ELECTRONIC SUPPORTING INFORMATION

Mid-Infrared spectroscopy and machine learning for postconsumer plastics recycling

Nicholas Stavinski,[a] Vaishali Maheshkar,[b] Sinai Thomas,[a] Karthik Dantu,*[b] and Luis Velarde*[a]

* Corresponding Author

[a] Department of Chemistry, University at Buffalo, State University of New York, Buffalo, New York, 14260, United States.

[b] Department of Computer Science and Engineering, University at Buffalo, State University of New York, Buffalo, New York, 14260, United States.

Table of Contents

Section 1. Plastics Database

Table S1. Database sample descriptions and totals. Overview of database plastics used in this study. Samples are organized by their resin code, appearance, and color. Images depict the color and transparency for each type of plastic object. **Not all individual plastics samples are shown in Table S1. Only a representative image is shown for clarity. Full image library may be available upon request**

Plastic Type	Object	Color	Total	Image
		Black/dark	2	
		Blue	1	
		Brown	2	
		Green	1	
		Grey	1	
		Opaque	19	
		Purple	1	
HDPE	Bottle	Red	1	

		White	51	
		Yellow	3	
	Container/cup /bowl/plate	Opaque	1	
		White	3	
	Film/wrapping /bag	Clear	45	
		Opaque	6	
		White	6	
	Lid/cap	Black/dark	1	
		Blue	1	
		Clear	1	
		Opaque	2	
		Orange	5	

		White	14	
	Bottle	Yellow	4	
	Container/cup /bowl/plate	Red	1	
		Brown	1	
		Clear	112	
		Green	2	
		Greenish blue	11	
		Opaque	2	
		Reddish orange	1	
		White	26	
	Film/wrapping /bag	Yellow	1	
LDPE	Lid/cap	Clear	3	

		White	1	
		Yellow	1	
		Brown	1	
		Clear	61	
	Bottle	White	1	
		Blue	1	
	Container/cup /bowl/plate	Clear	72	
	Hard packaging mold	Clear	8	
PET	Lid/cap	Clear	23	
		Opaque	1	
	Bottle	Orange	3	
PP	Container/cup /bowl/plate	Black/dark	23	

		Clear	29	
		Opaque	17	
		White	61	
	Film/wrapping /bag	Opaque	1	
	Hard packaging mold	Black/dark	1	
		Blue	8	
		Clear	11	
		Green	2	
		Opaque	2	
		Red	3	
	Lid/cap	White	5	
PS	Container/cup /bowl/plate	Black/dark	15	

		Blue	2	
		Brown	15	
		Clear	21	
		Green	15	
		Opaque	16	
		Pink	15	
		Red	10	
		White	40	
	Flat shape/eating utensil	Clear	3	
		White	5	
	Hard packaging mold	Blue	1	
		Orange	1	

		Black/dark	3	
		Brown	4	
	Lid/cap	Opaque	1	

Section 2. ATR-FTIR Spectra

FigureS1a-e: ATR-FTIR spectra corresponding to machine learning results in **Table 1** and **Figure 2**. These spectra are most reflective of the true chemical composition of mixed postconsumer plastic waste. No baseline corrections (accounting for CO_2 or H_2O), normalizations, or other manipulations were performed.

Figure S1a. ATR-FTIR spectra of #1 PET postconsumer plastics (red) and a virgin polymer for reference (black). 501 spectra are shown.
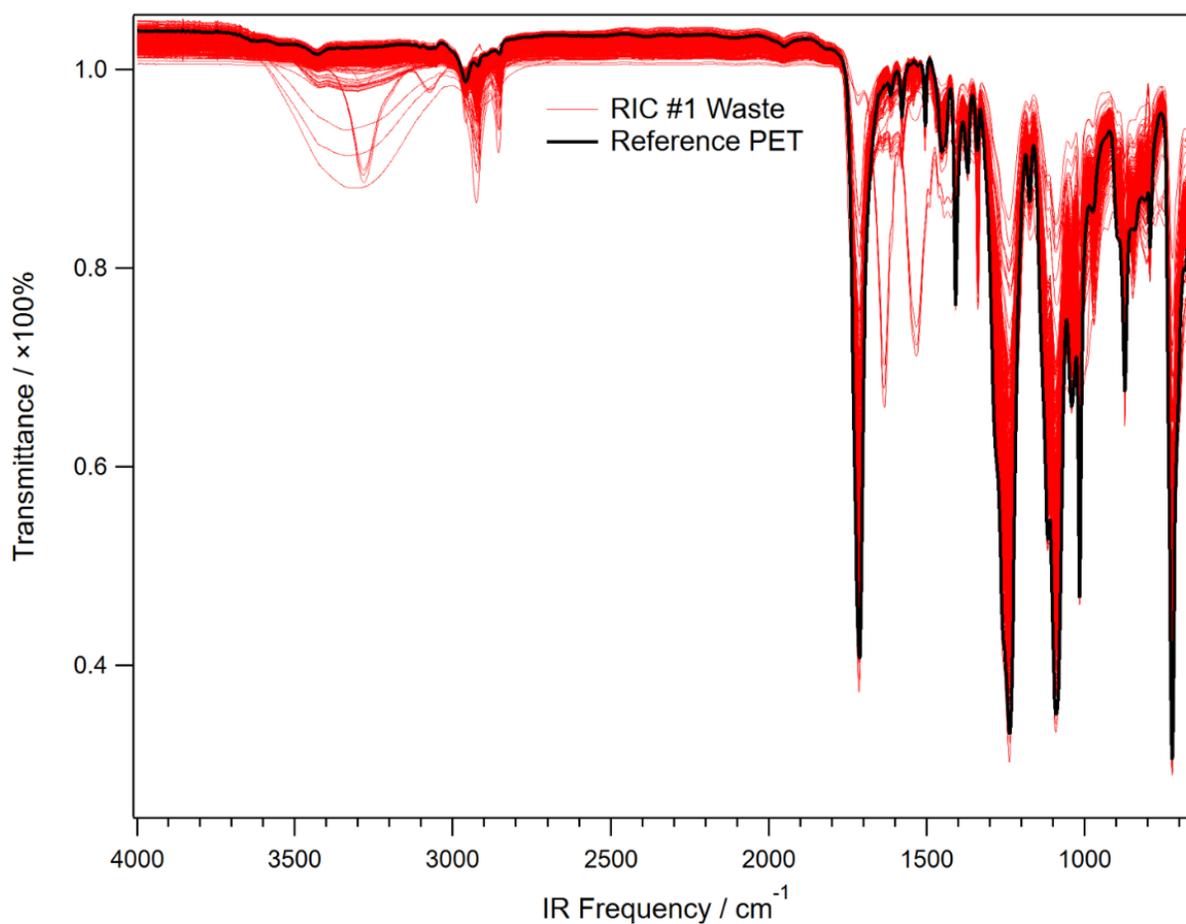

Fig. S1b ATR-FTIR spectra of #2 HDPE postconsumer plastics (red) and a virgin polymer for reference (black). 501 spectra are shown.
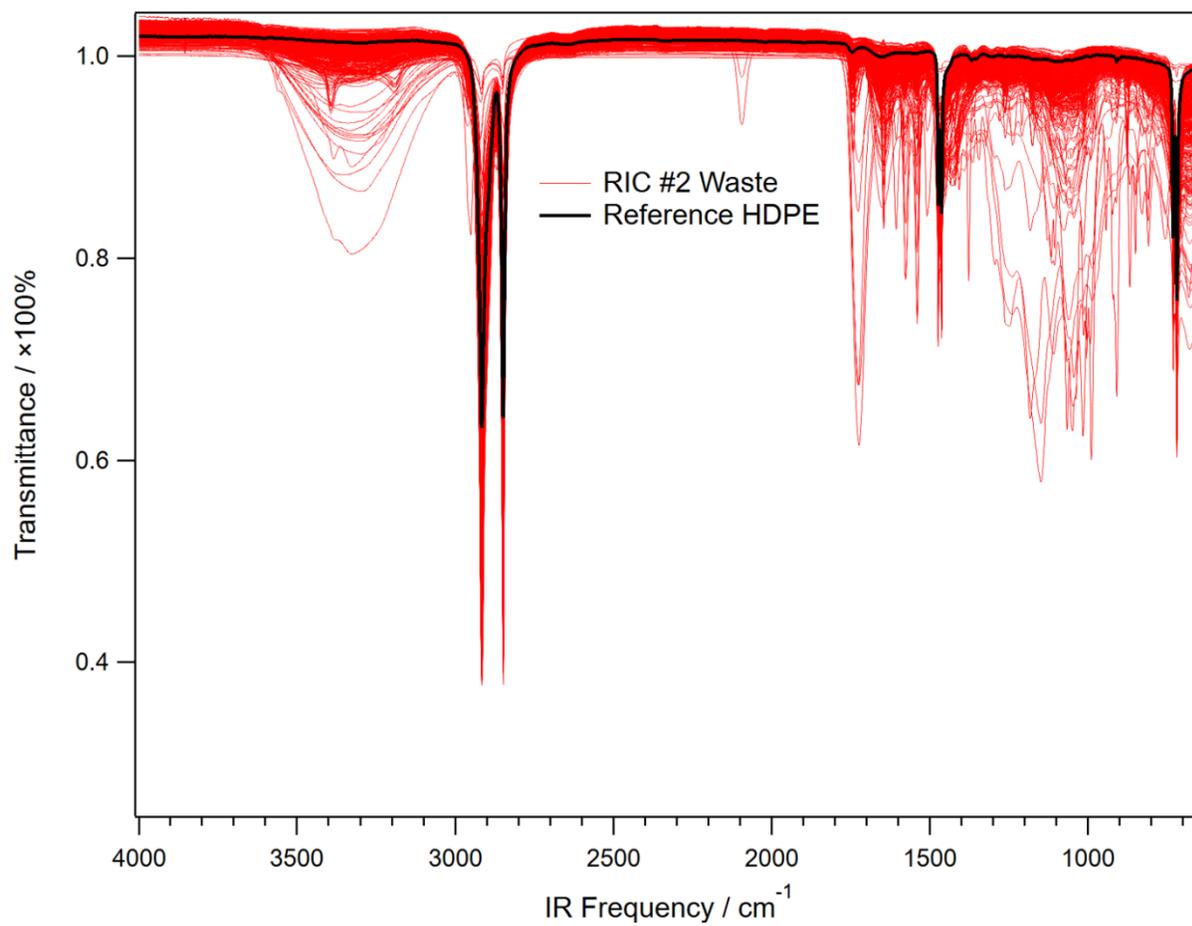
Fig. S1c ATR-FTIR spectra of #4 LDPE postconsumer plastics (red) and a virgin polymer for reference (black). 501 spectra are shown.
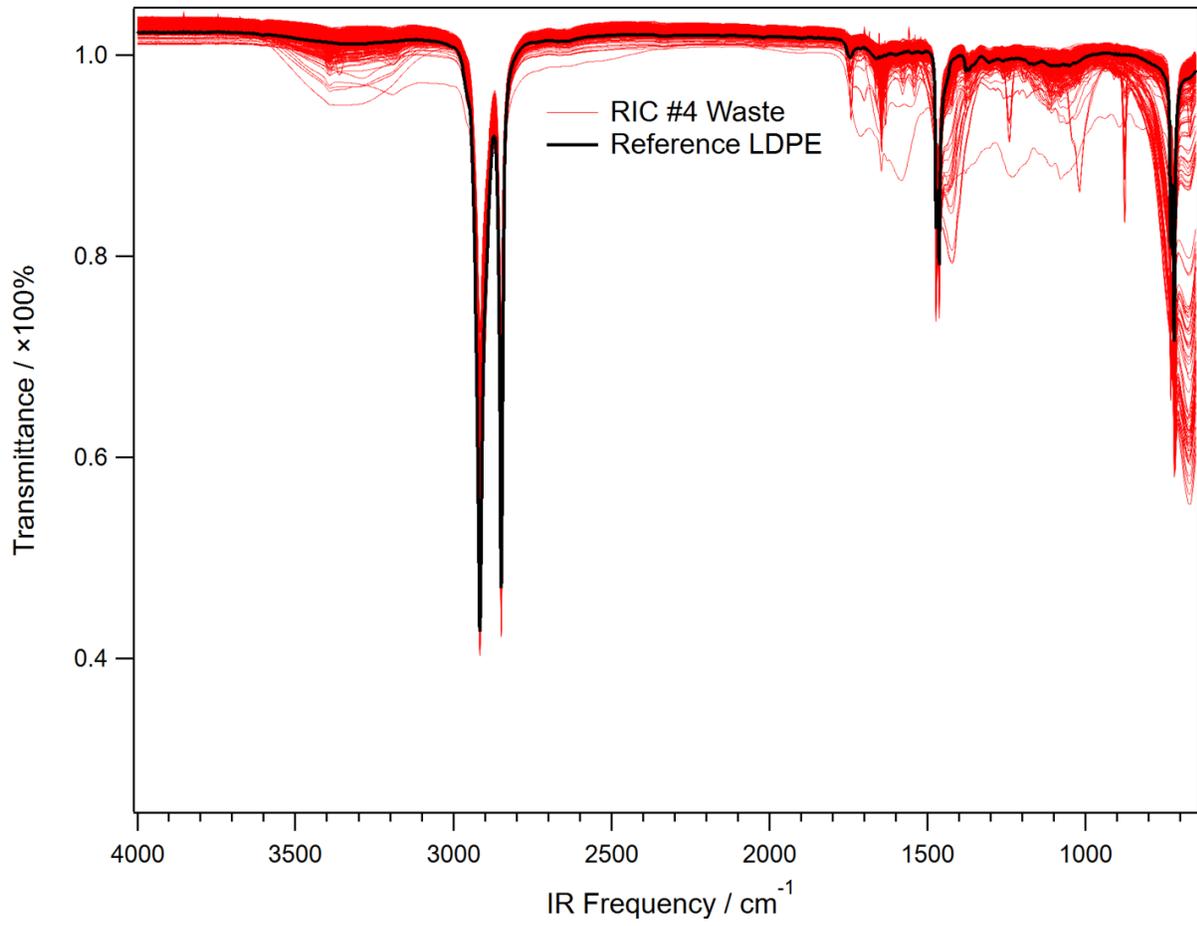
Fig. S1d ATR-FTIR spectra of #5 PP postconsumer plastics (red) and a virgin polymer for reference (black). 501 spectra are shown.
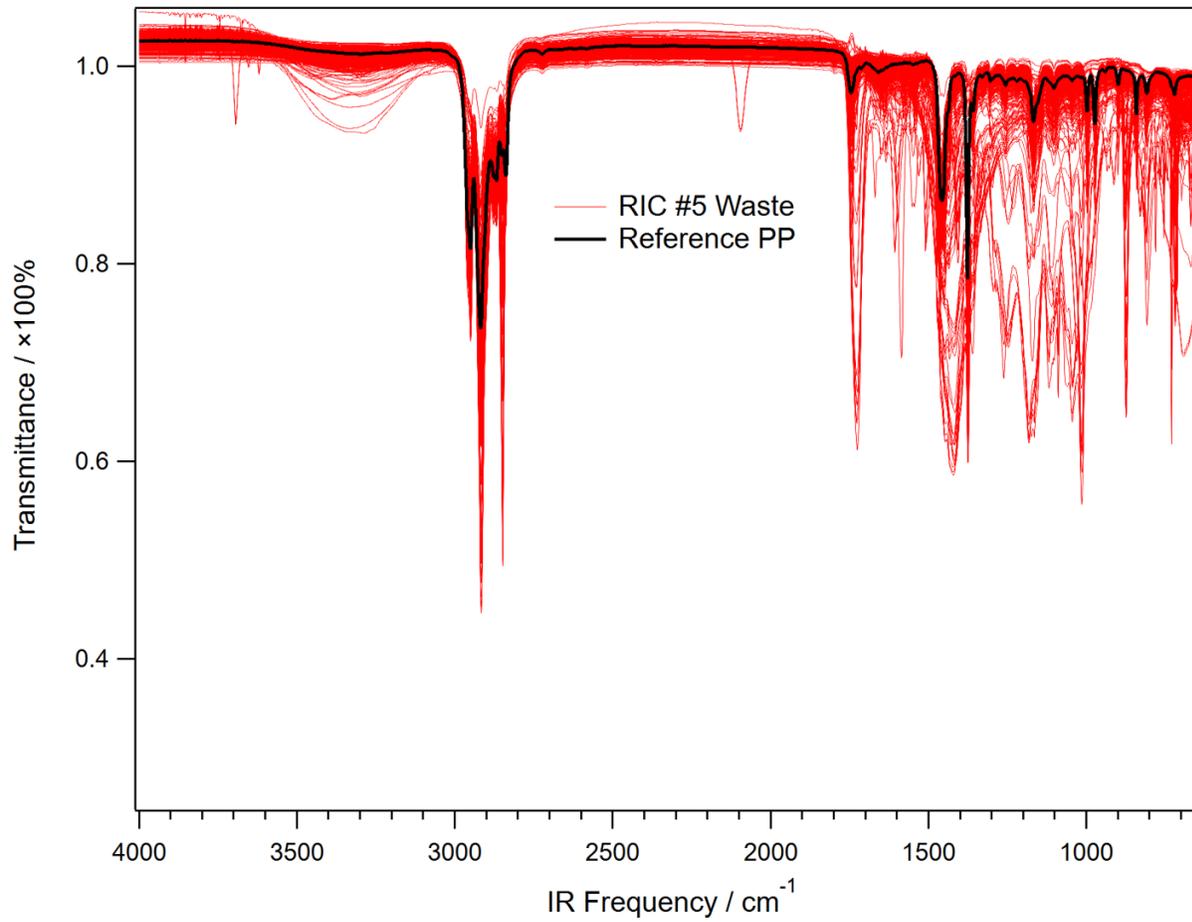
Fig. S1e ATR-FTIR spectra of #6 PS postconsumer plastics (red) and a virgin polymer for reference (black). 501 spectra are shown.
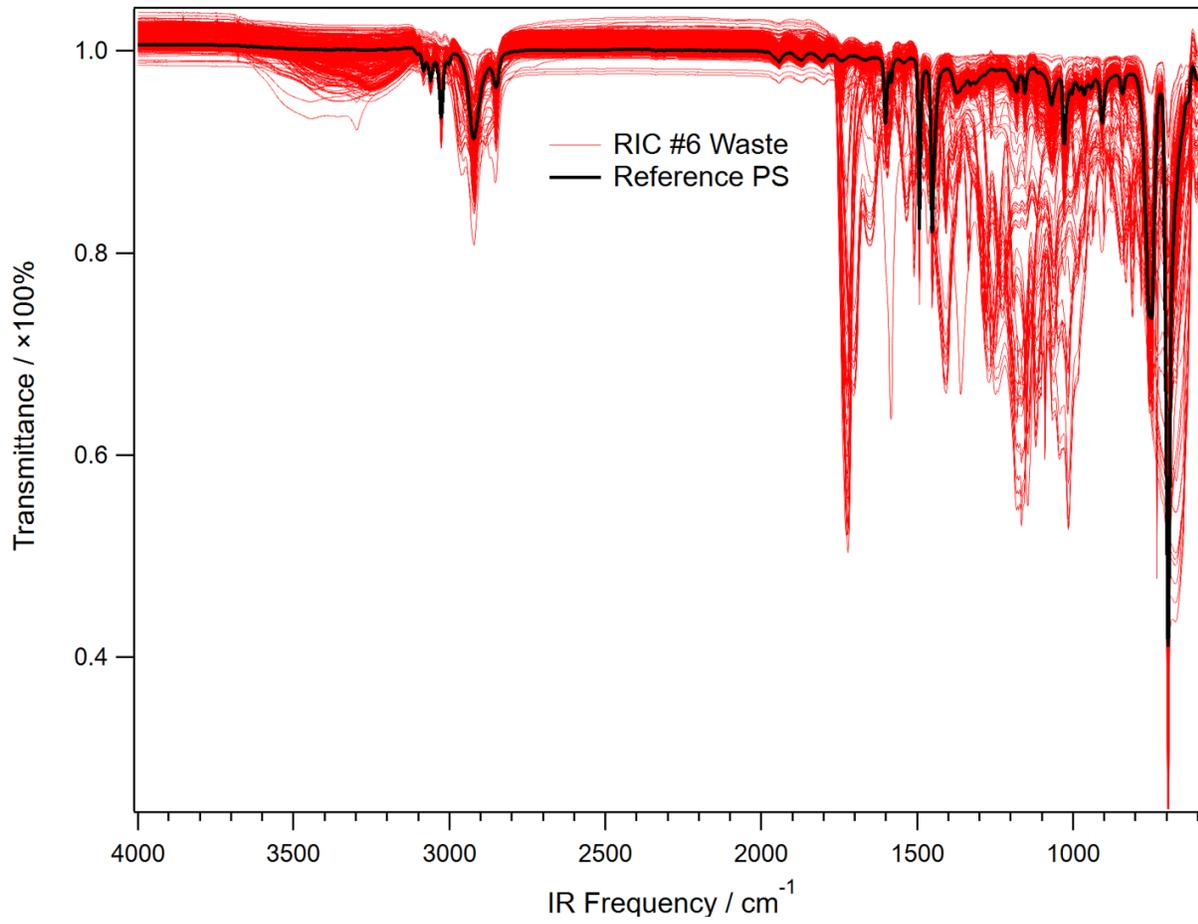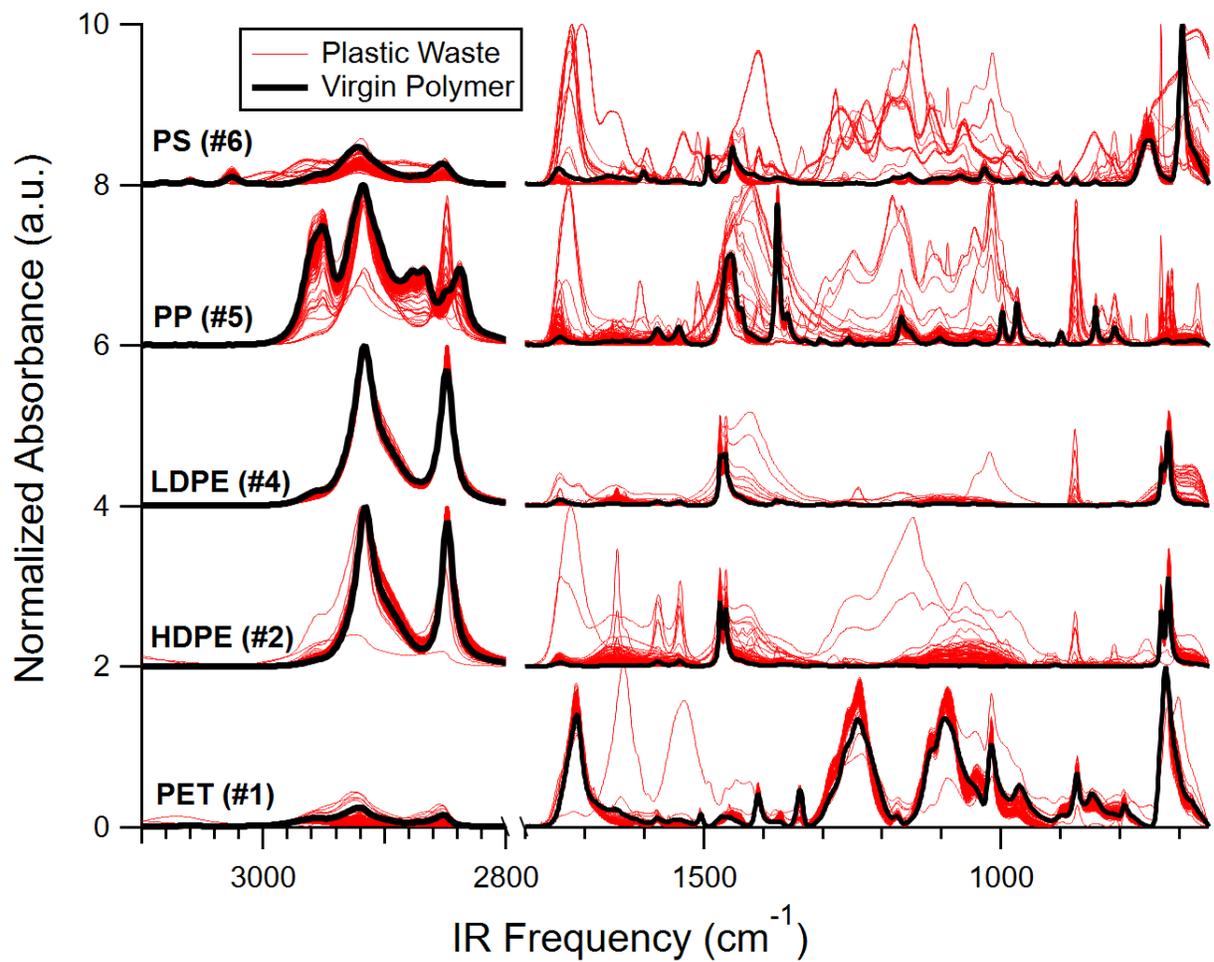
Fig. S2 Processed spectra (Methods 2.2) of MIR database plastics. Plastic waste (red). Virgin reference polymer (black). The x-axis is split for observation of the regions-of-interest. Only 1/3rd of the database spectra are shown for clarity.

Section 3. Machine Learning

IMPORTANT DEFINITIONS:

"<u>Unprocessed Spectra</u>" = Spectra were acquired from 4000 to 650 cm^{-1} as percent transmittance spectra without baseline correction (10 iterations, 64 baseline points, and excluded CO_2 bands), normalization, or conversion to Absorbance. These acquisitions provide the most realistic glimpse of chemical heterogeneity within mixed postconsumer plastics because they include O-H and C-O mid-infrared bands corresponding to not only water and carbon dioxide, respectively, but also polymer additives.

"<u>Processed Spectra</u>" = Spectra were acquired from 4000 to 650 cm^{-1} and processed using OPUS 7.5. Each spectrum contains 3474 data points, where each point represents the intensity in percent transmittance (Fig. 2) or absorbance units (Fig. 4) at a given wavenumber. The raw spectra were processed by converting from percent transmittance to absorbance, applying a concave rubberband baseline correction (10 iterations, 64 baseline points, and excluded CO_2 bands), and performing a minimum/maximum normalization. These spectra are the "cleanest" in this study, but it should be understood that achieving results using this processing method would contribute additional mathematical steps in a real-world, practical mid-infrared sorting technology (e.g., adding additional time and computational load to the process).

3a. Autoencoders

This study's dataset consists of 2,505 samples and each sample has 3,474 features. We have used the Multilayer Perceptron autoencoder model. We have used the Keras functional API for defining the model. The encoder is defined to have 2 hidden layers, the first layer has two times the number of inputs (6,948) and the second layer with the same number of inputs (3,474). The bottleneck layer has the same number of inputs (3,474). The decoder has a similar structure in reverse, it has 2 hidden layers, the first layer has the number of inputs in the dataset (3,474) and the second layer has double the inputs (6,948). The output layer will have the same number of output nodes as the input data which is 3,474. Linear activation is used to output numeric values. Batch normalization and leaky ReLU activation is used by the model for learning. Adam optimizer and mean squared error loss function are used.

Fig. S3 Learning curve of autoencoder model applied to full MIR (4000-650 cm^{-1}) of all 5 plastic types. Y-axis is loss that approaches zero over increased epochs (x-axis). Model shows sufficient fit for reconstruction.
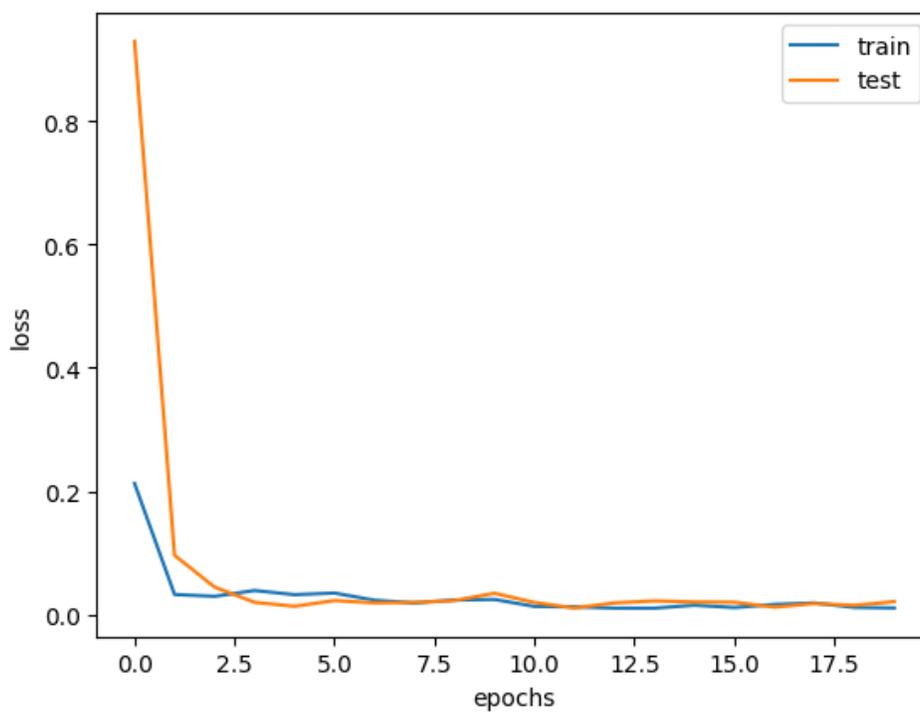
i.   **ML Autoencoders Results using Unprocessed Spectra**

See manuscript **Table 1.** and **Fig. 3**

ii.   ML Autoencoders Results using Processed Spectra

Table S1b Machine learning performance of spectra pre-processed using autoencoders (Fig. S2)

Algorithm	Full Mid-Infrared (4000-650 cm^{-1})		C-H Stretching (2990-2820 cm^{-1})		Fingerprint (1500-650 cm^{-1})	
	Accuracy (%)	τ (ms)	Accuracy (%)	τ (ms)	Accuracy (%)	τ (ms)
RF	100	125.807	100	70.001	100	53.144
KNN	98.842	3548.462	97.684	428.233	99.161	876.026
SVM	95.807	0.058	93.450	0.050	94.728	0.057
LR	96.526	99.298	95.248	5.424	96.366	26.249

Figure S4 Confusion matrices of ML models applied across five plastic types and the mid-infrared regions-of-interest. True and predicted label accuracies are highlighted in blue along the diagonal of each confusion matrix. Trained data were processed using the methods in 2.2 and 2.3 (Figure S2).
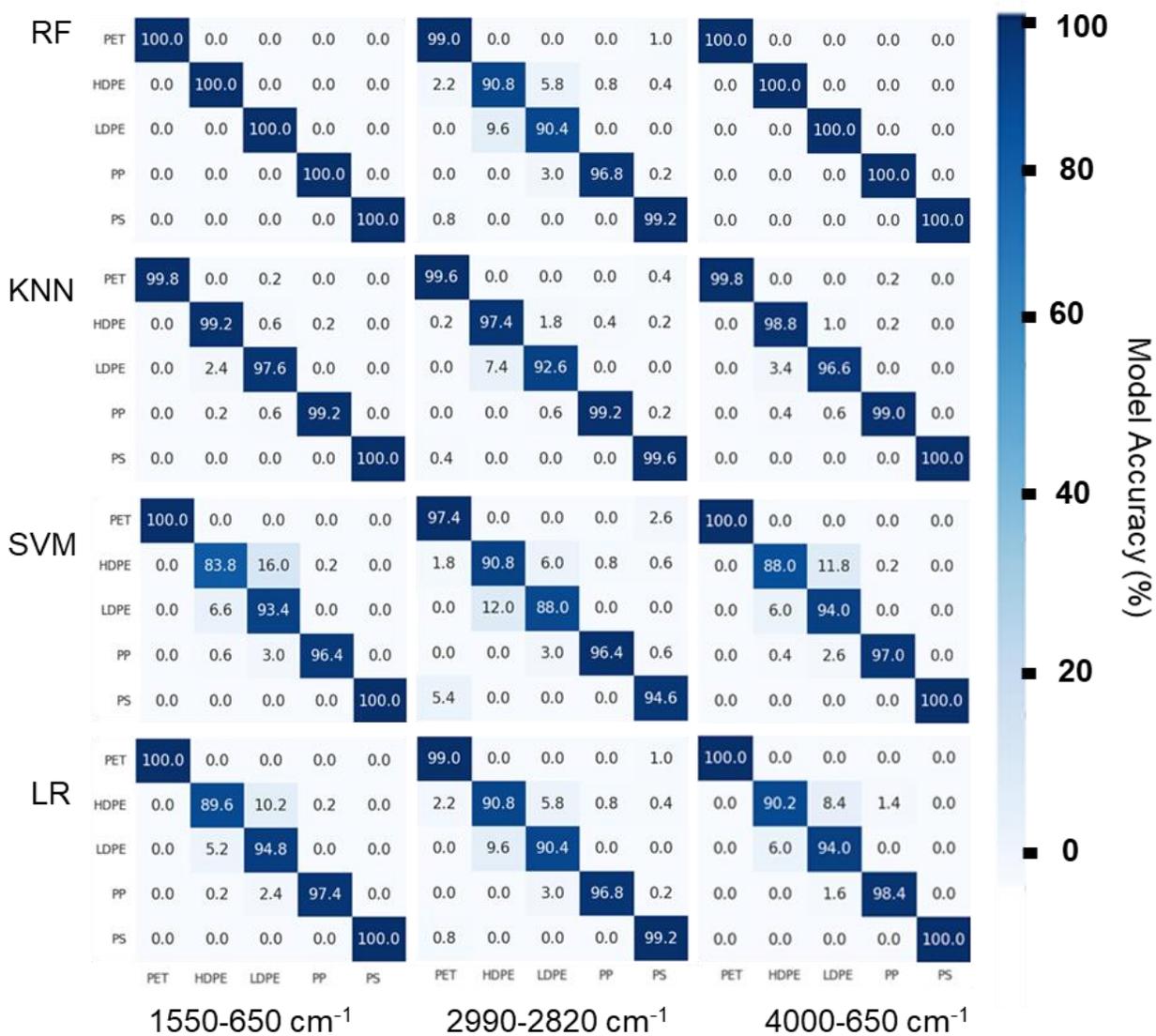
3b. Principal Component Analysis (PCA) was evaluated as a pre-processing technique on the spectral dataset and Random Forest algorithm was applied producing an accuracy of 98%.

3c. Standard Machine Learning Classifiers (without Autoencoders)

    i.   1D-CNN Implementation to the MIR Plastics Database

Convolutional neural networks (CNNs) are deep learning (DL) methods that have become of interest to the growing plastics recycling research community for their extensive learning architecture and ability to extract features from images for classification.[1-3] One-dimensional CNNs were applied to this study's database to assess the viability of DL techniques for plastic waste sorting. CNNs are traditionally used for high dimensional dataset such as images, whereas this dataset consists of 1-dimensional MIR spectra. Additionally, the performance of DL improves over increased dataset sizes. Relative to the size of conventional image datasets, the performance of DL was surpassed by the standard classifiers in this study (Table S5 and Figure S2). Future academic work should assess real-world plastics sorting systems using both standard classifiers and DL methods, as optimized sorting technologies that are later developed for industrial purposes may utilize a combination of automated ML schemes depending on the sample-of-interest and application.

Using a one-dimensional CNN (1D-CNN) architecture, spectral features were extracted for classification. The MIR spectra are 1-dimensional data files consisting of 3474 features and are converted to a 3474-element vector of features. The 1D-CNN architecture consists of 64 filters with kernel size of 3. These filters are convolved with the input vector to produce a scalar value which indicates the presence or absence of the pattern the filter is identifying. ReLu activation

function is a nonlinear function used to transform the summed weighted input from the node into the activation of the node or output for that input. Dropout of 0.5 is selected to drop the randomly selected neurons during training. Max pooling downsampled the input by calculating the maximum value for patches of feature maps thus reducing the dimensionality of the vector. To tune the performance of the model, the hyperparameters of the 1D-CNN architecture were applied according to the sklearn Application Programming Interface.

ii.   ML Results using Unprocessed Spectra

Table S1c Machine learning performance of unprocessed spectra. Spectra corresponding to these results are found in Figure S1a-e.

	$4000 - 650$ cm^{-1}		$2990 - 2820$ cm^{-1}		$1500 - 650$ cm^{-1}	
	Accuracy (%)	τ (ms)	Accuracy (%)	τ (ms)	Accuracy (%)	τ (ms)
RF	90.615	61.58	69.848	67.23	79.673	51.80
KNN	82.907	2137.97	83.546	242.20	81.110	795.92
SVM	79.633	1300	76.757	282.71	96.006	419.78
LR	96.126	141.27	89.297	24.44	94.369	29.29
1D-CNN	94.424	12370.5	88.947	4558.3	87.679	5681.1

Figure S5 Confusion matrices of ML models applied across five plastic types and the mid-infrared regions-of-interest. True and predicted label accuracies are highlighted in blue along the diagonal of each confusion matrix (Figure S1a-e).
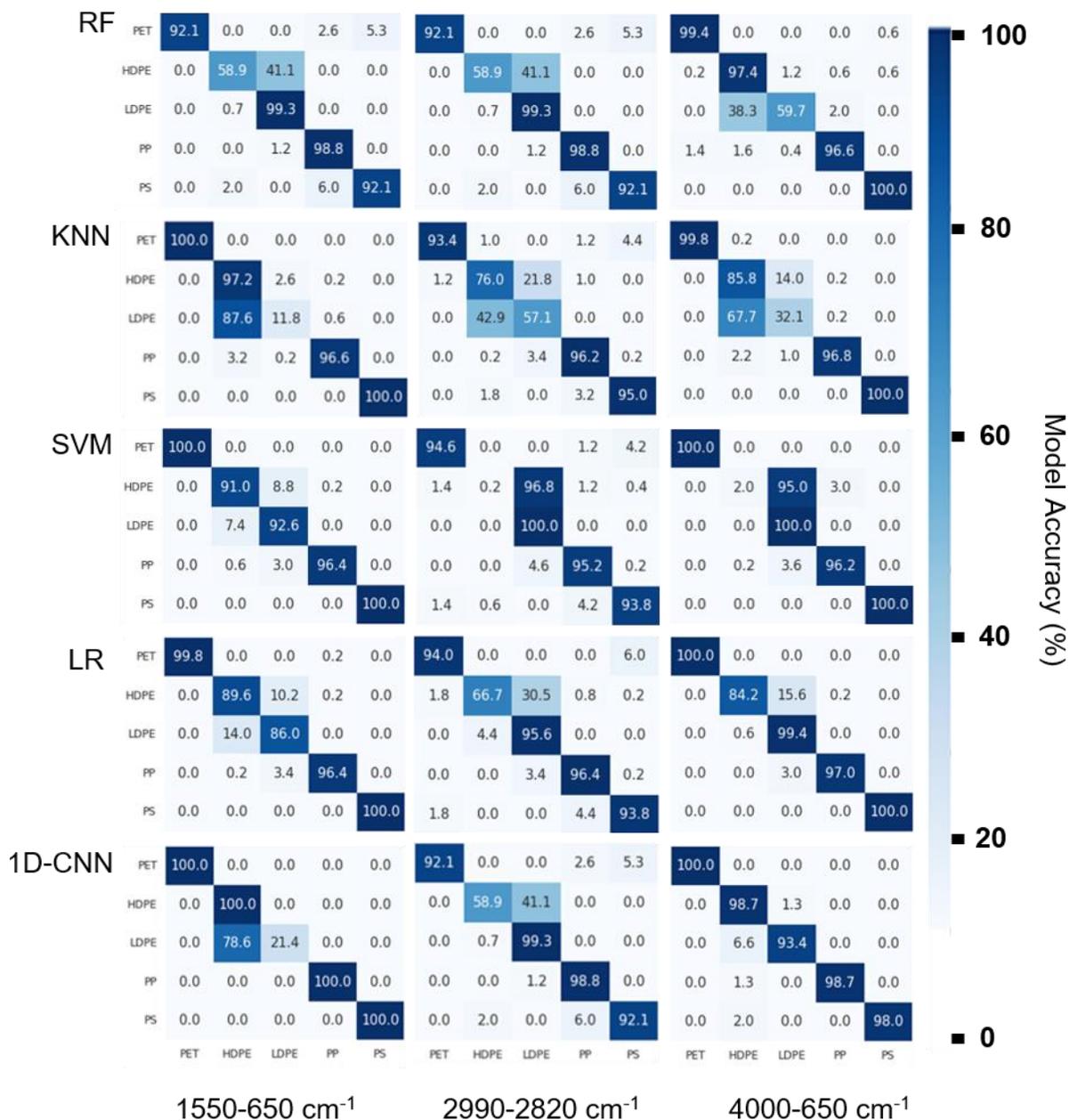
iii.   ML Results using Processed Spectra

Table S1d Classification accuracies and prediction times for mid-infrared spectral regions.

Spectra were processed according to procedures specified in 2.2 and 2.3 (Figure S2).

	$4000 - 650$ cm^{-1}		$2990 - 2820$ cm^{-1}		$1500 - 650$ cm^{-1}	
	Accuracy (%)	τ (ms)	Accuracy (%)	τ (ms)	Accuracy (%)	τ (ms)
RF	98.60	17.33	97.41	16.13	98.00	17.40
KNN	97.34	25.97	96.01	8.070	98.40	11.61
SVM	98.60	1.470	94.01	0.9400	96.41	0.9200
LR	97.00	0.9600	93.01	0.4400	95.41	0.4600
1D-CNN	96.90 (+/- 1.39)	6.000	91.73 (+/- 2.95)	2.000	92.08	2.000

Figure S6 Confusion matrices of ML models applied across five plastic types. True and predicted label accuracies are highlighted in blue along the diagonal of each confusion matrix.
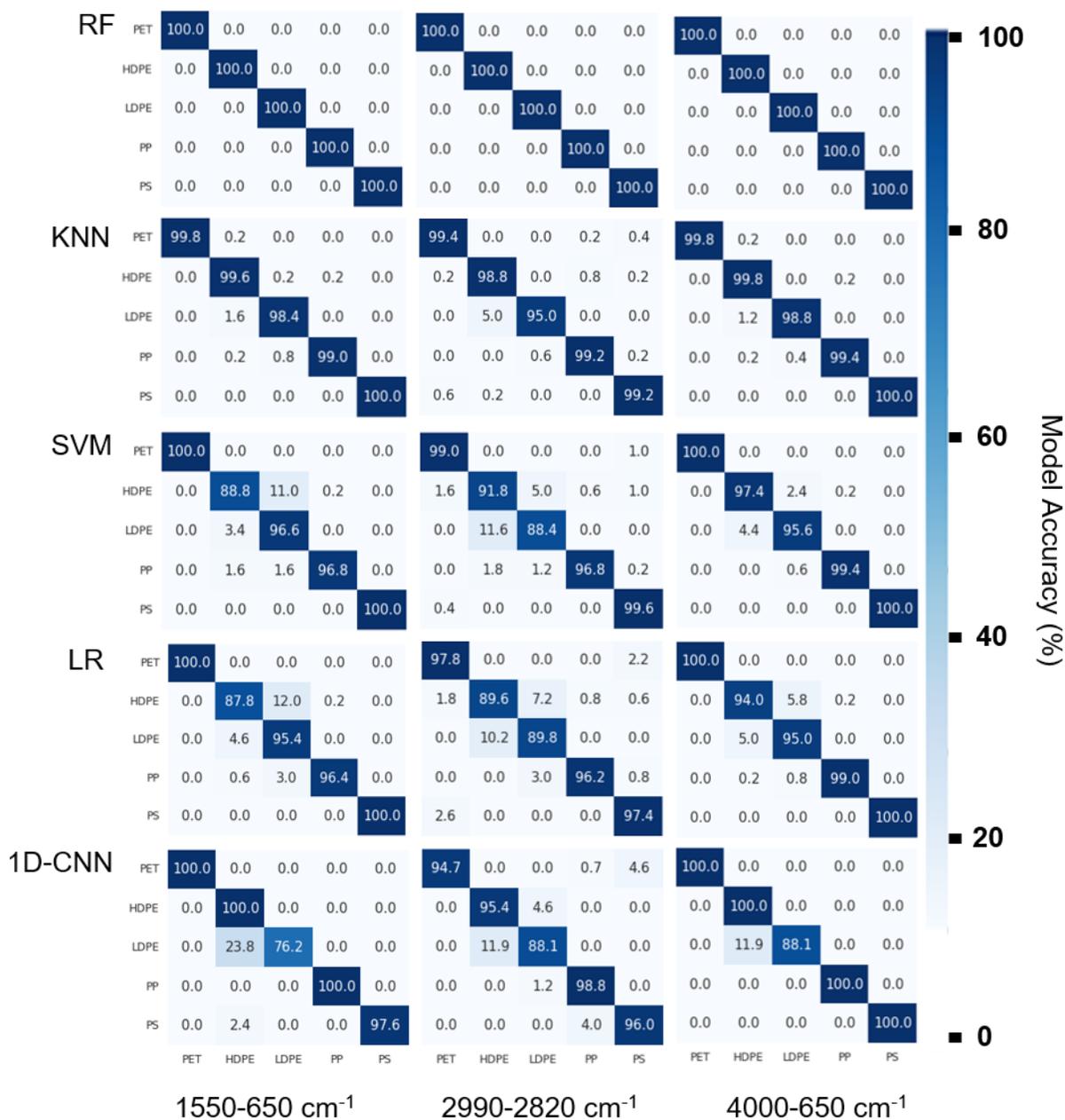

Section 4. HDPE and LDPE Classification

Table S1e Machine learning performance of unprocessed (percent transmittance, Fig. 3) HDPE and LDPE spectra using autoencoders.

Algorithm	Full Mid-Infrared (4000-650 cm^{-1})		C-H Stretching (2990-2820 cm^{-1})		Fingerprint (1500-650 cm^{-1})	
	Accuracy (%)	τ (ms)	Accuracy (%)	τ (ms)	Accuracy (%)	τ (ms)
RF	100	30.694	100	18.095	100	18.067
KNN	97.602	552.581	97.602	152.058	97.502	116.718
SVM	97.802	0.038	96.503	0.039	97.103	0.043
LR	97.003	7.179	96.903	2.385	97.302	2.372

Figure S7 Confusion matrices of ML models applied to HDPE and LDPE postconsumer plastics MIR spectra. True and predicted label accuracies are highlighted in blue along the diagonal of each confusion matrix.
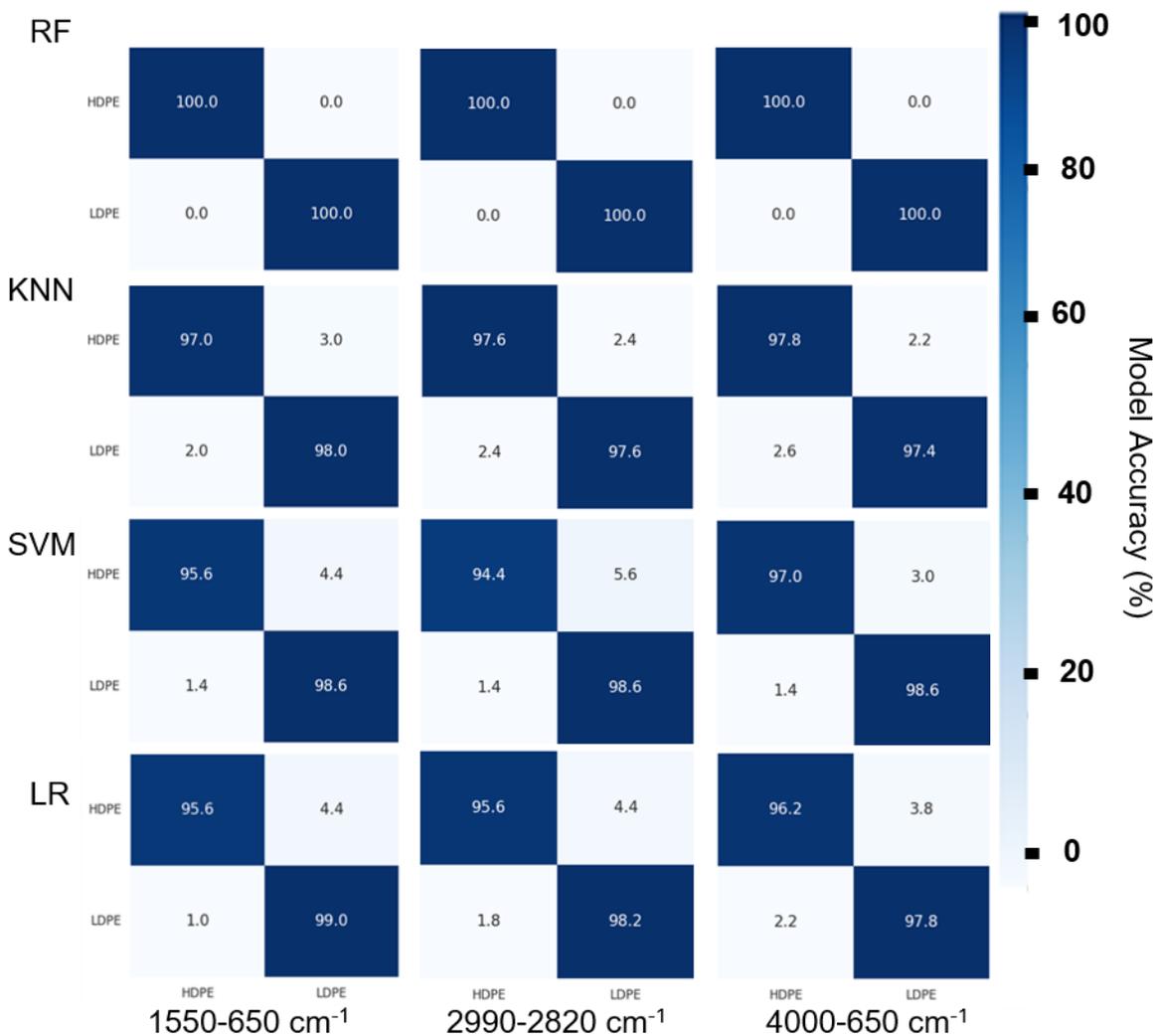
Section 5. Black Plastics Classification

Figure S8 Confusion matrix for black plastics. Accuracy is 100% using Random Forest without any pre-processing technique.
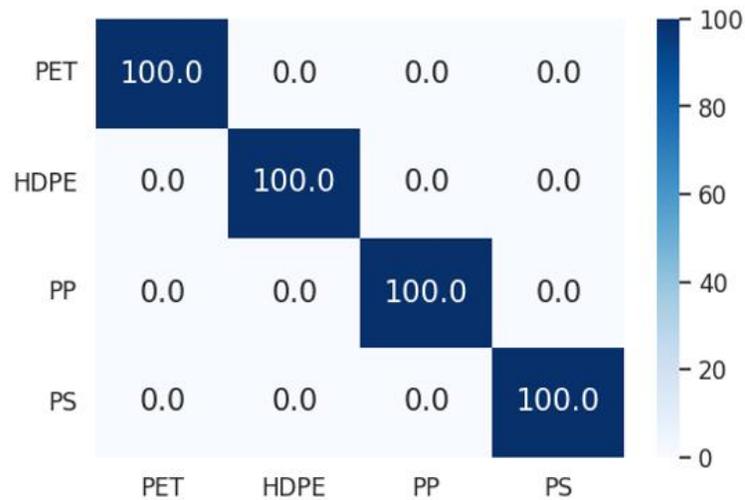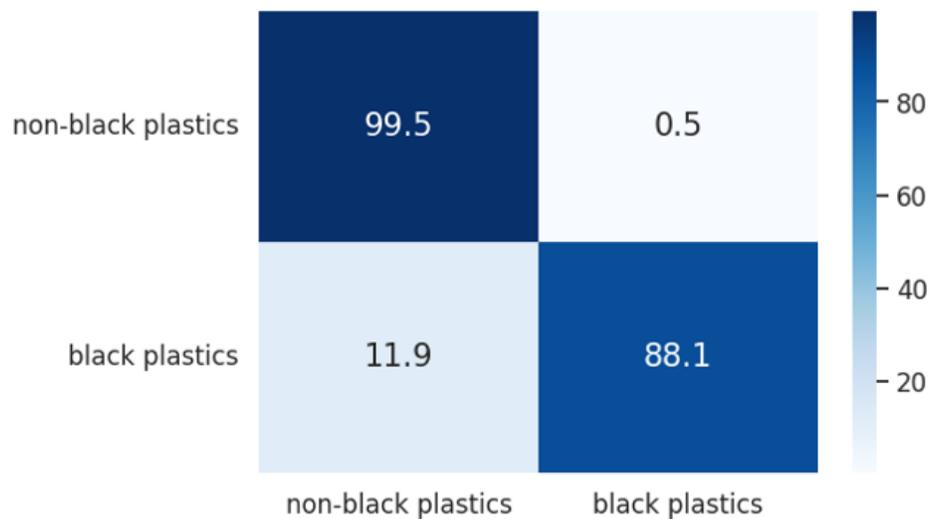

Figure S9 Confusion matrix for black plastics. Accuracy is 99.779% using Random Forest without any pre-processing technique.

Section 6. References

(1) Zinchik, S.; Jiang, S.; Friis, S.; Long, F.; Høgstedt, L.; Zavala, V. M.; Bar-Ziv, E. Accurate Characterization of Mixed Plastic Waste Using Machine Learning and Fast Infrared Spectroscopy. *ACS Sustainable Chem. Eng.* **2021**, 9 (42), 14143–14151. https://doi.org/10.1021/acssuschemeng.1c04281.

(2) Zhou, K.; Oh, S.-K.; Pedrycz, W.; Qiu, J.; Fu, Z.; Ryu, B.-G. Design of Data Feature-Driven 1D/2D Convolutional Neural Networks Classifier for Recycling Black Plastic Wastes through Laser Spectroscopy. *Advanced Engineering Informatics* **2022**, 53, 101695. https://doi.org/10.1016/j.aei.2022.101695.

(3) Long, F.; Jiang, S.; Adekunle, A. G.; M Zavala, V.; Bar-Ziv, E. Online Characterization of Mixed Plastic Waste Using Machine Learning and Mid-Infrared Spectroscopy. *ACS Sustainable Chem. Eng.* **2022**, 10 (48), 16064–16069. https://doi.org/10.1021/acssuschemeng.2c06052.

(4) Signoret, C.; Caro-Bretelle, A.-S.; Lopez-Cuesta, J.-M.; Ienny, P.; Perrin, D. Alterations of plastics spectra in MIR and the potential impacts on identification towards recycling. *Resources, Conservation and Recycling* **2020**, 161, 104980. DOI: https://doi.org/10.1016/j.resconrec.2020.104980.