Leaf Classification for Plant Recognition using EfficientNet Architecture

Yagan Arun
Department of Computer Science and Engineering
St. Joseph's College of Engineering
Chennai, India

Viknesh G S
Department of Computer Science and Engineering
St. Joseph's College of Engineering
Chennai, India

Abstract— **Automatic plant species classification has always been a great challenge. Classical machine learning methods have been used to classify leaves using handcrafted features from the morphology of plant leaves which has given promising results. However, we focus on using non-handcrafted features of plant leaves for classification. So, to achieve it, we utilize a deep learning approach for feature extraction and classification of features. Recently Deep Convolution Neural Networks have shown remarkable results in image classification and object detection-based problems. With the help of the transfer learning approach, we explore and compare a set of pre-trained networks and define the best classifier. That set consists of eleven different pre-trained networks loaded with ImageNet weights: AlexNet, EfficientNet B0 to B7, ResNet50, and Xception. These models are trained on the plant leaf image data set, consisting of leaf images from eleven different unique plant species. It was found that EfficientNet-B5 performed better in classifying leaf images compared to other pre-trained models. Automatic plant species classification could be helpful for food engineers, people related to agriculture, researchers, and ordinary people.**

Keywords— Plant Leaf Recognition, Deep Learning, Transfer Learning, EfficientNet.

I. INTRODUCTION

The existence of plants and trees is crucial for the balance of nature (ecological balance). Plants and trees have medicinal values, provide us food, absorb carbon dioxide to synthesize oxygen for the survival of living beings, control environmental temperature and humidity levels, purify the air by absorbing toxic gases present in the environment, and provide shelter animals. So, it has always been of great interest and essential to study the morphology of plants. It is observed that there are hundreds of thousands of existing plant species that plant scientists often study. Often plant scientists and researchers rely upon their observation and identification skills to classify plants from leaves. So, a computer-aided solution that can automatically recognize plants from leaf images will be helpful. Recognition of plants will be helpful to plant scientists and ordinary people, and people related to agriculture.

Classification of leaves has already been implemented using the morphological feature or handcrafted features such as length and width of the leaf, the perimeter of the leaf, the hull area, the hull perimeter, and colour histograms which are being used by traditional machine learning classifiers such as K-Nearest Neighbors (KNN), Penalized Discriminant Analysis (PDA) and deep learning classifiers like Probabilistic Neural Network (PNN).

Morphological characteristic features of leaves such as shape [1,8,12-14], texture [1,8] and venation pattern [9] are often utilized for the classification of the leaves. Apart from handcrafted features of leaf, non-handcrafted features are being used as well. A deep learning classifier [2,4,6] learns the required features of a leaf to classify. Hence, it enables automatic extraction and classification of features, thus avoiding experts encoding features from the morphology of plants.

Advancement in central processing units (CPUs) and graphic processing units (GPUs) has made it possible to utilize modern high-performance methods to process raw data, which led to deep learning. So, to extract these non-handcrafted features, deep neural networks can be utilized. The earth is enriched with different types of valuable plants that play a significant role in research development. To automatically classify plant species using leaves, we utilize non-handcrafted features. Deep networks are efficient in learning complex patterns and features, so in this paper, we present the comparison of performances between selected pre-trained models and choose the best fit model for plant classification.

II. REVIEW OF LITERATURE

This section discusses methods used for classifying plant leaves using features extracted from the morphology of leaves (handcrafted features) and non-handcrafted features.

A. Features extracted from morphology of plants

Uluturk and Ugur (2012) [13] used seven base shape features such as eccentricity, form factor, minor axis length ratio of major axis length, convex hull length ratio of perimeter, extent, rectangularity, and perimeter ratio of major axis length, and three additional features which were extracted from two half regions after bisection of the leaf. Yang and Wei (2019) [15] introduced a new method for classifying plant leaves, and they used two matrices, a sign matrix and a triangle center distance matrix. Sign matrix is used to characterize the convex/concave property of the leaf shape, while the triangle center distance matrix represents the bending degree and spatial information of the leaf. Trishen et al. (2015) [12] used various shape features such as length and width of the leaf, area of the leaf, perimeter of the leaf, the hull area, the hull perimeter, colour histogram, and centroid base radial distance map for classification of leaves. The texture is also an essential component in the field of study of leaf classification. The leaves are being classified based on the distribution of pixels in leaf images. Beghin et al. [1] used the Sobel filter to capture dissimilarities of the macro-texture (pattern formed by the

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 11, November-2021

venation) of the leaves. Mallah et al. [8] used a rotationally invariant version of the Gabor filter to classify leaves texture features. Venation pattern is also a necessary feature used for the classification of plant leaves as they define the shape and structure of the leaf. Mónica et al. [9] segmented the veins from leaf images of three different legume species using unconstrained hit-or-miss transform (UHMT), which is an extension of hit-or-miss transform (HMT) and measured vein and areole features.

B. Non-handcrafted features

Beikmohammadi and Faez (2018) [2] presented leaf classification for plant recognition using transfer learning methodology. A combination of the pre-trained neural network and a regressor was being used to obtain high accuracy on Flavia and LeafSnap datasets. MobileNet is being used as a neural network along with logistic regression. Habiba et al. (2019) [4] proposed a system for recognising Bangladeshi plant leaves using deep learning. They prepared a dataset that had leaf images of 8 different types. These eight leaf types were trained on VGG16, VGG19, Resnet50, InceptionV3, Inception-Resnetv2, and Xception. Accuracy, precision and recall were calculated, and it was found that VGG16 had a better recognition rate than others. Jeon and Ree (2017) [6] presented a system for plant leaf recognition using a convolution neural network. GoogLeNet transfer learning model was used for leaf recognition to get an accuracy of 94% with 30% of leaf damage. Two different models were created by adjusting the network depth and were trained on the Flavia dataset with eight different leaf types. The inception module in GoogLeNet was used for the feature extraction.

III. RESEARCH METHODOLOGY

The proposed methodology for plant leaf classification aims to analyse and compare the performances of AlexNet, ResNet-50, EfficientNet, and Xception pre-trained networks. All the layers are frozen except the last few layers (top three to five layers) for retaining the major of weights learned during training on the ImageNet dataset. These layers of each pre-trained network will be mapped to a dense layer and then to a softmax layer which will consist of 11 neurons.

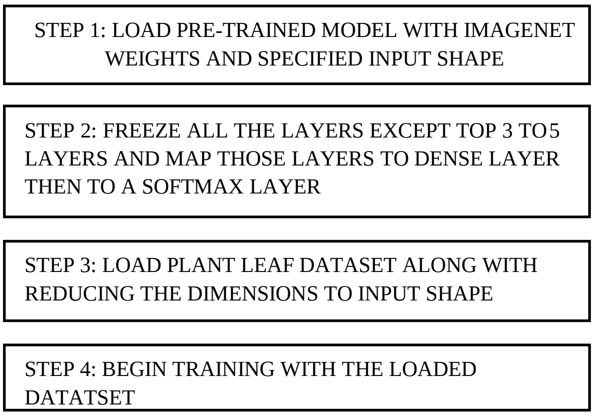

Fig. 1. Steps involved in training any pre-trained network using keras neural net work library

Various performance parameters are being calculated after training each network. Precision, recall, and f1-score are the primary metrics that are being used to compare the models. The network which performs the best is being chosen as the best fit model for plant leaf classification.

IV. TRANSFER LEARNING MODELS

Transfer learning is utilized for the extraction and classification of non-handcrafted features. Deep convolution neural networks consist of convolution layer, pooling layer, and fully connected layers, which are majorly used in image processing and computer vision tasks. In contrast, transfer learning focuses on storing knowledge gained from solving a problem and applying it to a new problem of a similar type. For example, a classifier trained to classify apples can classify oranges, which is the core idea of transfer learning. A wide variety of pre-trained convolution neural networks such as AlexNet, ResNet, EfficientNet, and Xception have been considered to extract and classify features.

A. AlexNet

The architecture of AlexNet proposed by Krizhevsky et al. (2012) [7] consists of eight layers in which five of them are convolutional, three of them are fully-connected layers, and finally, the softmax layer, this architecture consists of 60 million parameters and utilizes ReLU non-linearity function as activation function instead of tanh function. Since AlexNet consists of 60 million parameters, overfitting is resolved through data augmentation and introducing dropout layers. The AlexNet architecture receives an input image size of 227 x 277. The final softmax layer consists of 11 neurons; each value from these neurons represents the ratio of the input image to the corresponding image class.

B. EfficientNet

Efficient Networks proposed by Tan and Le (2019) [11] are based on scaling efficiently depth wise, width wise and resolution wise. The common way to scale CNN is along with their depth. EfficientNet model has obtained a top accuracy of 84.33% with 66M parameters in the ImageNet classification problem. EfficientNet groups eight different models right from B0 to B7. The scaling dimensions that are taken into considerations in EfficientNet are 1. Depth(d), 2. Width(w) and 3. Resolution(r). The networks are being scaled through the compound scaling method. Compound scaling uses a coefficient φ for scaling networks depth, width and resolution uniformly

$$d = \alpha^{\varphi}, w = \beta^{\varphi}, r = \gamma^{\varphi} \qquad (1)$$

$$st.\ \alpha \cdot \beta^2 \cdot \gamma^2 \approx 2,\ \ where\ \alpha \geq 1, \beta \geq 1, \gamma \geq 1$$

From (1) α, β and γ are constants that can be determined through a small grid search and φ is a user-defined coefficient that controls how much more resources are available for scaling the model and α, β and γ specifies how much more of these extra resources should be assigned for scaling the model. FLOPS of a normal convolution operation is $\propto d, w^2, r^2$ i.e. when the model is scaled along depth wise the FLOPS doubles but when the model is scaled along width or resolution wise the FLOPS becomes four times the base network. The operations performed in convolution layers do not change with scaling network architectures, so a baseline network can be defined and can be further scaled through compound scaling. The baseline architecture used is quite similar to M-NASNet. Since eight different networks were tested, each network ranging from B0 to B7 has a different input image size. The last layer of each network is mapped to a softmax layer with 11 neurons.

C. ResNet

ResNet, short for Residual Network proposed by He et al. (2016) [5], is a classical neural network utilized majorly for

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 11, November-2021

computer vision tasks. ResNet-50 is specifically used for the classification of plant leaves. ResNet is based on the concept of skipping connections; it mitigates the vanishing gradient problem and allows the model to learn an identity function that ensures higher layers will perform as well as low layers. ResNet uses ReLU non-linearity as the activation function. Fig.2 shows the residual block or the identity block used in ResNet. Theoretically, the training error should decrease as more layers are added to a neural network, but practically the error increases after a point. However, ResNet does not suffer from this problem; residual blocks used mitigates this issue. The ResNet-50 architecture receives an input image size of 224 x 224 and contains 50 layers, with the last layer being mapped to a softmax layer with 11 neurons.

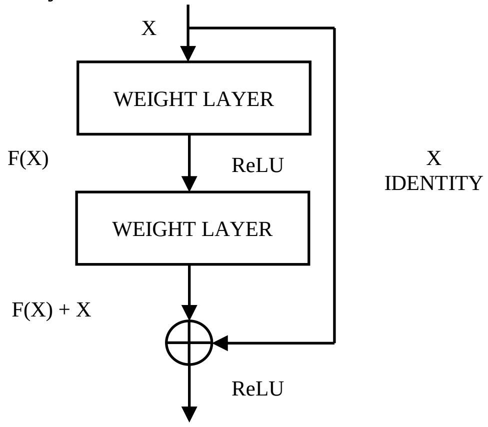

Fig. 2. Residual block in ResNet

D. *Xception*

Xception network proposed Chollet (2017) [3], an extreme version of inception network is a modified depthwise separable convolution. The original depthwise separable convolution has a depthwise convolution and then a pointwise convolution. Depthwise convolution is the channel-wise n×n spatial convolution. Pointwise convolution is the 1x1 convolution to change the dimension. However, the modified version has a pointwise convolution and then a depthwise convolution. Hence the architecture is a linear stack of the modified depthwise separable convolution along with residual connections. The Xception network receives an input image size of 299 x 299 and contains 71 deep layers. The last layer is being mapped to a softmax layer with 11 neurons.

TABLE I. PARAMETERS USED TO TRAIN PRE-TRAINED NETWORKS

Model Name	**Input Size**	**Batch Size**	**Learning Rate**
AlexNet	227 x 277	100	0.001
ResNet-50	224 x 224	100	0.01
Xception	299 x 299	100	0.001
EfficientNet B0	224 x 224	100	0.001
EfficientNet B1	240 x 240	100	0.001
EfficientNet B2	260 x 260	100	0.001
EfficientNet B3	300 x 300	100	0.001
EfficientNet B4	380 x 380	100	0.001
EfficientNet B5	456 x 456	100	0.001
EfficientNet B6	528 x 528	100	0.001
EfficientNet B7	600 x 600	100	0.001

Adam optimizer was used to optimize the unfrozen layers in the pre-trained networks.

$$w_t = w_{t-1} - \eta\left(\frac{\widehat{m_t}}{\sqrt{\widehat{v_t}}+\varepsilon}\right) \qquad (2)$$

$$where\ \widehat{m_t} = \frac{m_t}{1-\beta_1^t}, \widehat{v_t} = \frac{v_t}{1-\beta_2^t}$$

Equation (2) is the mathematical representation of Adam optimizer, where w_t is the weights at the time t, w_{t-1} is the weights at time $t-1$ (previous weights), η is the step size, $\widehat{m_t}$ and $\widehat{v_t}$ are bias corrected weight parameters and ε is a small positive constant. Utilizing Adam optimizer resulted in faster convergence of the gradients with default values of $\beta_1 = 0.9$ and $\beta_2 = 0.999$.

V. EXPERIEMENT AND ANALYSIS

A. *Experimental setup*

All the models in this experiment were trained with GPU support on a Google cloud environment running Debian Linux operating system with Intel(R) Xenon(R) CPU at 2.20Ghz and Nvidia Tesla K80 GPU. The GPU was utilized for faster training of selected pre-trained networks. Keras neural network library with TensorFlow as backend and PyTorch was used to train, test, analyze, and compare the pre-trained networks.

B. *Experimental dataset*

Siddharth et al. (2019) [10] have created a plant leaf dataset with a total of 12 different types of plant leaves, out of which 11 classes were utilized. Those leaves are Alstonia Scholaris, Arjun, Basil, Chinar, Guava, Jamun, Jatropha, Lemon, Mango, Pomegranate, and Pongamia Pinnata. The dataset contains both healthy and unhealthy leaves of each plant type. Only the entire dataset's healthy leaves were extracted and utilized for training the selected pre-trained networks. The data was augmented to increase the number of images. Then the dataset was split into train set, validation set, and test set with a ratio of 80%, 10%, and 10%, respectively. Fig.3 shows 11 different types of plant classes used for training the pre-trained networks.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 11, November-2021

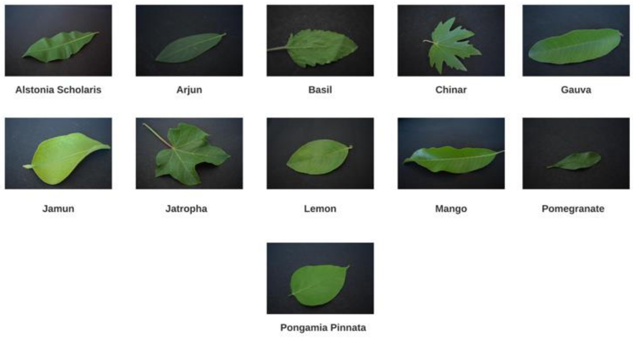

Fig. 3. Dataset of 11 diferrent plant leaf classes

C. *Performance analysis*

Pre-trained networks were analysed and compared using three primary metrics they are precision, recall, and f1-score. These metrics were calculated using true positive (TP), true negative (TN), false positive (FP), and false negative (FN) obtained from the confusion matrix produced by each pre-trained network. Multi-class classification is being performed for this plant classification problem, so a true positive is the total number of correctly classified plant leaf images from each plant class. In contrast, a true negative is the total number of correctly classified plant leaf images from all other plant classes except the relevant plant class. A false positive is the total number of misclassified plant leaf images in all other plant classes except the relevant plant class, while a false negative is the total number of misclassified plant leaf images from the relevant plant class. Precision tells what portion of positives is truly positive, and recall tells what actual positives are correctly classified.

$$precision = \frac{TP}{TP + FP}, \; recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1\ score = 2 \cdot \frac{precision \cdot recall}{precision + recall} \quad (4)$$

TABLE II. PERFORMANCE METRICS OF SELECTED PRE-TRAINED NETWORKS

Model	Precision	Recall	F1 - score	Accuracy
EfficientNet B0	96.27	95.37	95.82	99.24
EfficientNet B1	95.60	94.55	95.07	99.10
EfficientNet B2	96.39	95.48	95.93	99.26
EfficientNet B3	97.51	97.02	97.27	99.50
EfficientNet B4	95.62	94.99	95.30	99.14
EfficientNet B5	98.43	97.96	98.19	99.75
EfficientNet B6	96.29	95.76	96.02	99.27
EfficientNet B7	96.35	96.03	96.19	99.30
ResNet 50	97.78	97.13	97.46	99.53
Xception	98.01	97.68	97.85	99.62
AlexNet	96.52	96.28	96.40	99.27

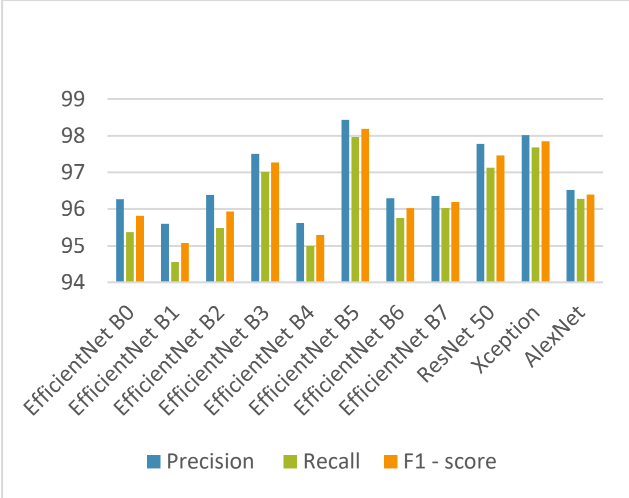

Fig. 4. Comparative analysis of selected pre-trained networks

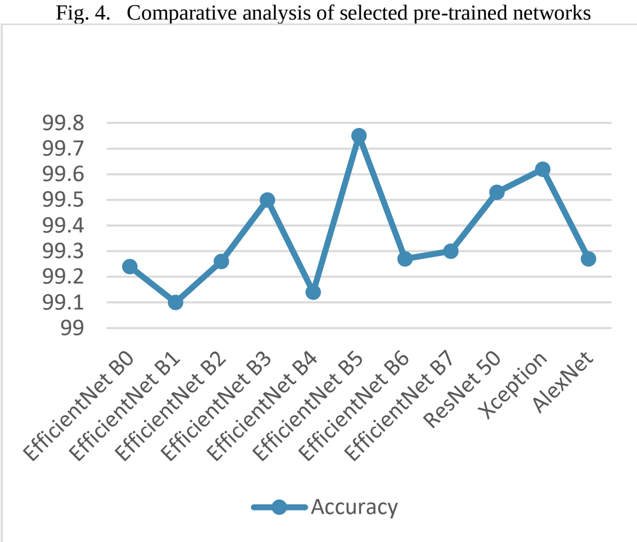

Fig. 5. Accuracy of selected pre-trained networks

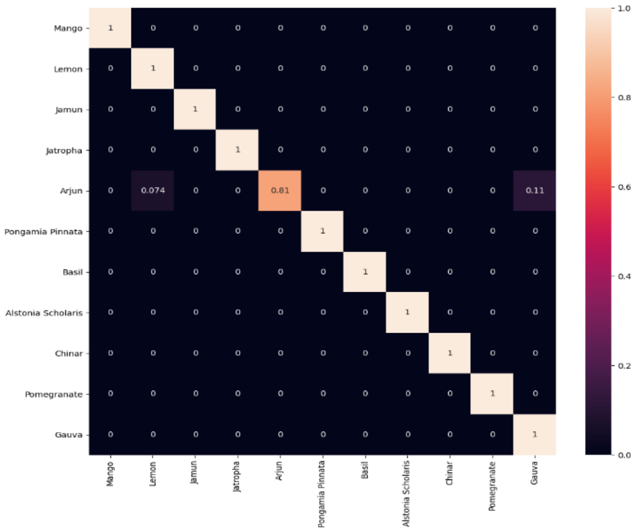

Fig. 6. Confusion matrix of EfficientNet-B5

As seen in Fig.4 and Fig.5, almost all the pre-trained networks gave the results close to each other, but the EfiicientNet B5 network produced the best precision and recall value. Hence the f1 score for EfficientNet B5 is higher than any other pre-trained network considered for this experiment. Also, almost all the pre-trained networks produced accuracy close to

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 11, November-2021

each other. However, EfficientNet B5 has produced the best accuracy compared to the rest of the networks. Fig.6 shows the confusion matrix table used to compare and analyse the performance of the EfficientNet B5 network on a set of test data fabricated from the plant leaf database provided by Siddharth et al. (2019) [10]. Each row of the matrix represents the actual plant class, while each column of the matrix represents the predicted plant class.

VI. ARCHITECTURE OF THE PROPOSED SYSTEM

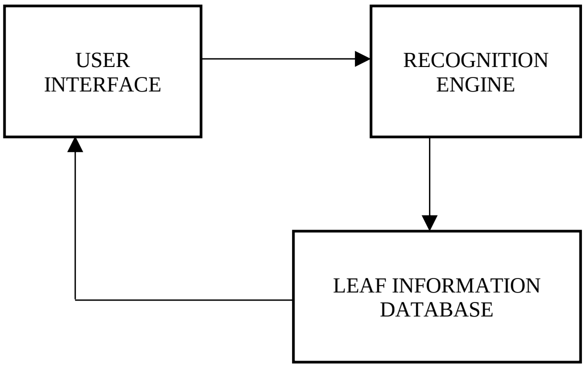

Fig. 7. Simple block digram of the proposed system

The user interface consists of a camera that lets the user capture leaf images. The leaf image captured by the user is then sent to the recognition engine active in a cloud environment that runs EfficientNet-B5 for feature extraction and feature classification. Once the recognition system identifies the leaf, it is then sent to the leaf information database to retrieve the information about the plant. This information is augmented to the user in real-time.

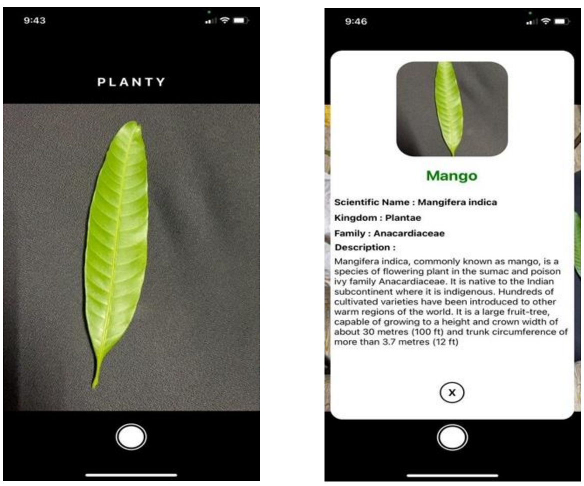

Fig. 8. Results of the plant recognition system.

VII. CONCLUSION

Deep learning methods especially transfer learning, has become so popular for computer vision problems recently. In this experiment, we considered 11 different pre-trained models (AlexNet, EfficientNet B0 to B7, ResNet-50, and Xception), which were analyzed and compared on plant leaf dataset fabricated by Siddharth et al. (2019) [10], which was modified to 11 different plant leaf classes to classify. From analyzing the performances of different networks, it was found that EfficientNet B5 performed better than other selected pre-trained networks; hence from our study for this problem and this dataset, we conclude EfficientNet B5 to be the best fit model for the classification. The main motive of this experiment was to find an alternative to feature extraction from the morphology of plants and training those vectors with classical machine learning algorithms. So, through this method, the use of physical features calculated from the morphology of plants are mitigated, no preprocessing of leaf image is required. The features will be extracted and be classified by the pre-trained networks. A dedicated system has been developed explicitly using the EfficientNet B5 architecture. Therefore, could be helpful to people in agriculture, researchers, and even to ordinary people.

REFERENCES

[1] Beghin, T., Cope, J.S., Remagnino, P., Barman, S., 2010, “Shape and texture based plant leaf classification”, International conference on advanced concepts for intelligent vision systems, pp.345-353.
[2] Beikmohammadi, A., Faez, K., 2018, “Leaf Classification for Plant Recognition with Deep Transfer Learning”, Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS), pp.21-26.
[3] Chollet, F., 2017, “Xception: Deep Learning with Depthwise Separable Convolutions”, Proceedings of the IEEE conference on computer vision and pattern recognition.
[4] Habiba, S.U., Islam, M.K., Ahsan, S.M.M., 2019, “Bangladeshi Plant Recognition using Deep Learning based Leaf Classification”, International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2), pp.1-4.
[5] He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian, 2016, “Deep residual learning for image recognition”, Proceedings of the IEEE conference on computer vision and pattern recognition, pp.770-778.
[6] Jeon, W., Rhee, S., 2017, “Plant Leaf Recognition Using a Convolution Neural Network”, The International Journal of Fuzzy Logic and Intelligent Systems, pp.26-34.
[7] Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012, “ImageNet classification with deep convolutional neural networks”, Advances in neural information processing systems, pp.1097-1105.
[8] Mallah, C., Cope, J., Orwell, J., 2013, “Plant leaf classification using probabilistic integration of shape, texture and margin features”, Signal Processing, Pattern Recognition and Applications.
[9] Mónica, G.L., Rafael, N., Roque, M.C., Miriam, R.A., Carina, G., Pablo, M.G., 2014, “Automatic classification of legumes using leaf vein image features”, Pattern Recognition, pp.158-168.
[10] Siddharth, S.C., Ajay, K., Uday, P.S., Madhav Institute of Technology & Science, 2019, “A Database of Leaf Images: Practice towards Plant Conservation with Plant Pathology”, Mendeley Data.
[11] Tan, M., Le, Q., 2019, “Efficientnet: Rethinking model scaling for convolutional neural networks, International Conference on Machine Learning”, pp.6105-6114.
[12] Trishen, M., Mahess, R., Somveer, K., Sameerchand, P., 2015, “Plant Leaf Recognition Using Shape Features and Colour Histogram with K-nearest Neighbour Classifiers”, Procedia Computer Science, pp.740-747.
[13] Uluturk, K., Ugur, A., 2012, “Recognition of leaves based on morphological features derived from two half-regions”, International Symposium on Innovations in Intelligent Systems and Applications, pp.1-4.
[14] Yang, C., Wei, H., 2019, “Plant Species Recognition Using Triangle-Distance Representation”, IEEE Access, pp.178108-178120.