ECTI Transactions on Computer and Information Technology
Journal homepage: https://ph01.tci-thaijo.org/index.php/ecticit/
Published by the ECTI Association, Thailand, ISSN: 2286-9131

AI-Based Smart Identification of Medicinal Plants Using Vision Transformer and CatBoost for Biodiversity and Healthcare

Faisal Firdous[1], Deepak Gupta[2] and Hemant Sood[3]

ABSTRACT

In most countries, medicinal plants are crucial remedies for disease treatment. Even though the majority are edible, ingesting the incorrect herbal plant can have fatal consequences. It is essential to accurately identify these plants not only for safe usage by individuals but also for various real-time applications like aiding biodiversity conservation, supporting farmers in recognizing local herbs, and also preserving indigenous systems. Numerous automatic methods for identifying medicinal plants have been developed; however, most of them are severely limited, either by the relatively small number of plant species they support or by the fact that they rely on manual visual segmentation of plant leaf surfaces. This means that instead of being easily recognized in their natural environments, which frequently include complicated and chaotic backgrounds, they are snapped against a plain background. Deep learning-based techniques have advanced significantly in recent years. Still, they are trained on data that isn't always fully reflective of the intra-class and inter-class variances among the plant species in consideration. The paper approaches this issue by integrating the hybrid model of a pre-trained vision transformer with a CatBoost classifier tuned with Optuna. The vision transformer model is trained with the Indian medicinal plant dataset with the five most commonly used species. The hybrid model is compared with the deep learning models regarding precision, recall, F1-score, accuracy, and execution time on the same dataset. Our proposed model achieves a training phase accuracy of 93%, which shows the improvement for automating the identification of medicinal plants. In conclusion, our proposed hybrid model reveals enhanced accuracy, improved reliability, and reduced false positives in automating the identification of medicinal plants, contributing effectively to healthcare applications and biodiversity.

Article information:

Keywords: Artificial Intelligence, Image Processing, Vision Transformer (ViT), Phenotype Characters, Medicinal Plants

Article history:
Received: August 28, 2025
Revised: October 10, 2025
Accepted: November 6, 2025
Published: December 6, 2025 (Online)

DOI: 10.37936/ecti-cit.2026201.263491

1. INTRODUCTION

Plants provide food, fiber, shelter, fuel, and medicinal materials for human life on earth [1]. Natural products, especially from herbs, are considered safe for the environment and, in most cases, have a better safety profile compared with synthetic drugs. The use of traditional medicinal herbs such as tulsi, neem, alovera, turmeric, and ginger is above geographical boundaries, as their applications have been found all over different regions and are mainly used in India. These traditional medicines have survived until this date by curing many common seasonal health disorders and possess an impressive safety record with no side effects. Medicinal plants in India and worldwide have been playing a highly essential role in traditional medicine, and they are one of the most basic protections for human health [2]. Drawing heavily from botanical sources, traditional medicine forms

[1,2] The authors are with the Computer Science & Engineering and Information Technology, Jaypee University of Information Technology, Waknaghat, Solan 173234, India., Email: faisalparray39@gmail.com and deepak.vd@gmail.com

[3] The author is with the Department of Biotechnology & Bioinformatics, Jaypee University of Information Technology, Waknaghat, Solan 173234, India, Email: hemant76sood@gmail.com

[1] Corresponding author: faisalparray39@gmail.com

an encapsulation of vital protection measures for human life. The use of leaves of medicinal plants goes even beyond medicinal use and reaches into domains that include culinary uses, fragrant purposes, and olfactory and gustatory intensifications in India and worldwide [3].

Plants are of great importance to the future of the human race, as they provide a source of sustenance and oxygen. Several species are used in therapeutics, folk medicine, and the pharmaceutical industry. The National Cancer Institute and World Health Organization have estimated that 80 percent of people worldwide rely on herbal medicines for some part of their primary health care [4].

Ethnobotany is the study of how various cultures use plants. Only 1/4 of the world's estimated 250,000 higher plant species have been studied for their medicinal potential. The obvious conclusion is that we could potentially lose cures for many diseases if plant species become extinct before they are studied. Numerous plant species are becoming extinct, with 123 already classified as extinct and 37 extinct in the wild. Many others are under susceptible categories such as Critically Endangered (3,325 species), Endangered (6,063 species), and Vulnerable (7,072 species). These types of threats arise from various factors such as population decline, reduced reproduction, habitat destruction (both natural and human-induced), loss of pollinators, overexploitation, and diminishing genetic diversity [5, 6]. A striking example is the rosy periwinkle, which is native to Madagascar. The plant has been used in folk medicine for centuries, but its effectiveness in treating certain cancers was discovered only in the 20th century [7]. Currently, this plant is endangered, and it is uncertain whether it can be conserved by following the effective practices.

People's trust in the long-established medicine system, which withstood the test of time due to its affordability [8]. In addition, biotechnological methods are generally utilized to conserve and propagate the medicinal plants, along with the utilization of information from Ayurveda to use herbs by scientific intervention for treating the ailments, which is enough to signify the role of traditional systems of medicine. Today, the global scenario has made it necessary to protect the knowledge of indigenous systems of medicine in various countries, as it directly includes the intellectual property, protection of plant resources, and traditional knowledge of a particular country. This, in turn, has important socioeconomic implications. Trade in medicinal plants and derived products has become an important economic venture in many developing countries. Unfortunately, there has been large-scale exploitation and unconstrained use of medicinal plants for commercial benefits, which has led to the depletion of species and has brought some of them to the verge of extinction [9]. Therefore, effective approaches for monitoring and classifying medicinal plant species are crucial for ensuring their usage and conservation. Field-based and traditional approaches are error-prone, time-consuming, and more often require expert knowledge. Thereby, there is a need for solutions that can assist practitioners, farmers, policymakers, and researchers in early and accurate recognition of medicinal plant species in real time.

However, there are numerous technological and logistical obstacles to the identification of automatic medicinal plant species that require a robust response. Moreover, many similar or different species of the same genus exist or evolve due to natural variation, which may be recorded with the proposed study. Under environmental conditions, many factors influence the development of plants, which include mutations at the genetic or morphological level influenced by biotic and abiotic components in the environment. So, identification of the true plant is a prerequisite for their authentic product development as well as for evaluating their actual population status. Current research focuses on utilizing leaves to identify medicinal plant species. Image processing methods are being used by researchers on a greater scale to identify plants based on pictures of their leaves. The classification and identification of leaves of medicinal plants can be greatly aided by artificial intelligence.

1.1 Contributions of the Paper

The main goal of this paper is to try to identify the Indian medicinal plants with a hybrid model of ViT and CatBoost with Optuna. It also presents the use of transformer models for the identification of plants with better accuracy as compared to baseline deep models. Additionally, it also discusses the benefits of using a hybrid model. The contributions of this paper are highlighted as follows:

i. A dataset has been collected from various online sources for those species that need to be identified for many uses.

ii. A new hybrid model integrating the ViT with CatBoost is developed; the transformer leverages the capabilities of feature extraction and CatBoost for robust classification.

iii. Optuna-based tuning is utilized to optimize CatBoost parameters, which enhances the model's computational efficiency and classification accuracy.

iv. A comparison of various widely used deep learning models along with the proposed hybrid model was performed and rigorously tested on an Indian medicinal plant dataset.

v. Model interpretability is improved using SHAP analysis to identify the most influential features that contribute to accurate medicinal plant identification, which provides deeper insight into decision-making behavior.

The rest of this paper is organized as follows: Section 2 offers a literature survey of the pertinent literature, while Section 3 outlines the suggested methodology. Sections 4 and 5 provide further details on the testing classification report and outcome analysis, respectively. The conclusion and future work are covered in Section 6.

2. LITERATURE SURVEY

An extensive overview of existing research will be offered in this section, which encompasses deep learning, machine learning, and transformer models, with the intent of identifying their limitations, capabilities, and gaps in research and thus adhering to the basis for the proposed method.

P. Singla *et al.* [10] developed a web-based application by using the deep learning model for recognition of medicinal plants and also conveys the alerts to farmers if there is any ailment. Several deep convolutional neural network models were compared to the suggested model. The suggested model demonstrated categorization binary outcomes with 99.39% accuracy, 0.0361 loss, 0.989 precision, and 0.984 recall, in contrast to the deep CNN models.

S. Chulif *et al.* [11] determined the herbarium-field triplet loss network and proposed a model that defines the mapping between the real-world and herbarium domains. The cross-domain plant identification challenges of PlantCLEF 2020 and 2021 are quite similar. In results, it has been proved that the network can differentiate rare species as well as has an ability to generalize without field images. In the test set of the PlantCLEF 2020 and PlantCLEF 2021 species with few training field images, the HFTL network achieved a mean reciprocal rank score of 0.108 and 0.158, respectively.

H. K. Diwedi *et al.* [12] proposed a framework of an enhanced convolutional neural network with transfer learning using upgraded ResNet50. This approach adopted PTL by expanding the ReNet50 framework to deploy for feature extraction. Optimized support vector machine was applied for classification. Based on the Indian medicinal plants database, a list of publicly available medicinal plant species is prepared. The modified ResNet50+OSVM model yielded 96.8% accuracy during testing and 98.5K. Pankaja *et al.* [13] developed a new approach to the organization of plant leaves, which is the recognition of a class through the incorporation of the Whale Optimization Algorithm and Random Forest. The datasets considered were Swedish and Flavia leaf specimens. The results show an accuracy rate as high as 97.58%, along with better efficiency in terms of execution time than previous methodologies.

M. Sharma *et al.* [14] suggested an intelligent approach to leaf recognition of Indian medicinal plants. The input features of the classifier models here are heterogeneous attributes gathered from the leaves of Indian medicinal herbs, and the results indicate that the Random Forest classifier with a hybrid feature vector manages to get a very low PFA of 0.02%, along with precision, accuracy, and sensitivity greater than 99%. As compared to the existing models in use, this model shows outstanding results, with up to 3% performance enhancement, and has made some significant improvements in recognition and categorization accuracy concerning the Indian medicinal plant leaves.

S. Kavitha *et al.* [15] advanced a deep learning model to identify herb plants using a vision-based intelligent approach. In this study, 500 photographs were compiled for each therapeutic plant. To increase the sample size, the data were resized and supplemented. The MobileNet deep learning model was selected for the automatic recognition of medicinal leaves. The accuracy of the deep learning model for the proper identification of medicinal herbs was 98.3

D. T. N. Nhut *et al.* [16] proposed various deep learning approaches, including ViT and Bidirectional Encoder Image Transformer (BEiT), and among all the models, BEiT achieves the highest accuracy on the VNPlant-200 dataset. The limitation in this paper is that the author uses the images with less resolution.

I. Pacal [17] introduces a multi-axis ViT model for the classification of maize diseases, and the dataset consists of 4 classes. To boost the accuracy, the Global Response Normalization (GRN)-based MLP from the ConvNeXtV2 architecture was adapted instead of the MLP in the MaxViT architecture. Three datasets were utilized to create the enhanced and large amount of data, and the model gives the accuracy of 99%, which has been demonstrated as exceedingly effective for practical applications in agriculture.

R. K. Rachman *et al.* [18] introduced a ViT Base (B) transfer learning model for the recognition of rice disease and achieved an accuracy of 97%, which surpasses the deep learning model EfficientNetV2 B0. It is mentioned that the ViT is a promising solution for integrating cutting-edge AI into sustainable agricultural practices, ultimately contributing to improved crop management and yield.

The conclusion from the related research on plant species categorization emphasizes how the use of machine learning and deep learning approaches has significantly advanced the subject. Researchers have shown the effectiveness of support vector machines, deep convolutional neural networks, and novel techniques, including progressive transfer learning and hybrid feature vectors, in several investigations [19]. Impressive accuracy rates have been achieved by these strategies, often outperforming conventional classification techniques. The analysis also highlights how crucial cross-domain identification is, especially in difficult situations like diagnosing rare illnesses or species. The possibility for efficiently tackling such difficulties is shown by the creation of specialized models.

Overall, the simultaneous use of these efforts highlights how machine learning and deep learning have a revolutionary effect on the categorization of plant species, providing the possibility of improved precision, effectiveness, and scalability in further studies. The possibility for improvements in the categorization of plant species is still bright, as scientists continue to hone current techniques and investigate cutting-edge strategies. These developments will provide priceless information for the fields of ecological protection, agriculture [20], and medical research. The growth and development of plants under natural conditions are usually influenced by many abiotic (light, temperature, humidity, etc.) and biotic (bacteria, fungi, etc.) factors, which induce variations in the plant populations, so for tracking the original or true population data, the Indian medicinal plant dataset input would be more appropriate or logically utilized in the current scenario.

3. METHODOLOGY

In our recommended approach, the primary objective is to accomplish automatic categorization of medicinal plants utilizing a hybrid model consisting of a pre-trained ViT architecture and a CatBoost classifier [21] model enhanced with Optuna. The Indian medicinal plant dataset is utilized to compile medicinal plant species [22]. Although conventional CNN models have been trained over varied datasets, they produce more false positive rates with low accuracy [23]. Furthermore, most of these datasets contain a smaller number of images and species. To get rid of the above problems, data preprocessing techniques such as image enhancement, segmentation, and augmentation are in use. As the dataset increases, it is difficult to obtain better performance from the CNN models [24]. To overcome this problem, the ViT model has been used, as transformers provide a few advantages, such as the ability to capture long-range dependencies and adapt to varying input sizes while ensuring parallel processing, and will best fit the related tasks of images. The major objective of this paper is to carry out automatic classification of medicinal plants using hybrid architecture of a combination of a ViT with a CatBoost classifier. The database used is of Indian medicinal plants that is publicly available [22]. Existing CNN-based models decrease the performance for the same dataset. Therefore, an advanced hybrid model has been developed that achieves higher accuracy classification while reducing execution speed. Image preprocessing, the process of segmentation, the feature extraction process, and categorization are the four phases of the suggested method. The first step is to collect datasets of various species. Subsequently, various preprocessing steps were applied to the existing dataset. The dataset is trained using the pre-trained ViT model, following principal component analysis (PCA) to minimize the features that have been extracted, and the attributes are categorized using the CatBoost classifier. Fig. 1 depicts the proposed hybrid model of ViT and CatBoost. The proposed model, designed for the identification of Indian medicinal species, follows these key steps:

```
Input:
Dataset D of leaf images with labels
Parameters: IMG_SIZE = 256×256, BATCH_SIZE = 32,
SEED = 99, PCA_k
Output:
Trained CatBoost model M
Performance metrics: Accuracy, F1-Score, Confusion Matrix

Function Preprocess(D)
  For each image in D do
      Resize image to IMG_SIZE
      Normalize pixel values to range [0, 1]
Shuffle D using SEED
Split D into:
      - 80% training set → D_train
      - 20% validation set → D_val
      - 10% test set (Separate) → D_test
Return D_train, D_val, D_test
Function ExtractFeaturesViT(D)
Load pre-trained Vision Transformer model ViT
(excluding final classification layer)
For each image in D do
      x ← ViT(image)
Return feature matrix X
Function ApplyPCA(X, k)
   Fit PCA on X to retain k components
   Return transformed matrix X_pca
Function TrainCatBoost(X_train, y_train)
Initialize CatBoost model with tuned hyperparameters
   Fit model on (X_train, y_train)
   Return trained model M
Main:
   D_train, D_val, D_test ← Preprocess(D)
   X_train ← ExtractFeaturesViT(D_train.images)
   X_val ← ExtractFeaturesViT(D_val.images)
X_test ← ExtractFeaturesViT(D_test.images)
   X_train_pca ← ApplyPCA(X_train, PCA_k)
   X_val_pca ← ApplyPCA(X_val, PCA_k)
   X_test_pca ← ApplyPCA(X_test, PCA_k)
M ← TrainCatBoost (X_train_pca, D_train.labels)
y_pred ← M.predict(X_test_pca)
   Evaluate predictions: Accuracy, F1-Score, and
   Confusion Matrix on D_test
```

3.1 Data Collection

The dataset contained species of Indian medicinal plants, from which some samples are shown in Fig. 2. The dataset is gathered across multiple regions in Karnataka and Kerala, India. Features such as varied resolutions, different lighting, various backdrops, and different seasons of the year were all included in the datasets. The datasets included 5900 photos of 40 different plant species and individual leaf photos of 80 different plant species, totalling 6900 samples

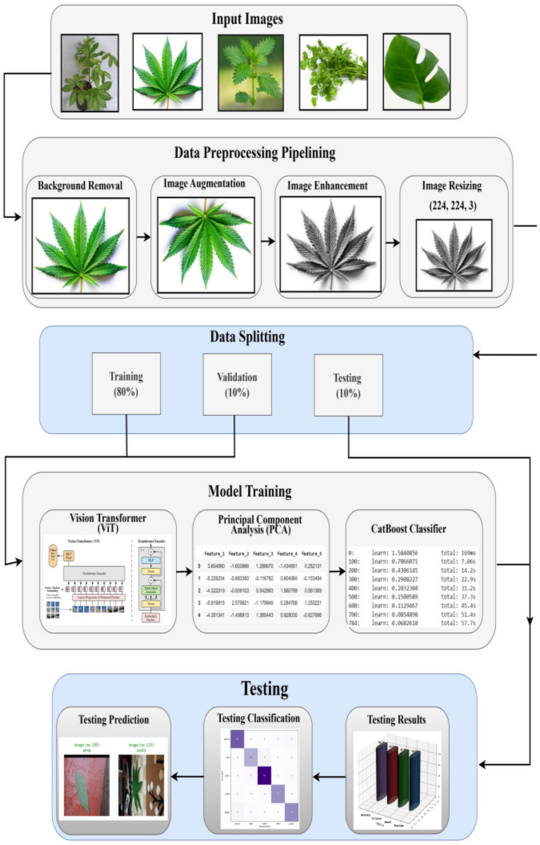

Fig.1: *Research flow of real-time medicinal plant identification.*

that were captured with smartphones under real-time conditions. For this research, a total of five species of leaf images have been meticulously selected, as enumerated in Table 1. The final image quality is typically poor because of the hand and/or camera motion. To ensure the acquisition of high-quality images and improve the accuracy and performance of the subsequent models used in the study, image preprocessing has been carried out, taking these parameters into account. The data distribution is 80% training and 20% validation, as shown in Fig. 3.

Although the Indian medicinal plant dataset consists of 40 species, for this paper only 5 species were considered due to various reasons, such as the sufficient number of available qualitative images, the widespread use of these medicinal species in various healthcare systems, and the visual diversity in leaf morphology. By focusing on mostly common and distinctive species such as alovera, amla, ginger, coffee, and castor, the aim is to develop a reliable baseline for model development and performance.

3.2 Data Pre-processing

An extensive data preparation pipeline is implemented once image acquisition is finished to get images ready for further analysis. Specifically intended to enhance the quality and appropriateness of images for activities that come after, this preprocessing regimen consists of a set of stages such as image enhancement, resizing, background removal, and augmentation. By improving the dataset's interpretability, robustness, and generalizability, these preprocessing methods together provide an adequate foundation for further analysis and model building.

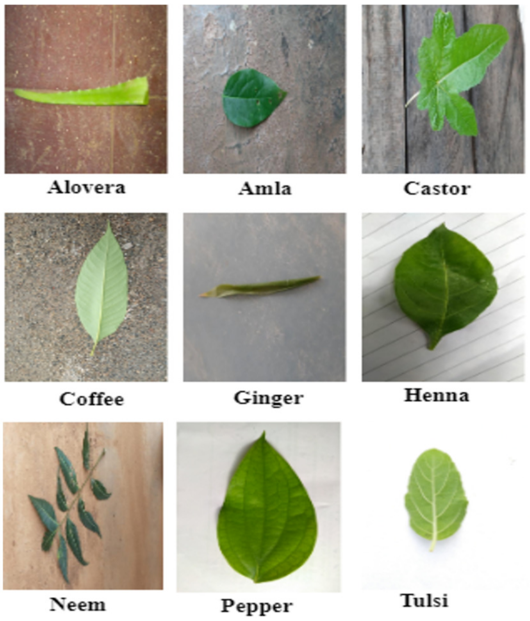

Fig.2: *Samples of Indian medicinal leaf images.*

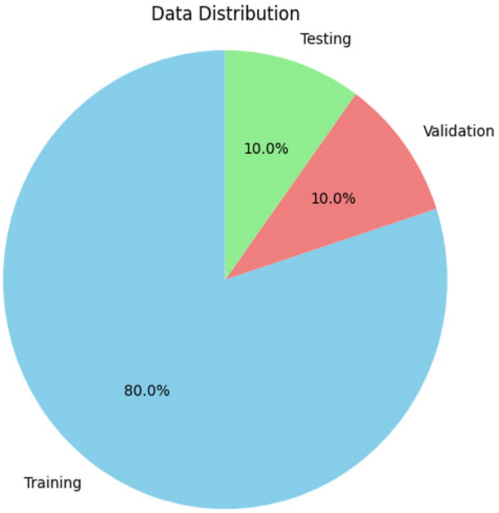

Fig.3: *Data distribution.*

3.3 Image Augmentation

The Indian medicinal plant dataset contains images of medicinal plants from the states of Karnataka and Kerala, India. A total of five species have been considered, but the data is insufficient to identify the actual species of medicinal plant. To overcome these issues and balance the data, data augmentation is performed with all the medicinal plant species. Initially the resolution of every image was 256 × 256 pixels. The incorporation of diversification into images via the methodology of image augmentation significantly enhances the generalizability and overall efficacy of machine learning and deep learning-based classification models. Five types of augmentation had been performed, which included flipping left to right, flipping up and down, brightness, contrast, and saturation of images. The range of contrasted images was set to 0.2 to 0.4, saturation was set to be 2 to 6, the brightness of images was increased to 0.1, and flipping was left to right and up to down. Table 1 displays the total number of photos before augmentation, and after performing the augmentation, the data has been increased five times, which increases the accuracy of the model.

Table 1: *List of medicinal plant species with botanical names and total number of images.*

Common Name	Botanical Name	Total Images	Augmented Data
Aloe vera	Aloe_barbadensis_miller	118	590
Amla	Phyllanthus_emblica	67	335
Castor	Ricinus_communis	129	645
Coffee	Coffea	83	415
Ginger	Zingiber_officinale	118	590
Total		515	2575

3.4 Image Enhancement

Image enhancement is a crucial step in the image processing process that aims to improve the quality and interpretability of the input image so that the resulting output image is both more appropriate and informative than the original. Image enhancement is a carefully considered procedure that aims to clarify hidden information and improve the image's region of interest, making it more useful for further analysis and interpretation. This process improves the image's visual clarity and detail, which greatly increases the efficiency and precision of later image-based tasks and analysis.

3.5 Image Resizing

The resizing of images during preprocessing is a crucial step aimed at ensuring compatibility with the diverse requirements of deep learning models. Given the varied specifications of these models about image dimensions, standardization through resizing to a uniform size becomes imperative. Following augmentation, all images were again resized to a uniform dimension of 256 × 256 pixels, facilitating consistent input in input size for the ViT model, which requires fixed-size inputs for optimal performance.

3.6 ViT Model Construction

Transformers can be instantly utilized in imagery, with the least required alterations. To accomplish so, an image is to be separated into patches and deliver the sequences of linear embedded data of these patches as an input to a Transformer [26]. Image patches are handled like tokens in an NLP implementation. Fig. 4 shows the ViT model architecture.

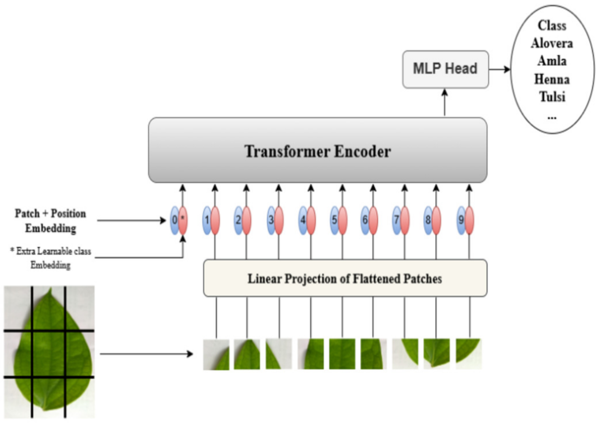
Class
Alovera
Amla
Henna
Tulsi
...
MLP Head
Transformer Encoder
Patch + Position Embedding
* Extra Learnable class Embedding
0
*
1
2
3
4
5
6
7
8
9
Linear Projection of Flattened Patches

Fig.4: *ViT Model Architecture.*

$$Z_0 = [x_{class}; x_1^p E; x_2^p E; \cdots; x_N^p E] + E_{pos}, E \in R^{P^2 C * D}, E_{pos}, \in R^{(N+1)*D} \tag{1}$$

Equation (1) pertains to the input image's class token, patch embedding, and position embedding (E denotes embedding).

$$Z'1 = MSA(LN(z_1^{(-1)}) + z_1^{(-1)}, \quad l = 1, \ldots, L \tag{2}$$

Equation (2) states that for each layer from 1 to L (the entire number of layers), there is a Multi-Head Attention layer (MSA) encompassing a LayerNorm Layer (LN).

$$Z'1 = MSA(LN(Z_1')) + Z_1', l = 1, \ldots, L \tag{3}$$

According to Equation (3), each layer from 1 to L (the overall number of layers) is associated with a Multilayer Perceptron layer (MLP) that wraps a LayerNorm layer (LN).

$$y = LN(Z^0{}_1) \tag{4}$$

According to Equation (4), the last layer L, the output y, is the zeroth token of z enclosed within a LayerNorm layer (LN).

Vision transformers attain great performance when pre-trained at the appropriate scale and translated to challenges with fewer data points [27]. During the model development phase, the vit keras library is used to create ViT architecture, especially ViT-B16. The activation function softmax, class count, and picture size were the configuration parameters used to instantiate this architecture, which is well-known for its efficiency in image classification tasks. By using pre-learned representations during pretraining, the model's capacity to generalize across a variety of picture datasets is improved. To meet the demands of the classification assignment, the ViT-B16 model is modified. To accept image inputs with a size of 256×256, an input layer with the proper proportions is constructed. The input layer is fitted with the ViT model, and the output tensor is then flattened to make it ready for the next fully connected layer. With GELU activation functions, two thick layers were added to the model architecture to aid in feature translation and abstraction. With a softmax activation function to calculate class probabilities, the final dense layer had five output units corresponding to the class labels. Softmax activation is calculated by equation 5.

$$\text{Softmax}(zi) = \frac{e^{zi}}{\sum_{j=1}^{K} e^{zj}} \tag{5}$$

According to equation 5, zi is the output logit for class i, and K is the total number of classes, and in our case, 5 classes have been considered. To get the model ready for training, compilation and model optimization were done. For consistent and effective training, the AdamW optimizer is selected with a 0.001 learning rate and 0.01 weight decay. For multi-class classification tasks, the loss function is tuned to sparse categorical cross entropy calculated using equation 6.

$$\mathcal{L}_{CE} = -\sum_{i=1}^{N} \sum_{c=1}^{C} y_{i,c} \log \hat{y}_{i,c} \tag{6}$$

where $y_{i,c}$ gives the true label represented by 1, meaning sample i belongs to class c, otherwise 0, and $\hat{y}_{i,c}$ predicted the probability of sample i for class c. To evaluate the model's performance in a comprehensive manner, accuracy and sparse top-k categorical Accuracy was used as an evaluation metric. This carefully designed model architecture and optimization plan provide a solid basis for reliable and efficient picture categorization, preparing the ground for the training and assessment stages that follow.

3.7 Training Process

For the classification of medicinal species, our proposed hybrid model includes CatBoost algorithm due to its ability to handle categorical data most efficiently, prevent overfitting through ordered boosting, and deliver superior accuracy with minimal parameter tuning for small to medium-sized datasets. It delivers superior accuracy with minimal parameter tuning. Compared with other boosting algorithms like LightGBM and XGBoost, CatBoost has demonstrated better generalization and faster convergence in tabular and image-derived feature data. To enhance the performance of the model, the Optuna framework was employed for hyperparameter tuning. It provides a flexible and efficient optimization process with features like automated search space reduction and pruning unpromising trails. While other methods like random search and grid search are widely used, Optuna' Bayesian strategy with fewer iterations allows faster convergence toward optimal configurations, which makes it suitable for resource-constrained experimentation.

To increase the training accuracy and speed of the hybrid model, various methods and parameters have been utilized. During preprocessing, data augmentation techniques were adopted to get a large amount of data for recognition purposes. Various transformations have been applied, like blurring, rotation, scaling, flipping, etc. This helps to increase the diversity of data for the training set. In addition to these methods, various parameters, which include batch size, learning rate, number of epochs, weight decay rate, activation functions, optimizer, loss, metrics, and input image dimension, can substantially leverage the performance of the model. Adjusting the learning rate can impact a model's accuracy and speed. To overcome the overfitting, prevent abrupt loss divergences, and improve model stability, a weight decay parameter is used. For the ViT model, an image size of 256 × 256 is utilized for both training and testing the data. Various hyperparameters were considered for achieving the better performance of the model as mentioned in Table 2.

Table 2: *Hyperparameters used in the CatBoost model.*

Hyperparameter	Value
Image Size	256 × 256 pixels
Optimizer	AdamW
Learning Rate	0.001
Weight Decay	0.01
Activation Function	GELU
Final Activation Function	Softmax
Epochs	5
Loss Function	Sparse Categorical Cross Entropy

Important parameters like the learning rate are 0.001 for the ViT model for feature extraction, and for classification, the learning rate is 0.0165 for the CatBoost classifier. To prevent overfitting and help models learn more complex patterns from data, L2 leaf regression of 0.01294 rates for classification and a weight decay rate of 0.001. The total iterations are 785 for classification with a depth of 7, and to assign the random weights to objects, the Bayesian bootstrap type is used. To control the amount of noise, random strength parameters are taken as 2.61×10^{-6}. Bagging temperatures of 5.25 were utilized to control the randomness of the data sampling process, which is used in building the model's trees during gradient boosting. Bagging temperature is calculated using equation 7.

$$P(x_i) \propto \exp\left(-\frac{\Delta L(x_i)}{T}\right) \qquad (7)$$

where $\Delta L(x_i)$ is the loss difference if sample i is used and T is the bagging temperature.After 5 epochs, the proposed model achieves an accuracy of 93.3% with a train loss value of 0.0143. The results demonstrate the proposed hybrid model performs better for the Indian medicinal plant dataset.

To evaluate and monitor the training process of our proposed model, the training accuracy and loss values were recorded for each epoch. Fig. 5 illustrates the learning curve and testing over 5 epochs. The graph shows the consistent increase in training and testing accuracy, which indicates effective model learning. With these results, our proposed model depicts the optimization efficiency and supports the convergence.

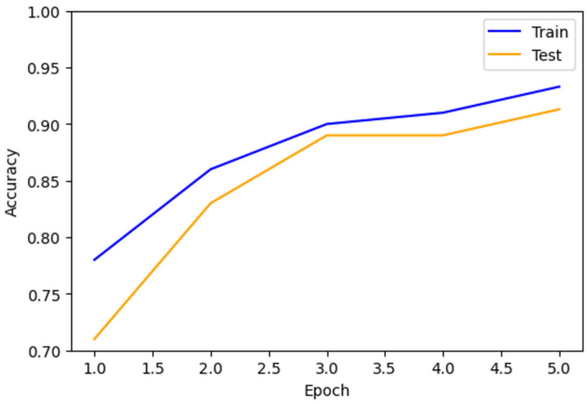

Fig.5: *Learning curve showing training and testing accuracy over 5 epochs for our proposed model.*

3.8 Experimental Setup

The proposed hybrid model was evaluated and implemented in a GPU-based environment to ensure efficient testing and training performance. Experiments were performed in Google Colab, which offers access to NVIDIA Tesla T4 GPUs having the processing capabilities for deep learning tasks. Python was used as the primary programming language along with some libraries such as Keras, TensorFlow, and Matplotlib. Table 3 represents the software tools, operating system, hardware configuration, and dataset utilized for training and processing the dataset.

Table 3: *Software and hardware environment used for model training and processing.*

Category	Description
Software Tools	– Python 3.6.0
	– TensorFlow / Keras (for ViT)
	– CatBoost (Python Library)
	– Optuna (for Hyperparameter Optimization)
	– Matplotlib / SHAP (for Visualization)
Operating System	Google Colab environment (Linux backend)
Hardware Configuration	NVIDIA Tesla T4 GPU
	16 GB GPU memory
	12 GB RAM (Colab runtime)
Dataset	Indian Medicinal Plant Leaf Dataset 5 classes

4. CLASSIFICATION

After model training, the saved model and weights were loaded to facilitate feature extraction, ensuring the preservation of the trained model's architecture and learned parameters for subsequent analysis and deployment. Subsequently, a feature extraction model is constructed using the loaded model, with a specific focus on extracting features from a designated layer pivotal for capturing high-level representations of the input data. These features play a crucial role in downstream analysis tasks such as classification or clustering. Leveraging the constructed feature extraction model, features were extracted from the training dataset by passing it through the model. This process yielded a set of representative features encapsulating meaningful information from the input images, serving as the foundation for subsequent analysis and model evaluation. The significance of feature extraction in enhancing model interpretability and performance across various applications is underscored.

4.1 Dimensionality Reduction

After features were extracted using the built model, principal component analysis (PCA) is used to minimize the number of dimensions in the feature space while maintaining all the required data. For a variance ratio of around 0.99, a total of 45 main components are chosen. Variance is calculated using equation 8, where λ_i is the eigenvalue for the principal component i, k is the number of selected components, and n is the total number of features.

$$Variance\ Retained = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{n} \lambda_i} \qquad (8)$$

Interpretability and computational efficiency are improved by this dimensionality reduction approach, which makes it easier to explore underlying patterns and connections within the feature space. After fitting the features obtained to the PCA model, a transformed feature representation is generated as a consequence of the PCA transformation. In order to enable further analysis and interpretation, the modified characteristics were then arranged into a structured framework, with each primary component. A structured dataset is created by compiling the modified features that resulted, which included the reduced-dimensional representation brought about by PCA. For further in-depth investigation of the underlying structure and trends in the data, this dataset provides a basis for further research activities. In addition to mitigating the curse of dimensionality, many things change the underlying structure of the feature space by using PCA for dimensionality reduction. All things considered, the study approach is more effective overall because of the reduced-dimensional feature representation that PCA produces, which improves model generalization and allows for more efficient computing.

4.2 Classification CatBoost Model

In the experimentation phase, a CatBoost classifier model, fine-tuned with the Optuna optimization framework, is deployed to classify the extracted features. The model's hyperparameters were meticulously tuned to enhance performance and robustness, as mentioned in Table 4.

Table 4: *Hyperparameters used in the CatBoost model.*

Hyperparameter	Value
Learning Rate	0.0165
Iterations	785
Maximum Tree Depth	7
L2 Leaf Regularization	0.01294
Bootstrap Type	Bayesian Bootstrap
Random Strength	2.61×10^{-6}
Bagging Temperature	5.25

With iterations set to 785 and the learning rate optimized to 0.0165. Additional parameters, including depth, L2 leaf reg, bootstrap type, random strength, and bagging temperature, were carefully selected to optimize model effectiveness and generalization. Following model training on the extracted features from the training dataset, the test set underwent preprocessing to prepare it for prediction. This involved applying the same feature extraction technique utilized in the training phase, followed by dimensionality reduction using PCA to align the feature space with the trained model's input requirements. The transformed test features were then organized into a structured format, mirroring the feature column names from the training dataset, to facilitate prediction using the trained CatBoost classifier model. Predictions were generated for the test set samples using the trained model, providing insights into the model's performance on unseen data. Employing this rigorous approach to model deployment and test set processing can ensure the robustness and reliability of their classification system, thus enhancing the credibility and validity of the research findings.

5. TESTING

During the testing stage, the model's effectiveness was determined by calculating the forecasts and related performance metrics. Accuracy, precision, and recall were used to measure classification performance across different species. Additionally, the Mean Squared Error (MSE) was computed separately to compute the average squared difference between the model's predicted probabilities and the true class labels. The proposed model achieved an accuracy score of 93.3%, a weighted F1-score of 0.933, and an MSE of 0.55649, indicating both high classification consistency and low prediction error. These useful metrics are used to assess how well the trained model can accurately predict the leaf categories from the input visuals. By examining these metrics, it can better comprehend the benefits and drawbacks of the model, which will be useful for future changes and improvements to the categorization system.

5.1 Performance Metrics

The confusion matrix is used to record the performance assembled by the classifier for the desired evaluation, which is a table of n x n with n classes. By equating the predicted classes with the actual classes, it conveys assumptions about the rendition of the classifier. It includes various principles like true negative, true positive, false positive, and false negative. Overall, the performance of the classifier is considered using the accuracy and can be figured utilizing the following formula, as depicted in Equation (9).

$$Accuracy = \frac{(Number\ of\ Correct\ predictions)}{(Total\ no\ of\ predictions)} \tag{9}$$

Precision is the proportion of samples that are in the positive category that the classifier expected to be in the positive group. Precision provides information with respect to the capability of the classifier in correctly recognizing the actual positive samples. The formula for precision is given in equation (10). Of all the positive predictions the classifier made, precision provides the percentage of how correct the positive predictions were. Equation 10 calculates this most important statistic for assignments in which the goal is to be sure that positive predictions are correct and to minimize false positives.

$$Precision = \frac{(TruePositive)}{(TruePostive + FalsePositive)} \tag{10}$$

Recall, also referred to as sensitivity or the true positive rate, is the ratio of the actual number of samples that, in reality, belong to the positive class and are correctly classified as positive by the classifier. It informs about the capability of the classifier in identifying each positive sample. Recall can be estimated using Equation 11 as the formula. Recall represents the percentage of true positive predictions out of all actual positive samples, and it is represented by this symbol. This statistic is important when the goal is to reduce false negatives and make sure that the positive forecasts are accurate. A high recall will mean that most of the positive samples within the dataset are being captured by this classifier.

$$Recall = \frac{(TruePositive)}{(TruePostive + FalseNegative)} \tag{11}$$

Recall and accuracy can be summarized into one metric, the so-called F1-score, which gives a general view of the performance of the classifier. Considering accuracy, or the ability to correctly identify positive samples, and recall, or the ability to include all the positive samples, it makes for a much more thorough evaluation. Equation 12 is the formula to get the F1-score. The F1-score ranges between 0 and 100% and is given in percentage format. A successful classification strategy selects a good balance between recall and precision when the F1-Score is high.

$$F1 - Score = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} \tag{12}$$

5.1.1 Cross-Validation Performance

To evaluate the generalizability and robustness of our proposed model, a 5-fold cross-validation was carried out on the Indian medicinal plant dataset. In each fold, 80% of the data was utilized for training and 20% for validation to ensure that every sample was presented to both training and evaluation phases. Additionally, 10% of the entire dataset was taken as an independent test set that was not involved in model training or validation. The results corresponded to the mean and standard deviation across the 5 folds, while the test set was utilized for final model testing. Table 5 summarized the average performance across folds and their standard deviations.

Table 5: *Demonstrates the average performance across folds and standard deviations.*

Metric	Mean (%)	Std. Dev (%)
Accuracy	92.8	±0.47
Precision	91.5	±0.51
Recall	90.7	±0.56
F1-Score	91.1	±0.50

These results of our proposed hybrid model persistently outperform well across various subsets of the dataset, with low standard deviation values, which indicates reliable and stable predictions. In balancing the precision and recall for medicinal plant classification, the high average F1 score highlights the model's effectiveness.

5.2 Results and Discussion

A comprehensive classification report with specific metrics for every class is generated during the assessment of the model's performance in classification. Key performance indicators for each class in the dataset are compiled in the classification report, and they include precision, recall, F1-score, and support. The parameters were set to guarantee that each class in the report is accurately labelled, improving the report's readability and clarity. The model's performance can be evaluated class-by-class with the help of the classification report, which offers a detailed examination of the model's features and drawbacks in several categories. Accuracy, recall, precision, and F1-scores are the measures considered to examine the testing outcome. The accuracy, recall, precision, and F1-score of the CatBoost model are 93.3%, 90%, 92%, and 91%, respectively, making it effective in the recognition of the Indian medicinal plants, with a total average score of 93% across all measures. Fig. 6 and Table 6 show the metrics score applied to test data of the CatBoost classifier as a bar graph.

Table 6: *CatBoost classifier model outcome.*

	Precision	Recall	F1-score	Support
Alovera	0.91	0.94	0.93	53
Amla	0.91	0.88	0.89	33
Castor	0.95	0.96	0.95	72
Coffee	0.97	0.97	0.97	38
Ginger	0.93	0.88	0.90	43

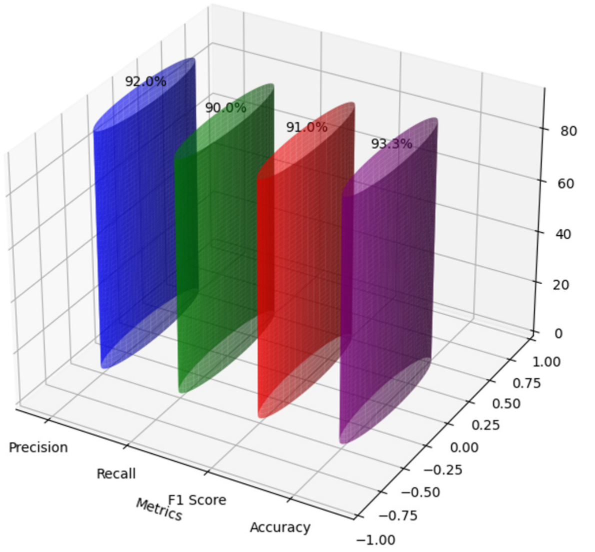

Fig.6: *Model performance of CatBoost classifier.*

To demonstrate the classification findings, a confusion matrix is created as part of the model performance assessment. The confusion matrix makes it easier to see any misclassifications or patterns in the prediction errors by giving a thorough overview of the model's predictions across several categories. Moreover, the confusion matrix's visual representation is improved by using a color map to indicate the degree of categorization errors shown in Fig. 7. The confusion matrix is easier to understand because of its color-coded form, which makes it easier to spot areas with greater error rates.

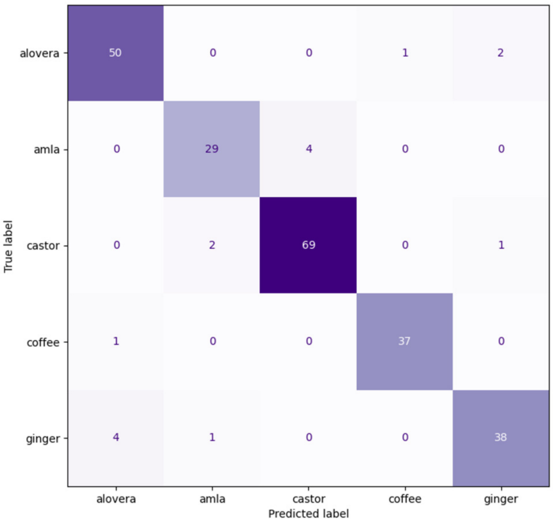

Fig. 7: *Confusion matrix of CatBoost classifier.*

5.2.1 Comparative Analysis with Standard Models

Comparative analysis of various baseline deep learning algorithms such as ResNet50, EfficientNetB0, and MobileNetV2 has been performed using transfer learning from ImageNet pretrained weights with our proposed model with training accuracy and loss. Our model shows the highest accuracy of 93.3%, which outperforms the others among all metrics like precision, recall, and F1-score. The final classification layers were substituted with a dense output layer of five neurons and a softmax activation function. All the models were fine-tuned on the same number of epochs with the optimizer as AdamW, a learning rate of 0.001, and a batch size of 32. Table 7 presents the comparison of the performance of the proposed ViT + CatBoost hybrid model against three widely used deep learning models: ResNet50, EfficientNetB0, and MobileNetV2. Fig. 8 shows the comparative analysis of classification performance across baseline models and the proposed ViT + CatBoost hybrid model.

Table 7: *Demonstrates the comparison of our proposed model with other deep learning models.*

Model	Accuracy	Precision	Recall	F1-Score
ResNet50	89.6	0.88	0.87	0.87
EfficientNetB0	91.2	0.90	0.89	0.89
MobileNetV2	90.3	0.89	0.88	0.88
ViT+CatBoost (Proposed)	93.3	0.92	0.90	0.91

5.3 Testing Predictions

To assess test predictions, a function is established to use those generated predictions for five randomly chosen images from the test set; also, their corresponding predictions were generated. To make the prediction findings more understandable and clearer, this color-coded annotation approach is used as shown in Fig. 9. The function used the predicted categories to iteratively go through the chosen photos and assess how accurate the model's predictions were concerning the ground truth labels. The pictures were then presented with the titles that corresponded to them, giving an illustration of how well the model performed on data that had not yet been viewed.

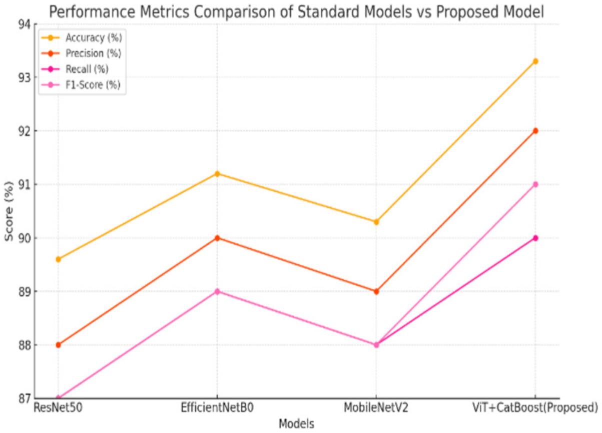

Fig. 8: *Comparative analysis of classification performance across baseline models and the proposed ViT + CatBoost hybrid model.*

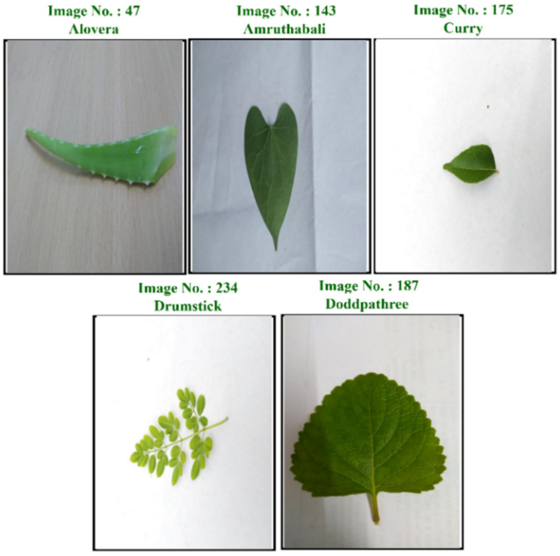

Fig. 9: *Random images from the test dataset.*

The efficiency and precision of the model in classifying leaves from test photos were evaluated using this systematic assessment technique, providing important information about the model's generalization potential and possible areas for improvement. The SHapley Additive explanation (SHAP) has been implemented for the model explanation, feature importance, and interpretability of the proposed model. The SHAP analyzes the contribution of independent feature variables to the model's predicted output. Fig. 10 plots show the top 8 features that most sig-

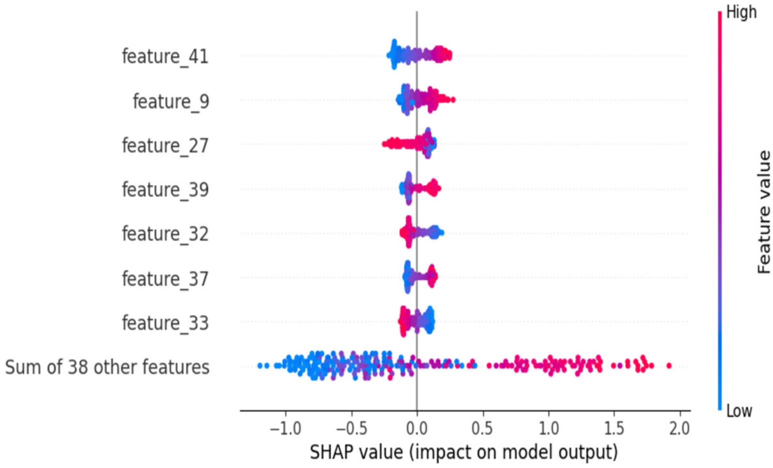

Fig.10: *SHAP-based feature importance values derived from PCA-transformed ViT embeddings used by the CatBoost classifier.*

nificantly influence the predictions of the CatBoost model, allowing us to gain insights into the model's decision-making process.

6. CONCLUSION AND FUTURE WORK

Despite the increasing availability of researchers and public apps, accurate recognition of medicinal plants is still a challenge. Most of the datasets do not provide features like size, color, shape, and inter-class domain. CNN-based learning models do not provide better results for species classification, as the data to be provided is inadequate. This paper presents a hybrid model based on the Transformer and CatBoost classifier tuned with Optuna, in which images were selected to evaluate the model's performance thoroughly, and accuracy and sparse top-k categorical accuracy were used as evaluation metrics. The classical CatBoost hyperparameters are tuned with Optuna to achieve a better performance for the recognition. Image preprocessing, background removal, and resizing are performed to enhance the input image. The first step is to take the leaf images from the Indian medicinal plant dataset and then perform preprocessing; after that, the model will extract the features, and classification will be done. The hybrid model, ViT, and CatBoost classifier model perform better than other existing models in the training and testing stages, according to the overall simulation results. Using the Indian medicinal plant dataset, various comparative analyses have been done, and it gives an accuracy of 93%, and other parameters like precision, recall, and F1-score are consistently higher.

However, our study has demonstrated the identification of medicinal plant leaves by testing on a few desirable plant species supplemented with more datasets, and future work will be extended for catering to a large number of commercially important medicinal plant species of higher altitudes as well. Also, incorporating hyperspectral imagery and expanding to real-time mobile application deployment.

DECLARATIONS

The authors have no relevant financial or non-financial interests to disclose. The authors have no conflicts of interest relevant to the content of this article. All authors certify that they have no affiliations with or involvement in any organization or entity with any financial or non-financial interest in the subject matter or materials discussed in this manuscript.

AUTHOR CONTRIBUTIONS

Conceptualization, F.F. and D.G.; methodology, F.F.; software, F.F.; validation, D.G.; formal analysis, F.F.; investigation, F.F.; data curation, H.S.; writing—original draft preparation, F.F.; writing—review and editing, F.F., D.G. and H.S.; visualization, F.F.; supervision, D.G. All authors have read and agreed to the published version of the manuscript.

References

[1] L. Shahmiri, P. Wong and L. S. Dooley, "Accurate Medicinal Plant Identification in Natural Environments by Embedding Mutual Information in a Convolution Neural Network Model," *2022 IEEE 5th International Conference on Image Processing Applications and Systems (IPAS)*, Genova, Italy, pp. 1-6, 2022.

[2] O. A. Malik, N. Ismail, B. R. Hussein and U. Yahya, "Automated Real-Time Identification of Medicinal Plant Species in Natural Environments Using Deep Learning Models—A Case Study from the Borneo Region," *Plants*, vol. 11, no. 15, p.1952, 2022.

[3] R. Azadnia, M. M. Al-Amidi, H. Mohammadi,M. A. Cifci, A. Daryab and E. Cavallo, "An AI-

based approach for medicinal plant identification using deep CNN based on global average pooling," *Agronomy*, vol. 12, no. 11, p.2723, 2022.

[4] B. Bhattacharjee, K. Sandhanam, S. Ghose, D. Barman and R. K. Sahu, "Market overview of herbal medicines for lifestyle diseases," *Role of Herbal Medicines*, pp. 597-614, 2024.

[5] R. Gowthami, N. Sharma, R. Pandey and A. Agrawal, "Status and consolidated list of threatened medicinal plants of India," *Genetic Resources and Crop Evolution*, vol. 68, no. 6, pp. 2235-2263, 2021.

[6] P. Mehta, K. Bisht, K. C. Sekar and A. Tewari, "Mapping biodiversity conservation priorities for threatened plants of the Indian Himalayan Region," *Biodiversity and Conservation*, vol. 32, no. 7, pp. 2263-2299, 2023.

[7] C. G. Yedjou, J. Grigsby, A. Mbemi, D. Nelson, B. Mildort, L. Latinwo and P. B. Tchounwou, "The management of diabetes mellitus using medicinal plants and vitamins," *International Journal of Molecular Sciences*, vol. 24, no. 10, p. 9085, 2023.

[8] M. A. Kiflie, D. P. Sharma and A. H. Mesfin, "Deep learning for medicinal plant species classification and recognition: a systematic review," *Frontiers in Plant Science*, vol. 14, p. 1286088, 2024.

[9] J. Yue, W. Li and Y.-Z. Wang, "Superiority verification of deep learning in the identification of medicinal plants: Taking Paris polyphylla var. varyunnanensis as an example," *Frontiers in Plant Science*, vol. 12, p.752863, 2021.

[10] P. Singla, V. Kalavakonda and R. Senthil, "Detection of plant leaf diseases using deep convolutional neural network models," *Multimedia Tools and Applications*, vol. 83, pp 64533-64549, 2024.

[11] S. Chulif, S. H. Lee, Y. L. Chang and K. C. Chai, "A machine learning approach for cross-domain plant identification using herbarium specimens," *Neural Computing and Applications*, vol. 35, no. 8, pp. 5963–5985, 2023.

[12] H. K. Diwedi, A. Misra and A. K. Tiwari, "CNN-based medicinal plant identification and classification using optimized SVM," *Multimedia Tools and Applications*, vol. 83, no. 11, pp. 33823–33853, 2024.

[13] K. Pankaja and V. Suma , "Plant leaf recognition and classification based on the whale optimization algorithm (WOA) and random forest (RF)," *Journal of the Institution of Engineers (India): Series B*, vol. 101, no. 5, pp. 597–607, 2020.

[14] M. Sharma, N. Kumar, S. Sharma, S. Kumar, S. Singh and S. Mehandia, "Medicinal plant recognition using heterogeneous leaf features: an intelligent approach," *Multimedia Tools and Applications*, vol. 83, pp. 51513-51540, 2023.

[15] S. Kavitha, T. S. Kumar, E. Naresh, V. H. Kalmani, K. D. Bamane and P. K. Pareek, "Medicinal plant identification in real-time using a deep learning model," *SN Computer Science*, vol. 5, p. 73, 2023.

[16] D. T. N. Nhut, T. D Tan, T. N. Quoc and V. T. Hoang, "Medicinal plant recognition based on vision transformer and BEiT," *Procedia Computer Science*, vol. 234, pp. 188–195, 2024.

[17] I. Pacal, "Enhancing crop productivity and sustainability through disease identification in maize leaves: Exploiting a large dataset with an advanced vision transformer model," *Expert Systems with Applications*, vol. 238, p. 122099, 2024.

[18] R. K. Rachman, D. R. I. M. Setiadi, A. Susanto, K. Nugroho and H. M. M. Islam, "Enhanced vision transformer and transfer learning approach to improve rice disease recognition," *Journal of Computing Theories and Applications*, vol. 1, no. 4, pp. 446–460, 2024.

[19] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proceedings of the 36th International Conference on Machine Learning*, Long Beach, California, PMLR 97, pp. 6105-6114, 2019.

[20] S. P. Mohanty, D. P. Hughes, and Marcel Salathé, "Using deep learning for image-based plant disease detection," *Frontiers in Plant Science*, vol. 7, p. 01419, 2016.

[21] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush and A. Gulin, "CatBoost: unbiased boosting with categorical features," in *Proceedings of the 32nd International Conference on Neural Information Processing Systems (NIPS'18)*, pp. 6639-6649, 2018.

[22] P. B. R and N. S. Rani, "DIMPSAR: Dataset for Indian medicinal plant species analysis and recognition," *Data in Brief*, vol. 49, p. 109388, 2023.

[23] A. Kaya, A. S. Keceli, C Catal, H. Y. Yalic, H. Temucin and B. Tekinerdogan, "Analysis of transfer learning for deep neural network-based plant classification models," *Computers and Electronics in Agriculture*, vol. 158, pp. 20–29, 2019.

[24] A. Krizhevsky, I. Sutskever and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, 2012.

[25] D. Gupta and R. Rani, "A study of big data evolution and research challenges," *Journal of Information Science*, vol. 45, no. 3, pp. 322–340, 2019.

[26] F. Firdous, S. Bashir, S. Z. Rufai and S. Kumar, "OpenAI ChatGPT as a Logical Interpreter of code," *2023 2nd International Conference on Edge Computing and Applications (ICECAA)*, Namakkal, India, pp. 1192-1197, 2023.

[27] A. Dosovitskiy *et al.*, "An image is worth 16×16 words: Transformers for image recognition at scale," *arXiv preprint* arXiv:2010.11929, 2020.

Faisal Firdous is an Assistant Professor in the Department of Computer Science and Engineering at Jaypee University of Information Technology, Solan, Himachal Pradesh, India, currently pursuing a Ph.D. in Computer Science and Engineering with research interests spanning Artificial Intelligence, Deep Learning, Computer Vision, and Image Processing, focusing on medicinal plant recognition, explainable AI, and healthcare applications; he holds an M.Tech in Computer Science and Engineering and a B.Tech in the same field. He has experience of more than 3 years in academics.

Deepak Gupta is an Assistant Professor in the Department of Computer Science & Engineering and Information Technology at Jaypee University of Information Technology, Waknaghat, India. He holds advanced degrees in Computer Science & Engineering and Physics. Before joining academia, he spent nearly a decade in the IT industry, where he held various roles in software product development and program management. In academia, he has gained extensive experience in teaching, research, and academic administration. His research interests include big data analytics, cybersecurity, artificial intelligence, and programming languages.

Hemant Sood is a Professor in the Department of Biotechnology and Bioinformatics at Jaypee University of Information Technology, Waknaghat, India, specializing in plant tissue culture, genetic transformation, and the development of micropropagation and cell culture technologies for high-value medicinal plants; her research emphasizes transferring these technologies to farmers to enhance their socio-economic status and holds three granted patents as the lead inventor; recognized among 125 Women Luminaries in STEM by CII, she was also honored with the Dr. Sarvepalli Radhakrishnan Distinguished Associate Professor & Researcher Award 2023 for her contributions to plant biotechnology.