Deep Learning to Identify Plant Species

[1]Fayas Rasheed, [2]Liz George,

[1]PG Scholar, [2]Asst.Professor of St. Joseph's College of Engineering and Technology,
[1]Department of Computer Applications,
[1]St. Joseph's College of Engineering and Technology, Palai, India

Abstract: Deep learning is the method that has the ability to develop precise models for recognizing images. Deep learning has shown potential for automating the process of identifying various plant species from images of their leaves, flowers, and fruits. To reduce noise and improve the plant features, the input images undergo pre-processing. The deep learning model is then trained using a sizable labelled dataset of plant images. Once trained, the model can accurately recognize the plant species from new images. While automated plant classification systems typically rely on leaf shape as the main feature for identification, leaves also have other characteristics that can contribute to more precise classification, such as their texture, vein patterns, and color. This technology has the potential to be applied in various fields, such as agriculture, botany, and environmental conservation, to help identify and monitor different plant species in their natural surroundings. However, as with any deep learning application, the quality of the training data and the neural network architecture design are essential factors that can have a significant impact on the system's performance.

Index Terms: Deep learning, Species, Architecture *(key words)*

I. INTRODUCTION (HEADING 1)

Identifying and classifying the vast array of vascular plant species, estimated at 391,000, presents a challenging task for botanists and experts. To address this issue, there is an increasing demand for computerized systems to aid in plant identification and classification. Deep learning has emerged as a promising solution to recognize plant species based on their visual characteristics. Deep learning, which is a subfield of artificial intelligence, is presently a prevalent and widely utilized approach across numerous domains, such as biology, computer vision, medicine, and speech recognition. The advancement of deep learning technology in computer vision has extended its application to plant species identification, with significant implications for biodiversity preservation and natural resource management. Accurate identification of plant species is crucial in plant breeding, conservation, and ecological research. Historically, features were created through manual engineering or crafting. Nevertheless, a recent development in machine learning has shown that learned representations are more efficient and effective. Conventional methods for plant species identification rely on the manual inspection of physical traits like leaves, flowers, and fruits, which can be time-consuming, subjective, and errorprone, especially for large datasets. Deep learning provides an automated and precise alternative to identify plant species by analyzing a large volume of plant images. The topic will cover various deep learning model architectures, such as convolutional neural networks (CNNs), and the techniques used to train and fine-tune these models. It will also address the challenges faced in the application of deep learning to plant species identification, including data labelling, data augmentation, and model architecture selection. The objective is to establish a convolutional neural network (CNN) algorithm that can identify various types of plants and to determine the most effective means to enhance the speed and precision of this system. Deep learning, a contemporary AI methodology, offers a resilient structure for supervised learning that can adeptly map input vectors to output vectors, even in extensive datasets. The deep learning architecture can be categorized into various types, such as Convolutional Neural Network. Unlike conventional machine learning methods, deep learning can extract more detailed information.

II. LITERATURE REVIEW

Deep Learning for Plant Species Classification Using Leaf Vein Morphometric [1]: This study utilized a deep learning approach for the purpose of classifying plant species using data from leaf vein morphometry. To obtain the desired results, leaf images were first pre-processed, and then features were extracted using three distinct CNN models such as D-Leaf, fine-tuned AlexNet and pretrained AlexNet. The features that were obtained were subjected to classification using five different techniques, namely: k-NearestNeighbor (k-NN), CNN, ANN, Support Vector Machine (SVM) and Naïve-Bayes (NB). The testing accuracy of D-Leaf was determined to be 94.88 percent, while the accuracies of AlexNet and fine-tuned AlexNet were 93.26 percent and 95.54 percent, respectively. These findings suggest that D-Leaf may be a successful method for identifying different plant species.

Fine-tuning Deep Convolutional Networks for Plant Recognition [2]: Using a convolutional neural network (CNN), the scientists created a plant recognition system through deep learning techniques. In order to accomplish this, the CNN was pre-trained using a vast dataset with 1.8 million of images, and then a fine-tuning approach was utilized to transfer the acquired recognition abilities to the plant identification task. The CNN architecture included 2 fully-connected and 5 convolutional layers, along with an added prediction layer to derive classification scores. Firstly, a network was trained and fine-tuned using this architecture. Then, after identifying the appropriate validation set parameters, the same system was fine-tuned again using all available images According to the study, fine-tuning was found to be a successful method for transferring recognition skills learned from general domains to the particular task of identifying plants. These findings offer a promising approach for designing automated plant identification systems using deep learning techniques.

Research on Artificial Intelligence: Deep Learning to Identify Plant Species [3]: This research investigates the use of deep learning techniques to identify plant species. The study aims to test whether increasing the learning epochs or extending the learning image amount could improve the ML test speed and recognition accuracy, and determine the best method to improve the convolutional neural network (CNN) in real life. The researchers utilized Pycharm, Tensorflow, Numpy, Keras, and other tools to implement the study. The parameters of the study involved storing the learning set address, 3 convolutional Layer parameters, and epoch numbers. The loading part of the process included setting up an empty image list and storing array information. The transformer part of the process involved converting the image to a matrix array using cv2.In the graphics derive part, two optimal graphics were generated: accuracy and loss. A smaller loss corresponds to a higher accuracy for the models. The researchers then varied the test size and training set ratio to investigate the impact on validation accuracy. They found that increasing the training set size and reducing the test size resulted in higher validation accuracy. For example, with a test size of 0.3 and a training set: test set ratio of 7:3, the validation accuracy was 0.5948.With a test size of 0.5 and a training set: test set ratio of 1:1, the validation accuracy was 0.3015 and with a test size of 0.7 and a training set: test set ratio of 3:7, the validation accuracy was 0.0586.Finally, the study explored the effect of increasing the learning epochs on testing time, finding that as the epoch number increases, the accuracy is directly proportional to epoch times. The results suggest that increasing the learning epoch number and extending the training image set could improve the ML test speed and recognition accuracy. The testing time for 3 epochs was 19 seconds, for 7 epochs it was 18 seconds, and for 15 epochs it was 17 seconds. Based on this observation, it can be concluded that as the number of learning epochs increases, the model accuracy tends to increase, and this increase is directly proportional to the epoch times.

The Analysis of Plants Image Recognition Based on Deep Learning and Artificial Neural Network [4] : The article discusses the use of digital image analysis for plant leaf recognition, and various methods for segmenting and extracting features from leaf images. This study assesses the effectiveness of several classification algorithms in identifying plant species using leaf characteristics. The algorithms compared include KNN-based neighborhood classification and SVM-based support vector machine. The article suggests a novel approach to enhance the precision of leaf identification by utilizing a neural network classification format that relies on the backpropagation error algorithm. The results of the experiment show that this method is effective and has the potential to be applied in various fields, such as identification of plant, improved variety identification, and ecological monitoring of plant. The article also highlights some limitations of the current deep learning algorithms, such as the need for large amounts of data as training samples, difficulty in recognizing leaves in complex backgrounds, and lack of a unified evaluation criterion [7]. According to the article, in order to create a thorough system for identifying plant species, it is necessary to develop a single database of plant leaves and establish a reliable evaluation criterion that relies on image processing analysis, pattern recognition and machine vision. In general, it offers significant perspectives on the present status of plant leaf identification techniques utilizing digital image analysis, and emphasizes the possibilities for further investigations in this area [5].

III. PLANTS SPECIES IDENTIFICATION

Collecting a large and diverse dataset of images of plant species that are to be identified and pre-processing of data is performed. Cleaning and pre-processing the data, which involves removing any irrelevant information and noise from the images, resizing the images to a consistent size, and normalizing the pixel values [6]. The step augmentation is performed by applying various image transformation techniques such as rotations, flipping, and scaling to create new images. The training data preparation step includes dividing the dataset into training, validation, and testing sets. The training set is used to train the deep learning model, while the validation set is used to fine-tune the hyper parameters of the model. The testing set is used to evaluate the performance of the trained model. Selecting an appropriate deep learning architecture for the plant species identification task and training the model on the training dataset. CNNs are commonly utilized for this task due to their effectiveness in image classification tasks. Then evaluating the performance of the trained model on the testing dataset using appropriate performance methods such as accuracy, recall and precision. After the evaluation performed the trained model will deploy to make predictions on new and unseen plant images.

1. FEATURES OF SPECIES IDENTIFICATION

Accuracy: Deep learning models have shown high accuracy in identifying plant species from images, often outperforming traditional manual identification methods. Models can be trained on large datasets of images to learn the distinguishing features of different plant species, and then use this knowledge to make accurate predictions about the species of a given image [10]. This can reduce errors and increase efficiency in plant species identification, making it a promising tool for researchers, conservationists, and anyone working in the field of plant biology.

Efficiency: Deep learning models can process large amounts of data much faster than traditional manual identification methods. Once a deep learning model has been trained on a large set of plant images, it can automatically find the species of a new image in a matter of seconds, reducing the need for manual intervention and speeding up the identification process. This can save a significant amount of time and resources, making plant species identification more efficient and accessible for researchers and conservationists.

Automation: One of the key benefits of using deep learning for plant species identification is its ability to automate the identification process. Once a deep learning model has been trained on a large dataset of plant images, it can be used to automatically identify the species of new images without the need for manual intervention. This can save a significant amount of time and resources, especially when processing large amounts of data. Furthermore, the ability to automate the identification process can make it more accessible to researchers, conservationists, and anyone working in the field of plant biology who may not have specialized expertise in plant identification.

Scalability: Deep learning models can be trained on large datasets of plant images and can be scaled up to accommodate even larger datasets, making it easier to monitor plant populations and biodiversity. This scalability is a significant advantage of deep learning-based plant species identification over traditional manual identification methods, which may become impractical or impossible to use when dealing with large datasets. The ability to scale the identification process can make it more efficient and comprehensive, allowing researchers and conservationists to monitor plant populations and biodiversity at a much larger scale than would otherwise be possible.

Adaptability: Deep learning models are highly adaptable and can be trained on a wide range of plant species and environments, making them versatile and suitable for a variety of applications. With enough training data, deep learning models can learn the distinguishing features of different plant species and identify them accurately, regardless of the environment in which they were photographed. This adaptability is a significant advantage of deep learning-based plant species identification over traditional manual identification methods, which may require specialized expertise in certain plant species or environments. The versatility and adaptability of deep learning models can make plant species identification more accessible and practical for researchers, conservationists, and anyone working in the field of plant biology.

2. CHALLENGES FACED BY SPECIES IDENTIFICATION

While deep learning has emerged as a technique with high potential for plant species identification, there are several challenges associated with this approach [5]. Developing a plant species identification system based on deep learning poses a highly significant challenge, which consist of [1]The availability of large and diverse datasets of images. In some cases, the availability of labeled data for specific plant species can be limited, which can make it challenging to train a deep learning model to identify these species accurately.

Also one of the other challenges are [2]The variations in plant appearance due to factors such as lighting, growth stages, and camera angle can make it difficult for deep learning models to identify plant species accurately. Therefore, developing models that can generalize to these variations is a significant challenge.

Other challenges occurring during the process is the [3] Overfitting of the data, which occurs when a deep learning model is trained too well on the training set of data and becomes too specialized in identifying specific features of that dataset. As a result, the model may not perform well on new, unseen images. Preventing overfitting is critical to ensure that the model can generalize to new data and make accurate predictions.

Training deep learning models requires [4] significant computational resources, including high-performance computing systems and specialized hardware such as graphics processing units (GPUs). These resources can be expensive, limiting the accessibility of deep learning-based plant species identification systems to researchers and conservationists.

Deep learning models are often considered black boxes, making it difficult to understand how they arrive at their predictions. This lack of [5] interpretabilities can be a challenge when trying to understand why a model is making incorrect predictions or when attempting to improve the model's performance.

3. APPLICATIONS OF IDENTIFYING DIFFERENT TYPES OF PLANTS

Identifying different types of plants is an important area of study with numerous practical applications in various fields. Applications of plant species identification consist of:

[1]*Biodiversity Conservation:* Identifying different types of plants is essential for the conservation of plant biodiversity. By identifying different plant species, scientists and conservationists can create species inventories, monitor plant populations, and develop effective conservation strategies.

[2]*Agriculture and Horticulture:* Plant species identification is important in agriculture and horticulture for identifying plant diseases, pests, and weeds. This can help farmers and growers take appropriate measures to control or prevent the spread of diseases and pests, as well as prevent the growth of invasive species.

[3]*Forestry:* Plant species identification is important in forestry for managing and conserving forest ecosystems. This can help foresters to identify and manage invasive species, and monitor the growth and health of different tree species.

[4]*Medicinal Plants:* Plant species identification is important in the field of medicine for identifying the specific plants with medicinal properties. This can help scientists and researchers to develop new drugs and therapies.

[5]*Botany and Plant Biology:* Plant species identification is fundamental to botany and plant biology, enabling the classification, description, and study of different plant species.

[6]*Environmental Science*: Plant species identification is important in environmental science for monitoring environmental changes and assessing the impact of human activities on the environment. By identifying different plant species, scientists can determine the ecological health of a particular ecosystem.

[7]*Food Industry:* Plant species identification is important in the food industry for identifying edible plants and fruits, which can help in food production and product development.

IV. METHODOLOGY

A) *ACNN:* It is a popular method that has shown remarkable success in image recognition tasks The process entails feeding an image through multiple convolutional and pooling layers to acquire knowledge about characteristics and traits that are linked to distinct types of plants. The output from the CNN is then used to classify the image into the appropriate species.

B) *ANN:* These are another type of deep learning model that can be used for plant species identification [8]. They consist of interconnected layers of artificial neurons and are trained using a backpropagation algorithm. ANNs can be used to identify patterns and features in images, and they can be used for both classification and regression tasks.

C) *Naive Bayes*: An algorithm that uses probability and can be utilized for the identification of plant species. It involves calculating the probability of an image belonging to each species based on the features in the image. The species with the highest probability is then chosen as the classification result.

D) *SVM:* It involves finding a hyperplane that separates the different species in the feature space [9]. The SVM algorithm tries to maximize the margin between the hyperplane and the nearest data points of each species.

E) *KNN:* A machine learning algorithm that doesn't rely on specific parameters can be employed to recognize different plant species. It involves finding the k-nearest neighbors of an image in the feature space and then classifying the image based on the majority vote of the neighbors.

V. CONCLUSION

Plant species identification using deep learning is a method that involves training computer algorithms, such as convolutional neural networks, to recognize patterns and features in images of plants. This approach has demonstrated a great potential for accurately identifying plant species based on their images and has several advantages over traditional methods of identification. Deep learning models offer a significant advantage in that they can learn features and patterns in images without manual intervention, thereby minimizing the necessity for time-consuming and potentially ineffective manual feature engineering. Additionally, deep learning models can handle large amounts of data and can identify subtle differences in plant features that may be challenging for humans to detect. However, deep learning models require a considerable amount of training data to perform well, and they can be affected by variations in lighting, background, and image quality. Another limitation is that they may not always provide explanations for their decisions, which can be problematic in certain situations, such as in legal or ethical disputes. In conclusion, the application of deep learning for plant species identification has the potential to revolutionize the field of plant science and agriculture by providing an accurate and efficient method for identifying plant species. Nevertheless, it is crucial for researchers to continue refining this approach and addressing its limitations and challenges to ensure its effectiveness and usefulness in various contexts.

VI. ACKNOWLEDGMENT

First and foremost, praises and thanks to the God, the Almighty, for His showers of blessings throughout my seminar preparation and for giving me the courage and wisdom to take up the seminar and complete it successfully.

I am grateful to Dr. V.P Devassia, Principal, St. Joseph's College of Engineering and Technology, for giving an opportunity to be a part of this prestigious institution. I take this opportunity to express my sincere gratitude to Mr. Anish Augustine, HOD- In charge, Department of MCA, Ms. Liz George, Assistant Professor, St. Joseph's College of Engineering and Technology, Palai, who provided me the atmosphere to complete this seminar report successfully.

REFERENCES

1. W. Dan, Z. Jianghua, and Nurbayi, ''Study on plant leaf identification in arid areas,'' Comput. Eng. Appl., vol. 55, no. 13, pp. 129–133, 2019.
2. M.Rahali, H.Loukil, and M.S.Bouhlel,'' Analysis of image compression approaches using wavelet transform and Kohonen's network,'' in Proc. Int. Workshop Soft Comput. Appl. Cham, Switzerland: Springer, 2016, pp. 474–486
3. Bengio, Y.: Deep learning of representations for unsupervised and transfer learning. Unsupervised and Transfer Learning Challenges in Machine Learning, Volume 7 p. 19 (2012)K. Elissa, "Title of paper if known," unpublished.
4. Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., Darrell, T.: Decaf: A deep convolutional activation feature for generic visual recognition. arXiv preprint arXiv:1310.1531 (2013)
5. E. Aptoula and B. Yanikoglu, ''Morphological features for leaf based plant recognition,'' in Proc. IEEE Int. Conf. Image Process., Sep. 2013, pp. 1496–1499.
6. C. Pornpanomchai, S. Rimdusit, P. Tanasap, and C. Chaiyod, "Thai herb leaf image recognition system (THLIRS)," Kasetsart J. (Nat. Sci.), vol. 45, pp. 551–562, 2011.
7. X. F. Wang, D. S. Huang, J. X. Du, H. Xu, and L. Heutte, "Classification of plant leaf images with complicated background," Appl. Math. Comput., vol. 205, no. 2, pp. 916–926, Nov. 2008, doi: 10.1016/j.amc.2008.05.108.
8. N. Duong-Trung, L. D. Quach, M. H. Nguyen, and C. N. Nguyen, "A combination of transfer learning and Deep Learning for Medicinal Plant Classification", Proceedings of the 2019 4th International Conference on Intelligent Information Technology - ICIIT '19, 2019, doi: 10.1145/3321454.3321464.
9. S. H. Lee, C. S. Chan, P. Wilkin, and P. Remagnino, "Deep-plant: Plant identification with convolutional neural networks," in Proc. IEEE Int. Conf. Image Process., Sep. 2015, pp. 452–456.
10. Saleem, G., Akhtar, M., Ahmed, N., et al.: 'Automated analysis of visual leaf shape features for plant classification', Comput. Electron. Agric., 2019, 157, pp. 270–280