Ecological Informatics 82 (2024) 102683

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

Medicinal and poisonous plants classification from visual characteristics of leaves using computer vision and deep neural networks

Rahim Azadnia [a,*], Faramarz Noei-Khodabadi [a], Azad Moloudzadeh [b], Ahmad Jahanbakhshi [c], Mahmoud Omid [a,*]

[a] Department of Biosystems Engineering, Faculty of Agriculture, University of Tehran, Karaj, Iran
[b] Department of Plant Biology, Faculty of Science, University of Urmia, Urmia, Iran
[c] Department of Biosystems Engineering, University of Mohaghegh Ardabili, Ardabil, Iran

ARTICLE INFO

Keywords:
Medicinal plants
Image classification
Machine learning
Deep learning
Data augmentation
Fast AutoAugment

ABSTRACT

Poisonous plants are the third largest category of poisons known globally, which pose a risk of poisoning and death to humans. Currently, the identification of medicinal and poisonous plants is done by humans using experimental methods, which are not accurate and are associated with many errors, and also the use of laboratory methods requires experts and this method is very costly and time-consuming. Therefore, distinguishing between medicinal and poisonous plants is very important using emerging, non-destructive, fast and accurate methods such as computer vision and artificial intelligence. In this study, we propose a robust and generalized model using spatial attention (SA) and channel attention (CA) modules for the classification of different plants. A dataset containing 900 confirmed images of three plant classes (oregano, poisonous and weed) was used. The attention mechanisms enhance efficiency of deep learning (DL) networks by allowing them to precisely focus on all relevant input elements. In order to enhance the performance of the proposed model, the CA was implemented based on four pooling operations including global average pooling-based CA (GAP-CA), mixed pooling-based CA (Mixed-CA), gated pooling-based CA (Gated-CA), and tree pooling-based CA (Tree-CA) operations. The results showed that the DL model based on Tree-CA had promising performance and outperformed other state-of-the-art models, achieving the values of 99.63%, 99.38%, 99.52%, 99.74%, and 99.42%, for accuracy, precision, recall, specificity, and F1-score, respectively. The findings support our proposed attention model's success in identifying medicinal plants from similar poisonous plants. Recent advancements in computer-based technologies and artificial intelligence enable automatic detection of medicinal and poisonous plants, revolutionizing traditional identification methods.

1. Introduction

Medicinal plants have a special value and importance in ensuring the health of communities both in terms of treatment and prevention of diseases as they have fewer side effects compared to chemical drugs. In addition to medicinal uses, these plants can also be used as food and to prepare drinks and cosmetics (Azadnia and Kheiralipour, 2021).

Origanum vulgare (*Lamiaceae*) is a perennial medicinal plant found in America, Asia, Europe, and North Africa (Kintzios, 2002). In traditional medicine, it is used as a remedy for several health problems including skin sores, muscle pain, asthma, cramping, indigestion, common cold and upset stomach (Ansari et al., 2022; Bora et al., 2022; Pezzani et al., 2017). This plant, which can be found almost everywhere in the world, has different species, some of which are very similar to non-edible and poisonous plants. Poisonous plants that resemble medicinal and edible ones are considered a serious threat for livestock farmers and humans all over the globe since they have brought about much damage and injuries to them (Panter et al., 2019). Consumption of poisonous plants causes

Abbreviations: CA, Channel attention; DL, Deep learning; RNN, Recurrent neural network; FAA, Fast AutoAugment; GELU, Gaussian error linear unit; FC, Fully connected; SELU, Scaled exponential linear unit; VGG, Visual geometry group; TP, True positive; FN, False negative; FP, False positive; TN, True negative; SA, Spatial attention; CNN, Convolutional neural network; DA, Data augmentation; CWA, Channel-wise attention; ReLU, Rectified linear unit; GAP, Global average pooling; SCAM, Spatial channel attention module; CPU, Central processing unit; GPU, Graphics processing unit; GB, Gigabyte; GHz, Gigahertz.

* Corresponding authors.
E-mail addresses: rahim.azadnia@ut.ac.ir (R. Azadnia), omid@ut.ac.ir (M. Omid).

https://doi.org/10.1016/j.ecoinf.2024.102683
Received 20 October 2023; Received in revised form 9 June 2024; Accepted 9 June 2024
Available online 10 June 2024
1574-9541/

mouth burning, skin burning, diarrhea, headache, blurred vision, nausea and many other diseases (Tamilselvan et al., 2014). Therefore, the production and distribution of fake, low-quality and toxic medicinal plants in the whole world can result in irreparable effects (Pushpa et al., 2024; Chukwuma et al., 2015).

A complete knowledge of the environment and the region in which we live is so important as such knowledge can have a significant impact on our future, both in terms of economy and public health. For example, a farmer should know the plants that grow inside and outside the field in order to avoid contamination of the products during harvest. Similarly, a rancher should be able to identify poisonous plants that are harmful to livestock. Every year, there are numerous reports of poisoning and, in more severe cases, death of humans and animals due to consumption of or contact with poisonous plants. Such plants have also caused serious financial loss in some cases. Considering the fact that poisonous plants can cause a lot of damage to humans in general and livestock farmers in particular, possible risks should be minimized by increasing our knowledge about such plants and using smart equipment to detect them. Today, due to the lack of advanced equipment and sufficient information, it is very difficult for local people to accurately distinguish between medicinal plants and other non-edible poisonous ones. Therefore, the development of medicinal plant identification tools is of particular importance. Currently, botanists identify different species of medicinal plants using traditional botanical classification methods which are mostly based on experience. Using these traditional methods is usually time-consuming and tiring and requires an expert in the field. Recently, the development of fast and accurate computer-based technologies and the use of artificial intelligence have enabled automatic detection of various medicinal and poisonous plants. By using these new and non-destructive techniques, the traditional ways of identifying medicinal plants can be converted into mechanical and intelligent methods. Researchers have reported that identification and classification of medicinal plants with the help of computers, machine learning algorithms and image recognition based on deep learning (DL), are the most accurate and reliable detection methods (Azadnia et al., 2022a; Islam et al., 2023; Roopashree et al., 2022).

Traditional machine learning methods for plant recognition focus on features such as plant shape (Aakif and Khan, 2015; Saleem et al., 2019), texture (Naresh and Nagendraswamy, 2016; Sood and Singh, 2020), venation (Ghasab et al., 2015; Larese et al., 2014), or considering them jointly (Yang, 2021; Ahmad et al., 2021; Azadnia and Kheiralipour, 2021). Nevertheless, these methods are now outdated and outperformed by the latest advancements in DL (Wani et al., 2020). Recently, classification of medicinal plants (Azadnia et al., 2022a; Bodhwani et al., 2019; Tiwari et al., 2022) and identification of plant diseases (Ali et al., 2024; Chen et al., 2022; Joshi et al., 2021; Sethy et al., 2020) have been successfully conducted through deep neural networks.

Some species of *Origanum vulgare* (*Lamiaceae*) are very similar to inedible and poisonous plants in terms of shape, color and texture, and thus are difficult to distinguish from useful species even for experienced experts (Fig. 1).

Therefore, it is important to take into account several features and place emphasis on the features that are efficient for different herbs. Performance of deep neural networks will be enhanced when the attention mechanism enables them to precisely focus on all relevant input elements. This method has become a popular technique in DL for text classification, image interpretation, and sentiment analysis (Wani et al., 2020). Models like attention mechanisms have high accuracy, and most of these mechanisms can be jointly trained with a basic model, such as a recurrent neural network or a convolutional neural network (CNN), using a regular back-propagation algorithm. In DL models, feature extraction takes place hierarchically through CNNs from a global perspective. At the same time, the attention mechanism focuses on the significant information of an image, improving the performance of CNN models through emphasis on that important information (Brauwers and Frasincar, 2021; Hassanin et al., 2024).

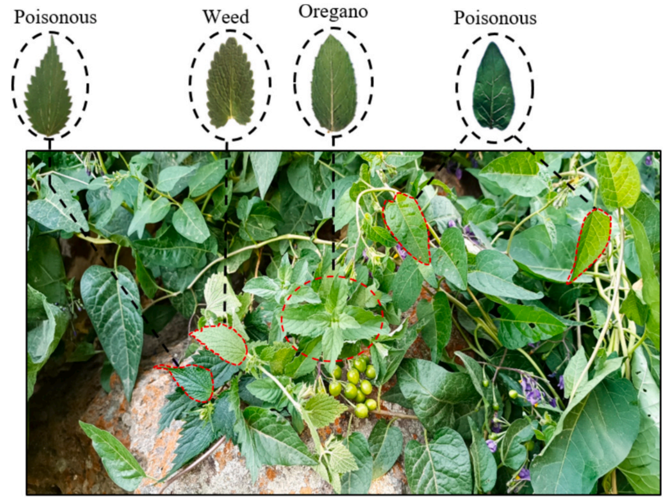

Fig. 1. Different plant images under the complex field environment.

Different attention mechanisms have specific applications. Channel attention (CA) and spatial attention (SA) can be mentioned as two of the most important applications. However, both methods have their own limitations as some important features may be ignored in each of them. Therefore, it would be wise to use a combination of both attention mechanisms in order to obtain comprehensive and accurate information that helps us identify different types of medicinal and non-edible (poisonous) plants. Recently, the use of attention mechanisms to identify plants (Brindha and Gopi, 2022; Turgut et al., 2022; Wang et al., 2021), weeds (Hasan et al., 2023; Wang et al., 2022; Zhang et al., 2022a) and fruits (She et al., 2022; Yang et al., 2022; Lu et al., 2022; Xue et al., 2020; Azadnia and Kheiralipour, 2022) has become the focus of many researchers in the field of agriculture. This paper is aimed at evaluating a network that combines CA and SA modules. Such a network integrates both CA and SA in order to enhance our ability to distinguish medicinal plant from other non-edible or even perilous ones. As far as we know, this is the first piece of study that attempts to propose an efficient computer vision approach for detecting medicinal plants and telling them apart from similar poisonous plants through the application of attention modules. It is hoped that the following goals will be realized through this research:

1) A model (CSA-Herb) will be proposed in order to evaluate the most important locational information in the images of medicinal plants. The model will extract accurate and important locational and channel information for the classification of medicinal plants.
2) Different CA modules will be provided on the basis of pooling operations to aggregate the channel information of the feature map.
3) To realize the aim of image recognition, a dataset of similar medicinal and poisonous plants will be gathered and augmented through a Fast AutoAugment approach.

2. Material and methods

2.1. Sample description

In this research, the leaves of three categories of plants (i.e. medicinal plants, weeds and poisonous plants) were identified and collected by an herbalist in the Iranian northwestern cities of Salmas, Khoy and Urmia. The collected samples were so similar in appearance that it was very difficult for ordinary people to identify them from each other visually. Images of 300 leaf samples from each class (900 images in total) were collected for this research. The leaf samples were placed in zipped plastic bags and taken to the laboratory for imaging immediately after harvesting. A mobile phone camera (Redmi Note 11S, Xiaomi

Corporation, China, 108MP camera) was used to photograph the samples. The background of the images, focus and camera angle were fixed. However, the photographing distance and the lighting conditions of the environment during the photographing of the samples were not constant. Therefore, the imaging was carried out in non-controlled light conditions. Fig. 2 shows some pictures taken from the samples.

2.2. Augmentation protocol

Deep learning (DL) has become an advanced method in the field of computer science, but despite their high learning capacity, models based on DL often face a challenge called overfitting. To overcome the overfitting problem, deep neural networks must be supported by a significant amount of labeled data. Data augmentation (DA) technique is used as a basic method to quantitatively increase and diversify training data (Momeny et al., 2021; Momeny et al., 2023). DA is a technique in which the training set is artificially increased by creating modified copies of a dataset using the existing data. Using a designed set of DAs instead of simple random transformations during data training can significantly improve the generalization ability of the network (Hsia et al., 2022; Jahanbakhshi et al., 2021a; Momeny et al., 2022). As the name suggests, automatic DA methods work automatically, and help us avoid the performance limitations created by the human exploration methods. Fast AutoAugment (FAA) is an enhanced Auto augmentation method presented by to automatically search for optimal DAs methods. FAA can minimize the computational complexity of DL and significantly improve performance in the image classification process. The method consists of controllers, augmenters, and base model. The controller samples the data augmentation policies from the searched space. The images of the dataset will be transformed with the new policy by the augmenter. A baseline CNN model is then applied to train the augmented images created with the new policy (Faryna et al., 2024; Jahanbakhshi et al., 2021c).

As Fig. 3 shows, the data are first divided into five equal folds in the proposed FAA-based approach. After that, policies will augment each fold without repetition. Then, the CNN model starts to process the folds. In the next step, the optimal augment policies are evaluated and recognized. The output of each CNN is controlled by a Bayesian optimizer. Weak policies are then removed by the controller and strong ones are selected. The process continues to find the optimal policies for data augmentation.

2.3. The proposed approach (SCAM-Herb model)

CNN-based methods have better performance in identifying plants, but still, based on data analysis and some previous research (Wang et al., 2020, Azadnia et al., 2023), texture characteristics are very important. Therefore, more attention should be paid to the modules that consider texture features. A SCAM-Herb model based on deep recognition structure is proposed in this study to address these problems in plant image recognition, especially for those with similar features. The proposed module architecture is constructed based on the ResNeSt model, which is a ResNet variant with a CA and SA blocks (He et al., 2016; Zhang et al., 2022b). The proposed model employs a group of convolution layers from the ResNeXt network with a channel-wise attention (CWA) mechanism while maintaining the ResNet network structure (Xie et al., 2017). The CWA mechanism uses the ResNeSt block to facilitate information interaction between cross feature map groups. To put it more simply, the SCAM-Herb model is primarily built on the ResNeSt structure. Additionally, it uses SA along with CWA for split attention. This allows the model focus on both the position and importance of the most crucial information in the feature maps. Therefore, we can receive important information through a variety of receptive fields. In this study, we use a similar configuration to ResNeSt, along with the combined improved attention module. Fig. 4 indicates the proposed model block diagram.

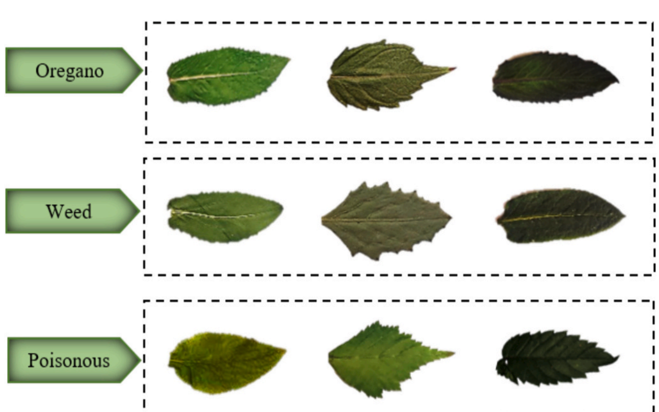

Fig. 2. Images of three different classes of plants.

The h, w, and c in Fig. 4 are the height, width, and the number of feature map channels, respectively. 'r' is the number of splits within a cardinal group, and 'k' is the number of cardinal groups, which is defined as follows:

$$\widehat{U}^k = \sum_{j=R(k-1)+1}^{Rk} U_j \tag{1}$$

In the proposed module, global and local information are acquired through combination of CA and SA. The proposed CA model was proposed by changing the split module used in the ResNeSt model. So, different max-pooling layers are employed in CA module to emphasize important texture features in images. The following sections will comprehensively explain both the CA and SA modules.

2.3.1. Proposed channel attention (CA) module

The proposed model is based on the ResNeSt network, with the difference that a SA is added to the CWA in the split attentions section. In fact, SA was applied to help the model focus more on where and why the key feature maps are needed. In addition to the crucial information of the feature maps, it can be of great importance to obtain global information about the images. Moreover, the complexity of texture features can be simplified by eliminating inefficient features and retaining relevant ones. To this end, various pooling layers are applied in the split-attention module to efficiently extract important texture features and global information while removing unwanted noise information. The following is a summary of the proposed CA module structures:

2.3.1.1. Mixed-pooling based CA module (Mixed-CA). Mixed-pooling is a technique developed by combining Max-pooling and average-pooling layers (Lee et al., 2016). Here, in the mixed-CA module, we used a mixed-pooling layer in the SA of the ResNeSt network to remove unwanted information while accurately extract key texture features and global information. The mixed-pooling layer in this module is given below:

$$f_{mix}(x) = a_l.f_{max}(x) + (1 - a_l).f_{avg}(x) \tag{2}$$

where $a_l \in [0,1]$ is the value that determines the combination of max pooling and average pooling. In the mixed-CA model, the Gaussian Error Linear Unit (GELU) function was utilized for the activation function.

2.3.1.2. Gated-pooling based CA module (Gated-CA). In the gated-CA model, gated-pooling layer is used in SA. Like the mixed-pooling layer, the gated-pooling layer consists of the max-pooling and average-pooling layers, with the difference that gated-pooling is a responsive strategy that directly learns the mixture propagation, which is determined after the learning process is completed. In addition, in this method a gated mask will be learned. The gated-pooling method is

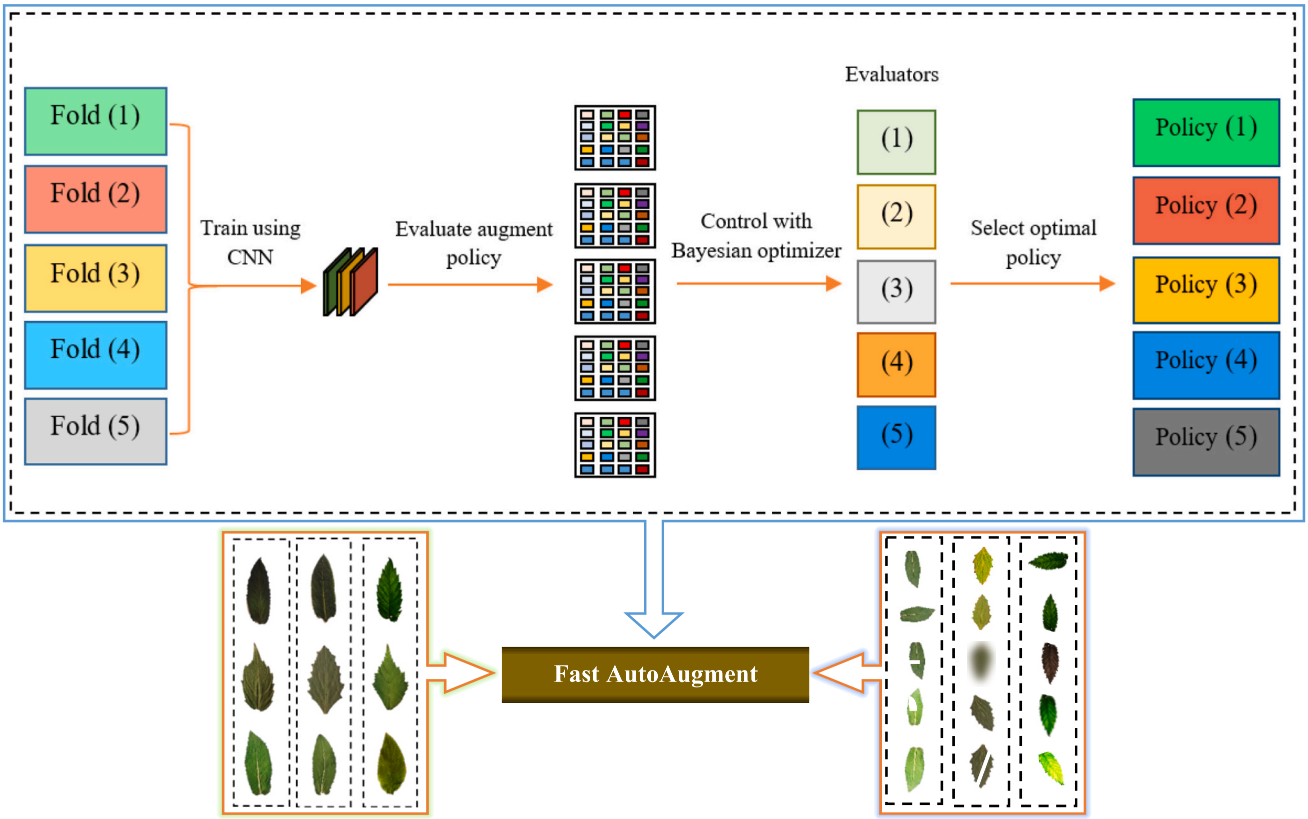

Fig. 3. FAA-based data augmentation architecture.

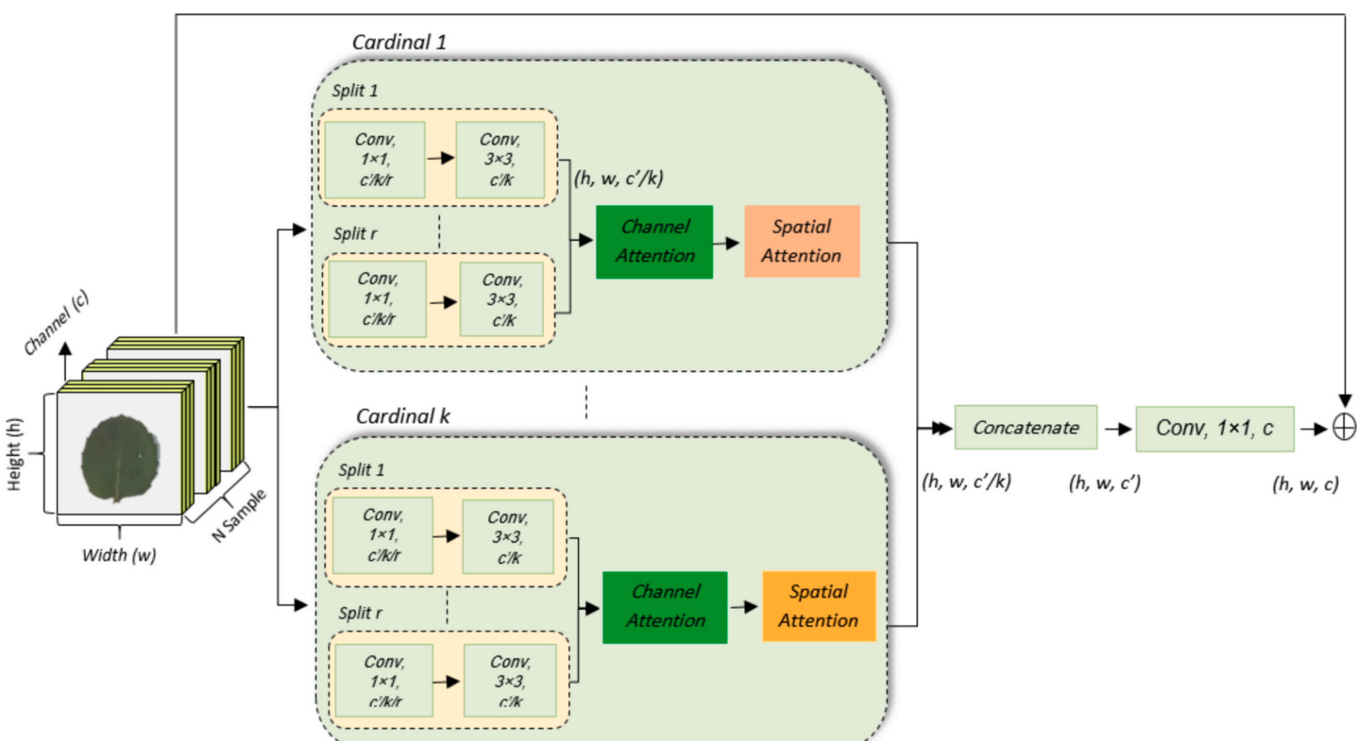

Fig. 4. The proposed SCAM-Herb architecture.

defined as follows:

$$f_{gate} = \sigma(\omega^T x) f_{max}(x) + (1 - \sigma(\omega^T x)) f_{avg}(x) \quad (3)$$

$$\sigma(\omega^T x) = 1/(1 + exp\{-\omega^T x\}) \quad (4)$$

where $\sigma(\omega^T x) \in [0, 1]$ is the responsive mixing proportion, x is used to denote the values in the region being pooled and ω is the internal node of gating mask. The Rectified Linear Unit (ReLU) activation function was used in module.

2.3.1.3. Tree-pooling based CA module (Tree-CA). In Tree-CA model the tree-pooling layer was used in SA section. One of the natural generalizations of pooling operations is to let the pooling operations themselves go through the learning process. These pooling layers are separate from convolutional layers and perform pooling separately in each channel. This channel separation is used to introduce a low number of parameters as compared to a convolutional layer. In the tree-pooling we learn pooling filters and also learn to responsively combine those learned filters. Both aspect of this learning occurs in the binary trees, which have predefined levels unlike traditional decision trees where each leaf is associated with a trained hybrid filter. According to this study, a level 2 tree (Leaf nodes "1", "2" and internal node of "3") is defined as follows:

$$f_{tree}(x) = \sigma(\omega_3^T x) v_1^T x + (1 - \sigma(\omega_3^T x)) v_2^T x \quad (5)$$

where v_1 and v_2 are leaf node pooling filters.

2.3.1.4. Global average pooling (GAP) based CA module (GAP-CA). In a typical CNN model, the convolutional layers are usually accompanied by one or more fully connected (FC) layers and are finally followed by a softmax layer. The feature map is converted to a one-dimensional vector using these FC layers. These FC layers retain all of the spatial information, resulting in too many model parameters and overfitting. To solve this problem, Lin et al. (2013) proposed a method in which the FC layer was replaced by the GAP layer. The GAP layer converts each feature map into a single number by averaging the numbers within the h × w matrix. This reduces the spatial dimensions of the tensors, thus reducing the number of parameters and preventing the occurrence of overfitting. The Scaled Exponential Linear Unit (SELU) activation function was used in this attention module. A schematic representation of the proposed CA modules is shown in Fig. 5.

2.3.2. Proposed spatial attention (SA) Module

Each part of the plant images contains distinct information. For example, edge position information is more important than information in other areas. The principal aim of SA is to highlight significant features. This attention module identifies important image regions by distributing weight across spatial positions. So, the features of the desired area are highlighted, preventing the extraction of noisy features. A SA map is generated using the inter-spatial relationship among the features. In fact, SA focuses more on where important information is located and is a kind of complement to CA. In this study, gated-pooling along the channel axis was used to calculate SA, and to create an efficient feature descriptor. The structure of the SA module is shown in Fig. 6.

The gated-pooling along the axis of the channel is used to identify regions containing useful information (Zagoruyko and Komodakis, 2016). After defining a feature descriptor, a convolutional layer was used to create a SA map $M_s(F) = R^{H \times W}$ for highlighting the important regions. In fact, we extracted CA of the feature map by the max-pooling and average-pooling operations in the gated method and created two 2D maps. The constructed maps were connected by an optimal convolutional layer in order to create a 2D SA map. The SA is calculated as follows:

$$\begin{aligned} M_{s1}(F) &= \sigma\left(f^{7\times7}\left([\sigma(\omega^T x).f_{max}(F); 1 - \sigma(\omega^T x).f_{max}(F)]\right)\right) \\ &= \sigma\left(f^{7\times7}\left(\left[F^s_{max}; F^s_{avg}\right]\right)\right) \end{aligned} \quad (6)$$

where σ is a sigmoid activation function and $f^{7\times7}$ is a convolution operation with a filter size of 7 × 7. The final SA weight in the main feature map is computed through the following equation:

$$M_{s2}(F) = W \times F \quad (7)$$

where W shows the weight obtained by the SA module, and F is the input feature map. Therefore, the presented model extracts various features based on local and global information, which facilitates the interpretation of the original image and increases the classification accuracy.

2.4. Data split

How to split the data for training and testing of the proposed models is reported in Table 1.

2.5. Evaluation of classifiers

The performance of the proposed model was compared with other state-of-the-art models. To do this, the collected dataset was trained and evaluated using different residual structures, Inception-v3, VGG-11, VGG-19, Alex Net, GoogleNet, and Squeeze Net models. In this research, classification operations of images were carried out in MATLAB R2016a software. The proposed models were simulated in the deep learning toolbox on a system equipped with an Intel(R) Core i7-5500U, CPU 3.0 GHz with 64 GB of RAM, and Nvidia GeForce 1070 GPU.

2.6. Evaluation criteria of classifiers

In order to evaluate the performance of all image classification models, the criteria of accuracy, precision, recall, specificity, and F1-score were used based on eqs. 8 to 12 (Azadnia et al., 2022b; Azizi et al., 2024; Marhamati et al., 2024).

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (9)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (10)$$

$$\text{Specificity} = \frac{TP}{TP + FN} \quad (11)$$

$$F1 - score = 2 \times \frac{Percision \times Recall}{Precision + Recall} \quad (12)$$

3. Results

In this study, the fast auto-augmentation (FAA) method was applied to improve the data. An improved search strategy based on density matching allows FAA to discover effective augmentation policies. Thus, in this study, 80% of the total data (8640 images) were used for training and 20% (180 images) were used randomly to test the proposed SCAM-Herb model (Fig. 7).

In Fig. 8, the performance of the presented models in different epochs is reported. As can be seen, the performance of the proposed models increases with the number of epochs and the number of training data. According to Fig. 8a, the SCAM-Herb model based on Tree-CA module with 50% of dataset and 100 epochs achieved an accuracy of 96.64%. With 90% of the trained dataset, this model was able to diagnose

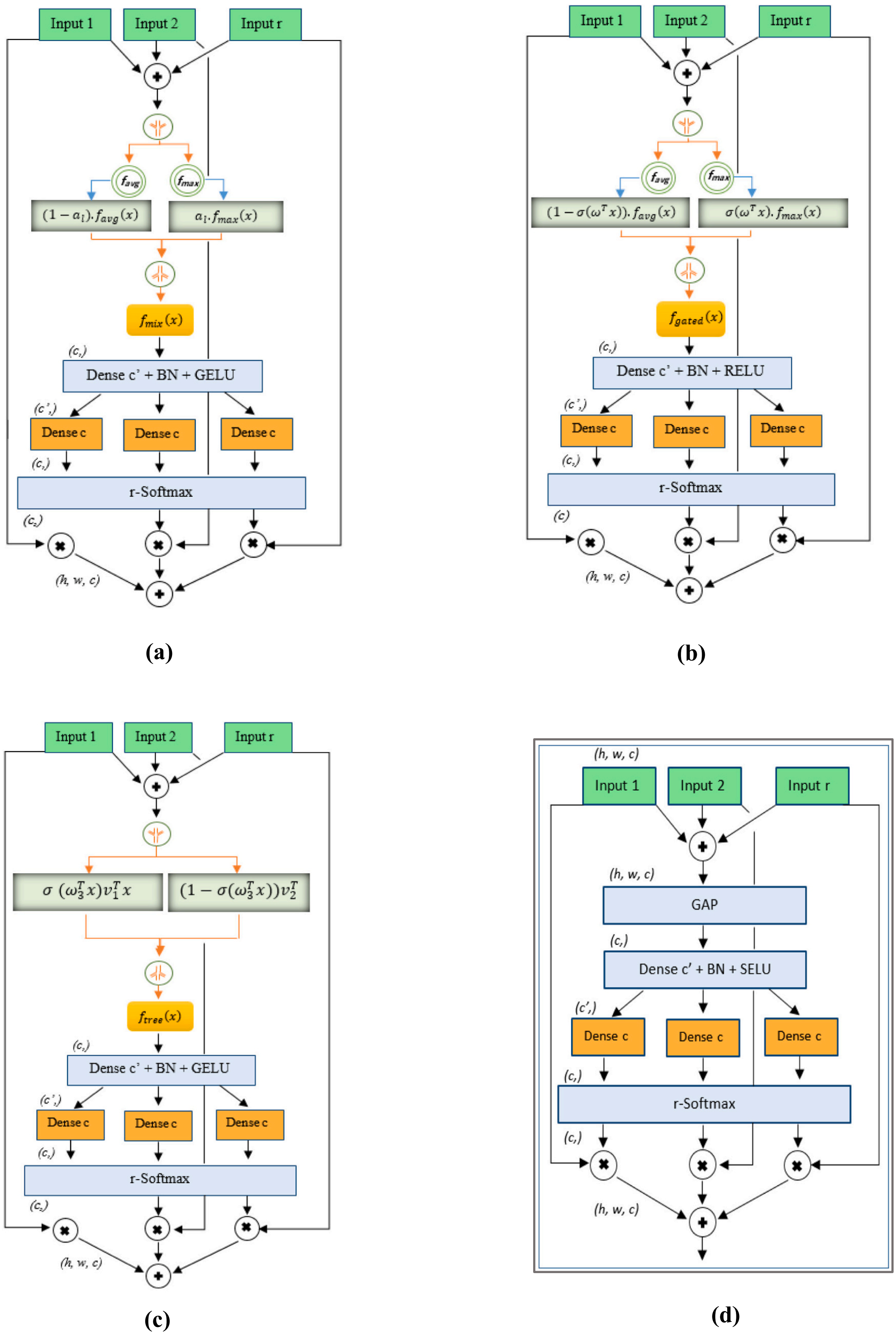

Fig. 5. Schematic representation of improved split attention used in the proposed CA module: a) Mixed-CA, b) Gated-CA, c) Tree-CA, and d) GAP-CA.

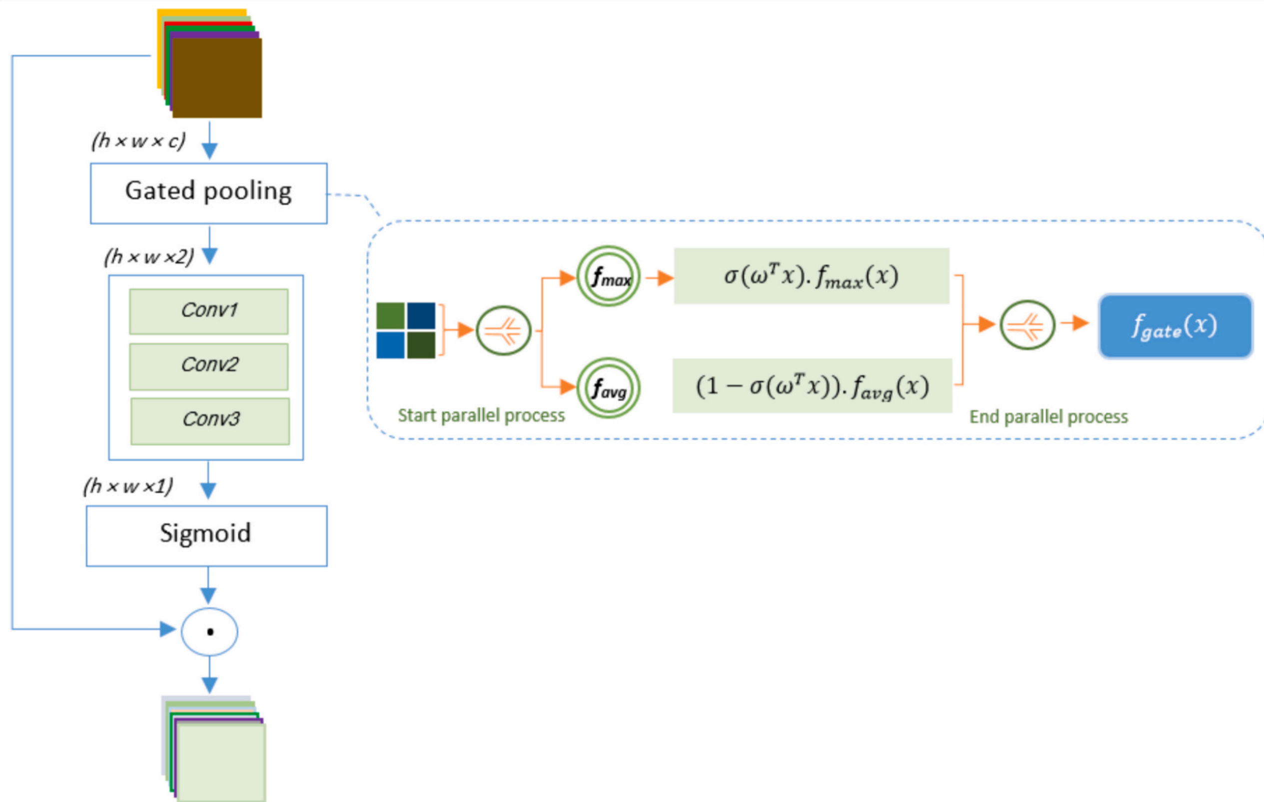

Fig. 6. The structure of SA module.

Table 1
Data splitting in the proposed models.

Factors	Data splitting percentage	Number of original images	Number of augmented data	Total number of images after data augmentation
Train	80%	720	7920	8640
Test	20%	180	–	180
Total	100%	900	7920	8820

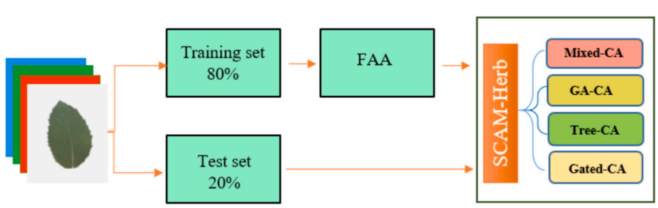

Fig. 7. Data splitting in the proposed model.

different plants with 99.78% accuracy, demonstrating the excellent performance of the proposed model in focusing and quickly extracting important features and a high recognition rate. Also, the SCAM-Herb model based on gated-CA module was able to achieve a high accuracy of 99.52% with 90% of training data (Fig. 8b). Moreover, the SCAM-herb model based on mixed-CA and GAP-CA modules with 50% of the training data achieved accuracies of 94.15% and 93. 23%, respectively, which performed relatively poorly compared to the two aforementioned models (Fig. 8c and d). However, with increasing the training data, the accuracy of the two models increased to 99.38% and 99.43%, respectively. Notably, when more than 80% of the data were used to train the models, the accuracy of the models was nearly constant or less variable. Therefore, 80% of the total data was sufficient for training the models, and the remaining 20% was used to test the model and improve the generalization performance of the models.

The confusion matrix for evaluating the proposed models based on the test data is shown in Fig. 9. The results show that according to the tree-CA module, all medicinal (oregano) and poisonous plants are correctly classified in their classes. In the class related to weeds, one item is wrongly graded in the class of medicinal plant (oregano) (Fig. 9a). The model based on the gated-CA module has misclassified an image from the medicinal plant (oregano) class and an image from the weed class in other classes (Fig. 9b). The proposed model based on mixed-CA has wrongly classified the number of five images in all classes. Two images were related to medicinal plants, two were related to weeds, and one was related to poisonous plants that were wrongly classified in other classes (Fig. 9c). The model based on GAP-CA presented a weaker performance than other models (such as tree-CA, gated-CA, and mixed-CA). This model wrongly graded 10 images, including four images of medicinal plants, three images of weeds, and three images of poisonous plants (Fig. 9d). This could be attributed to the different leaf appearance characteristics of each class, as many of the leaves have sharp and broad edges that were similar to each other. Another possible reason is the presence of thin and angular veins, which caused the proposed models to misclassify a number of samples.

It should be noted that the incorrectly identified samples had similar and homogeneous characteristics in terms of color, shape, and texture, posing challenges to the proposed model. Some misclassified examples are shown in Fig. 10.

Several statistical metrics including accuracy, precision, recall, specificity, and F1-score were calculated to evaluate the efficacy of proposed model and their results are reported in Table 2. The proposed models such as Tree-CA, Gated-CA, Mixed-CA and GAP-CA have been able to classify plant images into three different classes with an average accuracy of 99.63%, 99.07%, 98.13%, and 96.23% respectively, which shows the excellent success of all classifications for grading these plants. Also, the results of Table 2 show that the tree-CA model had the highest

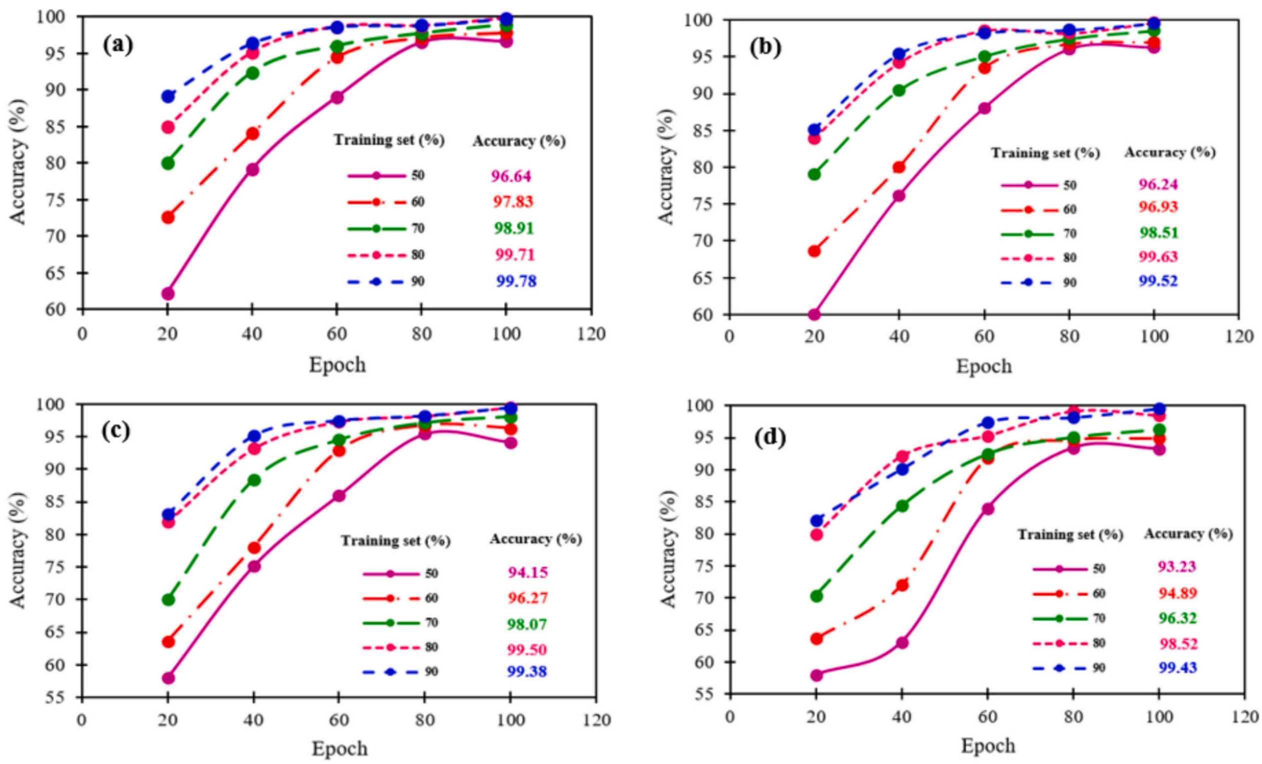

Fig. 8. Accuracy of training sets at different epochs by SCAM-Herb model based on (a) Tree-CA, (b) Gated-CA, (c) Mixed-CA, and (d) GAP-CA modules.

performance in plant classification with an average accuracy of 99.63%.

The weighted average of recall and precision is used to determine the F1-score. The higher the F1-score value for a model (closer to 100%), the more accurate the performance of that model (Hadipour-Rokni et al., 2023; Jahanbakhshi et al., 2021b). The results in Table 2 show that the proposed SCAM-Herb models such as Tree-CA (99.42%), Gated-CA (98.87%), Mixed-CA (96.92%), and GAP-CA (94.47%) respectively, they have the best performance and the highest value in the F1-score criterion.

The SCAM-Herb model based on different CAs was able to classify all three different plant classes with an excellent recognition rate. However, the SCAM-Herb model based on the Tree-CA module outperformed other proposed models with an average accuracy of 99.63%, indicating the ability of the proposed model to extract local and global features from plants. It should be noted that the GELU activation function was used in the tree-CA model, which contributed to the superiority of this model over other proposed models.

Fig. 11 clearly shows the feature distribution and mask performance of the SCAM-Herb model to explain the presented model. Based on this, the proposed model tends to focus on global information in deeper layers. As a result, these layers are more concentrated on the covering areas of the object. But the initial layers only consider the local features. Consequently, the combination of the CA and SA modules is one of the main objectives of the study, which combines local and global features to enhance the recognition rate of the proposed model. It should be noted that the areas that tend to be red have high activation values.

4. Discussion

The primary objective of this study was to distinguish medicinal plants from poisonous ones by utilizing a DL-based attention model. Baseline DL algorithms are faced with obstacles like high data requirements and parameter optimization. In some cases, the pooling layers used in these algorithms may remove important information related to images. Most importantly, extracting important features in images is one of the crucial factors in increasing model performance. One effective method involves performing pre-processing operations to extract features from the region of interest. On the other hand, pre-processing operations on images may lead to a complicated DL model. For instance, Hasan et al. (2023) proposed an image patch-based DL technique to classify weed. Choosing the appropriate number and type of patches from plant images can significantly impact the model's accuracy and complexity. Hence, an attempt has been made to develop an attention mechanism approach. The main objective was to detect and extract key features and local information from plant images, without the requirement for preliminary pre-processing procedures. This approach reduces training time and optimizes the model's complexity.

To demonstrate the superiority of our attention module, we performed a comparative analysis with several models with similar structures, namely ResNet, ResNeXt, and ResNeSt. Our evaluation of these models considered various performance metrics such as accuracy, precision, recall, specificity, and F1-score. The results show the effectiveness of our attention module in improving the classification performance of medicinal and poisonous plants. According to the data presented in Table 3, the integration of both CA and SA modules greatly enhanced the model's ability to accurately identify plants. Meanwhile, the ResNeSt101 + SCAM model is slightly superior to the ResNeSt50 + SCAM model with the lowest number of parameters and showed the best performance among the other models with accuracy, precision, recall, specificity, and F1-score of 99.56%, 99.53%, 99.76%, 99.54%, and 99.64%, respectively. Also, the ResNeSt50-SCAM model had a satisfactory performance with an accuracy rate of 99.12%.

The collected dataset was trained and tested with several state-of-the-art models such as VGG-16, VGG-19, GoogleNet, AlexNet, Inception-v3, and Squeeze Net to further evaluate and compare the proposed model (Table 4). As can be seen in Table 4, the proposed ResNeSt101 + CSAM model is superior to other DL models with 99.56% accuracy and 21.30 M parameters. Nevertheless, lighter models with fewer parameters showed poorer performance compared to the proposed model. For example, the Squeeze Net and GoogLeNet models

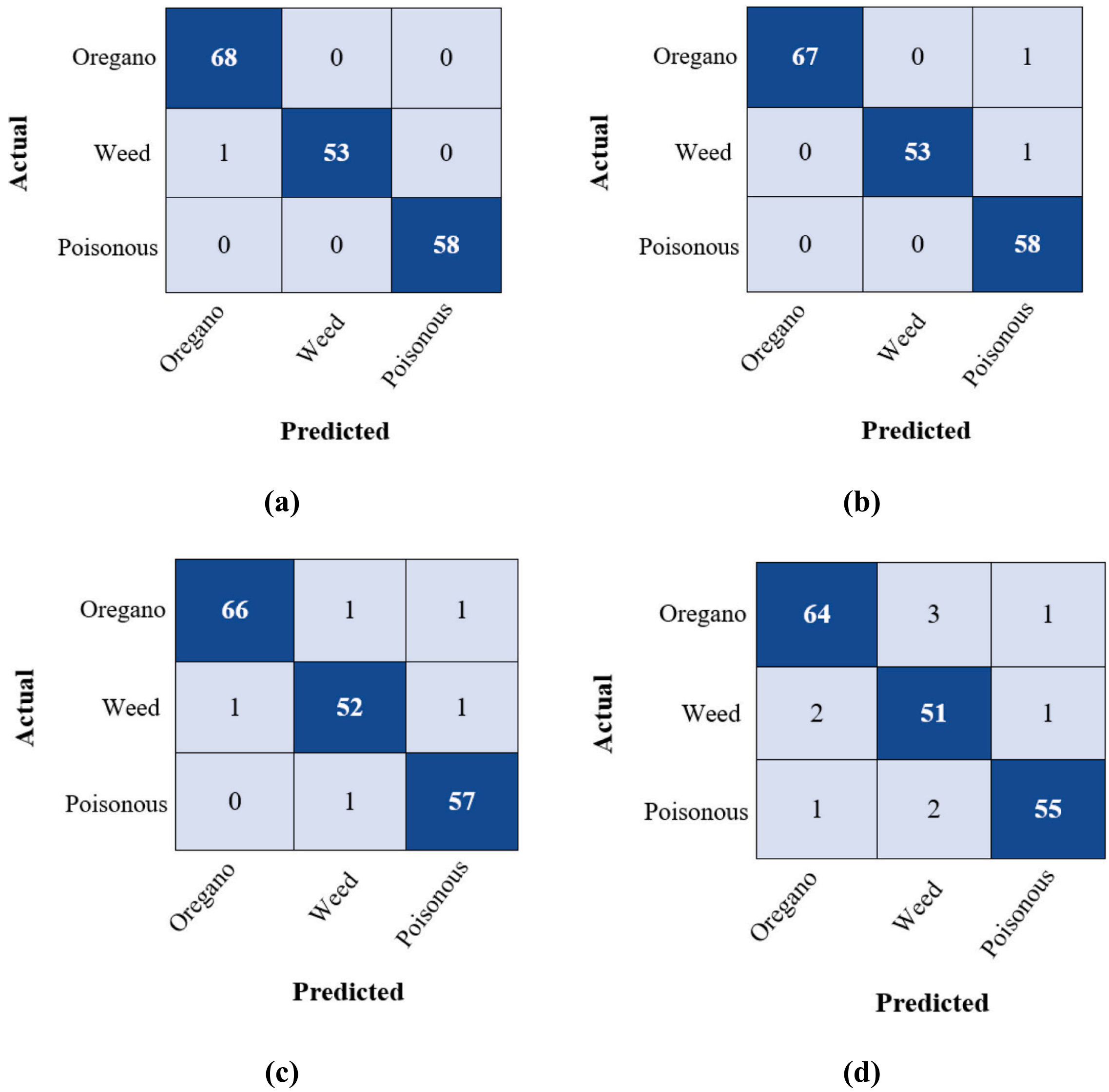

Fig. 9. Confusion matrix for the prediction of three classes of plants by the proposed SCAM-Herb model based on (a) Tree-CA, (b) Gated-CA, (c) Mixed-CA, and (d) GAP-CA modules.

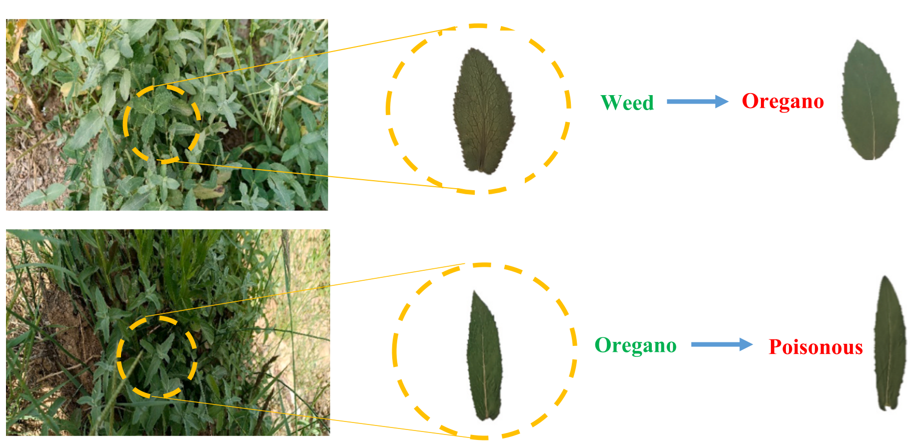

Fig. 10. Illustration of confused plant images by proposed SCAM-Herb model.

achieved 96.07% and 96.32% accuracy, which their performance is weaker than the proposed model. In addition, deeper models such as VGG-16 and VGG-19 have shown good performance with an accuracy of over 98%, but their number of parameters is large and therefore they are not recommended for online detection systems. One of the main and important priorities in the development of an intelligent plant

Table 2
Comparison of classification performance of proposed CSAM-Herb models using evaluation criteria.

SCAM-Herb models	Evaluation criteria	Classes			
		Medicinal Plant	Weed plant	Poisonous plant	Average per class
Tree-CA	Accuracy	99.44	99.44	100	**99.63**
	Precision	100	98.15	100	99.38
	Recall	98.55	100	100	99.52
	Specificity	100	99.21	100	99.74
	F1-score	99.27	99.06	100	99.42
Gated-CA	Accuracy	98.89	99.44	98.89	**99.07**
	Precision	98.53	98.15	96.67	97.78
	Recall	100	100	100	100
	Specificity	99.11	99.21	98.36	98.89
	F1-score	99.26	99.06	98.31	98.87
Mixed-CA	Accuracy	98.31	97.76	98.31	**98.13**
	Precision	95.65	96.29	98.28	96.74
	Recall	98.50	96.29	96.61	97.13
	Specificity	98.20	98.14	99.16	98.50
	F1-score	97.05	96.29	97.43	96.92
GAP-CA	Accuracy	96.04	95.51	97.14	**96.23**
	Precision	94.11	94.44	94.82	94.45
	Recall	95.52	91.07	96.49	94.36
	Specificity	96.63	97.54	97.46	97.21
	F1-score	94.81	92.98	95.64	94.47

identification system is the implementation of a lightweight model. The performance of our proposed model (SCAM-Herb) showed that in addition to its unique structure, it is a lightweight model that has high accuracy in identifying medicinal and poisonous plants.

The performance of DL models in plant classification and identification is influenced by various factors, including symptom representation, covariate shift, image background, disorders with similar symptoms, and concurrent disorders. These factors reduce the accuracy of DL models to a large extent (Barbedo, 2018). The challenges posed by similar plants are among the most problematic of these disorders, and they severely limit the performance of DL algorithms. For instance, Dourado-Filho and Calumby (2021) argued that it was difficult to achieve better accuracy in classifying similar plants using a primary CNN model. They enhanced the model's ability by replacing the softmax layer with a support vector machine classifier.

It can be difficult even for botanists to distinguish between medicinal and poisonous plants due to their similar appearance. To address these issues, we proposed an attention network that combined from two SA and CA modules. The ResNeSt network served as the backbone of the proposed model and was complemented by a SA module that helps in identifying key features and focuses on object location in images without image pre-processing operations. Additionally, the integration of the CA module makes it possible to extract global information from feature maps, allowing images of similar plants to be distinguished with high precision.

One of the significant points in this study compared to other hybrid attention mechanisms such as CAMB (Woo et al., 2018) and DA-Net (Fu et al., 2019) is that fine-grained structures are not utilized in split attention modules to capture subtle variations across various image categories. As well, the challenge related to prioritization in the arrangement of attention models has also been solved. Moreover, the ResNeSt structure is utilized to address the issues of gradient vanishing or explosion in training deep CNN models. This model proves to be more efficient than traditional CNN models in performing a diverse range of

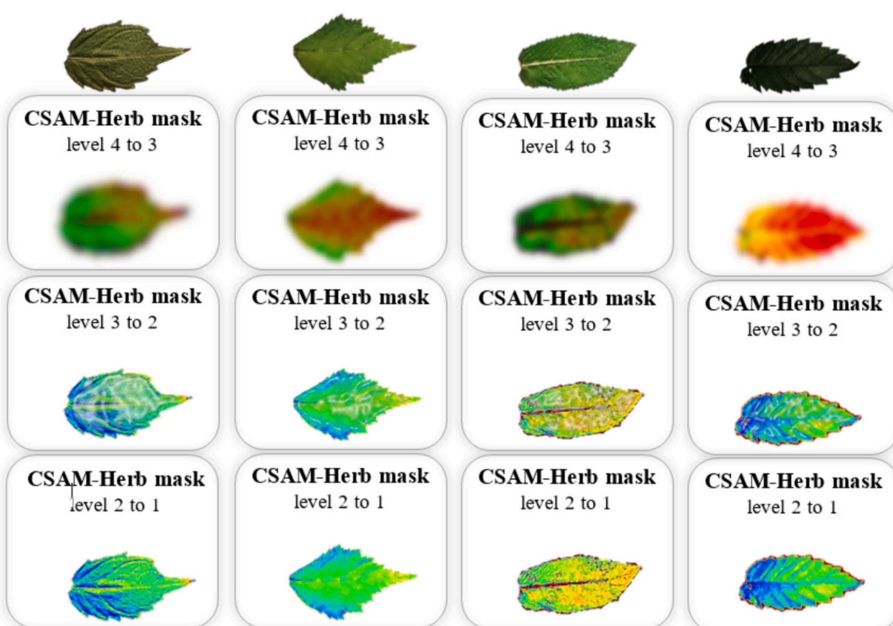

Fig. 11. A visual representation of attention activation.

Table 3
Comparing the performance of the proposed model with different residual structures.

Models	Parameter (M)	Accuracy	Precision	Recall	Specificity	F1-score
ResNet50	25.5	96.55	96.41	97.09	96.47	96.75
ResNet101	44.5	97.91	97.65	97.31	97.69	97.48
ResNeXt50	25	97.73	97.82	97.66	97.57	97.74
ResNeXt101	44.5	98.77	98.87	98.74	98.93	98.81
ResNeSt50 + SCAM	25.6	99.12	99.27	99.38	99.31	99.33
ResNeSt101 + SCAM	21.3	99.56	99.53	99.76	99.54	99.64

Table 4
Comparing the performance of the proposed model with other state-of-the-art models.

Models	Parameter (M)	Accuracy	Precision	Recall	Specificity	F1-score
VGG-16	129.1	98.42	97.09	97.62	98.13	96.36
VGG-19	140	98.59	98.33	97.78	98.47	98.05
GoogleNet	7	96.32	96.55	96.43	97.19	96.49
AlexNet	60	97.25	97.14	96.29	96.89	96.71
Inception-v3	22.4	98.5	97.81	96.7	97.03	97.25
Squeeze Net	4	96.07	96.37	96.61	96.53	96.48
ResNeSt101 + CSAM	21.3	**99.56**	99.53	99.76	99.54	99.64

tasks. Compared to a regular network block, the ResNeSt block includes an additional pathway between the input and output, allowing the network to simply learn the residual of multi-level resolution features (Zhang et al., 2022b).

The results showed that our proposed model improves the automatic identification of medicinal, poisonous and weed plants and can play an effective role in their grading. This approach can be very beneficial in various fields such as precision agriculture, smart agriculture, providing healthy products to the market and maintaining consumer health. In similar studies, researchers have reported that effective identification of plant species can be used to obtain plant information for use in iNaturalist or contributing to the conservation and study of plant biodiversity (Barhate et al., 2023).

One of the limitations of this study could be the presence of occluded leaves. In this study, the presence of partially or completely covered leaves is not considered, which can affect the accuracy of plant species classification. To address this limitation, future directions could involve developing new algorithms that can account for occlusion, such as using depth information or multi-view images to reconstruct the complete shape of the leaves (Barhate et al., 2023).

5. Conclusion

The incorrect identification of medicinal, poisonous and weedy plants not only causes irreversible damage to humans and livestock, but also has a negative impact on the economy. The combination of artificial intelligence and machine vision techniques has the potential to greatly aid humans in this context. Using conventional machine learning methods has always been associated with challenges. This research introduces a novel approach for distinguishing between medicinal, poisonous, and weed plants. By combining SA and CA modules, successful classification of plant samples was achieved with an accuracy of over 99%. The SCAM-Herb model was implemented and tested based on four CA modules, including tree-CA, gated-CA, mixed-CA, and GAP-CA. After evaluating, the tree-CA module-based model was chosen to effectively remove unwanted and noisy information, identify key features, and extract valuable textural and global information. In order to evaluate the proposed model in more details, the results were compared with other state-of-the-art models. The experimental results indicated that the proposed model (ResNeSt101 + SCAM) outperformed all other predictive models, with accuracy, precision, recall, specificity, and F1-score values of 99.56%, 99.53%, 99.76%, 99.54%, and 99.64%, respectively. In our future study, we aim to expand the dataset of plants to include more varieties. Also, a practical smartphone-based app will be developed to identify various plant species.

CRediT authorship contribution statement

Rahim Azadnia: Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization, Writing – original draft. **Faramarz Noei-Khodabadi:** Software, Methodology, Data curation, Writing – original draft. **Azad Moloudzadeh:** Writing – original draft, Resources, Methodology, Funding acquisition, Data curation. **Ahmad Jahanbakhshi:** Conceptualization, Investigation, Methodology, Validation, Writing – review & editing. **Mahmoud Omid:** Writing – review & editing, Supervision, Software, Methodology, Formal analysis, Conceptualization.

Declaration of competing interest

None.

Data availability

Data will be made available on request.

References

Aakif, A., Khan, M.F., 2015. Automatic classification of plants based on their leaves. Biosyst. Eng. 139, 66–75.
Ahmad, N., Asif, H.M.S., Saleem, G., Younus, M.U., Anwar, S., Anjum, M.R., 2021. Leaf image-based plant disease identification using color and texture features. Wirel. Pers. Commun. 121 (2), 1139–1168.
Ali, A.H., Youssef, A., Raja, M.A., Abdelal, M., 2024. An ensemble of deep learning architectures for accurate plant disease classification. Eco. Inform. 102618.
Ansari, M., Kargar, S., Eslami, M.A., Falahati, R., Albenzio, M., Caroprese, M., Kanani, M., 2022. Potential benefits of early-life supplementation of liquid feed with fennel (Foeniculum vulgare) seeds or oregano (Origanum vulgare) leaves on growth, health, and blood metabolites in Holstein dairy calves. J. Dairy Sci. 105 (8), 6639–6653.
Azadnia, R., Kheiralipour, K., 2021. Recognition of leaves of different medicinal plant species using a robust image processing algorithm and artificial neural networks classifier. J. Appl. Res. Med. Aromat. Plants 25, 100327.
Azadnia, R., Kheiralipour, K., 2022. Evaluation of hawthorns maturity level by developing an automated machine learning-based algorithm. Eco. Inform. 71, 101804.
Azadnia, R., Al-Amidi, M.M., Mohammadi, H., Cifci, M.A., Daryab, A., Cavallo, E., 2022a. An AI based approach for medicinal plant identification using deep CNN based on global average pooling. Agronomy 12 (11), 2723.
Azadnia, R., Fouladi, S., Jahanbakhshi, A., 2023. Intelligent detection and waste control of hawthorn fruit based on ripening level using machine vision system and deep learning techniques. Results Eng. 17, 100891.
Azadnia, R., Jahanbakhshi, A., Rashidi, S., Bazyar, P., 2022b. Developing an automated monitoring system for fast and accurate prediction of soil texture using an image-based deep learning network and machine vision system. Measurement 190, 110669.
Azizi, H., Asli-Ardeh, E.A., Jahanbakhshi, A., Momeny, M., 2024. Vision-based strawberry classification using generalized and robust deep networks. J. Agricult. Food Res. 15, 100931.
Barbedo, J.G., 2018. Factors influencing the use of deep learning for plant disease recognition. Biosyst. Eng. 172, 84–91.
Barhate, D., Pathak, S., Dubey, A.K., 2023. Hyperparameter-tuned batch-updated stochastic gradient descent: plant species identification by using hybrid deep learning. Eco. Inform. 75, 102094.
Bodhwani, V., Acharjya, D.P., Bodhwani, U., 2019. Deep residual networks for plant identification. Proc. Comp. Sci. 152, 186–194.
Bora, L., Avram, S., Pavel, I.Z., Muntean, D., Liga, S., Buda, V., Danciu, C., 2022. An up-to-date review regarding cutaneous benefits of Origanum vulgare L. essential oil. Antibiotics 11 (5), 549.
Brauwers, G., Frasincar, F., 2021. A general survey on attention mechanisms in deep learning. IEEE Trans. Knowl. Data Eng. 35 (4), 3279–3298.
Brindha, G.J., Gopi, E.S., 2022. Masking technique based attention mechanism for off-type identification in plants. Mach. Learn. Appl. 8, 100282.
Chen, H.C., Widodo, A.M., Wisnujati, A., Rahaman, M., Lin, J.C.W., Chen, L., Weng, C.E., 2022. AlexNet convolutional neural network for disease detection and classification of tomato leaf. Electronics 11 (6), 951.
Chukwuma, E.C., Soladoye, M.O., Feyisola, R.T., 2015. Traditional medicine and the future of medicinal Plants in Nigeria. J. Med. Plants Stud. 3 (4), 23–29.
Dourado-Filho, L.A., Calumby, R.T., 2021. An experimental assessment of deep convolutional features for plant species recognition. Eco. Inform. 65, 101411.

Faryna, K., van der Laak, J., Litjens, G., 2024. Automatic data augmentation to improve generalization of deep learning in H&E stained histopathology. Comput. Biol. Med. 170, 108018.
Fu, J., Liu, J., Tian, H., Li, Y., Bao, Y., Fang, Z., Lu, H., 2019. Dual attention network for scene segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3146–3154.
Ghasab, M.A.J., Khamis, S., Mohammad, F., Fariman, H.J., 2015. Feature decision-making ant colony optimization system for an automated recognition of plant species. Expert Syst. Appl. 42 (5), 2361–2370.
Hadipour-Rokni, R., Asli-Ardeh, E.A., Jahanbakhshi, A., Sabzi, S., 2023. Intelligent detection of citrus fruit pests using machine vision system and convolutional neural network through transfer learning technique. Comput. Biol. Med. 155, 106611.
Hasan, A.M., Diepeveen, D., Laga, H., Jones, M.G., Sohel, F., 2023. Image patch-based deep learning approach for crop and weed recognition. Eco. Inform. 78, 102361.
Hassanin, M., Anwar, S., Radwan, I., Khan, F.S., Mian, A., 2024. Visual attention methods in deep learning: an in-depth survey. Inform. Fusion 108, 102417.
He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.
Hsia, C.H., Chang, T.H.W., Chiang, C.Y., Chan, H.T., 2022. Mask R-CNN with new data augmentation features for smart detection of retail products. Appl. Sci. 12 (6), 2902.
Islam, S., Ahmed, M.R., Islam, S., Rishad, M.M.A., Ahmed, S., Utshow, T.R., Siam, M.I., 2023. BDMediLeaves: a leaf images dataset for Bangladeshi medicinal plants identification. Data Brief 50, 109488.
Jahanbakhshi, A., Abbaspour-Gilandeh, Y., Heidarbeigi, K., Momeny, M., 2021a. Detection of fraud in ginger powder using an automatic sorting system based on image processing technique and deep learning. Comput. Biol. Med. 136, 104764.
Jahanbakhshi, A., Abbaspour-Gilandeh, Y., Heidarbeigi, K., Momeny, M., 2021b. A novel method based on machine vision system and deep learning to detect fraud in turmeric powder. Comput. Biol. Med. 136, 104728.
Jahanbakhshi, A., Momeny, M., Mahmoudi, M., Radeva, P., 2021c. Waste management using an automatic sorting system for carrot fruit based on image processing technique and improved deep neural networks. Energy Rep. 7, 5248–5256.
Joshi, R.C., Kaushik, M., Dutta, M.K., Srivastava, A., Choudhary, N., 2021. VirLeafNet: automatic analysis and viral disease diagnosis using deep-learning in Vigna mungo plant. Eco. Inform. 61, 101197.
Kintzios, S.E. (Ed.), 2002. Oregano: The Genera Origanum and Lippia. CRC press.
Larese, M.G., Namías, R., Craviotto, R.M., Arango, M.R., Gallo, C., Granitto, P.M., 2014. Automatic classification of legumes using leaf vein image features. Pattern Recogn. 47 (1), 158–168.
Lee, C.Y., Gallagher, P.W., Tu, Z., 2016. Generalizing pooling functions in convolutional neural networks: Mixed, gated, and tree. In: Artificial Intelligence and Statistics. PMLR, pp. 464–472.
Lin, M., Chen, Q., Yan, S., 2013. Network in Network. arXiv preprint. *arXiv:1312.4400*.
Lu, S., Chen, W., Zhang, X., Karkee, M., 2022. Canopy-attention-YOLOv4-based immature/mature apple fruit detection on dense-foliage tree architectures for early crop load estimation. Comput. Electron. Agric. 193, 106696.
Marhamati, M., Dorri, B., Imannezhad, S., Hussain, M.A., Neshat, A.A., Kakmishi, A., Momeny, M., 2024. Patient's airway monitoring during cardiopulmonary resuscitation using deep networks. Med. Eng. Phys. 104179.
Momeny, M., Latif, A.M., Sarram, M.A., Sheikhpour, R., Zhang, Y.D., 2021. A noise robust convolutional neural network for image classification. Res. Eng. Des. 10, 100225.
Momeny, M., Jahanbakhshi, A., Neshat, A.A., Hadipour-Rokni, R., Zhang, Y.D., Ampatzidis, Y., 2022. Detection of citrus black spot disease and ripeness level in orange fruit using learning-to-augment incorporated deep networks. Eco. Inform. 71, 101829.
Momeny, M., Neshat, A.A., Jahanbakhshi, A., Mahmoudi, M., Ampatzidis, Y., Radeva, P., 2023. Grading and fraud detection of saffron via learning-to-augment incorporated inception-v4 CNN. Food Control 147, 109554.
Naresh, Y.G., Nagendraswamy, H.S., 2016. Classification of medicinal plants: an approach using modified LBP with symbolic representation. Neurocomputing 173, 1789–1797.
Panter, K.E., Welch, K.D., Gardner, D.R., 2019. Poisonous plants: Biomarkers for diagnosis. In: Biomarkers in Toxicology. Academic Press, pp. 627–652.
Pezzani, R., Vitalini, S., Iriti, M., 2017. Bioactivities of Origanum vulgare L.: an update. Phytochem. Rev. 16, 1253–1268.
Pushpa, B.R., Rani, N.S., Chandrajith, M., Manohar, N., Nair, S.S.K., 2024. On the importance of integrating convolution features for Indian medicinal plant species classification using hierarchical machine learning approach. Eco. Inform. 81, 102611.
Roopashree, S., Anitha, J., Mahesh, T.R., Kumar, V.V., Viriyasitavat, W., Kaur, A., 2022. An IoT based authentication system for therapeutic herbs measured by local descriptors using machine learning approach. Measurement 200, 111484.
Saleem, G., Akhtar, M., Ahmed, N., Qureshi, W.S., 2019. Automated analysis of visual leaf shape features for plant classification. Comput. Electron. Agric. 157, 270–280.
Sethy, P.K., Barpanda, N.K., Rath, A.K., Behera, S.K., 2020. Deep feature based rice leaf disease identification using support vector machine. Comput. Electron. Agric. 175, 105527.
She, J., Zhan, W., Hong, S., Min, C., Dong, T., Huang, H., He, Z., 2022. A method for automatic real-time detection and counting of fruit fly pests in orchards by trap bottles via convolutional neural network with attention mechanism added. Ecol. Inform. 70, 101690.
Sood, M., Singh, P.K., 2020. Hybrid system for detection and classification of plant disease using qualitative texture features analysis. Proc. Comp. Sci. 167, 1056–1065.
Tamilselvan, N., Thirumalai, T., Shyamala, P., David, E., 2014. A review on some poisonous plants and their medicinal values. J. Acute Dis. 3 (2), 85–89.
Tiwari, V., Joshi, R.C., Dutta, M.K., 2022. Deep neural network for multi-class classification of medicinal plant leaves. Expert. Syst. 39 (8), e13041.
Turgut, K., Dutagaci, H., Rousseau, D., 2022. RoseSegNet: An attention-based deep learning architecture for organ segmentation of plants. Biosyst. Eng. 221, 138–153.
Wang, Q., Cheng, M., Huang, S., Cai, Z., Zhang, J., Yuan, H., 2022. A deep learning approach incorporating YOLO v5 and attention mechanisms for field real-time detection of the invasive weed Solanum rostratum Dunal seedlings. Comput. Electron. Agric. 199, 107194.
Wang, P., Liu, J., Xu, L., Huang, P., Luo, X., Hu, Y., Kang, Z., 2021. Classification of Amanita species based on bilinear networks with attention mechanism. Agriculture 11 (5), 393.
Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., Hu, Q., 2020. ECA-net: efficient channel attention for deep convolutional neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11534–11542.
Wani, M.A., Bhat, F.A., Afzal, S., Khan, A.I., 2020. Advances in Deep Learning. Springer.
Woo, S., Park, J., Lee, J.Y., Kweon, I.S., 2018. Cbam: convolutional block attention module. In: Proceedings of the European conference on computer vision (ECCV), pp. 3–19.
Xie, S., Girshick, R., Dollár, P., Tu, Z., He, K., 2017. Aggregated residual transformations for deep neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1492–1500.
Xue, G., Liu, S., Ma, Y., 2020. A hybrid deep learning-based fruit classification using attention model and convolution autoencoder. Complex Intell. Systems 1–11.
Yang, C., 2021. Plant leaf recognition by integrating shape and texture features. Pattern Recogn. 112, 107809.
Yang, R., Hu, Y., Yao, Y., Gao, M., Liu, R., 2022. Fruit target detection based on BCo-YOLOv5 model. Mob. Inf. Syst. 2022 (1), 8457173.
Zagoruyko, S., Komodakis, N., 2016. Paying more attention to attention: improving the performance of convolutional neural networks via attention transfer. *arXiv preprint. arXiv:1612.03928*.
Zhang, H., Wang, Z., Guo, Y., Ma, Y., Cao, W., Chen, D., Gao, R., 2022a. Weed detection in peanut fields based on machine vision. Agriculture 12 (10), 1541.
Zhang, H., Wu, C., Zhang, Z., Zhu, Y., Lin, H., Zhang, Z., Smola, A., 2022b. Resnest: Split-attention networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2736–2746.