A Minor Project Report on

**Plant Leaves Recognition**

**Using Image Processing**

Submitted in Partial Fulfillment of the

Requirements for Degree of

**Bachelor’s of Engineering in Software Engineering**

Under Pokhara University

**Submitted By:**

Aasma Sharma (15780)

Sanjaya Shrestha (15761)

**Supervisor:**

**Himal Acharya**

Date:

16 Dec 2018

Department of Software Engineering

**NEPAL COLLEGE OF**

**INFORMATION TECHNOLOGY**

Balkumari, Lalitpur (*PU Affiliated*)

ACKNOWLEDGEMENT

We had taken efforts in this project. However, it would not have been possible without the kind support and help of many individuals and organizations. We would like to extend our sincere thanks to all of them.

We are highly indebted to Nepal College of Information Technology for their guidance and constant supervision as well as for providing necessary information regarding the project & also for their support in completing the project.

We would like to express our deepest gratitude towards our supervisor **Er. Himal Acharya** for his kind co-operation, encouragement and supervision which helps us in completion of this project.

We would like to express our special gratitude and thanks to **Dr. Roshan Chitrakar** for giving us this opportunity to explore our idea and implement them as projects.

Our thanks and appreciations also go to our colleague in developing the project and people who have willingly helped us out with their abilities.

Aasma Sharma (15780)

Sanjaya Shrestha (15761)

December 2018

ABSTRACT

*Since recent decades, digital image processing, image analysis and machine vision have been sharply developed, and they have become a very important part of artificial intelligence and the interface between human and machine grounded theory and applied technology. These technologies have been applied widely in industry and medicine, but rarely in realm related to agriculture or natural habitats.*

*The rapid development in the artificial intelligence is rarely directed towards the plants or natural habitats which are dominantly present in the world. Only few steps are seems to be taken in this direction. Agriculture, farming is the main occupation that keeps us alive so we need to have such technology which can recognize plants and its usage so the conservation can be done. The resources will not go waste.*

*So we would like to propose to develop such a software which can recognize plant by its leaves and can inform the usage, conditions, habitat etc. We will use image processing, deep learning so by taking leaves we will recognize the plants.*

**Keywords: Image Processing, image analysis, machine vision, artificial intelligence**

# Table of Contents

[I. Table of Contents III](#_Toc532679678)

[II. List of figures IV](#_Toc532679679)

[III. INTRODUCTION 1](#_Toc532679680)

[A. PROBLEM STATEMENT 1](#_Toc532679681)

[B. PROJECT OBJECTIVES 1](#_Toc532679682)

[C. SIGNIFICANCE OF STUDY 1](#_Toc532679683)

[D. SCOPE AND LIMITATIONS 1](#_Toc532679684)

[IV. LITERATURE REVIEW 2](#_Toc532679685)

[A. REVIEW 2](#_Toc532679686)

[B. LEAF RECOGNITION APPROACH 2](#_Toc532679687)

[1. Image Pre-Processing 2](#_Toc532679688)

[2. Feature Extraction 2](#_Toc532679689)

[3. Principal Component Analysis (PCA) 3](#_Toc532679690)

[4. K-NN Classification 3](#_Toc532679691)

[5. Leaf Snap 4](#_Toc532679692)

[V. METHODOLOGY 5](#_Toc532679693)

[A. SOFTWARE DEVELOPMENT LIFE CYCLE 5](#_Toc532679694)

[1. Waterfall Model 5](#_Toc532679695)

[2. Tools and Technologies Used 6](#_Toc532679696)

[3. Time Schedule 7](#_Toc532679697)

[B. SYSTEM FLOWCHART 8](#_Toc532679698)

[1. Plant Snap Samples Collection 8](#_Toc532679699)

[2. Image Acquisition 8](#_Toc532679700)

[3. Image Processing 8](#_Toc532679701)

[4. Feature Extraction and Classification 8](#_Toc532679702)

[5. Plant Feature Database 9](#_Toc532679703)

[VI. Conclusion 10](#_Toc532679704)

[VII. Recommendations 10](#_Toc532679705)

[VIII. References 11](#_Toc532679706)

# List of figures

[Fig 1: Waterfall Model 5](#_Toc532679707)

[Fig 2: Leaf database overview 7](#_Toc532679708)

[Fig 3: Gantt Chart 7](#_Toc532679709)

[Fig 4: System proposed approach 8](#_Toc532679710)

[Fig 5:Plant species with their class labels. 9](#_Toc532679711)

[Fig 6: Features stored in excel file 9](#_Toc532679712)

# **INTRODUCTION**

Plant Leaf Recognition using Image Processing, is a project primarily focused to identify the plant on the basis of leaves features. It uses different image processing algorithms to extract features of leaves from image. It is a desktop app capable of accepting an image input and identifying the plants. This section explains more about the project scope and objectives.

## **PROBLEM STATEMENT**

The scope of AI and machine learning is increasing, people are being interested more on robotics which implements computer vision in huge amount. Basically, computer vision deals with the analyzing the environment on basis of image and video processing. The development of robotics or computer vision towards agricultural field is still behind, Projects are rarely done and implemented. This project can be a very good initiative to develop the scope of computer vision and to increase the products on agricultural field. There are many plants around us but we are unaware about the existence, value and use of those plants. By this project now people can identify the plants and their use.

## **PROJECT OBJECTIVES**

The main objective of this project is to provide a system capable of recognizing plant through it’s leaves and to encourage further development of computer vision in light of agriculture and environment.

## SIGNIFICANCE OF STUDY

The study aims to find a good platform to know about the local plant around us. Plants are abundant around us having knowledge of those plants and their properties will be very helpful. Trekkers, Student, Re-searchers will find this helpful as they can use it in their field.

## SCOPE AND LIMITATIONS

The main scope is to provide a platform to identify plants efficiently. The application is specially targeted to students, researchers, biologist, trekkers, who needs to study about the plants their use properties.

The limitation of this projects are:

1. User can know about certain plants and species only, as the app cannot cover all the plants around the globe.

2. This app can only identify the morphological differences of plants not the differences of their genetic characteristics.

3. The system is limited to accept input images with green images on white background.

# LITERATURE REVIEW

This section is about the literature review of project regarding its different topic and a overview of theoretical basis of project. The projects will be using different algorithms libraries and concept which will be explained here.

## **REVIEW**

There are a number of existing literatures that experimented on various methods of classifying plant leaves.

Chaki and Parekh [1] proposed an algorithm by extracting a leaf’s shape using Centroid-Radii and Moments Invariant models. They used Hu’s moments for pre processing and normalization of images which produced features that are invariant to translation rotation , and scaling. From this , the centroid are computed and the radii length (or the distance from centroid to boundary points) are obtained at regular angle intervals. Their system produced 97.19% accuracy.

Du et al [2] used a moving median center (MMC) hypersphere classifier in their system. The MMC main algorithm is to find an initial center and have that “hop” from point to point, simulating hyper-spheres. Their experiment yielded 91% accuracy.

Gwo, Wei and Lei [3] did rotary matching of edges to accommodate leaf edge variation. This is carried out by detecting feature points, rotary matching these to find the optimal composition, and modeling the leaf using Viterbi training algorithm. Their system is highly dependent on the retrieval of good feature points.

Wu et al [4] used Problastic Neural Network (PNN) classifier. They use basic geometrical features to extract 12 digital morphological features and applied Principal Component Analysis(PCA). They reported a result of 90% accuracy

## LEAF RECOGNITION APPROACH

This approach consists of three phases namely image pre-processing, feature extraction and classification.

### Image Pre-Processing

This is a feature extraction phase involved in this approach.

#### Units

The digital image of 24 bit RGB image with resolution 720 by 920 pixels. An RGB image is firstly converted into grayscale image. Equation 1 is used to convert RGB value of pixel into its grayscale value.

Gray=0.2989 \* R + 0.85870 \* G + 0.1140\*B (1)

#### Boundary Enhancements

The margin of leaf is highly focused in this pre-processing step. Convolving the image with a Laplacian filter of 3 X 3 spatial mask.

### Feature Extraction

This approach uses a 5 common Digital Morphological Features (DMFs), based on statistical properties of the Intensity histograms of grayscale transformations of the original RGB images.

#### Average Intensity

Average intensity is defined as the mean of the intensity image, m.

#### Average Contrast

Average contrast is the standard deviation of the intensity image,

#### Smoothness

Smoothness is defined as and measures the relative smoothness of the intensities in a given region, For a region of constant intensity, R takes the value 0 and R approaches 1 as regions exhibit larger disparities in intensity values is generally normalized by (L-1)2 to ensure that R € [0,1].

#### Third Moment

µ3is a measure of the intensity histogram’s skewness. This measure is generally normalized by (L-1)2 like smoothness.

#### Uniformity

Defined as uniformity,s maximum value is reached when all intensity levels are equal.

#### Entropy

It is a measure of entropy randomness.

### Principal Component Analysis (PCA)

The main aim of PCA is topresent information of original data as the linear integration of certain linear irrelevant variables. PCA alters the data to a new co-ordinate system such that the greatest variance by any projection of data comes to lie on the first coordinate, the second greatest variance on the second coordinate, and so on. Each coordinate is called principalcomponent.

### K-NN Classification

K-NN algorithm is one of the simplest classification algorithm and is also most used learning algorithms. KNN is a non parametric, lazy learning algorithm. Its purpose is to use a database in which the data points are separated into several classes to predict the classification of new sample point.

K-NN Algorithm is based on feature similarity: How closely out-of-sample features resemble our training set determines how we classify a given data point. It is used for classification-the output is a class membership(predicts a class-a discrete value). An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k-nearest neighbors. It can also be used for regression -output is the value for the object (predicts continuous values). This value is the average (or median) of the values of its k nearest neighbors. [6]

#### Algorithm

A case is classified by a majority vote of its neighbors, the case being assigned to the class most common amongst its K nearest neighbors measured by a distance function. If K=1, then the case is simply assigned to the class of its nearest neighbors.

**Distance Functions**

1. Euclidean distance
2. Manhattan
3. Minkowski

It should also be noted that all three distance measures are only valid for continuous variables. In the instance of categorical variables the Hamming distance must be used. It also brings up the issue of standardization of the numerical variables 0 and 1 when there is a mixture of numerical and categorical variables in the dataset.

**Hamming Distance**

#### Value of K

Choosing the optimal value for K is best done by first inspecting the data. In general, a large K value is more precise as it reduces the overall noise but there is no guarantee. Cross-validation is another way to retrospectively determine a good K value by using an independent dataset to validate the K vale. Historically, the optimal K for most datasets has been between 3-10. That produces much better result than 1 NN. [7]

### Leaf Snap

Leaf Snap is a series of electronic field guide developed by researchers from Columbia University, the University of Maryland, and the Smithsonian Institution. The free mobile apps use visual recognition software to help identify tree species from photographs of their leaves. They contain beautiful high-resolution images of leaves, flowers, fruits, petioles, seeds and bark.

The original Leafsnap currently includes trees found in the Northeastern United States and Canada, and will soon grow t include the trees of the entire continental United States. The high-resolution images in the original app were created by the conservation organisation. [8]

# METHODOLOGY

This section includes the brief description of stages of development of Plant Leaf Recognition projects.

## SOFTWARE DEVELOPMENT LIFE CYCLE

We divide the course of development in two section implementing two development cycle. The first one is Image processing and second is classification of image. The first section implements Waterfall Model of development cycle where system becomes capable of interpreting images and extracting features.

### Waterfall Model



Fig 1: Waterfall Model

Waterfall model is the first widely used in software development to ensure success of project. It is also referred to as a linear sequential life cycle model. It is very simple to understand and use. In a waterfall model, each phase must be completed before the next phase can begin and there is no overlapping of phases.

#### Requirement Analysis

In this phase we worked to gather the required features and possibilities of final version of project. We configured the output and input process, architecture and block diagram of the overview of project. The information includes the references site, review,similar products, features,algorithms.

#### System Design

The above research materials suggest us different approach of doing the projects, In this phase we discussed about the requirements for developing the projects. The overall system architecture is discussed in this period. The system requirement like Pre-processing algorithm, classifier, dataset, language, platform, limitations etc are discussed. The sequential steps of development is designed.

#### Implementation and Coding

Here in this period we begin to implement algorithm and creating the system. Initially we developed a system capable of loading images and resizing it, so now the system can use the image. Then the processing algorithm is implemented which results in preparation of images for feature extraction. According to feature the algorithm and library is used which gives s feature values.

#### Integration and Testing

The individually coded feature wise code or class is now merged together to work combinedly in system. Simultaneously we run through the testing series. The performance of system is checked whether the criteria is meeting or not. The changes or conversion we made is relevant or not. Likewise the feature wise code is integrated to a system.

#### Deployment and Maintenance

The work done is then evaluated and required maintenance is done. The issues like problem in using and understanding the system solved or updated.

### Tools and Technologies Used

The development becomes easy and possible by the use of several enterprise application software, tools and libraries which are discussed here.

#### OpenCV

Open Source Computer Vision libraries, is most used and a core of this project. The inbuilt functions and class becomes really helpful for acceleration of this projects. The image processing and K-NN classifier is done using OpenCV.

#### Dataset

We used the dataset called “leaf” which provides us 40 different plant species with their scientific name and eaf specimens available by species. It consists of leaf image photographed over a coloured background using Apple iPAD device. The 24bit RGB image with resolution 720X920 pixels. Binary versions are also provided for simple leaves.[8]

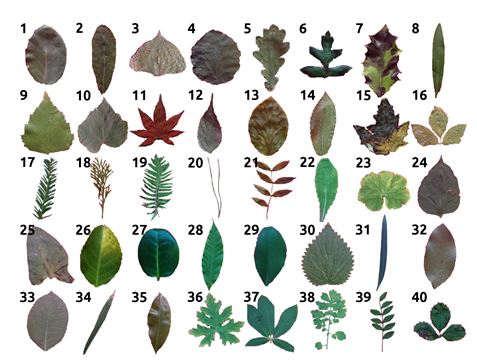


Fig 2: Leaf database overview

#### Eclipse IDE

For coding and running application we use eclipse IDE as this is suggested by Oracle for Java. The environment of eclipse is easy to use and debug our code and helps in importing packages online.

### Time Schedule

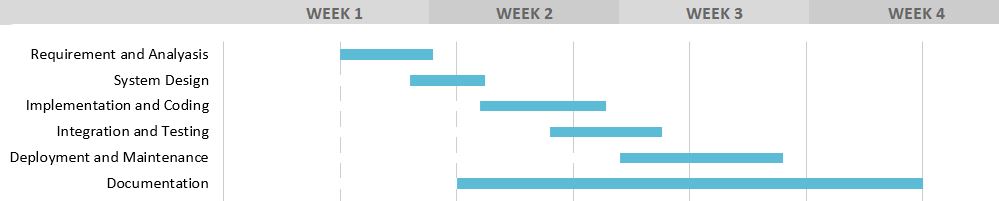


Fig 3: Gantt Chart

The project was completed working in these 6 sections as illustrated in Gantt chart. The time and work is divided as per the requirement and complexity. Simultaneously the research was also going on to find out and implement efficient way of developing.

## SYSTEM FLOWCHART

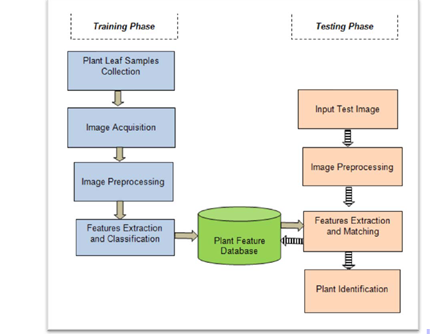


Fig 4: System proposed approach

### Plant Snap Samples Collection

Searching through Internet we found many leaves datasets. But all are not applicable, Since they provide a lot of species and images which we could not handle in current scenarios. So the dataset “leaf” [8] is used for training purpose.

### Image Acquisition

The image or samples taken is now prepared for supplying to system. The samples may be of big size and more colorful. In this we reduce the size of images to 720 by 920 pixels and into 24 bit images. The operation is carried out by implementing the OpenCV inbuilt libraries.

### Image Processing

This consists of processing image to extract features. First we convert the image to grayscale image and then processed to extract features. The images are prepared to extract features like Average Contrast, Average Intensity, Smoothness, Third Moment, Uniformity, Entropy. All the operations are carried out by using OpenCV libraries.

### Feature Extraction and Classification

The used database consists of 20 classes with different lab specimens images. The feature extracted is then stored in a excel file with its class labels 1 to 20



Fig 5:Plant species with their class labels.

### Plant Feature Database

Currently instead of database we had used excel file for the feature storage.

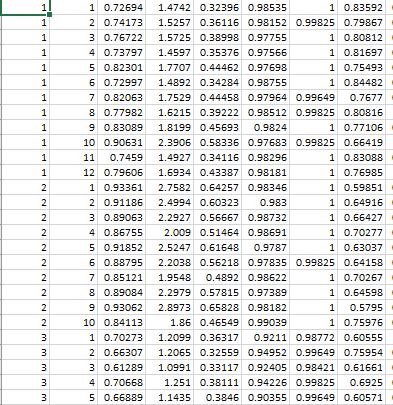


Fig 6: Features stored in excel file

This is the excel file which consists of the feature extracted from image. The first column indicates the class label which ranges from 1 to 20. The second column indicates the specimen no which are used for feature extraction. The 6 features extracted are then sequentially stored column wise in the sheet.

# Conclusion

The motivation that lead us to develop the Plant Leaf Recognition system is to initiate the development path of in field of Deep Learning and AI. Since the agriculture sector is lagging behind in todays world developing a good and advanced technology may lead agriculture forward. The population growth is increasing rapidly and by the side the agricultural production is becoming challenging due to environment changes and lack of human interest. By the 2050 A.D data shows us the occuerence of Global Food Crisis. If the current trend is not controlled then it may be more sooner. We believe developing this kind of system not only attracts the technology but also increases the interest of people on agriculture.

# Recommendations

Due to time limit many features that can be added to the system is not touched. This sections highlighted the features that can be added to make a system more reliable and usable.

1. System accepts only a simple leaves images which can be integrated to make acceptable all possible leaves type.
2. There are more species around the world, Including all possible species will makes system more reliable.
3. System is limited to a leaves input only but plant have other distinct features also which will make easier for classification
4. System is limited within a desktop which seems inapplicable, integrating to a android or mobile app will further increase its scope.

# References

[1] J. Chaki and R. Parekh,”Plant Leaf Recognition using shape based features and neural network classifiers,” International Journal of Advanced Computer Science and Applications, vol. 2, no. 10, p. 7, 2011.

[2] J. -X. Du, X. -F. Wang, and G.-J Zhang, “Leaf shape based plant species recognition,” *Elsevier – Applied Mathematics and Computation 2007.*

[3] C.-Y. Gwo, C.-H. Wei, and Y. Li “Rotary matching of edge features for leaf recognition “ *Elsevier – Computers and Electronics in Agriculture.*

[4] S.Wu, F. Bao, E. Xu. Y.-X.Wang, Y.-F Chang, and Q.-L. Xiang, “A leaf recognition algorithm for plant classification using problastic neural network,” IEEE, 2007

[6] <https://medium.com/@adi.bronshtein/a-quick-introduction-to-k-nearest-neighbors-algorithm-62214cea29c7> [Online]

[7] <https://www.saedsayad.com/k_nearest_neighbors.htm> [Online]

[8] <http://leafsnap.com/> "Leafsnap: A Computer Vision System for Automatic Plant Species Identification," Neeraj Kumar, Peter N. Belhumeur, Arijit Biswas, David W. Jacobs, W. John Kress, Ida C. Lopez, João V. B. Soares, Proceedings of the 12th European Conference on Computer Vision (ECCV), October 2012