

Received August 31, 2018, accepted September 23, 2018, date of publication October 4, 2018, date of current version October 29, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2873689

Hyperspectral Identification and Classification of Oilseed Rape Waterlogging Stress Levels Using Parallel Computing

JI'AN XIA¹, YUWANG YANG¹, HONGXIN CAO², WEIXIN ZHANG², LEI XU¹, (Member, IEEE), QIAN WAN², YAQI KE¹, WENYU ZHANG², DAOKUO GE², AND BO HUANG¹, (Senior Member, IEEE)

¹School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China

²Institute of Agricultural Information, Jiangsu Academy of Agricultural Sciences, Nanjing 210014, China

Corresponding authors: YuWang Yang (yuwangyang@njust.edu.cn) and HongXin Cao (caohongxin@hotmail.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61671244 and Grant 61773206, in part by the Key Research and Development Program of Jiangsu Province under Grant SBE2018310371, in part by the National Defense Basic Scientific Research Project under Grant JCKY2018* * *303, and in part by the Natural Science Foundation of Jiangsu Province of China under Grant BK20170131.

ABSTRACT This paper aims to evaluate the feasibility and performance of two related applications in agricultural big data: hyperspectral imaging and parallel computing. After selecting two oilseed rape varieties (NingYou 22 and NingZa 19) as the objects of study, we captured hyperspectral images of these siliques following their exposure to three different waterlogging stress levels (0, 3, and 6 days). The machine learning library for Spark was used to realize both artificial neural network (ANN) and support vector machine (SVM) classification algorithms, and to conduct hyperspectral classification analysis of the oilseed rape siliques under the various levels of waterlogging stress on the parallel computing platform. From the classification data sets, 70% of the data were randomly selected for training, and the remaining 30% were used for prediction. The experimental results indicate that, when the hyperspectral image of a region of interest (400–1000 nm) was extracted and combined with the spectrum image data, the oilseed rape waterlogging detection model based on the Spark parallel computing framework was feasible and efficient. For the multi-class classification problem, the accuracy of the ANN algorithm was superior to that of the SVM, but its convergence time exceeded that of the SVM algorithm. Using the ANN and SVM algorithms for binary classification of the samples from both varieties, the results indicate that the performance of the SVM algorithm was superior in terms of the binary classification problem. Meanwhile, of the two oilseed rape varieties, the NZ 19 waterlogging samples yielded better classification results. Five optimal wavebands (512, 621, 689, 953, and 961 nm) were selected as the inputs for the classification algorithm. The results show that the classification accuracy of full-waveband hyperspectral imaging was slightly higher than that of optimal waveband imaging, while the ANN algorithm was more accurate than the SVM algorithm. Finally, the three indices of speedup, scaleup and sizeup were used to evaluate the operation performance of the hyperspectral data set algorithm based on the Spark parallel computing platform.

INDEX TERMS Rapeseed siliques, hyperspectral image parallel computing, machine learning, spark, classification algorithm.

I. INTRODUCTION

At present, oilseed rape is the second most widely produced oil-bearing crop worldwide, with its yield surpassed only by that of soybeans. The global oilseed rape yield is expected to reach 74.28 million tons during the 2017–2018 growth season, and plans are underway in several countries to increase

oilseed rape crops to satisfy the robust supply and export demand [1]. Rapeseed oil boasts the highest oleic acid and linoleic acid levels among various vegetable oils, and its saturated fat content is less than half that of soybean oil, which is widely regarded as a wholesome cooking oil. Additionally, the residual rapeseed dregs following oil extraction

may be used as animal feed. Rapeseed oil can also be used to manufacture clean, renewable fuel—biodiesel, which has the potential for broad application.

Agricultural production is highly dependent on weather conditions and, as one of the main abiotic stresses, waterlogging is a severe problem that increasingly afflicts agricultural production. With continuous precipitation and high underground water levels, oilseed rape is very likely to become waterlogged, and this may cause a decline in the number of effective branches, siliques and particles per plant, ultimately leading to a decline in overall yield (if it occurs during the crop's growth) [2]. In particular, during the plant's reproduction phase, if waterlogging stress exceeds the self-adjustment threshold value of the oilseed rape's body, part of the photosynthetic apparatus of the leaves may suffer irreversible damage, which will ultimately affect normal growth and precipitate a decline in oilseed rape yield and quality [3]. Furthermore, when soil water content is too high the field humidity is elevated, the transmission and proliferation of various germs and diseases will be promoted, including downy mildew, club root and weeds, resulting in secondary disasters associated with waterlogging [4].

A hyperspectral image has more than 100 spectral bands and thus may form a continuous spectrum. In addition to the 2D plane graph, each image pixel contains a spectral dimension together they form a 3D data cube, which contains rich visual and spectral information; these data may be applied in the development of a complex model for detecting, classifying, and identifying the materials, textures, and components in the spectral image [5]. Hyperspectral image technology can be used to simultaneously obtain spatial and spectral information. As such, it has attracted significant attention and emerged as a robust tool for researchers evaluating agricultural crops according to various indices. Other applications of hyperspectral image technology include testing fruit quality and nutrient content [6]–[8], detecting plant biomass, diseases and insect pests [9], [10], and seed identification [11].

Cloud computing has also attained prominence in the agricultural sector, particularly in the field of precision agriculture. The collection of high volumes of diverse agriculture data is ongoing [12], and high-performance computation based on parallel computing is highly applicable in this multidisciplinary, data-intensive field, facilitating the collection, analysis, mining, and prediction of agricultural big data. By combining prior knowledge with machine learning and deep learning methods, data mining and decision-making abilities can be enhanced; furthermore, the method may also facilitate faster and more precise agricultural management, improve decision quality, reduce the occurrence of asymmetric information, and produce other benefits [13], [14]. Hitherto, big data and parallel computing technology have been applied to various aspects of precision agriculture, including crop quality management, monitoring and decision-making with respect to the greenhouse environment, and facilitation of crop performance [15]–[17].

In this paper, we propose a method for detecting and classifying crop waterlogging by combining hyperspectral images and parallel computing technology, and apply these techniques to actual production scenarios. The methods involved in this analysis are as follows: ROI selection, data pre-processing, format conversion of hyperspectral data, use of the Machine Learning Library (MLlib) to realize artificial neural network (ANN) and support vector machine (SVM) algorithms, multi-class and binary classification analyses on hyperspectral data and optimal wavebands, and evaluation of the accuracy of the classification performance achieved using parallel computing.

The rest of this article is organized as follows: in Section II, the objectives and methods of the experiment are introduced. The machine learning algorithm and experimental details are described in Section III. Analyses of the experiment results, and of the performance of the parallel computing framework, are presented in Section IV. Finally, Section V presents the conclusions based on the analyses and findings.

II. RELATED WORK

This section discusses three elements of the design in this research: (1) the design of the oilseed rape waterlogging experiment; (2) the design of the hyperspectral image collection system; and (3) design scheme of the parallel computing platform.

A. OILSEED RAPE WATERLOGGING EXPERIMENT SCHEME

The oilseed rape waterlogging experiment was conducted at the experimental farm of Jiangsu Academy of Agricultural Sciences, at north latitude 32.03° and east longitude 118.87°. Two oilseed rape varieties which cultivated by Jiangsu academy of agricultural sciences were selected for seeding and planting: NingYou 22 (NY 22 non-hybrid variety) and NingZa 19 (NZ 19 hybrid variety). Comparison of the two varieties was done to optimize the experimental results. The waterlogging experiment involved seedbed seedlings and pool-culture waterlogging. During the 2017/2018 growth season, both oilseed rape varieties tested in this study (i.e., NY 22 and NZ 19) were used in a water-control pool-culture test. During the early flowering period (March 7, 2018), the waterlogging level was determined at 0, 3 and 6 days. The sowing date was October 2, 2017, and the transplanting date was November 3, 2017. The planting scheme of the experiment is depicted in Fig. 1, in which C2 and C3 represent NY 22 and NZ 19, respectively; I, II and III refer to the experimental areas under waterlogged conditions for 0, 3, and 6 days, respectively. The waterlogging site is depicted in Fig. 2. A water pipe was employed to continuously inject water into the experimental areas for 0, 3, and 6 days; no area was sealed at the bottom, and the water pipe simulated the waterlogging conditions that may be expected under normal growth conditions in the field. The pool-culture area measured $3.77\text{m} \times 2.77\text{ m} \approx 10.44\text{ m}^2$, with rows spaced 0.4 m apart and plants spaced 20 cm apart. The oilseed rape siliques were sampled during the reproduction phase (April 16, 2018),

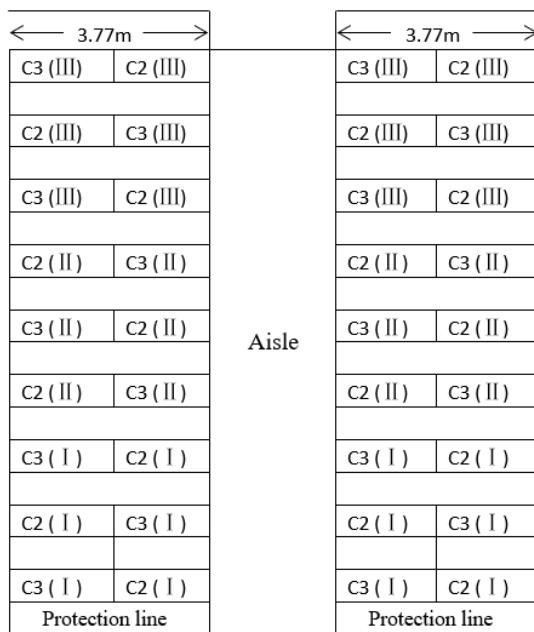


FIGURE 1. Planting scheme of waterlogging experiment.



FIGURE 2. Site of waterlogging stress experiment.

as follows: 20 fresh oilseed rape siliques samples that were devoid of disease, insect pests and damage were randomly collected from each area; in total, there were 36 groups of 20 samples, 18 for each rape variety. Thus, there were 720 siliques samples in total ($n = 360$ per rape variety), and hyperspectral images of fresh siliques samples were acquired in the laboratory.

B. THE HYPERSPECTRAL IMAGING SYSTEM

The camera used to capture the hyperspectral images was the Resonon Pika XC camera. The hyperspectral imaging

system consisted of: (1) a Pika XC hyperspectral imaging camera; (2) a linear translation stage; (3) a mounting tower; (4) a lighting rig; and (5) a software-based control system. The hyperspectral imaging camera has a spectrum collection scope of 400–1,000 nm, with a spectral resolution of 2.5 nm; it also has 240 spectral channels and 1,600 spatial channels. A high-performance laptop (Dell Latitude E6450, i7-4610M CPU @3.00 GHz, installed memory 8 GB, 64-bit Windows 7 operating system) and SpectrononPro software were used to obtain the hyperspectral images, the dimensions of which were $1,600 \times 800 \times 240$ pixels. The hyperspectral image was saved in band-interleaved-by-line (.bil) format with a header file.

Prior to collecting the hyperspectral data, the parameters of the collection system were calibrated. First, a standard test whiteboard (BaSO_4) was used to determine the frame rate and integration time of the hyperspectral camera; then, paper was printed with black lines to adjust the camera focal length; the dark current of the hyperspectral camera and reflectivity of the reference whiteboard were ascertained to eliminate dark current noise and complete automatic calibration of the instrument; finally, paper was printed with circular patterns to adjust the aspect ratio of the camera. Following the calibration, 10 siliques samples collected from the same area were placed onto the linear translation platform for analysis of the hyperspectral images. The collection parameters were established as follows: the distance between the linear platform and camera was 21 cm; the angle between the light source and sample was 30 degrees; the integration time of the camera was 12.01 ms; the frame rate was 31.23 Hz; the number of scanning lines was 800; and the scanning speed was 0.23 cm/s.

Both the oilseed rape siliques sampling and hyperspectral image collection were completed at the Institute of Agricultural Information, Jiangsu Academy of Agricultural Sciences.

C. THE PARALLEL COMPUTING PLATFORM

Huawei routing and switching devices, and a Dell Rack Mount Server, were used to assemble the parallel computing platform. As illustrated in Table 1, one router and one switch were used for network connection and data exchange; six servers (Master, Nodes 1–5) formed the parallel computing cluster based on the Spark framework; the cluster was connected to the Internet (IP address: 202.119.84.75:6800) and functioned in standalone mode, that is, the Master node was responsible for resource allocation and job scheduling while Nodes 1–5 executed the computation tasks as the worker nodes. The Master node also processed tasks as a worker node. The allocation of nodes within the parallel computing cluster, and the operating environment, are detailed in Table 2. The cluster nodes were set up using the CentOS 6.8 operating system; Spark 2.2.0 was used as the parallel computing framework for the entire cluster; and the Hadoop Distributed File System (HDFS) was used to distribute storage and read files. Cloudera Manager 5 was used to monitor the functionality of the entire cluster; IntelliJ IDEA was used as the

TABLE 1. Parallel computing platform structure.

	Name	Model	Quantity
1	Switch	HUAWEI S3700	1
2	Router	HUAWEI AR1200	1
3	Cluster nodes (Running nodes)	Dell C2100	1+5
4	IP address Open port	202.119.84.75 6800	1
5	Transmission media	Twisted-pair Cat 6	

TABLE 2. Spark cluster configuration and running environment.

	Name	Model or version
1	CPU	Inter Xeon 5600
2	Memory	Dell ECC-DDR3 8GB
3	Operating System	CentOS 6.8
4	Java Development Kit (JDK)	JDK 1.8.0
5	Hadoop	2.7.5
6	Hadoop Administration tool	Cloudera Manager 5
7	Spark	2.2.0
8	Scala	2.11.8
9	Integrated Development Environment (IDE)	IntelliJ IDEA Community 2016

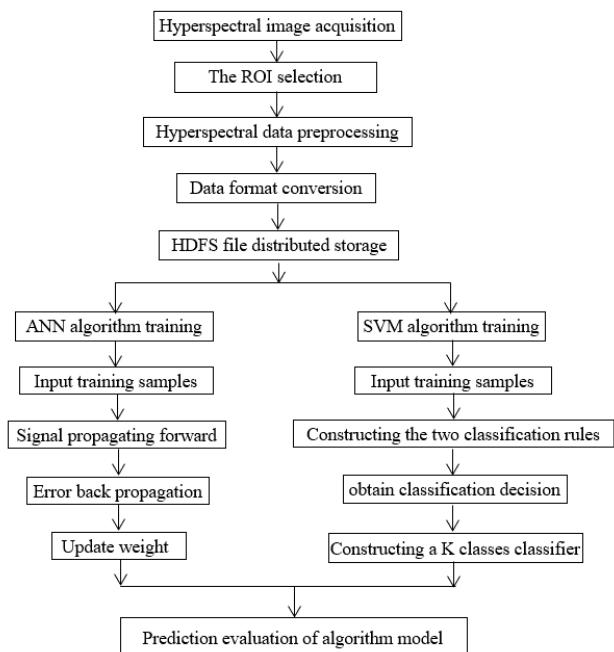
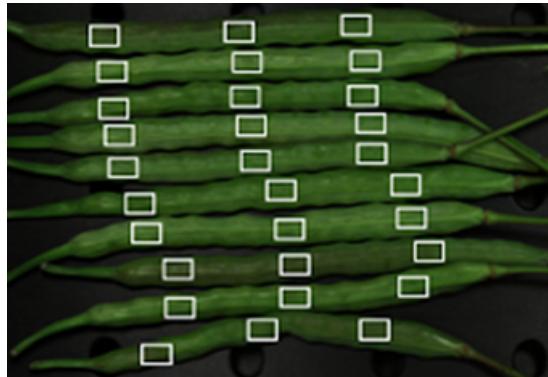
integrated development environment (IDE); Scala 2.118 was used as the programming language; and MLlib for Spark was used to program and execute the classification algorithms. Classification analysis was performed on the collected hyperspectral data and images.

Development and testing of the parallel computing platform was performed at the School of Computer Science and Engineering, Nanjing University of Science and Technology.

III. PROBLEM STATEMENT AND ALGORITHM MODEL

The key stages in the experimental procedure are presented in Fig. 3; the hyperspectral images of the siliques samples were acquired in the laboratory. As Fig. 4 illustrates, three ROIs on the hyperspectral image of each siliques sample were selected, and each ROI had an area of 58×40 (2,320) pixels. Then, 3×10 pixel ROIs were drawn for extracting sample information from the image. The original hyperspectral data exhibited noise and drift, and the 5-point Savitzky-Golay method was applied for smoothing and denoising [18], [19].

We used our in-house program to convert the hyperspectral data to LibSVM format: the category label was added to each training data set, so that the classification algorithm could train the classification model. The HDFS was used for distributed storage and management of the training and verification datasets. The Master node was used as the NameNode to manage the entire HDFS file system and directory tree,

**FIGURE 3.** Flow diagram of experiment.**FIGURE 4.** Regions of interest for a sample.

and one copy was stored in all of the other nodes. Via the Spark platform, the ANN and SVM algorithms were used for classification modeling of the hyperspectral data. For the training set, 70% of the data were randomly selected, with the remaining 30% used as the prediction set to analyze the classification results of both algorithms and the classification performance of the hyperspectral images data, based on the Spark platform.

Figure 5 presents the mean reflectance spectra of six types of experimental sample. It is evident that, in the blue waveband at 440 nm and the red waveband at 680 nm, the chlorophyll of the oilseed rape siliques exhibited strong radiation absorption ($> 80\%$), shown by the two red valleys. The absorption between these two wavebands (near 550 nm) declined, forming a green peak (25–45% absorption). Reflectivity increased (65–85%) between the visible light band

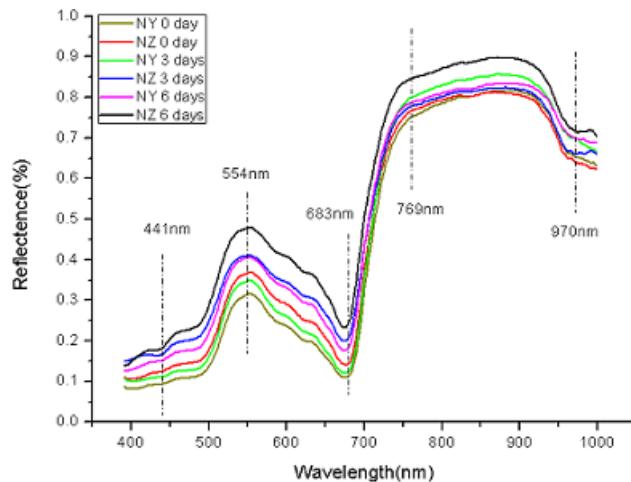


FIGURE 5. Mean reflectance spectra of six types of experimental samples.

and near-infrared band (near 760 nm), and formed a red edge. Between 780–900 nm, the reflectivity of the green plant spectrum was maintained at 60%–85%. The reflectance spectrum of 800–1,000 nm is related to the cell structure and water content of the siliques [20]. Sclerotinia rot of colza was previously diagnosed based on similar spectra curves in a hyperspectral image of oilseed rape stems (wavelength range: 384–1,034 nm) [21]; the stem had a spectral curve similar to that of the corresponding leaves [22]. Another study evaluated seed output, detecting similar spectral curves on a hyperspectral image of oilseed rape leaves [23]. In our study, the siliques and stems contained similar natural pigments, such as chlorophyll, xanthophyll and carotenoid; thus, we found that the siliques and leaves exhibited similar hyperspectral curves in the visible light and near-infrared bands.

Additionally, as the waterlogging level increased, so too did the reflectivity of both varieties of siliques in the visible light and near-infrared bands (400–1,000 nm). The reflectivity of the NZ 19 siliques exceeded that of NY 22, and a higher waterlogging level was associated with higher reflectivity. The green peak near the green waveband at 550 nm, and the red valley near the red waveband at 680 nm, were the major peaks within the visible light spectrum for plant chlorophyll. Differences in the green peaks and red valleys of siliques under different waterlogging levels may be attributable to differences in chlorophyll content.

A. NEURAL NETWORK ALGORITHM

A multi-layer feedforward neural network algorithm was realized by adopting the MLlib for Spark. IntelliJ IDEA 2016 was used as the IDE, and Scala was the programming language. Hadoop 2.7.5 was used as the distributed storage framework; Spark was used as the distributed computation framework for compiling and executing the algorithm.

The feedforward neural network had a three-layer structure. As illustrated in Fig. 6, the first layer was the input layer, and the number of nodes on the input layer corresponded to

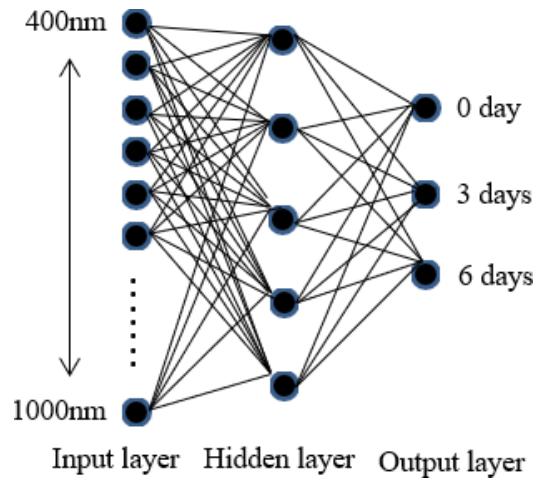


FIGURE 6. Schematic diagram of the feedforward neural network.

the number of sample characteristics; the second layer was the hidden layer, for which the node number was set manually; the third layer was the output layer, wherein the node number corresponded to the target number of sample characteristics. During forward propagation of the signal, the input layer nodes were input into the hidden layer and the output nodes of the hidden layer were used as the input to the output layer [24], [25].

The neural network parameters included the weights and bias terms of the input layer and the hidden layer. w_i^k was the connection weight between the i^{th} unit on the k^{th} layer and the i^{th} unit on the $(k+1)^{th}$ layer; b_k was the bias term on the k^{th} layer. The signal was transmitted forward, and the node output of output layer was as follows:

$$y(x) = f_k \left(w_i^k \cdots f_2 \left(w_2^k f_1 \left(w_1^k x + b_1 \right) + b_2 \right) \cdots + b_k \right) \quad (1)$$

The sigmoid function was used as the activation function for the intermediate layer. The output neuron computed the output value of the entire neural network, according to the input value and activation function.

$$f(x_i) = \frac{1}{1 + e^{-x_i}} \quad (2)$$

During the backpropagation of errors, the partial derivative of the objective function, which has various neuronal weights, was computed layer by layer from the output layer. The weights and threshold values of various layers were adjusted using the gradient descent algorithm, and the weight value was updated until the final output value of the network approximated the expected value after modification. Because the vanishing gradient problem tends to arise in the sigmoid function, the SoftMax function was applied to the gradient descent algorithm for dealing with multiple classification problems during backpropagation.

$$f(x_i) = \frac{e^{x_i}}{\sum_{i=1}^n e^{x_i}} \quad (3)$$

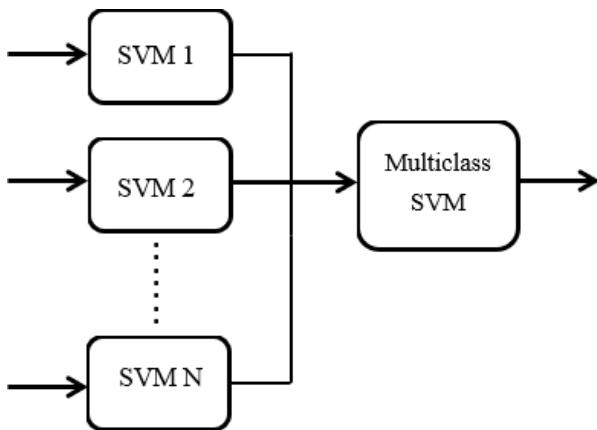


FIGURE 7. Schematic diagram of support vector machine based on the one-against-rest.

B. SUPPORT VECTOR MACHINE ALGORITHM

The SVM algorithm was initially designed for application to binary classification problems. For handling multi-class problems, we constructed a suitable SVM to achieve multi-class classification, as shown in Fig. 7. We used MLlib for Spark and realized the one-against-rest (OAR) classifier via multiple binary classifiers [26]. During assembly of the training sample, the samples of one oilseed rape type were classified into one class, and the samples of the other oilseed rape type were classified into another class. Thus, for samples of k categories, k SVM functions were built and, during the classification process, the unknown samples were classified into a class with the highest classification function value [27].

Assuming n training data $(x_1, y_1), \dots, (x_n, y_n)$, the i^{th} SVM trainer was required to solve the following problem:

$$\begin{aligned} \min_{\omega, b, \xi} \quad & \frac{1}{2} \omega^T \omega + C \sum_{i=1}^l \xi_i \\ \text{subject to} \quad & y_i (\omega^T \phi(x_i) + b) \geq 1 - \xi_i \\ \xi_i \geq 0, \quad & i = 1, \dots, l \end{aligned} \quad (4)$$

In which ω is a high-dimension vector, C is the penalty function, and $\phi(x_i)$ and C maps x_i to high-dimensional space. In Formula (4), the minimum value of $\frac{1}{2}\omega^T\omega$ refers to the maximize term $\frac{1}{2}\|\omega\|^2$, which represents the margin between two groups of data; $C \sum_{i=1}^l \xi_i$ represents the penalty term, which is used to minimize training errors. SVM primarily solves the balance problem between the regularization term and the training error [26]. Following the resolution of Formula 4, k classification decision functions may be obtained:

$$(\omega^1)^T \phi(x) + b^1, \dots, (\omega^k)^T \phi(x) + b^k \quad (5)$$

During the SVM classification, the unknown sample x was classified into the category with the largest decision function

value.

$$x = \arg \max_{i=1 \dots k} \left((\omega^i)^T \phi(x) + b^i \right) \quad (6)$$

By building multiple SVM classifiers, we can realize the multi-class SVM classifier.

C. CLASSIFICATION MODEL BASED ON SPARK

The hardware and software in a data center together comprise the “cloud” Internet computing platform. The purpose of parallel computing is to provide reliable and scalable services based on virtualization, massive data management, and the deployment of resources to perform tasks while ensuring quality of service (QOS) [28].

Spark is a distributed parallel computing framework based on Resilient Distributed Datasets (RDDs) and memory computing, which was developed with support from The Apache Software Foundation. As Spark boasts the characteristics of rapidity, ease of use, high performance and scalability, it has emerged as an important parallel computing platform.

Spark is also equipped with a special MLlib that optimizes support for machine learning algorithms; in logistic regression tests, its computing speed is 100 times higher than that of Hadoop’s MapReduce model [29]. The advantages of Spark are highly suited to realizing multiple iterations in machine learning, and this has significantly improved the operational speed of algorithms.

D. EVALUATION OF CLASSIFICATION RESULTS

1) ACCURACY

In the below formula 7, TP refers to the number of positive samples and TN to the number of negative samples; N is the total number of samples.

$$\text{Accuracy} = \frac{TP + TN}{N} \quad (7)$$

Higher accuracy denotes better classification results.

2) F-SCORE

Precision is the ratio between the number of accurately classified samples and the number of actually detected samples in the classified samples; Recall is the ratio between the number of accurately classified samples and the total number of samples. In the below formula 8 for F-Score, P represents Precision and R represents Recall.

$$F\text{-Score} = \frac{(\alpha^2 + 1)(P \times R)}{\alpha^2(P + R)} \quad (8)$$

The F-Score is the weighted harmonic mean of Precision and Recall. Generally, $\alpha = 1$

$$F1\text{-Score} = \frac{2 \times (P \times R)}{P + R} \quad (9)$$

A higher F1 score indicates a more effective classification algorithm.

TABLE 3. Three-class classification results of the ANN and SVM prediction model.

ANN	600 input layers 1hidden layer 3 output layers					SVM	Iteration times 80				
NY 22	0	3	6	Accuracy	F1-Score	NY 22	0	3	6	Accuracy	F1-Score
0	31	5	0	86.11%	86.11%	0	29	7	0	80.55%	84.70%
3	5	29	2	80.55%	78.26%	3	3	28	5	77.78%	71.72%
6	0	4	32	88.89%	91.25%	6	0	6	30	83.33%	83.98%
Overall	36	38	34	85.18%	85.20%	Overall	32	41	35	80.55%	80.13%
NZ 19						NZ 19					
0	34	2	0	94.44%	95.42%	0	32	4	0	88.89%	93.32%
3	1	32	3	88.89%	87.47%	3	1	30	5	83.33%	83.33%
6	0	3	33	91.66%	91.66%	6	0	2	34	94.44%	90.36%
Overall	35	37	36	91.71%	91.51%	Overall	33	36	39	88.89%	89.01%

TABLE 4. Three-class classification results for all samples combined.

ANN	600 input layers 1hidden layer 3 output layers					SVM	Iteration times 130				
	0	3	6	Accuracy	F1-Score		0	3	6	Accuracy	F1-Score
0	65	7	0	90.27%	90.84%	0	61	11	0	84.72%	89.03%
3	7	63	2	87.50%	85.29%	3	4	66	2	91.66%	81.98%
6	0	6	66	91.66%	94.22%	6	0	12	60	83.32%	89.51%
Overall	72	76	68	89.81%	90.11%	Overall	65	89	62	86.56%	86.84%

IV. EXPERIMENT AND RESULT ANALYSIS

A. HYPERSPECTRAL MULTI-CLASS DETECTION OF SILIQUES WATERLOGGING

Prior to hyperspectral classification of the oilseed rape siliques, 70% of the data were used for training, and the remaining 30% for prediction. To minimize errors in classification and the running time, each hyperspectral sample was analyzed 50 times; the mean value was used for classification of the sample and the running time was recorded.

Tables 3 present the classification results for both oilseed rape varieties, obtained using the ANN and SVM algorithms, respectively. The results indicate that for the NY 22 variety, the ANN algorithm had a classification accuracy of 85.18% and an F1 score of 85.20%; the SVM algorithm had a classification accuracy of 80.55% and an F1 score of 80.13%. For the NZ 19 variety, the ANN algorithm had a classification accuracy of 91.71% and an F1 score of 91.51%; and the SVM algorithm had a classification accuracy of 88.89% and an F1 score of 89.01%. Finally, the samples of both varieties were combined (720 samples in total) for the classification of waterlogging levels. Again, 70% of the data were used for training, and 30% for verification. The ANN and SVM algorithms were used for classification training and prediction, and the training results are presented in Table 4.

The ANN algorithm had an accuracy of 89.81% and an F1 score of 90.11%, while the SVM algorithm had an accuracy of 86.56% and an F1 score of 86.84%. For multi-class problems, the accuracy of the ANN algorithm was therefore superior overall to that of the SVM algorithm. Previously, various classification and discrimination methods were used to

analyze hyperspectral images of farmed salmon, including a computer vision system (CVS), partial least squares discriminant analysis (PLS-DA), the SVM algorithm, and the random forest (RF) algorithm. The results indicate that the discrimination analysis method was well-suited to the classification and identification of hyperspectral images, and the SVM algorithm yielded the highest classification accuracy [30]. Furthermore, support vector machine and neural network were used for hyperion hyperspectral image for classification of cluttered units in an ophiolite set, the results show that the SVM and NN method have overall accuracies of 52% and 65% respectively [31]. In the literature [32], MATLAB was used previously for machine learning via ANN and SVM, to analyze hyperspectral images of peaches with different degrees of chilling injury; the results indicated that, for multi-class problems, the accuracy of the ANN algorithm was superior to that of the SVM algorithm. Our results demonstrate that, using the parallel computing platform and hyperspectral data to solve a multi-class problem involving, the ANN algorithm exhibited higher accuracy than the SVM algorithm. The feedforward neural network fitted the function relation between features and targets via the activation function, and the multi-layer neural network with the hidden layer provided excellent classification results. Multi-class classification can be realized by creating multiple binary SVM classifiers using the OAR method; when the volume of data is high, better generalization results can be obtained through data regularization, which can in turn minimize generalization errors. The ANN and SVM algorithms were used to classify and discriminate two varieties of oilseed rape siliques (NY 22 and NZ 19).

TABLE 5. Binary-class classification results of samples of different oilseed rape varieties.

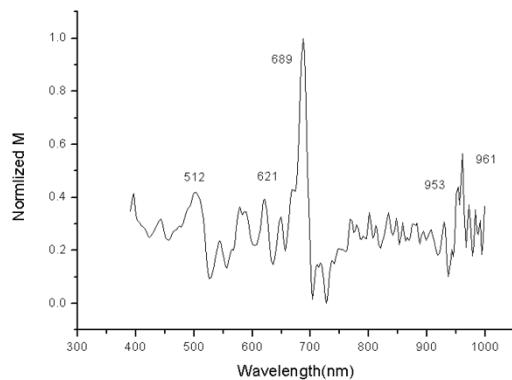
	600 input layers 1hidden layer 2 output layers			
ANN	NY 22	NZ 19	Accuracy	F1-Score
NY 22	91	17	84.25%	86.78%
NZ 19	10	98	90.74%	87.82%
Overall	101	115	87.49%	87.30%
SVM	Iteration times 80			
NZ 19	14	94	87.03%	90.03%
NY 22	102	6	94.44%	91.05%
Overall	116	100	90.73%	90.54%

Both algorithms were applied to all siliques samples for binary classification and discrimination; 70% of the data were used for training, while 30% for prediction. The results are as shown in Table 5. The ANN algorithm had an accuracy of 87.49% and an F1 score of 87.30%, and the SVM algorithm had an accuracy of 90.73% and an F1 score of 90.54%; thus, for the binary classification problem of oilseed rape siliques, the accuracy of the SVM algorithm was higher than that of the ANN algorithm. Moreover, owing to simplification of the classification model, SVM could be applied to obtain a better hyperplane and compute the classification decision function, so that the accuracy of the SVM algorithm was significantly improved.

B. CLASSIFICATION AND DISCRIMINATION OF OPTIMAL WAVEBANDS

Because the hyperspectral image contains a large amount of information, it is time-consuming to model and verify the classification algorithm. Currently, extraction of the optimal waveband of a spectral image appears to be the most effective method for reducing data volume. In this study, the optimal waveband was determined by obtaining the maximum M value of the original average spectrum and conducting normalization processing. As shown in the following diagram (Fig. 8), wavebands with wavelength M values of 512, 621, 689, 953 and 961 nm were selected as the optimal wavebands. The corresponding greyscale single waveband is illustrated in Fig. 9. Applying the same experimental method, three ROIs were selected from each siliques sample, and each ROI had an area of 58×40 (2,320) pixels. For each image, 3×10 ROIs were then drawn for extracting sample information from the image.

The optimal waveband was used for classification of the hyperspectral data of oilseed rape waterlogging. The results obtained using optimal wavebands for classification and discrimination are shown in Tables 6. For the NY22 oilseed rape variety, the ANN algorithm had a prediction accuracy of 80.56% and an F1 score of 80.68%, and the SVM algorithm had a prediction accuracy of 81.48% and an F1 score of 81.48%. For the NZ 19 variety, the ANN algorithm had a prediction accuracy of 90.73% and an F1 score of 90.35%, and the SVM algorithm had a prediction accuracy of 87.35% and an F1 score of 87.17%. It is clear, therefore, that for the classification and discrimination of optimal wavebands,

**FIGURE 8.** Optimal waveband selected according to normalized M values.

the ANN algorithm had superior average accuracy and F1 scores to those of the SVM algorithm. In a previous study, five optimal wavebands were identified, and the ANN algorithm was used to classify and discriminate hyperspectral images of Red Delicious apples that had sustained chilling injury, with the average classification accuracy being 98.4% [33]. Elsewhere, eight optimal wavebands were identified for hyperspectral images of chilling injuries sustained by peaches; in that study, the ANN algorithm was used for classification and the classification accuracy was 95.8% [34]. Our experimental results indicate that, the classification accuracy of full-waveband hyperspectral imaging was slightly higher than that of optimal waveband imaging.

We combined the optimal waveband data of both varieties of siliques, and used the SVM and ANN algorithms for classification and analysis; the results are presented in Table 7. The ANN algorithm had an accuracy of 84.25% and an F1 score of 84.25%, and the SVM algorithm had an accuracy of 79.16% and an F1 score of 79.12%. Again, the performance of the ANN algorithm was superior to that of the SVM algorithm. Additionally, we observed that the hyperspectral optimal waveband classification result for the NZ 19 variety was better than that of the NY 22 variety, as was the case in the full-waveband analysis. The classification results indicate that using the optimal waveband for classification and modeling can reduce the hyperspectral data volume, and the complexity of the classification model. Although the optimal waveband denotes key data within the training set, it also minimizes the volume of classification and modeling data. Therefore, its accuracy was decreased in comparison to that of the full-waveband classification.

C. PERFORMANCE EVALUATION OF PARALLEL COMPUTING

1) SPEEDUP

Speedup refers to the ratio between the time taken to run a task in the stand-alone versus parallel environment, and is used to measure the parallel computation performance and results of a parallel system. By maintaining the same dataset volume, system performance can be optimized by altering the number

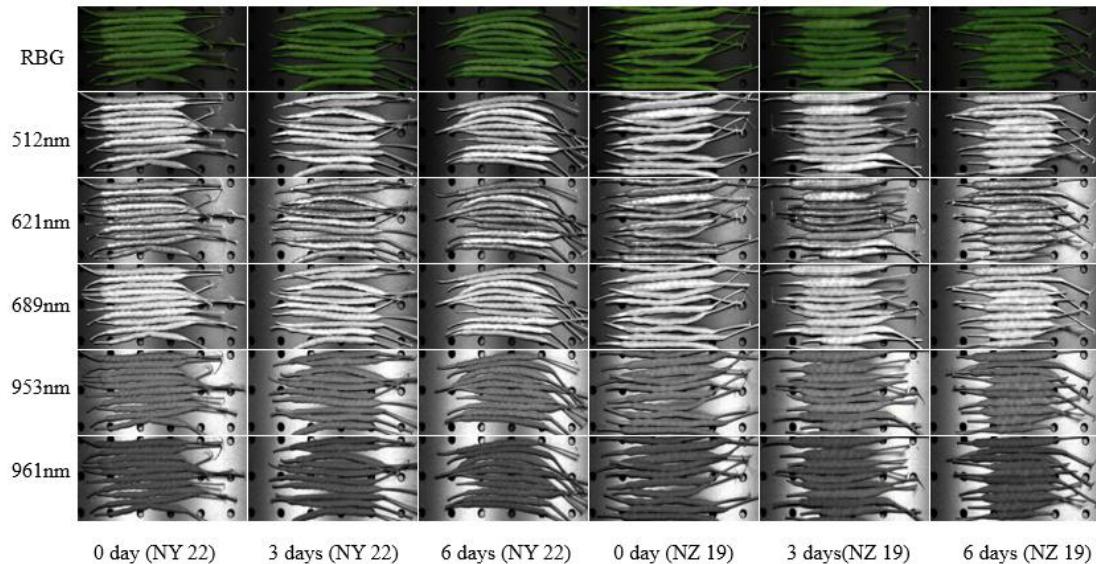


FIGURE 9. Individual band hyperspectral grayscale images of the waterlogging samples.

TABLE 6. Three-class classification results: optimal wavebands identified by the ANN and SVM algorithm.

ANN	5 input layers 1hidden layer 3 output layers					SVM	Iteration times 25				
	NY 22	0	3	6	Accuracy	F1-Score	NY 22	0	3	6	Accuracy
0	29	6	1	80.55%	86.47%	0	31	5	0	86.11%	89.80%
3	2	30	4	83.33%	74.83%	3	2	27	7	75.00%	72.94%
6	0	8	28	77.78%	80.74%	6	0	6	30	83.33%	81.70%
Overall	31	44	33	80.56%	80.68%	Overall	33	38	37	81.48%	81.48%
NZ 19						NZ 19					
0	33	3	0	91.66%	92.47%	0	30	6	0	83.33%	87.71%
3	2	31	3	86.11%	86.11%	3	2	33	1	91.66%	82.22%
6	0	2	34	94.44%	92.47%	6	0	5	31	86.11%	91.60%
Overall	35	36	37	90.73%	90.35%	Overall	32	44	32	87.35%	87.17%

TABLE 7. Three-class classification results: optimal wavebands for the two oilseed rape varieties.

ANN	5 input layers 1hidden layer 3 output layers					SVM	Iteration times 60				
	0	3	6	Accuracy	F1-Score		0	3	6	Accuracy	F1-Score
0	64	6	2	88.89%	90.08%	0	61	8	3	84.72%	85.30%
3	3	58	11	80.55%	79.94%	3	7	52	13	72.22%	72.70%
6	3	9	60	83.33%	82.72%	6	3	11	58	80.55%	79.38%
Overall	70	73	73	84.25%	84.25%	Overall	71	71	74	79.16%	79.12%

of nodes participating in the computation. The computational formula for speedup is:

$$\text{Speedup}(n) = \frac{T_1}{T_n} \quad (10)$$

Where T_1 is the execution time when running an algorithm on a single node, and T_n is the execution time when n nodes are running the same algorithm on the same dataset. The ANN and SVM algorithms were run for all of the collected

hyperspectral datasets when there were 1–6 nodes, and the algorithm running time was recorded. Each hyperspectral dataset had a size of 43 MB, and the total dataset size was 3.09 GB (43 MB \times 72).

According to the running time, the SVM algorithm was more cost-efficient than the ANN algorithm. With the increase in node number, speedup of both algorithms increased linearly with the increase in nodes; speedup exhibited a continuously increasing trend, indicating that

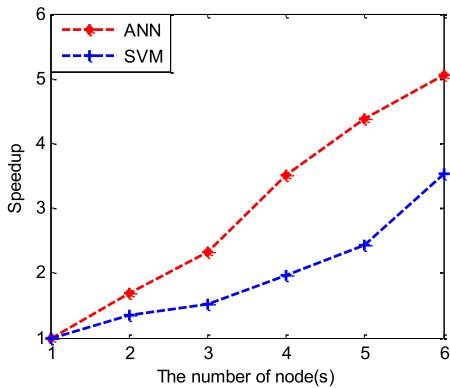


FIGURE 10. Speedup by number of nodes.

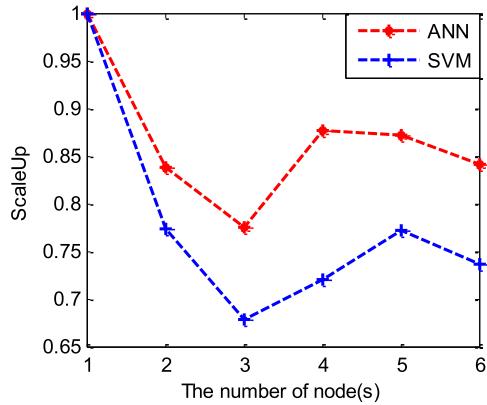


FIGURE 11. Scaleup by number of nodes.

both algorithms had excellent linear speedup performance when applied to hyperspectral big datasets. Additionally, Fig. 10 indicates that the neural network training time exceeded that of the SVM algorithm in a single-node environment, such that the ANN algorithm exhibited superior speedup performance versus the SVM algorithm. However, regarding the algorithm running times, the SVM algorithm was faster than the ANN algorithm.

2) SCALEUP

Scaleup refers to the ability of an algorithm to improve its performance proportional to an increase in node number. The computational formula is:

$$\text{Scaleup}(n) = \frac{\text{Speedup}(n)}{n} \quad (11)$$

Where n is the node number, and $\text{Speedup}(n)$ represents the speedup of n nodes [35].

As Fig. 11 illustrates, when the node number was between 1 and 3, the scaleup of both algorithms gradually declined. This indicates that with an increase in node number, Spark would spend more time on task scheduling and communication. When the node number exceeded 3, the scaleup of both algorithms increased significantly. When the node number was 4 or 5, the scaleup of the ANN algorithm declined slightly, while the SVM algorithm exhibited a slight increase in scaleup; when the node number was 6, the scaleup of

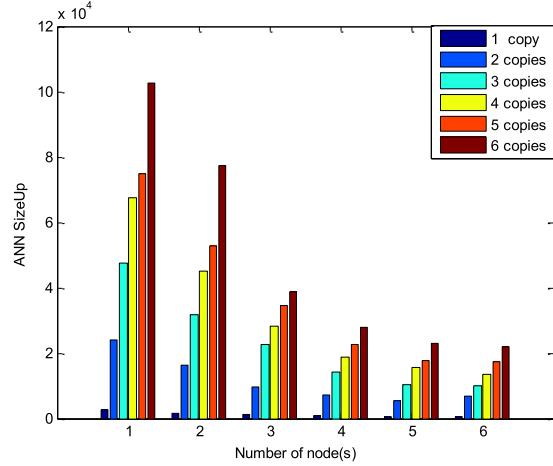


FIGURE 12. Sizeup of artificial neural network (ANN) algorithm.

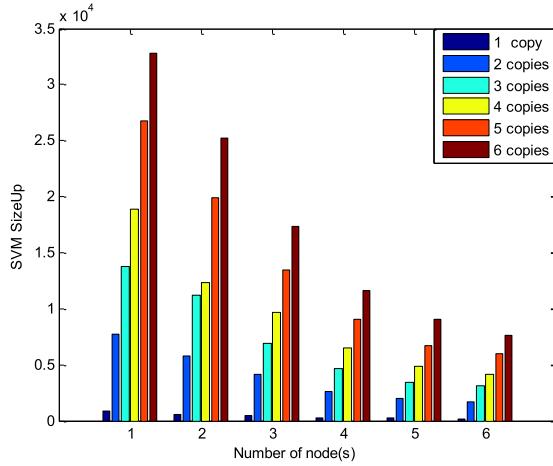


FIGURE 13. Sizeup of support vector machine (SVM) algorithm.

both algorithms exhibited a significant decline. This indicates that the modeling and prediction time of classification algorithms can be significantly reduced when the node number is increased, where this is associated with the greater time requirements of the system.

3) SIZEUP

Sizeup refers to alterations in parallel algorithm performance in association with an increase in data volume. When the node number is unchanged, sizeup can be calculated by changing the size of the dataset. The computational formula is as follows:

$$\text{Sizeup}(n) = \frac{T_{ndb}}{T_{db}} \quad (12)$$

where T_{db} refers to the execution time when the system applies the algorithm to the original dataset, and T_{ndb} refers to the time required to run the same algorithm following expansion of the original dataset by n time [36]. To test the sizeup performance of big datasets on the Spark platform, we expanded the original hyperspectral dataset, making six copies having data volumes of 3.09, 6.19, 9.28, 12.38,

15.48 and 18.57 GB, respectively [37]. Both the ANN and SVM algorithms were tested on the various nodes.

As Figs. 12 and 13 illustrate, the sizeup of the ANN algorithm was significantly greater than that of the SVM algorithm, indicating that the SVM algorithm was more time-efficient for big data classification. When the node number was at least three, the sizeup of the ANN algorithm exhibited significant decline, thus indicating that its efficiency was improved significantly such that the computational performance of the nodes could be better-exploited. Compared to the ANN algorithm, the sizeup of the SVM algorithm was more stable and, with increased data volume, the consumption time exhibited linear growth.

V. CONCLUSION

In this study, parallel computing technology was used to classify hyperspectral images of oilseed rape siliques under different waterlogging conditions, and the experimental results verify the feasibility and efficiency of combining parallel computing technology and machine learning methods for analyzing agricultural hyperspectral data. Among the multi-class classification algorithms, the classification accuracy and F1 score of the ANN algorithm were higher than those of the SVM algorithm; among the binary classification algorithms, the classification accuracy and F1 score of the SVM algorithm were higher than those of the ANN algorithm. Therefore, for multi-class classification on the parallel platform, the ANN algorithm is more suitable, while the SVM algorithm is better suited to binary classification problems. The classification model was assembled by selecting the optimal waveband, and the results indicated that, for the classification of hyperspectral optimal wavebands, the accuracy of the ANN algorithm was superior to that of the SVM algorithm. When hyperspectral datasets were run on the parallel platform, both algorithms exhibited excellent linear speedup performance. When the number of nodes was less than three, the sizeup of both algorithms showed a decreasing trend; with a continuous increase in node number, the sizeup of the ANN algorithm showed a slight decrease, while the sizeup of the SVM algorithm showed a slight increase. When the node number reached six, the sizeup of both algorithms declined. When the node number was at least three, the sizeup of the ANN algorithm decreased significantly, but no major change was detected in the sizeup of the SVM algorithm. With the increasingly wide application of hyperspectral technology in various academic fields, hyperspectral data mining and pattern recognition based on parallel computing technology and machine learning methods provide new research and development venues combining computer science with other academic disciplines.

REFERENCES

- [1] United States Department of Agriculture. (May 2018). *Oilseeds: World Markets and Trade, Global Oilseed Demand Growth Forecast to Out-pace Production*. [Online]. Available: <https://apps.fas.usda.gov/psdonline/circulars/oilseeds.pdf>
- [2] A. Gobin, "Weather related risks in Belgian arable agriculture," *Agricul. Syst.*, vol. 159, pp. 225–236, Jan. 2017.
- [3] Y.-H. Lee, K.-S. Kim, Y.-S. Jang, J.-H. Hwang, D.-H. Lee, and I.-H. Choi, "Global gene expression responses to waterlogging in leaves of rape seedlings," *Plant Cell Rep.*, vol. 33, no. 2, pp. 289–299, Feb. 2014.
- [4] M. Xu et al., "The effect of waterlogging on yield and seed quality at the early flowering stage in *Brassica napus L.*," *Field Crop Res.*, vol. 180, pp. 238–245, Aug. 2015.
- [5] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.
- [6] M. Huang and R. Lu, "Apple mealiness detection using hyperspectral scattering technique," *Postharvest Biol. Technol.*, vol. 58, no. 3, pp. 168–175, Aug. 2010.
- [7] M. Shuaibu, W. S. Lee, J. Schueller, P. Gader, Y. Ki, Hong, and S. Kim, "Unsupervised hyperspectral band selection for apple Marssonina blotch detection," *Comput. Electron. Agricult.*, vol. 148, pp. 45–53, May 2018.
- [8] J. Li, W. Huang, X. Tian, C. Wang, S. Fan, and C. Zhao, "Fast detection and visualization of early decay in citrus using Vis-NIR hyperspectral imaging," *Comput. Electron. Agricult.*, vol. 127, pp. 582–592, Sep. 2016.
- [9] N. Caporaso, M. B. Whitworth, M. S. Fowler, and I. D. Fisk, "Hyperspectral imaging for non-destructive prediction of fermentation index, polyphenol content and antioxidant activity in single cocoa beans," *Food Chem.*, vol. 258, pp. 343–351, Aug. 2018.
- [10] U. Knauer, A. Matros, T. Petrovic, T. Zanker, E. S. Scott, and U. Seiffert, "Improved classification accuracy of powdery mildew infection levels of wine grapes by spatial-spectral analysis of hyperspectral images," *Plant Methods.*, vol. 13, pp. 1–15, Jun. 2017.
- [11] S. Shrestha, M. Knapič, U. Žibrat, L. C. Deleuran, and R. Gislum, "Single seed near-infrared hyperspectral imaging in determining tomato (*Solanum lycopersicum L.*) seed quality in association with multivariate data analysis," *Sens. Actuators B, Chem.*, vol. 237, pp. 1027–1034, Dec. 2016.
- [12] V. Q. Nguyen, S. N. Nguyen, and K. Kim, "Design of a platform for collecting and analyzing agricultural big data," *J. Digit. Contents Soc.*, vol. 18, no. 1, pp. 149–158, Jan. 2017.
- [13] J. Woodard, "Big data and Ag-Analytics: An open source, open data platform for agricultural & environmental finance, insurance, and risk," *Agricul. Finance Rev.*, vol. 76, no. 1, pp. 15–26, May 2016.
- [14] R. Lokers, R. Knapen, S. Janssen, Y. van Randen, and J. Jansen, "Analysis of big data technologies for use in agro-environmental science," *Environ. Model. Softw.*, vol. 84, pp. 494–504, Oct. 2016.
- [15] J. Xia, Y. Yang, H. Cao, C. Han, D. Ge, and W. Zhang, "Visible-near infrared spectrum-based classification of apple chilling injury on cloud computing platform," *Comput. Electron. Agricult.*, vol. 145, pp. 27–34, Feb. 2018.
- [16] S. Vatari, A. Bakshi, and T. Thakur, "Green house by using IOT and cloud computing," in *Proc. Int. Conf. IEEE Int. Conf. Recent Trends Electron., Inf. Commun. Technol.(RTEICT)*, May 2016, pp. 246–250.
- [17] P. P. Jayaraman, A. Yavari, D. Georgakopoulos, A. Morshed, and A. Zaslavsky, "Internet of Things platform for smart farming: Experiences and lessons learnt," *Sensors*, vol. 16, no. 11, p. 1884, Nov. 2016.
- [18] C. Ruffin and R. L. King, "The analysis of hyperspectral data using Savitzky–Golay filtering-theoretical basis. I," in *Proc. Int. Conf. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jun./Jul. 1999, pp. 756–758.
- [19] J. Luo, K. Ying, and J. Bai, "Savitzky–Golay smoothing and differentiation filter for even number data," *Signal Process.*, vol. 85, no. 7, pp. 1429–1434, Feb. 2005.
- [20] N. M. Merzlyak, A. E. Solovchenko, and A. A. Gitelson, "Reflectance spectral features and non-destructive estimation of chlorophyll, carotenoid and anthocyanin content in apple fruit," *Postharvest Biol. Technol.*, vol. 27, no. 2, pp. 197–211, Feb. 2003.
- [21] W. Kong, C. Zhang, W. Huang, F. Liu, and Y. He, "Application of hyperspectral imaging to detect *sclerotinia sclerotiorum* on oilseed rape stems," *Sensors*, vol. 18, no. 1, pp. 123–140, Jan. 2018.
- [22] C. Zhang, F. Liu, W. Kong, and Y. He, "Application of visible and near-infrared hyperspectral imaging to determine soluble protein content in oilseed rape leaves," *Sensors*, vol. 15, no. 7, pp. 16576–16588, Jul. 2015.
- [23] X. Zhang and Y. He, "Rapid estimation of seed yield using hyperspectral images of oilseed rape leaves," *Ind. Crops Products*, vol. 42, no. 1, pp. 416–420, Mar. 2013.
- [24] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 2, pp. 513–529, Apr. 2012.

- [25] N.-Y. Liang, G.-B. Huang, P. Saratchandran, and N. Sundararajan, "A fast and accurate online sequential learning algorithm for feedforward networks," *IEEE Trans. Neural Netw.*, vol. 17, no. 6, pp. 1411–1423, Nov. 2006.
- [26] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Trans. Neural Netw.*, vol. 13, no. 2, pp. 415–425, Mar. 2002.
- [27] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27–54, Apr. 2011.
- [28] M. Armbrust *et al.*, "A view of cloud computing," *Commun. ACM*, vol. 53, no. 4, pp. 50–58, 2010.
- [29] Apache Software Foundation. (Feb. 2018). *Apache Spark is a Unified Analytics Engine for Large-Scale Data Processing*. [Online]. Available: <http://spark.apache.org>
- [30] J.-L. Xu, C. Riccioli, and D.-W. Sun, "Comparison of hyperspectral imaging and computer vision for automatic differentiation of organically and conventionally farmed salmon," *J. Food Eng.*, vol. 196, pp. 170–182, Mar. 2017.
- [31] B. Bahrambeygi and H. Moeinzadeh, "Comparison of support vector machine and neural network classification method in hyperspectral mapping of ophiolite mélange—A case study of east of Iran," *Egyptian J. Remote Sens. Space Sci.*, vol. 20, no. 1, pp. 1–10, Feb. 2017.
- [32] Y. Sun *et al.*, "Hyperspectral reflectance imaging combined with chemometrics and successive projections algorithm for chilling injury classification in peaches," *LWT*, vol. 75, pp. 557–564, Jan. 2016.
- [33] G. ElMasrya, N. Wang, and C. Vigneault, "Detecting chilling injury in Red Delicious apple using hyperspectral imaging and neural networks," *Postharvest Biol. Technol.*, vol. 52, no. 1, pp. 1–8, Nov. 2008.
- [34] L. Pan, Q. Zhang, W. Zhang, Y. Sun, P. Hu, and K. Tu, "Detection of cold injury in peaches by hyperspectral reflectance imaging and artificial neural network," *Food Chem.*, vol. 192, pp. 134–141, Feb. 2016.
- [35] D. S. Linthicum, "Approaching cloud computing performance," *IEEE Cloud Comput.*, vol. 5, no. 2, pp. 33–36 Mar. 2018.
- [36] M. Y. Hai Zhang, "Performance comparison of clustering algorithms in spark," *Comput. Sci.*, vol. 44, no. 6A, pp. 414–418 Jun. 2017.
- [37] I. Mavridis and H. Karatza, "Performance evaluation of cloud-based log file analysis with Apache Hadoop and Apache Spark," *J. Syst. Softw.*, vol. 125, pp. 133–151, Mar. 2017.



JIAN XIA was born in Nanjing, Jiangsu, China. He is currently pursuing the Ph.D. degree with the School of Computer Science and Engineering, Nanjing University of Science and Technology, China. His research interests include agricultural big data cloud computing, and classification of hyperspectral data.



YUWANG YANG received the B.S. degree from North Western Polytechnical University in 1988, the M.S. from the University of Science and Technology of China in 1991, and the Ph.D. degree from the Nanjing University of Science and Technology in 1996. He is currently a Professor with the School of Computer Science and Engineering, NUST. His research interests are high performance computing, machine learning, and intelligent system.



HONGXIN CAO received the B.S. and M.S. degrees from Northwestern Agricultural University in 1985 and 1988, respectively, and the Ph.D. degree from Nanjing Agriculture University in 1997. He is currently a Professor with the Institute of Agricultural Information, Jiangsu Academy of Agricultural Sciences. His research interests are research and application for the key technology of informatization agriculture.



WEIXIN ZHANG received the M.S. degree in information agriculture from Nanjing Agriculture University, China, in 2013. She is currently a Research Assistant with the Institute of Agricultural Information, Jiangsu Academy of Agricultural Sciences, Nanjing, China. Her research interests include the areas of agricultural model.



LEI XU received the bachelor's, master's, and Ph.D. degrees in communication and information system from the Nanjing University of Aeronautics and Astronautics, China, in 2006, 2009, and 2012, respectively. He is currently a Full Professor with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China. His research interests include network analysis, network security, and big data analysis.



QIAN WAN was born in Zhumadian, Henan, China. She is currently pursuing the degree with Nanjing Agricultural University. She is currently doing experiments on rapeseed yield in Jiangsu Academy of Agricultural Sciences.



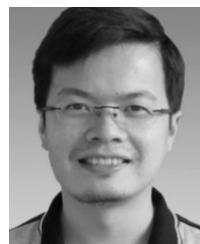
YAQI KE received the master's degree in the discipline of phonology from Nanjing Agriculture University, China, in 2018. She is currently a Research Assistant with the Nanjing University of Science and Technology, Nanjing, China. Her research interests include the areas of agricultural model, agricultural informatics, and machine learning.



WENYU ZHANG was born in Wuxi, Jiangsu, China. He received the Ph.D. degree from the College of Agriculture, Nanjing Agricultural University, China, in 2011. He is currently an Associate Professor with the Institute of Agricultural Information, Jiangsu Academy of Agricultural Sciences. His research interests include crop modeling, smart agriculture, and agricultural IoT.



DAOKUO GE received the bachelor's degree from Nanjing Agricultural University in 2002. He is currently a Professor with the Institute of Agricultural Information, Jiangsu Academy of Agricultural Sciences. His research interests include agricultural big data, crop growth simulation models, and hyperspectral diagnostics.



BO HUANG received the B.S. and Ph.D. degrees from the Nanjing University of Science and Technology (NUST), Nanjing, China, in 2002 and 2006, respectively. He joined NUST in 2007, where he is currently an Associate Professor with the School of Computer Science and Engineering. His research interests include intelligent transportation and AI systems. He has published 50+ in the above areas.

• • •