Classification and Identification of Tea Diseases Based on Improved YOLO V7 Model of MobileNeXt

Yuxin Xia
Yunnan Agricultural University
Wenxia Yuan
Yunnan Agricultural University
Shihao Zhang
Yunnan Agricultural University
Qiaomei Wang
Yunnan Agricultural University
Xiaohui Liu
Yunnan Agricultural University
Houqiao Wang
Yunnan Agricultural University
Yamin Wu
Yunnan Agricultural University
Chunhua Yang
Yunnan Agricultural University
JiaYi Xu
Yunnan Agricultural University
Lei Li
Yunnan Agricultural University
Junjie He
Yunnan Agricultural University
Zhiyong Cao
Yunnan Agricultural University
Zejun Wang
Yunnan Agricultural University
Zihua Zhao
Yunnan Agricultural University
Baijuan Wang (wangbaijuan2023@163.com)
Yunnan Agricultural University

Article

Keywords:

Posted Date: November 1st, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3494258/v1

Additional Declarations: No competing interests reported.

Classification and Identification of Tea Diseases Based on Improved YOLO V7 Model of MobileNeXt

Yuxin Xia[1], Wenxia Yuan[2], Shihao Zhang[1], Qiaomei Wang[2], Xiaohui Liu[2], Houqiao Wang[2], Yamin Wu[2], Chunhua Yang[2], Jiayi Xu[2], Lei Li[2], Junjie He[2], Zhiyong Cao[3], Zejun Wang[2], Zihua Zhao[2], Baijuan Wang[2,*]

1 College of Mechanical and Electrical Engineering, Yunnan Agricultural University, Kunming 650201, China;
2 College of Tea Science, Yunnan Agricultural University, Kunming 650201, China;
3 College of Information Engineering, Yunnan Agricultural University, Kunming 650201, China;
* Corresponding: wangbaijuan2023@163.com (B.W.).

To overcome the constraints associated with conventional approaches used in the classification and detection of tea diseases, which are characterized by their limited accuracy and sluggish responsiveness, this study introduces an enhanced YOLOv7 lightweight model algorithm integrated with MobileNeXt. This refinement not only bolsters the model's capacity for extracting and processing features but also effectively lightens the computational load, expedites recognition, and integrates a dual-layer routing attention mechanism visual converter to enhance the capture of crucial details and textures within disease images. Consequently, these enhancements lead to improved model performance and computational efficiency, ensuring precise and rapid identification of tea diseases. Furthermore, this model incorporates the more appropriate SIoU as the loss function, mitigating losses, minimizing omissions, and reducing misclassifications, thus resulting in superior recognition, even in complex image backgrounds. Based on the training outcomes, the enhanced model attains Precision, Recall and mean Average Precision scores of 93.5%, 89.9%, and 92.1%, respectively, marking substantial enhancements of 5.06%, 2.16%, and 2.91% compared to the original YOLOv7 model. Additionally, the model's size is reduced by 19.12%, and its detection speed accelerates by 11.13%. This improved model excels in accurately and expediting.

Tea culture boasts a rich historical legacy and experiences substantial global demand. In line with the enhancement of public health awareness[1], tea has gained popularity due to its inherent antioxidant attributes and the potential health advantages it offers. Abundant in polyphenols and antioxidants, tea contributes to the neutralization of free radicals, mitigates oxidative stress, and promotes cellular well-being[2].Nonetheless, the growth of tea trees is often marred by the frequent occurrence of diseases, which exert a considerable detrimental impact on the tea industry[3]. This adversity extends beyond the mere reduction in yield and quality, as it also inflicts severe economic losses upon tea farmers and producers[4].

Yunnan has emerged as a Chinese province characterized by remarkable biodiversity, owing to its diverse climatic and topographical features. This unique setting has fostered the development of a distinct plateau mountain tea industry[5]. However, the long-standing warm and humid climate in the region has created an environment conducive to the proliferation of tea diseases. These diseases manifest in diverse forms, presenting considerable challenges in terms of prevention and control, ultimately leading to diminished tea yield and quality[6]. To ensure effective cultivation management and the sustainable development of Yunnan's tea gardens, the timely and accurate identification of tea diseases has become an imperative task to mitigate losses.

In the realm of traditional target detection algorithms, reliance primarily rests on the analysis of color, texture, and related attributes[7]. Such methods predominantly encapsulate surface-level characteristics while falling short in their capacity to effectively extract more profound features. This limitation becomes notably pronounced in intricate scenes, giving rise to issues pertaining to diminished robustness and limited generalization, thus impeding the ability to fulfill the precise disease identification requirements.

With the advancement of computer technology, the utilization of machine learning and image processing technology has become a significant method for achieving automatic detection and identification of crop diseases, leading to the automation of disease diagnosis in crops[8-10]. Maria et al.[11] employed machine learning and transfer learning methods to recognize diseases in cauliflower. By comparing various transfer learning frameworks, they achieved the best recognition accuracy of 90.08%. Pallathadka et al.[12] employed PCA for feature extraction and histogram equalization for preprocessing to achieve automated rice leaf disease detection using a machine learning-based approach with a multiclass support vector machine, naive Bayes, and CNN. Javidan et al.[13] employed novel image processing algorithms and SVM for the diagnosis and classification of grape leaf diseases, including black measles, black rot, and leaf blight. Sun et al.[14] introduced a novel algorithm that combines SLIC with SVM to improve the extraction of significant tea leaf disease images within complex backgrounds. This approach accomplishes the extraction of significant tea leaf disease images. Das et al.[15] applied SVM, Random Forests, and Logistic Regression to classify various types of leaf diseases, aiming to identify different types of leaf diseases. However, conventional image feature extraction often relies on manual design, limiting its applicability and generalization. This poses challenges for practical use and comprehensive pest detection research in various natural settings. Hence, further research on image-based pest detection technology is of paramount importance.

In recent years, with the rapid development of deep learning technology, it has been widely applied in various fields, including autonomous driving[15, 16], fault detection[17, 18],medical decision making[19-21]and agriculture[22, 23]. Soeb et al.[24] proposed an AI-based solution for tea disease detection by creating their own tea leaf dataset. They utilized the YOLOv7 network to achieve rapid recognition and detection of tea diseases. Zhang et al.[25] employed an edge device detection method for one bud and two leaves of tea based on ShuffleNetv2-YOLOv5-Lite-E. This approach achieved model reduction and maintained accuracy by replacing a lighter-weight backbone network and channel pruning. Latha et al.[26] introduced a convolutional neural network model that utilized convolutional layers to extract features from input images in the dataset. They achieved the recognition of eight categories of tea leaf diseases and pests by using the output layer[27].

In light of the aforementioned challenges, this paper is dedicated to the investigation of tea disease identification within intricate settings. The study aims to enhance the original YOLOv7 network model to address these issues. First, the incorporation of MobileNeXt inverted residual blocks[28, 29] into the YOLOv7 backbone network is introduced to optimize the model. This optimization serves to diminish model parameters and facilitate a lightweight model improvement. Subsequently, in pursuit of flexible computing resource allocation and heightened content awareness, the visual converter of a dual-layer routing attention visual converter[30, 31] was introduced to bolster the efficiency of tea disease recognition within complex environments. Last, the loss function is replaced with the more appropriate SIoU[32] to reduce regression loss and expedite model convergence. These endeavors seek to enhance the classification and identification of tea diseases[33], specifically tea bituminous coal disease, zonate spot, anthracnose, and leaf blight. The insights derived from this study aim to provide valuable guidance for the construction of a recognition model better suited for the context of tea diseases within a natural environment.

Methods

Tea disease image dataset

The image dataset employed in this study for tea disease analysis was procured from tea gardens located at Yunnan Agricultural University in Kunming, Yunnan Province, and the Hekai tea garden base in Menghai County, Xishuangbanna. These images were captured during two distinct periods, namely mid-November 2022 and mid-June 2023. To ensure the richness and diversity of the dataset for robust model performance in target detection, data acquisition was carried out under varying light intensities and weather conditions. Notably, the images were acquired within natural environmental conditions, without imposing restrictions on imaging distances, employing an OPPO R15 mobile phone as the acquisition device. In total, 1823 image datasets were amassed, featuring a resolution of 4608 pixels by 3456 pixels, and were originally stored in JPG format.

During the collection of tea disease data, the four diseases, namely bituminous coal disease, anthracnose, zonate spot, and leaf blight, were distinctly classified, and redundant data were systematically removed. This refinement process resulted in the selection of 231, 293, 145, and 187 high-quality initial disease datasets for each respective condition. To mitigate the risk of over-fitting during the model training process, attributed to the limited sample size within the image dataset, this study judiciously employed data augmentation techniques, including enhancement color, flipping, random rotation and enhancement contrast as shown in Fig.1. These strategies were systematically implemented to expand the dataset in a balanced and reasoned manner. Following these enhancements, the total number of images for the four types of diseases reached a substantial count of 4280.

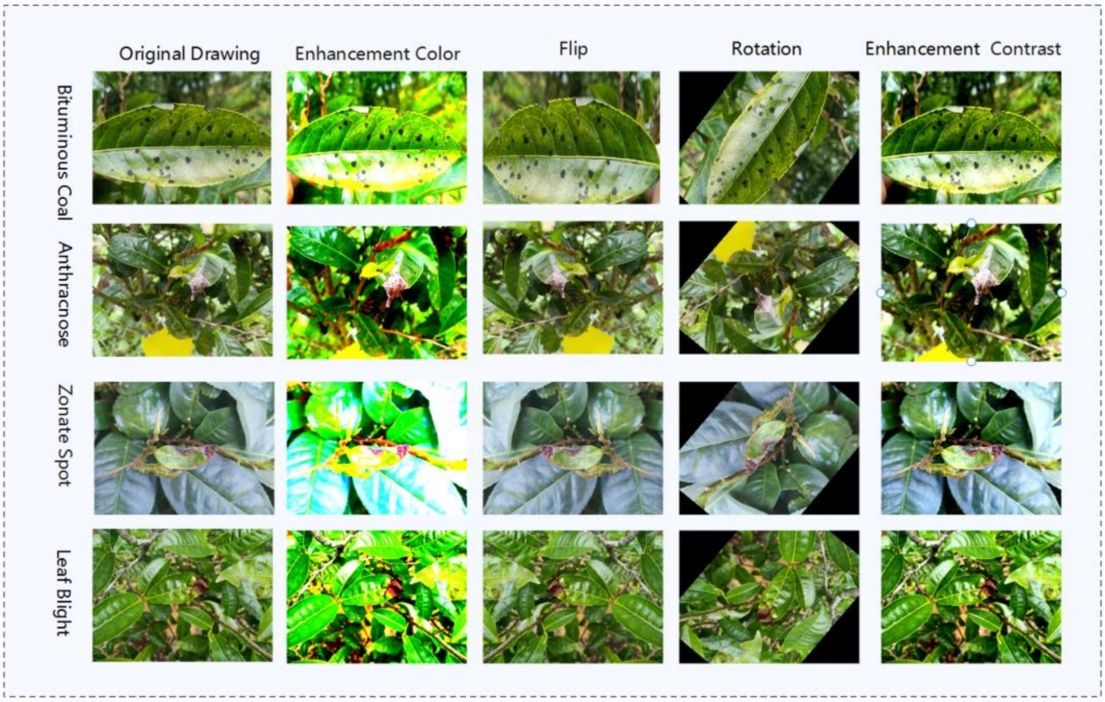

Figure 1. Data enhancement contrast diagram

In this investigation, the labeling of tea disease images was meticulously executed using the LabelImg manual annotation tool, and the resultant labels were subsequently organized into XML and TXT file datasets. Notably, throughout the labeling process, emphasis was placed on positioning the tea disease segments within the center of the bounding boxes. The aforementioned datasets were subsequently partitioned into training, testing, and validation sets following a ratio of 6:2:2. This partitioning resulted in a training set comprising 2568 samples, a test set consisting of 856 samples, and an equally sized validation set of 856 samples. These datasets were instrumental in the training and evaluation of the model in this study.

Improved YOLOv7 algorithm

The YOLOv7 network[34] architecture is delineated into four primary components: Input, Backbone, Neck, and Head. By reformulating the target detection challenge as a regression problem, this approach enhances efficiency while preserving the precision of object detection[35]. The initial input image undergoes a normalization process to conform to the model's input specifications. Subsequently, feature extraction is performed through the backbone network. Finally, the extracted feature map is processed to yield predicted bounding box positions and category information[36], thereby improving the model's capacity to capture objects of varying sizes .

MobileNeXt Lightweight network

To address the challenges posed by the substantial size, parameter count, and demanding hardware requirements of the original YOLOv7 network model, this study introduces a lightweight MobileNeXt structure to enhance the YOLOv7 backbone network. The architectural representation of MobileNeXt is consists of a 3×3 convolution layer, eight SandGlass[37] Block modules, an average pooling layer, and a 1×1 convolution layer. The feature extraction process begins with the initial extraction of shallow features using a 3×3 ordinary convolution. Subsequently, eight SandGlass Block modules are sequentially connected to construct a deep-level disease feature extraction network. The inclusion of an average pooling layer compresses the two-dimensional feature image into one-dimensional form. Lastly, the extracted features undergo further refinement and fusion, resulting in improved image classification performance, a reduction in computational demands and model parameters, and a consequent decrease in model complexity and calculation time.

The SandGlass Block represents an hourglass residual module, characterized by its distinctive placement of depthwise convolutions at both the outset and conclusion. This architectural choice is deliberate and seeks to maintain the spatial depth convolution layer of the channel. By interleaving pointwise convolutions between the two depth convolution layers, the aim is to augment the channel count, thus ensuring that the depth convolution transpires within a high-dimensional space, thereby enriching the expression of features. Simultaneously, the manipulation of larger channel tensors by means of depth kernels serves to significantly curtail the parameter count, achieving the objective of lightweight design. The specific formula is as follows:

$$\hat{G} = \emptyset_{1,p}\emptyset_{1,d}(F)G = \emptyset_{1,d}\emptyset_{2,p}(\hat{G}) + F \quad (1)$$

where $\emptyset_{i,p}$, denotes a 1×1 convolution and $\emptyset_{i,d}$denotes a deep convolution.

The SandGlass Block structure as shown in Fig. 2 adeptly capitalizes on the strengths of high-dimensional network convergence acceleration while harnessing the computational benefits offered by deep separable convolution. Through the incorporation of deep convolution layers functioning within high-dimensional space, it facilitates the acquisition of more intricate feature expressions. This approach also effectively diminishes the memory access demands associated with high-dimensional feature maps, thereby enhancing hardware execution efficiency.

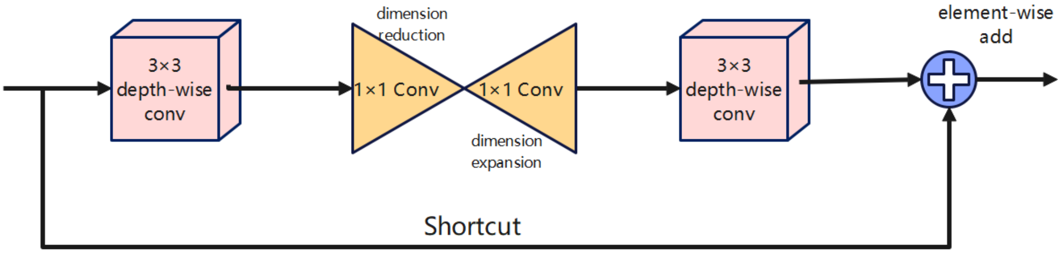

Figure 2. SandGlass network structure

Furthermore, the SandGlass Block design guarantees the efficient transmission of a greater volume of information from the lower layers to the upper layers, bolstering gradient backpropagation. By introducing two depthwise separable convolutions, it further encodes additional spatial information, thereby contributing to the overall efficacy of the structure.

In convolutional neural networks, the pooling layer is a commonly employed technique to reduce the dimensionality of image features, thereby enhancing classification accuracy, processing speed, as well as the model's training speed and generalization capacity. The average pooling layer functions by segmenting the input feature map into uniform regions, subsequently downsampling the feature map, and computing the average value of the constituent elements within each region. The resulting average value is then assigned as the output value of that respective region, thereby diminishing the size of the feature map and the number of parameters. This process effectively trims the feature map size while retaining the essential feature information. The specific computational procedure is delineated as follows:

Firstly, the input data and the size of the pooling kernel are defined[38]. It is assumed that the input data are $X \in R^{H*W}$, the size of the pooling kernel is k×k, and the step size is s. Then we define the output characteristic graph. The size of the output characteristic graph $Y \in R^{H^{-1}*W^{-1}}$:

$$H' = \left[\frac{H-k}{s}\right] + 1, W' = \left[\frac{W-k}{s}\right] + 1 \tag{2}$$

The average pooling pool (X) is defined as:

$$pool(x)_{i,j} = \frac{1}{k^2}\sum_{m=0}^{k-1}\sum_{n=0}^{k-1} X_{i*s+m,j*s+n} \tag{3}$$

Visual converter of the dual-layer routing attention mechanism

To bolster the model's proficiency in extracting and representing target features, particularly when confronted with intricate scenarios and diminutive targets, this study introduces a visual converter of dual-layer routing attention mechanism. This mechanism combines the self-attention mechanism from Transformers[39] with the feature pyramid network of BiFPN[40] (Bi-directional Feature Pyramid Network) to elevate target detection performance. Notably, this mechanism excels in bidirectional encoding, allowing it to incorporate both global and local information. Its fundamental concept involves the selection of pertinent key-value pairs and the judicious retention of a limited portion of the routing area. This retention facilitates the preservation of fine-grained information while concurrently reducing computational demands. The Bi-level Routing Attention (BRA) mechanism encompasses several components. Initially, the feature map is subdivided into non-overlapping regions of size S×S and subjected to linear mapping, as described by the following formula:

$$Q = X^r W^q,\ K = X^r W^k,\ V = X^r W^v \tag{4}$$

Subsequently, the attention weight is computed for the coarse-grained Tokens, and solely the top-k regions are selected as relevant regions for engagement in fine-grained operations. This procedure is mathematically articulated as follows:

$$A^r = Q^r (K^r)^T \tag{5}$$

$$I^r = topkIndex(A^r) \tag{6}$$

Lastly, the most pertinent top-k coarse-grained region for each Token is employed as both the key and value for their involvement in the ultimate operation. To accentuate locality, a deep convolution is applied to the value. The associated formula is expressed as follows:

$$K^g = gather(K, I^r) \tag{7}$$

$$V^g = gather(V, I^r) \tag{8}$$

$$O = Attention(Q, K^g, V^g) + LCE(V) \tag{9}$$

Building upon the aforementioned BRA mechanism, the initial phase adopts overlapping Patch embedding. Subsequently, the second to fourth stages augment the channel count and implement patch merging blocks to downsize the input spatial resolution. This sequence is followed by feature transformation, effectively diminishing the computational load of the model and enhancing its operational efficiency. The specific structure is shown in Fig.3.

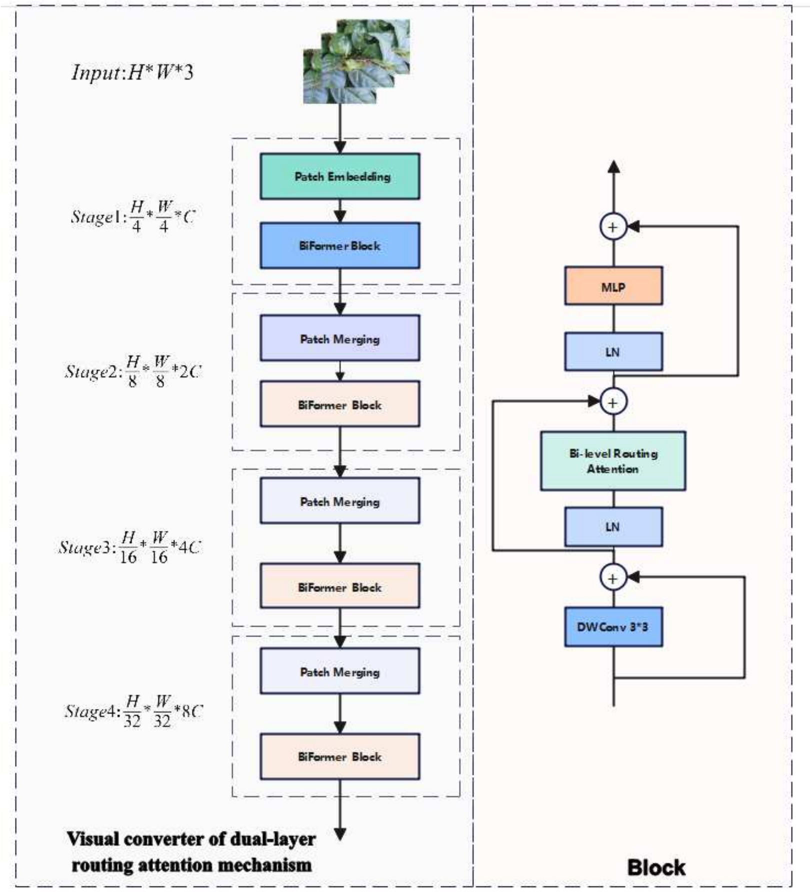

Figure 3. Structure diagram of attention mechanism of dual-layer routing

Loss function

IoU (Intersection over Union) denotes the ratio between the intersection and union of the actual box and the predicted box[41]. It serves as a crucial metric for gauging the dissimilarity between predictive and target information. In the original YOLOv7 algorithm, the CIoU loss function is employed for the computation of location loss. The precise mathematical formulation is presented as follows:

$$L_{CIoU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \tag{10}$$

$$\alpha = \frac{v}{1 - IoU + v} \tag{11}$$

$$v = \frac{4}{x^2}\left(arctan\frac{w^{gt}}{h^{gt}} - arctan\frac{w}{h}\right)^2 \tag{12}$$

In these equations, the variables are as follows: "b" represents the predicted box, $b^{gt}$the true box, $\rho^2(b, b^{gt})$signifies the Euclidean distance between the center points of the predicted box and the true box, "c" denotes the diagonal distance of the minimum enclosing region that can encompass both the predicted box and the true box, "v" represents the aspect ratio consistency, and "α" serves as the balance parameter. While the CIoU accounts for the bounding box's intersection area, the distance between center points, and the aspect ratio, it does not consider the angular misalignment between the predicted and actual boxes. Hence, to enhance the efficacy of tea disease detection, this study replaces the original loss function with SIoU[42], which leads to an improved regression accuracy for the predicted boxes. SIoU treats targets with varying degrees of clarity uniformly, introduces angle loss, and adjusts the predicted box along the shortest regression path. This approach aims to provide a more stable IoU loss and effectively address the issue of gradient instability. The specific mathematical expressions are detailed as follows:

$$\Lambda = 1 - 2sin^2\left(arcsin(x) - \frac{\pi}{4}\right) \tag{13}$$

$$\Delta = \sum_{t=x,y}(1 - e^{-\gamma p_t}) \tag{14}$$

$$\Omega = \sum_{t=w,h}\left(1 - e^{-w^t}\right)^{\theta} \tag{15}$$

$$L_{SIoU} = 1 - IoU + \frac{\Delta+\Omega}{2} \tag{16}$$

In the aforementioned equation, θ represents the angular loss, Δ signifies the distance loss when considering the angle loss, and Ω characterizes the shape loss.

The specific structure of the improved YOLOv7 is as follows

By implementing the aforementioned refinements to the YOLOv7 model, we achieved model lightweighting and an improved capacity for extracting disease features, resulting in enhanced convergence speed. The specific network structure is depicted as show in Fig.4:

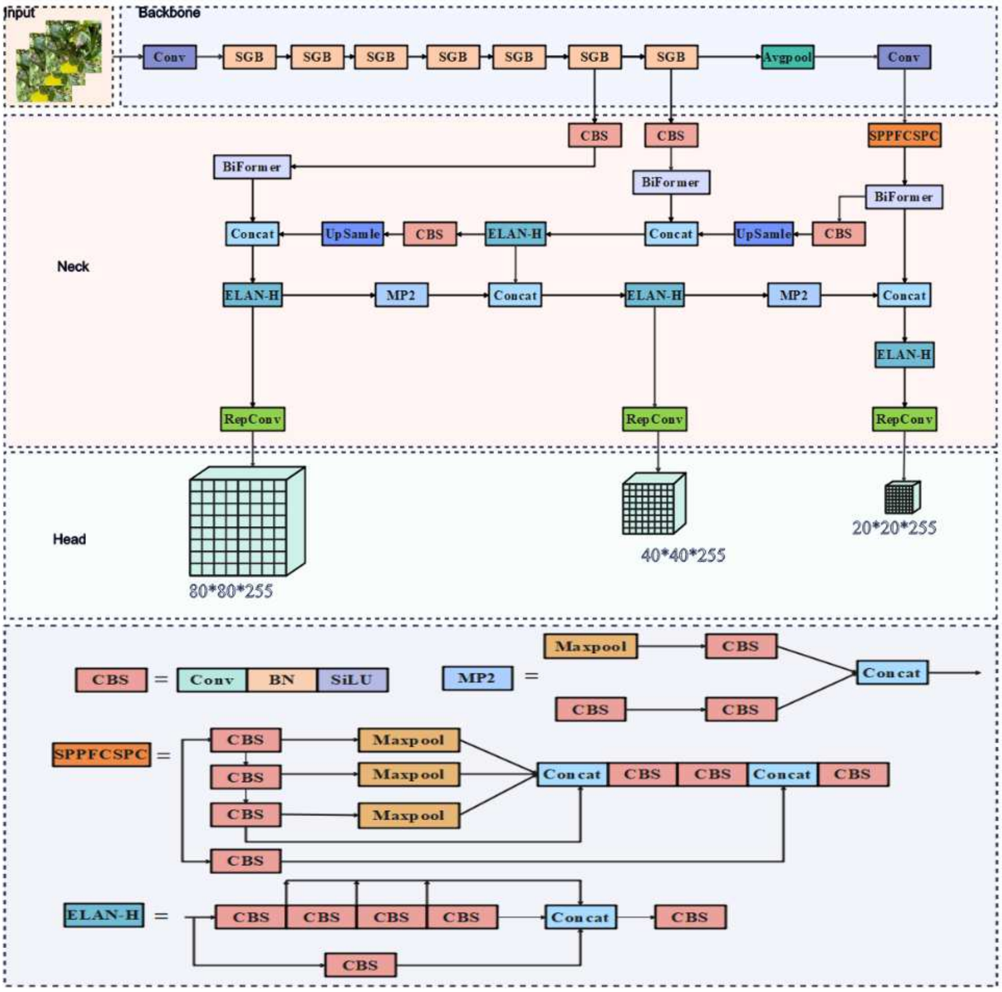

Figure 4. Structure diagram of improved YOLOv7

Selection of evaluation indicators

In assessing the performance of target detection category models, the P value, R value, and mAP are typically chosen evaluation metrics. Here, the area under the Recall-Precision curve (AP) is pivotal[43], and mAP serves as the average of individual category AP values, offering a comprehensive evaluation metric. The mathematical representation is articulated as follows:

$$P = \frac{TP}{TP+FP} \quad (17)$$

$$R = \frac{TP}{TP+FN} \quad (18)$$

$$mAP = \int_0^1 P * (R)dR \quad (19)$$

In this context, TP (True Positives) represents the count of accurately identified samples, FP (False Positives) corresponds to the number of incorrectly identified positives, and FN (False Negatives) signifies the count of missed samples.

Experimental Environment

In this study, the training model operated on the Windows 11 operating system, utilized an Intel(R) Core (TM) i9-12900H 2. 50 GHz processor, 16GB of memory, and an NVIDIA GeForce RTX 3060 graphics card. The deep learning framework employed was PyTorch, integrated within a CUDA 11.1 virtual environment, and made use of CUDNN 8.0. The programming language utilized was Python 3.7.

Results

The loss function serves a dual purpose as a vital metric for quantifying the disparity between predicted outcomes and actual results, and as a fundamental criterion for appraising model quality. A smaller loss function value signifies a closer alignment between the model's predictions and the actual results, indicative of superior model performance. In this study, the enhanced YOLOv7 model underwent training, with a total of 400 training iterations (Epochs). As illustrated in the subsequent figure, it is evident that during the initial stages of training for the improved YOLOv7 model, the loss function exhibits a more rapid descent in gradient speed compared to the original YOLOv7 model. Simultaneously, as training progresses, after approximately 150 rounds, there is a notable deceleration in the decline rate of the loss function. As shown in Fig.5, by the 300th round, it becomes apparent that the loss curve gradually stabilizes, and the loss function commences convergence. The final results reveal that the enhanced model consistently maintains an overall loss value of approximately 2.8%, whereas the original YOLOv7 model stabilizes at approximately 3.9%. The improved model demonstrates clear superiority, as its loss value is significantly lower than that of the unimproved model, affirming its strong performance.

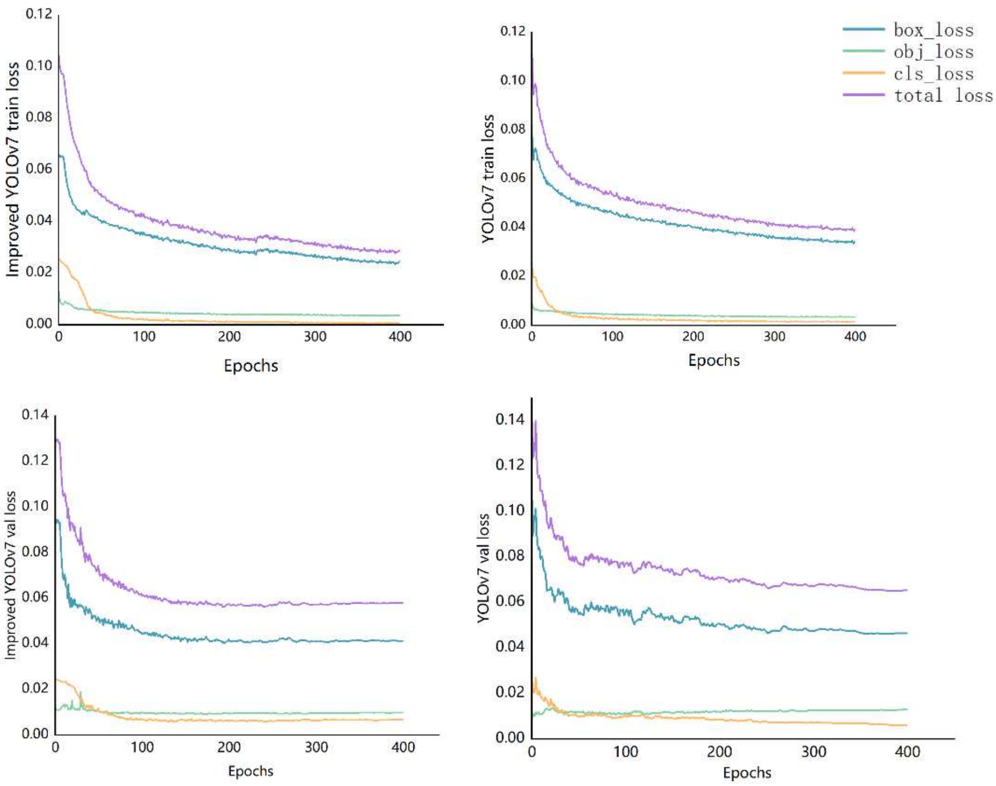

Figure 5. Comparison of loss function curves

To comprehensively assess the model's strengths and weaknesses, three key metrics, namely the P value, R value, and mAP, were selected to encompass a holistic evaluation of the model's classification performance across different target categories. AP reflects the average accuracy of a specific category across varying IoU thresholds. AP provides a comprehensive evaluation of the model's positional and predictive accuracy. The AP value is contingent on the model's accuracy and recall rate and is essentially the area under the Precision-Recall Curve for a particular category. In this curve, the horizontal axis signifies the recall rate, while the vertical axis represents the accuracy rate. mAP represents the average value across all categories' AP scores, offering a consolidated assessment of the model's overall performance. It serves as a pivotal criterion for assessing deep learning models, offering a comprehensive measure of detection model performance. These metrics collectively play an instrumental role in gauging the model's relative merits and shortcomings. The running result of the model is shown in Fig.6.

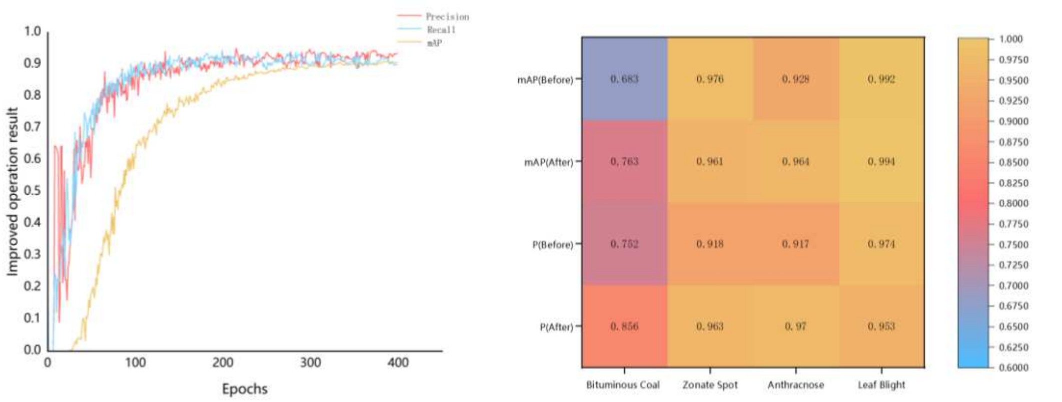

Figure 6. Results of improved model accuracy P, R and mAP(left), Comparison of P and mAP before and after improvement(right)

The graphical representation above unmistakably illustrates that the enhanced YOLOv7 model exhibits superior learning capabilities and enhanced recognition accuracy. Evidently, the P value, R value, and mAP indices demonstrate a sustained and pronounced upward trajectory. As the number of training rounds increased, these metrics gradually stabilized, entering a state of steady fluctuations. Based on the outcomes of model training, the improved YOLOv7 model consistently delivers commendable recognition performance, accurately identifying four distinct types of tea diseases. The model's enhanced recognition results for diseases in complex environments are visualized in the following figure.

From the comparison results of the above graph, it can be seen that only leaf blight is slightly lower than the original model, and the P values of other category recognition are increased by 13.83 %, 4.9 % and 5.78 % respectively. mAP represents the sum of the average accuracy of all categories and the average value obtained. The model in this study has high accuracy in prediction and can better identify the type of disease. The improved YOLOv7 model had only zonate spots lightly lower than the original model, and the other three diseases increased by 11.71 %, 3.88 %, and 0.2 % respectively. The recognition accuracy of the four diseases before and after improvement is shown in Fig.7.

Figure 7. Results of disease identification

To further validate the competitive edge of the enhanced model proposed in this study, the detection and recognition performance of various models was compared for tea bituminous coal disease, anthracnose, leaf spot, and leaf blight. The comparative findings are presented in following Tab.1.At the same time, to ensure result reliability, the original YOLOv7, YOLOv7-tiny, Faster R-CNN, and SSD network models were trained and tested on an identical dataset, while maintaining uniformity in the training platform configuration.

Model comparison	mAP	Examination Speed /FPS	Model Volume /MB
Improved YOLOv7	92.1	41.15	60.5
YOLOv7	89.5	37.03	74.8
YOLOv7-tiny	87.3	42.02	53.2
Faster R-CNN	81.6	7.34	99.6
SSD	86.4	17.42	83.7

Table 1. Comparison of results of different models

Based on the outcomes detailed in the above table, it is evident that the enhanced YOLOv7 model attains an average accuracy that surpasses that of YOLOv7, YOLOv7-tiny, Faster R-CNN, and SSD by 2.9, 5.5, and6.6 percentage points, respectively. While the detection speed and model size of the improved YOLOv7 model may be marginally inferior to those of the YOLOv7-tiny model, the enhanced YOLOv7 model boasts an average accuracy increase of 5.5 %. Consequently, its detection performance is superior and more practical for real-world applications in tea garden disease detection.

Discussion

In this research, we introduced an advanced algorithm built upon the YOLOv7 model to address the issues of low disease recognition rates and slow model response times within tea gardens. Our algorithm incorporates the MobileNeXt module to streamline the backbone network of the YOLOv7 model. Simultaneously, we integrated a double-layer routing attention visual converter and replaced the original loss function with the more suitable SIoU function, which is particularly effective in complex backgrounds. Following these enhancements, the model has demonstrated superior performance compared to the original version, realizing remarkable advantages.

Compared to the YOLOv7 model, our improved version achieved a 5.1 % increase in the recognition accuracy of the four diseases. Moreover, the final total loss of the enhanced YOLOv7 network consistently remains below 2.8 %, marking a 1.1 % reduction compared to the original YOLOv7 network. Additionally, our improvements have significantly reduced the prediction box position loss, prediction box confidence loss, and classification loss. Externally verified data have shown that the improved YOLOv7 model operates at a frame rate 4.12 HZ higher than the original YOLOv7. The mAP in actual detection has increased by 2.6 %, all while reducing the model's volume by 19.12 % and boosting the detection speed by 11.13 %.

However, it is important to acknowledge certain limitations in this study. The complex environment of tea gardens may still harbor other disease types. Subsequent research should focus on augmenting the quality and diversity of collected data to enhance the model's accuracy and generalization capabilities. This endeavor paves the way for deploying the model on mobile terminals, offering genuine contributions to tea garden disease detection and the intelligent tea industry.

Data Availability

Our relevant data are allowed by Yunnan Agricultural University and related bases. All data generated or analysed during this study are available in the Github repository. Links to the code and datasets are provided in the below hyperlinked text. Code and dataset of Improved YOLOv7 project: https://github.com/anqi99/yolov7.git.

References

1. Xue, Z., Xu, R., Bai, D. & Lin, H. Yolo-Tea: A Tea Disease Detection Model Improved by Yolov5. *Forests*. **14**, 415 (2023). https://doi.org/10.3390/f14020415

2. Lee, L. K. & Foo, K. Y. Recent Advances On the Beneficial Use and Health Implications of Pu-Erh

Tea. *Food Res. Int.* **53**, 619-628 (2013). https://doi.org/10.1016/j.foodres.2013.02.036

3. Chakruno, P., Banik, S. & Sumi, K. Important Diseases of Tea (Camellia Sinensis L.) And their Integrated Management. *Diseases of Horticultural Crops: Diagnosis and Management*: Apple Academic Press, 2022:119-138.
4. Sanlier, N., Atik, 0. & Atik, A. A Minireview of Effects of White Tea Consumption On Diseases. *Trends Food Sci. Technol.* **82**, 82-88 (2018). https://doi.org/10.1016/j.tifs.2018.10.004
5. Wei G, Zhou R. Comparison of machine learning and deep learning models for evaluating suitable areas for premium teas in Yunnan, China. *Plos one*, 2023, 18(2): e0282105.
6. Hu, G., Wu, H., Zhang, Y. & Wan, M. A Low Shot Learning Method for Tea Leaf’S Disease Identification. Comput. Electron. Agric. 163, 104852 (2019).
7. Sankaranarayanan, A. C., Veeraraghavan, A. & Chellappa, R. Object Detection, Tracking and Recognition for Multiple Smart Cameras. *Proc. Ieee*. **96**, 1606-1624 (2008). DOI: 10.1109/JPROC.2008.928758
8. Wang, H., Gu, J. & Wang, M. A Review On the Application of Computer Vision and Machine Learning in the Tea Industry. *Front. Sustain. Food Syst.* **7**, 1172543 (2023). https://doi.org/10.3389/fsufs.2023.1172543
9. Xing, W. et al. Suitability Evaluation of Tea Cultivation Using Machine Learning Technique at Town and Village Scales. *Agronomy*. **12**, 2010 (2022). https://doi.org/10.3390/agronomy12092010
10. Cui, Q., Yang, B., Liu, B., Li, Y. & Ning, J. Tea Category Identification Using Wavelet Signal Reconstruction of Hyperspectral Imagery and Machine Learning. *Agriculture*. **12**, 1085 (2022). https://doi.org/10.3390/agriculture12081085
11. Maria, S. K. et al. Cauliflower Disease Recognition Using Machine Learning and Transfer Learning. *Smart Systems: Innovations in Computing: Proceedings of SSIC 2021*: Springer, 2022:359-375.
12. Pallathadka, H. et al. Application of Machine Learning Techniques in Rice Leaf Disease Detection. *Materials Today: Proceedings*. **51**, 2277-2280 (2022). https://doi.org/10.1016/j.matpr.2021.11.398
13. Javidan, S. M., Banakar, A., Vakilian, K. A. & Ampatzidis, Y. Diagnosis of Grape Leaf Diseases Using Automatic K-Means Clustering and Machine Learning. *Smart Agricultural Technology*. **3**, 100081 (2023). https://doi.org/10.1016/j.atech.2022.100081
14. Sun, Y., Jiang, Z., Zhang, L., Dong, W. & Rao, Y. Slic_Svm Based Leaf Diseases Saliency Map Extraction of Tea Plant. *Comput. Electron. Agric.* **157**, 102-109 (2019).
15. Das, D., Singh, M., Mohanty, S. S. & Chakravarty, S. Leaf Disease Detection Using Support Vector Machine. *2020 International Conference on Communication and Signal Processing (ICCSP)*: IEEE, 2020:1036-1040. DOI: 10.1109/ICCSP48568.2020.9182128
16. Almalioglu, Y., Turan, M., Trigoni, N. & Markham, A. Deep Learning-Based Robust Positioning for All-Weather Autonomous Driving. *Nat. Mach. Intell.* **4**, 749-760 (2022).
17. Almalioglu Y, et al. Deep learning-based robust positioning for all-weather autonomous driving. *Nature Machine Intelligence*, **4**(9): 749-760 (2022).
18. Shen, S. et al. A Physics-Informed Deep Learning Approach for Bearing Fault Detection. *Eng. Appl. Artif. Intell.* **103**, 104295 (2021). https://doi.org/10.1016/j.engappai.2021.104295
19. Alakus, T. B. & Turkoglu, I. Comparison of Deep Learning Approaches to Predict Covid-19 Infection. *Chaos, Solitons & Fractals*. **140**, 110120 (2020). https://doi.org/10.1016/j.chaos.2020.110120
20. Iqbal, R., Maniak, T., Doctor, F. & Karyotis, C. Fault Detection and Isolation in Industrial Processes Using Deep Learning Approaches. *Ieee Trans. Ind. Inform.* **15**, 3077-3084 (2019). DOI: 10.1109/TII.2019.2902274

21. Sahoo, A. K., Pradhan, C. & Das, H. Performance Evaluation of Different Machine Learning Methods and Deep-Learning Based Convolutional Neural Network for Health Decision Making. *Nature Inspired Computing for Data Science*. 201-212 (2020).
22. Anami, B. S., Malvade, N. N. & Palaiah, S. Deep Learning Approach for Recognition and Classification of Yield Affecting Paddy Crop Stresses Using Field Images. *Artificial Intelligence in Agriculture*. **4**, 12-20 (2020). https://doi.org/10.1016/j.aiia.2020.03.001
23. Tiwari, D. et al. Potato Leaf Diseases Detection Using Deep Learning. *2020 4th international conference on intelligent computing and control systems (ICICCS)*: IEEE, 2020:461-466.
24. Soeb, M. J. A. et al. Tea Leaf Disease Detection and Identification Based On Yolov7 (Yolo-T). *Sci. Rep.* **13**, 6078 (2023).
25. Zhang, S. et al. Edge Device Detection of Tea Leaves with One Bud and Two Leaves Based On Shufflenetv2-Yolov5-Lite-E.*Agronomy*. **13**, 577 (2023).
26. Latha, R. S. et al. Automatic Detection of Tea Leaf Diseases Using Deep Convolution Neural Network. *2021 International Conference on Computer Communication and Informatics (ICCCI)*: IEEE, 2021:1-6. DOI: 10.1109/ICCCI50826.2021.9402225
27. Sankaranarayanan, A. C., Veeraraghavan, A. & Chellappa, R. Object Detection, Tracking and Recognition for Multiple Smart Cameras. *Proc. Ieee*. **96**, 1606-1624 (2008). DOI: 10.1109/JPROC.2008.928758
28. Chun-man, Y., Xiang, Z. & Qingpeng, W. Face Expression Recognition Based On Improved Mobilenext. (2022). https://doi.org/10.21203/rs.3.rs-2270472/v1
29. Jia, L. et al. Mobilenet-Ca-Yolo: An Improved Yolov7 Based On the Mobilenetv3 and Attention Mechanism for Rice Pests and Diseases Detection. *Agriculture*. **13**, 1285 (2023). https://doi.org/10.3390/agriculture13071285
30. Park, J., Kim, J. & Kim, C. Biformer: Learning Bilateral Motion Estimation Via Bilateral Transformer for 4K Video Frame Interpolation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023:1568-1577.
31. Zhu, L., Wang, X., Ke, Z., Zhang, W. & Lau, R. W. Biformer: Vision Transformer with Bi-Level Routing Attention. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023:10323-10333.
32. Zhu, J. et al. Fire Detection in Ship Engine Rooms Based On Deep Learning. *Sensors*. **23**, 6552 (2023).
33. Hu, G., Yang, X., Zhang, Y. & Wan, M. Identification of Tea Leaf Diseases by Using an Improved Deep Convolutional Neural Network. *Sustainable Computing: Informatics and Systems*. **24**, 100353 (2019). https://doi.org/10.1016/j.suscom.2019.100353
34. Wu, D. et al. Detection of Camellia Oleifera Fruit in Complex Scenes by Using Yolov7 and Data Augmentation. *Applied Sciences*. **12**, 11318 (2022).
35. Zhao, K., Zhao, L., Zhao, Y. & Deng, H. Study On Lightweight Model of Maize Seedling Object Detection Based On Yolov7. *Applied Sciences*. **13**, 7731 (2023).
36. Wang, Y., Wang, H. & Xin, Z. Efficient Detection Model of Steel Strip Surface Defects Based On Yolo-V7. *Ieee Access*. **10**, 133936-133944 (2022).
37. Beazley, K. Operation Sandglass: Old History, Contemporary Lessons. *Security Challenges*. **4**, 23-43 (2008).
38. Kasagi, A., Tabaru, T. & Tamura, H. Fast Algorithm Using Summed Area Tables with Unified Layer Performing Convolution and Average Pooling. *2017 IEEE 27th International Workshop on*

Machine Learning for Signal Processing (MLSP): IEEE, 2017:1-6.

39. Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P. S. & Vaughan, J. W. Advances in Neural Information Processing Systems 34. (2021).
40. Chen, J., Mai, H., Luo, L., Chen, X. & Wu, K. Effective Feature Fusion Network in Bifpn for Small Object Detection. *2021 IEEE international conference on image processing (ICIP)*: IEEE, 2021:699-703. DOI: 10.1109/ICIP42928.2021.9506347
41. Zheng, Z. et al. Distance-Iou Loss: Faster and Better Learning for Bounding Box Regression. *Proceedings of the AAAI conference on artificial intelligence*, 2020:12993-13000.
42. Gevorgyan, Z. Siou Loss: More Powerful Learning for Bounding Box Regression. *Arxiv Preprint Arxiv:2205.12740*. (2022). https://doi.org/10.48550/arXiv.2205.12740
43. Wang, C., Bochkovskiy, A. & Liao, H. M. Yolov7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023:7464-7475.

Acknowledgements

This research was funded by Development and demonstration of intelligent agricultural data sensing technology and equipment in plateau mountainous areas(202302AE09002001), Study on the Yunnan Menghai County Smart Tea Industry Science and Technology Mission (202304Bl090013) and National Natural Science Foundation (32060702).

Author Contributions

Conceptualization, writing—original draft preparation, Y.X. and W.Y.; methodology, S.Z. , Q.W. and X.L; software, C.Y., H.W., Y.W, and J.X.; formal analysis, J.H., L.L., and Z.C.; investigation, Z.W. and Z.Z; conceptualization, writing—review and editing, funding acquisition, B.W. All authors have read and agreed to the published version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to B.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.