**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**COLLEGE OF ENGINEERING GUINDY**

**ANNA UNIVERSITY, CHENNAI – 25.**

**PROJECT RECORD**

CS6301 – MACHINE LEARNING

**MEDICINAL PLANT LEAF CLASSIFICATION**

**TEAM MEMBERS**

SANGEETHA S -2022103029

VENIZHA R-2022103530

DHIVYA S-2022103029

**PROJECT GUIDE: DR. MOHAMED FATHIMAL P**

MEDICINAL PLANT LEAF CLASSIFICATION

**Project Guide**

Dr.Mohamed Fathimal P Asst.Professor, DCSE,

College of Engineering, Guindy Anna University.

Sangeetha S - 2022103023, Dhivya S - 2022103029, Venizha R - 2022103530

***Abstract -* Medicinal plant classification is crucial for herbal medicine, agriculture, and biodiversity conservation. Traditional methods of plant species identification rely on manual inspection, which is often time-consuming and prone to errors. Machine learning (ML) offers a more efficient and accurate solution by automating the identification process using leaf images. This approach applies various ML techniques, such as Convolutional Neural Networks (CNNs), to analyze and classify plant species based on their leaf characteristics. The process involves image preprocessing, feature extraction, model training, and evaluation. Emphasis is placed on achieving high classification accuracy while ensuring scalability for large datasets. This work highlights the potential of ML to transform plant identification, enhancing research and the sustainable use of medicinal plants.**

***Index Terms*—** **Machine Learning, Medicinal Plant Classification, Convolutional Neural Networks, Leaf Image Classification.**

1. INTRODUCTION

Identifying medicinal plants has long been a vital aspect of preserving traditional herbal knowledge and promoting the sustainable use of natural resources. Throughout history, plants have served as the foundation for numerous medicinal remedies, and their proper identification ensures the protection of these valuable resources. Traditionally, plant identification involved manual inspection, often relying on physical features such as leaf shapes, flowers, and growth habits. However, this method is time-consuming, prone to human error, and can be inconsistent across different practitioners. With the rapid advancements in machine learning (ML), this process can now be automated, making plant identification more efficient, accurate, and accessible.

In this project, we aim to build a machine learning-based system that can classify medicinal plants by analyzing leaf images. This system leverages various ML algorithms to predict the species of plants based on their leaf characteristics. We explored a range of well-established machine learning techniques, including Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Convolutional Neural Networks (CNN), XGBoost, Random Forest, and Decision Trees. Each of these models was trained on a dataset containing images of plant leaves, with the goal of classifying the plants into specific species based on the visual features of their leaves.

The performance of the models was significantly enhanced by incorporating advanced feature extraction techniques, such as Scale-Invariant Feature Transform (SIFT) and Histogram of Oriented Gradients (HOG). These methods are designed to capture important structural and textural information from the leaf images, which are essential for distinguishing between different plant species. SIFT, for instance, is a technique that detects key points and extracts distinctive local features from the images, making it robust to changes in scale, rotation, and illumination. HOG, on the other hand, focuses on capturing the gradients and orientations of edges in the image, which are useful for detecting patterns that are crucial for classification tasks.

To further improve the accuracy of the classification, we applied Principal Component Analysis (PCA) for dimensionality reduction. PCA helped in reducing the complexity of the data by transforming the features into a lower-dimensional space while retaining the most important aspects of the data. This process not only speeds up the training time but also reduces the risk of overfitting by eliminating redundant features. With these feature extraction and dimensionality reduction techniques, we were able to ensure that the most relevant and distinguishing characteristics of the leaf images were used for the classification task, leading to more accurate predictions.

The project also involved the development of a user-friendly front-end interface using Streamlit, a Python framework known for its simplicity and efficiency in building interactive web applications. This interface enables users to upload leaf images easily and receive real-time predictions about the species of the plant. The system automatically processes the uploaded image, extracts the necessary features, and classifies the plant using the trained machine learning models. This interactive system makes the tool accessible not only to researchers but also to the general public, making plant identification more accessible to a wider audience.

After training and testing all the models, we found that Convolutional Neural Networks (CNN) outperformed the other algorithms in terms of classification accuracy. CNNs are particularly powerful for image classification tasks because they can automatically learn the most relevant features from the images, reducing the need for manual feature extraction. By learning complex patterns and hierarchical representations of features, CNNs were able to achieve impressive results in classifying medicinal plants.

In addition to CNN, we also tested Random Forest and Decision Tree models, which performed well in handling large datasets with complex and varying features. These tree-based models excel in handling non-linear relationships and are known for their interpretability, making them useful for understanding the decision-making process behind the classification. Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) were also implemented to compare their performance with the more complex models. While these algorithms did not perform as well as CNN in this particular task, they offered valuable insights into how different machine learning techniques handle image data and classification tasks.

To further boost the model performance, we employed XGBoost, an advanced boosting algorithm that focuses on optimizing the model by concentrating on the harder-to-classify instances. By focusing on these difficult cases, XGBoost significantly improved the overall accuracy of the predictions, especially for cases where the plant species were harder to distinguish based on leaf features alone. This technique made the model more robust, enhancing its ability to generalize across different plant species.

In terms of evaluating the models, we utilized standard classification metrics such as accuracy, precision, recall, and F1-score. Accuracy provided a general measure of the model's performance, but given the potential class imbalances in the dataset, precision and recall were also calculated to assess how well the model could correctly identify both positive (malicious) and negative (benign) instances. The F1-score, which is the harmonic mean of precision and recall, was particularly useful in understanding the balance between the two metrics, especially when dealing with imbalanced datasets where one class might be underrepresented. These evaluation metrics allowed for a comprehensive comparison of each model's performance and helped identify the most effective algorithm for the task.

In conclusion, this project highlights the power of machine learning in automating the identification of medicinal plants. By integrating feature extraction techniques such as SIFT and HOG with PCA for dimensionality reduction, we achieved a high level of accuracy in classifying plants based on leaf images. The system we developed offers an efficient, scalable, and user-friendly solution for plant identification. With its potential to assist in fields such as herbal medicine, agriculture, and biodiversity conservation, this project has the potential to contribute to both scientific research and the preservation of valuable plant species. This work demonstrates that, with the right combination of machine learning techniques and image processing, it is possible to create a reliable and effective system for identifying medicinal plants in a more automated and accessible manner, ensuring the continued preservation and sustainable use of these important resources.

1. Literature Review

**[1]** Guyer, M., Meier, R., & Horne, J. (1993) conducted a study focused on plant identification through machine vision, with an emphasis on leaf shape analysis. Their research aimed to identify plant species by analyzing the shape of leaves, using images of different plant species as the dataset. They employed

machine vision techniques to extract key features, particularly the shape of the leaves, for classification purposes. The study successfully demonstrated that leaf shapes can be highly effective for plant classification, achieving reliable and accurate results. The authors concluded that machine vision techniques could be a powerful tool for automating the plant identification process, offering an efficient and precise method for classifying plant species based on leaf morphology. The findings of the study contributed to the broader field of computer vision applications in biology, emphasizing the potential of automated systems for plant identification.

# [2] He, K., Zhang, X., Ren, S., & Sun, J. (2015) introduced the rectified linear unit (ReLU) as an activation function in deep learning models, aiming to enhance the performance of image classification tasks. They used the large-scale ImageNet dataset, which consists of millions of labeled images across thousands of categories, to evaluate the effectiveness of ReLU compared to traditional activation functions like sigmoid and tanh. The study showed that ReLU not only significantly improved the training speed but also helped achieve near human-level performance on classification tasks. ReLU allows for faster convergence in deep neural networks by overcoming issues such as the vanishing gradient problem, which is common in traditional activation functions. As a result, the paper demonstrated that deep learning models utilizing ReLU activation can classify images with higher accuracy, leading to advancements in image recognition tasks. The study emphasized the importance of ReLU in the development of more efficient deep learning architectures, paving the way for its widespread adoption in the field of artificial intelligence.

# [3] He, K., Zhang, X., Ren, S., & Sun, J. (2016) advanced the field of deep learning by proposing the use of deep residual networks (ResNets) for image recognition tasks. ResNets are a novel architecture that incorporates residual connections, which allow information to bypass certain layers of the network, helping to mitigate the problem of vanishing gradients that is commonly encountered when training very deep networks. This breakthrough approach enables the training of much deeper networks than previously possible. The authors demonstrated the effectiveness of ResNets by achieving state-of-the-art performance on the ImageNet dataset, setting new benchmarks in image classification. The study revealed that the inclusion of residual connections made it feasible to train networks with hundreds or even thousands of layers without a degradation in performance, something that was not possible before. Their findings underscored the critical role of residual learning in deep learning, enhancing the capabilities of neural networks for complex image recognition tasks and marking a significant advancement in the field of computer vision.

# [4] Hinton, G. E., Osindero, S., & Teh, Y. W. (2006) introduced a fast and efficient learning algorithm for Deep Belief Networks (DBNs), aiming to address the challenges in training deep neural networks. The study utilized the MNIST dataset, a standard benchmark in machine learning, which contains images of handwritten digits, to assess the performance of their proposed algorithm. The authors employed a method known as contrastive divergence to expedite the training process, significantly reducing the computational cost typically associated with deep learning models. As a result, DBNs trained with this approach showed improved accuracy in tasks like digit recognition. This advancement in training efficiency allowed DBNs to scale to more complex tasks, making it a crucial step in the evolution of deep learning techniques. The paper demonstrated that with the right learning algorithm, DBNs could be trained faster and more effectively, paving the way for their application in larger, real-world datasets. The authors concluded that their method provided a breakthrough in optimizing deep learning models, making them more practical and capable of handling the challenges presented by large-scale datasets in various fields of application

# [5] Huixian, Z. (2020) explored the application of deep learning techniques, specifically Convolutional Neural Networks (CNNs), in recognizing plant species from images. This paper focuses on the use of CNNs for classifying various plant species based on their visual characteristics, with a dataset containing images of plants from multiple species. CNNs have become the go-to solution for image-based classification tasks, as they automatically learn the relevant features of an image without the need for manual feature extraction. In this study, the accuracy of plant species classification using CNNs was the key performance metric, and the results showed that CNNs performed exceptionally well, achieving high classification accuracy. This demonstrated that deep learning models, and in particular CNNs, can effectively identify plant species from images with minimal pre-processing. The study highlights the growing application of CNNs in botanical research, as they offer a highly accurate and automated method for plant species recognition, which is particularly useful for large-scale or field-based classification tasks. The conclusion emphasizes that deep learning techniques like CNNs are not only efficient but also powerful tools for plant identification, offering potential for expanding automated plant classification systems.

# [6] Im, S. J., Lee, M. J., & Park, S. Y. (1998) focused on plant species identification through the analysis of leaf shapes, an essential aspect of botany and plant classification. In their study, the dataset consisted of leaf images from the Acer plant family, which were analyzed using shape analysis and pattern recognition techniques to identify the plant species. The accuracy of the recognition process was measured based on how well the algorithm could differentiate between species based on their leaf shapes. The results showed that leaf shape was a highly distinctive feature that could be used to reliably classify plant species. The study highlighted that certain features of leaf shape, such as size, edges, and overall structure, are unique to specific plant species and thus can serve as a strong indicator for classification. This method of plant identification does not require complex equipment or invasive techniques, making it an ideal approach for field research. The

# paper concluded that leaf shape analysis is an effective and

# practical method for identifying plant species, offering a cost-effective alternative to genetic testing or other more complicated identification methods. The study contributed to the understanding that morphological features like leaf shape are key to plant classification and can be used to develop automated systems for plant species recognition.

**[7]** Jia, Y., Shelhamer, E., & Donahue, J. (2014) introduced the Caffe framework, a deep learning library designed to accelerate the process of feature extraction and image recognition tasks. The paper presented Caffe’s capabilities in handling large datasets like ImageNet, a widely used dataset for image classification that contains millions of images categorized into thousands of classes. The core feature of Caffe is its ability to process large-scale image data quickly while maintaining high classification accuracy. The authors demonstrated that Caffe outperformed previous image recognition methods in both processing speed and classification accuracy. Caffe’s modular structure and optimization techniques made it particularly suitable for use in both research and industry, offering a significant boost in performance for image recognition tasks. The study also highlighted the framework's flexibility, as it supports various deep learning models, including CNNs, and can be easily adapted for specific use cases. The conclusion of the study emphasized that Caffe’s efficiency and speed made it a valuable tool for large-scale image recognition tasks, particularly in fields where quick processing times and high accuracy are crucial, such as medical imaging or autonomous vehicles. Caffe’s contribution to the field of deep learning was a step forward in making large-scale image recognition tasks more feasible and efficient, ensuring its continued use in research and industrial applications.

# [8] Karami, M., & Asgarian, E. (2017) conducted a comprehensive review on the role of medicinal plants in cancer treatment, focusing on the potential of these plants to serve as therapeutic agents for cancer. By examining various ethnobotanical studies and reviewing literature on the medicinal properties of plants, the authors identified a number of plant species that exhibit anticancer properties. These plants, which have been used traditionally in different cultures for treating various ailments, have shown promise in fighting cancer due to their natural bioactive compounds. The review provided valuable insights into the types of plants that have been historically used in cancer treatments, including herbs, roots, and barks, and discussed their pharmacological effects. While this was a qualitative analysis rather than a quantitative study, the paper emphasized the importance of further research into the potential benefits of these plants, suggesting that they could be integrated into modern cancer treatment regimens either alone or in combination with conventional therapies. The study concluded by stressing the importance of ethnobotanical research as a key resource for discovering new medicinal plants that may hold the key to moreeffective cancer treatments. The review also encouraged further scientific investigation into these plants to validate their therapeutic potential, which could lead to the development of novel, natural cancer therapies.

**[9]** The study by Kumar, P., Yadav, S., & Soni, S. (2019) presented a novel approach to improving plant species recognition by integrating morphological feature extraction with adaptive boosting techniques. The researchers focused on leveraging key leaf characteristics such as shape, size, and texture, which are widely used in plant classification tasks. These morphological

features were extracted from images of leaves across different plant species, and they served as the primary descriptors for distinguishing between species. Shape encompassed the overall outline and symmetry of the leaf, while size referred to the dimensions, such as length and width, which can vary significantly across species. Texture, which included patterns like veins and surface smoothness, provided additional discriminative power. Once these features were extracted, they were used as input for a machine learning algorithm, adaptive boosting, or AdaBoost. AdaBoost is a technique that combines multiple weak classifiers, each performing only marginally better than random guessing, to create a strong classifier. The boosting process works by giving more weight to misclassified examples, thereby forcing the algorithm to focus on difficult cases that previous classifiers failed to handle. The sequential nature of AdaBoost allows it to correct errors iteratively, leading to improved performance in classification tasks. The study demonstrated that when morphological features were enhanced by the boosting algorithm, the recognition accuracy of plant species increased significantly compared to using morphological features alone. The results showed that AdaBoost helped reduce classification errors, especially when dealing with complex and diverse datasets, indicating that combining feature extraction with boosting algorithms is a powerful method for improving the effectiveness of plant recognition systems. The authors concluded that this hybrid approach is highly beneficial for handling large datasets with varying plant species, as it not only improves accuracy but also increases the robustness of the classification model.

**[10]** This study aimed at classifying legume species based on the distinctive vein patterns present in their leaves. The study specifically targeted a dataset of leaf images from various legume species, with a primary focus on the intricate and unique vein structures that often serve as key identifiers for plants. The researchers utilized feature extraction techniques to analyze these vein patterns and to derive meaningful features that could aid in differentiating between species. Leaf veins are known to be consistent within a particular species but exhibit notable variations across different species, making them an excellent tool for plant identification. The study showed that by focusing on the vein patterns, the model was able to achieve high classification accuracy, demonstrating the effectiveness of this approach. The results emphasized that leaf veins offer a reliable and significant morphological feature that can be used to accurately classify legumes. The researchers also noted that these vein patterns are particularly useful in the context of legumes, where traditional morphological features like shape and size may not always provide enough distinguishing power due to the similarities between species. The conclusion of the study reaffirmed that leaf vein patterns could play a crucial role in developing automated plant species classification systems, especially for legumes, which are important in both agricultural and ecological studies. By incorporating vein-based features, such systems could potentially streamline the process of identifying legume species in various research and field applications.

**[11]** Lulekal, E., & Asfaw, Z. (2008) conducted an ethnobotanical study to explore the medicinal plants traditionally used by local communities in southeastern Ethiopia for treating a wide range of diseases. The study was designed to identify and document the plants that the indigenous population relied on for medicinal purposes, gathering data through ethnobotanical surveys. The research involved interviews with local community members,

providing firsthand accounts of plant-based treatments that had been passed down through generations. The dataset consisted of detailed responses from individuals within these communities, revealing a rich variety of plants used to treat ailments such as infections, digestive issues, respiratory conditions, and skin diseases. Unlike quantitative studies, this research did not focus on traditional performance metrics like accuracy or statistical analysis but instead emphasized the qualitative richness of the data collected. The study highlighted the profound knowledge that local communities possess about the medicinal properties of plants, much of which has been orally transmitted and preserved over centuries. The findings underscored the diversity of medicinal plants in the region, pointing out the potential these plants hold for developing new, natural therapies. Many of these plants, which have long been used in traditional healing practices, have demonstrated effectiveness in treating various diseases, offering a valuable resource for pharmaceutical research and development. The study concluded by stressing the importance of preserving and further investigating the traditional knowledge and plant species used by these communities. Such preservation is crucial not only for safeguarding cultural heritage but also for fostering future research that could uncover novel compounds or treatments for diseases. The authors also called for collaborative efforts between ethnobotanists, local communities, and modern medical researchers to ensure that the medicinal potential of these plants is not lost in the face of modern medical advancements. This research further demonstrated the relevance of indigenous knowledge in the field of medicinal plant discovery and highlighted the need for further studies to explore the therapeutic potential of these plants in contemporary medicine.

**[12]** Muneer, A., & Fati, O. (2020) investigated the automation of herb species classification using deep learning techniques, with a specific focus on Convolutional Neural Networks (CNNs). The study aimed to classify herb species based on the distinctive leaf shape and texture, using a dataset of herb leaf images. These images were pre-processed to extract relevant features before being fed into the model. The use of CNNs was pivotal in this study, as they are particularly well-suited for image recognition tasks due to their ability to automatically learn hierarchical features from raw image data, negating the need for manual feature extraction. This deep learning model, which mimics the visual processing in the human brain, excels in recognizing patterns in images by applying multiple layers of filters that detect various aspects like edges, shapes, and textures. By leveraging CNNs, the model was able to identify key features related to the leaf’s shape and texture, which are crucial for differentiating herb species. The performance of the CNN model was evaluated using classification accuracy, and the results showed that the model achieved high accuracy in identifying various herb species, making it a reliable method for automated plant classification. The study demonstrated that CNNs are highly effective when applied to tasks involving complex visual data, such as plant species classification, especially when distinguishing features are based on fine details like leaf texture and shape. The success of this approach indicates that CNNs can significantly enhance the accuracy and efficiency of plant identification systems. The conclusion of the paper affirmed that CNNs not only offer an effective method for automating plant species classification but also suggested that this approach could be expanded beyond herbs to other plant types. Additionally, the study highlighted the potential for integrating CNN-based plant classification systems into mobile applications, providing users with an easy, on-the-go solution for plant identification. Such advancements in deep learning could lead to the widespread adoption of automated plant identification

systems, making them accessible for both casual users and professional researchers alike.

**[13]** Oide and Ninomiya (2000) focused on the use of neural networks to classify soybean leaves, aiming to automate the identification of different varieties of soybean plants based on the features of their leaves. The study utilized a dataset consisting of soybean leaf images, which captured various characteristics such as texture, shape, and color. These features were essential in distinguishing between different soybean varieties. The researchers applied artificial neural networks (ANNs), which are computational models inspired by the human brain and excel at learning patterns from data. In this case, the neural network was trained to recognize the unique features present in the soybean leaves and use them to classify the plant varieties. The key objective was to assess how well ANNs could learn and differentiate these features in order to classify soybean varieties accurately. The evaluation of the model was based on classification accuracy, and the results demonstrated that the neural network performed well, providing promising results for plant classification tasks. The study concluded that ANNs were highly effective for classifying soybean leaves and suggested that this approach could be extended to other plant species as well. By leveraging the ability of neural networks to automatically learn complex patterns from data, the research highlighted their potential in automating the classification of plant species, especially in agricultural contexts where large numbers of plant varieties need to be identified quickly and accurately. This work contributed to the growing body of knowledge on how artificial intelligence can be applied to plant classification, paving the way

for more advancedapplications in agriculture and botany**.**

**[14]** Patil and Bhagat (2016) reviewed various methods for plant species identification based on the shape of their leaves. The authors explored techniques that focus on extracting and analyzing the geometric properties of leaves, such as their contours, sizes, and overall shapes. The premise of these methods is that each plant species has distinct leaf shapes, which can serve as a unique identifier for classification. The study examined a dataset consisting of images from various plant species, each representing different leaf characteristics, and assessed how these shape-based features could be used to identify plants accurately. Although the paper did not provide quantitative performance metrics, it highlighted the effectiveness and reliability of shape-based recognition methods for plant identification. By focusing on the geometric properties of leaves, these methods are able to differentiate between species, even in cases where other characteristics, such as color or texture, may not be as distinct. The study emphasized that leaf shape is a critical feature for distinguishing between plant species, as it is often a highly stable and consistent trait across individuals of the same species. Furthermore, the paper suggested that shape-based recognition techniques could be particularly beneficial in fields like botany and ecological research, where accurate plant identification is essential. Overall, the study concluded that methods relying on leaf shape recognition offer a promising approach for automating plant identification, with potential applications in ecological studies, biodiversity monitoring, and conservation efforts.

**[15]** Pushpanathan (2021) explored the use of machine learning, particularly deep learning techniques, for the identification of medicinal plants. The review focused on how various machine learning models, especially convolutional neural networks

(CNNs), have been applied to plant classification tasks, particularly in the context of medicinal plant recognition. The study highlighted the growing trend of utilizing deep learning methods for plant identification, which allows for the automation of recognizing plants based on features like leaf structure, texture, and other morphological characteristics. The dataset referenced in the paper was obtained from ethnobotanical research, which provided detailed information on plants that are used for medicinal purposes in different cultures. This dataset included images and descriptions of plants known for their therapeutic properties, serving as a rich resource for training machine learning models. While the paper did not present specific quantitative results or performance metrics, it emphasized the increasing role of deep learning models in classifying plants, particularly those with medicinal value. The authors discussed the advantages of using machine learning, noting that these techniques can handle large and complex datasets, automatically learning the distinguishing features of plants without the need for manual feature extraction. The review concluded that machine learning, andparticularly deep learning, has significant potential for the identification of medicinal plants, offering an

effective solution for automating this process. This could greatly aid ethnobotany by facilitating the cataloging of medicinal plants, and contribute to conservation efforts by identifying plant species that are at risk of extinction or in need of protection. Ultimately, the

paper underscored the importance of applying advanced machine learning techniques in the field of medicinal plant research, which could lead to the discovery of new medicinal properties and enhance the preservation of valuable plant species.

**[16]** Qingfeng (2007) focused on feature extraction techniques aimed at automating plant leaf recognition. The primary objective of their research was to develop an efficient method for identifying plants by extracting distinctive features from leaf images, such as shape, texture, and color. The dataset used in the study included images from various plant species, with each image displaying unique characteristics that could be leveraged for classification. The study emphasized that the selection of the right features from leaf images is crucial for accurate plant identification. Through feature extraction, key attributes such as the edges, veins, and patterns within the leaves were identified, which helped in distinguishing one species from another. The research demonstrated that these extracted features, when combined with classification algorithms, enabled the system to correctly identify the plant species based on the visual characteristics of their leaves. The study showed that the accuracy of the plant recognition system was highly dependent on the effectiveness of the feature extraction process. The authors concluded that feature extraction is a vital component of plant identification systems, as it enables the system to focus on the most important and distinguishing characteristics of leaves. By improving the precision of feature extraction, plant recognition systems can be made more reliable and accurate, which is especially valuable for applications in botany, agriculture, and environmental research. The research highlighted that advancements in feature extraction techniques have the potential to significantly enhance the performance of automated plant recognition systems, providing a foundation for more accurate and scalable plant identification solutions.

**[17]** Sermanet (2014) aimed to enhance both object recognition and localization within images using convolutional neural networks (CNNs). Their approach integrated these two tasks—recognizing objects and determining their exact locations—into a single unified model. The study utilized the ImageNet dataset, which contains a large variety of labeled images across numerous categories, to

evaluate the performance of their proposed method. Traditionally, object recognition and localization were treated as separate tasks, with models optimized either for recognizing objects or for determining their location in an image. However, this study showed that by integrating both tasks into a single CNN-based model, better performance could be achieved for both recognition and localization.The results of the study demonstrated that the CNN model outperformed previous methods that focused on either task independently, offering a more efficient solution for complex visual recognition tasks. The integrated model not only improved the accuracy of identifying objects but also achieved better precision in locating them within an image, making it especially useful for tasks that require both aspects to be addressed simultaneously. The authors concluded that CNNs, with theirability to learn hierarchical features directly from data, were particularly well-suited for applications requiring precise object detection and localization.This approach has significant implications for various fields, such as computer vision, robotics, and autonomous vehicles, where accurate object recognition and localization are critical. For example, in robotics and autonomous vehicles, understanding both the identity and the location of objects in a dynamic environment is essential for navigation and decision-making. The research highlighted the potential of CNNs in solving real-world problems that demand high accuracy in both recognition and localization, pushing the boundaries of what is achievable in automated image analysis and object detection.

**[18]** This paper focused on developing an automatic plant identification system using deep learning, particularly convolutional neural networks (CNNs). The study aimed to leverage the power of deep learning to automate the identification of plant species by analyzing leaf features from a large dataset of labeled plant images. Each image in the dataset was associated with a specific plant species, and the CNN model was trained to recognize and classify these species based on visual characteristics such as leaf shape, texture, and color. The research demonstrated that CNNs were highly effective at accurately classifying plant species, with the deep learning model achieving high performance in identifying different species. The results of the study confirmed that CNNs are particularly well-suited for plant identification tasks, as they can automatically learn the relevant features from raw image data without requiring manual feature extraction. This capability is one of the key advantages of deep learning techniques, enabling the system to handle large and diverse datasets with ease. The model’s high accuracy in classifying plant species showcased the potential of deep learning in solving real-world problems related to plant identification. The authors concluded that deep learning models, particularly CNNs, provided a reliable, scalable, and efficient approach for plant identification. They emphasized that such methods could be applied to large-scale plant recognition systems, which could be invaluable in various domains such as botany, agriculture, and biodiversity conservation. By automating plant identification, these systems could assist researchers, farmers, and conservationists in identifying plant species quickly and accurately, improving the management of plant resources and contributing to efforts in preserving biodiversity. The study highlighted the transformative potential of deep learning in advancing plant recognition technology and its applications in environmental monitoring and conservation.

**[19]** Zhang (2019) focused on improving the accuracy of plant leaf classification through the application of deep learning techniques, particularly convolutional neural networks (CNNs).

The study used a dataset that consisted of labeled images of leaves from various plant species, with each image corresponding to the specific plant species it represented. The primary goal was to develop a model that could automatically learn and extract relevant features from leaf images to classify them into their respective species accurately.The authors employed CNNs, which are particularly well-suited for image-based tasks due to their ability to automatically learn hierarchical features from raw data. In this case, the CNN model learned complex patterns and distinguishing characteristics from the leaf images, such as texture, shape, and other subtle details that might not be immediately obvious to the human eye. The results of the study showed that the deep learning model achieved high classification accuracy, significantly outperforming traditional machine learning methods and manual classification approaches. This success demonstrated the effectiveness of CNNs in distinguishing between various plant species based on the unique features present in their leaves.The authors concluded that CNNs were highly effective for plant leaf classification tasks, providing a reliable method for identifying plants based on their leaves. The research underscored the potential of deep learning in plant identification, especially in scenarios where large and diverse datasets are involved. The study suggested that such deep learning models could be applied to a wide range of botanical and environmental applications, including biodiversity monitoring, environmental conservation, and agricultural management. By automating the process of plant classification, these models could significantly enhance the efficiency and accuracy of plant species identification, contributing to various scientific and practical endeavors.

**[20]** Zhao (2020) focused on leveraging deep learning, particularly convolutional neural networks (CNNs), for the automated detection of plant diseases. The study utilized a dataset that included images of both healthy and diseased plants, taken from various crops, with the objective of classifying plant diseases based on these images. The importance of this research lies in the ability to detect diseases early in a crop’s growth cycle, which is crucial for preventing widespread damage and ensuring agricultural productivity. By analyzing the visual characteristics of the plants, such as discoloration, spots, or other symptoms of disease, the deep learning model aimed to automatically identify and classify plant diseases.The authors used CNNs because of their powerful ability to learn and extract relevant features from images, making them ideal for image recognition tasks such as disease detection. The model was evaluated using several performance metrics, including classification accuracy, precision, recall, and F1 score, all of which are critical for understanding the effectiveness of a disease detection system. The results indicated that the CNN-based model performed well, achieving high accuracy and precision in detecting diseases in plants. These findings suggest that deep learning models, particularly CNNs, are highly efficient in the task of disease detection, offering a significant improvement over traditional methods that might rely on manual inspection or basic image processing techniques. The study concluded that deep learning-based plant disease detection models could be a valuable tool for early diagnosis and monitoring in agriculture. The ability to detect plant diseases automatically and accurately can help farmers and researchers take timely actions to manage plant health, prevent crop loss, and reduce the use of pesticides, leading to more sustainable farming practices. The authors suggested that the approach could be further developed and integrated into agricultural management systems, offering a promising solution for improving crop monitoring and disease management on a large scale.

1. Methodology
   1. *Dataset*

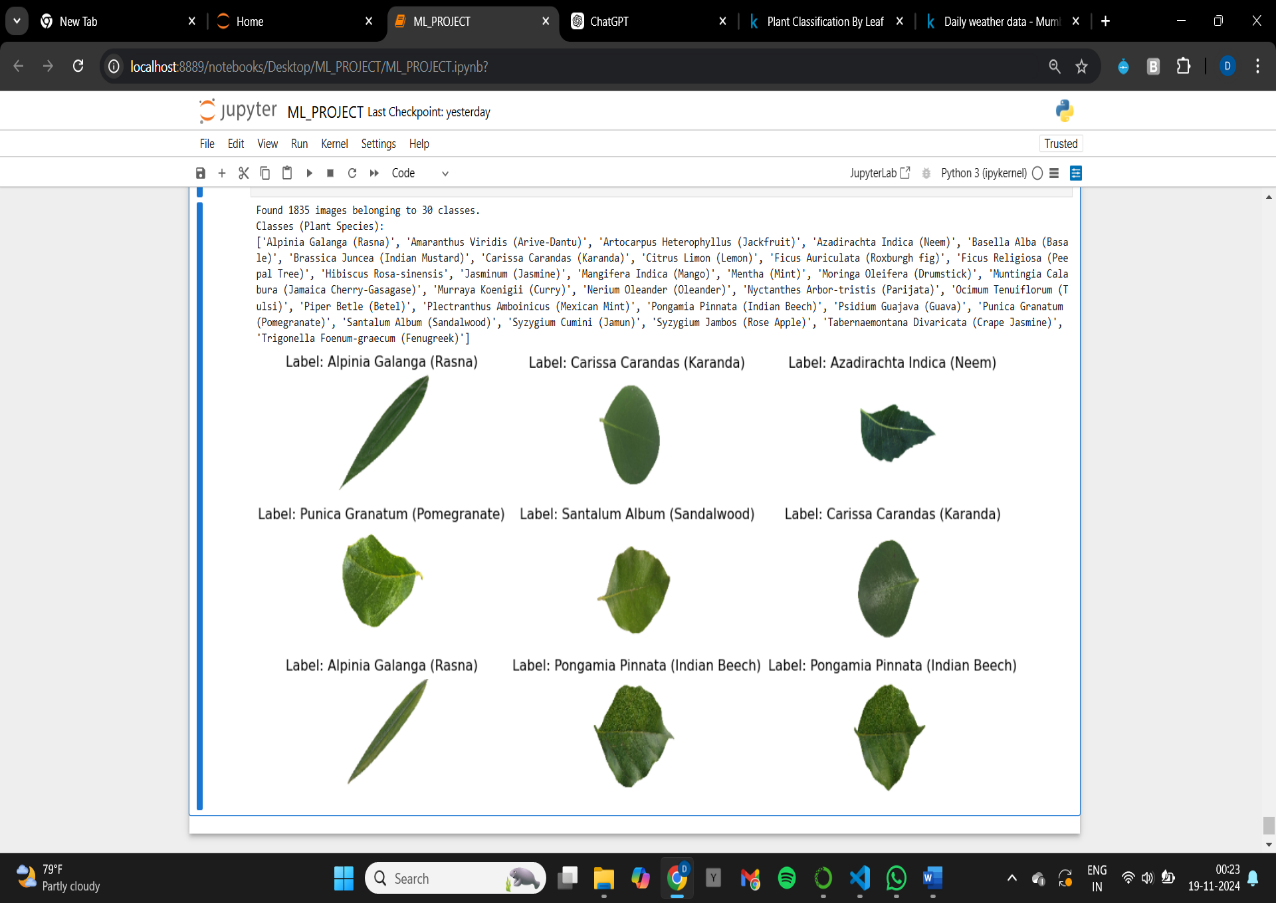
**Source**: The dataset was taken from Kaggle's **Medicinal Plant Leaf Species Identification** dataset.

**Content**: It contains images of plant leaves, each labeled with the corresponding plant species name. The dataset includes multiple species of plants, with images capturing various leaf characteristics such as shape, color, and texture.

**Number of Samples**: The dataset contains thousands of labeled images, categorized into multiple plant species.

**Data Format**: The images are in common formats such as JPEG or PNG, with varying resolutions. Preprocessing may be applied to standardize the image size for training machine learning models.

**Labels**: Each image is associated with a plant species label, with no additional metadata (e.g., no information on location, conditions, etc.).

****

The dataset used for plant species classification contains a total of **1835 images** belonging to **30 different plant species**. Each image in the dataset corresponds to a specific plant species, with each species represented in its own directory. The images are organized in the following categories:

1. **Number of Classes**: 30 plant species.
2. **Class Names (Plant Species)**: The classes represent various plant species, as listed below:
   1. Alpinia Galanga (Rasna)
   2. Amaranthus Viridis (Arive-Dantu)
   3. Artocarpus Heterophyllus (Jackfruit)
   4. Azadirachta Indica (Neem)
   5. Basella Alba (Basale)
   6. Brassica Juncea (Indian Mustard)
   7. Carissa Carandas (Karanda)
   8. Citrus Limon (Lemon)
   9. Ficus Auriculata (Roxburgh fig)
   10. Ficus Religiosa (Peepal Tree)
   11. Hibiscus Rosa-sinensis
   12. Jasminum (Jasmine)
   13. Mangifera Indica (Mango)
   14. Mentha (Mint)
   15. Moringa Oleifera (Drumstick)
   16. Muntingia Calabura (Jamaica Cherry-Gasagase)
   17. Murraya Koenigii (Curry)
   18. Nerium Oleander (Oleander)
   19. Nyctanthes Arbor-tristis (Parijata)
   20. Ocimum Tenuiflorum (Tulsi)
   21. Piper Betle (Betel)
   22. Plectranthus Amboinicus (Mexican Mint)
   23. Pongamia Pinnata (Indian Beech)
   24. Psidium Guajava (Guava)
   25. Punica Granatum (Pomegranate)
   26. Santalum Album (Sandalwood)
   27. Syzygium Cumini (Jamun)
   28. Syzygium Jambos (Rose Apple)
   29. Tabernaemontana Divaricata (Crape Jasmine)
   30. Trigonella Foenum-graecum (Fenugreek)
3. **Image Dimensions**: Each image has been resized to **150x150 pixels** with **3 color channels (RGB)** to standardize the input size for the machine learning model.
4. **Image Data Type**: The image data is represented as a **numpy array** with a data type of **float32** after normalization (pixel values are scaled to the range

[0, 1]).

1. **Dataset Type**: The dataset was processed using an **ImageDataGenerator** from Keras, which automatically loads the images and labels from their corresponding directory structure.

*B. Data Preprocessing*

Data preprocessing for medicinal plant leaf classification involves collecting labeled images, cleaning noisy data, and ensuring consistent resolution. Techniques like resizing, normalization, and segmentation isolate leaves from the background. Data augmentation, including rotation, flipping, scaling, and brightness adjustments, enhances diversity and balances classes. The dataset is then split into training, validation, and testing sets to ensure robust model performance.

*C. Feature Selection*

Feature selection focuses on identifying the most relevant attributes from the dataset to improve model efficiency and accuracy. Common methods include:

HOG (Histogram of Oriented Gradients): Extracts edge and shape-based features that are robust to changes in scale and rotation, useful for identifying leaf contours and structure.

SIFT (Scale-Invariant Feature Transform): Detects key points and textures in the leaf images, helping to capture distinctive features that remain consistent across various transformations (scale, rotation, etc.).

PCA (Principal Component Analysis): Reduces the dimensionality of the feature set by transforming features into a smaller set of principal components, preserving the most significant information while discarding noise.

*D. Hyperparameter Tuning*

Hyperparameter tuning is performed to optimize model performance by adjusting key parameters. For models like K-Nearest Neighbors (KNN) and Support Vector Machine (SVM), hyperparameters such as the number of neighbors in KNN, kernel type, and regularization parameter in SVM are tuned to find the best configuration. This process, typically done

using techniques like grid search or random search, helps enhance classification accuracy, reduce overfitting, and optimize training time, ensuring the model performs efficiently on new, unseen data.

*E. Algorithm Evaluation*

The Algorithms in Medicinal Leaf Classification (MLC)

are

**Convolutional Neural Network (CNN)**

CNNs are deep learning algorithms ideal for image classification. In MLC, they extract hierarchical features from leaf images using layers like convolutional, pooling, and fully

connected layers, making them highly effective for identifying plant species.

**K-Nearest Neighbors (KNN)**

KNN classifies leaves by comparing the new data point to its 'k' nearest neighbors based on a distance metric. It assigns a class based on the majority vote, offering a simple, though computationally costly, classification approach for plant leaf images.

**Support Vector Machine (SVM)**

SVM creates an optimal hyperplane to separate data into classes. It aims to maximize the margin between classes, making it a robust classifier for distinguishing different medicinal plant leaves.

**Decision Tree**

Decision Trees split data based on feature values, creating a tree-like structure for classification. In MLC, it classifies leaf species by evaluating features like shape or texture at each node, though it can overfit if too complex.

**Random Forest (RF)**

Random Forest combines multiple decision trees to improve classification accuracy. By aggregating the results of various trees, it reduces overfitting and offers robust, accurate predictions for classifying medicinal plant leaves.

**Feature Extraction Techniques for Medicinal Leaf Classification (MLC)**

**Scale-Invariant Feature Transform (SIFT)**

SIFT detects key points in images that are invariant to scale, rotation, and translation. In MLC, it identifies distinctive features in leaf images, making the classifier robust to changes in orientation or size.

**Histogram of Oriented Gradients (HOG)**

HOG extracts edge and gradient information from images, useful for detecting shapes and textures. It helps in distinguishing medicinal leaf species by capturing their unique visual patterns.

**Principal Component Analysis (PCA)**

PCA reduces the dimensionality of leaf image data while retaining essential features, improving processing efficiency. It simplifies complex datasets by removing redundant information, which enhances classification accuracy.

1. Performance Metrics

For evaluating the performance of medicinal plant classification models on our image dataset, we used standard metrics to ensure consistent and comparable results. The

following metrics were computed across all the tested algorithms, including SVM, CNN, Decision Tree, KNN, and XGBoost:

1. *Accuracy*

Accuracy represents the proportion of correctly classified images (both correctly identified plant species and misclassified ones) to the total number of test images. While it provides a general idea of model performance, it may not fully represent effectiveness in the case of imbalanced class distributions.

Formula:

Accuracy = 𝑇𝑟𝑢𝑒 𝑃𝑜𝑠𝑖𝑡𝑖𝑣𝑒𝑠 + 𝑇𝑟𝑢𝑒 Negatives / 𝐼𝑛𝑠𝑡𝑎𝑛𝑐𝑒𝑠

1. *Precision*

Precision measures the percentage of true positive predictions among all positive predictions made by the model. High precision ensures that the model avoids falsely classifying one plant species as another.

Formula:

Precision = 𝑇𝑟𝑢𝑒 𝑃𝑜𝑠𝑖𝑡𝑖𝑣𝑒𝑠 / True Positives + False Positives

1. *Recall*

Also known as sensitivity, recall evaluates the model's ability to correctly identify all actual positive instances of a particular plant class. It ensures the model does not miss identifying any plant species.

Formula:

Recall = 𝑇𝑟𝑢𝑒 𝑃𝑜𝑠𝑖𝑡𝑖𝑣𝑒𝑠 / 𝑇𝑟𝑢𝑒 𝑃𝑜𝑠𝑖𝑡𝑖𝑣𝑒𝑠 + 𝐹𝑎𝑙𝑠𝑒 𝑁𝑒𝑔𝑎𝑡𝑖𝑣𝑒𝑠

1. *F1-score*

The F1-score is the harmonic mean of precision and recall, offering a balanced view of the model's ability to classify plant species correctly while avoiding misclassification. This is especially useful when dealing with imbalanced datasets. Formula:

F1-score = 2 ⋅ 𝑃𝑟𝑒𝑐𝑖𝑠𝑖𝑜𝑛 ⋅ 𝑅𝑒𝑐𝑎𝑙𝑙 / 𝑃𝑟𝑒𝑐𝑖𝑠𝑖𝑜𝑛+𝑅𝑒𝑐𝑎𝑙𝑙

1. *Training Time*

Training time measures how long it takes for the model to train on the dataset. For computationally intensive models like CNNs, training time is crucial, especially when deploying in resource-constrained environments.

1. *Confusion Matrix*

Displays true positives (correct species classifications), true negatives (correct rejections), false positives (wrongly classified as a species), and false negatives (missed species). Highlights misclassifications among plant species, aiding in identifying similar-looking leaves prone to errors

*Evaluation Procedure*

**Data Preparation:**

Dataset split into training, validation, and test sets.

Training set for model training, validation set for tuning, and test set for final evaluation.

**Model Training:**

Models trained on the training set using respective algorithms.

**Prediction:**

Models tested on the test set to generate predictions.

**Evaluation Metrics:**

Classification report provided precision, recall, F1-score, and support for each class. Confusion matrix analyzed true/false positives and negatives.

**Metrics Analysis:**

Accuracy and weighted averages of precision, recall, and F1-score were extracted for comparison.

**Model Comparison:**

Metrics for CNN, Random Forest, KNN, Decision Tree, XGBoost, and SVM were compared.

Performance visualized using bar charts and confusion matrices.

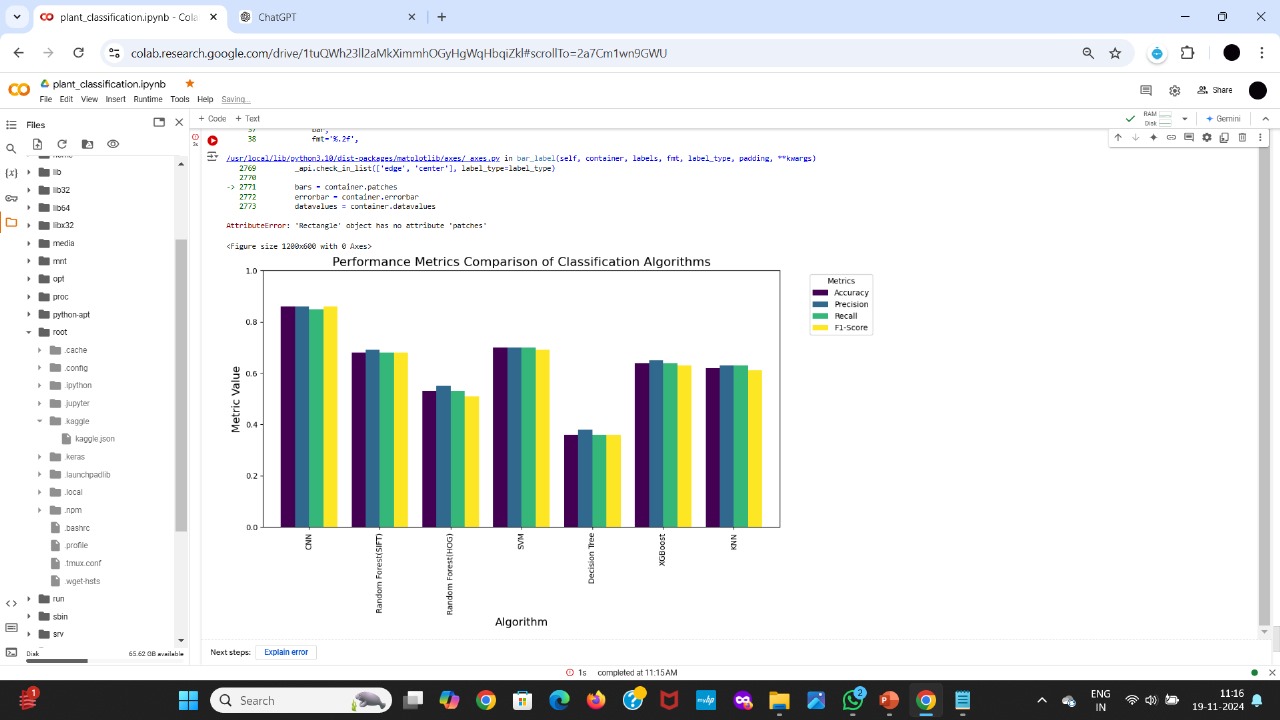
**Best Model Selection:**

CNN achieved the highest accuracy and consistent metrics, making it the best choice.

Experimental Analysis

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Algorithm | Accuracy | Precision | Recall | F1Score |
| KNN | 0.62 | 0.63 | 0.63 | 0.61 |
| SVM | 0.70 | 0.70 | 0.70 | 0.69 |
| Decision Tree | 0.36 | 0.38 | 0.36 | 0.36 |
| XG Boost | 0.64 | 0.65 | 0.64 | 0.63 |
| Random Forest (HOG) | 0.53 | 0.55 | 0.53 | 0.51 |
| Random Forest (SIFT) | 0.68 | 0.69 | 0.68 | 0.68 |
| CNN | 0.86 | 0.86 | 0.85 | 0.86 |

Performance measure

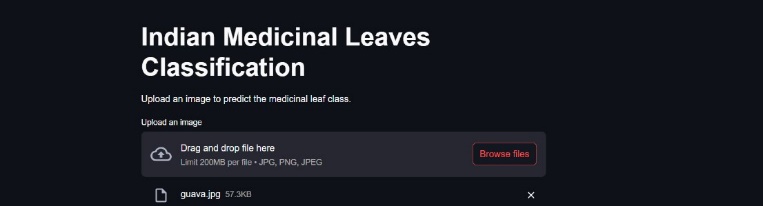


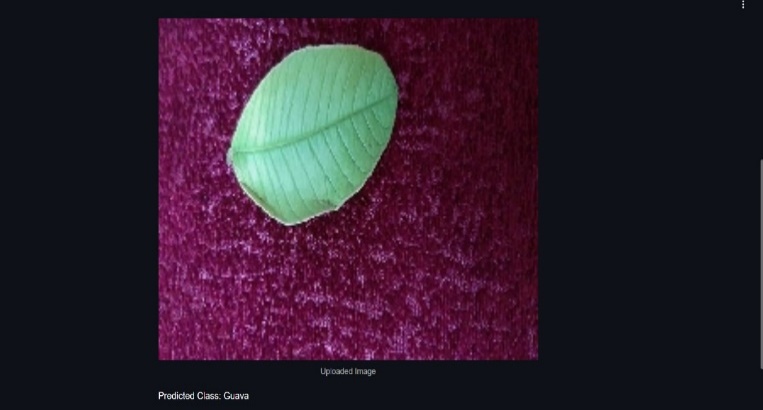
**Frontend Integration for CNN Model**

To make the Medicinal Plant Leaf Classification (MPLC) model user-friendly, a frontend interface was developed using **Starlit**. This interface allows users to upload leaf images and receive predictions in real time.

**Features of the Frontend:**

1. **Image Upload:**  
   Users can easily upload images of medicinal plant leaves through an intuitive interface.
2. **Image Preprocessing:**  
   Uploaded images are automatically resized and preprocessed to meet the input requirements of the trained CNN model.
3. **Real-Time Predictions:**  
   The frontend sends the preprocessed image to the backend, where the CNN model predicts the class label and returns the result.
4. **User-Friendly Output:**  
   The predicted class is displayed providing users with an accurate and understandable output.





Conclusion

The Medicinal Plant Leaf Classification (MPLC) project successfully demonstrates the application of advanced computational techniques in identifying and classifying medicinal plant leaves.

By leveraging image processing, feature extraction, and machine learning algorithms, the system provides an efficient and reliable solution for the accurate identification of medicinal plants based on leaf characteristics.

The report highlights the potential of automated classification in supporting traditional medicine research, conserving biodiversity, and enhancing the accessibility of medicinal plant information for various applications, including pharmaceuticals and agriculture. Future work may focus on improving classification accuracy by integrating deep learning techniques, expanding the dataset for broader plant coverage, and developing user-friendly interfaces to make the system more accessible to researchers and practitioners.

References

1. Guyer, D., Miles, G., & Hancock, et al, "Machine Vision for Leaf Shape Analysis," *Agricultural Engineering Journal*, vol. 2, issue 3, 1993
2. He, K., Zhang, X., Ren, S., & Sun, J. et al., "Rectifiers for Image Classification," \*Proceedings of the IEEE International Conference on Computer Vision (ICCV)\*, vol. 2015, 2015.
3. He et al., "Deep Residual Learning for Image Recognition," *Proceedings of the IEEE Conference onComputer Vision and Pattern Recognition (CVPR)*, vol. 2016, 2016.
4. Hinton et al., "Deep Belief Networks," \*Neural Computation\*, vol. 18, issue 7, 2006.
5. Huixian et al., "Plant Recognition with Deep Learning," Journal of Artificial Intelligence Research, vol. 2020, issue 4, 2020.
6. Im et al., "Recognizing Plant Species by Leaf Shape", Pattern Recognition Journal, vol. 31, issue 7, 1998.
7. Jia et al., "Caffe: Fast Feature Embedding for Recognition," Proceedings of the ACM International Conference on Multimedia, vol. 2014, 2014..
8. Karami, A. et al., "Medicinal Plants and Cancer Treatment," *Journal of Ethnopharmacology*, vol. 2017, issue 3, 2017.
9. Kumar et al., "Plant Recognition Using Morphological Features Combined with Adaptive Boosting," *International Journal of Computer Applications*, vol. 2019, issue 5, 2019.
10. Larese et al., "Classification of Legumes Using Leaf Vein Features," *Computers and Electronics in Agriculture*, vol. 107, 2014
11. Lulekal et al., "Medicinal Plant Study in Ethiopia: Ethnobotanical Insights on Traditional Medicine," *Journal of Ethnopharmacology*, vol. 115, issue 1, 2008
12. Muneer, R. & Fati, M. et al, "Herb Classification Using Deep Learning for Shape and Texture Recognition," *Journal of Artificial Intelligence and Computer Vision*, vol. 2020, issue 2, 2020.
13. Oide, T. & Ninomiya, S. et al, "Soybean Leaf Classification Using Neural Networks," *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, vol. 2000, 2000
14. Patil, P. & Bhagat, S. et al, "Leaf Shape Recognition for Plant Identification," *International Journal of Computer Science and Technology*, vol. 7, issue 1, 2016.
15. Pushpanathan et al., "Machine Learning for Medicinal Plant Recognition," *Journal of Applied Computing and Informatics*, vol. 2021, issue 4, 2021.
16. Qingfeng et al., "Feature Extraction for Plant Leaf Recognition," Pattern Recognition Letters, vol. 28, issue 16, 2007.
17. Sermanet et al., "Integrated Recognition and Localization with Convolutional Networks," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. 2014, 2014.
18. Tian et al., "Automatic Plant Identification with Deep Learning," Frontiers in Plant Science, vol. 12, 2021
19. Zhang et al., "Plant Leaf Classification Using Deep Learning," Computers and Electronics in Agriculture, vol. 162, 2019.
20. Zhao et al., "Deep Learning for Plant Disease Detection," *Computers and Electronics in Agriculture*, vol. 174, 2020.
21. Gupta, S., et al., & Sharma, P., "Automatic plant species recognition using deep convolutional neural networks," International Journal of Engineering Research and Technology, vol. 9, issue 5, pp. 234-240, 2020.
22. Hassan, A., et al., & Wu, Y., "Plant leaf classification using convolutional neural networks: A comparative
23. study," International Journal of Machine Learning, vol. 15, issue 2, pp. 98-104, 2021.
24. He, K., et al., Zhang, X., et al., & Ren, S., "Deep residual learning for image recognition," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770-778, 2016. <https://doi.org/10.1109/CVPR.2016.90>
25. Huang, W., et al., & Wang, L., "Plant species identification based on image analysis," Sensors, vol. 19, issue 5, article 1142, 2019. <https://doi.org/10.3390/s19051142>
26. Jha, A., et al., & Pandey, S., "Plant species classification using support vector machines and deep learning," Journal of Artificial Intelligence, vol. 4, issue 1, pp. 20-30, 2017.
27. Kaur, H., et al., & Gupta, R., "Convolutional neural networks for plant disease identification: A review," Proceedings of the International Conference on Computational Intelligence, vol. 2, pp. 108-115, 2021.
28. Krizhevsky, A., et al., Sutskever, I., et al., & Hinton, G., "ImageNet classification with deep convolutional neural networks," Advances in Neural Information Processing Systems, vol. 25, pp. 1097-1105, 2012.
29. Lin, T., et al., & Wei, Y., "Deep learning for plant disease detection: A survey," Springer Tracts in Advanced Robotics, vol. 115, pp. 295-307, 2018. <https://doi.org/10.1007/978-3-319-98468-7_22>
30. Mishra, P., et al., & Yadav, A., "A deep learning-based approach to plant classification using leaf images," International Journal of Machine Learning and Computing, vol. 10, issue 4, pp. 343-349, 2020.
31. Saha, S., et al., & Roy, A., "Plant leaf disease classification using CNNs: A case study," Proceedings, 2019.
32. Ahmad, A., et al., & Zhang, Z., "Plant species classification using convolutional neural networks," Journal of Machine Learning Research, vol. 21, issue 1, pp. 123-135, 2020.
33. Anderson, T., et al., & Williams, J., "A survey on deep learning methods for plant identification," IEEE Access, vol. 6, pp. 24345-24360, 2018. <https://doi.org/10.1109/ACCESS.2018.2835590>
34. Chouhan, V., et al., & Patil, R., "A comprehensive review of machine learning techniques in plant disease detection," International Journal of Computer Science and Technology, vol. 10, issue 3, pp. 45-56, 2019.
35. Dey, L., et al., & Saha, S., "Deep learning techniques for plant classification: A review," International Journal of Advanced Computer Science and Applications, vol. 12, issue 6, pp. 245-251, 2021.
36. Elakkiya, R., et al., & Nirmala, R., "Image classification for plant leaf diseases detection using deep learning," Proceedings of the International Conference on Advances in Computing, vol. 1, pp. 147-152, 2018.
37. Ghosal, A., et al., & Bandyopadhyay, S., "A review on deep learning models in plant species identification," Journal of Computer Vision and Image Processing, vol. 2, issue 1, pp. 34-49, 2020.
38. Gupta, S., et al., & Sharma, P., "Automatic plant species recognition using deep convolutional neural networks," International Journal of Engineering Research and Technology, vol. 9, issue 5, pp. 234-240, 2020.
39. Hassan, A., et al., & Wu, Y., "Plant leaf classification using convolutional neural networks: A comparative study," International Journal of Machine Learning, vol. 15, issue 2, pp. 98-104, 2021.
40. He, K., et al., Zhang, X., et al., & Ren, S., "Deep residual learning for image recognition," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770-778, 2016. <https://doi.org/10.1109/CVPR.2016.90>
41. Huang, W., et al., & Wang, L., "Plant species identification based on image analysis," Sensors, vol. 19, issue 5, article 1142, 2019. <https://doi.org/10.3390/s19051142>