wrongly predicted. The confusion matrix in Fig. 6 shows that the hybrid model is not able to distinguish the difference between the Healthy class and Pp2 class, i.e., a notable number of samples belonging to these two classes are misclassified as each other. This result indicates that the slight difference in the very early stage of infection is not properly distinguishable from uninfected, i.e., healthy leaves. In the early stage of infection, denoted as Pp3 in our class scheme, a similar result is observable, i.e., some samples have been misclassified as Healthy, Pp2, and Pp4–5. Also for class Pp4–5, a couple of samples have been misclassified to the nearby classes. This observation could be caused by the difficulty in manually labeling the leave data at hand. In our study, the definition of each class for the ASR disease is based on the severity scale of the visible symptoms on the leaves' surface. It is considered a continuous severity scale that has no distinct 'gap' between two nearby classes. As a consequence, images at the 'edge' of one part of the scale might be recognized as instances from the directly neighboring classes on the scale. For example, an image that is manually detected as having 4 % visible symptoms on the leaf area has been labeled as Pp4–5, but it could have been classified as Pp3 too, because its features are highly similar to this class, and vice versa.

In the confusion matrix shown in Fig. 6, a small amount of the images in the classes Healthy and Pp4–5 have been confused with each other and even with class Pp6–8. These outliers could be due to limitations during the image preprocessing. In the process of taking HSI photos, fishing rope is used to fix the leaf samples. Those fishing ropes are thus also included in the images. Although background masking has been applied to remove the interference of the images' background, those ropes spanned above the leaves' surface are difficult to remove with standard methods. Furthermore, the distribution of the appearance of fish ropes in each image is uneven, which potentially disturbs the training of the model.

Another disturbing factor is caused by overlapping leaves. Because of the limited size of the HSI dataset, the images that contain overlapping leaves must have also been utilized in the model training. Considering the spectral signature plot in Fig. 9(a), it can be seen that both overlapping healthy and overlapping Pp4–5 leaves have higher reflectance factors compared to healthy and Pp4–5 leaves in wavelengths ranging from 720 nm and 900 nm. In the wavelength between 520 and 580, the reflectance factor of overlapping healthy leaves is slightly higher than healthy leaves, while the overlapping Pp4–5 is lower than Pp4–5 leaves,

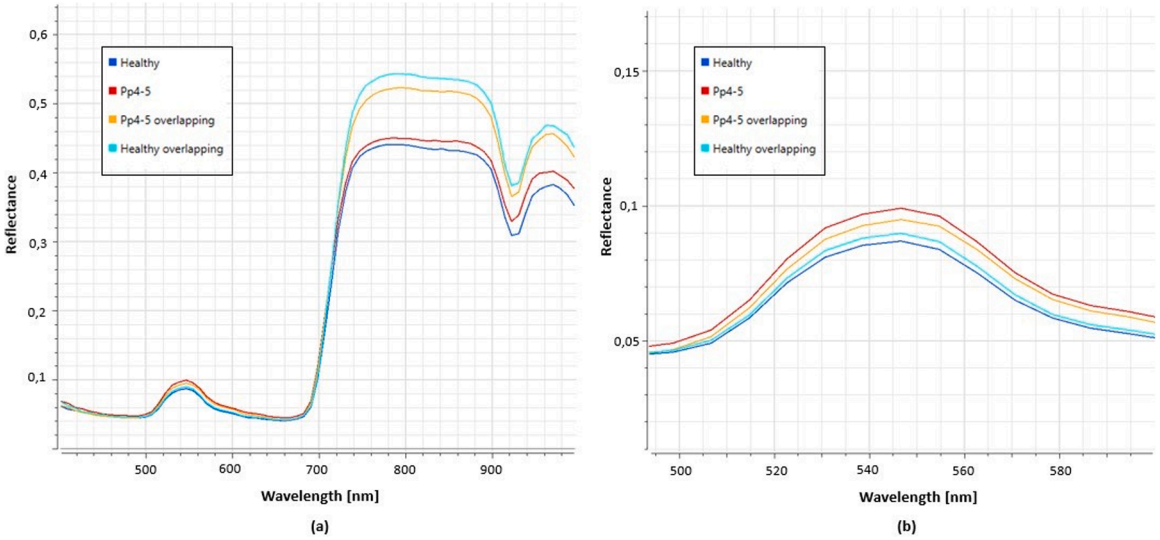


Fig. 9. Spectral signature plots of the healthy leaves, overlapping healthy leaves, leaves falling in the class Pp4–5, and Pp4–5 image of overlapping leaves in (a) full spectrum range and (b) spectrum ranging from 500 nm to 580 nm.

see Fig. 9(b). Our results align with the findings in [21,23]. According to these studies, the reflectance factor of healthy leaves in the range of 750 to 900 nm is slightly lower than infected leaves but it doesn't match our result when comparing healthy overlapped leaves with the Pp4–5 overlapped leaves. This can be explained by the interference of the leaf-overlapping which increase the reflectance factor in the spectral range from 750 to 900 because the overlapping part is expected to have a higher light scattering phenomenon [59]. In summary, the observed limitation of misclassification of some samples in Pp4–5 as instances of Pp3 and Pp2 is assumed to be due to the existence of images with overlapping leaves as well as the applied narrow severity scale. This demonstrates the difficulty of the task of early disease detection of fungal pathogens at the leaf scale.

The main limitation of this study is found to be the small size of the HSI dataset for training. The purpose of the study is to use CNN to recognize two fungal soybean diseases at a leaf scale. However, the limited size of the dataset makes it necessary to include images of overlapping leaves to ensure enough data for training. On the one hand, it introduces a sort of noise into the dataset while on the other hand, it can also be seen to a certain extent as simulating the canopy of soybean plants under field conditions.

In future work, the acquisition of hyperspectral canopy images collected from field or greenhouse trials is pursued. Considering the current state of practical disease control of soybean cultivation operated in the fields, several more influencing factors need to be taken into account. The application of CNN-based disease prediction models to hyperspectral images of plant stands collected by UAVs or agricultural machines constitutes another branch of future research. Besides, methods of dimension reduction and weighting channels' contribution should be considered to further improve the model's performance.

5. Conclusion

This study proposes and compares different CNN architectures for the early detection of fungal soybean diseases based on hyperspectral images of plant leaves. Our experiments reveal that a hybrid model combining 3D convolutional layers with succeeding 2D convolutional layers achieves the best recognition performance compared to pure 3D-CNNs and pure 2D-CNNs. An increasing depth of the CNNs has been found to not always increase the models' predictive performance. Contrarily, a proper depth of the convolutional neural networks has to be optimized to achieve viable classification performance. The finally proposed hybrid CNN model demonstrates viable performance in detecting the ASR disease on infected leaves in the early-mid infection stages. When tasked to recognize the disease in the early stages of infection (DAI3), which usually shows only 0.5 % to 2 % of symptomatic leaf area, the class-specific recall shows only a score of 0.48. After analyzing the distribution of the predictions in the confusion matrix, the hybrid model however is still found to show promising potential in detecting the disease in this early stage, because the majority of misclassified samples have been classified to the nearby classes which are naturally difficult to differentiate, even for humans. On the downside, the hybrid model revealed a lack of capability to reliably detect the very early stage of the infection with ASR, where no symptoms are observable on the leaves' surface. Furthermore, the investigated CNN models are not able to detect the SSR disease utilizing our class scheme for differentiating several infection stages.

The utilization of image augmentation by means of including common geometric transformation and a random masking technique is found to be advantageous when extending the size of small HSI datasets. It thus demonstrates the promising potential for further research on deep neural networks applied for automated disease monitoring based on non-destructive sensing methods in agricultural fields.

Ethics statement

Not applicable: This manuscript does not include human or animal research.

CRediT authorship contribution statement

Chieh Fu Hsiao: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Georg Feyrer:** Writing – review & editing, Validation, Supervision, Software, Methodology, Conceptualization. **Anthony Stein:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

Funding: This research was conducted in the NOcsPS project funded by the Federal Ministry of Education and Research, grant number 031B0731A. Responsibility for the content of this publication lies with the authors. Furthermore, we thank Prof. Dr. Ralf Vögele and his team of the Department of Phytopathology for original data collection and sharing. Additionally, we thank Prof. Dr. Joachim Müller for reviewing the master's thesis of the first author.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.atech.2025.101023.

Data availability

Data will be made available on request.

References

[1] T.W. Allen, C.A. Bradley, A.J. Sisson, E. Byamukama, M.I. Chilvers, C.M. Coker, A. A. Collins, J.P. Damicone, A.E. Dorrance, N.S. Dufault, P.D. Esker, T.R. Faske, L. J. Giesler, A.P. Grybauskas, D.E. Hershman, C.A. Hollier, T. Isakeit, D.J. Jardine, H. M. Kelly, R.C. Kemerait, N.M. Kleczewski, S.R. Koenning, J.E. Kurle, D.K. Malvick, S.G. Markell, H.L. Mehl, D.S. Mueller, J.D. Mueller, R.P. Mulrooney, B.D. Nelson, M.A. Newman, L. Osborne, C. Overstreet, G.B. Padgett, P.M. Phipps, P.P. Price, E. J. Sikora, D.L. Smith, T.N. Spurlock, C.A. Tande, A.U. Tenuta, K.A. Wise, J. A. Wrather, Soybean yield loss estimates due to diseases in the United States and Ontario, Canada, from 2010 to 2014, Plant Health Prog. 18 (1) (2017) 19–27, https://doi.org/10.1094/php-rs-16-0066.

[2] E. Silva, J.P. Da Graça, C. Porto, R.M.D. Prado, E. Nunes, F.C. Marcelino-Guimarães, M.C. Meyer, E.J. Pilau, Untargeted metabolomics analysis by UHPLC-MS/MS of soybean plant in a compatible response to Phakopsora pachyrhizi infection, Metabolites 11 (3) (2021) 179, https://doi.org/10.3390/metabo11030179.

[3] C.A. Bradley, T.W. Allen, A.J. Sisson, G.C. Bergstrom, K.M. Bissonnette, J. Bond, E. Byamukama, M.I. Chilvers, A.A. Collins, J.P. Damicone, A.E. Dorrance, N. S. Dufault, P.D. Esker, T.R. Faske, N.M. Fiorellino, L.J. Giesler, G.L. Hartman, C. A. Hollier, T. Isakeit, T.A. Jackson-Ziems, D.J. Jardine, H.M. Kelly, R.C. Kemerait, N.M. Kleczewski, A.M. Koehler, R.J. Kratochvil, J.E. Kurle, D.K. Malvick, S. G. Markell, F.M. Mathew, H.L. Mehl, K.M. Mehl, D.S. Mueller, J.D. Mueller, B. D. Nelson, C. Overstreet, G.B. Padgett, P.P. Price, E.J. Sikora, I. Small, D.L. Smith, T.N. Spurlock, C.A. Tande, D.E.P. Telenko, A.U. Tenuta, L.D. Thiessen, F. Warner, W.J. Wiebold, K.A. Wise, Soybean yield loss estimates due to diseases in the United States and Ontario, Canada, from 2015 to 2019, Plant Health Prog. 22 (4) (2021) 483–495, https://doi.org/10.1094/php-01-21-0013-rs.

[4] J.G.A. Barbedo, A review on the main challenges in automatic plant disease identification based on visible range images, Biosyst. Eng. 144 (2016) 52–60, https://doi.org/10.1016/j.biosystemseng.2016.01.017.

[5] A. Jain, S. Sarsaiya, Q. Wu, Y. Lu, J. Shi, A review of plant leaf fungal diseases and its environment speciation, Bioengineered. 10 (1) (2019) 409–424, https://doi.org/10.1080/21655979.2019.1649520.

[6] R.N. Raid, K. Pernezny, N. Havranek, J. Sanchez, B. Saddler, Weather-based forecasting systems reduce fungicide use for early blight of celery, Crop. Prot. 27 (3–5) (2008) 396–402, https://doi.org/10.1016/j.cropro.2007.07.006.

[7] M. Javaid, A. Haleem, I.H. Khan, R. Suman, Understanding the potential applications of artificial intelligence in agriculture sector, Adv. Agrochem. 2 (1) (2022) 15–30, https://doi.org/10.1016/j.aac.2022.10.001.

[8] M. Krichen, Convolutional neural networks: a survey, Computers 12 (8) (2023) 151, https://doi.org/10.3390/computers12080151.

[9] J. Lu, L. Tan, H. Jiang, Review on convolutional neural network (CNN) applied to plant leaf disease classification, Agriculture 11 (8) (2021) 707, https://doi.org/10.3390/agriculture11080707.

[10] C. Nguyen, V. Sagan, M. Maimaitiyiming, M. Maimaitijiang, S. Bhadra, M. T. Kwasniewski, Early detection of plant viral disease using hyperspectral imaging and deep learning, Sensors 21 (3) (2021) 742, https://doi.org/10.3390/s21030742.

[11] V. Singh, N. Sharma, S. Singh, A review of imaging techniques for plant disease detection, Artif. Intell. Agric. 4 (2020) 229–242, https://doi.org/10.1016/j.aiia.2020.10.002.

[12] A. J, J. Eunice, D.E. Popescu, M.K. Chowdary, J. Hemanth, Deep Learning-based leaf disease detection in crops using images for agricultural applications, Agronomy 12 (10) (2022) 2395, https://doi.org/10.3390/agronomy12102395.

[13] Ü. Atila, M. Uçar, K. Akyol, E. Uçar, Plant leaf disease classification using EfficientNet deep learning model, Ecol Inf. 61 (2021) 101182, https://doi.org/10.1016/j.ecoinf.2020.101182.

[14] L. Giannoni, F. Lange, I. Tachtsidis, Hyperspectral imaging solutions for brain tissue metabolic and hemodynamic monitoring: past, current and future developments, J. Opt. 20 (4) (2018) 044009, https://doi.org/10.1088/2040-8986/aab3a6.

[15] M.E. Rieker, M.A. Lutz, A. El-Hasan, S. Thomas, R.T. Voegele, Hyperspectral imaging and selected biological control agents for the management of fusarium head blight in spring wheat, Plants 12 (20) (2023) 3534, https://doi.org/10.3390/plants1220353.

[16] S. Thomas, M.T. Kuska, D. Bohnenkamp, A. Brugger, E. Alisaac, M. Wahabzada, J. Behmann, A. Mahlein, Benefits of hyperspectral imaging for plant disease detection and plant protection: a technical perspective, J. Plant Dis. Prot. 125 (1) (2017) 5–20, https://doi.org/10.1007/s41348-017-0124-6.

[17] C. Wang, B. Liu, L. Liu, Y. Zhu, J. Hou, P. Liu, X. Li, A review of deep learning used in the hyperspectral image analysis for agriculture, Artif. Intell. Rev. 54 (7) (2021) 5205–5253, https://doi.org/10.1007/s10462-021-10018-y.

[18] Y. Li, H. Zhang, Q. Shen, Spectral–Spatial classification of hyperspectral imagery with 3D convolutional neural network, Remote Sens. 9 (1) (2017) 67, https://doi.org/10.3390/rs9010067.

[19] S.K. Roy, G. Krishna, S.R. Dubey, B.B. Chaudhuri, HybridSN: exploring 3-D–2-D CNN feature hierarchy for Hyperspectral image classification, IEEe Geosci. Remote Sens. Lett. 17 (2) (2019) 277–281, https://doi.org/10.1109/lgrs.2019.2918719.

[20] S. Ghaderizadeh, D. Abbasi-Moghadam, A. Sharifi, N. Zhao, A. Tariq, Hyperspectral image classification using a hybrid 3D-2D convolutional neural networks, IEEe J. Sel. Top. Appl. Earth. Obs. Remote Sens. 14 (2021) 7570–7588, https://doi.org/10.1109/jstars.2021.3099118.

[21] D.C. Santana, J.D. De Queiroz Otone, F.H.R. Baio, L.P.R. Teodoro, M.E.M. Alves, Da Silva, C.A. Junior, P.E Teodoro, Machine learning in the classification of asian rust severity in soybean using hyperspectral sensor, Spectrochim. Acta Part Mol. Biomol. Spectrosc. 313 (2024) 124113, https://doi.org/10.1016/j.saa.2024.124113.

[22] R.H. Furlanetto, M.R. Nanni, M.S. Mizuno, L.G.T. Crusiol, C.R. Da Silva, Identification and classification of Asian soybean rust using leaf-based hyperspectral reflectance, Int. J. Remote Sens. 42 (11) (2021) 4177–4198, https://doi.org/10.1080/01431161.2021.1890855.

[23] J. Feng, S. Zhang, Z. Zhai, H. Yu, H. Xu, DC2Net: an Asian soybean rust detection model based on hyperspectral imaging and deep learning, Plant Phenomics. 6 (2024), https://doi.org/10.34133/plantphenomics.0163.

[24] Y. Chen, D. Wang, Two convenient methods to evaluate soybean for resistance to sclerotinia sclerotiorum, Plant Dis. 89 (12) (2005) 1268–1272, https://doi.org/10.1094/pd-89-1268.

[25] S.N. Wegulo, X.B. Yang, C.A. Martinson, Soybean cultivar responses to sclerotinia sclerotiorum in field and Controlled environment studies, Plant Dis. 82 (11) (1998) 1264–1270, https://doi.org/10.1094/pdis.1998.82.11.1264.

[26] A. Ranjan, N.M. Westrick, S. Jain, J.S. Piotrowski, M. Ranjan, R. Kessens, L. Stiegman, C.R. Grau, S.P. Conley, D.L. Smith, M. Kabbage, Resistance against sclerotinia sclerotiorum in soybean involves a reprogramming of the phenylpropanoid pathway and up-regulation of antifungal activity targeting ergosterol biosynthesis, Plant Biotechnol. J. 17 (8) (2019) 1567–1581, https://doi.org/10.1111/pbi.13082.

[27] A. Brugger, F.I. Yamati, A. Barreto, S. Paulus, P. Schramowsk, K. Kersting, U. Steiner, S. Neugart, A. Mahlein, Hyperspectral imaging in the UV range allows for differentiation of sugar beet diseases based on changes in secondary plant metabolites, Phytopathology 113 (1) (2022) 44–54, https://doi.org/10.1094/phyto-03-22-0086-r.

[28] V.T. Franceschi, K.S. Alves, S.M. Mazaro, C.V. Godoy, H.S. Duarte, E.M. Del Ponte, A new standard area diagram set for assessment of severity of soybean rust improves accuracy of estimates and optimizes resource use, Plant Pathol. 69 (3) (2020) 495–505, https://doi.org/10.1111/ppa.13148.

[29] D.A. Da Silva, C.L. Hamawaki, B.C.M. Juliatti, L.D.S. Nascimento, O.T. Hamawaki, D.L. Borges, F.C. Juliatti, A.P.O. Nogueira, An automatic phytopathometry system for chlorosis and necrosis severity evaluation of asian soybean rust infection, Comput. Electron. Agric. 192 (2021) 106542, https://doi.org/10.1016/j.compag.2021.106542.

[30] J. Li, X. Li, Y. Yan, Unlocking the potential of data augmentation in contrastive learning for hyperspectral image classification, Remote Sens. 15 (12) (2023) 3123, https://doi.org/10.3390/rs15123123.

[31] J.M. Haut, M.E. Paoletti, J. Plaza, A. Plaza, L. Plaza, Hyperspectral image classification using random occlusion data augmentation, IEEe Geosci. Remote Sens. Lett. 16 (11) (2019) 1751–1755, https://doi.org/10.1109/lgrs.2019.2909495.

[32] E.M. Dogo, O.J. Afolabi, B. Twala, On the relative impact of optimizers on convolutional neural networks with varying depth and width for image classification, Appl. Sci. 12 (23) (2022) 11976, https://doi.org/10.3390/app122311976.

[33] M. Yang, Physiological disorder diagnosis of plant leaves based on full-spectrum hyperspectral images with convolutional neural network, Horticulturae 8 (9) (2022) 854, https://doi.org/10.3390/horticulturae8090854.

[34] Y. Chang, T. Tan, W. Lee, L. Chang, Y. Chen, K. Fan, M. Alkhaleefah, Consolidated convolutional neural network for hyperspectral image classification, Remote Sens. 14 (7) (2022) 1571, https://doi.org/10.3390/rs14071571.

[35] L.A. Varga, J. Makowski, A. Zell, Measuring the ripeness of fruit with hyperspectral imaging and deep learning, in: 2021 International Joint Conference on Neural Networks (IJCNN), IEEE, 2021, pp. 1–8, https://doi.org/10.1109/IJCNN52387.2021.9533728.

[36] C. Garbin, X. Zhu, O. Marques, Dropout vs. batch normalization: an empirical study of their impact to deep learning, Multimed. Tools. Appl. 79 (19) (2020) 12777–12815, https://doi.org/10.1007/s11042-019-08453-9.

[37] C. Bircanoğlu, N. Arıca, A comparison of activation functions in artificial neural networks, in: In *2018 26th signal processing and communications applications conference (SIU)*, IEEE, 2018, pp. 1–4, https://doi.org/10.1109/SIU.2018.8404724.
[38] E.C. Seyrek, M. Uysal, A comparative analysis of various activation functions and optimizers in a convolutional neural network for hyperspectral image classification, Multimed. Tools. Appl. 83 (18) (2024) 53785–53816, https://doi.org/10.1007/s11042-023-17546-5.
[39] Hinton, G.E. (2012). Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint* arXiv:1207.0580.
[40] X. Fang, T. Zhen, Z. Li, Lightweight multiscale CNN model for wheat disease detection, Appl. Sci. 13 (9) (2023) 5801, https://doi.org/10.3390/app13095801.
[41] J. Cohen, A coefficient of agreement for nominal scales, Educ. Psychol. Meas. 20 (1) (1960) 37–46, https://doi.org/10.1177/001316446002000104.
[42] M. Sokolova, G. Lapalme, A systematic analysis of performance measures for classification tasks, Inf. Process. Manag. 45 (4) (2009) 427–437, https://doi.org/10.1016/j.ipm.2009.03.002.
[43] C.S. De Oliveira, L.M. Lião, G.B. Alcantara, Metabolic response of soybean plants to sclerotinia sclerotiorum infection, Phytochemistry 167 (2019) 112099, https://doi.org/10.1016/j.phytochem.2019.112099.
[44] A.J. Peltier, C.A. Bradley, M.I. Chilvers, D.K. Malvick, D.S. Mueller, K.A. Wise, P. D. Esker, Biology, yield loss and control of sclerotinia stem Rot of soybean, J. Integr. Pest. Manag. 3 (2) (2012) 1–7, https://doi.org/10.1603/ipm11033.
[45] R.N. Mutava, S.J.K. Prince, N.H. Syed, L. Song, B. Valliyodan, W. Chen, H. T. Nguyen, Understanding abiotic stress tolerance mechanisms in soybean: a comparative evaluation of soybean response to drought and flooding stress, Plant Physiol. Biochem. 86 (2014) 109–120, https://doi.org/10.1016/j.plaphy.2014.11.010.
[46] G.L. Hartman, E.D. West, T.K. Herman, Crops that feed the World 2. Soybean—worldwide production, use, and constraints caused by pathogens and pests, Food Secur. 3 (1) (2011) 5–17, https://doi.org/10.1007/s12571-010-0108-x.
[47] M.R. Pereira, F.N.D. Santos, F. Tavares, M. Cunha, Enhancing host-pathogen phenotyping dynamics: early detection of tomato bacterial diseases using hyperspectral point measurement and predictive modeling, Front. Plant Sci. 14 (2023), https://doi.org/10.3389/fpls.2023.1242201.
[48] C. Luo, X. Li, L. Wang, J. He, D. Li, J. Zhou, How does the data set affect cnn-based image classification performance?, in: In *2018 5th international conference on systems and informatics (ICSAI)* IEEE, 2018, pp. 361–366, https://doi.org/10.1109/ICSAI.2018.8599448.
[49] F. Chen, J.Y. Tsou, Assessing the effects of convolutional neural network architectural factors on model performance for remote sensing image classification: an in-depth investigation, Int. J. Appl. Earth. Obs. Geoinf. 112 (2022) 102865, https://doi.org/10.1016/j.jag.2022.102865.
[50] A. Agrawal, N. Mittal, Using CNN for facial expression recognition: a study of the effects of kernel size and number of filters on accuracy, Vis. Comput. 36 (2) (2019) 405–412, https://doi.org/10.1007/s00371-019-01630-9.
[51] S. Shahinfar, P. Meek, G. Falzon, How many images do I need?" understanding how sample size per class affects deep learning model performance metrics for balanced designs in autonomous wildlife monitoring, Ecol. Inf. 57 (2020) 101085, https://doi.org/10.1016/j.ecoinf.2020.101085.
[52] D. Han, Q. Liu, W. Fan, A new image classification method using CNN transfer learning and web data augmentation, Expert. Syst. Appl. 95 (2018) 43–56, https://doi.org/10.1016/j.eswa.2017.11.028.
[53] V.H. Phung, E.J. Rhee, A high-Accuracy model average ensemble of convolutional neural networks for classification of cloud image patches on small datasets, Appl. Sci. 9 (21) (2019) 4500, https://doi.org/10.3390/app9214500.
[54] L.A. Varga, M. Messmer, N. Benbarka, A. Zell, Wavelength-aware 2D convolutions for hyperspectral imaging, in: 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2023, https://doi.org/10.1109/wacv56688.2023.00378.
[55] K. Nagasubramanian, S. Jones, A.K. Singh, S. Sarkar, A. Singh, B. Ganapathysubramanian, Plant disease identification using explainable 3D deep learning on hyperspectral images, Plant Method. 15 (1) (2019), https://doi.org/10.1186/s13007-019-0479-8.
[56] D. Jung, J.D. Kim, H. Kim, T.S. Lee, H.S. Kim, S.H. Park, A hyperspectral data 3D convolutional neural network classification model for diagnosis of gray mold disease in strawberry leaves, Front. Plant Sci. 13 (2022), https://doi.org/10.3389/fpls.2022.837020.
[57] W. Li, C. Chen, M. Zhang, H. Li, Q. Du, Data augmentation for hyperspectral image classification with deep CNN, IEEe Geosci. Remote Sens. Lett. 16 (4) (2018) 593–597, https://doi.org/10.1109/lgrs.2018.2878773.
[58] M. Xu, S. Yoon, A. Fuentes, D.S. Park, A comprehensive survey of image augmentation techniques for deep learning, Pattern. Recognit. 137 (2023) 109347, https://doi.org/10.1016/j.patcog.2023.109347.
[59] M. Deepak, S. Keski-Saari, L. Fauch, L. Granlund, E. Oksanen, M. Keinänen, Leaf canopy layers affect spectral reflectance in silver birch, Remote Sens. 11 (24) (2019) 2884, https://doi.org/10.3390/rs11242884.