Chapter 3

Plant Foliar Disease Identification Model (PFDIM)

3.1 Introduction

Detection of plant leaves is an interesting area of pattern recognition because each plant leaves have different shapes and sizes. Many algorithms are proposed to recognized plant leaf areas. Mostly five steps are performed in any pattern recognition system namely pre-processing, segmentation, feature extraction, classification and recognition. Based on this general model of pattern recognition system, for this research work Plant Foliar Disease Identification Model (PFDIM) is proposed. Numerous components and sub-components of this model are described in detail in this chapter. Leaf Diseases Recognition System Engine (LDRSE) is a core component of this model which accepts image of leaf. Various processing steps required are describe in subsections of this chapter.

3.1.1. Selection of crop

For this proposed study researcher has selected mung bean plant of legume category of crop. Figure 3.1 shows classification of various crop categories.

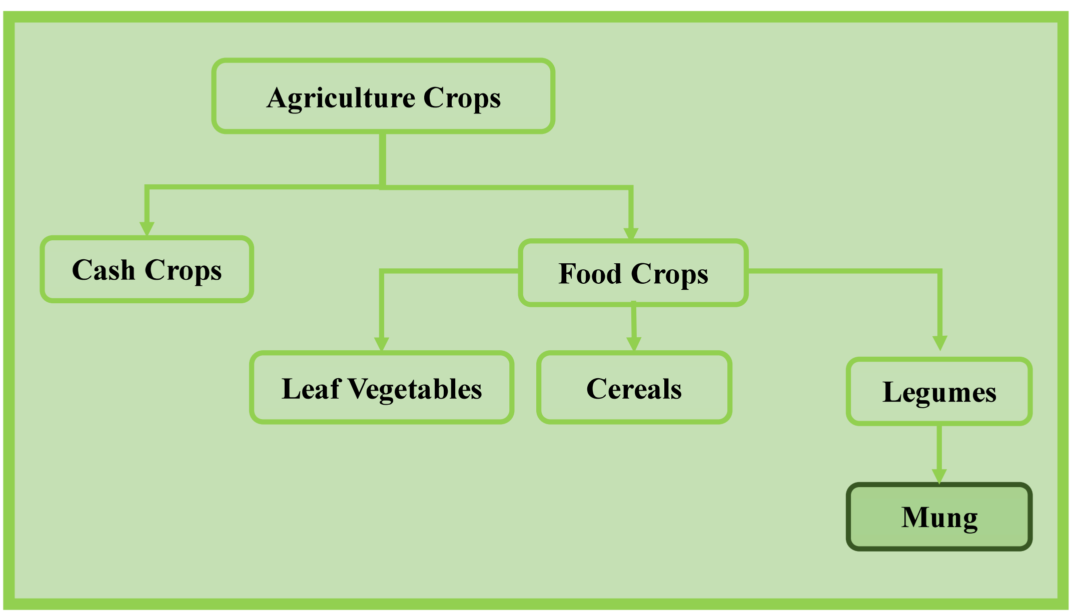

Figure 3.1: Classification of Crop

3.1.2. Plant Organ

Plants have special organs like roots, stem, leaves, veins, and reproductive organs like a flower. Figure 3.2 represents various important visible organs of mung bean plant.

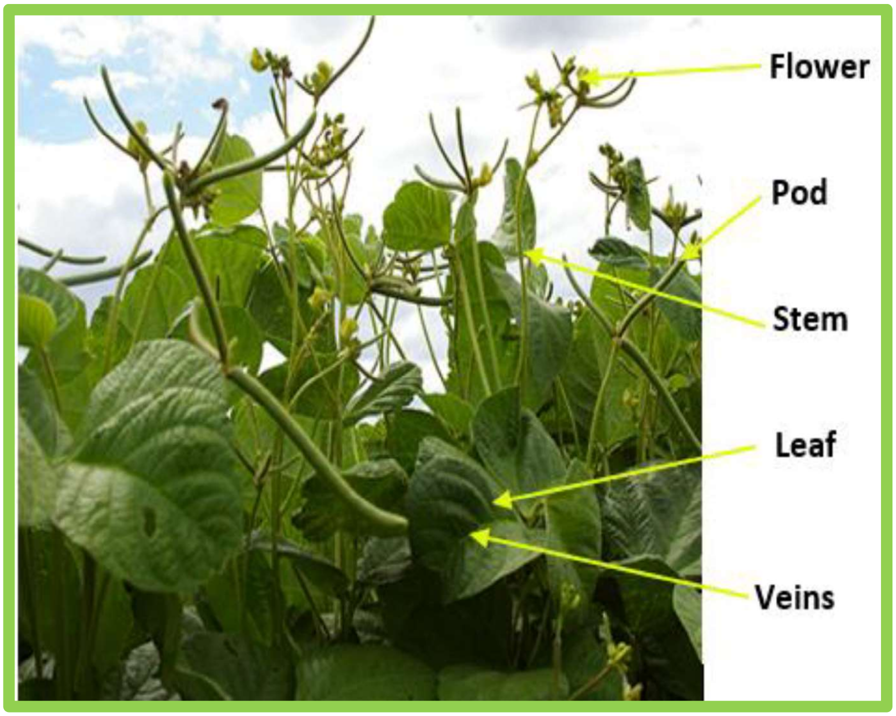

Figure 3.2: Plant Organs

Apart from all this organs researcher has selected leaf organ of a mung bean plant for research study.

3.2 PLANT FOLIAR DISEASES IDENTIFICATION MODEL - (PFDIM)

PFDIM is a proposed model which is intended to do the following tasks:

- Collecting image data.
- Separating Leaf image file into group of images
- Pre-processing image dataset to making it suitable for feature extraction.

- Classification.
- Display leaf result image with disease name or healthy category.

Here Figure 3.3 shows a diagram of the proposed model for Leaf diseases identification system.

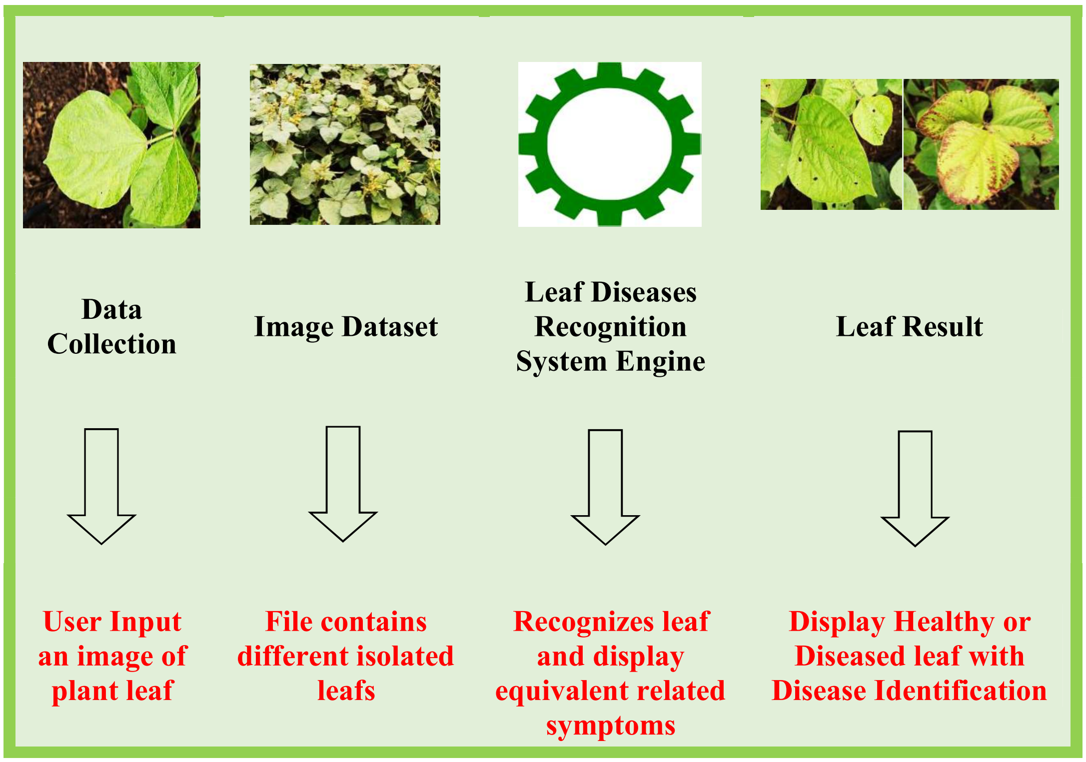

Figure 3.3: Plant Foliar Diseases Identification Model - (PFDIM)

The PFDIM includes four steps process to obtain the desired outcome. They are

(a) Data Collection (Collecting Leaf images)
(b) Image data set
(c) Leaf Diseases Recognition System Engine
(d) Leaf Result

Leaf Diseases Recognition System Engine is a core part of entire Plant Foliar Diseases Identification Model - (PFDIM) which does task of isolating, pre-processing, feature extraction and classification to obtain desired outcome.

(a) Data Collection: Many datasets of various plants are readily available which includes PlantVillage (PlantVillage, n.d.), Flavia (Flvia, n.d.), Swedish (Swedish, n.d.), New Plant Disease (New Plant Disease, n.d.) and Coffee-dataset (Coffee-dataset) etc. For this study a custom dataset formed as there is no standard dataset available for Mung bean plant leaf organ.

The critical need for precise plant disease identification is a standard dataset of plant organ images. The dataset creation consists of stages as follows:

- Plant Selection
- Capturing Images
- Dataset Creation.

For this research, the Mung bean plant is under consideration as it is a local crop of the South Gujarat Region. In the present study, the leaf dataset consists of three types of diseased and a healthy Mung bean leaf images; these are Cercospora Leaf Spot, Yellow Mosaic Virus, and Powdery Mildew. Images were captured from The Navsari Agriculture University at Navsari, Gujarat, India and various farm fields of nearer villages for reflective study.

(b) Image Data Set: Image data set is where the leaf with different symptoms were stored. It includes steps like capturing the image, store that image into storage medium for further processing.

The dataset mainly accounts for two different environments: controlled environment and uncontrolled environment. An image from the controlled environment contains a single mung leaf at its center and a white background i.e. no noise. On the other hand, an image from the uncontrolled environment, along with the mung leaf contains background noise like ground, stems, other mung leaves etc. Leaves were digitally captured in a controlled and uncontrolled environmenst using Oppo A5 13MP and MI Note 8 Pro 64MP smartphones. Figure 3.4 shows the sample images from each category in controlled and uncontrolled environment.

Controlled Environment **Uncontrolled Environment**

(a) Cercospora Leaf Spot

(b) Healthy

(c) Powdery Mildew

(d) Yellow Mosaic Virus

Figure 3.4: Images of the healthy and diseased leaves in (a) controlled and (b) uncontrolled environment

The Dataset consists of 1307 images which include 367 healthy and 940 diseased leaves. The diseases considered are Cercospora Leaf Spot, Powdery Mildew, and Yellow Mosaic Virus. Table 3.1 represents the list of numbers of images category wise.

Environment	Cercospora	Healthy	Powdery Mildew	Yellow Mosaic
Controlled	224	211	225	223
Uncontrolled	102	156	41	125

Table 3.1: Category wise number of images

Accuracy of application is highly dependent on dataset.

(c) Leaf Diseases Recognition System Engine

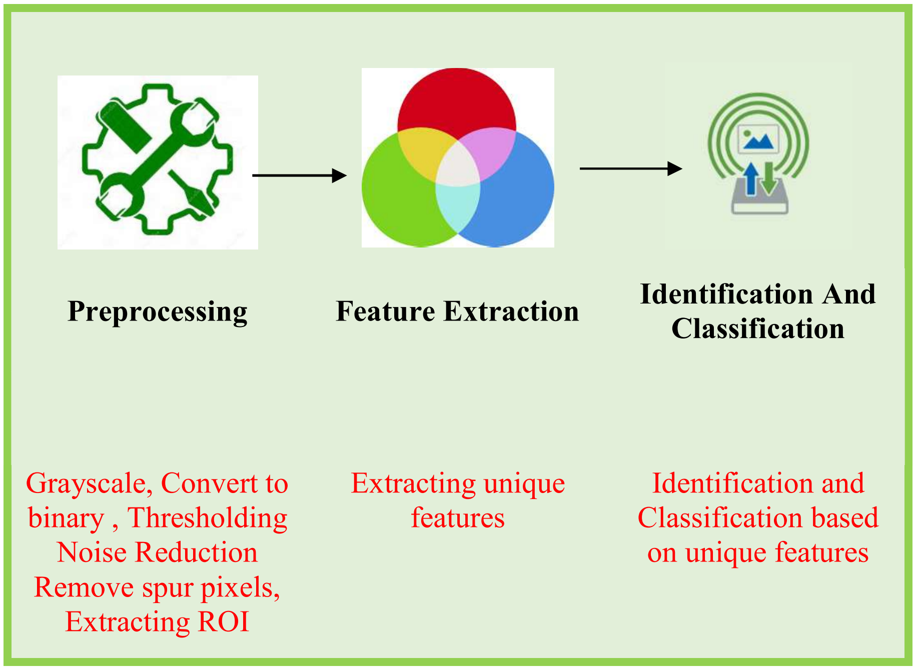

Figure 3.5: Leaf Diseases Recognition System Engine

Leaf Diseases Recognition System Engine (LDRSE) consist of total three stages

(i) Pre processing

(ii) Feature Extraction

(iii) Identification and Classification

(i) Pre-processing: After image acquisition the pre-processing phase takes place. In this phase, image enhancement is done. For this various operations are carried out in a series: RGB image Acquisition and color transformation, normalization/ resize of image size, masking green pixels, Segmentation, and morphological operations, and extracting ROI. This phase makes changes in the image and makes it appropriate for segmentation.

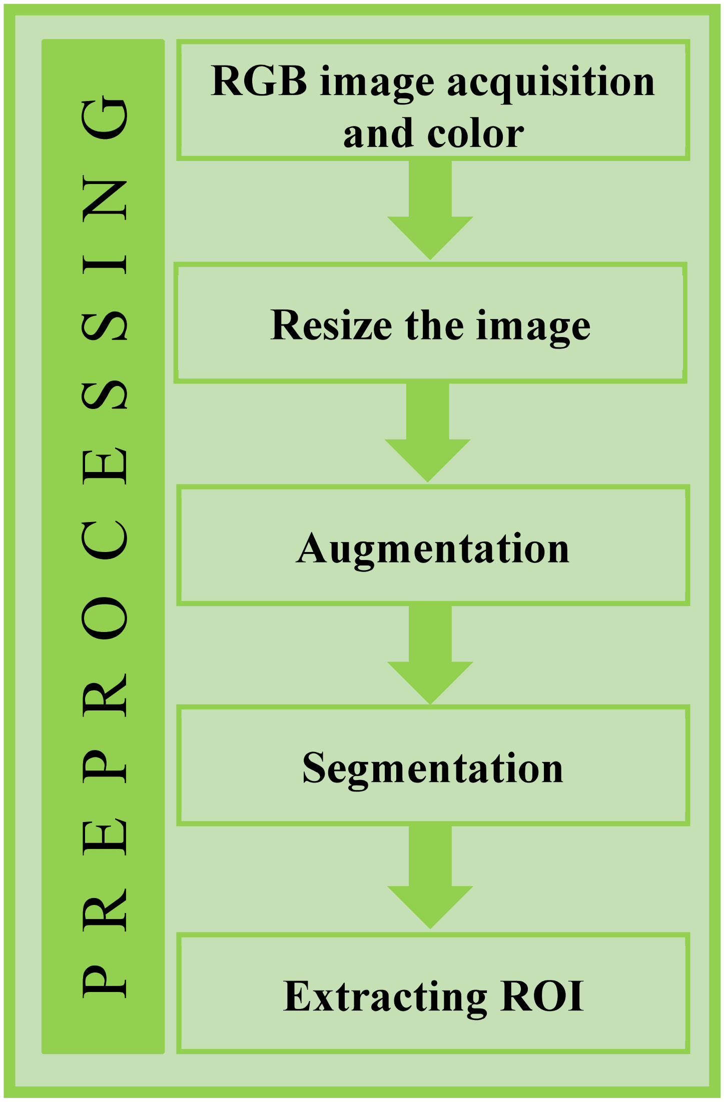

Figure 3.6: Stages involved in Pre-Processing

a) Image Acquisition & color transformation

The RGB image is taken as input and convert the image into color space representation. Color space facilitates the color specification in a standard accepted way.

b) Resize the image

Resizing the image is part of pre-processing. We can dynamically resize and form an image when it is inputted.

c) Augmentation

In this step researcher is trying to create synthetic data and try to increase training dataset size artificially by creating altered forms of images in the dataset. Image data augmentation is used to increase the training dataset in order to increase the performance and capability of the model to simplify. One can need it when training data is complex and having a very less samples. It artificially forms training images through combination of multiple changes in small amount. Transformation consist of methods like: Rotations, Translations / Shifts, Shearing, Variations in scale, Horizontal/Vertical flips.

d) Segmentation

During the segmentation phase image is divided into several segments so that the analysis process becomes easy. Segmentation is used for locating objects in the image and to detect bounding lines of the image, background subtraction.

Various techniques are available for performing segmentation like region-based, edge-based, threshold-based, model-based, and feature-based clustering, etc.
Image segmentation is the first step in image analysis and pattern recognition it is a critical and essential step and is one of the most difficult tasks in image processing, as it determines the quality of the final result of the analysis (Jagtap et al., 2014).

This phase gives an accurate classification of diseases.

e) Extracting ROI

After segmentation Region Of Interest (ROI) is extracted. In the present work ROI is the diseased part. Next significant features are extracted and used to decide the significance of the given sample.

(iii) Feature Extraction:

The purpose of feature extraction using connected components reduced the image data by measuring certain properties of each segmented region (Adam et al., n.d.). Feature extraction aims to describe the pattern by means of a minimum number of features that are effective in discriminating pattern classes (Singh, 2012). Once the image is pre-processed it is given as input for feature extraction, to extract structural features Feature extraction is the process of describing or collecting important image characteristics for advanced analysis and classification. In this phase, the unique characteristics of objects or groups of objects are collected. Based on this information classification is done and to do this from each and every class a set of characteristics is extracted that will distinguish it from other classes.

(iv) Classification and Recognition:

Classification is the phase where training and testing take place. It is where the decision takes place using features extracted from the previous phase. Classification defines classifying an image to some group which is having unique characteristics. After classifying an image in a particular group based on some local features extracted of symptoms is identified to the one particular symptom.

(d) Leaf Result: Final outcome of the Plant Foliar Diseases Identification Model - (PFDIM) is to recognize the Leaf is healthy or diseased. After recognition of the disease leaf, the next step is to find out what type of disease is on the leaf.

Following branching pattern shown in figure 3.7, used for classification of images to identify specific disease.

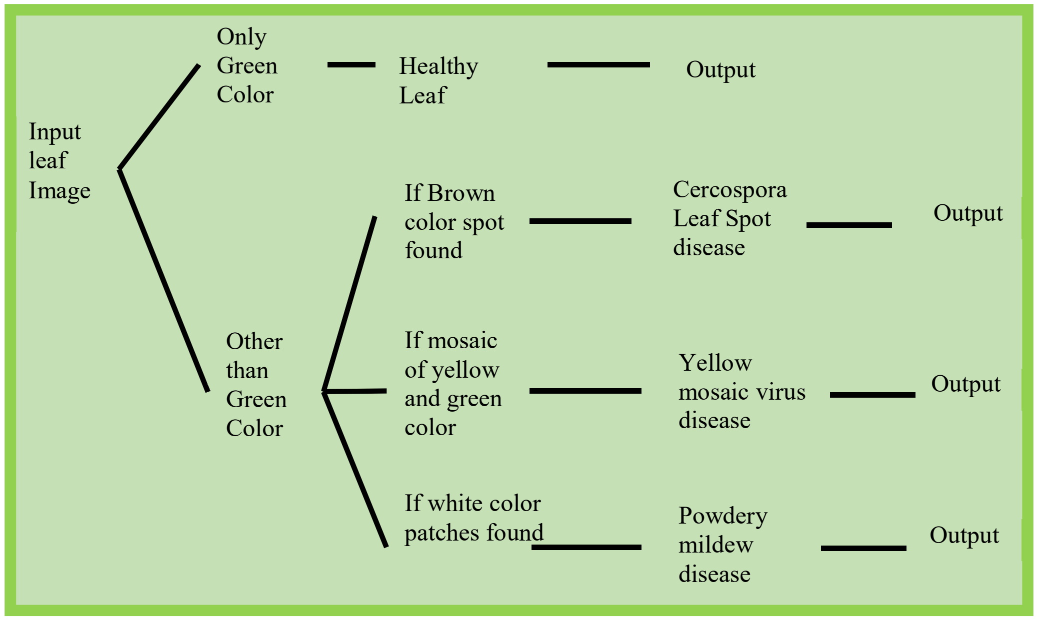

Figure 3.7: Branching Pattern for analysis of images

3.3 Hypothesis

H1: The study hypothesizes that the SVM model may perform better than KNN, AdaBoost, DTC, GaussianNB and LogisticRegression.

H2: The study hypothesizes that the Custom CNN (Custom Convolutional Neural Network) may perform superior than SVM (Support Vector Machine), KNN (K-Nearest Neighbors), Decision Tree Classifier (DTC), AdaBoost (Adaptive Boosting), GaussianNB (Gaussian Naive Bayes), and Logistic Regression on the chosen dataset.

H3: Controlled environment may have better accuracy compared to uncontrolled environment.

References

[1] Adam Sparks, N. P. Castilla, and CM Vera Cruz, “Brown spot and Blast”, Rice Knowledge Bank, International Rice research institute, IRRI

[2] Coffee-dataset, https://drive.google.com/open?id=15YHebAGrx1Vhv8-naave-R5o3Uo70jsm

[3] Flavia, Flavia, A Leaf Recognition Algorithm for Plant Classification using PNN (sourceforge.net)

[4] M. S. M. H. Mr. Sachin B. Jagtap, "Agricultural Plant Leaf Disease Detection and Diagnosis Using Image processing Based on Morphological Feature Extraction," IOSR Journal of VLSI and Signal Processing (IOSR-JVSP), vol. 4, no. 5, sep-2014.

[5] New Plant Disease, https://www.kaggle.com/vipoooool/new-plant-diseases-dataset

[6] PlantVillage, https://plantvillage.psu.edu/.

[7] S. Singh, "Use of Gabor Filters for Recognition of Handwritten Gurmukhi Character," International Journal of Advanced Research in Computer Science and Software Engineering, vol. II, no. 5, p. 234, May 2012.

[8] Swedish, Swedish Leaf Dataset (liu.se)