A Rapid Recognition Method for Rice False Smut based on HOG Features and SVM Classification

Naila S¹, b, Yu JJ ¹, b, Yang N ¹, a, Kashif H², d, Tang J³ and Wang AY ³
¹School of Electrical and Information Engineering, Jiangsu University, Zhenjiang, China
²School of Microelectronics and Communication Engineering, Chongqing University, China
³State Key Laboratory of Rice Biology, China National Rice Research Institute, Hangzhou, China

¹yangn7410@163.com, ¹15950500921@163.com, ¹nailashaheen10@yahoo.com,
²kashifhussain180@yahoo.com

Abstract. A rapid detection method for the detection of false smut is introduced based on HOG features and SVM classification to overcome the drawbacks of false smut detection like time wasting and inaccurate recognition. In this paper, original images of false smut used, which directly taken from the field of rice. First of all, the image false smut disease rice acquired by image preprocessing was based on the color threshold to obtain the (ROI) region of interest from healthy and unhealthy pictures of rice. Secondly, the gamma-correction performed on the extracted ROI unhealthy and background region of healthy rice images. The HOG features extracted to form a high-dimensional HOG feature vector. Third, the PCA analysis performed on high-dimensional HOG features, and the original 1764-dimensional feature vector reduced to a two-dimensional vector that can represent most vectors to achieve the purpose of dimension reduction. Finally, the dimension-reduced HOG feature vector sent to the SVM classifier for training. The expected recognition accuracy under the binomial kernel function is 80.6%, and the actual test set recognition is 78.56%.

1. Introduction
Rice false smut caused by fungus is one of the most severe diseases in rice. The disease occurs in more than 40 countries, especially in rice-growing countries in Asia, such as China [1]. In recent years, false smut has a significant impact on rice production, which has also impacted agricultural economic development [2]. Therefore, it is necessary to propose a method for the rapid identification of false smut disease of rice.

At present, there are two main types of image recognition methods for crop diseases. The first type is weak machine learning classification recognition based on image segmentation. This method has the following steps form completion as preprocessing, segmenting, and feature extraction from the ROI (region of interest).

This image recognition method commonly includes the adaptive threshold method [3], unsupervised color clustering method [4], a primary morphological method [5], watershed algorithm [6], etc. This type of technique mainly recognizes different diseases through the difference in color, size, and shape of the crop lesion area. It requires high-definition picture quality and stable light and other conditions.
and only used in a steady environment such as a laboratory, and its practicality is not high. The other type is image recognition based on deep learning. This method can summarize the characteristics of the recognition content by itself and is more suitable for recognition in complex environments. Standard deep learning image recognition methods include convolutional neural network (CNN) method [7], artificial neural network (ANN) method [8], and backpropagation neural network (BPNN) method [9], etc. However, the accuracy of disease recognition based on deep learning depends on the number and quality of samples. It is challenging to establish an extensive sample disease image database in a short time. For large sample images, labeling and classification based on these images are also very time-consuming since the image recognition method based on deep learning requires a massive number of sample libraries (generally more than five thousand).

This paper chose weak machine learning combined with a custom threshold for image segmentation as the method of rice false smut recognition. (the above two methods’ speed are all faster than the manual recognition speed [10]).

In this paper, the false smut feature recognition method based on Histogram of Oriented Gradient (HOG) and Support Vector Machine (SVM) techniques used on false smut (diseased rice) images taken from the rice field as research object under natural conditions to overcome the traditional method problems and to improve the ease of use and practicality of detection method based on machine vision.

2. Materials and methods

2.1. Experimental materials

This article takes rice as a test object and captures images by using iPhone Xs Max mobile phones. A total of 725 photos of false smut obtained. After the image preprocessing and segmentation, 225 pictures used as recognition samples, and the remaining 500 pictures used as training samples. Cut the lesions (rice false smut balls) into training positive testers, and take the non-lesions as negative samples, that is, a total of 500 positive samples and negative samples (the ratio of positive trials to negative samples is 1.5: 1). The experimental environment is in MATLAB 2016b. The first picture of false smut, as shown in Figure 1.

![Figure 1. Original false smut diseased rice picture](image)

2.2. Image preprocessing

An image segmentation algorithm based on the color threshold used to segment the diseased area of the rice false smut image. The region of interest (ROI) in the rice smut disease image is gray and brown, and the background area is yellow and green. The image segmented by using the HSV channel. The specific algorithm described in Figure 2. The ROI for rice aspergillosis obtained after the final segmentation presented in Figure 3 and Figure 4.
Figure 2. Image segmentation based on the color threshold flow chart.

Figure 3. Segmented rice false smut ball.

Figure 4. Segmented health rice ears.

2.3. HOG features
HOG feature can maintain good invariance to image geometric and optical deformation [11]. Therefore, this paper selects the HOG feature as the input feature of the rice false smut image. The image direction gradient, gradient amplitude, and gradient direction are obtained by equation (1), (2), (3), and (4):

\[ G_x(x, y) = H(x+1, y) - H(x-1, y) \]

(1)

\[ G_y(x, y) = H(x, y+1) - H(x, y-1) \]

(2)

\[ G(x, y) = \sqrt{G_x(x, y)^2 + G_y(x, y)^2} \]

(3)

\[ \alpha(x, y) = \tan(G_x(x, y)/G_y(x, y)) \]

(4)

where \( G_x(x, y) \), \( G_y(x, y) \), \( G(x, y) \) and \( \alpha(x, y) \) respectively represent the horizontal gradient, vertical gradient, pixel value, gradient amplitude, and gradient direction at the pixel point \((x, y)\) of the input image.

2.4. SVM methods
This paper uses the SVM classification algorithm to process the obtained HOG feature vector. The goal of SVM is to find a hyperplane. Two problems of classification solved by using a hyperplane. First, to find sample points of classification closest to the hyperplane and second to find the distance between sample point and hyperplane. So, the distance from the sample point and hyperplane will be maximized.

Therefore, this paper uses SVM to classify the disease of rice (unhealthy ear) false smut and healthy ears of rice to achieve recognition.

The interference of sunlight in the field involved, so that’s why this paper does not use the conventional SVM linear kernel function but uses the quadratic kernel function in the nonlinear kernel, the expression corresponding to the quadratic polynomial kernel function published in equation (5).

\[ K(x, x') = (a + r x^T x')^2 \]

(5)

where \( r \) and \( a \) are adjustable parameters.

3. Results and discussion
After extracting HOG features from rice false smut images, this paper obtained the preliminary classification results of SVM, as shown in Figure 5. The purple points in Figure 5 represent positive samples, that is, feature vectors of the rice false smut image; the cyan points represent negative samples, that is, feature vectors of healthy rice ear images. However, it can be found from Figure 5 that it is difficult to distinguish between diseased rice false smut and healthy rice that can only characterize by using the HOG feature. Therefore, this paper continues to use PCA principal component analysis for...
For a large number of high-dimensional HOG feature vectors, the PCA dimension reduction selected in this paper can achieve a specific effect, as shown in Figure 6. This paper selects the first ten-dimensional vectors in the HOG and 1764-dimensional feature vector for component analysis. Results show that the primary two-dimensional vectors can better represent all feature vectors, the contribution of the first feature vector is 18.1%, and the participation of the second feature vector is 7.9%. Therefore, in this paper, the 1764-dimensional HOG feature vector is reduced to a 2-dimensional vector to characterize false smut rice disease.

After the PCA analysis, the SVM method, and the binomial kernel function used to classify diseased rice false smut and healthy rice, and the results revealed in Figure 7. Although there may be errors in very few data points, the results of SVM classification can still distinguish rice false smut disease clearly. To further improve the reliability of the recognition and verify the results, this paper uses ten-fold cross-validation for multiple recognitions, and the prediction SVM classification accuracy is 80.6%. Finally, this paper conducted numerous tests on 225 test sets, and the average recognition rate was 78.56% (under the premise that the SVM model has trained, the recognition speed is 6s). The recognition result of one of the rice false smut images exposed in Figure 8 (red box used to mark the diseased rice false smut).
4. Conclusions
In summary, an approach that can distinguish diseased rice false smut and health rice based on the HOG-SVM method demonstrated. First, the image of rice false smut disease acquired by image preprocessing was based on the color threshold to obtain the region of interest (false smut diseased rice ball) and background region (healthy rice ear). Second, the gamma-correction performed on the extracted ROI and background area; the HOG features dig out to form a high-dimensional HOG feature vector. Third, the PCA analysis executed on high-dimensional HOG features, and the original 1764-dimensional feature vector reduced to a two-dimensional vector that can represent most vectors to achieve the purpose of dimensionality reduction. Finally, the dimensionality-reduced HOG feature vector sent to the SVM classifier for training. The expected recognition accuracy under the binomial kernel function is 80.6%, and the actual test set recognition is 78.56%. Experimental results show that the method proposed in this paper is relatively fast and accurate. This method will play a progressive role in the identification of false smut disease of rice or other crop diseases in the future.

Acknowledgments
We sincerely thank the National key R & D plan tasks (grant no. 2019YFC1606600), the Jiangsu Agricultural Science and Technology Independent Innovation Project (grant no. CX(18)3043), the Zhenjiang Key R & D Plan (grant no. NY2019013, SH2019012), the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD) and the Outstanding Youth Science Foundation of Jiangsu Province (grant no. BK20180099) for funding this study.

Reference
[1] Mebeasellassie A, Li LY, Feng AQ, Zhu XY, and Li JX 2018 Current Plant Biology 15 38-43.
[2] Wu N, Jiang HB, Bao YD, Zhang C, et al. 2020 Sensor. Actuat. B-Chem 308.
[3] Ding GS, Dong FZ, and Zou H 2019 Appl. Soft. Comput 84.
[4] Khang S, Nor A M I and Wei Hong Lim 2013 Appl. Soft. Comput 13 2017-2036.
[5] Ihssan S. M, Amjed A F and IsamAbu Q 2019 Engineering Science and Technology, an International Journal 22 1027-1034.
[6] Sun Q, Zheng JX, and Li C 2019 Powder. Technol 356 295-303.
[7] Dimitrios K, Vasileios A, Tania S, and Nikolaos G 2020 Comput. Netw 168.
[8] Sevcan A K and Hamidullah B 2018 J. Mol. Struct, 1156, 255-263.
[9] Bouchra N, Aouatif A, Mohammed N, and Nabil H 2019 Procedia Computer Science 148 116-125.
[10] He T and Li XF 2019 J. Vis. Commun. Image R 65.
[11] Muhammet F A, Akif D, Kadir S, and Meryem A M 2020 Measurement 158.