J Inf Process Syst, Vol.16, No.2, pp.435~446, April 2020
https://doi.org/10.3745/JIPS.02.0133

ISSN 1976-913X (Print)
ISSN 2092-805X (Electronic)

JOURNAL OF INFORMATION PROCESSING SYSTEMS JIPS

Identification of Tea Diseases Based on Spectral Reflectance and Machine Learning

Xiuguo Zou*, Qiaomu Ren*, Hongyi Cao*, Yan Qian*, and Shuaitang Zhang*

Abstract

With the ability to learn rules from training data, the machine learning model can classify unknown objects. At the same time, the dimension of hyperspectral data is usually large, which may cause an over-fitting problem. In this research, an identification methodology of tea diseases was proposed based on spectral reflectance and machine learning, including the feature selector based on the decision tree and the tea disease recognizer based on random forest. The proposed identification methodology was evaluated through experiments. The experimental results showed that the recall rate and the F1 score were significantly improved by the proposed methodology in the identification accuracy of tea disease, with average values of 15%, 7%, and 11%, respectively. Therefore, the proposed identification methodology could make relatively better feature selection and learn from high dimensional data so as to achieve the non-destructive and efficient identification of different tea diseases. This research provides a new idea for the feature selection of high dimensional data and the non-destructive identification of crop diseases.

Keywords

High Dimensional Data, Machine Learning, Spectral Reflectance, Tea Diseases

1. Introduction

Tea has a long history in China to act as one of the ancient drinks and a major cash crop. Tea has the functions of clearing heat, detoxifying, and relieving fatigue, etc., thereby gaining much popularity among consumers [1]. The main producing areas of tea in China are featured by a warm climate with a humid environment. However, such climatic conditions are conducive to the breeding of pathogens. Besides, tea may also have diseases in the course of transportation and storage, resulting in a significant decline in the quality and production of tea. It has become a hot and required topic about identifying the tea diseases in the early stage [2,3]. Conventional disease identification methods include manual methods and physicochemical methods. The manual methods identify tea diseases through visual and tactile senses, which requires experienced experts or agricultural workers. However, the results of this method vary significantly among different examiners. Moreover, examiners may be fatigued after long-time observation, which will lead to a decrease in the efficiency and accuracy of the identification [4]. The physicochemical method identifies the tea diseases with the techniques of chemistry and molecular biology, for example, fluorescence immunoassay and polymerase chain reaction (PCR) method [5].

Manuscript received June 26, 2018; first revision October 31, 2018; second revision January 8, 2019; third revision January 13, 2020; accepted February 12, 2020.

Corresponding Author: Xiuguo Zou (xiuguozou@gmail.com)

* College of Engineering, Nanjing Agricultural University, Nanjing, China (xiuguozou@gmail.com, m18951155511@163.com, 1041928804@qq.com, 37193240 @qq.com, zhangshuaitang2014@163.com)

However, physicochemical identification is a destructive method in most cases, meaning that it will destroy the object being examined. Besides, this method is time-consuming and requires professional skills [4]. Therefore, it is urgent to develop an efficient and non-destructive method to identify crop diseases. Recently, researchers have been extensively exploring image recognition technology, computer technology [6], laser technology, and hyperspectral imaging technology to identify crop diseases [4]. For example, Qin et al. [7] segmented the image of alfalfa through the K-means clustering algorithm and linear discriminant analysis, and the naive Bayes method and support vector machine (SVM) have also been used to establish disease identification model. Chen et al. [8] employed wavelet transform and textural matrix analytical calculation to enhance the image of wheat disease and retrieve the disease image by image matching. Tian and Li [9] used chromaticity moments as eigenvector to identify cucumber disease based on the SVM method. Chai and Wang [10] used the Bayesian discriminant method to identify early blight, late blight, and leaf mold of tomato by image processing and pattern recognition technologies. Wei [11] segmented and marked the tea images and classified tea quality through the HSI (hue, saturation, intensity) color model. Based on the color and shape of tea and tea stem, Chen [12] classified the tea through the proposed multi-feature and multiple classifiers derived from SVM and Bayesian classifiers. Hyperspectral technology integrates the advantages of spectrum identification and image identification that can acquire the internal and external information of the object and lead to its extensive application to monitoring the growth and identifying the diseases of crops. For example, Li [13] proposed a non-destructive method for measuring tea quality based on machine vision and spectrum technology, through internal component measurement of tea and information diagnosis of tea tree. Chen et al. [14] established a neural network model to examine the tea quality based on hyperspectral data of tea. Peng et al. [15] employed the spectrum technology in the rapid examination of tea plant growth and tea quality. Zhao et al. [16] proposed an efficient method for detecting the slight damage of fruits using spectral imaging technology. Bravo et al. [17] adopted spectral reflectance based on visible light and near-infrared band to diagnose the stripe rust of wheat in the early stage. Leckie et al. [18] detected the aphids' violation of pine tree by the spectroscopic data such as visible light and near-infrared bands.

In this research, using the data from the spectral reflectance of tea, the feature selector was used to remove the irrelevant and redundant data from the high dimensional data, to avoid the Hughes phenomenon [19]. Based on the selected spectral reflectance, a recognizer of tea disease was built to achieve the non-destructive and efficient tea disease identification.

2. Materials and Methods

2.1 Experimental Materials

The tea leaves with disease and healthy tea leaves used in the experiment were acquired from Pingshan Forest Park, Luhe District, Nanjing. The samples were packed into a sealing bag once they were collected. The sealing bag was put into the refrigerator to keep the leaves fresh, and the experiment was carried out in the laboratory in an immediate manner. After the screening and processing by agricultural experts, 80 leaf samples with anthracnose, 72 leaf samples with brown leaf spots, 80 leaf samples with tea white stars, and 60 healthy leaf samples were selected for the experiment. The images of the samples are shown in Fig. 1.

(a) (b) (c)

Fig. 1. Images of tea leaves with three kinds of diseases: (a) anthracnose, (b) brown leaf spot, and (c) tea white star.

2.2 Experimental Device

The experimental device used in the experiment was a hyperspectral imaging system, as shown in Fig. 2. The device was composed of spectrograph ImSpector V10E, CCD camera GEV-B1621M, optical halogen lamp, camera obscura, control cabinet, electric displacement console, and computer, etc. The spectrum of the hyperspectral camera was between 358 nm and 1,021 nm, and the spectral resolution was 2.8 nm.

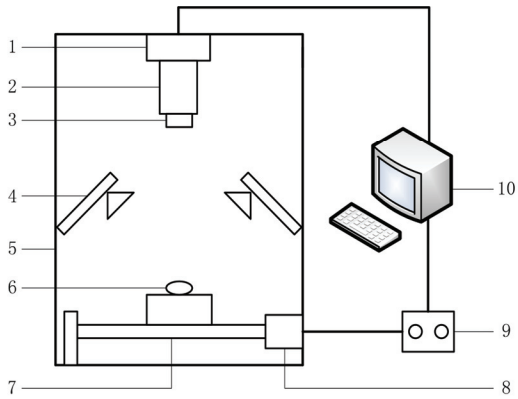


Fig. 2. Hyperspectral imaging system (1=camera, 2=spectrometer, 3=shot, 4=halogen light source, 5= black box, 6=tea sample, 7= electric stage, 8=stepper motor, 9=mobile platform controller, 10=computer).

2.3 Data Collection

Data collection was performed according to the following steps:

Step 1: acquire the hololeucocratic calibration image W by collecting standard white calibration board with 99% of reflectivity.

Step 2: acquire the holomelanocratic calibration image D from the image behind the lens cover.

Step 3: perform the data collection for all leaf samples and put the samples into the objective table and adjust the table to the appropriate location. The 616-dimensional original hyperspectral image I with a wavelength of 358–1,021 nm was obtained using the hyperspectral image capture software (Spectral Image) [20].

The parameter settings for the above steps are presented in Table 1.

Table 1. Parameter setting of the spectral data acquisition system

Parameter name	Value
Image resolution (pixel)	1632 × 1415
Exposure time (ms)	50
Electric displacement stage speed (mm/s)	1.21
Object distance (mm)	720

2.4 Data Processing

The data processing was performed using the following configurations: Computer with RAM of 16 GB and CPU of Intel Core i5-6500, which installed Excel 2010, MATLAB 2016a (MathWorks, Natick, MA, USA) and ENVI 5.3 (Exelis Visual Information Solutions Inc., Boulder, CO, USA).

2.4.1 Image correction

In order to eliminate the interference noise in the process of data collection, the original hyperspectral image I is corrected by using Eq. (1), and the corrected image is recorded as R.

$$R=(I-D)/(W-D) \tag{1}$$

where R is the corrected image, I is the original hyperspectral image acquired by the hyperspectral system, D and W are introduced in Section 2.3.

2.4.2 Relative spectral reflectance of ROI region

Each pixel in a hyperspectral image corresponds to the spectral information of a full-wave band. According to the average distribution of the disease spots in the sample, a region of 200×200 pixels in the center of the leaf was selected as the region of interest (ROI). The average spectral reflectance of ROI was extracted from the 80 leaves with anthracnose, 72 leaves with brown leaf spots, 80 leaves with tea white stars, and 60 healthy leaves, respectively. The results are shown in Fig. 3.

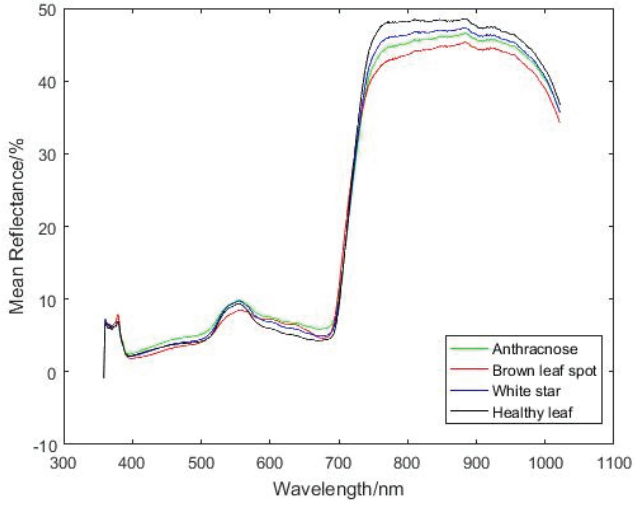


Fig. 3. Relative spectral reflectance of the blade.

2.5 Research Methods

The One-vs-All method was used to transform the multi-classification problem into a binary classification problem. The first class of multiple classes was marked as positive classes, and all other classes were marked as negative class. Similarly, the second, third, and fourth classes were all treated in this way.

In both the training and testing process, 5-fold cross-validation was used to evaluate the learning performance. In order to ensure the stability and accuracy of the experimental results, the 5-fold cross-validation was repeated ten times, and their average was used as the evaluation index. Besides, the original data were classified into each fold according to the sample ratio of 8:7:8:6, and the distribution of each fold data was kept by that of the original sample, to ensure that each class data was trained to improve the performance of the methodology (Fig. 4).

The evaluation indexes used in this research included the identification accuracy, recall rate, and F1 score. Through the One-vs-All method, 12 evaluation indexes could be obtained from four categories (Fig. 5).

$\langle x_{i,1}, x_{1,2}, \cdots, x_{i,n-1}, x_{i,n}, y_{real,i}, y_{predict,i} \rangle$

i equals the number of samples, ranging from 1 to 292

n equals the dimensions of data

$x_{i,n}$ represents the n-th feature of the i-th sample

$y_{real,i}$ represents the true label of the i-th sample

$y_{predict,i}$ represents the predicted label of the i-th sample

Each $y_{real,i}$ and $y_{predict,i}$ only belong to one class among Anthracnose, Brown leaf spot, White star and Healthy leaf

Fig. 4. Sample space and related definitions.

Anthracnose will be recorded as A, Brown leaf spot will be recorded as B, White star will be recorded as W and Healthy leaf will be recorded as H.

if $y_{predict,i} \in A$

A will be referred to positive samples, B、 W and H will be referred to negative samples

if $y_{real,i} \in A$

TP(True Positive) ++

else

FP(False Positive) ++

$$\textbf{\textit{Precision}} = \frac{TP}{TP + FP} \quad (2)$$

$$\textbf{\textit{Recall}} = \frac{TP}{TP + FN} \quad (3)$$

TP+FN: The number of positive sample

$$F1 = \frac{2 * \mathrm{Pr}ecision * \mathrm{Re}call}{\mathrm{Pr}ecision + \mathrm{Re}call} \quad (4)$$

The evaluations of other classes can be computed as same as the above process

Fig. 5. Definitions and descriptions of evaluations.

2.5.1 Feature selection based on decision tree

The data obtained in this research were hyperspectral, and each original sample had 616 features. Each sample often had irrelevant and redundant features, which not only reduced the learning rate and increased the training time but also declined the overall performance of the classifier.

The decision tree has been extensively employed as a suitable feature selection method to divide the subset of samples according to information entropy, which is more suitable for small sample data [21, 22]. The ID3 decision tree was used for selecting the feature from the whole feature space.

The original sample, 616-dimensional original data, was used to build the tea disease recognizer with the decision tree as the classifier. In addition, 10-time 5-fold cross-validation was used to train the tea disease recognizer based on the original sample and decision tree.

The sample after feature selection was obtained from the original sample using the feature selector based on the decision tree. According to the information metric of the decision tree, the dimension number of features was reduced from 616 dimensions to 16 dimensions. The results of the feature selection are displayed in Table 2.

Table 2. The result of features after feature selection

Feature	Wave band (nm)	Feature	Wave band (nm)
1	679.559143	9	698.119080
2	378.627686	10	457.556244
3	358.492340	11	598.305847
4	393.832397	12	700.305603
5	368.540039	13	376.606995
6	868.905396	14	390.784546
7	504.615356	15	516.212524
8	695.933167	16	368.540039

The tea disease recognizer was built using the sample after feature selection and using decision tree algorithms as the classifier. The tea disease recognizer based on the original data and the decision tree and the one based on the selected data and the decision tree were obtained using 10-time 5-fold cross-validation. The above process is shown in Fig. 6.

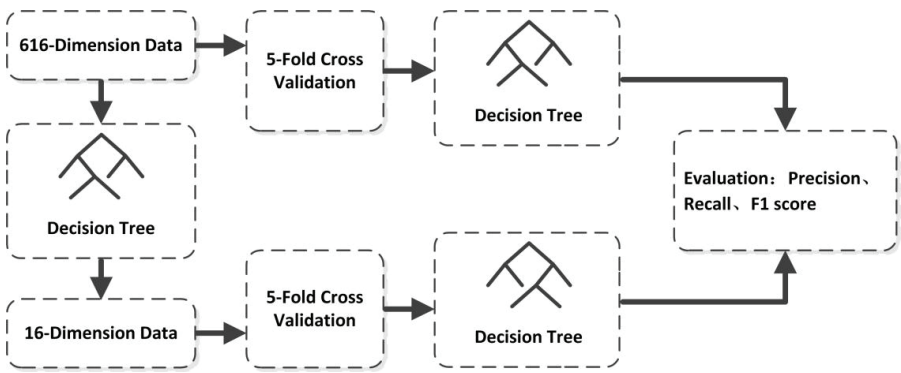


Fig. 6. Flowchart of feature selector based on original data and decision tree and feature selector based on data after feature selection and decision tree.

2.5.2 Identification of tea diseases based on random forest

Classification is an essential component of machine learning. Traditional classifiers include SVM algorithm [23], naive Bayesian algorithm [24], K-nearest-neighbor algorithm [25] and decision tree algorithm [26,27], etc. However, these classifiers are prone to cause an over-fitting problem, sometimes resulting in reduced accuracy. Therefore, many scholars used multiple models to improve the performance of machine learning, where weak classifiers were used to build strong classifiers. These methods are called ensemble learning [28].

Random forest algorithm, proposed by Breiman [29], integrates the Bagging ensemble learning theory [30] and random subspace method [31] in a dynamic way. The basic classifier in the random forest is the decision tree, and the random forest consists of several decision trees obtained by ensemble learning and training. The output results of all the basic classifiers formulate the final classification results [32].

The sample after feature selection constructed the tea disease recognizer using the random forest as the classifier. In the whole process, 10-time 5-fold cross-validation was used. Before the training, we set the number of individual basic classifiers in the random forest as 500. Finally, training was performed on the tea disease recognizer based on the selected data and the random forest. Fig. 7 presents the workflow of the feature selector and the tea disease recognizer.

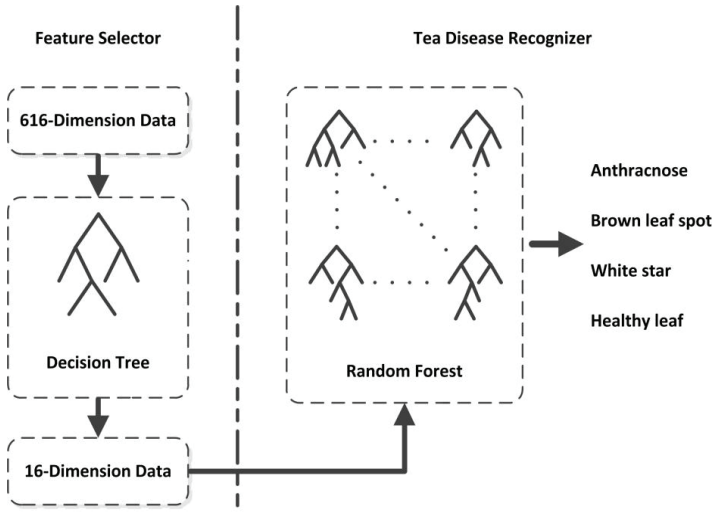


Fig. 7. Flowchart of feature selector based on decision tree and tea disease recognizer based on random forest.

3. Results and Discussion

3.1 Feature Selection based on Decision Tree

By comparing Tables 3 and 4, after the feature selection, the same learning strategies and verification methods were adopted. Each evaluation index after feature selection was superior to that before feature selection, which indicated that the selected features retained some properties of the original features and reduced the noise caused by irrelevant and redundant features.

Table 3. Ten-times average evaluations from tea disease recognizer based on original data and decision tree

Class	Precision	Recall	F1 Score
Anthracnose	0.71	0.70	0.70
Brown leaf spot	0.69	0.71	0.70
White star	0.74	0.80	0.77
Healthy leaf	0.69	0.68	0.68

Table 4. Ten-times average evaluations from tea disease recognizer based on data after feature selection and decision tree

Class	Precision	Recall	F1 Score
Anthracnose	0.76	0.72	0.74
Brown leaf spot	0.71	0.80	0.74
White star	0.78	0.78	0.78
Healthy leaf	0.78	0.71	0.71

3.2 Identification of Tea Diseases based on Random Forest

By comparing Tables 4 and 5, it was found that the identification accuracy, recall rate, and F1 score in Table 5 were increased by 10%, 4%, and 8%, respectively, with the maximum increasing by 23%, 7%, and 15%, respectively, compared with Table 4. The random forest algorithm successfully improved the performance of the classifier, which made the classification results more accurate.

Table 5. Ten-times average evaluations from tea disease recognizer based on data after feature selection and random forest

Class	Precision	Recall	F1 Score
Anthracnose	0.79	0.74	0.76
Brown leaf spot	0.94	0.85	0.89
White star	0.77	0.85	0.81
Healthy leaf	0.94	0.74	0.82

3.3 F1 Score Distribution in Multiple Cases

The relationship between F1 score, identification accuracy, and recall rate was determined by Eq. (4). The identification accuracy and recall rate were adopted to evaluate the performance of the classification model better. The box plot was used to visualize the distribution of F1 scores in various cases. The upper edge and the lower edge of the box plot represent the maximum and minimum values of F1 score, respectively. The discrete points represent the outliers in the data, and the upper and lower edges of the box represent the upper quartile and the lower quartile, respectively, where the horizontal line represents the median. The box plots of F1 score are shown in Fig. 8. Fig. 8(a), (b), (c) and (d) show the identification of the distribution of F1 scores for anthracnose, brown leaf spot, tea white plot, and healthy leaves using the tea disease recognizer based on the original data and decision tree, the one based on selected data and decision tree, and the one based on the selected data and random forest.

It can be observed from Fig. 8(a) that the distribution of F1 score (marked A3) of anthracnose using the recognizer based on the selected data and the random forest was superior to the other two cases (A1

and A2). In terms of identifying brown leaf spot, tea white star, and healthy leaf, the optimal distribution of F1 score was achieved in the tea disease recognizer based on selected data and random forest.

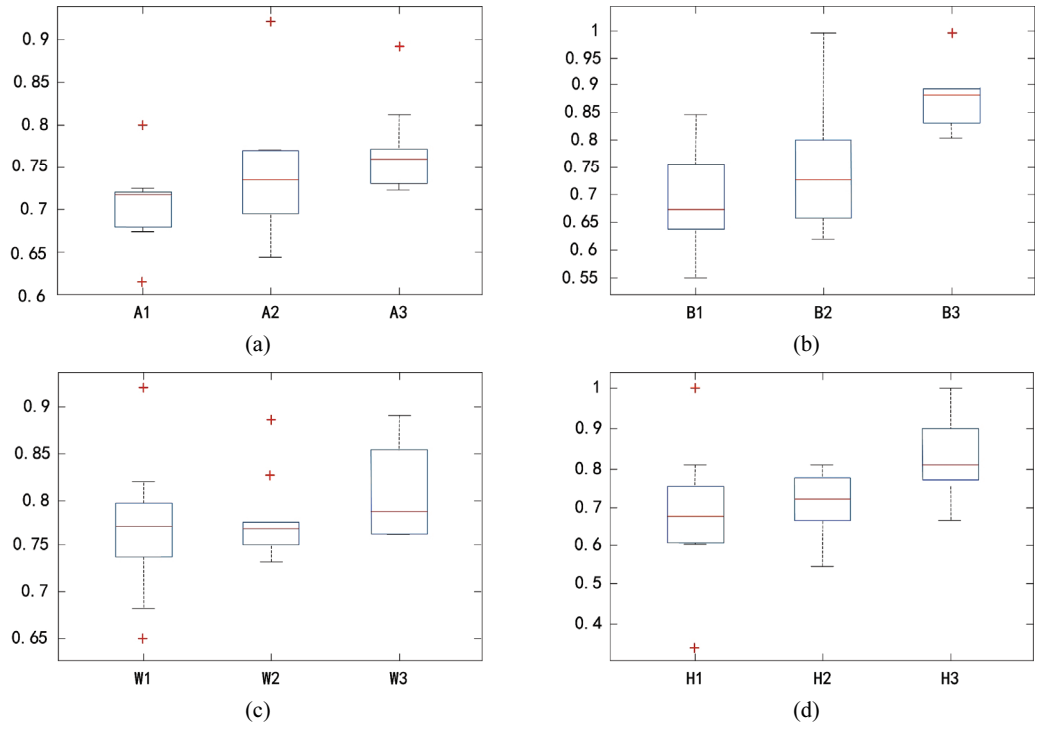


Fig. 8. The situations of F1 distribution under the tea disease recognizer based on original data and decision tree, the one based on selected data and decision, and the one based on selected data and random forest, respectively (from left to right): (a) anthracnose, (b) brown leaf spot, (c) white star, and (d) healthy leaf.

3.4 Discussion and Future Work

The experimental results showed that the feature selection strategy based on the decision tree was fully able to reduce the dimension of high-dimensional data. Besides, it was also shown that the decision tree method performed well in being a good classification strategy and in feature selection. The tea disease recognizer based on random forest could effectively learn the information from training data, and then identify the diseases that tea might have. From the evaluation indexes of identification accuracy, recall rate, and F1 score, the best experimental results were achieved in many experiments under the methodology with feature selector based on decision tree and the tea disease recognizer based on random forest, laying a foundation for the high-efficiency and non-destructive identification of crop diseases.

Considering that F1 score can be calculated by the identification accuracy and recall rate, the performance of the classification model was well evaluated by F1 score in this research. Besides, the difference in index value could not reflect the advantages and disadvantages of the model.

In addition, there is a lack of public datasets for the comparison of similar research work in the field, such as the ImageNet dataset [33] and the COCO dataset [34] for deep learning research. Some similar research often used different data, making the results less comparable.

With the continuous development of artificial intelligence and embedded technology, machine learning algorithms can run on embedded system platforms in real-time. Transplanting the implementation algorithms proposed in this paper into plant protection drones can better reduce the impact of tea diseases on tea quality and yield, and reduce the economic losses caused by tea diseases.

4. Conclusion

The hyperspectral images were obtained, and the relative spectral reflectance of sensitive bands in ROI was extracted as the feature. The decision tree was used as a feature selection method to remove irrelevant or redundant features. The 16-dimensional features were selected from 616-dimensional features, and the decision tree was used as the classifier to learn features before and after feature selection. The F1 score was increased by an average of 3% when using the decision tree for feature selection, indicating the good ability of the decision tree in feature selection.

Compared with the tea disease recognizer based on original data and decision tree and the one based on selected data and decision tree, the increased performance was observed in the one based on selected data and random forest. In the end, the average F1 score of tea disease identification was over 80%.

Acknowledgement

This paper is supported by the Fundamental Research Funds for the Central Universities of China (No. KYTZ201661), China Postdoctoral Science Foundation (No. 2015M571782), and Jiangsu Agricultural Machinery Foundation (No. GXZ14002), University Student Entrepreneurship Training Program of Jiangsu Province (No. 201810307031T).

References

[1] X. Wan, D. Li, Z. Zhang, T.Xia, T. Ling, and Q. Chen, “Research advance on tea biochemistry,” *Journal of Tea Science*, vol. 35, no. 1, pp. 1-10, 2015.

[2] Z. Luo, X. Cai, L. Bian, Z. Li, and Z. Chen, “Advanced in research and application of sex pheromone of tea (Camellia sinensis) pest,” *Journal of Tea Science*, vol. 36, no. 3, pp. 229-236, 2016.

[3] P. Peng, X. Wang, Y. Xiao, and Y. Xu, “Review and prospect on forecasting of tea pests and diseases,” *Southwest China Journal of Agricultural Sciences*, vol. 23, no. 5, pp. 1742-1745, 2010.

[4] S. Cheng, “Fast detection methods for crop disease infection period using spectral and imaging technology,” Ph.D. dissertation, Zhejiang University, Hangzhou, China, 2014.

[5] Y. Hong, J. Li, H. M. Wang, and K. Zhao, “Progress in real-time quantitative PCR technique,” *International Journal of Epidemiology and Infectious Disease*, vol. 33, no. 3, pp. 161-166, 2006.

[6] W. Benhabib and H. Fizazi, “A multi-objective TRIBES/OC-SVM approach for the extraction of areas of interest from satellite images,” *Journal of Information Processing Systems*, vol. 13, no. 2, pp. 321-339, 2017.

[7] F. Qin, D. X. Liu, B. D. Sun, L. Ruan, Z. H. Ma, and H. G. Wang, “Image recognition of four different alfalfa leaf diseases based on deep learning and support vector machine,” *Journal of China Agricultural University*, vol. 22, no. 7, pp. 123-133, 2017.

[8] B. Chen, X. Guo, and X. Li, "Image diagnosis algorithm of diseased wheat," *Transactions of the Chinese Society for Agricultural Machinery*, vol. 40, no. 12, pp. 190-195, 2009.

[9] Y. Tian and C. Li, "Research on recognition of cucumber disease based on image processing in sunlight greenhouse," *Journal of Agricultural Mechanization Research*, vol. 28, no.2, pp. 151-153, 2006.

[10] Y. Chai and X. Wang, "Recognition of greenhouse tomato disease based on image processing technology," *Techniques of Automation and Applications*, vol. 32, no. 9, pp. 83-89, 2013.

[11] X. Wei, "Study on quality grading of tea by digital image processing techniques," *Journal of Anhui Agricultural Sciences*, vol. 40, no. 7, pp. 4251-4253, 2012.

[12] S. Chen, "The study of tea-leaf and tea-stalk image recognition and classification based on multi-features and multi-classifiers," M.S. thesis, Anhui University, Hefei, China, 2014.

[13] X. Li, "Research on nondestructive determination of tea quality based on machine vision and spectroscopy techniques," Ph.D. dissertation, Zhejiang University, Hangzhou, China, 2009.

[14] Q. Chen, J. Zhao, J. Cai, and V. Saritporn, "Estimation of tea quality level using hyperspectral imaging technology," *Acta Optica Sinica*, vol. 28, no. 4, pp. 669-674, 2008.

[15] J. Y. Peng, X. L. Song, F. Liu, Y. D. Bao, and Y. He, "Fast detection of Camellia sinensis growth process and tea quality informations with spectral technology: a review," *Guang pu xue yu guang pu fen xi (Guang pu)*, vol. 36, no. 3, pp. 775-782, 2016.

[16] J. Zhao, J. Liu, Q. Chen, and V. Saritporn, "Detecting subtle bruises on fruits with hyperspectral imaging," *Transactions of the Chinese Society for Agricultural Machinery*, vol. 39, no. 1, pp. 106-109, 2008.

[17] C. Bravo, D. Moshou, J. West, A. McCartney, and H. Ramon, "Early disease detection in wheat fields using spectral reflectance," *Biosystems Engineering*, vol. 84, no. 2, pp. 137-145, 2003.

[18] D. G. Leckie, E. Cloney, and S. P. Joyce, "Automated detection and mapping of crown discolouration caused by jack pine budworm with 2.5 m resolution multispectral imagery," *International Journal of Applied Earth Observation & Geoinformation*, vol. 7, no. 1, pp. 61-77, 2005.

[19] B. Mojaradi, H. Abrishami-Moghaddam, M. J. V. Zoej, and R. P. Duin, "Dimensionality reduction of hyperspectral data via spectral feature extraction," *IEEE Transactions on Geoscience & Remote Sensing*, vol. 47, no. 7, pp. 2091-2105, 2009.

[20] S. Zhang, Z. Wang, and X. Zou, Y. Qian, and L. Yu, "Recognition of tea disease spot based on hyperspectral image and genetic optimization neural network," *Transactions of the Chinese Society of Agricultural Engineering*, vol. 33, no. 22, pp. 200-207, 2017.

[21] Y. Wang and J. Li, "Analysis of feature selection and its impact on hyperspectral data classification based on Decision Tree Algorithm," *Journal of Remote Sensing*, vol. 11, no. 1, pp. 69-76, 2007.

[22] X. B. Yang, and J. Zhang, "Decision tree and its key techniques," *Computer Technology and Development*, vol. 17, no. 1, pp. 43-45, 2007.

[23] S. F. Ding, B. J. Qi, and H. Y. Tan, "An overview on theory and algorithm of support vector machines," *Journal of University of Electronic Science & Technology of China*, vol. 40, no. 1, pp. 2-10, 2011.

[24] M. A, "Research and application on naïve Bayes classification algorithm," M.S. thesis, Dalian University of Technology, Dalian, China, 2014.

[25] Y. Sang, "Research of classification algorithm based on k nearest neighbor," M.S. thesis, Chongqing University, Chongqing, China, 2014.

[26] S. Feng, "Research and improvement of decision trees algorithm," *Journal of Xiamen University (Natural Science)*, vol. 46, no. 4, pp. 497-500, 2007.

[27] Y. Zhang, and J. Cao, "Decision tree algorithms for big data analysis," *Computer Science*, vol. 43, no. 6A, pp. 374-383, 2016.

[28] K. Fang, J. Wu, J. Zhu, and B. Shia, "A review of technologies on random forests," *Statistics & Information Forum*, vol. 26, no. 3, pp. 32-38, 2011.

[29] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5-32, 2001.

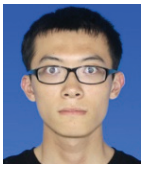
[30] L. Breiman, “Bagging predictors,” *Machine Learning*, vol. 24, no. 2, pp. 123-140, 1996 .
[31] T. K. Ho, “The random subspace method for constructing decision forests,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 832-844, 1998.
[32] S. Dong and Z. Huang, “A brief theoretical overview of random forests,” *Journal of Integration Technology*, vol. 2, no. 1, pp. 1-7, 2013.
[33] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, et al, “ImageNet large scale visual recognition challengem” *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211-252, 2015.
[34] T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollar, and C. Lawrence Zitnick, “Microsoft coco: common objects in context,” in *Computer Vision – ECCV 2014*. Cham: Springer, 2014, pp. 740-755.

Xiuguo Zou **https://orcid.org/0000-0002-8074-7555**

He is Ph.D. and Associate professor. He received the doctor degree in Nanjing Agricultural University (China) in 2013. He currently works in Nanjing Agricultural University. His interests and research are focused on image processing and pattern recognition. He has authored over 20 technical journals in the area of image processing and pattern recognition.

Qiaomu Ren **https://orcid.org/0000-0002-0843-9355**

He received B.S. degrees in college of Engineering from Nanjing Agricultural University in 2019. Since September 2019, he is as a graduate student at Southeast University, majoring in computer science. His current research interests include big data processing based on machine learning and deep learning.



Hongyi Cao **https://orcid.org/0000-0003-2419-2396**

He is an undergraduate student at Nanjing Agricultural University, majoring in automation. His current research interests include big data processing based on machine learning and deep learning.

Yan Qian **https://orcid.org/0000-0001-8350-9352**

She is Ph.D. and Associate professor. She received the doctor degree in Nanjing Agricultural University (China) in 2014. She currently works in Nanjing Agricultural University. Her interests and research are focused on machine learning and deep learning. She has authored over 10 technical journals in the area of machine learning and deep learning.

Shuaitang Zhang **https://orcid.org/0000-0001-7547-5439**

He received M.S. degrees in college of Engineering from Nanjing Agricultural University in 2018. Since July 2018, he has been an engineer in a high-tech company. His current research interests include image processing and pattern recognition.