

Dynamic ensemble selection of convolutional neural networks and its application in flower classification

Zhibin Wang^{1,2}, Kaiyi Wang^{1,2}, Xiaofeng Wang^{1,2}, Shouhui Pan^{1,2}, Xiaojun Qiao^{1,2*}

(1. Information Technology Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China;

2. National Engineering Research Center for Information Technology in Agriculture, Beijing 100097, China)

Abstract: In recent years, convolutional neural networks (CNNs) have achieved great success in image classification. However, CNN models usually have complex network structures that tend to cause some related problems, such as redundancy of network parameters, low training efficiency, overfitting, and weak generalization ability. To solve these problems and improve the accuracy of flower classification, the advantages of CNNs were combined with those of ensemble learning and a method was developed for the dynamic ensemble selection of CNNs. First, MobileNet models pre-trained on a public dataset were transferred to flower datasets to train thirteen different MobileNet classifiers, and a resampling strategy was used to enhance the diversity of individual models. Second, the thirteen classifiers were sorted by a classifier sorting algorithm, before ensemble selection, to avoid an exhaustive search. Finally, with the credibility of recognition results, a classifier subset was dynamically selected and integrated to identify the flower species from their images. To verify the effectiveness, the proposed method was used to classify the images of five flower species. The accuracy of the proposed method was 95.50%, an improvement of 1.62%, 3.94%, 22.04%, 13.77%, and 0.44%, over those of MobileNet, Inception-v1, ResNet-50, Inception-ResNet-v2, and the linear ensemble method, respectively. In addition, the performance of the proposed method was compared with five other methods for flower classification. The experimental results demonstrated the accuracy and robustness of the proposed method.

Keywords: flowers, classification, convolutional neural network, dynamic ensemble selection

DOI: 10.25165/j.ijabe.20221501.6313

Citation: Wang Z B, Wang K Y, Wang X F, Pan S H, Qiao X J. Dynamic ensemble selection of convolutional neural networks and its application in flower classification. *Int J Agric & Biol Eng*, 2022; 15(1): 216–223.

1 Introduction

Classification of flower species plays an important role in the informatization of flower management^[1,2]. The results of identification can greatly improve the efficiency of retrieving information related to flower species. However, accurate classification of flower species remains difficult because of the similarities between different flower species, and differences between flowers in the same species.

Manual classification, whereby flower species are assessed by specialists who perform naked-eye observations according to their experience and knowledge, is laborious, time-consuming, and has low recognition accuracy^[3]. Because of the rapid development of image processing and artificial intelligence technology, researchers have utilized various image-based classification methods to overcome these obstacles. Hus et al.^[4] developed an interactive system for recognizing flower images taken by digital cameras.

Guru et al.^[5] investigated the suitability of texture features and designed a system for flower classification, and then used the features for training and classification with a probabilistic neural network. These results demonstrated that the combination of multiple features vastly improves the classification accuracy, from 35% for the best single feature to 79% for the combination of all features. Cheng et al.^[6] proposed an attribute-based method for flower recognition, which extracted a series of visual attributes from a given set of flower images and generalized them to new images with possibly unknown flowers. Soleimanipour et al.^[7] developed a vision-based hybrid approach, using a hybrid of the Viola-Jones detector and multi-template matching, for highly accurate and effective identification of Anthurium flower cultivars. Their results indicated that the technique had acceptable performance in detecting the spadix region and achieved more than 99% classification accuracy. In general, the accuracy of traditional image-based classification techniques depends largely on manually designed features to express the characteristics of flower images. Because this is a difficult process, the accuracy and generalization of classification are weak in new application scenarios.

In recent years, deep learning, a kind of efficient technology for feature representation learning, has been rapidly developed and successfully applied in many fields^[8-10]. Owing to their outstanding ability to learn complex and robust feature representations, the accuracy of flower image classification has been greatly improved by employing deep neural networks, particularly convolutional neural networks (CNNs). Cibuk et al.^[11] applied a CNN-based hybrid method to the classification of flower species and showed that their proposed method achieved

Received date: 2020-11-27 **Accepted date:** 2021-07-19

Biographies: **Zhibin Wang**, PhD, Senior Engineer, research interest: machine learning and image processing, Email: wangzb@nercita.org.cn; **Kaiyi Wang**, PhD, Professor, research interest: breeding informatization, Email: wangky@nercita.org.cn; **Xiaofeng Wang**, PhD candidate, research interest: big data and internet of things, Email: wangxf@nercita.org.cn; **Shouhui Pan**, PhD, Associate Professor, research interest: management information system, Email: pansh@nercita.org.cn.

***Corresponding author:** **Xiaojun Qiao**, PhD, Professor, research interest: intelligent decision-making system and intelligent agricultural equipment. Information Technology Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China. Tel: +86-10-51503348, Email: qiaoxj@nercita.org.cn.

96.39% and 95.70% accuracy for the Flowers17 and Flowers102 datasets, respectively. Guo et al.^[12] classified flowers using the Tabu_Genetic algorithm together with a CNN, which consisted of 14 layers, and it achieved 78.81% accuracy. Toğaçar et al.^[13] proposed a hybrid method that used four CNN models, AlexNet, GoogLeNet, ResNet-50, and VGG-16, for feature extraction, and the classification performance was measured using an SVM classifier. Their results showed that the intersection of the features obtained using the feature selection methods improved the classification performance, and the overall accuracy obtained was 98.91%. To improve the accuracy of flower image recognition, Hiary et al.^[14] adopted the FCN and VGG-16 to construct a novel two-step deep-learning classifier to distinguish flowers of a wide range of species and achieved at least 97% classification accuracy. In the study of Mitrovic et al.^[15], numerous network models were implemented for flower classification. They found that the AlexNet model, with a uniform sigmoid function for allocating initial weights, produced the best classification results. Although deep-learning-based methods have achieved great success in flower classification, there are still some limitations: it is difficult to select appropriate network structures, parameters, and algorithms, and there is a requirement for long training times for optimal recognition performance in practice.

Unlike classification approaches that use a single fixed model, ensemble learning provides methods that overcome the limitations mentioned above, to some extent. Ensemble learning uses a set of learners and applies rules to integrate the learning results, to obtain better performance than a single learner^[16]. The effectiveness of ensemble learning has been widely demonstrated in a variety of applications^[17-19]. Bae et al.^[20] proposed a modified m-CNN model that integrated images and text in multi-view learning for flower classification. Experimental results demonstrated that the proposed algorithm achieved superior performance, compared with other data fusion methods. Huang et al.^[21] presented a flower classification framework based on ensemble CNNs and demonstrated its effectiveness on the Flowers102 dataset. In the author team's previous work^[22], a CNN ensemble method was developed for flower classification. Experimental results showed that the method proposed by authors had a better generalization ability and higher recognition rate than the single classifiers. Notably, after obtaining multiple learners, most ensemble algorithms employ all of them to constitute an ensemble. However, both theoretical and empirical studies have shown that, instead of using a whole ensemble, a proper subset of an ensemble can often achieve better generalization performance^[23-25]. Another clear advantage of a selective ensemble is that storage cost is reduced and efficiency is improved because fewer individual learners need to be stored and used for the classification. However, the selection process is not easy, and choosing appropriate learners remains something of an art.

To overcome these problems and achieve better recognition performance, the advantages of CNNs and ensemble learning were combined and a method was developed using dynamic ensemble selection of CNNs for flower classification. The main contributions of this study are 1) MobileNet models pre-trained on the ILSVRC-2012-CLS image classification dataset are adopted as single classifiers. Adopting these pre-trained models as the single classifiers can avoid the problems of selecting optimal single-network model parameters and designing appropriate network structures; 2) CNNs have complicated mechanisms that tend to cause overfitting, with poor generalization ability.

However, in the proposed method, by randomly generating multiple single classifiers with a simple structure and integrating their outputs, the accuracy and generalization of the recognition algorithm can be greatly improved; 3) To compromise between ensemble accuracy and efficiency, a dynamic ensemble selection method was developed. With the credibility of recognition results, an optimal classifier subset was dynamically selected and integrated to identify each test sample.

2 Flower image datasets

Two datasets of flower images were used to evaluate the performance of the proposed method. One dataset^[26] contained 3670 flower images from five species and was divided into a training set with 3320 images and a validation set with 350 images. The other dataset contained 2670 images of the same five flower species as the first dataset. This dataset was divided into a test set with 1600 images and a validation set with 1070 images. The images in the second dataset were obtained in one of two ways. Some images were captured in the field using digital cameras or mobile phones; to ensure the robustness of the proposed method, images of the same species were taken under various illumination conditions from different angles and featuring different flowers. The other images were collected from the Internet, for example, the Flowers17 and Flowers102 datasets from the Visual Geometry Group at the University of Oxford and the Subject Dataset of China Plant^[27]. Therefore, there was no uniform image size or resolution in the datasets. Before classification, all the flower images were scaled to the same size (224×224 pixels) and then converted to TensorFlow native TFRecord format.

In this study, flower images were collected from different sources, different regions, and different seasons to represent a wide range of scenarios so that they provided a challenge for testing the performance of the proposed method. To evaluate the diversity of each base classifier in the ensemble, a new validation set with 1420 images was used which was constructed by combining the validation sets in the two flower image datasets. Representative images and the image datasets used in this study are shown in Figure 1 and Table 1, respectively.

Table 1 Flower image datasets used in this study

Flower species	Training set	Validation set	Test set
Daisy	561	259	272
Dandelion	821	299	349
Rose	573	286	348
Sunflower	625	276	293
Tulip	740	300	338
Total	3320	1420	1600

3 Methods

The framework of the proposed method is depicted in Figure 2. The method has four main modules: pre-training, fine-tuning, dynamic ensemble selection, and output. First, MobileNet models pre-trained on the ILSVRC-2012-CLS image classification dataset are used as single classifiers for feature extraction. Second, the models are transferred to flower datasets to train thirteen different MobileNet classifiers, and a resampling strategy is used to enhance the diversity of individual models. Third, the thirteen classifiers are sorted by a classifier sorting algorithm, before ensemble selection, to avoid an exhaustive search. Finally, with the credibility of recognition results, a classifier subset is selected dynamically and integrated to identify the flower species.



Figure 1 Representative flower image examples used in this study

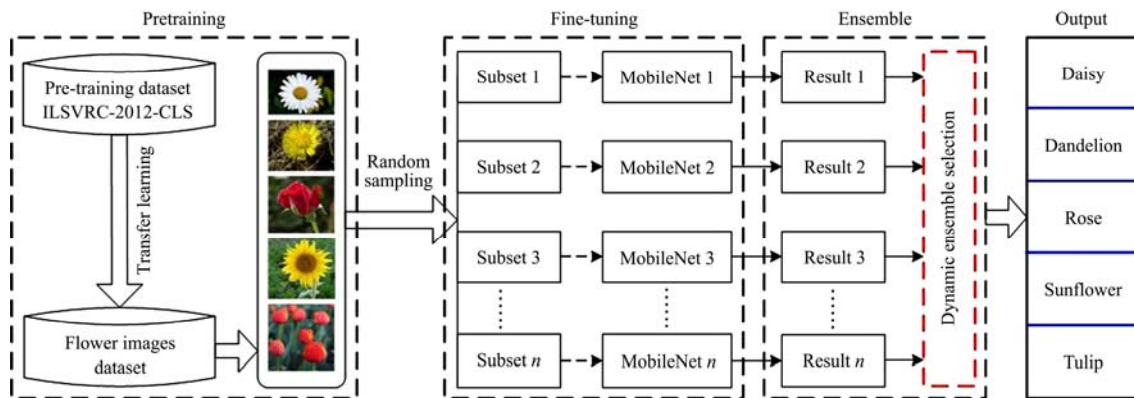


Figure 2 Structure of the proposed method for flower classification

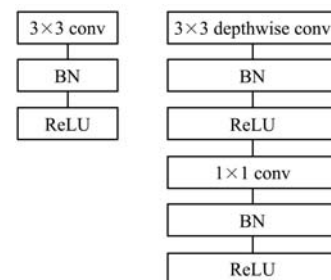
3.1 MobileNet

MobileNet, which was developed for use in mobile and embedded systems by Howard et al.^[28], can achieve a good balance between performance and computational cost. The MobileNet architecture is based on depthwise separable convolutions, followed by a pointwise convolution with a 1×1 convolution layer. In the standard convolution layer, each kernel is applied to all channels of the input image, whereas depthwise convolution is applied to each channel separately. This approach significantly reduces the number of parameters, compared with standard convolutions with the same depth. A comparison between standard convolution and depthwise separable convolution in MobileNet is shown in Figure 3.

3.2 Dynamic ensemble selection

Dynamic ensemble selection is an ensemble learning paradigm in which one or more base classifiers are selected for each query instance to be classified^[29-31]. In dynamic selection, the aim is to

select the most competent classifiers for any given query sample. However, finding the optimal subset of classifiers entails searching in the space of all classifier combinations, and the computational complexity increases exponentially with the number of classifiers.



Note: Conv: Convolution; BN: Batch Normalization; ReLU: Rectified Linear Unit.

Figure 3 Comparison between standard convolution and depthwise separable convolution in MobileNet models

A method was proposed for the dynamic ensemble selection of CNNs, which combines the procedures of classifier selection and integration. A flowchart of the proposed method is shown in Figure 4. The proposed method has two stages: classifier sorting and dynamic classifier ensemble selection. First, all classifiers are sorted by the proposed classifier sorting method. Classifiers are

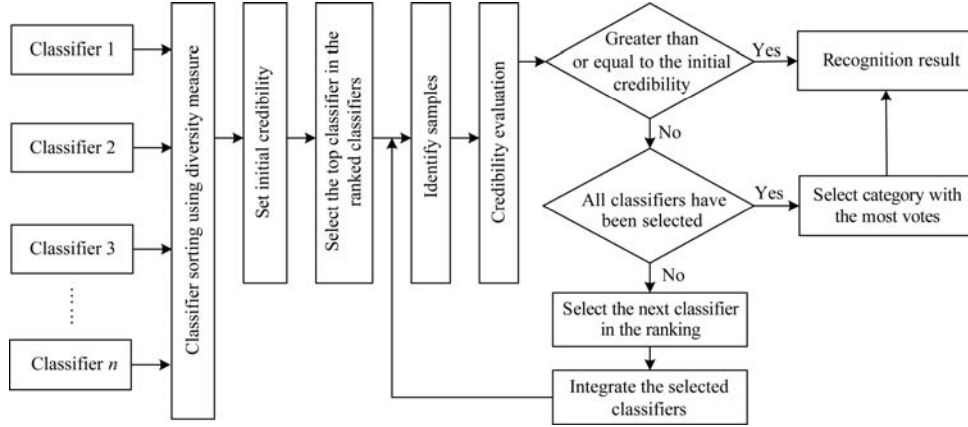


Figure 4 Flowchart of the proposed method for dynamic classifier ensemble selection

3.2.1 Diversity measurement

Diversity among the base classifiers is generally considered to be important when constructing a classifier ensemble. There are several diversity measures for classifier members^[32]. The disagreement measure was used. For binary classification, this measure is defined as the ratio between the number of observations on which one classifier is correct and the other is incorrect to the total number of observations. The relationship between two base classifiers C_i and C_j is shown in Table 2.

Table 2 Relationship between two classifiers C_i and C_j

	C_j correct (1)	C_j wrong (0)
C_i correct (1)	N^{11}	N^{10}
C_i wrong (0)	N^{01}	N^{00}

If N is the total number of samples, for a two-classifier measure, N^{11} indicates the number of times both classifiers are correct, N^{00} indicates the number of times that both classifiers are incorrect, and N^{10} and N^{01} indicate the number of times that only the first or only the second classifier is correct, respectively, and $N = N^{01} + N^{10} + N^{11} + N^{00}$. The disagreement measure (dis) between the two base classifiers C_i and C_j is

$$\text{dis}_{i,j} = \frac{N^{01} + N^{10}}{N^{11} + N^{00} + N^{01} + N^{10}} \quad (1)$$

The diversity of a whole set of L base classifiers (Dis) is the disagreement measure, which is defined as the average over all pairs of base classifiers (Equation (2)).

$$\text{Dis} = \frac{2}{L(L-1)} \sum_{i=1}^L \sum_{j=i+1}^L \text{dis}_{i,j} \quad (2)$$

Therefore, the diversity increases with the value of the disagreement measure.

3.2.2 Classifier sorting

A static classifier sorting method (Algorithm 1) was developed by combining the selection of classifiers with the disagreement measure. The flowchart of Algorithm 1 is shown in Figure 5.

In Algorithm 1, the classifier with the highest recognition accuracy on the validation set is ranked first, and then the next classifier is selected from the remaining candidate classifiers to rank second. The selection criterion is that the classifier set composed of the newly chosen classifier and the previously

then dynamically selected, one by one, from the sorted sequence to be integrated to identify the test sample. The number of selected classifiers is determined by the credibility of recognition results of the test sample. Details of the proposed method are presented in the following sections.

selected classifier(s) maximizes the value of Dis. The selection step is repeated until all the candidate classifiers have been sorted. Thus, Algorithm 1 transforms the problem of classifier selection to a problem of classifier sorting, and thereby avoids the process of classifier searching and improves the efficiency of classifier selection.

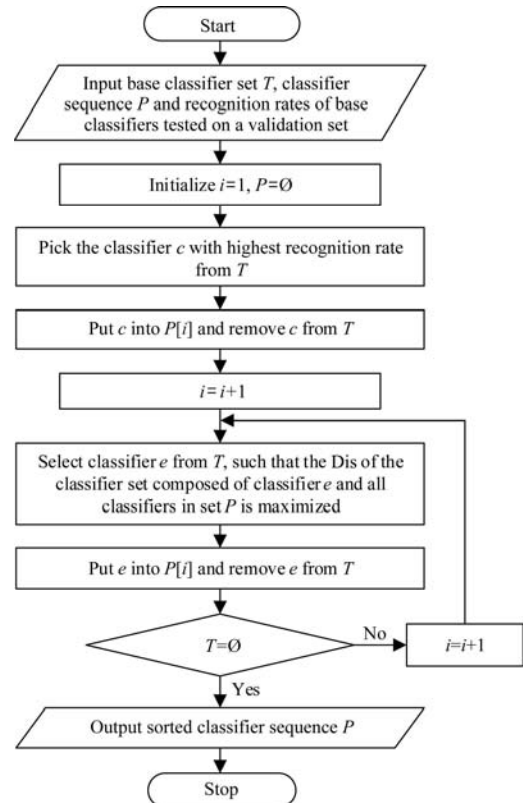


Figure 5 Flowchart of the classifier sorting algorithm

With the sorted classifier sequence, the top classifiers in the sorted classifier sequence are dynamically selected as the optimal classifier subset and integrated to identify each test sample. When only one classifier is needed, the classifier with the highest recognition rate on the validation set is selected for each query instance to be classified. However, when N classifiers are needed, the top N classifiers in the sorted classifier sequence are selected as

the optimal classifier subset. Algorithm 1 transforms the problem of classifier selection to a problem of classifier sorting, and thereby avoids the process of classifier searching and improves the efficiency of classifier selection.

3.2.3 Dynamic classifier ensemble selection

After all base classifiers are sorted, an optimal classifier subset is selected dynamically from the sorted classifier sequence P and integrated to identify the test sample. The method for dynamic classifier ensemble selection (Algorithm 2) is presented in Figure 6.

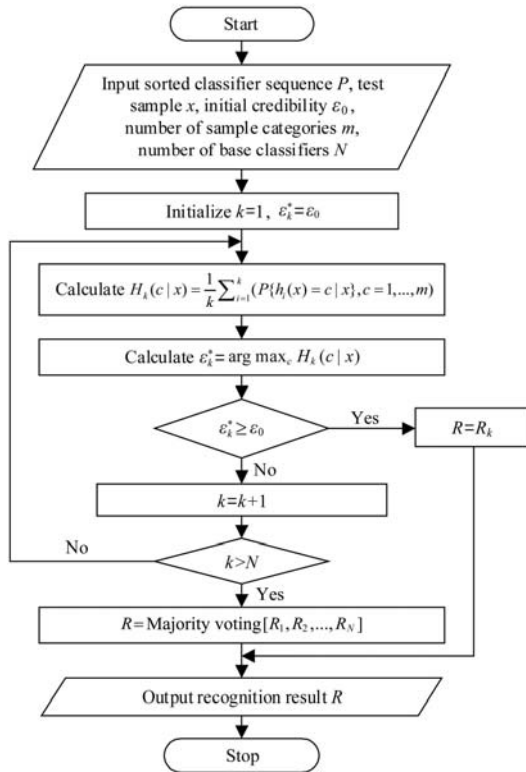


Figure 6 Flowchart of the dynamic classifier ensemble selection algorithm

In Algorithm 2, ε_0 is the initial credibility and its value is set, according to practical application needs, in the range $[0, 1]$. $P\{h_i(x)=c|x\}$ is the posterior probability of sample x being classified as label c by the i -th classifier. $H_k(c|x)$ is the average posterior probability of sample x being classified as label c by a linear ensemble of k classifiers, and the corresponding classification result is R_k . m is the number of sample categories. The value of $H_k(c|x)$ is $(\varepsilon_1, \varepsilon_2, \varepsilon_3, \dots, \varepsilon_m)^T$, where each dimension component represents the credibility of its corresponding category after ensemble recognition and ε_k^* is the maximum credibility. The posterior probability is computed by using Bayes' theorem^[33]. Consider the specific task of flower classification, which is the focus of this study. Each input image comprises an array of pixel intensity values, and the desired output is a posterior probability distribution over all the categories of flower species. For a classifier whose output is not a posterior probability distribution, the output should be transformed to a posterior probability^[34].

Algorithm 2 proceeds as follows. First, the initial credibility ε_0 is set according to the required recognition accuracy in practical application. Second, one or more base classifiers for an optimal classifier subset are selected dynamically from the sorted classifier sequence P and integrated to identify test sample x ; if $\varepsilon_k^* \geq \varepsilon_0$, the category of ε_k^* corresponds to the recognition result R_k of sample x . Third, if all classifiers have been selected and the recognition credibility requirement remains unsatisfied, the identification

results $[R_1, R_2, \dots, R_N]$ of each integration are voted on, and the category with the most votes is considered to be the category of sample x . Therefore, for each test sample, there is a specialized optimal classifier subset, which is dynamically selected and integrated to identify the test sample. That is, the recognition is targeted.

3.3 Evaluation criteria

The performance of the proposed classification method was evaluated according to accuracy. Accuracy (%) is a good metric for measuring the proportion of correctly classified instances over all the samples in a test set. Given the number of correctly classified instances (NC) and the number of all the samples in the test set (NA), accuracy is defined as

$$\text{Accuracy} = \frac{\text{NC}}{\text{NA}} \times 100\% \quad (3)$$

4 Results

All the experiments were performed using TensorFlow in the Python programming environment on a computer with an Intel® Core™ i7-7700HQ (2.8 GHz) processor, 16 GB of memory (RAM), NVIDIA GeForce 940MX, and Windows 10 operating system.

4.1 Single classifier generation

MobileNet models were selected as the single classifiers to classify flower images. MobileNet was pre-trained on the ILSVRC-2012-CLS image classification dataset and the pre-trained weights were fine-tuned using the following parameters: initial learning rate 0.01, number of training iterations 500. The Adam algorithm was used as the optimizer, and the output layer used the softmax classifier to convert the output result to a probability in the range $[0, 1]$.

To construct a good ensemble, the bootstrap resampling method was used to create different training sets, so that each classifier in the ensemble was trained with a different training set, to ensure diversity. Moreover, different batch sizes were used for single-network training because this approach achieved better recognition performance in the previous studies of authors^[22]. Thirteen MobileNet single classifiers were generated randomly and the recognition accuracy of every single classifier is shown in Table 3.

Table 3 Classification results were obtained by single classifiers with different batch sizes

Classifier	Batch size	Seconds/iteration	Accuracy/%	
			Validation set	Test set
MobileNet_1	32	10.68	90.07	89.50
MobileNet_2	50	17.52	88.87	88.06
MobileNet_3	80	29.04	94.72	93.69
MobileNet_4	100	37.20	93.10	92.25
MobileNet_5	130	48.00	92.18	90.06
MobileNet_6	150	55.80	94.01	92.19
MobileNet_7	180	67.80	92.61	92.00
MobileNet_8	200	74.48	91.27	89.69
MobileNet_9	220	84.48	91.76	91.00
MobileNet_10	250	97.08	92.96	91.87
MobileNet_11	280	109.44	92.61	91.75
MobileNet_12	300	118.32	93.59	92.50
MobileNet_13	345	191.16	95.00	93.88

The results show that MobileNet_13 had the highest recognition rate and MobileNet_2 had the lowest recognition rate on both the test and validation sets.

4.2 Dynamic ensemble selection

The diversity among the single classifiers in Table 3 was

measured on the validation set using the disagreement method. All the classifiers were sorted according to the Dis measure, defined in Equation (2), using Algorithm 1. The diversity Dis and the classifier ranking results, using Algorithm 1, are shown in Table 4. The second column shows, for each classifier, the diversity (Dis) of the set comprising that classifier and all previously selected (i.e., higher-ranked) classifiers. For example, the diversity of {MobileNet_13, MobileNet_10}, the set of the two highest-ranked classifiers, is 0.0268.

Table 4 Diversity and ranking of the single classifiers

Classifiers	Dis	Ranking
MobileNet_1	0.0320	13
MobileNet_2	0.0293	3
MobileNet_3	0.0290	6
MobileNet_4	0.0321	12
MobileNet_5	0.0315	11
MobileNet_6	0.0290	4
MobileNet_7	0.0306	10
MobileNet_8	0.0295	8
MobileNet_9	0.0292	7
MobileNet_10	0.0268	2
MobileNet_11	0.0302	9
MobileNet_12	0.0287	5
MobileNet_13	0	1

The classifier ranking was used with Algorithm 2 to dynamically select a classifier subset and integrate them to identify each flower image in the test set.

The proposed method was compared with several frequently used CNN models, such as MobileNet, Inception-v1, ResNet-50, Inception-ResNet-v2, and the linear ensemble method. In the experiments, the initial credibility was set as $\epsilon_0=1$. The results of the classification methods are shown in Table 5.

Table 5 Classification accuracy was obtained by various single classifiers, linear ensemble, and the proposed method

Methods	Accuracy/%	
Single model	MobileNet_13	93.88
	Inception-v1	91.56
	ResNet-50	73.46
	Inception-ResNet-v2	81.73
Ensemble model	Linear ensemble	95.06
	Proposed method	95.50

The classification accuracy of the proposed method was 95.50%, an improvement of 1.62%, 3.94%, 22.04%, 13.77%, and 0.44% over the accuracies of MobileNet_13, Inception-v1, ResNet-50, Inception-ResNet-v2, and the linear ensemble method, respectively. The proposed algorithm had a higher recognition rate because an optimal classifier subset was dynamically selected and integrated to identify each flower image. Therefore, the recognition was more targeted, and therefore more accurate. The linear ensemble method used the same subset of classifiers to identify all images, therefore, lacked focus. The results also revealed redundancy among the multiple classifiers. If all classifiers were integrated directly, some invalid decisions would be fused and the final recognition results would be affected. Among all the single classifiers, ResNet-50 had the lowest recognition rates. For the limited training set, a complex network structure was more likely to lead to overfitting and poor recognition performance. However, in the dynamic ensemble selection method, by randomly generating multiple single classifiers with a simple structure and integrating their outputs, the accuracy and

generalization of the recognition algorithm were greatly improved.

4.3 Parameter influence

Initial credibility (ϵ_0) is an important parameter that affects the performance of the proposed dynamic ensemble selection method. To study the influence of initial credibility on the results of the integration algorithm, three initial credibility values were used.

Table 6 Influence of initial credibility on recognition rate of the proposed method

ϵ_0	Number of samples that exceeded the initial credibility	Proportion of test set/%	Accuracy/%
1	836	52.25	95.50
0.90	1466	91.63	95.13
0.80	1495	93.44	94.94

The results showed that the recognition rate of the proposed method was highest with the highest initial credibility value, but the number of identified samples whose credibility values exceeded the initial credibility was lowest (Table 6). Conversely, the recognition rate was lowest with the lowest initial credibility value, but the number of identified samples whose credibility values exceeded the initial credibility was highest. For example, for $\epsilon_0=1$, 836 samples in the test set (52.25%) had credibility values that exceeded the initial credibility. For other samples, classified with Algorithm 2, if one or a few classifiers have credibility values exceeding the initial credibility, there is no need to select and integrate any more classifiers; otherwise, classifiers are added in turn for ensemble recognition until all the classifiers have been selected. If all the classifiers have been selected but the initial credibility has not been reached, the recognition results of all integrations are voted on, and the category with the most votes is considered to be the result of sample identification. Therefore, the dynamic ensemble selection method has better pertinence than a single classifier, and the credibility of each recognition result is assessed, which ensures both the reliability of the recognition results and the accuracy of recognition. However, this means that more classifiers are involved in the integration, leading to long recognition times. Similarly, for $\epsilon_0=0.8$, 1495 samples in the test set (93.44%) had credibility values that exceeded the initial credibility. This result shows that the recognition accuracy for most samples satisfied the requirements of credibility. The remaining samples that were difficult to identify required the use of the dynamic integration recognition method. Therefore, the number of classifiers to be integrated was small and the recognition time was short; however, the accuracy of the corresponding recognition results was relatively low. In practical application, the appropriate initial credibility should be set according to the application scenario, taking into account both the recognition accuracy and efficiency.

5 Discussion

CNNs have been used widely to improve the accuracy of image classification. However, there are still some limitations, as discussed in Section 1. To overcome these difficulties, the advantages of CNNs and ensemble learning were combined and developed a method for the dynamic ensemble selection of CNNs. The experimental results, as shown in Table 5, demonstrated that the proposed model had the best accuracy, compared with the single-CNN models, MobileNet_13, Inception-v1, ResNet-50, Inception-ResNet-v2, and the linear ensemble method. In addition, the performances of the proposed approach in this study with that of previous methods were compared, as shown in Table 7.

Table 7 Comparison with similar studies using the same dataset

Study	Year	Model/Method	Accuracy/%
Guo et al. ^[12]	2019	CNN & Tabu_Genetic algorithm	78.46
Luus et al. ^[35]	2019	CNN & Dimension reduction methods & Semi-supervised learning	79.33
Chen et al. ^[36]	2019	CNN & SVM & Random forest	90.00
Rahman et al. ^[37]	2018	HOG & HSV & SVM	87.00
Toğaçar et al. ^[13]	2019	CNN & Feature selection	98.91
The authors' Work	2020	MobileNets & Dynamic ensemble selection	95.93±0.45

In experiments, the flower dataset was used in the study of Toğaçar et al.^[13] was also used to test the performance of the proposed method. Twenty percent of the images in each class were randomly selected for testing, and five groups of experiments were conducted. The classifier ranking shown in Table 4 was used to dynamically select a classifier subset, which was integrated to identify the flower images in the test sets.

The results in Table 7 show that the proposed method still achieved good classification results. The mean and standard deviation of the accuracy was (95.93±0.45)%. Furthermore, the results demonstrated that the proposed method in this study is accurate and robust. Notably, the method proposed by Toğaçar et al.^[13] had the highest recognition accuracy (98.91%). In their study, the AlexNet, GoogLeNet, ResNet-50, and VGG-16 CNN models were used for feature extraction. In the next step, the efficient features were selected using f-regression and multiple inclusion criteria. As a result, two new feature sets were created with the mentioned feature selection methods and then the intersecting features of these two clusters were extracted. These features were then classified by the SVM method and achieved 98.91% classification accuracy. The intersection of the features obtained by feature selection methods contributed to the classification performance. For the other classification methods

in Table 7, detailed classification performance analysis has been discussed in Reference [13].

Diversity is a necessary condition for high generalization capability in classifier ensembles. In this study, two methods of ensuring diversity were adopted. First, the bootstrap resampling method was used to create different training sets, so that each classifier in the ensemble was trained with a different training set. Second, the homogeneous ensemble technique was used in this proposed method: multiple identical MobileNet models were selected as the single classifiers to be integrated. Nevertheless, the results in Table 7 show that, in the study of Toğaçar et al.^[13], using different CNN models (such as AlexNet, GoogLeNet, ResNet-50, and VGG-16) for feature extraction could obtain better recognition results than the others. Therefore, the heterogeneous ensemble technique will be adopted to ensure diversity among single classifiers in future work.

Effectively identifying and analyzing materials are key procedures for breeding novel crop varieties because of the large quantities of materials and their combinations. Currently, breeding information management systems have been developed and applied, which can both provide more (and more comprehensive) breeding information and improve the accuracy and reliability of breeding decisions^[38-40]. A crop trait information acquisition system was developed by the authors previously (Seed Breeding Cloud Platform, <http://ebreed.com.cn/>; in Chinese)^[41], to effectively improve breeding information management. Furthermore, the wide application of mobile phones provides convenience for real-time and on-field management. The acquisition system has been applied in the breeding of many crops. In future research work, the authors plan to combine the dynamic ensemble learning method with the acquisition system to realize the precise management of flower breeding information. The design of a planned flower breeding management information system is shown in Figure 7.

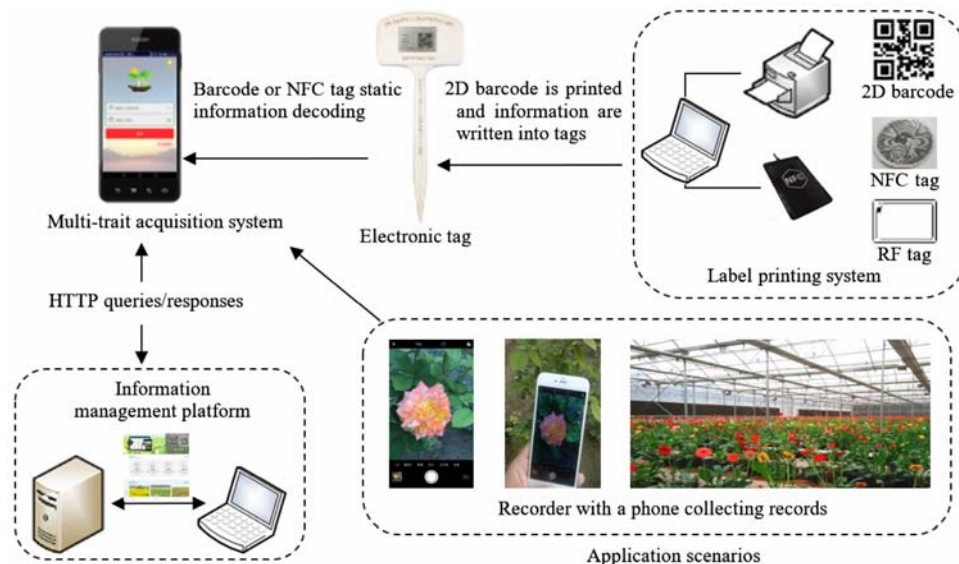


Figure 7 Design of flower breeding management information system

Flower images can be captured using mobile devices, and then they can be linked using the breeding management information system to display growth stages, flower colors and characteristics, and types of disease symptoms or insect infestations. Subsequent image processing, for example, accurate monitoring of flower growth status, determination of disease progression, and recognition of disease types can be implemented using the captured images.

6 Conclusions

In this study, a method for the dynamic ensemble selection of CNNs for flower image classification is described. Pre-trained MobileNet models were used as single classifiers for feature extraction, and thirteen different single MobileNet classifiers were generated randomly, and then dynamically selected and integrated to identify flower species. The initial credibility was adopted to

ensure the reliability of classification results. Classification experiments were performed using images of five flower species. The accuracy of the dynamic ensemble selection method was 95.50%. By comparing the performance results of this proposed method with those of MobileNet, Inception-v1, ResNet-50, Inception-ResNet-v2, the linear ensemble method, and several previous flower classification methods, the accuracy and effectiveness of the proposed method were demonstrated.

Diversity is an important factor for improving the performance of ensemble models. In the future, the heterogeneous ensemble technique will be adopted to ensure diversity among single classifiers. The authors plan to combine the dynamic ensemble learning method with the Seed Breeding Cloud Platform to realize the precise management of flower breeding information.

Acknowledgements

This study was financially supported by the National Key R&D Program of China (Grant No. 2019YFD1101100), the National Natural Science Foundation of China (Grant No. 61403035), and the Science & Technology Innovation Ability Construction Project of Beijing Academy of Agriculture and Forestry Science (Grant No. KJCX20211003). The authors are appreciative of the comments from the anonymous reviewers who provided helpful suggestions on improving this work.

[References]

- [1] Han J, Jin C, Wu L. Research on accuracy of flower recognition application based on convolutional neural. *Advances in Intelligent Systems and Computing*, 2019; 965: 224–232.
- [2] Yuan P, Li W, Ren S, Xu H. Recognition for flower type and variety of chrysanthemum with convolutional neural network. *Transactions of the CSAE*, 2018; 34(5): 152–158. (in Chinese)
- [3] Lee H, Hong K. Automatic recognition of flower species in the natural environment. *Image and Vision Computing*, 2017; 61: 98–114.
- [4] Hsu T, Lee C, Chen L. An interactive flower image recognition system. *Multimedia Tools and Applications*, 2011; 53(1): 53–73.
- [5] Guru D S, Kumar Y H S, Manjunath S. Textural features in flower classification. *Mathematical and Computer Modelling*, 2011; 54: 1030–1036.
- [6] Cheng K, Tan X. Sparse representations based attribute learning for flower classification. *Neurocomputing*, 2014; 145: 416–426.
- [7] Soleimanipour A, Chegini G R. A vision-based hybrid approach for identification of Anthurium flower cultivars. *Computers and Electronics in Agriculture*, 2020; 174: 1–8.
- [8] LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*, 2015; 521: 436–444.
- [9] Wang W, Yang Y, Wang X, Wang W, Li J. Development of convolutional neural network and its application in image classification: A survey. *Optical Engineering*, 2019; 58(4): 1–9.
- [10] Khan A, Sohail A, Zahoora U, Qureshi A S. A survey of the recent architectures of deep convolutional neural networks. *Artificial Intelligence Review*, 2020; 53: 5355–5516.
- [11] Cibuk M, Budak U, Guo Y, Cevdet Ince M, Sengur A. Efficient deep features selections and classification for flower species recognition. *Measurement*, 2019; 137: 7–13.
- [12] Guo B, Hu J, Wu W, Peng Q, Wu F. The Tabu_Genetic algorithm: A novel method for hyper-parameter optimization of learning algorithms. *Electronics*, 2019; 8(5): 579. doi: 10.3390/electronics800579.
- [13] Toğaçar M, Ergen B, Cömert Z. Classification of flower species by using features extracted from the intersection of feature selection methods in convolutional neural network models. *Measurement*, 2020; 158: 107703. doi: 10.1016/j.measurement.2020.107703.
- [14] Hiary H, Saadeh H, Saadeh M, Yaqub M. Flower classification using deep convolutional neural networks. *IET Computer Vision*, 2018; 12(6): 855–862.
- [15] Mitrovic K, Milosevic D. Flower classification with convolutional neural networks. *23th International Conference on System Theory, Control and Computing*. Sinaia: IEEE, 2019; pp.845–850.
- [16] Dong X, Yu Z, Cao W, Shi Y, Ma Q. A survey on ensemble learning. *Frontiers of Computer Science*, 2020; 14: 241–258.
- [17] Li J, Wu S, Liu C, Yu Z, Wong H. Semi-supervised deep coupled ensemble learning with classification landmark exploration. *IEEE Transactions on Image Processing*, 2020; 29: 538–550.
- [18] Zvarevashe K, Olugbara O. Ensemble learning of hybrid acoustic features for speech emotion recognition. *Algorithms*, 2020; 13(3): 70. doi: 10.3390/a13030070.
- [19] Yang X, Xu Y, Quan Y, Ji H. Image denoising via sequential ensemble learning. *IEEE Transactions on Image Processing*, 2020; 29: 5038–5049.
- [20] Bae K I, Park J, Lee J, Lee Y, Lim C. Flower classification with modified multimodal convolutional neural networks. *Expert System with Applications*, 2020; 159: 113455. doi: 10.1016/j.eswa.2020.113455.
- [21] Huang B, Hu Y, Sun Y, Hao X, Yan C. A flower classification framework based on ensemble of CNNs. *PCM 2018. Lecture Notes in Computer Science*, 2018; 11166: 235–244.
- [22] Wang Z, Wang K, Wang X, Pan S. A convolutional neural network ensemble for flower image classification. *9th International Conference on Computing and Pattern Recognition*. Xiamen: ACM, 2020; pp.225–230.
- [23] Li H, Wang X, Ding S. Research and development of neural network ensembles: a survey. *Artificial Intelligence Review*, 2018; 49(4): 455–479.
- [24] Akbari E, Dahlan H M, Ibrahim R, Alizadeh H. Hierarchical cluster ensemble selection. *Engineering Applications of Artificial Intelligence*, 2015; 39:146–156.
- [25] Yin X, Huang K, Hao H. DE²: dynamic ensemble of ensembles for learning nonstationary data. *Neurocomputing*, 2015; 165: 14–22.
- [26] Flower dataset. Available: <http://download.tensorflow.org/data/flowers.tar.gz>. Accessed on [2020-02-18].
- [27] Subject Database of China Plant. Available: <http://www.plant.csdb.cn/photo>. Accessed on [2020-02-20].
- [28] Howard A G, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, et al. MobileNets: efficient convolutional neural networks for mobile vision applications. 2017; arXiv: 1704.04861.
- [29] Yan Y, Yin X C, Wang Z B, Yin X W, Yang C, Hao H W. Sorting-based dynamic classifier ensemble selection. *12th International Conference on Document Analysis and Recognition*, Washington: IEEE, 2013; pp.673–677. doi: 10.1109/ICDAR.2013.138.
- [30] Cruz R M O, Sabourin R, Cavalcanti G D C. Dynamic classifier selection: recent advances and perspectives. *Information Fusion*, 2018; 41: 195–216.
- [31] Alam K M R, Siddique N, Adeli H. A dynamic ensemble learning algorithm for neural networks. *Neural Computing and Applications*, 2020; 32: 8675–8690.
- [32] Tang E, Suganthan P N, Yao X. An analysis of diversity measures. *Machine Learning*, 2006; 65(1): 247–271.
- [33] Bishop C M. *Pattern recognition and machine learning*. Springer Science+Business Media, LLC. 2006; 68p.
- [34] Liu C. Classifier combination based on confidence transformation. *Pattern Recognition*, 2005; 38: 11–28.
- [35] Luus F, Khan N, Akhalwaya I. Active learning with TensorBoard, 2019; arXiv: 1901.00675.
- [36] Chen B, Liu J, Sun J, Liu J. Flowers classification via deep learning models. Available: <http://noiselab.ucsd.edu/ECE228/Posters/Group40.pdf>. Accessed on [2019-07-25]
- [37] Rahman M A, Kuswari A Y. Klasifikasi Jenis Bunga menggunakan SVM dengan Fitur HSV dan HOG. *Yukiyu Science*. Available: <http://eprints.mdp.ac.id/id/eprint/2657>. Accessed on [2020-03-16].
- [38] Han Y Y, Wang K Y, Liu Z Q, Pan S H, Zhao X Y, et al. Golden seed breeding cloud platform for the management of crop breeding material and genealogical tracking. *Computers and Electronics in Agriculture*, 2018; 152: 206–214.
- [39] Yue Y, Zhao G, Sun R. Breeding data service platform based on the new architecture of cloud technology. *3th Advanced Information Technology, Electronic and Automation Control Conference*, Chongqing: IEEE, 2018; pp.1457–1463.
- [40] Chen S, Gao H, Li L, Liu Y. An information acquisition and management system for maize breeding. In: *6th IEEE International Conference on Software Engineering and Service Science*, Beijing: IEEE, 2015; pp.807–813. doi: 10.1109/ICSESS.2015.7339179.
- [41] Han Y Y, Wang K Y, Liu Z Q, Pan S H, Zhao X Y, et al. A crop trait information acquisition system with multitag-based identification technologies for breeding precision management. *Computers and Electronics in Agriculture*, 2017; 135: 71–80.