Journal of Information & Optimization Sciences
ISSN 0252-2667 (Print), ISSN 2169-0103 (Online)
Vol. 44 (2023), No. 4, pp. 717–723
DOI : 10.47974/JIOS-1270

DCGAN-based deep learning approach for medicinal leaf identification

Silky Sachar *
Anuj Kumar †
Department of Computer Science and Applications
Panjab University
Chandigarh
India

Abstract

Though visual identification of plants seems easier for the trained botanists or agriculturists, the automated identification of plants using leaf images still remains a challenging task. The proper identification of plants forms the most important phase as it leads to usage of plants for various purposes. In this paper, we have manually collected about 30 leaves per species belonging to five medicinal plant species. The dataset was created using the scans of the adaxial and abaxial sides of the leaves. As the small number of images makes it difficult for the Convolutional neural network to learn the features, we have augmented the dataset using Deep Convolutional Generative Adversarial Networks (DCGAN). This paper shows that the low-quality images obtained by the scanner could be effectively augmented using the DCGAN thus increasing the variance in the dataset. A comparison of proposed versions of deep learning models namely VGG16, ResNet50 and DenseNet 121 is presented. To validate the results obtained, 5-Fold-Cross validation was used.

Subject Classification: *68T07.*

Keywords: *DCGAN, Deep learning, Medicinal leaves, Convolutional neural networks.*

I. Introduction

The beginning of life on earth marks the association of plants with the human beings. The evolution of plant based medicines have produced the Ayurvedic of the India subcontinent, the Native American and the Amazonian of the North and South Americas respectively and many other

* *E-mail:* **Silky.sachar@gmail.com (Corresponding Author)**
† *E-mail:* **Anuj_gupta@pu.ac.in**

local systems in Asia and Africa[1]. A leaf identification system would not only be used by the trained botanists but also by the common man to make utmost use of the plants growing in their surroundings for their health benefits.

Deep learning finds applications in many other areas like leaf disease detection, emotion recognition, assistance for visually impaired, pneumonia detection [2]–[5]among other. For this paper, we have first manually collected the leaf image dataset and then scanned the adaxial and abaxial sides of the leaves to obtain the digitized images. Our main contribution lies in exploring the application of deep learning in two areas namely image augmentation using DCGAN which not only generates images identical to the input images, but also helps to add variety in the image dataset and development of deep learning models from scratch, based on the architecture of VGG16, ResNet50 and DenseNet121.

The rest of the paper is organized as follows: Section II is composed of the related work in the field. Section III explains the identification process in detail. Section IV concludes the paper.

II. Related Work

The main process of image processing involves acquisition of image data and applying basic pre-processing algorithms to the images so that their features can be highlighted and can be easily learnt.

Instead of using all the input images at a single scale, authors in propose to recognize plants by down-sampling the images into low-resolution images with bilinear interpolation operations. This helps the CNN model to learn discriminative features at different depths. The

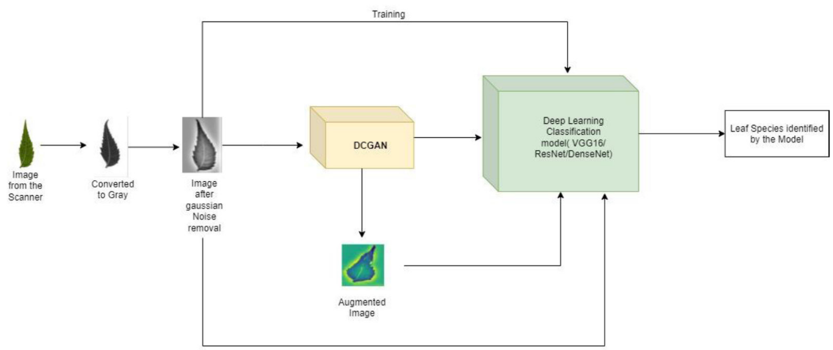

Figure 1

The main process of medicinal leaf identification

learned features were then concatenated from the channel view. Finally, all discriminative features were aggregated and the experiment proved to achieve superior results in freely available MalayaKew and Leaf Snap datasets. Transfer learning is a widely used technique, adopted in the field of image recognition. It allows researchers to use the pre-trained models which have been trained on high-end GPU's. A comparison of deep learning approaches was depicted in and a CNN based model D-Leaf was also proposed. Feature extraction capabilities of AlexNet, fine-tuned AlexNet and D-Leaf were compared. The features obtained were classified using SVM, Artificial Neural Network, k-Nearest Neighbors, Naïve Bayes and CNN and results were compared. The D-Leaf achieved an accuracy of 94.88% which was almost same as AlexNet and fine-tuned AlexNet.

III. Identification Process

i. *Image data augmentation using DCGAN:*

The purpose of the generator model in DCGAN is to generate synthetic images by adding random noise into the original input images so that the discriminator is not able to identify whether the image was the original or generated by the generator. The DCGAN mainly consists of generator network (G) and the discriminator network (D). Network G generates images based on the random noise denoted by z. Network D receives input x representing an image. The output of D(x) represents the probability of the picture being real. At first, the network G generates a very poor-quality image. As there is generation of discriminators, they are able to accurately categorize the generated images against the real images. The network D acts as a binary classifier which outputs 0 for the synthetic image and 1 for the real image. As the training continues, the second-generation generator is able to produce images with slight better quality which appears like the generated images were real compared to the images generated by the first-generation generator. Then second-generation discriminator is trained which accurately identifies images generated by the second-generation generator. As the generations grow to third, fourth fifth and so on the discriminator is not able to accurately distinguish between the synthetic and the real images. The objective function of GAN is as follows:

$$min_G max_D V(D,G) = E_{x \sim P_{data(x)}}[log D(x)] E_{z \sim P_{z(z)}}[\log(1 - D(G(z)))] \quad (1)$$

Where x is the real image, D(x) is the probability of discriminating x as a real image by the network D. G(z) is a sample image generated by

G using the noise z. D(G(z)) represents the probability of discriminating G(z) as a real sample by D. DCGAN allows the GAN, which is a part of unsupervised learning to use the CNN which is part of supervised learning. This makes the DCGAN to produce better quality sample images and to improve the speed of convergence.

In our experiments, we found out that Upsampling proves to provide better outputs than the Convolution transpose layers and is fast as well as effective in learning the predictions. Relu is used as the activation layer. This process when carried out layer to layer is able to generate the leaf features, as the number of channels decrease. Finally, the tanh activation layer is able to generate the image.

The Discriminator model consists of convolution layers followed by Batch Normalization and Leaky Rely layers. This model receives an input image and results in only 1 output "real" or "fake".

ii. *Building Image Identification Network*

Three deep learning models based on the network architecture of VGG16, Resnet50 and Densenet121 were developed from scratch and the performance was validated using 5-Fold Cross validation technique. The base network architectures are discussed below:

VGG16: Developed by Simonyan and Zisserman [7], this CNN model was an improvement of the AlexNet model. The authors focused on creating convolution layers of 3x3 filter size with stride of 1 using same padding and maxpooling layer of filter size 2x2 of stride 2. In the end there are 2 fully connected layers followed by softmax for classification.

ResNet50: This model was developed with a vision to stack more layers to a deep learning network and still overcome the problem of vanishing gradients [8]. While training deep neural networks, the gradients are back-propagated to the previous layers and repetitive multiplication makes the gradient so small that it vanishes soon after which leads to saturation in the performance of the CNN. The base of Residual nets is the formation of skip connections where the original input is connected to the output of the convolution block.

DenseNet 121 [9]: The Dense networks are the deeper networks but are more efficient to train as there is shorter connection between the layers as the first layer is connected to 2^{nd}, 3^{rd}, 4^{th},5^{th} and so on and then the second layer is connected to the 3^{rd}, 4^{th} ,5^{th} and so on. This architecture enables maximum information flow in the network. DenseNet is different from

ResNets as in former the features are combined through concatenation and in the latter the features are combined using summation.

The 3 models mentioned above taken as base models were implemented without their weights and trained from scratch on the augmented leaf dataset which consisted of original scanned images plus the augmented images in each category. In each experiment standard input image size of 224x224 was used. The output from the base model was obtained and flattened. This flattened input was then connected to two dense layers consisting of 256 and 128 neurons followed by "ReLu" activations. The final softmax layer with 5 neurons to represent 5 classes in the dataset was used as the classification layer. Stochastic Gradient Descent optimizer was sued with learning rate of 1e-4 and momentum of 0.9 in all the models. These models also show the robustness of the leaf dataset augmentation technique used. These models were trained in five trials for the same dataset. In each trial the dataset was shuffled and 5-Fold cross validation technique was applied. The results are shown in Table 1.

Table 1

Performance of deep learning models on the augmented image data

Model Name	**Fold**	**Accuracy**	**Mean Accuracy**	**Mean Standard deviation**
VGG16	Fold1	97.71	96.13	1.22
	Fold2	94.95		
	Fold3	96.79		
	Fold5	96.77		
	Fold5	94.47		
ResNet50	Fold1	96.33	96.60	0.544
	Fold2	96.33		
	Fold3	95.87		
	Fold5	97.24		
	Fold5	97.24		
DenseNet	Fold1	97.25	**97.51**	0.47
	Fold2	97.71		
	Fold3	97.71		
	Fold5	96.77		
	Fold5	98.16		

IV. Conclusion

In this paper, we have explored the application of deep learning in the field of medicinal leaf recognition. It is also intended to provide a guideline for the researchers who are new in the field. After carefully acquiring the leaf images from 5 medicinal plant species over a period of 8 months, the images were pre-processed for the augmentation and classification. This paper shows how efficiently deep learning can be used to generate synthetic images from the low-quality input images obtained from the scanner. To mitigate the problem of small dataset as in this case, we used DCGAN to generate images for each class. The images thus obtained then added to original dataset to increase its size three times. These images served as input to three Deep learning models created from the base architectures of VGG16, ResNet and DenseNet121 individually. All three models were trained on the same dataset to compare their classification capabilities while keeping the input image size and optimizer same. The results were evaluated using the 5-Fold cross validation. It can be observed from the experiments that the best mean accuracy of 97.51% was reported across the five folds from the DenseNet architecture model.

References

[1] N. Mamedov, "Medicinal Plants Studies: History, Challenges and Prospective," *Med. Aromat. Plants*, vol. 01, no. 08, 2012, doi: 10.4172/2167-0412.1000e133.

[2] B. M. Joshi and H. Bhavsar, "Plant leaf disease detection and control: A survey," *J. Inf. Optim. Sci.*, vol. 41, no. 2, pp. 475–487, 2020, doi: 10.1080/02522667.2020.1734295.

[3] A. Dureja and P. Pahwa, "Medical image retrieval for detecting pneumonia using binary classification with deep convolutional neural networks," *J. Inf. Optim. Sci.*, vol. 41, no. 6, pp. 1419–1431, 2020, doi: 10.1080/02522667.2020.1809096.

[4] D. Virmani, C. Gupta, P. Bamdev, and P. Jain, "iSeePlus : A cost effective smart assistance archetype based on deep learning model for visually impaired," *J. Inf. Optim. Sci.*, vol. 41, no. 7, pp. 1741–1756, 2020, doi: 10.1080/02522667.2020.1812198.

[5] V. Gupta, S. Juyal, G. P. Singh, C. Killa, and N. Gupta, "Emotion recognition of audio/speech data using deep learning approaches,"

J. Inf. Optim. Sci., vol. 41, no. 6, pp. 1309–1317, 2020, doi: 10.1080/02522667.2020.1809089.

[6] J. Hu, Z. Chen, M. Yang, R. Zhang, and Y. Cui, "A multiscale fusion convolutional neural network for plant leaf recognition," *IEEE Signal Process. Lett.*, vol. 25, no. 6, pp. 853–857, Jun. 2018, doi: 10.1109/LSP.2018.2809688.

[7] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, pp. 1–14, 2015.

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016, vol. 2016-December, doi: 10.1109/CVPR.2016.90.

[9] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 2261–2269, 2017, doi: 10.1109/CVPR.2017.243.

Received September, 2022