IJRAR.ORG E-ISSN: 2348-1269, P-ISSN: 2349-5138

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYTICAL REVIEWS (IJRAR) | IJRAR.ORG

An International Open Access, Peer-reviewed, Refereed Journal

MEDICINAL PLANT CLASSIFICATION USING MACHINE LEARNING

Ms. Jayashree H [1] and Dr. Shashidhar Kini K [2]

[1]Student, Dept. of MCA, Srinivas Institute of Technology ,Valachil ,Mangaluru, Karnataka, India

[2] Head of the Department, Dept. of MCA, Srinivas Institute of Technology Valachil, Mangaluru , Karnataka,India

Abstract - Traditional methods for identifying medicinal plants are slow, error-prone, and reliant on botanical expertise, risking patient safety and hindering drug discovery. This paper presents an advanced medicinal plant classification system utilizing Convolutional Neural Networks (CNNs). This system addresses the shortcomings of traditional identification methods by achieving accuracy of 89.68%. It is trained with a comprehensive dataset comprising 30 unique medicinal plant species , each species represented by 500 distinct leaf images. This allows the system to accurately classify and identify plants based on complex morphological traits. Additionally, the system provides information on the medicinal uses of the identified plants. A user-friendly interface, developed with Flask, enhances accessibility and usability. Future enhancements will include expanding the database to incorporate more plant species and refining the system's capacity to handle diverse environmental conditions, thereby offering valuable support for botanists, researchers, and healthcare professionals.

Keywords - Convolutional Neural Networks (CNNs), Leaf Images, Morphological Traits, Medicinal Uses.

I. INTRODUCTION

The Medicinal Plant Bracket design is designed to advance the fields of botany and pharmacology through the development of a sophisticated system for relating and grading medicinal shops.This design addresses three primary objects easing the identification of medicinal shops, enabling medicine discovery, and enhancing point birth ways.

Precise recognition of plants is essential for researchers., herbalists, and pharmacologists. This project utilizes state-of-the-art ML algorithms and advanced image processing techniques to classify plants situate on their morphological and chemical characteristics. By creating a comprehensive and precise classification system, the project aims to streamline the identification process, making it faster and more reliable.

Enabling drug discovery is another key objective. By accurately classifying medicinal plants, the project helps in identifying species with potential therapeutic compounds. This can accelerate the discovery and development of new drugs, offering significant benefits to the pharmaceutical industry and healthcare.

Feature extraction is essential for understanding the unique properties of medicinal plant. The project employs advanced data analysis tools to extract critical features from plant data, providing researchers with valuable insights into the biochemical and morphological traits that contribute to their medicinal properties.

Medicinal plants, valued for their therapeutic properties, have been integral to traditional medicine for millennia. Recent advancements in computational methods, particularly in image processing, have shown promise in plant classification. CNNs, known for their effectiveness in image analysis, are pivotal in this context, offering significant potential for automated classification of the plants based on leaf characteristics like shape, size, colour, and texture.

This research paper presents a system for accurate classification of medicinal plants using Convolutional Neural Networks (CNNs), addressing limitations of traditional identification methods.

II. LITERATURE REVIEW

Dileep M.R et al. presents a CNN-based model called AyurLeaf for classifying medicinal plants using leaf features like shape, size, color, and texture. The dataset used in this study comprises samples from 40 medicinal plants commonly found in Kerala. By employing a neural network model based on AlexNet, the system reached a classification accuracy of 96.76% using 5-fold cross- validation. This research seeks to address the lack of expert taxonomists by offering an automated method for the precise identification and classification of medicinal plants, thereby helping to preserve traditional medicinal knowledge.

Begue et al. developed an system for identifying 24 medicinal plants using computer vision and ML techniques. Leaves were photographed, and features such as length, width, and colour were extracted. Random forest classifier achieved 90.1% accuracy, outperforming other methods. This study, the first to create a image dataset for medicinal plants in Mauritius, aims to enhance public knowledge and conservation efforts through a proposed web-based recognition system.

Lee et al. (2021) explored the application of metalearning in CNNs for medicinal plant classification. By fine-tuning pre-trained models on a specialized dataset, they achieved notable improvements in classification accuracy. The study demonstrated that transfer learning is highly effective in domains with limited data, showcasing its power to enhance model performance and accuracy in identifying medicinal plants, thus providing a valuable technique to improve classification outcomes in constrained datasets.

Wang et al. (2020) introduced a model that integrates CNNs with SVMs for medicinal plant classification. Their approach utilized CNNs for feature extraction and SVMs for classification, demonstrating enhanced accuracy compared to traditional methods. The study highlighted that combining CNNs with SVMs effectively leverages the strengths of both techniques, resulting in improved classification performance for identifying medicinal plants, showcasing a promising advancement in plant classification methodologies.

R. Geerthana, P et al. from the Department of Computer Science & Engineering, Velammal College of Engineering and Technology, Tamil Nadu, India, developed a medicinal plant identification system using Deep Learning. Their system, trained on a dataset of 58,280 images of five Indian medicinal plant species, achieved a accuracy 96.67%using CNN. The model utilizes features such as leaf texture, shape, and color, demonstrating high effectiveness for plant classification and early warning applications.

J. Samuel Manoharan, et al., from the Department of Electronics andCommunication Engineering at Sir Isaac Newton College of Engineering and Technology, Nagapattinam, India, proposed an algorithm for herbal plant identification that incorporates dimension factors to address dataset incompleteness. Their method, integrating image segmentation with machine learning and a two-stage authentication process, improved recognition accuracy and reduced classification errors. This approach enhances robustness in detecting herbal leaves, overcoming limitations in existing algorithms and improving detection rates.

III. METHODOLOGY

A. Convolutional Neural Network (CNN)

A CNN is a deep learning architecture used for tasks such as image classification. CNNs are trained to categorize images into one of many predefined classes. They are particularly effective for image processing tasks, including detection, segmentation, and classification. Unlike traditional neuralnetworks, CNN automatically identifies main features without needing explicit supervision. A CNN contains multiple layers that progressively transform the input into the output, with each hidden layer learning increasingly complex features. This hierarchical feature learning permit CNNs to excel in recognizing patterns and making accurate classifications.

B. Dataset

The dataset used in this project includes 30 unique medicinal plant species, each represented by 500 distinct leaf images. This comprehensive dataset is crucial for the CNN to accurately identify and classify a wide range of medicinal plants. By incorporating 15,000 images in total, the system can learn from diverse examples of leaf morphology, capturing subtle variations between species. This diversity ensures that the CNN can generalize well to new, unseen images, enhancing its classification accuracy and reliability. The dataset thus provides a robust foundation for developing a system capable of effective medicinal plant identification.

C. Workflow

The workflow for this project starts with the data collection stage. This involves gathering a diverse dataset of medicinal plant pictures from different sources, like botanical gardens, online databases, and field photography. Ensuring that the dataset includes vast range of species and different environmental conditions is crucial. Each image should be accurately labeled with the corresponding plant species, as this will form the basis for supervised learning during model training. The Varity and quality of collected data significantly influence the model's effectiveness and generalization capabilities.

Following data collection, the next step is data preprocessing, which prepares the images for input into the CNN Preprocessing steps include resizing all images to a uniform size normalizing pixel values to a standardized range and applying data expansion techniques such as rotations, flips, and zooms to grow the dataset's variability and robustness.

The third stage involves the development of the CNN architecture tailored for medicinal plant classification. This process includes designing a network with multiple layers, such as convolutional Layers for extracting feature, pooling layers for dimensionality reduction. and fully connected layers used for final classification.

In the fourth stage, the preprocessed dataset used to in the CNN model. This involves feeding the images into the model, adjusting weights and biases through backpropagation, and iterating over multiple epochs to minimize the loss function. The dataset is typically split into training and validation sets to monitor the model's performance and prevent overfitting. Hyper parameters such as rate of learning, batch size, and also number of epochs were fine-tuned to optimize the training process.

The final stage focuses on evaluating the trained model and deploying it for practical use. A confusion matrix is generated to visualize the model's predictions versus actual labels, highlighting any misclassifications. Once the model achieves satisfactory performance, it is deployed as webapplication using frameworks like Flask. This application provides a userfriendly interface where users can upload images of medicinal plants and receive real-time classifications.

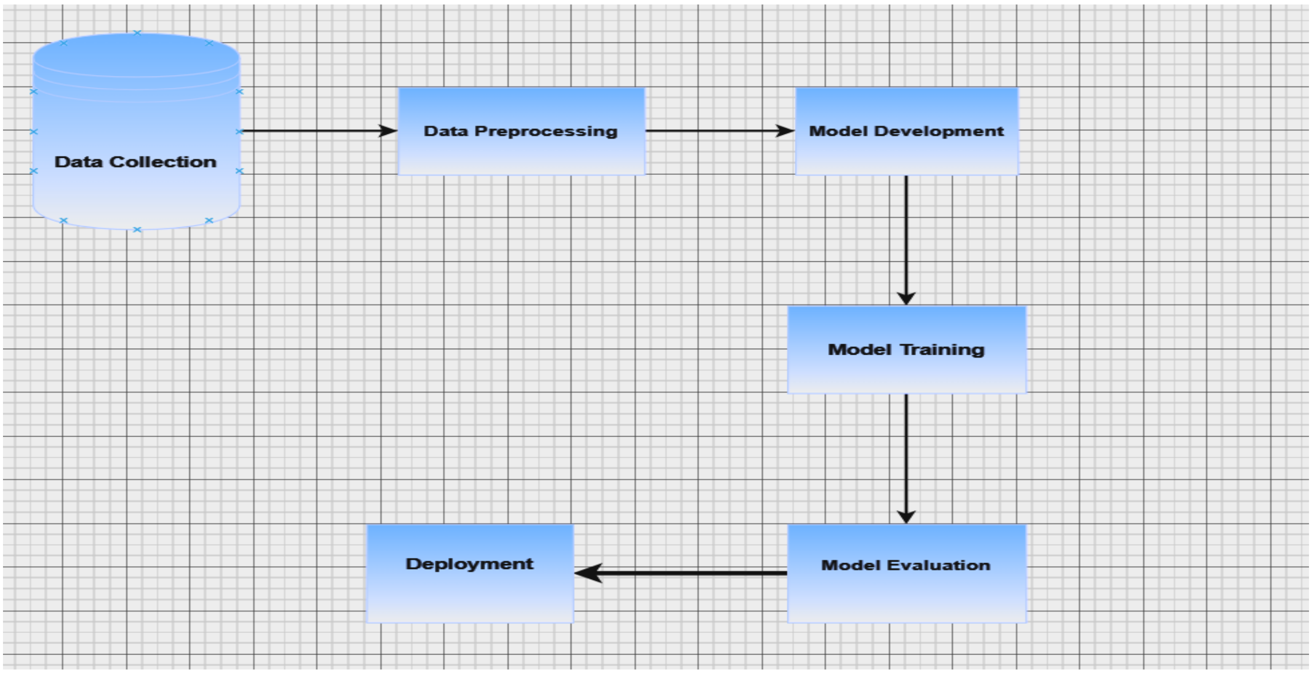

Fig 1 Workflow Diagram

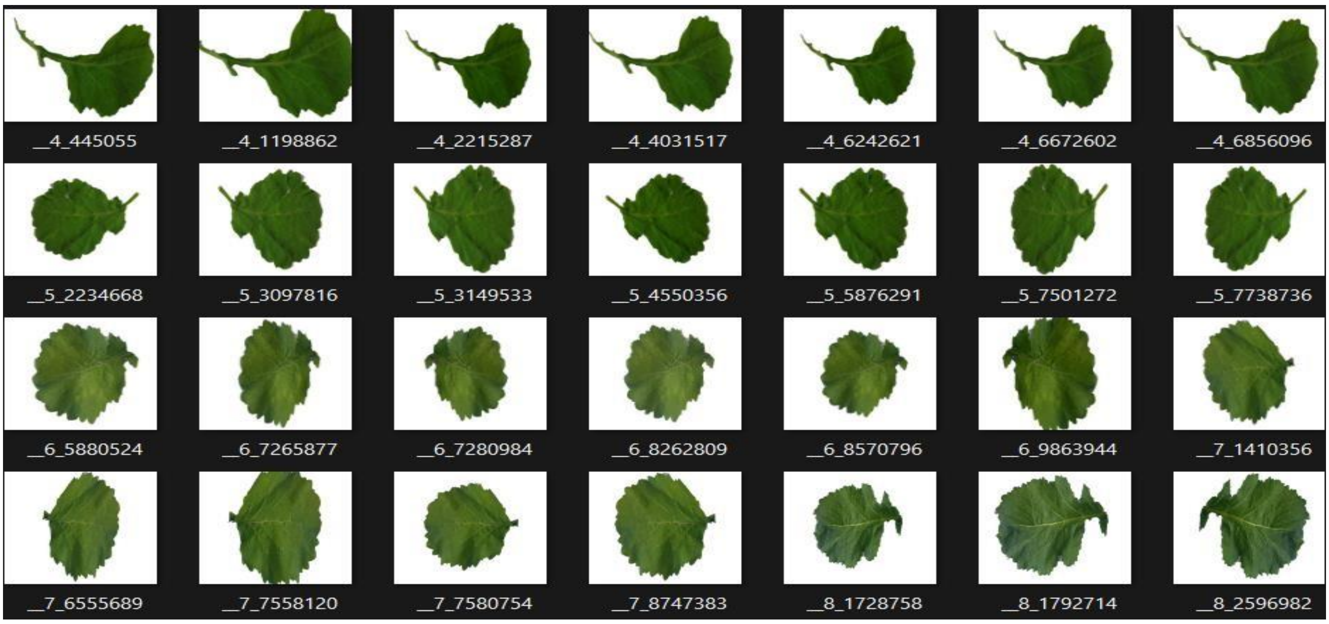

Fig 2 .Indian mustard leaf with different orientation

Figure 3 displays a grid of 20 randomly selected images from a dataset. The images arranged in a 4x5 layout, with each image coming from different folders within the dataset. Each subplot is labeled with the name of the folder from which the image was chosen. This visualization helps to provide a quick overview of the variety and content of the images in dataset.

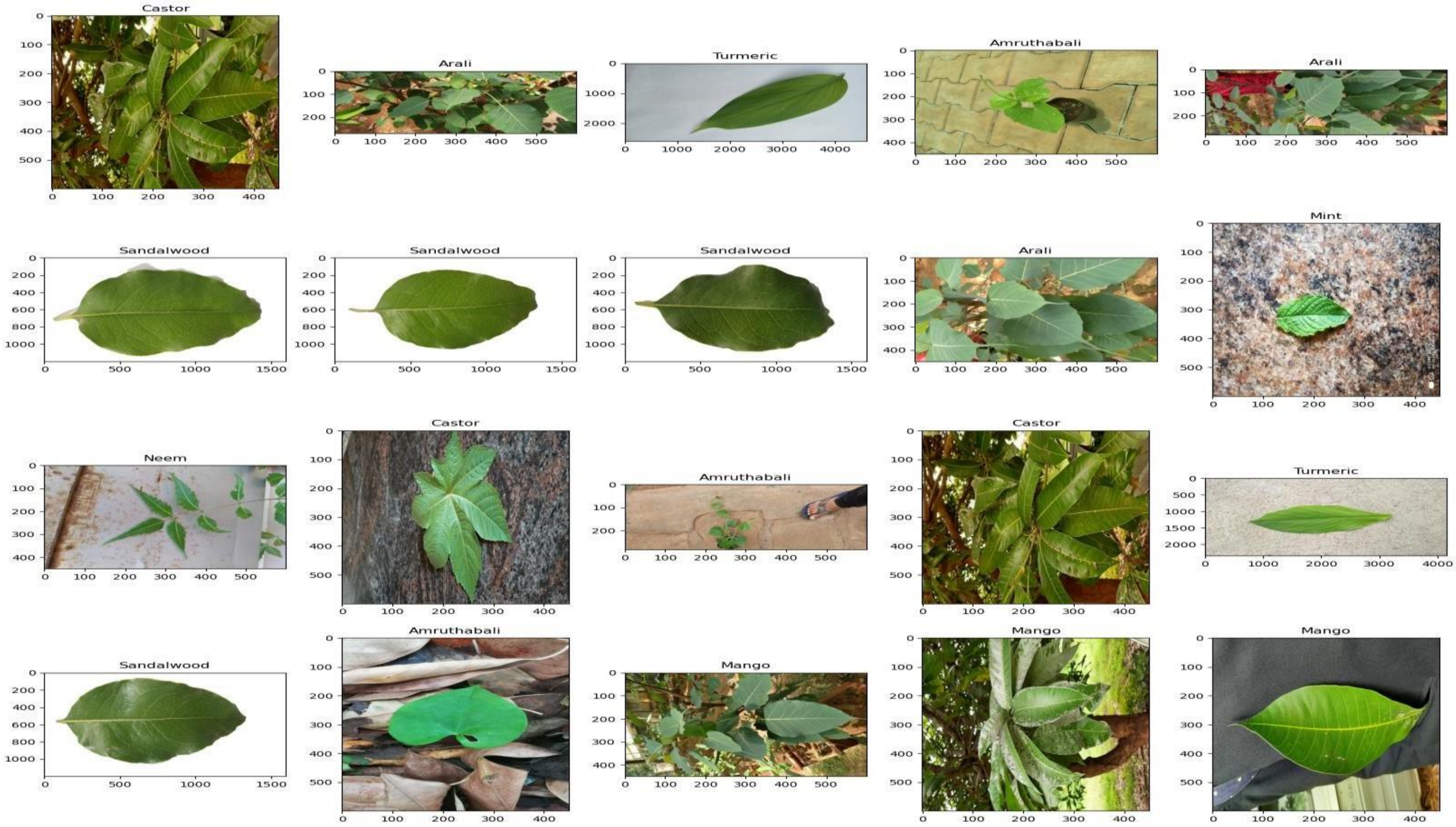

Fig 3. Image Samples with Folder Labels

IV. RESULTS AND DISCUSSION

ACCURACIES PLOT

The graph illustrates the preparation and validation accuracy of the model for classifying medicinal plant leaves over four epochs. The blue line, representing training accuracy, shows a consistent increase, indicating effective learning from the training data. The orange line, representing validation accuracy, rises initially but then slightly declines after the first epoch, suggesting potential over fitting as the model continues to train. The close alignment of the validation and training accuracy lines indicates that the model has good generalization performance. However, the slight drop in validation accuracy suggests a need for further tuning or regularization to maintain consistent performance and prevent over fitting in future epochs.

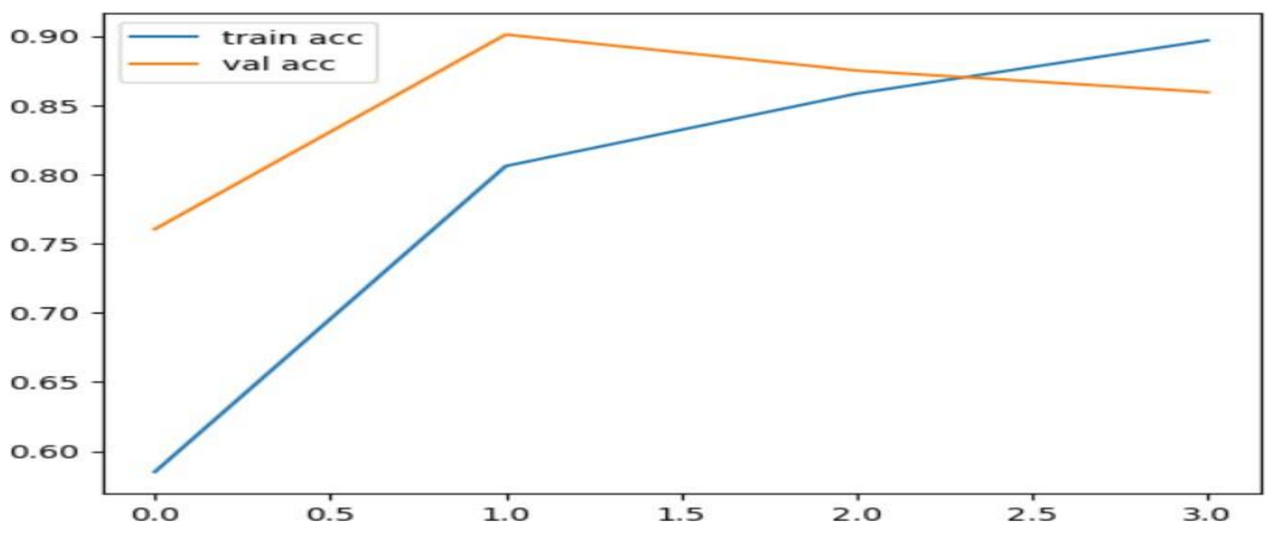

Fig 4. Accuracies Plot

LOSS PLOT

The graph illustrates the loss of a CNN model for classifying medicinal plant leaves over several epochs. Training loss and loss in validation are plotted, where the blue line shows training loss and the orange line indicates validation loss. As training loss consistently decreases, indicating that model is effectively learning and minimizing errors on the training data. Initially, the validation loss also decreases, suggesting improved model performance on unseen data. However, a slight increase in validation loss after the first epoch hints at over fitting, where the model starts to memorize training data instead of generalizing well. Regularization techniques and more data might be necessary to address this issue.

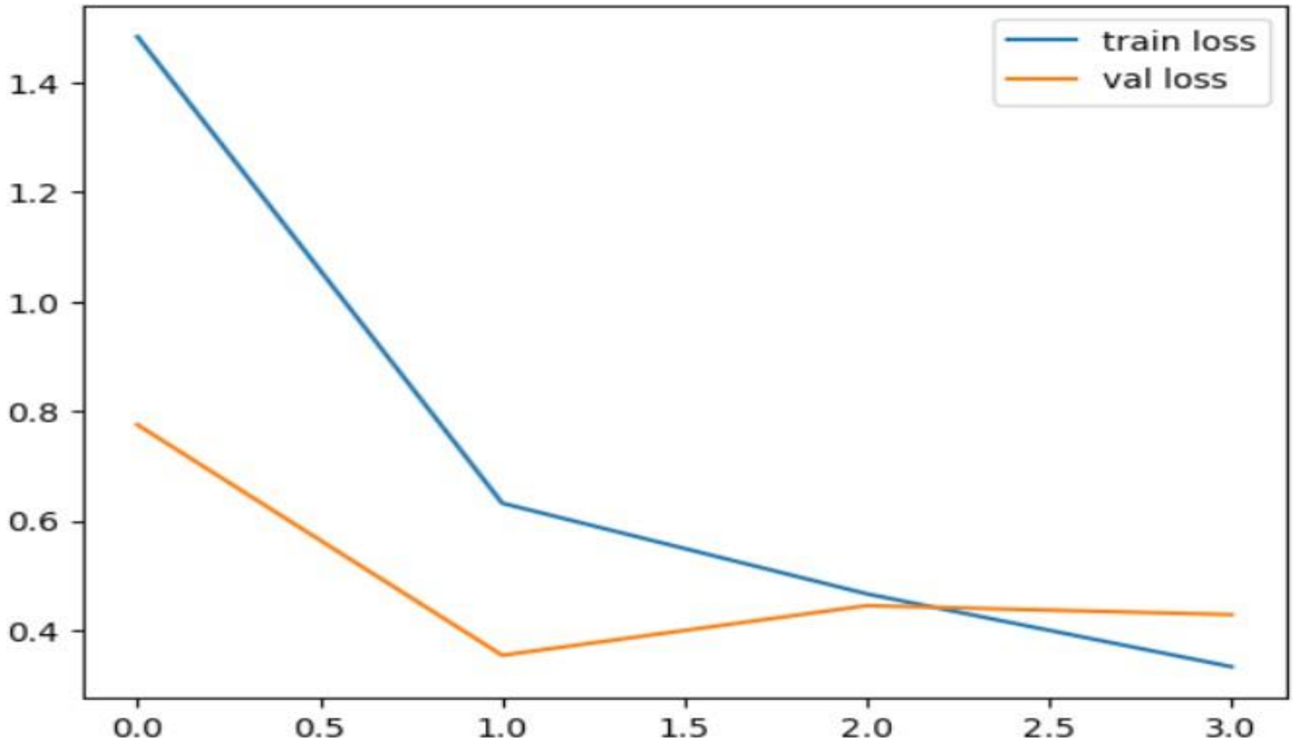

Fig 5. Loss Plot

V. CONCLUSION AND FUTURE WORK

The CNN-based medicinal plant classification system successfully identifies plants with 89.68% accuracy, providing valuable information on medicinal uses. The user-friendly web application enhances accessibility and practical use. Expanding the dataset to include more species, refining the model for diverse conditions, and incorporating additional features will further improve accuracy and robustness.

REFERENCES

[1] Begue, Adams, Venitha Kowlessur, Upasana Singh, Fawzi Mahomoodally, and Sameerchand Pudaruth. "Automatic recognition of medicinal plants using machine learning techniques." *International Journal of Advanced Computer Science and Applications* 8, no. 4 (2017): 166-175.

[2] Geerthana, R., P. Nandhini, and R. Suriyakala. "Medicinal Plant Identification Using Deep Learning." *International Research Journal on Advanced Science Hub* 3 (2021): 48-53.

[3] Dileep, M. R., and P. N. Pournami. "AyurLeaf: a deep learning approach for classification of medicinal plants." In *TENCON 2019-2019 IEEE Region 10 Conference (TENCON)*, pp. 321-325. IEEE, 2019.

[4] Hu, Renchuan, Chunrui Lin, Weibin Xu, Yan Liu, and Chunlin Long. "Ethnobotanical study on medicinal plants used by Mulam people in Guangxi, China." *Journal of ethnobiology and ethnomedicine* 16 (2020): 1-50.

[5] of selected medicinal plants leaf using image processing. In 2012International Conference on Machine Vision and Image Processing(MVIP), pages 5–8, Dec 2012.

[6] Du, M., Zhang, S. and Wang, H., 2009. Supervised Isomap for PlantLeaf Image Classification. 5th International Conference on EmergingIntelligent Computing Technology and Applications, Ulsan, SouthKorea, 627-634.