TEXTURE BASED FEATURES APPROACH FOR CROP DISEASES CLASSIFICATION AND DIAGNOSIS-A RESEARCH

Namrata Ghatol[1], Dr. G.P. Dhok [2]

[1]ME Student, Dept of Electronics and Telecommunication
[2]Associate Professor, Sipna College of Engineering and Technology, Amravati.

Abstract - *Agriculture is the main element of economic growth in developing countries. Plant diseases cause major economic and production losses as well as curtailment in both quantity and quality of agricultural production. Now a day's, for supervising large field of crops there has been increased demand for plant leaf disease detection system. The critical issue here is to monitor the health of the plants and detection of the respective diseases. In this project, Image processing is used to detect and classify crop diseases based on the texture features approach. The images are taken through a high resolution digital camera and after preprocessing, these are then run through the various machine learning algorithms and classified based on their color and texture features. In this project basically two machine learning algorithms are used namely K-Nearest Neighbours, Support Vector Machine. The implementation will be done using MATLAB.*

Key Words: *Classification, Textures features approach, Machine learning, K-Nearest Neighbours, Support Vector Machine*

1. INTRODUCTION

Agriculture is an integral part of the economy of a country. Especially in developing countries like India.. A large number of factors are responsible for the contributions by the agriculture sector to be this low like low literacy rates among farmers, bad quality seed and availability of resources like water. One of the most critical factors is the diseases that the crops contract. High quality crop production is the big challenge for farmers. . The crop production rates are directly proportional to the each day plant growing progress. So plant disease detection is very important. Also the external appearance of agricultural products is the main quality attribute. The outer appearance greatly affects their scale value and customer's buying behaviour.

Therefore, disease diagnosis and correct treatment essential for the healthy crop production process as early as possible. The farmer's wrong diagnosis of crop disease causes insecticides spray inappropriately. Various image processing techniques can be significantly applied to observe the crop growth progress and disease diagnosis.

In this project, the diseases are classified using an image of the leaf taken by a high resolution camera. As symptoms in most cases are noticed on the leaves, Color and texture features are extracted from the image and passed through the machine learning algorithm for classification. Machine learning-based detection and recognition of plant diseases can provide extensive clues to identify and treat the diseases in its very early stages. Comparatively, visually or naked eye identification of plant diseases is quite expensive, inefficient, inaccurate and difficult. Automatic detection of plant diseases is very important to research topic as it may prove the benefits in monitoring large fields of crops, and thus automatically detect the symptoms of diseases as soon as they appear on plant leaves. This project focuses on diseases detection and Classification of crop species like Chilli, Soyabean, cotton, corn(maize), orange, mango, sunflower, peanut etc. using image processing techniques. These plants largely produced crops in India. Improving the productivity of these crops can significantly reduce the food deficiency and can contribute towards improvement in health care. Hence, these crops are taken as the crops of interest.

2. LITERATURE REVIEW

S. Arivazhagan et al [1], proposed an approach where, first conversion of an image from RGB to HSI is done and the green pixels are masked using threshold values. The proposed system is a software solution for detection of plant diseases. The developed processing scheme consists of four main steps, first a color transformation structure for the input RGB image is created, then the green pixels are masked and removed using specific threshold value followed by segmentation process, the texture statistics are computed for the useful segments, finally the extracted features are passed through the classifier .Texture analysis is then done using color co-occurrence matrix (SGDM). The image is then classified using either minimum distance criterion or SVM classifier which has 86.77% and 94.74% accuracy respectively.

Loyce Selwyn pinto, proposed an approach [2] In which, Image processing is used to detect and classify sunflower crop diseases based on the image of their leaf. The images are taken through a high resolution digital camera and after preprocessing, are subjected to k-means clustering to get the diseased part of the leaf. These are then run through the various machine learning algorithms and classified based on their color and texture features. A comparison based on accuracy between various machine learning algorithms is done namely K-Nearest Neighbors, Multi-Class Support Vector Machine, Naive Bayes and Multinomial Logistic Regression to achieve maximum accuracy. The proposed methodology is able to classify diseases of the sunflower crop in a very accurate and efficient way.

Manisha A. Bhange et al [3], proposed three methods for extracting features, histogram for color, erosion concept morphology for obtaining boundaries of the images and color coherence vector to classify pixels. This approach majorly consists phases namely image preprocessing, feature extraction, clustering, training and classification. color coherence vector features and color morphology are used for feature extraction. K-means clustering is used for segmentation and SVM is used for classification of the images the accuracy is 81%. The disease considered here is the bacterial blight of the pomegranate leaf.

It is difficult to determine the accurate disease in noisy image. Image should be noise free for processing. Therefore, noise reduction techniques and image enhancement are required for desirable processing. Valliammai and Geethaiakshmi [4] have found that the appropriate feature extraction of leaf can be possible if input image is noise free. The leaf vein edges not exactly visible in Gaussian noise method. The speckle noise affected the leaf size, shape and pattern .Therefore, Gaussian and speckle noise removal techniques are essential to restore the noise free leaf images for further process. These Hybrid filter method is developed to eliminate the noise, improve the quality of image and thereby produces better results compared to other traditional filters.

Implementation of RGB and Gray scale images in plant leaves disease detection –comparative study by Padmavathi and Thangadurai [5] have given the comparative results of RGB and Gray scale images in leaf disease finding process. In detecting the infected leaves, color becomes an important feature to find the disease intensity. They have considered Gray scale and RGB images and used median filter for image enhancement and segmentation for extraction of the diseased portion which are used to identify the disease level. The plant disease recognition model, based on leaf image classification, by the use of deep convolution networks have developed. 13 kinds of diseases are identified from the healthy leaves with the capability to differentiate leaves from their surroundings.

Rupesh G. Mundada et al [6] have proposed an approach where images are converted from RGB to Gray scale first, followed by resizing and filtering them. This approach proposes a software prototype system for early pest detection. Images of the infected leaf are captured by a camera and processed using image processing techniques to detect presence of pests. This approach is mainly used to detect whiteflies, aphids on the affected crops at their early stages. Feature extraction with features like contrast and entropy is performed. Classification is done using a Support Vector Machine.

Bindushree H B et al [7], proposed an approach where the processed image is first is segmented using k means clustering .out of the three clusters created one of the clusters contains the disease affected area and image features are extracted from the particular cluster using Gray Level Co-occurrence Matrix (GLCM). These features are later fed into support vector machines (SVM).The final classification results from SVMs indicate whether the leaf in the image dataset is healthy or disease affected. The results using SVM are obtained from various kernels such as linear, polynomial, quadratic, RBF .

P. Revathi, M. Hemalatha [8] worked on classification of diseases in cotton leaves. Authors have considered six types of diseases in the cotton plant for classification. Based on advanced computational techniques the significance of this work design is to reduce the time, cost and complexity. To identify the affected region of a leaf the author has used Enhanced Particle Swarm Optimization (EPSO) for feature selection. For calculating the edge, color, texture variance for feature analysis of the diseased part Skew divergence is used. The result obtained using skew divergence and EPSO technique is 98%.

Hrishikesh P. Kanjalkaret al [9], proposed the approach where, the image is first converted from RGB to HIS. Segmentation is done using connected components labeling, thresholding is used to avoid unwanted regions. 11 features are used in this approach and classification is done using back propagation neural network. The accuracy of this method is 83%.

Dheeb Al Bashish proposed an approach where [10], the images are segmented using K-means clustering and are then converted from RGB to HSI. The color co-occurrence texture analysis method is used using spatial grey level dependence matrices. Features are calculated from H and S components. The neural network used here is a feed forward back-propagation with 93% of overall success

Prof. Sanjay B. Dhaygude et al,[11], have proposed an approach in which firstly by color transformation structure RGB is converted into HSV space because HSV is a good color descriptor. Masking and removing of green pixels with pre-computed threshold level. Then in the next step segmentation is performed using 32x32 patch size and obtained useful segments. These segments are used for texture analysis by color co-occurrence matrix. Finally if texture parameters are compared to texture parameters of normal leaf

3. ARCHITECTURAL SYSTEM

For detection and diagnosis of the various crop diseases by taking into consideration the texture and colour of leaves of respective crops following system is used.

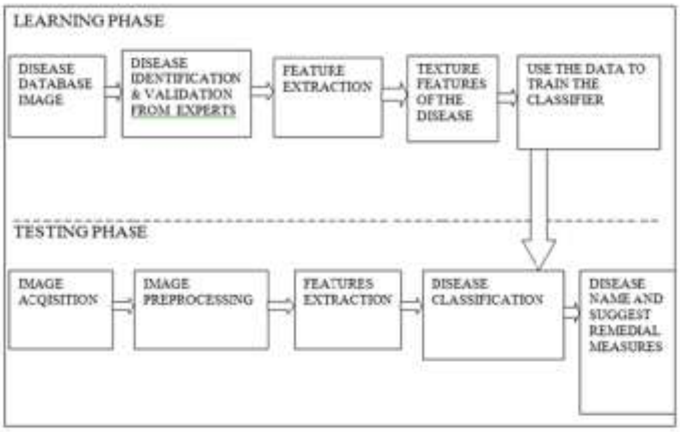

Fig -1: Architectural system

System's architecture basically consists of two phases namely learning phase and testing phase.

LEARNING PHASE-

STEP 1-DISEASE DATABASE IMAGE AND IDENTIFICATION-

In this method, diseased images of plants are captured through the high-resolution camera to create the required training database. This database has different types of images which comprises of healthy, diseased crop images and images are stored in jpeg format. Validation of images are done by experts thus making the training as accurate as possible. To collect this database we are taking help from Dr. punjabrao Deshmukh KrishiVidyapeeth (PDKV),Amravati, Maharashtra.

STEP 2- FEATURE EXTRACTION (TEXTURE BASED)-

Feature extraction is the important part to gracefully predict the infected region. In feature extraction method, various attributes of the segmented image are extracted. Texture oriented features are calculated such as contrast, correlation, energy, homogeneity and mean. The segmented image is then converted into a gray scale image. Statistical texture based features are extracted using Gray Level Co-occurrence Matrix (GLCM). Thus the gray scale image is converted to GLCM matrix so that the following features can be extracted.

STEP 3-CLASSIFICATION (TRAINING)-

In training phase, the classification of diseased leaves images into various categories is done by experts. Then the training dataset is made and stored.

*TESTING PHASE***-**

STEP 1-IMAGE ACQISITION-

The first step is to capture the sample of diseased leaves to be tested from the digital camera. The images are in RGB form.(red, blue, green).Images are taken in controlled environment and are stored in the JPEG format.

STEP 2-IMAGE PREPROCESSING

Image preprocessing is performed on images to highlight the important features of an image and make the image more suitable for use in particular application. The purpose of data preprocessing is to eliminate the noise in the image, so as to adjust the pixel values. It contains various steps like –image enhancement, conversion of RGB image into gray scale image, image resizing, image segmentation etc.

STEP 3-FEATURE EXTRACTION

In this step, various attributes of the segmented image are extracted same as in testing phase above. Texture based features are extracted using Gray Level Co-occurrence Matrix (GLCM). Thus the gray scale image is converted to GLCM matrix so that texture oriented features are calculated such as contrast, correlation, energy, homogeneity and mean.

STEP 4-DISEASE CLASSIFICATION-

In testing processes, For classification features extracted from training leaves are compared with those extracted from testing leaves. The image is then classified based on the matched features.

STEP 5-IDENTIFICATION OF DISEASE NAME AND SUURES-

After the classification the disease name along with remedial measures for treatment based on the classification are displayed to the user.

4. FEATURE EXTRACTION

Transforming the input data into the set of features is called feature extraction. The main goal of feature extraction is to obtain the most relevant information from the original data. Feature extraction is the important part to gracefully predict the infected region. In feature extraction method, various attributes of the segmented image are extracted. Texture oriented features are calculated such as contrast, correlation, energy, homogeneity and mean. The segmented image is then converted into a gray scale image .Statistical texture based features are extracted using Gray Level Co-occurrence Matrix (GLCM).

Thus the gray scale image is converted to GLCM matrix so that the following features can be extracted. Mean, Standard Deviation and Coarseness are taken from the segmented image before conversion.

A. Contrast

It is the measure of the intensity contrast between a pixel and its neighbor over the whole image.

$$\sum_{i,j} |i-j|^2 \, p(i,j)$$

B. Energy

It is the sum of squared elements in the GLCM.

$$\sum_{i,j} p(i,j)^2$$

C. Mean

It is the average value of the elements along different dimensions of an array.

$$\frac{1}{i*j}\sum_{i,j} p(i,j)$$

D. Homogeneity

It is the measure of the closeness of the distribution of elements in the GLCM to the GLCM diagonal.

$$\sum_{i,j}\frac{p(i,j)}{1+|i-j|}$$

E. Standard Deviation

It computes the standard deviation of the values in matrix or array.

$$s=\left(\frac{1}{n}\sum_{i-1}^{n}(x_i-\bar{x})^2\right)^{\frac{1}{2}}$$

Where,

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$$

The goal is to generate features that exhibit high information-packing properties:

- Extract the information from the raw data that is most relevant for discrimination between the classes
- Extract features with low within-class variability and high between class variability
- Discard redundant information.

5. CLASSIFICATION

In training phase, the training images of diseased leaves are classified by experts. After classification the dataset is made and stored. This trained dataset is further used to train the classifier for classification of diseased leaves to be tested.

In testing processes, For classification features extracted from training leaves are compared with those extracted from testing leaves. The image is then classified based on the matched features. The two Classifiers are: K-Nearest Neighbors (KNN) and Support Vector Machine (SVM).These techniques are selected due to the reason that in many real applications these classifiers have performed well and also for the fact that these classifiers differ in complexity and speed.

A. K-Nearest Neighbors-

K-Nearest Neighbors is a simple supervised classifying technique. The K-nearest neighbours algorithm (kNN) is a nonparametric method used for classification. The input consists of the *k* closest training examples in the feature space. In Knn classification, the output is a class membership. 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 (*k* is a positive integer, typically small). If $k = 1$, then the object is simply assigned to the class of that single nearest neighbor.

kNN is a type of instance based learning, or lazy learning. The KNN algorithm is among the simplest of all machine learning algorithms. The training examples are vectors in *a* multidimensional feature space, each with a class label. The training phase of the algorithm consists only of storing the feature vectors and class labels of the training samples. In the classification phase, *k* is a user defined constant, and an unlabeled vector (a query or test point) is classified by assigning the label which is most frequent among the *k* training samples nearest to that query point

B. Support vector machine (SVM)-

Support Vector Machine is a complex classifier as compared to KNN. It was originally developed for linear classification but later modified for multi class classification. The Support vector machine comes in the category of supervised learning .The SVM used for regression and classification. But it is popularly known for classification. It is a very efficient classifier. In this every object or item is represented by a point in the n- dimensional space. The value of each feature is represented by the particular coordinate. Then the items divided into classes by finding hyper-plane as shown in the figure. The diagram shows support Vectors that represent the coordinates of each item.

In machine learning, support vector machines (SVMs) are supervised learning models with associated learning algorithms that analyse data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier (although methods such as Platt scaling exist to use SVM in a probabilistic classification setting). An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall. In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces.

6. RESULTS AND DISCUSSION

The proposed work is implemented on Intel Core processor i5, 4GB RAM Laptop configuration and operating system is windows 7. MATLAB R2016a software is used to write the programming code. In this we used Image processing toolbox and the database of diseased crop images is collected by taking help from Dr. Punjabrao Deshmukh Krishi Vidyapeeth (PDKV),Amravati,Maharashtra.

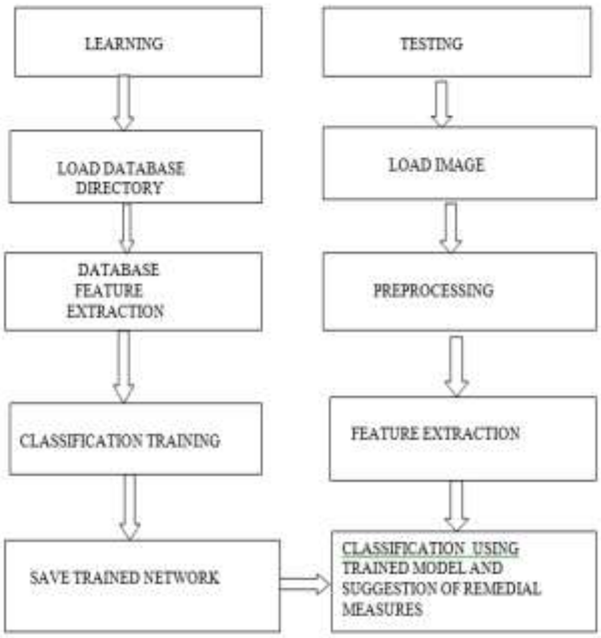

Fig -1: System Flowchart

The feature set consisting of Contrast, Energy, Mean, Homogeneity, Standard Deviation and Coarseness are used and classification is done on the basis of these. The different classifiers are compared against the each other on the basis of Accuracy. Accuracy of the classifiers have been calculated by using the formula. We have taken samples of each type as the testing data. The testing images are taken under natural settings.

Accuracy = (Number of correct predictions/ Total number of samples) *100

Following tables shows the accuracy of the 2 machine learning classifiers and shows the comparison between them for respective crop samples

Table -1: Accuracy of SVM and KNN for cotton and maize

CROP	DISEASE	SVM	KNN
Cotton	Bacterial Blight	93.45	93.58
Cotton	Target Spot	91.23	90.36
Cotton	Powdery Mildew	95.11	95.78
Maize	Gray Leaf Spot	96.37	96.12
Maize	Rust	93.18	92.34
Maize	Downey Mildew	91.46	92.34

Table -2: Accuracy of SVM and KNN for orange and mango

CROP	DISEASE	SVM	KNN
Orange	Citrus Canker	94.32	95.67
Orange	Magnesium Deficiency	98.21	97.76
Orange	Sooty Mold	95.11	95.49
Mango	Algal Leaf Spot	96.37	97.05
Mango	Anthracnose	92.18	92.48
Mango	Sooty Mold	97.46	95.33

Table -3: Accuracy of SVM and KNN for sunflower and peanut

CROP	DISEASE	SVM	KNN
Sunflower	Downey Mildew	95.77	95.11
Sunflower	Powdery Mildew	93.45	92.23
Sunflower	Alternaria Leaf Spot	92.27	93.56
Peanut	Rust	96.37	95.76
Peanut	Cerespora Leaf Spot	92.44	94.89
Peanut	Web Blotch	91.46	93.33

7. CONCLUSION

An application of texture analysis in detecting and classifying the plant leaf diseases has been explained in this project. The image will be taken through high resolution digital camera and in preprocessing stage the system will perform segmentation of leaf and will analyse it with feature extraction algorithms. The machine learning algorithms like K-Nearest Neighbors and Support Vector Machine, Random Forest will be used for feature extraction purpose. Plant diseases will be detected based on their texture.Various parameters like accuracy, precision, sensitivity, error rate will be considered . By this method , the crop disesases will be identified at the initial stage itself and remedial measures will be suggested to solve the respective problem .The accuracy and low cost of the classification allow for an effective automatic surveillance. The accuracy can be increased by using various enhancement techniques and by increasing the size of training dataset.

8. FUTURE SCOPE

In the future, the proposed methodology can be integrated with other to be developed methods for disease identification and classification using color and texture analysis to develop an expert system for early crop foliar disease warning and administration,where the disease type can be identified by color and texture analysis and the severity level estimation can be performed. The performance of the system can be improved in the future by using advanced background separation methods to separate the leaf object from a complex background. The similar methodology can be applied to other plant foliar infections and early warning can be provided.

REFERENCES

[1] S. Arivazhagan, R. NewlinShebiah, S. Ananthi, S. Vishnu Varthini, “Detection of unhealthy region of plant leaves and classification of leaf diseases using texture features”, AgricEngInt: CIGR Journal Vol. 15, No.1, PP 211-217, March, 2013

[2]Loyce Selwyn Pinto, Argha Ray, M.Udhayeswar Reddy, Pavithra Perumal, Aishwarya P. “Crop Disease Detection using Texture Analysis” IEEE International Conference On Recent Trends In Electronics Information Communication Technology, May 20-21, 2016, India

[3] Ms. Manisha A. Bhange, Prof. H. A. Hingoliwala, “Detection of Bacterial Blight on Pomegranate Leaf”,International Journal on Recent and Innovation Trends in Computing and Communication, Volume: 3 Issue: 6, ISSN: 2321-8169 3682 - 3685, PP 3682-3685, June 2015

[4] Valliammai, N.and Geethaiakshmi,S.N.,“Multiple noise reduction using hybrid method for leaf recognition”, 2012 International Conference on Devices, Circuits and Systems (ICDCS), 15-16 March 2012, Coimbatore, India.

[5]Padmavathi,K. and Thangadurai, K.,“Implementation of RGB and grayscale images in plant leaves disease detection – comparative study”, Indian Journal of Science and Technology, Vol. 9,pp. 1-6, 2016.

[6] Rupesh G. Mundada, Dr. V. V. Gohokar, “Detection and Classification of Pests in Greenhouse Using Image Processing”, IOSR Journal of Electronics and Communication Engineering (IOSR-JECE) e-ISSN: 2278-2834, p-ISSN: 2278-8735.Volume 5, Issue 6, PP 57-63,Mar. - Apr. 2013

[7] Bindushree H B, Dr. Sivasankari G G, “Detection of Plant Leaf Disease Using Image Processing Techniques”, International Journal of Technology Enhancements and Emerging Engineering Research, Vol 3, Issue 04, ISSN:2347-4289, PP 125-128

[8] P.Revathi and M.Hemalatha, “Classification of Cotton Diseases Using Cross Information Gain Minimal Resource Allocation Network Classifier with Enhanced Particle Swarm Optimization,” Journal of Theoretical and Applied Information Technology, vol. 60, no. 1, February 2014

[9]Mr. Hrishikesh P. Kanjalkar, Prof. S.S.Lokhande, “Detection and Classification of Plant Leaf Diseases using ANN”, International Journal of Scientific & Engineering Research, Volume 4, Issue 8, ISSN 2229-5518, PP 1777-1780,August-2013

[10]Dheeb Al Bashish, Malik Braik, SuliemanBani-Ahmad, “A Framework for Detection andClassification of Plant Lead and Stem Diseases”, 2010 International Conference on Signal and Image Processing,PP 113-118, IEEE

[11] Prof. Sanjay B. Dhaygude, Mr.NitinP.Kumbhar “Agricultural plant Leaf Disease Detection Using Image Processing”, International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering Vol. 2, Issue 1,PP 599-602, ISSN: 2278 – 8875,, January 2013

[12] Al-Bashish, D., M. Braik, and S. Bani-Ahmad. 2011. Detection and classification of leaf diseases using K-means-based segmentation and neural networks based classification. Information Technology Journal, 10(2): 267-275.

[13] Sachin D. Khirade, A. B. Patil, “Plant Disease Detection Using Image Processing”, 2015 International Conference on Computing Communication Control and Automation, IEE

[14] Harshal Waghmare,Radha Kokare. “Detection and Classification of Diseases of Grape Plant Using Opposite Colour Local Binary Pattern Feature and Machine Learning for Automated Decision Support System” 2016 3rd International Conference on Signal Processing and Integrated Networks (SPIN)

[15]Sumit Nema, Bharat Mishra “Advance App Design Methods of Leaf Disease Detection using Image Processing Approach –A Review”International Journal of Innovative Research in Science,Enginnering and Technology vol.6,Issue 7,July 2017

[16] Gavhale,K. R.andGawande,U., “An Overview of the research on plant leaves disease detection using image processing techniques,” IOSR J. of Computation Engineering (IOSR-JCE), Vol. 16, pp. 10-16, 2014.