

Received May 4, 2020, accepted May 9, 2020, date of publication May 15, 2020, date of current version June 1, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2994913

Online Sorting of the Film on Cotton Based on Deep Learning and Hyperspectral Imaging

CHAO NI¹, ZHENYE LI¹, XIONG ZHANG¹, XINYAN SUN¹, YUPING HUANG¹, LING ZHAO²,
TINGTING ZHU¹, AND DONGYI WANG³

¹College of Mechanical and Electronic Engineering, Nanjing Forestry University, Nanjing 210037, China

²College of Mechanical and Automotive Engineering, Liaocheng University, Liaocheng 252000, China

³Bio-Imaging and Machine Vision Laboratory, Fischell Department of Bioengineering, University of Maryland, College Park, MD 20740, USA

Corresponding authors: Yuping Huang (837047549@qq.com), Tingting Zhu (tingting_zhu2018@163.com), and Dongyi Wang (dywang@umd.edu)

This work was supported in part by the Six Talent Peaks Project in Jiangsu Province under Grant 013040315, in part by the China Textile Industry Federation Science and Technology Guidance Project under Grant 2017107, in part by the National Natural Science Foundation of China under Grant 31570714, and in part by the Startup Foundation for Talented Scholars of Nanjing Forestry University (NJFU) under Grant 163040129.

ABSTRACT Mulch film is usually mixed in with cotton during machine-harvesting and processing, which reduces the cotton quality. This paper presents a novel sorting algorithm for the online detection of film on cotton using hyperspectral imaging with a spectral region of 1000 - 2500 nm. The sorting algorithm consists of a group of stacked autoencoders, two optimization modules and an extreme learning machine (ELM) classifier. The variable-weighted stacked autoencoders (VW-SAE) are constructed to extract the features from hyperspectral images, and an artificial neural network (ANN), which is one optimization module, is applied to optimize the parameters of the VW-SAE. Then, the extracted features are input in the ELM to classify four types of objects: background, film on background, cotton and film on cotton. The ELM is optimized by a new optimizer (grey wolf optimizer), which can adjust the hidden nodes and parameters of the ELM simultaneously. A group of experiments was carried out to evaluate the performance of the proposed sorting algorithm using cotton that was provided by a Xinjiang municipality cotton ginning company. The experimental results show that the VW-SAE can improve the classification accuracies by approximately 15 %. The overall recognition rate of the proposed algorithm is over 95 %, and its recognition time is comparable to some state-of-the-art methods.

INDEX TERMS Cotton, sorting system, plastic film, deep learning, hyperspectral imaging, grey wolf optimizer, variable-wise weighted stacked autoencoder.

I. INTRODUCTION

Cotton is one of the most important crops in the world. As the main cotton-producing province in China, Xinjiang has widely applied mulch film covering technology to retain the soil moisture, to maintain the soil structure and to prevent pests [1]. However, mulch film is often mixed with cotton during the machine-harvesting and machine-processing steps, which results in reduced cotton quality. Some techniques have been developed for detecting foreign matter in cotton, such as electrostatic separation, ultrasonic detection and computer vision detection [2], [3]. Electrostatic separation is a rudimentary method that utilizes the charge characteristics to distinguish the film from cotton. However, it is affected by many uncertainties, such as voltage, and it results in

The associate editor coordinating the review of this manuscript and approving it for publication was Mu-Yen Chen.

poor stability and sensitivity, which limit its applications. Ultrasonic sensors identify plastic film according to different densities between plastic film and cotton. However, the relatively low speed for ultrasonic transmission has resulted in a slow identification process [4]. With the development of computer technologies, computer vision techniques with the advantages of low costs, fast speed and consistency have been widely used in foreign matter detection [5]–[11]. Existing computer vision techniques depend on the color differences to distinguish foreign matter and cotton. However, it is difficult to detect foreign matters such as plastic film, which has good photopermeability.

Hyperspectral imaging is an emerging technology that integrates spectroscopy and imaging to obtain both the spectral and spatial information from objects simultaneously [12]. It can detect the chemical compositions and structural features in the spatial domain simultaneously. Hyperspectral

imaging has been applied in agricultural and food inspection since 1990s, but few studies have reported on cotton quality assessments using hyperspectral imaging. Guo [13], [14] reported that hyperspectral imaging in reflectance mode over the spectral range of 400-1000 nm was capable of detecting white and transparent polypropylene fiber, black human hair, and black and transparent PE mulching film from cotton with overall recognition accuracies of 73.2 % and 75.3 % for the training and testing sets, respectively. Moreover, most of other studies employed a hyperspectral imaging system covering a wavelength range longer than 1,000 nm. The Fortier Channel [15] applied Fourier transform near-infrared (FT-NIR) spectroscopy with a wavelength range of 1,100 – 2,400 nm to distinguish the individual types of cotton trash with a 97 % overall prediction accuracy for trash components. However, it is an offline algorithm that focuses on cotton trash samples (hulls, leaves, seed coats, and stems). On the other hand, this method indicates that extending the detection wavelengths beyond 1,000 nm is necessary in order to obtain more useful sample information that is difficult to distinguish from similar color matters.

Currently, multivariate statistical methods and machine learning methods, such as the partial least squares regression (PLSR) [16], the multiple linear regression (MLR) [17], the support vector machine (SVM) [18] and the artificial neural network (ANN) [19], are frequently used to improve hyperspectral signal classification results. While the PLSR and MLR conduct linear analysis between spectra and samples, they are not suitable for parsing complicated mapping relationships, such as the nonlinearity between spectra and samples. Although the SVM could establish nonlinear relationship for samples and spectra, its results depend on kernel functions. The ANN was developed to extract the nonlinear and complex features of samples. However, this method is generally considered a shallow learning approach with a model structure with one hidden layer [19]. Deep learning has been developed to improve the conventional ANN [20]–[22]. It can more greatly learn the hierarchical feature representations and extract the input information layer by layer to represent different levels of nonlinearities [23]–[25].

In this research, a hyperspectral imaging technique coupled with deep learning was used to classify the film from seed cotton. The proposed algorithm integrates an improved weighted stacked autoencoder, the grey wolf optimizer and an extreme learning machine (ELM) to build the classification models for recognizing the seed cotton and film. Different from the classic autoencoder architecture, in the weighted stacked autoencoder, the self-encoded features were weighted based on their corresponding correlation with the network output [20]. Next, the advanced features from the weighted stacked autoencoder are used as the input for the ELM. ELM is a single-hidden layer feedforward ANN. Instead of using gradient descent algorithm, ELM utilizes the concept of random mapping and Moore–Penrose generalized inverse to optimize the network weight values. The established model can ensure not only the smallest training error but also better

generalization ability compared to the conventional gradient descent optimization algorithm. The corresponding training time of ELM is dramatically decreased [26]. However, as an ANN network, the number of hidden neurons of ELM and its first layer parameters from random mapping indeed affect the regression performance of the ELM model [27]. To select the best combination of these parameters, the grey wolf algorithm is applied into our application which shows better ANN optimization results compared to other classic metaheuristics in previous studies [28], [29].

Specifically, the integrated model is used as the final classifier to identify the film and cotton. Therefore, the specific objectives of current research include the following:

- Develop a sorting machine for the online detection of film from seed cotton based on hyperspectral imaging system and deep learning;
- Develop a fuzzy factor to adjust the weights based on the correlation coefficient of the inputs (original signal) and the outputs of the stacked autoencoder to extract the more representative features; and
- Use the grey wolf optimizer for the first time to determine the neurons and weights of the extreme learning machine to achieve higher classification accuracy.

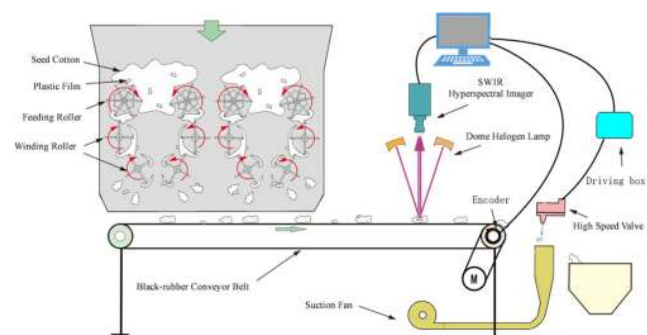


FIGURE 1. Schematic of hyperspectral imaging and sorting system.

II. MATERIALS AND METHODS

A. MATERIALS AND DATA COLLECTION

A Xinjiang municipality cotton ginning company provided approximately 10 kg of seed cotton that was mechanically harvested from the south of the Xinjiang municipality, China. Trained workers from the company picked out the film from the unginned cotton. At last, 49 pieces of various sized films were singled out. The mixture of films and seed cottons were fed into our sorting machine/system, and they were used to construct the experimental and testing datasets in this paper. The schematic of our machine is shown in Figure 1. The seed cotton is loaded from the top inlet of the feeding room. To improve the efficiency and accuracy of plastic film sorting, there are two kinds of rollers that are designed for seed cotton. The main function of the top rollers is load bearing and cotton feeding, and the shaft diameter is larger in order to achieve higher strength and stiffness. During feeding, the rotation speed is approximately 3-6 r/min with a slow but stable conveyance. The aim of the bottom rollers is to disperse the

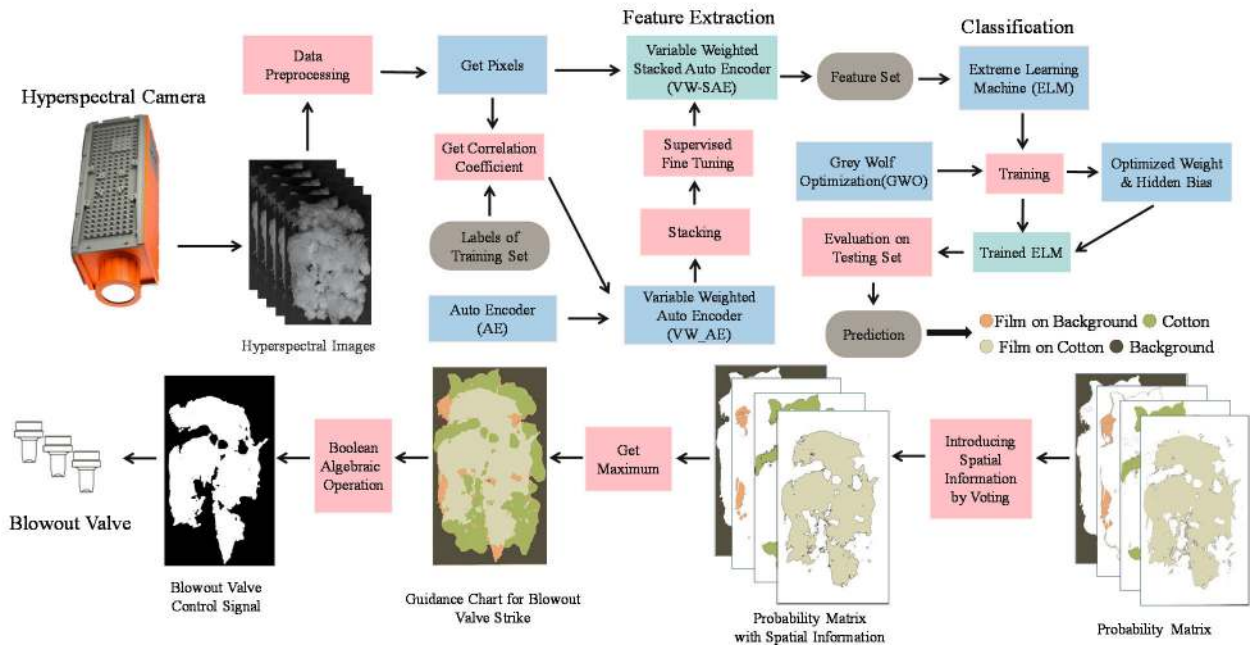


FIGURE 2. Structural chart of sorting algorithm for detecting plastic film from cotton.

clusters of entangled cotton, and their rotation speed is set as 200 r/min to achieve a better cleaning effect. Because the teeth of the adjacent winding rollers interact with the same ones of the top feeding rollers, the cotton is conveyed layer by layer. The teeth on the winding rollers can pull seed cotton from the top feeding rollers once the fiber is hooked during high-speed rotation. Because of the centrifugal force, the winding rollers will project cotton onto the black-rubber conveyor belt. In the same way, some mixed films and cotton are scattered by the impact of the teeth. Then, the seed cotton is separated and transported on the black-rubber conveyor belt. The width of the belt is approximately 2 m, and the material of the belt is black rubber to minimize the background reflected light. A servo-motor drives the conveyor belt at 2 m/s, and the encoder produces the speed pulse.

The hyperspectral imaging system is placed on the belt. The high-speed hyperspectral camera (Spectral Camera SWIR, SPECIM Spectral Imaging Ltd., Finland) was used to acquire the hyperspectral images. The spatial pixels of the camera number 384, and the spectral range is 1000 to 2500 nm with 288 spectral bands. A 15-mm lens that was designed for optimized performance from 900 to 2500 nm was utilized to achieve about a 5.2-nm pixel resolution, and the camera's field-of-view was approximately 2 m. The external illumination was equipped with two lines of dome halogen lamps to light the scene, and the dome halogen lamps can realize omnidirectional lighting to overcome the darker areas that may result in occlusion. Before practical use, a white reference plate was put on the belt to adjust the white balance and to fix the brightness value. The acquisition board on the computer was connected to the camera using a Camera Link cable. The board receives the encoder pulses

and sends a trigger signal to synchronize the frame rate of the camera with the speed of the belt. Normally, the frame rate is approximately 390 fps with 2 m/s as the belt speed.

The sorting system is located in the front of the belt. The high-speed valves with nozzles are arranged in a line, which are exactly aligned with the belt. There are 48 nozzles, and the width of each nozzle is approximately 41.6 mm. Under normal conditions, the seed cotton can fly through the inlet of the trash removal box into the storage box due to the effect of inertia. In contrast, the separated films will be absorbed into the trash removal box due to the combined action of big air friction and small inertia. Once the computer recognizes the films on the belt, it will count the number of pulses of the encoder for synchronization. When the films are under the valves, the computer will give a trigger signal to the corresponding valves to eject the films, and the films will be sucked in the trash removal box for vacuum aspiration. The online recognition algorithm is performed on an Nvidia GTX1060 GPU with 6 GB of DRAM.

B. SORTING ALGORITHM FOR FILM AND COTTON

The process of the proposed sorting algorithm is shown in Figure 2, and it includes three main parts: 1) the features are extracted by the variable-wise weighted stacked autoencoder, 2) detection by the extreme learning machine with the grey wolf optimizer, and 3) postprocessing.

1) VARIABLE-WISE WEIGHTED AUTOENCODER

The basic structure of an autoencoder is an unsupervised neural network with one hidden layer, and it consists of an input layer, a hidden layer and an output layer. The goal of the autoencoder is to reconstruct the original input (x_i) as

accurately as possible in the output layer (\tilde{x}_i). Here, we stack a group of autoencoders to construct a deep network [30] to reconstitute the inputs in order to extract the appropriate spectral features (Y) from hyperspectral images. However, in the spectrum analysis domain, it is well known that not all wavelength variables have the same importance for the output in the NIR spectra [31]. Some wavelength variables even have a negative influence on the regression result [32]. Wavelength selection can, to some degree, eliminate the negative influence from these wavelengths, but there still are some disadvantages. First, wavelength ranges with comparatively high noise might carry useful relevant information for prediction, and simply removing such wavelengths would spoil the multichannel advantage of the model to some extent. Second, some wavelength combinations might represent some data information, and the selection of an individual wavelength might cause the loss of useful information. Therefore, some research developed variable-wise weighted methods to assign continuous nonnegative values to wavelengths rather than directly eliminating unimportant wavelengths [33]. Thus, the method promises to preserve the useful information that is hidden among the noise, to retain the multichannel advantage and to reduce the influence of the negative features with small positive or zero weights.

In this paper, a variable-wise weighted stacked autoencoder [20] is adopted to extract the high-level features and to reduce the dimensions of the data. To extract the more representative features, a fuzzy factor, which is different from that in [20], is used to adjust the weights based on the correlation coefficient of the inputs (original signal, $x_{i(j)}$) and outputs (reconstructed signal, $\tilde{x}_{i(j)}$) of the weighted stacked autoencoder, which is named the VW-AE. Therefore, the target (loss function, J_λ) of the VW-AE can be expressed as follows:

$$J_\lambda(A, \tilde{A}, b, \tilde{b}) = \frac{1}{2N} \sum_{i=1}^N \sum_{j=1}^d \lambda_j (x_{i(j)} - \tilde{x}_{i(j)})^2 \quad (1)$$

where \tilde{A} is a $d \times d_h$ weight matrix, \tilde{b} is the bias vector for the output layer, and $\lambda_{(j)}$ is the weight of the j -th variable. $\lambda_{(j)}$ is set as follows:

$$\lambda_{(j)} = \frac{f(|CC_{(j)}|) \times |CC_{(j)}|}{|CC|_{max}} \quad (2)$$

where $f(|CC_{(j)}|)$ is a unipolar sigmoid function of $|CC_{(j)}|$, which is assigned a fuzzy weight to adjust the scale of $|CC_{(j)}|$. $|CC_{(j)}|$ is the correlation coefficient of the j -th variable, and it is calculated as follows:

$$|CC_{(j)}| = \frac{\sum_{i=1}^N (x_{i(j)} - \bar{x}_j)(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_{i(j)} - \bar{x}_j)^2 \sum_{i=1}^N (y_i - \bar{y})^2}} \quad (3)$$

$$|CC|_{min} = \text{Min}\{|CC_{(j)}|\} \quad j = 1, 2, \dots, d$$

$$|CC|_{max} = \text{Max}\{|CC_{(j)}|\} \quad j = 1, 2, \dots, d$$

where \bar{x}_j and \bar{y} are the means of the j -th variable and the output, respectively, and $Y = [y_1, y_2, \dots, y_N]^T$ is the output that is connected with the input X .

In equation (2), the variables that are highly related to the output are given large weights, and the fuzzy weight assigns the large $|CC_{(j)}|$ a corresponding larger weight and the small $|CC_{(j)}|$ a much smaller weight, which is different from the normalization in [20]. By training the VW-AE, the reconstruction should be more accurate for the output-relevant variables, and the hidden features are more relevant to the output.

2) EXTREME LEARNING MACHINE OPTIMIZED BY THE GREY WOLF OPTIMIZER

Compared with traditional classification algorithms, an ELM has the advantages of a strong generalization ability and fast learning speed [34]. Some scholars have applied ELMs to hyperspectral image classification and achieved better performance than other classification algorithms [35]. An ELM usually uses a single-layer feedforward network. Its basic structure includes an input layer, a hidden layer and an output layer, as shown in Figure 3.

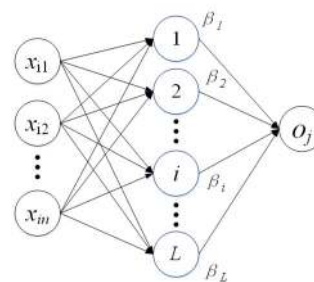


FIGURE 3. Structure of an ELM.

A single hidden layer neural network can be expressed as follows:

$$\sum_{i=1}^L \beta_i g(W_i \cdot X_j + b_i) = o_j, \quad j = 1, 2, \dots, N \quad (4)$$

where $X_i = [x_{i1}, x_{i2}, \dots, x_{im}]^T$ is the input for ELM, which is the extracted features from VW-SAE in our experiments, L is the number of hidden nodes, N is the number of training samples, $W_i = [w_{i1}, w_{i2}, \dots, w_{im}]^T$ is the weights of the inputs, β_i is the weight of the outputs, b_i is the bias of the i -th hidden node, $g(x)$ is an activation function, and o_i is the output of the ELM. In the experiment, the range of W_i is $[-1, 1]$, and the range of b_i is $[0, 1]$. The learning process achieves the lowest errors between the true values (or targets) t_i and outputs o_i , which is expressed as follows:

$$\sum_{j=1}^N ||o_j - t_j|| = 0 \quad (5)$$

In other words, it finds the parameters, including β_i , W_i and b_i , that make equation (5) tenable.

$$\sum_{i=1}^L \beta_i g(W_i \cdot X_j + b_i) = t_j, \quad j = 1, 2, \dots, N \quad (6)$$

Assuming H is the output of the hidden nodes, and β is the weight of the outputs, equation (6) can also be expressed as follows:

$$H\beta = T \tag{7}$$

where T is the expected output.

$$H(W_1, \dots, W_L, b_1, \dots, b_L, X_1, \dots, X_N) = \begin{bmatrix} g(W_1.X_1 + b_1) & \dots & g(W_L.X_1 + b_L) \\ \vdots & \ddots & \vdots \\ g(W_1.X_N + b_1) & \dots & g(W_L.X_N + b_L) \end{bmatrix}_{N \times L} \tag{8}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L \times m} \tag{9}$$

$$T = \begin{bmatrix} T_1^T \\ \vdots \\ T_N^T \end{bmatrix}_{N \times m} \tag{10}$$

Once the input weights W_i and bias b_i of the ELM are determined, the output matrix of the hidden nodes H is definite. In this way, the ELM model with one hidden layer can be transformed into a linear system $H\beta = T$, and the weights of outputs β can be calculated as follows:

$$\hat{\beta} = H^+T \tag{11}$$

where H^+ is the generalized inverse of matrix H , and T is the expected output.

There are two main steps of the GWO: 1) encircling prey and 2) hunting by imitating the grey wolf in nature. When encircling prey, the first three attacking wolves (GW_α , GW_β and GW_δ) can move to any place in order to guarantee they are around the target, which means that the three attacking wolves are those transversal vectors in P which make the corresponding ELMs achieve the first three best performance. Then, the rest of the other wolves update their positions according the best three. In other words, the other transversal vectors (rows) in P are updated and closer to the three selected wolves. The details of the GWO are introduced in reference [36]. Note that the neuron existence flag vector in [36] has been simplified into a numerical variable L in the array (P), because that the performance of ELM is only related to the number of hidden neurons, instead of its existence flag ordering in the binary coding. The best three wolves are selected based on the objective function introduced the next paragraph.

The biggest advantage of the GWO is that the maintained strategy of handling the exploration and exploitation in the search process ensures that the most appropriate parameters and number of hidden nodes of the ELM can be found at the same time with the objective function (F) as:

$$F = (1 - acc) + \gamma \frac{L - L_D}{L_U - L_D} \tag{12}$$

where acc is the whole accuracy of classification, L is the number of the hidden nodes, L_U is the upper limit of the number of the hidden nodes, L_D is the lower limit of the number of the hidden nodes, and γ is the parameter for adjusting accuracy and model complexity. γ , L_U and L_D are set as 0.76, 5, and 200 in our experiments. More specifically, the objective function consists of two parts: 1) the first term is the classification accuracy and 2) the second term means the complexity of the ELM. The fewer hidden nodes mean the simpler ELM. Therefore, the final target is to achieve the least F , which meets the higher accuracy and the fewer number of hidden nodes at the same time. The ELM that is optimized by the GWO is named the GWO-ELM in this paper.

3) POSTPROCESSING

The output of the GWO-ELM can generate four probability matrixes, which represent the classification probability for ‘Film on Background’, ‘Cotton’, ‘Film on Cotton’ and ‘Background’. Considering illumination variations, imaging noise, and dirt on the cotton, some pixels will be misclassified. In this postprocessing step, we will simply combine the spatial information to achieve better classification performance than the pixelwise method. The detailed steps are as follows. For the probability map of each class, a 5×5 uniform kernel will be used to convolute with the map. It based on the classification probability of each pixel being highly related to its neighboring probability information. Considering that no prior knowledge of cotton/film shapes was available in practice, in the application, a simple and fixed filter kernel

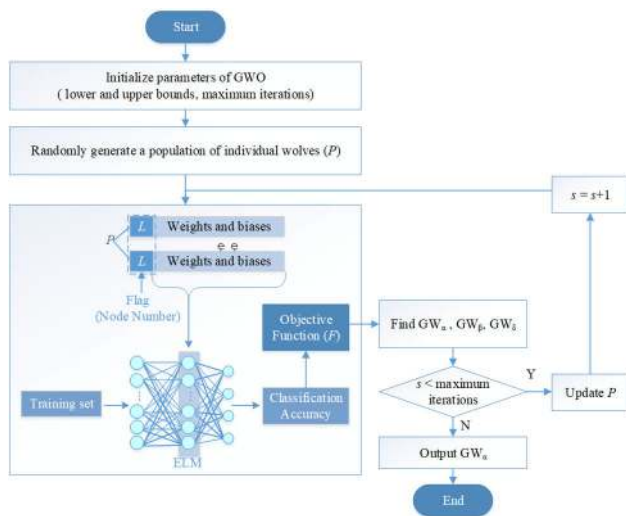


FIGURE 4. Flow chart of GWO optimization process.

Since both the number of hidden nodes and weights impact the entire performance of an ELM with one hidden layer, the grey wolf optimizer (GWO) is used to simultaneously optimize the number of hidden nodes and weights of the ELM in order to achieve higher classification accuracy. The overall workflow of GWO optimization process is shown in Figure 4.

was used. The four updated probability maps will be normalized to ensure that the sum of the classification probability of each pixel equals one. The pixel label will be determined as the class with the maximized normalized probability.

III. RESULTS AND DISCUSSION

A. DETERMINATION OF MODEL PARAMETERS AND STRUCTURE

To construct and test the proposed sorting model, 107 hyperspectral images with the size of $538 \times 384 \times 288$ (height \times width \times wavelength) were collected. 21 hyperspectral images were randomly selected to construct the training set and the remaining images were used as the testing set. All the pixels in each image were labeled as one of four categories: background, film on background, seed cotton and film on seed cotton. Each pixel in any hyperspectral image corresponds to 288-dimensional data of spectral information, which was set as a sample. Finally, 223872 samples were labeled, which meant that there were 55968 samples for each type in the training set, which were used to determine the parameters and structure of the proposed sorting model.

TABLE 1. Classification results of the GWO-ELM with different feature extraction methods.

Method	mRMR	VW-SAE
Background	0.9752	0.9944
Film on background	0.6989	0.8628
Cotton	0.8897	0.9763
Film on cotton	0.9818	0.9897
Accuracy	0.8864	0.9558

First, the weights and biases of the VW-SAE were randomly generated. The weights and biases of the VW-SAE in each layer were updated using layer-by-layer pretraining technology and the gradient descent algorithm based on the root mean square error loss function, and their parameters were determined by the 10-fold cross-validation method. After each layer of the VW-SAE’s training was completed, a two-layer neural network was used as an optimizer to fine-tune the VW-SAE as a whole. Through the experiments, we found that the proposed VW-SAE was composed of three VW-AEs with 144, 72 and 36 neurons, respectively. The sigmoid was used as the activation function in order to acquire the ideal spectral features for classification. Finally, 36 high-level features are extracted by the VW-SAE and set as inputs of the ELM. Table 1 lists the results of the GWO-ELM with the feature inputs extracted from hyperspectral image based on the VW-SAE and the traditional minimum redundancy maximum relevance algorithm (mRMR) [37]. For the mRMR, the highest classification accuracy is acquired when 10 features are extracted from 288 wavelengths, whose results are listed in the table. Detailed mRMR experiment results are listed in the supplementary material. In comparison, VW-SAV can extract more appropriate information than the

mRMR, especially for distinguishing film on background and cotton.

Here, the ELM adopted a single hidden layer neural network, and the activation function was the sigmoid in order to conduct nonlinear classification. The weights, biases and number of hidden nodes of the ELM were simultaneously optimized by the GWO algorithm. Recent studies show better optimization performance of GWO for training perceptrons [28]. In the experiment, the number of hidden layer neurons of the ELM was set to 14. A comparative experiment was carried out to further valid the optimization results, and Table 2 lists the results with different number of hidden nodes of the ELM. It is clear that the ELM with 14 hidden nodes can achieve the best classification results, especially the type of film on background.

TABLE 2. Classification results of the different number of hidden nodes of the ELM.

Number of hidden nodes	10	14	20
Background	0.9652	0.9944	0.9835
Film on background	0.7995	0.8628	0.8586
Cotton	0.9647	0.9763	0.9896
Film on cotton	0.9897	0.9897	0.9897
Accuracy	0.9298	0.9558	0.9554

To further prove effectiveness of GWO method to our specific dataset, we compared its results with two classical optimization algorithms, genetic algorithm (GA) and particle swarm optimization (PSO) algorithm. The final comparison results are shown in Table 3, and GWO performs the best among the three optimization algorithms.

TABLE 3. Comparison results for different optimization algorithms.

Optimization algorithm	Number of hidden nodes	Overall Accuracy
GWO	14	0.9558
GA	14	0.9421
PSO	10	0.9537

B. SPECTRAL ANALYSIS AND PROCESSING

Figure 5 shows the relative reflectance for the four types of objects (film on cotton, background, cotton and film on background), where the dotted line represent the actual spectral range of every type, and solid line represents its mean, respectively. The patterns for the relative mean reflectance for film on cotton and cotton and for the background and film on background were similar throughout the entire spectral region because of the weak reflectance of the transparent film. There is a noticeable difference between the relative mean reflectance for the cotton and background, which suggests that it may be easy to recognize cotton on the black belt.

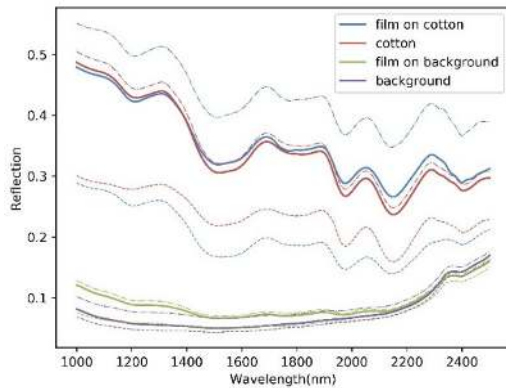


FIGURE 5. The relative reflectances for the four types of sample, where dotted lines represent the actual spectral ranges for four types, and solid lines represent their means, respectively.

However, it is clear that the spectral scale of film on cotton covers the whole spectral scale of cotton, and that there is overlap between the spectral scales of film on background and background in Figure 5. Therefore, it is hard to sort the film only with specific reflectance.

Since the spectra of the film on cotton and film on background hugely differ, the samples are first classified into four types in our work: film on cotton, film on background, cotton and background. Table 4 lists the classification results of the different models for the four types and the three types (film, cotton and background). Obviously, the results of the four types from the different models are better than those of the three types.

TABLE 4. Classification results of the different models for four types or three types.

Models	Four types	Three types
VW-SAE+GWO-ELM	0.9558	0.816
VW-SAE+ANN	0.9448	0.796
SAE+GWO-ELM	0.9310	0.769
SAE+ANN	0.9023	0.750
GWO-ELM	0.8098	0.646
ANN	0.8617	0.673

*Four types: background, film on background, cotton and film on cotton.
Three types: background, film and cotton.

C. ANALYSIS OF MODEL PERFORMANCE

Table 5 summarize the classification results for background, film on background, cotton and film on cotton using five combinations of mathematical models and a traditional machine learning model (ANN). Overall, both the grey wolf optimization extreme learning machine (GWO-ELM) and the artificial neural network (ANN) coupled with the combination of variable weighting and the stacked autoencoder (VW-SAE) can recognize the objects very well except for film on background that had a recognition rate of 0.8628 for the

VW-SAE +GWO-ELM and 0.8506 for the VW-SAE+ANN. When the variable weighting was not used, for either the GWO-ELM or the ANN models coupled with stacked autoencoders, each corresponding object obtained lower recognition rates excluding the background. Notably, the classification results for film on background achieved an almost 6 % reduction using the SAV+GWO-ELM model and the recognition rate achieved about 18 % reduction using the SAV+ANN model, which suggest that the variable weighting algorithms have significant effects for different models, but they have no influence on the background classification results. In addition, the identification of film on background achieved a classification accuracy of 0.4178 for the GWO-ELM model and the identification achieved a classification accuracy of 0.5794 for the ANN model. In other words, when the variable weighting coupled with a stacked autoencoder algorithm was used, the classification results for each of the objects could be improved over those of the GWO-ELM model and ANN model. Meanwhile, comparing to the ANN model, the variable-weighted stacked autoencoder algorithm has more effect on improving the classification accuracy of film on background for GWO-ELM model. In addition, Table 5 also shows that the single ANN model achieved the optimal recognition rate for film on background compared to other models. Overall, for all combinations of discrimination models, the recognition rate for background and film on cotton could reach approximately 0.99, while for the recognition of film on background, the variable-weighted stacked autoencoder algorithm could improve the classification results over those of the single GWO-ELM and ANN models.

To intuitively observe the recognition results, the established models in Table 5 are used to classify each pixel for a hyperspectral image with the size of $538 \times 384 \times 288$ (height \times width \times wavelength). Figure 6(a) showed that three pieces of film in the raw picture can be observed. Each model can recognize the background very well, which may be due to the large spectral differences. As shown in Figure 6(h) and 6(g), the ANN and GWO-ELM cannot well recognize the film on the background, and the GWO-ELM model also cannot ideally identify the film on cotton. After using the stacked autoencoder algorithm, the recognition rates for film on background and film on cotton both significantly improve, shown as Figure 6(e) and 6(f). Meanwhile, combining the VW-SAE with the GWO-ELM and ANN models can enhance the recognition performances for film on background, as shown in Figure 6(c) and 6(d). It can be observed that the generated pseudocolor maps have the same conclusion with Table 5, which indicates that the established models are reliable for online detection.

Table 6 presents the recognition accumulative time and overall classification accuracy for each model. The stacked autoencoder algorithm can improve the overall classification accuracy by 4.3 % for the ANN and by 11.2 % for the GWO-ELM, while the combination of variable weighting and the stacked autoencoder can further improve the classification results by 4.25% and 2.48 % for the SAE+ANN

TABLE 5. Classification results of different models for the four types: background, film on background, cotton and film on cotton.

Models		Background	Film on background	cotton	film on cotton
VW-SAE+GWO-ELM	Background	0.9944	0.0043	0.013	0.000
	film on background	0.0120	0.8628	0.0989	0.0264
	Cotton	0.0086	0.0006	0.9763	0.0146
	film on cotton	0.0002	0.0030	0.0071	0.9897
VW-SAE+ANN	Background	0.9915	0.0070	0.0015	0.000
	film on background	0.000	0.8506	0.0918	0.0576
	Cotton	0.0088	0.0008	0.9594	0.0310
	film on cotton	0.0000	0.0031	0.0192	0.9777
SAE+GWO-ELM	Background	0.9931	0.0036	0.0033	0.000
	film on background	0.0651	0.8071	0.0933	0.0345
	Cotton	0.0076	0.0006	0.9556	0.0362
	film on cotton	0.0000	0.0134	0.0184	0.9681
SAE+ANN	Background	0.9901	0.0082	0.0017	0.0000
	film on background	0.1140	0.6673	0.2082	0.0105
	Cotton	0.0006	0.0039	0.9535	0.0420
	film on cotton	0.0000	0.0004	0.0014	0.9982
GWO-ELM	Background	0.9876	0.0113	0.0012	0.000
	film on background	0.1129	0.4178	0.3209	0.1484
	cotton	0.0156	0.0157	0.8701	0.0986
	film on cotton	0.0000	0.0095	0.0268	0.9638
ANN	background	0.9838	0.0107	0.0055	0.000
	film on background	0.0404	0.5794	0.2833	0.0969
	cotton	0.0138	0.0102	0.8975	0.0786
	film on cotton	0.0000	0.0011	0.0126	0.9862

* VW: variable weighted, SAE: stacked autoencoder, GWO-ELM: grey wolf optimization extreme learning machine, and ANN: artificial neural network

and SAE+GWO-ELM, respectively. These results demonstrate that the algorithm that combines variable weighting and the stacked autoencoder provides a significant and positive improvement over both the ANN and GWO-ELM models. Although the classification accuracy improved using the combination of variable weighting and the stacked autoencoder, the process time increased by 45.5 % for the GWO-ELM with an accumulative recognition time of 1.44 s for an image with the size of $538 \times 384 \times 288$. However, compared with the ANN, the accumulative recognition time could decrease by 35.4 % for the combination of the variable weighting and the stacked autoencoder coupled with the GWO-ELM, which can be used for online detection.

Finally, the classification results of the VW-SAE+GWO-ELM are combined as film and nonfilm, as shown

in Figure 7(c). These results will trigger the valve to separate the film from cotton.

The study demonstrates that the variable-weighted stacked autoencoder algorithm coupled with the GWO-ELM can achieve better classification results with limited computational time increases, which could meet the online detection requirement. As shown in Figure 5, the spectra that were obtained from hyperspectral imaging had similar absorption peaks to other studies [38]. Nevertheless, it is slightly tough to visually identify cotton and film on cotton, as well as background and film on background, since the patterns of their spectra were similar, which encouraged us to develop a new algorithm to recognize these four objects. The stacked autoencoder algorithm can enhance the classification results for the four types of objects based on a single GWO-ELM

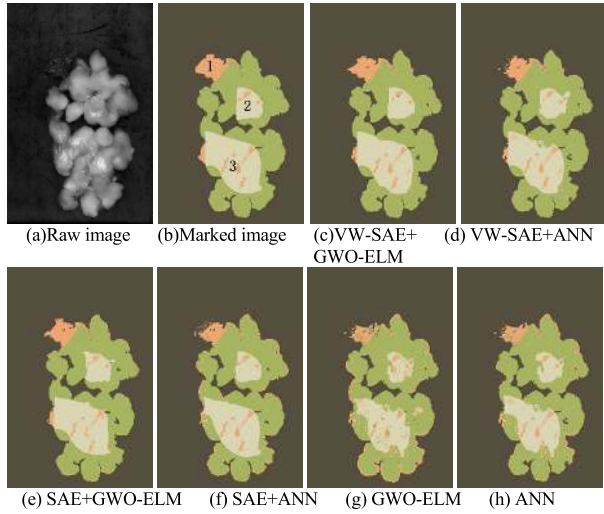
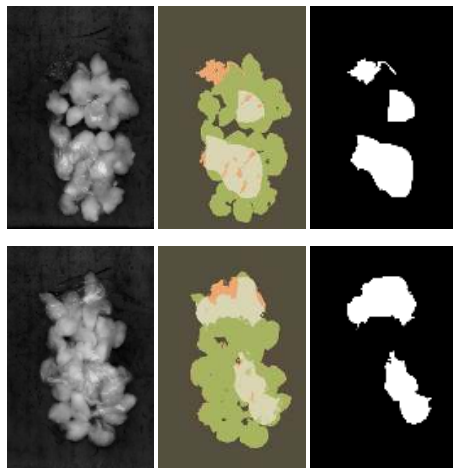


FIGURE 6. Classification results for each model based on pseudocolor maps.

TABLE 6. Accumulative recognition time and averaged recognition rates for each of the models with a 538 × 384 × 288 image.

Models	Recognition time for an image (s)	Overall classification accuracy (%)
VW-SAE+GWO-ELM	1.44	95.58
VW-SAE+ANN	2.23	94.48
SAE+GWO-ELM	1.44	93.10
SAE+ANN	2.23	90.23
GWO-ELM	0.99	80.98
ANN	1.37	86.17



(a)Raw images (b) Results of four types (c) Final results

FIGURE 7. Final outputs of the proposed sorting algorithm.

or ANN model. However, the variable-weighted stacked autoencoder algorithm could obviously improve the overall recognition results. The recognition rates for film on background for all of models were relatively low, which could be due to the influence of the strong absorption of the black background. Although the variable-weighted stacked autoencoder coupled with the GWO-ELM provided similar results to the variable-weighted stacked autoencoder

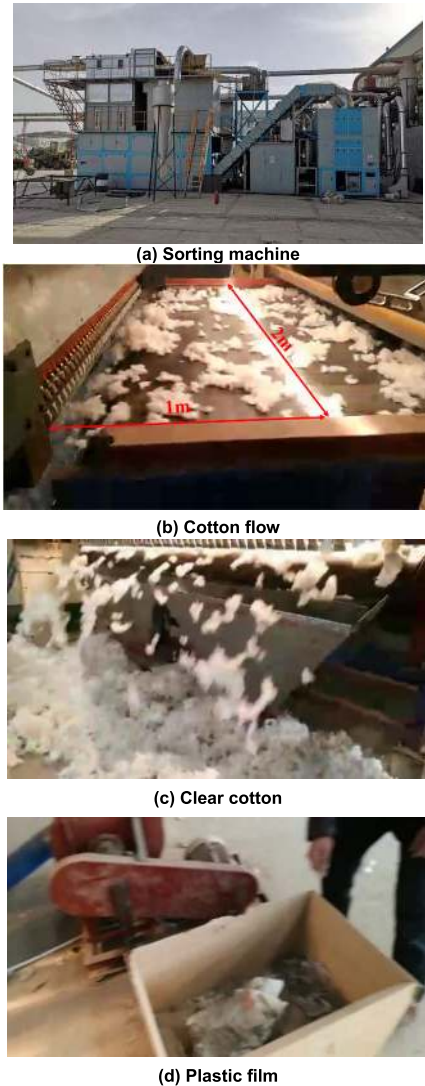


FIGURE 8. Sorting system and process for cotton.

coupled with the ANN, the recognition time for each pixel was faster, which could assess cotton more efficiently and save costs. Moreover, the average recognition rate for the variable-weighted stacked autoencoder coupled with the GWO-ELM model was also comparable to Mengyun Zhang’s [32] and Ruoyu Zhang’s [38] laboratory studies. These studies applied hyperspectral imaging in the transmittance and reflectance modes over the spectral range of 900 - 1700 nm to inspect foreign matter on the surface of cotton. However, the thin films in our studies are much hard to detect compared to the foreign matters, such as plant bract and leaves.

D. APPLICATION OF PROPOSED SORTING SYSTEM

The proposed method has been integrated into the sorting system with a Nvidia GTX1060 GPU and tested by two companies from Shandong Province, China and put into production in Xinjiang municipality, China. The algorithm is implemented by deep learning library Keras supported with Nvidia parallel computing platform CUDA,

and the GPU-accelerated library cuDNN. The mixture of films and seed cottons was fed into our proposed sorting machine/system (Figure 8(a)), and as shown in Figure 8, the machine can separate the films and seed cottons in real time. It should be noted that 384 pixels in a hyperspectral image is about 2m of the conveyor belt in Figure 8(b). In practice, considering the hyperspectral camera is the line scan camera, the position of its illumination light is removeable as shown in Figure 8(b). We can easily reduce the distance between the lights and the nozzles and then adjust accordingly the height of hyperspectral images to achieve better sorting result. After several field experimentations, the distance between the camera and the sorting nozzle is set as about 1m as shown in Figure 8(b). Figure 8(c) and (d) show the clear cotton and plastic film separated by the sorting system. In-field running results show that the proposed machine can process 3 tons of mixtures per hour, and sorting accuracy can achieve up to 95%.

IV. CONCLUSION

The study has developed a new sorting algorithm for the online detection of film on cotton using hyperspectral imaging over the spectral range of 1000 - 2500 nm. The results showed that the single ANN and GWO-ELM cannot recognize film on background, and using variable-weighted stacked autoencoder algorithms to extract features can provide positive effects for both the GWO-ELM and ANN by recognizing film on background at up to 86 % accuracy. The recognition rates for film on background are relatively low for all of the models, which may be due to the influence of the absorption of the dark background. The combination of variable weighting and the stacked autoencoder coupled with the GWO-ELM is determined to be the optimal model with an overall classification accuracy of 95.58 % for online detection due to the lower recognition time of 1.44 s per image. The proposed method has great potential to achieve online detection for the recognition of film on cotton.

On the other hand, the facility costs of this proposed sorting system is a little expensive, due to the integrated GPU for computation acceleration. With the future hardware developments, the equipment cost is expected to be further reduced. The proposed algorithm has a good application prospect and the similar idea can be applied into other fields, such as wheat and stalk separations.

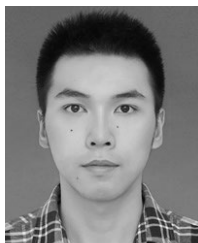
REFERENCES

- [1] Y. Meng, J. Song, Y. Lan, G. Mei, Z. Liang, and Y. Han, "Harvest aids efficacy applied by unmanned aerial vehicles on cotton crop," *Ind. Crops Products*, vol. 140, Nov. 2019, Art. no. 111645.
- [2] K. Z. Guo Shuxia, Z. Ruoyu, G. Wensong, A. Honglei, and C. Tengda, "Separation test of electrostatic separating device for machine-harvested seed cotton and plastic film residue," *Trans. Chin. Soc. Agricult. Eng.*, vol. 27, pp. 6–10, Dec. 2011.
- [3] H. Zhang and D. Li, "Applications of computer vision techniques to cotton foreign matter inspection: A review," *Comput. Electron. Agricult.*, vol. 109, pp. 59–70, Nov. 2014.
- [4] Q. Zhang, L. Hua, and J. Yao, "Development of detection system for cotton plastic covering using ultrasonic," *Appl. Mech. Mater.*, vols. 411–414, pp. 1439–1444, Sep. 2013.
- [5] L. Chang, "The detecting system of foreign fibers in cotton based on DSP," M.S. thesis, Hefei Univ. Technol., Hefei, China, 2006.
- [6] M. P. Millman, M. Acar, and M. R. Jackson, "Computer vision for textured yarn interlace (nip) measurements at high speeds," *Mechatronics*, vol. 11, no. 8, pp. 1025–1038, Dec. 2001.
- [7] A. Abouelela, H. M. Abbas, H. Eldeeb, A. A. Wahdan, and S. M. Nassar, "Automated vision system for localizing structural defects in textile fabrics," *Pattern Recognit. Lett.*, vol. 26, no. 10, pp. 1435–1443, Jul. 2005.
- [8] M. A. Lieberman, C. K. Bragg, and S. N. Brennan, "Determining gravimetric bark content in cotton with machine vision," *Textile Res. J.*, vol. 68, no. 2, pp. 94–104, Feb. 1998.
- [9] L. Zhang, M. Levesley, A. Dehghani, and T. King, "Integration of sorting system for contaminant removal from wool using a second computer," *Comput. Ind.*, vol. 56, nos. 8–9, pp. 843–853, Dec. 2005.
- [10] L. Zhang, A. Dehghani, Z. Su, T. King, B. Greenwood, and M. Levesley, "Development of a mechatronic sorting system for removing contaminants from wool," *IEEE/ASME Trans. Mechatronics*, vol. 10, no. 3, pp. 297–304, Jun. 2005.
- [11] L. Zhang, A. Dehghani, Z. Su, T. King, B. Greenwood, and M. Levesley, "Real-time automated visual inspection system for contaminant removal from wool," *Real-Time Imag.*, vol. 11, no. 4, pp. 257–269, Aug. 2005.
- [12] R. Lu and Y. R. Chen, "Hyperspectral imaging for safety inspection of food and agricultural products," *Proc. SPIE*, vol. 3544, pp. 121–133, Jan. 1999.
- [13] J. Guo, Y. Ying, J. Li, X. Rao, Y. Kang, and Z. Shi, "Detection of foreign materials on surface of ginned cotton by hyper-spectral imaging," *Trans. Chin. Soc. Agricult. Eng.*, vol. 28, pp. 126–134, Jan. 2012.
- [14] J. Guo, "Detection of trashes in combed cotton using hyper-spectral images," *Trans. Chin. Soc. Agricult. Machinery*, vol. 43, no. 12, pp. 197–203, 2012.
- [15] C. A. Fortier, J. E. Rodgers, M. S. Cintrón, X. Cui, and J. A. Foulk, "Identification of cotton and cotton trash components by Fourier transform near-infrared spectroscopy," *Textile Res. J.*, vol. 81, no. 3, pp. 230–238, Feb. 2011.
- [16] M. Ma, S. Khatibisepehr, and B. Huang, "A Bayesian framework for real-time identification of locally weighted partial least squares," *AIChe J.*, vol. 61, no. 2, pp. 518–529, Feb. 2015.
- [17] K. J. Preacher, P. J. Curran, and D. J. Bauer, "Computational tools for probing interactions in multiple linear regression, multilevel modeling, and latent curve analysis," *J. Educ. Behav. Statist.*, vol. 31, no. 4, pp. 437–448, Dec. 2006.
- [18] H. Kaneko and K. Funatsu, "Application of online support vector regression for soft sensors," *AIChe J.*, vol. 60, no. 2, pp. 600–612, Feb. 2014.
- [19] L. Fortuna, P. Giannone, S. Graziani, and M. G. Xibilia, "Virtual instruments based on stacked neural networks to improve product quality monitoring in a refinery," in *Proc. IEEE Instrum. Meas. Technol. Conf.*, vol. 56, May 2005, pp. 1889–1893.
- [20] X. Yuan, B. Huang, Y. Wang, C. Yang, and W. Gui, "Deep learning-based feature representation and its application for soft sensor modeling with variable-wise weighted SAE," *IEEE Trans. Ind. Informat.*, vol. 14, no. 7, pp. 3235–3243, Jul. 2018.
- [21] X. Wang, Y. Wang, Z. Wang, C. Wang, and Y. Li, "Hand gesture recognition using sparse autoencoder-based deep neural network based on electromyography measurements," in *Proc. Nano-, Bio-, Info-Tech Sensors, 3D Syst. II*, Mar. 2018, Art. no. 105971.
- [22] T. Akilan, Q. M. J. Wu, A. Safaei, and W. Jiang, "A late fusion approach for harnessing multi-CNN model high-level features," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2017, pp. 566–571.
- [23] A. Wang, Y. Wang, and Y. Chen, "Hyperspectral image classification based on convolutional neural network and random forest," *Remote Sens. Lett.*, vol. 10, no. 11, pp. 1086–1094, Nov. 2019.
- [24] P. Nie, J. Zhang, X. Feng, C. Yu, and Y. He, "Classification of hybrid seeds using near-infrared hyperspectral imaging technology combined with deep learning," *Sens. Actuators B, Chem.*, vol. 296, Oct. 2019, Art. no. 126630.
- [25] Zhu, Zhou, Gao, Bao, He, and Feng, "Near-infrared hyperspectral imaging combined with deep learning to identify cotton seed varieties," *Molecules*, vol. 24, no. 18, p. 3268, 2019.
- [26] W. Zheng, X. Fu, and Y. Ying, "Spectroscopy-based food classification with extreme learning machine," *Chemometric Intell. Lab. Syst.*, vol. 139, pp. 42–47, Dec. 2014.
- [27] 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.

- [28] S. Mirjalili, "How effective is the grey wolf optimizer in training multi-layer perceptrons," *Appl. Intell.*, vol. 43, no. 1, pp. 150–161, Jul. 2015.
- [29] Q. Al-Tashi, H. M. Rais, S. J. Abdulkadir, S. Mirjalili, and H. Alhussian, *A Review of Grey Wolf Optimizer-Based Feature Selection Methods for Classification/Evolutionary Machine Learning Techniques*. Singapore: Springer, 2020, pp. 273–286.
- [30] Y. Lecun, Y. Bengio, and G. E. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [31] J. Foulk, C. Fortier, J. Rodgers, and D. Whitelock, "Near-infrared classification of cotton lint, botanical and field trash," *J. Cotton Sci.*, vol. 16, no. 1, pp. 72–79, 2012.
- [32] M. Zhang, C. Li, and F. Yang, "Classification of foreign matter embedded inside cotton lint using short wave infrared (SWIR) hyperspectral transmittance imaging," *Comput. Electron. Agricult.*, vol. 139, pp. 75–90, Jun. 2017.
- [33] L. Xu, J.-H. Jiang, H.-L. Wu, G.-L. Shen, and R.-Q. Yu, "Variable-weighted PLS," *Chemometric Intell. Lab. Syst.*, vol. 85, no. 1, pp. 140–143, Jan. 2007.
- [34] H. Al-Behadili, A. Grumpe, C. Dopp, and C. Wohler, "Extreme learning machine based novelty detection for incremental semi-supervised learning," in *Proc. 3rd Int. Conf. Image Inf. Process. (ICIIP)*, Dec. 2015, pp. 230–235.
- [35] U. Ergul and G. Bilgin, "HCKBoost: Hybridized composite kernel boosting with extreme learning machines for hyperspectral image classification," *Neurocomputing*, vol. 334, pp. 100–113, Mar. 2019.
- [36] H. Faris, S. Mirjalili, and I. Aljarah, "Automatic selection of hidden neurons and weights in neural networks using grey wolf optimizer based on a hybrid encoding scheme," *Int. J. Mach. Learn. Cybern.*, vol. 10, no. 10, pp. 2901–2920, 2019.
- [37] Y.-F. Gao, B.-Q. Li, Y.-D. Cai, K.-Y. Feng, Z.-D. Li, and Y. Jiang, "Prediction of active sites of enzymes by maximum relevance minimum redundancy (mRMR) feature selection," *Mol. BioSyst.*, vol. 9, no. 1, pp. 61–69, 2013.
- [38] R. Zhang, C. Li, M. Zhang, and J. Rodgers, "Shortwave infrared hyperspectral reflectance imaging for cotton foreign matter classification," *Comput. Electron. Agricult.*, vol. 127, pp. 260–270, Sep. 2016.



CHAO NI was born in Nanjing, Jiangsu, China, in 1979. He received the B.S. degree in automation from the Nanjing University of Science and Technology, Nanjing, in 2001, and the Ph.D. degree in control theory and control engineering from Southeast University, Nanjing, in 2008. He is currently an Associate Professor with Automation Department, Nanjing Forestry University, China. From October 2017 to November 2018, he was a Visiting Scholar with the University of Maryland, College Park, USA. His research interests include artificial intelligence in industrial application, data processing, and spectroscopy analysis.



ZHENYE LI was born in Yangzhou, Jiangsu, China, in 1997. He received the bachelor's degree in engineering from Nanjing Forestry University, in 2019, where he is currently pursuing the master's degree. His research interests include the application of intelligent algorithms in industry, image data processing, and hyperspectral analysis.



XIONG ZHANG was born in Yangzhou, Jiangsu, China, in 1995. He received the B.S. degree in automation from the College of Mechanical and Electronic Engineering, Nanjing Forestry University, Nanjing, in 2017, where he is currently pursuing the M.S. degree in control theory and control engineering with the College of Mechanical and Electronic Engineering. His research interests include artificial intelligence and image processing.



XINYAN SUN was born in Maanshan, Anhui, China, in 1996. He received the B.S. degree in automation from Nanjing Forestry University, Nanjing, China, in 2019, where he is currently pursuing the master's degree in control science and engineering. His research interests include artificial intelligence in data processing and image processing.



YUPING HUANG received the B.S. and Ph.D. degrees in agricultural engineering from Nanjing Agricultural University, Nanjing, China, in 2011 and 2018, respectively. She had spent two years and seven months as a Visiting Student for her Ph.D. dissertation research at Michigan State University, USA, during which she developed a novel spatially resolved spectroscopy (SRS) systems and a noncontact semi-transmittance multichannel spectroscopic systems for quality evaluation of horticultural products.

Since 2018, she has been a Lecturer with the College of Mechanical and Electronic Engineering, Nanjing Forestry University. Her research was focused on nondestructive quality evaluation of agricultural products using spectroscopy and image analysis.



LING ZHAO was born in Jinan, Shandong, China, in 1980. He received the B.S. degree in agricultural mechanization and automation from Shandong Agricultural University, Tai'an, in 2002, and the Ph.D. degree in machinery manufacturing and automation from Beihang University, Beijing, China, in 2009. He is currently an Associate Professor with the School of Mechanical and Automotive Engineering, Liaocheng University, China. His research interests include mechanical design, structural optimization, and CFD/FEM.



TINGTING ZHU received the Ph.D. degree in pattern recognition and artificial intelligence from the School of Automation, Southeast University, in 2019. She is working with the College of Mechanical and Electronic Engineering, Nanjing Forestry University, China. She achieved a fellowship jointly awarded by the Fonds de Recherche du Québec—Nature et Technologies (FRQNT) and the China Scholarship Council and studied as a Visiting Student at the Department of Atmospheric and Oceanic Sciences, McGill University, Canada, from 2017 to 2018. Her current research interests include machine learning, data processing and modeling, renewable energy generation forecast, and climate feedback.



DONGYI WANG received the B.S. degree from Electrical Engineering Department, Fudan University, China. He is currently pursuing the Ph.D. degree with Bioengineering Department, University of Maryland, College Park, USA. He was a Visiting Student at Electrical Engineering Department, Chinese University of Hong Kong, in 2014. His research interests include computer vision, artificial intelligence, hyperspectral imaging analysis, food, and healthcare automation.

...