Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 5, No.3, May : 2024

PLANT SPECIES IDENTIFICATION

G. Saranya Assistant Professor, Computer Science and Engineering Narasaraopeta Engineering College (*Autonomous*) Narasaraopet , India dasarisaranya4@gmail.com
G. M. Karthik Reddy. Student Computer Science and Engineering Narasaraopeta Engineering College (*Autonomous*) Narasaraopet , India karthikreddi4444@gmail.com
D. N. V. Charan Reddy.Student Computer Science and Engineering Narasaraopeta Engineering College (*Autonomous*) Narasaraopet , India charan172003@gmail.com
M.Ratnaker Student Computer Science and Engineering Narasaraopeta Engineering College (*Autonomous*) Narasaraopet , India ratnakar.muppuri@gmail.com

Abstract –
Since plants are fundamental to human survival, conserving biodiversity depends on the creation of a plant database and identification system. Researchers have been primarily focused on plant identification based on leaf classification for the past 20 years. Convolutional neural network (CNN) modeling for plant identification is suggested. In such a network, overfitting is a serious problem, though. To deal with this issue, a regularization procedure similar to dropout is a viable choice. Furthermore, on the same database, parameter fine-tuning and transfer learning have produced positive results. The acquired validation accuracy serves as a proxy for the convolutional neural network's performance in the proposed system. This study assessed the ability of two well-liked boosting techniques, Adaboost.M1 and Logit Boost algorithms, to improve the plant classification performance of a classifier: Support Vector Machine (SVM).

Our method for Plants Species Identification Convolutional Neural Network (CNN) and ResNeXt101-32x4d model, providing experts with an easy-to-use interface to upload photos and get precise predictions, we get 99% accuracy.

Keywords –
Plants, Deep learning, Python flask, convolutional neural network (CNN), Plant species prediction.

I. INTRODUCTION

In many fields, including agriculture, medicine, industry, and environmental conservation, plant species variety is critical. In addition to being essential for providing food, medicine, and raw materials, plants are also responsible for preserving the equilibrium of atmospheric gases. However, a number of plant species have become extinct more quickly as a result of human activities including pollution, urbanization, and deforestation. This has had a negative impact on ecosystems and human health. It is obvious that plant species need to be protected, but the large number of species—many of which are still unknown—makes traditional identification techniques extremely difficult. Tools for automatically identifying different species of plants by utilizing image-based techniques present a viable answer to this problem. With the help of these technologies, anyone—even non-experts—can identify different species of plants by only taking and processing pictures with their mobile devices. The goal of this project is to automate the identification of plant species using deep learning architectures, namely convolutional neural networks (CNNs). CNNs are an attractive method for improving plant species identification since they have shown impressive performance in a variety of image-based domains. Through the utilization of deep learning, this research seeks to advance the creation of effective and user-friendly instruments for plant species identification, which will support conservation initiatives and encourage wider involvement in the preservation of biodiversity.

II. LITERATURE SURVEY

He et al. [1] study "Deep Residual Learning for Image Recognition" presents ResNet, a potent neural network design. ResNet uses skip connections to address the issue of training very deep neural networks. The network can learn residual functions thanks to these connections, which facilitates the training of incredibly complex models with more than 1000 layers. ResNet's ability to efficiently learn from massive datasets like ImageNet allowed it to produce outstanding results on tasks like picture classification.[2] It employs a coarse-to-fine method based on the botanical taxonomy to determine the genus and species of plants. The global and local aspects of leaf picture complement each other in the two-view depiction. To make the method adaptable to new plant species and less reliant on a large number of training data, a deep measure based on Siamese convolutional neural networks is employed.

[3] A methodology for crop disease identification and classification utilizing deep convolutional neural networks (CNNs) and transfer learning is presented by Vangaveeti et al. Their strategy is comparable to our plant species research in that it uses deep learning techniques for image classification tasks, albeit being centered on crops. Their work shows how pre-trained models can increase classification accuracy in agricultural applications by utilizing transfer learning. The study's focus on disease identification is in line with the overarching objective of our research, which is to improve the monitoring and management of plant health.

[4] Taufik et al. present a study on plant type identification using leaf images and convolutional neural networks (CNNs). Their research shares common ground with our plant species project, as both aim to classify plants based on visual attributes. By leveraging CNNs, their approach demonstrates the efficacy of deep learning techniques in accurately identifying plant types from leaf images. This study underscores the potential of image-based classification methods for plant species recognition and monitoring, aligning with our project's objectives.

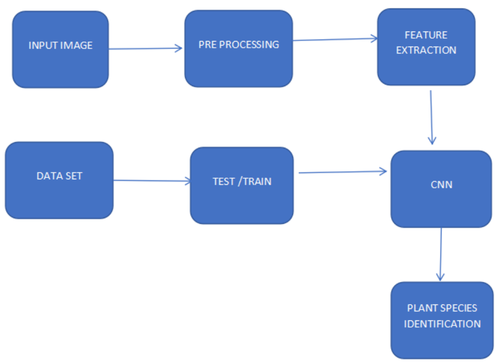
[5] In the fields of computer biology and bioinformatics, Chaudhury and Barron offer a technique for identifying plant species from obscured leaf photos. Their method tackles problems with clouded leaf images, which is a frequent problem in tasks involving plant classification. Their research improves plant species recognition accuracy through the use of sophisticated computational approaches, which is in line with the objectives of our project. This work provides insightful information for our own endeavors by highlighting the significance of strong algorithms for accurately recognizing plant species even in the presence of occlusions.

III. METHODOLOGY

This section will cover every step of training and developing the model, from gathering data to employing the most potent deep learning models to train our own data. Figure I represent the overall process of methodology.

The procedure begins with the input image, which is a representation of the raw data that was fed into the system for analysis. Subsequently, the input data undergoes preprocessing, which includes operations like scaling, data format conversion, and normalization, to prepare it for model training. Once the data has undergone preprocessing, it is separated into multiple subsets for testing, validation, and training.

The ResNext101_32× 4d model learns how to map input pictures to the correct labels using the training dataset. In order to prevent over fitting and modify hyper parameters, the model's performance is evaluated concurrently with the validation dataset.



Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 5, No.3, May : 2024

Fig. 1. Overall model process

A. Dataset Analysis

We downloaded the datasets from the Kaggle website over the internet, however prior plant species datasets are unquestionably necessary in order to perform any kind of forecast. The following link contains the dataset that we have gathered:
https://www.kaggle.com/datasets/xhlulu/leafsnap-dataset

The Leaf snap dataset, accessible through Kaggle, provides an extensive set of photos depicting various plant species, each linked to its scientific name. The collection includes nearly 26,000 images. The two primary categories of this supervised dataset are "images" and "segmented." Photographs of several plant species taken in field and laboratory settings are included in the 'images' category, offering a wide range of backgrounds, lighting situations, and viewpoints.

In the meantime, the "segmented" folder is made up of leaf form outlines from different plant species, obtained from both field and laboratory locations. This dataset provides researchers with an abundance of visual data that may be used to train strong machine learning models for tasks involving the identification of plant species. The utility of the dataset is increased with the addition of segmented leaf forms, which offer more data for possible feature extraction and analysis. Preprocessing, model creation, and evaluation stages of the project will be made easier with the help of this dataset analysis, which lays the groundwork for understanding the features and extent of the Leaf snap dataset.

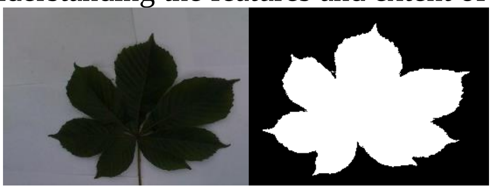


Fig.2.1. Original and Segmented image taken in field

Fig.2.2. Original and Segmented image taken in lab

B. Data Visualization

As seen in Figure 3, a graphic representation of the dataset is created to provide a quick comprehension and visualization of the various plant species. The Figure 3 displays percentage of different plant species that were utilized in the dataset (64 Plant Species).

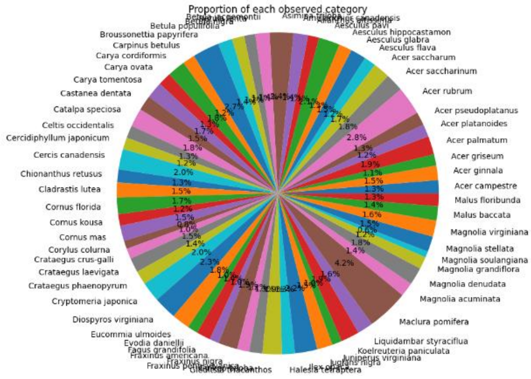


Fig. 3. Dataset Visualization

Figure 4 provides a detailed breakdown of the diverse variety of Plant species that are present in our dataset.The many different aspects of leaves are emphasized, such as their size, shape, colour, texture, orientation, background noise, image quality, veins and margins, dataset diversity, and intra-class variability. Researchers can better understand the type of plant represented by using this representation, which offers crucial context for the dataset's composition.

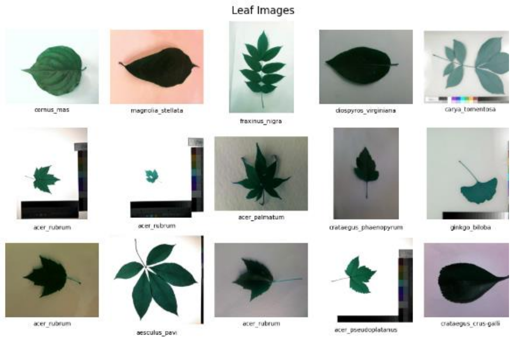


Fig. 4. Representation of various plants leaves in Dataset

C. Pre Processing Techniques

Gathering and examining data from a variety of sources, such as image repositories and botanical databases, is what the first phases entail. Understanding the structure, format, and potential issues such as missing values or class imbalances are the goals of this exploratory phase. The cleaning and preparation phase is essential for guaranteeing data consistency and integrity after data gathering. To do this, duplicates must be eliminated, missing data must be handled, numerical characteristics must be standardized, and categorical variables must be transformed into numerical representations for additional analysis.

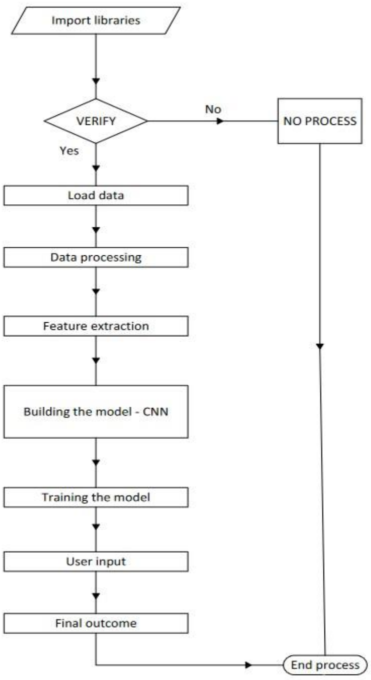


Fig. 5. The entire pre-processing procedure.

Next, more preparation measures are needed if the dataset contains plant photos. To improve dataset variety, images are resized and standardized to a common size and methods such as data augmentation are used. A key function is feature extraction, which is the detection of pertinent features—like leaf form, texture, or color—that define a species of plant. To evaluate the model, the dataset is then split into training and testing subsets, making that each subset reflects a variety of plant specie's distributions. Model robustness can be increased and class imbalances can be reduced by utilizing strategies like data augmentation and balancing. Lastly, data visualization methods are used to help with modeling decisions by providing insights into the important features of the preprocessed information.

D. Evaluation of Model

The ResNeXt101-32x4d model's efficacy in plant categorization tasks, a thorough examination of its performance metrics and diagnostic tools is required this strategy is inherited by ResNets [1] which stack modules of the same topology. This simple rule reduces the free choices of hyperparameters, and depth is exposed as an essential dimension in neural networks. The bottleneck width of the ResNeXt101-32x4d model is equal to 4, and its cardinality is 32.

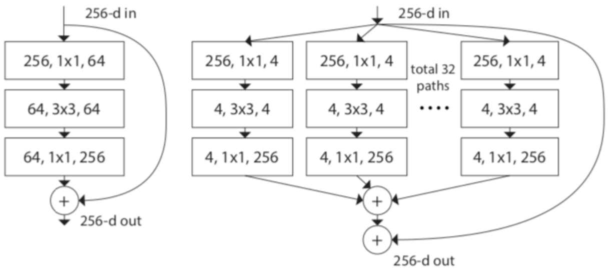


Fig. 6. Difference between ResNet bottleneck block and ResNeXt bottleneck block
Left: A ResNet block [1].
Right: Approximately the same complexity ResNeXt block with cardinality = 32. (# in channels, filter size, # out channels) is the representation of a layer.
The accuracy of the ResNeXt101-32x4d model for plant categorization, a number of actions must be taken during evaluation. First, the dataset is separated into sections for testing, validation, and training. Metrics like loss and accuracy are used to track the model's performance after it has been trained using optimization approaches.
the ResNeXt101-32x4d model by testing its accuracy in categorizing plant species, evaluating its performance on fresh data, and training it on a subset of the available data.

E. Accuracy

For the purpose of this experiment, two techniques were used to identify plant species. The plant species are identified and predicted by the system through the use of classification algorithms. Following identification, the system presents pertinent data together with the name of the identified plant species. With this capacity, our system can accurately identify plant species with a 99.2% accuracy rate in real-world circumstances. Figure 7.1.1 shows the accuracy values that our model was able to obtain, and Figure 7.1.2 shows the loss values that were seen when the model was being trained.

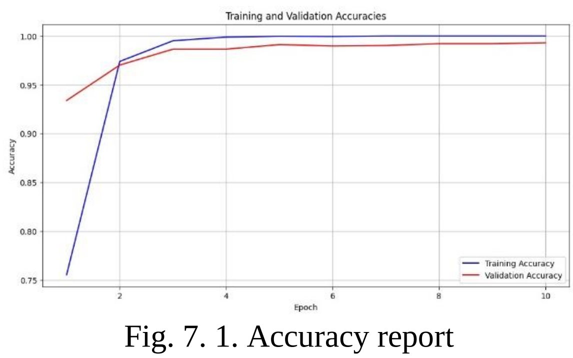


Fig. 7. 1. Accuracy report

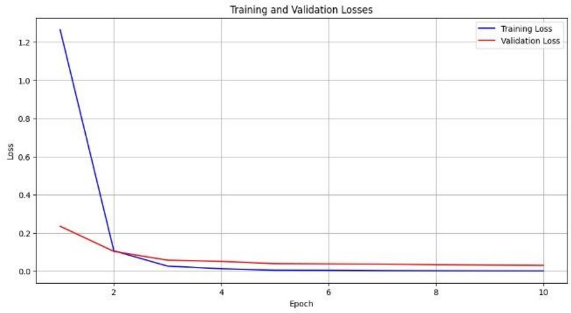


Fig. 7. 2. Loss report

Model	Existing Work	Accuracy
Siamese CNN	96%	-
ResNeXt	-	98%

IV. CONCLUSION

The results presented in this study highlight the potential of CNNs as effective instruments for automating tasks related to plant species identification, providing notable advantages over manual techniques in terms of efficiency, accuracy, and scalability. We think that as long as deep learning and computer vision techniques keep getting better, innovation in this area will keep growing and eventually help us comprehend and protect plant biodiversity globally.

V. FUTURE SCOPE

The identification of plant species with CNNs has the potential to significantly transform a number of fields, including agriculture, biodiversity conservation, and environmental science. To realize this potential and solve related issues, interdisciplinary cooperation, ongoing research, and innovation are essential.

VI. RESULT

A. Input image

To predict the name of the plant, the user must upload or submit an image of a leaf from that species. Figure 8 below shows the image of the user interface.
After selecting the image user have to click the upload image button.

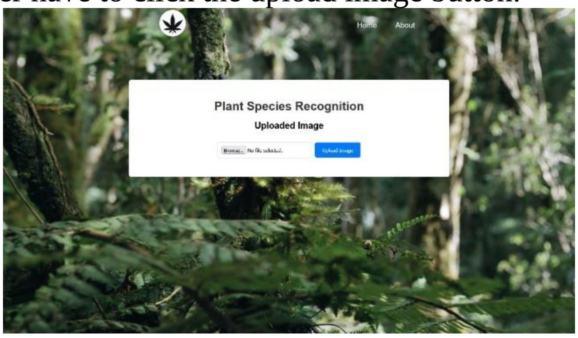


Fig. 8. Uploading the Image

B. Displaying Predicted Image

The figure 9 will displays the predicted result of the uploaded image after completion of the analysis with trained model.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 5, No.3, May : 2024

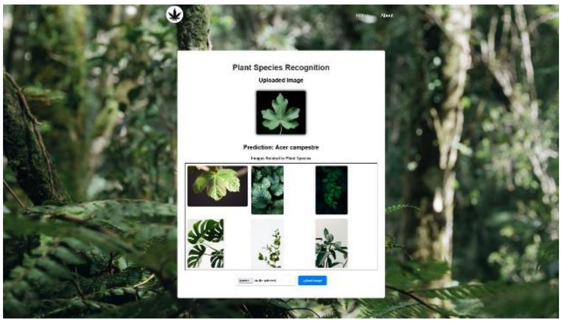


Fig. 9. Predicted Result

VII. REFERENCES

[1] He, X. Zhang, S. Ren, and J. Sun. Deep residuallearning for image recognition. In CVPR, 2016.
[2] V. M. Araújo, A. S. Britto Jr., L. S. Oliveira, and A. L. Koerich, ‘‘Two-viewfine-grained classification of plant species,’’ Neurocomputing, vol. 467, pp. 427–441, Jan. 2022.
[3] V. N. Suchir Vangaveeti, V. Kalaichelvi, A. K. Jha and A. John Joseph, "Crop Disease Detection and Classification using Transfer Learning and Deep Convolutional Neural Networks, Dubai, United Arab Emirates, 2023, pp. 179-184.
[4] I. Taufik, H. Nazilah, N. Lukman, Y. A. Gerhana, Jumadi and R. Permatasari, "Identification of Plant Types Based on Leaf Image Using Convolutional Neural Network (CNN), Makassar, Indonesia, 2023, pp. 1-5.
[5] A. Chaudhury and J. L. Barron, ‘‘Plant species identification from occluded leaf images,’’ IEEE/ACM Trans. Comput. Biol. Bioinf., vol. 17, no. 3, pp. 1042–1055, May 2020.
[6] X. Wang, W. Du, F. Guo, and S. Hu, ‘‘Leaf recognition based on elliptical half Gabor and maximum gap local line direction pattern,’’ IEEE Access, vol. 8, pp. 39175–39183, 2020
[7] J. Huixian, ‘‘The analysis of plants image recognition based on deep learning and artificial neural network,’’ IEEE Access, vol. 8, pp. 68828–68841, 2020.
[8] Z. Li, R. Guo, M. Li, Y. Chen, and G. Li, ‘‘A review of computer vision technologies for plant phenotyping,’’ Comput. Electron. Agricult., vol. 176, Sep. 2020, Art. no. 105672.
[9] T. Feilong, H. T. Yew, F. Wong and R. R. Porle, "Advancements for Improved Plant Disease and Pest Identification, Miri Sarawak, Malaysia, 2024, pp. 354-358.
[10] V. Sharma, K. Chaurasia and A. Bansal, "Optimizing Species Recognition in Medicinal Plants: A Comprehensive Evaluation of Deep Learning Models, Noida, India, 2024, pp. 738-743.
[11] A. Srivastava, P. K. Saini, K. Kumar, S. Tiwari and N. Garg, "Potato Leaf Disease Detection Using Dense Net-CNN, Bengaluru, India, 2024, pp. 648-653.
[12] P. Pandey and R. Patra, "A Real-time Web-based Application for Automated Plant Disease Classification using Deep Learning, Ahmedabad, India, 2023, pp. 230-235.
[13] C. S, M. F. Begam, C. N. S and G. G. S, "Medicinal Plants Attribute Detection by Deep learning Image Processing Techniques, Bangalore, India, 2023, pp. 1-5.
[14] A. S. Kini, P. K. V and S. N. Pai, "Leaves Classification Through Image Segmentation and Convolutional Neural Network, Greater Noida, India, 2021, pp. 631-634.