Thermal Imaging-Assisted Infection Classification (BoF) for Brinjal Crop

Shubhangi Verma, O. P. Singh, Sachin Kumar, and Sumita Mishra

Abstract In the development of economy, agriculture has always played an important role for different nations; since it is considered to be the main source of income, food, and employment to rural populations in the country, owing to diversified geographical locations, environmental conditions, and pest attacks, it is of prime importance to devise technological-assisted methods to monitor and provide early remedial actions for the damage and infections to the crop. Algorithm proposed focuses on health monitoring of brinjal crop using digital thermal imaging. Paper aims to identify the plant disease by analyzing thermal images of brinjal leaves. Infrared images are rich in important hidden details that are not visible due to their low contrast and blurring. Experiment was conducted on two sets of images, first set comprising of healthy and infected thermal images and second comprising of normal RGB capture of healthy and infected images; 30 to 35 images per crop per set were acquired, total dataset analyzed had 1160 images, and the process of identification was implemented via bag of features (BoF), under the umbrella; feature extraction was carried out by SIFT operator, and classification was performed using classification MLSTSVM. Simulation was implemented using MATLAB 2018b. Results showed that duration of the process was less for RGB images by a margin of approximately 6 secs, but the accuracy efficiency achieved was more for thermal images by margin of 3%, having 87% in all. From the results, it can be concluded that however duration required for the identification was more for thermal images but still percentage accuracy is more for thermal images; thus, thermal image-assisted algorithm can be employed for crops in remote scenarios where accuracy plays a vital role.

Keywords Brinjal plant · Thermal images · RGB · BoF · Accuracy

S. Verma · O. P. Singh · S. Kumar (✉) · S. Mishra
Amity University, Lucknow campus, Lucknow, India
e-mail: skumar3@lko.amity.edu

O. P. Singh
e-mail: opsingh@amity.edu

S. Mishra
e-mail: smishra3@lko.amity.edu

H. Sharma et al. (eds.), *Communication and Intelligent Systems*, Lecture Notes in Networks and Systems 204, https://doi.org/10.1007/978-981-16-1089-9_5

1 Introduction

In the development of economy, agriculture has always played an important role of different countries since it is the main source of income, food, and employment to rural populations in the country. USA, China, and India are some top countries in food exporters which produce more food together than European Union put together. Brazil is in the fourth place which is more toward the production of sugarcane, soybeans, and beef. India stands second largest food producer in terms of intake of total calorie, and if total value of agriculture production is measured, India falls in fourth position followed by China, USA, and Brazil. Agriculture in India contributes to 19.5% of the GDP [1–6].

In order to increase the agriculture's contribution toward GDP, it is essential for us to take necessary actions toward the development in agriculture sector by helping rural communities in understanding the India's situation of leading crop production. As India tops in producing some vegetables like potato, tomato, sugarcane, and many more, monitoring health of these crops is an important factor in order to increase its productivity and its contribution in India's economy as well. Here, we are specifically focusing on brinjal plants since it is poor man's vegetable and popular with low-income consumers and small-scale farmers which are grown mostly in all states of our country, and India is a second largest producer after China. In the year 2017–2018, West Bengal tops the list in production of brinjal and produced its total share of 23.69%, whereas Haryana is in 10th position and its share is 2.50% as shown in Fig. 1 [4–6].

Monitoring the health of the crops is an important step toward the growth in productivity. There are certain factors which should be kept in mind while monitoring

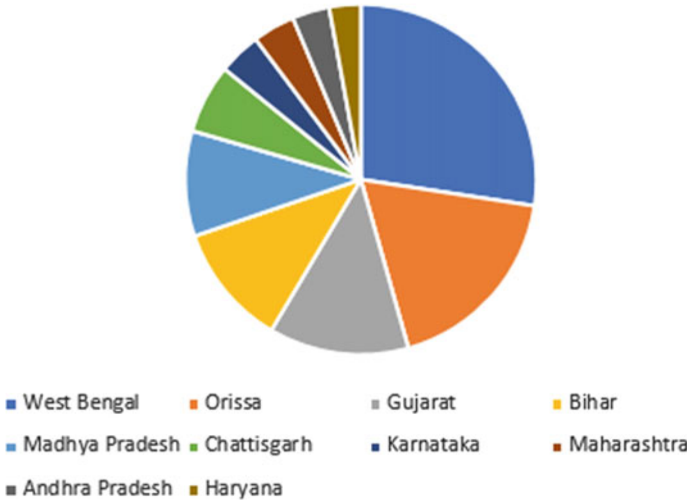

Fig. 1 Top 10 leading brinjal producers of India [4–6]

crop's growth like light, water, temperature, nutrients, diseases, and many more. Early detection of whatever issues associated with any one of the above factors can help us to monitor the growth in an efficient way and without wasting our resources in later stage. In this paper, we are specifically focusing on identification of diseases in brinjal plants [5, 7–9].

The paper is arranged as follows: Sect. 2 presents the motivation for conducting the study, Sect. 3 presents the material and methods employed for implementing the algorithm adopted for classification of thermal and RGB images for disease identification in brinjal plant, Sect. 4 discusses in brief the results obtained from the implementation conducted, and finally conclusion is presented in Sect. 5.

2 Motivation

Over the years, diversified techniques have evolved for early identification and cure of infection in crops. D.M. Bulanon et al. presented the paper on the thermal temporal variation in citrus fruit as a technique for enhancing fruit identification; in this work, canopy was examined after every twenty-four-hour period employing a thermal infrared camera and samples were segregated successfully in the dataset employing image processing algorithms at the instant of highest temperature variation; the results obtained gave improved results as compared to existing algorithms. Model presented a combination of thermal and RGB datasets with learning algorithms to remotely identify plants infected with disease. Shan-e-Ahmed Raza et al. obtained a group of features from the sample data employing global and local mathematical processing and displayed that combining normal and thermal identification methods provided depth information and improved the accuracy and precision of disease identification in infected plants [8–11].

IR sensor applications include detecting drought phenotype samples for studies from rice fields and grape farms, samples have issues of sensitivity, such as plant temperature variations of 15 °C happening between portions of shaded area and complete sunlight. Hamlyn G. Jones et al. employed the techniques in data analytics; displaying about variation, may be identified despite visible variations in soil water content using normalization methods [11]. Carmen M. Ortiz-Bustos et al. studied BGF emission by benign sunflower plant leaves and employed a combination of thermal and BGF imaging techniques for the identification of the diseases. Lower BGF was identified in parasitized samples via leaf expansion, and low pigment concentration was identified at final time [12].

Zulkifli Bin Husin et al. presented a technique for early identification of chili infection via leaf feature monitoring. Image dataset was acquired and processed to identify the infection level of the sample. Method ensures that pesticides or insecticides are applied only if the samples are infected. The system proposed is extremely cost-effective and inexpensive [13]. With farming process, processing of image is employed to identify the infectious diseases on samples. Ashwani Kumar et al. employed ANN for the purpose. Grapes and apples were the samples used for the algorithm. The technique proposed employed two datasets, first one for training of infected samples and the second one for testing of queried sample images. Work presents algorithms for the spread of infection and counting of fruits. MATLAB 2018b was employed for the implementation [14]. Automated disease identification is achieved employing multiple AI algorithms. The model proposed has three blocks, i.e., color segmentation from sample dataset, disease segmentation from segregated samples, and classification of infection. Segmentation via colors of leaf is a preprocessing block, this segments out unnecessary information that is not required for identification, and this helps in segmenting grape leaf regions in the dataset. P. Kumsawat et al. presented the model which displays the output segmented sample, which is subjected to filtering process and permits the system to study disease through color features. Model proposed can successfully categorize the dataset into three classes, rust infection, scab infection, and benign samples [15]. S. Bani-Ahmad et al. presented a soft solution for automated identification of plant diseases. Model developed comprises four main components. First is the process of segmentation; after this phase, two steps are there; first, identify region of interest, i.e., green colored, and in second regions are masked depending upon threshold value defined, employing Otsu's technique; second step involves subdivision into clusters with RGB values defining region of interest for elimination. Algorithm developed efficiently detects and classifies the infected region [16].

Disease identification comprises steps such as acquiring image dataset, preprocessing, segregating, feature extraction, and benign or infected classification. A.B. Patil et al. presented techniques for detection of infections in plants employing health of leaves and feature extraction from these leaves for disease identification [17]. The automated platform for rice infection based on the feature extraction of different rice plants is presented by Jaya Sil et al. Dataset of images is obtained from live images acquired from camera, rice crop was taken for the work, and images acquired are processed to be fed to NN for classification of infected regions [18]. Kumar et al. classified soil into different categories [19] and also developed a method for disease identification in plants [20].

Owing to potential dependency and importance crops play in world's economy, it is important to monitor and save crops timely for healthy production. Therefore, there is a potential requirement of pioneering and out of box results in the area, supported by relevant research results. Work proposed is an effort to increase productivity through early and precise disease detection with thermal images and thus plant for early remedial action, and the process is executed by analyzing thermal images.

3 Material and Methods

In crops, diseases can be classified as biotic and abiotic diseases. Biotic diseases originate from living organisms. They can be caused by fungi, bacteria, and viruses. Abiotic disease originates from non-living substances like hail, spring frosts, weather

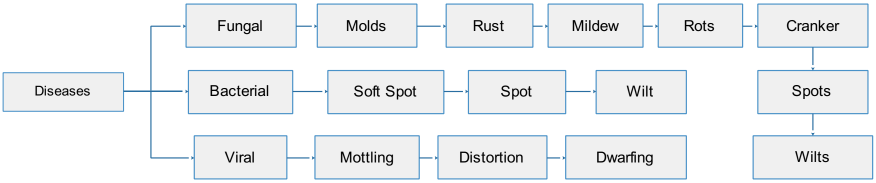

Fig. 2 Categorical division of diseases

conditions, burning of chemicals, etc. Figure 2 depicts categorical division of diseases.

Before extracting the features of diseases, it is important for us to know about the diseases caused in brinjal plant. Table 1 shows the different diseases' name along with its symptoms and images of the respective disease in plant [4–6, 17–20].

Table 1 Diseases in brinjal plant

S. No.	Disease	Symptoms	Images
1	Bacterial wilt	Leaf surface wilting, yellowing of foliage	
2	Cercospora leaf spot	Leaf spots by chlorotic lesions and irregular in shape	
3	Alternaria leaf spot	Cracks appear in leaf with concentric rings	
4	Damping off	Collapse of seed lines and spread through fungus present in soil	
5	Tobacco mosaic virus	Leaves are deformed and develop blisters in the advance case	
6	Collar rot	Caused by the accumulation of water around stem	

3.1 Experimental Setup

Experimental setup is a combination of hardware and software modules, both functioning in together for brinjal plant disease detection, and modules are as follows.

3.1.1 Hardware

The images are captured by FLIR one camera, which can be interfaced with the smartphone. The smartphone which we have used in this work is Motorola G6. The reason to choose this phone is that it has C-type USB slot so that this camera can be easily interfaced. Other smartphones with same feature can also be selected for this proposed model. This camera helps us to take the thermal images, and we have option to choose the palette of our choice. In the proposed work, we have chosen rainbow HC palette as it helps us to see the minute temperature difference.

3.1.2 Software

Thermal images acquired are cleaned to reach optimum-level acceptable quality for further processing, and the process is known as preprocessing of images and is done using MATLAB software. This software allows us to implement various filter algorithms, creation of user interfaces, segmentation techniques, and data analysis procedures. In preprocessing of images, techniques like image enhancement and noise filters are used to clean image, features linked with affected portions are extracted via segmentation techniques to process for disease identification, in data analysis standard parameters are correlated with segmented portions to conclude about infection in affected regions, and the processes mentioned are incorporated on a single platform through user interface built in MATLAB.

3.2 Proposed Methodology

For the detection of diseases, the following set procedure has been taken; the first step is thermal image acquisition, the second step is preprocessing, the third step is segmentation, and the fourth and last step is disease identification. Figure 3 shows the detailed flowchart of the procedure taken in this work.

3.2.1 Image Acquisition

The sample image was collected from a crop nursery, located at A block, Indira Nagar, Lucknow, images were acquired from 30 brinjal plants, images were captured in two

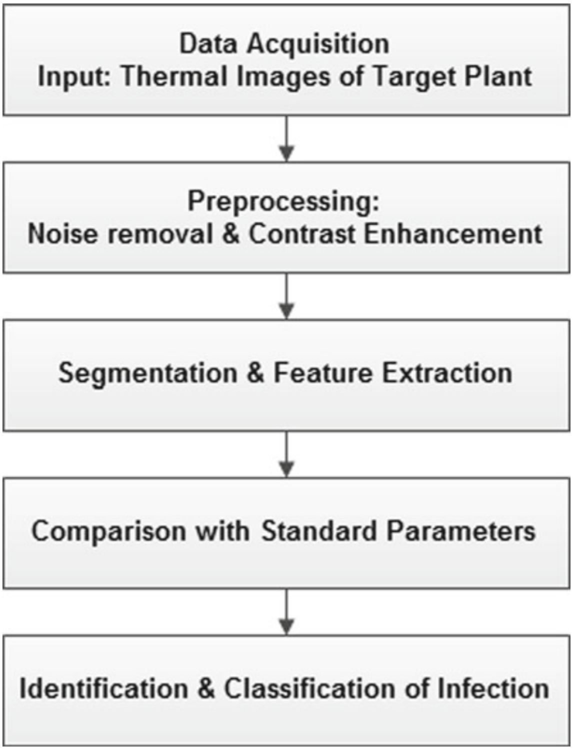

Fig. 3 Methodology proposed

sets, first set comprising healthy and infected thermal images and second comprising normal RGB capture of healthy and infected images, and 40 to 45 images per crop per set were acquired, total dataset analyzed had 2160 images; Fig. 4 depicts thermal samples collected.

3.2.2 Thermal Image Preprocessing

Images acquired were subjected to preprocessing steps that involved cleaning and contrast enhancement of acquitted thermal image of brinjal leaves, in cleaning images are first processed for noise removal and image degradation, this is achieved by the adaptive median filter and wiener filter, after noise removal dataset is further processed to improve image contrast, and dualistic sub-image histogram equalization is employed for the process as depicted in Fig. 5. The processed dataset comprising two sets each having healthy and infected categories is then fed for classification to BoF.

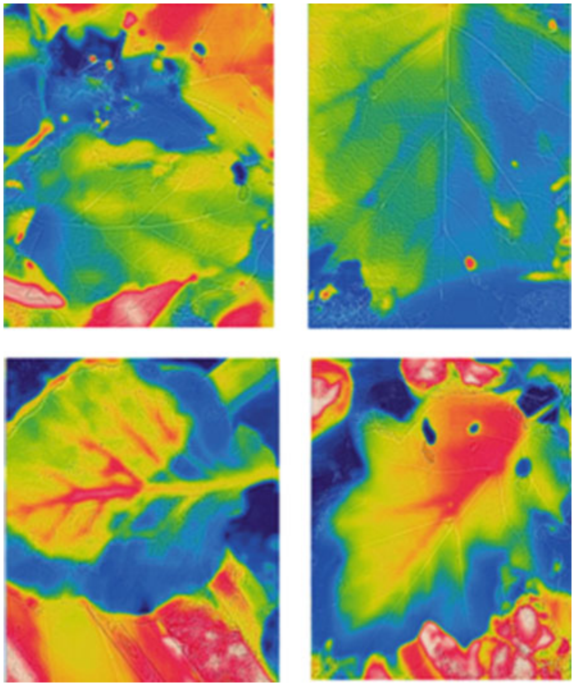

Fig. 4 Sample of thermal dataset acquired

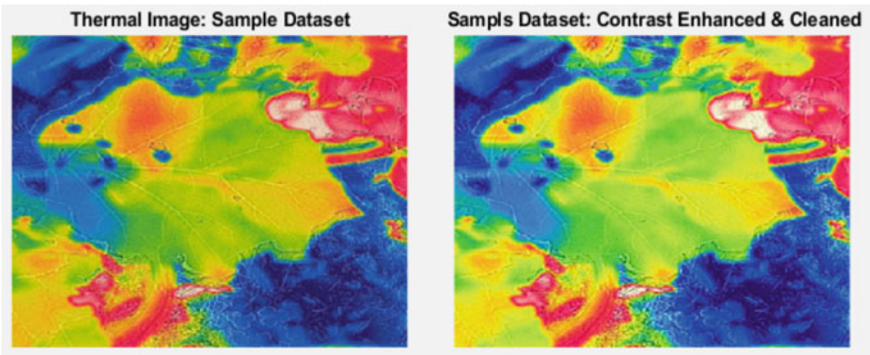

Fig. 5 Preprocessed image

3.3 Classification: BoF

Bag of feature has been employed for disease classification of the brinjal plant; two datasets one for thermal images and other for RGB images of healthy and infected leaves are passed to the training model, one set at a time to train the network. Once trained, the classifier is capable of efficiently classifying infected crops. BoF primarily involves five steps, feature extraction, codebook creation, encoding feature, feature pooling, and learning and classification phase.

The primary objective is to identify features which is achieved by the quantization of local descriptors of images in the dataset according to visual vocabulary. The vocabulary is formulated by clustering a large volume of local descriptors employing the K-means algorithm. The algorithm processes the training data as input and clusters of similar datasets as output. The individual cluster would be described by one visual feature. Images will now be depicted as a BoF. Features are extracted in two phases; in first, patches are sampled employing their interest points [17] or densely with a regular grid. In second, the extraction of features and descriptors is obtained. The dataset with thermal capture of healthy and infected samples is first fed, and implementation process can be summarized as follows:

1. Acquiring Data: The sample image was collected from a crop nursery, located at A block, Indira Nagar, Lucknow, images were acquired from 25 brinjal plants, images were captured in two sets, first set comprising healthy and infected thermal images and second comprising normal RGB capture of healthy and infected images, 30 to 35 images per crop per set were acquired, total dataset analyzed had 1160 images, and dataset obtained was cleaned and labeled as healthy and infected. The process was first performed on thermal dataset.
2. Feature Identification and Segregation: Objective is to segregate a representative group of images with the most significant information in the data. After identifying the critical features in each dataset, computation of vector takes place which describes the features. The tasks of object recognition and image categorization are completed using SIFT descriptor. SIFT descriptor consists of feature extraction and detection both. Figure 6 depicts creation of BoF for thermal healthy and infected images and extraction of features.
3. Formulation of Codebook and Quantization: In encoding phase, the local descriptor is obtained in a new form by utilizing visual vocabulary.

    (a) Codebook formulation: K-means clustering—Construction of visual vocabulary: A vector quantization technique is utilized. In this, N descriptors are partitioned into K clusters; here, each descriptor is associated with a cluster with closest mean. Trial-and-error iterations lead to $K = 500$ that gave optimum result, as K need not be sufficiently large and not too small too for a feature set of magnitude 1,587,788. K-means clustering of healthy and infected thermal images is depicted in Fig. 7.
    (b) Encoding: BoF methodology for encoding converts the local descriptor into a more adapted form through codebook [1, 12]. Soft assignment algorithm has been utilized in this case; it keeps more information about the original image features.

4. Classification Phase: During categorization of image, the primary objective is to automatically annotate images with predefined groups. Image labels are predicted as soon as descriptors are extracted using a set of classifiers. BoF

```
Creating Bag-Of-Features.

-------------------------

* Image category 1: Thermal Images-Healthy Leaf

* Image category 2: Thermal Images-Infected Leaf

* Selecting feature point locations using the Grid method.

* Extracting SURF features from the selected feature point locations.

** The GridStep is [8 8] and the BlockWidth is [32 64 96 128].


* Extracting features from 292 images in image set 1...done. Extracted 7992368 features.

* Extracting features from 233 images in image set 2...done. Extracted 8778432 features.


* Keeping 80 percent of the strongest features from each category.
```

Fig. 6 Feature extraction of healthy and infected thermal images

```
* Using K-Means clustering to create a 500 word visual vocabulary.

* Number of features        : 1587788

* Number of clusters (K)     : 500


* Initializing cluster centers...100.00%.

* Clustering...completed 60/100 iterations (~39.87 seconds/iteration)...converged in 60 iterations.


* Finished creating Bag-Of-Features
```

Fig. 7 K-means clustering of healthy and infected thermal images

employed SVM classifier from machine learning toolbox for classification. For the task, we have two categories of labeled dataset as healthy and infected.

Several "one-versus-all" classifications are performed for discriminating the multi-class problem. One-versus-all MLSTSVM—In this methodology, training of each class is done with rest of the other classes. Assume the scenario of K class, and OVA MLSTSVM classifier solves K-linear equations and generates K non-parallel hyperplanes, one hyperplane for each class.

Hyperplanes for the classification of images are obtained such that data points of each class lie in the close affinity of its respective hyperplane and maintain a clear separation. ith hyperplane is obtained as given in Eq. 1:

$$f_i = (w_i x) + b_i = 0 \tag{1}$$

where $w_i \in R_n$ refers to the normal vector to the hyperplane and $b_i \in R$ refers to the bias term. The objective function of ith linear OVA MLSTSVM classifier is expressed as given in Eq. 2:

$$\min(w_i,\ b_i, \xi_i)\frac{1}{2}\|X_i w_i + e_{i1} b_i\| + \frac{c_i}{2}\xi_i^T \xi_i$$

$$\text{s.t.}(Y_i w_i + e_{i2} b_i) + \xi_i = e_{i2} \tag{2}$$

where c_i denotes penalty parameter, $e_{i1} \in R_{1i}$ and $e_{i2} \in R_{1\text{-}1i}$ are two vectors of one's, and ξ_i is the slack variable. OVAMLSTSVM classifier decision function that classifies samples as healthy or infected is given in Eq. 3:

$$d(x) = \arg \min_{i=1\ldots K} \frac{|w_i x + b|}{\|w_i\|} \tag{3}$$

5. **Performance of Model**: The algorithm trained the image category classifier for 2 categories—healthy and infected, following which encoding was done for each category and finished training the category classifier. Confusion matrix and accuracy achieved for the images with BoF are depicted in Fig. 8.

The procedure for classification was repeated for RGB dataset of healthy and infected images; Fig. 9 depicts the confusion matrix and accuracy achieved for RGB images.

4 Result and Discussion

The experiment for the methodology proposed was conducted to estimate efficiency of disease identification through thermal images as compared to RGB images; for the experiment, dataset was collected from a crop nursery located in Lucknow, images were acquired from 25 brinjal plants, images were captured in two sets, first set comprising healthy and infected thermal images and second comprising normal RGB capture of healthy and infected images, 30 to 35 images per crop per set were acquired, total dataset analyzed had 1160 images, and the process of identification was implemented via BoF, under the umbrella; feature extraction was carried out by SIFT operator, and classification was performed using classification MLSTSVM; results obtained are displayed in Table 2; strongest features were retained for the K-means clustering before classification. Classification took 60 iterations for thermal images whereas 51 iterations for RGB images; duration of the process was less for RGB images by a margin of approximately 6 secs, but the accuracy efficiency achieved was more for thermal images by margin of 3%, having 87% in all. From the results,

```
Evaluating image category classifier for 2 categories.

-------------------------------------------------------

* Category 1: Thermal Images-Healthy Leaf

* Category 2: Thermal Images-Infected Leaf

* Evaluating 292 images from category 1...done.

* Evaluating 233 images from category 2...done.

* Finished evaluating all the test sets.

* The confusion matrix for this test set is:

                                   PREDICTED

KNOWN                     | Thermal Images-Healthy Leaf   Thermal Images-Infected Leaf

-----------------------------------------------------------------------------------------

Thermal Images-Healthy Leaf    | 0.92                0.08

Thermal Images-Infected Leaf   | 0.18                0.82


* Average Accuracy is 0.87.
```

Fig. 8 Confusion matrix and accuracy achieved for thermal images with BoF

it can be concluded that however duration required for the identification was more for thermal images but still percentage accuracy is more for thermal images. The algorithm proposed can be very useful for more accurate disease identification of crops with a little compromise on duration elapsed for identification.

5 Conclusion

Damages caused due to erratic environmental conditions, diseases, and pest attacks cause substantial losses in the yield and quality of vegetables produced worldwide. Food and Agriculture Organization (FAO) report says that more than half world's population depends on agriculture for their survival. Paper proposes an algorithm to identify infected leaves and thus plant for early remedial action, the process is executed by analyzing thermal and RGB images of Brinjal leaves, and BoF has been employed for classification procedure. Simulation was implemented using MATLAB 2018b. Classification took 60 iterations for thermal images whereas 51 iterations for RGB images; duration of the process was less for RGB images by a margin of approximately 6 secs, but the accuracy efficiency achieved was more for thermal images by margin of 3%, having 87% in all. From the results, it can

```
Evaluating image category classifier for 2 categories.

-------------------------------------------------------

* Category 1: Healthy Leaves

* Category 2: Infected Leaves

* Evaluating 236 images from category 1...done.

* Evaluating 291 images from category 2...done.

* Finished evaluating all the test sets.

* The confusion matrix for this test set is:


                   PREDICTED

KNOWN           | Healthy Leaves   Infected Leaves

--------------------------------------------------------

Healthy Leaves   | 0.81          0.19

Infected Leaves  | 0.13          0.87


* Average Accuracy is 0.84.
```

Fig. 9 Confusion matrix and accuracy achieved for RGB images with BoF

Table 2 Obtained parameters values for thermal and RGB images

Type of image/algorithm parameters	Healthy leaf features extracted	Infected leaf features extracted	Strongest features employed for K-means	Iterations	Duration (s)	Accuracy (%)
Thermal	7,992,368	8,778,432	1,587,788	60	39.87	87
RGB	9,180,288	9,297,064	1,475,302	51	33.22	54

be concluded that however duration required for the identification was more for thermal images but still percentage accuracy is more; thus, thermal image-assisted algorithm can be employed for crops in remote scenarios where accuracy plays a vital role. The proposed model can also be used for other crop plants as well with minor modifications in methodology.

References

1. Alston JM, Pardey PG (2014) Agriculture in the global economy. J Econ Perspect 28(1):121–146
2. Wen D, Ren A, Ji T et al (2020) Segmentation of thermal infrared images of cucumber leaves using K-means clustering for estimating leaf wetness duration. Int J Agric Biol Eng 13(3):161–167
3. Zeng X, Miao Y, Ubaid S, Gao X et al (2020) Detection and classification of bruises of pears based on thermal images. Postharvest Biol Technol 161. https://doi.org/10.1016/j.postharvbio.2019.111090
4. www.apeda.in/agriexchange/
5. www.kisansuvidha.com/brinjal-diseases/
6. www.indiastat.com/agriculture-data/2/stats.aspx
7. Lydia MS, Aulia I, Jaya I et al (2019) Preliminary study for identifying rice plant disease based on thermal images. J Phys Conf Ser 1566; 4th international conference on computing and applied informatics 2019 (ICCAI 2019), 26–27 November 2019, Medan, Indonesia
8. Haris KM, Kumar S, Mishra S, Pragya (2018) Precision crop monitoring for sugarcane and corn. Int J Latest Trends Eng Technol 9(3):289–293. e-ISSN: 2278-621X
9. Kumar S, Mishra S, Khanna P, Pragya (2017) Precision sugarcane monitoring using SVM classifier. Procedia Comput Sci 122:881–887. Information Technology and Quantitative Management (ITQM 2017), Elsevier
10. Raza S-e-A, Prince G et al (2015) Automatic detection of diseased tomato plants using thermal and stereo visible light images. PLoS Journal
11. Jones HG, Serraj R, Loveys BR et al (2009) Thermal infrared imaging of crop canopies for the remote diagnosis and quantification of plant responses to water stress in the field. Funct Plant Biol J
12. Ortiz-Bustos CM et al (2017) Use of blue-green fluorescence and thermal imaging in the early detection of sunflower infection by the root parasitic Orobanche cumana Wallr. Advances in oil crops research. Front Plant Sci
13. Bin Husin Z, Bin Abdul Aziz AH et al (2012) Feasibility study on plant Chili disease detection using image processing techniques. In: 2012 third international Conference on Intelligent Systems Modelling and Simulation
14. Jhuria M, Kumar A, Borse R (2013) Image processing for smart farming: detection of disease and fruit grading. In: Proceedings of the 2013 IEEE second international conference on image information processing (ICIIP-2013)
15. Meunkaewjinda A, Kumsawat P, Attakitmongcol K et al (2008) Grape leaf disease detection from color imagery using hybrid intelligent system. In: Proceedings of ECTI-CON 2008
16. Al-Hiary H et al (2011) Fast and accurate detection and classification of plant diseases. Int J Comp Appl (0975 – 8887) 17(1)
17. Khirade SD, Patil AB (2015) Plant disease detection using image processing. In: IEEE conference
18. Phadikar S, Sil J (2008) Rice disease identification using pattern recognition. In: Proceedings of 11th international conference on computer and information technology (ICCIT 2008), 25–27 Dec 2008, Khulna, Bangladesh
19. Kumar S, Sharma B, Sharma VK, Poonia RC (2018) Automated soil prediction using bag-of-features and chaotic spider monkey optimization algorithm. Evol Intell, 1–12. https://doi.org/10.1007/s12065-018-0186-9.
20. Kumar S, Sharma B, Sharma VK, Sharma H, Bansal JC (2018) Plant leaf disease identification using exponential spider monkey optimization. Sustain Comput Inf Syst 28. https://doi.org/10.1016/j.suscom.2018.10.004