Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 6.199

IJCSMC, Vol. 8, Issue. 8, August 2019, pg.6 – 16

Classification and Segmentation Approach for Plant Disease Detection

Narinder Kaur[1]; Prabhjot Kaur[2]
[1]Universal Group of Institutions, Lalru, Punjab, India, kachooraneenu@gmail.com
[2]Universal Group of Institutions, Lalru, Punjab, India, directoracdm@ugichd.edu.in

Abstract: The infections in plants leafs can be detected with the help of plant disease recognition system. Different phases such as textural feature scrutiny, segmentation and classification are involved in plant infection or disease recognition. This study utilizes the KNN classifier along with GLCM algorithm for the plant disease detection. The projected method primarily takes pre-processed picture in form of input. In the subsequent step, texture feature scrutiny is performed with the help of GLCM algorithm. The area based segmentation is executed through K-mean clustering and KNN classifier is implemented for the infection forecasting. MATLAB programming is used for the implementation of projected approach. The reproduction outcomes demonstrate precision ratio up to 97 percent.
KEYWORDS: SVM, KNN, GLCM, K-means , Region based Segmentation

1. Introduction

The food necessities of whole human kinds are being fulfilled by only the Agricultural domain. In several developing countries like India around 70% of the entire population depends on farming and produce various types of fruits and vegetable harvest in their farms. Though, for the cultivation of best possible yield and quality harvest, several high quality techniques are needed. Therefore, the detection of plant infection is extremely imperative [1]. The farmers face various issues in agriculture therefore the recognition of plant diseases is a tricky procedure. An investigational scrutiny is identified as potato plant pathology in which the infectious syndromes of the plants are considered. The disease detection application is implemented for the up gradation of farming regions. The infection managing system provides assistance in the presentation of precise therapy suggestions. In the present scenario, the three mechanisms of

device apparition scheme such as processor, digital camera and application software are utilized. A lot of algorithms are incorporated inside the application software. The environmental science involves numerous applications like digital picture processing and picture scrutiny tools for proposing approaches in the microelectronics and processor structure [2]. The different concerns related to conventional cinematography are also solved in this study. In the plant disease detection technique, the descriptions gathered from the infinitesimal to telescopic arrays are improved and examined by utilizing this instrument. The environmental science incorporates numerous applications. Several issues like untimely and belatedly diseases, bug harm, and spin viral infections are responsible for fewer yields of potatoes. The picture or image processing is the preliminary phase in picture investigation and prototype detection and this is also considered the most important phase of image processing [3]. This phase is used for the determination of concluding outcomes of scrutiny. The image segmentation procedure is used for the portioning of a picture into displaced sections.

2. Literature Review

R. Meena Prakash, et.al (2017) reviewed numerous image processing approaches for the identification of plant diseases [4]. The main objective of this study was the implementation of image analysis and classification methods for the detection and classification of the leaf infections. The projected structure mainly included four dissimilar sections. These were identified as picture processing, segmentation of leaf with the help of k-means clustering for the recognition of contaminated regions, characteristic withdrawal and categorization of infections. This methodology utilized arithmetical Gray-Level Co-Occurrence Matrix (GLCM) aspects and categorization processes for the retrieval of textural features. The projected approach exploited Support Vector Machine (SVM) classifier for providing enhanced feature extraction.

Kawaljit kaur, et.al (2017) proposed an inclusive research work relevant to numerous ailments recognized inside the fruits crops [5]. An automatic technique was developed for the identification of fruit infections and this technique also reduced the amount of time required for the detection of fruit diseases. The noise was the main cause of image distortion. Therefore this approach also utilized noise removal scheme. The proposed research work scrutinized that affliction infection affected mostly fruit harvest. The performance investigation was carried out with the help of indistinct lead picture for the identification of infections inside the fruit yields.

The particular valued scrutiny was implemented inside the image processing systems for the detection of diseases in the preliminary phase.

Chaitali G. Dhaware, et.al (2017) proposed a novel approach for the identification and classification of automated leaf diseases from several plant leaves with the help of image processing technique [6]. This scheme was implemented with the help of sensible demand as the descriptions were moved unswervingly because of the least hard work of farmers. Therefore, this approach reduced the hard work of farmers involved in farming. The farmer captured the picture of plant leaf with the help of appropriate mobile camera. This picture was applied to the DSS mechanism without making any supplementary effort.

Pranjali B. Padol, et.al (2016) projected a novel approach which used SVM classification mechanism for the recognition and categorization of infections occurring inside the grape plant leaves [7]. Several techniques like resizing, thresholding and Gaussian sieving were implemented in the image processing approach. The K-means clustering practice was utilized for the execution of grape leaf segmentation. This practice used textural and shade aspects for clustering. In addition, SVM classification method was used for the identification of infection kind of grape leaf. The proposed work used two dissimilar groups of grape plant leaf named as Downy Mildew and Powdery Mildew experiment. The projected scheme showed a precision rate of 88.89% for both of these classes.

Preetha Rajan, et.al (2016) projected a novel approach for the identification of vermin within the harvest [8]. The proposed approach made the use of image processing for this purpose. Several experiments were performed with the help of this scheme on the basis of various plant pictures collected from different origins. This projected technique proved beneficial for the timely recognition of vermin inside the harvest. Therefore, with the help of proposed approach, the use of insect killers in farming meadows could be diminished. This measure also proved cost effective and nature protecting. The simulations outcomes demonstrated that this scheme was proved effortless and competent for different purposes. The proposed approach also improved the time and correctness level than various other physical schemes.

2.1 Support Vector Machine

The Support Vector Machines (SVMs) are considered as a set of pertinent supervised learning methods which are employed for the implementation of cataloging or classification and regression. This approach minimizes the experiential categorization fault and maximizes geometric margin concurrently. The support vector machine classifier is also identified as Maximum Margin Classifier because of this attribute. This approach is relied on the Structural risk Minimization (SRM) as well. The support vector machine classifier maps the key vector in the vicinity of higher dimensional space where maximal unraveling hyperplane is produced. Two analogous hyperplanes are developed on every face of the hyperplane for the partitioning of information. The separating hyperplane is a kind of hyperplane which is used for the maximization of the detachment amid two analogues hyperplanes [11]. It is implicated here that the bigger is the gap or remoteness amid these analogous hyperplanes, the superior is the simplification fault of the classifier. The considered information points are applied in the form of Equation (5).

$$\{(x1, y1), (x2, y2), (x3, y3), \ldots., (xn, yn)\} \quad \ldots (5)$$

In this equation,$yn = 1/-1$, is a stable entity which represent the class to which point *xn* relates. The variable *n* represents the amount of patterns included. Each xn represents the p-dimensional genuine vector. The scaling is imperative for the safeguarding of variables having huge variance. The training data is analyzed by utilizing the partitioning of hyperplane in the given equation which comprises:

$$w.x + b = 0 \qquad \ldots (6)$$

In this equation, the scalar value is represented by vector 'b' and vector 'w' represents the p-dimensional vector. The vector 'w' is demonstrated perpendicular to the separating plane. The addition of offset parameters 'b' is performed for the increasing of gap or distance. The hyperplane is enforced to go through the origin in the absence of vector 'b' because of which the resolution can be limited. This approach also involves parallel hyperplanes which is shown by the given equations:

$$w.x + b = 1 \qquad \ldots (7)$$

$$w.x + b = -1 \qquad \ldots (7)$$

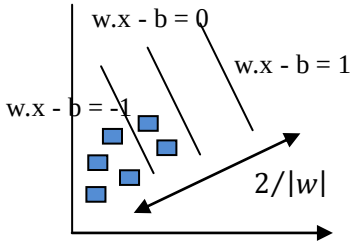


Figure 1: Maximum margin hyperplanes for a SVM trained with samples from two classes

3. Proposed Methodology

The method is projected for the plant infection discovery. The stages included in the plant infection recognition methods are given below:-

1. Pre-Processing Stage: - The picture is used in the form of input in the preliminary stage and then this image or picture is transformed into the grayscale picture.

2. Textural Feature Analysis: - In this stage, the GLCM algorithm is used for the scrutiny of input picture's texture features. In order to computer the textural features, the GLCM algorithm along with statistical texture study is implemented on the experiential amalgamation of strengths. This computation is performed with the help of intensities at exacting locations which are pertinent to one other inside a picture. There different intensity ends or pixels occur inside every arrangement are used for the classification of first-order, second-order and higher-order figures. A technique named Gray Level Co-occurrence Matrix (GLCM) is used for the extraction of second order textural features [9]. This method has been implemented in various approaches. The GLCM matrix consist various rows and columns and these are equal to the number of gray levels inside a picture. The relation frequency by which two pixels are divided through a pixel detachment is represented by (Δx, Δy) and their intensities are given in the form of *i* and *j* for a constituent and definite as$P(i, j|\Delta x, \Delta y)$. Amid the gray levels *i* and *j* at'd' dislocation detachment and (θ) angle, the matrix constituent P(i, j | d, θ) can be measured to have second order arithmetical likelihood values. In case of input picture which has G gray levels from 0 to G – 1 and M × N neighborhood, the strength at sample m is taken as f(m, n) for neighborhood line n. Therefore,

$$P(i, j|\Delta x, \Delta y) = WQ(i, j|\Delta x, \Delta y) \qquad (1)$$

Here,

$$W = \frac{1}{(M-\Delta x)(N-\Delta y)} \qquad (2)$$

$$Q(i,j|\Delta x,\Delta y) = \sum_{n=1}^{N-\Delta y}\sum_{m=1}^{M-\Delta x} A \qquad (3)$$

And

$$A = \begin{cases} 1 \; if \; (m,n) = 1 \; and \; f(m+\Delta x), n+\Delta y) = j \\ 0 \quad elsewhere \end{cases} \qquad (4)$$

Feature	**Mathematical Formulation**
Contrast	$\sum_i \sum_j (i-j)^2 g_{ij}$
Correlation	$\frac{\sum_i \sum_j (ij) g_{ij} - u_x u_y}{\sigma_x \sigma_y}$
Energy	$\sum_i \sum_j {g_{ij}}^2$
Homogeneity	$\sum_i \sum_j \frac{1}{1+(i-j)^2} g_{ij}$
Mean, M	$\sum_{i=0}^{L-1} g(i) P(g(i))$

Standard Deviation, S	$\sqrt{\sum_{i=0}^{L-1}(g(i)-M)^2 P(g(i))}$
Skewness	$\frac{1}{S^3}\sum_{i=0}^{L-1}(g(i)-M)^3 P(g(i))$
Entropy	$\sum_{i=0}^{L-1} P(g(i))\log_2 P(g(i))$
Kurtosis	$\frac{1}{S^k}\sum_{i=0}^{L-1}(g(i)-M)^k P(g(i))$
RMS	$\sqrt{\frac{1}{L*L}\sum_{i=0}^{L-1}\sum_{j=0}^{L-1}(g(i,j)-I)^2}$

TABLE I. MATHEMATICAL FORMULATION OF FEATURES [10]

3.1 Region Based Segmentation

The k-mean clustering is implemented for the area based segmentation in this stage. The stages included in the k-mean clustering algorithm are given below:

1. Input Data: - The picture used in the form of input is acknowledged as the information and requires segmentation to be applied.
2. Calculate Arithmetic mean: - The mathematical mean of the input information is computed for defining the centered point of the information.

3. Formation of Segments: - In this step, Euclidian remoteness is computed for the focal point. The points having analogous remoteness are clustered in single cluster while others are clustered in the second cluster.

4. Selection of ROI: - The segmented picture is chosen in the form of ROI and applied as input to subsequent stage of classification or categorization.

3.2 Classification of Disease

The classification phase classifies the input picture into distinct infections. In this phase, the input is utilized in the form of ROI which is chosen in the last stage. An approach named KNN classifier is used in the sample detection procedure for the classification of entities on the basis of neighboring training instances acknowledge inside the feature margin. In this example-relied learning algorithm, merely local estimation of the purpose is present and whole calculations are held upon till the completion of classification. The KNN is considered the simplest classifier amid all the other machine learning algorithms. An entity is categorized with the help of bulk vote of associates [12]. The uncommon class amid its k nearest associates is identified in the form of entity. This entity is allocated to the class of its adjacent neighbor in case when k=1 in a direct manner. The NN approach is applied for both categorization and regression prognostic concerns. During the evaluation of any methods several significant features like easiness in the output interpreting, time calculation and the accessible prognostic control are acknowledged. The $knn( )$ function utilizes the Euclidean distance delinquency. The equation utilized for the calculation of Euclidean distance is given below:

$$D(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 \ldots . + (p_n - q_n)^2} \quad \ldots . (8)$$

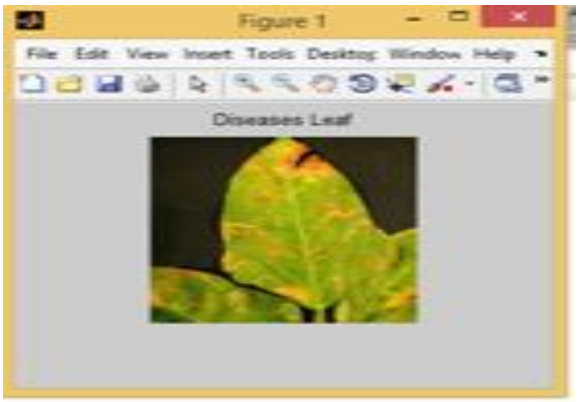
In this equation, the subjects having ‘n’ properties is compared are represented in the form of ‘p’ and ‘q’.

The selected numbers of neighbors are utilized for the calculation of KNN. These neighbors are selected by the parameter ‘k’. The recital of KNN algorithm is influenced on huge level through the selection of the k value. The maximization of k value reduces the variance produced because of the arbitrary fault. The suitable selection of k value is used for the maintenance of balance amid over-fitting and under-fitting. The infection name is given by the output of the KNN classifier.

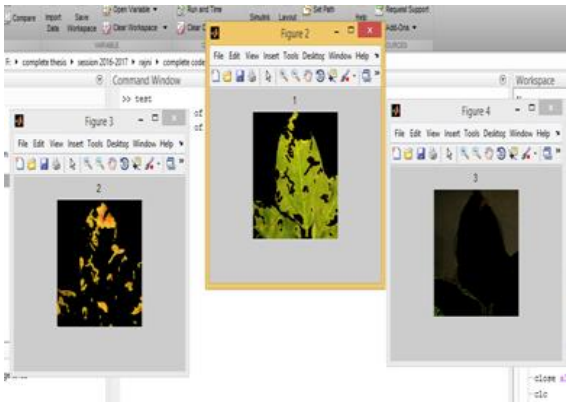
4. Results and Evaluation

With the help of socially available data sample named as "Plant Village", a folder of approximately 40 descriptions of potato leaves is collected and utilized for the evaluation of projected algorithm. The database or folder utilizes pictures of 10 normal leaves and 30 infectious leaves for performing the research. The whole record is divided into two dissimilar sets throughout the tests. The training set comprises total 24 images (60%) and testing set comprises total 16 images (40%). The classification is executed with the help of KNN classifier that comprises linear Kernel. The performance of the projected classifier is evaluated with respect to some parameters like correctness, compassion, recall, and F1-score. The proposed approach shows the precision rate of approximately 97% for categorization.

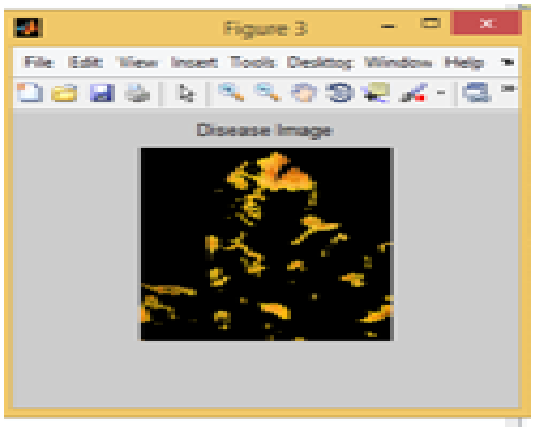
The implemented performance measures are represented by the table II. The assessments are performed amid potato and other class only since there is not much task projected on infection discovery of potatoes. The table II represents the performance of KNN classification approach.



a)Input Image



b) Segmented Image



c. Diseased Image

As shown in figure 1 a) is the key picture for plant infection discovery b) Execute k-mean region relied segmentation c) Contaminated picture

Class	Precision (%)	Recall (%)	F1-Score (%)
0: Leaf Minar	89	94	92
1: Mosnic virus	97	93	95
2: non-diseased	98	98	98
Average/total	97	97	97

TABLE II: PERFORMANCE MEASURES OF CLASSIFICATION

5. Conclusion

This investigative study concludes that infections or diseases occurring in the plants can be recognized with the help of plant disease discovery technique. This research study utilizes GLCM algorithm for the scrutiny of texture features. The region relied segmentation is performed with the help of k-mean clustering approach and infections are predicted with the help of KNN classification model. The MATLAB programming is used for the simulation of the projected approach and figures and tables are used for the demonstration of outcomes. The reproduction outcomes depicted that the proposed approach showed precision rate of approximate 97 percent for infection forecasting.

References

[1] Aakanksha Rastogi, Ritika Arora, Shanu Sharma, "Leaf Disease Detection and Grading using Computer Vision Technology &Fuzzy Logic", 2015 2nd International Conference on Signal Processing and Integrated Networks (SPIN).

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

[3] Harshal Waghmare, Radha Kokare, Yogesh Dandawate, "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).

[4] R. Meena Prakash, G. P. Saraswathy, G. Ramalakshmi, "Detection of Leaf Diseases and Classification using Digital Image Processing", 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS).

[5] Kawaljit kaur, Chetan Marwaha, "Analysis of Diseases in Fruits using Image Processing Techniques", International Conference on Trends in Electronics and Informatics ICEI 2017.

[6] Chaitali G. Dhaware, Mrs. K.H. Wanjale, "A Modern Approach for Plant Leaf Disease Classification which Depends on Leaf Image Processing", 2017 International Conference on Computer Communication and Informatics (ICCCI -2017).

[7] Pranjali B. Padol, Prof. Anjali A. Yadav, "SVM Classifier Based Grape Leaf Disease Detection", 2016 Conference on Advances in Signal Processing (CASP).

[8] Preetha Rajan, Radhakrishnan. B, Dr. L. Padma Suresh, "Detection And Classification Of Pests From Crop Images Using Support Vector Machine", 2016 International Conference on Emerging Technological Trends [ICETT].

[9] Asad Parvez, Anuradha C. Phadke, "Efficient Implementation of GLCM based Texture Feature Computation using CUDA Platform", International Conference on Trends in Electronics and Informatics, ICEI 2017.

[10] Monzurul Islam, Anh Dinh, Khan Wahid, "Detection of Potato Diseases Using Image Segmentation and Multiclass Support Vector Machine", 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE).

[11] Xiaomu Duan, Tong Zhao, Tan Li, Jinxin Liu, Liang Zou, Li Zhang, "Method for diagnosis of on-load tap changer based on wavelet theory and support vector machine", 2017, The 6th International Conference on Renewable Power Generation (RPG).

[12] Zhongheng Zhang, "Introduction to machine learning: k-nearest neighbors", Ann Transl Med 2016, 4(11):218.

[13] Hu YH, Ping XW, Xu MZ, Shan WX, He Y, "Detection of Late Blight Disease on Potato Leaves Using Hyperspectral Imaging Technique," PubMed, 36(2), 2016, pp. 515-519.

