A Method for Sugarcane Disease Identification Based on Improved ShuffleNetV2 Model

Yanlei Xu
Jilin Agricultural University
Zhenhuan Lu
Jilin Agricultural University
Yuting Zhai
Jilin Agricultural University
Yubin Jiao
Army Academy of Armored Force
Yang Zhou (zhouyang@jlau.edu.cn)
Jilin Agricultural University

Research Article

Keywords: Computer vision, sugarcane, plant disease, multi-scale, attention mechanism

Posted Date: August 25th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3282256/v1

Additional Declarations: No competing interests reported.

Abstract

Rapid and accurate identification of sugarcane diseases is an important way to improve sugarcane yield. Therefore, this study proposes an improved model based on ShuffleNetV2 network (Im-ShuffleNetV2) for sugarcane disease identification. Firstly, we incorporated the ECA (Enhanced Channel Attention) attention mechanism into ShuffleNetV2, enhancing the network's ability to extract features and detect sugarcane lesion areas. Secondly, a new multi-scale feature extraction branch and Transformer module have been introduced, further improving the independent learning ability of the network. Finally, a large number of numerical results have demonstrated the advantages of the proposed model in terms of parameter size and sugarcane disease identification accuracy. Just as Im-ShuffleNetV2 only has a parameter of 0.4MB, it has significant advantages over parameters such as EfficientV2-S (55.6MB), MobileNetV2 (8.73MB), MobileViT XX small (3.76MB), FasterNetT2 (52.4MB), AlexNet (55.6MB), and MobileNetV3 Large (16.2MB). In addition, compared with the ShuffleNetV2 network, the accuracy has improved by 3.4%. This model not only improves the accuracy of sugarcane leaf disease detection, but also demonstrates the advantage of lightweight, providing valuable reference for future research in the field of sugarcane.

1 Introduction

Sugarcane is extensively grown in tropical and subtropical regions and serves as a crucial raw material for global sugar production, accounting for approximately 90%. The cultivation and processing of sugar are directly linked to the sustainable development of the sugar industry and are of great significance for enhancing people's quality of life. However, the diseases encountered during the sugarcane planting process are also on the rise, which significantly impacts the sustainable growth of the planting industry and results in substantial economic losses for farmers. Therefore, only by enhancing the prevention and control of sugarcane diseases can we ensure the smooth progress of agricultural production and ultimately achieve sustainable development in terms of economic benefits, social benefits, and environmental protection. Currently, the identification of sugarcane diseases is primarily carried out through manual detection, which has drawbacks including low efficiency in recognition, subjective judgment, and high labor costs, ultimately leading to a decrease in the scale of crop cultivation [1−4].

In recent years, with the continuous advancement of information technology and agricultural intelligence, deep learning has made remarkable strides in crop protection. By training on vast amounts of image data, it can swiftly and accurately identify, monitor, and predict crop diseases to provide farmers with more precise disease prevention and control programs. This facilitates efficient, rapid, and healthy agricultural development while effectively promoting modernization and sustainable agriculture [5−9]. In contrast to conventional machine learning techniques that rely on manually crafted features for texture recognition, Convolutional Neural Networks (CNNs) [10] employ end-to-end architectures to automatically extract image features, leading to enhanced classification accuracy and recognition speed of the network. However, many convolutional neural networks currently increase the number of network layers and sacrifice significant computing resources to improve recognition performance. Consequently, issues such as slow network training speed and an excessive number of parameters arise. For example, ResNet

[11], ResNeXt [12], and other networks have deep structures that deliver excellent recognition performance but cannot be widely used on mobile devices by most farmers due to their large number of parameters and calculations.

Yingying Wang et al. [13] propose a maize disease recognition model that utilizes an attention mechanism and achieves a recognition accuracy of 99.35%. Bin Liu et al. [14] demonstrate an improved deep learning model that achieves a recognition accuracy of 97.22% in identifying various grape diseases, including black rot. Yu et al. [15] propose a multi-step optimization approach for the ResNet architecture, achieving a high accuracy of 95.70% in detecting various apple diseases. Despite making significant progress in disease identification accuracy, it is currently unsuitable for deployment on mobile or embedded devices due to the absence of lightweight processing that reduces network parameters.

Therefore, to address the issues of recognition accuracy and model size faced by deep models in practical applications, this paper proposes an improved ShuffleNetV2[16] (Im-ShuffleNetV2) for sugarcane disease identification. This proposed network integrates the attention mechanism ECA and Transformer module to augment both channel attention and global context feature extraction capabilities of the network, thereby enabling a more focused analysis of sugarcane disease features. Furthermore, through multi-branch improvement, the fundamental unit can capture diverse scale information and enhance its ability to discern sugarcane disease textures. Consequently, this network enhances accuracy in identifying sugarcane diseases while simultaneously minimizing parameter count and memory usage.

2 Methods and Work

2.1 DataSet

In this study, kaggle's public sugarcane disease data set was used, including 5 different disease types, about 500 pieces each, namely Redrot, rust, Mosaic, Yellow leaf disease and healthy state. After screening, they were respectively Redrot (187), Rust (196), Mosaic (207), Yellow leaf disease (241) and Healthy state (209). The data set is divided in a 7:3 ratio. To avoid overfitting during training, the data set was enhanced three times, such as horizontal image flip, random 360-degree rotation, noise, random black box occlusion, random clipping, and contrast enhancement, etc., and the data was enhanced in the form of random enhancement combination. The enhanced training set contained 2920 images, including 524 for Redrot, 552 for rust, 580 for Mosaic, 676 for Yellow leaf disease, and 588 for health status. The test set contained 940 images, including 224 for Redrot, 232 for rust, 248 for Mosaic, 288 for Yellow leaf disease and 248 for health status. As shown in Fig. 1.

2.2 Improved Algorithm

In this study, the proposed improved ShuffleNetV2 network mainly consists of three parts. Firstly, ShuffleNetV2 was adopted as the backbone of the network, and its basic modules were optimized based on the idea of multi-scale feature fusion [18] to further improve the efficiency of extracting sugarcane disease features from the network. Secondly, the ECA attention mechanism has been added to make the network pay more attention to sugarcane disease areas, thereby further improving the accuracy and stability of network recognition. Finally, the Transformer module was added to enhance the learning ability of the network and further improve its accuracy in identifying sugarcane diseases. The proposed network structure diagram is shown in Fig. 2.

To effectively capture global information from the feature graph context and enhance he efficiency of the network, this study incorporates the ECA attention mechanism into the down sampling unit to augment its feature recognition capability while minimizing network size expansion. The Stage2, Stage3, and Stage4 modules are combined with enhanced basic units to boost the network's feature extraction capacity. Additionally, a Transformer module is introduced after the Stage2 module. By processing sugarcane disease feature maps using the Transformer module, more comprehensive global texture information can be obtained, enabling better discrimination between noise and disease-related texture features.

2.3.1 ShuffleNetV2 network backbone improvements

In this study, the main model selected for the sugarcane disease recognition task was the convolutional neural network ShuffleNetV2. This model was aimed at achieving better accuracy while maintaining a lightweight and deployable model. It incorporates two core operations: depth-separable convolution [19] and channel shuffle [16]. These operations not only ensure the model's lightweight nature but also maintain its efficient ability to extract feature information.

Figure 3 shows the structures of ordinary convolution and depth-separable convolution. In ordinary convolution, since the previous layer produces results with multiple channels, a corresponding number of filters is required for the convolution operation. Multiple convolution cores are used to convolve feature graphs from multiple channels, and their results are added together to generate a feature map with multiple channels. On the other hand, in depth-separable convolution, the feature graph is decomposed into single-channel feature graphs for each channel and then convolved one by one through channel-by-channel convolution [20].

Finally, point-by-point convolution [20] is used to convolve the resulting feature map with a filter of size 1 that has the same number of channels as in the previous layer to obtain the final required feature map. This technique reduces the number of calculations and parameters, improving the training efficiency of CNN’s. The input feature map is divided into several groups, and then the channel information between different groups is interleaved to enhance the model's representation ability without compromising network accuracy. As shown in Fig. 4, two input feature maps are divided into two groups. The channel information between these groups is then interleaved to generate new feature maps which are outputted as new feature information. This process improves the expressive capability of the model while maintaining network accuracy.

The original ShuffleNetV2 network backbone consists of a basic unit and a down sampling unit. As shown in Fig. 5, in the basic module (Fig. 5(a)), after Channel Split, the channel is divided into two parts, with each part having half the number of channels as the input feature map. The left branch carries out identity mapping processing, while the right branch undergoes 1×1 convolution followed by 3×3 channel-by-channel convolution and 1×1 point-by-point convolution. Concatenation and channel mixing are performed on both branches to ensure full integration of feature information from both branches. In the subsampling module (Fig. 5b), there is no channel division; instead, the feature map is directly inputted into both left and right branches. Both parts then undergo 1×1 convolution and 3×3 channel-by-channel convolution operations with a step size of 2. Finally, the two branches are merged together, and their channels mixed. Corrected sentence: This paper mainly improves the basic unit of the ShuffleNetV2 network. The feature extraction of the right branch is primarily completed by 3×3 depth separable convolution, which can maintain good accuracy while greatly reducing model size. However, considering that the dataset used in this study is mainly from a natural environment and contains a lot of noise, it may be difficult to extract feature information. To address this issue, we propose an improved basic unit shown in Fig. 5c. By incorporating multi-scale feature fusion concepts, branches with different convolution kernel sizes are added to perform multidimensional feature extraction of image information. This allows for extracting features and textures related to sugarcane local diseases at different scales and improving sugarcane disease recognition accuracy. On the right branch, we remove the original 3×3 depth separable convolution and 1×1 convolution and divide channels again into two branches based on channel division - namely 1×1 convolution and 3×3 convolution - with a channel division coefficient of four. This allows for extracting features and textures related to sugarcane local diseases at different scales and improving sugarcane disease recognition accuracy. On the right branch, we remove the original 3×3 depth separable convolution and 1×1 convolution and divide channels again into two branches based on channel division - namely 1×1 convolution and 3×3 convolution - with a channel division coefficient of four.

2.3.2 ECA attention mechanism

Due ECA [17] is an efficient channel attention mechanism proposed by Qi long Wang et al in 2020. In the CNN network, by strengthening the information interaction and convergence between adjacent channels, feature extraction and selection become more targeted and distinguishable, thereby improving model performance.

Compared with other attention mechanisms, the ECA attention mechanism can effectively avoid the impact of reduced input channels on channel attention learning and has higher computational efficiency and fewer model parameters. Because this study used datasets from natural environments with complex environmental backgrounds and noise effects, these factors can have a negative impact on recognition tasks. Especially in complex environmental backgrounds, useless feature information can sometimes mask the information of sugarcane disease features, resulting in the inability of the network to extract sugarcane disease feature information. By introducing ECA attention mechanism, information exchange between channels is strengthened, enabling the network to focus its attention on sugarcane disease features and suppressing the impact of complex background noise on network recognition accuracy.

As shown in Fig. 6, it is a schematic diagram and location of the ECA attention mechanism. ECA is located below the 1×1 convolution of the left branch in the down sampling module(b), which uses a 1-dimensional kernel size of K to facilitate information exchange between channels. Firstly, without changing the channel dimension, the ECA attention mechanism first adjusts the feature map resolution through a global average pool of size 1×1. Then, it utilizes 1D convolution with adaptive kernel size obtained from dimension C to achieve information exchange between channels, as shown in Eq. (a). Finally, the weight coefficients of the output feature map are obtained through the sigmoid function and multiplied with the weight coefficients of the original feature map to further enhance the feature extraction ability of sugarcane diseases.

$$K = \left| \frac{\log_2 C + b}{\gamma} \right|_{\text{odd}}$$

1

Where C represents the number of channels, b and γ It is a hyperparameter with values set to 1 and 2, respectively. In addition, odd represents K as odd.

2.3.3 Transformer module

The Transformer [21] model is a network model based on the self-attention mechanism, which is widely used in natural language processing and has unique advantages in vision. It compares with traditional convolutional neural networks and can better extract global context information from sugarcane disease feature maps. This enables easier learning of long-term dependencies and enhances its ability to extract sugarcane disease features while capturing relationships between different locations for more detailed and hierarchical feature representation. Besides, due to its use of the self-attention mechanism, Transformer can directly learn relevant features of sugarcane diseases, reducing attention scope and mitigating noise background interference in sugarcane disease recognition. Consequently, computational complexity is reduced as well, addressing one limitation faced by convolutional neural networks. Considering that the environmental background of the sugarcane disease dataset used in this paper is complex and noisy, traditional convolutional neural networks are unable to effectively learn local structures and global dependencies due to limited receptive fields and fixed kernel sizes. They are highly susceptible to noise interference and complex backgrounds. As a result, these networks struggle to identify characteristic textures and local information related to sugarcane diseases; hence it becomes necessary to introduce Transformer models as a solution.

The structure of the Transformer module is depicted in Fig. 7. With fewer parameters, the Transformer module can model both local and global information in the input tensor. Firstly, it extracts local feature information through local modeling [22]. Then, it comprehensively captures feature information through global modeling [22]. Next, the new feature map is fused with the original one to obtain more information. Finally, a 1×1 ordinary convolution is performed for dimensional adjustment.

Local modeling: First, a channel-by-channel convolution with a kernel size of 3 is used to extract and encode the local spatial information of the input feature map. Then, a 1×1 point-by-point convolution is employed to project the tensor into a high-dimensional space by learning the linear combination of input channels, which facilitates subsequent global modeling expansion.

Global modeling: The purpose is to enable the network to learn the global characteristics of spatial induction bias. Its global feature space encoding is mainly implemented by Transformer. However, since Transformer was initially designed for sequence-to-sequence tasks, it cannot be directly applied in the visual domain. Therefore, before being fed into Transformer, the feature map needs to be transformed into non-overlapping patches using an Unfold operation [22]. For each patch, Transformer is used to encode the relationships between patches. To ensure that the Transformer module retains both inter-patch relationships and intra-patch spatial order of pixels, all patches are then restored back to form a feature graph through a Fold operation [22]. Subsequently, a 1×1 general convolution is employed to project the feature map into a low-dimensional space and concatenate it with the original image. Finally, local spatial information of this concatenated feature map is further encoded using 3×3 convolution to extract more diverse spatial texture features.

3 Results and Analysis

The Adam [23] optimizer is used in the experiment with an initial learning rate of 0.01 and 100 training epochs using the cross-entropy loss function. The experiment is conducted on a Windows10 64-bit operating system, using Python version 3.7 and Pytorch framework for coding. The server utilizes an Intel Core (TM) i7-7820XCPU processor with 32GB memory and NVIDIA GeForce GTX TITAN X graphics processor (GPU) with 12GB memory.

3.1Evaluation Indicators

The improved ShufflenetV2 model is being used to further verify its recognition effect on the sugarcane disease dataset in this study. Under the same experimental conditions, the designed improved model is being compared with other deep learning models using the sugarcane disease image dataset of this study. The specific results can be seen in Fig. 7 and Table 1. To comprehensively evaluate the performance of the improved model, this paper adopts indicators such as Accuracy, Precision, Recall, and F1score to measure its ability in identifying sugarcane disease datasets. Accuracy represents the proportion of correctly predicted samples by the model out of total samples; precision represents the proportion of positively predicted samples that are truly positive; recall represents the proportion of positively predicted samples that are correctly identified as positive; F1 score is a comprehensive evaluation metric calculated as a harmonic average between accuracy and recall rates.

Additionally, network complexity and resource consumption are measured by their respective number and memory usage.

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN)$$

2

$$\text{Precision} = TP / (TP + FP)$$

3

$$\text{Recall} = TP / (TP + FN)$$

4

$$F1score = 2 * (Precision * Recall) / (Precision + Recall) \quad (5)$$

Where, TP (True Positives) and TN (True Negatives) represent the number of samples that the model predicted as positive and are actually positive and the number of samples that the model predicted as negative and are actually negative, respectively. FP (False Positives) and FN (False Negatives) represent the number of samples that the model predicted as positive but are actually negative, and the number of samples that the model predicted as negative but are actually positive, respectively.

3.2 The Comparison Results of the Initial Algorithm

According to the results in Table 1, AlexNet [18] and MobileVIT XX Small [22] performed poorly in terms of accuracy, with 87.90% and 85.88% respectively. Therefore, the low recognition accuracy limits the practical application of these networks. Networks such as EfficientNetV2 S [24] and FasterNetT2[25] performed well in terms of accuracy, at 90.48% and 92.10%, respectively. Besides, networks such as EfficientNetV2 S and FasterNetT2 performed well in terms of accuracy, reaching 90.48% and 92.10%, respectively. Unfortunately, EfficientNetV2 S and FasterNetT2 have larger parameter sizes and memory usage compared to ShuffleNetV2, making them unsuitable for mobile deployment. The accuracy of the lightweight network MobileNetV2 can reach 91.69%, with a parameter size of 2.20MB and a memory usage of 8.73MB; The accuracy of MobileNetV3 Large is 90.08%, the parameter size is 4.20MB, and the memory usage is 16.20MB; ShuffleNetV2 achieved an accuracy of 90.48%, with a parameter size of 0.30MB and a memory usage of 1.46MB; The accuracy of the improved model Im-ShuffleNetV2 is 93.88%, with a parameter size of 0.40MB and a memory usage of 1.78MB. It is worth mentioning that, especially the improved ShuffleNetV2 network we studied, although it showed a slight increase in parameter size and memory usage compared to ShuffleNetV2, with an increase of 0.1MB and 0.32MB respectively, it showed a good improvement in accuracy, reaching 93.88%. In addition, as shown in Fig. 8, when the network is trained for approximately76 epochs, the validation set accuracy of the improved ShuffleNetV2 network is the highest compared to other networks, and the fluctuation is relatively small. These quantitative results demonstrate that the proposed model Im-ShuffleNetV2 has good performance advantages in accuracy, parameter size, and memory usage compared to other networks.

Table 1 Performance comparison of different models.

Model	Accuracy /%	Precision /%	Recall /%	F1score /%	parameters /MB	memory /MB
AlexNet	87.90	87.77	87.39	87.50	14.60	55.60
EfficientNetV2 S	92.10	92.16	92.09	92.09	20.20	77.80
FasterNetT2	90.48	90.84	90.33	90.31	13.70	52.40
MobileNetV2	91.69	91.85	91.56	91.63	2.20	8.73
MobileNetV3 Large	90.08	91.23	90.70	90.72	4.20	16.20
MobileVIT XX Small	85.88	85.75	85.76	85.66	1.00	3.76
ShuffleNetV2	90.48	90.47	90.60	90.50	0.30	1.46
Im-ShuffleNetV2	93.88	94.22	94.23	94.21	0.40	1.78

3.3 Ablation experiment

The ablation experiment is an important method for evaluating the importance of neural networks and has been widely used in the field of deep learning research. The core idea is to gradually remove components of the neural network to verify their contributions to the overall network performance. In this study, three different optimization strategies (ECA attention mechanism, Transformer module, and network backbone) were introduced to improve the accuracy and performance of the ShuffleNetV2 network. The results of ablation experiment are presented in Table 2, S represents the original ShuffleNetV2 network, T represents the Transformer module, P represents the network backbone improvement, and E represents the ECA attention mechanism. S + P + E + T represents adding the Transformer module, network backbone improvement, and ECA attention mechanism to the original ShuffleNetV2 network.

Through the analysis of the experimental results in Table 2, significant performance improvements were achieved when each optimization strategy was applied separately. Specifically, after applying the ECA attention mechanism optimization, the ShuffleNetV2 achieved an accuracy improvement of 0.6% and a F1 score improvement of 1.59%. This indicates that the introduction of ECA attention mechanism effectively improved the network's classification accuracy and overall performance. Similarly, after optimizing ShuffleNetV2 using the Transformer module, the accuracy improved by 1.2%, and the F1 score increased by 0.59%. This demonstrates the positive effect of the Transformer module in improving network performance. Finally, through the improvement of the network backbone, the accuracy of ShuffleNetV2 increased by 1.6%, and the F1 score increased by 1.37%. This means that the improvement of the network backbone successfully improved the classification accuracy and performance of the network. In summary, significant improvements in accuracy and F1 scores were achieved by applying each optimization strategy separately in ShuffleNetV2, providing strong evidence and direction for optimizing network performance. Additionally, there was only a small increase in parameter volume of 0.1MB when adding the Transformer module, and in terms of memory usage, there was no impact when adding ECA attention mechanism, while the other two strategies only increased by 0.11MB and 0.32MB, respectively, indicating that the three optimization strategies will slightly increase the model size.

To demonstrate that combining the three optimization strategies results in stronger performance than using them separately, a comprehensive experiment was also conducted, applying all three optimization strategies simultaneously to the network. Specifically, combining the ECA attention mechanism with the network backbone improvement resulted in an accuracy improvement of 2.2% and a F1 score improvement of 2.8%. Combining the ECA attention mechanism with the Transformer module resulted in an accuracy improvement of 1.8% and a F1 score improvement of 2.42%. Combining the network backbone improvement with the Transformer module resulted in an accuracy improvement of 3% and a F1 score improvement of 2.93%.

Finally, the simultaneous application of ECA attention mechanism, network backbone improvement, and Transformer module resulted in an accuracy improvement of 3.4% and a F1 score improvement of 3.71%. Regarding parameter volume and memory usage, the experimental results showed the following changes: Experiment Group (S + E + P) had no changes in parameter volume and memory usage (increased by 0MB), Experiment Group (S + E + T) had an increase in parameter volume of 0.1MB and memory usage increase of 0.32MB, Experiment Group (S + E + P) had an increase in parameter volume of 0.1MB and memory usage increase of 0.11MB, and Experiment Group (S + T + P + E) had an increase in parameter volume of 0.1MB and memory usage increase of 0.11MB. Overall, by comprehensively using these three optimization strategies, the network's accuracy and F1 score have been significantly improved, while the increase in parameter volume and memory usage relative to the network's base size is relatively small. These experimental results demonstrate that these three optimization strategies have a good effect on improving the network's ability to obtain sugarcane disease features when used in combination.

Table 2
Comparison of ablation experiments.

Model	Accuracy/%	F1-score/%	Parameters/M	memory/M
S	90.48	90.50	0.30	1.46
S + P	92.08	92.09	0.30	1.57
S + E	91.08	91.09	0.30	1.46
S + T	91.68	91.87	0.40	1.78
S + T + P	93.28	93.43	0.40	1.57
S + T + E	92.28	92.92	0.40	1.78
S + P + E	92.68	93.30	0.30	1.57
S + P + E + T	93.88	94.21	0.40	1.57

3.4 Classification model verification

As shown in Fig. 9, confusion matrix is used to analyze the recognition accuracy of the Im-ShuffleNet model. It can be observed that the classification values on the diagonal are the largest, indicating that the model can correctly recognize sugarcane diseases. However, in the confusion matrix, 12 Healthy leaf images were misclassified as Mosaic disease. It is speculated that some Mosaic disease images have less obvious texture features, leading the network to misidentify Healthy leaves as Mosaic disease leaves. In addition, in the Rust disease classification, 7 images were mistakenly identified as Redrot disease leaves, and 8 images were mistakenly identified as Mosaic disease leaves. One possible reason is that the texture, color and other features of the disease are not sufficiently prominent in some Rust disease images, while the similarity of the image backgrounds may also affect the classification results. However, overall, the model has high accuracy and robustness, as shown by the confusion matrix visualization analysis.

To further demonstrate the superior performance of the proposed model, the classification performance comparison results of ShuffleNetV2 and Im-ShuffleNetV2 on the validation set are recorded in Table 3. As shown in Table 3, the proposed Im-ShuffleNetV2 has a significant advantage over ShuffleNetV2 in terms of Precision, Recall, and F1 score. For instance, in recognizing Yellow leaf disease and Mosaic disease, the F1 score of Im-ShuffleNetV2 is 96.51% and 91.84%, respectively, which is 4.5% and 8.4% higher than those of ShuffleNetV2 (96.51% and 83.44%, respectively). This is mainly attributed to the addition of ECA attention mechanism and Transformer module in the proposed model, which allows the network to accurately focus its attention on the features of sugarcane diseases while extracting global feature textures, better differentiating noise, Yellow leaf disease, and Mosaic disease. In summary, all these results indicate a significant advantage of our model in the classification and recognition of sugarcane diseases.

Table 3 Comparison of classification performance between ShuffleNetV2 and Im-ShuffleNetV2 on validation sets.

Model	Category	Precision /%	Recall /%	F1score /%
	Redrot	95.48	94.20	94.48
	Rust	86.64	92.24	89.35
	Yellow leaf	94.18	89.93	92.01
ShuffleNetV2	Healthy	91.41	94.35	92.86
	Mosaic	84.65	82.26	83.44
	Redrot	96.86	96.48	96.64
Im-ShuffleNetV2	Rust	91.45	92.24	91.84
	Yellow leaf	96.68	96.18	96.51
	Healthy	92.94	95.56	94.23
	Mosaic	92.98	90.73	91.84

4 Conclusion

In summary, we propose an improved ShuffleNetV2 (Im-ShuffleNet) by integrating ECA attention mechanism, network backbone enhancement and Transformer module into the original architecture, which performs better in extracting features of sugarcane diseases. Compared with other networks, the proposed ShuffleNetV2 network achieves higher accuracy, recall and F1 score, which are 93.88%, 94.22% and 94.21%, respectively. Meanwhile, the number of parameters and memory usage of the proposed network are only 0.4MB and 1.57MB, making it a more balanced choice for disease recognition on mobile devices. Therefore, our model will provide more choices for the recognition of sugarcane diseases and the deployment of deep models on small devices. In future research, we will collect more sugarcane disease images for model training in order to improve the model's generalization ability and make it applicable to the recognition of more types of sugarcane diseases. In addition, we will further optimize the network to improve recognition accuracy and reduce the model size.

Declarations

Ethical Approval

not applicable

Competing interests

The authors declare no conflict of interest.

Authors' contributions

Conceptualization, Y.X. and Z.L.; Methodology, Y.X. and Y.Z. (Yuting Zhai); Validation, Z.L. and Y.Z. (Yang Zhou); Writing-Original Draft Preparation, Z.L.; Writing-Review & Editing, Z.L. and Y.Z. (Yuting Zhai); Supervision, Y.X. (Yanlei Xu) and Y.J. (Yubin Jiao); Funding Acquisition, Y.X. (Yanlei Xu). All authors have read and agreed to the published version of the manuscript.

Funding

Science and Technology Development Plan Project of Changchun (Grant number 21ZGN28); the Jilin Provincial Science and Technology Development Plan Project (Grant number 20210101157JC); the Jilin Provincial Science and Technology Development Plan Project (Grant number 20230202035NC); Jilin province science and technology development plan project (Grant number YDZJ202301ZYTS408).

Availability of data and materials

not applicable

References

1. Guzman, M.; Victoria, J. Diagnosis of five sugarcane diseases using the same leaf sample. In Proceedings of the International Society of Sugar Cane Technologists. Proceedings of the XXIV Congress, Brisbane, Australia, 17-21 September 2001. Volume 2, 2001; pp. 462-464.
2. Militante, S.V.; Gerardo, B.D.; Medina, R.P. Sugarcane disease recognition using deep learning. In Proceedings of the 2019 IEEE Eurasia conference on IOT, communication and engineering (ECICE), 2019; pp. 575-578.
3. Li, X.; Li, X.; Zhang, S.; Zhang, G.; Zhang, M.; Shang, H. SLViT: Shuffle-convolution-based lightweight Vision transformer for effective diagnosis of sugarcane leaf diseases. Journal of King Saud University-Computer and Information Sciences 2023, 35, 101401.
4. Hossain, M.I.; Ahmad, K.; Siddiqui, Y.; Saad, N.; Rahman, Z.; Haruna, A.O.; Bejo, S.K. Current and prospective strategies on detecting and managing colletotrichumfalcatum causing red rot of sugarcane. Agronomy 2020, 10, 1253.
5. Bao, W.; Zhao, J.; Hu, G.; Zhang, D.; Huang, L.; Liang, D. Identification of wheat leaf diseases and their severity based on elliptical-maximum margin criterion metric learning. Sustainable Computing: Informatics and Systems 2021, 30, 100526.
6. Khan, A.I.; Quadri, S.; Banday, S.; Shah, J.L. Deep diagnosis: A real-time apple leaf disease detection system based on deep learning. computers and Electronics in Agriculture 2022, 198, 107093.
7. Pan, Q.; Gao, M.; Wu, P.; Yan, J.; AbdelRahman, M.A. Image classification of wheat rust based on ensemble learning. Sensors 2022, 22, 6047.
8. Shao, M.; He, P.; Zhang, Y.; Zhou, S.; Zhang, N.; Zhang, J. Identification Method of Cotton Leaf Diseases Based on Bilinear Coordinate Attention Enhancement Module. Agronomy 2022, 13, 88.

9. Liu, B.; Tan, C.; Li, S.; He, J.; Wang, H. A data augmentation method based on generative adversarial networks for grape leaf disease identification. IEEE Access 2020, 8, 102188-102198.
10. Wang, Y.; Chen, Y.; Wang, D. Context Data Fusion Model Enlightened Multi-Scale Capsule Network for Fruit Diseases Identification. International Journal of Pattern Recognition and Artificial Intelligence 2022, 36, 2257006.
11. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2016; pp. 770-778.
12. Xie, S.; Girshick, R.; Dollár, P.; Tu, Z.; He, K. Aggregated residual transformations for deep neural networks. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2017; pp. 1492-1500.
13. Wang, Y.; Tao, J.; Gao, H. Corn disease recognition based on attention mechanism network. Axioms 2022, 11, 480.
14. Liu, B.; Ding, Z.; Tian, L.; He, D.; Li, S.; Wang, H. Grape leaf disease identification using improved deep convolutional neural networks. Frontiers in Plant Science 2020, 11, 1082.
15. Yu, H.; Cheng, X.; Chen, C.; Heidari, A.A.; Liu, J.; Cai, Z.; Chen, H. Apple leaf disease recognition method with improved residual network. Multimedia Tools and Applications 2022, 81, 7759-7782.
16. Ma, N.; Zhang, X.; Zheng, H.-T.; Sun, J. Shufflenet v2: Practical guidelines for efficient cnn architecture design. In Proceedings of the Proceedings of the European conference on computer vision (ECCV), 2018; pp. 116-131.
17. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient channel attention for deep convolutional neural networks. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020; pp. 11534-11542.
18. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2015; pp. 1-9.
19. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.-C. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2018; pp. 4510-4520.
20. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861 2017.
21. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. Advances in neural information processing systems 2017, 30.
22. Mehta, S.; Rastegari, M. Mobilevit: light-weight, general-purpose, and mobile-friendly vision transformer. arXiv preprint arXiv:2110.02178 2021.
23. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 2014.

24. Tan, M.; Le, Q. Efficientnetv2: Smaller models and faster training. In Proceedings of the International conference on machine learning, 2021; pp. 10096-10106.
25. Chen, J.; Kao, S.-h.; He, H.; Zhuo, W.; Wen, S.; Lee, C.-H.; Chan, S.-H.G. Run, Don't Walk: Chasing Higher FLOPS for Faster Neural Networks. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023; pp. 12021-12031.

Figures

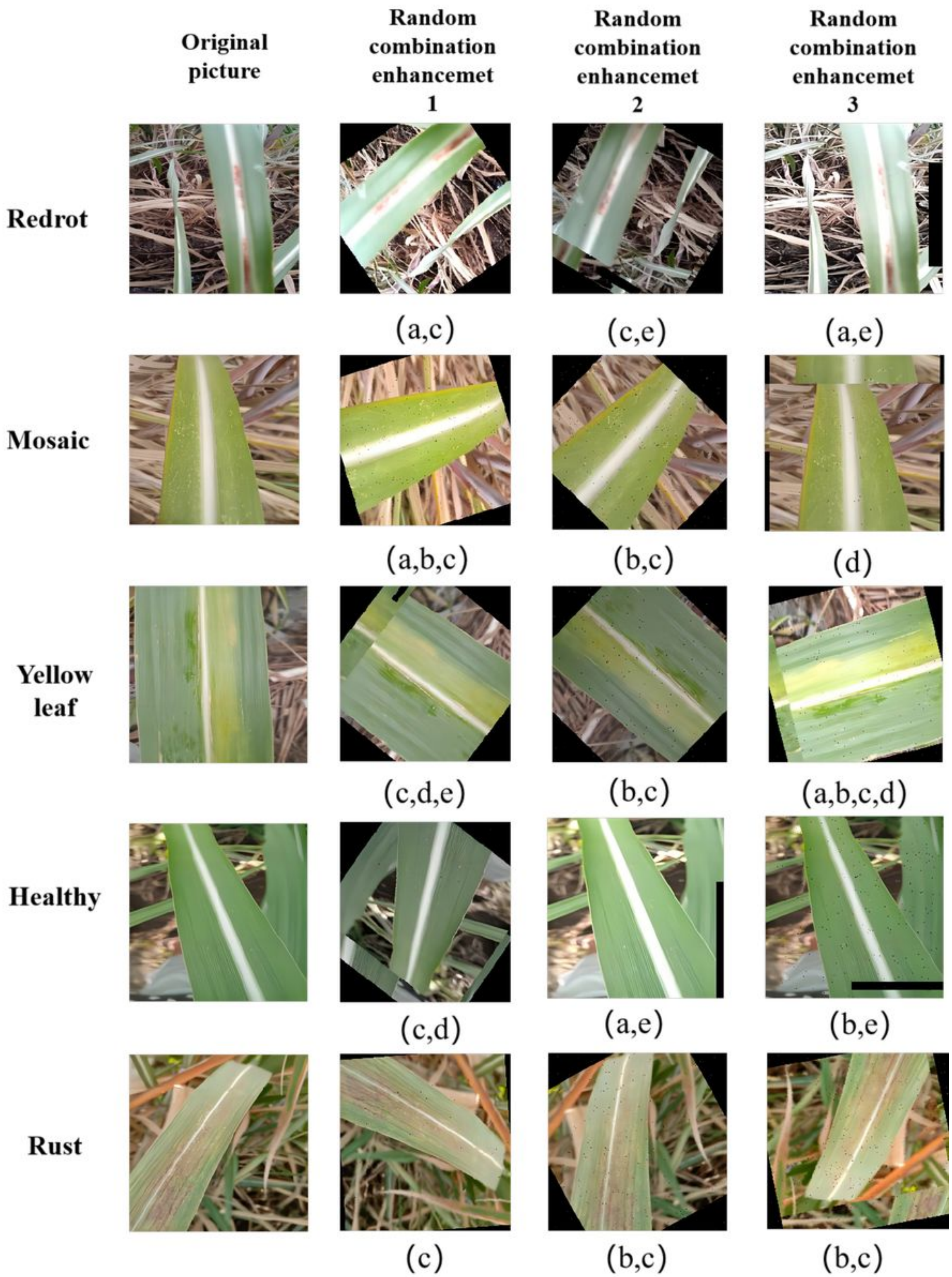

Figure 1

A partially randomized amalgamation of sugarcane disease datasets and augmented data: (a) Brightness enhancement and Random rotation; (b) Random rotation and Random black box occlusion; (c) Brightness enhancement and Random black box occlusion; (d) Brightness enhancement, Salt-and-pepper noise, and Random rotation; (e) Salt-and-pepper noise and Random rotation; (f) Random cropping; (g) Random cropping, Random black box occlusion, and Random rotation; (h) Salt-and-pepper noise and

Random rotation; (i) Salt-and-pepper noise, Brightness enhancement, Random cropping, and Random rotation; (j) Salt-and-pepper noise and Random rotation, (k) Brightness enhancement and Random black box occlusion; (l) Salt-and-pepper noise and Random rotation; (m) Random rotation; (n) Salt-and-pepper noise and Random rotation; (o) Salt-and-pepper noise and Random rotation.

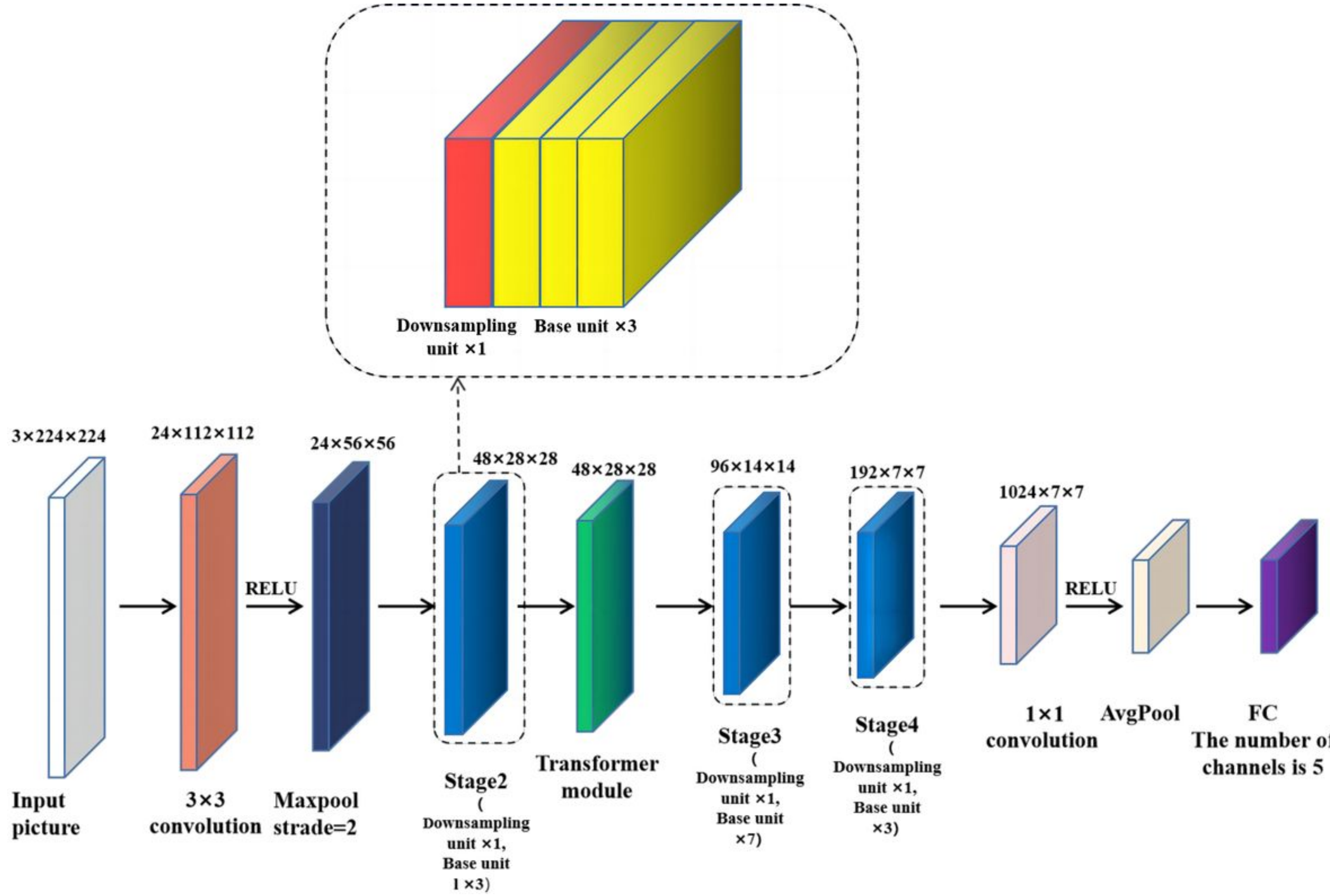

Figure 2

The network architecture of the proposed model.

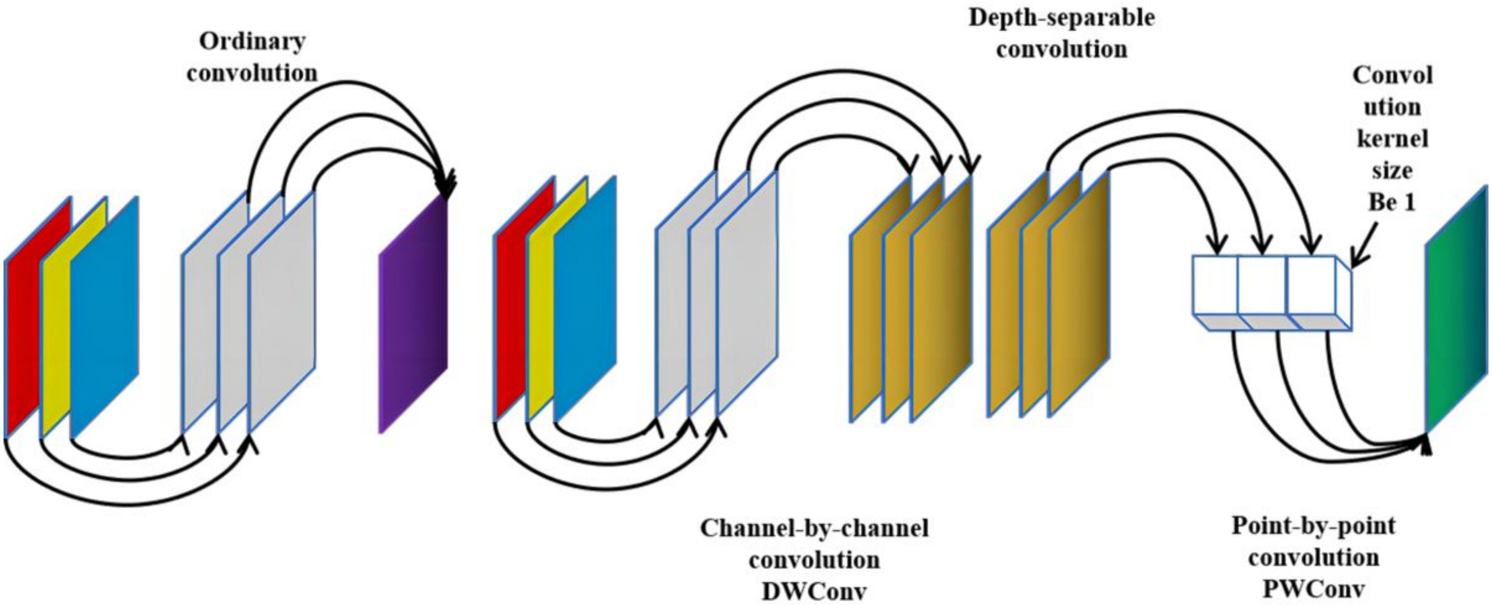

Figure 3

Ordinary convolutional sample image and depth-separable convolutional sample image.

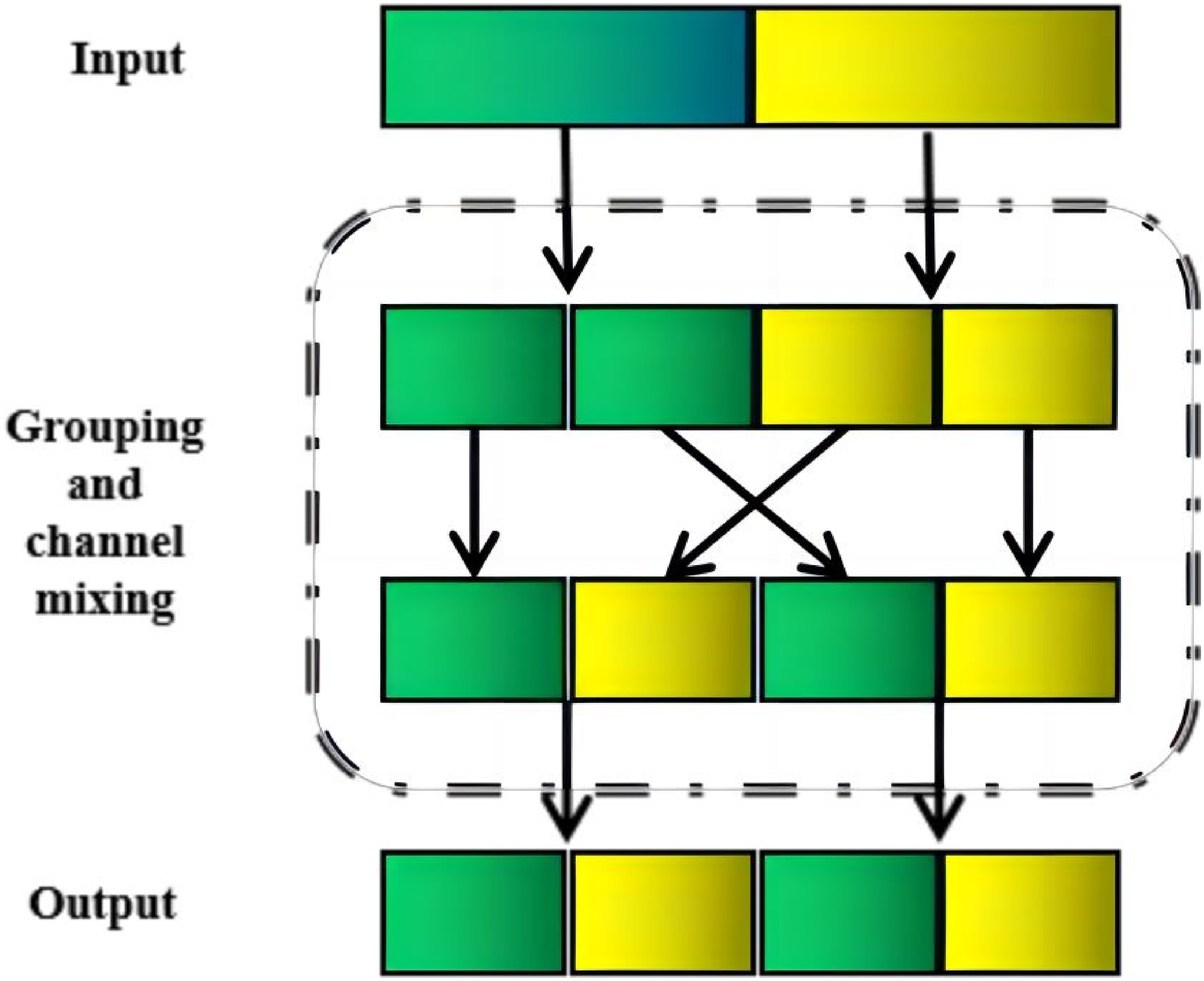

Figure 4

Example of channel shuffle image.

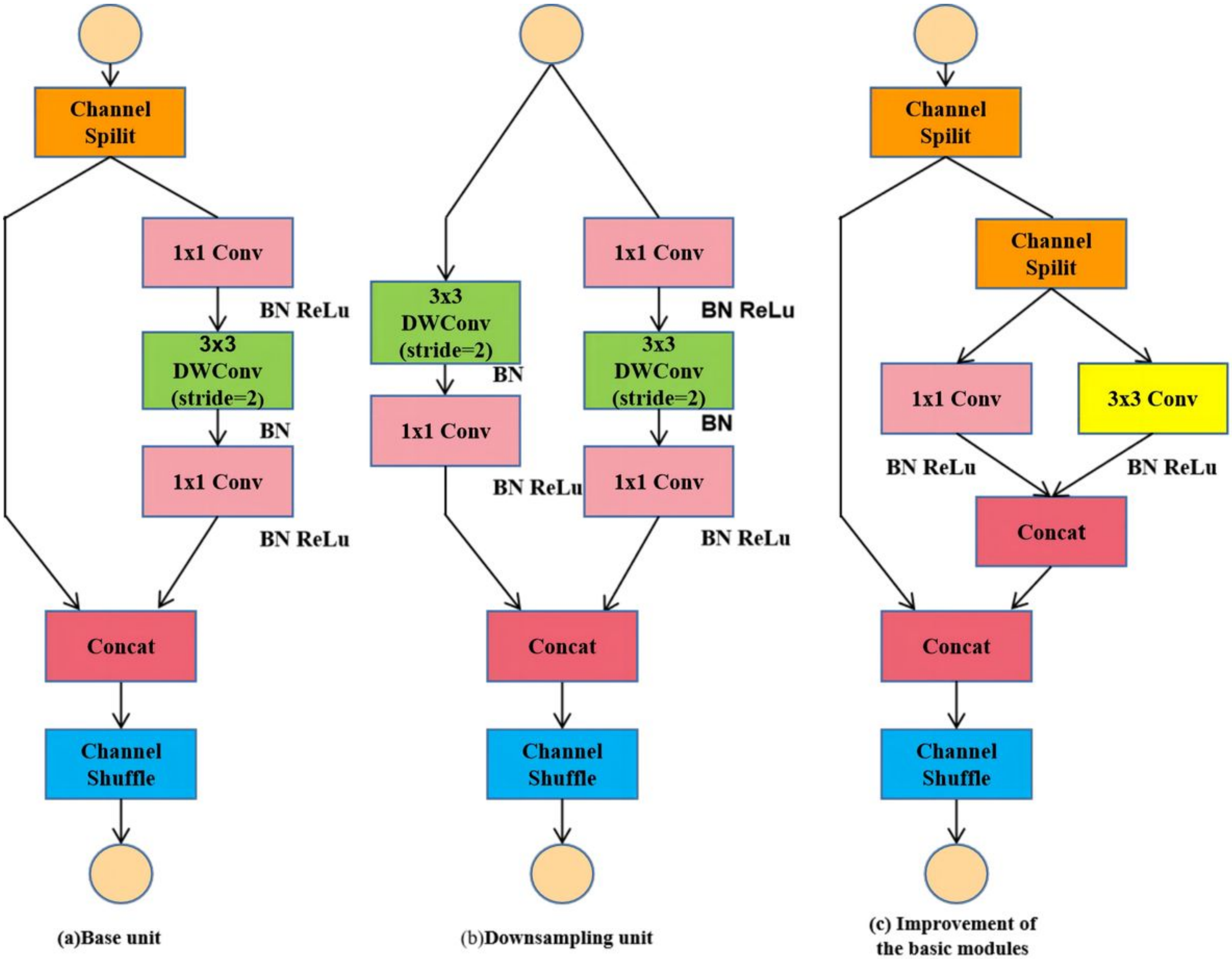

Figure 5

ShuffleNetV2 network backbone and its improved module schematic diagram.

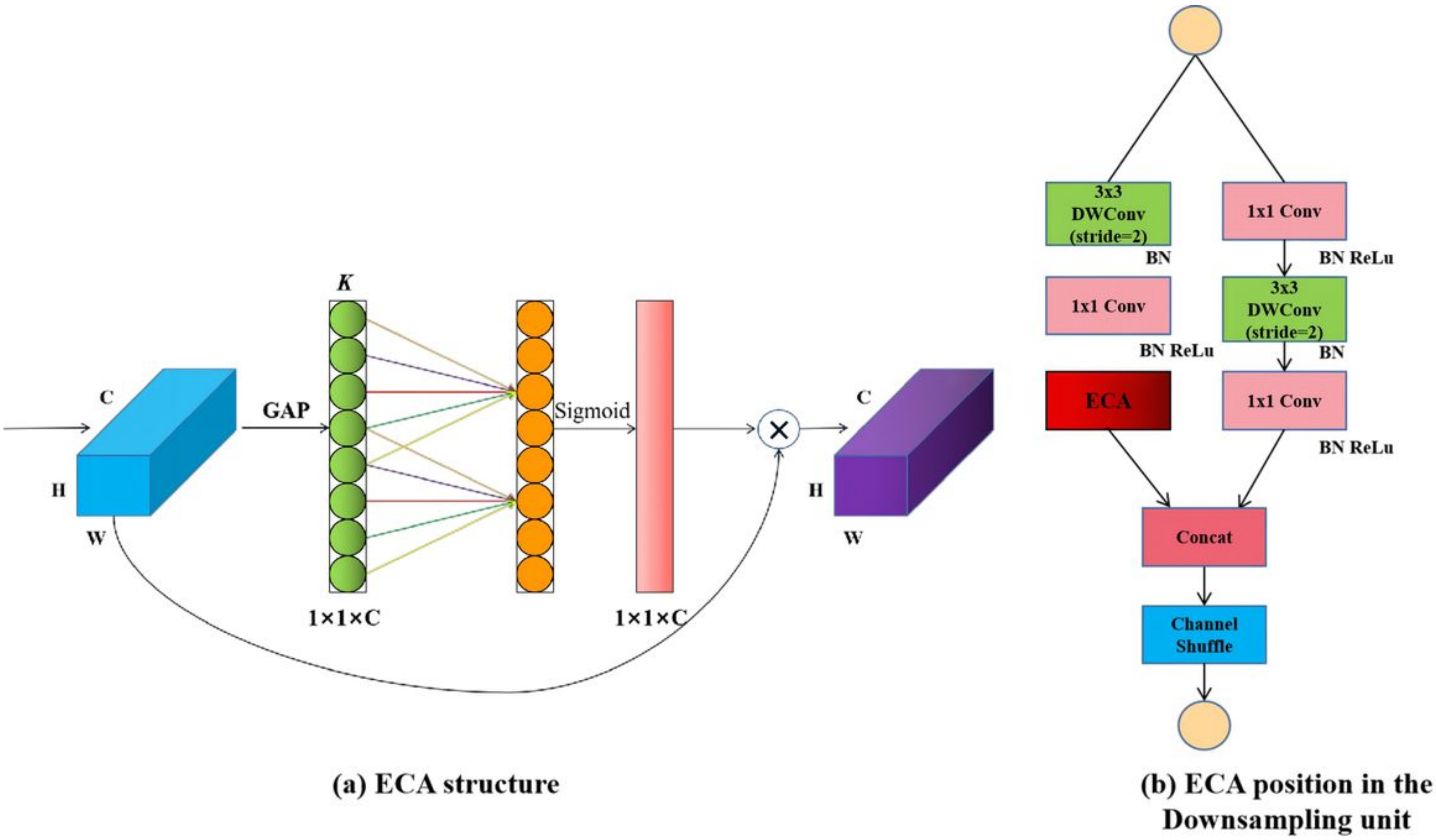

Figure 6

Structural schematic diagram and location of ECA attention mechanism.

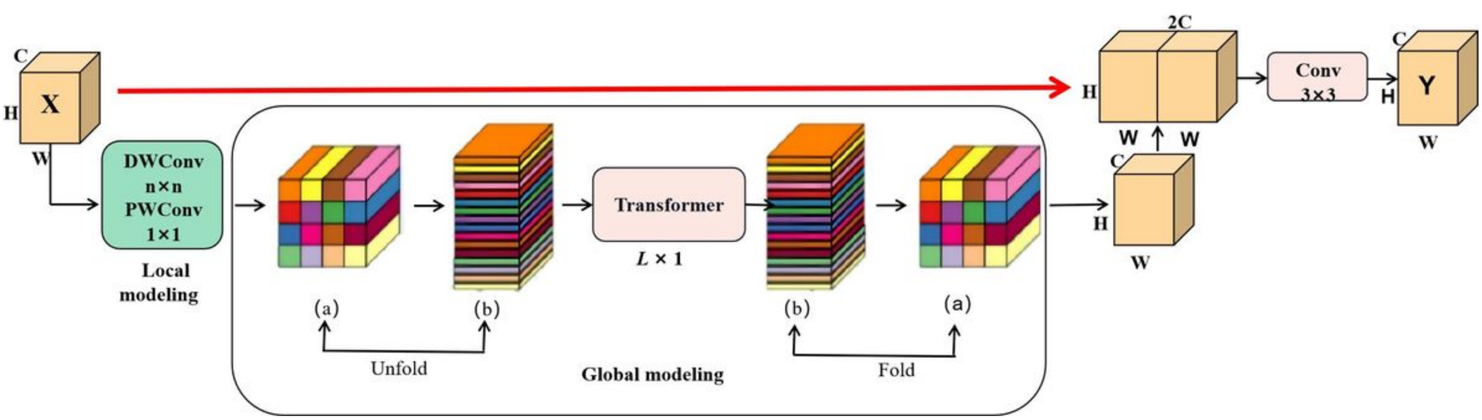

Figure 7

Transformer module structure diagram.

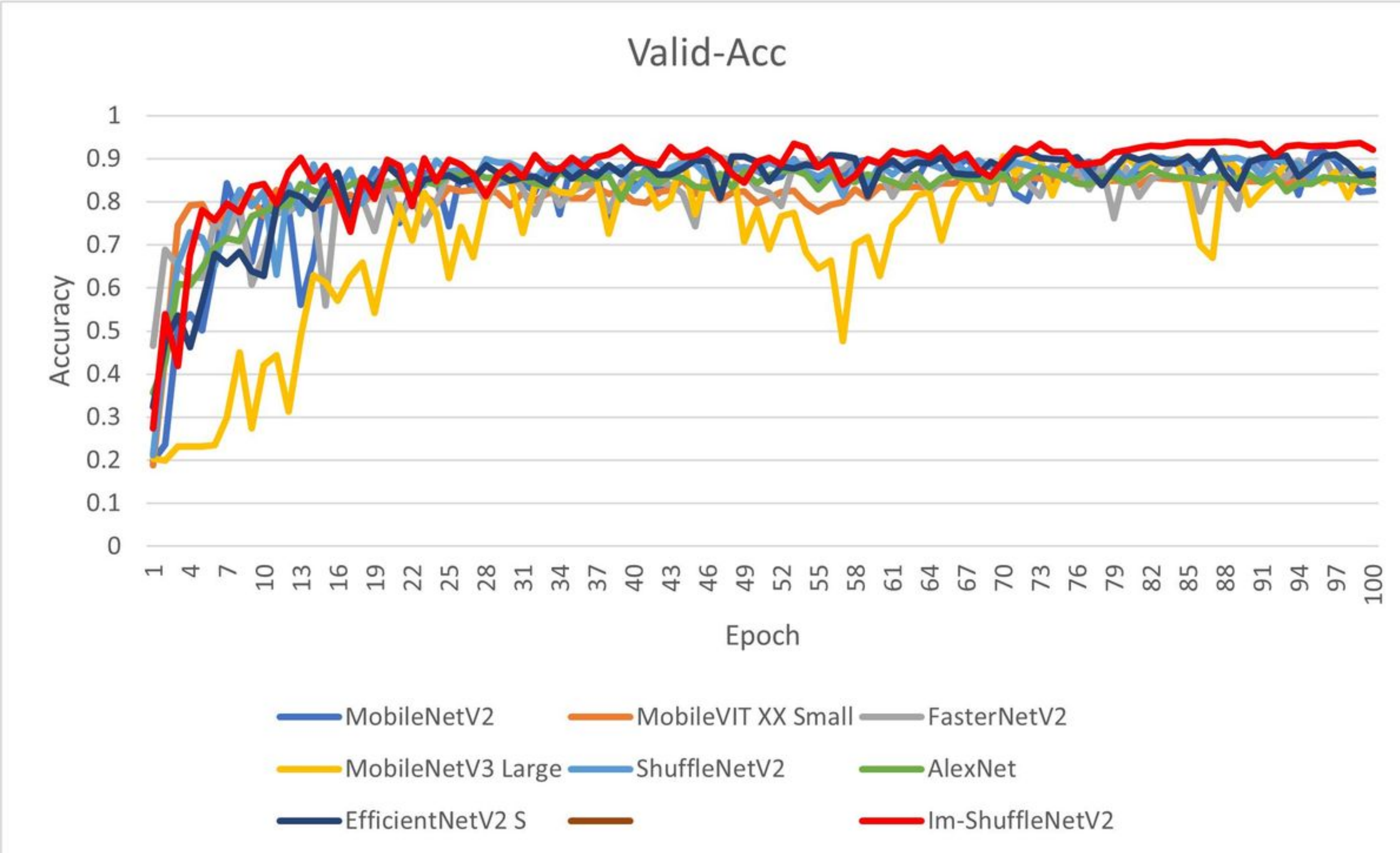

Figure 8

Comparison of accuracy rates of different models.

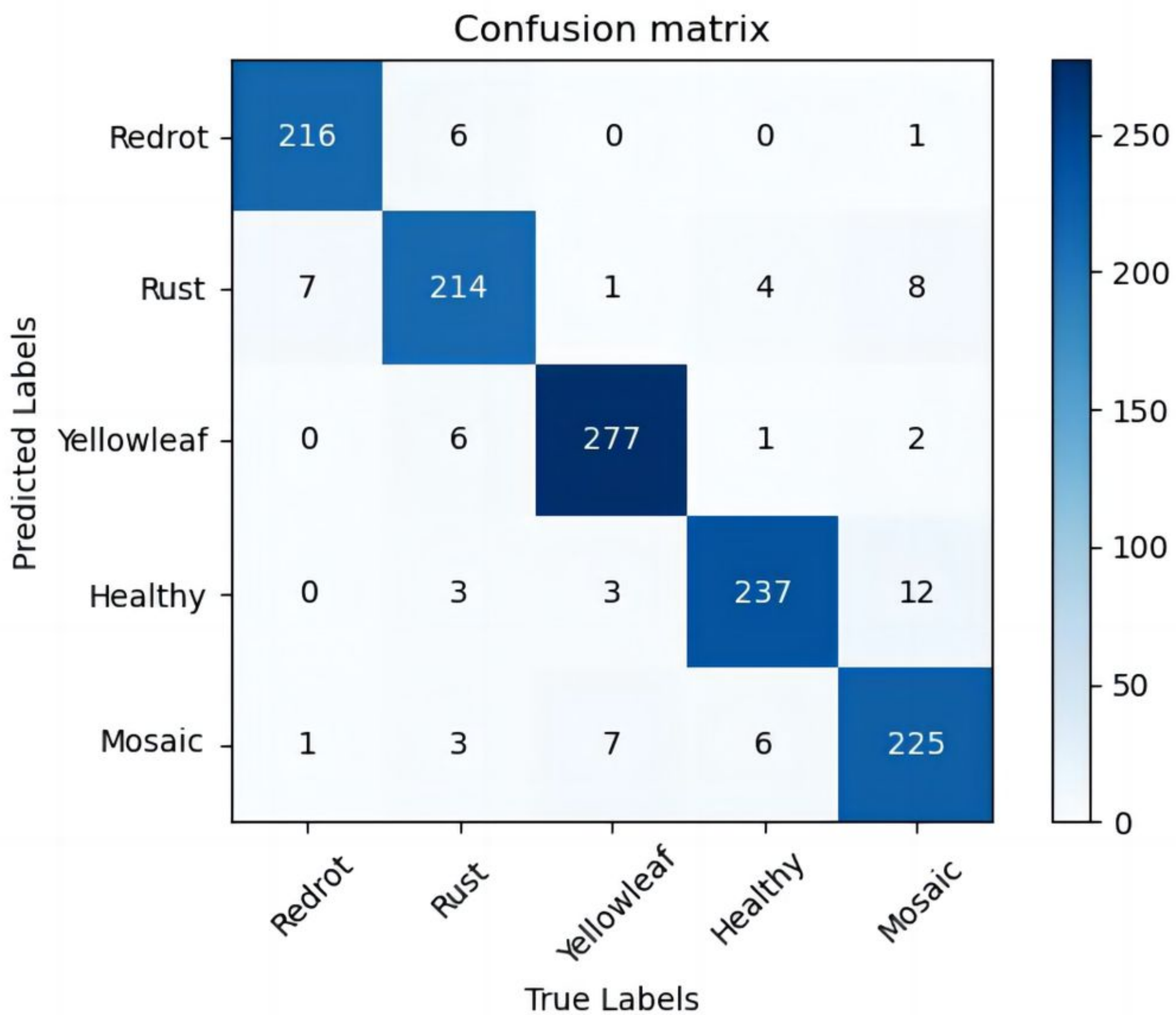

Figure 9

Improved network test set confusion matrix.