



Non-destructive detection and classification of textile fibres based on hyperspectral imaging and 1D-CNN

Jiadong Huang, Hongyuan He^{*}, Rulin Lv, Guangteng Zhang, Zongxian Zhou, Xiaobin Wang

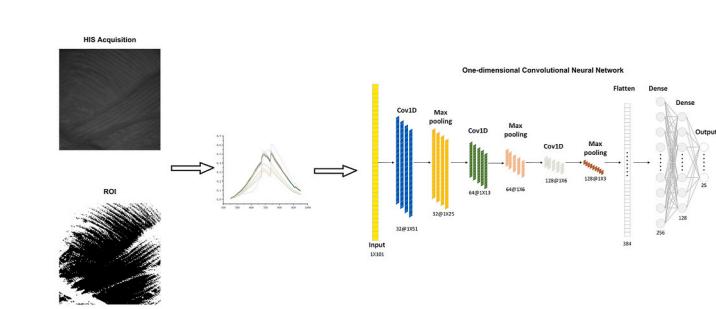
School of Criminal Investigation, People's Public Security University of China, Beijing, China



HIGHLIGHTS

- Hyperspectral technique was proposed to classify and identify textile fibers.
- The spectra were preprocessed by pixel fusion method.
- One-dimensional convolutional neural network model was built to classify hyperspectral data.
- Based on hyperspectral data, the traditional machine learning method and neural network method are compared.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Hyperspectral imaging
Textile fibres
One-dimensional convolutional neural network
Classification

ABSTRACT

Textile fibre is very common in daily life, and its classification and identification play an important role in textile recycling, archaeology, public security, and other industries. However, traditional identification methods are time-consuming, laborious, and often destructive to the samples. In order to quickly, accurately, and nondestructively classify and recognize textile fibres, this study established a textile fibre classification and recognition method based on hyperspectral imaging (HSI) and a one-dimensional convolutional neural network (1D-CNN) model. Hyperspectral images of 25 kinds of commercial textile fibres were collected and denoised by pixel fusion. Four traditional machine learning classification models, k-nearest neighbors (KNN), support vector machine (SVM), random forest (RF), and partial least squares-discriminant analysis (PLS-DA), were used to identify the data. The results show that RF has the highest classification accuracy, reaching 91.4%. Then a back propagation neural network (BPNN) model and a one-dimensional convolutional neural network (1D-CNN) model were constructed and compared with the traditional machine learning methods. The results show that the 1D-CNN models have 97.9% and 98.6% accuracy on the training and test sets, respectively. The precision (Pr), sensitivity (Se), specificity (Sp), and F1 score (F1 score) of the models reached 98.7%, 98.6%, 99.9%, and 98.6%, respectively, which were significantly better than the four traditional machine learning models. It seems that 1D-CNN combined with the HSI technique may be a potential method in the detection and classification of textile fibres.

* Corresponding author. School of Criminal Investigation, People's Public Security University of China, Xicheng Muxidi South 1st, Beijing, 100038, China.
E-mail address: 13311296819@189.cn (H. He).

1. Introduction

Textile fibres are the raw materials for textile products, which can be divided into natural fibres and synthetic fibres [1]. Natural fibres include plant fibres and animal fibres, such as cotton and wool. Synthetic fibre is a silk thread formed by squeezing raw fibre material out of air and water through a spinneret hole, such as lyocell, polyethene terephthalate (PET), and polypropylene (PP) [2]. Textile fibre species identification plays a vital role in textile recycling [3], archaeology [4], public security [5], and other industries. Textile fibres may become necessary material evidence at crime scenes and play a key role in detecting old cases [6]. Tracing the origin of textile fibre not only helps to solve the case but also provides strong evidence for the lawsuit. Non-destructive and efficient identification of textile fibres in waste fabrics is also significant in reusing textile materials [7]. Therefore, it is of great practical importance to develop a fast and accurate textile fibre identification method.

Traditional textile fibres detection methods include visual inspection, microscopic observation, and combustion. With the continuous improvement and updating of detection technology, the current detection methods for textile fibres are chromatography [8–10] and spectral methods. Chromatography is destructive to textile fibre samples and cannot maintain the samples' integrity. Spectral methods is a non-destructive, rapid detection method. Raman spectroscopy [11], terahertz time-domain spectroscopy [12], Infrared spectroscopy [13–15], and other spectral detection techniques have been applied to the identification of textile fibres. Jin et al. [11] combined Raman spectroscopy and a principal component analysis and support vector machine (PCA-SVM) model to identify different graphene oxide reinforced textiles. Sun et al. [13] used near-infrared spectroscopy(NRS) combined with a new identification method of recently regularized subspace, which was successfully applied to identify cashmere textiles from adulterated wool textiles; Peets et al. [15] used Fourier transform infrared (FT-IR) spectroscopy with attenuated total reflection (ATR) to identify different types of textile fibres and mixtures. However, using an ATR-FT-IR spectrometer, the spectral quality is poor, absorption intensity is low, baseline noise is high, and very simple mathematical processing (PCA) was used in the paper. Therefore, it is impossible to distinguish textile fibres whose components are very similar [3].

Hyperspectral imaging (HSI) is characterized by multiple bands and high spectral resolution. HSI has been widely used in food detection and agriculture [16–18], such as food classification [19], ingredient detection [20], pesticide content detection [21]. In the field of criminal technology, HSI has been applied to the identification of physical evidence such as bloodstains [22], explosive residues [23], semen stains [24], and stamp-pad ink [25]. However, no relevant study has classified textile fibres through hyperspectral. In the field of hyperspectral image classification, hyperspectral image classification is usually achieved by machine learning methods, such as SVM, RF, and other traditional machine learning methods. However, traditional machine learning methods have low computational efficiency and accuracy for hyperspectral images with large data volumes. In recent years, deep learning methods represented by convolutional neural networks (CNN) have high accuracy and generalization [26]. CNN model has been well applied in image recognition, hyperspectral image recognition, spectral analysis [27,28].

Therefore, in order to carry out a more scientific, effective, accurate, and reasonable identification study of textile fibre evidence, this paper collected hyperspectral images of 25 textile fibres and constructed a one-dimensional convolutional neural network (1D-NN) model to classify and identify the hyperspectral data. Moreover, the 1D-CNN model was compared with the BPNN model and four traditional machine learning models.

Table 1
Classification of textile fibre samples.

Plant Fibre	ID ^a	Animal Fibre	ID	Synthetic Fibre	ID
cotton	7	wool	2	acrylic fibres	1, 4, 16
long-staple cotton	13	camel hair	3, 18, 19	bamboo carbon fibre	5, 6, 8, 12, 14, 15
tencel linen	10	milk cotton	9, 11, 17, 23	silk	21
soybean fibre	22			cotton	
bamboo fibre	20, 24, 25				

^a Classification label of the sample.

2. Materials and methods

2.1. Experimental samples

In this study, 25 common commercially available textile fibres of different compositions were purchased, and the samples were classified as shown in Table 1. Detailed information on the samples can be found in Table S1.

2.2. Hyperspectral image acquisition and correction

A staring-type hyperspectral imaging system(Wayho Technology Co.) was used for hyperspectral image acquisition of 25 textile fibre samples. The spectral resolution of the system is 1 nm. In order to obtain clear and undistorted images, the textile fibres were placed in the centre of the instrument platform, the camera was focused on the samples from directly above, and there were two 50 W halogen light sources on the left and right sides of the camera, and the irradiation angle was 45° from the camera. The spectral range of the hyperspectral camera was set as 450–950 nm, and the acquisition interval was 5 nm. There were 101 bands in total, and the camera resolution was 2048 × 2046 pixels.

To eliminate the noise caused by the varying intensity of the light source, the image acquired by hyperspectral was corrected in dark and white. A dark reference image (I_D) was obtained by turning off the light source and covering the camera lens entirely with an opaque cover. The white reference image (I_W) was obtained using a white PTFE tile with a reflectivity close to 100%. Then, the original hyperspectral image (I_{sample}), the white reference image (I_W) and the dark reference image (I_D) are brought into formula (1) to calculate the corrected image (I):

$$I = \frac{I_{sample} - I_D}{I_W - I_D} \quad (1)$$

2.3. Spectral data extraction

In order to extract the spectral data, the region of interest (ROI) was defined. Fig. 1(a) showed that the whole sample region in the hyperspectral image was defined as ROI for a whole fibre sample. Fig. 1(b) showed that the whole sample region of a thread-like textile fibre sample was defined as ROI for a thread-like textile fibre sample.

Pixel-level spectra contained significant noise, and this study used a pixel fusion method [25]which averaged 10 adjacent pixels in the ROI to reduce noise. The pixel fusion method is less computationally intensive and allows greater independence between pixels. The spectrum of each pixel on the ROI after pixel average was taken as the data set, and the spectral data set was marked. Spectral data and labels constituted instances of data analysis, and a total of 600,404 instances were obtained, with an average ROI of 25,016 per sample. 70% of the dataset was randomly used as the training set, which had 420,282 instances; the Validation set and test set were respectively 15%, and both contained 90061 instances. For all the models, we trained them 5 times, took the

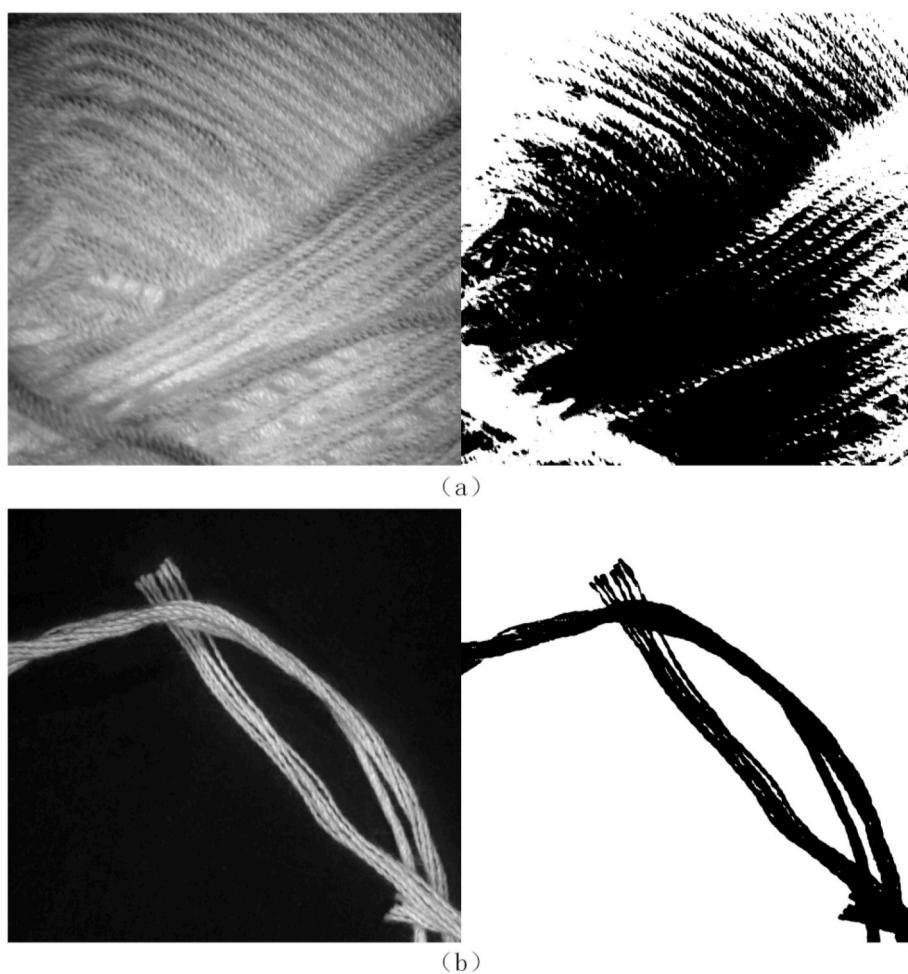


Fig. 1. Hyperspectral image ROI area (a) whole textile fibres (b) thread-like textile fibres.

average accuracy of the 5 training sessions as the final accuracy, and ensured that each training was randomly divided into the training set, validation set, and test set. In the model training phase, the training set was used for network training, and the validation set was used to accelerate network convergence. Eventually, the model was trained and evaluated by the training and test sets.

2.4. Classification model

2.4.1. Traditional machine learning methods

The K nearest neighbor (KNN) algorithm is one of the simplest and most widely used classification methods [29]. The method was first developed by Cover and Hart in 1968. It predicts a sample by the majority voting principle for K samples close to that. Due to its simplicity and effectiveness, the method has been widely used in text classification,

image recognition, and spectral analysis. The number of neighbors (K) is an important parameter that is crucial to the performance of the KNN model [30]. The advantages of the KNN algorithm are that it is easy to understand, adjust the model, and performs well in solving multi-classification problems [31].

Support Vector Machine (SVM) is a machine learning algorithm based on statistical learning theory [32], which is widely cited for classification problems. SVM algorithm has good generalization performance and is not easy to produce over-fitting, so it has obvious advantages in solving practical problems [33]. It is a supervised learning algorithm based on VC dimensional theory and structural risk minimization. It is a supervised learning algorithm based on VC dimensional theory and structural risk minimization and has an excellent theoretical foundation. SVM algorithm finds the optimal hyperplane in the data space by training the samples to form support vectors to distinguish different classes of samples as much as possible [34].

Random Forest (RF) is a decision tree-based classification learner, where each tree is constructed using different bootstrap samples. At the same time, a subset of explanatory variables is randomly selected at each node, and then the classification result is decided according to the number of votes per tree [35]. RF models are insensitive to noise and can handle many input variables (high-dimensional) [36]. RF has shown high accuracy in classification problems, such as origin identification [37] and variety identification [38].

Partial least squares discriminant analysis (PLS-DA) is a supervised classification algorithm based on partial least squares regression (PLSR) [39]. PLS-DA is the method of choice for classification. It is a robust algorithm for predicting and describing models [40]. PLS-DA is a

Table 2
Classification model parameters table.

Methods	Parameters
KNN	algorithm = 'auto', n_neighbors = 3;
SVM	kernel = 'linear', C = 1.0;
RF	n_estimators = 300, max_depth = None, max_features = 'auto';
PLS-DA	n_components = 60, scale = True, copy = True;
BPNN	optimizer = Adaptive Moment Estimation (Adam), loss = categorical_crossentropy, metrics = ['accuracy'], batch_size = 256, epochs = 100;
CNN	optimizer = Adaptive Moment Estimation (Adam), learning rate = 0.001, loss = categorical_crossentropy, metrics = ['accuracy'], batch_size = 128, epochs = 100;

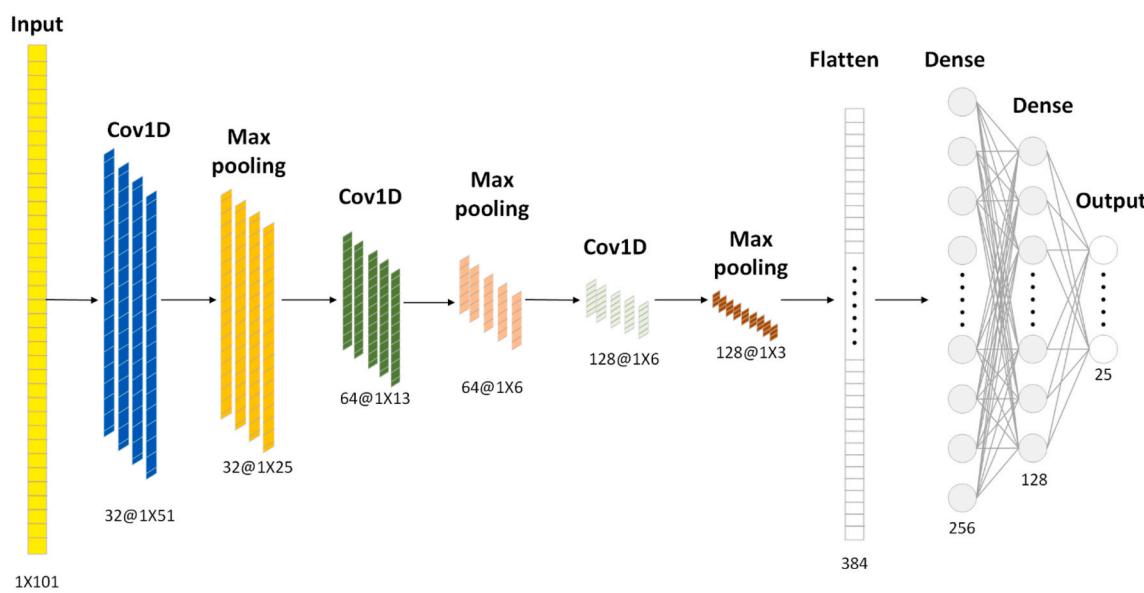


Fig. 2. Structure and related parameters of 1D-CNN. The model consists of 3 convolution layers, 3 pooling layers, 1 flatten layer and 2 dense layers, with the name of each layer at the top, the size of the output data, and the size of the convolution kernel at the bottom.

supervised chemometric method that is exploratory and quantitative.

Therefore, this paper used four common machine learning methods, KNN, SVM, RF, and PLS-DA, to analyze the hyperspectral data of fibre samples.

A grid search algorithm was used for hyperparameter optimization of the machine learning model. The KNN model with 1, 3, and 5 neighbors was searched. SVM was used with a linear kernel function, and penalty term C was searched for four values of 0.01, 0.1, 1, and 10. RF did not set the maximum depth, and the number of decision trees was searched between 100, 200, 300, and 400. The number of components of the PLS-DA classifier was searched from 1 to 60. The specific parameters of the finalized KNN, SVM, RF, and PLS-DA models and their values are shown in Table 2.

2.4.2. Neural networks

In this paper, we focused on the classification and recognition of textile fibre hyperspectral data by 1D-CNN and compared it with BPNN. Back propagation neural network (BPNN) is a widely known neural network in machine learning, with data onward transmission and error feedback propagation [41]. BPNN is a neural network algorithm that continuously reduces the error by optimizing the parameters using a gradient descent algorithm through the error between the computed value and the actual value. As the gradient descent algorithm is adopted, the optimization of parameters almost stops when BPNN encounters some flat gradient regions. This makes the model converge slowly and inefficiently. Table 2 showed the parameters of the BPNN model and the 1D-CNN model.

Convolutional neural network (CNN) is an important deep learning method that automatically learns features from data and generalizes the results to anonymous data of the same type. The method is widely used in image classification, natural language processing, and spectral recognition. CNN is structurally similar to BPNN, but CNN has unique hidden layers: convolutional layer and pooling layer. The convolutional layer is used for automatic feature extraction; the pooling layer reduces the number of parameters and retains the main features; the fully-connected layer can integrate and distinguish features.

The structure of the 1D-CNN of this paper is shown in Fig. 2. The CNN model extracted features from the preprocessed hyperspectral data by convolution and pooling operations. This study's proposed 1D-CNN model contained one input layer, three one-dimensional convolutional layers (Cov1D), three max-pooling layers, one flatten layer, two dense

layers, and one output layer.

In this model, the input layer received an HSI vector of size 101 and then passed through three convolutional layers and three pooling layers. In the first convolution operation, the size of the convolution kernel was 3, the stride was 2, and the number of convolution kernels was 32; the size of the second convolution kernel was 3, the stride was 2, and the number of convolution kernels was 64; the size of the third convolution kernel was 3, the stride was 1, and the number of convolution kernels was 128. The padding of each convolution layer was set to 1. Padding was used to add an appropriate width and height to each side of the input feature to ensure that the output feature has the same size as the input. The activation function of the convolution layer usually uses the rectifying linear unit function (ReLU function). However, The ReLU function ignores the features of negative signals [42]. This paper normalised the spectral data and conformed to the standard normal distribution. Therefore, the tanh function was used as the activation function of the convolution layer. Moreover, the tanh function takes values between -1-1, which is more commonly used than the sigmoid function. The tanh activation function was formulated as follows.

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

After each convolutional layer, the maxi-pooling region was selected through the maxi-pooling layer to reduce the number of parameters, reduce the calculation cost, prevent overfitting, and improve the overall performance and accuracy of the network. The pool size of the maxi-pooling layer was 2, and the stride was 2.

Finally, the output of the pooling layer (multidimensional data) was converted into a one-dimensional vector through the flatten layer and transmitted to the fully connected layer. 2 fully connected layers had 256 and 128 neurons, respectively. The output layer used a softmax activation function to provide predictions of textile fibre categories. 25 neurons in this layer produced 25 outputs, representing 25 categories, respectively.

In this paper, the loss function of the 1D-CNN model was cross-entropy [43]. The cross-entropy function found the error between the predicted and actual values. Then the error was returned layer by layer, and the weights were updated using the Adam optimizer with an initial learning rate of 0.001 [44]. The updated parameters continued to repeat the above process until the loss function value was minimized and the network was completed. The cross-entropy function can be defined

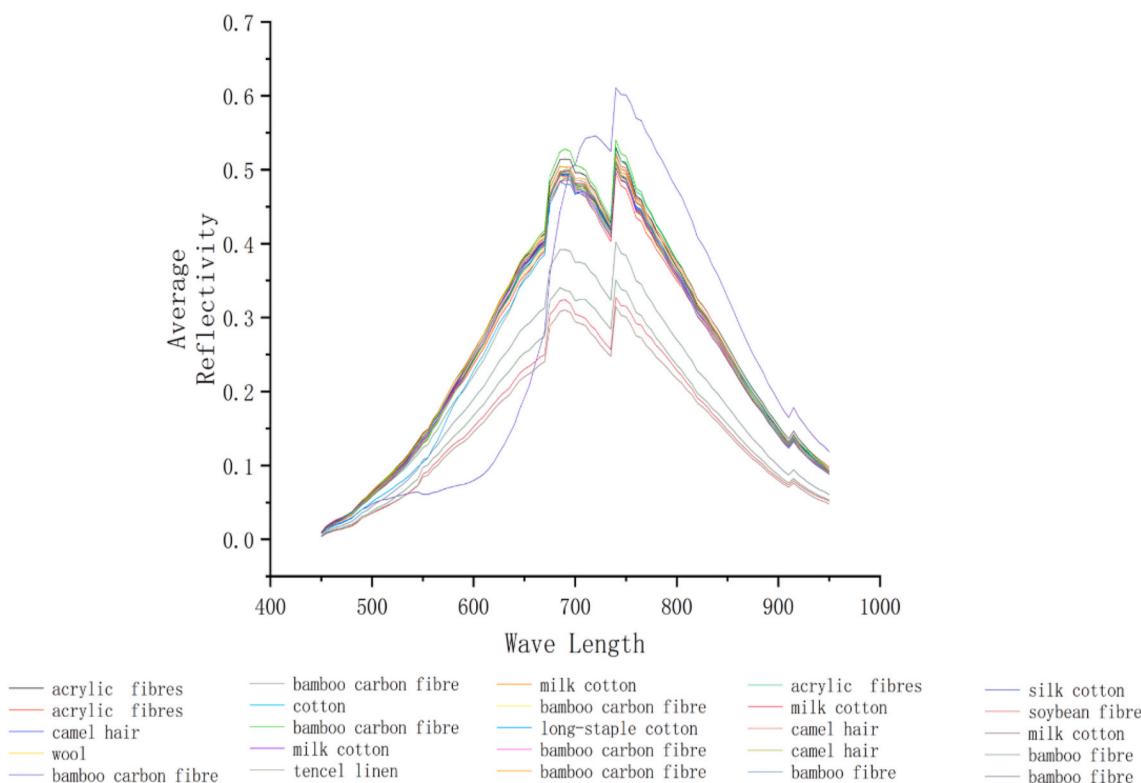


Fig. 3. Average reflection spectrum of 25 textile fibres.

using the following formulae:

$$H(y, \hat{y}) = -\frac{1}{n} \sum_{i=1}^n (y_i \log \hat{y}_i + (1-y_i) \log(1-\hat{y}_i)) \quad (3)$$

where n is the number of categories; y_i and \hat{y}_i is the actual value and the predicted value of textile fibre.

2.5. Model evaluation methods

In this study, accuracy(Acc), precision(Pr), sensitivity(Se), specificity(Sp), and F1 score were used to evaluate the classification model. The specific calculation formulas of the model evaluation methods were shown in formulas (4)-(8):

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (4)$$

$$Pr = \frac{TP}{TP + TN} \times 100\% \quad (5)$$

$$Se = \frac{TP}{TP + FN} \times 100\% \quad (6)$$

$$Sp = \frac{TP}{TN + FP} \times 100\% \quad (7)$$

$$F1 = \frac{2TP}{2TP + FP + FN} \times 100\% \quad (8)$$

In the above equation, TP indicates the number of correctly classified positive samples; TN indicates the number of correctly classified negative samples; FP indicates the number of negative samples misclassified as positive samples; FN indicates the number of positive samples misclassified as negative samples.

Table 3
Classification accuracy of traditional ML model.

Model	ACC% _{test}	Training time	Predict time	Run time
KNN	66.9	1s	12min	12min1s
SVM	61.6	21min18s	1s	21min19s
PLS-DA	76.6	1min18s	1s	1min19s
RF	91.4	42min25s	7s	42min32s

2.6. Software

In this work, the 1D-CNN model was developed in Python (3.6.13) using the Keras library (v2.5.0) and TensorFlow (v2.5.2) backend implementations. All experiments were performed on a computer with specifications of (CPU) Intel (R) Core (TM) I7-11800H 2.30 GHz, (RAM) 16 GB (2×8 GB) DDR4, NVIDIA RTX 3060 GPU.

3. Results and discussion

Fig. 3 showed the average reflectance spectra for each pixel point of the hyperspectral images of 25 textile fibres. It can be seen from the figure that the average reflectance showed a rapidly increasing trend in the 550–670 nm. All spectra had a clear absorption peak at 740 nm. The average emissivity decreased rapidly in the band from 740 to 950 nm, and there was a small absorption peak at 915 nm. The results showed that except for 5 textile fibres that could be separated from the average reflectance spectrogram, the remaining 20 types of textile fibres had slight differences, and the shapes of the reflectance spectra were very similar, which posed a great challenge and difficulty for the intuitive classification of the samples. Therefore, the hyperspectral data need to be analyzed with the help of machine learning.

3.1 Classification results of traditional machine learning

The textile fibre dataset was classified by traditional machine

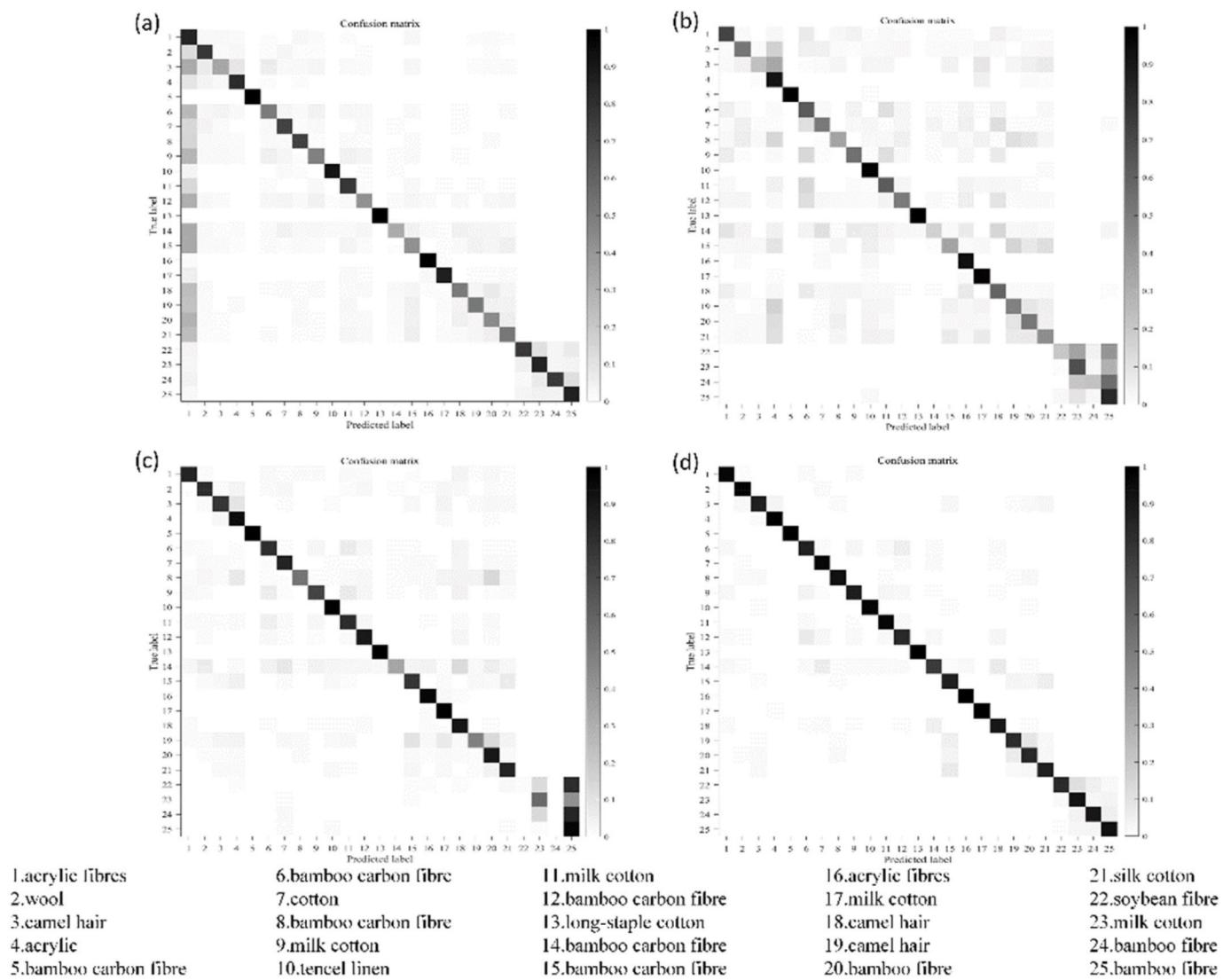


Fig. 4. Confusion matrix of traditional ML model (a) KNN (b) SVM (c) PLS-DA (d) RF.

Table 4
Effect of the number of decision trees on the RF model.

Number of trees	The accuracy rate of the test set	Training time
100	90.9	13min10s
200	91.2	25min44s
300	91.4	42min25s
400	91.2	34min22s

learning (ML) models, and the model classification accuracy and confusion matrix were shown in Table 3 and Fig. 4, respectively. The results showed that the random forest (RF) model had the highest accuracy rate of 91.4% among the four traditional machine learning models, showing a better performance. The parameters of the random forest model were discussed in this study.

In the RF model, the number of decision trees affected the accuracy and training time of the model. As shown in Table 4, with the increase in the number of decision trees, the training time of the RF model gradually became more extended. When the number of decision trees was taken as 300, the accuracy of the test set reached the highest, 91.4%, and the training time was also the longest, 42min25s. When the number of decision trees was taken as 400, the model's accuracy was no longer improved compared to when it was taken as 300, which indicated that

Table 5
Effect of batch size on the neural network model.

Batch Size	BPNN		1D-CNN			
	ACC _{val}	ACC _{test}	Training time	ACC _{val}	ACC _{test}	Training time
32	76.7	72.4	35min06s	98.2	98.3	55min24s
64	83.8	82.8	34min18s	98.7	98.4	43min01s
128	86.7	85.0	21min01s	98.5	98.6	29min16s
256	87.5	87.9	15min01s	98.1	98.2	23min27s
512	83.8	83.6	10min22s	97.4	97.5	19min07s
1024	81.1	49.4	8min01s	91.5	91.6	18min32s

relying on the traditional machine learning method can no longer improve the classification accuracy.

3.2. Training results of the neural network model

The hyperspectral data sets of textile fibres were input into the BPNN model and 1D-CNN model. Table 5 shows the effect of batch size on both models. It was found that the training time of both models decreased as the batch size increased. For the BPNN model, when the batch size was taken as 256, the accuracy of both the validation set and the test set was the highest, reaching 87.5% and 87.9%, respectively, and then the

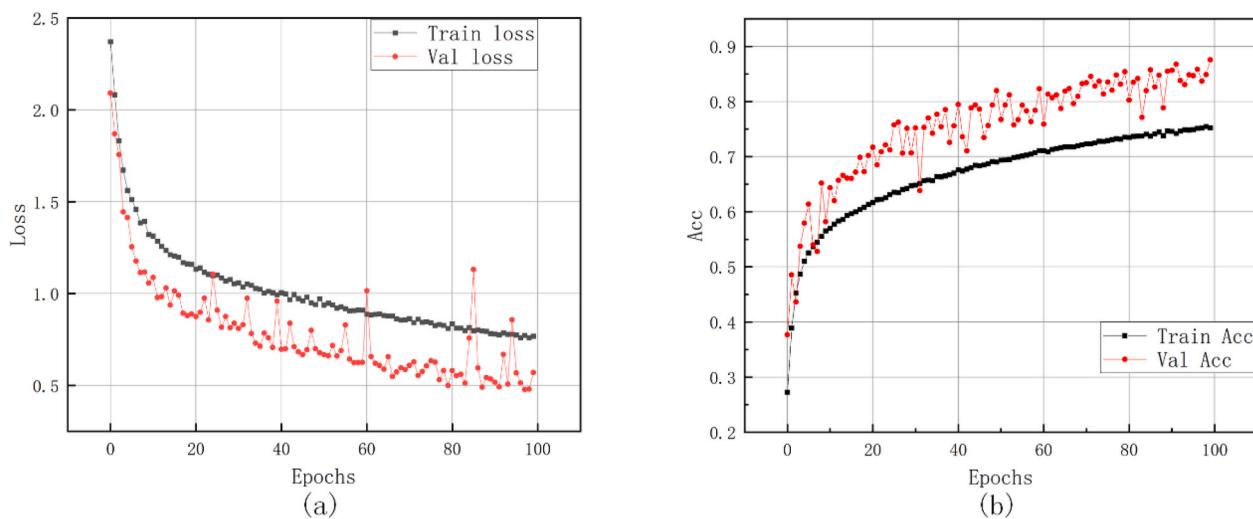


Fig. 5. Visualization of the training process during 100 epochs of the BPNN model. (a: loss function curve of the learning process; b: accuracy of the learning process).

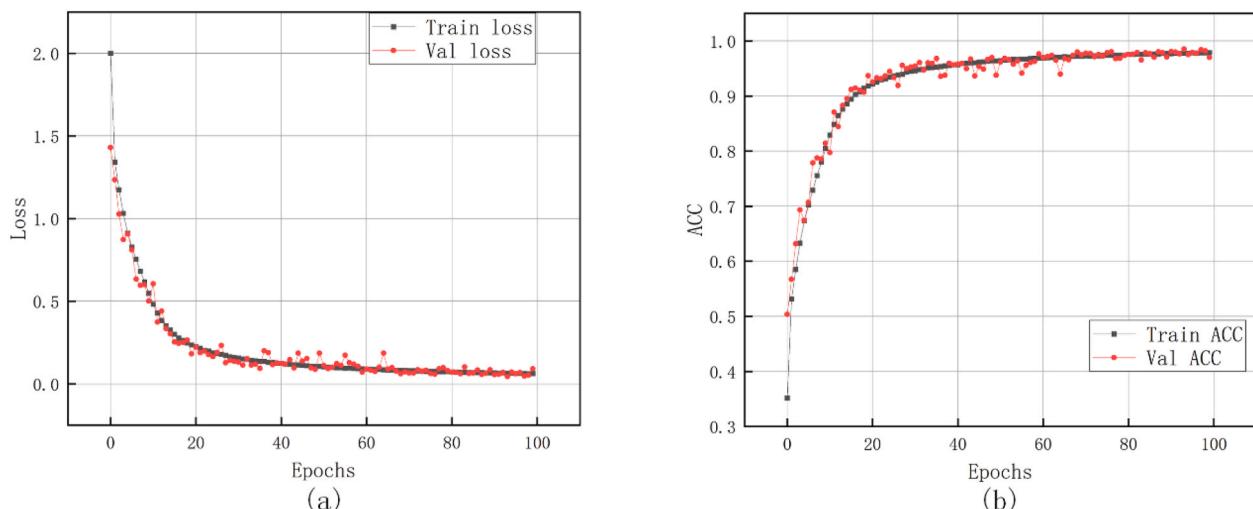


Fig. 6. Visualization of the training process during 100 epochs of the 1D-CNN model. (a: loss function curve of the learning process; b: accuracy of the learning process).

accuracy started to decrease; for the 1D-CNN model, when the batch size was 128, the accuracy of the validation set was the highest, reaching 98.6%. Therefore, we set the batch size to 256 for the BPNN model and 128 for the 1D-CNN.

From Table 3, Table 4, and Table 5, we found that the training time of both the traditional machine learning model and the neural network model was positively related to the complexity of the model, such as the number of decision trees of the RF model and the batch size of the neural network affected the complexity of the model, but the increase in the complexity of the model did not necessarily improve the classification accuracy of the samples. In Table 5, as the batch size of the 1D-CNN model decreased, the training time of the model kept increasing. However, after the batch size reached 128, decreasing the model's batch size decreased the model's test accuracy. Which indicated that the model would take too long to train because the batch size is too small, thus leading to overfitting of the model and not achieving optimal classification results. Therefore, it was important to test and optimize different classification models for different sample data.

The two models were both trained in 100 epochs, and the loss function curves and accuracy curves of the training were shown in Fig. 5 and Fig. 6, respectively. The abscissa of the two figures indicated the

number of iterations in training. The ordinate of Figs. 5 (a) and Fig. 6(a) was the value of the loss function, and the ordinate of Figs. 5 (b) and Fig. 6(b) was the overall accuracy. As can be seen from Figs. 5 and 6, the training process of the BPNN model fluctuated seriously. In contrast, the training process of 1D-CNN was smoother, with a lower loss function value and higher classification accuracy. Fig. 6(a) showed the loss functions of the training and validation sets of 1D-CNN converge to 0.061 and 0.044, respectively. Fig. 6(b) showed that the final classification accuracies of the training and validation sets reach 97.9% and 98.6%, respectively. The results showed that the 1D-CNN model had good stability and efficiency for classifying hyperspectral textile fibre data.

Fig. 7 shows the confusion matrix of the classification results of the BPNN and 1D-CNN models. The figure shows that the classification results of the BPNN model were significantly worse than those of the 1D-CNN model. Most of the predicted samples in the 1D-CNN model were classified as actual classes correctly, and only a few samples had slight misclassification. Therefore, the actual hyperspectral textile fibre types can be identified based on the results of the dataset distribution in the 1D-CNN model. The study results showed that choosing the appropriate neural network was vital. The classification model constructed by 1D-

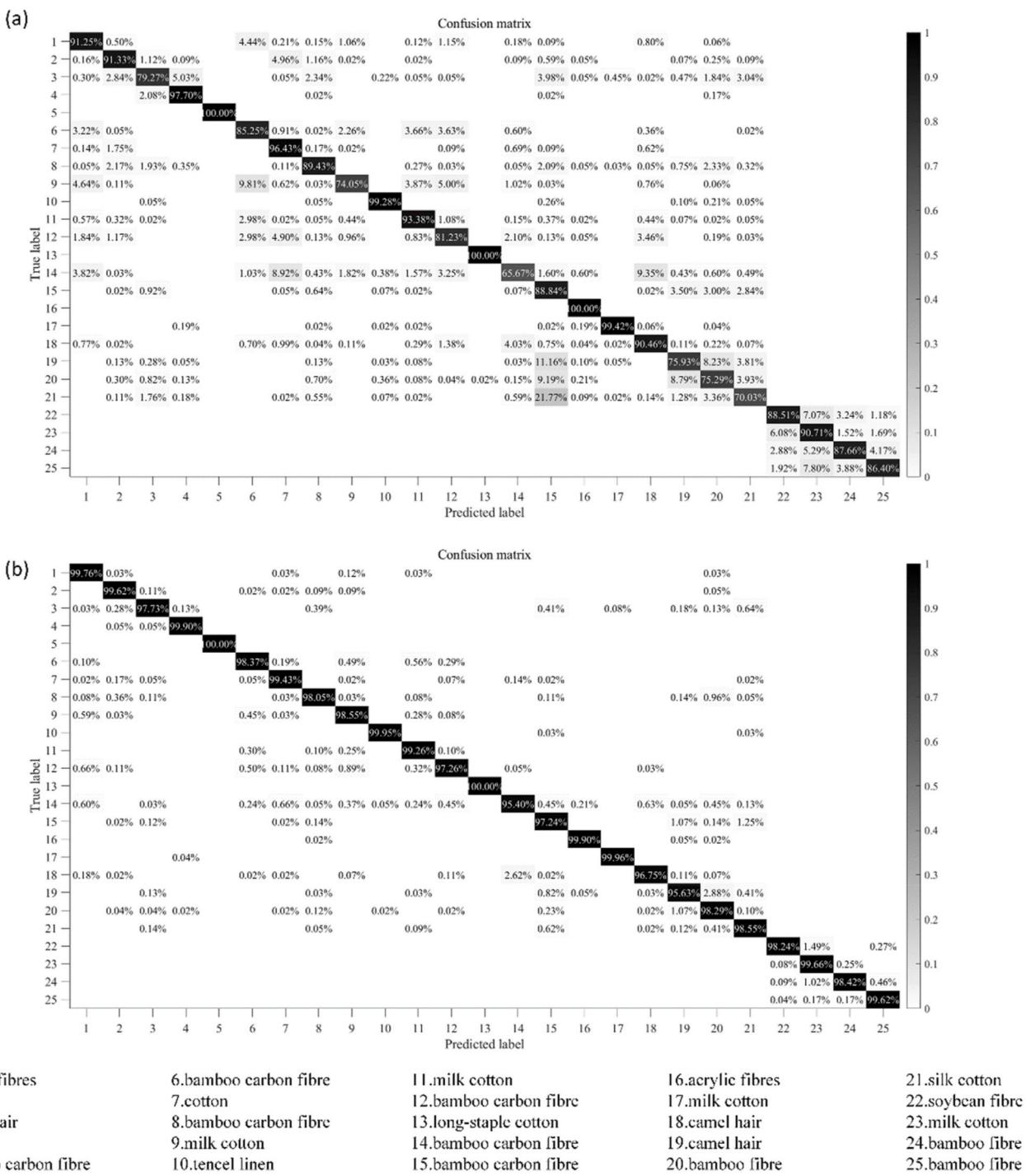


Fig. 7. Classification results of confusion matrix model (a) BPNN (b) 1D-CNN.

CNN was more accurate than BPNN. The accuracy of the training and test sets was above 98.0% after 100 epochs of training, while the BPNN model was basically below 88.0% after 100 epochs of training.

3.3. Model evaluation and analysis

In this part, the classification of textile fibres was based on four traditional machine learning models and two neural network models. The training epochs of the neural network models were all 100 epochs. Table 2 shows other important parameters of the models. Firstly, the samples were classified using the traditional machine learning models. Then the same samples were recognized using the neural network

models. Finally, the traditional machine learning and neural network models were evaluated respectively through the five model evaluation indicators mentioned earlier in this paper, and the results are shown in Table 6. It was found that the ACC, Pr, Se, Sp, and F1 values of the RF model reached 91.4%, 91.4%, 91.2%, 99.6%, and 91.3%, respectively, which were higher than those of the BPNN model and had better classification efficiency. The precision of the RF model was 91.4%, which was higher than sensitivity, indicating that the model had better classification results for true positives than true negatives. The ACC, Pr, Se, Sp, and F1 values of the 1D-CNN model reached 98.6%, 98.7%, 98.6%, 99.9%, and 98.6%, respectively, which were all higher than those of the other five models, indicating that the classification model constructed

Table 6
Classification model evaluation.

	Model	ACC % _{train}	ACC % _{test}	Pr % _{test}	Se % _{test}	Sp % _{test}	F1 socre % _{test}
Traditional ML models	KNN	82.3	66.9	76.1	68.9	98.6	70.4
	SVM	61.3	61.6	63.0	59.6	98.4	58.2
	RF	100	91.4	91.4	91.2	99.6	91.3
	PLS- DA	80.6	80.6	80.0	74.7	99.2	73.5
Neural network	BPNN	75.5	87.9	88.1	87.9	99.5	87.7
	1D- CNN	97.9	98.6	98.7	98.6	99.9	98.6

by the 1D-CNN model was more accurate and had more vital classification ability of 25 textile fibres.

4. Conclusion

In this study, we proposed a method to detect and classify textile fibres using the 1D-CNN model. Firstly, hyperspectral images of textile fibres of 25 different types were collected. Then the ROI was extracted from the hyperspectral images, and the ROI was fused by 10 × 10 pixels. A total of 600,404 instances were obtained. The instances were divided into the train, validation, and test sets. Training and predicting instances by building 1D-CNN models and the 1D-CNN model were compared with KNN, SVM, RF, PLS-DA, and BPNN models. The results showed that the accuracy(98.6%), precision(98.7%), sensitivity(98.6%), specificity (99.9%), and F1 score(98.6%) of the 1D-CNN model were significantly higher than those of KNN, SVM, RF, PLS-DA, and BPNN, with higher accuracy and robustness.

In summary, hyperspectral combined with the 1D-CNN model is a non-destructive, fast, and accurate textile fibre classification and identification method. The next step is to acquire more high-quality images of textile fibre samples and build a more comprehensive database. Meanwhile, in order to further improve the recognition accuracy and classification precision, the neural network model structure and algorithm need to be further optimized, and the network depth can be appropriately increased. In addition, we hope to develop a regression model for quantitative analysis of textile fibre content components using hyperspectral imaging in the future.

CRediT authorship contribution statement

Jiadong Huang: Methodology, Investigation, Writing – original draft. **Hongyuan He:** Conceptualization, Methodology, Investigation, Resources, Project administration, Funding acquisition. **Rulin Lv:** Writing – review & editing, Supervision. **Guangteng Zhang:** Investigation, Supervision. **Zongxian Zhou:** Writing – review & editing. **Xiaobin Wang:** Conceptualization, Resources, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgements

This study was supported by the Opening Project of Shanghai Key Laboratory of Criminal Scene Evidence (No.2021XCWZK05) and Fundamental Research Funds of People's Public Security University of

China (No.2021JKF217).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.aca.2022.340238>.

References

- P.P. Meleiro, C. García-Ruiz, Spectroscopic techniques for the forensic analysis of textile fibers[J], *Appl. Spectrosc.* 51 (4) (2016) 278–301.
- J. Zhou, L. Yu, Q. Ding, et al., Textile fiber identification using near-infrared spectroscopy and pattern recognition[J], *Autex Res. J.* 19 (2) (2019) 201–209.
- J.-R. Riba, R. Cantero, T. Canals, et al., Circular economy of post-consumer textile waste: classification through infrared spectroscopy[J], *J. Clean. Prod.* 272 (2020), 123011.
- D. Quintero Balbas, G. Lanterna, C. Cirrincione, et al., Non-invasive identification of textile fibres using near-infrared fibre optics reflectance spectroscopy and multivariate classification techniques[J], *Eur. Phys. J. Plus* 137 (1) (2022) 1–15.
- R. Powell, W. Van Bronswijk, J. Coumbaros, Enhancing the evidential value of textile fibres: Part 1: development of a spectral database and evaluative comparison strategy[J], *Forensic Sci. Int.* 287 (2018) 54–62.
- R. Powell, P. Collins, G. Horsley, et al., Enhancing the evidential value of textile fibres Part 2: application of a database-driven fibre comparison strategy to a cold-case investigation[J], *Forensic Sci. Int.* 325 (2021), 110894.
- W. Li, Z. Wei, Z. Liu, et al., Qualitative identification of waste textiles based on near-infrared spectroscopy and the back propagation artificial neural network[J], *Textil. Res. J.* 91 (21–22) (2021) 2459–2467.
- L.M. Petrick, T.A. Wilson, W. Ronald Fawcett, High-performance liquid chromatography-ultraviolet-visible spectroscopy-electrospray ionization mass spectrometry method for acrylic and polyester forensic fiber dye analysis[J], *J. Forensic Sci.* 51 (4) (2006) 771–779.
- T.G. Schotman, X. Xu, N. Rodewijk, et al., Application of dye analysis in forensic fibre and textile examination: case examples[J], *Forensic Sci. Int.* 278 (2017) 338–350.
- C. Mourí, R. Laursen, Identification of anthraquinone markers for distinguishing Rubia species in madder-dyed textiles by HPLC[J], *Microchim. Acta* 179 (1–2) (2012) 105–113.
- H. Jin, L. Zhu, S. Jin, et al., Raman spectroscopy analysis of graphene oxide-enhanced textiles[J], *J. Raman Spectrosc.* 52 (4) (2021) 843–848.
- M. Naftaly, J. Molloy, G. Lanskii, et al., Terahertz time-domain spectroscopy for textile identification[J], *Appl. Opt.* 52 (19) (2013) 4433–4437.
- X. Sun, H. Yuan, C. Song, et al., A novel drying-free identification method of cashmere textiles by NIR spectroscopy combined with an adaptive representation learning classification method[J], *Microchem. J.* 149 (2019), 104018.
- K. Saito, T. Yamagata, M. Kanno, et al., Discrimination of cellulose fabrics using infrared spectroscopy and newly developed discriminant analysis[J], *Spectrochim. Acta Mol. Biomol. Spectrosc.* 257 (2021), 119772.
- P. Peets, K. Kaupmees, S. Vahur, et al., Reflectance FT-IR spectroscopy as a viable option for textile fiber identification[J], *Herit. Sci.* 7 (1) (2019) 1–10.
- A. Khan, M. Munir, W. Yu, et al., A review towards hyperspectral imaging for real-time quality control of food products with an illustrative case study of milk powder production[J], *Food Bioprocess Technol.* 13 (5) (2020) 739–752.
- W. Che, L. Sun, Q. Zhang, et al., Pixel based bruise region extraction of apple using Vis-NIR hyperspectral imaging[J], *Comput. Electron. Agric.* 146 (2018) 12–21.
- Y. Zhao, C. Zhang, S. Zhu, et al., Shape induced reflectance correction for non-destructive determination and visualization of soluble solids content in winter jujubes using hyperspectral imaging in two different spectral ranges[J], *Postharvest Biol. Technol.* 161 (2020), 111080.
- A. Gong, S. Zhu, Y. He, et al., Grading of Chinese Cantonese sausage using hyperspectral imaging combined with chemometric methods[J], *Sensors* 17 (8) (2017) 1706.
- C. Zhang, H. Jiang, F. Liu, et al., Application of near-infrared hyperspectral imaging with variable selection methods to determine and visualize caffeine content of coffee beans[J], *Food Bioprocess Technol.* 10 (1) (2017) 213–221.
- W. He, H. He, F. Wang, et al., Non-destructive detection and recognition of pesticide residues on garlic chive (*Allium tuberosum*) leaves based on short wave infrared hyperspectral imaging and one-dimensional convolutional neural network [J], *J. Food Meas. Char.* (2021) 1–11.
- M. Zulfiqar, M. Ahmad, A. Sohaib, et al., Hyperspectral imaging for bloodstain identification[J], *Sensors* 21 (9) (2021) 3045.
- M.Á.F. De La Ossa, J.M. Amigo, C. García-Ruiz, Detection of residues from explosive manipulation by near infrared hyperspectral imaging: a promising forensic tool[J], *Forensic Sci. Int.* 242 (2014) 228–235.
- C.S. Silva, M.F. Pimentel, J.M. Amigo, et al., Detecting semen stains on fabrics using near infrared hyperspectral images and multivariate models[J], *TrAC, Trends Anal. Chem.* 95 (2017) 23–35.
- S. Wang, H. He, R. Lv, et al., Classification Modeling Method for Hyperspectral Stamp-pad Ink Data Based on One-dimensional Convolutional Neural network[J], *Journal of Forensic Sciences*, 2021.
- C.-Y. Wang, T.-S. Ko, C.-C. Hsu, Interpreting convolutional neural network for real-time volatile organic compounds detection and classification using optical emission spectroscopy of plasma[J], *Anal. Chim. Acta* 1179 (2021), 338822.

- [27] W. Ng, B. Minasny, M. Montazerolghaem, et al., Convolutional neural network for simultaneous prediction of several soil properties using visible/near-infrared, mid-infrared, and their combined spectra[J], *Geoderma* 352 (2019) 251–267.
- [28] Z. Liu, W. Li, Z. Wei, Qualitative classification of waste textiles based on near infrared spectroscopy and the convolutional network[J], *Textil. Res. J.* 90 (9–10) (2020) 1057–1066.
- [29] K. Huang, S. Li, X. Kang, et al., Spectral–spatial hyperspectral image classification based on KNN[J], *Sens. Imag.* 17 (1) (2016) 1–13.
- [30] R. Dong, J. Wang, S. Weng, et al., Field determination of hazardous chemicals in public security by using a hand-held Raman spectrometer and a deep architecture-search network[J], *Spectrochim. Acta Mol. Biomol. Spectrosc.* 258 (2021), 119871.
- [31] R.M. Balabin, R.Z. Safieva, Biodiesel classification by base stock type (vegetable oil) using near infrared spectroscopy data[J], *Anal. Chim. Acta* 689 (2) (2011) 190–197.
- [32] D. Wang, J. Jiang, J. Mo, et al., Rapid screening of thyroid dysfunction using Raman spectroscopy combined with an improved support vector machine[J], *Appl. Spectrosc.* 74 (6) (2020) 674–683.
- [33] Y.-Y. Yuan, S.-T. Wang, J.-Z. Wang, et al., Rapid detection of the authenticity and adulteration of sesame oil using excitation-emission matrix fluorescence and chemometric methods[J], *Food Control* 112 (2020), 107145.
- [34] S. Pan, H. Zhang, Z. Li, et al., Classification of Ginseng with different growth ages based on terahertz spectroscopy and machine learning algorithm[J], *Optik* 236 (2021), 166322.
- [35] K.Y. Peerbhay, O. Mutanga, R. Ismail, Random forests unsupervised classification: the detection and mapping of solanum mauritianum infestations in plantation forestry using hyperspectral data[J], *IEEE J. Sel. Top. Appl. Earth Obs. Rem. Sens.* 8 (6) (2015) 3107–3122.
- [36] X. Deng, Z. Liu, Y. Zhan, et al., Predictive geographical authentication of green tea with protected designation of origin using a random forest model[J], *Food Control* 107 (2020), 106807.
- [37] M.H. Kabir, M.L. Guindo, R. Chen, et al., Geographic origin discrimination of millet using vis-NIR spectroscopy combined with machine learning techniques[J], *Foods* 10 (11) (2021) 2767.
- [38] M. Ciulu, E. Oertel, R. Serra, et al., Classification of unifloral honeys from SARDINIA (Italy) by ATR-FTIR spectroscopy and random forest[J], *Molecules* 26 (1) (2020) 88.
- [39] J. Zeng, Y. Guo, Y. Han, et al., A review of the discriminant analysis methods for food quality based on near-infrared spectroscopy and pattern recognition[J], *Molecules* 26 (3) (2021) 749.
- [40] A. Sharma, R. Verma, R. Kumar, et al., Chemometric analysis of ATR-FTIR spectra of fingernail clippings for classification and prediction of sex in forensic context[J], *Microchem. J.* 159 (2020), 105504.
- [41] S. Kassaymeh, S. Abdullah, M. Al-Laham, et al., Salp swarm optimizer for modeling software reliability prediction problems[J], *Neural Process. Lett.* 53 (6) (2021) 4451–4487.
- [42] G. Yu, B. Ma, J. Chen, et al., Nondestructive identification of pesticide residues on the Hami melon surface using deep feature fusion by Vis/NIR spectroscopy and 1D-CNN[J], *J. Food Process. Eng.* 44 (1) (2021), e13602.
- [43] D. Ma, L. Shang, J. Tang, et al., Classifying breast cancer tissue by Raman spectroscopy with one-dimensional convolutional neural network[J], *Spectrochim. Acta Mol. Biomol. Spectrosc.* 256 (2021), 119732.
- [44] R. Haghi, E. Pérez-Fernández, A. Robertson, Prediction of various soil properties for a national spatial dataset of Scottish soils based on four different chemometric approaches: a comparison of near infrared and mid-infrared spectroscopy[J], *Geoderma* 396 (2021), 115071.