DIFFERENTIATION OF PLASTICS BY COMBINING RAMAN SPECTROSCOPY AND MACHINE LEARNING**

Y. Yang, W. Zhang, Zh. Wang, Y. Li*

School of Information Science and Engineering at Lanzhou University, Lanzhou, Gansu, China; e-mail: liyuee@lzu.edu.cn

We combined Raman spectroscopy with machine learning for the classification of 11 plastic samples. A confocal Raman system with an excitation wavelength of 532 nm was used to collect the Raman spectral data of plastic samples and principal component analysis was used for feature extraction. The prediction models of plastic classification based on three machine learning algorithms are compared. The results show that all three machine learning algorithms are able to classify 11 plastics well. This indicates that the combination of Raman spectroscopy and machine learning has great potential in the rapid and nondestructive classification of plastics.

Keywords: *Raman spectroscopy, plastic classification, principal component analysis, random forest, support vector machine, neural networks, machine learning.*

ДИФФЕРЕНЦИАЦИЯ ПЛАСТМАСС СОЧЕТАНИЕМ МЕТОДОВ СПЕКТРОСКОПИИ КОМБИНАЦИОННОГО РАССЕЯНИЯ СВЕТА И МАШИННОГО ОБУЧЕНИЯ

Y. Yang, W. Zhang, Zh. Wang, Y. Li*

УДК 535.375.5:678.5/.6

Школа информатики и инженерии Ланьчжоуского университета, Ланьчжоу, Ганьсу, Китай; e-mail: liyuee@lzu.edu.cn

(Поступила 31 мая 2021)

Классификация образцов пластмасс выполнена на основе объединения КР-спектроскопии с машинным обучением. Конфокальная КР-система с длиной волны возбуждения 532 нм использована для сбора КР-спектральных данных образцов, анализ главных компонент применен для выделения их признаков. Проведено сравнение прогностических моделей классификации пластмасс на основе трех алгоритмов машинного обучения. Показана удовлетворительная классификация 11 видов пластмасс с помощью рассматриваемых моделей.

Ключевые слова: *спектроскопия комбинационного рассеяния света, классификация пластмасс, метод анализа главных компонент, метод случайного леса, метод опорных векторов, нейронные сети, машинное обучение.*

Introduction. People's demand for plastics is rising every year and today plastic products can be seen everywhere [1–5]; however, plastics can cause great harm to the environment due to their own characteristics. Recycling has become an urgent problem to be solved [6, 7]. Waste plastics are diverse in nature, and the manual sorting of such a feedstock is time-consuming, labor-intensive, and inefficient; therefore, automated methods for their separation are used. The complex physical and chemical properties of the plastic raw materials made by the melting of mixed plastics are unstable, and usually cannot be used as high-quality or even normal raw materials, which is a huge waste of plastic resources [8]. Waste plastics are usually recycled through four steps: collection, separation, processing, and marketing [9]. In the process of separation,

**Full text is published in JAS V. 89, No. 4 (http://springer.com/journal/10812) and in electronic version of ZhPS V. 89, No. 4 (http://www.elibrary.ru/title_about.asp?id=7318; sales@elibrary.ru).

the effective identification and classification of plastic waste has become more and more important. The classification methods for plastic composition verification include–traditional appearance based, density, pyrolysis and heating, and spectral [9, 10]. Compared with the traditional plastic classification methods, the spectral classification method has the advantages of simple operation, fast analysis and high reliability. Common spectral classification methods include infrared (IR) spectroscopy, laser-induced breakdown spectroscopy (LIBS), X-ray absorption spectroscopy, and Raman spectroscopy [9, 11–15]. However, infrared spectroscopy does not distinguish colored or opaque plastics effectively because of its limited penetration [16]. LIBS is a commonly used method for plastic spectrum classification, but its reliability and stability need to be improved [17]. Although X-ray absorption spectroscopy has good penetrability and stability, it requires large equipment and high-voltage support, which is not conducive to the rapid classification of plastics [9].

Raman spectroscopy is a technique based on Raman scattering theory to characterize the molecular structure of matter. Compared with other spectral analysis techniques, Raman spectral analysis has many unique advantages and characteristics such as simple operation, high sensitivity, and the possibility of online fast, nondestructive testing [18]. Due to its unique characteristics, Raman spectroscopy has been widely used in various fields of production and life [19]. Whether the plastics are colorless or colored, classification can be correctly achieved using Raman spectroscopy [20]. In recent years, the application of Raman spectroscopy in the classification of polymerization has been increasing gradually. Chen Hesheng et al. [21] carried out Fourier Raman spectral analysis on several kinds of plastics, and the classification of plastics can be accomplished intuitively by comparing spectral images. Allen et al. basically achieved the classification of PET, PS, PP, and six other plastics by using principal component analysis and the *k*-nearest neighbor method combined with Raman spectroscopy [22]. Tsuchida and Kawazumi et al. used Raman spectroscopy to extract the Raman characteristic peak of plastics and established the identification model. The recognition model realizes the recognition and classification of PP, PS, and ABS, which fully demonstrated the feasibility of Raman spectroscopy for plastic identification [23]. Due to the limited plastics involved, plastics classification based on Raman spectroscopy has not been used in practical applications by recyclers.

Here, we combine principal component analysis (PCA) and machine learning (ML) to build classification models for the classification of 11 types of plastics. We use PCA to reduce the dimension of the data set while retaining as much original information as possible [24], then Machine learning algorithms are trained to learn with the dimension-reduced data. Machine learning technology is a general term for all kinds of learning algorithms. The idea behind machine learning is to find patterns and structures by training existing data, and then make decisions or predictions about new data [25]. The ML algorithm has been widely used in various kinds of spectral recognition due to its fast, efficient high-recognition rate [26]. Random forest (RF), support vector machine (SVM), and back propagation neural network (BPNN) are the three most popular machine learning algorithms. Therefore, in this study, PCA is used for feature extraction, and the three commonly used machine learning algorithms are used to build different prediction models for these features. The prediction results show that the prediction model based on Raman spectroscopy combined with machine learning is classified as having good performance for plastics. This method provides a new choice for the sorting and recycling of waste plastics.

To summarize, our method can accurately identify polyethylene (PE), polyethylene terephthalate (PET), polypropylene (PP), polystyrene (PS), polyvinyl chloride (PVC), acrylonitrile butadiene styrene copolymers (ABS), polyamide (PA), thermoplastic elastomer (TPE), thermoplastic polyurethanes (TPU), thermoplastic rubber (TPR), and thermoplastic vulcanizate (TPV), etc. Additionally, the PA of various additives can also be identified. The background fluorescence from colored, crushed, or dirty samples can be significantly reduced by using a confocal system. Owing to the features of our method (combining machine learning and PCA), this method can easily be extended to include more polymers.

Material and methods. Eleven commonly used plastic samples including PE, PET, PP, PS, PVC, ABS, PA, TPE, TPU, TPR, and TPV are considered. The plastic samples come from two sources: granular raw materials purchased on the market and plastic products collected in daily life (listed in the electronic supplementary materials). There are several sources for each type of plastic.

Acquisition of Raman spectra. Raman spectra (Raman shift: 300–3,200 cm^{-1}, the integration time: 3 s) of all plastic samples were acquired at the 532 nm excitation level (Verdi V-6, laser power 10 mW) on an Alpha confocal Raman microscope system coupled to WITec's Project 5.0 software (WITec, Germany) [27]. All spectra were collected using a Zeiss LD EC Epiplan-Neofluar 50× objective lens with NA 0.55 (Nikon, Japan) at a temperature of 25°C. Table 1 lists the details of all plastic samples.

TABLE 1. Plastic Data Set

Plastic type	Sample source	Sample number	Spectra number
PE	Four milk bottle, Drink caps, Plastic bags	20	75
PET	Soft drink bottles, Food packaging bags	13	50
PP	Toy materials, Plastic containers, Packaging bags	18	75
PS	Stationery materials, Interior decoration materials, Instrument shell, Foam board	17	60
PVC	Granular raw material, Piping materials, Floor materials	7	25
ABS	Granular raw material, Electrical parts, Sporting goods materials	10	50
PA	Granular raw material, Plastic cable ties, Ropes, Gloves	22	90
TPE	Granular raw material, Insoles, Adhesive tape	7	50
TPU	Granular raw material, Thin film, Mobile phone case, Basketball shoe parts	9	50
TPR	Granular raw material, Toy materials, Sports equipment, Luggage wheels	9	50
TPV	Granular raw material, Plastic handle, Plastic pipe	8	50
Total		140	625

Data preprocessing. We processed all acquired Raman spectra in accordance with the following three steps using WITec Project 5.0 software. Firstly, the background of every spectrum is subtracted by fitting a 4-order polynomial to the masked spectrum (all Raman peak areas were removed from the mask). After that, the baseline-corrected Raman spectra are extracted for further processing. Secondly, all extracted spectra are smoothed using the Savitzky–Golay linear filter with the 9-point window (the first order). Finally, all Raman spectra are normalized to unit length.

Principal component analysis. After preprocessing, principal component analysis is used for data dimensional reduction and extraction of the characteristic parameters. By using an orthogonal transformation, it can convert the original data into a set of linearly uncorrelated variables, identified as the principal components (PCs). PCA can transform high-dimensional feature input variables into low-dimensional output variables and compress the original data while keeping as much information as possible. PCA is a commonly used data dimensionality reduction method; especially for high-dimensional data, a small low-dimensional space can be used to reflect the original high-dimensional space. Here, the Kaiser rule is used to determine how many PCs should be retained.

The Kennard Stone method. In order to avoid overfitting, the Kennard Stone method is used to randomly divide the calibration and validation sets at a ratio of 4:1 [28]. The algorithm aims to select a diverse set of samples for the training set. Firstly, two samples were found with the longest Euclidian distance and were put into the training set. Then, we separately calculated the Euclidean distance between each remaining sample and each known sample in the training set, and found two samples with the maximum and minimum Euclidean distance and put them into the training set. Finally, the aforementioned steps were repeated until we reached the required number of samples.

Machine learning (ML). Three classification models of plastic Raman spectral data were established based on RF, the SVM algorithm with a five-fold cross-validation strategy, which was used for the model's evaluation, and BPNN, respectively. These three classical and widely used algorithms were selected to compare the prediction performance of different branches of machine learning in this study [29].

RF is a nonlinear machine learning model based on decision trees. It is an ensemble learning algorithm that integrates multiple decision trees to complete the prediction [30]. The prediction result of RF on the classification problem is the vote of all the prediction results of the decision tree. The training set of each decision tree is set by the Bootstrap sampling method. In addition, the features used in training each node of each decision tree are also partial features sampled from all eigenvectors. Therefore, an RF can classify multidimensional data and eliminate overfitting to some extent. Because the decision tree is used as the weak learner, RF has the advantage of less computation and simple implementation. At the same time, since the RF algorithm is based on the decision tree algorithm, the choice of the decision tree algorithm has an important impact on the accuracy of the prediction model established by the RF algorithm. In this study, the class and regression tree (CART) is determined as the decision tree algorithm of the RF model. Compared with the traditional ID3 algorithm and C4.5 algorithm, the Gini coefficient is a characteristic of the CART algorithm, which has the advantage of small computation and high efficiency. In addition, the number of decision trees has a great impact on the performance of the RF classification model. On the basis of the CART

algorithm, the number of decision trees is set to increase from 50 to 1,000 intervals. Considering the computational complexity and classification accuracy of the algorithm, 100 decision trees are selected to form the random forest, and the RF toolbox of MATLAB is used to implement the RF algorithm.

SVM is a supervised learning algorithm based on statistical learning theory and the minimum structural risk principle [31]. It shows many unique advantages in solving small sample nonlinear and high-dimensional pattern recognition problems, and has been widely used in statistical classification and regression analysis [32]. The type of kernel function has a great influence on the classification accuracy of the SVM model. As the kernel function of the radial basis function (RBF) has fewer parameters and is suitable for the classification of multidimensional data, the commonly used RBF is utilized as the kernel function of the SVM prediction model in this study. In addition, the selection of the penalty parameter C and the kernel function g has a major influence on the classification ability of SVM. We chose the widely used grid optimization to determine the parameters of the SVM model. The establishment and analysis of the SVM mo-del was performed in MATLAB using the LIBSVM 3.14 toolbox developed by Chang and Lin [33].

BPNN also known as multilayer perceptron, is an efficient feedforward network algorithm [34]. The main idea is that the signal propagates forward and the error propagates backward. By setting the hidden layer, it can realize nonlinear segmentation based on the original logistic regression. Determining the number of neurons in each layer is the first task when establishing a neural network prediction model. The number of neurons in the input layer and the output layer is determined according to the input characteristic variables and the classification categories. In addition, the selection of the activation function also plays a key role in the accuracy of the classification model. We chose the classical hyperbolic tangent function as the activation function of neurons. In this study, a commonly used three-layer neural network prediction model was built using the neural network toolbox in MATLAB.

Here, we used four statistical parameters to evaluate the performance of the model. Sensitivity (also known as recall), precision, and specificity are defined as follows:

$$\text{Sensitivity} = \frac{TP}{TP + FN} \times 100\% , \tag{1}$$

$$\text{Precision} = \frac{TP}{TP + FP} \times 100\% , \tag{2}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \times 100\% , \tag{3}$$

where TP is the true positive decision, i.e., correctly identified; FP is the false positive decision, i.e., incorrectly identified; TN is the true negative decision, i.e., correctly rejected; FN is the false negative decision, i.e., incorrectly rejected.

Accuracy is defined as a ratio of correctly classified samples to total samples. Thus,

$$\text{Accuracy} = \frac{TN + TP}{TP + TN + FN + FP} \times 100\% . \tag{4}$$

Another important parameter is the $F1$-score, which measures classification efficiency. The $F1$-score, also known as a balanced F-score, is defined as a balanced average of accuracy and recall. We had hoped that both the precision and recall would be very high, but in fact they were contradictory. The $F1$-score takes into account both the precision rate and the recall rate so that the two can reach the highest score at the same time and achieve a balance. The $F1$-score is defined as

$$F1 = 2 \times \frac{\text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \times 100\% . \tag{5}$$

Results and discussion. *Raman spectral analysis.* Figure 1 shows the comparison of the normalized average Raman spectra of 11 plastics. From the collected Raman spectra, it can be observed that there are certain differences in the Raman spectra of different types of plastics, with some different characteristic peaks which are the key to identifying different types of plastics. Plastics are polymers with analogous structures, so they present some Raman peaks similar to each other; therefore, an efficient classification algorithm is needed to identify them.

Main Raman peaks (cm^{-1}) of 11 plastics are follows: PE 1063, 1129, 1170, 1295, 1369, 1416, 1440, 1461, 2721, 2848, 2886; PET 632, 701, 797, 858, 999, 1100, 1116, 1182, 1295, 1416, 1461, 1614, 1731, 2907, 2965, 2998, 3081; PP 323, 399, 458, 529, 810, 842, 975, 1029, 1153, 1330, 1369, 1440, 1461, 2721, 2848, 2886, 2960; PS 620, 797, 1031, 1153, 1182, 1330, 1440, 1602, 2848, 2907, 2973, 3052; PVC 362, 636,

694, 1029, 1100, 1170, 1259, 1330, 1440, 1612, 1722, 2874, 2917, 2935, 3081; ABS 620, 999, 1031, 1153, 1308, 1440, 1585, 1602, 1666, 2854, 2907, 2998, 3058; TPE 620, 842, 999, 1153, 1295, 1440, 1602, 2721, 2848, 2886, 2917, 3052; TPU 632, 858, 975, 1182, 1440, 1602, 1727, 2874, 2917, 2949, 3052; TPR 620, 797, 999, 1031, 1440, 1585, 1602, 1666, 2848, 2907, 2998, 3052; TPV 399, 810, 842, 975, 1063, 1153, 1295, 1369, 1440, 2721, 2848, 2886, 2917, 2935; PA 1063, 1129, 1295, 1381, 1440, 1461, 2874, 2917.

Raman spectroscopy essentially reflects the vibrational mode of organic molecules. Table 2 lists the tentative assignments of the main Raman peaks found in plastics [21].

TABLE 2. Raman Peaks Assignment

Peak positions, cm^{-1}	Assignment
362	C–C bending
636, 694	C–Cl stretching
620, 632, 701, 797, 810, 975, 1029, 1063, 1100, 1129, 1153, 1170 1182	C–C stretching-
842	CH_3 bending
858	1,4-disubstituted benzene ring
999	The triangular ring of benzene breathes
1031	In-plane bending in benzene ring
1116, 1602, 1585, 1614	C–C stretching in benzene ring
1259, 1295, 1330, 1369, 1381, 1416, 1440, 1461	CH_2 bending
1666	C=C symmetric stretching
1731	C=O stretching
2721, 2848, 2874, 2886, 2917, 2935, 2949, 2965, 2973, 2998	CH_2 stretching
2907	CH stretching
2960	CH_3 stretching
3052, 3081	CH stretching in benzene ring

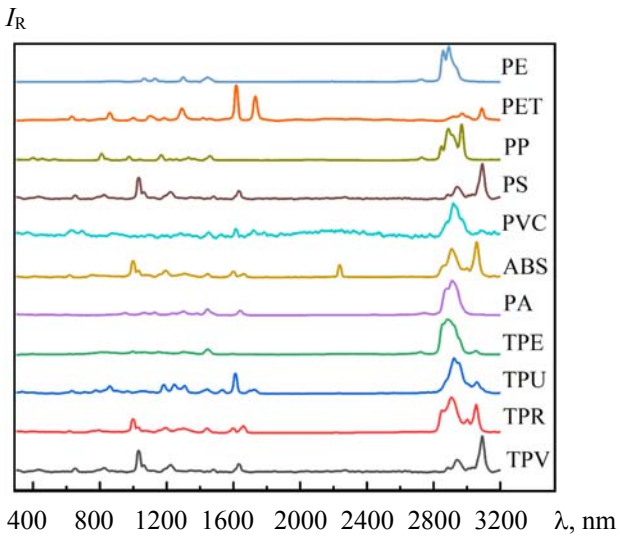

Fig. 1. Comparison of the normalized mean Raman spectra of 11 plastics.

Feature extraction. We calculated 625 spectra and selected the first three PCs for the preliminary analysis of the classification of plastics: 44.99, 18.61, 14.54, 7.89, 3.02, 2.32, 1.45, 1.17. The contribution rates of the first three PCs are 44.99, 18.61, and 14.54%, respectively, and the cumulative contribution rate is 78.14%. The three-dimensional scatter plots of the first three PCs are shown in Fig. 2. It can be seen that PE, ABS, TPU, and TPR are clearly distinguished from each other, and the other seven types of plastics have overlapping areas. This indicates that the selection of the PC quantity plays an important role in the accuracy of the prediction model. The contribution of each PC and the cumulative contribution plotted are shown in Fig. 3. The eigenvalues of the first four PCs are larger than 1, and their cumulative contribution rate

reaches 86.03%. Considering the Kaiser rule, the first four principal components were chosen. The cumulative contribution of other principal components does not make for a lot of contribution, suggesting that it is impossible to visualize the classification of plastics simply by constructing a three-dimensional (3D) scatter plot.

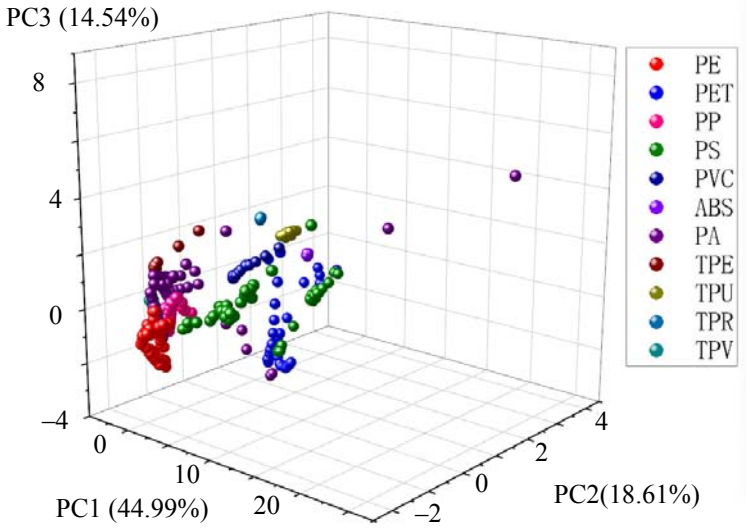

Fig. 2. Three-dimensional scatter plot of the first three PCs.

Machine learning (ML) classification. The four PCs extracted from PCA features were used to establish classification models based on three ML algorithms (RF, SVM, and BPNN). To avoid overfitting, the Kennard Stone method was used to randomly divide the calibration and validation sets at a ratio of 4:1.

Firstly, the CART algorithm as the basic decision tree algorithm is chosen, and set up with 100 decision trees to establish a prediction model of the RF. Secondly, the RBF kernel function combined with the grid optimization algorithm to establish the SVM classification model is selected. In order to avoid overfitting, the calibration set of the SVM algorithm adopts the five-fold cross-validation strategy to evaluate the model's performance. The range of the grid optimization search parameters C and g are [–2, 4] and [–4, 6]. Then a three-layer BP neural network with the hyperbolic tangent activation function is selected to establish the classification model of the BPNN algorithm. Since the first four principal components with a cumulative contribution rate of 86.03% are selected for feature extraction, as shown in Fig. 3, the number of nodes in the input layer and output layer mentioned earlier is related to the dimension and category of the data, so the number of input nodes is set to four. Considering the values calculated by the three empirical formulas for calculating the number of hidden layer nodes, combined with the complexity of the network structure and the size of the error

$$l = n - 1, \tag{6}$$

$$l = \sqrt{m+n} + a, \tag{7}$$

$$l = \log_2 n, \tag{8}$$

(where n, l, m are the numbers of nodes in the input layer, hidden layer, and output layer, respectively, a is the adjustment constant, and the value range is generally 0–10), the node deletion method and the expansion method are used to initially determine the number of hidden layer nodes to be 3–16, and then in each case, model evaluation is performed 20 times under each node, and finally 13 is selected as the number of hidden layer nodes, with the highest accuracy rate of 95.04%. In order to improve the accuracy of classification, a binary code is applied to 11 plastic samples, thus the number of neurons in the output layer is determined to be four. Finally, after repeated research and training samples, other training parameters of the neural network to obtain more accurate classification results are determined. The parameters of the BP neural network are follows: input layer neurons 7; output layer neurons 4; hidden layer neurons 13; training error 0.01; maximum iterations 1000.

In order to more accurately compare and increase the reliability of experiments, a different calibration and validation set is selected each time, and all the algorithms are run 20 times. Table 3 shows the overall performance (calibration, validation, and times for the classification of the entire test dataset) of the predic-

tive model using the three ML algorithms. The results show that all three algorithms have good prediction models. Those based on the three ML algorithms show good accuracy, indicating that the ML algorithms can classify plastic samples more correctly. The SVM algorithm performs best in the calibration and validation set. The average classification accuracy reached 99.58 and 98.5% for the calibration and validation set, respectively. At the same time, the SVM algorithm is proven to have efficient classification performance for small sample data. On the other hand, the RF algorithm has the lowest average CPU time, requiring only 0.04s, which also proves its fast training ability and low computational cost. For the BPNN algorithm, due to the small number of samples, its prediction model accuracy and CPU running time are not the best; however, in general, its classification performance is also good, especially regarding the stability of the calibration set. If a large number of plastic samples need to be trained and classified, the BPNN algorithm is a good choice.

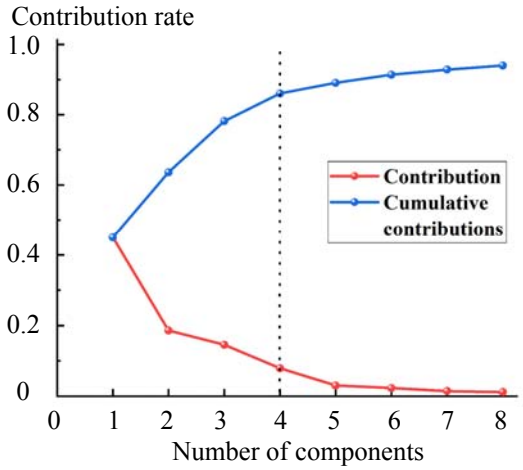

Fig. 3. The contribution rate varies with the number of PCs.

TABLE 3. Algorithm Accuracy among RF, SVM, and BPNN Models
(run 20 times with four components)

Algorithm	Calibration, %	Validation, %	Time, s
RF	98.12±1.30	97.92±1.52	0.04±0.01
SVM	99.60±2.83	98.56±0.88	5.53±0.53
BPNN	95.04±1.64	93.28±2.30	0.93±0.58

N o t e s. The setup of computer: CPU AMD R5 2600x; Memory 16 GB; Graphics MSI GTX1060 (6 GB).

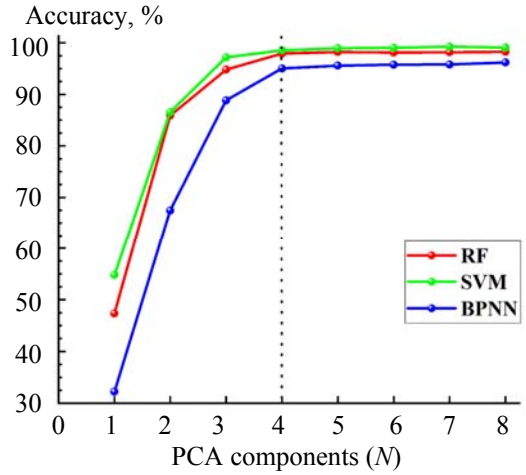

Fig. 4. Comparison of the prediction accuracy of different machine learning algorithms.

Figure 4 shows the average accuracy of the three ML algorithms under a different number of principal components. It can be clearly seen that the average accuracy of each algorithm is high enough when the first four principal components are selected as feature input variables, but when more principal components are selected, the average accuracy is seldom improved. Therefore, the first four principal components are selected to establish a prediction model to reduce the computational complexity of spectral data while maintaining high classification accuracy. This advantage will be more evident for prediction model establishment from large sample data.

For multiclass classification, in order to evaluate the specific classification performance of each category, the sensitivity, specificity, and *F*1-scores of 11 types of plastics are calculated as shown in Table 4. The *F*1-score is one of the equilibrium evaluation parameters, but the other three parameters can also reflect the performance of the relevant classification model. In general, the higher the *F*1-score value, the better the performance of the classifier and the higher the reliability of the classifier. It can be seen from Table 4 that PET and PA have the lowest *F*1-scores and the lowest sensitivity. All the other nine classes have high *F*1-scores for three prediction models. The reason for that may be that various strong Raman peaks from additives are observed in the PET and PA samples. For example, PA with fiberglass additives presents strong Raman signals at 935 and 1,635 cm^{-1}, while pure PA does not have these two peaks. To some extent, this indicates that the ML algorithms combined with Raman spectroscopy can have a good classification effect on plastics.

TABLE 4. Calculation of Plastic Statistical Parameters under Different Machine Learning Algorithms

Algorithm	Plastic type	Sensitivity, %	Specificity, %	Accuracy, %	*F*1-score, %
RF	PE	100.00	100.00	100.00	100.00
	PET	93.50	99.57	99.08	94.06
	PP	99.67	100.00	99.96	99.83
	PS	97.50	99.78	99.56	97.71
	PVC	100.00	99.58	99.60	95.53
	ABS	100.00	100.00	100.00	100.00
	TPE	99.00	99.87	99.80	98.76
	TPR	100.00	99.74	99.76	98.57
	TPU	100.00	99.96	99.96	99.76
	TPV	100.00	99.83	99.84	99.05
	PA	91.67	99.39	98.28	93.82
SVM	PE	100.00	100.00	100.00	100.00
	PET	95.50	99.35	99.04	93.96
	PP	100.00	100.00	100.00	100.00
	PS	100.00	99.87	99.88	99.44
	PVC	100.00	99.96	99.96	99.55
	ABS	100.00	100.00	100.00	100.00
	TPE	99.50	100.00	99.96	99.74
	TPR	100.00	100.00	100.00	100.00
	TPU	100.00	100.00	100.00	100.00
	TPV	100.00	99.65	99.68	98.22
	PA	92.78	99.58	98.60	94.98
BPNN	PE	99.33	99.27	99.27	97.23
	PET	86.50	98.91	98.91	87.24
	PP	99.67	97.86	97.86	92.92
	PS	91.25	99.78	99.78	94.29
	PVC	93.00	100.00	100.00	95.97
	ABS	100.00	99.74	99.74	98.59
	TPE	100.00	99.61	99.61	97.88
	TPR	100.00	99.83	99.83	99.09
	TPU	100.00	99.48	99.48	97.27
	TPV	100.00	98.91	98.91	94.32
	PA	77.22	99.72	96.48	85.70

Conclusions. Raman spectra of 11 plastics are obtained rapidly by Raman spectroscopy. We apply machine learning methods to achieve good classification since the characteristic peaks in the plastic spectrum are partially overlapped. We use principal component analysis to preliminarily analyze the spectral data and select the optimal principal component quantity to establish the prediction model of the three machine learning algorithms. The high accuracy of the plastic classification model calibration set based on three machine learning algorithms is achieved. This study shows that 11 types of plastics can be well classified by Raman spectroscopy combined with machine learning algorithms. The fingerprint characteristics of Raman spectroscopy and the automation of the machine learning algorithms allow for prompt classification with high reliability. Automatic sorting of waste plastics can be achieved in the process of recycling. In addition, this work provides a scientific basis for establishing the Raman spectrum database of plastics and performing the quantitative analysis of mixed plastics.

This work was supported by the National Natural Science Foundation of China (grant No. 61405083) and the Fundamental Research Funds for the Central Universities (grant No. 2018-129).

REFERENCES

1. A. L. Andrady, M. A. Neal, *Philos. Trans. Roy. Soc. B*, **364**, 1977–1984 (2009).
2. R. C. Thompson, S. H. Swan, C. J. Moore, et al., *Philos. Trans. Roy. Soc. B*, **364**, 1973–1976 (2009).
3. R. Geyer, J. R. Jambeck, K. L. Law, *Sci. Adv.*, **3**, No. 7, e1700782 (2017).
4. ICIS, Chemical, Business, Plastics Europe (2019).
5. D. K. Barnes, F. Galgani, R. C. Thompson, et al., *Philos. Trans. Roy. Soc. B*, **364**, 1985–1998 (2009).
6. C. M. Rochman, M. A. Browne, B. S. Halpern, et al., *Nature*, **494**, 169–171 (2013).
7. S. Gregoire, M. Boudinet, F. Pelascini, et al., *Anal. Bioanal. Chem.*, **400**, 3331–3340 (2011).
8. Y. V. Vazquez, S. E. Barbosa, *Waste Manage*, **53**, 196–203 (2016).
9. Q. Wang, X. Wu, L. Chen, et al., *Appl. Spectrosc.*, **71**, No. 11, 2538–2548 (2017).
10. S. Pongstabodee, N. Kunachitpimol, S. Damronglerd, *Waste Manage*, **28**, No. 3, 475–483 (2008).
11. R. Leitner, H. Mairer, A. Kercek, *Real-Time Imag.*, **9**, No. 4, 245–251 (2003).
12. Q. Zhao, M. Chen, *Anal. Lett.*, **48**, No. 2, 301–307 (2014).
13. Y. Yu, L. B. Guo, Z. Q. Hao, et al., *Opt. Express*, **22**, No. 4, 3895–3901 (2014).
14. R. Lenz, K. Enders, C. A. Stedmon, et al., *Mar. Poll. Bull.*, **100**, No. 1, 82–91 (2015).
15. B. K. Chabuka, J. H. Kalivas, *Appl. Spectrosc.*, **74**, No. 9, 1167–1183 (2020).
16. V. K. Unnikrishnan, K. S. Choudhari, S. D. Kulkarni, et al., *RSC Adv.*, **3**, 48 (2013).
17. E. Tognoni, G. Cristoforetti, *Opt. Laser Technol.*, **79**, 164–172 (2016).
18. Y. Xi, Y. Li, Z. Duan, et al., *Appl. Spectrosc.*, **72**, No. 12, 1752–1763 (2018).
19. J. Wang, K. Liu, S. Jin, et al., *Appl. Spectrosc.*, **74**, No. 2, 130–159 (2020).
20. W. Cowger, A. Gray, S. H. Christiansen, et al., *Appl. Spectrosc.*, **74**, No. 9, 989–1010 (2020).
21. H. S. Chen, *Plast. Sci. Technol.*, **40**, No.6, 69–72 (**2012**).
22. V. Allen, J. H. Kalivas, R. G. Rodriguez, *Appl. Spectrosc.*, **53**, No. 6, 672–681 (1999).
23. A. Tsuchida, H. Kawazumi, A. Kazuyoshi, et al., In: *Sensors, IEEE Christchurch*, New Zealand, October 25–28, 2009, 1473–1476 (2009).
24. I. T. Jolliffe, J. Cadima, *Philos. Trans. Roy. Soc. A*, **374**, 20150202 (2016).
25. D. M. Camacho, K. M. Collins, R. K. Powers, et al., *Cell*, **173**, No. 7, 1581–1592 (2018).
26. C. Berghian-Grosan, D. A. Magdas, *Talanta*, **218** (2020).
27. Y. Li, R. Shen, H. Wu, et al., *Spectrochim. Acta A*, **225**, 117483 (2020).
28. R. W. Kennard, L. A. Stone, *Technometrics*, **11**, No. 1, 137–148 (1969).
29. I. M. Nolasco Perez, A. T. Badaro, S. J. Barbon, et al., *Appl. Spectrosc.*, **72**, No. 12, 1774–1780 (2018).
30. A. L. Boulesteix, S. Janitza, J. Kruppa, et al., *Wires Data Min. Knowl.*, **2**, No. 6, 493–507 (2012).
31. Y. Yu, Y. Lin, C. Xu, et al., *Biomed. Opt. Express*, **9**, No. 12, 6053–6066 (2018).
32. D. Wang, J. Jiang, J. Mo, et al., *Appl. Spectrosc.*, **74**, No. 6, 674–683 (2020).
33. C. C. Chang, C. J. Lin, *ACM Trans. Intell. Syst. Technol.*, **2**, No. 3, 1–27 (2011).
34. N. M. Nawi, R. S. Ransing, M. N. B. M. Salleh, et al., In: *Proc. Database Theory and Application, Bio-Science and Bio-Technology*, Jeju Island, South Korea, December 13–15, 177–188 (2010).