Linear Vector Quantization for the Diagnosis of Ground Bud Necrosis Virus in Tomato

Kaveri Umesh Kadam¹, R. B. Dhumale², N. R. Dhumale³, P. B. Mane⁴, A. M. Umbrajkaar⁵ and A. N. Sarwade⁶

¹Department of Electronics & Communication Engineering, Jamia Millia Islamia, New Delhi, India
²,³AISSMS Institute of Information Technology, Pune, India
⁴Dr. D. Y. Patil, Institute of Engineering, Management & Research, Akurdi, Pune
⁵Department of Electronics & Communication Engineering, Jamia Millia Islamia, New Delhi, India

*Correspondence: R.B. Dhumale; rbd.scoe@gmail.com; Tel.: +91-9763463366

ABSTRACT - In this varying environment, a correct and appropriate disease diagnosis including early preclusion has never been more significant. Our study on disease identification of groundnut originated by Groundnut Bud Necrosis Virus will cover the way to the effective use of image processing approach in agriculture. The difficulty of capable plant disease protection is very much linked to the problems of sustainable agriculture and climate change. Due to the fast advancement of Artificial Intelligence, the work in this paper is primarily focused on applying Pattern Recognition based techniques. The purpose is to determine the grade of disease to control by developing a model for the selection of bud blight disease caused by GBNV in tomatoes. The images are classified according to the grade of the disease. Different methods have been applied to make a proper diagnosis by bringing clarity in the diagnostic results. Linear Vector Quantization works well than, Radial Basis Function, Back Propagation Neural Network and Support Vector Machine.

Keywords: Groundnut Bud Necrosis Virus, Artificial Intelligence, Pattern Recognition, Grade of the Disease.

INTRODUCTION

China and India. Tomatoes can be grown in different geographical regions in open grounds or greenhouses, and the fruit can be cropped by labor-intensive or automatic means. In certain situations like rejuvenation, pruning, weeding, irrigation, and frost protection; this crop plant can be continuously recurring or semi- recurring, but commercially it is believed a once a year [2].

The main restrictions in the production of tomatoes are high price of production, deficient of postharvest facilities, useless & uneven supply chain, lack of knowledge regarding simply writing this "market survey for product demand, the unavailability of disease-free good quality seeds, deprived agronomic practices, aboriginal weeding techniques and short of suitable plant safety measures for the control of insect or pests as well as diseases [3]. These challenges result in low productivity, poor quality, and high wastages [4]. The important thing is different diseases in plants cause production and financial losses as well as a decrease in both quality and quantity of tomato production [5].

The disease of groundnut originated by Groundnut Bud Necrosis Virus (GBNV) is usually spread in tomato strains and has caused severe damage to many other crops in Australia, India, Nepal, China, Thailand, and the USA [6]. The scientific name of GBNV is Peanut Bud Necrosis Virus. The virus is transmitted by a Thrips Palmy Worm as a vector. It was the first time noticed in India in 1968 [7]. It infects a wide range of economically significant crop plants, including tomatoes. The key signs and symptoms of disease includes necrotic ring spots on the leaves and necrotic patches on the fruits, necrosis of young growing buds, bronzing of the leaf with brown necrotic lesions and in severe cases wilting and stunting of seedlings. This causes 50-90% damage to the crop [8].

Various diseases can be affected by the production of tomato and it reduces the production capacity [9]. Maintaining the healthy status of tomatoes or avoiding the effects of diseases is important [10]. Tomatoes are cultivated in different climates in different parts of the world. In such different climates, if a manual diagnosis of the disease is performed, it is possible to produce false and misleading results [11]. The visual symptoms become visible; samples are collected and examined via the traditional approach [12]. Traditional techniques of disease identification include various immune systems, which require laboratory setups and expert technicians [13], [14]. Although this is a perfect technique, there are constraints to its implementation; it is a costly, labor-intensive, and, significantly time-consuming approach [15]. In addition to laboratory experiments, the diagnostician should also examine biological and chemical characteristics, symptom distribution and
variability, the specificity of the host, to determine the causative factors [16].

The Knowledge-Based Disease Diagnosis System is included to diagnose the plant disease. In which technicians take samples to the laboratory and diagnose the disease [17]. However, this diagnosis can take time because the village lacks such a laboratory [18]. Therefore, the motivation of this project is to provide farmers with a diagnosis of the disease at a low cost and in less time [19]. The Image Processing (IP) based system will work in such a way, farmers can capture the image of the crop in their fields and upload it on the internet using Android app to know what disease has happened to their crop [20]. Therefore, that they can get an advantage in the village and do not always have to go to the city. Different diseases are encountered during the production of tomatoes. The disease must be identified precisely [21].

Table 1: Survey on Classification Techniques

<table>
<thead>
<tr>
<th>Comparison parameters</th>
<th>Decision tree</th>
<th>Naive Bayes</th>
<th>k-NN</th>
<th>SVM</th>
<th>ANN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning speed</td>
<td>Common</td>
<td>Top</td>
<td>Top</td>
<td>Worst</td>
<td>Top</td>
</tr>
<tr>
<td>Classification speed</td>
<td>Top</td>
<td>Top</td>
<td>Worst</td>
<td>Good</td>
<td>Good</td>
</tr>
<tr>
<td>Performance with presence of missing value</td>
<td>Average</td>
<td>Good</td>
<td>Good</td>
<td>Best</td>
<td>Best</td>
</tr>
<tr>
<td>Performance with non-relevant features</td>
<td>Average</td>
<td>Good</td>
<td>Average</td>
<td>Good</td>
<td>Good</td>
</tr>
<tr>
<td>Noise tolerance</td>
<td>Good</td>
<td>Average</td>
<td>Average</td>
<td>Good</td>
<td>Good</td>
</tr>
<tr>
<td>Performance on discrete/binary attributes</td>
<td>Good</td>
<td>Average</td>
<td>Average</td>
<td>Worst</td>
<td>Good</td>
</tr>
<tr>
<td>Tolerance with parity problems</td>
<td>Good</td>
<td>Worst</td>
<td>Average</td>
<td>Average</td>
<td>Best</td>
</tr>
<tr>
<td>Clarity on Classification prediction</td>
<td>Best</td>
<td>Best</td>
<td>Average</td>
<td>Worst</td>
<td>Average</td>
</tr>
<tr>
<td>Handling of model parameter</td>
<td>Average</td>
<td>Best</td>
<td>Average</td>
<td>Worst</td>
<td>Best</td>
</tr>
<tr>
<td>Overall accuracy</td>
<td>Good</td>
<td>Worst</td>
<td>Good</td>
<td>Best</td>
<td>Good</td>
</tr>
</tbody>
</table>

The work has proposed for classifying various types of rice diseases by extracting features from the infected areas of the rice plant images [28]. Fermi energy-based segmentation technique has been suggested to separate the infected area of the image from its surroundings [29]. Based on the field experts’ opinions, the signs of the disease are illustrated with features like color, shape as well as the location of the infected segment as well as extracted by developing new algorithms [30]. To shrink the difficulty of the classifier, important features are selected using Rough Set Theory (RST) to minimize the loss of information [31]. As a final point using selected features, a rule-based classifier has been implemented that covers every diseased rice plant image as well as gives better results compared to conventional classifiers [32].

The study conducted to computerize the Plant Disease Identification (PDI) on the significance of IPT includes taking of six varieties of plant disease infected leaves along with classification using a decision tree [33]. In the feature extraction phase, a total of ten intensity based statistical features were calculated for every plant disease [34]. The concluding feature vector consists of ten features; these selected features are proposed to the decision tree for classification. As a whole 78% detection accuracy acquiesced by a decision tree [35]. A method for plant leaf disease recognition as well as classification using K-nearest neighbor (KNN) classifier has been proposed [36]. The texture features are extracted from the leaf disease images for classification [37]. K-NN classifier was used to classify Alternaria, anthracnose’s, bacterial blight, leaf spots, and cankers of different plant species. The proposed method can fruitfully identify and distinguish the selected diseases with 96.76% accuracy.

A KNN-SVM based PDI scheme by using the method of feature extraction with the aid of color, shape as well as Gabor filter feature extraction model, clustering by applying KNN-SVM based classification has been proposed. It is estimated that the Mean Squared Error (MSE) and the accuracy level of the proposed work are improved than the conventional classification tree method. The MSE of conventional, as well as the proposed method, is 0.29 and 0.009 correspondingly, which means that the error rate has decreased as well as the accuracy is 97.69% and 91.77% correspondingly, demonstrates progress in efficiency [38].
The developed software system for automatic PDI handing out system has five main steps; a color transformation organization for the input RGB image is generated, and then noise removal, image segmentation, lastly the classification is calculated by giving various features i.e. size, color, proximity, and Average Centroids Distance. Experimental end outcomes on a database of 4 dissimilar diseases verify the robustness of the proposed approach.

The research work is proposed to achieve the following objectives to develop a tomato disease diagnosis process with improved performance.

1. To develop a model for the selection of bud blight disease caused by GBNV in tomato.
2. To make an efficient use of image processing techniques for bud blight disease in tomato.
3. To develop a cost-effective app, this will be user friendly.
4. Strategies to improve economic loss, quality, and quantity of tomatoes.
5. To suggest chemicals for disease control in tomato.

## 2. METHODOLOGY
The key to the analysis is the selection of tomato fruits. Many researchers acquired images that are constantly illuminated. For the purposes of analysis, mainly color or gray images are considered. This chapter discusses sample tomato selection methods, sampling time, site selection, and detailed analysis of the methodology used in this proposed work to diagnose GNBV disease.

Site choice is essential. Similar varieties, rootstock as well as similar age, vitality, and health plots on the similar soil region are picked. To assist with the diagnosis, two samples should be taken when one is healthy and one should be affected by disease. Tomatoes are cultivated in Pune, Nashik, Beed, as well as Narayangaon close to Pusegaon. In Narayangaon alone, an area of about 18,000 acres is under cultivation of tomatoes, resulting in about 1400 tons of produce a day. In Pimpalgaon, around 40,000 hectares of tomatoes are cultivated. Nashik is the main supplier of tomatoes in most parts of the country. Pimpalgaon supplies tomatoes to other markets like Delhi, Madhya Pradesh, Assam and Haryana. Moreover, tomatoes are exported to countries like Bangladesh and Pakistan. The samples required for this project have been collected from Pune Market Yard during March 2022.

Experiments were conducted on tomatoes divided into two groups, varying the level of maturity in the local market. Images of collected samples of tomatoes are classified into five groups. It was classified according to the grade of the disease. Each grade typically had 70 images as samples. Some of these image samples are shown in table 2.

The direct methods consist of molecular techniques and serological methods. Where Indirect Methods consist of biomarkers for disease detection and Plant Properties or Stress-Based Disease Detection. In this work, an imaging technique is proposed. It has been used as a machine learning-based GNBV disease diagnosis method. This has the following main steps and is shown in figure 1.

- Image acquisition,
- Resize Image,
- Image Preprocessing,
- Feature Extraction,
- Feature Selection,
- Disease Classifier,
- Image Segmentation,
- Severity Calculation.

When all above steps are completed, it is important to find the area of the infected area. The grade is determined by the infected area. The infected areas in this work are classified into five sections. These sections are given in table 3.

The boundaries state the bottom and the top 1% of all pixel values. In this work, only the colour intensity values are improved. The intensity values are mapped in gray-scale images of random size into new values in 300X400. The RGB model, which is based on the initials of the primary colours, consists of three mixed primary colours, red, green, and blue. Compute the chosen feature given below. This computation employs only the values in the GLCM. Energy, Entropy, Contrast, Homogeneity, Correlation, Shade and Prominence.

**Artificial Neural Network (ANN)** is used as a method of basic classification to diagnose the disease. In this work, different classification methods have been used to diagnose the disease. These methods have been used for quantitative studies. However, all these methods are based on machine learning algorithms.

Learning Vector Quantization (LVQ) is used to classify patterns. It consists of output units and each output unit represents a specific class. The unit represents the class with the unit weight vector as the reference or codebook vector. This is a unique instance of a supervised learning-based competitive net. It is discovered that the output units are positioned to estimate the decision surfaces of the current Bayesian classifier at the training procedure stage. In this case, the network is given a series of training patterns with predetermined distributions of the referenced vector. After training, the LVQ net classifies an input vector whose weight is substantially similar to an input vector belonging to the same class. By conducting LVQ boundaries, inaccurate classification is reduced. LVQ is a supervised learning algorithm that looks like a KSOFM net, while KSOFM is an unsupervised learning algorithm.

The structure of LVQ is shown in figure 2. The following are some of the parameters used in an LVQ training process:

| X | Input Training Vector (X₁, X₂, X₃,...,Xₙ,...,X₂ₐ,Xₙₐ) |
| T | Target Output Vector (T₁, T₂, T₃,...,Tₙ,...,T₂ₐ,Tₙₐ) |
| Wⱼ | Weights connecting between hidden layer to output layer |
| Cⱼ | Bias of Hidden Unit |

Website: www.ijeer.forexjournal.co.in
### Table 2: Database prepared for tomato samples

<table>
<thead>
<tr>
<th>Grade</th>
<th>Infected Area</th>
<th>Image</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade 0 (Healthy)</td>
<td>00%-05%</td>
<td><img src="image" alt="Healthy Tomato" /></td>
</tr>
<tr>
<td>Grade I</td>
<td>01-25%</td>
<td><img src="image" alt="Grade 1 Tomato" /></td>
</tr>
<tr>
<td>Grade II</td>
<td>26-50%</td>
<td><img src="image" alt="Grade 2 Tomato" /></td>
</tr>
<tr>
<td>Grade III</td>
<td>51-75%</td>
<td><img src="image" alt="Grade 3 Tomato" /></td>
</tr>
<tr>
<td>Grade IV</td>
<td>75% above</td>
<td><img src="image" alt="Grade 4 Tomato" /></td>
</tr>
</tbody>
</table>

### Table 3: Grading as per infected area

<table>
<thead>
<tr>
<th>Grades</th>
<th>Infected area in (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade 0 (Healthy)</td>
<td>00-05</td>
</tr>
<tr>
<td>Grade 1</td>
<td>06-25</td>
</tr>
<tr>
<td>Grade 2</td>
<td>26-50</td>
</tr>
<tr>
<td>Grade 3</td>
<td>50-75</td>
</tr>
<tr>
<td>Grade 4</td>
<td>76-100</td>
</tr>
</tbody>
</table>
Linear Vector Quantization for Diagnosis of Ground Bud Disease

Figure 1: Machine Learning based GNBV Disease Diagnosis Method
The algorithm is given below. The purpose of the algorithm is to find the output unit near the input vector in which each output unit is a known class.

**Step-0** Initialize the reference vector using the following steps.
- First, take the "m" training vector from the set of training vectors as weight vectors; use the remaining vectors for training.
- Assign the initial weights.
- K-means clustering method.
- Set initial learning rate (η).

**Step-1** Do step 2- step 6 if the stop condition is wrong.

**Step-2** Do step 3- step 4 for every x-valued training input vector.

**Step-3** The Euclidean distance be calculated; for \( i = 1 \text{ to } n, j = 1 \text{ to } m \),
\[
D_{ij} = \sum_{i=1}^{n} \sum_{j=1}^{m} (x_j - w_{ij})^2
\]
When \( D_{ij} \) is the least, identify the successful unit index J.

**Step-4** Utilize the following conditions to update the weights on the winning unit, \( W_j \). If \( T = C_j \) then \( W_{j(new)} = W_{j(old)} + \eta(x - W_{j(old)}) \)
If \( T \neq C_j \) then \( W_{j(new)} = W_{j(old)} - \eta(x - W_{j(old)}) \)

**Step-5** Slow down the rate of learning η.

**Step-6** Check whether the network is in a stopping state.

The clustering technique in food image processing is a great option for properly classifying objects into different groups by sharing common properties. Data segmentation is the usual technique for the analysis of statistical data in which unsupervised learning is used to set in K-subsets. Straightforward and very fast K-means technology is used to divide the image by grouping the pixels present in the image. Practically attractive K-means are used for a variety of clustering methods. The cluster center divides the input dataset into K-clusters by the center of change of the optimization of each cluster. The distance between the clustering input and the center is calculated and inserted into the adjacent cluster. Implemented a repetitive version of the K-means algorithm to separate the tomato image into two divisions like the infected part and a good part. The algorithm of image segmentation is given below.

**Step 0:** Convert Image from RGB to CIELAB color space.
**Step 1:** Euclidean measured the difference between the two colors using distance metrics and divided them into 'a * b * c' spaces using K-means clustering.
**Step 2:** From the K-mean clustering results drawn in Step 1, apply the name, i.e., label each pixel of the tomato image as an infected or well pixel. And after naming, calculate how many pixels are in each section.
**Step 3:** To create different images based on the number of clusters, pixels will have to be separated in the image using the given name. Use color to show images in two colors that are contagious and good.

### 3. RESULTS AND DISCUSSION

In this proposed work, the techniques utilized for feature extraction are GLCM. During the research, 350 images are considered from the image dataset. Each image is preprocessed, furthermore the GLCM is created. Texture features are computed and the attached regions are employed to take out the shape features. The feature significances achieved from every image are used for training and testing of the classifier. Out of 350 images, the features of only 6 images are shown in table 4.

The texture features are computed exactly by GLCM. These features are named as F1 to F6 for further analysis of the proposed method.

Learning vector quantization (LVQ), established by Kohonen, is a very intuitive and easy yet controlling classification scheme which is very smart for a lot of reasons. The technique is executed; treaty with multilevel things that are categorized;
moreover, dissimilar to waste alternative sorting schemes such as feed-forward networks or support vector machines (SVMs), the framework of intuitive assignment of class data to prototypes requires clearing the LVQ system. For these causes, LVQ has been used in a variety of educational and professional applications such as image analysis, bioinformatics, telecommunications, robotics, and more. The structure of LVQ trained for GBNV disease diagnosis is shown in figure 3. Best training performance of LVQ is obtained at epoch 28 of 0.084673 is shown in figure 4. Percentage of correctly classified patterns using LVQ algorithm is given in table 5.

Table 4: Features Extracted for Selected Images

<table>
<thead>
<tr>
<th>Image</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F4</th>
<th>F5</th>
<th>F6</th>
<th>F7</th>
<th>F8</th>
<th>F9</th>
<th>F10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Img-1</td>
<td>0.9582</td>
<td>0.1255</td>
<td>0.9783</td>
<td>0.2573</td>
<td>52.101</td>
<td>80.813</td>
<td>3.7319</td>
<td>8.7552</td>
<td>5006.0</td>
<td>2.9486</td>
</tr>
<tr>
<td>Img-2</td>
<td>0.9636</td>
<td>0.2923</td>
<td>0.9350</td>
<td>0.2298</td>
<td>41.325</td>
<td>64.578</td>
<td>4.0625</td>
<td>8.4992</td>
<td>2783.7</td>
<td>4.1647</td>
</tr>
<tr>
<td>Img-3</td>
<td>0.9688</td>
<td>0.1408</td>
<td>0.9791</td>
<td>0.3178</td>
<td>53.018</td>
<td>82.872</td>
<td>3.7778</td>
<td>7.6539</td>
<td>3863.5</td>
<td>3.0045</td>
</tr>
<tr>
<td>Img-4</td>
<td>0.9628</td>
<td>0.1772</td>
<td>0.9703</td>
<td>0.2580</td>
<td>57.779</td>
<td>79.905</td>
<td>4.2828</td>
<td>8.4053</td>
<td>3312.7</td>
<td>2.5442</td>
</tr>
<tr>
<td>Img-5</td>
<td>0.9652</td>
<td>0.1103</td>
<td>0.9857</td>
<td>0.2624</td>
<td>71.637</td>
<td>91.445</td>
<td>4.5543</td>
<td>10.230</td>
<td>7348.8</td>
<td>1.9785</td>
</tr>
<tr>
<td>Img-6</td>
<td>0.9735</td>
<td>0.0950</td>
<td>0.9885</td>
<td>0.4169</td>
<td>47.629</td>
<td>82.231</td>
<td>3.1947</td>
<td>6.4566</td>
<td>4080.1</td>
<td>3.2221</td>
</tr>
</tbody>
</table>

Table 5: Percentage of correctly classified patterns using LVQ algorithm

<table>
<thead>
<tr>
<th>Techniques</th>
<th>Training Accuracy</th>
<th>Testing Accuracy</th>
<th>Overall Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Back-Propagation Neural Network (BPNN)</td>
<td>84.38 %</td>
<td>81.09 %</td>
<td>82.73 %</td>
</tr>
<tr>
<td>Radial Basis Functions (RBFs)</td>
<td>94.50 %</td>
<td>93.58 %</td>
<td>94.04 %</td>
</tr>
<tr>
<td>Linear Vector Quantization (LVQ)</td>
<td>97.33 %</td>
<td>98.73 %</td>
<td>97.53 %</td>
</tr>
<tr>
<td>Support Vector Machine (SVM)</td>
<td>84.72 %</td>
<td>83.28 %</td>
<td>84.00 %</td>
</tr>
</tbody>
</table>
Table 7: Comparative results based upon Specificity and Sensitivity

<table>
<thead>
<tr>
<th>Techniques</th>
<th>Specificity</th>
<th>Sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Back-Propagation Neural Network (BPNN)</td>
<td>8.3</td>
<td>8.3</td>
</tr>
<tr>
<td>Radial Basis Functions (RBFs)</td>
<td>8.5</td>
<td>8.5</td>
</tr>
<tr>
<td>Linear Vector Quantization (LVQ)</td>
<td>9.3</td>
<td>8.9</td>
</tr>
<tr>
<td>Support Vector Machine (SVM)</td>
<td>7.7</td>
<td>8.7</td>
</tr>
</tbody>
</table>

![Figure 5: Performance evaluation of Disease classifiers](image)

![Figure 6: Performance evaluation based on Specificity and Sensitivity](image)

4. SIMULATION RESULT

Agriculture is the backbone of our Indian economy. Considering the importance of the problem related to agricultural production, the objective of this work is to contribute to the development of agriculture by developing favorable applications related to agricultural development. Modern farming uses the latest technologies to help farmers get disease information at low cost and easily accessible. Modern methods have been developed to diagnose fruit, leaf, or branch diseases of different crops, fruits, and flowers. It uses a digital image processing method. The features are extracted from the captured image and used as input to the classifier. Machine Learning algorithms are widely used to implement such classifiers. The purpose is to determine the grade of disease to control by developing a model for the selection of bud blight disease caused by GBNV in tomatoes. The images are classified according to the grade of the disease. Each grade typically had 70 images as samples. This work encourages discussion on the diagnosis of GBNV disease on tomatoes, by assessing the data in terms of the proposed work in materials and methods. Different methods have been applied to make a proper diagnosis by bringing clarity in the diagnostic results. LVQ works well than RBF, BPNN, and SVM with an overall classification accuracy of 97.53%.

REFERENCES


