scientific reports

Check for updates

OPEN

Image segmentation for pest detection of crop leaves by improvement of regional convolutional neural network

Xianchuan Wu[1], Yuling Liu[1], Mingjing Xing[1], Chun Yang[2] & Shaoyong Hong[1]✉

Pest detection is important for crop cultivation. Crop leaf is the main place of pest invasion. Current technologies to detect crop pests have constraints, such as low efficiency, storage demands and limited precision. Image segmentation is a fast and efficient computer-aided detection technology. High resolution image capture solidly supports the crucial processes in discerning pests from images. Study of analytical methods help parse information in the images. In this paper, a regional convolutional neural network (R-CNN) architecture is designed in combination with the radial bisymmetric divergence (RBD) method for enhancing the efficiency of image segmentation. As a special application of RBD, the hierarchical mask (HM) is produced to endorse detection and classification of the leaf-dwelling pests, offering enhanced efficiency and reduced storage requirements. Moreover, to deal with some mislabeled data, a threshold variable is introduced to adjust a fault-tolerant mechanism into HM, to generate a novel threshold-based hierarchical mask (TbHM). Consequently, the hierarchical mask R-CNN (HM-R-CNN) model and the threshold-based hierarchical mask R-CNN (TbHM-R-CNN) model are established to detect various types of healthy and pest-invasive crop leaves to select the regional image features that are rich in pest information. Then simple linear iterative clustering (SLIC) method is incorporation to finish the image segmentation for the classification of pest invasion. The models are tuned and optimized, then validated. The most optimized modeling results are from the TbHM-R-CNN model, with the classification accuracy of 96.2%, the recall of 97.5% and the F1 score of 0.982. Additionally, the HM-R-CNN model observed appreciable results second only to the best model. These results indicate that the proposed methodologies are well-suited for training and testing a dataset of plant diseases, offering heightened accuracy in pest classification. This study revealed that the proposed methods significantly outperform the existing techniques, marking a substantial improvement over current methods.

Keywords Pest detection, Crop leaves, Image segmentation, Regional convolutional neural network (R-CNN), Hierarchical mask (HM), Threshold-based hierarchical mask (TbHM)

Agriculture is a vital component of any country's economy. The number of crops produced annually is quite important for productivity. Crops were susceptible to a variety of pests, leading to significant economic waste and loss of labor resources[1]. Over 300 different types of pests or worms can cause harm to crops, either directly or indirectly. Over the years, crops become more susceptible to pests due to weather changes and infrequent rainfall[2]. Pests have killed or destroyed more than half of all crops produced. Many farmers employ traditional pest management techniques, such as routine pesticide spraying based on set schedules, rather than the actual presence of pests in the paddy fields[3]. However, the widespread use of insecticides, pesticides, and nematicides can be detrimental to the environment and living organisms as they can contaminate crops and devastate entire farms. Once an infestation or infection has reached a certain level, traditional pest control methods like insecticides, water, and smoke treatments become ineffective. Therefore, regular monitoring is necessary to detect and correctly identify bugs or pests, enabling the implementation of preventive measures to halt their proliferation[4]. But manual monitoring is labor-intensive and time-consuming. Technology plays a significant role in advancing agriculture and aiding farmers. By utilizing the latest agricultural informationization technology, pests or bugs can be rapidly detected at an early stage.

[1]School of Data Science, Guangzhou Huashang College, Guangzhou 511300, China. [2]School of Accounting, Guangzhou Huashang College, Guangzhou 511300, China. ✉email: shy2002021@163.com

Agricultural informationization raises the key issues of enhancing productivity and improving food quality while minimizing costs and maximizing profits. in recent years, Nonchemical pest detection and disease control are on high demand in many departments. However, there are currently no intelligent methods or frameworks to guarantee a timely and efficient detection manner. Modern advancements and improvements in computer science have offered various forms of support to humanity, facilitating our daily tasks in multiple ways and enhancing efficiency[5]. Computer vision approaches have become increasingly significant. This allows us to minimize the time spent on various activities. Intelligent visual sensing computation method was studied to estimate the bacteria inoculation status in tropical maize, which involves assessing the color properties of field leaf images captured under varied lighting conditions[6,7].

With the increasing development of intelligent visual sensors globally and recent advancements in deep learning within the computer vision field[8,9], innovative mechanism emerged to identify pests and diseases on the leaves and trunks of trees, utilizing a smartphone as the image capturing device[10]. When assessed on a collection of images sourced from reliable web domains the model maintained a low accuracy score because the conditions for image capture distinct from those used for training[11]. Nevertheless, this research provoking that, the development of improved materials and methodologies for fully automated pest detection and extraction systems is necessary to yield superior outcomes. Data augmentation was employed to circumvent overfitting issues typically encountered when training deep learning models on limited datasets[12]. The findings indicate that the automated processing within the proposed method is more efficient compared to manual procedures. A fresh learning phase for the insect training set becomes necessary when new biotic stressors are presented[13].

Deep learning technique is beneficial to be implemented for categorize the pests and diseases, feeding the dataset into the discrimination model and subsequently processing it with the feature learning framework[14]. This kind of technique is widely recognized that deep learning yields superior results with larger datasets; hence, it is feasible to train on the entire dataset based on the feedforward networks, encompassing both original and augmented images. An advanced technique that merges the deep convolution learning network with some metaheuristic algorithms to standardize plant photographs[15]. Metaheuristic method was previously validated for image segmentation, in application to distinguish wheat leaves from the intricate background[16]. However, particularly for extensive datasets, the insights derived from a single layer feedforward network are insufficient to accurately depict the input data.

An architecture grounded in deep learning was proposed using convolutional neural networks to categorize multiple pests within a plant disease dataset[17]. The system effectively anticipates by integrating the system into real-world business operations, a larger dataset will be accumulated, offering an improved opportunity to assess the system's performance and potentially enhance its accuracy rate. Despite the numerous physical similarities shared among different insect species, pinpointing and categorizing them remains a challenge. This issue is further compounded when working with insects that have been previously captured in traps. To this end, some researchers have suggested to use the LeNet structure[18]. Nevertheless, their synthesized results from the analyses highlight several areas for improvement: (1) To enhance the overall precision of the methodologies and materials[19]; (2) SLFN alone is inadequate for accurately representing the input data of extensive datasets[20]; (3) The model's capacity is limited to detecting a restricted array of plant diseases[21]; (4) The numerous shared physical attributes among various insect species complicate their localization and classification[22]. Consequently, it becomes essential to devise a novel convolutional system for pest detection and analysis, in capability of accurately identifying diseases while surmounting these existing limitations.

Regional convolutional neural network (R-CNN) is a special kind of deep learning structural framework based on the convolutional architecture[23]. Training the network weights on a large dataset and fine-tuning the pre-trained network weights on a smaller dataset facilitates efficient data analysis within the dataset[24]. Pertaining to image feature extraction, R-CNN is a pre-trained network constructed upon extensive datasets, adept at effectively deriving stratified features from images. R-CNN may be on the basis of ResNet or ImageNet structure[25], to outline the features that facilitate the correlation of patterns and structures within plant disease diagnostic images, adepting at effectively deriving stratified features from images[26]. Thereby it is functional to differentiate the healthy and diseased plants. For color pictures, a series of layered images is also generated by dissecting the color, brightness, and saturation elements of various segments of the image within the RGB color domain[27]. The efficacy of the R-CNN framework can be evaluated on the plant disease dataset by retraining an existing model on the image dataset, to form a transfer learning network for image segmentation.

A state-of-art algorithm for image segmentation is called simple linear iterative clustering (SLIC). It is an adaptation of the k-means approach for generating superpixels, offers high regional accuracy with minimal inter-region overlap[28]. In essence, SLIC performs clustering of pixels within a five-dimensional space that includes color and image plane dimensions, enabling the rapid creation of compact and nearly uniform hyperpixels. It works effectively with characterization of low demand of computational exhaustion. The combination of R-CNN and SLIC is prospective in efficient to optimize the cluster selection for good image segmentation. However, the SLIC-based methodologies need to associate the computational load with distance calculations is diminished by confining the search space to a predefined region equivalent in size to the superpixel[29]. Also, in parallel process with deep estimation, SLIC yields its application within the realm of hyperpixel generation remains under-researched[30]. Therefore, mask strategy is introduced in this paper to enhance the segmentation effects incorporating the network extracted feature sets. Thus, the mask R-CNN algorithm is proposed for regional identification in image segmentation operation.

Radial bisymmetric divergence (RBD) method generates a regional mask to select partial data from the full-range raw image. RBD is a brand-new method used for pest segmentation, and it is validated easy-to-cooperate with network models[31]. The RBD-based mask R-CNN model can perform in multiple tactics. Hierarchical mask (HM) is a popular option for detecting pests within a specific area. with a focus on rapid processing. By HM concept, the image data initially undergoes pre-processing to eradicate superfluous distortions. It is in advance

on compiling a high-quality dataset encompassing various healthy and diseased leaf categories, alongside diverse pest conditions, and its applicability has been proved[32]. HM plays an important role to save memory usage, and to raise the model prediction accuracy[33].

Hierarchical mask R-CNN (HM-R-CNN) gives a simple way to mask the image data, but image segmentation depends on an indirect computation of the image matrics for pest detection though it is of multiple merits. Challenges exist in the fact that it is hard to track the variation of pests in crop leaves throughout the crop production process, because pests' activity is diverse[34]. Occasion also occurs that some inaccurately labeled data can result in flawed learning and diminish the efficacy of pest identification methods[35]. What is more essential to address these issues is to raise a threshold-endorsed analytical method. Confirming that network architecture is fast and effective for image segmentation, the method can be improved with a fault-tolerated technique to create a threshold-based hierarchical mask (TbHM), so as to refine the architecture as for suiting the various complex situations of pest variations. In this concept, the threshold-based hierarchical mask R-CNN (TbHM-R-CNN) architecture is introduced. The TbHM-R-CNN is proposed by modification of the hierarchical HM-R-CNN, threshold adjustment can be set in accompanied with parameter tuning and for result estimations.

In this work, SLIC-based segmentation techniques can achieve superior identification results with the support of feature selection by R-CNN architecture. By optimizing the effects of image segmentation, the main contribution is as follows,

- For the precise detection of insect pests, the HM-R-CNN has been proposed to enhance image data by minimizing unwanted distortions during image preprocessing and to achieve high accuracy in a short duration.
- To obtain various types of healthy and diseased leaves, the TbHM-R-CNN is introduced. It integrates mask R-CNN with a transfer learning mechanism, enabling it to accurately recognize pests and differentiate them into distinct categories.

The rest of the paper is organized with the Methodology section, the Application case section and the Conclusion section. The Methodology section includes the contexts of subsection 1: to describe the image preprocessing procedures. Subsection 2: to introduce the mask segmentation method of HM-R-CNN method. Subsection 3: to introduce the TbHM-R-CNN method for transfer learning. Subsection 4: to present the model evaluation indicators. The section of Application case is to apply the proposed methods to the dataset, including the parts of sample collection, model configuration, performance metrics of the proposed methods and the comparative effect evaluations to some existing methods. Finally, the conclusions are drawn in the conclusion section.

Methodologies

Essential to pest management is the consistent monitoring of pest populations throughout the crop production process. The majority of current pest control techniques necessitate labor-intensive and time-consuming manual examinations. The proposed methods are based on the R-CNN architecture as well as the RBD technique. Initially, the image data undergoes pre-processing to eradicate superfluous distortions. The HM-R-CNN approach incorporates one of the RBD mask mode with the conventional R-CNN structure, to find the informative features. Then SLIC method is utilized to finish the image segmentation work based on the selected feature variables. Moreover, compiling a high-quality dataset encompassing various healthy and diseased leaf categories, alongside diverse pest conditions, can prove challenging. To effectively categorize various types of healthy and diseased leaves, the TbHM-R-CNN architecture is introduced with the consideration of a threshold control on the tunable parameters and the prediction results.

Figure 1 illustrates the general application of the proposed methods based on R-CNN architecture. Prior to the application of the filter layer to the input image, the present input set and its corresponding weight values, alongside the anticipated output values and weights for the subsequent layer, are stored as data. The dataset will subsequently undergo filtering and feature extraction by R-CNN. Once the data is filtered, the image is segmented. This framework represents a derivative of the mask R-CNN instance segmentation model, commencing with a convolutional layer and culminating in one or more convolution layers. The feature extraction process is executed to discern pests utilizing the HM technique, so that the HM-R-CNN model is established for selecting informative features from the raw input images, in which the stacked convolution is designed. Moreover, by introducing a threshold variable, the TbHM-R-CNN model is built up to activate the fault-tolerant mechanism, so that the model can be modified as more suitable for a wide-range applicant stability and robustness. Following feature selection, SLIC method is applied for segmentation, to pinpoint the pests by tuning the model parameters on the characteristics of leaf images, attaining high regional precision and efficiency in detecting the affected areas of diseased leaves. Successively, the resultant segmented regional images are further delivered for transfer learning and detection application.

The whole modeling workflow is suitable for a data application of training and validation. The training dataset is employed to train the segmentation model and fine-tune it for improved performance. The subsequent stage involves transfer learning for classification, where a pre-trained CNN model (can be in LeNet, ResNet, VGG or any other possible structure[36]) is chosen for validation classification, with dropping out the abnormal predictions. With minimal modifications, the model would be capable of classifying pests based on image data. Upon this classification, pest identification is prospectively executed. The aim of mask R-CNN modeling and its calculation is to showcase the capacity of neural networks to identify the shape and orientation of plant maladies, to discern the intrinsic characteristics of disease propagators within leaves.

Image preprocessing and the RBD operation

Image preprocessing involves the enhancement of image data by diminishing superfluous distortions for subsequent analysis. This process does not alter the quantity of image data[37]. Certain attributes of an image

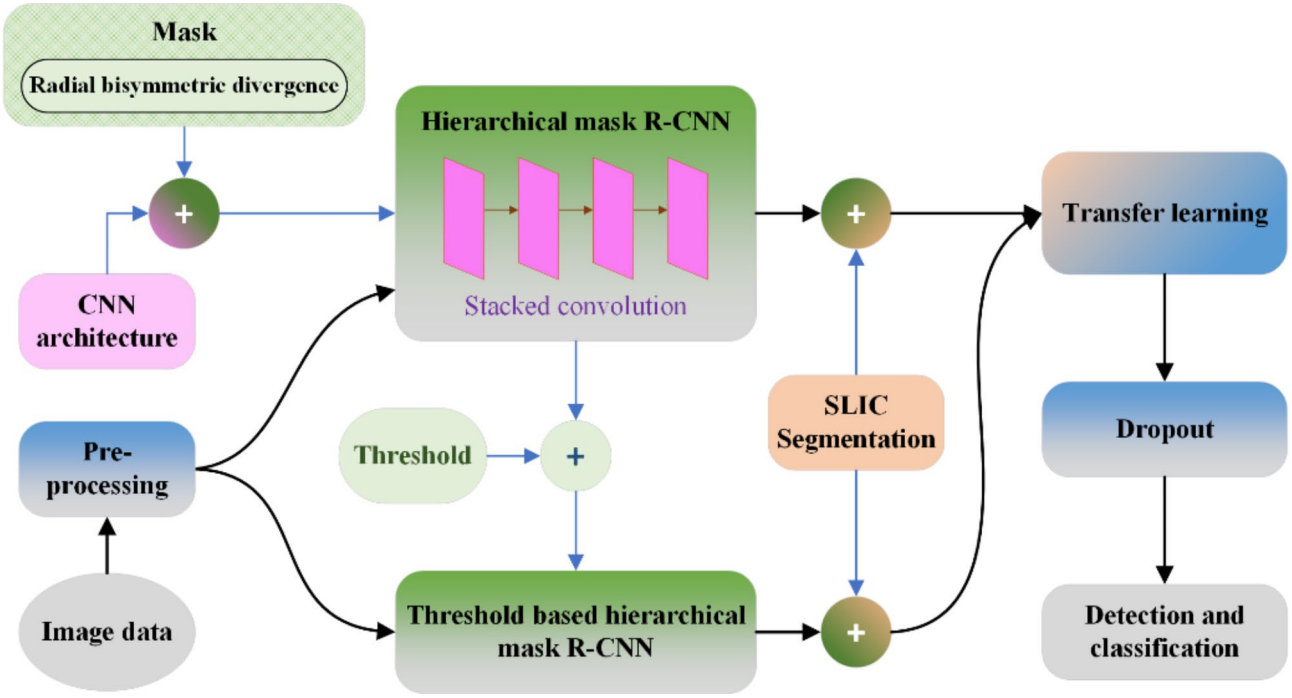

Fig. 1. The flowchart for image segmentation by proposed methods based on R-CNN architecture.

include regional features, grayscale covariance matrix, mean, standard deviation, entropy, contrast, and energy. These attributes are essential for generating the dataset necessary during the leaf detection process. Employing image preprocessing lessens computational complexity and boosts algorithmic efficiency[38].

Prior to extracting valuable insights into leaf status, it is imperative to pre-process the leaf images first. These images are captured and preserved in RGB coloring format. The data obtained is refined through a three-stage pre-processing protocol designed to enhance image quality. The main task of pre-processing is to eliminate the Gaussian noise or white noise from the leaf image, allowing for a more accurate disease detection using the filtered image[39]. The image pre-processing course is performed in mathematic procedures.

A function $f(x,\ y)$ is set to represent the RGB value at a point $(x,\ y)$ in an x- y coordinate system. The change in their pixel values from a relative is represented as the Eqs. (1)-(2)

$$\frac{\partial f}{\partial x} = f\left(x+1, y\right) - f(x, y) \tag{1}$$

$$\frac{\partial f}{\partial y} = f\left(x, y+1\right) - f(x, y) \tag{2}$$

where $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ represent the 1st -order partial derivatives of $f(x,\ y)$ on x and on y, respectively. Moreover, the rate of increment takes place in the x or y direction is indicated by the 2nd -order derivatives, which are formulated in Eq. (3) to (4), so that the differential form of $f(x,\ y)$ can be shown in Eqs. (5),

$$\begin{aligned}\frac{\partial^2 f}{\partial x^2} &= \left(f\left(x+1, y\right) - f\left(x, y\right)\right) - \left(f\left(x, y\right) - f\left(x-1, y\right)\right) \\ &= f\left(x+1, y\right) + f\left(x-1, y\right) - 2f\left(x, y\right).\end{aligned} \tag{3}$$

$$\begin{aligned}\frac{\partial^2 f}{\partial y^2} &= \left(f\left(x, y+1\right) - f\left(x, y\right)\right) - \left(f\left(x, y\right) - f\left(x, y-1\right)\right) \\ &= f\left(x, y+1\right) + f\left(x, y-1\right) - 2f\left(x, y\right).\end{aligned} \tag{4}$$

$$\begin{aligned}D^2 f &= \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \\ &= f\left(x+1, y\right) + f\left(x-1, y\right) + f\left(x, y+1\right) \\ &\quad + f\left(x, y-1\right) - 4f\left(x, y\right).\end{aligned} \tag{5}$$

where $\frac{\partial^2 f}{\partial x^2}$ and $\frac{\partial^2 f}{\partial y^2}$ represent the 2nd-order partial derivative values of $f(x,\ y)$ on x and on y, respectively. $D^2 f$ represents the whole differential value of $f(x,\ y)$.

Radial bisymmetric divergence (RBD) is to define a radial bisymmetric differencing operation to modify the rate of increment takes place in axis directions. The differential calculations are expressed as the same as in Eqs. (1)-(2). The RBD modification requires to verify convergence and determine the edge points instead of using

axial differencing. The adjustment ensures that the sequence of values converges towards a stable equilibrium point, maintaining the desired balance in the system. then the 2nd -order derivatives are re-defined as,

$$\frac{\partial^2 f}{\partial x^2} = \begin{cases} a \cdot \Delta x + k \cdot (f(x+1,y) - f(x,y)), & for\ x \geq 0, \\ a \cdot \Delta x + k \cdot (f(x,y) - f(x-1,y)), & for\ x < 0, \end{cases} \tag{6}$$

$$\frac{\partial^2 f}{\partial y^2} = \begin{cases} b \cdot \Delta y + k \cdot (f(x,y+1) - f(x,y)), & for\ y \geq 0, \\ b \cdot \Delta y + k \cdot (f(x,y) - f(x,y-1)), & for\ y < 0, \end{cases} \tag{7}$$

where Δx and Δy represent the minimal variation in x and y axes, respectively, k is an adjusting parameter to be linked in network training; it is called the divergence rate. This divergence rate indicates that the branching point will be ascertained by the unique characteristics and fluctuations of the flow across the contour regional parts.

The quantity of filters employed within the network architecture is contingent upon the divergence rate (k). The convergence velocity at the edge points is relatively low; exponential function is employed to enhance the speed, which is designed as,

$$\frac{\Delta v}{v} = -k_x e^{\frac{v_x}{x}} + k_y e^{\frac{v_y}{y}} \tag{8}$$

where v_x and v_y are the velocity of convergence along the x and y axes, respectively; k_x and k_y are the corresponding adjusting parameter; $\frac{\Delta v}{v}$ is the gradient in whole differential, and it is measured and moved forward according to the direction of the slope. In calculation, the change of $\frac{\Delta v}{v}$ is quantified by the intensity of dataflow in a particular direction.

Table 1 lists the slope metric of $\frac{\Delta v}{v}$, the convergence of v, as well as the paramters of k, k_x and k_y due to the application of Eqs. (6), (7) and (8), originating from the raw image pixel points. The feasibility of RBD mask strategy applied to R-CNN was examined to adopt the challenges for classification of pest detection.

The mask-endorsed HM-R-CNN

HM-R-CNN is a proposed network architecture that represents a derivative of mask R-CNN instance segmentation model in combined use of RBD technique. Typically, the architecture comprises a series of stacked convolution layers, supplemented by fully connected layers[40].

The input layer receives a matrix composed of image elements, shown as $\boldsymbol{x} \in R^{n \times d}$ in size of n, and in dimension d. The convolution layer uses different convolution to check the input matrix for convolutional operation, and extracts the local features of the input matrix, namely,

$$CON = f(w\boldsymbol{x} + b), \tag{9}$$

where x is the convolution kernel window word vector matrix, w is the weight matrix, b is the bias, and $f(\cdot)$ is the activation function to output the feature of CON. There initialize two pooling layers positioned after two convolution layers, for extracting the feature vector graphs, and sending the information to the full connection layer for classification.

Suppose that the HM-R-CNN receives an input matrix of dimensions $d_1 \times d_2 \times d_3$, encompassing RGB color channels. According to the hierarchical principle, a given image $\boldsymbol{x} = \{x_1, x_2 \ldots, x_i \ldots, x_j \ldots, x_n\}$ was divided into several regions, and different characters in different divided regions were distinguished. For example, in image $\boldsymbol{x} = \{x_1, x_2 \ldots, x_i \ldots, x_j \ldots, x_n\}$, there are two target elements of x_i and x_j. This image will take two different targets as the center and divide the image into two independent regions $r_1 = \{w_i \ldots, t_1 \ldots, w_{h+i-1}\}$ and $r_2 = \{w_j \ldots, t_2 \ldots, w_{h+j-1}\}$ with size h. Thus the convolutional computation is streaming into two channels, and Eq. (10) is reconstructed as

$$CON_r = f(r_1\boldsymbol{x} + b_1) \otimes f(r_2\boldsymbol{x} + b_2), \tag{10}$$

where CON_r is the marked feature extracted by the combined convolution of the two hierarchical channels, b_1 and b_2 are the constant intercept in network regressions.

By dividing regions, the convolution layer yields outputs in three types of class, location, and regional features, with softmax constituting an additional layer. Conversely, the randomized mask R-CNN architecture comprises c layers and possesses a significantly reduced parameter count, in comparison to the randomized

Masking size	k	k_x	k_y	v	$\frac{\Delta v}{v}$
32×32	4	0.9	0.75	122	0.172
64×64	8	0.9	0.75	135	0.267
128×64	12	1.2	1.00	135	0.400
256×128	24	1.2	1.00	144	0.542
256×256	36	1.5	1.25	144	0.667

Table 1. Convergence velocity of RBD masking in slope metrics.

mask R-CNN configuration. The inception module concurrently gathers a suite of features through parallel $p \times p$ convolutions (usually applied by 3×3, 5×5, and 7×7), along with a parallel maximum pooling layer. To ensure practicality of the implementation, a $p \times p$ convolution is employed with the intent of diminishing the dimensionality. There the function was computed by mapping from a $d_1 \times d_2 \times d_3$ data to a $p \times p$ result at each layer. Next, the final fully connected layer encompasses the feature outputs, which correspond to the total training effect of all different regions, in dimension agreement with the number of categories in the dataset, and compiled with a rejection by a dropout ratio of 0.2, thus to convey the feature information to the softmax layer. Technically, the hierarchical mask R-CNN has been presented and its code released in GitHub website[41].

With specific designs in our adapted version of the HM-R-CNN, the image is divided into regions with specific objects as the center. The network is composed of two convolution layers measuring $256 \times 256 \times 3$ and $256 \times 128 \times 3$, two pooling layers in dimensions of $128 \times 64 \times 3$ and $64 \times 64 \times 3$, one fully connected layer, and incorporate a softmax unit as linear or nonlinear activation for feature output.

According to the principle of computational complexity estimation, when the mask architecture is constructed by a total of c layers for convolution, pooling flattening and in the fully connected perception, the computational load would be estimated up to $O(d_1 \times d_2 \times d_3 \times c \times l/(p \times p))$ confronting a convolution filter length of l. For example, when confronting the input image with the dimension of $256 \times 256 \times 3$, it is valued as $d_1 = 256$, $d_2 = 256$ and $d_3 = 3$. The convolutional network is designed with 14 layers for convolution and pooling calculation, i.e. $c = 14$. And the filter length is set as $l = 20$, aiming to output a feature in dimension of 3×3 ($p = 3$). In this case, the computation load is of 6 million digit-level operations. On the contrary, if using conventional methods for extracting the features without mask, the computational load will go stead at $O(d_1 \times d_2 \times d_3 \times c \times l)$. The example computation load is up to 55 million. When confronting the dynamic big data situation, the input dimension is changing, the convolutional network structure is diverse, thus leading to a large amount of computational work load for HM-R-CNN.

The mask-enhanced TbHM-R-CNN

The TbHM-R-CNN architecture is designed on the basic structure of R-CNN (see Fig. 2). The input image undergoes dynamic training for classification. The initial two convolution channels of the 256-pixel embedding space form the input image. This is followed by two successive pooling layers measuring 128×64 and 64×64 . The pooling features are available for mask selection by HM technique. Notably, the formatted three-variable RGB color consistency is maintained across the provided examples (shown as the " $\times 3$" in each step through the procedure). The special design of TbHM is to introduce a threshold value to debug the training in the convolution layers and the pooling layers. The crux of the threshold learning process lies in fault-tolerant for screening the regional masks. Then the selected regional structured data is characterized for flattening to finish the R-CNN procedure, and the selected feature data is further delivered for modeling to reach the goal of classification or prediction.

In the procedure, the segmented regional pixels represent a smaller energy value, which correlates with a reduced gradient value change (i.e. the abovementioned $\frac{\Delta v}{v}$ in Eq. (8)) in the convolutional processes. Prior to the application of a filter layer to the input image, the present input set al.ongside its weight values, as well as the projected output values and the ensuing layer's weights, are preserved as a fuzzy determinant data for the impending filtration operation[42]. Then the gradient values can be recovered during the latter stages of the pooling layer's threshold debugging. Subsequent regional mask layer will exhibit the diagonal or horizontal coding patterns due to the threshold factor.

The TbHM-R-CNN model is originated from HM-R-CNN, and a threshold parameter is introduced to make a refinement. As abovementioned, Eq. (10) is the characterized calculus of HM-R-CNN; to this end, the

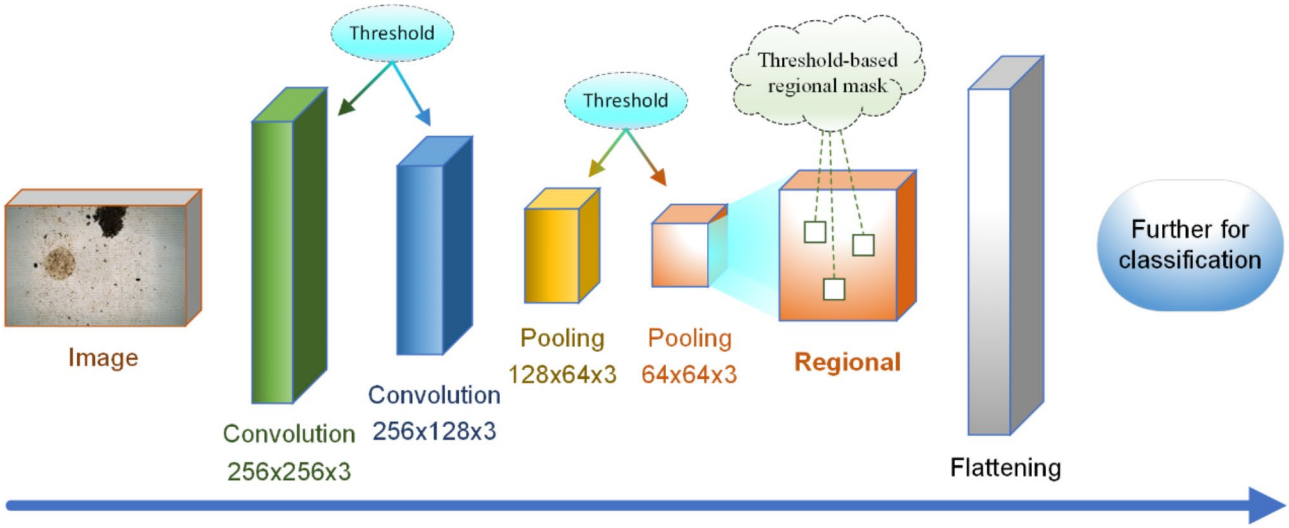

Fig. 2. Image classifier by regional segmentation based on specific design of TbHM-R-CNN model.

convolutional calculation of TbHM-R-CNN uses Eq. (11) instead of Eq. (10), while the other steps are the same as those of HM-R-CNN,

$$CON_r^{(\theta)} = f\left((r_1 + \theta_1)\boldsymbol{x} + b_1\right) \otimes f\left((r_2 + \theta_2)\boldsymbol{x} + b_2\right), \tag{11}$$

where $\theta_1 = (\theta_{c1}, \theta_{p1})$ and $\theta_2 = (\theta_{c2}, \theta_{p2})$ represent the thresholds applied in the two convolution layers and the two pooling layers, $CON_r^{(\theta)}$ is the renewal output features from the combined convolution of two hierarchical channels, with the adjustment of the threshold-based factor.

Indicators for model evaluation

Classification problems need to be quantitatively evaluated using the following typical indicators: Accuracy (ACC), Precision (PRE), Recall (REC) and F1 score[43]. They are calculated on the basis of a confusion matrix. For a binary classification problem, the confusion matrix provides the counts for True Positive (*TP*), False Positive (*FP*), True Negative (*TN*) and False Negative (*FN*). Then the mentioned indicators are formulated as follows,

$$\mathrm{ACC} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{12}$$

$$\mathrm{PRE} = \frac{TP}{TP + FP}, \tag{13}$$

$$\mathrm{REC} = \frac{TP}{TP + FN}, \tag{14}$$

$$\mathrm{F1} = \frac{2 \times \mathrm{PRE} \times \mathrm{REC}}{\mathrm{PRE} + \mathrm{REC}}. \tag{15}$$

When facing a multi-classification problem, weighted-average method is prevalent[44]. By considering the category imbalance, different weights are given to different categories (the weights are determined based on the true distribution ratio of the category), and each category is multiplied by the weights and then summed.

Application case

The dataset

As it is well known that pests and diseases are serious in crops in early March, around the time of Insects Awaken (One of the 24 solar terms in Chinese traditional sayings). Samples were collected in five consecutive sunny days at the Insects Awaken duration. Images were taken in the afternoon of each day between the hours of 15:00 and 17:00. A total of 3015 images of pest invasive plant leaf samples were collected from the 10 different kinds of agricultural plants. Each kind occupies a certain percentage of the total number (See Table 2). According to plant subjects, they were assigned for a total of 28 class labels. Each class label indicates a crop-disease pair and a test was performed to predict the crop-disease pair using only the plant leaf image. For the experiment, all the leaves were laid under a microscope to show the insect images. Technically, the images have been reduced to 432×288 pixels in all of the approaches described in this research (see examples in Fig. 3). Both model optimization and predictions were performed on these downscaled images.

In our experiments, various segmentation techniques, including SLIC, Mask R-CNN, HM-R-CNN, and TbHM-R-CNN, were employed on a specific collection of diseased leaves. The dataset was in RGB color format, with the test data comprising multiple fresh images. All experiments were transitioned to a version of the plant disease dataset where the leaves were segmented, and all superfluous background information was eliminated.

Common name	Scientific name / Latin name	No. of samples (percentage occupation)		No. of class labels
Cassava	Manihot esculenta Crantz	181	(6%)	2
Cunninghamia	Cunninghamia lanceolata (Lamb.) Hook.	241	(8%)	3
Ficus	Ficus microcarpa L. f.	452	(15%)	3
Litchi	Litchi chinensis Sonn.	362	(12%)	4
Orange	Citrus reticulata Blanco	452	(15%)	4
Rice plant	Oryza sativa L.	302	(10%)	3
Shrubs	Shrubs	241	(8%)	2
Sugar cane	Saccharum officinarum L	362	(12%)	3
Sweet potato	Ipomoea batatas (L.) Lam.	302	(10%)	3
Yam Bean	Pachyrhizus erosus (Linn.) Urb.	120	(4%)	1

Table 2. Names, class labels and the collecting numbers of the crop plant leaf samples.

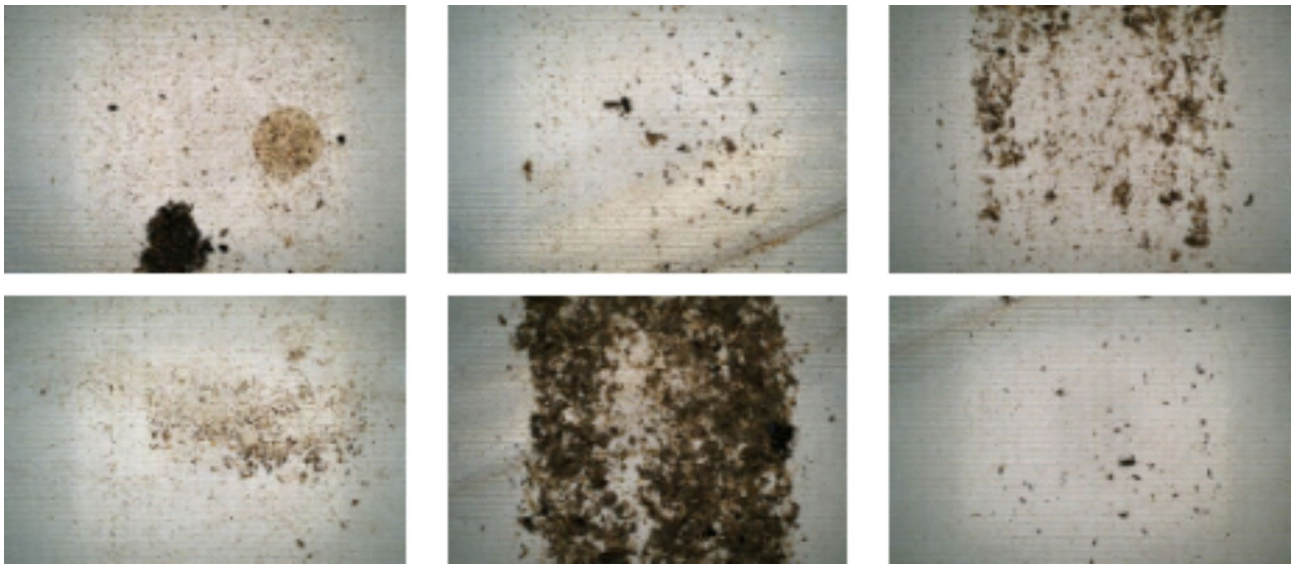

Fig. 3. The example of the plant leaf samples under the microscope view.

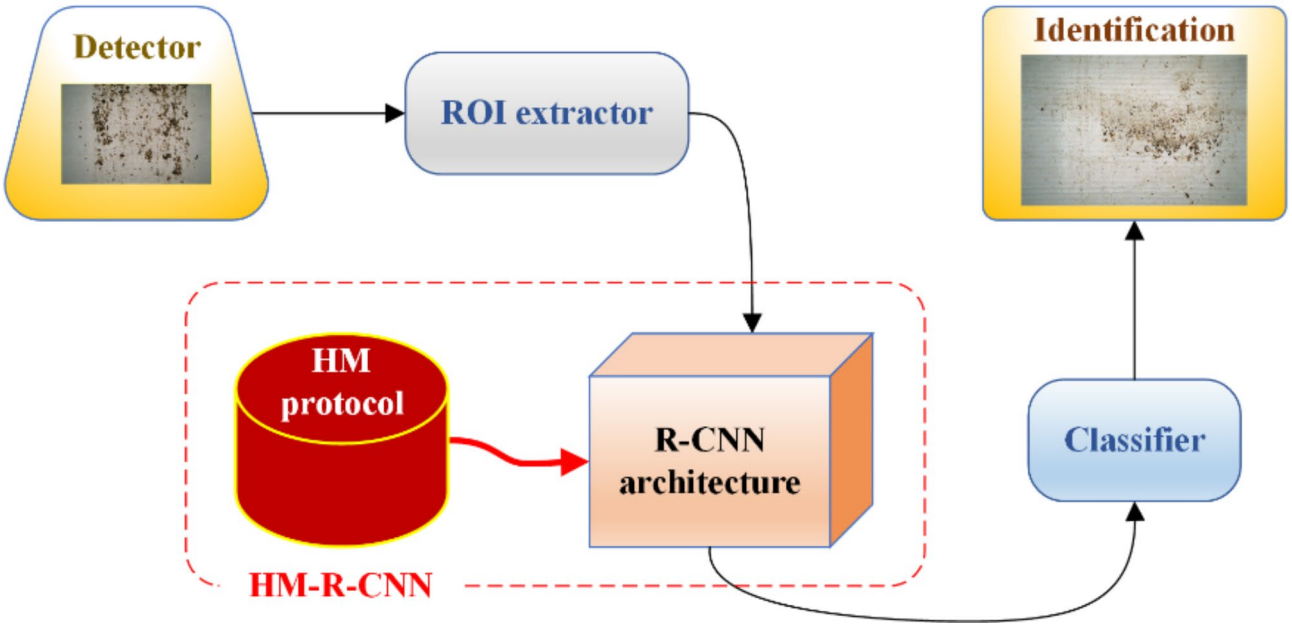

Fig. 4. The HM-R-CNN classifier for pest detection.

Architecture	Feature extractor	Bounding box
R-CNN	ResNet-50	Endorse the temporary parameter valuing of $[W_s, h_N]$
Mask R-CNN	Radial bisymmetric divergence	
HM-R-CNN	Hierarchical mask	
TbHM-R-CNN	Hierarchical mask + Threshold	

Table 3. Details for deep learning by different R-CNN-based architectures.

This step was crucial because the standardized data collection procedure could potentially introduce undesirable variability into the dataset.

The computing code was programed in MATLAB R2022a to execute the automated segmentation of the dataset, and performed it on a 64-bit operating system using an Intel(R) Pentium(R) G3220 CPU with 3 GHz and 8 GB of RAM. Our method involves generating an overlay of images by dissecting the brightness, and saturation elements across various sections of the image within the microscope domain. An additional phase in the process enables us to swiftly correct any dormant images that are especially conspicuous within certain subsets of the dataset, thereby eradicating another source of systematic error. In data analysis and machine learning, the concealed or imper latent features or patterns pertain to the concealed or imperceptible dimensions within the data. Latent images encapsulate intricate details about focal points of interest, potential zones of heightened pest or disease incidence, or other pertinent characteristics that are imperative for subsequent stages in the workflow. To enhance the data's integrity and precision, rectifying the latent images involves identifying and mitigating issues associated with these occult patterns or systematic flaws present in subsets of the dataset.

Model configuration for data analysis

To train the neural network, a comprehensive dataset depicting the 3015 images of diseased leaves was amassed. The images served for model training and evaluation. 65% of the samples were for modle training and 35% for evaluation. The evaluation samples were not included in the model training process. The noise was expunged using preprocessing methods. Thereafter, feature extraction methodologies and machine learning protocols were implemented to facilitate the recognition of pests.

Several R-CNN-derived object recognition methodologies have been utilized; each customized to fulfill the prerequisites of regional localization and enhance the efficacy of instance-oriented segmentation. Extending on this concept, the forthcoming discourse will present an innovative technique for partitioning segmentation areas specifically for regions of leaves affected by disease. The mask R-CNN handles a mask yielding independently from classification and bounding box forecasting. With the assist of cutting-edge techniques, HM-R-CNN stands out as a streamlined hierarchical bisectional segmentation approach. The TbHM-R-CNN offers distinct outputs for each candidate object, which is more suitable in practical for class designations, boundary box adjustments, class characteristics and softmax distributions.

Figure 4 illustrates the proposed pest detection classifier using the HM-R-CNN architecture. In the basic R-CNN model, the image data was pre-treated by ROI to extract a region in dimension of $256 \times 256 \times 3$. The convolutional network was designed with 14 layers for convolution, pooling, flattening and for fully connected structure calculations. And the convolution filter was set as in length of 20. In application, different dimensions of mask (such as using the 3×3, 5×5, and 7×7 size convolutions) were applied to the data to form the HM-R-CNN models. Consulting to Fig. 4, it is easy to switch to the TbHM-R-CNN mode by suiting a threshold valuable in the HM protocol (the red cylinder in the figure).

In details, the basic R-CNN architecture is constructed with 2 convolution layers paired with 2 pooling layers. The convolution filters were randomly initialized as a 0–1 vector in the length of 20. The filter slid over the input vectors with a stride of 2. For every sliding pace, the local sum of 20 variables were calculated with the filter weights, to produce the convolved feature data. The filters are tunable during the computing process. The functions are used to calculate the output results of the model to obtain the image segmentation results of the targeted images, as shown in the following,

$$y = \mathrm{softmax}(\boldsymbol{W}_s \boldsymbol{h}_N + b) \tag{16}$$

where $\boldsymbol{W}_s$ is the weight matrix, b is the bias, $\boldsymbol{h}_N$ is the last neurons of hidden layer and output in for R-CNN. In training process, the back propagation training mode was employed to optimize the model, and through the whole connection layer of each output is refined by minimizing the model with regularization. The formula is given as follows,

$$loss = -\sum_{i \in D} \sum_{j \in C} \widehat{y}_i^j \log y_i^j + \lambda \|\theta\|^2 \tag{17}$$

where D is the data set of training set, C is the category of dataset, y is the prediction category of the targeted images, $\widehat{y}$ is the actual category, and $\lambda\|\theta\|^2$ is a regularization term.

After the classification by HM-R-CNN, relevant pixels in the direction of each region were dispersedly identified, to make the output images formed in nuanced bi-symmetry. The method demonstrates that a 4-neighborhood cluster can effectively cluster equivalent sets of pixels, even along arbitrary directions. Then segmentation was easy to operate by SLIC, and SLIC computes the distance from each cluster center to the pixels within a $2S \times 2S$ region exclusively. This process surpasses previous techniques in terms of rapidity, diminishes memory consumption, attains elevated regional precision, and enhances the efficiency of the segmentation algorithm.

To raise the model's applicable capabilities, transfer learning was integrated to the mask R-CNN architecture, and substitute the modulation layer within the convolutional architecture with a collection of features exhibiting varying processing speeds. The TbHM-R-CNN model was also built up by fitting a threshold for parameter tuning. Referring to Eq. (11), θ_1 and θ_2 are loaded in adaptive mode, aiming to tune the resultant output features adjust for the tolerance of 5% bias in classification accuracy.

The detail settings of structural parameters are shown in Table 3. The R-CNN model was constructed as in an ResNet-50 structure. Radial bisymmetric gradients were utilized for the mask R-CNN framework, and also for the HM and TbHM models. All the CNN architectures used the same setting of bounding box for

the endorsement of temporary valuing of the parameters $\boldsymbol{W}_s$ and $\boldsymbol{h}_N$. The comparative results of tuning parameters were temporary stored in the box, and a common softmax unit is typically applied as the terminal activation function to normalize the output of the neural network.

Performance metrics of proposed system

Different algorithms were applied to infected plant leaves, and a comparative analysis of the regional accuracy of each strategy was conducted. For R-CNN and mask R-CNN, the subsequent experimental configurations were executed employing the R-CNN paradigm. The instance segmentation technique was validated utilizing the proposed HM-R-CNN and TbHM-R-CNN methodologies depicted in Fig. 5. It can be seen from Fig. 5 that the R-CNN-based deep learning models are able to get stronger intensity than SLIC to recognize the pixel data. This would help enhance the ability to recognize the pixel's information. The experiments were conducted with a deep learning framework that incorporates a progressive bilateral gradient feature set.

The method known as TbHM is a threshold-based hierarchical mask R-CNN technique. TbHM-RCNN constitutes a more expeditious object detection algorithm. The TbHM-RCNN approach yields the image displayed in Fig. 6. This illustration showcases the solitary detection capability of the mask R-CNN technique, which is more specifically to pinpoint the feature information of pest infested regions. Based on these features, it is easy to further calculate the pest insect area of the leaf, in aid to estimate the severe extend of pest detection.

A variety of leaf-based pests and diseases have been discerned and quantified for classification purposes. To tackle analogous issues, models or memory networks that have been trained on a substantial dataset are referred to as pre-trained models. Characteristics from mask R-CNN and HM-R-CNN are amalgamated to construct a novel pre-trained model. With rooted in the mask, the transfer learning features substitute the modulation layer substitute the modulation layer within the convolutional stack with a velocity difference feature set, as elucidated in the following section. Table 4 illustrates that mask R-CNN surpasses HM-R-CNN and TbHM-R-CNN in precision when detecting the leaves under the microscopic view, requiring merely a single lens to photograph the foliage. Table 4 tells that, in general, the mask R-CNN technique is deemed satisfactory with all classification accuracy over 0.6. In comparison, the common mask R-CNN with radial bisymmetric divergence observed a basic line for classification, the HM-R-CNN model had got an enhanced effect, and the TbHM-R-CNN obtained the highest accuracy of over 0.7 for all the three views. Herein, it can be concluded that the single-shot technique of mask R-CNN is synergized with the training dataset of transfer learning.

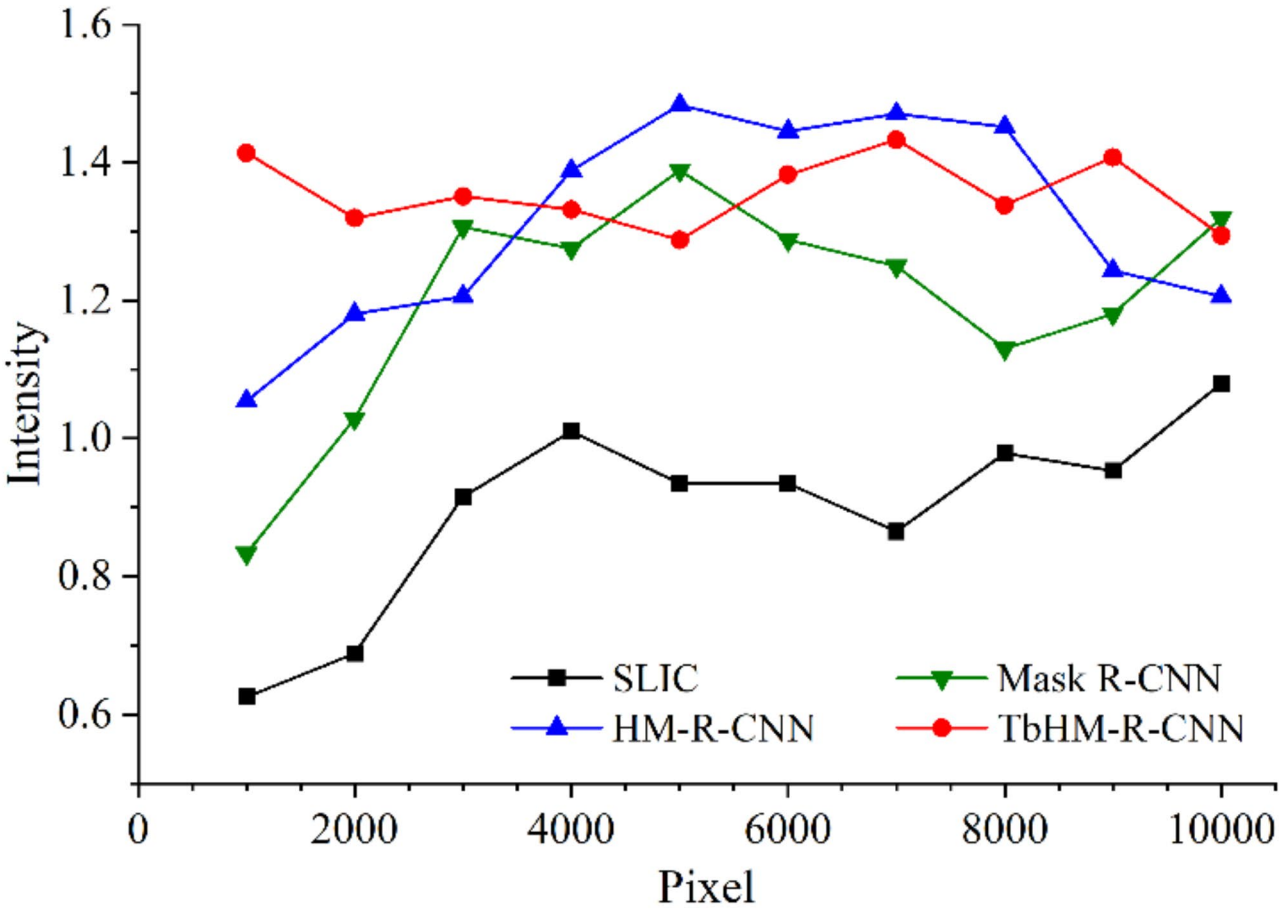

Fig. 5. Image segmentation intensity for different methods.

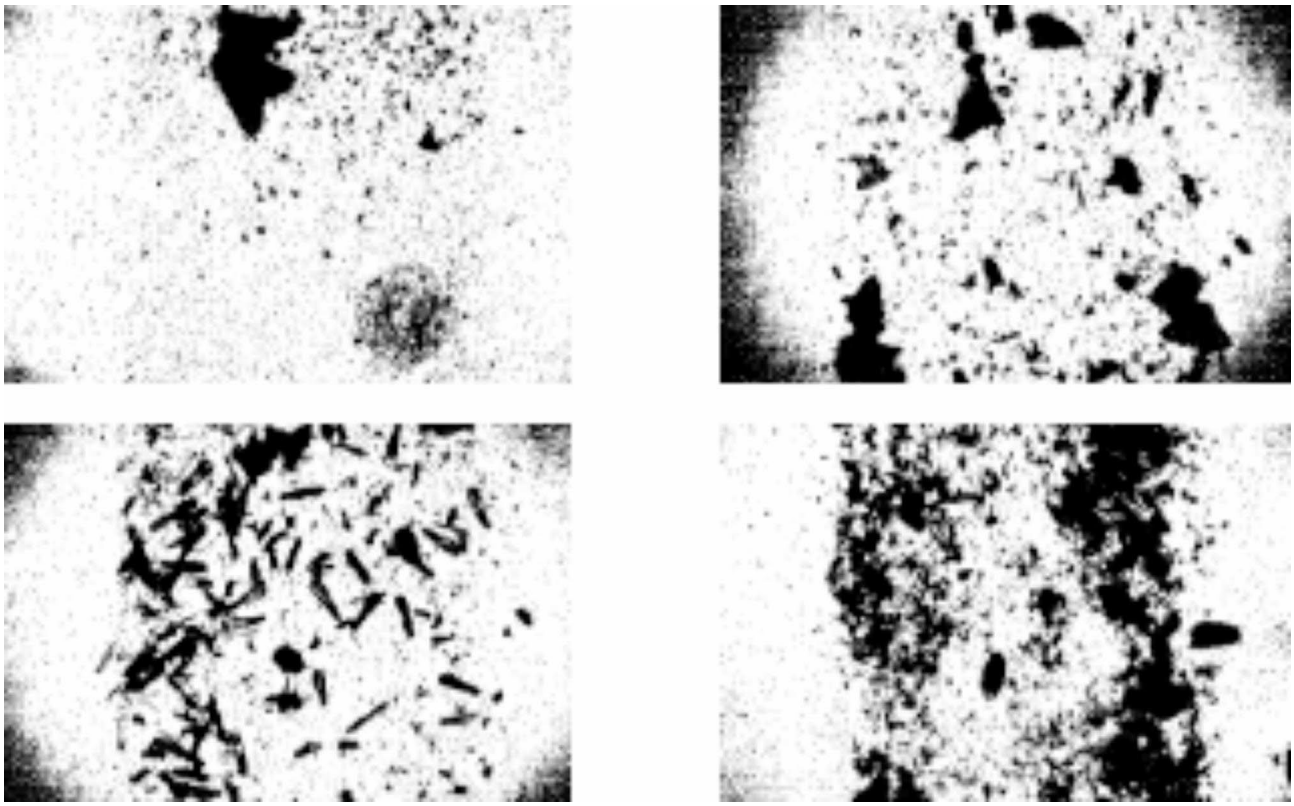

Fig. 6. Example images for feature converted by TbHM-RCNN technique.

	Mask *R*-CNN	HM-*R*-CNN	TbHM-*R*-CNN
Bacterial canker	0.6126	0.6807	**0.7109**
Gray mold	0.6875	0.7429	**0.8205**
Whitefly	0.6432	0.7135	**0.7408**

Table 4. Classification accuracy by different R-CNN-based techniques through three microscopic views.

With the mask, the R-CNN method applied to different regional areas of the image by rational segmentation. For the mask R-CNN, HM-R-CNN and TbHM-R-CNN algorithms, regional efficacy is assessed based on the number of successful points that correspond to actual regions. Pinpointing specific areas within an image segmentation where objects may occur and extracting the characteristics of these zones through CNN facilitates the prediction of their classes and bounding boxes, which will subsequently aid in further feature extraction. Figure 7 depicts the regional accuracy curves for segmentations based on different counts of pixels. The figure show that the regional search went in the top accuracy-to-pixel ratio when the count of pixels fall in around 0.1, which indicated that using about 10% of the pixels of the whole image can deduce a relatively high classification accuracy. When concerning on the comparison of methods, the TbHM-R-CNN model obviously performed prior than the other two models. These results validated that the regional mask which is screen out by radial bisymmetric divergence method is able to improve the signal-to-noise ratio for information selection, and that the TbHM is the comparatively best approach to aid the R-CNN architecture for feature selection in the deep learning progress for pest detection in crop leaves.

Comparison of the proposed model

The efficacy of the proposed method is validated by benchmarking it against the outcomes of conventional techniques such as ANN, SVM, KNN and Naïve Bayes[45]. Also, some prevalent deep learning methods are taken into the comparison, including conventional CNN and DNN-SAR[46], the ResNet-50, GoogleNet, and AlexNet structures[47]. Their respective performances were evaluated according to diverse evaluation metrics. Figure 8 exhibits the comparison of accuracy rates among various classification methodologies. The classification accuracies of conventional techniques such as ANN, SVM, KNN, Naive Bayes, and conventional CNN are recorded at 80.8%, 82.1%, 75.3%, 72.6%, and 88.3% respectively. In contrast, the proposed models achieve significantly higher classification accuracies, up to 91.5%, 94.1% and 96.2% for mask R-CNN, HM-R-CNN, TbHM-R-CNN, respectively, which surpass those of the existing methods. Table 5 presents a comparative

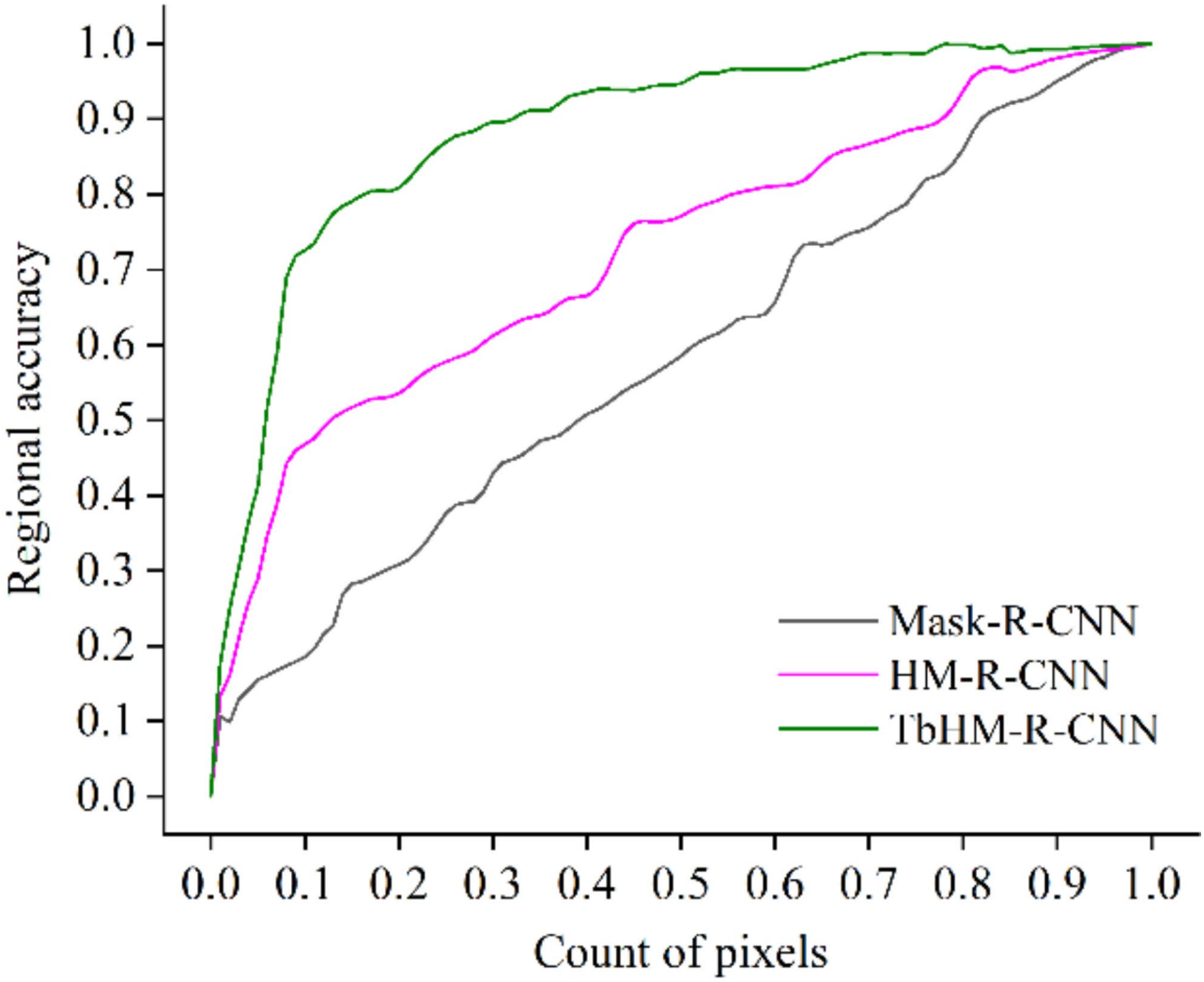

Fig. 7. Regional accuracy for different segmentation by different mask-R-CNN models and its corresponding applied variants/pixels.

analysis of precision, recall, and F1 score of the proposed models against existing models. Established models such as DNN-SAR, ResNet-50, GoogleNet, and AlexNet report precision values of 80.4%, 78.3%, 82.6%, and 79.9% respectively, whereas the proposed model achieves an enhanced precision value to 96.3%, 95.6% and 98.9% for Mask R-CNN, HM-R-CNN and TbHM-R-CNN, respectively. Additionally, the proposed models boast superior recall values compared to its counterparts, along with the F1 scores of over 95%.

In summary, a novel image segmentation method for leaf pest detection system was introduced, utilizing a segmentation approach founded on radial bisymmetric divergence R-CNN. This system demonstrates adept capabilities in detecting and classifying pests and diseases effectively. The outcomes reveal that the system achieves an impressively high accuracy of 96.2% in identifying and categorizing the infected areas of leaves, which is pivotal for safeguarding crops swiftly. The proposed model exhibits an outstanding recall value of 97.5 and an F1 score of 0.982, outperforming existing models. These metrics corroborate that the suggested system surpasses other current techniques in performance.

Conclusions

The early detection of pests and diseases in plants is crucial for preventing crop losses. In this study, prevalent leaf pests were successfully detected and classified using a bisymmetric hierarchical segmentation R-CNN based on deep learning techniques. Key feature sets included both the age and coloration of the leaves. Investigations have substantiated that the segmentation method employed is efficacious in discerning the infected areas of the blade, particularly at low velocity gradient values. The mask R-CNN based approach has proven to be more dependable and precise overall. While the SLIC-based segmentation method offers a robust localization effect, the outcomes indicate that it yields comparable efficiency results to the cluster-based method, provided there is an appropriate selection of clustering and training parameters.

The suggested technique employs radial bisymmetric gradients and transfer learning to achieve competitive results. Successful identification of infected areas on crop leaves was achieved by the HM-R-CNN and TbHM-R-

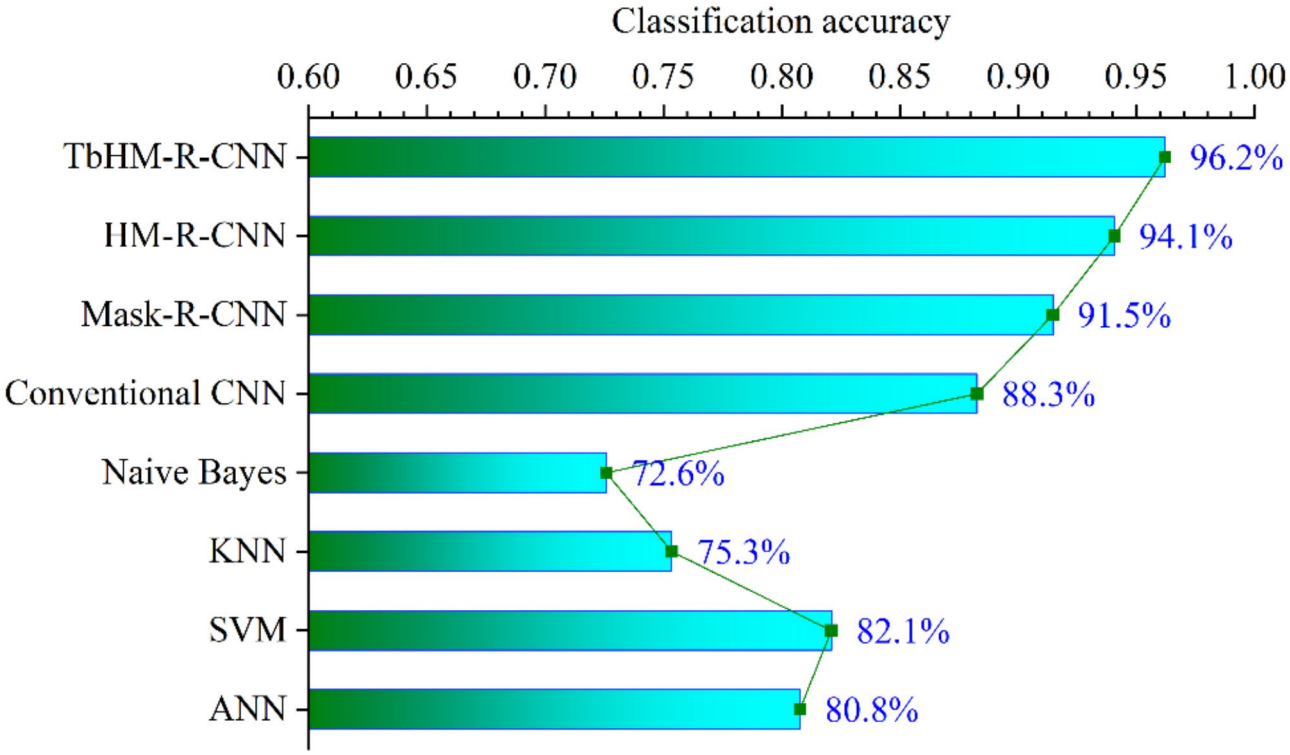

Fig. 8. Comparison of accuracy on the proposed methods vs. some existing classification techniques.

	Precision (PRE)	Recall (REC)	F1 score
DNN-SAR	80.4%	76.9%	0.730
ResNet-50	78.3%	68.4%	0.786
GoogleNet	82.6%	85.4%	0.840
AlexNet	79.9%	76.5%	0.782
Mask R-CNN	96.3%	94.6%	0.954
HM-R-CNN	95.6%	96.5%	0.960
TbHM-R-CNN	**98.9%**	**97.5%**	**0.982**

Table 5. Comparison of Precision, Recall, and F1 score for different CNN deep learning architectures.

CNN methods. In comparison with current methods, the proposed method boasts a highest accuracy of 96.2%, precision of 98.9%, recall of 97.5%, and an F1 score of 0.982. When purchasing the high accuracy, the established model is at the cost of model transfer stability. However, the pinpointing and identifying various regions or sites on the leaves that are compromised or afflicted by a disease is much related to the stability and robustness of the established model. To address this issue, feature-based hybrid detection methods can potentially be further ameliorated through the application of ensemble techniques (such as bagging, boosting, or stacking) and subsequently consolidating predictions from multiple models. The results derived from the ensemble approach prospectively surpass those yielded by an individual model. Such an approach can significantly aid in the development of automated pest control systems within the context of digital agriculture. This constitutes a critical following study in the management and detection of plant diseases.

Data availability

The datasets used during the current study available from the corresponding author on reasonable request.

Received: 9 April 2024; Accepted: 4 October 2024
Published online: 15 October 2024

References

1. Mohd Hanafiah, K. et al. Impact of Malaysian palm oil on sustainable development goals: co-benefits and trade-offs across mitigation strategies. *Sustain. Sci.* **17**, 1639–1661 (2022).

2. Pal, S., Samanta, S. & Banerjee, A. Seasonal incidence of pulse aphid (Aphis craccivora Koch.) and its natural enemies on field pea (Pisum sativum L.) in relation to some abiotic factors in Alluvial Zone of West Bengal. *Legume Res.* **46**, 1680–1685 (2023).
3. Kalpna, Hajam, Y. A. & Kumar, R. Management of stored grain pest with special reference to Callosobruchus maculatus, a major pest of cowpea: a review. *Heliyon*. **8**, e08703 (2022).
4. de Groot, Pocock, M., Bonte, M. J. O. & Fernandez-Conradi, J. Valdés-Correcher, E. Citizen science and monitoring forest pests: a beneficial alliance? *Curr. For. Rep.* **9**, 15–32 (2023).
5. Cossich, V. R. A., Carlgren, D., Holash, R. J. & Katz, L. Technological breakthroughs in sport: current practice and future potential of artificial intelligence, virtual reality, augmented reality, and modern data visualization in performance analysis. *Appl. Sci.* **13**, 12965 (2023).
6. Yassue, R. M., Galli, G., Fritsche-Neto, R. & Morota, G. Classification of plant growth-promoting bacteria inoculation status and prediction of growth-related traits in tropical maize using hyperspectral image and genomic data. *Crop Sci.* **63**, 88–100 (2023).
7. Wei, D. P., Chen, J. Q., Luo, T., Long, T. & Wang, H. B. Classification of crop pests based on multi-scale feature fusion. *Comput. Electron. Agric.* **194**, 106736 (2022).
8. Bjekic, M. et al. Wall segmentation in 2D images using convolutional neural networks. *PeerJ Comput. Sci.* **9**, 1–18 (2023).
9. Aljohani, M. et al. An automated metaheuristic-optimized approach for diagnosing and classifying brain tumors based on a convolutional neural network. *Results Eng.* **23**, 102459 (2024).
10. Apacionado, B. V. & Ahamed, T. Sooty mold detection on citrus tree canopy using deep learning algorithms. *Sensors. (Basel Switzerland)*. **23**, 8519 (2023).
11. Bischoff, V., Farias, K., Menzen, J. P. & Pessin, G. Technological support for detection and prediction of plant diseases: a systematic mapping study. *Comput. Electron. Agric.* **181**, 105922 (2021).
12. Chakrabarti, P., Sarkar, S. & Basu, P. Pesticide induced visual abnormalities in Asian honey bees (Apis cerana L.) in intensive agricultural landscapes. *Chemosphere*. **230**, 51–58 (2019).
13. Zou, B. et al. Comparison of direct and indirect determinations of dynamic ventilation rate in a modern dairy free stall barn. *Int. J. Agric. Biol. Eng.* **13**, 41–46 (2020).
14. Zhao, S., Sun, X. & Gai, L. Data enhancement and multi-feature learning model for pest classification. *J. Intell. Fuzzy Syst.* **45**, 5409–5421 (2023).
15. Taji, K. et al. An ensemble hybrid framework: a comparative analysis of metaheuristic algorithms for ensemble hybrid CNN features for plants disease classification. *IEEE Access.* **12**, 61886–61906 (2024).
16. Nagoor, S. & Jinny, S. V. A dual fuzzy with hybrid deep learning architecture based on CNN with hybrid metaheuristic algorithm for effective segmentation and classification. *Int. J. Inf. Technol.* **15**, 531–543 (2023).
17. Shafik, W., Tufail, A., Liyanage, C. D. S. & Apong, R. A. A. H. M. Using a novel convolutional neural network for plant pests detection and disease classification. *J. Sci. Food. Agric.* **103**, 5849–5861 (2023).
18. Arrechea-Castillo, D. A., Solano-Correa, Y. T., Muñoz-Ordóñez, J. F., Pencue-Fierro, E. L. & Figueroa-Casas, A. Multiclass land use and land cover classification of Andean sub-basins in Colombia with Sentinel-2 and deep learning. *Remote Sens.* **15**, 2521 (2023).
19. Rastogi, S., Kumari, V., Sharma, V. & Ahmad, F. J. Colorimetric detection of oxytocin in bottle gourd using cysteamine functionalized gold nanoparticle (Cys-AuNPs). *Food. Anal. Methods*. **15**, 2972–2983 (2022).
20. Sulistyo, S. B., Wu, D., Woo, W. L., Dlay, S. S. & Gao, B. Computational deep intelligence vision sensing for nutrient content estimation in agricultural automation. *IEEE Trans. Autom. Sci. Eng.* **15**, 1243–1257 (2018).
21. Elshazly, E. H. et al. Identifying the anti-MERS-CoV and anti-HcoV-229E potential drugs from the Ginkgo biloba leaves extract and its eco-friendly synthesis of silver nanoparticles. *Molecules*. **28**, 1375 (2023).
22. Napolitano, C. et al. Synthesis and insecticidal activity of N-(5-phenylpyrazin-2-yl)-benzamide derivatives: elucidation of mode of action on chitin biosynthesis through symptomology and genetic studies. *Pestic. Biochem. Physiol.* **199**, 105771 (2024).
23. Saeed, F. et al. A robust approach for industrial small-object detection using an improved faster regional convolutional neural network. *Sci. Rep.* **11**, 23390 (2021).
24. Bukumira, M. et al. Carrot grading system using computer vision feature parameters and a cascaded graph convolutional neural network. *J. Electron. Imaging.* **31**, 061815 (2022).
25. Ni, L. L. et al. Monthly precipitation prediction at regional scale using deep convolutional neural networks. *Hydrol. Process.* **37**, e14954 (2023).
26. Prasath, B. & Akila, M. IoT-based pest detection and classification using deep features with enhanced deep learning strategies. *Eng. Appl. Artif. Intell.* **121**, 105985 (2023).
27. Cai, J. Z., Wu, J., Guo, C. L., Cai, S. H. & Cai, C. B. Ultrafast multi-slice chemical exchange saturation transfer imaging scheme based on segmented spatiotemporal encoding. *Magn. Reson. Imaging.* **60**, 122–129 (2019).
28. Qiao, Y., Jiao, L. & Hou, B. High-quality depth up-sampling based on multi-scale SLIC. *Electron. Lett.* **54**, 494–496 (2018).
29. Liu, W. P., Wang, J. F. & Chen, G. N. Research on image segmentation of multiphoton microscopic cell image based on SLIC super-pixels method. *Basic. Clin. Pharmacol. Toxicol.* **125**, 28 (2019).
30. Bommisetty, R. M., Prakash, O. & Khare, A. Video superpixels generation through integration of curvelet transform and simple linear iterative clustering. *Multimed. Tools Appl.* **78**, 25185–25219 (2019).
31. Liu, Y. Y. et al. Identification and diagnosis of mammographic malignant architectural distortion using a deep learning based mask regional convolutional neural network. *Front. Oncol.* **13**, 1119743 (2023).
32. Mei, Y., Dai, Y. T. & Lei, Y. The influence of hierarchical masks on masked repetition priming: evidence from event-related potential investigation. *Front. Hum. Neurosci.* **13**, 70 (2019).
33. Wang, B. Q. & Yang, F. L. Lightweight and privacy-preserving hierarchical federated learning mechanism for artificial intelligence-generated image content. *J. Real-Time Image Process.* **21**, 149 (2024).
34. Habiyaremye, A. & Korina, L. Indigenous knowledge systems in ecological pest control and post-harvest rice conservation techniques: sustainability lessons from Baduy communities. *Sustain*. **13**, 9148 (2021).
35. Zhang, Z. Y., Zhao, P., Jiang, Y. & Zhou, Z. H. Learning from incomplete and inaccurate supervision. *IEEE Trans. Knowl. Data Eng.* **34**, 5854–5868 (2022).
36. Ahadullah, M., Sapar, S. H., Al-Saidi, N. M. G. & Said, M. R. M. Competitive improvement of the time complexity to encode fractal image: by applying symmetric central pixel of the block. *IEEE Access.* **9**, 5028–5045 (2021).
37. Bakkouri, S. & Elyousfi, A. An adaptive CU size decision algorithm based on gradient boosting machines for 3D-HEVC inter-coding. *Multimed. Tools Appl.* **82**, 32539–32557 (2023).
38. Bhujade, V. G., Sambhe, V. & Banerjee, B. Digital image noise removal towards soybean and cotton plant disease using image processing filters. *Expert Syst. Appl.* **246**, 123031 (2024).
39. Elumalai, G. & Ganesan, M. Deep learning based hand wrist segmentation using Mask R-CNN. *Int. Arab. J. Inform. Technol.* **19**, 785–792 (2022).
40. GitHub. hlydecker/hierarchical-rcnn (Public). https://github.com/hlydecker/hierarchical-rcnn
41. Nakach, F. Z., Zerouaoui, H. & Idri, A. Binary classification of multi-magnification histopathological breast cancer images using late fusion and transfer learning. *Data Technol. Appl.* **57**, 668–695 (2023).
42. Song, X. F. et al. A clinical decision model based on machine learning for ptosis. *BMC Ophthalmol.* **21**, 169 (2021).
43. Karanikola, A., Liapis, C. M. & Kotsiantis, S. Investigating cluster validation metrics for optimal number of clusters determination. *Intell. Decis. Technol.* **15**, 809–824 (2021).

44. Saleem, M. H., Potgieter, J. & Arif, K. M. Automation in Agriculture by Machine and Deep Learning techniques: a review of recent developments. *Precis. Agric.* **22**, 2053–2091 (2021).
45. Mukhopadhyay, A. K., Majumder, S. & Chakrabarti, I. Systematic realization of a fully connected deep and convolutional neural network architecture on a field programmable gate array. *Comput. Electr. Eng.* **97**, 107628 (2022).
46. Kuo, J. K., Wu, J. J., Huang, P. H. & Cheng, C. Y. Inspection of sandblasting defect in investment castings by deep convolutional neural network. *Int. J. Adv. Manuf. Technol.* **120**, 2457–2468 (2022).
47. Maeda-Gutiérrez, V. et al. Comparison of convolutional neural network architectures for classification of tomatoplant diseases. *Appl. Sci.* **10**, 1245 (2020).

Acknowledgements

This work was supported by Special project in key areas of ordinary universities and colleges in Guangdong province (2023ZDZX4069) and Guangdong provincial education science planning project (2022GXJK378).

Author contributions

Xianchuan Wu: methodology, investigation, validation, data assessment and curation, writing: original draft review and editing. Yuling Liu: data assessment and curation, formal analysis, methodology, writing: draft review and editing of the manuscript. Mingjing Xing: conceptualization, methodology, formal analysis, validation. Chun Yang: investigation, participant management, data assessment and curation. Shaoyong Hong: conceptualization, methodology, supervision, investigation, participant management, data assessment and curation, project administration, writing: original draft review and editing of the manuscript. All authors read and approved the final manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.H.