Real-time Disease Detection in Rice Fields in the Vietnamese Mekong Delta

Thanh Tinh NGUYEN1,2, Ricardo OSPINA3, Noboru NOGUCHI3, Hiroshi OKAMOTO1 and Quang Hieu NGU1

1 Graduate School of Agriculture, Hokkaido University, Kita-9, Nishi-9, Kita-Ku, Sapporo, Hokkaido, 060-8589, Japan
2 Department of Agricultural Engineering, Can Tho University, Ninh Kieu District, Can Tho City, Vietnam
3 Research Faculty of Agriculture, Hokkaido University, Kita-9, Nishi-9, Kita-Ku, Sapporo, Hokkaido, 060-8589, Japan

(Received May 11, 2020; Accepted November 16, 2020)

This study introduces an image processing method capable of performing real-time detection of two common diseases, leaf blast (LB) disease and bacterial blight (BB) disease, in the paddy fields of the Vietnamese Mekong Delta (VMD). The input images were recorded with an RGB camera. The discrimination of the diseases on rice leaves was obtained by an image processing method based on the extraction of texture and color features from disease lesions, in conjunction with either the Gaussian Naïve Bayes classifier or the K-Nearest Neighbors (KNN) algorithm, to classify the disease into various categories. Both methods perform real-time detection of LB and BB disease in the early stages of development with uncontrolled light conditions in rice fields. Our results show that Gaussian Naïve Bayes is simple but effective, with a shorter processing time and higher detection accuracy than KNN.

Keywords: Gaussian Naïve Bayes, K-Nearest Neighbors, machine vision, real-time detection, smart agriculture.

INTRODUCTION

In rice cultivation, the majority of pesticides are used focusing on three main problems: herbicides for weed management, fungicides for disease management, and insecticides for pest management. The Food and Agricultural Organization reports ratios of pesticide application in rice fields in Vietnam on 2002 of 25.3% herbicides, 32.6% fungicides, and 40.3% insecticides (FAO, 2005). Fungal leaf blast (LB) disease and bacterial leaf blight (BB) disease are common and well-known diseases found in the rice fields of Vietnam.

The rice blast fungus <i>Pyrillium oryzae</i> causes a fungal disease common in rice fields (Francisco and Zahrul, 2003). Depending on the site of the symptoms, the rice blast disease is referred to as leaf blast, collar blast, node blast or neck blast. In the early stages of LB, the lesions on the leaf blade are elliptical or spindle-shaped, with brown borders and gray centers, as shown in Fig. 1a.

The bacterium <i>Xanthomonas oryzae</i> causes a bacterial disease (Francisco and Zahrul, 2003) characterized by a water-soaked lesion that usually starts at the leaf margins, a few centimeters away from the tip, and spreads towards the leaf base. The affected areas increase in length and width, and become yellowish to light brown due to dryness, with a yellowish border between dead and green areas of the leaf, as shown in Fig. 1b.

Several studies have reported that LB and BB are the most harmful rice diseases and have caused yield losses. For example, rice disease resulted in a yield reduction of 1–10% from 456 farmer’s fields surveyed across tropical Asia on during 1987–1997 (Savary et al., 2000), and rice yield losses ranging from 50% to 85% have been reported in the Philippines by the International Rice Research Institute (IRRI, 2020). In Vietnam, grain yield losses of 38.21% to 64.57% due to neck blast disease have been reported (Hai et al., 2007).

Thus, plant protection focusing on managing diseases and controlling the amount of fungicides to be applied has been an important part of research over the last few years. The literature includes many reports on the detection of rice diseases. Ks and Sahayadhas (2018) report on the prediction of early symptoms of BB and brown spots on rice plants by separating leaf color, signs and illumination from different color channels. This algorithm makes it easy to perform final feature analyses, however, it cannot predict diseases with symptoms in similar colors. Bakar et al. (2018) describe an integrated method for the detection of LB using three categories: infection stage, spreading stage, and worst stage. This is possible by analyzing the Hue, Saturation and Value color spaces with multi-level thresholding, and identifying classified regions of interest during image segmentation. This technique successfully detects the disease based on images taken in uncontrolled environments, however, it is not suitable for the detection of other diseases with similar features. In another study, Islam et al. (2018) present the Gaussian Naïve Bayes method to classify the disease based on the percentage of RGB values of the affected portion using image processing. This method has successfully detected three rice diseases, brown spot, rice bacterial blight, and rice blast, and has a fast processing time and high accuracy, however, it cannot detect a disease with similar color features but a different shape.
Some recent research shows robustness in the detection of many kinds of diseases on rice leaves using deep learning convolutional neural network (CNN) algorithms (Lu et al., 2017; Mique and Palaoag, 2018). These algorithms show high accuracy, however, their application and testing are limited to an individual leaf, and the processing time is not good for real-time application. Thus, current research applications for rice disease detection focus exclusively on the development of a tool to help humans evaluate diseases. To solve these problems that are a simple algorithm, short image processing time and high detection accuracy. The present study used an algorithm based on the mini-bounding rectangular blob as a shape feature, and the mean values of three color channels as the color features for red, green and blue (RGB) color spaces and hue, saturation and intensity (HSI) color spaces. These features were used by the Gaussian Naïve Bayes classifier and the K-Nearest Neighbors (KNN) algorithm for the real-time detection of the two dataset training diseases, LB and BB, which are native to the rice fields of the VMD. In addition, the processing time and accuracy of each method were estimated as the primary parameters for selecting the method to be used in the development of a smart sprayer for rice fields. Besides, detection LB and BB diseases, which normally appear at the same time in the field, is helpful for precise selection of the kind of fungicide for effective treatment.

MATERIALS AND METHODS

The present research addresses early stage disease detection, in which small lesions appear on the leaves. Images were collected during the early stage of disease from several places in the VMD, including Dinh Thanh An Giang, Co Do Can Tho, and Chau Thanh Trà Vinh. A normal RGB camera, iPhone 6S rear-facing camera with 12-megapixel (4,032×3,024 pixels) not using flashlight was used to capture images of the rice fields under uncontrolled illumination conditions such as mornings and afternoons on different days (light intensity during capture around 500–90,000 lx). All input images were stored in a computer and then used for training and testing the image processing detection of LB and BB. The output data can be used by agriculture experts to take further action.

The dataset images of the two diseases used for training cover a single lesion on the leaf, an example shown in Fig. 2. For the image processing, this research used C++ programming language (Microsoft Visual Studio Community 2017) and OpenCV (OpenCV-3.4.1) on Windows 10 running in a laptop computer (Lenovo, ThinkPad X220). Through image processing techniques such as background subtraction, finding the contours, and the application of the masking method, it is possible to extract the basic characteristics of the disease lesions, including the average value of each color channel (RGB) and the ratio between the height and width of the minimum rectangle bounding blob. The transformation of RGB color spaces to HSI color spaces (Nnolim, 2015) can be derived using Eqs. (1) to (6). All feature values of each image were used to calculate the parameters for each feature required by the classifier based on the Gaussian Naïve Bayes classifier and the KNN algorithm. Unlike the training images, the test images include many disease lesions of different sizes and growth stages, as well as uncontrolled light conditions, so that we would be able to estimate the applicability of the program.

$$r = \frac{R}{R+G+B}, \quad g = \frac{G}{R+G+B}, \quad h = \frac{G}{R+G+B}$$

$$h \in [0, \pi] \text{ for } b \leq g$$

$$h = \cos^{-1}\left(0.5\frac{r-g+(r-b)}{[(r-g)^2+(r-b)(b-g)]^{0.5}}\right)$$

$$h \in [\pi, 2\pi] \text{ for } h > g$$

$$h = 2\pi - \cos^{-1}\left(0.5\frac{r-g+(r-b)}{[(r-g)^2+(r-b)(b-g)]^{0.5}}\right)$$

$$s = 1 - \text{MIN}(r, g, b)$$

$$i = \frac{R+G+B}{3\times255}$$

$$H = h\times180/\pi, \quad S = s\times100, \quad I = i\times255$$

Where:

- $R$ : Red channel color value
- $G$ : Green channel color value

Fig. 1 Rice leaf disease. (a) Leaf blast disease. (b) Bacterial blight disease.

Fig. 2 Example of dataset training images. (a) LB dataset training. (b) BB dataset training.
The image processing algorithm for the detection of the two kinds of diseases using the Gaussian Naïve Bayes classification algorithm and the RGB color space and shape is shown in Fig. 3. The Gaussian Naïve Bayes classifier is a simple probabilistic classifier based on applying Bayes' theorem. Gaussian Naïve Bayes considers each and every feature variable as an independent variable. This classifier can be trained very efficiently in supervised learning and requires only a small amount of training data that are necessary for classification. The input image in the RGB color space collected with the RGB camera and shown in Fig. 3a was resized to a fixed resolution (450×600 pixels) to improve the memory storage capacity and to reduce computational time. A threshold filter was used to convert the RGB image into a binary image in which the white pixels correspond to disease and the black pixels to a healthy rice leaf. The setting value of the appropriate threshold filter was adjusted and selected manually for the training images. Figure 3b shows the resulting binary image; it still has noise from the motion between the camera and the object, improper shutter opening, atmospheric disturbances and misfocusing (Archana and Amit, 2014). This noise was smoothed by applying the morphological operations of erosion and dilation. The erosion operation eliminated small isolated white pixel areas, and the dilation operation merged areas where large groups of white pixels were found next to each other. The erosion operation used a rectangular shape structuring element of 2×2 pixels. The dilation operation made white areas denser and eliminated breaks using a rectangular shape structuring element of 1×1 pixels. These morphological operations transformed the binary image shown in Fig. 3b into the smoothed result shown in Fig. 3c.

The next step was to find the contours from the image shown in Fig. 3c (Bradski and Kaehler, 2008; Brahmbhatt, 2013). The contours produced the characteristics of the bounding blob, namely: the height $h$, the width $w$, the area of the minimum bounding rectangle $S_r$, and the ratio $w/h$. The resulting ratios are shown in Fig. 3d. Parallel to finding the characteristics of the bounding blob, the subtraction of the background from the original RGB image took place based on the contours; all the background outside of the contour was removed using a masking method. It was then possible to calculate the RGB mean values inside each contour. The mean values of the three channels are shown in Fig. 3e. Finally, the classifier used was based on the Gaussian Naïve Bayes classifier (Mitchell, 1997) and is described by Eqs. (7) to (11) with the training dataset in the RGB color space. The average of each feature given by Eq. (7) and the variance given by Eq. (8) were calculated from the dataset for training and were then applied to Eq. (9) to calculate the likelihood of each feature. Equation (9) was applied to each contour in the image with red mean value $R$, green mean value $G$, blue mean value $B$ and ratio $K$. Equation (10) is the prediction value for blast disease and Eq. (11) is the prediction value for blight disease. The comparison between Posterior (Blast) and Posterior (Blight) was used to decide the kind of disease based on the higher posterior value. If Posterior (Blast) is higher, then a blast disease was detected. If Posterior (Blight) is higher, then a blight disease was detected. The result is shown in Fig. 3f.

\[
\mu = \frac{1}{n} \sum_{i=1}^{n} x_i, \quad (7)
\]

\[
\text{var} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \mu)^2, \quad (8)
\]
The image processing algorithm for the detection of the two kinds of diseases by the Gaussian Naïve Bayes classification algorithm from the HSI color space and shape is shown in Fig. 4. Similar to the method used in Fig. 3, the input image in the RGB color space shown in Fig. 4a was resized to a fixed resolution (450 x 600 pixels) to improve memory storage capacity and reduce computational time. The transformation from the RGB to the HSI color space was obtained by using Eqs. (1) to (6); the result of the transformation is shown in Fig. 4b. A threshold filter was used to convert the HSI image into a binary image. The contours produced the characteristics of the bounding blob: $h$, $w$, $S_r$, and $w/h$. The resulting ratios are shown in Fig. 4d. The background in the HSI image was subtracted by using a masking method, and the HSI mean values were then calculated inside each contour. The mean values of the three channels are shown in Fig. 4e. Finally, the Gaussian Naïve Bayes classifier was used with the training dataset in the HSI color space. The result is shown in Fig. 4g.

Figure 5 shows an explanation of the feature extraction method. Figure 5a shows a detail of the minimum rectangle bounding blob analysis from Fig. 3d and Fig. 4e, which is a calculation of the $w/h$ ratio as a feature of texture and of the area $S_r$ as a characteristic used to filter noise. Figure. 5b illustrates how to extract the mean values (average values) of the three RGB colors from a single disease lesion using the masking method shown in Fig. 3e, and Fig. 5c illustrates how to extract the mean values (average values) of the three HSI colors from a single disease lesion using the masking method shown in Fig. 4f. In each disease lesion corrected from the contour, the masking method was applied to remove the background and keep the color inside the contour only, then the rectangle bounding function was used to crop each single disease lesion into small images. This helped to reduce the processing time to calculate the average value for each color channel inside the contour. Equations (12) and (13) explain how to obtain the mean value for each color channel. The mean (average) value of each color channel can be calculated as the sum of the pixels in a single-color channel divided by the sum of the pixels (not including pixels of value zero).

$$p(\text{Red} | R) = \frac{1}{\sqrt{2\pi \cdot \text{var}_R}} e^{-\frac{(\text{Red} - \overline{\text{Red}})^2}{2 \cdot \text{var}_R}}$$

$$p(\text{Green} | G) = \frac{1}{\sqrt{2\pi \cdot \text{var}_G}} e^{-\frac{(\text{Green} - \overline{\text{Green}})^2}{2 \cdot \text{var}_G}}$$

$$p(\text{Blue} | B) = \frac{1}{\sqrt{2\pi \cdot \text{var}_B}} e^{-\frac{(\text{Blue} - \overline{\text{Blue}})^2}{2 \cdot \text{var}_B}}$$

$$p(\text{Ratio} | K) = \frac{1}{\sqrt{2\pi \cdot \text{var}_K}} e^{-\frac{(\text{Ratio} - \overline{\text{Ratio}})^2}{2 \cdot \text{var}_K}}$$

$$P(\text{Posterior} \text{Blast}) = p(\text{Red} | R) \cdot p(\text{Green} | G) \cdot p(\text{Blue} | B) \cdot p(\text{Ratio} | K)$$

$$P(\text{Posterior} \text{Blight}) = p(\text{Red} | R) \cdot p(\text{Green} | G) \cdot p(\text{Blue} | B) \cdot p(\text{Ratio} | K)$$

Where:
- $\overline{\text{Red}}$, $\overline{\text{Green}}$, $\overline{\text{Blue}}$: Average value of each feature
- $\text{var}_R$, $\text{var}_G$, $\text{var}_B$: Variance value of each feature
- $R$: Red channel color mean value
- $G$: Green channel color mean value
- $B$: Blue channel color mean value
- $K$: Bounding rectangle width to height ratio $K = w/h$
DISEASE DETECTION IN RICE

Fig. 5 Feature subtraction. (a) Minimum bounding analysis. (b) RGB mean value analysis. (c) HSI mean value analysis.

The KNN method of classifying objects requires only two parameters to tune, k and the distance metric, to achieve sufficiently high classification accuracy. Thus, in KNN-based implementations, it is critical to find the best choice of k and the distance metric to compute the nearest. Generally, larger values of k reduced the effect of noise but made boundaries between classes less distinct. The special case in which the class is predicted to be the class of the closest training sample (i.e., when \( k = 1 \)) is called the nearest neighbor algorithm. In the present study, the different tested values of k were 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10, and the distance metrics were the Euclidean distances. A brief explanation of the Euclidean distance metric is described by Eqs. (14) to (17) as follows.

\[
\begin{align*}
    r &= \frac{R - \text{min}R}{\text{max}R - \text{min}R} \\
    g &= \frac{G - \text{min}G}{\text{max}G - \text{min}G} \\
    b &= \frac{B - \text{min}B}{\text{max}B - \text{min}B} \\
    k &= \frac{K - \text{min}K}{\text{max}K - \text{min}K}
\end{align*}
\]

(14)

\[
\begin{align*}
    r_t &= \frac{R_t - \text{min}R_t}{\text{max}R_t - \text{min}R_t} \\
    g_t &= \frac{G_t - \text{min}G_t}{\text{max}G_t - \text{min}G_t} \\
    b_t &= \frac{B_t - \text{min}B_t}{\text{max}B_t - \text{min}B_t} \\
    k_t &= \frac{K_t - \text{min}K_t}{\text{max}K_t - \text{min}K_t}
\end{align*}
\]

(15)

\[
\begin{align*}
\text{Distance (color)} &= \sqrt{(r - r_t)^2 + (g - g_t)^2 + (b - b_t)^2} \\
\text{Distance (color-shape)} &= \sqrt{(r - r_t)^2 + (g - g_t)^2 + (b - b_t)^2 + (k - k_t)^2}
\end{align*}
\]

(16)

(17)

Where:
- \( r, g, b, k \) temporary values of red, green, blue and ratio of each testing dataset
- \( r_t, g_t, b_t, k_t \) temporary values of red, green, blue and ratio of each training dataset
- \( \text{min}R, \text{min}G, \text{min}B, \text{min}K \) minimum values of red, green, blue and ratio of training dataset
- \( \text{max}R, \text{max}G, \text{max}B, \text{max}K \) maximum values of red, green, blue and ratio of training dataset
- \( R_0, G_0, B_0, K_0 \) features values of red, green, blue and ratio of each training dataset
- \( R, G, B, K \) features values of red, green, blue and ratio of each testing dataset

The KNN algorithm consists of two steps: a training step and a testing step. In the training step, the training examples were vectors (each with a class label) in a multi-dimensional feature space. In this step, the feature vectors and the class labels of the training samples were stored. In the testing step, \( k \) was a user-defined constant, and a test point was classified by assigning a label, which was the most recurrent among the \( k \) training samples nearest to that query point. In other words, the KNN method compared query points based on their distance to points in the training dataset. This is a simple yet effective way of classifying new points.
The minimum and maximum values of each feature in the training dataset were applied to Eqs. (14) and (15) to calculate temporary values, namely: \( r_c, g_c, b_c, r_k, g_k, b_k \) and \( k \). These values were then applied to Eq. (16) to calculate the distance using only the color feature; Eq. (17) uses both color and shape features. In each testing component, the distance was calculated for all the components in the training data, and the assignment for each class which has the shortest distance then takes place.

**RESULTS AND DISCUSSION**

In the present research, a total of 116 images were used for training: 58 depicting blast diseases and 58 depicting bacterial blight diseases. The training images were carefully selected, taking different illumination, different lesion sizes and different growth stages of the rice into consideration. A set of 25 images (including a total of 110 single-lesion diseases) was used for detection and accuracy evaluation. For the background subtraction with the RGB color space method, both the green and blue channels can be used to remove the healthy leaves by a threshold filter method. However, the experimental results show that the threshold filter using only the blue channel is more effective, because no rice disease affecting the green leaves (healthy leaves) turn them into blue color, the disease turns the green leaves into a yellowish to light brown or lesions (healthy leaves) turn them into blue color, the disease turns the green leaves into a yellowish to light brown or lesions (healthy leaves) turn them into blue color, the disease turns the green leaves into light brown or lesions (healthy leaves) turn them into blue color, the disease turns the green leaves into light brown or lesions (healthy leaves) turn them into blue color, the disease turns the green leaves into light brown.

Table 1 refers to results using the RGB color space and Table 2 refers to results using the HSI color space. In each color space, the accuracy and processing time were calculated with different features applied to the Gaussian Naïve Bayes equation. In this study, we separated the features in three ways: using the color feature \((R, G, B)\) only, using the shape feature \((K)\) only, and using both color and shape features \((R, G, B, K)\) combined. Our results show that detection accuracy was increased when the color and shape features were combined, and the RGB color space yielded better detection accuracy than the HSI color space. Increasing the number of features for the classifier also increased the processing time slightly (by only 0.01 seconds); this could be acceptable to achieve a higher accuracy criterion.

Table 3 shows the results when using the KNN classifier, summarizing the detection accuracy with different \(k\) values (from 1 to 10) for the color feature only, and for the color and shape features using both the RGB and HSI color spaces. Similar to the Gaussian Naïve Bayes classifier, the KNN classifier also increased detection accuracy when combining the color feature (RGB or HSI) and the shape feature.

The detection results presented in Table 4 suggest that the Gaussian Naïve Bayes classifier method using both color and shape features (RGBK) was more appropriate for the development of a smart sprayer system because it had a shorter image processing time. This can reduce the

**Table 1** Results of Gaussian Naïve Bayes classifier with RGB color space.

| Color feature (RGB) | Shape feature (K) | Color and shape features (RGBK) |
|---------------------|------------------|-------------------------------|
| Average detection accuracy (%) | 82.7 | 78.2 | 90 |
| Average processing time per image (s) | 0.23 | 0.21 | 0.24 |

**Table 2** Results of Gaussian Naïve Bayes classifier with HSI color space.

| Color feature (HSI) | Shape feature (K) | Color and shape features (HSIK) |
|---------------------|------------------|-------------------------------|
| Average detection accuracy (%) | 80.0 | 78.2 | 83.6 |
| Average processing time per image (s) | 0.23 | 0.21 | 0.24 |

**Table 3** Average detection accuracy of KNN classifier with RGB and HSI color spaces.

| Color and shape features | \(k = 1\) | \(k = 2\) | \(k = 3\) | \(k = 4\) | \(k = 5\) | \(k = 6\) | \(k = 7\) | \(k = 8\) | \(k = 9\) | \(k = 10\) |
|-------------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| RGB (%) | 77.27 | 85.45 | 77.27 | 79.09 | 79.09 | 77.27 | 78.18 | 76.36 | 77.27 | 78.18 |
| RGBK (%) | 80.0 | 80.0 | 81.82 | 81.82 | 81.82 | 80.91 | 80.0 | 80.91 | 80.91 | 80.91 |
| HSI (%) | 74.54 | 72.72 | 72.27 | 75.45 | 76.36 | 73.63 | 73.63 | 72.72 | 71.82 | 72.72 |
| HSIK (%) | 80.0 | 80.91 | 80.91 | 80.91 | 80.91 | 81.82 | 81.82 | 80.82 | 80.0 | 80.0 |

**Table 4** Comparison between Gaussian Naïve Bayes and KNN.

| | Gaussian Naïve Bayes | KNN |
|----------------|------------------|-----|
| Average detection accuracy (%) | 80.0 | 80.8 |
| Average detection accuracy (%) | 83.6 | 80.9 |
| Processing time (s) | 0.24 | 0.26 |
responding time from the image capture and processing to decide which nozzle of the sprayer should be activated with correct timing into the treatment area, thus increasing the accuracy of the sprayer. Moreover, it had an average detection accuracy (90%) greater than that of the KNN classifier method (80.8%). The rational explanation for these results is that, in the present study, the accuracy of both methods depended on the size of the dataset used for training. Accuracy could be improved by using a bigger dataset with a large variance in the illumination conditions, but this also resulted in longer training and detection times, as shown in Table 5. For the KNN method algorithm in particular, which calculated the distance with each component in the training dataset, the processing time becomes longer and the method then becomes harder to use for real-time detection applications. The image processing time for individual images using the Gaussian Naïve Bayes classifier was 0.24 seconds, which makes it possible to apply this method for the real-time detection of diseases and spraying of fungicides according to the speed of a machine working in rice fields under the same conditions as transplanting (Sato et al., 1996). However, when the lighting conditions changed significantly, the B (blue) filter setting needed to be changed as well. Another disadvantage of the Gaussian Naïve Bayes method is that, in the present study, accurate disease detection depended on the method being applied during the early stages of disease symptoms. The highest accuracy was obtained when disease lesions were not covered by leaves and the lesions did not blend.

Figure 6 shows two examples of the present experimental results of disease detection by using Gaussian Naïve Bayes method. The detected LB disease is shown in red and the detected BB in yellow. The disease detection shown in Fig. 6 corresponds to images captured under uncontrolled conditions: multiple leaves, different sizes of lesions, stages of growth, orientations and light conditions, and complex backgrounds. Figure 6a shows fungal leaf blast disease in the tillering stage of the rice plant, while Fig. 6b shows bacterial blight disease in the flowering stage of the rice plant. Almost all disease lesions were detected successfully and classified correctly according to the kind of disease. However, some disease lesions could not be detected successfully. This problem was caused by the background subtraction, which failed to isolate all disease lesions and show them completely.

Table 5 Comparison of processing times between Gaussian Naïve Bayes and KNN when increasing the training dataset.

| Method               | Training dataset: 116 | Training dataset: 232 |
|----------------------|-----------------------|-----------------------|
|                      | Training and testing processing time (s) |                          |
| Gaussian Naïve Bayes method | 0.24         | 0.28                 |
| Dataset test: 110    |                        |                       |
| KNN method           | 0.26               | 0.34                 |
| Training and testing processing time (s) |                          |
| Dataset test: 110    |                        |                       |

Figure 6 LB and BB disease detection results by Gaussian Naïve Bayes method. (a) LB disease in the tillering stage. (b) BB disease in the flowering stage (LB diseases is shown in red and BB diseases in yellow).

VMD is the biggest rice cultivate in Vietnam, which is 3.8 million hectares including 12 provinces. Especially, some provinces can yield triple-cropped rice per year such as An Giang, Dong Thap, Vinh Long, Can Tho, Hau Giang.
and Tra Vinh. Therefore, the amount of fungicide application in a province with high density of crop rice becomes high on a year basis. By successfully detecting two common dangerous diseases in this study, a smart sprayer machine will be developed; which can spray only on a specific treatment area, which can help to reduce the amount of the fungicides by 70–80%. This helps to reduce the chemical residue effects in the environment and to increase the income of the farmers in this area.

CONCLUSIONS

This study demonstrates that the Gaussian Naïve Bayes classifier method is better than the KNN algorithm method for detecting two kinds of diseases, rice blast disease and bacterial blight disease, in the early stages under uncontrolled illumination conditions. This study also discusses the effects of color features (RGB, HSI) and shape features on the detection accuracy of the two methods. Both methods are able to detect both kinds of disease, though with different levels of accuracy and different processing times. The Gaussian Naïve Bayes method has the advantages of high accuracy and a low processing time. Although the present results were satisfactory, however, this method might yield lower accuracy when disease lesions become wide and blend, or when lighting conditions change.

ACKNOWLEDGEMENTS

The authors thank Dr. Nguyen Thanh Tam (Mekong Delta Development Research Institute, Can Tho University) for sharing his knowledge about the kinds of diseases in rice fields. This study was funded in part by Can Tho University Improvement Project VN14-P6, supported by a Japanese Official Development Assistance (ODA) loan.

REFERENCES

Archana, S., Amit, C. 2014. Comprehensive review of denoising techniques in image restoration. Intern. J. Innov. Res. Comput. Commun. Eng. 2: 4597–4602.

Bakar, M. A., Abdullah, A., Rahim, N. A., Yazid, H., Misman, S., Masnan, M. 2018. Rice leaf blast disease detection using multi-level colour image thresholding. J. Telecomm. Electron. Comput. Eng. (JTEC) 10: 1–6.

Bradski, G., Kaehler, A. 2008. Learning OpenCV. O’Reilly Media, Inc., Sebastopol, p 222–264.

Brahmbhatt, S. 2013. Practice OpenCV. Apress Media, LLC, New York, p 67–93.

Cover, T., Hart, P. 1967. Nearest neighbor pattern classification. IEEE Trans. Inf. Theory 13: 21–27.

FAO. 2005. Food and Agriculture Organization of the United Nations regional office for Asia and the Pacific. http://www.fao.org/3/aD40c/aD40c0b.htmlContents. Avaiable at 2. Country reports, 2.17 Viet Nam.

Francisco, E., Zahirul, I. 2003. Diagnosis of Common Diseases of Rice. IRRI, Phillipines, p 13–22.

Hai, L., Kim, P. V., Du, P. V., Thuy, T. 2007. Grain yield and grain-milling quality as affected by rice blast disease (Pyricularia grisea), at My Thanh Nam, Cai Lay, Tien Giang. Omonrice 15: 102–107.

IRRI. 2020. Disease- and pest-resistant rice. https://www.irri.org/disease-and-pest-resistant-rice. Accessed April 20, 2020.

Islam, T., Sah, M., Baral, S., RoyChoudhury, R. 2018. A faster technique on rice disease detection using image processing of affected area in agro-field. Paper presented at the 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), 62–66.

Ks, A., Sahayadhas, A. 2018. Automatic rice leaf disease segmentation using image processing techniques. Intern. J. Eng. Technol. 7: 182–185.

Lan, Y., Yi, S., Zeng, N., Liu, Y., Zhang, Y. 2017. Identification of rice diseases using deep convolutional neural networks. Neurocomputing 267: 378–384.

Maque, E., Palaasong, T. 2018. Rice Pest and Disease Detection Using Convolutional Neural Network. Int. Conf. Inf. Sci. Syst.
DISEASE DETECTION IN RICE

Mitchell, T. 1997. Machine Learning. McGraw-Hill, New York, p 154-199.

Nnolim, U. A. 2015. Design and implementation of novel, fast, pipelined HS2RGB and log-hybrid RGB2HSI colour converter architectures for image enhancement. Microprocess. Microsyst. 39: 223-236.

Sato, J., Shigeta, K., Nagasaka, Y. 1996. Automatic operation of light tractor with sprayer in paddy fields. JARQ (Japan) 30: 239-244.

Savary, S., Willocquet, L., Elazegui, F. A., Castilla, N. P., Teng, P. S. 2000. Rice pest constraints in tropical Asia: quantification of yield losses due to rice pests in a range of production situations. Plant Dis. 84: 357-369.