A novel deep learning-based chemical image identification method of infrared spectroscopy using external perturbation

Electronic Supplementary Information (ESI)

Table of Content

Sample Preparation

First, all the textile sample was continuously dried in a vacuum oven at 105 °C for 3 h to obtain a dried sample, and then they were placed in a chamber with relative humidity (RH) of 100% and a constant temperature of 20 °C where the sample was subjected to moisture absorption from air. Samples were taken out at certain adsorption times to prepare samples with variant water contents. The sample was weighed using an analytical balance with an accuracy of 0.1 mg and the moisture regain (water content) of the sample was determined according to the following Equation under standard conditions (25 °C and RH 65%):

$$\text{Moisture regain} = (Wc - Wd) / Wd * 100\%$$

Where Wc and Wd are the weight of the sample after water absorption and dry weight, respectively. According to the Chinese national standard GB/T 9994-2018[1], the maximum moisture regains of the sample is set to 16.3% (%, w/w). Four different moisture contents were prepared for each sample, including 0, 5.4, 11.2, and 16.3 (%, w/w), and a total of 936 independent samples with specific moisture contents were obtained.

Spectral Collection

The diffuse reflectance near-IR spectrum of the sample was collected under constant temperature and humidity conditions using a Nicolet Antaris II FT-NIR spectrometer equipped with an integrating sphere attachment. The built-in gold foil was used to capture the background spectrum. Each sample of about 0.5 m^2 in size was folded into 4-6 layers, and then placed directly on the window of the integrating sphere and pressed with the iron cube to make it in close contact with the surface. The parameters of the spectral acquisition are resolution 4 cm^{-1}, scan number 32 and the spectral range 10,000-4000 cm^{-1}. Each spectral acquisition takes approximately 1 min, and three duplications of measurement are performed for each sample to suppress random noise, the average spectrum is calculated as the sample spectrum.

Data Preprocessing

Discrete wavelet transformation ('db4' filter), derivative[2](including 1st and 2nd derivative) and multiple scatter correction (MSC)[3] methods are successively used to pre-process the spectrum to respectively eliminate high-frequency noise, baseline drift,

and light scattering. The sample is projected into the 3-D principal component spectral space using principal component analysis(PCA) to visualize the spatial distribution of the samples. A chemical image of the sample was prepared using two-dimensional correlation spectral (2DCOS) analysis. Classification models are created using Soft Independent Modelling by Class Analogy (SIMCA) and Support Vector Machine (SVM), respectively. GoogLeNet models were implemented by Tensorflow and Keras library. Among them, the GoogLeNet classification operation was performed on the Python platform, and the others were programed in MATLAB R2017b. The program codes of spectral processing and identification were written by ourselves. The process was performed on a Windows PC with 16 GB of RAM and an Nvidia Geforce GTX 1080Ti graphics card.

Chemical Image Construction

Place any sample in a precisely designed humid environment and gradually increase the moisture content of the sample. An infrared(IR) spectrum x is acquired each time the water content is changed. A series of dynamic moisture-dependent spectra acquired corresponding to m different water contents constitute a spectral matrix X_{m*n}, where n represents the number of wavelength points. Through the 2D-COS analysis of X_{m*n}, a synchronous map Φ and an asynchronous map Ψ are obtained, and the calculation formula is as follows:

$$\hat{A} = X - \bar{X}$$

$$\Phi = \hat{A} * \hat{A}^T$$

$$\Psi = \hat{A} * N * \hat{A}^T$$

Where $\hat{A}$ is the dynamic spectrum, which is the difference between the original spectrum X and the reference spectrum, and the average spectrum $\bar{X}$ of the original spectrum is usually used as the reference spectrum. N is a Hilbert-Noda matrix, and the calculation formula is as follows:

$$N = \begin{cases} 0 & i = j \\ \dfrac{1}{\pi(j - i)} & i \neq j \end{cases}$$

Among them, i and j are the order of $\hat{A}$ and $\hat{A}^T$, respectively.

Synchronous map, which is symmetric with respect to the diagonal line, presents in-phase evolution of peak intensities with perturbation. In the asynchronous map, an asynchronous cross-peak can appear only if the intensities of two spectral profiles change out of phase with each other with the profile on both sides of the diagonal line being asymmetric[4]. They may detect the subtle difference of the moisture-induced behavior of the O-H group in a molecule under water perturbation, as a result, much new information that cannot be readily acquired from conventional static spectra could be obtained.

Due to the symmetry of the two-dimensional correlation spectrum, there is redundant information for the synchronous 2DCOS map. Considering they have the same dimensions, the upper triangular portion of the synchronous map and the lower triangular part of the asynchronous graph are combined into a fused two-dimensional correlation map, that is, a chemical image. It should be noted that the main diagonal line of the synchronous map (usually called the power spectrum) is physically meaningful. Therefore, it is retained in the fusion correlation map, while the data on the main diagonal line of the asynchronous map is abandoned.

GoogLeNet Transfer Learning

The GoogLeNet pre-trained model has obtained excellent image feature extraction ability (i.e. weight) by training in a big database. In this paper, a small number of chemical images were used to train only the subsequent fully connected and classified layers with its deep convolution layer preserved. The number of filters of the fully connected layer(FC), as well as the category number of the classification layer, is equal to the number of types of the current chemical images. The weights between layers can be determined by an iterative process learning using a backpropagation (BP) algorithm[5]. The objective function is minimized using mean square error (MSE), which is regularized using the L2 norm to avoid overfitting. The formula is as follows:

$$\text{Loss} = \frac{1}{N}\sum_{n=1}^{N}[(y_n - \hat{y}_n)^2] + \lambda\|w\|^2$$

Where N is the number of samples, y and ŷ respectively represents measured values and predicted values, λ is the regularization parameter and w is a weight matrix.

In addition, the activation function of the FC is the leaky rectified linear function (ReLu) function[6], the batch normalization (BN)[7] is used to accelerate the retraining

process. The BP algorithm combined with the Sgdm optimizer[8] is employed to find the local minimum of the objective function. Weight initialization is achieved by variance scaling to prevent from gradient vanishing.

The model evaluation index is the prediction accuracy, namely:

$$\text{Accuracy} = NA / NT * 100\%$$

where NA and NT are the number of correct predictions and the total number of validation samples, respectively. Given that external validation samples are independent with training samples on which the model is calibrated, the prediction accuracy of validation samples is an objective metric for evaluating classification models.

The input of the 141st Dropout layer is randomly set to zero to prevent overfitting. Change its random probability to 65%. The 142nd fully connected layer and the 144th Classification Output layer contain information for combining network extracted features into Probability of belonging to each class, loss value and predicted affiliation. Both the number of filters in the FC and the category number of the classification layer are set to equal to 2, that is, category of the current chemical image. And increasing the learning rate factor of the FC to obtain a faster learning rate. Because the Sigmoid activation function is more suitable for the two-class problem than the Softmax activation function, it is used before the Output layer[9]. It is worth noting that the image input layer limits the size of the input image to 224*224*3 pixels, so the chemical image is converted to the RGB (Red-Green-Blue) image with the corresponding size before the training process. The neural network training process requires the setting of various hyper-parameters. Among them, InitialLearnRate specifies the initial step size in the negative gradient direction of the loss function, MiniBatchSize is the size of the training set subset used in each iteration, and MaxEpochs represents the maximum epoch number used for training. There are currently no general rules to tune them. In this study, these hyperparameters, InitialLearnRate, 1e-4, MiniBatchSize, 15, MaxEpochs, 20 were determined by the trial-and-error method.

Spectral Analysis

Figure S1(A) shows the near-IR spectra of all dried cashmere and cashmere-wool blended textiles. There is a serious drift of the spectral baseline between samples. The main reasons include the light scattering effect caused by the difference in sample texture and the absorption caused by the color difference. The 2nd derivative

preprocessing is performed in Figure S1(A) to obtain Figure S1(B), which effectively eliminates the baseline drift and improves the apparent resolution of the spectral characteristic peaks. It can be seen that on the one hand, the near-IR spectra of cashmere textiles and cashmere and wool-blend textiles contain rich composition information. On the other hand, the near-IR spectra of the two textiles are very similar. The case of cotton textiles vs mercerized cotton textiles is also similar. The near-IR spectrum and the 2nd derivative spectrum of the dried samples are shown in Figure S2(A) and (B). The main characteristic absorption bands of these four samples are shown in Table S1.

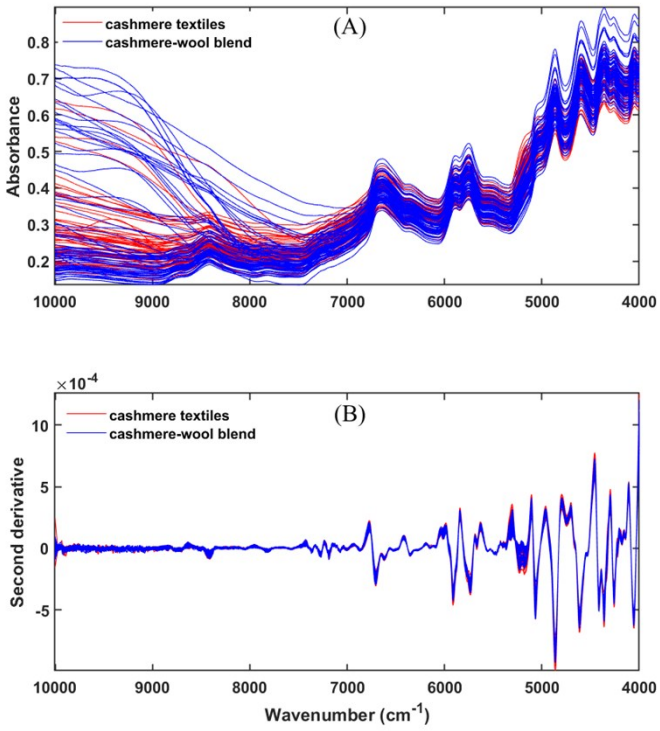

Fig. S1 (A) Raw NIR spectra of the dried samples and (B) their 2nd derivative spectra, the blue color for cashmere-wool blend textiles and the red for pure cashmere textiles.

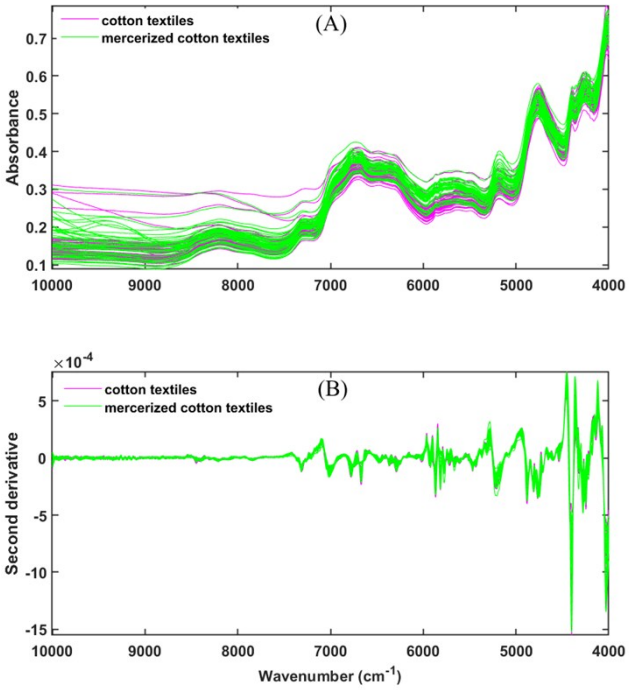

Fig. S2 (A) Raw NIR spectra of the dried samples and (B) their 2nd derivative spectra, the pink color for cotton textiles and the green for mercerized cotton samples.

Table S1. Assignments of NIR spectra of dried textile samples.

Peak position / cm^{-1}	Assignment
8410-8177	3ν(CH)
6715	2ν(OH) of textile
6314	υ(OH) +δ(OH) of cellulose
5980-5800	νs(CH) and νas(CH2) of textile
5612	2ν(CH)
4752	2υ(OH) multimers of cellulose
4393	O-H/C-H of cellulose
4270	υ(CH2) +δ(CH2)
4015	υ(C-H)+ υ(C-C)

PCA Analysis

PCA analysis of cashmere textiles and cashmere-wool blends as well as pure wool textiles, and cotton textiles and mercerized cotton textiles were performed. For cashmere textiles and cashmere-wool blends as well as pure wool textiles, the cumulative contribution of the first three principal components was 89.82% (PC1 equals 51.31%, PC2 and PC3 account for 28.35 and 10.16%, respectively). Regarding cotton and mercerized cotton textiles, the cumulative contribution of the first three principal components is 89.82% (PC1, PC2, PC3 accounted for 60.29, 16.33 and 12.16%, respectively), which indicates that the first three principal components can basically reflect the information contained in the sample data set. The scores of all the sample spectra on the first three principal components were plotted to obtain their distribution trend in the 3-D principal component spectral space, as shown in Figure S3. As can be seen from Figure S3(A), the distribution of cashmere textiles, cashmere-wool blends, and pure wool textile samples is different, but there is a serious overlap between them. The same for cotton and mercerized cotton textiles.

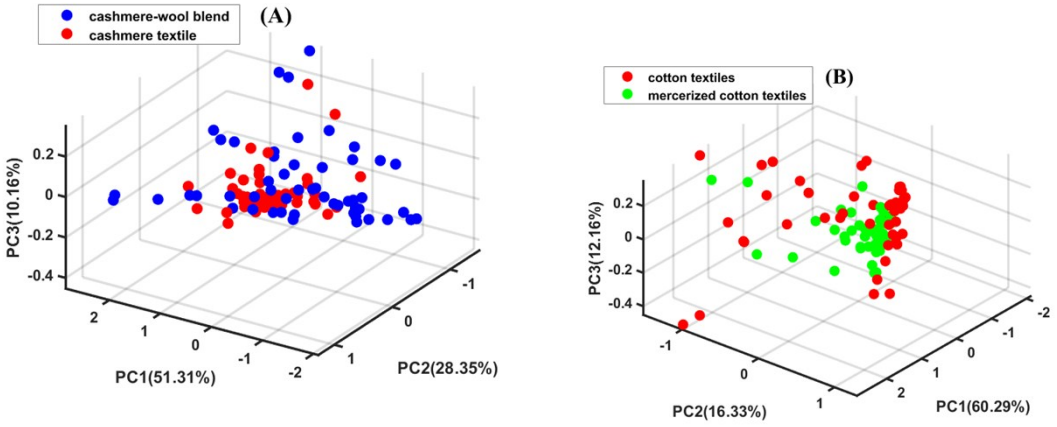

Fig. S3 The three-dimensional score plots of PC1, PC2, and PC3 for dry textile samples. (A) the plot for the cashmere vs cashmere-wool blend samples, (B) the plot for the cotton vs mercerized cotton samples.

Comparison of Discrimination Models

SIMCA model. Among the linear classification methods commonly used in qualitative analysis of IR spectroscopy, SIMCA may be the best one and has many practical applications[10]. In SIMCA, the principal component spectral space (submodel)

of each type of sample is first established, and then the distance between the measured sample and the spectral components of various principal components is calculated, and the attribution is determined according to the distance. In the SIMCA model established here, the different spectral preprocessing methods are used respectively, and leave-one-out cross-validation (LOOCV) is combined with the Q value[11] vs *Hotelling* T^2 map[11] to determine the optimal principal component number of each class model. Several SIMCA models for drying cashmere and cashmere-wool blended and pure wool and dried cotton-mercerized cotton textiles were established one by one. The statistical parameters of these models are shown in Tables S2 and S3. It can be seen that compared with the original spectrum, after using three pretreatment methods, the prediction accuracy of the model is significantly improved. Among them, the 2nd derivative pretreatment method achieved the best results: the prediction accuracy of cashmere and cashmere-wool blended and pure wool textiles were 60.48 and 63.33%, respectively, and the prediction accuracy of cotton and mercerized cotton textiles was 68.89 and 71.02%. Therefore, the MSC method is better than the 1st derivative but not as good as the 2nd derivative method. Compared with the derivative processing method, MSC has a stronger effect in eliminating light scattering, which not only can effectively eliminate the light scattering effect caused by the difference in textile texture, but also can eliminate the light scattering effect caused by the difference in microscopic diameter between cashmere and wool on which the discrimination partly depend. Eliminating this part of the information is detrimental to modeling. It is reasonable to interpret the effect that MSC is not as good as the 2nd derivative.

Table S2. The statistics of the SIMCA models respectively using the raw spectra and the pretreated spectra by different pretreating methods.

Preprocessing	Number of latent variables		Accuracy / %	
	Cashmere	Blend[a]	Cashmere	Blend[a]
Raw	4	7	46.19	36.67
MSC	7	5	50.71	56.67
1st derivative	5	6	55.71	60

2nd derivative	5	9	60.48	63.33

a. Blend denotes the cashmere-wool blended textiles.

Table S3. The statistics of the SIMCA models respectively using the raw spectra and the pretreated spectra by different pretreating methods

Preprocessing	Number of latent variables		Accuracy / %	
	Cotton	MC[a]	Cotton	MC[a]
Raw	4	6	51.71	52.67
MSC	8	7	55.19	56.67
1st derivative	9	11	65.71	61.22
2nd derivative	4	7	68.89	71.02

a. MC denotes the Mercerized cotton textiles

SVM model. SVM is a commonly used method for discriminating two types of samples by IR spectral discriminant analysis. It is generally believed that SVM is superior to linear classification discriminating methods in its ability to solve nonlinear problems[13]. SVM uses kernel functions to map linearly indivisible raw data into separable higher dimensional spaces. In this study, a radial basis function (RBF) is used as a kernel function to generate an optimal decision function through cross-validation to avoid overfitting. The grid search is used to optimize both the penalty parameter and the kernel coefficient, and their ranges are limited to the following two lists, [0.01, 0.1, 1, 10, 100, 1000] and [0.01,0.001, 0.0001]. Table S4 lists the statistics of the SVM model. The comparison shows that the model performance is similar to that of the SIMCA model, which indicates that the SVM can not effectively distinguish different types of textiles with high similarity.

Table S4. The statistics of the SVM models

Kernel coefficient	Preprocessing	Penalty coefficient	Accuracy / %			
			Blend[a]		MC[b]	
			Cashmere	Cotton		
0.0001	2nd derivative	0.1	66.67	65.71	-	-
0.001	2nd derivative	1.0	-	-	70.09	72.51

a. Blend denotes the cashmere-wool blended textiles.

b. MC denotes the Mercerized cotton textiles.

Moisture-dependent Spectra

We applied an external moisture perturbation to the sample to enhance the distinguishable features for different types of samples. Samples with the moisture content of 5.4, 11.2, and 16.3 (w/w, %) were prepared on the basis of the dried ones, that is, a single sample produces multiple moisture-containing samples and their near-IR spectra were collected simultaneously. A single textile sample per class was randomly selected, and their water-induced spectra are compared in Figure S4. It can be seen that the near-IR spectra of the wet samples are dominated by two broad water peaks compared to those of the dried sample. Among them, the 7100-6800 cm^{-1} band is attributed to ν1+ν3 (ν1 is symmetric stretching, ν3 is asymmetric stretching mode) [14], while the 5150-4950 cm^{-1} band with the largest change in absorbance is attributed to ν2+ν3 (ν2 is the bending vibration mode)[15]. According to the literature[16], the latter band can reflect the state of hydrogen bonds in the humid textile matrix, including rich information on the water absorption pattern of the sample. It can be concluded that the introduction of water perturbation significantly increases the amount of spectral data and the amount of information, and magnifies the near-IR spectral difference between different types of textiles.

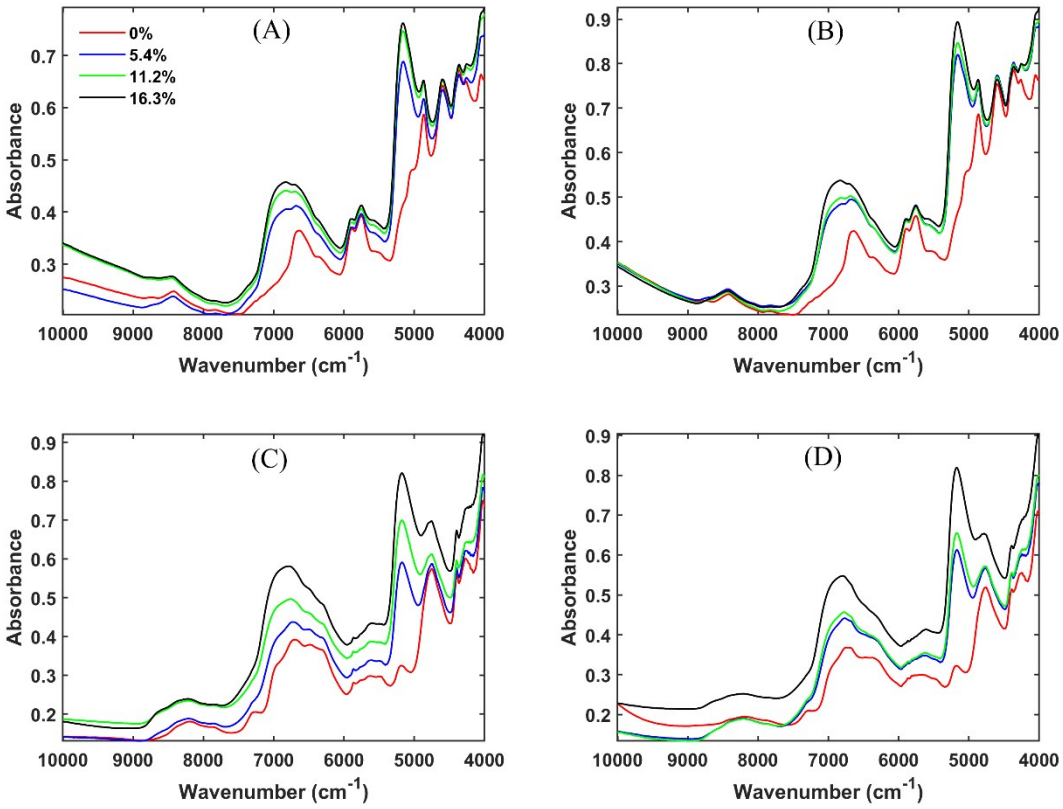

Fig. S4 Moisture-dependent NIR spectra for a randomly selected (A) cashmere, (B) cashmere-wool blend, (C) cotton, (D) mercerized cotton textiles.

Reference

1. Triantafillou, E., Zemel, R., & Urtasun, R. Advances in Neural Information Processing Systems. 2017, 2255-2265.
2. Kitamura, K., Hozumi, K. Anal. Chim. Acta. 1985, 172, 111-118.
3. Isaksson, T., & Næs, T. Appl. Spectrosc. 1988, 42(7), 1273-1284.
4. Noda, I. Appl. Spectrosc.1993, 47(9), 1329-1336.
5. Leonard, J., Kramer, M. A. Comput. Chem. Eng. 1990, 14(3), 337-341.
6. Maas, A. L., Hannun, A. Y., & Ng, A. Y. Proc. Icml. 2013, 30(1): 3.
7. Ioffe, S., Szegedy, C. International Conference on Machine Learning. 2015,448-456.
8. Kim, S. M., Shin, J., Baek, S., & Ryu, J. H. J. Coastal Res. 2019, 90(sp1), 302-309.
9. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Rabinovich, A. Proceedings of the IEEE conference on computer vision and pattern recognition. 2015, 1-9.
10. Yang, I. C., Tsai, C. Y., Hsieh, K. W., Yang, C. W., Ouyang, F., Lo, Y. M., & Chen, S. J. Food Drug Anal. 2013, 21(3), 268-278.
11. Agelet, L. E., Ellis, D. D., Duvick, S., Goggi, A. S., Hurburgh, C. R., & Gardner, C. A. J. Cereal Sci. 2012, 55(2), 160-165.

12. Bylesjö, M., Rantalainen, M., Cloarec, O., Nicholson, J. K., Holmes, E., Trygg, J. J. Chemometr. 2006, 20(8-10), 341-351.
13. Luts, J., Ojeda, F., Van de Plas, R., De Moor, B., Van Huffel, S., & Suykens, J. A. Anal. Chim. Acta. 2010, 665,129–145.
14. Buijs, K., & Choppin, G. R. (1963). Near-IR studies of the structure of water. I. Pure Water. The Journal of Chemical Physics, 39(8), 2035-2041.
15. Bonner, O. D., Choi, Y. S. J. Phys. Chem. 1974, 78(17), 1723-1727.
16. Franks, F. Water in Crystalline Hydrates Aqueous Solutions of Simple Nonelectrolytes (Vol. 2). Springer Science & Business Media. 2013.