Identification of Medicinal Plants using Hybrid Model

Dr. Poonam Lambhate
Department of Computer Engineering
JSPM's JSCOE
jscoeit@gmail.com

Ms. Gauri Girme
Department of Computer Engineering
JSPM's JSCOE
girmegauri18@gmail.com

Ms. Kulthe Rachana
Department of Computer Engineering
JSPM's JSCOE
rachana.kulthe04@gmail.com

Ms. Rutugandh Gaikwad
Department of Computer Engineering,
JSPM's JSCOE
rutugandhgaikwad@gmail.com

Ms. Vaishnavi Deshmukh
Department of Computer Engineering
JSPM's JSCOE
vaishud150604@gmail.com

Abstract - Plants play a vital role in our lives by providing various facilities such as food, oxygen, medicines, etc. They are valuable for their medicinal properties and contain substances that are used in medicines. However, many such plants are being destroyed due to many environmental challenges such as global warming, pollution, and Limited government support for the research on medicinal plants. Manual identification of plants requires manual observation and expertise. This issue must be addressed by creating an automated identification system for medicinal plants. The hybrid model of CNN and SVM is to be implemented in the system. This hybrid model of CNN and SVM will extract features and classify them based on various features such as the color, shape, and texture of the leaf of a medicinal plant. This system will help identify the medicinal plants effectively and accurately and display information about them if they are identified as medicinal plants.

Keywords – Plants, Medicinal plants, CNN, Hybrid model, SVM, Accuracy, Identification

I. INTRODUCTION

(Identification of Medicinal Plants using Hybrid Model)

Ayurveda is an ancient practice of medicine in India and has its roots in Vedic times approximately 5000 years ago. The main constituents of Ayurvedic medicines are plant leaves and other parts of plants like roots, bark, etc. More than 8000 plants of Indian origin are of medicinal value. Over 80% of plants used for ayurvedic formulations are collected from forests and wastelands whereas the remaining are cultivated in agricultural lands. With the extensive use of medicinal plants, identifying them is a challenging task as they require manual observation which can be time-consuming and may lead to errors in the identification of plants. This is uncertain for those with limited knowledge of the botanical field which may lead to false identification of plants and outcomes. We aim to train the Machine Learning model on a dataset of known medicinal plants and use them to classify the plants. This recognition based on leaves depends on finding exact descriptors and extracting the feature vectors from them. Then the feature vectors of the training samples are compared with the feature vectors of the test sample to find the degree of similarity using an appropriate classifier.

Overview: The system has been developed to automate the identification process of medicinal plants. The system first takes an image of plants as input and identifies it using the model trained by training and testing datasets. After successful identification, whether the plant is medicinal or not the system fetches information such as name, uses, and applications of medicinal plant from the database of plant information and displays it on GUI.

II. LITERATURE SURVEY

The paper[1] presents a technique for automatic identification of medicinal plants using machine learning. The authors have implemented a random forest algorithm that uses color, texture, and geometric features to classify medicinal plant species. The proposed approach aims to automate the identification process, which has traditionally been done manually. The authors review prior work on medicinal plant classification using image processing and machine learning techniques, and demonstrate the effectiveness of their approach through experiments

on a dataset of plant leaves.
- The technique uses random forest algorithm and extracts color, texture, and geometric features from plant leaf images
- Prior work on medicinal plant classification using image processing and machine learning is reviewed
- Experiments show the effectiveness of the proposed approach in accurately classifying medicinal plant species undefined.

The paper[2] presents a novel approach to medicinal plant identification through computer vision and deep learning. It introduces a custom dataset called DeepHerb, with 2,515 leaf images of 40 Indian medicinal herbs. The research compares various pre-trained neural networks (VGG16, VGG19, InceptionV3, Xception) for feature extraction and classification using Artificial Neural Network (ANN) and Support Vector Machine (SVM). The Xception-ANN model achieved 97.5% accuracy. The paper also describes "HerbSnap," a mobile app utilizing this model for real-time plant identification.
- The DeepHerb dataset was specifically designed for the study, containing 2,515 leaf images from 40 medicinal plant species, making it a valuable resource for future research.
- The paper uses Bayesian optimization to fine-tune the SVM hyperparameters, improving the model's accuracy.
- The research leverages transfer learning to overcome the challenge of small datasets, utilizing pre-trained CNNs to achieve high classification accuracy without needing massive data.

The paper[3] focuses on using digital image processing techniques to identify plant species using features like shape, color, and texture of leaves. The study applies methods like Gaussian filtering, K-means clustering, and Principal Components Analysis (PCA) for feature extraction. Support Vector Machine (SVM) is used as the primary classifier, achieving an accuracy of 95.17%. The system effectively enhances plant identification using machine learning, but future research could address challenges with smaller leaves.
- The study uses techniques like Gaussian filtering for noise reduction and K-means clustering for image segmentation.
- PCA (Principal Components Analysis) is used to extract key features like shape, color, and texture from plant leaf images
- SVM is the primary machine learning classifier, achieving 95.17% accuracy.

Future work could improve the detection of small and immature leaves.

The paper[4] describes an approach to automatically identify medicinal plants using a deep learning (DL) model. The research focuses on six plant species: Betel, Curry, Tulsi, Mint, Neem, and Indian Beech, with 500 images per plant collected from Kaggle. After applying pre-processing techniques like resizing and augmentation, the MobileNet DL model was used for classification. The model achieved 98.33% accuracy. The model has a 97%+ score in metrics like precision, recall, and accuracy, and the app provides easy access to medicinal plant information.
- Six medicinal plants (Betel, Curry, Tulsi, Mint, Neem, Indian Beech) were used from a Kaggle dataset, with 500 images per speciesMobileNet model was used for classification, achieving 98.33% accuracy
- Data augmentation techniques like flipping, rotation, and color manipulation were applied to increase training samples

III. METHODOLOGY

The system focuses on designing and implementing a machine learning-based system to identify medicinal plants based on leaf characteristics. It encompasses various phases, including data collection, model development, and system deployment, aimed at creating a reliable tool for plant identification. Implement a hybrid machine learning model combining Convolutional Neural Networks (CNNs) and Support Vector Machines (SVM), to classify plants based on leaf features. The system focuses on identifying medicinal plants using visual leaf characteristics only, and it may not extend to other plant parts (e.g., flowers, stems). This system will provide a tool for automated, accurate identification of medicinal plants, aiding in research, healthcare, and conservation efforts.

A. ARCHITECTURE MODEL

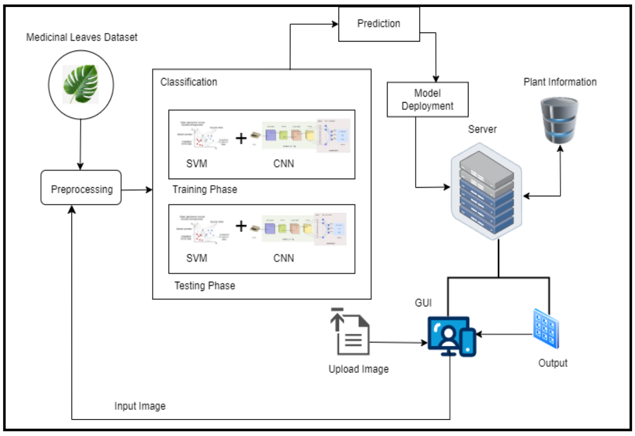

Fig. 3.1 Architecture Diagram

1) Data Acquisition and Preprocessing: The system collects images of medicinal leaves to create training and testing datasets. These images are preprocessed to prepare them for analysis. The preprocessing includes steps like resizing, normalization, and feature extraction.

2) Model Training: The system employs the combination of two classification algorithms:

a) Support Vector Machine(SVM): This is a classification algorithm that finds an optimal hyperplane to separate data point classes.

b) Convolutional Neural Network(CNN): This algorithm is for image classification tasks. They learn hierarchical features from input images.

The preprocessed images are used to train both SVM and CNN models. This process involves adjusting parameters to minimize errors between predicted output and actual labels.

The use of both SVM and CNN models is a hybrid approach where SVM handles smaller datasets and lower dimensional feature spaces and CNN handles large datasets and complex image features.

The combination of CNN and SVM helps to improve accuracy as CNN can extract features and SVM can classify them.

3) Model Testing: Once the models are trained, they are evaluated using a separate set of images i.e. training dataset. It helps assess the model's accuracy and generalization ability.

4) Model Deployment: The model is deployed on the server. This allows the system to be accessible remotely.

5) User Interface(GUI): Users can upload an image of a leaf through GUI. The uploaded image is preprocessed and fed into a hybrid model. The system predicts the plant species based on the output of the model. The system displays predicted plant species and relevant plant information, such as medicinal properties and use.

B. MATHEMATICAL MODEL

$$y = \text{Softmax}\,(W_{FC}\,\text{Pooling}\,(\text{ReLU}\,(W_{Conv} * X + b_{Conv})) + b_{FC})$$

where,
X: Input data
$W_{conv:}$ Weights of the convolutional layer
b_{Conv}: Bias of the convolutional layer
ReLU: Activation function
Pooling: Poling operation
W_{FC}: Weights of fully connected layers

1) Convolution Operation ($W_{Conv} * X + b_{Conv}$)
   - X is the input to the convolutional layer, typically an image or a feature map.
   - W_{Conv} represents the weights of the convolutional filters, and b_{Conv} is the bias term.
   - The convolution operation (denoted by "*") is applied to the input, where the filters (W_{Conv}) scan the input and detect local features (edges, textures, patterns, etc.).
2) ReLU Activation (ReLU(...))
   - After the convolution, a Rectified Linear Unit (ReLU) activation function is applied. This introduces non-linearity into the model by setting all negative values to zero while leaving positive values unchanged.
   - ReLU helps the network learn complex patterns by allowing the model to focus on positive activations and ignore irrelevant information.
3) Pooling Operation (Pooling(...))
   - After the ReLU activation, a Pooling operation is applied to reduce the spatial dimensions of the feature maps, which reduces computational complexity and helps the model focus on the most important features.
4) Fully Connected Layer (W_{FC} * ... + b_{FC})
   - The output of the pooling operation is flattened into a vector and passed to a Fully Connected (FC) layer.
   - W_{FC} represents the weights of the fully connected layer, and b_{FC} is the bias term.
5) Softmax Output
   - Finally, the output of the fully connected layer is passed through a Softmax function, which converts the raw scores into probabilities.

C. ALGORITHM

A Convolutional Neural Network (CNN) is a specialized machine learning algorithm widely employed in computer vision tasks. Specifically designed for processing grid-like data, such as images or videos, CNNs have revolutionized tasks like image classification, feature detection, and image recognition. At the core of a CNN are convolutional layers, where learnable filters scan the input image to detect distinctive features like edges, textures, or shapes. For a given input image, I, and a filter (or kernel) K, the convolution operation at a particular position (i, j) is defined as:

$$S(i, j) = \sum m \sum n\, I(i+m, j+n) \cdot K(m, n)$$

where:
- S (i, j) is the value of the resulting feature map at position (i, j).
- I (i+m, j+n) represents the pixel value of the input image at position (i+m, j+n).
- K (m, n) is the weight of the kernel at position (m, n).
- m and n are the dimensions of the filter.

This equation essentially describes how the filter slides across the image, multiplying and summing the overlapping values of the filter and the image to create an activation map. This process helps detect specific features like edges or textures in the image.

IV. RESULTS AND ANALYSIS

A. RESULT

1) Images before training

The Fig. 4.1 shows the collection of various leaf images in the dataset. Each image is labeled with its corresponding plant species.

Fig 4.1. Images before training

2) Images after training

Fig. 4 .2 shows the results of the CNN model trained on a dataset of leaf images. Here, the model classified the images based on their visual features, which it had learned from the training data.

Fig. 4.2. Images after training

3) Accuracy graph

Fig. 4.3. shows the performance of the model likely a neural network, during its training process. The x-axis represents the number of training epochs, while the y-axis indicates the accuracy of the model.

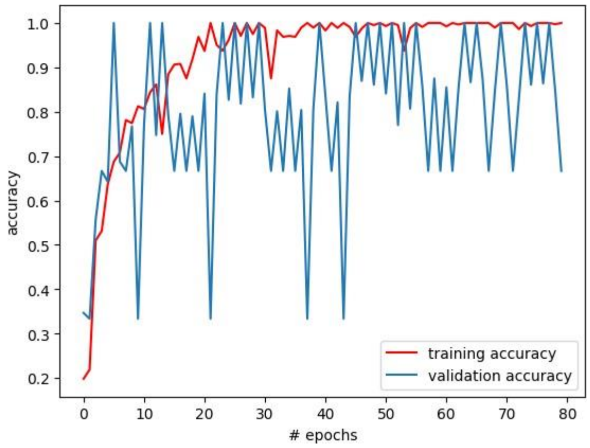

Fig 4.3. CNN accuracy graph

The two curves on the graph represent:

i) Training Accuracy: This curve shows how accurately the model predicts the current class of training data. As the model is trained, it typically improves its accuracy on training data.

ii) Validation Accuracy: This curve shows how accurately the model predicts the correct class of validation dataset, which is a separate set of data not used for training.

B. OUTPUT

Fig. 4. 4 shows the output of the medicinal plants identification system where users can upload an image of a leaf and then process it by a trained Convolutional Neural Network(CNN) model.

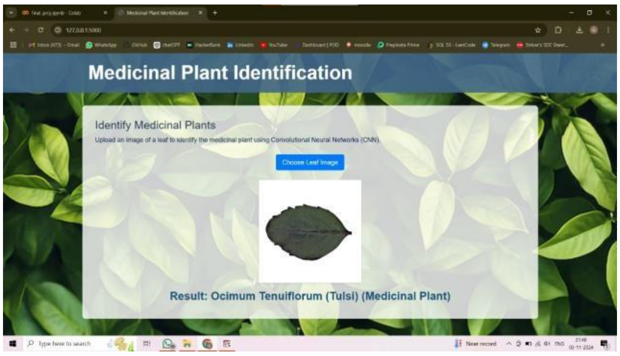

Fig. 4.4. Output

This model predicts the plant species and displays the result, including its common and scientific names and medicinal properties.

V. CONCLUSION

The automated identification and classification of medicinal plants using machine learning represent a significant advancement in both botanical science and herbal medicine. By leveraging the power of technology, this approach enhances the accuracy and efficiency of plant identification, overcoming the limitations associated with traditional methods that often rely on expert knowledge and are susceptible to human error. Machine learning not only facilitates rapid and reliable identification of medicinal plants but also democratizes access to this knowledge, empowering non-experts to engage with plant identification through user-friendly applications.

VI. ACKNOWLEDGMENT

We would like to express our special thanks of gratitude to our guide and head of the Computer Department Dr. Poonam Lambhate for their able guidance and support. We would also like to extend our gratitude to the Principal Mr. R. D. Kanphade for providing us with all the facilities that was required. We would also like to acknowledge with much appreciation the role of our staff. We sincerely thank our parents and our friends who have been always helping and encouraging us.

VII. REFERENCES

[1] Anu Paulson, S. Ravishankar.AI Based Indigenous Medicinal Plant Identification, 2020 Advanced Computing and Communication Technologies for High-Performance Applications (ACCTHPA), July 202.

[2] Luciano D.S. Pacifico; Larissa F.S. Britto; Emilia G. Oliveira; Teresa B. Ludermir. Automatic Classification of Medicinal Plant Species Based on Color and Texture Features, IEEE 2019

[3] C. Sivaranjani; Lekshmi Kalinathan; R. Amutha; Ruba Soundar Kathavarayan. Real-Time Identification of Medicinal Plants using Machine Learning Techniques. IEEE 2019

[4] Perez, A.J., Lopez, F., Benlloch, J.V. and Christensen, S., 2000. Color and shape analysis techniques for weed detection in cereal fields. Computers and electronics in agriculture, 25(3), pp.197-212

[5] Vijayashree, T. and Gopal, A., 2015. Classification of Tulsi Leaves Based on Texture Analysis

[6] A. and Sharma, S., 2013. An automatic leaf recognition system for plant identification using machine vision technology. International Journal of Engineering Science and Technology, 5(4), p.874

[7] Im, C., Nishida, H. and Kunii, T.L., 1998, November. A Hierarchical Method of Recognizing Plant Species by Leaf Shapes. In MVA (pp. 158- 1Im, C., Nishida, H. and Kunii, T.L., 1998, November. A Hierarchical Method of Recognizing Plant Species by Leaf Shapes. In MVA (pp. 158- 161)

[8] Dr. Sandip D. Satav* Dr. Shailesh M Hambarde1 Dr. Aparna S Hambarde2 Dr. Poonam D Lambhate3 Mrs. Aarti S Satav3. Utilizing A Hybrid Gru Intrusion Detection System To Improve The Security Of IoT-Based Smart Farming. Journal For Basic Sciences. 2024/10

[9] Dr. Sandip D. Satav* Dr. Poonam D Lambhate1 Dr. Shailesh M Hambarde2 Dr. Aparna S Hambarde3 Mrs. Aarti S Satav. Using Random Forests And Internet Of Things Sensors To Optimize Predictive Maintenance In Smart Factories. Computer Research and Development. 2024/10

[10] Dr. Sandip D. Satav1 Dr. Poonam D Lambhate2 Dr. Shailesh M Hambarde3 Mrs. Aarti S Satav4. WSN Technology-Based IoT-Based Smart Agriculture System For Insect And Pest Identification, Crop Status Prevention. Journal of Technology. 2024/7

[11] Dr. Poonam D Lambhate1 Dr. Shailesh M Hambarde2 Mrs. Aarti S Satav3. Smart Cities Can Generate Efficient Power From Garbage Through The Integration Of IoT And LSTM-MLP. Journal of Informetrics 2024

[12] Dr. Sandip D. Satav* Dr. Shailesh M Hambarde1 Dr. Poonam D Lambhate2 Mrs. Aarti S Satav3. An Effective Allocation Scheme for Resources in Edge Computing enabled Networks with Uncertain Network Status. Dastavej Research Journal 2024

[13] Sarish R. Chavan Dr. Poonam D. Lambhate, Vaishnavi P. Gayke, Amos N. Chavan, Srushti H. Moghe. Detection of Arrhythmia using Single-Lead ECG and Deep Neural Network. International Journal of Advanced Research in Science, Communication and Technology 2023/5