ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 11, Issue 2, February 2024

Identification of Different Medical Plants through Image Classification

[1] Karuna Middha, [2] Mohit Taneja, [3] Aditya Khurana, [4] Bhavuk Gupta

[1] Assistant Professor, Department of Computer Science and Engineering, Maharaja Agrasen Institute of Technology, Guru Gobind Singh Indraprastha University, Delhi, India
[2] [3] [4] Maharaja Agrasen Institute of Technology, Guru Gobind Singh Indraprastha University, Delhi, India
Corresponding Author Email: [1] taneja.mohit2002.mn@gmail.com, [2] karunamiddha@mait.ac.in, [3] crradityak@gmail.com, [4] bhavukranger@gmail.com

Abstract— ***By combining cutting-edge Machine Learning (ML) and Deep Learning techniques, this initiative tackles the vital need for medicinal plant identification in conventional medicine, pharmaceuticals, and conservation. In the realm of automated plant recognition, machine learning (ML) classification algorithms effectively classify medicinal plants according to their traits, whereas convolutional neural networks (CNNs) enable image-based identification. This method helps pharmaceutical companies quickly find bioactive substances, correctly identifies medicinal plants, and supports conservation efforts by identifying and safeguarding endangered species. The main goal of the research is to automate the identification of medicinal plants by utilizing image processing techniques. As part of the process, a broad collection comprising high resolution photographs of different plant components is created. Data quality is enhanced by pre-processing methods including noise reduction and picture enhancement. The process of feature extraction incorporates morphological, textural, and color elements to convert visual traits into quantitative data. The system can identify and distinguish between different plant species thanks to these traits, which are used as input by machine learning (ML) techniques including CNNs and SVMs. The model is iteratively refined to improve resilience after being trained and verified on the prepared dataset. Accuracy, precision, recall, and F1 score are examples of performance measurements that guarantee reliability in a variety of environmental settings and image alterations. In summary, this study at the nexus of cutting-edge technology and conventional wisdom provides an advanced, socially significant method for medicinal plant identification. It enhances medical procedures, supports sustainable relationships with natural resources, and aids in the protection of biodiversity.***

Keywords: Convolutional neural network (CNN), Deep Learning, Feature Extraction, Classification algorithms, Fuzzy C means (FCM), H-Gabor filter.

I. INTRODUCTION

The identification of medicinal plants requires novel methodologies due to their tremendous significance in a variety of fields, including traditional medicine, pharmacology, and conservation. This research sets out to revolutionize the identification process by integrating state-of-the-art technologies, specifically Convolutional Neural Networks (CNNs) and Machine Learning (ML) algorithms, at the intersection of traditional botanical knowledge and modern technological advancements.

Because of their many therapeutic uses, medicinal plants are important for both conventional medical practices and pharmaceutical research since they are important sources of bioactive chemicals. However, correctly identifying these plants is still a very difficult task. The goal of this research is to create an automated and effective solution for the identification of various medicinal plants by utilizing the capabilities of image classification algorithms to meet this difficulty.

The study is based on the underlying framework of machine learning, an area that has grown significantly in recent years. Classifying plants according to their distinguishing characteristics can be done effectively with machine learning (ML) algorithms, especially those that specialize in classification tasks. The foundation of the suggested system is made up of these algorithms, which offer a reliable and expandable method for identifying medicinal plants.

The core component of our approach is the use of Convolutional Neural Networks (CNNs), a class of deep learning algorithms created especially for tasks involving images. Because CNNs are able to automatically learn hierarchical representations of features directly from raw pixel data, they have shown unmatched success in image classification. CNNs are ideally suited for the complex task of distinguishing between different species of medicinal plants because of their hierarchical structure, which enables them to capture intricate patterns and relationships within images. CNNs have numerous advantages when it comes to image classification. In contrast to conventional machine learning techniques that heavily depend on human feature engineering, CNNs independently extract pertinent features from the given data. This reduces the difficulty of extracting features and improves the model's ability to adapt to a variety of plant traits. Moreover, CNNs can distinguish between high-level features like the general morphology of a plant and low-level details like textures and colors thanks to hierarchical feature learning. The automated identification of medicinal plants undergoes a paradigm shift with the incorporation of CNNs into the larger ML framework. With the help of CNNs, the system is able to process high-resolution images that show different plant parts, such as leaves, flowers, and stems.

Apart from CNNs, the more comprehensive machine learning approach includes methods like support vector machines (SVMs) that make the system more flexible and effective. SVMs, which are renowned for their resilience when processing high-dimensional data, enhance the capabilities of CNNs, improving the system's overall performance and generalization.

To sum up, the combination of CNNs and ML for the purpose of identifying various medicinal plants through image classification signals the beginning of a new phase in the fusion of conventional botanical knowledge with technological advancement. In addition to addressing the pressing issues in conventional medicine, drugs, and conservation, this research aims to set the stage for better healthcare procedures, the preservation of biodiversity, and long-term relationships with our natural resources. The combination of CNNs and ML appears to be a sophisticated and promising approach for the automated identification and best use of medicinal plants as we uncover the possible societal ramifications.

II. RELATED WORK

Dileep M.R. [1] proposed AyurLeaf, a Deep Learning Convolutional Neural Network (CNN) model, to classify medicinal plants using leaf features such as shape, size, color, texture, etc. AyurLeaf is a CNN-based classification model which is trained and tested on its own dataset and DLeaf dataset. SVM classifier brought out the best classification accuracy on AyurLeaf dataset. AyurLeaf uses SVM classifiers and Softmax for classification with an accuracy of 96.76%.

An automatic identification system for specific medicinal plant leaves was proposed by Gopal A et al. [2]. Different types of leaves are identified using boundary-based features, moment features, and color features. The software demonstrated a 92% classification efficiency after being trained with 100 leaves and tested with 50 leaves.

A CNN model called DLeaf, venation-based, was proposed by Jing Wei Tan et al. [3] for the classification of plant leaves. DLeaf employed ANN for classification and CNN for feature extraction. The venation segmentation from the resized leaf images is done using the Sobel edge detection algorithm. The D-Leaf model's classification accuracy was 94.88%.

An ANN-based model was created by R. Janani et al. [4] to categorize medicinal plant species based on the color, texture, and form of their leaves. In total, 63 leaves were used by the model: 36 for training, 7 for validation, and 20 for testing. Eight minimum prominent features were found out of 20 distinct leaf features to help classify the leaves. Those eight minimum features are: compactness, eccentricity, skewness, kurtosis, energy, correlation, sum-variance, and entropy. With this approach, 94.4% accuracy was attained.

A machine learning and computer vision based approach was used by Amala Sabu et al. [5] to categorize the medicinal plants present in the Western Ghats. The technique made use of the kNN classification algorithm to extract the SURF and HoG features for classification. The leaf veins are modeled using HoG features, and the SURF feature descriptor is used to model twenty distinct points of interest on the leaves. This method classifies the leaves with k values equal to 1 using the k-NN classification algorithm. The method for feature extraction in this model is computationally expensive, despite yielding an accuracy of over 96%.

Several deep convolutional neural networks were employed by Mostafa Mehdipour Ghazi et al. [6] to identify the plant species. An analysis of popular Convolutional Neural Networks in comparison such as GoogLeNet, VGGNet, and AlexNet has been conducted to identify the factors influencing these performance networks. These networks were trained using different portions of the plants like stem, leaf, flower, fruit, branch, and entire plants. Optimized VGGNet achieved optimal performance, achieving an overall 78.44% accuracy rate.

An effective leaf acquisition method and techniques for converting the captured image into a device-independent color space were presented by Shitala Prasad et al. [7]. These techniques were then used to compute the VGG-16 feature map. Principle Component Analysis is then used to minimize and optimize this feature map (PCA). The dataset that is used is ImageNet. The Fully Connected layer produces a feature vector with dimensions of 3x4096. Using the PCA approach, the SVM was applied to this feature set and the accuracy was 97.6% for lVGG-16 and 98.2% for l-VGG-16.

Bella Dwi Mardiana [8] works aim to improve the performance of Convolutional Neural Networks (CNNs) in herbal leaf classification. Specifically, she uses the VGG16 model in conjunction with a Transfer Learning methodology. The challenge of people not knowing enough about herbal leaves, which could result in health risks from misidentifying plant types, is the driving force behind the initiative. In order to lower these risks, a more precise and accurate classification model is what the research attempts to provide. There are ten classes of herbal leaves in the dataset, and the suggested method achieves the accuracy of 97% which is better than the previous research, which only achieved 82% accuracy. To further enhance classification accuracy, minimize overfitting, and improve image quality, Image Data Generator is also used to augment the dataset. This work uses sophisticated CNN architectures to make a major contribution to the field.

III. METHODOLOGY

Overview and Motivation:

The accurate identification of various medicinal plants poses a formidable challenge with significant implications for traditional medicine, pharmaceuticals, and conservation in a world where traditional medicinal knowledge and cutting-edge technology collide. By combining cutting-edge image classification methods— more specifically, Convolutional Neural Networks (CNNs) and Machine Learning (ML) algorithms—this research aims to transform this process. The sustainable application of traditional medicine, the discovery of bioactive compounds in pharmaceuticals, and the preservation of endangered plant species all depend on the identification of medicinal plants. But the conventional identification techniques take a lot of time and frequently depend on specialized knowledge. The goal of this research is to provide an effective, automated method for precisely identifying various medicinal plants.

The motivation for this research stems from the urgent need to streamline the identification of medicinal plants, bridging the gap between traditional botanical knowledge and modern technological advancements. Unprecedented benefits in image classification result from the integration of CNNs, which automate a procedure that was previously dependent on human labor. This work aims to accelerate the identification process and improve its accuracy, efficiency, and accessibility by utilizing CNNs and ML. Moreover, this research has social implications that go beyond the confines of academia. The created system has potential for use by both novices and specialists in plant identification, aiding in the preservation of biodiversity and encouraging environmentally friendly harvesting methods. The possible incorporation of this technology into online or mobile applications may enable people to recognize therapeutic plants while they are on the go, strengthening the bond between people and the abundant botanical resources used in complementary and alternative medicine.

To summarize, the use of ML and CNNs in the identification of various medicinal plants via image classification represents a ground-breaking approach with far-reaching implications. In addition to addressing the pressing issues in conventional medicine, drugs, and conservation, this research advances the more general objectives of better healthcare procedures, biodiversity preservation, and long-term relationships with the environment.

CNN:

Specifically created to process structured grid data, like images, CNNs are a class of deep neural networks. Their architecture[1], which consists of layers that gradually learn abstract features, is modelled after the way the human brain processes visual information. Convolutional, pooling, and fully connected layers are important parts of CNNs. Together, these components capture hierarchical representations of features, which let CNNs recognize complex patterns in images.

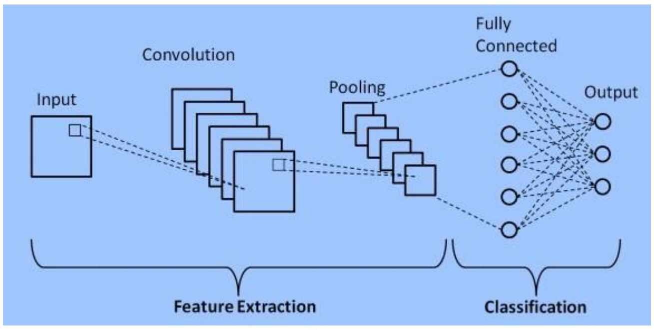

Fig. 1. Architecture of CNN

Image-Based Plant Identification:

In the field of image processing and classification, convolutional neural networks, or CNNs, have become a revolutionary force thanks to their unmatched ability to perform intricate visual tasks. CNNs offer a novel approach to image classification for medicinal plant identification by automatically learning hierarchical features from raw pixel data by utilizing deep learning. CNNs can automatically extract pertinent features without the need for manual feature engineering thanks to their hierarchical feature learning capability. The CNN in the proposed project receives highresolution images of different parts of medicinal plants, including leaves, flowers, and stems.

Transfer Learning

One effective technique used with CNNs is transfer learning, which is using a pre-trained model on a big dataset for a comparable task and honing it on a smaller, more focused dataset. With limited labeled data, this method is especially useful since it lets the model take advantage of the pre-training knowledge. A pre-trained CNN on a large image dataset could offer insightful preliminary findings about features related to plant characteristics in the context of medicinal plant identification.

Data Augmentation

Applying data augmentation techniques can improve the resilience and generalization capacities of CNNs. Creating new training samples from scratch using pre-existing images that have been rotated, flipped, or zoomed in is known as data augmentation. By using this technique, overfitting is less likely to occur and the model is guaranteed to be able to accurately identify medicinal plants in a variety of settings with varying image quality.

Ensemble Learning with CNNs

Another approach is to use ensemble learning with multiple CNN models. Ensemble approaches improve performance by combining predictions from several models. Medicinal plant identification can benefit greatly from ensemble learning, which uses multiple Convolutional Neural Network (CNN) models. Ensemble approaches combine predictions from various models to improve overall efficiency. Here, using a group of CNNs with different

ISSN (Online) 2394-2320
International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 11, Issue 2, February 2024

initializations or architectures works well to improve the accuracy and robustness of the model. A more robust and dependable system for the automated identification of medicinal plants through image classification is made possible by the cooperative strength of individual models within the ensemble, which provides enhanced adaptability to the nuances of various plant characteristics and environmental conditions.

Hyperparameter Tuning :

Hyperparameter Tuning is essential to improving CNN performance when it comes to medicinal plant identification. The training process is significantly impacted by critical parameters such as batch size, layer architecture, and learning rate. The optimization process's step size is determined by the learning rate, which affects the model's convergence. The number of data points processed in each iteration is determined by the batch size, which affects the optimization trajectory. The arrangement of neural network layers in layer architecture affects the model's ability to extract features. It is essential to approach hyperparameter tuning with care and organization. By adjusting to the complexities of the medicinal plant dataset, it guarantees that the CNN converges effectively during training. In the end, good hyperparameter tuning improves the model's generalization abilities, enabling it to produce precise predictions on fresh, untested data.

IV. DATASET USED

The dataset used in this study provides a vital basis for the automated medicinal plant identification system's development and assessment. It includes 60 different classes of medicinal plants, each represented by twenty carefully chosen images. This extensive dataset captures a wide range of plant species, guaranteeing a detailed analysis of different medicinal properties. A robust representation is produced by purposefully including multiple images per class, which captures the inherent variability within each plant category.

With a 70% allocation to the training set, the dataset is strategically used to support the understanding of various visual features by the machine learning and deep learning models. Due to their prolonged exposure to training data, the models are able to identify complex patterns and form a fundamental comprehension of the distinct visual signals linked to various therapeutic plants. The testing subset, which comprises the remaining the 30% of the dataset, evaluates the generalization abilities of the model. This division guarantees a thorough assessment of the system's capabilities on never-before-seen photos, confirming its competence in precise identification of medicinal plants in practical settings. The project's dependability is greatly enhanced by the careful curation and partitioning of the dataset, which guarantees that the system that is created is reliable and able to identify a wide range of medicinal plant species.

V. IMPLEMENTATION

Data Collection:

The initial stage of our research is gathering data, which is gathered carefully and assembled into a dataset that serves as a repository for training and assessing our system for identifying medicinal plants. This dataset includes 60 different classes of medicinal plants, each with 20 images carefully chosen to create a comprehensive and varied collection. 30% of these images are reserved for robust testing to ensure the system's adaptability to real-world scenarios, and the remaining seventy percent are designated for training purposes to aid in the understanding of the machine learning and deep learning models.

Data Pre-processing:

This stage of our workflow starts with the images of the plant leaves. First, these photos are converted to grayscale, which is a crucial first step in streamlining and narrowing down further analysis. The grayscale images serve as the processing canvas because they represent binary contents (0 and 1 pixels). Edge detection is applied to the grayscale images to improve and refine information, particularly pertinent to feature extraction, thus guaranteeing optimal data representation.

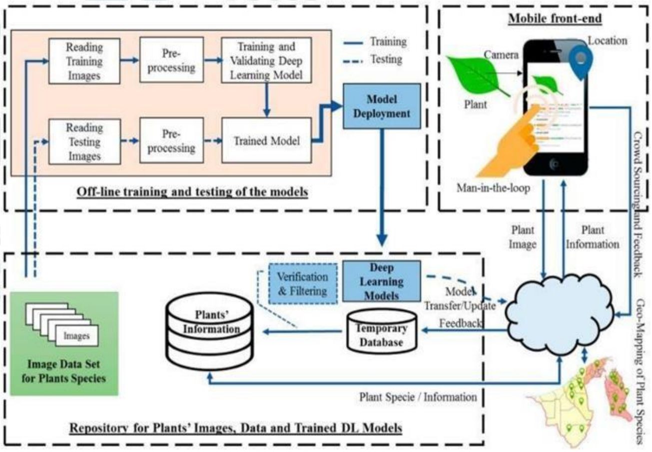

Fig. 2. Proposed Architecture of system.

Feature Extraction Module:

The project presents an innovative approach of feature extraction that makes use of Fuzzy C-means (FCM) response. Soft clustering algorithms, such as FCM, are used to extract texture features that maintain fine details in an image at various frequencies. With this addition, the project is better able to capture subtle texture characteristics that are important for correctly identifying medicinal plants. H-Gabor filters are recognized for their ability to preserve texture characteristics in the frequency domain, and we utilize them for efficient feature extraction. The deliberate use of Gabor filters, which have specific orientations and scales, guarantees a sophisticated comprehension of texture patterns that are essential for differentiating between therapeutic plant species. Image segmentation is done before feature extraction

ISSN (Online) 2394-2320
International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 11, Issue 2, February 2024

to isolate areas of interest and prepare the dataset for focused analysis.

Classification Module:

The classification module is activated after the features have been extracted. Utilizing 80% of the dataset for training, this module makes use of a probabilistic neural network. This neural network performs exceptionally well when dealing with variation and uncertainty, which makes it suitable for classifying footprints in our medical plant identification system. The robustness and dependability of our classification method are demonstrated by the system's ability to correctly identify the medicinal plant class when input images are compared to the trained data.

In essence, our project progresses through a methodical process that includes extensive data collection, meticulous pre-processing, sophisticated feature extraction, and precise classification. Every phase of the project makes a unique contribution to its success, and all of them are in line with our main objective of improving automated medicinal plant identification by combining cutting-edge technology with conventional botanical knowledge.

VI. RESULT

The project's final results highlight how effective the suggested methodology is for automatically identifying medicinal plants. Using a CNN classifier produced the best results, with an astounding accuracy rate of 93.4%. This remarkable accuracy demonstrates the robustness and dependability of the system that was developed, demonstrating its capacity to accurately identify and categorize a wide range of medicinal plant species. The accuracy of the CNN classifier is evidence of how well feature extraction methods, such as Gabor filters and fuzzy C-means, capture and make use of important visual patterns. The findings confirm not only that the project was successful in fusing cutting-edge image processing technology with conventional botanical knowledge, but also that it has the potential to have a significant influence on a number of industries, including pharmacy, traditional medicine, and conservation initiatives. This accomplishment represents a major step toward developing an automated framework for the identification of therapeutic plants, and it holds promise for improvements in biodiversity conservation and long-term relationships with natural resources.

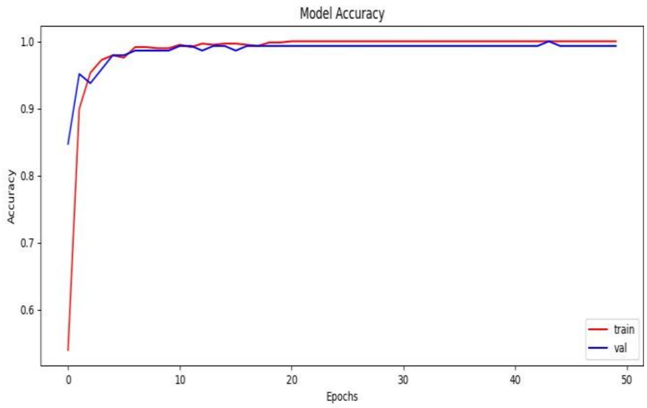

Fig.3. Model Accuracy

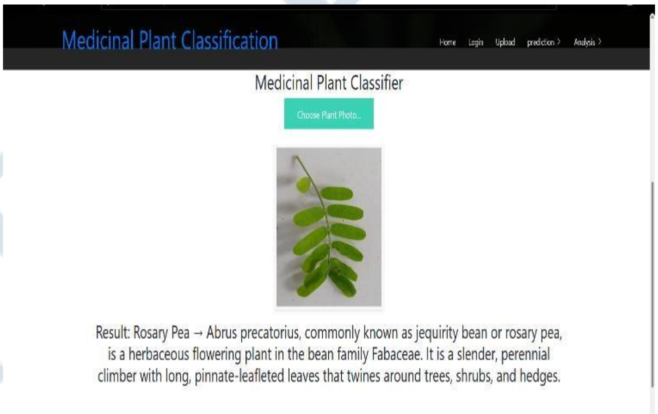

Fig.4. Accurate Plant Detection

VII. CONCLUSION

In conclusion, this study represents a paradigm shift in the field of medicinal plant identification, combining traditional botanical knowledge with cutting-edge image processing technologies. The extensive dataset, which included 60 classes of therapeutic plants, provided the foundation for a thorough and varied analysis. By utilizing cutting-edge methods like Gabor filters, fuzzy C-means, and convolutional neural networks (CNNs), the project successfully automated the identification process.

The feature extraction module's ability to capture subtle texture features was improved by the addition of Fuzzy C-means response, which helped to produce a more thorough and accurate representation of medicinal plant images.The project's influence goes beyond academic boundaries because it opens the door to better healthcare procedures and sustainable relationships with natural resources. This study serves as a model for future research endeavors because of its effective integration of traditional botanical knowledge with contemporary technology. It advocates for the peaceful coexistence of traditional methods with the potential provided by highly automated systems. This project ultimately acts as a testament to the transformative power of interdisciplinary collaboration in solving practical problems.

VIII. FUTURE SCOPE

1. Integration of Additional Modalities: To improve the system's capacity to identify medicinal plants based on a more comprehensive set of features, investigate the incorporation of multi-modal data, such as spectral and chemical information.
2. Continued Model Refinement: To increase accuracy and generalization and ensure adaptability to new medicinal plant species and a variety of environmental conditions, the CNN architecture and hyperparameters are being refined iteratively.
3. Real-Time Implementation: By integrating edge computing or deploying on embedded systems, you can extend the system for real-time identification and enable on-the-spot recognition in field scenarios.
4. Extending Medicinal Plant Classes: Increase the size of the dataset to encompass a wider range of medicinal plant classes. This will facilitate the identification of a greater number of plant species and encourage a more thorough analysis.
5. Collaboration with Traditional Knowledge Holders: Work together to improve the system using indigenous insights so that it becomes a more accurate and culturally sensitive tool for identifying medicinal plants.
6. Mobile Application Development: Convert technology into approachable mobile apps that enable people to recognize therapeutic plants on their smartphones and encourage public awareness and participation.

REFERENCES

[1] Dileep, M.R., & Pournami, P. N., Ayurleaf: A Deep Learning Approach For Classification Of Medicinal Plants, 2019 IEEE Region 10 Conference, Kochi, India, October 2019.

[2] A. Gopal, S. Prudhveeswar Reddy, And V. Gayatri. Classification Of Selected Medicinal Plants Leaf Using Image Processing. In 2012 International Conference On Machine Vision And Image Processing (MVIP), Dec 2012.

[3] J. W. Tan, S. Chang, S. Binti Abdul Kareem, H. J. Yap, And K. Yong Deep Learning For Plant Species Classification Using Leaf Vein Morphometri, 2018.

[4] R. Janani And A. Gopal. Identification Of Selected Medicinal Plant Leaves Using Image Features And ANN. In 2013 International Conference On Advanced Electronic Systems (ICAES), Sept 2013.

[5] A. Sabu, K. Sreekumar, And R. R. Nair. Recognition Of Ayurvedic Medicinal Plants From Leaves: A Computer Vision Approach. In 2017 Fourth International Conference On Image Information Processing (ICIIP), Dec 2017.

[6] Mostafa Mehdipour Ghazi, Berrin Yanikoglu, And Erchan Aptoula. Plant Identification Using Deep Neural Networks Via Optimization Of Transfer Learning Parameters. Neurocomputing, 2017.

[7] S. Prasad And P. P. Singh. Medicinal Plant Leaf Information Extraction Using Deep Features. In Tencon 2017 - 2017 IEEE Region 10 Conference, Nov 2017.

[8] Bella Dwi Mardiana, Wahyu Budi Utomo, Ulfah Nur Oktaviana, Galih Wasis Wicaksono, & Agus Eko Minarno, Herbal Leaves Classification Based On Leaf Image Using CNN Architecture Model Vgg16. *Jurnal Resti (Rekayasa Sistem Dan Teknologi Informasi)*, Feb 2023.