Plant Species Identification

Ms. Bandalgi B.S.[1] Ms. Doke S.D. [2] Ms. Bidbag V.S. [3] Mr. Gadade Y.B.[4] ,

Guide: 1.Prof.Hosale A.P.
Department of Computer Engineering

A.G.Patil Insttitue of Technology Solapur, India

Abstract:

We focus on plant species identification as it is a classic and hot issue. In tradition plant species identification the samples are scanned specimen and the background is simple. However, real-world species recognition is more challenging. We first systematically investigate what is realistic species recognition and the difference from tradition plant species recognition. To deal with the challenging task, an interdisciplinary collaboration is presented based on the latest advances in computer science and technology.

Plants play an irreplaceable role in our world and they have direct effect in many domains such as agriculture, climate, ecological system and so on. Besides, they are the main source of food for human survival and development. Many problems such as habitat degradation, global warming, ecosystems destruction,environment worsen, species extinction, and so on have something to do with plant protection. Plant species identification is the prerequisite for protection.

There have been many research related to the issue. Method based on image classification is now considered to help improve the plant taxonomy. It is o the most promising solutions among those related research work, as discussed in And it has been a long term hot research issue.

Considering flowers and fruits of plants are seasonal, some researchers believe that leaves are more suitable for identification. In the early time, leaves are frequently used for computer-aided plant species classification. Most image based identification methods and evaluation data proposed were based on leaf.

Key Words:- Plant Species Identification,Feature Extraction,SVMs.

1. Introduction:

Plants are a fundamental part of life on our planet. They give us oxygen to breathe, food, medicine and plenty of other things which make our lives worth living. They are the backbone of all life [2].
Accurate identification of plant species is vital for various scientific and practical applications. Traditional identification methods rely on botanical expertise and physical examination, which can be time-consuming and error-prone. Advances in machine learning and image processing have led to the development of automated systems world for there is an unlimited number of plant species. Hence, the task of plant identification is limited to a very small number of people. However, plant species knowledge is necessary for various purposes such as identifying a new or rare species, balancing of the ecosystem, medicinal purposes, agricultural industry, etc.

[2]. To be able to achieve these objectives, automation of plant species identification is a necessity [3].

There are enormous plant species in the world, which is nearly 390,000 [4] in number, and each year, new species are reported in different parts of the world [5].capable of identifying plant species using images of leaves, flowers, or other plant parts.

Accurate identification of plant species is a critical component in a variety of disciplines including botany, agriculture, ecology, and conservation biology. Traditional identification techniques rely on expert knowledge and detailed morphological analysis, which can be time-consuming, error-prone, and inaccessible to non-specialists. These limitations have driven the development of automated plant identification systems that utilize modern computational techniques.

In recent years, machine learning (ML), especially deep learning, has emerged as a powerful tool for image-based plant species classification. Convolutional Neural Networks (CNNs), a class of deep learning models particularly well-suited for visual pattern recognition, have shown significant success in processing complex image data for identifying plant traits such as leaf shape, vein structure, and color patterns.

Several datasets have facilitated the training and evaluation of such models. For example, the Leafsnap dataset [6] contains high-resolution images of leaves from tree species found in the Northeastern United States, and has become a standard benchmark for leaf classification. Other datasets such as PlantCLEF and Flavia provide broader taxonomic diversity and environmental variation, making them valuable resources for evaluating model generalizability.

Numerous studies have demonstrated the efficacy of CNN-based approaches. [7] implemented a deep learning framework called Deep-Plant and achieved high accuracy in identifying plant species from leaf images. Similarly, [8] applied deep learning to digitized herbarium specimens, highlighting its potential for large-scale biodiversity assessments.

This research builds on these foundations to develop and evaluate a CNN-based model for plant species identification using publicly available datasets. By leveraging data augmentation and transfer learning techniques, the study aims to improve classification accuracy and robustness across species and imaging conditions. The findings have implications for creating accessible plant ID tools and advancing automated ecological monitoring systems.Manual identification is often time-consuming and inefficient [9], even the expert taxonomists take a considerable amount of time to identify a plant species. Since the traditional identification methods are strenuous,

there arises a need to automate the process of species identification. As a result, researchers have tried to develop automated plant species identification and classification systems which can serve the purpose of species recognition to some extent. A few of these are discussed in the next

section.[1]

Identifying plants through their leaves is a thoroughly pursued endeavor that has widely varying applications ranging from ecology, horticulture, disease identification, rare plant preservation in plants to medicinal applications in Ayurveda and various plant bases medical systems. Our purpose in this project is to identify

plant species digitally using the image of a single leaf through neural networks. We will approach our project using Keras, Tensorflow, and Convolutional Neural Networks. Fore mentioned approach gives satisfactory results with high accuracy.

2.Literature Review:

Wu et al. [1] have proposed on of the earliest plant identification system. In their scheme, they have created their own dataset named Flavia, which has been used by various other researchers as standard dataset for their work. It consists of 1907 leaf images of 32 different plant species. In their study, they extracted 5 basic geometric and 12 digital morphological features based on shape and vein structure from the leaf images. Further, principal component analysis (PCA) was used to reduce the dimensions of input vector to be fed to the probabilistic neural network (PNN) for classification. They used a three layered PNN which achieved an average accuracy of 90.32%. Wang et al. [2[proposed a robust method for leaf image classification by using both global and local features. They used shape context (SC) and SIFT (Scale Invariant Feature Transform) as global and local features respectively. Knearest neighbor (k-NN) was used to perform classification on ICL dataset which achieved an overall accuracy of 91.30%.

1. Shift Toward AI in Plant Species Identification

To overcome these limitations, researchers have turned to Artificial Intelligence (AI), especially Deep Learning using Convolutional Neural Networks (CNNs). These models automatically learn relevant features from leaf images—such as shape, vein structure, color, and texture—enabling high accuracy in classification.

Studies such as Lee et al. (2017)[11] and Rehman et al. (2019)[12] have shown that CNN-based models outperform traditional machine learning techniques in plant species identification.

2. CNN Architectures and Performance

Several CNN architectures have been used in plant classification:

AlexNet (Krizhevsky et al., 2012):[13]

One of the first deep models applied to plant leaf classification, showing high accuracy with large datasets.

VGGNet (Simonyan & Zisserman, 2014):[14]

Used for its deep architecture and small convolution filters; effective in capturing leaf shape and edges.

ResNet (He et al., 2015):[15]

Introduced residual learning, enabling deeper networks without performance degradation; improved classification for similar-looking species.

Inception & Inception-ResNet (Szegedy et al.):[16]

Combined parallel convolutions of varying sizes to capture multi-scale leaf features, improving model accuracy and robustness.

These deep learning models have shown superior performance in classifying plant species, making them ideal for applications involving large and diverse image datasets.

3.Need of this project:

1. Biodiversity Conservation:

- Accurate identification is essential for understanding and protecting biodiversity.
- It allows for better monitoring of plant populations and habitats, which is crucial for conservation efforts.
- By identifying species, scientists can better understand their ecological roles and develop strategies to protect them from threats like habitat loss and climate change.

2. Ecological Monitoring:

- Species identification is fundamental for tracking changes in plant communities and ecosystems.
- It enables scientists to understand how different species interact with each other and with their environment.
- This knowledge is vital for managing ecosystems effectively and predicting the impact of human activities on plant life.

3. Safe Use of Plant-Based Products:

- Accurate identification is crucial for the safe use of plant-based natural health products.
- It helps distinguish between beneficial and potentially harmful plants, ensuring that people use them correctly and safely.
- For example, knowing the scientific name and plant parts of a medicinal plant can help avoid confusion and potential toxicity.

4. Research and Education:

- Plant species identification is a foundational skill for botanists and other scientists studying plants.
- It provides a basis for understanding plant diversity and evolution.
- It also helps researchers to develop new methods for plant identification, such as using computer vision and machine learning.

5. Communication and Nomenclature:

- Accurate identification allows for clear communication about plant species, both within scientific communities and with the public.
- Using scientific names (genus and species) avoids confusion caused by common names, which can vary geographically and be used for multiple species.
- This is particularly important in fields like forensics, where accurate identification is crucial for legal proceedings.

4.Proposed System:

The goal is to develop a machine learning-based application that identifies plant species using images of leaves, flowers, or entire plants. The system can be trained on a dataset of labeled plant images and deployed via a web or mobile interface.

- System Components:-

1. Data Collection
   - Collect or use existing datasets (e.g., PlantCLEF, LeafSnap, Flavia).
   - Each data point should contain:
     - Image of the plant (leaf, flower, etc.)
     - Label (species name)

2. Data Preprocessing
   - Resize and normalize images
   - Augment dataset (rotate, flip, zoom)
   - Encode labels (One-hot or LabelEncoder)

3. Feature Extraction
   - Option A: Use a pretrained CNN (like VGG16, ResNet, MobileNet) for transfer learning.
   - Option B: Train your own CNN from scratch if dataset is large enough.

4. Model Building
   - Use Keras/TensorFlow
   - Use classification layers (Softmax at the end)
   - Evaluate with accuracy, precision, recall, F1-score

5. Training
   - Train/validation/test split (e.g., 70/20/10)
   - Use early stopping and model checkpoints
   - Hyperparameter tuning (learning rate, batch size, etc.)

6. Evaluation
   - Confusion matrix
   - Per-class accuracy
   - Visualize predictions vs. actual labels

5.Advantages

1. High Accuracy and Consistency

- ML models (especially deep learning with CNNs) can achieve high accuracy in identifying species, often outperforming human experts in consistency.
- Reduces human error and subjectivity in plant identification.

2. Scalability

- Can be trained on thousands of plant species and easily updated with new data.
- Capable of identifying rare or region-specific species if properly trained.

3. Fast and Real-Time Predictions

- Once trained, the model can provide instant results.
- Useful in real-world applications like agriculture, botany, and environmental monitoring.

4. Cost-Effective

- Reduces reliance on human experts or laboratory tests for plant classification.
- A smartphone with a camera and the app can be enough for field identification.

5. Accessibility

- Can be deployed on mobile or web platforms, making it widely accessible to:
  - Farmers
  - Botanists
  - Environmental researchers
  - Hobbyists and students

6. Automation

- Enables automated plant monitoring systems in smart agriculture.
- Can integrate with drones or robots for autonomous species tracking.

7. Learning and Educational Tool

- Useful for students and researchers to learn plant taxonomy.
- Can include additional information such as habitat, medicinal use, and conservation status.

8. Customizability

- Can be adapted to different domains:
  - Agriculture: Crop species
  - Forestry: Tree identification
 - Urban environments: Weed detection

6.Methodology:

The methodology for solving the problem for Crop Disease Prediction could include the following steps:

1) **Input Image:** Read an image into the workspace, using the imread command. The example reads one of the sample images included with the toolbox, an image, and stores it in an array named I . imread infers from the file that the graphics file format is Tagged Image File Format (TIFF).

2) **Preprocessing:**

   Image Resize: Resizing an image is done using software to add or subtract pixels, and is called resampling. When an image is resampled it increases or decreases the width and height of the image in pixels. There are other ways to resize an image, such as by cropping it to a smaller size. Hence in this process, image is resized into 256 X 256, i.e. the number of pixels in the row is equal to 256 and then the number of pixel in the column is equal to the 256.

3) **Feature Extraction:** Local binary patterns (LBP) is a type of visual descriptor used for classification in computer vision. LBP is the particular case of the Texture Spectrum model proposed in 1990.

4) **Classification:** "Support Vector Machine" (SVM) is a supervised machine learning algorithm that can be used for both classification or regression challenges.

   ➢ **Architecture Diagram:**
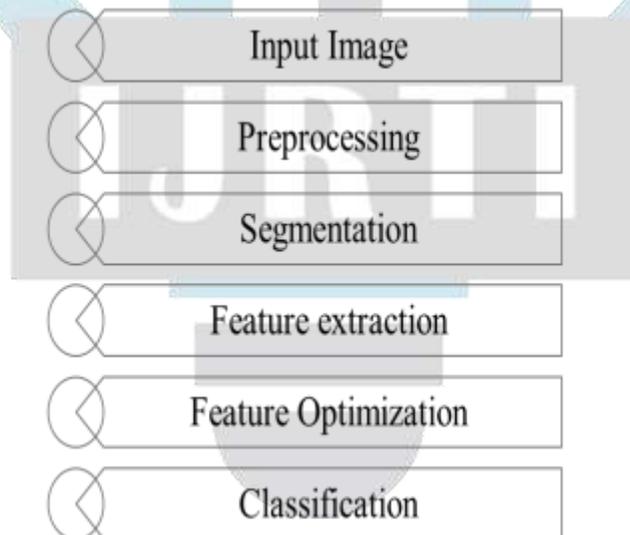
7.Conclusion:

The proposed plant species identification system using Python and machine learning offers a powerful, efficient, and accessible solution for accurate plant classification. By leveraging deep learning techniques, especially convolutional neural networks (CNNs), the system can analyze plant images and predict species with high precision.

This approach not only reduces human error and speeds up the identification process but also opens up new possibilities for large-scale biodiversity monitoring, agricultural automation, and educational tools. Its flexibility allows it to be deployed across web and mobile platforms, making it usable in both field and academic settings. As machine learning models continue to improve and more plant image datasets become available, this system can evolve to support more species, integrate with real-time sensing technologies, and contribute significantly to fields like botany, ecology, and smart farming.

8.Future work:

In the future, plant species identification using Python and machine learning can explore several promising directions to improve accuracy, scalability, and usability. One key area is the enhancement of data collection and annotation, including the integration of high-quality, multi-modal datasets that combine images of leaves, flowers, stems, and fruits with metadata such as GPS coordinates and environmental conditions. Advanced machine learning techniques, such as transformer-based models (e.g., Vision Transformers), few-shot learning, and semi-supervised learning, can be leveraged to improve model performance, especially in recognizing rare or unseen species.

Furthermore, deploying lightweight models optimized for mobile and edge devices would make plant identification accessible in real-time and in offline settings, which is crucial for use in remote field locations. Domain adaptation techniques and transfer learning can further enhance generalization to field images that differ from lab-collected data. Lastly, developing interactive, educational tools using frameworks like Streamlit or Flask, and establishing robust evaluation metrics and benchmarks, would contribute significantly to both scientific research and public engagement in plant biodiversity.

9.References:

[1]Surleen Kaur , Prabhpreet Kaur,"Plant Species Identification based on Plant Leaf Using Computer Vision and Machine Learning Techniques," Journal of Multimedia Information System VOL. 6, NO. 2, June 2019 (pp. 49-60): ISSN 2383-7632

[2] S. H. Lee, C. S. Chan, P. Wilkin, P. Remagnino, "Deepplant: Plant identification with convolutional neural networks," in Proceedings of the 2015 IEEE International Conference on Image Processing (ICIP), pp. 452-456, 2015

[3] J. Wäldchen, P. Mäder, "Plant species identification using computer vision techniques: A systematic literature review," Archives of Computational Methods in Engineering, vol. 25, no. 2, pp. 507-543, 2018.

[4] J. Wäldchen, M. Rzanny, M. Seeland, P. Mäder, "Automated plant species identification—Trends and future directions," PLoS computational biology, vol.14, no. 4, pp. 1-19, 2018.

[5] D. Shreya, "How many plant species are there in the world? Scientists now have an answer," MONGABAY NEWS AND INSPIRATION FROM NATURE'S FRONTLINE, May 12, 2016, news.mongabay.com/2016/05/many-plants-world-scientists-may-now-answer/

[6]Kumar, N., Belhumeur, P. N., Biswas, A., Jacobs, D. W., Kress, W. J., Lopez, I. C., & Soares, J. V. B. (2012). Leafsnap: A Computer Vision System for Automatic Plant Species Identification. European Conference on Computer Vision.

[7]Lee, S. H., Chan, C. S., Mayo, S. J., & Remagnino, P. (2017). Deep-plant: Plant identification with convolutional neural networks. Neurocomputing, 222, 131–141.

[8]Carranza-Rojas, J., Goeau, H., Bonnet, P., Mata-Montero, E., & Joly, A. (2017). Going deeper in the automated identification of Herbarium specimens. BMC Evolutionary Biology, 17(1), 181.

[9]S. G. Wu, F. S. Bao, E. Y. Xu, Y.-X. Wang, Y.-F.Chang, Q.-L. Xiang, "A leaf recognition algorithm for plant classification using probabilistic neural network," In Proceedings of 2007 IEEE International Symposium on Signal Processing and Information Te, pp. 11-16, 2007.
[10] D. J. Bora, "Importance of image enhancement techniques in color image segmentation: a comprehensive and comparative study," arXiv preprint arXiv:1708.05081, 2017.

[11] S. Lee, K. Kim, and Y. Kim,
"Deep neural networks for plant species classification using leaf images,"
International Journal of Computer Vision and Signal Processing, vol. 7, no. 1, pp. 1–6, 2017.

[12] A. Rehman, M. S. Khan, M. Tariq, and S. Abbas,
"Plant species identification using deep convolutional neural networks,"in International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), 2019.

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton,
"ImageNet classification with deep convolutional neural networks,"in Advances in Neural Information Processing Systems (NeurIPS), vol. 25, pp. 1097–1105, 2012.

[14] K. Simonyan and A. Zisserman,

"Very deep convolutional networks for large-scale image recognition,"arXiv preprint arXiv:1409.1556, 2014.

[15] K. He, X. Zhang, S. Ren, and J. Sun,

"Deep residual learning for image recognition,"

in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), pp. 770–778, 2016.

[16] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi,

"Inception-v4, Inception-ResNet and the impact of residual connections on learning,"

in Proc. AAAI Conf. Artificial Intelligence, vol. 31, no. 1, pp. 4278–4284, 2017.