Neuro-kNN classification system for detecting fungal disease on vegetable crops using local binary patterns

Jagadeesh D. Pujari, Rajesh Yakkundimath*, Abdulmunaf S. Byadgi

(S.D.M. College of Engineering & Technology, Dharwar – 580 002, INDIA, K.L.E. Institute of Technology, Hubli – 580 030, INDIA, University of Agricultural Sciences, Dharwar – 580005, INDIA)

Abstract: This paper describes the behavior of classifiers for identification and classification of fungal disease symptoms found on vegetable crops. Symptoms of fungal disease, namely, anthracnose, powdery mildew, rust, downey mildew, early blight and late blight found on specific type of vegetable crop are considered for recognition and classification. The way the disease analysis is done considering both sides (front and back portions) of the leaves has been addressed. The analysis of the fungal disease present on the leaves is made in the early stage before it damages the whole leaf and subsequently the plant. The Local Binary Patterns (LBP) extracted from disease affected leaves is used as input to the classifiers. An integrated classification system “Neuro-kNN” has been proposed, of which multilayer BPNN classifier is used for training purpose and k-Nearest Neighbor (k-NN) classifier for testing purpose. The recognition accuracy is observed using Artificial Neural network (ANN) and Neuro-kNN classifier methods. The average classification accuracy is found to be 84.11% for the test samples using ANN. The average classification accuracy has increased to 91.54% using Neuro-kNN classifier. The work finds application in automatic recognition fungal disease found on vegetable crops by the service robots in the real world.

Keywords: Fungal disease, vegetable crops, Local Binary Patterns, Artificial Neural Network, k-Nearest Neighbor


1 Introduction

Computers have made impact in all the spheres of life through their tremendous technological developments in terms of more powerful and flexible computing devices. The potentials of computer and communication technologies are explored in science, engineering, medicine, commerce, law, etc. The field of agriculture and horticulture is not an exception. Computer vision applications are slowly making their way in the fields of agriculture and horticulture. Computers have been deployed in inspection of food products, packing of products, interpretation of grain and crop images in the agro-food industry, etc. The major advantage of using computers is that they are more accurate, precise and efficient as compared to human beings in the real world.

Today India ranks second worldwide in farm output. Agriculture is still the largest economic sector and plays a major role in socioeconomic development of India. Agriculture is the means of livelihood of almost two thirds of the workforce in India. India has over 210 million ha of farm land. In vegetables production, India is only behind China with an annual production of 87.53 million t from 5.86 million ha having a share of 14.4% to the world production. A large variety of vegetable crops are grown in India, of which onion, tomato, beans, bengal gram (chickpea), legumes and oilseeds are the major ones.

Adoption of high yielding cultivars and suitable production technologies has largely contributed for higher production and productivity. As such, several important decisions regarding safe practices for the production and processing of vegetables have been made in the recent
past. One of the main concerns is proper disease control. Early detection of diseases is a major challenge in horticulture/agriculture science. Every year, large quantities of chemicals are used as fungicides to control various diseases common to vegetables, thus evoking serious concern from environmentalists over deteriorating groundwater quality. Likewise, farmers are also concerned about the huge costs involved in these activities and severe profit loss. Early detection will help farmers to avoid huge loss. Technology support would help them in this aspect by cutting on cost of pesticides. The cost intensity, automatic correct identification and classification of diseases based on their particular symptoms are very useful to farmers and also agriculture scientists.

Several key technologies incorporating concepts from image processing and artificial intelligence were developed by various researchers in the past to tackle this situation. In all these techniques, digital images are acquired in a given domain using digital camera and image processing techniques are applied on these images to extract useful features that are necessary for further analysis.

The symptoms of fungal disease found on vegetable crops differ in color, shape, and size according to the cause. Machine vision techniques are used in this to solve problems of features extraction and analysis which include features of color, size, shape, and surface texture. An abnormal symptom is an indication to the presence of the disease, and hence, can be regarded as an aid in diagnosis. The diagnosis of fungal disease symptoms may cause some confusion due to the similarities in shape, size, and color but only an expert could recognize it. The first step in fighting against these symptoms is the adequate recognition of their presence.

In order to bring in the technology of a computer vision in automating this activity, the connected literature survey has been carried out to know the state-of-the-art. Blazquez (1990) described spectroscopic measurements of color infrared photographs for detection of late blight disease on tomato. Lefebvre et al. (1993) presented the problem in automating pulp sampling of potatoes such as their shape, color, and texture in order to detect viral diseases. Niemira et al. (1999) presented a method for quantification of late blight in potato tuber tissue using a digital scanner and image analysis software. Pydipati and Burks (2006) used a HSI color features based on Color Co-occurrence Matrix in conjunction with statistical classification algorithms, which is used to identify diseased and normal citrus leaves under laboratory conditions. Huang (2007) presented an application of neural network and image processing techniques for detecting and classifying phalaenopsis seedling diseases. Maliappis (2008) introduced a computer management for cultivation of low-tech greenhouses. The designed Greenhouse Intelligent Management System (GIMS) using web-based application is provided with personalized consultation to its users. Boissard et al. (2008) presented a strategy based on advances in automatic interpretation of images applied to leaves of roses. Camargo and Smith (2008) presented an image-processing based method that identifies the visual symptoms of plant diseases, from an analysis of colored images. Cui et al. (2009) explored feasible methods for quantifying soybean rust severity. Images of soybean leaves with different rust severity are collected using portable spectroradiometer and multi-spectral CDD camera. Gulhane and Gurjar (2011) introduced identifying and diagnosing disease appearing on cotton leaf. Bauer et al. (2011) proposed three methods of automatic classification of leaf diseases based on high-resolution multi-spectral stereo images. Al-Hiary et al. (2011) evaluated a software solution for automatic detection and classification of plant leaf diseases. Guru et al. (2011) presented a novel algorithm to classify tobacco seedling diseases present on leaves. Al Basheesh et al. (2011) designed and implemented image processing based software solution for automatic detection and classification of plant diseases. Ananthi and Varthini (2012) designed a software solution for automatic detection and computation of texture statistics for plant leaf diseases. Revathi et al. (2012) focused on new technological strategies using mobile captured symptoms of cotton leaf spot images and categorized the diseases using Homogeneous Pixel Counting Technique for Cotton Diseases Detection (HPCCD) algorithm. The
Neuro-kNN classification system for detecting fungal disease on vegetable crops

A classifier is trained to achieve intelligent farming, including early identification of diseases in the groves, selective fungicide application, etc. Chaudhary et al. (2012) compared the effect of CIELAB, HSI, and YCbCr color space in the process of disease spot detection. Patil and Raj kumar (2012) described a method for extraction of color and texture features of diseased leaves of maize. Kanjalkar and Lokhande (2013) proposed a software solution for automatic detection and computation of plant leaf diseases. Rathod et al. (2013) provided various methods used to study leaf disease detection using image processing. He et al. (2013) discussed the extraction methods of the damaged portion from the cotton leaf image in order to measure the damage ratio of the cotton leaf affected by disease or pests. Arivazhagan et al. (2013) proposed a software solution system for automatic detection and classification of plant leaf diseases. Bandi et al. (2013) evaluated the performance of various classifiers in detecting diseased citrus leaves.

From the literature survey, it is revealed that image processing techniques were applied for recognition and classification of different types of diseases in plants. However, not much work connected with identification and classification of fungal disease symptoms found on vegetable crops has been observed. It is realized that the image processing techniques could extend to the identification of fungal disease symptoms affected on specific type of vegetable crop. Although image recognition has attracted many researchers in the area of pattern recognition of plant’s leaves that is used in diagnosing the disease present on leaves, there is less attention given for disease diagnosis appearing on front and back portions of the plant’s leaves. The present work addresses how the fungal disease analysis is done considering both sides (front and back portions) of the leaves of vegetable crops. The analysis of the fungal disease present on the leaves of vegetable crops is detected in the early stage before it damages the whole leaf and subsequently the plant.

A lot of work has been done to automate the visual inspection of the diseases by machine vision with respect to color, texture, and shape. Instead of considering state-of-the-art color and texture feature separately, this work introduces a texture feature derived from the color images to validate the accuracy and efficiency. The identification of fungal disease symptoms affecting leaves of vegetable crops is done using Local Binary Patterns. An integrated image processing system has been proposed aiming at the identification and classification of fungal disease symptoms. The development of such an intelligent system is justified by its speed and reliability of the response of the system.

2 Proposed methodology

In the present work, tasks like image acquisition, preprocessing, segmentation, feature extraction, feature reduction, and classification are carried out. The classification tree is given in Figure 1. The detailed block diagram of adopted methodology is shown in Figure 2.

2.1 Image set

The vegetable crops like beans (*Phaseolus vulgaris*), bengal gram/chickpea (*Cicerarietinum*), soybean (*Glycine max*), sunflower (*Helianthus annuus*), and tomato (*Solanum lycopersicum*) affected by fungal disease are considered for the present work. The chosen symptoms of fungal disease found on vegetable crops
considered for recognition and classification are beans anthracnose, beans powdery mildew, bengal gram anthracnose, bengal gram powdery mildew, soybean powdery mildew, soybean rust, sunflower powdery mildew, sunflower rust, sunflower downey mildew, tomato early blight, and tomato late blight. The image samples were obtained from the department of plant pathology, at the University of Agricultural Sciences, Dharwad, India. Figure 3 shows the image samples belonging to various classes. Table 1 shows scientific classification of fungal symptoms affected on each vegetable type.

![Image samples](https://example.com/image.png)

**Figure 3** Images showing visual symptoms caused by fungal disease

<table>
<thead>
<tr>
<th>Fungal symptom</th>
<th>Causal organism</th>
<th>Family</th>
<th>Order</th>
<th>Class</th>
<th>Subdivision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beans anthracnose</td>
<td>Colletotrichum lindemuthianum</td>
<td>Phyllachoraceae</td>
<td>Phyllachorales</td>
<td>Sordariomycetes</td>
<td>Incertae sedis</td>
</tr>
<tr>
<td>Beans powdery mildew</td>
<td>Erysiphe betae</td>
<td>Erysiphaceae</td>
<td>Erysiphales</td>
<td>Leotiomycetess</td>
<td>Leotiomycetidae</td>
</tr>
<tr>
<td>Bengal gram anthracnose</td>
<td>Colletotrichum dematium</td>
<td>Glomerellaceae</td>
<td>Glomerellales</td>
<td>Sordariomycetes</td>
<td>Incertae sedis</td>
</tr>
<tr>
<td>Bengal gram powdery mildew</td>
<td>Leveillula taurica</td>
<td>Erysiphaceae</td>
<td>Erysiphales</td>
<td>Leotiomycetess</td>
<td>Leotiomycetidae</td>
</tr>
<tr>
<td>Soyabean powdery mildew</td>
<td>Microsphaera diffusa</td>
<td>Erysiphaceae</td>
<td>Erysiphales</td>
<td>Leotiomycetess</td>
<td>Leotiomycetidae</td>
</tr>
<tr>
<td>Soyabean rust</td>
<td>Phakopsora pachyrhizi</td>
<td>Phakopsoraceae</td>
<td>Uredinales</td>
<td>Uredinomycetes</td>
<td>Incertae sedis</td>
</tr>
<tr>
<td>Sunflower powdery mildew</td>
<td>Erysiphe cichoracearum</td>
<td>Erysiphaceae</td>
<td>Erysiphales</td>
<td>Leotiomycetess</td>
<td>Leotiomycetidae</td>
</tr>
<tr>
<td>Sunflower rust</td>
<td>Puccinia helianthi</td>
<td>Puccinaceae</td>
<td>Uredinales</td>
<td>Uredinomycetes</td>
<td>Incertae sedis</td>
</tr>
<tr>
<td>Sunflower downey mildew</td>
<td>Plasmopara halstedii</td>
<td>Peronosporaceae</td>
<td>Peronosporales</td>
<td>Oomycetes</td>
<td>Incertae sedis</td>
</tr>
<tr>
<td>Tomato early blight</td>
<td>Alternaria solani</td>
<td>Pleosporaceae</td>
<td>Pleosporales</td>
<td>Dothideomycetes</td>
<td>Pleosporomycetidae</td>
</tr>
<tr>
<td>Tomato late blight</td>
<td>Phytophthora infestans</td>
<td>Pythiaceae</td>
<td>Peronosporales</td>
<td>Oomycota</td>
<td>Peronosporomycetida</td>
</tr>
</tbody>
</table>

### 2.2 Image acquisition

For image acquisition, a color camera (DXC-3000A, Sony, Tokyo, Japan) was used. The camera had a zoom lens of 10-120 mm focal length and a 72 mm close-up lens set. To provide a rigid and stable support and easy vertical movement, the camera was mounted on a stand. The illumination source was a 32-W fluorescent bulb with a 305 mm diameter and a rated voltage of 230 V. The 72-mm close-up lens was used to achieve a spatial resolution of 0.064 mm/pixel in horizontal and vertical directions. The images acquired were 400x400 pixels in size. To keep distance between camera and image sample constant, we used vertical supports available on the camera. The setup used to obtain the image samples is shown in Figure 4.

![Image acquisition setup](https://example.com/image.png)

**Figure 4** Image acquisition setup

As the image samples were captured on the whole leaf with the background, the quality of the image was decisive for the results of analysis, affecting both the ability to detect features under analysis and precision of subsequent measurements. Background subtraction was done in order to reduce the scene complexities. The
result of the system heavily depended upon the efficient working of the image segmentation method. In order to segment foreground from background in the image, Chan-vase segmentation method was used.

2.3 Segmentation

The Chan-vase method was based on the active contour model, the Mumford-Shah $R^1$ functional and the Osher-Sethian level set method. Given a grayscale image in the Chan-Vase method $[I: \Omega: R^p (p = 2, 3)]$, a curve ‘C’ was defined that represents a partition $(\Omega)$ in terms of two regions $\Omega_{in}$ and $\Omega_{out}$. These regions gave an optimal piece-wise constant approximation of the given image. The contour minimizes the energy, as given by the Equation (1).

$$E(C, c_1, c_2) = \lambda_1 I_{\Omega_{in}}(I(x) - c_1)^2 - \lambda_2 I_{\Omega_{out}}(I(x) - c_1)^2) + \mu \text{ length}(C)$$

(1)

where, $c_1$ and $c_2$ are the average intensities in the regions $\Omega_{in}$ and $\Omega_{out}$ of an image respectively. The parameter $\lambda i$ represents an external energy and $\mu$ controls the size of the item. Algorithm 1 gives the steps involved Chan-Vase method based on segmentation process. Figure 5 gives a sample input image and the corresponding output image after using Chan-vase active contour method.

**Algorithm 1: Chan-vase method**

**Input:** Input image, initialization of number of regions, $(n=0)$ and maximum number of iterations.

**Output:** Segmented image

**Start**

Step 1: Set $n=n+1$

Step 2: Compute $c_1$ and $c_2$ inside and outside the active contour for the current level set function.

Step 3: Evolve level-set function.

Step 4: Repeat until the solution is stationary or $n>n_{max}$ represents the maximum number of iterations.

Step 5: Go to the next pixel to be examined.

**Stop.**

2.4 Feature extraction

One of the best performing texture descriptors and widely used in various computer vision applications was Local Binary Patterns (LBP) (Ojala et al., 2002). It was proven to be highly discriminative and its invariance to monotonic gray level changes and computational efficiency made it suitable for demanding image analysis tasks. Given a pixel in the input image, LBP was computed by comparing it with its Neighbors. LBP feature considers only signs of local differences (i.e. difference of each pixel with its neighbors). The LBP operator was designed for texture description as given in Equation (2). The operator assigned a label to every pixel of an image by thresholding the neighborhood of each pixel with the center pixel value and considering the result as a binary number. The texture descriptor was used to build several local descriptions of the image and combining them into a global description.

$$LBP_{N,R} = \sum_{v_n} s(v_n - v_c)2^n, s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

(2)

The $LBP_{N,R}$ indicates notation for LBP operator. The subscript represents using the operator in a $(P, R)$ neighborhood, $v_c$ is the value of the central pixel, $v_n$ is the value of its neighbors, ‘$R$’ is the radius of the neighborhood and ‘$N$’ is the total number of neighbors. Suppose the coordinate of $v_c$ is $(0,0)$, then the coordinates of $vn$ are $(Rcos(2\pi n/N), Rsin(2\pi n/N))$. The values of neighbors that are not present in the image grids may be estimated by interpolation. Let the size of image is $IJ$. After the LBP code of each pixel is computed, a histogram is created to represent the texture image given in Equation (3).

$$H(k) = \sum_{i=1}^{I} \sum_{j=1}^{J} f(LBP_{N,R}(i,j)k), ke[0,k]$$

(3)

where, ‘$K$’ is the maximal LBP code value. In this experiment the value of ‘$N$’ and ‘$R$’ are set to ‘8’ and ‘1’ respectively to compute the LBP feature. Figure 6 gives...
2.4.1 Feature reduction

The feature vector size obtained from LBP was too large to give as input to the classifier which results in lower classification accuracy. In machine learning, during the training of the classifiers, if the number of image features is large, it can lead to ill-posing and over fitting, and reduce the generalization of the classifier. One way to overcome this problem is to reduce the dimensionality of features. The dimensionality of features is reduced with the combination of LBP and Principal Component Analysis. PCA will reduce the length of the feature vector. PCA transfers a set of correlated variables into a new set of uncorrelated variables and maps the data into a space of lower dimensionality. After applying PCA, the feature vector reduced from 12610×16 to 20×20=400. Algorithm 2 depicts steps involved in reduction of LBP features using PCA.

**Algorithm 2: LBP feature reduction using Principal Component Analysis**

**Input:** Images of fungal disease symptoms affected on vegetables.

**Output:** Reduced LBP features.

**Start**

Step1: Collect the images of fungal disease symptoms affected on vegetable leaves

Step2: Apply Chan-vase segmentation technique

Step3: Obtain the mean value D of the LBP feature set ‘D’

Step4: Subtract the mean value D from D. From these values a new matrix ‘M’ is obtained

Step5: Calculate the co-variance matrix; C=MMT

Step6: Calculate the eigenvectors and eigenvalues of the covariance matrix

Step7: Choose components and form a feature vector

Feature Vector = (eig1, eig2, eig3 ………eign)

Step8: Derive the new LBP feature set

Final feature = Row feature vector x Row data adjust

**Stop.**

2.5 Classifier

A multilayer feed forward neural network with back propagation learning was used to recognize the different fungal disease symptoms affecting leaves of vegetable crops. Multilayer feed-forward neural networks were the most commonly used neural networks for object identification and classification. The layers of neurons between the inputs and the “output layer” were called “hidden layer”. The Back Propagation Neural Network (BPNN) was simple and effective to implement and found suitable for a wide range of machine learning applications. The number of neurons in the input layer corresponded to the number of input features and the number of neurons in the output layer corresponded to the number of classes. The number of nodes in the hidden layer is calculated using the Equation (4).

$$n = \frac{(I + O)}{2} + y^{0.5}$$  \hspace{1cm} (4)

where, $n$= number of nodes in hidden layer; $I$ = number of inputs feature; $O$ = number of outputs, and $y$ = number of inputs pattern in the training set.

K-Nearest Neighbor (kNN) is the simplest of all classification algorithms. This is a method of classifying patterns based on the class label of the closest training patterns in the feature space. Every time as a test pattern is to be classified, it has to be compared with all the training patterns, to find the closest pattern. An object was classified by a majority vote of the class of its neighbors. The object was assigned to the class most common amongst its k nearest neighbors. The ‘K’ value is chosen by using the Equation (5).

$$K = \sqrt{\text{number of feature sets}}$$  \hspace{1cm} (5)
The study uses eleven output nodes and four-hundred input nodes corresponding to eleven chosen categories of image samples, and the chosen four-hundred image features respectively. The study considers 990 image samples (90 samples of each type). Among the image samples considered, 440 are front leaves (40 samples of each type) and 550 are back leaves (50 samples of each type). The image samples are divided into two halves and one half is used for training and other for testing. The training and testing are carried out with LBP features.

The process of recognition and classification is given in Algorithm 3.

Algorithm 3: Recognition and classification of fungal disease symptoms

**Input:** Images of fungal disease symptoms found on vegetables crops.

**Output:** Recognized and classified images.

**Start**

Step 1: Collect the images of fungal disease symptoms

Step 2: Apply Chan-vase segmentation technique

Step 3: Extract LBP features

Step 4: Perform feature reduction using Principal Component Analysis

Step 5: Train the BPNN with reduced LBP features

Step 6: Accept test images and perform Step 2 through Step 4

Step 7: Recognize and classify the images using BPNN classifier

**Stop.**

3 Results and discussion

All the algorithms used in this work were implemented using MATLAB 7.10. The image samples were divided into two halves and one half was used for training and the other was for testing. The percentage of recognition and classification accuracy was defined as the ratio of correctly recognized image samples to the total number of test image samples. The percentage accuracy is calculated as given by Equation (6).

\[
\text{Percentage accuracy} = \frac{\text{Correctly recognized image samples}}{\text{Total number of test image samples}} \times 100
\] (6)

The histogram classification accuracies are plotted as shown in Figure 7. From Figure 7, it was observed that the maximum recognition and classification accuracy of 93% was found with sunflower downey mildew and minimum accuracy of 66% was found with bengal gram powdery mildew for front portion of leaves. The maximum recognition and classification accuracy of 96% was found with soybean rust and minimum accuracy of 72% was found with beans powdery mildew for back portion of leaves. The maximum recognition and
classification accuracy of 94% was found with soybean rust and minimum accuracy of 80% was found with bengal gram anthracnose for both front and back portions of leaves.

The histogram classification accuracies are plotted as shown in Figure 8. It was observed that the maximum recognition and classification accuracy of 96% was found with tomato early blight and minimum accuracy of 84% was found with sunflower powdery mildew for front portion of leaves. The maximum recognition and classification accuracy of 96% was found with beans anthracnose and minimum accuracy of 84% was found with sunflower powdery mildew for back portion of leaves. The maximum recognition and classification accuracy of 100% was found with beans powdery mildew and tomato late blight respectively and minimum accuracy of 90% was found with sunflower powdery mildew and sunflower downey mildew respectively for both front and back portions of leaves.

The performances of Artificial Neural Network using BPNN and Neuro-kNN classifiers were compared (Figure 9). The comparison result showed that average classification accuracies obtained with front, back, and both back and front portions of leaves of vegetable crops are 82.90%, 94.72% and 86.27% respectively using ANN classifier. The average classification accuracies obtained with front, back, and both back and front portions of leaves of vegetable crops were 89.90%, 90% and 94.72% respectively using Neuro-kNN classifier. It was observed average identification accuracies of 91.54% and 84.11% for the test samples using Neuro-kNN and ANN classifiers respectively. It was evident that the Neuro-kNN classifier gave better classification accuracy than the ANN classifier for recognition and classification of fungal disease symptoms affecting both sides of the leaves of vegetable crops.

Figure 10 gives the computation time required for the
considered work using the ANN and Neuro-kNN classifiers. The average computation time for the two classifiers was computed in seconds. It showed that Neuro-kNN had 13% speedup over the ANN classifier.

![Figure 9 Average classification accuracies of fungal disease symptoms using ANN and Neuro-kNN classifiers with LBP](image)

![Figure 10 Speedup of computation for ANN and Neuro-kNN classifiers](image)

4 Conclusions

In this paper we used the proposed methodology, LBP for identification and classification of fungal disease symptoms affecting both sides of the leaves of vegetable crops. It was observed from the results that the identification accuracy was 91.54% for the test samples using Neuro-kNN classifier. However, when ANN using BPNN was used for both training and testing, the identification accuracy was 84.11%. One interesting observation was that for LBP, one can train the classifier with ANN classifier and test using k-NN classifier. The whole idea behind testing the classifier using k-NN is to make testing time short compared to training. One can afford longer training time but not the testing time. The testing time needs to be short to achieve real-time applications. The experimental results indicated that the Neuro-kNN approach was a valuable approach compared to ANN approach which can significantly support an accurate recognition of fungal disease in a little computational effort.

For future study, further different neural network architectures, such as Support Vector Machine (SVM) and Fuzzy based classifiers can be used for classification. This work can be extended to classify visual symptoms affected by fungal disease found on commercial crops, cereal crops, and fruit crops. The work can also be extended to identify various diseases like viral, bacterial found on agriculture/horticulture crops.

Acknowledgements

We would like to express our thanks to Dr.B.S. Anami, principal, K.L.E.Institute of Technology, for his valuable suggestions.

References


Revathi, P., and M. Hemalatha. 2012. SMS based HPCCDD algorithm for the identification of leaf spot diseases, proceedings of the fourth international conference on signal and image processing. Lecture Notes in Electrical Engineering, Springer India.