Journal of Agricultural Engineering

https://www.agroengineering.org/

Residual attention based multi-label learning for apple leaf disease identification

Changjian Zhou, Zhenyuan Zhao, Wenzhuo Chen, Yuquan Feng, Jia Song, Wensheng Xiang

Please cite this article as doi: 10.4081/jae.2024.1595

Submitted: 15/10/2022
Accepted: 13/02/2023

Residual attention based multi-label learning for apple leaf disease identification

Changjian Zhou,[1,2] Zhenyuan Zhao,[2,3] Wenzhuo Chen,[2,3] Yuquan Feng,[2,3] Jia Song,[1] Wensheng Xiang[1,2,4]

[1]College of Life Science, Northeast Agricultural University, Harbin, China
[2]High-Performance Computing and Artificial Intelligence Laboratory, Northeast Agricultural University, Harbin, China
[3]School of Electrical and Information, Northeast Agricultural University, Harbin, China
[4]State Key Laboratory for Biology of Plant Diseases and Insect Pests, Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Beijing, China

Corresponding authors: Jia Song (songjia@neau.edu.cn) and Wensheng Xiang (xiangwensheng@neau.edu.cn)

CRediT authorship contribution statement: Changjian Zhou: conceptualization, methodology, software, writing original draft. Zhenyuan Zhao, Wenzhuo Chen and Yuquan Feng: methodology, coding; Jia Song and Wensheng Xiang: conceptualization, methodology, resources, supervision, reviewing.

Declaration of Competing Interest: the authors declare that they have no known competing financial interests of this paper.

Funding: this work was supported by the National Natural Science Foundation of China [grant number 32030090].

Abstract Recent studies suggest that plant disease identification via machine learning approach is vital for preventing the spread of diseases. Identifying multiple diseases simultaneous on a single leaf is one of the most irritating issues in agricultural production. However, the existing approaches are difficult to meet the requirements of production practice in accuracy or interpretability. Here, we present residual attention based multi-label learning framework (RAMDI), a method for predicting apple leaf diseases in natural environment. Built upon an attention based multi-label learning framework, the channel and spatial attention mechanisms are investigated and embedded in residual network for multi-label disease prediction, which takes advantage of channel-wise and spatial-wise attention weights. Experimental results indicate that the RAMDI achieves 0.976 accuracy, 0.986 F-score, and 0.979 mAPs, outperforms the existing state-of-the-art apple leaf disease identification models. RAMDI not only predicts multi-disease on a single leaf simultaneously, but also reveals the interpretability among positive predictions that contribute most to identify the key features that are significant for the leaf diseases. This method achieves the following two achievements. Firstly, it provides a solution for detecting multiple diseases on a single leaf. Secondly, this approach gains an interpretable understanding for apple leaf disease identification.

Key words: fruits; attention mechanism; machine learning; one-hot encoding.

1. Introduction

The United States Department of Agriculture (USDA) released a report that the global apple production is estimated to reach 81.8 million tons in 2021, a rise of 1.6% year-on-year (Alice, 2021). As one of the most valuable and popular fruits around the world, apple is processed into various foods or condiments. However, apple production is struggling with various disease intrusions, which restricts the improvement of apple yield and quality. Traditional manual identifying diseases is labor-intensive and time-consuming. Moreover, multiple diseases occurring on a single leaf concurrently creates a great challenge for precise identification (Zhou et al., 2021). To mitigate the strong dependence on human labor, it is significant to replace manual identification with automatic detection on behalf of the development of computational approaches.

Generally, symptoms of fruit diseases appear on the leaves first, which makes the leaf disease identification particularly important. Timely identifying disease on leaves prevents fruits from being invaded. To date, many silico prediction methods have been proposed to identify apple leaf diseases, mainly include conventional machine learning methods and various variants of deep learning approaches. The conventional machine learning approaches, such as image processing methods (Ayyub, et al., 2019), support vector machine (Chakraborty et al., 2021), ACS-LBP (Li et al., 2016), 2DSLDR (Shi et al., 2017), hybrid neural clustering (James, et al., 2021), HIoT (Pandiyan et al., 2020) and so on. These methods have achieved satisfactory results in apple leaf disease identification tasks. In addition, as the deep learning architectures achieve unprecedented performances in massive data processing, numerous of apple leaf disease identification models are released in recent years. Such as FCNN-LDA (Agarwal, et al., 2019), leaf spot attention network (Yu, et al., 2020), focus loss function method (Zhong et al., 2020), RegNet (Li et al., 2022), MEAN-SSD (Sun et al., 2021) and CA-ENet (Wang et al., 2021). Together, these works greatly advanced our understanding of the apple leaf disease identification in different species under various conditions. However, the existing methods suffered from the following limitations.

Firstly, most of the existing studies only focus on the single leaf with single disease, mainly including *Alternaria blotch*, *brown spots*, *gray spot*, *mosaic* and *rust*, but failed to support multiple diseases occurring on single leaf simultaneously by an integrated predictive model, and the study of the interplay between different diseases is limited. Ayyub et al. (2019) proposed an image processing-based apple disease identification approach. The traditional image processing methods such as image segmentation, feature extraction (color, texture and shape), feature combination and the support vector machine were employed to identify apple diseases. James et al. (2021) developed a Hybrid Neural Clustering (HNC) classifier by using the K-means to cluster the vector points, and fed them into a feed forward back propagation neural network to classify various apple fruit diseases. Pandiyan et al. (2020) designed a Heterogeneous Internet of things procedural (HIoT) system to point out leaf disease in an efficient manner. The IoT was identified as a repetitive and persistent space to find the impact gesture in leaf image, and it was used for real-time resembling apple leaf diseases. Li et al. (2022) presented a new lightweight convolutional neural network RegNet to identify 5 apple leaf diseases (*rust*, *scab*, *ring rot*, *panonychus ulmi*, and *healthy*) with a high accuracy. When given many multi-label disease instances, the independent-based predictions are likely to make a substantial proportion of false-positive or false-negative results in practice, therefore it should be considered with extra caution. One apple leaf disease may be accompanied by other diseases, such as *scab* and *frog eye leaf spot*, *rust* and *frog eye leaf spot*, etc. The highly reliable disease identification model should consider the coexistence of multiple diseases and their interdependence.

Secondly, most of the existing works rely on the public datasets or collecting data from the Internet, failing to fully take into account of the farmland environmental factors such as climate, humidity, temperature and illuminance, etc. Chakraborty, et al. (2021) and Sun et al. (2021) collected an apple leaf disease dataset under the experimental environment with simple background. Yu, et al. (2020) used the dataset with similar background of apple leaf disease. Zhong et al. (2020) adopted AI Challenger dataset for training models. Agarwal, et al. (2019) developed the apple leaf diseases identification method using PlantVillage dataset. However, the trained model is difficult to achieve the satisfactory prediction effect in real planting environment. It is crucial to take advantage of the data collected from field environment to minimize the potential technological bias whenever such datasets are available.

Lastly, most of the existing works pay more attention to improve identification accuracy but rarely provide a clear interpretation of the predictions. Although some works carefully interpret their designed architectures (Wang et al., 2021), few existing works give insightful analysis into the working process for individual predictions. The occurrence of each disease is often accompanied by a variety of factors, taking advances of interpretable attention mechanism, and visualizing them is helpful to understand the causal relationship. However, most of the existing models remain significant for positive predictions, which are of little help to understand the result forming mechanisms.

Based on the above reasons, it is a strong motivation to design a unified state-of-the-art deep architecture that supports multi-label disease identification using field planting environment dataset with integrating multiple technologies. The RAMDI is proposed here, an attention-based multi-label learning approach for prediction and interpretation multiple apple leaf diseases. Five categories of apple leaf diseases are supported by the presented model, including *scab*, *rust*, *powdery mildew*, *frog eye leaf spot* and *healthy*. To the best of our knowledge, when a type of disease occurred, it is possible accompanied by other categories of diseases at the same time. For example, when it suffered from the *frog eye leaf spot* disease, the *scab*, or *rust* is usually also occurred. The attention based multi-label learning of our method enables accommodation of the shared structure of different diseases

while fully exploiting their distinctive features. As the features of multi-disease on a single leaf are more difficult to handle, multiple attention mechanisms are investigated and integrated to capture the features of each disease and interpret every individual prediction.

2. Materials and methods

2.1 Raw data acquisition and preprocessing

2.1.1 Raw data acquisition

The development of an apple leaf disease identification architecture typically requires disease profiling data at base resolution for training and testing purposes. Part of the raw images were captured using digital cameras or smartphones in the orchards of Qi Xia, Yantai Shandong province, China from 2019 to 2021, and part of the instances were taken from many apple cultivars at Cornell AgriTech (New York, USA; Geneva, Switzerland) in 2019 to ensure the diversity of training data (Thapa et al., 2020), which covered most of the apple cultivars over widespread. A total of 27,883 raw images were obtained, prioritizing those derived from multi-label diseases and generated under different weather conditions in an orchard environment, as shown in Fig.1. After data cleaning, the invalid data of disease instances are eliminated, and 21,631 images are curated into one-hot encoding labels such as Fig.2, which describes the part of the implementation instances (including 4,624 *health images* (0), 4,831 *frog eye images* (1), 1,684 *powdery mildew images* (2), 2,860 *rust images* (3), 4,826 *scab images* (4), 1920 *rust* and *frog eye images* (1,3), 886 *scab* and *frog eye images* (1,4)).

2.1.2 Data preprocessing

The raw data were captured by cameras or smartphones with different channels or resolutions. To the best of our knowledge, the necessary condition for deep learning models to perform well is with massive data of uniform size. Consequently, all of the images were resized to 224×224×3, and the augmentation methods including geometrical and intensity transformations such as horizontal flipping, rotation, aspect ratio, contrast and brightness and noise were employed to mitigate overfitting. All of the instances are split into training set, validation set and test set with percentage ratios of 70%, 15% and 15%. The training set is carried out for training models, validation set is used for monitoring overfitting and optimizing hyperparameters. The overall performance of the models is assessed on test set.

2.2 Related works

This study presented an attention based multi-label learning deep architecture for multiple apple leaf diseases identification, which is closed to 2 branches such as attention mechanism and multi-label algorithm. A brief review that leads to the proposed methodology is given as follows.

2.2.1 Attention mechanism

Attention mechanism plays a crucial role in various vision tasks that invests more resources in specific target regions and ignores useless information around as humans do, thereby enhancing the association of the labels with the image regions. The visual attention mechanisms are briefly categorized into spatial attention, branch attention, channel attention and temporal attention according to data domain. Spatial attention mechanism, as an adaptive spatial region that solves the problem of "where to pay attention", predicts the most relevant and important spatial positions. SENet (Hu et al.,

2020), Non-Local (Wang et al., 2018), SASA (Ramachandran et al., 2019) and ViT (Dosovitskiy et al., 2021) are the most popular spatial attention approaches. Branch attention utilizes a multi-branch structure to solve "which to pay attention to" by a dynamic branch selection method, the important ones are selected from the different masked branches. Condconv (Yang et al., 2019) is the typical representative model using a branch attention method to increase the capability of networks efficiently. Channel attention mechanism is regarded as an object selection process that adaptively decide the weights of each channel to determine what need to pay attention to. There are various channel attention methods along with their development process respectively in recent years, such as SENet, global second-order pooling attention block (Gao et al., 2019), the lightweight style-based recalibration attention block (Lee et al., 2019), gated channel attention transformation (Yang et al., 2020), the efficient channel attention block, multi-spectral channel attention block, multi-spectral channel attention block (Qin et al., 2021) and so on. Temporal attention focusing on the problem of "when to pay attention", which is usually used in video processing.

2.2.2. Multi-label learning algorithm

Multi-label learning solves the problem that a single instance is associated with a set of labels simultaneously. The multi-label learning aims to predict the correct set of labels on a single instance as accurately as possible, which shows significant implications in practical applications such as scene analysis, image annotation and plant disease identification. As the high-dimension feature space and numerous noises existing in multi-label data, it is a high challenge to apply widely. In recent years, multi-label learning has been explored in learning separated-label and multi-label relationship using various techniques. The separated-label approach such as Binary Relevance adopts a classifier such as ANN or RBF to learn each class individually and evaluate the output in the target space. The multi-label relationship methods such as ML-ZSL (Lee et al., 2018) incorporates knowledge graphs to describe multi-label relations in the semantic label space. MRDM (Huang et al., 2021) uses the dependence of class labels to associate the data space with label space for multi-label feature selection. It is a great opportunity and challenge to predict plant disease using multi-label based approaches. For example, various diseases occurred on a single leaf as shown in Fig.3. How to identify these diseases simultaneously is one of the objectives to be solved in this work.

2.3. Evaluation Metrics

The evaluation metrics such as accuracy (Acc), precision (Pre), loss, recall, F-score (Fs), hamming score, hamming loss and mAPs were adopted in this work. Accuracy is utilized to measure the proportion of positive predictions in all samples. Precision is used as the proportion of positive predictions among positive samples determined by classifiers. Recall is introduced as the proportion of the positive predictions in the positive samples. F-score is used as the harmonic average of precision and recall, which is adopted as the evaluation metrics for early stopping and calculated by averaging each metric for multiple classes. Hamming loss is introduced in multi-label classification problem, which counts the number of misclassified labels. Hamming score returns the average accuracy of all samples, and the accuracy rate is the proportion of the number of positive predictions that correctly predicted labels to the truly positive labels. Finally, mAP is the abbreviation of mean average precision, which calculate the mean value of the average precision in all categories.

2.4. Proposed Method

2.4.1 Model architecture

In this work, a residual attention based multi-label architecture RAMDI was proposed for apple leaf disease prediction. The RAMDI framework is shown in Fig.4. Given a set of labelled base-resolution apple leaf disease instances, RAMDI learns the mapping between the image features and the disease category automatically. Once this mapping is learned, the RAMDI with residual attention mechanism enables us to interpret the model and extract the key features that contribute most to the positive prediction. There are 4 residual attention stages in RAMDI framework. Each stage contains 3 convolution layers with *BatchNormalization* and *ReLU* activation function, and the attention block is embedded in residual attention stage. The global average pooling layer is employed to condense a set of information, and then input the condensed features into full connection layer with *Sigmoid* activation function for prediction.

2.4.2 Residual attention block

In this work, the residual attention stage is designed for feature learning, and the channel attention and spatial attention mechanisms are investigated and embedded in the improved residual network for feature learning as shown in Fig.5. Two *pooling* methods such as *maxpooling* and *averagepooling* are utilized to condense the input feature information respectively, and the matrix addition operation is employed between the two condensed features with element-wise sum and fed into channel attention and spatial attention mechanisms respectively as demonstrated in (1).

$$G = Sigmond(G_C \oplus G_S) \tag{1}$$

where G denotes the attention weights, and the *Sigmond* activation function is implemented for normalizing the attention vectors. G_C and G_S are the expanded channel and spatial attention vectors respectively, which are indicated in (2) and (3).

$$G_C = E(CA(conv(Maxpool(F)) \oplus conv(Avgpool(F)))) \tag{2}$$

$$G_S = E(SA(conv(Maxpool(F)) \oplus conv(Avgpool(F)))) \tag{3}$$

where E denotes the *expand* operation, which ensures the same feature dimensions between the input and output vectors. CA and SA denote the channel attention and spatial attention operations, $conv$ denotes the *conv1D* operation, $Maxpool$ and $Avgpool$ denote the *maxpooling* and *averagepooling* operations respectively. $\oplus$ denotes the element-wise sum operation, and F denotes the input features. Together, the residual attention block can be stated as (4).

$$R = F \oplus (G \otimes F) \tag{4}$$

where R denotes the output features of residual attention block.

In the residual attention block, the original features are respectively input into 3 exploited branches including attention branch, weight multiplication branch and residual connection branch. Attention branch aims to trim invalid informative and enhance the key valuable feature weights. Weight multiplication is designed to condense the valuable features that contribute to the true positive predictions, and the residual connection branch is utilized to mitigate overfitting and learn the categorical features of apple leaf disease images. It is expected that these vectors can well synthesize the valuable information needed for each prediction branch.

In this work, the RAMDI model maps each vector to the probability of each disease type simultaneously. The proposed RAMDI model is optimized by weighting *binary cross-entropy* loss function in different tasks. Specifically, the two attention mechanisms are implemented to account for possible interaction of different diseases, and the *Sigmoid* activation function assigns the possible prediction results by one-hot encoding approach, which helps to extract diseases feature patterns in a dense manner and aims to generate high dimensional representations of it.

3. Experiment and results analysis

In this work, the Cent 7.5 Linux operating system with python language, pytorch 1.7 and cuda 10.1 framework was deployed, and the 4×NVIDIA 2080Ti GPUs were employed for accelerating computing in the experiments.

3.1 Implementation details

In general, the hyperparameters are optimized based on the validation sets, while the final prediction results are implemented on the test set. In this work, the *stratified 5-fold cross-validation* is introduced for training in the experiments, while K-fold cross-validation is employed to mitigate overfitting or vanishing gradient in deep learning architectures, especially for those who have small or imbalanced training set. The *stratified 5-fold cross-validation* is implemented by extracting the training set according to the proportion of categories, which take advantage of all the data categories while it is a small amount. The hyperparameters utilized in this work is exhibited in Table.1.

3.1.1 Loss function

In multi-label learning approaches, there are 2 state-of-the-art loss functions such as *Dice* and *binary cross -entropy* are employed in this work. The *Dice* loss is more immune to the data-imbalance issue, which attaches similar weights between false positives and false negatives. The *Dice* loss is adaptive in (5).

$$Dice\ loss = \frac{1}{n}\sum_{x}\left[1 - \frac{2p_{x1}q_{x1} + \rho}{p_{x1}^2 + q_{x1}^2 + \rho}\right] \tag{5}$$

where n denotes the sum of simples, x_i denotes the individual example, p_{xi} is the positive simple, ρ is a factor to both the nominator and the denominator. The *binary cross-entropy* loss function is adopted for multi-label learning with a robust manner which defined in (6).

$$\begin{cases} Y_{positive} = \log(p) \\ Y_{negative} = \log(1-p) \end{cases} \tag{6}$$

where $Y_{positive}$ denotes the positive prediction labels and $Y_{negative}$ is the negative prediction labels, and p is the positive prediction probability. The *binary cross-entropy* loss function could continuously reduce the cross entropy between outputs and the labels in the training process, so that the output of label 1 closer to 1 and label 0 closer to 0. In this work, the two loss functions are both implemented, and the experiment performance is demonstrated in **Section 3.2**.

3.1.2 Activation function

The *Sigmoid* activation function is utilized with one-hot encoding to adapt multi-label problems,

which demonstrated as (7).

$$S(x) = \frac{1}{1+e^{-x}} \tag{7}$$

where $S(x)$ is the output probability of the function and x is a linear vector, and $S(x)$ is tend to 0 or 1 when the positive or negative predictions are inputted. It is expected to activate the value of each prediction at once, and output the probability of each positive prediction respectively. In this work, the one-hot encoding method utilizes N-bit state register to encode N states, and each state has its own register bits, and only one of them is valid at any time. The workflow of *Sigmoid* activation function with one-hot encoding is illustrated in Fig.6.

3.2 Results and analysis

To comprehensively evaluate the performance of the proposed RAMDI model, the comparison approaches and evaluation metrics are stated here. To ensure a fair and interpretable comparison, the classic deep learning approach ResNet-50 was selected as the representative CNN model, and 2 different loss functions and 2 embedding methods were combined into 4 models for comparison. The experimental results on test set are detailed in Table.2. Taking Resnet-50 as the baseline, the 2 loss functions such as *Dice* and *binary cross-entropy* (*BCE)* were implemented respectively, the 2 embedding methods such as embedding in block (see Fig.4) and embedding in stage (see Fig.7) were designed in this work. Subsequently, the state-of-the-art attention models such as SENet (Hu et al., 2020), CBAM (woo et al., 2018), ECA (Wang et al., 2020), and Swin transformer (Liu et al., 2021) were introduced for comparison.

This work aims to design an interpretable classifier that could achieve an exciting performance in the identification of multi-label apple leaf diseases. Various combinations of the loss functions and attention embedding methods are investigated and the RAMDI with *binary cross-entropy* loss function achieved a satisfactory performance. Importantly, to assess the contribution of the attention embedding approach used in our model, the channel attention and spatial attention are implemented respectively and add the outputs as an attention block. This approach could take advantage of the two attention mechanisms with encouraging results. For pooling operation, the *maxpooling* and *averagepooling* are employed to trim the input features respectively and add the outputs as input into attention block, which retains valuable information to the greatest extent. Besides, the *Sigmoid* activation function and one-hot encoding were cooperated to execute the multiple labels assignment tasks.

4 Discussion

A residual attention based multi-label learning model was designed in this work. It predicted multiple occurring diseases on a single apple leaf simultaneously and present the interpretation that contributed to the predictions.

To fully exploit the inherent structure of the prediction models, two different embedding approaches and two loss functions were combined respectively. It was found that the *binary cross-entropy* loss function drastically improved the predictive performance. To deal with the dataset bias, the data augmentation and *stratified 5-fold cross-validation* training strategy was adopted for training stage, which increased the training effect effectively. It is encouraging that the overall performance of the proposed RAMDI model outperforms the conventional machine learning approaches and the existing start-of-the-art multi-label models.

4.1 Interpretability

A satisfactory model needs not only achieve high prediction accuracy, but also grasp interpretability from its internal structure. The proposed RAMDI model adopts residual concatenated attention mechanisms and multi-label prediction blocks to explain visually how the model makes specific expected decisions. Specifically, this section focuses on why the proposed approach is valued most while identifying different diseases, and acquired the kernels which contributed most in the positive predictions.

To the best of our knowledge, when multiple diseases occurred on a single leaf simultaneously, it is a disturbing job to predict disease with multiple symptom features using single-label machine approaches. The attention mechanisms such as spatial attention and channel attention are transplanted to address the problems of "what is the leaf disease" and "where is the leaf disease". Since the input images are capable of picking out specific elements from the features to make output, thus the two attention mechanisms give the model ability to determine and place weights on the relevant place of the input instances for each prediction work as expected. Consequently, it can be seen that the attention mechanisms effectively filter out the noisy information of the false prediction information by visualizing the attention tensor weights as shown in Fig.8, the most critical represent parts are identified in RAMDI model while making each prediction.

By calculating the gradient of output tensors with respect to their input, the residual connection keeps gradient stable convergence, which reflects the contributions of the input tensors to the output in the proposed RAMDI model. In addition, the *binary cross-entropy* loss function measures the contribution of each input instance to each disease prediction and assigns its contribution scores to the corresponding label in the output matrix. The threshold parameter is introduced to define whether the output is positive or not according to the contribution scores, which could visualize the attribution maps of the importance in the predictions.

4.2 limitations

Different from the single-label prediction models, multi-label learning faces more uncertainties and incline to close to the practical planting environment. The occurrence of one disease may be accompanied by others. identifying multiple diseases timely is of great practical significance to control the spread of diseases. However, it is impossible to address a general disease identification model the limited training data, and few-shot learning with multi-label may be a possible solution. In addition, the implicit variables relationship of different labels needs to be further explored to determine the weight of the corresponding labels. It is significant to analyze the correlation labels to improve the performance of prediction models. The attention mechanisms are employed to improve model performance in accuracy and robustness. However, the increase of computing resources makes it difficult to be deployed in mobile devices, a lightweight attention based multi-label model for plant leaf disease identification is expected.

It is important to note that, the proposed RAMDI model currently does not consider the distinct severity of different diseases as it is difficult to obtain disease severity data set. So even in the same setting, the false-positive predictions cannot vary substantially between the different severity of leaf diseases. The problem is partially due to the limited training data for severity analysis, although it is important to develop multi-label learning based plant leaf disease severity prediction models in real-world. This issue may be alleviated with the development of few-shot learning and digital twin technologies.

5. Conclusions

This work presented RAMDI, a residual attention based multi-label architecture for apple leaf disease identification. It is an encouraging finding that the combination of two attention mechanisms significantly improved the feature representation ability of apple leaf diseases. In addition, the residual concatenation between the input features and the full connection layer tensors improved the prediction performance to process multi-label classification tasks. Despite the existing models have achieved the state-of-the-art performances in single disease identification, there is still a large margin across the board between the laboratory results and practice value, the proposed RAMDI bridges the performance gap by giving a more dependable and interpretable approach for detecting multiple diseases, which provides a new insight for identifying multiple diseases simultaneously.

At present, the comprehensive information of plant diseases has not been fully exploited, especially the disease severity analysis, mobile terminal-based plant disease identification, and the correlation analysis of multiple diseases, and so on. It is expected to develop more robust plant disease identification methods to better serve the digital agriculture in future.

References

Agarwal, M., Kaliyar, R., Singal, G., Gupta, S., 2019. FCNN-LDA: A Faster Convolution Neural Network model for Leaf Disease identification on Apple's leaf dataset. 12th International Conference on Information & Communication Technology and System (ICTS), 246-251, https://doi.org/10.1109/ICTS.2019.8850964.

Alice W., 2021. World apple, grape, and pear production forecast to rise in 2021/22. https://www.mintecglobal.com/top-stories/world-apple-grape-and-pear-production-forecast-to-rise-in-2021/22.

Ayyub, S., Manjramkar, A., 2019. Fruit Disease Classification and Identification using Image Processing. 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2019, 754-758, https://doi.org/10.1109/ICCMC.2019.8819789.

Chakraborty, S., Paul, S., Rahat-uz-Zaman, M., 2021. Prediction of Apple Leaf Diseases Using Multiclass Support Vector Machine. 2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST), 147-151, https://10.1109/ICREST51555.2021.9331132.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N., 2021. An image is worth 16x16 words: Transformers for image recognition at scale. The Ninth International Conference on Learning Representations (LCLR).

Gao, Z., Xie J., Wang Q., Li, P., 2019. Global Second-Order Pooling Convolutional Networks. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 3019-3028. https://doi.org/10.1109/CVPR.2019.00314.

Hu, J., Shen, L., Albanie, S., Sun, G., Wu, E., 2020. Squeeze-and-Excitation Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 42, 8, 2011-2023. https://doi.org/10.1109/TPAMI.2019.2913372.

Huang, R., Wu, Z., 2021.Multi-label feature selection via manifold regularization and dependence maximization, Pattern Recognition,120. https://doi.org/10.1016/j.patcog.2021.108149.

James, G., Sujatha,S., 2021.Categorising Apple Fruit Diseases Employing Hybrid Neural Clustering

Classifier. Materials Today: Proceedings. https://doi.org/10.1016/j.matpr.2020.12.139.
Lee, C., Fang, W., Yeh, C., Wang, Y., 2018. Multi-label Zero-Shot Learning with Structured Knowledge Graphs, 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 1576-1585, https://doi.org/10.1109/CVPR.2018.00170.
Lee, H., Kim, H., Nam, H., 2019. Srm: A style-based recalibration module for convolutional neural networks. Preprint at: https://arxiv.org/abs/ 1903.10829.
Li, C., Peng, J., Zhang, S., 2016. Apple leaf disease identification method based on feature fusion and local discriminant mapping. Guangdong Agric. Sci. 43 (10), 134–139. https://doi.org/10.16768/j.issn.1004-874X.2016.10.024.
Li, L., Zhang, S., Wang, B., 2022. Apple Leaf Disease Identification with a Small and Imbalanced Dataset Based on Lightweight Convolutional Networks. SENSORS, 22(1), https://doi.org/10.3390/s22010173.
Liu, Z., Lin, Y., Cao, Y., Hu, H., Zhang, Z., Lin, S., Guo, B., Swin Transformer: Hierarchical Vision Transformer using Shifted Windows, 2021. [Online], https://arxiv.org/abs/2103.14030.
Pandiyan, S., Ashwin, M., Manikandan, R., Karthick, R, Anantah, R., 2020. Heterogeneous Internet of Things organization predictive analysis platform for apple leaf diseases recognition. Computer Communications, 154(12), 99-110, https://doi.org/10.1016/j.comcom.2020.02.054.
Qin, Z., Zhang, P., Wu, F., Li, X., 2021. Fcanet: Frequency channel attention networks. Preprint at: https://arxiv.org/abs/ 2012.11879.
Ramachandran, P., Parmar, N., Vaswani, A., Bello, I., Levskaya, A., Shlens, J., 2019. Stand-alone self-attention in vision models. Advances in Neural Information Processing Systems, 32. https://proceedings.neurips.cc/paper/2019/hash/3416a75f4cea9109507cacd8e2f2aefc-Abstract.html.
Shi Y., Huang, W., Zhang, S., 2017. Apple disease recognition based on two-dimensionality subspace learning. Comput. Eng. Appl. 53 (22), 180–184. ISSN 1002-8331, https://doi.org/10.3778/j.issn.1002-8331.1605-0073.
Sun, H., Xu, H., Liu, B., He, D., He, J., Zhang, H., Geng, N., 2021. MEAN-SSD: A novel real-time detector for apple leaf diseases using improved light-weight convolutional neural networks, Computers and Electronics in Agriculture. 189, https://doi.org/10.1016/j.compag.2021.106379.
Thapa, R., Zhang, K., Snavely, N., Belongie, S., Belongie, S., Khan, A., 2020. The Plant Pathology Challenge 2020 data set to classify foliar disease of apples, Applications in Plant Sciences, 8,9. https://doi.org/10.1002/aps3.11390.
Wang, P., Niu, T., Mao, YR., Zhang, Z., Liu, B., He, DJ., 2021. Identification of Apple Leaf Diseases by Improved Deep Convolutional Neural Networks with an Attention Mechanism. Frontiers in Plant Science, 12, https://doi.org/10.3389/fpls.2021.723294.
Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., Hu, Q., 2020. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 11531-11539. https://doi.org/10.1109/CVPR42600.2020.01155.
Wang, X., Girshick, R., Gupta A., He, K., 2018. Non-local Neural Networks. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 7794-7803, https://doi.org/10.1109/CVPR.2018.00813.
Woo, S., Park, J., Lee, Y., Kweon, I., CBAM: Convolutional Block Attention Module, 2018, [online], http://arxiv.org/abs/1807.06521.
Yang, B., Bender, G., Le, Q. V., & Ngiam, J.,2019. Condconv: Conditionally parameterized convolutions for efficient inference. Advances in Neural Information Processing Systems, 32.

Yang, Z., Zhu, L., Wu, Y., Yang, Y., 2020. Gated channel transformation for visual recognition. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 11 794–11 803. https://doi.org/10.1109/CVPR42600.2020.01181.

Yu, H., Son, C., 2020. Leaf Spot Attention Network for Apple Leaf Disease Identification. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 229-237, https://doi.org/10.1109/CVPRW50498.2020.00034.

Zhong, Y., Zhao, M., 2020. Research on deep learning in apple leaf disease recognition, Computers and Electronics in Agriculture,168, https://doi.org/10.1016/j.compag.2019.105146.

Zhou, C., Zhou, S., Xing, J., Song, J., 2021. Tomato Leaf Disease Identification by Restructured Deep Residual Dense Network. IEEE Access, vol. 9, pp. 28822-28831, https://doi.org/10.1109/ACCESS.2021.3058947.

Fig.1. Part of the collected raw apple leaf disease in orchard

Fig.2. Part of labelled apple leaf disease images

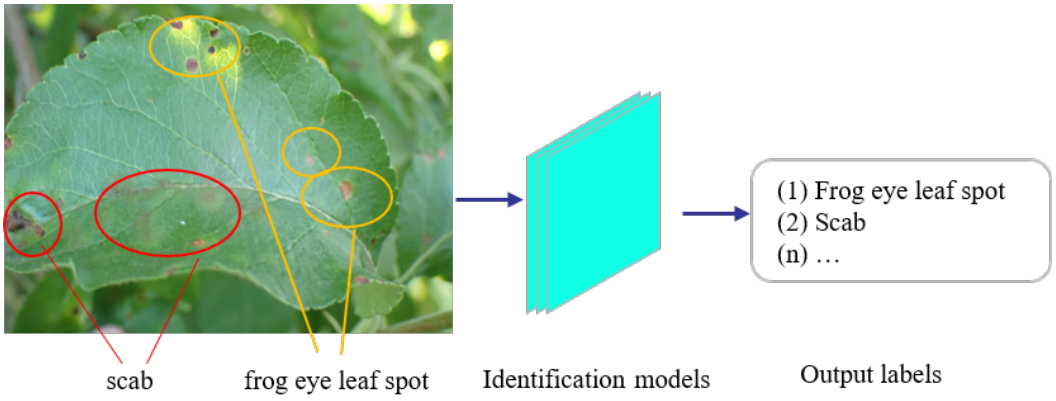

Fig.3. Multi-label learning for leaf disease identification

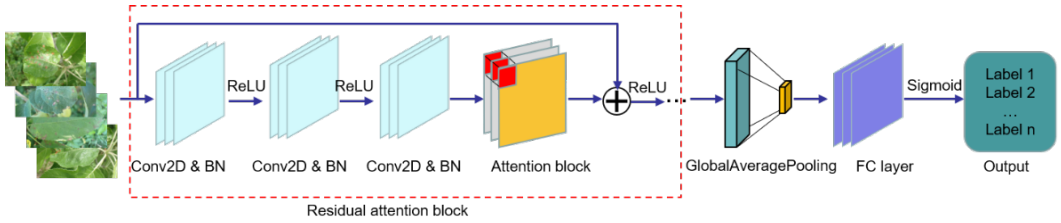

Fig.4. The architecture of RAMDI model

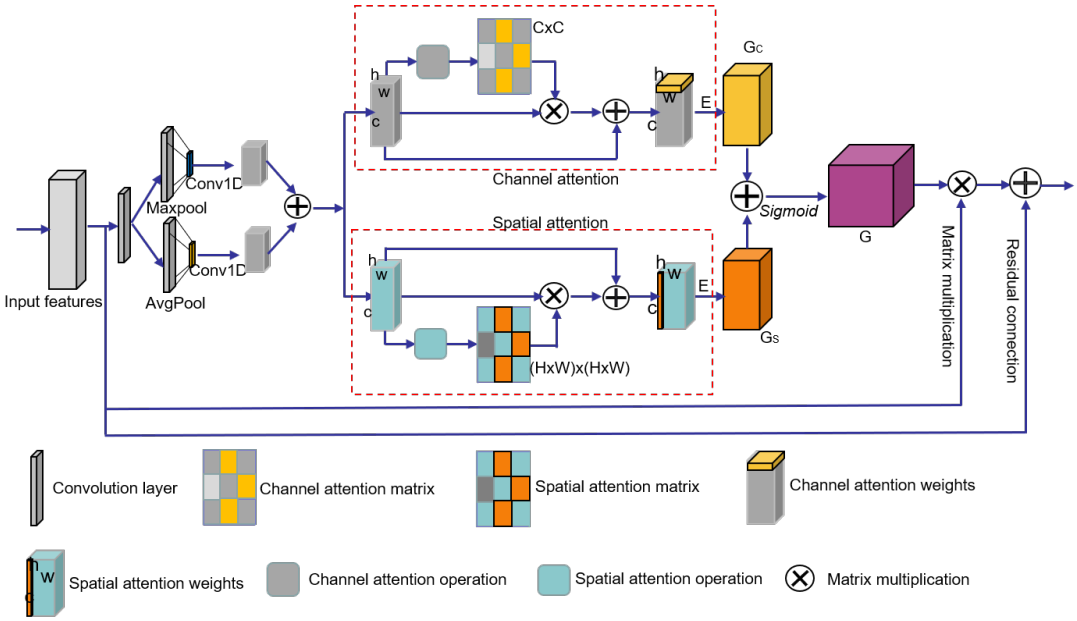

Fig.5. The architecture of residual attention block

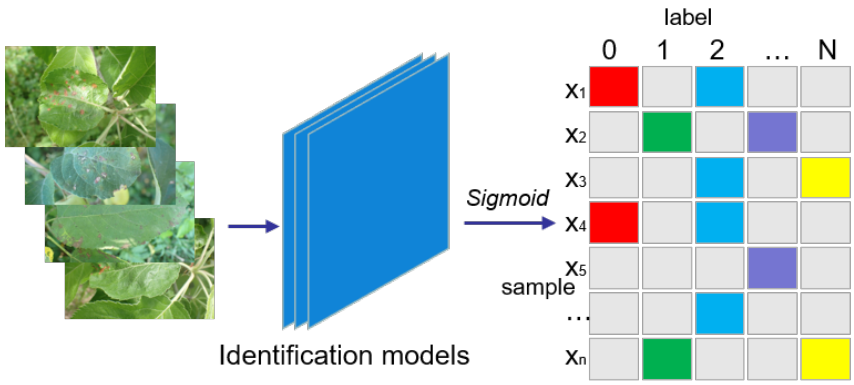

Fig.6. The flow chart of multi-label learning

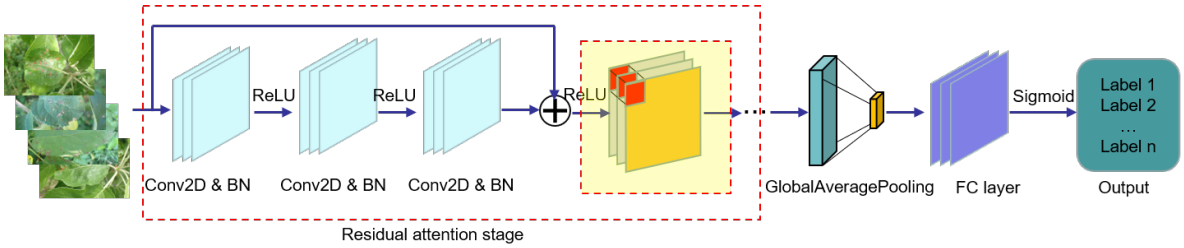

Fig.7. Attention mechanisms embedded in residual attention stage

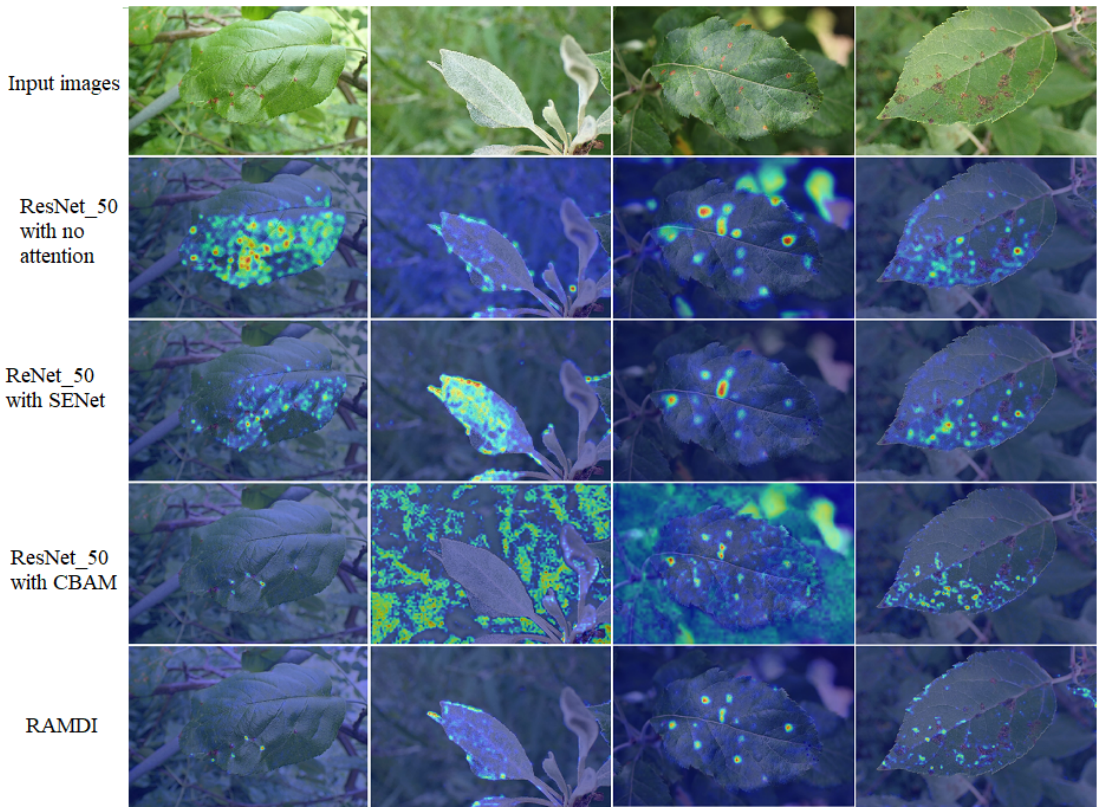

Fig.8. The visualization of different models

Table 1. The hyperparameters in the proposed method

Hyperparameter	Value
Training strategy	stratified 5-fold cross-validation
Optimization function	Adam
Loss function	Binary cross-entropy, Dice
Activation function	ReLU
Batch size	64
epochs	200

Table 2. Experimental results on test set

Approaches		Evaluation metrics							
		Acc	**Loss**	**Pre**	**Recall**	**Fs**	**HS**	**HL**	**mAPs**
Dice loss function	Resnet-50 (baseline)	0.959	0.020	0.982	0.984	0.980	0.975	0.0102	0.959
	Resnet-50_SENet_block	0.965	0.019	0.983	0.983	0.981	0.977	0.0096	0.961
	Resnet-50_SENet_stage	0.961	0.020	0.981	0.983	0.979	0.971	0.0104	0.959
	Resnet-50_CBAM_block	0.961	0.020	0.982	0.982	0.979	0.974	0.0107	0.954
	Resnet-50_CBAM_stage	0.945	0.029	0.972	0.975	0.970	0.963	0.0153	0.935
	Resnet-50_ECA_block	0.964	0.019	0.983	0.984	0.981	0.976	0.0097	0.964
	Resnet-50_ECA_stage	0.951	0.025	0.976	0.979	0.974	0.968	0.0130	0.947
	Resnet-50_ours_stage	0.954	0.026	0.975	0.979	0.974	0.969	0.0127	0.953
	Resnet-50_ours_block	0.954	0.024	0.978	0.980	0.976	0.970	0.0121	0.950
BCE loss function	Resnet-50 (baseline)	0.958	0.021	0.981	0.975	0.978	0.974	0.0078	0.972
	Resnet-50_SENet_block	0.966	0.021	0.985	0.982	0.983	0.980	0.0067	0.974
	Resnet-50_SENet_stage	0.971	0.019	0.985	0.979	0.981	0.978	0.0066	0.973
	Resnet-50_CBAM_block	0.969	0.022	0.980	0.978	0.978	0.975	0.0078	0.969
	Resnet-50_CNAM_stage	0.970	0.023	0.981	0.979	0.979	0.976	0.0077	0.971
	Resnet-50_ECA_block	0.974	0.021	0.985	0.982	0.983	0.980	0.0065	0.976
	Resnet-50_ECA_stage	0.973	0.021	0.985	0.981	0.982	0.979	0.0069	0.972
	Resnet-50_ours_stage	0.968	0.023	0.981	0.977	0.978	0.975	0.0079	0.968
	Resnet-50_ours_block	**0.976**	**0.019**	**0.988**	**0.985**	**0.986**	**0.983**	**0.0054**	**0.979**
Swin transformer		0.961	0.022	0.982	0.982	0.979	0.974	0.0117	0.954
Accuracy (Acc), Loss, Precision (Pre), Recall, F1-score(F1s), Hamming Score (HS), Hamming Loss (HL), and mAPs were all collected under the same experiment environment.									