Unsupervised learning approach for Plant Leaf Disease Detection

Manpreet Kaur[1] and Sanjay Singla[2]

Abstract- Disease occurred in plants and leaves reduce the adequacy and magnitude of crops productions. The identification of disease prevents damage of crops during growth, harvest and post-harvest. The disease is diagnosed in direct and indirect methods. In the direct methods the diseases caused by micro-organisms such as bacteria, fungi and viruses are directly recognised to provide exact identification of disease. In case of indirect method, the different features such as morphological changes, temperature change and organic compounds are identified for plant disease identification. Some of the major problems with traditional process were the constant observation of specialists which becomes costly in case of large farms. Moreover miles have to be travelled by farmers for contacting the experts with expensive cost. Automatic detection of plant disease covers large area of crops for damage and disease identification. The monitoring of crops can be done properly keeping the constraints for changes in temperature and organic compounds. The methodology makes the identification of the disease by extracting the important features for categorizing different disease occurred in the plant leaf.

I. Introduction

Plants have become an important resource. The diseases caused in plants have affected the plants in both quantity and quality. The growth of the crops requires continuous monitoring for finding the disease and its management through the experts. There are several diseases that affect the plants devastating economic, social and ecological losses. Among different kinds of diseases, some plants have disease symptoms that are not visible, whereas for other cases the signs can be detected with electromagnetic spectrum. Reduction in agricultural fertility can be due to a many factors among which damage caused by pests and micro-organisams plays a powerful role in crop losses. The failure in crop yield ranges between 20% and 40% due to pathogen infections. The advanced disease detection and prevention in crops is necessary in order to shrink the diseaseinduceddamage in crops during growth, harvest and post harvest processing, as well as to increase productivity andensure agricultural sustainability. Detection and identification of disease can be done using the two direct and indirect methods. The direct method identifies the disease with molecular and serological methods for high throughput analysis when wide number of samples are analysed. This method includes identification of the disease caused by micro-organisms such as bacteria, fungi and viruses. Whereas the parameters such as morphological change, transpiration rate change, climate change, and volatile organic compounds released by infected plants are recognised by the indirect method. Automatic detection of plant disease covers large area of crops for damage and disease identification. The monitoring of crops can be done properly keeping the constraints for changes in temperature and

[1] *Department of Computer Science Engineering IET Bhaddal, Ropar, Punjab, India*
[2] *Department of Computer Science Engineering IET Bhaddal, Ropar, Punjab, India*

organic compounds. The methodology makes the identification of the disease by extracting the important features for categorizing different disease occurred in the plant leaf.

Disease detection symptoms:

- **Underdevelopment of tissues or organs:** Symptoms for the underdevelopment of the plants are faulty root development, leaf abnormality, inadequate chlorophyll production that leads to improper development of flowers and fruits.
- **Overdevelopment of tissues or organs:** Some of the overdevelopmentof tissues include galls on roots, stems and lavish flowers.
- **Death of plant parts:** These symptoms affect the whole plant having shoot or leaf blights and fruits rots.
- **Variation of normal appearance:** In this examples include mosaic patterns of light and dark green color on leaves, and varied coloration in leaves and flowers.

Plants and their Diseases:

Powdery mildew– In this alteration of stems, flowers, leaves, and fruits takes place with a white to gray color surface mycelia layerwhich can be rubbed off.

Bacterial leaf spots– These spots are often angular. Color is usually consistent and plant pathogens are apparent. Tissue may initially apperas being water-logged but may become papery as it dries.

Mosaic and Ringspot- A random patchwork of green and yellow color over the surface of a leaf describe the mosaic and ringspot. Sometimesleaves may also become ill-shaped.

_____

Leaf Deformation– The normal shape and size of the leaves of the infected plant may be malformed. Leaves may be elongated, smaller size, or thickened	Fungous leaf spots– These spots are round and differ in size and occasionally prolongate on stems. Spots are not restricted by leaf veins.	Presence of Spores– Fromthe presence of spore on the leaf surface various fungal diseases can be simply described.

Figure: Plants and their diseases

Related work:

Arivazhagan S., Shebiah R. Newlin, Ananthi S., Varthini S.[1]proposed method for automatic classification and discovery of plant leaf diseases. The process initiates with converting the RGB image of the plant leaves into HIS color space representation for facilitating color specifications. H component is considered for the analysis. In the next step the masking of the green pixels are done with the thresholding value. After masking the infected portion of the leaf are segmented into equal size patches. Patch having greater than fifty percent of information among the segmented patches are considered for further analysis. The color co-occurrence is done with the Spatial Gray-level Dependence Matrix (SGDM). SGDM matrices measure the probability of pixel at one appropriate gray level will occur at adistinct distance and orientation from any pixel given thatpixel has a second appropriate gray level. The texture features from the H component is computed for classification. The classification is one with minimum distance criterion and support vector machine.

William John Orillo, Dela Jennifer, Leobelle, Jensen Sa Paul, Ira Valenzula [2]proposed the method that performs the image acquisition at initial step. The image contrast enhancement and noise removal is done with the HSV transformation. For adjusting the image intensity the Otsu's thresholding is done from gray level histogram. The image segmentation is done with thresholding to convert binary level image. The image is converted into histogram equalization using bins. The converted image performs masking to remove disease part of leaves. Different features are extracted based on fraction of leaf covered by disease, arithmetic mean of RGB color components, and standard deviation of RGB components and mean of HSV component. All the four components are analyzed with back propagation ANN.

Patil Sanjay B., Duhane T.M. [3] proposed a method that works in two stages: Feature Extraction and Feature Matching.In feature extraction the feature vectors fromimages are extracted and stored in feature database. If the distance between the featurevector of the query image and database image is smallenough compared with predefined threshold, then thecorresponding image in the database is considered as amatch to the query image.Asacolor feature mean value of hue histogram of the leaf isextracted. The shape features are extracted in the form ofmatching key points of the leaf image. These key points areextracted using SIFT algorithm. The texture features areextracted using our newly proposed LGGP (Local GrayGabber Pattern) algorithm. The mean value of thehistogram of the LGGP image is used as texture feature. Euclidean distance is performed for the feature retrieval from the database. The segmentation method is done for the identification of the disease, this is done with thresholding and region growth that is done with the K-means clustering that clusters all contains of green portion (Healthy) of the leaf into one cluster and another cluster contains diseased portion of the leaf.

Omrani Elham, Khoshnevisan Benyamin, Shamshirband Shahaboddin, Saboohi c Hadi, Anuar Nor Badrul , Md Nasir Mohd Hairul Nizam [4] proposed a method that initiates with the removal of noise from the image and transformation of RGB to L*a*b* color space transformation. To extract the affected region, K-means clustering is done. From the extracted infectious area of the leaf, color, shape and texture features are obtained for classification. Texture features are obtained with the wavelet and GLCM matrix. For the classification the ANN and SVM are applied for detecting the disease of the plant leaf.

SannakkiSanjeev S, Vijay S Rajpurohit Vijay S, Nargund V B, Kumar Arun R, YallurPrema S [5] proposed a method that discusses different drawbacks of the existing methods for finding the leaf diseases with accuracy percent. The proposed method is initiated with the removal of noise with the Gaussian filter. The K-Means clustering technique is performed for computing the segmentation work. The result of the segmentation the diseased regions are extracted as clusters. From the clustered result, the total leaf area and the diseased leaf area are calculated for further computation of the percentage infected. A triangular fuzzy membership function is defined for grading the disease and find the percentage of disease occurred.

Problem Definition and objectives
Problem Definition:
- Apart from the number of methods applied for the plant disease identification there are some methods which are too much specific. However, in some papers the proposed methods can deal with only one kind of plant species. The results for validation are not applied for fining the accuracy.
- Another important point is the lack of technical knowledge about the methods. The problems having simple approaches is always favourable but the complex solutions give better results and require high technical knowledge. However some problems for over fitting and under fitting remains in the sample set having low representations and bias value.
- The new methods have to be developed that would perform the identification of plant disease addressing the different types of plant diseases. Moreover, the image processing techniques needs to be applied in various dimensionality in order to find better results.

Objectives:
- To study the existing methods on plant disease detection.
- To design and implement a proposed methodology to proper identification of the plant disease.
- Analysing the result of proposed methodology on different types of plant species.
- Performance evaluation of proposed algorithm.

Proposed system:
- **Data Collection:** Firstly we need to collect the data. So the foremost step is the collection of data which is very necessary step.
- **Pre-Processing***:*Preprocessing is done using the median filter.
- **Color space HIS transformation:**Thecolor space transformation is performed with HIS transformation that will have H component for analysis as it has more dominance to the user observance.
- **Segmentation using FCM (Fuzzy C Means):**Segmentation divides the image into meaningful parts for improved analysis and understanding of the image. FCM is used as segmentation method as it makes a decision for clustering the image pixel value depending on its degree of membership obtained from its belongingness to the diseased on non-diseased part of the plant leaf.
- **Feature Extraction:** Feature extraction is done from the segmented region obtained after performing FCM. The different mean and variance of the pixels are considered for finding the features of the segmented portion.
- **Classification***:* Based on the features obtained, the Support Vector Machines (SVM) execute the classification for a given training set.

The step-by-step methodology is to be followed for plant leaf disease detection.Following diagram represent step by step procedure of implementation steps.
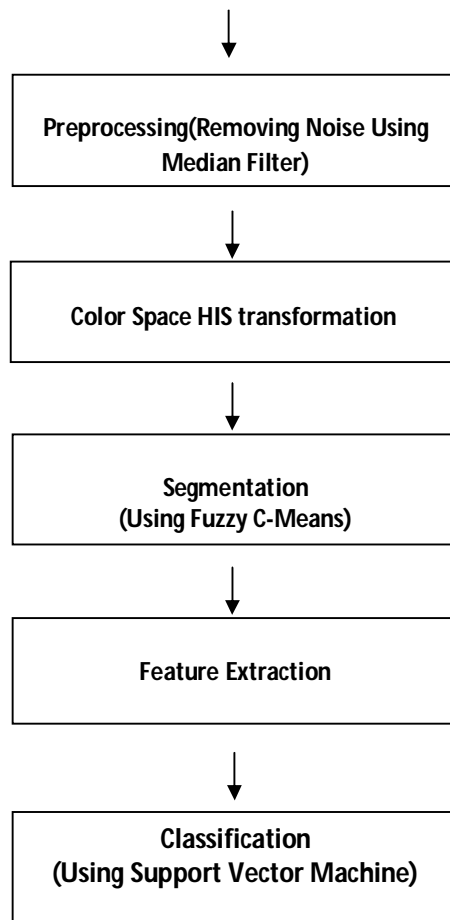
Data Collection

_____

```
                              ↓
        ┌─────────────────────────────────────┐
        │  Preprocessing(Removing Noise Using  │
        │            Median Filter)            │
        └─────────────────────────────────────┘
                              ↓
        ┌─────────────────────────────────────┐
        │    Color Space HIS transformation    │
        └─────────────────────────────────────┘
                              ↓
        ┌─────────────────────────────────────┐
        │            Segmentation              │
        │        (Using Fuzzy C-Means)         │
        └─────────────────────────────────────┘
                              ↓
        ┌─────────────────────────────────────┐
        │         Feature Extraction           │
        └─────────────────────────────────────┘
                              ↓
        ┌─────────────────────────────────────┐
        │            Classification            │
        │     (Using Support Vector Machine)   │
        └─────────────────────────────────────┘
```

Figure: Block Diagram

Results and Discussions

The experimental analysis is done on six different types of plant leaf disease images. A training data set of 6 images is considered for performing the experiment on given test image. The initial step performs the pre-processing on the input diseased leaf image. The input image is read from the console and filtered and then transformation of the input RGB image is done to the HIS color space. The median filter is performed for filtering. The median filter is a nonlinear filtering technique that removes the noise keeping the edges preserved. The result of the filtering is taken as input for the HSI transformation. The HSI component has is decoupled into chrominance (H and S component) and luminance (I Component). The HSI color space gives color description and is most informative component and is better for analysis as it has more dominance to the user observance. HIS transformation is then followed by calculating membership based Fuzzy C-Means algorithm performs the segmentation of the diseased regions resulting in better feature extraction. The feature values for the diseased images are computed with the mean and variance that identifies the plant disease properly.. The FCM algorithm optimizes the objective function resulting in better identification of the plant disease.
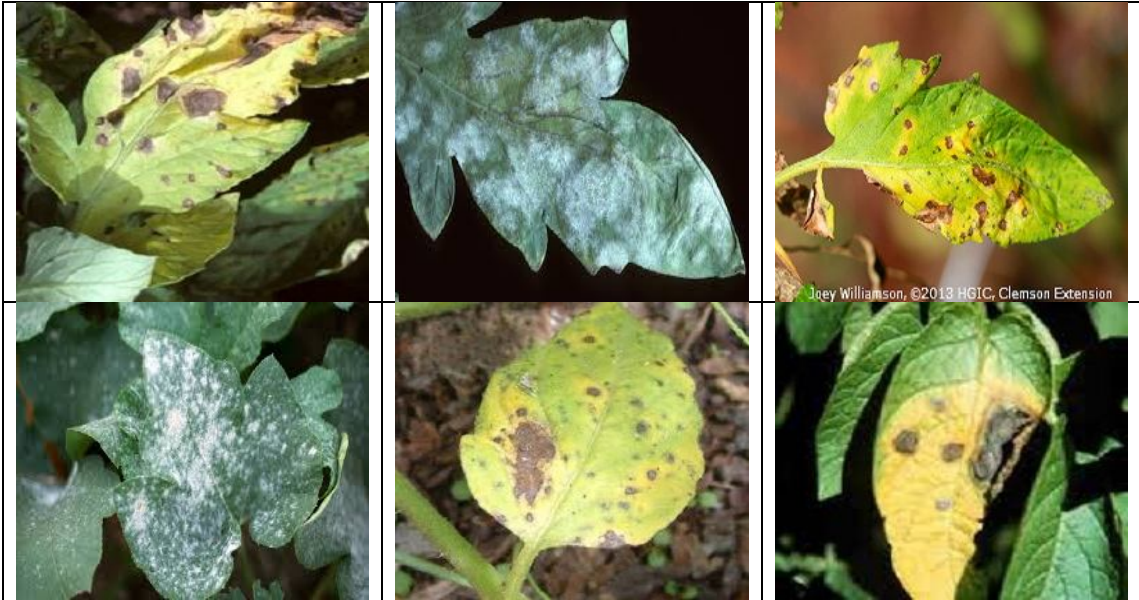
Training Images

Figure: Training Images

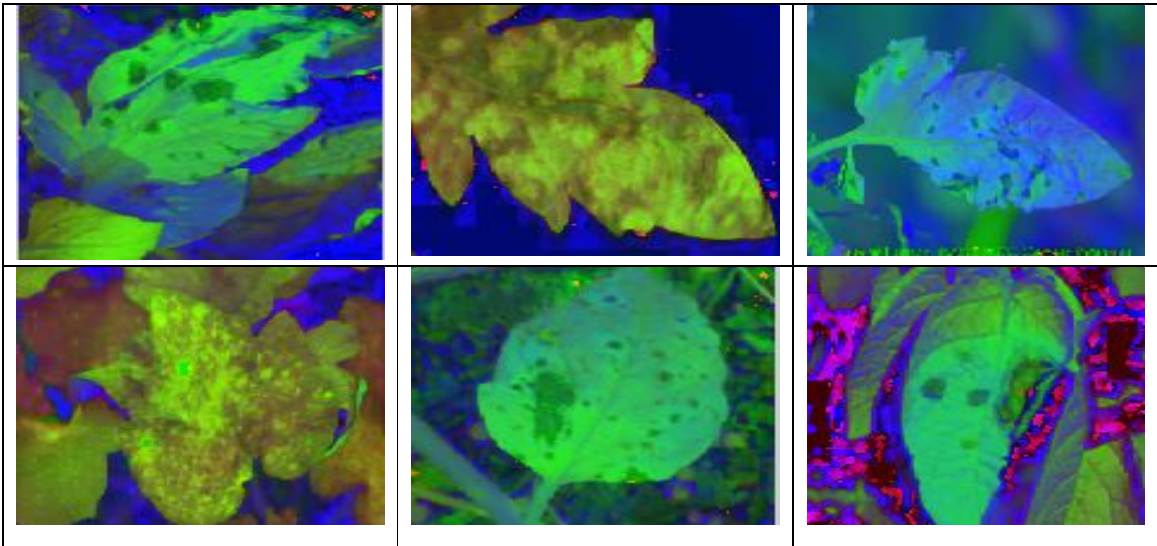**Test Images**

Figure: Test Images

HIS Images

Figure: HIS Images

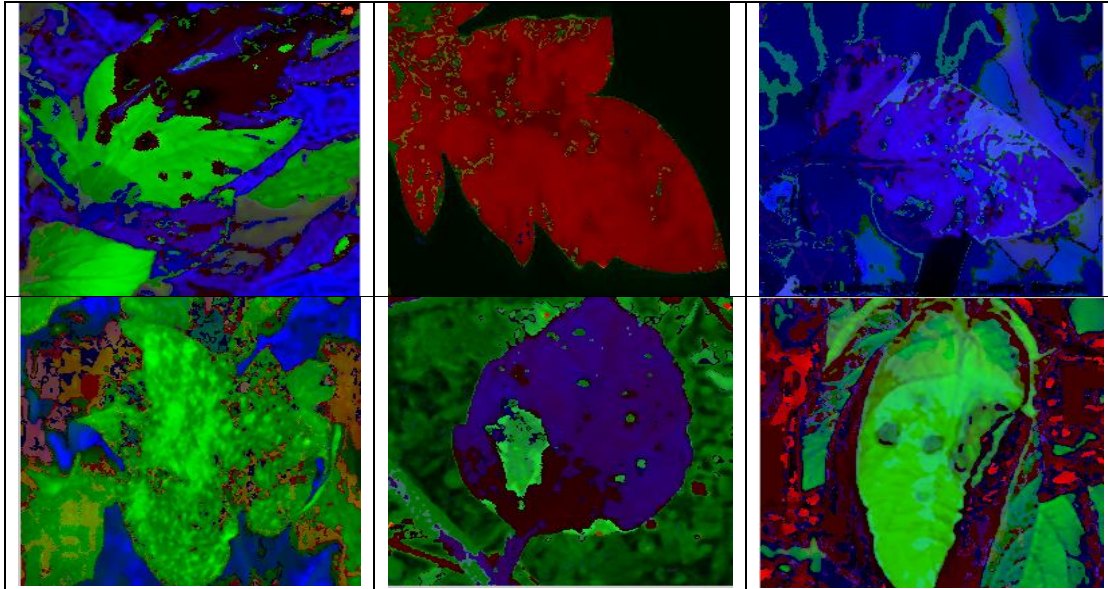**Segmented Images**

_____
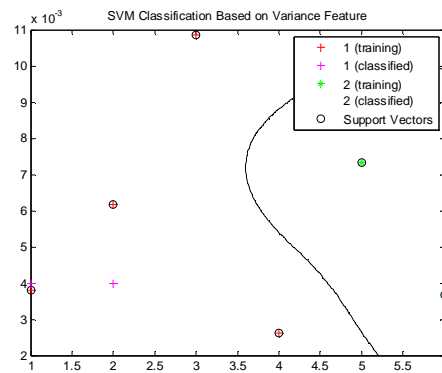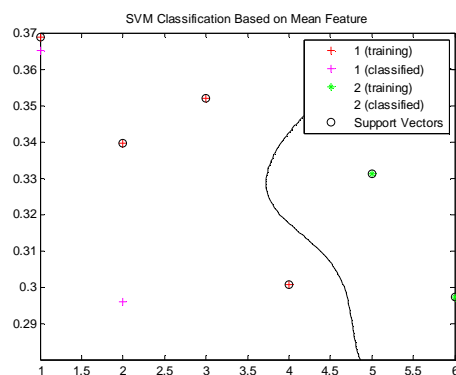

Figure: Segmented Images

Figure : SVM classification based on mean feature   **Figure: SVM classification based on variance feature**

Conclusion and Future scope

This work is the study of plant leaf diseases and various techniques for the finding the disease region. The various image processing techniques are deployed for finding in less computation time. The best techniques will be evaluated who have the maximum benefits from all these techniques..Inspite of the maximum benefits every technique has certain disadvantages. Automatic detection of plant disease covers large area of crops for damage and disease identification. The monitoring of crops can be done properly keeping the constraints for changes in temperature and organic compounds. The proposed methodology has been implemented successfully. The performance for the same is tested on a real set of leaf data. The result is quite convincing and immense adaptability in developing countries, where such information plays an essential role for enrichment in yield. The methodology makes the identification of the disease by extracting the important features for categorizing different disease occurred in the plant leaf. The proposed method is application of FCM with SVM classification for detection of plant leaf. Automatic detection and classification of plant leaf disease can be significantly supported by our experimental results. The proposed system provides 97.50% accuracy. The future scope of this work is to optimize the function with Evolutionary Computation.

References

[1]   Arivazhagan S, Shebiah R Newlin, Ananthi S, Varthini S Vishnu, 2013 "Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features".

[2]   Willam John Orillo, Dela Jennifer Dela, Leobelle, Jensen Sa Paul, Ira Valenzula, 2014 "Identification of Diseases in Rice Plant (OryzaSativa) using Back Propagation Artificial Neural Network

[3]   Patil Sanjay B., Dudhane T.M., 2015 "Leaf disease severity Detection using K-means clustering and CBIR".

[4]   OmraniElham, Khoshnevisan Benyamin, ShamshirbandShahaboddin, Saboohi c Hadi, Anuar Nor Badrul , Md Nasir MohdHairulNizam, 2014 "Potential of radial basis function-based support vector regression for apple disease detection".

[5]   SannakkiSanjeev S, Vijay S Rajpurohit Vijay S, Nargund V B, Kumar Arun R, YallurPrema S, 2011 "Leaf Disease Grading by Machine Vision and Fuzzy Logic",

[6]   Hiary H. Al, Ahmad S. Bani, Reyalat M., Braik M. and ALRahamneh Z., 2011 "Fast and Accurate Detection and Classification of Plant Diseases

[7]   Patil   Sanjay   B.,   BodheShrikant   K.,2011   "Leaf   Disease   Severity Measurement Using Image Processing".

[8]   Gurjar Ajay A., GulhaneViraj A.,2015 "Disease Detection On Cotton Leaves by Eigenfeature Regularization and Extraction Technique".