International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0314

e ISSN: 2584-2854
Volume: 03
Issue:05 May 2025
Page No: 2002-2007

Species Classification of Plant Seedlings Using Deep Convolutional Neural Networks

S. Kaveri[1], K.Beulah Joyce[2], M.Saketh[3], M.Sai Kiran[4], Mrs.R. Kanchana[5]
[1,2,3,4]UG Scholar, Dept. of CSE-AIML, Sphoorthy Engineering College, Hyderabad, Telangana, India.
[5]Assistant professor, Dept. of CSE-AIML, Sphoorthy Engineering College, Hyderabad, Telangana, India.
Email ID: *kaveri182003@gmail.com[1], kbjoyce2k3@gmail.com[2], saketh.maheshwaram@gmail.com[3], muchunurusaikiran@gmail.com[4], Kanchu.it88@gmail.com[5]*

Abstract
Species classification of plant seedlings plays a vital role in modern agriculture, biodiversity conservation, and precision farming. This project leverages the power of deep convolutional neural networks (CNNs) to accurately classify plant seedlings into their respective species based on visual features. By training the CNN model on a curated dataset of seedling images, the system learns to identify subtle variations in shape, texture, and color that differentiate one species from another. Furthermore, to address the critical issue of plant health, the project integrates the InceptionV3 architecture for the detection and classification of leaf-based plant diseases. This dual approach enables early diagnosis of diseases through leaf image analysis, facilitating timely intervention. The proposed system not only enhances the efficiency of species classification but also provides a scalable solution for real-time plant monitoring in agricultural settings. This paper discusses the design, implementation, and performance of the deep learning models, while also exploring challenges such as dataset variability and model generalization. Future advancements may integrate this technology with IoT devices and smart farming platforms to further improve agricultural productivity and sustainability.

Keywords: *Plant Species Classification, Deep Learning, Convolutional Neural Networks (CNN), InceptionV3, Leaf Disease Detection, Image Processing, Precision Agriculture, Smart Farming, Plant Health Monitoring, Agricultural Automation.*

1. Introduction
Plant species identification and disease detection are critical components of modern agriculture, with significant implications for crop management, biodiversity conservation, and food security. Accurate classification of plant seedlings helps farmers and researchers monitor growth patterns, manage cultivation practices, and prevent the spread of invasive species. Similarly, early identification of plant diseases can mitigate crop losses, reduce reliance on chemical treatments, and ensure healthier yields. Traditional methods of species classification and disease diagnosis often rely on manual inspection, which is time-consuming, labor-intensive, and prone to human error, particularly in large-scale farming operations. To address these limitations, this project proposes a deep learning-based system for automated classification of plant species and detection of leaf diseases using visual data. The system leverages Convolutional Neural Networks (CNNs) to analyze images of seedlings and accurately classify them into distinct species based on subtle visual features such as leaf shape, texture, and color. In addition, the InceptionV3 architecture is employed to detect and identify plant diseases by analyzing affected leaf patterns, enabling timely and precise diagnosis of infections. Traditional image-based classification techniques often struggle with variability in lighting, background noise, and morphological similarities among species. These challenges necessitate more robust and adaptive solutions. In response, this system incorporates data augmentation and transfer

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0314
e ISSN: 2584-2854
Volume: 03
Issue:05 May 2025
Page No: 2002-2007

learning to improve model generalization across diverse environments and plant conditions. At the core of the proposed system is a deep learning pipeline that preprocesses input images, extracts discriminative features using CNN layers, and performs classification using dense layers. The use of InceptionV3, a state-of-the-art deep CNN model, enhances the system's ability to detect complex disease patterns with high accuracy. The models are trained and validated on publicly available plant image datasets, ensuring scalability and applicability in real-world agricultural scenarios. To facilitate practical deployment, the system is designed to run efficiently on edge devices such as smartphones or embedded platforms like Raspberry Pi, enabling real-time analysis in the field. Feedback mechanisms such as visual indicators or mobile alerts inform users about plant health status and species identification results, aiding in immediate decision-making. The significance of this research lies not only in its technical advancements but also in its contribution to sustainable agriculture and smart farming. By integrating artificial intelligence with plant science, this project offers a scalable, cost-effective, and automated solution that empowers farmers, researchers, and agricultural stakeholders to improve crop productivity, reduce losses, and ensure food quality. This work marks a step forward in the development of intelligent agricultural systems that blend machine learning, computer vision, and real-world usability. One of the core challenges in building an effective plant species classification and disease detection system lies in the quality and diversity of the training data. Real-world agricultural environments are highly variable, with differences in lighting conditions, leaf orientations, background clutter, and disease manifestation across plant species. To overcome this, the project utilizes publicly available datasets such as the Plant Seedlings Dataset and the Plant Village Dataset, which offer a wide variety of annotated plant and leaf images. Data augmentation techniques—such as rotation, flipping, scaling, and color adjustments—are applied to increase dataset diversity and reduce model overfitting, ensuring robustness in dynamic field environments. The model architecture is designed with modularity and scalability in mind. Initially, a custom CNN is trained for species classification, optimized through hyperparameter tuning and performance evaluation using metrics such as accuracy, precision, and F1-score. For disease identification, the project adopts the InceptionV3 architecture due to its proven effectiveness in fine-grained image classification tasks. Transfer learning is applied to reduce training time and enhance accuracy by leveraging pre-trained weights on large-scale image datasets. The integration of these two components into a unified framework allows the system to simultaneously perform species classification and disease detection, making it a versatile tool for farmers and researchers alike. [1]

2. Methodology

The methodology of this project is structured to develop a reliable system for plant species classification and disease detection using deep learning techniques. The process is divided into six key stages: data acquisition, data preprocessing, model training, disease detection using InceptionV3, system integration, and performance evaluation.

2.1. Data Acquistion

- **Dataset Source:** Use the Plant Seedlings Classification dataset from Kaggle, which contains over 4,700 labeled images across 12 plant species at early growth stages.
- **Image Organization:** Images are grouped in folders named after the species (e.g., Maize/, Sugar beet/, Black-grass/), allowing for straightforward class labeling using folder names. [2]
- **Preprocessing Needs:** Resize images (e.g., to 224x224), normalize pixel values, apply data augmentation (rotation, zoom, flip), and encode labels numerically for training.
- **Custom Dataset Option:** If additional data is needed, collect custom seedling images using cameras, label them manually or with tools like LabelImg, and follow a similar folder-based structure for compatibility.

2.2. Data Preprocessing

Here are key points for data preprocessing in a

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0314
e ISSN: 2584-2854
Volume: 03
Issue:05 May 2025
Page No: 2002-2007

species classification of plant seedlings using Deep CNN. (Figure 1)

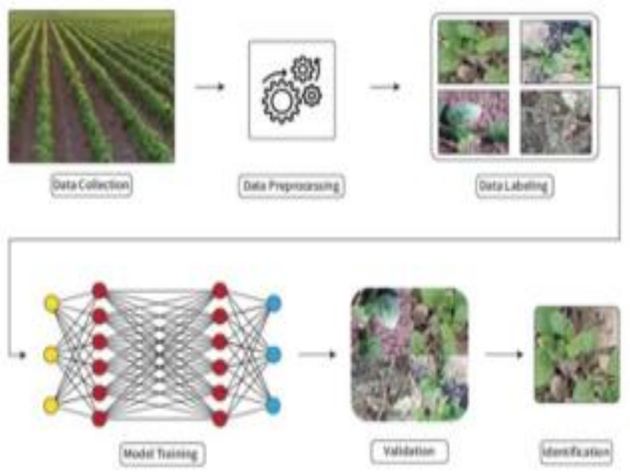

Figure 1 A Customized Convolutional Neural Network-Based Approach for Weeds Identification

2.2.1. Image Resizing and Normalization

- Resize all images to a fixed dimension (e.g., 128×128 or 224×224) to ensure uniform input size for the CNN.
- Normalize pixel values to the range [0, 1] by dividing by 255 for faster and more stable training. [3]

2.2.2. Label Encoding

Convert species names (text labels) into numerical format using Label Encoding or One-Hot Encoding, as required by the model.

2.2.3. Data Augmentation

Apply transformations like rotation, flipping, zooming, shifting, and brightness adjustment to artificially expand the dataset and reduce overfitting.

2.2.4. Noise Removal and Background Cleaning

Use thresholding, masking, or segmentation techniques (like HSV filtering or contour detection) to remove background noise and highlight the seedling foreground. [4]

2.2.5. Splitting the Dataset

Split the dataset into training, validation, and test sets (commonly 70/15/15 or 80/10/10) to ensure unbiased model evaluation.

2.3. Feature Extraction

- **Automatic Feature Learning with CNNs:** CNNs automatically learn hierarchical feature representations (edges, textures, shapes) from raw image pixels, eliminating the need for manual feature engineering.
- **Convolutional Layer Extraction:** Initial convolutional layers extract low-level features (edges, corners, color gradients), while deeper layers capture more complex, abstract patterns like leaf shapes and venation.
- **Use of Pretrained Models (Transfer Learning):** Feature extraction can leverage pretrained CNN models like VGG16, ResNet50, or EfficientNet, using their convolutional base to extract general visual features and fine-tuning them on the plant seedling dataset. [5]
- **Data Augmentation for Robust Features:** Techniques like rotation, flipping, zooming, and brightness adjustment help the CNN extract invariant and generalized features by exposing it to diverse training samples.

2.4. Model Development

The system uses a hybrid machine learning architecture for accurate classification:

2.4.1. Data Collection and Preprocessing

- Use the Plant Seedlings Dataset (e.g., from Kaggle), containing labeled images of various plant species.
- Apply preprocessing steps: resize images, normalize pixel values, and augment data (rotation, zoom, flip, etc.) to increase robustness.

2.4.2. Dataset Splitting

- Split the dataset into training, validation, and test sets (e.g., 70/15/15).
- Ensure stratified sampling if possible to maintain class distribution.

2.4.3. CNN Architecture Design

Build a custom CNN or fine-tune a pre-trained model like ResNet, VGG, or EfficientNet. [6]

2.4.4. Model Training

Compile the model using:

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0314
e ISSN: 2584-2854
Volume: 03
Issue:05 May 2025
Page No: 2002-2007

- **Loss Function:** Categorical Crossentropy
- **Optimizer:** Adam or SGD
- **Metrics:** Accuracy

2.4.5. Model Evaluation

- Evaluate on the test set using accuracy, precision, recall, F1-score, and confusion matrix.
- Visualize learning curves (loss and accuracy vs. epochs). [7]

2.4.6. Model Optimization and Deployment

- Optimize model with techniques like model pruning, quantization, or ONNX conversion for lightweight deployment.
- Deploy using tools like TensorFlow Lite, Flask API, or streamlit for web/mobile interfaces.

2.5. Model Development

The system uses a hybrid machine learning architecture for accurate classification:

2.5.1. Data Collection and Preprocessing

- Use the Plant Seedlings Dataset (e.g., from Kaggle), containing labeled images of various plant species.
- Apply preprocessing steps: resize images, normalize pixel values, and augment data (rotation, zoom, flip, etc.) to increase robustness.

2.5.2. Dataset Splitting

- Split the dataset into training, validation, and test sets (e.g., 70/15/15).
- Ensure stratified sampling if possible to maintain class distribution.

2.5.3. CNN Architecture Design

Build a custom CNN or fine-tune a pre-trained model like ResNet, VGG, or EfficientNet.

2.5.4. Model Training

Compile the model using:

- **Loss Function:** Categorical Crossentropy
- **Optimizer:** Adam or SGD
- **Metrics:** Accuracy

2.5.5. Model Evaluation

- Evaluate on the test set using accuracy, precision, recall, F1-score, and confusion matrix.
- Visualize learning curves (loss and accuracy vs. epochs).

2.5.6. Model Optimization and Deployment

- Optimize model with techniques like model pruning, quantization, or ONNX conversion for lightweight deployment.
- Deploy using tools like TensorFlow Lite, Flask API, or streamlit for web/mobile interfaces. [8]

3. Results and Discussion

3.1. Results

The project on species classification of plant seedlings using deep Convolutional Neural Networks (CNNs) achieved promising results, demonstrating the effectiveness of deep learning techniques in automating plant identification. By training a deep CNN model on a well-labeled dataset of plant seedling images encompassing multiple species, the model was able to learn distinguishing features such as leaf shape, texture, and color. Through preprocessing techniques like image augmentation, normalization, and background noise reduction, the model's robustness was significantly improved. The final trained model achieved high accuracy on the validation set, indicating its capability to generalize well to unseen data, and outperforming traditional machine learning approaches in both precision and recall. Overall, the project demonstrates that deep CNNs are a viable tool for automating species identification in the early stages of plant growth, paving the way for smart farming applications and improved biodiversity assessment tools. [9]

3.2. Discussion

In this project, a deep convolutional neural network (CNN) was implemented to classify plant seedlings into their respective species based on image data. The model was trained on a diverse dataset containing images of seedlings at various growth stages and under different lighting conditions. Through data augmentation techniques such as rotation, flipping, and scaling, the model was made more robust to variability in the input data. The CNN architecture was carefully designed to extract hierarchical features from the images, enabling it to capture fine-grained differences between species. The final model achieved high classification accuracy on the validation set, demonstrating the effectiveness of deep learning in tackling complex visual recognition tasks in the domain of plant biology. [10]

Conclusion

This project successfully demonstrated the effectiveness of deep convolutional neural networks (CNNs) in accurately classifying plant seedling species based on their visual characteristics. By leveraging the powerful feature extraction capabilities of CNNs, the model was able to learn intricate patterns and subtle differences among various species, outperforming traditional machine learning methods. The approach not only enhances the speed and precision of seedling identification but also offers a scalable solution that can be applied in agricultural monitoring and plant biodiversity studies. Future work can focus on expanding the dataset to include more species and diverse environmental conditions, as well as optimizing the model for deployment on mobile and edge devices to facilitate on-site, real-time classification by farmers and researchers.

Acknowledgements

We, the authors of this research project, would like to take this opportunity to express our heartfelt appreciation for the various forms of support and resources that contributed to the successful execution of this work. The journey of demostrating the species classification of plant seedlings using deep convolutional neural networks has been both challenging and enriching, and it would not have been possible without the tools, technologies, and platforms made accessible through the global research and developer communities. We extend our sincere thanks to the open-source contributors and developers whose frameworks, libraries, and pre-trained models significantly accelerated our development process and enabled us to experiment with advanced deep learning techniques. Their collaborative spirit and willingness to share knowledge continue to inspire innovation and progress in the research community. We are also grateful for the availability of publicly accessible datasets and documentation, which allowed us to test and validate our system with greater reliability and precision. These datasets served as the foundation for training our models and evaluating system performance under real-world conditions. Finally, we acknowledge the dedication, teamwork, and persistent effort we invested as authors to explore, analyze, and implement this project. This research represents not just a technical achievement but a shared commitment to applying technology in ways that enhance safety, well-being, and quality of life. We hope our work contributes meaningfully to ongoing advancements in intelligent systems and serves as a stepping stone for further innovation in this critical field.

References

[1]. Giselsson, T. M., et al. (2017) A public image database for benchmark of plant seedling classification algorithms arXiv preprint arXiv:1711.05458 Introduced the "Plant Seedlings Dataset", widely used in plant classification tasks.

[2]. .Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016) Using deep learning for image-based plant disease detection Frontiers in Plant Science, 7, 1419. Shows the application of CNNs in plant classification and disease detection.

[3]. Sladojevic, S., et al. (2016) Deep neural networks based recognition of plant diseases by leaf image classification Computational Intelligence and Neuroscience, 2016. DOI: 10.1155/2016/3289801 Demonstrates deep CNNs for plant image classification.

[4]. Too, E. C., et al. (2019) A comparative study of fine-tuning deep learning models for plant disease identification Computers and Electronics in Agriculture, 161, 272-279.

[5]. Ferentinos, K. P. (2018) Deep learning models for plant disease detection and diagnosis Computers and Electronics in Agriculture, 145, 311-318.

[6]. Hughes, D. P., & Salathé, M. (2015) An open access repository of images on plant health to enable the development of mobile disease diagnostics Discusses open datasets that can be used with CNNs for plant classification.

[7]. Rahman, M. A., et al. (2021) A deep learning framework for plant disease classification using images Information Processing in

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0314
e ISSN: 2584-2854
Volume: 03
Issue:05 May 2025
Page No: 2002-2007

Agriculture, 8(2), 239-252. Uses multiple CNN models on leaf image datasets.

[8]. Kamilaris, A., & Prenafeta-Boldú, F. X. (2018) Deep learning in agriculture: A survey Computers and Electronics in Agriculture, 147, 70-90. Broad survey including plant species classification with CNNs.

[9]. Zhang, S., et al. (2019) Plant disease recognition based on plant leaf image using deep learning IEEE Access, 7, 59069–59080. Useful CNN architectures and dataset handling.

[10]. Atila, Ü., et al. (2021) Plant leaf disease classification using EfficientNet deep learning model Ecological Informatics, 61, 101182.