Fusion of the deep networks for rapid detection of branch-infected aeroponically cultivated mulberries using multimodal traits

Osama Elsherbiny[1,2], Jianmin Gao[1*], Yinan Guo[1], Mazhar H. Tunio[1], Abdallah H. Mosha[1]

(1. *School of Agricultural Engineering, Jiangsu University, Zhenjiang 212013, China*;
2. *Agricultural Engineering Department, Faculty of Agriculture, Mansoura University, Mansoura 35516, Egypt*)

Abstract: Automatic diagnosis of diseases in aeroponically cultivated branches is crucial for enhancing the efficacy of root development and overall plant survivability during propagation. Deep learning and visible imaging offer potential for precise health assessment, despite challenges in feature selection and model design, impacting diagnostic accuracy and effectiveness. The primary objective of this study is to explore a hybrid deep network that integrates multimodal data, such as texture and color attributes, as well as image color modes, to accurately detect the presence of mildew on mulberry branches. The proposed framework incorporates a Convolutional Neural Network (CNN) and Gated Recurrent Units (GRU). Various color modes were utilized, including grayscale, RGB (Red-Green-Blue), HSV (Hue-Saturation-Value), and CMYK (Cyan-Magenta-Yellow-Black). The traits based on RGB consist of nineteen vegetation color indices (VIs) and six texture variables obtained from the gray-level co-occurrence matrix (GLCM). The outcomes demonstrated that the CNN_{CMYK}-GRU_f network effectively integrates CMYK image data and color-texture features for tracking mulberry branch health during aeroponic propagation. It achieved a validation accuracy (Ac) of 99.50%, with classification precision (Pr), recall (Re), and F-measure (Fm) at the same level. Additionally, it obtained an intersection over union (IoU) of 98.90% and a loss value of 0.034. This network exhibited superior performance compared to the model that relied solely on individual image attributes, surpassing other deep networks such as Vision Transformers (Ac=94.80%), Swin Transformers (Ac=89.80%), and Multi-Layer Perceptrons (Ac=88.30%). Thus, the proposed methodology is capable of precisely assessing the health of mulberry shoots, enabling the swift deployment of intelligent aeroponic systems. Furthermore, adapting the developed model for mobile platforms could enhance its accessibility and promote sustainable, efficient agricultural practices.

Keywords: mulberry twigs health, digital imagery, VIs-GLCM, top features selection, CNN-GRU, aeroponic system
DOI: 10.25165/j.ijabe.20251802.8666

Citation: Elsherbiny O, Gao J M, Guo Y N, Tunio M H, Mosha A H. Fusion of the deep networks for rapid detection of branch-infected aeroponically cultivated mulberries using multimodal traits. Int J Agric & Biol Eng, 2025; 18(2): 75–88.

1 Introduction

Mulberry, a key species in sericulture ecology, plays a vital role in the success of sericulture, which is a major part of agriculture worldwide[1]. However, pests and diseases, responsible for up to 40% of global food production losses, pose a threat to food security[2]. Despite the positive impact of pesticides on food production since the 1950s, they have harmful effects on both human health and the environment[3,4]. Climate change and environmental stress also endanger soils and crops, disrupting soil biodiversity and ecosystem services, while traditional farming practices degrade soil health and long-term productivity[5]. To address challenges in soil-based farming, innovations like hydroponics, substrate cultivation, and aeroponics have emerged as sustainable alternatives. These methods reduce water and fertilizer use and offer solutions to issues like soil pollution from heavy metals, showing potential for rapid progress in agriculture[6-8].

Aeroponics rapid propagation technology has emerged as a solution for breeding seedlings. It delivers nutrients and supports plant propagation by distributing a nutrient solution through nozzles[9]. This technique incorporates asexual reproduction, transmitting desirable traits from the mother tree to its offspring. By reducing dependency on high-quality seeds, aeroponics enables rapid growth and development, shortening the breeding cycle[10]. The success of plant growth in rapid propagation via aeroponics is strongly influenced by humidity levels within the systems[11,12]. Humidity fluctuations can affect plant growth, impair physiological functions, and increase susceptibility to diseases like mildew[13,14], potentially disrupting water uptake and overall growth. While traditional disease detection depends on expert observation, continuous monitoring is essential, as experience-based judgments may lead to misdiagnoses. In this light, the present work develops image-based methodologies for monitoring branch infections in aeroponically cultivated mulberries. High-resolution plant images are captured via advanced digital imaging techniques, processed through algorithms to identify potential stress or disease markers[15]. These methods utilize differences in color, texture, and shape, each of which may signal specific health issues in plants[16]. Digital imaging is a cost-effective, straightforward, and rapid solution[17,18]. In contrast, direct methods like spectral analysis or chlorophyll measurement, while reliable for plant health evaluation, often require expensive, specialized equipment, destructive sampling, and are less suited for large-scale monitoring[19]. Furthermore, such

Received date: 2023-11-19 **Accepted date:** 2025-02-25
Biographies: Osama Elsherbiny, Postdoctoral fellow, research interest: precision agriculture, Email: osama1@ujs.edu.cn, osama_algazeery@mans.edu.eg; **Yinan Guo**, MSc, research interest: atomization engineering, Email: 2212116048@stmail.ujs.edu.cn; **Mazhar H. Tunio**, Assistant Professor, research interest: irrigation engineering, Email: mazharhussaintunio@sau.edu.pk; **Abdallah H. Mosha**, PhD candidate, research interest: atomization engineering, Email: 5103200333@stmail.ujs.edu.cn.
***Corresponding author: Jianmin Gao**, Professor, research interest: soil and fog tilling, Address: No 301, Xuefu Road, Zhenjiang 212013, Jiangsu, China, Tel: +86-136-5528−2069, Email: gaojianminujs@163.com, 1000001903@ujs.edu.cn.

techniques demand expertise in both data acquisition and analysis[20]. By integrating artificial intelligence (AI), these techniques can detect subtle signs of disease that might otherwise go unnoticed, enabling the early identification of plant health problems[17].

The effective and accurate identification of diseases in mulberry branches cultivated aeroponically through AI depends on the consideration of several crucial factors. These encompass the selection of an optimal color space for image representation, the application of RGB vegetation indices (VIs) for advanced color analysis, the incorporation of texture descriptors derived from the gray level co-occurrence matrix (GLCM), the implementation of data augmentation strategies, and the integration of multiple deep learning (DL) networks. These considerations are vital for the development of an effective framework for precise disease identification in mulberry branches. Elsherbiny et al.[21] compared various image color modes, including CMYK (cyan-magenta-yellow-black), HSV (hue-saturation-value), RGB (red-green-blue), and grayscale, through deep networks to diagnose water status in wheat crops with IoT-based multimodal data. The findings demonstrated that deep neural networks relying on RGB images outperformed other modes, emphasizing the importance of selecting the right color mode for accurate plant condition assessment. The success of a digital image-based framework depends on the selection of handcrafted features, such as VIs and GLCM-based texture characteristics. Research highlights a strong link between plant diseases and the biochemical and physical changes in crops, often resulting in noticeable color variations in leaves. Generally, digital images are composed of pixels that merge the RGB color channels. These channels empower the calculation of VIs[22], which serve as valuable indicators for various factors, comprising canopy moisture content[23], chlorophyll levels[24], nitrogen content[25], and disease-related stresses[26]. RGB-based VIs, characterized by high sensitivity to subtle changes, facilitate efficient monitoring and enable targeted management strategies to optimize growth and boost crop yield[27]. GLCM, a variant implemented with RGB images, is applied to assess plant health. Mathew et al.[28] utilized GLCM for leaf symptom analysis and developed a composite voting model combining decision trees, support vector machines (SVM), and k-nearest neighbors (k-NN) for early disease detection. Bhimte and Thool[29] created an automated system for diagnosing cotton leaf diseases by extracting color and texture features and applying SVM for classification.

To address limited dataset sizes, data augmentation is vital. Geometric transformations (resizing, cropping, rotation, horizontal flipping) and intensity variations (contrast, brightness, color, noise) are employed to boost the dataset's diversity and size, ultimately strengthening the model's robustness and performance[30]. In addition, the integration of multiple deep networks, each meticulously trained, has been instrumental in boosting both the excellence and resilience of the expectation model. This methodology, previously deployed to assess plant characteristics, enhances model performance and precision. Jin et al.[31] indicated that a hybrid deep neural network combining 2D-Convolutional Neural Networks (CNNs) and bidirectional Gated Recurrent Units (GRUs) attained 0.74 accuracy in categorizing healthy and Fusarium-infected wheat heads. Ullah et al.[32] introduced a mixture network with EfficientNetB3 and MobileNet for tomato disease detection, showcasing strong feature extraction and reliable automated detection.

Image classification and DL approaches provide solutions to reduce mildew impact on mulberry branches and minimize pesticide and fertilizer use. By enabling accurate and timely health assessments, the implemented system has the potential to improve the survivability and yield of aeroponically grown mulberry shoots. Hence, the primary objectives of this research were: 1) to develop a novel hybrid deep network that leverages multimodal data - specifically, image data and characteristics derived from VIs and GLCM - aiming to advance mulberry branch health management in aeroponic cultivation; 2) to evaluate a range of deep networks, such as CNN, GRU, and hybrid combinations of both, while deploying different image color modes; 3) to identify top-level features of VIs and GLCM to optimize model behavior; and 4) to delineate the superior components of deep network architecture and enhance model interpretability by highlighting regions of interest for robust detection of health states in mulberry branches.

2 Materials and methods

2.1 Efficient aeroponic growth management system

2.1.1 System design overview

At Jiangsu University in Zhenjiang city, Jiangsu province, China, an aeroponics experiment was undertaken to study the rapid growth of mulberry branches. Figure 1 depicts the configuration of the regulated system and its hardware constituents. The system regulating humidity and temperature was composed of six components: one module for gathering data, another for analyzing data, an interface for human-machine interaction, a storage component, a mechanism for execution, and a unit for data relay. The system consisted of a central computer and a sensor for measuring humidity and temperature. Every half second, the sensor transmitted data to the computer, which then relayed it via a serial link to the Synchronous Transport Mode-32 (STM32) device for processing. If the mold growth rate stayed within acceptable limits, the computer forwarded the data to the uCOS-III message queue for subsequent calculations. The task associated with the Emergency Managers Weather Information Network (EMWIN) used message queue data to create live humidity and temperature visuals. When mildew levels exceeded a set threshold, the STM32 device sent commands to adjust the atomizing sheet's spray frequency and control the liquid pump. A touch-responsive TFT-LCD (thin-film transistor liquid crystal display) on the human-machine interface, powered by the uCOS-III platform, swiftly detected inputs and executed commands. Among the components of the data storage module were the SD card, STM32 FLASH, and the FatFs file system, which seamlessly collaborated to facilitate efficient data storage and retrieval. When activated, the system recorded humidity and temperature readings in a file named "dataTH.txt," with the SD card automatically cleared each time the system was powered on.

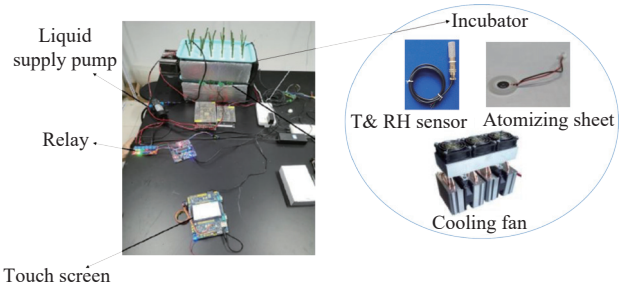


Figure 1 Schematic depicting the composition of the controlled system and its hardware components

In the FLASH memory of the STM32, there resided an array of crucial data: the repository of system text, thresholds for

temperature and humidity, rates defining mildew, intervals for spraying, and duration of each spraying event. Within the actuator module, an assemblage of elements was present, comprising a cooling fan, liquid supply pump, power supply, fogging tablet, lower computer relay, LED light board, and an alarm mechanism. The serial port facilitated the transfer of actuator control data from the STM32 to the STC89C52 microcontroller within this configuration. Upon receiving the data packet, the STC89C52 microcontroller decoded the contents to discern the necessary actions, which were subsequently executed in accordance with the provided specifications. Wireless communication was achieved by employing the ZigBee serial transmission module, thereby upgrading the conventional serial communication method.

2.1.2 System management operation

The operational mechanism was designed to operate in two modes: user-controlled and automated. In a controlled greenhouse, no heater was required due to natural temperature regulation. The system monitors aeroponic growth, comparing real-time humidity and temperature data to set reference points. When mold growth remains within the expected range, it automatically adjusts the cooler and atomizing sheet to maintain optimal conditions. During the automated regulation period, the LED fill light remains on. If the temperature exceeds the upper limit, the cooling fan activates to cool the device and shuts off once the temperature returns to the set range. A drop in temperature below the minimum triggers an alert for greenhouse staff. The fogging system activates when humidity falls below the threshold, operating until the maximum is reached, then shuts off. Rising mildew levels prompt the system to adjust cooling, misting, and water pumping to maintain optimal conditions.

The operational setup empowers the handler with precise control over each actuator, all of which begin in an off state. These actuators include the ventilation fan, fluid pump, misting panel, and fill light. Users can manually activate the pump to refill the nutrient reservoir based on mildew levels. The ventilation fan allows for manual temperature regulation, while the light's brightness can be adjusted. For the misting panel, users can set the spray duration and break intervals, with the system operating for 2 minutes followed by a 1-minute pause.

2.2 Data collection

The experiment was carried out on healthy mulberry cuttings, aged 1-2 years, characterized by strong metabolism, high energy, and rapid root development. Segments, 15-20 cm in length and containing 2-4 buds, were taken from the central and lower portions of the cuttings. These were disinfected in a potassium permanganate solution for 30 min, rinsed, and dried. The cuttings were then soaked in a rooting powder solution for 2-3 h, ensuring the base was immersed approximately 3 cm. Before placing the cuttings in a multi-layered incubator, any additional moisture was taken off. The incubator was designed with four distinct sections: a layer specifically for nutrient delivery, an area reserved for the plants, a segment with a misting sheet, and an illumination-focused light layer. In a span of 10 to 15 days, the cuttings displayed prominent callus growth and developed new roots. The incubator housed 24 mulberry branches (Figure 2), from which 20 samples were chosen due to their appropriateness for the study. Of these, 10 were healthy, while the other 10 displayed signs of mildew infection. Images of these branches were systematically taken every two days for a total of 40 days, leading to the creation of 21 separate image sets, summing up to 420 images. Captured under various lighting conditions, these images mainly concentrated on the detection and analysis of mildew, the disease being studied. The RGB images, oriented perpendicularly, were captured and saved in JPEG format, with a spatial resolution of 1080×1920 pixels, and were specifically focused on branches at a height of 20 cm.

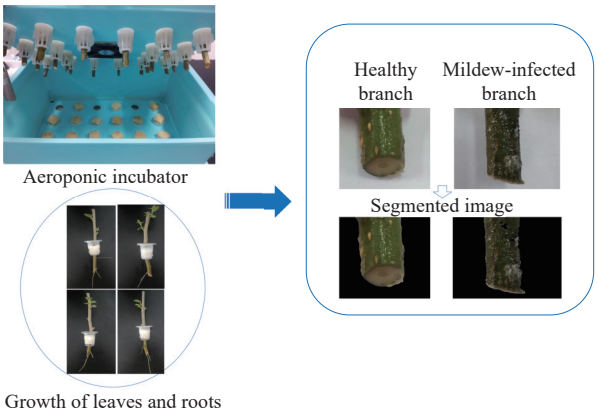


Figure 2 Healthy and mildew-infected branches in aeroponic incubator

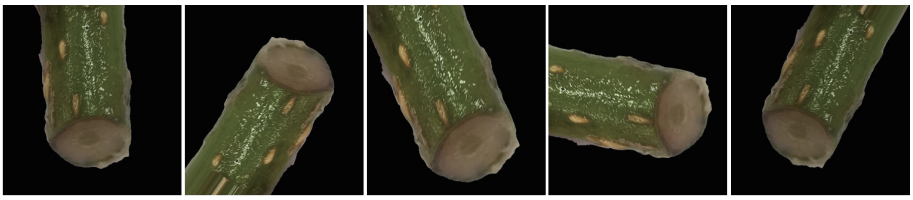
2.3 Color image mode

Color features can be derived by converting images into different color spaces, such as CMYK, RGB, HSV, and grayscale[33]. These transformed color representations yield valuable insights for analysis and application. The CMYK relies on the combination of four ink colors - cyan, magenta, yellow, and black - and is extensively utilized in the printing industry to reproduce full-color images. The RGB blends red, green, and blue light in various proportions, enabling the creation of a wide spectrum of colors. The HSV serves as a framework for color representation within the cylindrical coordinate system. It represents colors through three components: hue (type of color), saturation (vibrancy), and value (brightness). Grayscale images are composed of shades of gray, varying from black to white. Such color features underwent thorough scrutiny to identify the optimal input for the models.

2.4 Image preprocessing

This process consists of three stages: 1) segmenting the image to eliminate the background, 2) augmenting the data to increase the size of the training dataset, and 3) transforming the features by rescaling them through normalization. To precisely assess the branch's health, it must be isolated from the background through thresholding, then converted to grayscale and segmented into a binary image[34]. There are two potential pixel values: pixels with intensity values greater than or equal to the threshold value of 1 represent the foreground (white), while pixels with intensity values lower than or equal to the threshold value of 0 signify the background (black) and can be discarded. Afterward, data augmentation (Figure 3) enhanced learning by expanding the dataset to 4200 images and improving object differentiation under varied real-world conditions. Potential color VIs and GLCM features were also extracted for further analysis. Finally, normalization was applied to individual attributes (A) to account for variations in magnitude among different features. To standardize the characteristics (A_n), a computation was conducted by deducting the lowest image data point ($A_{\min}$) and subsequently dividing it by the range between the highest ($A_{\max}$) and lowest feature values, as delineated in the following equation:

$$A_n = \frac{A - A_{\min}}{A_{\max} - A_{\min}} \quad (1)$$



Note: The numbers in parentheses represent the parameters for each transformation, indicating the degree or intensity of the modification applied to the image.

Figure 3 Augmented RGB images of healthy branches

2.5 Feature extraction

2.5.1 Color RGB vegetation indices

VIs were integrated into RGB-based multimodal data to detect color shifts indicative of mulberry branch health, capturing variations in pigmentation, chlorophyll levels, and physiological changes related to infections like mildew. The relevant indices (Table 1) are capable of revealing subtle differences not discernible in raw images. By tracking shifts in plant coloration, VIs offer a quantitative gauge of mildew infection severity, which likely correlates with its progression, providing valuable insights for timely and effective disease management[35]. As outlined in Table 1, nineteen VIs were designated for their proven efficacy in characterizing plant health through RGB color spaces. These indices primarily focus on subtle variations in the red (r), green (g), and blue (b) bands, which are highly sensitive to changes in plant pigmentation and stress responses. To ensure accurate disease diagnosis, RGB color space percentages were extracted as sample features. The mean values of the RGB channels (r, g, b) were computed for each sample, forming the basis for deriving the 19 VIs. The derived indices were then utilized as input features for various DL models. The interaction between RGB color data and multimodal features was instrumental, as the VIs effectively correlated color variations with the plants' physiological condition. This integration markedly enhanced the model's capacity to predict infection severity with high precision. As highlighted by Zhu et al.[36], Peng et al.[37], and Elsherbiny et al.[17], the fusion of multimodal data has the potential to significantly enhance predictive accuracy in precision agriculture.

2.5.2 GLCM-based texture features

GLCM, known as grayscale correlation-based texture analysis, examines the connection between two pixels in an image at a defined distance. It derives texture features of mulberry branches by employing probabilistic attributes. In this work, six variations of GLCM were utilized, including dissimilarity (Di), contrast (Co), homogeneity (Ho), energy (En), correlation (Cor), and angular second moment (ASM). Di measures pixel pair distance, Co reflects image depth and texture, Ho indicates pixel value similarity, En assesses texture uniformity, Cor evaluates pixel association, and ASM describes image roughness and texture distribution. Hall-Beyer[51] outlined an explanation of the RGB-based GLCM variables. The probabilities associated with specific pixel value combinations, denoted as $P_{i,j}$, are calculated using row index i and column index j. The mean values for indices i and j are denoted as μ_i and μ_j, respectively. The standard deviations for values associated with the i and j indices are given as σ_i and σ_j, respectively. The variable N represents the total number of rows or columns.

$$\mathrm{Di} = \sum_{i,j=0}^{levels-1} P_{i,j} \cdot |i-j| \tag{2}$$

$$\mathrm{Co} = \sum_{i,j=0}^{levels-1} P_{i,j} \cdot (i-j)^2 \tag{3}$$

$$\mathrm{Ho} = \sum_{i,j=0}^{levels-1} \frac{P_{i,j}}{1+(i-j)^2} \tag{4}$$

Table 1 Explanation of the VIs derived from RGB images

Category of indices	Equation	Citations
Normalized blue (bn)	$b/(r+g+b)$	[38]
Normalized green (gn)	$g/(r+g+b)$	[38]
Normalized red (rn)	$r/(r+g+b)$	[38]
Red-blue ratio (RBRI)	r/b	[39]
Green-blue ratio (GBRI)	g/b	[40]
Green-red ratio (GRRI)	g/r	[41]
Woebbecke (WI)	$(g-b)/(r-g)$	[42]
Kawashima (IKAW)	$(r-b)/(r+b)$	[43]
Normalized difference (NDI)	$(rn-gn)/(rn+gn+0.01)$	[44]
Green-red vegetation (GRVI)	$(g-r)/(g+r)$	[43]
Excess blue vegetation (ExB)	$1.4\times bn-gn$	[45]
Excess green vegetation (ExG)	$2\times gn-rn-bn$	[45]
Excess red vegetation (ExR)	$1.4\times rn-gn$	[45]
Visible atmospheric resistance (VARI)	$(g-r)/(g+r-b)$	[46]
Vegetative (VEG)	$g/(ra\times b(1-a)), a=0.667$	[47]
Principal component analysis (IPCA)	$0.994\times\lvert r-b\rvert+0.961\times\lvert g-b\rvert+0.914\times\lvert g-r\rvert$	[48]
Color index of vegetation (CIVE)	$0.441\times r-0.881\times g+0.385\times b+18.78745$	[49]
Excess green minus excess red (ExGR)	$ExG-ExR$	[45]
Combination (COM)	$0.25\times ExG+0.3\times ExGR+0.33\times CIVE+0.12\times VEG$	[50]

$$\text{ASM} = \sum_{i,j=0}^{levels-1} P_{i,j}^2 \tag{5}$$

$$\text{En} = \sqrt{\text{ASM}} \tag{6}$$

$$\text{Cor} = \sum_{i,j=0}^{levels-1} P_{i,j} \left[\frac{(i-\mu_i)\cdot(j-\mu_i)}{\sigma_i \cdot \sigma_j} \right] \tag{7}$$

2.6 Superior feature selection based on SelectKBest with the Decision Tree Classifier

Feature selection streamlines the dataset by preserving essential features, addressing challenges such as computational burden and overfitting arising from the curse of dimensionality[52]. This research utilizes the SelectKBest approach with ANOVA F-value scores to prioritize features according to their relevance in forecasting the target variable[53]. The introduced approach combines feature selection with hyperparameter tuning for the Decision Tree Classifier (DT). By doing so, the model's parameters are fine-tuned for the selected feature subset, producing a model that is both efficient and accurate. The process of constructing decision trees from training datasets is termed as decision tree induction. These trees, resembling flowcharts, comprise root, decision, and leaf nodes: the root initiates the tree, decision nodes guide the path based on data attributes, and leaf nodes signify the final outcomes. During training, parameters like maximum depth (*Md*) were considered with values (2, 4, 6, 8, 10), minimum samples leaf (*Ms*) with (1, 2, 3, 4, 5), and maximum leaf nodes (*MLn*) with (2, 4, 6, 8, 10). The DT model exploited multimodal inputs, consisting of 19 VIs and 6 GLCMs, for enhanced feature selection. It assessed various quantities of selected features (*k*) through a pipeline integrating SelectKBest with a DT model. Within this setup, GridSearchCV identified the most fitting hyperparameters for the classifier, emphasizing the top *k* features based on accuracy. Figure 4 illustrates the fundamental steps involved in training the DT model, fine-tuning hyperparameters, and structuring features.

A feature selection approach using SelectKBest and DT

Import necessary libraries
Load data:
- Read the Excel file.
- Proposed features (*X*) and target labels (*Y*).

Initialize a dictionary to store results:
- *k* (number of features selected).
- Parameters of the decision tree.
- Avg. and Std. of training and test scores.
- Selected feature names and their percentage scores.

Define a parameter grid for the decision tree:
- Max_depth.
- Min_samples_leaf.
- Max_leaf_nodes.

Loop through each possible k value:
- **Create a pipeline consisting of:**
  - Feature selection using SelectKBest.
  - Decision tree classifier.
- **Perform grid search with 5-fold cross-validation:**
  - Fit the pipeline to the data.
  - evaluate the performance.
- **Extract results from GridSearchCV:**
  - Best parameters for the decision tree.
  - Avg. and Std. of training and test scores.
  - Selected feature names and their scores.
  - Calculate feature scores as a percentage.

Find the iteration with the highest mean test score :
- Identify the index of the iteration.
- **Extract and print details such as:**
  - Best *k*.
  - Hyperparameters of the best decision tree.
  - Mean training and test scores.
  - Selected feature names for this iteration.

End

Figure 4 Pseudo-code for training a decision tree with SelectKBest for selecting top variables

2.7 Proposed deep networks

2.7.1 Convolutional Neural Network (CNN)

The CNN holds the predominant position as a deep learning architecture for processing input data with two spatial dimensions. It follows a structure denoted by $h_e \times w_i \times C_n$, where h_e represents the height, w_i denotes the width, and C_n signifies the number of channels within the image. This network was fed with images in various color spaces (grayscale, CMYK, RGB, HSV). These images had corresponding values for the number of channels, with C_n being equal to 1, 4, 3, and 3 bands. As stated by Kamilaris and Prenafeta-Boldú[54], the CNN is widely recognized as the leading model in computer vision applications and has gained increasing significance in the field of agriculture. CNN models generally consist of three distinct layers in their architectural composition. The input layer serves as the primary data source, receiving multimodal inputs such as image data and 8 significant features chosen by the DT model. The hidden layers include components such as convolution, batch normalization, ReLU activation, max-pooling, and flattening. The output layer, through full connectivity, transforms the inputs from preceding layers into two separate categories: healthy branches and mildew-infected branches.

The CNN structure is depicted in Figure 5, whereas Table 2 outlines the key features of each layer within the CNN. Convolutional layers extract features from image slices using filters, enabling multiple transformations. These filters are customized to the input slices and undergo nonlinear processing to learn abstract features and introduce nonlinearity in the feature space[55]. Nonlinear processing generates different activation patterns, aiding in learning semantic differences across the dataset. To address covariate shift and normalize gradients during training, batch normalization layers are incorporated[56]. Pooling mechanisms decrease output complexity via max-pooling, capturing utmost essential attributes. The pool sizes chosen for optimal VI-GLCM features are (1,2), while for images, the selected pool size is (2,2). Training the CNN involves utilizing an Adam optimizer, a momentum value of 0.9000, a learning rate of 0.001, 350 epochs, and a batch size of 70. The output of the pooling layer is flattened into a 1D vector since subsequent dense layers only accept 1D vectors. The ultimate densely connected stages function as categorizers and utilize the softmax formula. These layers contain a limited number of neurons that receive input from a vector $a^{[i-1]}$ and produce a vector output$a^{[i]}$, specifically for the *j*th node of the *i*th layer. The weights $w_{j,l}$, with $n_{l-1} \times n_l$ parameters, represent the learned factors in the *l*th layer. Moreover, the bias of the layer is denoted as $b_j^{[i]}$, and the activation function is represented by $\psi^{[i]}$.

$$z_j^{[i]} = \sum_{l=1}^{n_{i-1}} w_{j,l}^{[i]} \cdot a_l^{[i-1]} + b_j^{[i]} \rightarrow a_j^{[i]} = \psi^{[i]} \cdot \left(z_j^{[i]}\right) \tag{8}$$

2.7.2 Gated Recurrent Units (GRU)

GRU is a specific type of Recurrent Neural Network (RNN) architecture designed to overcome certain limitations of traditional RNNs. It plays a crucial role in effectively handling time-series and nonlinear data[57,58]. The architecture of this network comprises an input layer denoted as *x*, a hidden layer represented as *h*, and an output layer labeled as *y*. The output layer and hidden layer can be computed via the following formula:

$$Y_t = g \cdot (S_t \cdot W_{hy}) \tag{9}$$

$$S_t = f \cdot (X_t \cdot W_{SX} + S_{t-1} \cdot W_{SS}) \tag{10}$$

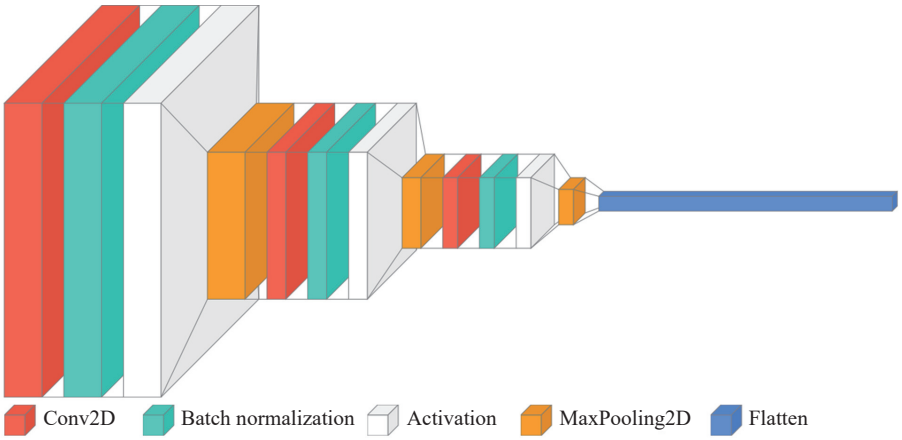


Figure 5 Comprehensive architecture of the CNN model

Table 2 Advanced layered design for the proposed CNN

Layer	Type	Input size	
		Image	Superior features
0	Input data	(50×50×C_n)	(1×8×1)
1	Conv2D	(50×50×C_n)	(1×8×1)
2	Batch Normalization	(50×50×512)	(1×8×512)
3	ReLU	(50×50×512)	(1×8×512)
4	MaxPooling2D	(50×50×512)	(1×8×512)
5	Conv2D	(25×25×512)	(1×8×512)
6	Batch Normalization	(25×25×256)	(1×8×256)
7	ReLU	(25×25×256)	(1×8×256)
8	MaxPooling2D	(25×25×256)	(1×8×256)
9	Conv2D	(12×12×256)	(1×8×256)
10	Batch Normalization	(12×12×128)	(1×8×128)
11	ReLU	(12×12×128)	(1×8×128)
12	MaxPooling2D	(12×12×128)	(1×8×128)
13	Flatten	(6×6×128)	(1×8×128)

In this work, the GRU model was evaluated for branch health detection, utilizing image data and first-level characteristics. The model's architecture is detailed in Table 3. To enhance prediction accuracy, it is crucial to incorporate various multimodal data, including images and 8 VIs-GLCM features extracted at different time intervals within the same quadrat. The model's process, depicted in Figure 6, involves the deployment of the previous output (h_{t-1}) and the current input (x_t). The reset gate determines the portion of information to be reset. The update gate is responsible for updating the GRU output. The candidate hidden layer is processed to yield the current output. Throughout the training process, the GRU gates (Z_t and r_t) and parameters (W_z, W_r, and W) are updated.

$$Z_t = \sigma \cdot (Wz \cdot [h_{t-1}, X_t]) \quad (11)$$

$$r_t = \sigma \cdot (Wr \cdot [h_{t-1}, X_t]) \quad (12)$$

$$h_i' = \tanh \cdot (W \cdot [r_t \times h_{t-1}, X_t]) \quad (13)$$

2.7.3 Advanced multimodal data-driven classification model

This study aimed to explore the top-selected features based on SelectKBest and DT model. Moreover, it sought to investigate the influence of multimodal variables on the performance of CNN and GRU models. The most successful hybrid model, depicted in Figure 7, was composed of two models: CNN_{CMYK} and GRU_f, which operate different multimodal traits. These properties, mined from various origins, include $F1$ (CNN_{CMYK}-related features derived from images) and $F2$ (GRU_f-based multimodal characteristics obtained from VI-GLCM). By integrating these diverse attributes, the hybrid model demonstrated robustness and achieved high accuracy in the classification of healthy conditions of mulberry branches.

Table 3 The deep network architecture proposed for the GRU model

Layer	Type	Input size	
		Image	Superior features
0	Input data	(2500×C_n)	(8×1)
1	GRU	(2500×C_n)	(8×1)
2	Dropout	(2500×64)	(8×64)
3	GRU	(2500×64)	(8×64)
4	Dropout	(2500×128)	(8×128)
5	GRU	(2500×128)	(8×128)

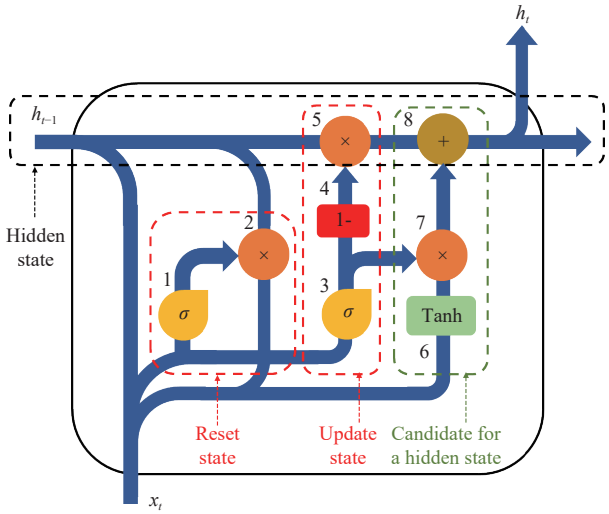


Figure 6 Configuration of GRU architecture

2.8 Processing of data and software tools

The dataset consists of 4200 images of mulberry branches, divided into two groups: 2100 images of healthy branches and 2100 of those infected with mildew. For model training, 80% of the data (3360 images) was applied, and the remaining 20% (840 images) served for evaluating the model's performance. The process of creating the model and analyzing the data took place on the Kaggle platform, which offers complimentary utilization of NVidia K80 GPUs within kernels. To accelerate the training of the proposed deep learning model, the GPU available on the Kaggle platform was utilized, resulting in a significant 12.5X speedup in kernel performance. The data analysis was performed on a PC with an

Intel Core i7-3630QM CPU running at 2.4 GHz and 8 GB of RAM. For the classification task, TensorFlow library version 2.4.1 was employed, and both CNN and GRU modules were utilized. The implementation was conducted entirely with Python version 3.7.10.

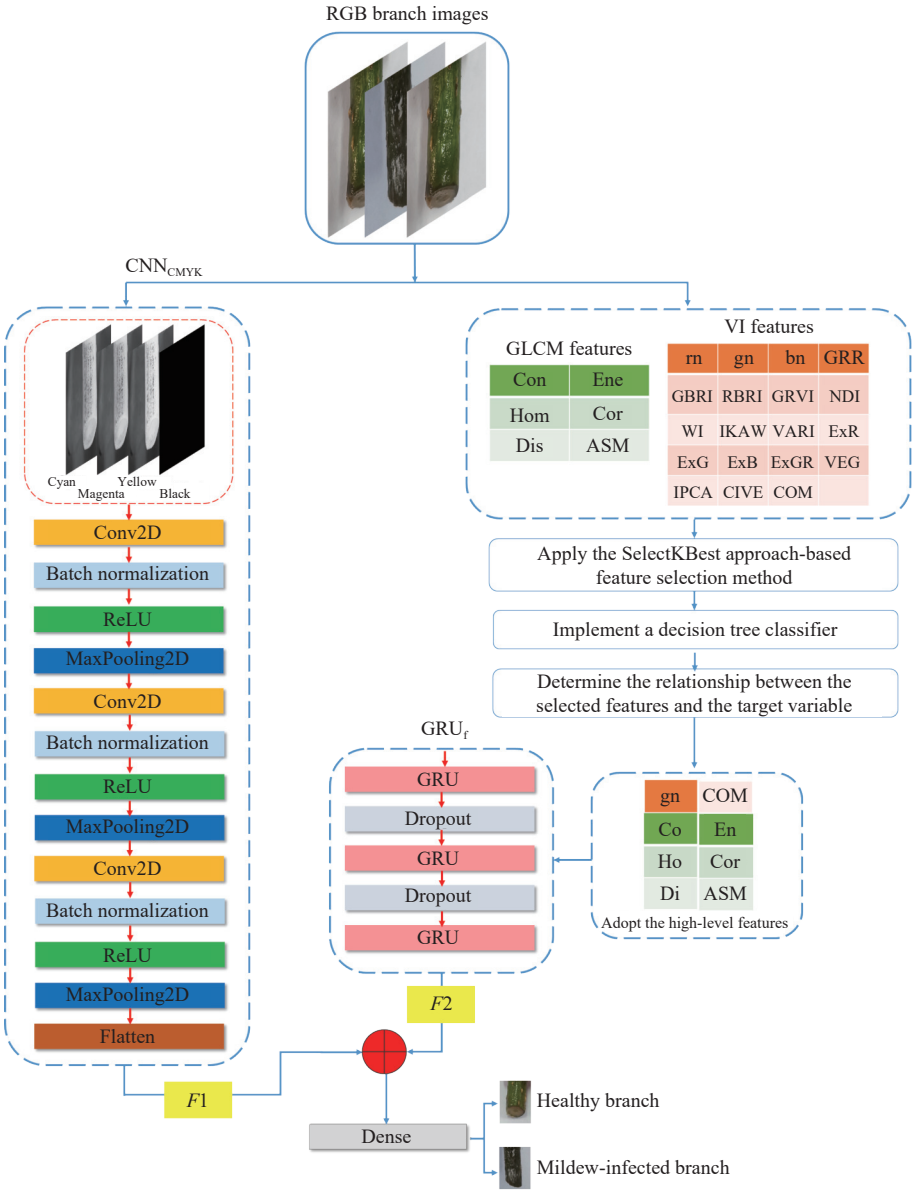


Figure 7 Configuring an optimal deep network architecture through multimodal data for detecting the health state of mulberry branches

2.9 Evaluation of detection performance

The performance evaluation of the proposed deep networks comprised a quantitative assessment based on measures like recall (Re), precision (Pr), intersection over union (IoU), overall accuracy (Ac), and F-measure (Fm). These measures were computed relied on the number of true-positives (TP), true-negatives (TN), false-positives (FP), and false-negatives (FN), thus providing a comprehensive evaluation of the networks' performance. The calculations for these measures are as follows:

$$\mathrm{Re}=\frac{\sum \mathrm{TP}}{\sum \mathrm{FP}+\sum \mathrm{TN}}\times 100\% \tag{14}$$

$$\mathrm{Pr}=\frac{\sum \mathrm{TP}}{\sum \mathrm{FP}+\sum \mathrm{TP}}\times 100\% \tag{15}$$

$$\mathrm{IoU}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}+\mathrm{FN}}\times 100\% \tag{16}$$

$$\mathrm{Ac}=\frac{\sum \mathrm{TP}+\sum \mathrm{TN}}{\sum \mathrm{FP}+\sum \mathrm{FN}+\sum \mathrm{TP}+\sum \mathrm{TN}}\times 100\% \tag{17}$$

$$\mathrm{Fm}=2\times\frac{\mathrm{Re}\times \mathrm{Pr}}{\mathrm{Re}+\mathrm{Pr}}\times 100\% \tag{18}$$

3 Results and discussion

3.1 Automated feature selection and hyper-parameter tuning through DT Model

As illustrated in Figure 8, the most relevant VIs and GLCM characteristics identified in each DT training iteration through SelectKBest are presented. The figure underscores the significance of various features across cycles, particularly RGB-derived VIs and GLCM-based textures. The VIs of highest importance are gn and

COM, exhibiting importances of 49.99% and 50.01%, respectively. In contrast, the fine-grained GLCM traits - Co, En, Ho, Cor, Di, and ASM - have importances of 1.09%, 61.15%, 0.57%, 37.78%, 1.02%, and 63.11%, respectively. Table 4 details the performance metrics for the DT model, covering the mean accuracy score from different cross-validation folds along with the standard deviation (SD) of these scores. The tuned hyperparameters for the DT model, specifically Md, Ms, and MLn, were ascertained. The DT model exhibited the highest efficacy when deploying traits from two VIs in the 18th iteration and six GLCMs in the first. The incorporation of two VIs enhanced the model's performance to a score of 0.790, surpassing the 0.770 achieved with 19 VIs. Additionally, the integration of six GLCM features significantly augmented the model's discriminative power, yielding a score of 0.667. This outstripped the performance of the feature sets labeled 1 through 5, with respective scores of 0.536, 0.594, 0.595, 0.614, and 0.612. These findings are consistent with Gaagai et al.[59], who emphasized the influence of two crucial factors on the behavior of machine learning (ML) models: the training-based filtering of semantic features and the selection of upper hyperparameters.

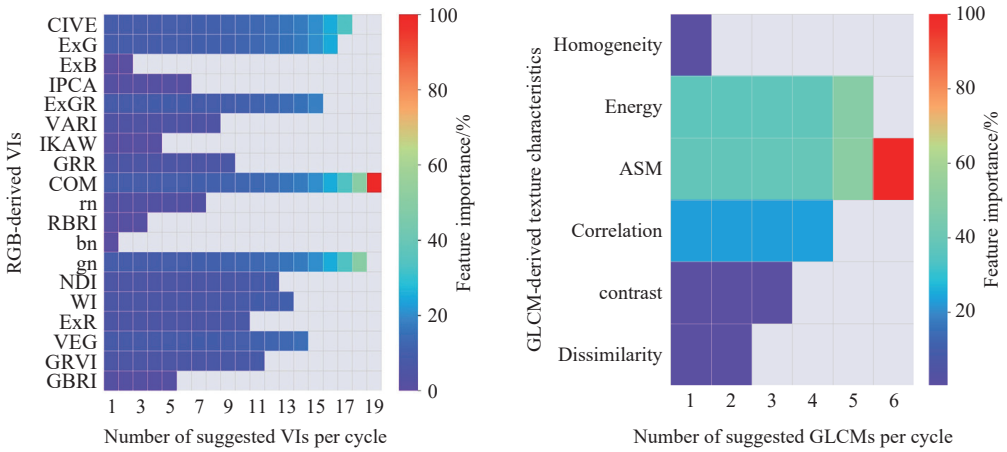


Figure 8 Assessment of feature importance for GLCM-derived texture and RGB color across various selection sizes

Table 4 Comparative analysis of DT performance with high-level RGB-based traits and GLCM variables under optimal hyperparameter configurations

RGB-based traits	Hyperparameters (Md, Ms, MLn)	High-level features	Train		Test	
			Score	SD	Score	SD
VI	(2, 1, 2)	gn, COM	0.805	0.025	0.790	0.091
GLCM	(8, 1, 10)	Co, En, Ho, Cor, Di, ASM	0.702	0.021	0.667	0.069

3.2 Integrated CNN models with multimodal characteristics

The CNN model exploited CMYK, RGB, HSV, and grayscale color space images as input data for analysis. Furthermore, this study appraised the effectiveness of the methodologies by investigating multimodal attributes, merging VIs with GLCM parameters while training the CNN model on image data. Table 5 summarizes a comparison of performance measures (Re, Pr, Fm, and IoU) to assess CNN models' efficacy in tracking branch health, accounting for accuracy and loss in training and validation. Incorporating the CMYK color space alongside multimodal features led to a significant boost in accuracy. The hybrid CNN_{CMYK}-CNN_f model unveiled superior performance (Re=0.979, Pr=0.979, Fm=0.979, IoU=0.968) compared to the standalone CNN_{CMYK} model (Re=0.969, Pr=0.969, Fm=0.969, IoU=0.937), showcasing exceptional classification capabilities. Moreover, in terms of validation scores, the model accomplished an accuracy of 0.979 and a loss of 0.051, outstripping the CNN_{CMYK} model (Ac=0.969, Ls=0.060). These outcomes align with Barbedo[60], who emphasized the potential of CNNs to extract high-level characteristics during training, making them effective tools for addressing plant pathology concerns. This highlights the significant role and effectiveness of CNNs in modern agriculture and the management of plant diseases.

Table 5 Outputs of CNN models utilizing various color spaces and multimodal features

Model	Image	f	Training			Validation			Performance			
			Ac	Ls	Tt	Ac	Ls	Tv	Re	Pr	Fm	IoU
CNN_{img}	Gray	-	1.0	8.365×10^{-5}	6.883	0.921	0.219	0.351	0.921	0.921	0.921	0.858
	RGB	No	1.0	2.944×10^{-6}	7.474	0.966	0.086	0.325	0.966	0.966	0.966	0.940
	HSV	-	1.0	3.346×10^{-5}	5.719	0.938	0.208	0.261	0.938	0.938	0.938	0.885
	CMYK	-	1.0	8.927×10^{-5}	8.385	0.969	0.060	0.321	0.969	0.969	0.969	0.937
CNN_{img}-CNN_f	Grey	-	1.0	1.212×10^{-4}	7.388	0.928	0.210	0.339	0.928	0.928	0.928	0.867
	RGB	Yes	1.0	3.359×10^{-6}	7.150	0.973	0.059	0.390	0.973	0.973	0.973	0.951
	HSV	-	1.0	5.798×10^{-5}	7.239	0.957	0.131	0.407	0.957	0.957	0.957	0.918
	CMYK	-	1.0	0.003	9.784	0.979	0.051	0.433	0.979	0.979	0.979	0.968

Note: Where Ls signifies the model loss, f refers to high-level multimodal characteristics of VIs and GLCM, Tt stands for the total training time (min), and Tv denotes the duration required to conduct a single-sample test (sec).

3.3 Fusion of GRU models with multimodal features

The diverse features, comprising VIs and GLCM, were simultaneously appraised during image training, considering them as multimodal traits, and were fed into the GRU models. Table 6 displays the marks related to performance metrics such as Re, Pr, Fm, and IoU. It also delineates the training and validation outcomes through metrics like Ls and Ac. The HSV-based GRU model outperformed other color image modes. Trained solely on image

data, it achieved the following performance metrics: Re (0.561), Pr (0.321), Fm (0.412), IoU (0.400), and Ac (0.561) with a loss of 0.667. Meanwhile, the combined GRU_{HSV}-GRU_f model slightly outperformed the HSV image-based model in identifying healthy mulberry stems. Its performance scores encompassed a Re of 0.653, a Pr of 0.429, an Fm of 0.501, and an IoU score of 0.494. During validation, it attained an accuracy of 0.653 with a loss of 0.582. These results correspond with those of Elsherbiny et al.[21], who explained the importance of conducting a thorough investigation into deep network structures and feature selection. Such an investigation is essential, as it greatly enhances overall performance and produces better outcomes.

Table 6 Outputs of GRU models deploying distinct color spaces and multimodal properties

Model	Image	f	Training			Validation			Performance			
			Ac	Ls	Tt	Ac	Ls	Tv	Re	Pr	Fm	IoU
GRU_{img}	Gray	-	0.572	0.671	32.561	0.559	0.681	1.770	0.559	0.311	0.399	0.391
	RGB	No	0.559	0.680	32.580	0.549	0.680	1.783	0.549	0.291	0.389	0.377
	HSV	-	0.573	0.668	31.571	0.561	0.667	1.759	0.561	0.321	0.412	0.400
	CMYK	-	0.560	0.683	32.891	0.530	0.687	1.189	0.530	0.288	0.379	0.371
GRU_{img}-GRU_f	Gray	-	0.635	0.609	32.855	0.620	0.621	1.690	0.620	0.395	0.471	0.462
	RGB	Yes	0.637	0.601	32.134	0.616	0.639	1.891	0.616	0.377	0.459	0.446
	HSV	-	0.668	0.571	30.966	0.653	0.582	1.844	0.653	0.429	0.501	0.494
	CMYK	-	0.659	0.595	32.591	0.609	0.625	1.490	0.609	0.365	0.433	0.426

3.4 Hybrid deep network with multimodal characteristics

Through the fusion of CNN and GRU models, this study successfully attained enhanced accuracy levels. The training procedure incorporated diverse space images and RGB-based multimodal features, including combinations of higher VIs and GLCM traits. The results in terms of the performance metrics of the hybrid deep network, including Re, Pr, Fm, and IoU, as well as the validation outputs of Ac and Ls, are presented in Table 7. The hybrid deep network, incorporating both CNN and GRU models with CMYK, exhibited outstanding performance compared to other feature combinations. The integration of the CNN model alongside CMYK color space processing and the GRU model utilizing advanced VIs-GLCM features demonstrated outstanding performance across multiple metrics. It achieved impressive scores in terms of Ac, Re, Pr, and Fm, all registering at an exceptional value of 0.995. Additionally, the model attained a notably high IoU score of 0.989, while maintaining a minimal loss of 0.034. These results underscore the model's accuracy in making predictions and its robust capability in localizing branch-infected mulberries. Due to the well-designed framework[61], the first-level model, known as the CNN_{CMYK}-GRU_f model, surpassed those models that rely solely on training images, like CNN_{img}-GRU_{img} (Ac=0.984, Re=0.984, Pr=0.984, Fm=0.984, IoU=0.977, and Ls=0.045). These findings highlight the importance of the features integrated into this combined network, demonstrating their capacity to appropriately estimate the health of the mulberry branch. These outcomes are similar to the findings of Jin et al.[31], who combined CNN and GRU models to classify healthy and diseased wheat heads based on hyperspectral data, achieving a satisfactory accuracy level of 0.743, exceeding the performance of single models.

Table 7 Outputs of the hybrid CNN-GRU model employing multiple color spaces and RGB-extracted variables

Model	Image	f	Training			Validation			Performance			
			Ac	Ls	Tt	Ac	Ls	Tv	Re	Pr	Fm	IoU
CNN_{img}-GRU_{img}	Gray	-	1.0	1.675×10^{-4}	38.582	0.912	0.260	1.933	0.912	0.912	0.912	0.839
	RGB	No	1.0	1.35×10^{-5}	37.989	0.970	0.116	1.929	0.970	0.970	0.970	0.939
	HSV	-	1.0	8.761×10^{-5}	38.591	0.979	0.112	2.024	0.979	0.979	0.979	0.947
	CMYK	-	1.0	0.003	38.484	0.984	0.045	2.134	0.984	0.984	0.984	0.977
CNN_{img}-GRU_f	Gray	-	1.0	2.316×10^{-4}	7.559	0.943	0.201	1.142	0.943	0.943	0.943	0.890
	RGB	Yes	1.0	8.581×10^{-6}	7.585	0.979	0.081	1.039	0.979	0.979	0.979	0.952
	HSV	-	1.0	6.971×10^{-5}	7.251	0.983	0.098	1.075	0.983	0.983	0.983	0.959
	CMYK	-	1.0	5.353×10^{-6}	7.088	0.995	0.034	1.065	0.995	0.995	0.995	0.989

3.5 Advanced model vs. current deep learning architectures

The performance of the top-tier hybrid network, as exemplified in Figure 9, is benchmarked against the most recent advancements in image classification technologies. Notably, this comprises Vision Transformers (ViT) introduced by Dosovitskiy et al.[62], Swin Transformers (Swin) developed by Liu et al.[63], and Multi-Layer Perceptrons (MLPMixer) as proposed by Tolstikhin et al.[64]. The comparative analysis presented in this study confirmed a significant improvement in the proposed approach, attaining superior image classification performance compared to earlier models. The planned model consistently demonstrated better performance in terms of Ac (99.5%), Re (99.5%), Pr (99.5%), Fm (99.5%), and IoU (98.9%) compared to the ViT (0.948, 0.948, 0.948, 0.900), the Swin (0.898, 0.898, 0.899, 0.897), and the MLPMixer (0.883, 0.883, 0.895, 0.882, 0.791). The proposed framework utilizes multimodal learning to integrate data from various sources, enabling effective feature extraction and facilitating the development of more robust models for informed agricultural decision-making[65-67].

3.6 Training curves of the deep networks

Numerous procedures have been undertaken to advance the effectiveness of categorization algorithms: 1) analyzing the characteristics of nominated advanced attributes during the training of a specific deep learning model, 2) fine-tuning parameters, 3) thoroughly exploring components within the deep network architecture, 4) considering different image color spaces, 5) integrating VIs with GLCM features, and 6) fusing various deep learning models. Together, these endeavors have played a pivotal role in advancing the learning curve of the higher-classification

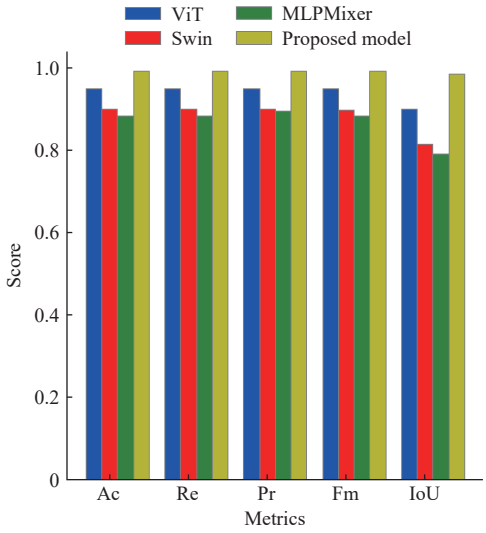
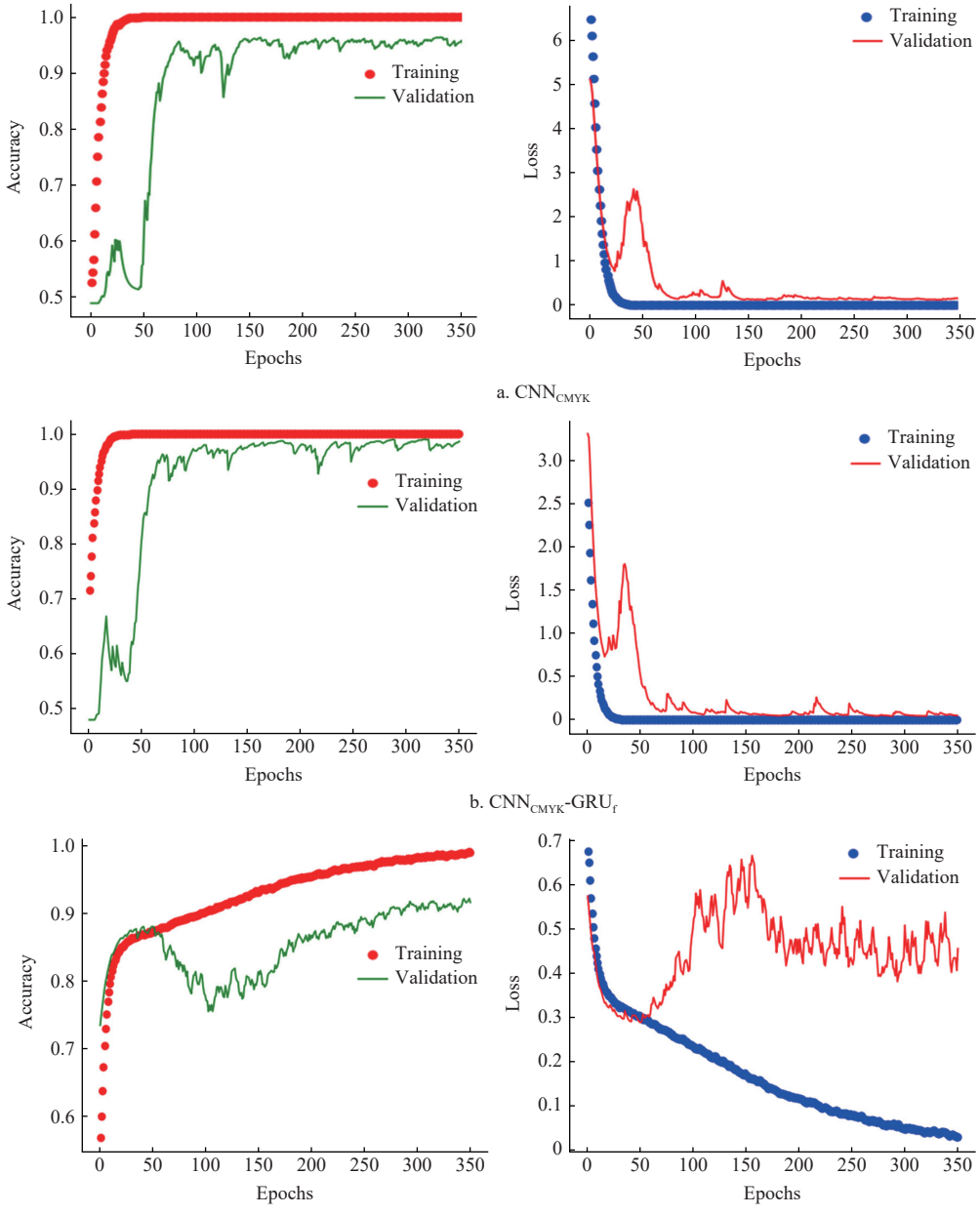


Figure 9 Evaluation of recent deep learning models in comparison with the proposed approach

model, as depicted in Figure 10. It proves the amended performance of the model by deploying upper variations mined from CMYK color space images during the training and validation phases. These specific attributes, coupled with RGB-extracted traits, exhibited a larger ability to evaluate the fitness status of branches compared to other color space images. As illustrated in Figures 10a-10c, the precision of both the instructional and verification datasets consistently improved as the number of epochs increased. Simultaneously, there was a steady decrease in the models' loss, interspersed with occasional increases followed by subsequent decreases. Additionally, it was observed that the training accuracy tended to surpass the validation accuracy. The advanced model, known as CNN_{CMYK}-GRU_f, demonstrated an outstanding learning curve, as revealed in Figure 10b. The incorporation of two models led to the successful establishment of the CNN_{CMYK}-GRU_f two-stage model framework. It achieved excellent classification performance, reaching 99.5% accuracy in terms of Re, Pr, and Fm, along with IoU of 98.9% and Ls of 0.034. Compared to other models, the CNN model based solely on image input (Figure 10a) achieved



a. CNN_{CMYK}

b. CNN_{CMYK}-GRU_f

c. ViT

Figure 10 Performance assessment measures (accuracy and loss) of deep networks for mulberry branch health detection

96.9% for Ac, Re, Pr, and Fm, with IoU=93.7% and Ls=0.060, while the ViT model (Figure 10c) attained 94.8% for Ac, Re, Pr, and Fm, with IoU=90.0% and Ls=0.186. The implementation of the superior model aligns with the anticipated outcomes delineated by Wakamori et al.[68], who underscored that the adoption of multimodal approaches markedly augments accuracy relative to single-modal methods. Elmetwalli et al.[69] validated the enhanced performance as well, noting that the adoption of early stopping during backpropagation training mitigated the risk of overfitting.

3.7 Confusion matrices and the interpretability of deep networks

As shown in Figure 11, the confusion matrices showcase the performance of the proposed CNN_{CMYK}, hybrid CNN_{CMYK}-GRU_f, and ViT models in identifying the health of mulberry twigs. In a collection of 840 images, the first, second, and third models exhibited approximately 30, 8, and 44 instances of misclassification, respectively. This analysis indicated that the CNN_{CMYK}-GRU_f model revealed a marked reduction in misclassifications, repeatedly exposing higher true-positive and true-negative rates, while concurrently minimizing both false negatives and false positives. To facilitate a more in-depth analysis of these models, saliency map visualizations were utilized to evaluate their interpretative transparency. As exposed in Figure 12, the CNN_{CMYK} model produced a widely scattered distribution of high-activation regions, suggesting that it relies on multiple features but lacks a clear focus on disease-relevant areas. This could contribute to its moderate misclassification rate. In contrast, the CNN_{CMYK}-GRU_f model generated more centralized and structured activations, indicating that the inclusion of GRU layers helped refine the model's focus on disease-prone regions, leading to a reduced misclassification rate. Conversely, the ViT model, despite its ability to capture global dependencies, exhibited a higher misclassification rate (44 instances) than the CNN_{CMYK}-GRU_f. The saliency maps indicate that while ViT provides better-distributed attention across the entire twig, it might have struggled with distinguishing subtle textural differences between healthy and unhealthy samples.

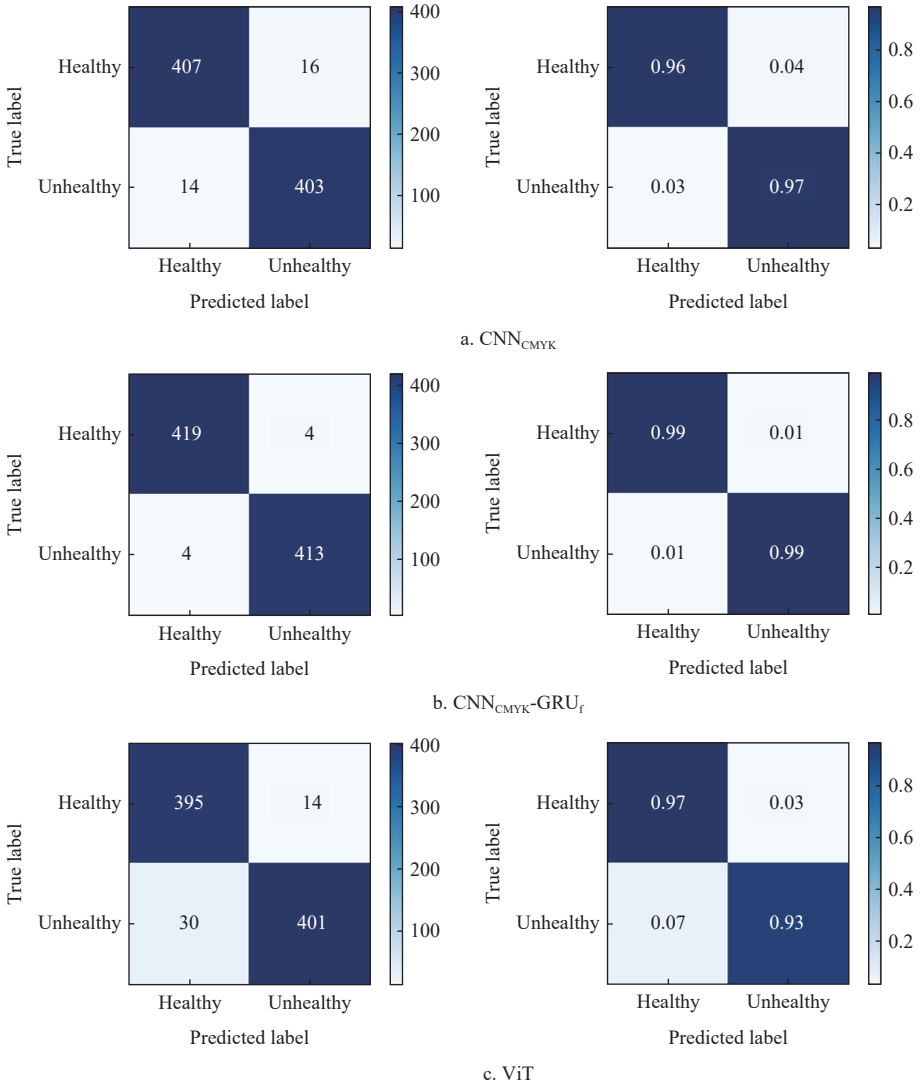


Figure 11 Confusion matrix for the assessment of mulberry branch health detection via the proposed models

In this study, the CNN_{CMYK}-GRU_f framework had superior accuracy in diagnosing the health status of mulberry twigs. Compared to prior research, the findings confirmed a higher level of precision, particularly in contrast to the approach introduced by Narimani et al.[70], who developed an IoT-based system for monitoring plant situations in an aeroponic greenhouse. The VGG-19, InceptionResNetV2, and InceptionV3 algorithms were utilized for Geranium plant disease detection. The outcomes disclosed that the VGG-19 algorithm attained a 92% accuracy in categorizing rust diseases in Geranium plants. Moreover, the methodology of this study outperformed that of Kurup et al.[71], who utilized capsule networks for plant infection detection and leaf classification. With a

dataset of 54 306 images from 14 crop species, they realized 94% accuracy for disease diagnosis and 85% for leaf classification with 2997 images from 11 species. Furthermore, the first-level model excels beyond the methodology introduced by Karlekar and Seal[72] in terms of soybean disease identification accuracy. Their two-stage approach involves extracting leaf components by eliminating complex backgrounds, followed by the application of a CNN for disease classification. In experiments conducted on the "Image Database of Plant Disease Symptoms" which encompasses 16 categories, they achieved an exceptional identification accuracy of 98.14%. The observations detailed herein align with the broader trend in agricultural research, which highlights the integration of multiple data sources into dynamic models for more effective plant status diagnosis[73], disease prediction[74,75], and soil analysis[76]. This approach enhances agricultural decision-making, optimizes crop management, and sets the stage for more efficient, data-driven solutions in modern agriculture.

This work contributes to the precision of detecting the health of mulberry twigs, facilitating timely intervention. It aids in the preservation of water and nutrients that might otherwise be inefficiently expended on diseased plants, thereby promoting robust root formation and increasing the efficiency of water absorption. Diseases like mildew can disrupt water uptake, emphasizing the necessity of this study's advanced deep learning framework for early detection of mildew presence on mulberry branches and the implementation of targeted remedial actions. These actions may involve adjusting water and nutrient supply[77], optimizing plant spatial arrangement to prevent pathogen transmission[78], and utilizing beneficial microbes or biostimulants to enhance plant defenses[79]. The results demonstrated the effectiveness of this innovative classification approach, potentially revolutionizing aeroponic cultivation through automated disease diagnosis. This advancement could lead to improvements in plant health, crop yield, and overall sustainability, while also opening doors for broader applications of deep learning and image analysis in agriculture, heralding a new era in sustainable crop disease management.

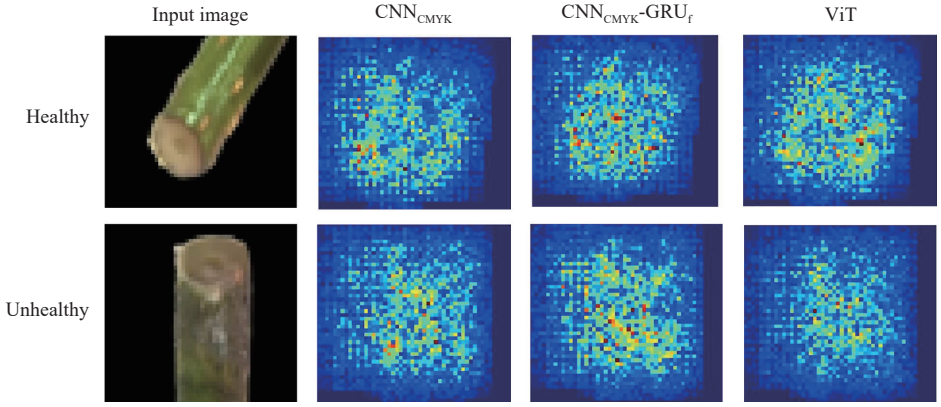


Figure 12 Saliency map visualization for model interpretability, highlighting regions of interest and comparing feature maps of healthy and unhealthy images across different models

3.8 Potential limitations and future outlook

Despite the promising outcomes established by the employed methodology, several constraints within this framework remain, which could be explored in future research endeavors. The current experiments were confined to mulberry branches. Extending the testing to other crops would significantly improve its applicability in precision agriculture, particularly for automated detection. The efficacy of the study strategy is contingent upon the incorporation of fusion traits, particularly those derived from VIs and GLCM. Given the inherent variability across datasets, incorporating supplementary features (like environmental factors, for instance temperature, humidity, light intensity, and carbon dioxide levels) may be crucial to ensuring consistent accuracy across different environments and crop types. Moreover, as the ensemble models were established via specific ML approaches such as CNN and GRU, exploring alternative configurations could yield valuable insights. With the continuous emergence of novel models and techniques such as transformer networks, periodic updates may be essential to sustain predictive performance. Although executed under controlled conditions, appraising the framework in semi-controlled greenhouse environments across seasons would offer crucial perspectives. Long-term trials of this nature could deepen the understanding of its performance under a broader range of operational conditions. Lastly, the robust framework of the established model on the mulberry dataset indicates its potential for adaptation to mobile platforms, such as smartphones or drones, for real-time disease detection in the field. Developing a lightweight version to lower computational demands while preserving accuracy would enhance its viability for use in precision agriculture.

4 Conclusions

This study proposes a novel hybrid deep network for precise mildew detection on aeroponically grown mulberry branches. The model combines a Convolutional Neural Network (CNN) with Gated Recurrent Units (GRU) and processes RGB images across a variety of color channels. It is trained with multi-dimensional attributes, including color-based vegetation indices (VIs) and texture features from the gray-level co-occurrence matrix (GLCM). The experimental outcomes demonstrated that the hybrid deep network (CNN$_{CMYK}$-GRU$_f$), which combined CMYK and advanced VIs-GLCM features, outperformed competing models. This model attained an impressive validation accuracy (Ac) of 99.50% with a corresponding loss (Ls) of 0.034. In contrast, the standalone CNN$_{CMYK}$ model attained Ac of 0.969 and Ls of 0.060. The GRU$_{CMYK}$ model, trained solely on image data, exhibited significantly inferior performance (Ac=0.530, Ls=0.687).

The suggested framework enhances model performance through optimal feature selection, architecture exploration, hyperparameter tuning, and multimodal data integration. This approach leads to efficient, nondestructive detection of branch infections in mulberry trees, enabling timely interventions and promoting tree recovery without labor-intensive processes. Beyond

application in mulberry cultivation, the advanced hybrid deep network presents significant potential for broader use across agricultural sectors. The ability to detect plant diseases early with high accuracy could streamline disease management protocols, reduce the need for chemical treatments, and promote sustainable agricultural practices. Moreover, this methodology could be adapted to detect a wider range of plant pathogens (such as rusts, blights, and molds) across various crops. In the future, integrating this model into real-time monitoring systems could provide farmers with accessible, accurate, and cost-effective diagnostic tools, contributing to improved food security and global agricultural productivity.

Acknowledgements

The authors acknowledge the financial support provided by the National Natural Science Foundation of China Program (Grant No. 51975255), the Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions (Grant No. PAPD-2018-87), and the Jiangsu Funding Program for Excellent Postdoctoral Talent (Grant No. 2023ZB888).

[References]

[1] Duragkar H A. Identifying diseases in mulberry leaves that affects silk production: A deep learning approach. PhD dissertation. Dublin: National College of Ireland, 2023; 17p.

[2] Food and Agriculture Organization of the United Nation (FAO). International plant protection convention. Plant Health and Food Security. 2017.

[3] Kim K H, Kabir E, Jahan S A. Exposure to pesticides and the associated human health effects. Science of the Total Environment, 2017; 575: 525–535.

[4] Knillmann S, Liess M. Pesticide effects on stream ecosystems. In: Schröter M, Bonn A, Klotz S, Seppelt R, Baessler C, editors. Atlas of ecosystem services: Drivers, risks, and societal responses. Springer. 2019; 211–214.

[5] Sofo A, Zanella A, Ponge J F. Soil quality and fertility in sustainable agriculture, with a contribution to the biological classification of agricultural soils. Soil Use and Management, 2022; 38(2): 1085–1112.

[6] Lakhiar I A, Gao J, Syed T N, Chandio F A, Buttar N A. Modern plant cultivation technologies in agriculture under controlled environment: A review on aeroponics. Journal of Plant Interactions, 2018; 13(1): 338–352.

[7] Luo Y H. Research and experiment on key technologies of root zone environment control in aeroponics cultivation system. PhD dissertation. Changsha: Hunan Agricultural University, 2020; 170p. (in Chinese)

[8] Kumar S P, Sampath B, Kumar R S, Babu B H, Ahalya N. Hydroponics, aeroponics, and aquaponics technologies in modern agricultural cultivation. In: Trends, Paradigms, and Advances in Mechatronics Engineering. 2023; pp.223–241. doi: 10.4018/978-1-6684-5887-7.ch012.

[9] Tokunaga H, Anh N H, Dong N V, Ham L H, Hanh N T, Hung N, et al. An efficient method of propagating cassava plants using aeroponic culture. Journal of Crop Improvement, 2020; 34(1): 64–83.

[10] Yu Y Y, Jia D D, Zhuang Z, Zhu C X, Liu A Q. Research progress on aeroponics for plants. Jiangsu Journal of Agricultural Sciences, 2019; 47(18): 38–42. (in Chinese)

[11] Tunio M H, Gao J, Shaikh S A, Lakhiar I A, Qureshi W A, Solangi K A, et al. Potato production in aeroponics: An emerging food growing system in sustainable agriculture for food security. Chilean Journal of Agricultural Research, 2020; 80(1): 118–132.

[12] Fasciolo B, Awouda A, Bruno G, Lombardi F. A smart aeroponic system for sustainable indoor farming. Procedia CIRP, 2023; 116: 636–641.

[13] Sadek N, kamal N, Shehata D. Internet of Things based smart automated indoor hydroponics and aeroponics greenhouse in Egypt. Ain Shams Engineering Journal, 2024; 15(2): 102341.

[14] Ghimire A, Dahal M, Karki R. Hydroponics: An innovative approach to urban agriculture. Nepalese Journal of Agricultural Sciences, 2023; 25: 89–98.

[15] Fernando S D, Gamage A, De Silva D H. Machine learning to aid in the process of disease detection and management in soilless farming. In: 2022 IEEE 7th International conference for Convergence in Technology (I2CT), Mumbai, India, 2022; pp.1–5. doi: 10.1109/I2CT54291.2022.9824206.

[16] Abbaspour-Gilandeh Y, Molaee A, Sabzi S, Nabipur N, Shamshirband S, Mosavi A. A combined method of image processing and artificial neural network for the identification of 13 Iranian rice cultivars. Agronomy, 2020; 10(1): 117.

[17] Elsherbiny O, Elaraby A, Alahmadi M, Hamdan M, Gao J. Rapid grapevine health diagnosis based on digital imaging and deep learning. Plants, 2024; 13(1): 135.

[18] Prey L, von Bloh M, Schmidhalter U. Evaluating RGB imaging and multispectral active and hyperspectral passive sensing for assessing early plant vigor in winter wheat. Sensors, 2018; 18: 2931.

[19] Elsherbiny O, Gao J, Ma M, Guo Y, Tunio M H, Mosha A H. Advancing lettuce physiological state recognition in IoT aeroponic systems: A meta-learning-driven data fusion approach. European Journal of Agronomy, 2024; 161: 127387.

[20] Li L, Zhang Q, Huang D F. A review of imaging techniques for plant phenotyping. Sensors, 2014; 14: 20078–20111.

[21] Elsherbiny O, Zhou L, He Y, Qiu Z J. A novel hybrid deep network for diagnosing water status in wheat crop using IoT-based multimodal data. Computers and Electronics in Agriculture, 2022; 203: 107453.

[22] Liu Y, Feng H, Yue J, Li Z, Yang G, Song X, et al. Remote-sensing estimation of potato above-ground biomass based on spectral and spatial features extracted from high-definition digital camera images. Computers and Electronics in Agriculture, 2022; 198: 107089.

[23] Zhou L, Zhang C, Taha M F, Qiu Z J, He Y. Determination of leaf water content with a portable NIRS system based on deep learning and information fusion analysis. Transactions of the ASABE, 2021; 64(1): 127–135.

[24] Li W, Sun Z, Lu S, Omasa K. Estimation of the leaf chlorophyll content using multiangular spectral reflectance factor. Plant, Cell & Environment, 2019; 42(11): 3152–3165.

[25] Taha M F, Abdalla A, ElMasry G, Gouda M, Zhou L, Zhao N, et al. Using deep convolutional neural network for image-based diagnosis of nutrient deficiencies in plants grown in aquaponics. Chemosensors, 2022; 10(2): 45.

[26] Zhou L, Xiao Q, Taha M F, Xu C J, Zhang C. Phenotypic analysis of diseased plant leaves using supervised and weakly supervised deep learning. Plant Phenomics, 2023; 5: 0022.

[27] Ruwanpathirana P P, Sakai K, Jayasinghe G Y, Nakandakari T, Yuge K, Wijekoon W M J C, et al. Evaluation of sugarcane crop growth monitoring using vegetation indices derived from RGB-based UAV images and machine learning models. Agronomy, 2024; 14(9): 2059.

[28] Mathew A, Antony A, Mahadeshwar Y, Khan T, Kulkarni A. Plant disease detection using GLCM feature extractor and voting classification approach. Materials Today: Proceedings, 2022; 58: 407–415.

[29] Bhimte N R, Thool V R. Diseases detection of cotton leaf spot using image processing and SVM classifier. In: 2018 Second international conference on intelligent computing and control systems (ICICCS), Madurai, India: IEEE, 2018; pp.340–344.

[30] Alomar K, Aysel H I, Cai X. Data augmentation in classification and segmentation: A survey and new strategies. Journal of Imaging, 2023; 9(2): 46.

[31] Jin X, Jie L, Wang S, Qi H J, Li S W. Classifying wheat hyperspectral pixels of healthy heads and Fusarium head blight disease using a deep neural network in the wild field. Remote Sensing, 2018; 10(3): 395.

[32] Ullah Z, Alsubaie N, Jamjoom M, Alajmani S H, Saleem F. EffiMob-Net: A deep learning-based hybrid model for detection and identification of tomato diseases using leaf images. Agriculture, 2023; 13(3): 737.

[33] Nery E W, Kubota L T. Sensing approaches on paper-based devices: A review. Analytical and Bioanalytical Chemistry, 2013; 405: 7573–7595.

[34] Yossy E H, Pranata J, Wijaya T, Hermawan H, Budiharto W. Mango fruit sortation system using neural network and computer vision. Procedia Computer Science, 2017; 116: 596–603.

[35] Feng W, Shen W Y, He L, Duan J Z, Guo B B, Li Y X, et al. Improved remote sensing detection of wheat powdery mildew using dual-green vegetation indices. Precision Agriculture, 2016; 17: 608–627.

[36] Zhu W J, Feng Z K, Dai S Y, Zhang P P, Wei X H. Using UAV multispectral remote sensing with appropriate spatial resolution and machine learning to monitor wheat scab. Agriculture, 2022; 12(11): doi: 10.3390/agriculture12111785.

[37] Peng Y, Zhao S Y, Liu J Z. Fused-deep-features based grape leaf disease diagnosis. Agronomy, 2021; 11(11): 2234.

[38] Netto A F A, Martins R N, de Souza G S, de Moura Araújo G, de Almeida S L H, et al. Segmentation of RGB images using different vegetation indices and thresholding methods. Nativa, 2018; 6(4): 389–394.
[39] Goddijn-Murphy L, Dailloux D, White M, Bowers D. Fundamentals of in situ digital camera methodology for water quality monitoring of coast and ocean. Sensors, 2009; 9(7): 5825–5843.
[40] Sellaro R, Crepy M, Trupkin S A, Karayekov E, Buchovsky A S, Rossi C, et al. Cryptochrome as a sensor of the blue/green ratio of natural radiation in Arabidopsis. Plant Physiology, 2010; 154(1): 401–409.
[41] Verrelst J, Schaepman M E, Koetz B, Kneubühler M. Angular sensitivity analysis of vegetation indices derived from CHRIS/PROBA data. Remote Sensing of Environment, 2008; 112(5): 2341–2353.
[42] Woebbecke D M, Meyer G E, Von Bargen K, Mortensen D A. Color indices for weed identification under various soil, residue, and lighting conditions. Transactions of the ASAE, 1995; 38(1): 259–269.
[43] Kawashima S, Nakatani M. An algorithm for estimating chlorophyll content in leaves using a video camera. Annals of Botany, 1998; 81(1): 49–54.
[44] Woebbecke D M, Meyer G E, Von Bargen K, Mortensen D A. Plant species identification, size, and enumeration using machine vision techniques on near-binary images. In: Optics in Agriculture and Forestry, International Society for Optics and Photonics, 1993; pp.208–219.
[45] Mao W H, Wang Y M, Wang Y Q. Real-time detection of between-row weeds using machine vision. In: 2003 ASAE Annual Meeting, American Society of Agricultural and Biological Engineers, 2003. doi: 10.13031/2013.15381.
[46] Gerardo R, de Lima I P. Applying RGB-Based vegetation indices obtained from uas imagery for monitoring the rice crop at the field scale: A case study in Portugal. Agriculture, 2023; 13(10): 1916.
[47] Hague T, Tillett N D, Wheeler H. Automated crop and weed monitoring in widely spaced cereals. Precision Agriculture, 2006; 7: 21–32.
[48] Saberioon M M, Amin M S, Anuar A R, Gholizadeh A, Wayayok A, Khairunniza-Bejo S. Assessment of rice leaf chlorophyll content using visible bands at different growth stages at both the leaf and canopy scale. International Journal of Applied Earth Observation and Geoinformation, 2014; 32: 35–45.
[49] Kataoka T, Kaneko T, Okamoto H, Hata S. Crop growth estimation system using machine vision. In: Proceedings 2003 IEEE/ASME international conference on advanced intelligent mechatronics (AIM 2003), Kobe, Japan: IEEE, 2003; pp.b1079–b1083.
[50] Guijarro M, Pajares G, Riomoros I, Herrera P J, Burgos-Artizzu X P, Ribeiro A. Automatic segmentation of relevant textures in agricultural images. Computers and Electronics in Agriculture, 2011; 75(1): 75–83.
[51] Hall-Beyer M. GLCM Texture: A Tutorial v. 1.0 through 2.7. Available: http://hdl.handle.net/1880/51900. Accessed on[2023-04-04].
[52] Latif M A, Afshan N, Mushtaq Z, Khan N A, Irfan M, Nowakowski G. Enhanced classification of coffee leaf biotic stress by synergizing feature concatenation and dimensionality reduction. IEEE Access, 2023; 100887–100906. doi: 10.1109/ACCESS.2023.3314590.
[53] Maheswari M U, Ramani R A. Comparative study of agricultural crop yield prediction using machine learning techniques. In: 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS), 2023; pp.1428–1433. doi: 10.1109/ICACCS57279.2023.10112854.
[54] Kamilaris A, Prenafeta-Boldú F X. Deep learning in agriculture: A survey. Computers and Electronics in Agriculture, 2018; 147: 70–90.
[55] Ubbens J R, Stavness I. Deep plant phenomics: a deep learning platform for complex plant phenotyping tasks. Frontiers in Plant Science, 2017; 8: 1190.
[56] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: Proceedings of the 32nd International Conference on International Conference on Machine Learning, 2015; pp.448–456.
[57] Bi L, Wally O, Hu G, Tenuta A U, Kandel Y R, Mueller D S. A transformer-based approach for early prediction of soybean yield using time-series images. Frontiers in Plant Science, 2023; 14: 1173036.
[58] Chung J, Gulcehre C, Cho K, Bengio Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint, 2014; In press. doi: 10.48550/arXiv.1412.3555.
[59] Gaagai A, Aouissi H A, Bencedira S, Hinge G, Athamena A, Heddam S, et al. Application of water quality indices, machine learning approaches, and GIS to identify groundwater quality for irrigation purposes: A case study of Sahara Aquifer, Doucen Plain, Algeria. Water, 2023; 15(2): 289.
[60] Barbedo J G A. Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification. Computers and Electronics in Agriculture, 2018; 153: 46–53.
[61] Iwana B K, Uchida S. An empirical survey of data augmentation for time series classification with neural networks. Plos one, 2021; 16(7): e0254841.
[62] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv: 2010.11929, 2020; In press. doi: 10.48550/arXiv.2010.11929.
[63] Liu Z, Lin Y T, Cao Y, Hu H, Wei Y X, Zhang Z, et al. Swin transformer: Hierarchical vision transformer using shifted windows. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada: IEEE, 2021; pp.10012–10022.
[64] Tolstikhin I O, Houlsby N, Kolesnikov A, Beyer L, Zhai X, Unterthiner T, et al. Mlp-mixer: An all-mlp architecture for vision. Advances in Neural Information Processing Systems, 2021; 34: 24261–24272.
[65] Summaira J, Li X, Shoib A M, Li S, Abdul J. Recent advances and trends in multimodal deep learning: A review. arXiv preprint arXiv: 2105.11087, 2021; In press. doi: 10.48550/arXiv.2105.11087.
[66] Zhong F M, Chen Z K, Zhang Y C, Xia F. Zero-and few-shot learning for diseases recognition of Citrus aurantium L. using conditional adversarial autoencoders. Computers and Electronics in Agriculture, 2020; 179: 105828.
[67] Zhu W J, Li J Y, Li L, Wang A C, Wei X H, Mao H P. Nondestructive diagnostics of soluble sugar, total nitrogen and their ratio of tomato leaves in greenhouse by polarized spectra–hyperspectral data fusion. Int J Agric & Biol Eng, 2020; 13(2): 189–197.
[68] Wakamori K, Mizuno R, Nakanishi G, Mineno H. Multimodal neural network with clustering-based drop for estimating plant water stress. Computers and Electronics in Agriculture, 2020; 168: 105118.
[69] Elmetwalli A H, Mazrou Y S A, Tyler A N, Hunter P D, Elsherbiny O, Yaseen Z M, et al. Assessing the efficiency of remote sensing and machine learning algorithms to quantify wheat characteristics in the Nile Delta Region of Egypt. Agriculture, 2022, 25; 12(3): 332. doi: 10.3390/agriculture12030332.
[70] Narimani M, Hajiahmad A, Moghimi A, Alimardani R, Rafiee S, Mirzabe A H. Developing an aeroponic smart experimental greenhouse for controlling irrigation and plant disease detection using deep learning and IoT. In: 2021 ASABE Annual International Virtual Meeting, Michigan: American Society of Agricultural and Biological Engineers, 2021. doi:10.13031/aim.202101252.
[71] Kurup R V, Anupama M A, Vinayakumar R, Sowmya V, Soman K P. Capsule network for plant disease and plant species classification. In Computational Vision and Bio-Inspired Computing (ICCVBIC 2019), Springer, 2020; pp.413–421. doi: 10.1007/978-3-030-37218-7_47.
[72] Karlekar A, Seal A. SoyNet: Soybean leaf diseases classification. Computers and Electronics in Agriculture, 2020; 172: 105342.
[73] Taha M F, Mao H, Mousa S, Zhou L, Wang Y, Elmasry G, et al. Deep learning-enabled dynamic model for nutrient status detection of aquaponically grown plants. Agronomy, 2024; 14: 2290.
[74] Zhang X D, Wang Y F, Zhou Z K, Zhang Y X, Wang X Z. Detection method for tomato leaf mildew based on hyperspectral fusion terahertz technology. Foods, 2023; 12(3): 535.
[75] Wang Y F, Li T Z, Chen T H, Zhang X D, Taha M F, Yang N, et al. Cucumber downy mildew disease prediction using a CNN-LSTM approach. Agriculture, 2024; 14: 1155.
[76] Zheng W G, Lan R P, Zhangzhong L, Yang L N, Gao L T, Yu J X. A hybrid approach for soil total nitrogen anomaly detection integrating machine learning and spatial statistics. Agronomy, 2023; 13(11): 2669.
[77] Mangaiyarkarasi R. Aeroponics system for production of horticultural crops. Madras Agricultural Journal, 2020; 107: 1–7.
[78] Kotzen B, Emerenciano M G, Moheimani N, Burnell G M. Aquaponics: Alternative types and approaches. In: Goddek S, Joyce A, Kotzen B, Burnell G M, editors. Aquaponics food production systems: Combined aquaculture and hydroponic production technologies for the future, Springer. 2019; pp.301–330.
[79] Riggio G M, Jones S L, Gibson K E. Risk of human pathogen internalization in leafy vegetables during lab-scale hydroponic cultivation. Horticulturae, 2019; 5(1): 25.