



Plant Species Identification Using Convolutional Neural Networks with Principal Component Analysis: An Interpretable Deep Learning Approach

Bhavay Jain¹, Gourav Tyagi², Shivam Garg³, Ms.Anisha⁴

Department of CSE-AI&DS,

Panipat Institute of Engineering and Technology, Samalkha, India

Mail ID: anisha.cse.ai-ds@piet.co.in

Abstract— Accurate identification of plant species plays a critical role in various domains such as agriculture, botany, environmental conservation, and medicine. Traditionally, plant identification relies heavily on expert knowledge and manual analysis, which can be time-consuming and error-prone. This project presents an automated system for plant species identification using advanced image processing and machine learning techniques. The proposed system leverages a deep learning-based Convolutional Neural Network (CNN) architecture to extract meaningful features from plant images, primarily focusing on leaves, flowers, and other identifiable parts. These features are then used to classify the plant species with high accuracy. The system has been trained and tested on a diverse dataset containing images of various plant species, addressing challenges like intra-species variability, similar morphological traits among species, and varying lighting conditions. To enhance performance and usability, techniques such as data augmentation, transfer learning using pre-trained models like MobileNet and ResNet, and optimization algorithms have been employed, and an unsupervised Gaussian Process Latent Variable Model combined with Principal Component Analysis (PCA) is used to make the CNN's decisions more biologically interpretable. This project aims to provide an accessible tool for researchers, students, farmers, and nature enthusiasts, contributing to biodiversity awareness and sustainable ecological practices.

Keywords: Plant Species Identification; Convolutional Neural Network; Principal Component Analysis; Gaussian Process Latent Variable Model; Deep Learning; Interpretable AI.

INTRODUCTION

A. Background of the Study

Artificial intelligence (AI) is mainly designed to make machines capable of thinking and acting like humans, and machine learning is a subsection of artificial intelligence. Deep learning delivers outstanding results even on big data, and is used in image recognition and computer vision through artificial neurons identical to human neurons. Plants are an important member of all natural life[1], and their official naming ensures that natural life is preserved and maintained; plants are essential for medicinal purposes, alternative energy sources such as biofuels, and household needs such as wood, clothing, food, and trapping. The current extinction trend is mainly the product of indirect and overt human activity, so generation of correct information on the geographical spread of plant species is crucial for the survival of future ecosystems.



Image recognition strategies have recently begun to emerge to simplify the plant inspection process. Adjustments in lighting, leaf movement, camera shaking, focus changes, and unexpected changes in camera settings all lead to incorrect estimates of plant species, so identifying plant species dependent on digital photographs remains a challenging problem. Multiple researchers have established the ability of a broad convolutional neural network (CNN) to outperform conventional object recognition strategies centred upon ordinary working light, texture, and shape characteristics[2]. With the widespread use of smartphones and applications such as PlantNet, millions of photographs of plants have been collected, and automated plant identification on mobile devices is essential for real-world ecological monitoring and tracking invasive alien plant species.

B. Statement of the Problem

Traditional classification methods depend mostly on preprocessing to eliminate complicated background information and increase the effectiveness of the classification model; manual feature engineering is incapable of dealing with large-scale data sets comprised of unrestricted pictures. Mobile-based plant photographs taken in natural environments vary greatly in terms of contributors, cameras, areas, and time of year, as well as individual plants and other characteristics, so existing models are still far from meeting the needs of a scenario with completely automated ecological monitoring. With the goal of overcoming these difficulties and being inspired by recent achievements in deep learning picture identification, this study recommends that Convolutional Neural Networks be used to create a plant recognition approach.

C. Aim and Objectives of the Study

The main objective of the study is the implementation of plant identification using machine learning and deep learning. Specifically, the study aims to explain the intuitive reasons for classification using CNN as a deep learning model based on a sample; to compare prediction accuracy using 5-fold cross-validation in 10 replicates using reshuffled data sets; and to compare model uncertainty using mutual information to get more details about the prediction process, while the error rate of the two approaches is compared with CNN using various parameters.

LITERATURE REVIEW

A. Concept of Plant Identification Using Deep/Machine Learning

Linear models in records, including straight or calculated regression, have the advantage of being white-box models in which the informative elements can be explored easily. Similarly, linear models such as Eigenfaces, relying on principal component analysis (PCA), have been utilized in computer vision; however, they suffer from a strong linear assumption, and this interpretability is not available for deep machine learning models used for image analysis. Explanatory strategies for such models include exploring learned representations, describing individual predictions, describing model behaviour, and reviewing explanatory elements based on representative examples, including hierarchical relevance propagation and GradCAM as ways to visualize the reasons for individual CNN predictions.

B. Shape and Surface-Based Plant Recognition

Earlier studies recognized diseases using geometric and histogram-based features extracted from segmented leaf regions with SVM classifiers; other researchers identified soybean diseases using colour and texture features, or used feed-forward neural networks with back-propagation to identify plant leaves and their diseases. A bacterial-foraging-optimization-based radial basis function neural network (BRBFNN) has also been used for identification and classification of plant leaf diseases, achieving rapid growth and improved accuracy through region-growing segmentation criteria[4].



C. Plant Identification Based on Deep Learning

Researchers have used AlexNet and GoogleNet CNN architectures to identify distinct plant diseases, while other studies employed a variety of CNN architectures to identify fifty-eight distinct plant diseases with high classification accuracy, examining the CNN architecture in real time. A nine-layer deep CNN trained on the PlantVillage dataset with data-augmentation techniques achieved a greater degree of precision than is typically associated with conventional machine learning[3]. Other researchers suggested a CNN architecture to recognize plant shapes from image sequences acquired at smart agro-stations, and transfer-learning models such as GoogleNet, VGGNet, and AlexNet[5] have also been evaluated for plant identification tasks using the LifeCLEF benchmark[6].

TABLE I. SUMMARY OF RELATED WORK ON CNN-BASED PLANT / LEAF CLASSIFICATION

Study	Technique(s) Used	Reported Result
Fuentes et al. (2017)	Region-based (Faster R-CNN / SSD) and reinforced convolutional neural networks	Accuracy rate of 83% and a precision rate of 99.49%
Ferentinos et al. (2018)	Multiple CNN architectures (AlexNetOWTBn, VGG) evaluated in real time	AlexNetOWTBn reached 99.53%, of which VGG reached 99.53% class accuracy
Ramcharan et al. (2019)	MobileNet as the location and classification backbone for an SSD model	80.6% precision on still pictures; 70.4% precision on video
Geetharamani & Pandian (2019)	A nine-layer deep convolutional neural network trained on PlantVillage	96.75% precision after data augmentation
Chen et al. (2020)	INC-VGGN transfer-learning architecture	96.75% precision
Li et al. (2020)	Deep CNN paired with SVM and RFC as shallow-learning classifiers	96.75% precision

METHODOLOGY AND SYSTEM DESIGN

This study's model evaluation is based on image data from various plant species from the kingdom Plantae. Different models were used to process the photos; the following subsections go over the data sets, processing pipelines for images, and the proposed deep learning technique.

A. Area of Study and Dataset

Images of plant species from photosynthetic eukaryotes were acquired and used in this project from the Swedish Botanic Garden. The dataset added up to 10,031 data sheets of 15 different plant examples, obtained mostly between 1996 and 2010, using a Canon EOS 5D Mark II digital SLR camera paired with an EF 16-35 L USM II focal lens. Imaging and feature extraction were carried out using ImageJ, an open-



source picture-handling tool, and VS Code. A selection of background photos, including other plants, dirt, and stones, was randomly included in the dataset in addition to the extracted plants, and a squared bounding box was used to retrieve all samples.

TABLE II. DATA CHARACTERISTICS

Plant Species	No. of Images
Ulmus carpinifolia	668
Acer	667
Salix aurita	669
Quercus	668
Alnus incana	667
Betula pubescens	665
Salix alba 'Sericea'	670
Populus tremula	668
Other 7 species (Ulmus glabra, Sorbus aucuparia, Salix sinerea, Populus, Tilia, Sorbus intermedia, Fagus silvatica)	~670 each

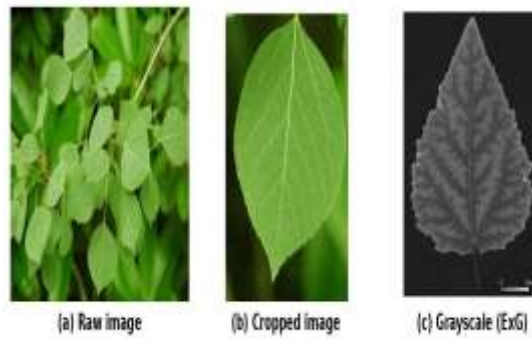


Fig. 1. Image pre-processing pipeline: (a) raw image, (b) cropped image, and (c) grayscale image after excess-green transformation.

Following the suggestions of Milioto et al. (2018), the RGB pictures were converted into excess-green images and then into grayscale by applying Equation (1):

$$I_{ExG} = 2.0 \cdot I_G - I_R - I_B \quad (1)$$

where I_R , I_G and I_B represent the RGB picture channels. In I_{ExG} , terrain and stones appear black while green pixels appear bright; this transformation was executed in OpenCV.



B. Gaussian Process Latent Variable Model

An unsupervised Bayesian Gaussian Process Latent Variable Model (B-GP-LVM)[9] was used to obtain an interpretable feature-extraction baseline, incorporating an automatic relevance determination kernel and variational inference. The GP-LVM addresses the challenge of reducing the D-dimensional matrix $Y \in \mathbb{R}^{(N \times D)}$ to $X \in \mathbb{R}^{(N \times d)}$ using the probabilistic mapping shown in Equation (2):

$$p(Y|X, \theta) = \prod_n N(Y[n, :] | 0, K_{NN} + \beta^{-1} I_N) \quad (2)$$

The kernel matrix therefore influences a set of N Gaussian Processes mapping from a latent feature space to the original image space, using a radial basis function kernel. To access a first estimate of the latent-space size, principal component analysis was used to examine the image vector-matrix Y, with the number of principal components explaining 78% of the variance used to approximate d, the number of latent plant characteristics.

C. Convolutional Neural Network with PCA

PCA[8] is a dimensionality-reduction procedure that transforms a set of features in a dataset into fewer elements called principal components while attempting to retain as much information from the original dataset as possible. CNN, also known as ConvNet, works on the basis of several hidden layers partitioned into convolutional, pooling, and fully-connected layers, capturing different elements of images through a series of filters. A fully-connected layer of 1024 neurons with a ReLU activation function was added to the VGG16 [11] model used in this investigation, followed by two neurons in a softmax output layer and a dropout layer with a 50% dropout rate, initialized using ImageNet weights.

Keras data augmentation was employed with the following parameters: rescale = 1/255, rotation range = 20, width shift range = 0.2, height shift range = 0.2, horizontal flip = True, and fill mode = 'nearest'. Categorical cross-entropy was used as the loss function, optimized using the Keras RMSProp optimizer with a batch size of 8, 50 epochs, and a learning rate of 10^{-4} . Images were resized and reshaped to [64×64×3] pixels prior to training.

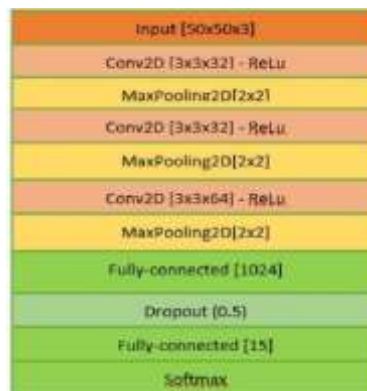


Fig. 2. The CNN architecture used for feature extraction and classification.

D. Model Evaluation Using Statistical Methods

The layer-wise relevance propagation (LRP) algorithm[10], together with latent representations of input characteristics from the GP-LVM, were used to code the significance of input characteristics in the



decision-making process of the machine learning algorithms studied. LRP, GradCAM, and GP-LVM variance characteristics are typically displayed as saliency maps in image analysis; relevant elements emerge in saliency maps as connected zones with positive metric values. A statistical test was used for pixel-wise, p-value-based thresholding of the saliency maps, with $N = 10^4$ randomisation repetitions, a Gaussian smoothing kernel of size $k = 5$, and a significance threshold of $\alpha = 0.001$ used to reliably and repeatably determine the positions of essential saliency-map features.

For categorization, the evaluated characteristics of the convolutional autoencoder and the B-GP-LVM were used together with N support vector machines using an RBF kernel. In a manner comparable to the CNN pipeline, ten iterations of five-fold cross-validation based on reshuffled data sets were used, with a Bayesian optimization approach used to optimize the hyperparameters of the kernel function and the SVM based on 10 initialization and 10 optimization points.

RESULTS AND DISCUSSION

A. Model Processing

Gaussian processes (GPs) are useful for image manipulation to obtain more accurate results, providing a Bayesian nonparametric framework for inference of highly nonlinear latent functions from observable data. The image dataset was of different dimensions and scales, so each image was resized to get a better image process and to produce a better result. The results obtained were contrasted with various assessment metrics to show how efficient the model is; each model utilized a comparative area of the same dataset.

B. Model Performance

The model was split in an 80-to-20 test-split ratio after several trials, using the full set of 10,031 images of 15 different plant species. The experiment was performed first using a plain CNN, after which PCA was applied to obtain better accuracy. The ReLU activation function and the Adam optimizer were used, with ReLU providing a faster training approach than alternative activation functions; a cross-entropy loss function was applied to minimize prediction loss. Fig. 3 shows the activation maps of the data resulting from the initialization of random weights, depicting exactly how each filter works after implementation; for the PCA and CNN combination, only the convolutional weights are changed, while other settings remain the same.

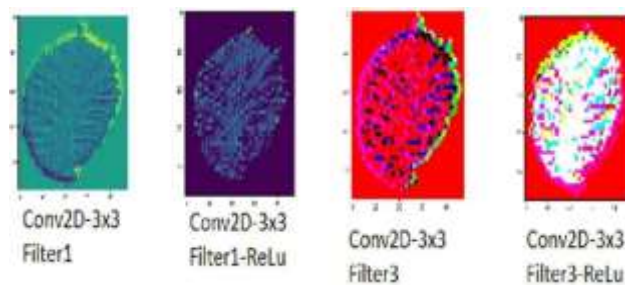


Fig. 3. Activation maps of the leaf data across successive convolutional filters.

Fig. 4(a) shows the reshaped size of the data after applying PCA, where the PCA weights are reshaped to give basic image patches; utilizing all the patches captures up to 100% of the variance. Fig. 4(b) shows



the reduced number of patches required to capture up to 90% of the variance. After various experiments varying the number of filters and kernels used, the application of PCA for optimizing the results shows how the features are correlated, with colour intensity indicating how close the images are and providing insight into more appropriate weight initialization than random weights.

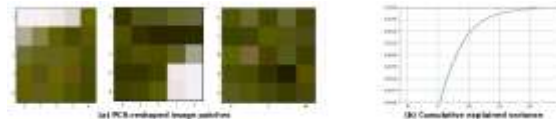


Fig. 4. (a) PCA-reshaped image patches and (b) cumulative explained variance captured against the number of principal-component patches.

The prediction accuracy is one solid parameter to show how accurate the model is. The accuracy of the model with PCA was recorded at 0.8985, the loss function at 0.4878, F1-score at 0.92, recall at 0.73, and precision at 0.86, as summarized in Table III. The research also found biologically incomprehensible justifications for the CNN's decisions, since these uninterpretable regions were used to accomplish high prediction accuracy during training; the discoveries show that pixels from different plants are used to accurately recognize the picture data, and the CNN's accuracy is based on visual areas associated with picture labelling and identification.

TABLE III. CLASSIFICATION PERFORMANCE METRICS

Metric	Plain CNN	CNN + PCA	Improvement
Test Accuracy	0.85 (approx.)	0.8985	Higher
Test Loss	0.55 (approx.)	0.4878	Lower
F1-Score	-	0.92	-
Recall	-	0.73	-
Precision	-	0.86	-



Fig. 5. Training accuracy of the plain CNN versus the CNN with PCA over ten epochs.



C. Discussion of Results

The study's findings show that PCA combined with CNN outperforms other contrasted techniques in prediction, identification accuracy, and F1-score, though there was no significant variation in overall CNN prediction performance for the dataset; the difference in error-rate prediction was considered related to leaf structure and the choice of CNN architecture. The study suggests that the higher accuracy of PCA with CNN is due to the 'Clever-Hans' mode of predictors, whereby image areas related to the image label account for high prediction accuracy rather than genuinely discriminative biological features; this effect was identified when pixels near the plant's real location were used to categorize it.

Unsupervised techniques are known to extract image regions that are interpretable, unlike CNN, which requires a per-sample model explanation; statistically viable properties for these models were discovered using Spearman's rank correlation test. The findings suggest that the PCA-with-CNN architecture is capable of recognizing complex coherences in visual material, though assessing an overall assertion for the hidden explanatory elements remains difficult due to the sample-based decision process. The suggested unsupervised models, on the other hand, generate saliency maps that depict the classification zones, with the explanatory parameters required for categorization known beforehand, allowing unnecessary features to be eliminated.

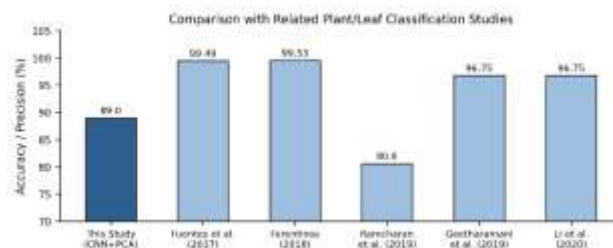


Fig. 6. Accuracy/precision comparison of the proposed model with related plant and leaf classification studies (Table I).

CONCLUSION AND RECOMMENDATIONS

The study utilized convolutional neural networks to classify and identify images based on several features that could not be understood biologically, in order to study plant identification through deep learning. The study found that unsupervised techniques perform comparably to supervised learning-based methodologies while avoiding the black-box model problem. Images of plant species from photosynthetic eukaryotes were acquired and used from the Swedish Botanic Garden to analyze and categorize various plant species; prediction accuracy and mutual information were examined in the experiments, and PCA was used to compare the prediction performance of the suggested unsupervised technique to that of the chosen CNN-based models.

When it comes to accuracy prediction and mutual information, PCA with CNN surpasses the other techniques tested, with accuracy increased to 89%; however, the study finds that CNNs must be used with caution, as they require an efficient learning strategy in order to procure good results. Unsupervised learning-based techniques, in which all employed image areas can be observed before categorizing, were found to be capable of solving the classification problem but performed poorly in terms of prediction accuracy, though the unsupervised process utilizes no elements that are biologically baffling.



A. Recommendations

- **Optimized Architectures:** Traditional CNN structures should be paired with optimization and customization methods to achieve appropriate accuracy, notwithstanding the complexity of real-world crop image datasets.
- **Underrepresented Crops and Diseases:** Crops with significant food and financial representativeness, such as grains and cereals, should be properly examined, and more methods are needed for identifying diseases caused by worms.
- **Multispectral and Hyperspectral Imaging:** Techniques such as multispectral and hyperspectral pictures can lead to novel agricultural solutions that increase productivity while reducing harm caused by biotic and abiotic agents.
- **Real-Time Video Extension:** The framework can be extended to a continuous video-feed system, allowing for unattended plant care and monitoring, alongside a system that treats and analyzes plant components for future research.

REFERENCES

- [1] J. S. Cope, D. Corney, J. Y. Clark, P. Remagnino, and P. Wilkin, "Plant species identification using digital morphometrics: A review," *Expert Systems with Applications*, vol. 39, no. 8, pp. 7562–7573, 2012.
- [2] K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis," *Computers and Electronics in Agriculture*, vol. 145, pp. 311–318, 2018, doi: 10.1016/j.compag.2018.01.009.
- [3] R. Geetharamani and J. A. Pandian, "Identification of plant leaf diseases using a nine-layer deep convolutional neural network," *Computers & Electrical Engineering*, vol. 76, pp. 323–338, 2019, doi: 10.1016/j.compeleceng.2019.04.011.
- [4] A. Fuentes, S. Yoon, S. C. Kim, and D. S. Park, "A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition," *Sensors*, vol. 17, no. 9, p. 2022, 2017, doi: 10.3390/s17092022.
- [5] M. M. Ghazi, B. Yanikoglu, and E. Aptoula, "Plant identification using deep neural networks via optimization of transfer learning parameters," *Neurocomputing*, vol. 235, pp. 228–235, 2017, doi: 10.1016/j.neucom.2017.01.018.
- [6] J. Chen, J. Chen, D. Zhang, Y. Sun, and Y. A. Nanekharan, "Using deep transfer learning for image-based plant disease identification," *Computers and Electronics in Agriculture*, vol. 173, Art. no. 105393, 2020, doi: 10.1016/j.compag.2020.105393.
- [7] A. Ramcharan, K. Baranowski, P. McCloskey, B. Ahmed, J. Legg, and D. P. Hughes, "Deep learning for image-based cassava disease detection," *Frontiers in Plant Science*, vol. 8, p. 1852, 2017, doi: 10.3389/fpls.2017.01852.
- [8] M. E. Tipping and C. M. Bishop, "Probabilistic principal component analysis," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 61, no. 3, pp. 611–622, 1999.
- [9] N. D. Lawrence, "Probabilistic non-linear principal component analysis with Gaussian process latent variable models," *Journal of Machine Learning Research*, vol. 6, pp. 1783–1816, 2005.
- [10] S. Bach, A. Binder, G. Montavon, F. Klauschen, K. R. Müller, and W. Samek, "On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation," *PLoS ONE*, vol. 10, no. 7, Art. no. e0130140, 2015, doi: 10.1371/journal.pone.0130140.
- [11] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. International Conference on Learning Representations (ICLR)*, 2015.