

## ORIGINAL RESEARCH

# Evaluation of image processing technique in identifying rice blast disease in field conditions based on KNN algorithm improvement by K-means

Mohammad Reza Larijani<sup>1</sup> | Ezzatollah Askari Asli-Ardeh<sup>1</sup>  | Ehsan Kozegar<sup>2</sup> | Reyhaneh Loni<sup>3</sup>

<sup>1</sup>Department of Biosystems Engineering, University of Mohaghegh Ardabili, Ardabil, Iran

<sup>2</sup>Department of Computer Sciences and Engineering, University of Guilan, Guilan, Iran

<sup>3</sup>Department of Biosystem Engineering, Tarbiat Modares University, Tehran, Iran

## Correspondence

Ezzatollah Askari Asli-Ardeh, Department of Biosystems Engineering, University of Mohaghegh Ardabili, Ardabil, Iran.  
Email: ezzataskari@uma.ac.ir

## Funding information

University of Mohaghegh Ardabili

## Abstract

Nowadays, rice farming is affected by various diseases that are economically significant and worthy of attention. One of these diseases is blast. Rice blast is one of the most important limiting factors in rice yield. The purpose of this study is the timely and rapid diagnosis of rice blast based on the image processing technique in field conditions. To do so, color images were prepared using image processing technique and improved KNN algorithm by K-means was used to classify the images in *Lab* color space to detect disease spots on rice leaves. Squared classification was based on Euclidean distance, and the Otsu method was used to perform an automatic threshold histogram of images based on shape or to reduce the gray level in binary images. Finally, to determine the efficiency of the designed algorithm, sensitivity, specificity, and overall accuracy were examined. The classification results showed that the sensitivity and specificity of the designed algorithm were 92% and 91.7%, respectively, in the determination of the number of disease spots, and 96% and 95.65% in determining the quality of disease spots. The overall accuracy of the designed algorithm was 94%. Generally, the results obtained showed that the above method has a great potential for timely diagnosis of rice blast.

## KEYWORDS

blast disease, image processing, K-means algorithm, KNN algorithm, rice

## 1 | INTRODUCTION

In scientific terms, rice is called *Oryza Sativa*. It is one of the most important and strategic crops and it is the food of more than two third of the population of the world. Rice is considered as a healthy and nutritious foodstuff, and now it provides 50% of the world's agricultural production and 20% of the energy needed by humans. The study of rice from the viewpoint of food security shows that over the past three decades, the use of rice in Iranian people's eating habits

has been increasing steadily, with per capita consumption ranging from 15–20 kg, 30–40 years ago, to higher than 37.4 kg in the last decade (Peyman, Bakhshipour Ziaratgahi, & Jafari, 2016).

However, population growth and growing demand for this product have placed Iran among the most important countries importing rice. Therefore, focusing on the increase in production per unit area is the most important goal of the country in the field of agriculture in order to apply all the factors affecting production. But rice farming is endangered by many pests and diseases that are economically

significant and need to be taken into consideration carefully (Agahi, Fotokian, & Younesi, 2012; Peyman et al., 2016). One of these rice diseases is blast. Blast is one of the most important limiting factors in plant performance. The maximum sensitivity to the blast disease is observed at the sprouting stage, that is when the rice plant starts to appear above the surface of the ground. Even resistant varieties are susceptible to this disease during the formation of flowering organs. Symptoms of the disease on the leaf appear first in the form of burned water points and then turn into rhombus-shaped spots of 1–3 cm in length which are tipped at the end. The spots are gray in the center and dark brown in the margin. In case of severe disease, all leaves of a plant may become dry. There are chemical methods to prevent this fungal disease, but the important point is the timely and accurate diagnosis of the disease (Kumar et al., 2003). Because of the very high importance of blast disease, extensive research has been conducted to control it from various aspects (diagnosis, use of pesticides and fungicides, effective factors in reducing the disease, heat and moisture). The cause of the spread of this disease in the rice fields in northern Iran has been overuse of nitrogen fertilizers (Khosravi, 2006). So far, with the management of plant nutrition, planting time, rotation plans, plant spacing, soil moisture, biological control, and chemical control, no considerable results have been achieved in the fight against this disease (Ashkani et al., 2016; Bakar et al., 2018). Considering the significance of the topic, it is very important to use the science of machine vision and image processing techniques, which today play a major role in precision agriculture. The machine vision, as a powerful and reliable tool, has been widely used in various industries, especially in agriculture. The main use of machine vision and image processing in agriculture involves controlling the status of agricultural land, precision agriculture (PA), controlling and supervising plants at the planting stage and controlling and monitoring the quality of agricultural products at the postharvest stage. The main reason for the ever-increasing expansion of the machine vision and image processing science in various branches of agricultural science is that in addition to identifying the shape, color, size and texture of the objects, these systems can also extract the numerical and quantitative characteristics of these objects (Azarmdel, Mohtasebi,

Jafari, & Muñoz, 2019; Jahanbakhshi & Kheiralipour, 2019; Kahar, Mutalib, & Abdul-Rahman, 2015; Pydipati, Burks, & Lee, 2006).

Nowadays, rapid development of computer processing technologies and creation of related software makes it possible for us to benefit from the advantages of artificial intelligence. One example of these technologies is the application of artificial neural networks and other algorithms that to some extent copy human brain functions, to solve problems in systems and modeling processes (Jahanbakhshi, Ghamari, & Heidarbeigi, 2017; Kaveh, Jahanbakhshi, Abbaspour-Gilande, Taghinezhad, & Moghimi, 2018). Samanta & Ghosh (2012) used an artificial neural network to detect pests on tea tree shrubs applying the CFS-based McCulloch-Pitts model for designing a disease identification algorithm. The results of the experiments based on the designed algorithm had a 100% precision in pest detection. Pydipati et al. (2006) identified four grapefruit-related diseases using image processing and isolation analysis techniques and applied the co-occurrence matrix for classification. They were able to classify the diseases with the precision of about 98.75%. Al-Saqr (2012) used artificial neural networks to detect tropical walnut mite pests. His test results showed with 100% accuracy that the detection algorithm can perform diagnostic operations. In addition, the pest analysis and detection time by the algorithm in question was 0.16 s. Kahar et al. (2015) used artificial neural network and fuzzy logic to categorize three types of rice disease, that is blast, rice pod rot, and rice leaf rot. The accuracy of their algorithm in the diagnosis of rice plant disease was 100%.

In modern agriculture, quick methods reviews, automated, cheap and accurate methods for diagnosing plant diseases are important. Timely and accurate diagnosis of disease in farms is one of the most important factors in controlling plant diseases. Also, the use of a method that can manage the whole farm to be online is very important. Therefore, the purpose of this research is to online management of rice fields using a quadcopter (helicopter) and image processing technique to identify rice blast disease in field conditions. Review of the related literature revealed that until now no studies have been conducted on detecting the pests and diseases of the rice plant to be online using a quadcopter and image processing technique.



**FIGURE 1** Imaging by quadcopter

## 2 | MATERIALS AND METHODS

### 2.1 | Field imaging and preparation of the images

Canopy color images of rice fields were prepared in RGB space by a PHANTOM 4 ADVANCED quadcopter, equipped with a 12-megapixel digital camera. The velocity of the quadcopter in rice fields was 0.5 m/s and its distance from the bushes was 90 cm (Figure 1). The images were prepared under different weather conditions almost four weeks after planting rice. Due to the filming of different parts of the farms, the dates of planting the rice bushes were not the same.

The stages of analysis of rice leaf images for detecting blast disease under field conditions based on the K-means and KNN algorithms are presented in Figure 2.

### 2.2 | Preprocessing images

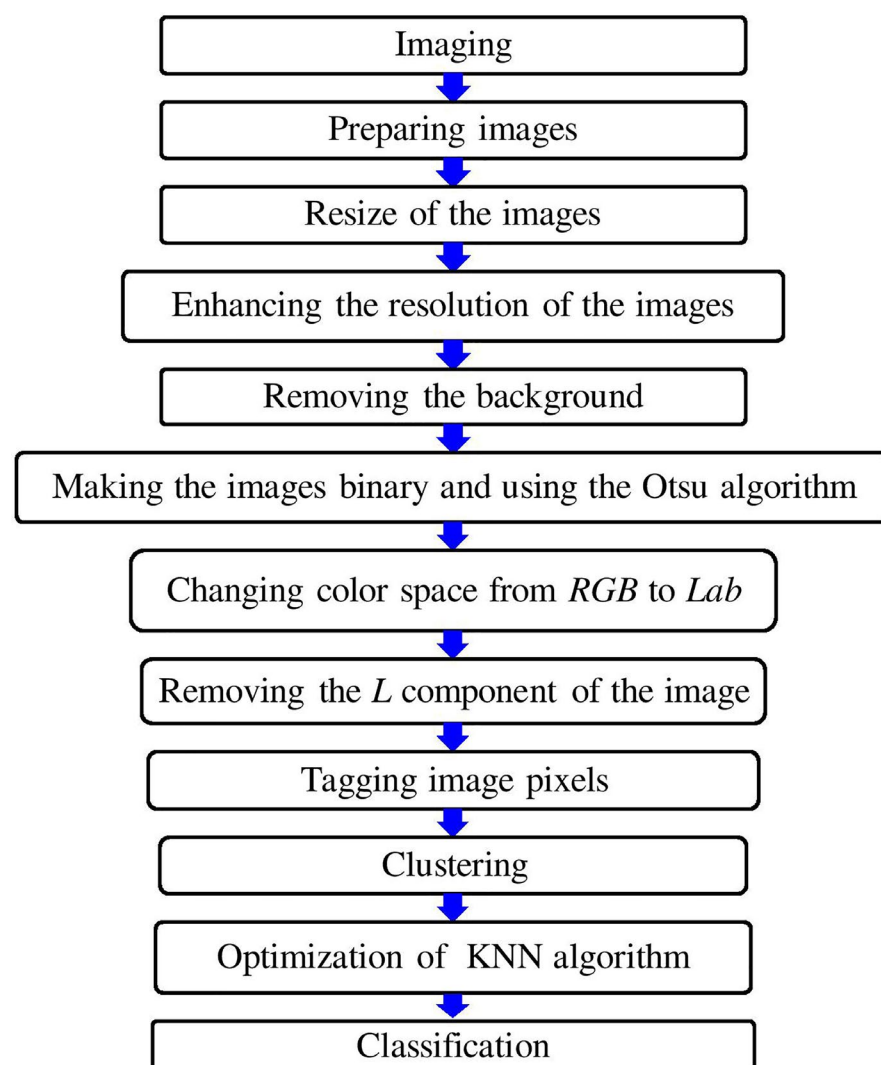
Image processing started after the imaging stage was completed. To process images, an application was coded in MATLAB R2012a. After the images were invoked by the program, the preprocessing

of the images began. Because the images were large and this reduced the speed of the analysis and processing of images, the images were converted to  $256 \times 256$  pixels so that the computing machine could perform image analysis in the shortest possible time. Improving the resolution of an image to better diagnose the disease and to raise the ability to diagnose the healthy surfaces from the diseased ones (points a, b, c and d in Figure 3). Therefore, the designed algorithm could better detect the disease on the plant leaf (Figure 3).

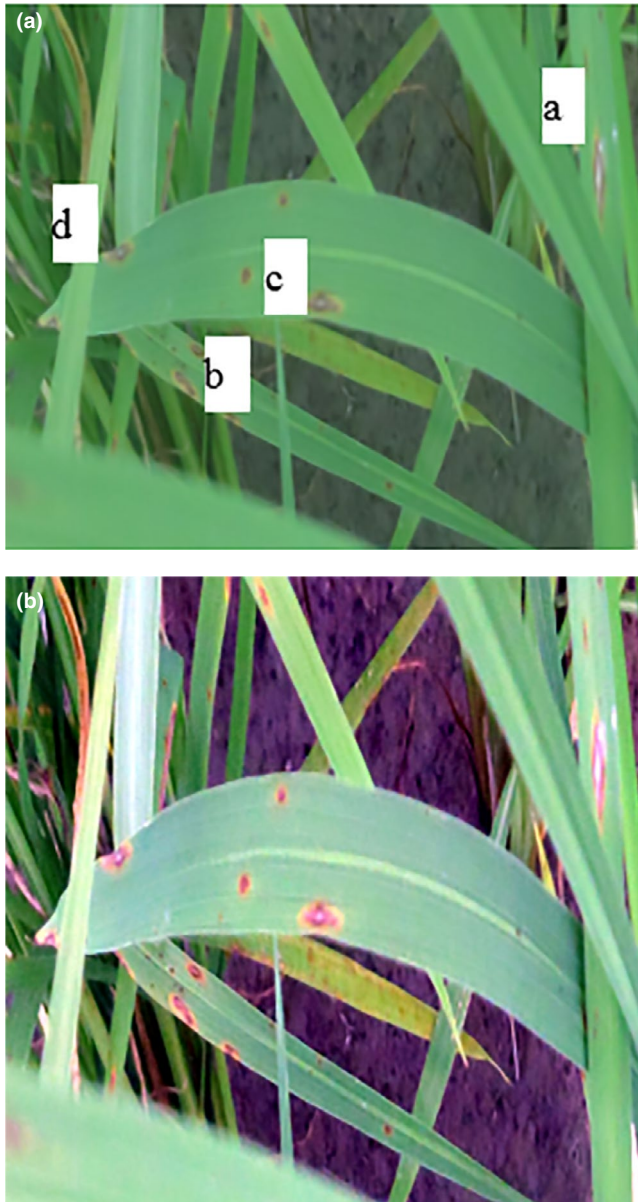
Because the images were taken at different times of the day and different weather conditions with different light intensities, the lighting and image alignment operations were then carried out in the next step. This allows the image to match with the image pixels.

### 2.3 | Image processing

In the next step, the separation algorithm was performed using the histogram profile extraction (Figure 4) to remove the probable background in the received images. The received images included rice bushes, stems, and green leaves and mud (predominantly green and gray colors). In this algorithm, the original image was



**FIGURE 2** Stages of analysis of rice leaf images

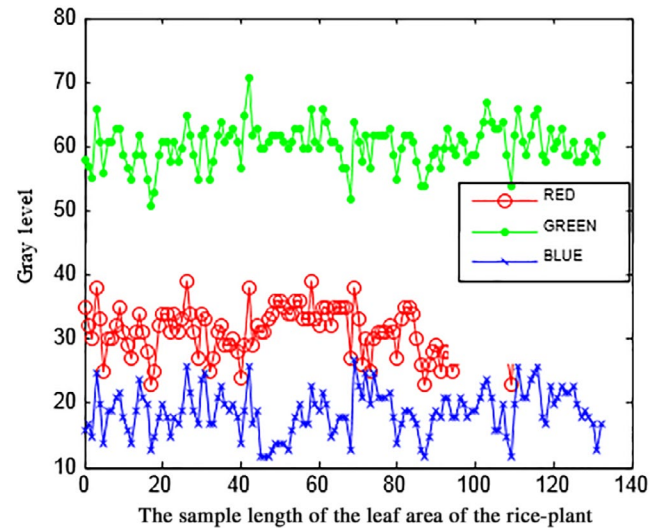


**FIGURE 3** (a) Input image of the algorithm and (b) High-resolution image

first split into three components: red, green, and blue, and then the rice plant was removed from the background of the image (Figure 5).

## 2.4 | Final processing

To perform segmentation, the received images must be transferred to the color space independent of the device. In a color space independent of the device, the coordinates used to determine the color determine the same color, regardless of the device used, and a color-dependent device is the space in which the resulting color depends on the equipment used to produce it (Kubat, 2015). Therefore, in this study, *Lab* color space was used which is widely used in machine vision. The three dimensions of this color space are the color brightness ( $L = 0$  indicates black color and  $L = 100$  shows white), its position



**FIGURE 4** Profile extracted from the rice leaf

relative to red, purple or green ( $a$ : negative values indicate green and positive values indicate violet), and its position relative to the yellow and blue ( $b$ : the negative values indicate a blue and the positive values indicate a yellow image).

The color space of the *Lab* turned into a color space suitable for machine vision due to its resistance to color variations. The conversion from *RGB* color space to *Lab* was performed in the form of relation (1):

$$L = 116 \times f\left(\frac{Y}{Y_n}\right) - 16 \quad (1)$$

$$a = 500 \times \left[ f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right) \right]$$

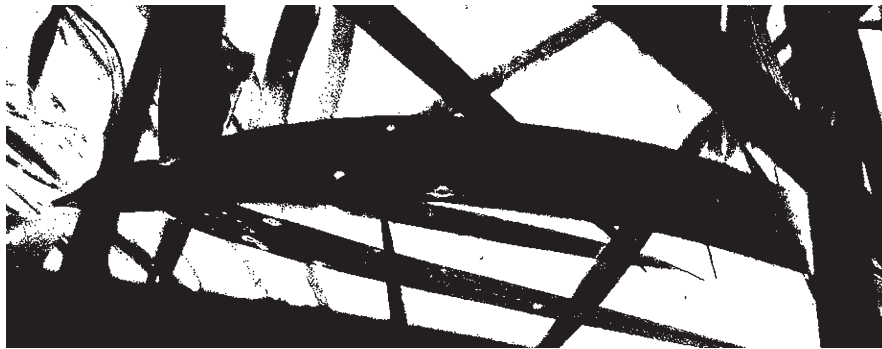
$$b = 200 \times \left[ f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right) \right]$$

Where

$$f(t) = \begin{cases} t^{1/3} & \text{if } t > \left(\frac{6}{29}\right)^3 \\ \frac{1}{3} \left(\frac{29}{6}\right)^2 t + \frac{4}{29} & \text{O.W.} \end{cases}$$

*Lab* color space is able to describe all colors visible by the human eye and is used as an independent color model. The *Lab* color space is derived from the original *XYZ* color space (Kubat, 2015). After removing the background of the images, the K-means method was used to segment images. Moreover, to remove the luminance effect, the component  $L$  of the images was also deleted and the clustering algorithm was applied only to the components  $a$  and  $b$ . The basis of clustering in the algorithm designed was squared Euclidean distance. In this algorithm, the Otsu method was used to carry out an automatic histogram of the threshold of images based on the shape or reduction of gray surface in binary images.





**FIGURE 5** Removing the background image of the rice bush

After converting the color space from *RGB* to *Lab*, the colors obtained from the previous step were clustered using the K-means method. In this algorithm (clustering algorithm), the colors which were similar to each other in the image were placed in a certain number of clusters up to a certain distance, and each of the clusters (3 clusters) was the final result. To use this method, the number of clusters and the spacing of each cluster from the other should be determined. For this purpose, relation (2) was used:

```
ab=double (lab_he (:, :, 2:3)); (2)
```

```
nrows=size (ab, 1);
```

```
ncols=size (ab, 2);
```

```
ab=reshape (ab, nrows*ncols, 2);
```

```
ncolors=3;
```

```
[cluster_idx clustercenter]=
```

```
kmeans (ab, ncolors,'distance', 'sqeuclidean','replicates', 3);
```

After specifying the clusters and spacing, an image was labeled by cluster index. Labeling of pixels using the results of the K-means method for each part of the image shows a cluster index (Figure 6).

After executing the image labeling algorithm, the next step is to obtain clustered images. This operation was carried out using equation (3). This stage is the main step in the rice blast disease diagnostic algorithm. The image of the diseased parts, the healthy parts of the plant, and pixels that do need to be processed will be obtained. Each of these images is a kind of cluster of the original image (Figure 7).

```
Segmented_images=cell (1, 3); (3)
```

```
rgb_label=repmat (pixel_labels, [1 1 3]);
```

```
for k=1:ncolors
```

```
color=he;
```

```
color (rgb_label~=k)=0;
```

```
segmented_images {k}=color;
```

After clustering, classification turn arrives. KNN should be improved for the following reasons:

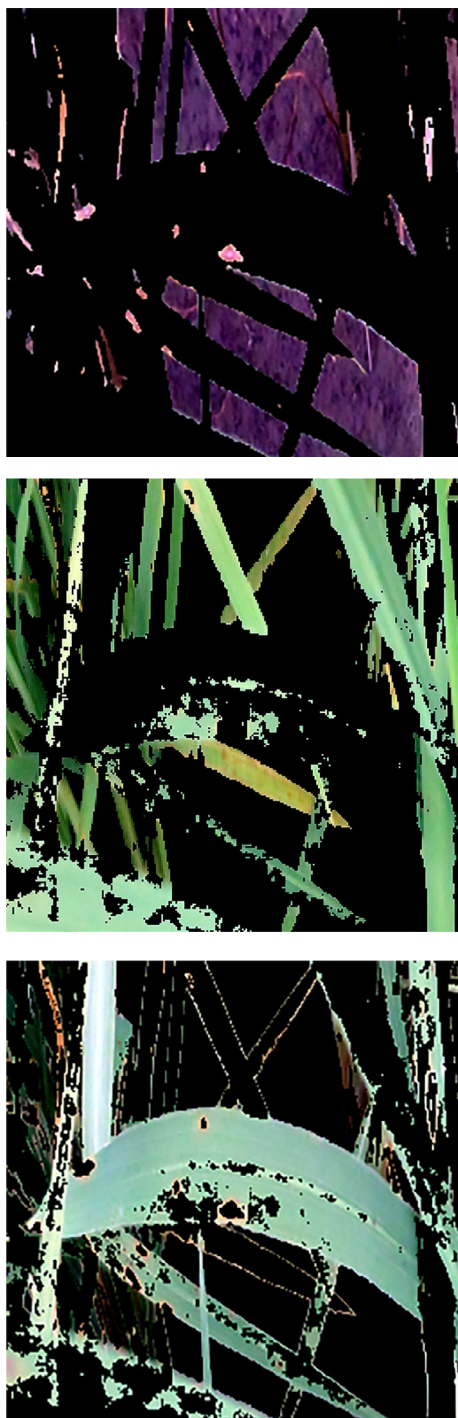
1. The KNN algorithm is slow since it reviews all the instances each time.
2. The algorithm is vulnerable to dimensionality.
3. The algorithm is sensitive to irrelevant and correlated attributes.
4. A wrong choice of the distance or the value of *k* degrades the performance.

The KNN algorithm is improved as follows and the classification operation begins:

K-means algorithm is used to form clusters, and the classification will be based on the centers of this new set of clusters. Thus, classifying a new instance into one of the *k* clusters instead of comparing it to the initial *n* instances divides the computation time of the algorithm by *k/n*. Finally, the distance between a given instance and the center of each cluster is restricted to significant attributes and weighted by their reliability coefficients.



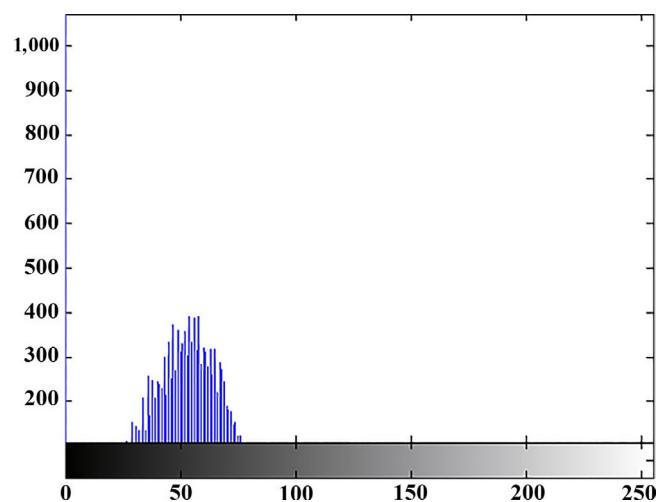
**FIGURE 6** Labeled image in *Lab* space



**FIGURE 7** Output images of the algorithm. Cluster 1: Identify the diseased points on the rice leaf and mark those points. Cluster 2: Light green spots on the leaf and stem of the plant and cluster 3: dark green spots on the plant leaf

### 3 | RESULTS AND DISCUSSION

In this research, with the improvement of KNN algorithm by K-means, a new, quick and accurate method for diagnosis of rice disease was developed using image processing. The results of this study showed that the dynamic range of gray surfaces should be increased to determine the damaged parts of the rice leaf. Therefore, the numeric



**FIGURE 8** Class I clustering histogram (determining the diseased points on the plant leaf)

data type of images changed from uint 8 type to double data type. Rewar, Singh, Chhipa, Sharma, and Kumari (2017), and Al Bashish, Braik, and Bani-Ahmad (2011) also used double images to diagnose various plant diseases under controlled, laboratory conditions and their input images were RGB.

Moreover, the *Lab* color space enabled us to separate the colors in the image and, using the K-means clustering algorithm, identify the diseased locations and the spots that have changed color on the rice leaf. The K-means method is one of the data mining techniques used in machine vision. Clustering is an uncontrolled learning method that does not rely on predefined categories or specific features as objectives, and places instances with the same amount of data together in one group. The objective function in the K-means clustering was calculated using equation (4):

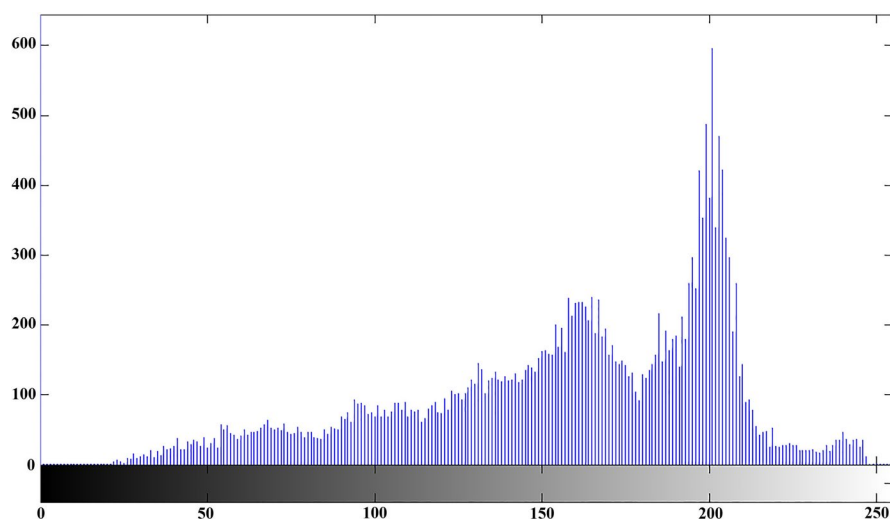
$$J = \sum_{j=1}^k \sum_{i=1}^n \|x_i^j - c_j\|^2 \quad (4)$$

where  $\|x_i^j - c_j\|$  is the measure of the distance between the points and  $c_j$  is the center of the  $j^{th}$  cluster. In the designed algorithm, at first  $K$  points were selected as the centers of the clusters. Then, each sample data were attributed to the cluster whose center had the smallest distance to that data. Finally, the average of the points belonging to each cluster was assigned to one cluster and for each cluster; a new point was calculated as the center.

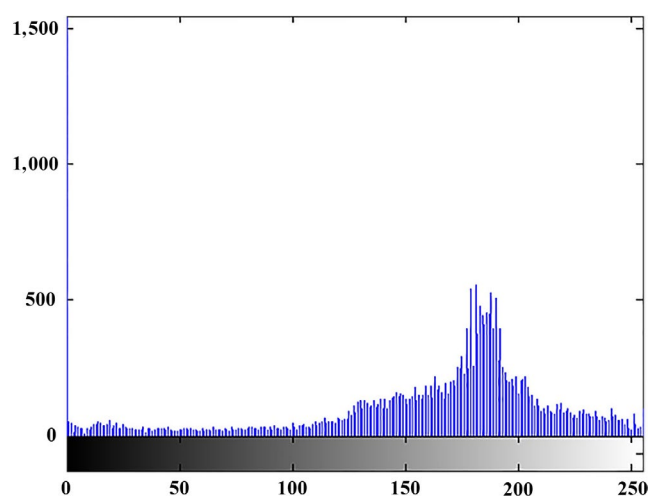
Figures 8, 9, and 10 illustrate the histogram of clustering of images in three classes, respectively. Gray surface area in class 1 clustering in Figure 8 is the threshold value for separating disease spots on rice leaves and other parts of the image. In the Figures 8, 9, and 10 the horizontal axis shows the gray surface level and the vertical axis the number of image pixels.

#### 3.1 | Accuracy of the designed algorithm

Two criteria of identification number and quality of blast spots were used to assess the ability and accuracy of the proposed algorithm. The



**FIGURE 9** Class II clustering histogram (light green)



**FIGURE 10** Class III clustering histogram (dark green)

results of the confusion matrix for the 500 sample images tested are given in Table 1. The main diagonal of the matrix shows the accurate results obtained by the algorithm (Improved KNN algorithm by K-means).

Three factors of sensitivity, specificity, and accuracy were investigated to determine the efficiency of the designed algorithm.

1. Sensitivity: the number of accurate categories per class/total number of each class.
2. Specificity:  $((\text{Number of inaccurate categories of each class} / \text{total number of other classes}) \times 100) - 100$ .
3. Total accuracy: total number of classes/total number of accurate categories in all classes.

The amount of each of the above-mentioned factors is shown in Table 2:

In similar studies, the accuracy of the diagnostic algorithm for various rice diseases was obtained by Venugoban and Ramanan

Output	determining the diseased spots	Number of disease spots	pixels other than the diseased spots
Number of diseased spots	20	460	0
The quality of determining the diseased spot	480	40	0
Pixels other than the diseased spots	0	0	500

**TABLE 1** The results of testing 500 sample images to determine the number and quality of blast spots

Image categories	Statistical factors		Total accuracy (%)
	Sensitivity (%)	Specificity (%)	
Number of diseased spots	92	91.70	94
The quality of determining the diseased spot	96	95.65	
Pixels other than the diseased spots	100	100	

**TABLE 2** statistical factors of sensitivity, specificity and total accuracy

(2014), Zainon (2012), and Qing et al. (2014) to be 90, 92.5, 69.76, and 85.20%, respectively. Compared to other researchers' investigations, this study reported a higher percentage of overall accuracy (94%) in field conditions. In addition, the time spent on the implementation of the algorithm designed to detect rice disease in one of the clusters was 5.894 s.

## 4 | CONCLUSION

The results of this study showed that image processing and machine vision have a high potential for determining blast disease in rice plant under field conditions. Moreover, the KNN machine learning method improved by K-means can be used as an effective diagnostic mechanism for the disease. This method is faster and less expensive than other methods and it is also nondestructive. The observations carried out show that the accuracy of the new method of machine vision is at an acceptable level and the results of this method can be relied upon. Besides that, by carefully managing the fields and by the timely prevention of the rice blast spread in the northern fields of the country, it is possible to avoid the overuse of chemical pesticides resulting in increased production costs, environmental pollution, and reduction in production per hectare.

## ACKNOWLEDGEMENTS

This study was conducted with the support of faculty members, engineers, and technicians at the Tonekabon-Chaparsar Rice Research Center (Mazandaran province, Iran). Thus, their support and cooperation in doing this research is appreciated.

## CONFLICT OF INTEREST

The authors have declared no conflict of interest.

## ETHICAL APPROVAL

This study does not involve any human or animal testing.

## INFORMED CONSENT

Written informed consent was obtained from all study participants.

## ORCID

Ezzatollah Askari Asli-Ardeh  <https://orcid.org/0000-0002-4048-2046>

## REFERENCES

Agahi, K., Fotokian, M. H., & Younesi, Z. (2012). Study of genetic diversity and important correlations of agronomic traits in rice genotypes (*Oryza sativa* L.). *Iranian Journal of Biology*, 25(1), 97–110. (In Persian).

- Al Bashish, D., Braik, M., & Bani-Ahmad, S. (2011). Detection and classification of leaf diseases using K-means-based segmentation and. *Information Technology Journal*, 10(2), 267–275.
- Al-Saqer, S. M. (2012). A robust recognition system for pecan weevil using artificial neural networks. *American Journal of Applied Sciences*, 9(9), 1347–1353.
- Ashkani, S., Rafii, M. Y., Shabanimofrad, M., Ghasemzadeh, A., Ravanfar, S. A., & Latif, M. A. (2016). Molecular progress on the mapping and cloning of functional genes for blast disease in rice (*Oryza sativa* L.): Current status and future considerations. *Critical Reviews in Biotechnology*, 36(2), 353–367.
- Azarmdel, H., Mohtasebi, S. S., Jafari, A., & Muñoz, A. R. (2019). Developing an orientation and cutting point determination algorithm for a trout fish processing system using machine vision. *Computers and Electronics in Agriculture*, 162, 613–629. <https://doi.org/10.1016/j.compag.2019.05.005>
- Bakar, M. A., Abdullah, A. H., Rahim, N. A., Yazid, H., Misman, S. N., & Masnan, M. J. (2018). Rice leaf blast disease detection using multi-level colour image thresholding. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, 10(1–15), 1–6.
- Jahanbakhshi, A., Ghamari, B., & Heidarbeigi, K. (2017). Assessing acoustic emission in 1055I John Deere combine harvester using statistical and artificial intelligence methods. *International Journal of Vehicle Noise and Vibration*, 13(2), 105–117. <https://doi.org/10.1504/IJNV.2017.087906>
- Jahanbakhshi, A., & Kheiralipour, K. (2019). Carrot sorting based on shape using image processing, artificial neural network, and support vector machine. *Journal of Agricultural Machinery*, 9(2), 295–307. (In Persian). <https://doi.org/10.22067/jam.v9i2.70579>
- Kahar, M. A., Mutalib, S., & Abdul-Rahman, S. (2015). *Early detection and classification of paddy diseases with neural networks and fuzzy logic*. In Proceedings of the 17th International Conference on Mathematical and Computational Methods in Science and Engineering, MACMESE (pp. 248–257).
- Kaveh, M., Jahanbakhshi, A., Abbaspour-Gilandeh, Y., Taghinezhad, E., & Moghimi, M. B. F. (2018). The effect of ultrasound pre-treatment on quality, drying, and thermodynamic attributes of almond kernel under convective dryer using ANNs and ANFIS network. *Journal of Food Process Engineering*, 41(7), e12868. <https://doi.org/10.1111/jfpe.12868>
- Khosravi, V. (2006). *The Relationship Between The Amount Of Silica In The Flag Leaf And The Percentage Of Cluster Blast Contamination In Different Rice Cultivars*. Research report. Publications Rice Research Institute – Department of mazandaran. P. 25. (In Persian).
- Kubat, M. (2015). *An introduction to machine learning*, Vol. 681. Cham: Springer.
- Kumar, K. K., Poovannan, K., Nandakumar, R., Thamilasari, K., Geetha, C., Jayashree, N., ... Balasubramanian, P. (2003). A high throughput functional expression assay system for a defence gene conferring transgenic resistance on rice against the sheath blight pathogen, *Rhizoctonia Solani*. *Plant Science*, 165(5), 969–976. [https://doi.org/10.1016/S0168-9452\(03\)00271-1](https://doi.org/10.1016/S0168-9452(03)00271-1)
- Peyman, S. H., Bakhshipour Ziaratgahi, A., & Jafari, A. (2016). Exploring the possibility of using digital image processing technique to detect diseases of rice leaf. *Journal of Agricultural Machinery*, 6(1), 69–79. (In Persian).
- Pydipati, R., Burks, T. F., & Lee, W. S. (2006). Identification of citrus disease using color texture features and discriminant analysis. *Computers and Electronics in Agriculture*, 52(1–2), 49–59. <https://doi.org/10.1016/j.compag.2006.01.004>
- Qing, Y., Xian, D. X., Liu, Q. J., Yang, B. J., Diao, G. Q., & Jian, T. A. N. G. (2014). Automated counting of rice planthoppers in paddy fields based on image processing. *Journal of Integrative Agriculture*, 13(8), 1736–1745. [https://doi.org/10.1016/S2095-3119\(14\)60799-1](https://doi.org/10.1016/S2095-3119(14)60799-1)
- Rewar, E., Singh, B. P., Chhipa, M. K., Sharma, O. P., & Kumari, M. (2017). Detection of infected and healthy part of leaf using image processing



- techniques. *Journal of Advanced Research in Dynamical & Control System*, 9(1), 13–19.
- Samanta, R. K., & Ghosh, I. (2012). Tea insect pests classification based on artificial neural networks. *International Journal of Computer Engineering Science (IJCES)*, 2(6), 1–13.
- Venugoban, K., & Ramanan, A. (2014). Image classification of paddy field insect pests using gradient-based features. *International Journal of Machine Learning and Computing*, 4(1), 1–5. <https://doi.org/10.7763/IJMLC.2014.V4.376>
- Zainon, R. (2012). *Paddy disease detection system using image processing*. (Doctoral dissertation, UMP). Faculty of Computer System and Software Engineering. P. 120.

**How to cite this article:** Larijani MR, Asli-Ardeh EA, Kozegar E, Loni R. Evaluation of image processing technique in identifying rice blast disease in field conditions based on KNN algorithm improvement by K-means. *Food Sci Nutr*. 2019;7:3922–3930. <https://doi.org/10.1002/fsn3.1251>