A survey of tree species classification algorithms using pre-trained CNN models

Harsha Aptekar[*1], Ashutosh Chillal[*2], Chinmay Ahire[*3], Yashodhan Agashe[*4],

Prof. Dr. C.A.Ghuge[*5]

[*1,2,3,4]Student, Artificial Intelligence & Machine Learning,

P.E.S Modern College of Engineering, Shivajinagar, Pune-05.

[*5]Professor, Artificial Intelligence & Machine Learning,

P.E.S Modern College of Engineering, Shivajinagar, Pune-05.

Abstract:

A core understanding of the biodiversity that surrounds us involves being able to accurately identify tree species. But having to check each plant manually could be time-intensive and tedious. In recent years, various Convolutional Neural Networks (CNNs) pretrained models such as ResNet50, InceptionV3, VGG16, MobileNetV3 and DenseNet121 have demonstrated their effectiveness across a range of tasks by leveraging large-scale datasets like ImageNet. This paper surveys the progress of pretrained CNN models, focusing on their use cases in previous research, their performance, and applicability to diverse image-based tree classification problems. This paper discusses the strengths and limitations of pretrained CNNs, highlighting their potential for further research and practical applications, to tackle species identification, ecological informatics and environmental conservation. Future directions emphasize the development of more efficient, adaptable, and scalable architectures to handle domain-specific challenges.

Keywords:

Plant species identification, Convolutional Neural Networks (CNN), Pretrained CNN models, ResNet50, InceptionV3, VGG16, DenseNet121, MobileNetV3

1.Introduction:

Protecting plant species, their habitats, and ecosystems in order to save them from going extinct is known as plant conservation. Because they provide oxygen, store carbon, provide food, and are the basis for biodiversity, plants are vital to ecosystems. However, many plant species are in danger of going extinct as a result of human activities such pollution, climate change, habitat degradation, deforestation, and the introduction of invasive species. Plant conservation contributes to biodiversity preservation, which enhances ecosystem resilience and yields important resources including materials, meals, and medications. There are more than 60,000 different kinds of trees. All life on Earth depends heavily on plants, which also give food and

oxygen, lessen environmental pollution, shelter a variety of birds, sustain agricultural output, and delicately balance our ecosystem. The appropriate maintenance of our ecology is our duty. The most important prerequisite for this is accurate plant identification since one needs to be aware of the advantages and the availabilities of trees [1]Manual identification will be time-consuming and prone to errors for non-experts, and it may need a lot of resources for professionals with domain knowledge [2]. Convolutional neural networks (CNNs), a type of deep learning approach, have transformed the field of plant species recognition in recent years by allowing for automatic and precise identification based on a variety of plant parts, such as leaves, flowers, and fruits [3] . Previous Neural networks , already been trained on sizable datasets, frequently for tasks like object detection or picture classification, are known as pretrained CNN (Convolutional Neural Network) models. VGGNet, ResNet, and Inception are a few examples. These models are useful for "transfer learning," in which time and computing resources are frequently saved by applying their acquired knowledge to new, related activities. These models allow us to either alter and retrain the final layers to categorize new image categories or freeze specific layers to operate as feature extractors [2]. By including specialized layers, they can also be modified for more complex tasks like segmentation and object detection. They provide high accuracy for tasks and generalize well even with little fresh data, which is why they are frequently utilized in industries like medical imaging, autonomous driving, and agriculture.

This paper is aimed to survey some classification models developed by other researchers in order to find out which model performs the best and what are the advantages and limitations of certain models. The structure of the paper is as follows: The background is covered in Section 2. The findings of related work have been discussed in section 3. Section 4 deals with discussions about the models and the future scope one might have furthering research in this field.

2. Background:

2.1 Convolutional Neural Network

While deep learning, a part of machine learning that automatically extracts features, machine learning is a subset of artificial intelligence (AI) that does it manually. The human brain served as the model for CNN [1]. A CNN's architecture is not the same as that of a standard neural network. In a CNN, neurons in one layer only connect to a small number of neurons in the subsequent layer, as opposed to all of the neurons in the subsequent layer in a typical neural network [4]. Essentially, there are two main parts: the classification layer and feature representation. CNN comprises three distinct layers: fully connected, pooling, and convolutional. The convolutional layers and pooling layers are used to extract features. Convolution generates what is effectively a features map as an output after applying a kernel of different sizes to the input image. A range of convolutional methods with different filter and step sizes are applied to the input image. The layer's output is finally obtained by adding the feature maps. Each convolutional layer is followed by an activation function to make the function non-linear. A few of the several kinds of activation functions are Tanh, Sigmoid or

Logistic, and RELU. The pooling layer then receives the feature map that was taken out of the convolutional layer.

A simplified representation of a CNN model is shown in Figure 1, where an input image of an car is sent to hidden layers that apply multiple filters to the image in order to extract pertinent features. This process is carried out multiple times until a final set of characteristics is produced and transmitted to the output layer. The output layer creates a probability distribution for each of the three classes—plane, bus, and automobile. The network then makes its decision based on the class with the highest probability. Each layer's filter weights are learned by a technique called back propagation, which modifies the weights to lessen the difference between outputs that are expected and those that are actually produced [5].

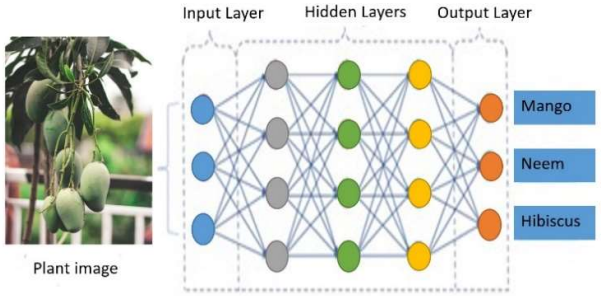

Figure 1. Simplified model of CNN [5]

2.1 Pretrained CNN models

With their amazing powers and ability to accelerate progress in image-related tasks, pre-trained Neural Network models have completely changed the field of computer vision. Developers and researchers can get remarkable outcomes while saving time and money by using pre-trained models. By using the knowledge of previously trained models, transfer learning enables a freshly constructed model to improve performance and generalize to new data without requiring a significant quantity of training data. Five of the top pre-trained CNN models will be examined in this article, along with their architectures, uses, and contributions to computer vision.

2.1.1 ResNet50

ResNet, short for Residual Network, is a ground-breaking CNN design that addressed the vanishing gradient issue by including skip links. Large-scale image categorization datasets like ImageNet have been used to train ResNet models. The ImageNet dataset was used to train one particular variation, ResNet-50. ResNet50 has demonstrated exceptional accuracy in a range of picture identification tasks thanks to its deep layers and residual connections, which enable it to learn complex representations. It has a wide range of applications, including medical imaging analysis, image style transfer, object identification, and image categorization [6].

2.1.2 VGG16

A well-liked CNN model that is commended for its effectiveness and simplicity is VGG16. Later, the ImageNet dataset—a sizable visual database with photos categorized into several

groups—was presented to the VGG16 model. 16 convolutional, fully linked layers with a fixed input size make up its structure. VGG16 is a reliable tool for picture categorization because of its deep layers and uniform architecture. The model's versatility has been used to the fields of visual question answering, medical imaging, and even art creation [7] [8].

2.1.3 Inception-v3

Inception-v3, a member of the Inception family of CNN models, places a strong emphasis on efficient feature extraction and processing efficiency. This model, which was first presented by Google, makes use of inception modules that include several concurrent convolutional layers of varying sizes. Once more, the ImageNet dataset was used to train Inception-v3. To learn the visual representations, the model has been extensively trained using the enormous dataset. Because of its architecture, Inception-v3 can collect both local and global characteristics, which makes it appropriate for a variety of applications like fine-grained picture categorization, object localization, and image recognition [9] [10].

2.1.4 MobileNet

The goal of MobileNet is to strike a compromise between accuracy and model size. Like other models, MobileNet was trained using the ImageNet dataset. Nonetheless, MobileNet is specifically made to be effective for embedded and mobile devices while preserving a respectable level of accuracy. MobileNet uses depthwise separable convolutions to significantly reduce the number of parameters while maintaining reasonable performance. Because of its lightweight design, MobileNet is perfect for applications with constrained processing power, such as augmented reality, real-time object identification, and mobile image recognition. [2] [11].

2.1.5 DenseNet121

Known for its effective and precise performance in image classification applications, DenseNet121 is a popular design in the Dense Convolutional Network family. The distinctive "dense connections" of DenseNet121's architecture let each layer to receive inputs from every layer before it, improving gradient flow and feature reuse. Compared to many other deep networks, such as ResNet, DenseNet121 is compact and efficient, requiring fewer parameters due to its architecture, which avoids duplicate feature learning. To keep the model lightweight and reduce dimensionality, DenseNet121 has 121 layers organized in dense blocks with transition layers between. Because of its small size and great accuracy, DenseNet121, which is frequently pretrained on big datasets like ImageNet, can be utilized in transfer learning or customized for particular applications, like as industrial inspections or medical imaging [6] [12].

3.Related work

In [1], the work uses a bespoke dataset of 10,000 photos, 1,000 for each variety of tree (e.g., mango, neem, and hibiscus), and provides a CNN model for detecting ten types of plants based on leaf images. A 13-megapixel Redmi 5A camera was used to take the pictures. Three fully connected layers and six convolutional layers make up the CNN architecture, which uses

Softmax to classify data across ten tree classes. Convolutional layers employ max pooling to minimize dimensions after using 32, 64, and 128 filters. Ten classification outputs are provided by the final Softmax layer, while the fully linked layers have 128 and 50 neurons. Four epoch settings were used for training: 100, 500, 2000, and 2500. Accuracy increased steadily, peaking at 99.40% on 2500 epochs. Twenty percent of the dataset was used for testing and validation, while the remaining 80 percent was used to train the model. The results demonstrate the effectiveness of the CNN model in automatically categorizing trees from leaf pictures. In order to increase the model's usefulness in botanical and ecological research, it might be expanded to categorize additional tree species or to include additional plant characteristics like flowers, fruits, and stems.

In [2],the use of transfer learning to enhance deep learning model performance in situations when labeled data is expensive and scarce is examined in this work. By utilizing knowledge from pre-existing models, transfer learning techniques—specifically, feature extraction (using pre-trained model activations as features) and fine-tuning (modifying final layers for new tasks)—allow for quicker and more effective training. The MobileNet architecture's effective depthwise separable convolutions, which lower computational complexity, make it especially noteworthy for mobile and embedded applications. The capabilities of MobileNet models were illustrated by experiments using the LeafSnap dataset, which included field photos of 184 different tree species. According to the study, MobileNetV2 initially obtained 84% validation accuracy and 97% training accuracy, which after fine-tuning stayed at 86%. MobileNetV3-Large attained 87% validation accuracy and 94% training accuracy after 100 epochs. The validation accuracy rose to 91% after adjustments. Additionally, a lightweight variant called MobileNetV3-Small showed its effectiveness by obtaining comparable training (92–94%) and validation accuracies (82–86%) with only 188,600 parameters and 60% less training time. The study's conclusions showed that MobileNetV3-based models outperformed MobileNetV2 in terms of accuracy and efficiency. This makes them ideal for deployment on smartphones and Internet of Things devices, especially in distant areas with limited network connectivity. It is noted that even higher accuracy gains could result with data augmentation. When all is said and done, MobileNetV3, particularly the Small version, is acknowledged as a cost-effective and time-efficient model that is ideal for real-world embedded applications.

[3] This research focuses on applying deep learning models for computer vision and recognition of Indian medicinal plant species. The 1,822 photos of 30 different plant species in the medicinal leaf dataset were divided 70:30 for training and testing. Five deep learning architectures are compared in the study: ResNet50, Inception v3, Xception, MobileNet, and DenseNet121. Inception v3 is chosen because of its higher accuracy. Every model has special advantages: DenseNet121 connects layers in a dense feed-forward manner, improving information flow; Xception replaces Inception modules with depthwise separable convolutions, optimizing parameter use and frequently outperforming Inception on larger datasets; MobileNet offers a lightweight architecture appropriate for mobile environments, utilizing depthwise separable convolutions to reduce parameters; and ResNet50 uses batch

normalization and residual connections to address vanishing gradients, enabling deeper networks. A stochastic gradient descent optimizer with a learning rate of 0.05 and momentum of 0.9 was used for 150 training epochs. The models were evaluated based on loss and validation accuracy. Inception v3 outperformed the others in terms of accuracy and fit for the classification requirements of the dataset, obtaining the maximum validation accuracy of 95%. Particularly for this specialized application of identifying plant species in their native habitats, the study showed that deep learning models perform better than conventional manual classification. The findings indicate potential for using Inception v3 or comparable models in mobile applications, offering a useful tool for identifying plants in the environment that could assist conservation initiatives, botanical research, and rural populations.

In [7], the VGG16 model is used in this study's two-stage transfer learning method to categorize Dipterocarpaceae tree species according to their trunk textures. First, two meticulously planned processes were used to update the VGG16 model, which had already been trained on ImageNet. Only a new classification layer was trained in the first stage, and feature layers were frozen to preserve the knowledge acquired from ImageNet. Within 11 epochs, this phase rapidly reached almost 100% accuracy on both the training and validation sets. In order to fine-tune the final three convolutional layers for the Dipterocarpaceae dataset, they were selectively unfrozen in the second stage. This allowed the model to adapt to this particular job while maintaining key elements from its initial training. With a batch size of 32 and up to 50 epochs, the training was conducted using Adam optimization and early stopping. In the first and second stages, the learning rates were 0.001 and 0.0001, respectively. Random zoom, scale, shear, and shift were among the data augmentation techniques employed to improve model generalization. When comparing various freezing and unfreezing approaches, the best results were achieved by unfreezing only the final layer, with an accuracy of 97.09% and a Macro F1 score of 97%.

This method supported conservation efforts by successfully and swiftly identifying Dipterocarpaceae species. The dataset will be enlarged in subsequent research, and sophisticated architectures for texture-based categorization will be investigated.

In [10], the study focuses on developing a plant recognition system using a dataset of images for two tree types: mango and Alstonia. The dataset contains 35 test images and 132 training images per tree type. For classification, the authors used a Convolutional Neural Network (CNN) and the pretrained Inception V3 model, known for high accuracy on large datasets like ImageNet. Key CNN hyperparameters included a depth of 1, a filter size of 3, a stride of 1, ReLU activation, the ADAM optimizer, and max pooling to reduce dimensionality. The CNN model achieved an accuracy of 80% in distinguishing between the two trees, while Inception V3 also performed well, surpassing 78.1% accuracy in prior studies on ImageNet. According to the study, it can be difficult to identify the Alstonia tree just by looking at pictures of its leaves because it looks like other species. In order to further increase model accuracy, future study will address this by adding photos of the bark and leaves taken at dawn to the dataset. The model workflow, which was implemented using TensorFlow, included preprocessing, feature extraction, and classification. In the end, it achieved an accuracy of more than 80%.

4.Discussions and future scope:

Various Convolutional Neural Network (CNN) models offer unique benefits and drawbacks in the field of picture classification, particularly for tasks like plant or tree identification. Smaller datasets or applications with limited resources can benefit from simple CNN architectures, which are typically built using a few convolutional and pooling layers and offer solid baseline results with high interpretability and lower computing costs. For example, a simple CNN model, such one with three fully connected layers and six convolution layers, is easy to design and can attain high accuracy (e.g., over 99% in some tasks) with enough training. However, because they lack the depth and feature extraction capability of more advanced networks, such models may have trouble scaling and generalizing to increasingly complicated datasets.

Pretrained architectures like Inception V3, DenseNet, and ResNet are examples of advanced models that offer significant advantages. The ability of Inception V3's architecture to capture both fine-grained and large-scale features makes it ideal for datasets with a wide range of characteristics. DenseNet models are particularly well-suited for jobs involving intricate patterns in very limited datasets because of their dense connection, which facilitates effective gradient flow and feature reuse. ResNet performs well with very deep networks and big datasets because it uses residual connections to solve the problem of disappearing gradients. These more sophisticated devices do, however, have drawbacks.

On difficult datasets, deeper networks may perform better than simpler CNNs in terms of accuracy and feature discrimination despite their speed. If incorrectly modified or applied to data that differs significantly from their initial training datasets, even strong pretrained models (like ImageNet) may overfit. This is somewhat aided by transfer learning and fine-tuning, which enable these models to adjust to particular tasks but may call for additional processing power and knowledge. In the future, hybrid models—particularly for edge and mobile applications—that blend pretrained networks with simpler CNN architectures might offer a balance between accuracy and efficiency. The resilience of plant recognition algorithms may also be improved by enlarging datasets and adding multimodal data, such as bark or flower photos for tree classification. High-performance models may become more widely available by including more effective topologies, such as MobileNet or EfficientNet, which are tailored for resource-constrained situations. We can anticipate more precise, adaptable, and user-friendly plant identification tools for a greater range of applications when these developments are combined with continuous advancements in model interpretability and feature visualization.

5.Conclusion:

With opportunities to balance model efficiency and complexity for a variety of applications, the future of plant and tree identification with CNNs is bright. Even though sophisticated models like DenseNet, ResNet, and Inception V3 are excellent at processing complex data and extracting features, they have significant computational requirements. Particularly for mobile

and edge devices, hybrid models that combine pretrained networks with simpler CNN structures or streamlined architectures like MobileNet may provide a workable solution. The resilience and adaptability of classification algorithms would be improved by growing datasets and include multimodal data (such as bark, leaves, and flowers), which would enable them to more effectively generalize across a variety of plant species. Enhancements to the model's interpretability would also increase user confidence and usability.

To conclude, the development in resource efficient plant recognition systems could serve a wide range of user inclusive of academia, agriculture and environmental conservation.

6.References

[1] I. Zarrin, ""Leaf based trees identification using convolutional neural network."," *Convergence in Technology (I2CT),* pp. 1-4, 2019.

[2] A. Hussain, ,. B. ., Barua , Osman, A., , Abozariba, R. and Asyhari, A. T, "Performance of mobilenetv3 transfer learning on handheld device-based real-time tree species identification.," *26th International Conference on Automation and Computing (ICAC),* pp. 1-6, 2021.

[3] Kavitha Kayiram, Prashant Sharma, Shubham Gupta and R. V. S. Lalitha, ""Medicinal Plant Species Detection using Deep Learning.," *2022 First International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT),* pp. 01-06, 2022.

[4] K. Simonyan, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556,* 2014.

[5] M. Krichen, "Convolutional neural networks: A survey.," *Computers 12, no. 8,* p. 151, 2023.

[6] He, Kaiming, Xiangyu Zhang, Shaoqing Ren and Jian Sun, "Deep residual learning for image recognition.," *Proceedings of the IEEE conference on computer vision and pattern recognition,,* pp. 770-778, 2016.

[7] Widians, Joan Angelina, Masna Wati, Novianti Puspitasari, Ummul Hairah and Ade Fiqri Tjikoa, "Texture-based Dipterocarpaceae trunk classification using two stage transfer learning of VGG16," *2023 International Conference on Electrical Engineering and Informatics (ICEEI),* pp. 1-4, 2023.

[8] S. K. H. R. G. a. J. S. Ren, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE transactions on pattern analysis and machine intelligence 39, no. 6,* pp. 1137-1149., 2016.

[9] C. V. V. S. I. J. S. a. Z. W. Szegedy, "Rethinking the inception architecture for computer vision.," *Proceedings of the IEEE conference on computer vision and pattern recognition,* pp. 2818-2826, 2016.

[10] M. a. A. C. Balipa, "Alstonia Tree Detection using CNN and Inception V3 Algorithms.," *2022 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS),* pp. 318-321, 2022.

[11] A. G. Howard, "Mobilenets: Efficient convolutional neural networks for mobile vision applications.," *arXiv preprint arXiv:1704.04861,* 2017.

[12] G. Z. L. L. V. D. M. a. K. Q. W. Huang, ""Densely connected convolutional networks," *Proceedings of the IEEE conference on computer vision and pattern recognition,* pp. 4700-4708, 2017.