PLANT SPECIES IDENTIFICATION

A PROJECT WORK SYNOPSIS

Submitted By:

KAKKIRALA LOHITH KUMAR 20BCS3880

ANKIT GHOSAL 20BCS6621

LALIT BISHT 20BCS6609

in the partial fulfilment for the award of the degree of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE ENGINEERING

Under the supervision of:

Ms. Merry K.P

NOVEMBER 2022

TABLE OF CONTENTS

ABSTRACT

Species knowledge is essential for protecting biodiversity. The identification of plants by conventional keys is complex, time consuming, and due to the use of specific botanical terms frustrating for non-experts. This creates a hard to overcome hurdle for novices interested in acquiring species knowledge. Today, there is an increasing interest in automating the process of species identification. The availability and ubiquity of relevant technologies, such as, digital cameras and mobile devices, the remote access to databases, new techniques in image processing and pattern recognition let the idea of automated species identification become reality. This paper is the first systematic literature review with the aim of a thorough analysis and comparison of primary studies on computer vision approaches for plant species identification. We identified 120 peer-reviewed studies, selected through a multi-stage process, published in the last 10 years (2005–2015). After a careful analysis of these studies, we describe the applied methods categorized according to the studied plant organ, and the studied features, i.e., shape, texture, color, margin, and vein structure. Furthermore, we compare methods based on classification accuracy achieved on publicly available datasets. Our results are relevant to researches in ecology as well as computer vision for their ongoing research. The systematic and concise overview will also be helpful for beginners in those research fields, as they can use the comparable analyses of applied methods as a guide in this complex activity.

1. INTRODUCTION

All life on Earth depends on plants, which are also a vital resource for people's health. Botanists can utilise this application for therapeutic purposes, but plant recognition is crucial for managing plant species in agriculture. The properties of each plant's leaf can be utilised to categorise the plant.

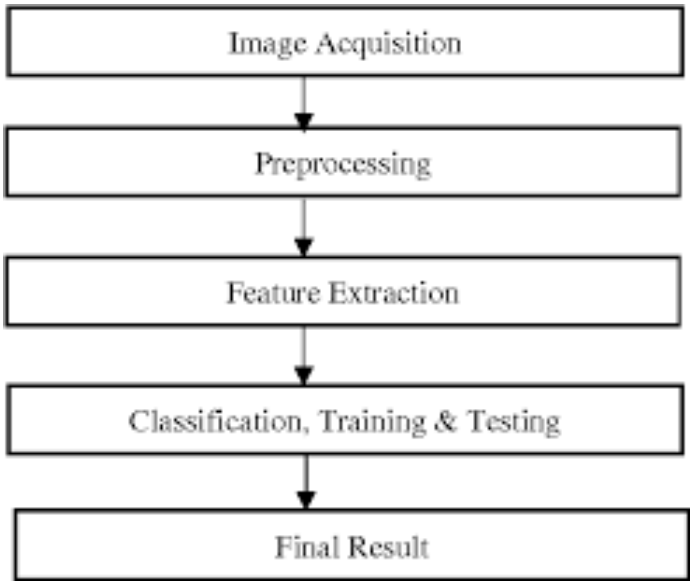
This study uses machine vision and digital image processing to propose a straightforward and computationally effective approach for identifying plants. Three stages make up the suggested methodology: pre-processing, feature extraction, and classification. The process of improving data pictures before computational processing is called pre-processing. Based on the colour and shape of the leaf image, features are extracted during the feature extraction phase.

Machine learning (ML) is the study of computer algorithms that improve automatically through experience and by the use of data. It is seen as a part of artificial intelligence. Machine learning algorithms build a model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are used in a wide variety of applications, such as in medicine, email filtering, and computer vision, where it is difficult or unfeasible to develop conventional algorithms to perform the needed task.

In addition to an informed, working definition of machine learning (ML), we detail the challenges and limitations of getting machines to ‘think,’ some of the issues being tackled today in deep learning (the frontier of machine learning), and key takeaways for developing machine learning applications for business use-cases. Machine learning is a method of data analysis that automates analytical model building. It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention.

1.1. PROJECT OVERVIEW

Our project is based on ML classification algorithm and through this project we will detect whether the person is suffering from Parkinson Disease or not for that, first we have to collect the dataset of the patients. So, from Kaggle we will download the plant species data set obtained from Machine Learning repository. These are stored into the python environment as Testing and Training datasets and imported using necessary packages. Python is an open source dynamic, high level, free and interpreted programming language. This supports object-oriented programming and procedural programming. Python is currently the most popular programming language for Machine Learning research and development. Jupyter notebook is an integrated development environment (IDE) primarily for the Python language, used in computer programming.



1.2. PROBLEM STATEMENT

Both the harmony of nature and the life of humans depend on plants. Nearly all animals, who are unable to produce their own food, rely on them as their primary source of nutrition and metabolic energy. Because plants are an essential component of life on Earth and provide humans with food and oxygen, it is important to study them.as well as other living things. Botanists' work can be automated by using a digital plant identification system to quickly characterise plant species without their knowledge. There are various freely accessible leaf picture databases, including the Image CLEF dataset, Leaf snap dataset, Intelligence dataset, and Flavia dataset. Using the Flavia dataset, the experiment's performance is assessed.

The study of finding, identifying, describing, categorising, and naming plants is known as plant taxonomy. It is one of the most important branches of taxonomy

Taxonomy and plant systematics are intricately intertwined. There isn't a definite difference between the two. In actuality, "plant taxonomy" is concerned with the actual handling of plant specimens, while "plant systematics" is concerned with the links between plants and their evolution, particularly at higher levels. But just as the objectives and strategies employed have changed, so too has the precise link between taxonomy and systematics.

1.3. PROBLEM FORMULATION

The project aims and objectives that will be achieved after completion of this project are discussed in this subchapter. The aims and objectives are as follows: The aim of this project is to compare various machine learning models in the successfully identify the correct plant species using the image of leaf

2. HARDWARE/SOFTWARE SPECIFICATION

2.1. HARDWARE SPECIFICATION

PC/LAPTOP:

A laptop, laptop computer, or notebook computer is a small, portable personal computer (PC) with a screen and alphanumeric keyboard Laptops typically have a clam shell form factor with the screen mounted on the inside of the upper lid and the keyboard on the inside of the lower lid, although 2-in-1 PCs with a detachable keyboard are often marketed as laptops or as having a "laptop mode".

Component	Minimum requirement
Processor	64-bit, four-core, 2.5 GHz minimum per core (If your dataset size is significantly larger than the medium dataset, we recommend 8 cores.)
RAM	16 GB
Hard disk	80 GB

2.2. SOFTWARE SPECIFICATION

2.2.1 ANACONDA /JUPYTER NOTEBOOK:

Jupyter Notebook is a web-based open-source application that is used for editing, creating running and sharing documents that contain live codes, visualization, text, and equations. Its core supported programming languages are Julia, R, and Python. Jupyter notebook comes with an IPython kernel that allows the programmer to write programs in python. There are over 100 kernels other than IPython available for use.

2.2.2 PYTHON:

A high-level, all-purpose programming language is Python. Code readability is prioritized in its design philosophy, which makes heavy use of indentation. Python uses garbage collection and has dynamic typing. It supports a variety of paradigms for programming, including functional, object-oriented, and structured programming.

3. LITERATURE SURVEY

Although a significant amount of research has been done studying various aspects of leaf identification in inventory systems, most of it deals with semi-automated systems. A state-of-the-art system which is fully automated and requires least human interaction is yet to be developed.

Pavan et al. proposed an algorithm for identification using multiclass classification based on color, shape volume and cell feature. They performed three stage comparisons: first stage compares redness, greenness, blueness index feature, second stage compare shape feature and the last stage compares cell feature and volume fraction feature. Experiment is performed on a sample of diverse collection of 1000 leaf and flower images. Limitations of this approach is that it semi-automatic approach and its recognition rate is up to 85% percent on an average.

Arun Priya the proposed approach consists of three phases such as pre-processing: transforming to grey scale and boundary enhancement, feature extraction: derives the common DMF from five fundamental features and classification: Support Vector Machine (SVM) classification for efficient leaf recognition. 12 leaf features which are extracted and orthogonalized into 5 principal variables are given as input vector to the SVM.

Valliamal et al. A probabilistic curve evolution method with particle filters is used to measure the similarity between shapes during matching process. The experimental results prove that the preferential image segmentation can be successfully applied in leaf recognition and segmentation from a plant image.

Dr. H.B.Kekre et al. the method of CBIR is discussed in this paper to filter images based on their content. In this paper feature vector is generated using color averaging technique, similarity measures and performance evaluation. Precision –Recall cross over plot is used as the performance evaluation measure to check the algorithm. The effect due to the size of database and number of different classes is seen on the number of relevancies of the retrievals.

Javed et al. used PNN to classify the plants with broad flat leaves. In this algorithm there were few select point where the user needs to specify the leaf blades and a base point according to which the image is then aligned and compared with other images based on some features like area, eccentricity, etc. They used 1200 sample leaves belonging to 30 different plants to train their system. This system is also semi-automatic and 91.41 percent accurate.

Arora. A et al. categorized the different images and used a variety of novel pre-processing methods such as shadow and background correction, petiole removal and automatic leaflet segmentation for identifying the leaf blobs. Also used complex network framework along with novel tooth detection method and morphological operations to compute several useful features. They used the PlantLeaves II dataset.

4. METHODOLOGY

In the identification of plant species , ML techniques have proven to be effective This study uses both unsupervised and supervised learning techniques to diagnose the image of leaf through classification prediction. Several approaches that entail clustering, reducing dimensionality, and learning of prediction approaches are used to create the pre-processing and feature extraction methods are used to get the image details

Data pre-processing, dimensionality reduction using random forest using ensembles of EM (Expectation -Maximization), and prediction using ensembles of all stages of the method that are utilized to predict that the leaf belongs to which plant species

5. RESULT

ML has been used for medical disease detection lately and particularly Parkinson's disease (PD) treatment. This can be explained by the convenient performance and accurate results of ML techniques. Classification of diseases is a significant type of predictive modelling. It is considered an important data mining approach because it clusters the population referring to a predetermined criterion. It is vital to compare the outcomes of various classification methods to decide which approach presents the best performance. Hence, the main goal of this research is to assess several approaches that are utilized for PD prediction and classification. Even though ML methods have been assessed in several studies separately, the evaluation of these methods based on various datasets makes it complex to perform an accurate comparison among the deployed methodologies. Hence, it is vital to evaluate these methods in one comparative study based on a chosen dataset

6. CONCLUSION AND FUTURE SCOPE

6.1. CONCLUSION

In this study, we have presented baseline automated identification techniques of plant species based on herbarium leaf images, using the pattern recognition approach. Five machine learning algorithms: Random Forest, SGD, KNN, XGB, and SVM have been used to build identification models. Both models achieved satisfactory results demonstrating their usefulness in identification tasks. The study presented here showed that automated classification of plant species that had similar leaf shapes is feasible based on leaf images. Although the SVM classifier has given us best results Though the developed system is not intended to replace human taxonomists, it may provide a rapid and easily accessible technique to identify plants with acceptable accuracy. We chose to work on species of plants as it is a large genus and species identification can be difficult especially to non-taxonomists

6.2. FUTURE SCOPE

In future work, we can focus on different techniques to identify the plant species using different datasets. In this research, we use different classifiers to identify the plant species. In the future we will use different parts of plants to identify the species of the plant

7. REFERENCES

1. Arora A., Gupta A., Bagmar N., Mishra S., Bhattacharya A.: A Plant Identification System using Shape and Morphological Features on Segmented Leaflets: Team IITK, CLEF 2012 In: CLEF (Online Working Notes/ Labs/Workshop). (2012)
2. Neeraj Kumar, Peter N. Belhumeur, Arijit Biswas, David W. Jacobs, W. John Kress, Ida C. Lopez, and João V. B. Soares, "Leafsnap: A Computer Vision System for Automatic Plant Species Identification", Computer Vision – ECCV 2012, 2012, pp 502-516.
3. Pavan Kumar Mishra, Sanjay Kumar Maurya, Ravindra Kumar Singh, Arun Kumar Misra, "A semi-automatic plant identification based on digital leaf and flower Images", International Conference On Advances In Engineering, Science And Management (ICAESM), 2012, pp. 68-73.
4. ArunPriya C., Balasaravanan T., Antony Selvadoss Thanamani, "An Efficient Leaf Recognition Algorithm for Plant Classification Using Support Vector Machine", Proceedings of the International Conference on Pattern Recognition, Informatics and Medical Engineering, 2012, pp. 428-432.
5. N.Valliammal, Dr. S.N.Geethalakshmi, "Automatic Recognition System Using Preferential Image Segmentation for Leaf and Flower Images", Computer Science & Engineering: An International Journal (CSEIJ), 2011, pp. 13-25.
6. Dr. H.B.Kekre, Mr. Dhirendra Mishra, Ms. Stuti Narula, Ms. Vidhi Shah, "Color Feature Extraction for CBIR", International Journal of Engineering Science and Technology (IJEST), 2011, pp. 8357-8365.
7. Javed Hossain, M. Ashraful Amin, "Leaf Shape Identification Based Plant Biometrics", Proceedings of 13th International Conference on Computer and Information Technology (ICCIT), 2010, pp. 458-463.
8. Jiazhi Pan and Yong He "Recognition of plants by leaves digital image and neural network", International Conference on Computer Science and Software Engineering, 2008, pp 906-910.
9. Stephen Gang Wu, Forrest Sheng Bao, Eric You Xu, Yu-Xuan Wang, Yi-Fan Chang, Qiao-Liang Xiang, "A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network", International Symposium on Signal Processing and Information Technology, IEEE, 2007, pp 11-16.
10. JiXiang Du, Xiao-Feng Wang and Guo-Jun Zhang, "Leaf shape based plant species recognition", Applied Mathematics and Computation 185 , 2007 pp 883-893. [11] Qingfeng Wu, Changle Zhou, and Chaonan Wang, "Feature Extraction and Automatic Recognition of Plant Leaf Using Artificial
11. Abu A, Lim LHS, Sidhu AS, Dhillon SK. 2013b. Semantic representation of monogenean haptoral bar image annotation. BMC Bioinf. 14(1): 48. doi: 10.1186/1471-2105-14-48 [Crossref], [PubMed], [Google Scholar]
12. Berg CC, Corner EJH, Jarrett FM. 2006. Moraceae-genera other than Ficus: Flora Malesiana-Series 1. Spermatophyta. 17(1):1–146. [Google Scholar]
13. Boser BE, Guyon IM, Vapnik VN. 1992. A training algorithm for optimal margin classifiers. In: Haussler D, editor. 5th annual ACM workshop on COLT. Pittsburgh (PA): ACM Press; p. 144–152. [Crossref], [Google Scholar]