

Research Article

BLSENet: A Novel Lightweight Bilinear Convolutional Neural Network Based on Attention Mechanism and Feature Fusion Strategy for Apple Leaf Disease Classification

Tianyu Fang,¹ Jialin Zhang,² Dawei Qi^(b),¹ and Mingyu Gao ^(b)

¹College of Mechanical and Electrical Engineering, Northeast Forestry University, Harbin 150040, China ²School of Computer and Information Engineering, Harbin University of Commerce, Harbin 150028, China ³School of Mechatronics Engineering, Harbin Institute of Technology, Harbin 150001, China

Correspondence should be addressed to Dawei Qi; qidw9806@126.com and Mingyu Gao; gmy0523@foxmail.com

Received 6 July 2023; Revised 23 January 2024; Accepted 27 January 2024; Published 6 February 2024

Academic Editor: Francesco Genovese

Copyright © 2024 Tianyu Fang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Accurate identification of apple leaf diseases is of great significance for improving apple yield. The lesion area of the apple leaf disease image is small and vulnerable to background interference, which easily leads to low recognition accuracy. To solve this problem, a lightweight bilinear convolutional neural network (CNN) model named BLSENet based on attention mechanism is designed. The model consists of two subnetworks, and each subnetwork is embedded with a Squeeze-and-Excitation (SE) module. By using the feature extraction ability of the two subnetworks and combining the bilinear feature CONCAT operation, the multiscale features of the image are obtained. Compared with the unimproved model LeNet-5 (84.63%), BLSENet has higher accuracy in the test set, which indicates that SE module and bilinear feature fusion have a positive effect on the performance of the model, and BLSENet has the ability to identify apple leaf diseases. The model has achieved the expected goal and can provide technical support for accurate identification and real-time monitoring of apple disease images.

1. Introduction

Apples are often attacked by diseases during the growth process [1]. Accurate identification of the types of diseases, timely prevention, and control are essential to improve the yield of apples [2–4]. At present, plant disease recognition has become an important research direction in the field of image recognition and intelligent agriculture.

Traditional machine learning algorithms need to classify images after extracting features [5]. The feature extraction process is time-consuming and labor-intensive, and the classification model has weak generalization ability and poor recognition effect [6]. Zhang et al. proposed an apple leaf disease identification method based on image processing technology and pattern recognition method. The RGB model was transformed into HSI, YUV, and grayscale, with background removal based on a specific threshold. The approach using region growing algorithm (RGA), genetic algorithm (GA), correlation-based feature selection (CFS), and support vector machine (SVM) achieved over 90% accuracy in recognizing various apple leaf diseases [7]. Bracino et al. proposed a machine learning model that can detect and classify the three most common apple diseases. The color and texture features of a single apple leaf image were extracted and selected. By comparing KNN, ANN, and GPR, it is determined that the GPR model with ARD squared kernel function is the best model [8]. Khan et al. employed contrast enhancement and a strong correlationbased segmentation method to segment images, optimizing the segmentation results through expectation maximization (EM). They utilized GA to extract features from the fused images and achieved significant classification accuracy using One-vs-All M-SVM [9]. Al-bayati et al. proposed a method for detecting apple leaf diseases using deep neural network (DNN). They employed Speeded Up Robust Features (SURF) for feature extraction, and the Grasshopper

Optimization Algorithm (GOA) is used for feature optimization; good results have been achieved [10]. However, the selection of this feature relies heavily on human experience and has great uncertainty. It requires specific data preprocessing to obtain better experimental results [11]. Because traditional feature extraction and recognition methods are not end-to-end operations, this is not conducive to rapid real-time detection in practical applications.

In recent years, as deep learning can automatically extract disease features and avoid manual dependence, a series of research results have been achieved in crop disease recognition. Rohini et al. proposed a model based on CNN to classify apple leaf images into diseased and undiseased. In the construction of the CNN model, the combination of convolution layer, ReLU, and max-pooling layer is considered. This task represents a binary classification problem. The proposed model is effectively implemented on the considered dataset with an accuracy of 91.11% [12]. Singh et al. used three pretrained CNN models to identify diseases in the Beans Leaf image dataset. In addition, different optimization techniques are used to highlight the performance differences of different CNN models. The experimental results show that the performance of EfficientNetB6 is better than other models, and the accuracy rate is 91.74% [13]. Kumar et al. proposed a strategy based on transfer learning, using the learned VGG-16/VGG-19 CNN network to estimate the severity of tomato leaf disease. In addition, the author performs hyperadjustment on the hyperparameters of the pretrained CNN model to improve its effectiveness. In order to evaluate the performance of the fine-tuned CNN model, the study measures the accuracy and loss values after multiple iterations on the training and validation datasets. Compared with another CNN model evaluated on the same dataset, VGG-16 shows higher classification accuracy (92.46%) [14]. Ding et al. proposed a new apple leaf disease recognition model named RFCA ResNet. This model has dual attention mechanism and multiscale feature extraction ability, which can reduce the adverse effects of complex background on recognition results. In addition, by combining the use of the class balance technique in conjunction with focal loss, the adverse effects of imbalanced datasets on classification accuracy can be effectively reduced. The RFB module can expand the receptive field and realize multiscale feature extraction. The accuracy of RFCA ResNet is 89.61%. It is superior to other methods and has good generalization performance, which has certain theoretical significance and practical value [15]. Gaikwad et al. used CNN to classify leaf disease. The author collected datasets from a real-time environment, with a total of 14181 images and 10 class labels. The experiment used 3 different versions of datasets: color, black and white, and grey images. These datasets are trained on AlexNet and SqueezeNet and use the same hyperparameters. The recognition accuracy of the two models is basically the same, and the classification accuracy of color images is 86.8% and 86.6%, respectively, indicating that color images are effective for classification [16]. In recent years, researchers have used various deep learning networks and frameworks for experiments. With the deepening of

research, it is currently the best choice to use deep learning to classify and identify apple leaf diseases [17]. Based on the aforementioned literature, we have discovered the diversity and complexity of the shape and color of diseases, which poses a challenge for achieving high-precision disease identification. While existing research encompasses various methods, including traditional feature extraction and deep learning techniques, the considerable variability in diseases has a notable impact on recognition accuracy. This diversity may result in existing models being unable to effectively capture and distinguish different disease features under certain circumstances, thereby limiting their practical applicability. To address this challenge, our focus has been on the multiscale extraction of disease features, incorporating methods such as multiscale feature fusion and employing more sophisticated deep learning architectures. These approaches aim to enhance the robustness of disease recognition systems by comprehensively capturing the complex characteristics of diseases. Therefore, this paper proposes a new CNN model, which can provide technical support for accurate identification and real-time monitoring of apple disease images.

In this paper, a bilinear classification model based on attention mechanism and feature fusion strategy named BLSENet is proposed for the classification of apple leaf diseases. The next arrangement and structure of this article are as follows. Firstly, the apple leaf disease dataset is presented, and the proposed network model BLSENet is introduced. Subsequently, the experimental results are described and analyzed in Section 3. The feasibility of the proposed model is verified by adjusting the model parameters and the ablation experiment of the model. Finally, the advantages and disadvantages of the proposed model are analyzed, and the future research direction is determined on this basis.

2. Methodology

2.1. Dataset. The dataset was collected from the College of Artificial Intelligence, Southwest University (as shown in Figure 1) [18–21]. The collected images encompass diverse diseases, each meticulously captured by skilled professionals using high-resolution cameras under appropriate lighting conditions to ensure image quality and clarity. Following the collection, the images underwent initial screening, retaining samples with representative disease features. To ensure data accuracy, each image was annotated for disease types by expert plant pathology specialists to guarantee precise and consistent labeling. The dataset contains nine types of apple leaf disease, including Health, Alternaria leaf spot, Brown spot, Frogeye leaf spot, Grey spot, Mosaic, Powdery mildew, Rust, and Scab.

The number of datasets used for the experiment is shown in Table 1. In the dataset, a total of 14582 images are included. 8754 images are randomly selected as the training set, 2913 pictures (accounting for 20% of the dataset) as the verification set, and the remaining 2915 pictures (accounting for 20% of the dataset) as the test set, as shown in Table 2.



FIGURE 1: Nine types of apple leaf disease: (a) Health, (b) Alternaria leaf spot, (c) Brown spot, (d) Frogeye leaf spot, (e) Grey spot, (f) Mosaic, (g) Powdery mildew, (h) Rust, and (i) Scab.

TABLE 1: Number of images in the apple leaf disease.

Category	Number
Health	516
Alternaria leaf spot	417
Brown spot	411
Frogeye leaf spot	3181
Grey spot	339
Mosaic	371
Powdery mildew	1184
Rust	2753
Scab	5410

TABLE 2: Division of datasets.

	Training set	Validation set	Test set	Total dataset
Quantity	8754	2913	2915	14582
Proportion	60%	20%	20%	100%

2.2. Model of Deep Convolutional Neural Network with Improvements

2.2.1. Multiscale Information Fusion Strategy. BLSENet is a bilinear CNN model. It is a new technology in fine-grained image recognition [22]. It has a good classification effect in terms of inability to distinguish category calculations with subtle visual differences [23]. The structure of BLSENet is shown in Figure 2. The input image is subjected to multiple Convolutions [24], Pooling [25], and BatchNormal [26] operations by two improved LeNet-5 CNNs, and two image features extracted by the CNN network are obtained. Then, the image features extracted by the CNN network are combined with the CONCAT operation to form the bilinear feature vector of the image [27]. Finally, the feature is classified by the fully connected layer classifier to obtain the probability of the identified category.



FIGURE 2: Structure of the proposed CNN.

LeNet, also known as LeNet-5, is a classical CNN proposed by Lecun [28]. It is one of the origins of modern CNNs. It has an input layer, two convolutional layers, two pooling layers, and three fully connected layers [29]. The improved LeNet-5 is used in BLSENet named A model; the two fully connected layers of A model are removed and replaced with SE modules. Then, a BatchNormal layer is added behind the first convolutional layer of the A model, which is named the B model. The B model is used as the upper branch network, the A model is used as the lower branch network, and two feature vectors named FC11 and FC21 with a dimension of 1×120 are output. The vector obtained by cascading FC11 and FC21 is named FC31, with a size of 1×240 . Subsequently, FC31 is reduced in dimension and a vector named FC32 with a dimension of 1×50 is obtained. Finally, the output of the fully connected layer is set to 9 to represent the category of leaf diseases.

2.2.2. Attention Mechanism Based on SE Module. The SE (Squeeze-and-Excitation Network) module is a computing unit; it can recalibrate the weight of the feature channel [30]. At the same time, the module can adaptively enhance the feature channel of the contrast information of the infrared image and suppress the irrelevant feature channel [31]. In this network, the SE module contains a Squeeze-and-Excitation operation. The training process is divided into two stages: the first stage is Squeeze and the second stage is Excitation.

Figure 3 shows the structure of the SE module. We hope to enhance the learning of convolution features by simulating the interdependence of channels so that the network can be sensitive to the information features that can be utilized in subsequent transformations. Therefore, our goal is to give it the opportunity to obtain global information, further improve the accuracy of the network by squeeze and excitation, and then send the filter to the next conversion. In recent years, SE modules have been widely used in deep learning to improve network performance. In many research fields, many network architectures use SE modules in the network to help improve the performance of the original network [32–35]. The structure is shown in Figure 3. This method is simple and easy to embed into the CNN framework, and the computational complexity increases little, but better results are obtained.

2.3. Model Training Details. The CNN model proposed in this paper is based on PyTorch which is an open-source deep learning library. The experimental process was carried out on a workstation equipped with the Intel(R) Core (TM) i9-10980XE CPU @ 3.00 GHz 3.00 GHz and the 24 GB NVIDIA GeForce RTX 3090 GPU. The experimental environment is shown in Table 3.

In the experiment, we set epoch = 200, batch size = 16, and initial learning rate = 0.0001 according to the experience, and Adam as the optimizer and cross-entropy as the loss function is used to train the network. The entire training parameters are shown in Table 4.

3. Experimental Results and Analysis

3.1. Training Results. The training result curve of the BLSENet network is shown in Figure 4. Accuracy is defined as the proportion of correctly classified samples by the model among all predictions. Loss is defined as the metric measuring the difference between the predictions of model and the actual labels during the training process, with the goal of minimizing this difference. It can be seen from Figure 4 that a total of 200 epochs were performed in the experiment. Finally, the accuracy rate on the test set is 93.58% and a good training effect is achieved.

On the Apple leaf disease dataset, the relationship between the loss value of the training set and the number of epochs is shown in Figure 4(a). The loss value on the training set decreases with the increase in the number of epochs, and the loss in the training set decreases from about 1.26 to about 0.34. The relationship between accuracy and the number of epochs is shown in Figure 4(b). As the number of epochs increases, the accuracy in the validation set gradually increases. The training set tends to be stable after the number of epochs is greater than 100, and its accuracy is about 93%.



TABLE 3: Experimental environment.

Hardwa	re environment		Software environment
CPU	Intel (R) Core (TM) i9-10980XE CPU @ 3.00 GHz 3.00 GHz	Environment configuration	PyTorch-GPU 1.8.0 + Python 3.7.10 + cuda 11.1 + cudnn 8.0.5
Graphics card	NVIDIA GeForce RTX 3090 (24G)	_	

TABLE 4: Training hyperparameters.

Hyperparameters name	Hyperparameters value	Hyperparameters range
Batch size	16	8, 16,32, 64
Learning rate	0.0001	0.00001, 0.0001, 0.001, 0.01
Epoch	200	100, 200, 300
Optimizer	Adam	SGD, AdaGrad, Adam
Loss function	Cross-entropy	—



FIGURE 4: Training results of BLSENet on the leaf dataset. (a) Loss curve. (b) Accuracy curve.

As shown in Figure 5, the results of the BLSENet model in the test dataset are analyzed and a confusion matrix is established. From the diagram, the model has good recognition accuracy for Brown spot, Frogeye leaf spot, Powdery mildew, Rust, and Scab, with accuracies of 95%, 90%, 94%, 93%, and 95%, respectively. This may be because there are a large number of images for Frogeye leaf spot, Powdery mildew, Rust, and Scab. The model can obtain sufficient training, resulting in a higher recognition rate. On the other hand, although there are not many images for Brown spot, which are comparable to Health, Alternaria leaf spot, Grey spot, and Mosaic, higher accuracy can be obtained. This may



FIGURE 5: Confusion matrix of BLSENet on the apple leaf disease dataset: the left figure represents the actual recognized number and the right figure represents the ratio of the identified quantity to the total quantity.

be because defects with a relatively large area are less likely to be affected by the background, making them easier to be correctly recognized by the model.

3.2. Comparison and Analysis of Experimental Results. After the model was established, epoch parameters were set to 100, 200, and 300 to select the appropriate value. The training results are shown in Figure 6. As can be seen from the figure, as epoch increases, both loss and accuracy show better results on the training set, but there is no significant difference between the three values. Then, the three parameter values were tested on the test dataset, and their accuracies were 90.22%, 93.58%, and 93.48%, respectively, as shown in Table 5. It can be observed that epoch = 200 has the best performance on the test dataset, and it does not consume a lot of training time. Therefore, considering the accuracy of the model based on the above analysis, the value of 200 was selected as the epoch.

To select the appropriate batch size parameter, batch size was set to 8, 16, 32, and 64. The training results are shown in Figure 7. As the batch size decreases, the network has better training results on the training set. The training results for batch size = 8 and batch size = 16 are similar. On the test dataset, their accuracies were 92.97%, 93.58%, 91.63%, and 90.57%, respectively, as shown in Table 6. It can be observed that batch size = 16 has the best performance on the test dataset. Therefore, considering the accuracy of the model based on the above analysis, the value of 16 was selected as the batch size.

To select the appropriate learning rate, three learning rates of 0.01, 0.001, 0.0001, and 0.00001 were tested on the test dataset. The training results are shown in Figure 8. From the figure, the training effect with a learning rate of 0.00001 is the worst because the learning rate is too small, which slows down the training efficiency. When the learning rate is 0.01, as can be seen from Figure 8(b), the accuracy curve is very unstable, which may be due to the learning rate being too large and making it difficult to find appropriate weight parameters. The training results with learning rates of 0.001 and 0.0001 are similar. Then, the four learning rates were tested on the test dataset, and the experimental results are shown in Table 7. Their accuracies were 91.60%, 91.34%, 93.58%, and 92.08%, respectively. Therefore, a learning rate of 0.0001 was selected based on the above analysis.

The optimization algorithm is very important for the performance of the model. The SGD [36], AdaGrad [37], and Adam [38] optimization algorithms were used to train the BLSENet in this paper, and their convergence speeds were compared. Figure 9 shows the training results of these three optimization algorithms. From the figure, the loss values using SGD and AdaGrad converge around 1.1 and the convergence effect is relatively poor as epoch increases. The accuracy values converge around 60%, which does not achieve the target accuracy. The results indicate that the model using the Adam algorithm has the fastest convergence speed and the best recognition effect. Then, the three optimizers were tested on the test dataset, and the experimental results are shown in Table 8. Therefore, Adam was selected as the optimization algorithm based on the above analysis.

The BLSENet is an improved model based on LeNet-5. To verify the improvement of the improved model compared to the original LeNet-5 model, BLSENet and LeNet-5 were compared on the test dataset. From Table 9, the recognition results of BLSENet are better than those of LeNet-5, with accuracies of 93.58% and 84.63%, respectively. The bilinear LeNet-5 combined with the SE module can improve the accuracy of LeNet-5 by 8.95%. Based on the above analysis, bilinear LeNet-5 combined with the SE module can improve the performance of the model.

In addition, we conducted some ablation experiments to analyze BLSENet. The results are shown in Table 9. As mentioned earlier, the concept of BLSENet is derived from LeNet-5 and combines the SE module. Therefore, we compared LeNet-5, LeNet-5 and LeNet-5, LeNet-5 + SE and



FIGURE 6: Training results of BLSENet with different epochs on the apple leaf disease dataset. (a) Loss curve. (b) Accuracy curve.

TABLE 5: Recognition results with different ep
--

Hyperparameters	Value	Number	Accuracy (%)
	100	2630	90.22
Epoch	200	2728	93.58
	300	2725	93.48

The bold values indicate the data with the best results.



FIGURE 7: Training results of BLSENet with different batch sizes on the apple leaf disease dataset. (a) Loss curve. (b) Accuracy curve.

Hyperparameters	Value	Number	Accuracy (%)
Batch size	8	2710	92.97
	16	2728	93.58
	32	2671	91.63
	64	2640	90.57

The bold values indicate the data with the best results.



FIGURE 8: Training results of BLSENet with different learning rates on the apple leaf disease dataset. (a) Loss curve. (b) Accuracy curve.

	TABLE 7:	Recognition	results	with	different	learning	rates.
--	----------	-------------	---------	------	-----------	----------	--------

Hyperparameters	Value	Number	Accuracy (%)
Learning rate	0.01	2670	91.60
	0.001	2680	91.34
	0.0001	2728	93.58
	0.00001	2684	92.08

The bold values indicate the data with the best results.

The bold values indicate the data with the best results.



FIGURE 9: Training results of BLSENet with different optimization algorithms on the apple leaf disease dataset. (a) Loss curve. (b) Accuracy curve.

TABLE 8: Recognition results with different optimization algorithms.

Hyperparameters	Name	Number	Accuracy (%)
	SGD	1820	62.44
Optimization algorithms	AdaGard	1861	63.84
	Adam	2728	93.58

8

Journal of Food Quality

Model	Number	Accuracy (%)
LeNet-5	2467	84.63
LeNet-5, LeNet-5 combined (bilinear)	2699	92.59
LeNet-5 + SE, LeNet-5 combined (bilinear)	2377	81.54
LeNet-5, LeNet-5 + SE combined (bilinear)	2700	92.62
BLSENet (bilinear)	2728	93.58





FIGURE 10: Training results of ablation experiment of BLSENet on the apple leaf disease dataset. (a) Loss curve. (b) Accuracy curve.

TABLE 10: Comparison of the	proposed method with	previous detection methods	for apple leaf disease
-----------------------------	----------------------	----------------------------	------------------------

Detection method	Accuracy (%)
MobileNet [17]	73.50
InceptionV3 [17]	75.59
ResNet152 [17]	77.65
INAR-SSD (SSD with inception module and rainbow concatenation) [39]	78.80
DenseNet-121 and regression function [40]	93.51
DenseNet-121 and multilabel classification function [40]	93.31
BLSENet	93.58

LeNet-5, LeNet-5 and LeNet-5 + SE, and Double LeNet-5 + SE (BLSENet). The training results are shown in Figure 10. Their training curves all reach similar results as epoch increases. To test the generalization ability of the models on the test dataset, these models were compared on the same test dataset, and the results are shown in Table 9. The experimental results show that the recognition ability of LeNet-5 is the worst, and the accuracy of LeNet-5 and LeNet-5, LeNet-5 and LeNet-5 + SE, Double LeNet-5 + SE (BLSENet) is similar, with an accuracy of between 92% and 94%. The recognition effect of LeNet-5 + SE and LeNet-5 is the worst, with a recognition rate of 81.54%. This is an interesting phenomenon, and we will continue to research it in the future.

Finally, the author presents a detailed table showcasing the accuracy achieved by the proposed neural network model in experiments and compares it with other models from relevant literature, as shown in Table 10. Excitingly, the table clearly demonstrates the outstanding performance of the proposed model in terms of accuracy, outperforming models from other literature and yielding the best results, approximately 93.58%. This finding not only highlights the superiority of the proposed model but also provides robust support for further advancements in the research field.

4. Conclusion

In this paper, we proposed an apple leaf disease recognition method called BLSENet based on attention mechanism, lightweight CNN, and bilinear CNN framework. By embedding the SE module into the end of LeNet-5 and combining it with bilinear pooling, BLSENet was constructed to extract image features of apple leaf diseases. BLSENet has higher accuracy in the test dataset compared with the unimproved model LeNet-5 (84.63%), which indicates that the SE module and bilinear feature fusion have a positive effect on the performance of the model and BLSENet has the ability to recognize apple leaf diseases accurately. The model has achieved the expected goal, which can provide technical support for accurate identification and real-time monitoring of apple disease images. In our future work, we will continue to focus on deep learning models capable of assessing the severity of apple leaf diseases. Simultaneously, we aim to deploy the model on devices such as unmanned aerial vehicles (UAVs) to achieve precise remote sensing for agricultural monitoring. This is a challenging task but is a pressing demand in the field.

Data Availability

The datasets, codes, and weight files used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (NSFC) under Contract no. 31570712.

References

- L. Pereira, A. Cunha, and C. Almeida-Aguiar, "Portuguese propolis from Caramulo as a biocontrol agent of the apple blue mold," *Food Control*, vol. 139, Article ID 109071, 2022.
- [2] H. Yu, X. Cheng, C. Chen et al., "Apple leaf disease recognition method with improved residual network," *Multimedia Tools and Applications*, vol. 81, no. 6, pp. 7759–7782, 2022.
- [3] R. Ding, Y. Qiao, X. Yang et al., "Improved ResNet based apple leaf diseases identification," *International Federation of Accountants-PapersOnLine*, vol. 55, no. 32, pp. 78–82, 2022.
- [4] K. Adem, M. M. Ozguven, and Z. Altas, "A sugar beet leaf disease classification method based on image processing and deep learning," *Multimedia Tools and Applications*, vol. 82, no. 8, pp. 12577–12594, 2023.
- [5] M. Gao, F. Wang, J. Liu et al., "Estimation of the convolutional neural network with attention mechanism and transfer

learning on wood knot defect classification," *Journal of Applied Physics*, vol. 131, no. 23, Article ID 233101, 2022.

- [6] M. Gao, F. Wang, P. Song, J. Liu, and D. Qi, "Blnn: multiscale feature fusion-based bilinear fine-grained convolutional neural network for image classification of wood knot defects," *Journal of Sensors*, vol. 2021, Article ID 8109496, 18 pages, 2021.
- [7] Z. Chuanlei, Z. Shanwen, Y. Jucheng, S. Yancui, and C. Jia, "Apple leaf disease identification using genetic algorithm and correlation based feature selection method," *International Journal of Agricultural and Biological Engineering*, vol. 10, no. 2, pp. 74–83, 2017.
- [8] A. A. Bracino, R. S. Concepcion, R. A. Bedruz, E. P. Dadios, and R. R. Vicerra, "Development of a hybrid machine learning model for apple (Malus domestica) health detection and disease classification," in *Proceedings of the 2020 IEEE 12th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)*, pp. 1–6, IEEE, Manila, Philippines, December 2020.
- [9] M. A. Khan, M. I. Lali, M. Sharif et al., "An optimized method for segmentation and classification of apple diseases based on strong correlation and genetic algorithm based feature selection," *IEEE Access*, vol. 7, pp. 46261–46277, 2019.
- [10] J. S. Al-bayati and B. B. Üstündağ, "Evolutionary feature optimization for plant leaf disease detection by deep neural networks," *International Journal of Computational Intelligence Systems*, vol. 13, no. 1, p. 12, 2020.
- [11] M. Gao, P. Song, F. Wang, J. Liu, A. Mandelis, and D. Qi, "A novel deep convolutional neural network based on ResNet-18 and transfer learning for detection of wood knot defects," *Journal of Sensors*, vol. 2021, Article ID 4428964, 16 pages, 2021.
- [12] V. Rohini and R. Jyothsna, "Disease detection in apple tree leaves using CNN algorithms," *Journal of Survey in Fisheries Sciences*, vol. 10, no. 4, pp. 1097–1101, 2023.
- [13] V. Singh, A. Chug, and A. P. Singh, "Classification of Beans leaf diseases using fine tuned CNN model," *Procedia Computer Science*, vol. 218, pp. 348–356, 2023.
- [14] R. Kumar, A. Chug, and A. P. Singh, "Plant leaf diseases severity estimation using fine-tuned CNN models," in *Proceedings of the 2023 6th International Conference on Information Systems and Computer Networks (ISCON)*, pp. 1–6, IEEE, Mathura, India, March 2023.
- [15] J. Ding, C. Zhang, X. Cheng et al., "Method for classifying apple leaf diseases based on dual attention and multi-scale feature extraction," *Agriculture*, vol. 13, no. 5, p. 940, 2023.
- [16] S. S. Gaikwad, S. S. Rumma, and M. Hangarge, "Fungi affected fruit leaf disease classification using deep CNN architecture," *International Journal on Information Technology*, vol. 14, no. 7, pp. 3815–3824, 2022.
- [17] C. Bi, J. Wang, Y. Duan, B. Fu, J. R. Kang, and Y. Shi, "MobileNet based apple leaf diseases identification," *Mobile Networks and Applications*, vol. 27, pp. 172–180, 2022.
- [18] D. Hughes and M. Salathé, "An open access repository of images on plant health to enable the development of mobile disease diagnostics," 2015, https://arxiv.org/abs/1511.08060.
- [19] J. Z. Feng and X. F. Chao, "Apple tree leaf disease segmentation dataset," *Science Data Bank*, 2022.
- [20] R. Thapa, K. Zhang, N. Snavely, S. Belongie, and A. Khan, "The Plant Pathology Challenge 2020 data set to classify foliar disease of apples," *Applications in plant sciences*, vol. 8, no. 9, Article ID 11390, 2020.

- [21] Q. Yang, S. Duan, and L. Wang, "Efficient identification of apple leaf diseases in the wild using convolutional neural networks," *Agronomy*, vol. 12, no. 11, p. 2784, 2022.
- [22] Y. Lyu, L. Jing, J. Wang, M. Guo, X. Wang, and J. Yu, "Siamese transformer with hierarchical concept embedding for finegrained image recognition," *Science China Information Sciences*, vol. 66, no. 3, Article ID 132107, 2023.
- [23] K. Song, X. S. Wei, X. Shu, R. J. Song, and J. Lu, "Bi-modal progressive mask attention for fine-grained recognition," *IEEE Transactions on Image Processing*, vol. 29, pp. 7006– 7018, 2020.
- [24] J. J. Liu, Q. Hou, M. M. Cheng, C. Wang, and J. Feng, "Improving convolutional networks with self-calibrated convolutions," in *Proceedings of the IEEE/CVF conference* on computer vision and pattern recognition, pp. 10096–10105, Seattle, WA, USA, June 2020.
- [25] Q. Hou, L. Zhang, M. M. Cheng, and J. Feng, "Strip pooling: rethinking spatial pooling for scene parsing," in *Proceedings of* the IEEE/CVF conference on computer vision and pattern recognition, pp. 4003–4012, Seattle, WA, USA, June 2020.
- [26] Y. Hou and L. Zheng, "Visualizing adapted knowledge in domain transfer," in *Proceedings of the IEEE/CVF conference* on computer vision and pattern recognition, pp. 13824–13833, Seattle, WA, USA, June 2021.
- [27] W. Wang, D. Tran, and M. Feiszli, "What makes training multi-modal classification networks hard?" in *Proceedings of* the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 12695–12705, Seattle, WA, USA, June 2020.
- [28] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradientbased learning applied to document recognition," *Proceedings* of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.
- [29] C. J. Burges, "Simplified support vector decision rules," *InIcml*, vol. 96, pp. 71–77, 1996.
- [30] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer* vision and pattern recognition, pp. 7132–7141, Salt Lake, UT, USA, June 2018.
- [31] Q. Xu, Y. Mei, J. Liu, and C. Li, "Multimodal cross-layer bilinear pooling for RGBT tracking," *IEEE Transactions on Multimedia*, vol. 24, pp. 567–580, 2022.
- [32] L. Yang, R. Y. Zhang, L. Li, and X. Xie, "Simam: a simple, parameter-free attention module for convolutional neural networks," in *Proceedings of the International conference on machine learning*, pp. 11863–11874, PMLR, Westminster, UK, July 2021.
- [33] N. Wang, S. Ma, J. Li, Y. Zhang, and L. Zhang, "Multistage attention network for image inpainting," *Pattern Recognition*, vol. 106, Article ID 107448, 2020.
- [34] S. Vandenhende, S. Georgoulis, and L. Van Gool, "Mti-net: multi-scale task interaction networks for multi-task learning," in *Proceedings of the Computer Vision–ECCV 2020: 16th European Conference*, pp. 527–543, Springer International Publishing, Glasgow, UK, August, 2020.
- [35] H. R. Joze, A. Shaban, M. L. Iuzzolino, and K. Koishida, "MMTM: multimodal transfer module for CNN fusion," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 13289–13299, Seattle, WA, USA, June 2020.
- [36] B. Woodworth, K. K. Patel, S. Stich et al., "Is local SGD better than minibatch SGD?" in *Proceedings of the International Conference on Machine Learning*, pp. 10334–10343, PMLR, Westminster, UK, November 2020.
- [37] K. Antonakopoulos, P. Mertikopoulos, G. Piliouras, and X. Wang, "AdaGrad avoids saddle points," in *Proceedings of*

the International Conference on Machine Learning, pp. 731–771, PMLR, Westminster, UK, June 2022.

- [38] S. Bock and M. Weiß, "A proof of local convergence for the Adam optimizer," in *Proceedings of the 2019 international joint conference on neural networks (IJCNN)*, pp. 1–8, IEEE, Budapest, Hungary, July 2019.
- [39] P. Jiang, Y. Chen, B. Liu, D. He, and C. Liang, "Real-time detection of apple leaf diseases using deep learning approach based on improved convolutional neural networks," *IEEE Access*, vol. 7, pp. 59069–59080, 2019.
- [40] Y. Zhong and M. Zhao, "Research on deep learning in apple leaf disease recognition," *Computers and Electronics in Agriculture*, vol. 168, Article ID 105146, 2020.