International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 10 Issue 6 ǁ 2022 ǁ PP. 1893-1898

Deep Learning for Plant Species Classification

B ANANDARAJ[1] P V SNEHA SREE[2], S REDDY PRASANNA[3], B THEJASWINI[4]

Assistant professor[1] Department of Computer Science & Engineering[1, 2, 3, 4, 5]
Madanapalle Institute of Technology & science, Chittoor district, AP, India.[1, 2, 3, 4, 5]

Abstract *- Biodiversity protection is critical, and we need to learn more about the species in order to do it. It's difficult to identify plant species using traditional hand-crafted traits. Non-experts have a hard time remembering botanical words. The concept of automatically identifying plant species is becoming a reality. In this case, machine learning and deep learning are crucial. As a result, we're using deep learning-based Convolutional Neural Networks (CNN) to extract features from leaf photos and classify plant species. Deep learning techniques trump all handcrafted techniques.*

Keywords: *Leaf vein; segmentation; feature extraction; CNN; classifier algorithms.*

Date of Submission: 15-06-2022 Date of acceptance: 30-06-2022

I. INTRODUCTION

1.1MOTIVATION

This project's objective was to develop an identification system that could recognise plants simply by glancing at their leaves. Users can now take pictures of leaves and upload them to a server using a smartphone app. The server does pre-processing and a feature extraction unit before a pattern matcher compares the data from this image with the data in the database to locate potential matches. There is a new technique suggested for counting multiple leaflets. The main justification for this choice is because some complex leaf species can have a wide range in the number, size, and shape of their leaflets. Consequently, a local description based on a certain number of flyers may be more realistic. We were limited to three leaflets in our technique, which were mechanically recovered from photographs using geometrical presumptions drawn from botany. `

1.2 PROBLEM DEFINITION

To accurately identify different plant species, machine learning will analyse leaf pictures and retrieved characteristics like form, edge, and texture. Leaves are a useful tool for identifying different plant species because of their size and distinctive characteristics. They also act as a fun primer on how to use methods that incorporate visual characteristics. To evaluate the effectiveness of classifiers in situations involving the classification of images, we will employ multiple classification strategies. Each approach of categorising pictures is based on a fundamental idea that is essentially the same. A digital camera is used to take the initial images of the surroundings. In order to extract crucial details required for subsequent analysis, image-processing techniques are used to the acquired images. Depending on the particular task at hand, a range of analytical discriminating processes are then used to classify the photos. This is the fundamental strategy for the proposed image processing-based plant identification in this study.

1.3 OBJECT OF THE PROJECT

- To create a design that achieves the desired result while reducing the output of undesirable materials.
- To create a design that complies with the demands of the end user.
- To produce the appropriate output volume.
- Creating the output in the wrong format and sending it to the appropriate recipient.
- To deliver the results on time so that you can decide well.

1.4 LIMITATIONS OF PROJECT

- This model stresses a tried-and-true technique that was created with the use of some machine learning algorithms.
- Here, the procedure doesn't work as accurately as it should given the algorithms that were employed, and the classification isn't as exact as it should be.

1.5 ORGANIZATION OF DOCUMENTATION

1.5.1 FEASABILITY

This stage involves analysing the project's viability and presenting a business proposal that includes a high-level project plan and some cost projections. During the system analysis phase, the proposed system's viability must be evaluated. This is done to ensure that the intended system won't put a strain on the company's finances. A feasibility study requires knowledge of the fundamental requirements of the System.

Three main factors are taken into account in the feasibility analysis:

- Economic Feasibility
- Technical Feasibility
- Social Feasibility

Economic Feasibility

To determine the system's financial impact on the organisation, this study is being done. The corporation is only permitted to spend a certain amount of money on the research and development of the system. Charges must be substantiated. Due to the fact that most of the employed technologies were widely available, the system that was built was also completed within budget. You only need to purchase the modified products.

Technical Feasibility

This study is being done to determine the technological viability or requirements of the system. No system should be developed that puts a significant burden on the existing technological resources. The current technological infrastructure will be strained as a result. This will lead to the client having unrealistic expectations. Since deploying it just requires a few or no tweaks, the planned system must have a low demand.

Social Feasibility

The purpose of the study is to determine how acceptable the system is to users. Instructions on how to use the system efficiently are part of this. The user must accept the system as a necessity rather than perceive it as a danger. Their level of acceptance is entirely due to their methods. He is the final person to use the system, thus it is important to give him some self-confidence so he can make suggestions.

II. RELATED WORK

With around 391,000 vascular plant species worldwide, there are a colossal number of plant species [1]. Machine learning, an area of artificial intelligence (AI) that is currently very popular and in usage, has been employed in a variety of fields, including biology, medicine, computer vision, speech recognition, and others [5]. A strong framework for supervised learning is provided by deep learning, a contemporary AI methodology [6]. It can effectively and quickly transfer an input vector to an output vector, even when the dataset is enormous [7]. The practise of image enhancement is used to draw attention to certain aspects of an image [8]. One of the key components of the plant identification system that may be used to classify the leaves based on their surface structure is texture. It is an irregular pattern of spatial distribution of various picture intensities that focuses primarily on each individual pixel of an image [9, 10]. To extract the vein structure, Cope et al. [11] presented a genetic algorithm (GA) and ant colony algorithm-based evolved vein classifier. Anami et al [12].'S proposal for a plant identification system based on a combination of colour and textural data. Ann automatic leaf identification method for legumes was developed by Larese et al. [13] based solely on the vein architecture. Simple measurements were made to the vein anatomy, and the veins were subsequently identified using a Random Forests method. On the Flavia and Foliage datasets, Kadir et al [14].'S offered a different approach. A CNN method was suggested by Lee et al. [15, 17] to identify 44 plant species that were purchased from the Royal Botanic Gardens of Kew, England. Using a Multilayer Perceptron (MLP) and an SVM, the retrieved features were then categorised. Leaf patches (D1) and the entire image (D1) was the two datasets that were utilised (D2). The accuracy for both datasets was greater than 97 percent. Additionally, researchers integrated local and global features together in and attained more than 91 percent accuracy. Furthermore, Sladojevic et al. [19] used CNN to recognise plant illnesses. For the purpose of identifying plants, Grinblat et al. offered a study based on the morphological patterns of leaf veins.

III. PROPOSED WORK

The identification of vast plant variants is made easier for botanists because to the taxonomy of plant species, which is thought to be crucial for maintaining the biosphere. Utilizing the contours method, the data transformation data pre-processing methodology is employed to convert the plant picture data set into a numerical data set with the essential features as attributes. Additionally, classification is performed using three machine learning algorithms, Support Vector Machine (SVM), Random Forest (RF), and k-Nearest Neighbor (k-NN), and the accuracy of these algorithms is compared. The algorithm that gives the best accuracy is then used for additional testing on real-time images. Fig. 1 depicts the planned work's workflow.

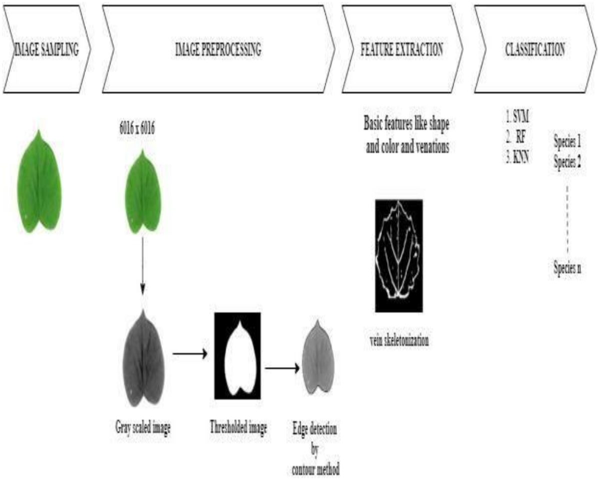

Fig. 1: Work flow diagram of proposed Plant Species classification using leaf veins.

IV. METHODOLOGY

The suggested study contains four basic processes, which are dataset collection, image pre-processing, feature extraction, and classification, as shown in Fig. 2. First, leaf samples were gathered, and photos from the Flavia dataset were obtained. Then, using a contours technique, the leaf pictures were pre-processed and input into the feature extraction step to extract the crucial data from the leaves. The retrieved characteristics were then trained and categorised using a variety of machine learning techniques.

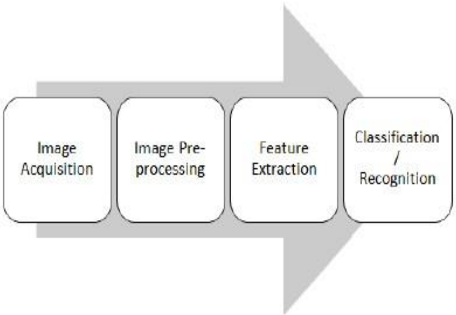

Fig.2: General Methodology for Automated Plant classification

A. Collecting Data-set:

The Flavia dataset, which has roughly 1907 pictures of 32 different species, was the source of the data for this project. Fig. 3 displays a few pictures from the Flavia dataset.

Fig. 3: images of 32 different species present in Flavia dataset. Here we show one image per Species.

B. Data pre-processing:

As the data we utilise may be vulnerable to disturbances, which may impact or supress the quality of the data, data pre-processing is crucial in machine learning. When a picture is acquired, noise is created by pixel values that do not accurately reflect the true intensities of the image. To emphasise or improve an image's key elements, it is necessary to first remove the background noise from the image. As a need for the data pre-processing inputs, the leaf images were reconfigured into square dimensions (m x m). To preserve the ratio of the leaf shape, the original 6016 x 4016 resolution photos were downsized to 1600 x 1200 resolutions. After that, the photos were kept in RGB format. For additional processing, this image with the backdrop removed is used. The pre-processed image and segmented images are shown in Fig. 4. Numerical data is converted from picture format (.jpg) to well-organized, well-organized data that increases data quality and shields programmes from possible landmines like null values, unexpected duplicates, wrong indexing, and incompatible format.

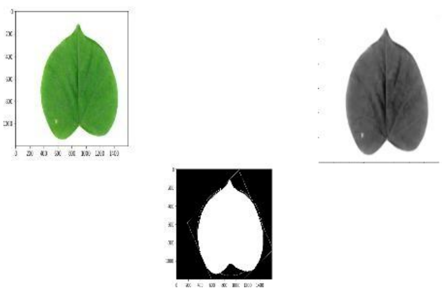

Fig. 4: Pre-processed image formats

C. Feature Extraction:

The various hand signs can be distinguished based on features. However, choosing a feature necessitates an accurate comprehension and interpretation of the derived feature values. To begin with, all photos were converted from RGB to greyscale. The region of interest (ROI) from the photos was then segmented out using Contour. Following segmentation, the pictures were postprocessed and skeletonized to ensure the creation of a clear image of the leaf. The vein morphological features were measured in order to extract the vein features from the segmented images. Inverse-difference-moments, end points, entropy, aspect ratio, areoles, rectangularity, and other factors were used to calculate the leaf area.

D. Classification:

An intelligent algorithm uses training data to recognise the distinctive characteristics of each particular plant species and classifies a fresh sample as belonging to the appropriate species. Classification is the final stage of an automated plant recognition system. Favoured machine learning techniques for identifying plants include Artificial Neural Networks (ANN), Support Vector Machines (SVM), and k-Nearest Neighbor (kNN). In this study, SVM, kNN, and Random Forest were utilised as three classification techniques (RF). Support Vector Machines (SVM), a supervised machine learning technology, are acknowledged as one of the powerful classification methods due to its excellent capacity for handling high dimensional space and data points that are not linearly spaced. When applied to feature mapped data, linear SVM can run quickly with little storage usage and enhance classification performance. In this study, linear SVM using a "One versus all" strategy was used because the dataset has multiple classifications.

A classification method known as k-NN categorises a sample using the majority opinion of its neighbours. The City block distance metric is used in this study to fix the number of neighbours at one. Without hyper-parameter adjustment, Random Forest is a flexible, user-friendly machine learning algorithm that frequently yields fantastic results. Because of how simple it is to use and how it can be applied to both grouping and relapsing tasks, it is also one of the calculations that is most frequently used. The appropriateness and utility of random forest calculation, along with a few other crucial aspects of it, can be derived from this context. Figure 5 illustrates the execution's flow.

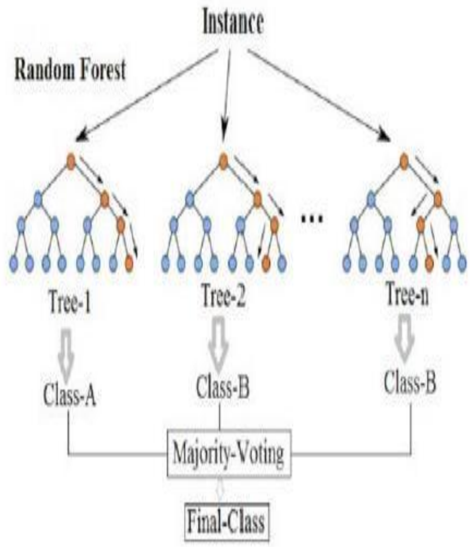

Fig.5: Random Forest Classification

V. RESULTS AND DISCUSSION

Following data pre-processing for performance evaluation, the data was divided into training and testing sets in a ratio of 70:30 and fitted to three distinct classifiers. We then expanded our work with testing by using some real-time photos of the different plant species. In TABLE I, the effectiveness of each classifier is discussed.

s.no	classifiers	Accuracy
1	Random Forest	90
2	Support Vector Machine	88
3	k-Nearest Neighbours	84

TABLE I: Classification accuracy of different classifiers

The best accuracy performance was noted by the Random Forest classifier, which attained 90%. However, the accuracy of the other classifiers was at least 85%. Therefore, RF is quite suited with contours for feature extraction model, according to this research. SVM and k-NN, however, performed worse with contours. However, the traditional feature extraction approach must manually extract each sort of feature, which takes a lot of time. For instance, if form features are taken into account in this study, distinct sets of algorithms will be needed for segmentation and then shape feature extraction. In order to create an automated system for classifying plant species, machine learning is more relevant and practical than traditional methods.

VI. CONCLUSION

The suggested methodology uses minimal computer resources, is effective, and doesn't call for further training. This architecture utilises multiple deep learning- and CNN-based techniques, as well as recent improvements in pre-trained models. The scalability of large applications with this paradigm is enormous due to the system's overall streamlining. We classified images of various plant species successfully using deep learning. In this instance, we trained CNN on a dataset of five different classifications of plant species. The user can test the system after the training by uploading an image and viewing the results in the various categories.

REFERENCES

[1]. Trishen Munisami, Mahess Ramsurn, Somveer Kishnah, Sameerchand Pudaruth, “Plant Leaf Recognition using Shape features and color Histogram with K-nearest classifiers”, Second International symposium on Computer Vision and Internet (VisionNet’15), Elsevier Procedia Computer Science 58(2015) 740-747.-2020.

[2]. Olfa Mzoughi, Itheri Yahiaoui, Nozha Boujemaa, Ezzeddine Zagrouba,” Multiple leaflets- based identificationapproachfor compound leaf species”, 2019

[3]. Guillermo L. Grinblat, Lucas C. Uzal, Mónica G. Larese, Pablo M. Granitto, “Deep learning for plant identification using vein morphological patterns”, Elsevier-2019

[4]. Sue Han Leea, Chee Seng Chana, ∗, Simon Joseph Mayo b, Paolo Remagninoc, “How deep learning extracts and learns leaf features for plant classification”,Elsevier,20183

[5]. Aravind Krishnaswamy Rangarajan, Raja Purushothaman*, Anirudh Ramesh,” Tomato crop disease classification using pre-trained deep learning algorithm”, Elsevier -2018.

[6]. Xihai Zhang, (Member, IEEE), Yue Qiao, Fanfeng Meng, Chengguo Fan And Mingming Zhang, "Identification of Maize Leaf Diseases Using Improved Deep Convolutional Neural Networks" IEEEAccess, 2019.

[7]. Ali Beikmohammadi, Karim Faez, "Leaf Classification for Plant Recognition with Deep Transfer Learning", 4th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS), IEE, 2018.

[8]. Jing Hu, Zhibo Chen∗, Meng Yang, Rongguo Zhang and Yaji Cui, "A Multi-Scale Fusion Convolutional Neural Network for Plant Leaf Recognition" IEEE, 2018

[9]. Jing Wei Tan, Siow-Wee Chang, Sameem Abdul-Kareem, Hwa Jen ap, Kien-Thai Yong, "Deep Learning for Plant Species Classification using Leaf VeinMorphometric",IEEE, 2018

[10]. Ali Beikmohammadi, Karim Faez, "Leaf Classification for Plant Recognition with Deep Transfer Learning", 4th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS), IEE, 2018.

[11]. Jing Hu, Zhibo Chen∗, Meng Yang, Rongguo Zhang and Yaji Cui, "A Multi-Scale Fusion Convolutional Neural Network for Plant Leaf Recognition" IEEE, 2018

[12]. Jing Wei Tan, Siow-Wee Chang, Sameem Abdul-Kareem, Hwa Jen ap, Kien-Thai Yong, "Deep Learning for Plant Species Classification using Leaf VeinMorphometric",

[13]. Voncarlos M. Araujo ´ ∗, Alceu S. Britto Jr∗§, Andre L. Brun ´ ∗, Luiz E. S. Oliveira†, Alessandro L. Koerich, "Fine-Grained Hierarchical Classification of Plant Leaf Images using Fusion of Deep Models", IEEE conf, 2018

[14]. Peng Jiang, Yuehan Chen, Bin Liu, Dongjian He, and Chunquan Liang, "Real-Time Detection of Apple Leaf Diseases Using Deep Learning Approach Based on Improved Convolutional Neural Networks", IEEE Access, 2019

[15]. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with Deep convolutional neural networks," in Advances in neural information processing systems,2018, pp. 1097-1105.

[16]. C.Szegedy et al., "Going deeper with convolutions," in Proceedings of the IEEE Conference on computer vision and pattern recognition, 2017, pp. 1-9.

[17]. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition, "In Proceedings of the IEEE conference on computer vision and pattern recognition, 2016,Pp. 770- 778.