Severity Identification of Potato Late Blight Disease from Crop Images Captured under Uncontrolled Environment

Sandika Biswas*, Bhushan Jagyasi*, Bir Pal Singh† and Mehi Lal‡
*TCS Innovation Labs Mumbai
Tata Consultancy Services, Yantra Park, Thane, India
Email: {biswas.sandika, bhushan.jagyasi}@tcs.com
†Central Potato Research Institute (CPRI), Simla, India
Email: directorcpri@gmail.com
‡Central Potato Research Institute Campus (CPRIC), Modipuram, Meerut, India
Email: mehilalonline@gmail.com

Abstract—**Plant disease management is an important factor in agriculture as it causes a significant yield loss in crops. Late Blight is the most devastating disease for Potato in most of the potato growing regions in the world. For optimum use of pesticide and to minimize the yield loss, the identification of disease severity is essential. The key contribution here is an algorithm to determine the severity of Potato Late Blight disease using image processing techniques and neural network. The proposed system takes images of a group of potato leaves with complex background as input which are captured under uncontrolled environment. In this proposed approach decorrelation stretching is used to enhance the color differences in the input images. Then Fuzzy C-mean clustering is applied to segment the disease affected area which also include background with same color characteristics. Finally we propose to use the neural network based approach to classify the disease affected regions from the similar color textured background. The proposed algorithm achieves an accuracy of 93% for 27 images captured in different light condition, from different distances and at different orientations along with complex background.**

Keywords—Leaf diseases, Enhancement, Segmentation, Fuzzy c-mean clustering, Neural Network

I. Introduction

Potato is one of the most highly consumed crops over the entire world. The late blight is the most important disease for potato, which is caused by a fungus-like organism, *Phytophthora infestans* in humid conditions with temperature within a specific range. This disease can destroy crops of the entire field, causing substantial yield loss, within a few days after the first lesion appears in the farm [1]. Therefore, early detection of this disease attack is essential for prevention of huge economic losses. The early Detection of disease requires continuous monitoring of the disease severity by an expert which is time consuming, tedious and some times not feasible due to a small number of experts and a larger number of farms. So, a fast, reliable, and automatic method is required to accurately quantify the disease severity.

In recent literature, the image processing techniques are being widely and efficiently used in agricultural field for disease detection and severity identification [2], [3]. Revathi et al. [4] proposed a Homogeneous Pixel Counting technique for Cotton Diseases Detection (HPCCDD) using leaf spot images captured by farmers through mobile. They used edge features of segmented disease spots extracted with homogenize edge operators like Sobel and Canny filters to classify the target regions and also provided a comparison based study for different classifiers and declared K-mean and Artificial Neural Networks as best classifiers with which they achieved 94% accuracy. Bashish et al. [2] proposed an image processing based framework for detection of plant leaf/stem diseases. They segmented the affected area using K-Means clustering technique and used GCLM features for classification among different cotton leaves diseases and healthy leaf using Back Propagation Neural Network. Revathi et al. [4] and Bashish et al. [2] considered single leaf images to obtain the disease severity which was specific to a particular leaf only.

Zulfikli et al. [5] and Sannakki et al. [6] considered images of cluster of leaves with complex background for chili and grape crops. Zulkifli et. al. [5] proposed an effective way for early detection of diseases in chili plants through leaf features inspection. They performed color clustering to extract the leaves from background and used color features to distinguish between healthy and diseased chili plant. Sannakki et.al. [6] proposed an image processing technique for grape disease identification and its severity detection from images captured with complex background. They applied anisotropic diffusion for noise removal after masking green pixels and used K-mean clustering for grape leaf disease segmentation from the healthy region of the plant. Although Zulkifli et al. and Sannakki et al. considered images with complex background, for these crops the damaged portion of the leaves have different color features from the background and hence it becomes easy to segregate.

In this paper, we use the images of cluster of leaves of potato crop, captured from farms under uncontrolled environment, to identify the disease severity due to Late Blight disease. Since the images used are captured from the farms under uncontrolled environment there is no prerequisite training provided to the farmers to capture the images. The key challenge for computing severity of Late Blight disease of Potato crop is to segment the disease affected vegetation

2014 IEEE Canada International Humanitarian Technology Conference - (IHTC)
978-1-4799-3996-1/14/$31.00 ©2014 IEEE

region mainly due to light brown lesions which have the same color features as of the soil appearing in the background. The main contribution of this paper is the combination of Fuzzy c-mean clustering and Neural Network based approach for achieving classifier of diseased portions of the leaves. The disease severity has been quantified in terms of total affected vegetation area in the image with cluster of leaves.

We envision to provide personalized advice to farmers using the mKRISHI® platform [7] coupled with information about disease severity computed using the proposed approach. The novel sensor network and human participatory sensing framework [1] will assist farmers to capture and send the images of their crops using participatory sensing mobile application. The optimum use of pesticides can hence be recommended by taking into consideration *disease risk* obtained from models (as discussed in [1]) and *disease severity* as computed using this novel image processing algorithm. The system also does not need any special training to the farmers as images considered here are captured at different angles, from different distances and in different light conditions which makes it more appealing and pragmatic in the given scenario.

II. MATERIAL AND METHOD

A. *Material*

In this paper, our aim is to obtain the severity of Late Blight disease from the images captured and sent by the farmers from the potato farms at various locations. The images captured by untrained farmers are not oriented and contain cluster of leaves with background visible in several segments. The estimation of the total infected areas in several segments would provide information about disease severity in their farms. The input images used in this paper are collected from the Central Potato Research Institute (CPRI) and its nearby farms of Meerut, Uttar Pradesh, in India. Images are captured using 12 mega pixel digital camera leading to the image size of 3000×4000 pixels stored in standard jpeg format. In order to make the algorithm computationally less exhaustive, the images have been down sampled in the preprocessing phase discussed in the subsequent part.

B. *Method*

In this paper, for obtaining the disease affected areas of leaves, the input images with complex background are sequentially taken through several processing steps. Fig.2 presents the proposed method which mainly consists of Preprocessing, Clustering, Cluster selection and Extraction of affected leaf area.

1) Preprocessing: The input images are of resolution 3000×4000 pixels. First the images are down sampled to its one eighth size in both height and width of the image to reduce the time complexity of the algorithm. Images are resized using bi-cubic interpolation method to preserve the image quality.

In the next step, the down sampled images are enhanced using de-correlation stretching to exaggerate the differences between different color bands of the image. Input images show highly correlated color channels as shown in Fig.3(a). Conventional enhancement of color images are based on independent contrast modification or stretches of three color channels R, G and B. But for highly correlated color channels conventional stretching is not an effective approach. De-correlation stretching is more suitable enhancement approach for these kind of images [8]. De-correlation stretching is a pixel-wise operation where color values in the enhanced image are defined as in equation (1),

$$I_o = Cov_i \times (I_i - m) + m\_target. \quad (1)$$

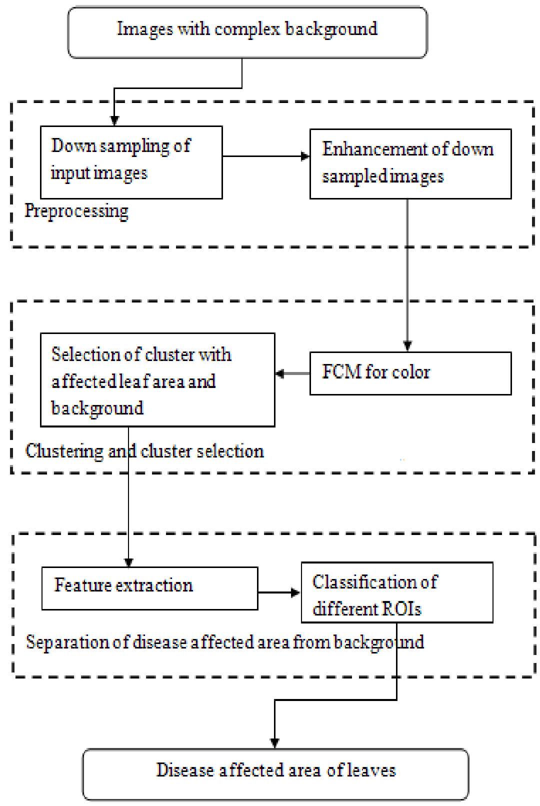
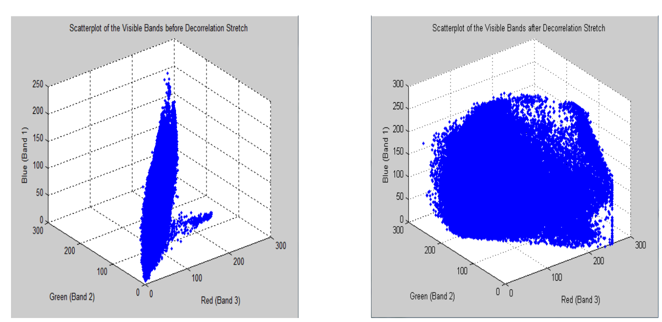


Fig. 1: Overall work flow of the proposed method



(a) Before enhancement **(b)** After enhancement

Fig. 2: Correlation between R, G and B bands of input and processed images

Where, I_i and I_o are 3×1 vectors having R, G and B values in input and enhanced images, respectively. m and m_target are also 3×1 vectors, which contain the mean of each band in the input image and the desired output mean in the output image respectively. In this experiment, m_target is kept same as m. Cov_i is a 3×3 matrix which specifies the band-to-band sample covariance of the image. Fig.3(b) shows color band scatter plot after de-correlation stretching of the images.

After decorrelation stretching, we apply device-independent color space transformation, which converts the RGB color values in the enhanced image to CIELAB color values. The Lab color space is device independent color space, which makes it easier to quantify the visual differences in colors present in a color image. In the following subsection, the enhanced image in CIELAB color space will be used for further processing.

2) Clustering and cluster selection: Clustering for color segmentation: Clustering is used for grouping of different regions or object in an image based on similarity in some properties so that objects or regions in the same cluster or group are of similar characteristics. In our problem, we aim to separate disease infected regions from the rest of the image. Fuzzy c-mean clustering is a common and most effective approach in image processing to separate out different objects in an image [9]. So, we have used FCM to separate the affected area from the healthy portion of the leaf. FCM partitions each image into three clusters among which one cluster contains the disease affected area along with background.

In the CIELAB color space, color information exists mainly in a and b color channels, so, a and b values are used for further processing. These two values are represented in a $M \times 2$ matrix, where M is the total number of pixels in the image. So, the input dataset X is defined as $M \times 2$ matrix, where each ith row is a sample point x_i. The modified fuzzy c-mean algorithm proposed by Bezdek et al. [10], [11] is composed of the following steps:

1) Image pixels x_1,...x_M are taken as input.
2) In output each image pixel is assigned to one particular cluster.
3) U=[u_{ij}] matrix, where u_{ij} is the degree of membership for the i^{th} sample of the input dataset x_i in j^{th} cluster. For the first iteration u_{ij} value are randomly initialized.
4) At l^{th} iteration, center vector $C(l)$=[c_j] is calculated from $U(l)$, where c_j is defined as ,

$$c_j = \frac{\sum_{i=1}^{M} u_{ij}^q x_i}{\sum_{i=1}^{M} u_{ij}^q} \tag{2}$$

Where, c_j is the j^{th} cluster and $1 \leq j \leq N_c$, where N_c is the number of clusters, m is exponent for the membership function matrix U which is selected as defined in Table 1.
5) In the next step $U(l)$ and $U(l+1)$ values are updated by equation (3) for the updated value of c_j.

$$u_{ij} = \frac{1}{\sum_{k=1}^{N_c} \left(\frac{||x_i - c_j||}{||x_i - c_k||} \right)^{\frac{2}{(q-1)}}} \tag{3}$$

Where,$1 \leq j \leq N_c$ and $1 \leq i \leq M$.

Parameters	**values**
Number of clusters (N_c)	3
Exponent for membership function matrix (q)	2.0
Maximum number of iterations	100
Minimum amount of improvement (ϵ)	1e-5

TABLE I: Different parameters used for Fuzzy-c-mean clustering

6) If $||U(l+1) - U(l)|| < \epsilon$ then the process is stopped; otherwise return to step 3, where ϵ is a very small value, which is defined as mentioned in Table 1 for this experiment.

Parameters used in fuzzy c-mean clustering for the proposed approach are specified in Table 1. Fig. 4(b), (c) and (d) represents the three clusters obtained from FCM method when applied to input image shown in Fig.4(a).

Selection of cluster of interest: Potato late blight disease causes black or dark brown lesions on the leaves and stems. The cluster with affected area is selected based on hue information of the clustered images. Each cluster is transformed from RGB color space to HSV color space. The cluster C_s with disease affected area, is selected using the following equations:

$$N_p(i) = \sum_{k=H_l}^{H_h} hist_i(j). \tag{4}$$

$$C_s = arg\{max(N_p(i))\}, \forall j : N_p(j) < N_p(i) \tag{5}$$

$1 \leq i \leq N_c$ and $1 \leq j \leq N_c$

Where, $hist_i$ is the histogram of hue component of i^{th} cluster. H_h and H_l represent upper and lower limit of hue for the color feature of late blight affected area of leaves, respectively. $N_p(i)$ represents total number of pixels in the specified hue range for ith cluster. In the next step, most green pixels are removed from the cluster of interest to extract the only affected areas. The pixels of healthier regions are removed from the cluster based on R and G values of each pixel. The pixels with R<G are retained in the final cluster image and for rest of the pixels all of the R, G and B channels are set to zero. From the three clusters Fig. 4(b), (c) and (d), the cluster that is selected having affected leaf part is shown in Fig.4(e).

3) Selection of affected leaf area from background: Using fuzzy-c-mean clustering we get a cluster having both the affected leaf part and background regions [Fig.3(d)]. Hence there is a need to separate the affected vegetation part from background. We observed that, though affected leaf area and background have same color features, they possess different texture characteristics. We proposed to use neural network to discriminate between affected vegetation part and background with same color texture. Here, we take each connected regions (ROIs) of C_s to classify as, whether it belongs to infected vegetation part or background.

Feature extraction: In this step, texture features are extracted from each connected components. Gray level co-occurrence matrix of hue component of the ROIs in HSV color space, is used to extract the texture features. Seven texture features: contrast, uniformity, maximum probability, homogeneity, inverse difference, difference variance, diagonal

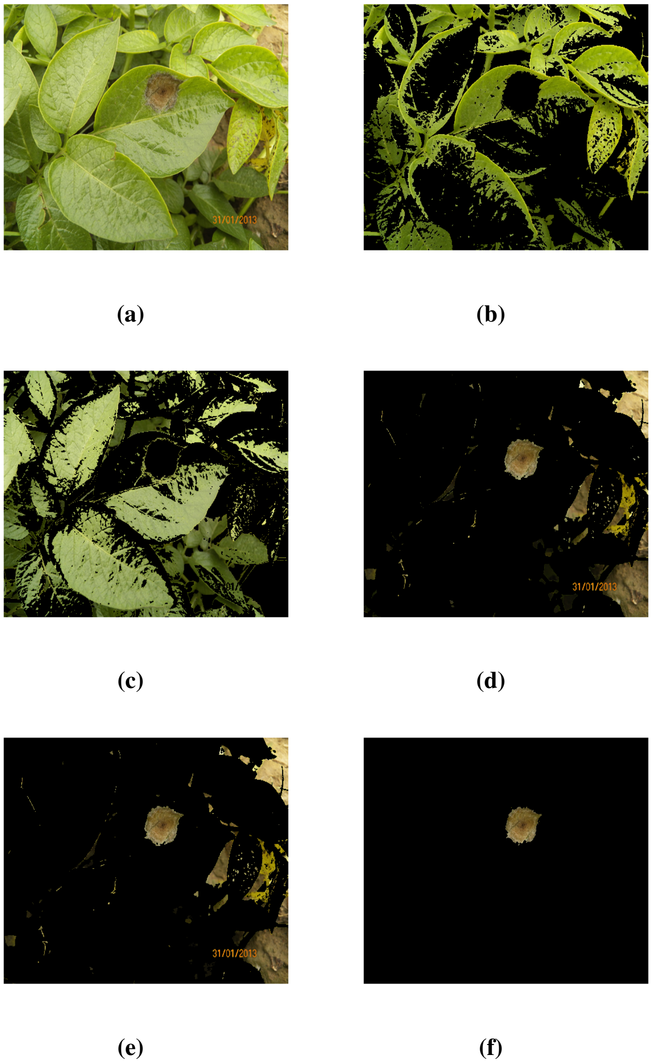


Fig. 3: (a) Original image, (b), (c), (d) Different clusters (e) Selected cluster with disease affected area, (f) Final cluster classified as affected vegetation part

variance [6] are used for classification using neural network based classifier.

Classification: Back propagation neural network with three layers having one input, one hidden and one output layer, is considered in the proposed method to classify the ROIs. Input layer has seven nodes for seven texture features and hidden layer consists of 15 nodes. The separated affected part using neural network approach is shown in Fig.4(f).

III. Results and Discussion

The dataset used for performance analysis of the proposed algorithm consists of 27 images. After applying FCM, for these 27 images we get total 340 connected regions (ROIs) including affected vegetation and background. Within total 340 ROIs, 70% of the ROIs are used to train the neural network, 15% are used for validation purpose and rest 15% are used for testing purpose. Training samples are selected randomly from the dataset. Using Back Propagation Neural Network

Phase	No. of samples	Accuracy
Training phase	238 (70%)	100%
Validation phase	51 (15%)	100%
Testing phase	51 (15%)	100%

TABLE II: Performance of back propagation neural network

we achieve 100% classification accuracy. Here, in this paper classification accuracy indicates the correct classification of disease affected area and background. Table 2 shows accuracy of the classifier during training, validation and testing phase.

Performance of the algorithm is also measured in terms of accuracy of the affected leaf area detection. Accuracy for i^{th} image is defined by equation (6).

$$Accuracy_i = (TP_i + 1 - FP_i)/2 \tag{6}$$

$$TP_i = \frac{\sum_{j=1}^{N_a} tp_i(j)}{\sum_{j=1}^{N_e} tp\_ground\_truth_i(j)} \tag{7}$$

where, TP_i and FP_i are true positive and false positive probabilities, respectively, for i^{th} image. N_a is total number of affected areas of leaf detected by the algorithm and N_e is total number of affected areas of leaf marked by expert. tp_i and $tp\_ground\_truth_i$ represents area of each affected region of leaves detected by algorithm and marked by experts respectively.

$$FP_i = \frac{\sum_{j=1}^{F_a} fp_i(j)}{\sum_{j=1}^{F_e} fp\_initial_i(j)} \tag{8}$$

where, F_a is total number of falsely detected connected regions by the algorithm at final phase and F_e is total number of falsely detected connected regions by the algorithm at final phase. fp_i and $fp\_initial_i$ represents area of each affected region detected by algorithm and marked by experts respectively.

In terms of affected vegetation area detection, the proposed method achieves accuracy of 93% averaged over the 27 images. Hence the proposed method achieves very high lesion detection accuracy and very high accuracy for the identification of the area of the lesions.

IV. Conclusion

In this paper an FCM clustering and neural network classification based approach is proposed to detect and quantify the severity for late blight disease of potato. Images of disease affected leaves are collected under uncontrolled environment to make the algorithm more robust and image background independent. The algorithm consists of mainly two steps: (a) Fuzzy c-mean clustering to separate the disease affected area along with background (b) to extract affected leaf area from background using neural network. The proposed approach achieves a very high accuracy in terms computation of the disease affected area which in turn results in accurate disease severity identification.

ACKNOWLEDGMENT

The authors would like to thank to Central Potato Research Institute for providing image dataset collected by their scientists which has been used in this experiments.

REFERENCES

[1] A. Pande, B. G. Jagyasi, and R. Choudhuri, "Late blight forecast using mobile phone based agro advisory system," *in Proc. third Internation Conference on Pattern Recognition and Machine Intelligence, PReMI 2009*, Dec 2009.

[2] D. Al Bashish, M. Braik, and S. Bani-Ahmad, "A framework for detection and classification of plant leaf and stem diseases," in *Signal and Image Processing (ICSIP), 2010 International Conference on*, pp. 113–118, IEEE, 2010.

[3] J. K. Patil and R. Kumar, "Advances in image processing for detection of plant diseases," *Journal of Advanced Bioinformatics Applications and Research*, vol. 2, no. 2, pp. 135–141, 2011.

[4] P. Revathi and M. Hemalatha, "Classification of cotton leaf spot diseases using image processing edge detection techniques," in *Emerging Trends in Science, Engineering and Technology (INCOSET), 2012 International Conference on*, pp. 169–173, IEEE, 2012.

[5] H. Zulkifli, A. Yeon, M. Shakaff, A. A. Abdul Halis, and S. Rohani, "Feasibility study on plant chili disease detection using image processing techniques," 2012.

[6] S. Sannakki, V. Rajpurohit, V. Nargund, and P. Kulkarni, "Diagnosis and classification of grape leaf diseases using neural networks," in *Computing, Communications and Networking Technologies (ICCCNT),2013 Fourth International Conference on*, pp. 1–5, July 2013.

[7] A. Pande, B. G. Jagyasi, S. Kimbahune, P. Doke, A. Mittal, D. Singh, and R. Jain, "Mobile phone based agro-advisory system for agricultural challenges in rural india," *IEEE Conference on Technology for Humanitarian Challenges*, Aug 2009.

[8] A. R. Gillespie, A. B. Kahle, and R. E. Walker, "Color enhancement of highly correlated images. i. decorrelation and hsi contrast stretches," *Remote Sensing of Environment*, vol. 20, no. 3, pp. 209–235, 1986.

[9] H. Le Capitaine, C. Frélicot, *et al.*, "A fast fuzzy c-means algorithm for color image segmentation," *Proceedings of European Society for Fuzzy Logic and Technology (EUSFLAT'2011)*, pp. 1074–1081, 2011.

[10] J. C. Bezdek, L. Hall, L. Clarke, *et al.*, "Review of mr image segmentation techniques using pattern recognition," *MEDICAL PHYSICS-LANCASTER PA-*, vol. 20, pp. 1033–1033, 1993.

[11] J. C. Bezdek, *Pattern recognition with fuzzy objective function algorithms*. Kluwer Academic Publishers, 1981.