A QUICK AND ACCURATE TOMATO LEAF DISEASE DISCOVERY AT EARLIER STAGE OF HARVESTING BY UTILIZING THRESHOLD SEGMENTATION AND RFO CLASSIFIER

1MR.G.BALRAM, 2DR.K.KIRAN KUMAR
1Research Scholar, Department of CSE, KLEF, Vaddeswaram, A.P.
2Professor, Department of CSE, KLEF, Vaddeswaram, A.P.
E-mail: 1balaramcse@cvsr.ac.in, 2kiran5434@kluniversity.in

ABSTRACT

In Agriculture, the prior period of harvesting feature gives splendid effectiveness. It lessens the collect disease rate with the ultimate objective that the economy has been balanced. The import and toll of agro things are extends because of the previous period of yield gathering technique. The various conventional methods can deal with a past development of harvesting, anyway improvement is required. Hence random forest optimization and threshold segmentation schemes have used to find the diseases in the harvests, moreover predicts the sort of plant-infirmity. In this investigation, the continuous prior period of tomato leaf illness and its contrasting compost is proposed. For this tomato crop leaf pictures has been assembled from various data bases like reap science, yes-modes, nelson. wisc datasets. Dependent upon singular thought, plants give the Rise in progress, yet by using the manual methods for cultivating, more creation rate is past the domain of creative mind. So IoT with front line Machine Learning methodologies and huge making sure about pre-processing techniques can uphold the sound cultivation. Until various procedures are delivered for agribusiness at a previous period of social occasion, anyway they have more imperatives. Current development doesn't work with past procedures; thusly, improvement is required for future sound agribusiness systems. In this investigation, threshold-based segmentation is used for pre-processing, and Random Forest optimization for classification of tomato leaf disease detection at prior stage. Proposed threshold segmentation RFO (TS-RFO) gives the 97.6% detection accuracy and 99% True certain rate 59.82PSNR, 0.9989SSIM, 0.0081MSE has been gotten; this is a splendid achievement stood out from existed systems.

Keywords: Crop Harvesting, Detection Of Plant Disease RFO, Classification, Threshold Segmentation

1. INTRODUCTION

The prior stage of crop harvesting is an important agriculture change for a future phase of production. The agricultural analyzer thinks that disease prediction plants are a significant role for farmers. This work can help the farm productivity at any time-varying nature of harvestings. In India most of the income rates depending on agriculture productivity; therefore, the prior stage of disease deduction in plants place a significant role. To predict plant disorders at an initial step is a better disease detection mechanism compared to conventional methods. The exited methods for plant disease detection is a natural observation by manual processes. This model can’t identify the exact problems of leaf disease. For this, a large amount of economic and human resources as needed for monitoring the plants. Some of the country’s formers had utilized the old conventional methods with expert suggestions. Due to this time has been consuming for evaluating the disease type and corresponding fertilizer. Plant infection expectation by visual nature is a more convoluted undertaking and a similar less proficient model. This model has a greater number of limitations, whereas auto-machine techniques can easily handle the above-discussed problems. They proposed in this research of tomato plant, disease identification method at a prior stage of harvesting. Various diseases related to tomato plants are listed as bacterial _spot, early_blight, late_blight, leaf_mold, septoria_leaf_spot, spider_mites, target_spot and yellow_leaf_curl_virus. These diseases may be observed in the tomato plant. Image processing and embedded system combination can help the
prior stage disease of tomato plant detection system.

Figure 1 explains that a general PCB diagram of Raspberry pi and pi cam devise. These two devices are identifying the tomato plant leaf diseases. Image processing is a significant analysis for determining the effected area of plant disease. The image processing further divided into steps, and this is image acquisition and segmentation. Image acquisition also classified into three sub-stages those are clipping, smoothing and enhancement. Coming to segmentation three more sub-stages are present, there are binarization, thresholding and segmentation. The above two main steps can quickly identify the tomato leaf disease. Image segmentation is one of the best procedure differentiation the image parts. In image segmentation, many models are performing the segmentation process; in that threshold segmentation method has been used in our approach.

The IoT can perform more application such as Agriculture, industrial and naval. In any field, IoT and machine learning tools might be providing excellent optimization outcomes, and modern approaches identify the crop diseases and corresponding fertilizers. Yet, these are less proficient, so profound learning and AI calculations have been used to improve the pace of acknowledgment and precision. Various agents have involved AI enhancement for plant illness ID and grouping. SVM- support vector machine, nural networks, FCNN, CNN, fuzzy logic and RFO.

2 RELATED METHODS

Various learning methods classify plants disease detection and classification. They can also categorize the plant disease type based on rank and accuracy. RGB to Gray conversion, histogram equalization, sensor calibration and segmentation are the various methods for identifying the crop disease. The nearest neighbours, DT- decision tree and GWO- Gray wolf optimization are the methods that can regulate crop monitoring. The below table 1.1 explains about various surveys of agriculture concerning IoT.
<table>
<thead>
<tr>
<th>ARTICLE NO</th>
<th>SURVEY ARTICLE METHODOLOGY OF COMMUNICATION</th>
<th>TECHNIQUE</th>
<th>KEY POINTS</th>
</tr>
</thead>
</table>
Leaf image  
Histogram equalization |
Image filtering |
| [3]       | Crop disease using pattern recognition | K-mean clustering | Crop disease based on AI  
Gaussian filtering  
Machine learning |
| [4]       | Unhealthy plant leaves detection | Image texture features | Discovery of harmful crop leaves detection  
Disease classification regression |
| [5]       | Image processing technique on plant disease detection | Conventional methods | Image acquisition  
Segmentation  
Histogram equalization |
| [6]       | Remote area plant Detection | Image training techniques | Plant health detection  
Fertilizers suggestion |
| [7]       | Advanced crop maintenance system | Machine learning | Image processing techniques  
on prior stage of plants |
| [8]       | Disease stage measurement | Geometrical tools | Leaf disease with image acquisition  
Segmentation mechanism |
| [9]       | Disease prevention | RGB colour transformation | Spot detection discrete wavelet transformation |
| [10]      | Image processing techniques with IoT | Leaf disease estimation | Region of interest disease detection classification |
Classification |
| [12]      | IoT agriculture | Adaptive image segmentation | Genetic algorithm image segmentation  
Disease classification |
| [15]      | Disease prediction | Artificial neural networks | Mechanical characteristics of agriculture sensor analysis |
The multi genetic algorithms for agriculture disease predictions and soil water retention characteristics are analyzed by using computational image processing techniques [16]. An integrated genetic program-based agriculture disorders estimation based on artificial intelligence technique can be applied on various datasets [17]. Estimation of agriculture mechanical properties and its functionality is explained with artificial intelligence methods in [18]. Medical image processing and computational parameters are used to debug the agriculture IoT methods of various crop datasets [19 20]. The above all implemented methods have more limitations and along with necessary improvements also needed. So, we are moving to advanced machine learning based tomato leaf disease identification system using threshold segmentation. [21,22]

3. METHODOLOGY

In this technique mainly random forest optimization (RFO) is applied for the classification stage, threshold-based segmentation has implemented for the preprocessing stage. The whole work can be partitioned into two stages one for preparing stage and the subsequent one is the trying stage. In this work, a tomato crop leaves are collected from various databases which are discussed above. Selected tomato leaves are used to identify the disease and corresponding fertilizers, this work can help for formers, import and exports of agro products etc. The brief explanation of tomato plant disease detection and classification is explained clearly in the below sections.

a) Pre-processing
b) Classification

3.1 Pre-processing At preprocessing, stage selected database images are composed for segmentation and image acquisition. Various segmentation methods are available but those have limited number of applications. So, threshold based segmentation mechanism is selected for training the tomato leaf images.

Figure 2: Tomato leaf images. a) Leaf image 1, b) leaf image 2.

Figure 2 explains about tomato leaves which are collected from available datasets. These leaves are healthy, and no disease can be identified like various leaves are collected for an experiment to identify the disease
Figure 3 Main Block Diagram

i) Image clipping

ii) Image smoothening

iii) Image enhancement

In image clipping input images has been clipped with the assistance of PCA (Principle Component Analysis) by utilizing back projection lost pixels technique (BPLPT) with the assistance of this strategy can personality the spots in the leaves. It can verify the background of crop pre-harvesting images and removes the waste layers. The clipped images are variable in nature and individually locating the disease spot by using below function that is shown in equation 1.

\[
clip_{img} = f(\text{row: row50 - 1, col: col + 70 - 1}) \quad (1)
\]

Because of this localization function, background pixels are lost, separate the main image. and background image. In the second step image smoothening has been performed with help of low pass filtering (LPF) by removing the spatial noise from selected crop image. This LPF usually adjusting the window operator, which can affect every pixel of the selected image at a time. The image varying nature and its characteristics values has been adjusted by particular function of a local region show in equation 2.

\[
K_{avj} = \frac{\text{low_filter}(f \text{ special('avj', 4),} J)}{255} \quad (2)
\]

In the third step, picture upgrade has been performed by utilizing histogram equalization technique. In this, the sharpness of the image and mast correlation adjustment parameters are implemented using the below mathematical computations.

\[
J = \text{histeq}(I, hgram)
\]
Journal of Theoretical and Applied Information Technology
© 2005 – ongoing JATIT & LLS
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The above all mathematical equations are train the selected tomato leaf and identify the particular disease clearly. But necessary improvements are needed this can possible by powerful segmentation mechanism.[23] In this scenario, threshold-based image segmentation has been implemented for accurate and efficient tomato plant leaf disease detection. The equation 4&5 are performed on tomato leaf image after the acquisition process. Because of segmentation clear effected area of disease is located the region of interest ROI can handle the infected diameter of the disease.

Algorithm 1: Threshold Segmentation

Step 1 : Input Image
Step 2 : Image clipping
Step 3 : Image histogram
Step 4 : Image smoothening
Step 5 : Segmentation

Using the below formulas segmentation has been performed.

\[
T_{xy} = a \sigma_{xy} + b m_{xy}
\]

\[
f(x, y) = \begin{cases} 
1, & \text{if } f(x, y) > T \\
0, & \text{if } f(x, y) \leq T
\end{cases}
\] ---- (4)

\[
g(x, y) = \begin{cases} 
a, & \text{if } f(x, y) > T_2 \\
b, & \text{if } T_1 < f(x, y) \leq T_2 \\
c, & \text{if } f(x, y) > T_1
\end{cases}
\] -- (5)

Here \(g(x, y)\) explains the threshold segmentation method. In this first threshold value has been fixed with the following technique. In the threshold value calculation first image height and width has been fixed. E.g 256 X 256, in the second step RGB pixel values is to be fix by using get r();, get g();, getb(); these three commands gives the total number of red, green, blue number of pixel values, average the red, blue, green pixels after that mean of RGB(); gives the threshold value \(T\). Using this threshold value differentiate the pixel tends to either may be a background or object.

3.2 Classification of Tomato Crop Disease

Tomato Crop images are selected from the database which is presented in Matlab 2015b local server. The selected harvest images related to various tomato leaf disease; every set consists of 15 samples; these are required for both training and testing. 75% of images are using for training the remaining 25% used for testing. The below block diagram figure 3 explains that preprocessing and classification of pre-crop harvesting disease recognition method. In the first step, crop images are selected after that pre-processing and classification has been performed again classification is divided into two steps that are prediction and description. The prediction process classifies the disease or regression has occurred. In the classifier with the help of logistic regression, support vector machine, k-nearest neighbour, naive Bayes methods are performed and at final proposed TSRFO compared with the existed method.
The final computational procedure can calculate the accuracy, overall accuracy average, type of diseases like bacterial leaf spot, leaf frogeye spot, sunburn disease, fungal disease and early scorch disease. According to disease nature fertilizers can apply on the infected leaf. The quality assessment parameters are MCC, accuracy, F1 score, sensitivity and specificity are calculated for identify the efficiency of system.

**Algorithm 2: RFO Random Forest Optimization**

1. Input dataset
2. Set No. of Classes = N
3. No. of Features = M
4. Determine No. of features by nodes m
5. Decision Tree = m<M
6. Select decision tree by random manner
7. Calculate the split data by using for loop.
8. End for

Using below formulas this RFO has been performed.

\[
v_{TSRFO} = \arg \max \left[ \sum_{i=1}^{m} P(x_i, v_i) \prod_{j=1, j \neq i}^n P(x_j|x_i v_i) \right]
\]

\[
W_i = \frac{\text{GainRatio}(A_i)}{\sum_{i=1}^{m} \text{GainRatio}(A_i)}
\]

\[
W_i = \sqrt{\text{indexrank}((A_i)+1)}
\]

\[
v_{DTWTSRFO} = \arg \max \left[ \sum_{i=1}^{m} P(x_i, v_i) \prod_{j=1, j \neq i}^n P(x_j|x_i v_i) \right]^{W_j}
\]
All mathematical steps of RFO optimization have been performed by using shown in above figure 4. In this, each step calculates the train data weights, tree establishment and growth of trees information are calculated. At the final stage, estimate the disease of tomato leaf and the error probability, the classification of the disease is followed by below figure 5.[24], with limited memory calculate aggregate value [25,26,27] are also used for improve the performance.

\[
\begin{align*}
X_1 &< 1.3 \\
X_2 &< 1.0
\end{align*}
\]

**Figure 5: weight calibration for the disease.**

Fig5 explains that normalization or disease classification graphical step based on weight form sensing RFO optimization method, which is discussed in the above section. Finally, parameters are calculated for quality assessment checking.
3.4 QUALITY ASSESSMENT

MCC (Matthews correlation coefficient, True positive rate, true false rate PSNR, MSE, SSIM all parameters are to be analysed with the help of below mathematical approach.

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP) \cdot (TP + FN) \cdot (TN + FP) \cdot (TN + FN)}}$$

(MCC: worst value = -1; best value = +1)

$$F_1 \text{ score} = \frac{2 \cdot TP}{2 \cdot TP + FN + FP}$$

($F_1 \text{ score} : \text{worst value} = 0; \text{best value} = 1$)

$$\text{sensitivity} = \frac{TP}{TP + FN}$$

(Above all equations are mainly used for calculating the performance estimation parameters, these are useful for comparing the designed application with existed methods)

4. RESULTS

**Figure 5b. Experimental Setup**

Fig 5b explains about experimental setup of agriculture crop prior stage maintenance system; this can used to identify the tomato tree diseases with accurate manner.

**Table 1: Type Of Tomato Disease**

<table>
<thead>
<tr>
<th>S No</th>
<th>Disease</th>
<th>Fertilizer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>&quot;bacterial_spot&quot;</td>
<td>2&quot; min 10% chlorine bleach</td>
</tr>
<tr>
<td>2</td>
<td>early_blight</td>
<td>Backing soda and oil</td>
</tr>
<tr>
<td>3</td>
<td>late_blight</td>
<td>Fungicidal sprays</td>
</tr>
<tr>
<td>4</td>
<td>leaf_mold</td>
<td>Polestar Food waste</td>
</tr>
<tr>
<td>5</td>
<td>septoria_leaf_spot</td>
<td>Vermicomposite</td>
</tr>
<tr>
<td>6</td>
<td>spider_mites</td>
<td>Phytoseiulus persimilis</td>
</tr>
<tr>
<td>6</td>
<td>target_spot</td>
<td>Fungicidal sprays</td>
</tr>
<tr>
<td>7</td>
<td>yellow_leaf_curl_virus</td>
<td>Myro fertilizers</td>
</tr>
</tbody>
</table>

Table 1 explains about, different diseases observed at tomato leaves; these are listed for the type of disease and corresponding fertilizer identification purpose.
Table 2: Accuracy Detection

<table>
<thead>
<tr>
<th>Samples of Diseases</th>
<th>Number of images used for training</th>
<th>Number of images used for testing</th>
<th>Detection accuracy/%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>MDC with K-Mean</td>
</tr>
<tr>
<td>bacterial _spot</td>
<td>15.00</td>
<td>10.00</td>
<td>80.000</td>
</tr>
<tr>
<td>tomato leaves</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>early_blight</td>
<td>15.00</td>
<td>14.00</td>
<td>92.850</td>
</tr>
<tr>
<td>tomato leaves</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>late_blight</td>
<td>15.00</td>
<td>10.00</td>
<td>90.000</td>
</tr>
<tr>
<td>tomato leaves</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>leaf_mold</td>
<td>15.00</td>
<td>12.00</td>
<td>83.330</td>
</tr>
<tr>
<td>tomato leaves</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>septoria_leaf_spot</td>
<td>15.00</td>
<td>10.00</td>
<td>90.320</td>
</tr>
<tr>
<td>tomato leaves</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>spider_mites</td>
<td>15.00</td>
<td>12.00</td>
<td>88.450</td>
</tr>
<tr>
<td>tomato leaves</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>target_spot</td>
<td>15.00</td>
<td>11.00</td>
<td>88.454</td>
</tr>
<tr>
<td>tomato leaves</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall accuracy</td>
<td></td>
<td></td>
<td>86.540</td>
</tr>
</tbody>
</table>

Table 2 explains that different methods for calculating the accuracy and crop leaf disease recognition using MDC, SVM, TSRFO on various tomato plant leaves, this has been examined by using the sample images from the database.

Table 3. Disease finding and pesticide.

<table>
<thead>
<tr>
<th>S No</th>
<th>Type of Tree</th>
<th>Disease</th>
<th>Results</th>
<th>Pest Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Tomato leaves data set_1</td>
<td>bacterial _spot</td>
<td>2” min 10% chlorine bleach</td>
<td></td>
</tr>
<tr>
<td>No.</td>
<td>Data Set Title</td>
<td>Problem Type</td>
<td>Solution</td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>--------------------------------</td>
<td>---------------</td>
<td>---------------------------</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Tomato leaves data set_1</td>
<td>early_blight</td>
<td>Backing soda and oil</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Tomato leaves data set_1</td>
<td>late_blight</td>
<td>Fungicidal sprays</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Tomato leaves data set_1</td>
<td>leaf_mold</td>
<td>Polestar Food waste</td>
<td></td>
</tr>
</tbody>
</table>
Table 3 explains that outcome that is obtained from the proposed TSRFO method and attains the disease recognition and corresponding pesticide. In this various sample, pre-harvesting images are selecting for input from database.

Table 4. Accuracy And Diseases

<table>
<thead>
<tr>
<th>leaf disease</th>
<th>bacterial _spot</th>
<th>bacterial _spot</th>
<th>bacterial _spot</th>
<th>bacterial _spot</th>
<th>bacterial _spot</th>
<th>bacterial _spot</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>bacterial _spot</td>
<td>15</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>93</td>
</tr>
<tr>
<td>early_blight</td>
<td>2</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>97.8</td>
</tr>
</tbody>
</table>
Table 4 explains that the average accuracy of some of the diseases, find with proposed methodology and obtain the accuracy is 97.6% which is better compared to conventional and modern models.

<table>
<thead>
<tr>
<th>Disease</th>
<th>Real</th>
<th>Predicted</th>
<th>True Positive</th>
<th>False Positive</th>
<th>True Negative</th>
<th>False Negative</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>late_blight</td>
<td>0</td>
<td>0</td>
<td>21</td>
<td>0</td>
<td>14</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>leaf_mold</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>21</td>
<td>0</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>septoria_leaf_spot</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>19</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>spider_mites</td>
<td>2</td>
<td>0</td>
<td>11</td>
<td>1</td>
<td>0</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>97.6</td>
<td></td>
</tr>
</tbody>
</table>

Fig 6 explains that various methods on selected tomato images; these are gain accuracy, True positive rate, MCC values. Here proposed method achieves the better improvement compared to existed methods.

**Figure 6. MCC, Accuracy, F1score Detection.**

<table>
<thead>
<tr>
<th>Method</th>
<th>MCC</th>
<th>accuracy</th>
<th>F1 score</th>
<th>sensitivity</th>
<th>specificity</th>
<th>PSNR</th>
<th>SSIM</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed TSRFO</td>
<td>+0.390</td>
<td>0.976</td>
<td>0.410</td>
<td>0.310</td>
<td>0.980</td>
<td>59.82</td>
<td>0.9989</td>
<td>0.0081</td>
</tr>
<tr>
<td>Decision tree (applied only to</td>
<td>+0.280</td>
<td>0.760</td>
<td>0.370</td>
<td>0.280</td>
<td>0.950</td>
<td>58.81</td>
<td>0.9981</td>
<td>0.0099</td>
</tr>
<tr>
<td>lung side &amp; platelet count)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 5 explains that quality assessment parameters in which MCC value is +0.390, accuracy 0.976, f1 score 0.410, sensitivity 0.310, specificity 0.980 and PSNR 59.82, SSIM 0.9989, MSE 0.0081 has achieved which is much accurate compared to all existed methods.

Fig 7 explains that peak to signal noise ratio, which is 59.82 has been achieved compared to existed methods; this is a better achievement. Based on these quality parameters like image quality and IoT data acquisition are estimated.

Table 6 ML Accuracy Comparison
Table 6 explains that various methods of accuracy comparison, in this many methods like logistic regression achieves 65%, SVM 40% KNN 67% and proposed TSRFO made 97.6% accuracy. Therefore implemented an application for agriculture prior stage crop harvesting method is more efficient and compete with adjusted methods.

Fig 8 explains that graphical representation of proposed Random forest classifier with threshold segmentation, this method is compared with conventional and modern methods conclude that TSRFO achieves good improvement at PSNR, MSE, SSIM, F1 score, etc.

5. CONCLUSION

In this work, a Tomato plant leaves prior stage disease prediction, and classification model is implemented. This application is beneficial for agriculture and healthy production. Conventional methods for tomato plant disease detection had achieved less accuracy, SSIM, MCC and F1 score. For any automatic application regarding IoT should improve the performance metrics. In this research TSRFO for tomato plant disease detection has attained more improvement and classify the disease at the prior stage. This work is very helpful for researchers who are deal with IoT and agriculture applications. The dataset which is obtained from crop science, nelson. wisc databases. These are utilized to classify the tomato crop leaf disease with accurate manner. At final achieves the 97.6% accuracy 59.82 PSNR, 0.0081 MSE as improved, this method also competes with existed methods.

REFERENCES


[22] Inthiyaz, Syed; Madhav, B. T. P.; Kishore, P. V. V., ‘Flower image segmentation with PCA fused colored covariance and gabor texture features based level sets’, AIN Shams Engineering Journal, 10.1016/j.asej.2017.12.007


