

<https://doi.org/10.1038/s44454-025-00022-0>

# Open-set convolutional neural network for infrared spectral classification of environmentally sourced microplastics

Check for updates

Junhao Xie , Aoife Gowen &amp; Jun-Li Xu

Microplastics (MPs) are considered a global emerging threat. Environmental MP monitoring is essential for understanding and managing MP pollution. Here, a convolutional neural network model was developed for the infrared spectral classification of MPs from the environment. During model development, data were collected from multiple sources (e.g. OpenSpecy) to ensure sufficient intra-class diversity, for improving generalization. Also, the uncertainty threshold and OpenMax (an open-set recognition technique) were applied to enable the model to handle unknown classes (classes not included in model training) that may be present in real-world environmental samples. Furthermore, a targeted data augmentation strategy was proposed to improve model performance across varying infrared spectral ranges. Results show that the targeted data augmentation improved classification accuracy by up to 4.6%. In the optimal uncertainty threshold range of  $0.87 \pm 0.01$ , the model achieved 93.1% accuracy on both the known-class test spectra (15,741 spectra) and the unknown-class test spectra (6,279 spectra). OpenMax also enabled the model to effectively handle the unknown-class test spectra. Notably, its ability to recognize unknown classes is more prominent at lower uncertainty thresholds. The strategies proposed here could potentially be extended to other fields, e.g., food safety, pharmaceutical quality control, where spectral classification is critical.

Microplastics (MPs) are plastic fragments and fibers with the largest dimension  $< 5$  mm. Their presence has been detected in various environments, including oceans<sup>1</sup>, land<sup>2</sup>, the atmosphere<sup>3</sup>, freshwater sources<sup>4</sup>, food contact materials<sup>5</sup>, food and so on. Alarming, these tiny particles have found their way into fauna<sup>6,7</sup>, flora<sup>8</sup>, and even the human body<sup>9–12</sup>. Due to their widespread distribution and associated risks, MPs are considered an emerging global threat. Consequently, the management and analysis of these emerging pollutants hold significant importance for consumers and governments. In the analysis of MPs, polymer identification is the most basic, yet indispensable step.

Infrared (IR) spectroscopy, particularly Fourier-transform infrared (FTIR) spectroscopy, is one of the most effective and popular tools for polymer identification due to the specificity of IR spectral peaks to major functional groups found in the molecular structure of polymers<sup>13</sup>. After obtaining the IR spectra of unknown particles and applying appropriate spectral preprocessing, such as baseline correction, smoothing, etc., identification of the unknown particles is typically achieved through (1) library search or (2) machine learning (ML) based classification. Library search is a process where the spectrum of an unknown particle (hereafter the unknown spectrum) is compared to each reference spectrum in the library using

similarity measures such as the Pearson correlation coefficient and spectral angle mapper. Once the reference spectrum most similar to the unknown spectrum is found (usually obtained using a threshold, or cut-off, such as a similarity greater than 70%), the label of that reference spectrum is assigned to the unknown spectrum. It is worth noting that the library search approach is a relatively slow method, due to the need for a one-by-one comparison process. It has been reported that hours were needed for the identification of 1.8 million spectra through library search<sup>14</sup>. Freely available MP spectral identification tools based on library searching include OpenSpecy (<https://openanalysis.org/openspecy/>) and siMPle (<https://simple-plastics.eu/>), which are widely used by the MP research community. ML based classification, in contrast, involves building a classification model from reference spectra, which is then used to categorize unknown spectra into predefined classes. These classes typically include common polymer types as well as materials associated with environmental MPs, such as sediment, plant, and animal debris<sup>15</sup>. Common algorithms for constructing such classification models include support vector machine (SVM), partial least squares-discriminant analysis (PLS-DA), random decision forest (RDF), and so on. In terms of throughput, ML based classification clearly outperforms library search. For instance, Hufnagl, et al.<sup>16</sup> reported that only

School of Biosystems and Food Engineering, University College Dublin, Belfield, Dublin, Ireland. ✉e-mail: [junhao.xie@ucdconnect.ie](mailto:junhao.xie@ucdconnect.ie)

about 5 min were needed to classify 1 million spectra using a RDF classifier. Considering this advantage, ML based classification appears more suited to demands for large-scale, routine monitoring of environmental MP pollution. One tool associated with ML based classification is the Bayreuth Particle Finder<sup>15</sup>.

In recent years, convolutional neural networks (CNNs) have shown outstanding performance in spectral classification tasks related to MP research. Like SVM and RDF, CNN is also a ML algorithm. However, CNNs are often considered “deeper” compared to these “traditional” algorithms, as CNN model training is more complex and comprehensive, allowing the model to effectively learn hierarchical and abstract features from the data<sup>17</sup>, even without feature engineering (e.g., spectral preprocessing like baseline removal)<sup>18,19</sup>. Consequently, CNN models tend to achieve higher classification accuracy. To illustrate, Ai et al.<sup>20</sup> used a CNN model to classify the hyperspectral data of uncontaminated soil, soil contaminated by a specific type of MP, and soil contaminated by multiple specific types of MPs, achieving a classification accuracy of 92.6%. In contrast, the decision tree and SVM models achieved only 87.9% and 85.6% accuracy, respectively, on the same task. Luo et al.<sup>21</sup> obtained spectral data of MP mixtures using surface-enhanced Raman spectroscopy (SERS), and applied a CNN model alongside five classification models trained with traditional ML algorithms to classify these SERS spectra, for identifying specific types of MPs within the mixtures. The CNN model achieved a classification accuracy of 99.4%, outperforming all five traditional ML models. Similar results, where CNN models demonstrated higher accuracies than traditional ML based models in MP spectral classification, have been reported in a number of relevant studies<sup>22–28</sup>.

Currently, many effective CNN models have been trained and applied for various spectral classification tasks in MP research. However, none appears to be particularly suitable for spectral classification of MPs from the environment. We argue that, in developing such models, the top priority should be to ensure strong generalization, along with the integration of appropriate strategies to identify both known classes (i.e., those seen by the model in training) and unknown classes (i.e., those not seen by the model during training).

Generalization in a model refers to its ability to perform well not only on the specific set of data it was developed on but also on entirely new, unseen data. One key factor that influences generalization is the intra-class variation/diversity in the data<sup>29,30</sup>. Typically, the richer the intra-class variation, the better the generalization, and limited intra-class variation often leads to poor generalization. For instance, if a model is trained solely on virgin polypropylene (PP) spectra (meaning very low intra-class diversity), it may struggle to accurately classify environmental PP spectra that might exhibit unexpected small peaks or slight peak shifts due to polymer aging<sup>24</sup>, and/or other reasons. Similarly, training exclusively on spectra with broad spectral ranges can compromise adaptability and might lead to misclassification when inputs are restricted to narrower spectral ranges.

Often, sufficient variation is not available in the training set; consequently, data augmentation is often used to generate enough data for model training, and this, in fact, is also a step towards enhancing generalization. In data augmentation, expected changes, such as slight peak shifts, offsets, noise, etc., are added to the spectra, but this level of increase in intra-class diversity is relatively limited and may not significantly boost generalization<sup>33</sup>. Ideally, the training set should include data from as many different sources as possible, covering spectra from virgin MPs, environmental MPs, MPs at different aging stages, MPs with various additives in different amounts, as well as spectra of different spectral ranges, spectra that are diffraction-limited, spectra collected by various instruments and in different modalities, and etc. Fortunately, this is now possible as MP research has been growing vigorously for 20 years<sup>31</sup>, and many datasets have been made publicly available<sup>32</sup>.

A major limitation of current CNN approaches for MP spectral classification is that they operate as closed-set systems, while classifying spectra of environmental MP samples is a task of open-set recognition. Therefore, the concept of open-set recognition should be integrated into the

development of CNN models for spectral classification of MPs from the environment. Specifically, CNN models for multi-class classification typically transform an input into a probability vector containing  $N$  probabilities, where  $N$  represents the number of known classes (i.e., classes included in model development, usually common plastic types in CNN models designed for MP spectral classification). Each probability in the vector reflects the likelihood that the input belongs to a specific known class, and the sum of all probabilities in this vector equals 1. In other words, an input is classified into one of the known classes, regardless of whether the input actually belongs to any of them. This is how the closed-set nature of CNN models can be problematic.

One can appreciate how many false positives could result from using a CNN model to spectrally identify MPs from the environment, as such MP samples are usually cocktails of particles of known classes (usually a few to ~20 polymer types) along with particles of unknown classes (classes not included in model training). It is a misconception to assume that methods for extracting MPs from environmental matrices (i.e., density separation, oxidative digestion, and enzymatic digestion) can address this issue. In density separation, MPs are extracted based on differences in density, which fundamentally cannot separate known classes from unknown classes. Similarly, oxidative digestion and enzymatic digestion are designed to remove portions of organic matter in the matrix, but they do not, in essence, eliminate unknown classes, leaving a sample that contains particles of known classes only. Clearly, an approach must be developed to enable CNN models to handle unknowns effectively. In this regard, several approaches might be useful, including: thresholding SoftMax probabilities, applying an open-set recognition technique such as OpenMax<sup>33</sup> and Class Anchor Clustering<sup>34</sup>, and combining both strategies, i.e., SoftMax thresholding + open-set recognition<sup>33</sup>. These methods allow the model to recognize, to some extent, inputs from previously unseen classes, thus preventing the model from forcefully assigning such inputs to one of the known categories. As a result, they enhance the model’s applicability and robustness in real-world scenarios.

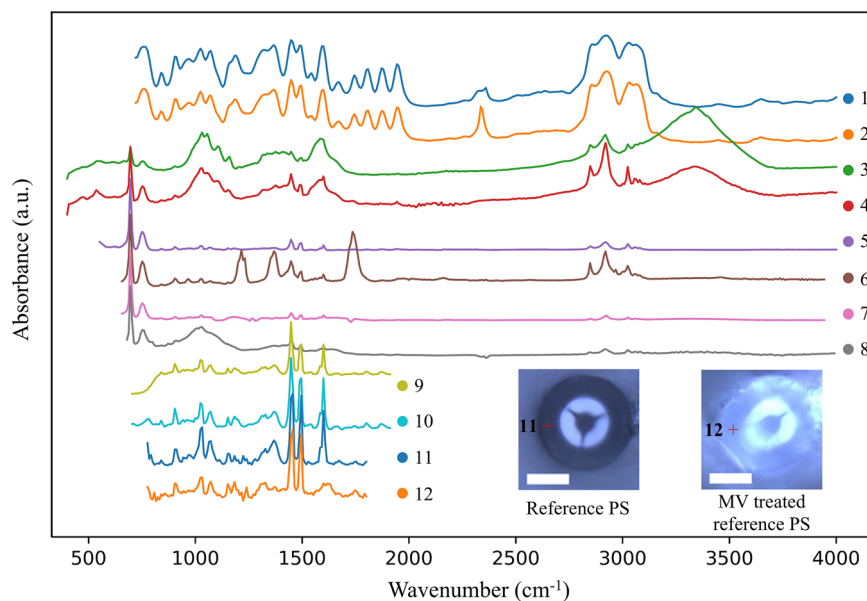
Here, we aim to train a one-dimensional CNN model for the classification of IR spectra of MPs. For increased intra-class diversity and hence improved generalization in the model, we supplemented our extensive laboratory-collected MP spectra with a diverse range of open-source spectral data. Additionally, we further augment the dataset by partitioning the data based on common spectral ranges to address the current issue of variation in spectral ranges due to the use of varied instrumentation. Furthermore, we integrated OpenMax into the CNN model, enabling open-set recognition. The open-set recognition performance of OpenMax was also compared against thresholding SoftMax probabilities alone.

## Results

### Intra-class spectral diversity

To achieve a well-generalized model, sufficient intra-class diversity is essential. In this study, spectral data were collected from different sources, which is expected to contribute to rich intra-class variability. Taking polystyrene (PS) as an example, this section visualizes the diversity within this class. Figure 1 presents 12 PS spectra selected from the remaining 1793 PS spectra after data cleaning, each labeled with an ID number that corresponds to detailed data information listed in Table 1. As seen, noticeable differences can be observed among these spectra in terms of both spectral signatures and wavenumber ranges. These variations are due to a combination of factors, including the use of different instruments with varying settings (for data collection), diverse source materials (e.g., environmental PS, consumer product PS, commercial reference PS, treated reference PS), and potentially other influences such as the degree of sample aging, surface contamination, and differences in sample collection environments. Note: Since Spectrum 11 and Spectrum 12 were collected from reference PS by us, we were able to include their corresponding optical images in the lower-right corner of Fig. 1. The red markings indicate the exact points where the spectra were collected. These two spectra, along with the optical images, demonstrate that

**Fig. 1 | Visualization of intra-class diversity within the PS category, based on 12 spectra selected from the data collected for this study.** Each spectrum is labeled with an ID number corresponding to Table 1. Optical images of the PS beads from which Spectra 11 and 12 were collected are shown in the lower-right corner. Scale bar is 20 microns. MV microwave. a.u. arbitrary units.



**Table 1 | Summary of data information for the PS spectra in Fig. 1**

ID	Instrument	Mode	Sample	Data source
1, 2	Nicolet iN10	Reflectance	N/A	Available on OpenSpecy; originally from Moore Institute for Plastic Pollution Research
3, 4	Nicolet 6700	ATR	PS MP collected from the Garonne River, France	Available on OpenSpecy; originally from de Carvalho et al. <sup>47</sup> .
5	N/A	N/A	N/A	Available on OpenSpecy; originally from Hawaii Pacific University polymer kit 1.0
6, 7	Nicolet 6700	ATR	Submerged PS plastic litter from Lake Tahoe, US	Available on OpenSpecy; originally from Davidson et al. <sup>48</sup> .
8	Nicolet iN10 MX	Transmittance	White foam	Available on OpenSpecy; originally from De Frond et al. <sup>49</sup> .
9, 10	Nicolet iS50	Transmittance	Commercial PS films	Collected in this study
11, 12	mIRage IR	Reflectance	(MV treated) reference PS beads	Collected in this study

the microwave treatment of PS beads not only altered their IR spectra but also led to physical changes, due to partial melting of the particles.

Figure 1 presents only a snapshot of all the PS spectra collected. Hence, in actuality, the intra-class diversity of PS is greater than what is shown in the figure. To quantitatively assess the intra-class diversity of both the 12 spectra in Fig. 1 and the full set of 1793 PS spectra, we calculated the Euclidean distance between each spectrum and the centroid of its corresponding set. We used the standard Euclidean distance from each spectrum  $x_i$  to its set centroid  $\mu_S$  (the component wise mean of spectra in that set):

$$\|x_i - \mu_S\|_2 = \sqrt{\sum_{j=1}^d (x_{ij} - \mu_{S,j})^2} \quad (1)$$

where  $j$  indexes the wavenumber dimension ( $d = 451$  in this study).

The mean of these distances was used as an indicator of intra-class diversity. Typically, the minimum mean distance is zero, indicating that all spectra within the set are identical. A larger mean distance reflects greater overall differences among the spectra, indicating a higher level of intra-class diversity. Calculations showed that the intra-class diversity of the 12 PS spectra in Fig. 1 is  $3.3 \pm 1.3$ , whereas the diversity across the full set of the 1793 PS spectra is  $4.5 \pm 2.0$ .

Zhu et al.<sup>35</sup> and Liu et al.<sup>23</sup> represent benchmark studies in CNN-based IR spectral classification of MPs. Therefore, comparing the intra-class diversity of our data with theirs is of particular relevance.

Focusing only on the classes that overlap with the known classes used in our study, the data used in Zhu et al.<sup>35</sup> were primarily collected from virgin plastic particles and plastic particles with additives using FPA-FTIR and ATR-FTIR. The dataset in Liu et al.<sup>23</sup> consisted of a small portion of spectra collected from virgin plastics and environmental plastics (such as plastic bottles, bags, and packaging) using FTIR, while the majority of their spectra originated from an early version of the OpenSpecy library.

It is important to emphasize that the earlier version of OpenSpecy used in their study differs substantially from the version we obtained in February 2024. The version we used contains a significantly larger number of spectra; for example, the earlier version includes only 3 acrylonitrile butadiene styrene (ABS) and 398 polyethylene (PE) spectra, whereas the more recent one contains 26 ABS and 2515 PE spectra.

Table 2 summarizes the comparison results. As shown, for classes shared across studies, our dataset consistently exhibits higher intra-class diversity.

### Effects of Type I data augmentation

Analysis of all the collected IR spectra revealed variations in spectral ranges. Some spectra have broader ranges, while others have narrower ones. In total, there are 16 distinct spectral ranges (Supplementary Table 1). One apparent reason for this is that the data were collected by different groups/individuals, and different IR instruments were used. For instance, spectra collected using an instrument equipped with a quantum cascade laser (e.g., The 8700 LDIR of Agilent, the Spero of Daylight Solutions) typically have spectral ranges

between 800 cm<sup>-1</sup> and 1800 cm<sup>-1</sup>, whereas those obtained using FTIR instruments can have lower wavenumbers reaching around 400–600 cm<sup>-1</sup> and upper wavenumbers extending up to 4000 cm<sup>-1</sup><sup>36</sup>. Given this variability, it is necessary to investigate whether the width of the spectral range affects the classification accuracy of the model. Essentially, this explores the impact of more or fewer spectral variables on model performance. See Supplementary Table 1 for the number of variables of each spectral range.

A 6-class CNN model was developed for the investigation described above. The six selected classes were PE, PET, PA, PP, PS, and PVC (see Supplementary Table 2 for the number of spectra used in training, validation, and testing). To examine the influence of spectral range on model performance, the valid spectral regions of the spectra in the test set of the 6-class model were deliberately limited to narrower ranges (then zero-padded back to 400–4000 cm<sup>-1</sup> to match the model's input format). When

the model was applied to these modified spectra, it was found that spectra with narrower spectral ranges (i.e., fewer spectral variables) were more prone to misclassification. Moreover, the narrower the spectral range, the higher the risk of misclassification. However, the application of Type I data augmentation (i.e., reusing long data. See the “Data augmentation” section for details) proposed here effectively mitigates this issue (Table 3). The reason for selecting only 6 classes rather than all 18 in this investigation was that these 6 classes contained a relatively larger number of long spectra (i.e., extend below 720 cm<sup>-1</sup> on the low-wavenumber side and beyond 3944 cm<sup>-1</sup> on the high-wavenumber side). Since Type I data augmentation is reusing long spectra, its implementation naturally requires the availability of such data. The more long spectra are included, the more meaningful and representative the evaluation is.

### Model performance

For clarity, this section focuses on two models. The first is a CNN model that was developed on 18 known classes, which we refer to as “the SoftMax model”. The second is a model developed by applying OpenMax to the first; and we refer to it as “the OpenMax model”. Further details of model development are provided in the “Model development” section.

After the cross-entropy loss stabilized at the 94<sup>th</sup> epoch (Supplementary Fig. 1), the SoftMax model was produced, achieving a classification accuracy of 97.3% on Test Set I (a subset of the test set, this subset contains known-class spectra only). However, as a closed-set model, it could not handle data from Test Set II (a subset of the test set, this subset contains unknown-class spectra only). When dealing with this limitation of a closed-set system, one approach that might be suggested is to include an “other” or “unknown” class during model training<sup>37</sup>. However, it is impractical to predefine and represent all possible unknown types in this way, as the variety of unknowns is inherently unbounded. An alternative approach is what we described in the “Methods” section of this study. Namely, applying an uncertainty threshold and defining any input whose maximum predicted probability falls below this threshold to be assigned to the unknown class (the label of this class is “unknown”). Figure 2 shows the influence of applying an uncertainty threshold to the SoftMax model. As observed, the accuracy on Test Set I gradually decreased to 90% as the threshold increased from 0 to 0.95, followed by a sharp drop to 0 when the threshold rose from 0.95 to 1. In contrast, the accuracy on Test Set II remained unchanged between thresholds of 0 and 0.2, then steadily increased from 0 to 100% as the threshold increased from 0.2 to 1. Notably, a threshold of 0.87 is where the classification accuracy for both Test Set I and Test Set II reached 93.1%. This threshold value of 0.87 was consistent with the result obtained from a 10-fold cross-validation on Validation Set II, which yielded a mean threshold of 0.87 ± 0.01, corresponding to the range where the accuracies for known and unknown classes were approximately balanced.

After applying OpenMax to the SoftMax model, the resulting model, referred to as the OpenMax model, was produced. In determining the optimal distance measure and tail size in the construction of the OpenMax model, a grid search was conducted, where the three distance measures (cosine distance, Euclidean distance, and Euclidean–cosine distance, which

**Table 2 | Intra-class diversity of the 18 known classes and comparison with previous studies**

Class name	This study (Mean ± Std)	Zhu, et al. <sup>35</sup> (Mean ± Std)	Liu, et al. <sup>23</sup> (Mean ± Std)
PET	4.6 ± 2.3	1.4 ± 1.1	1.2 ± 1.1
PU	4.4 ± 1.6	1.6 ± 0.6	3.4 ± 1.5
PC	4.1 ± 2.2	2.2 ± 0.9	3.5 ± 1.6
PE	3.9 ± 2.5	1.8 ± 0.9	1.0 ± 0.7
PP	3.5 ± 2.0	1.6 ± 0.6	1.4 ± 1.0
PVC	3.2 ± 1.5	1.4 ± 0.8	1.4 ± 0.9
polyamide	3.2 ± 2.2	2.0 ± 0.7	2.7 ± 0.8
ABS	2.3 ± 2.1	1.6 ± 0.6	1.3 ± 1.1
PMMA	2.0 ± 1.6	1.7 ± 0.6	3.0 ± 1.5
PEVA	6.1 ± 1.6		0.9 ± 0.8
PS	4.5 ± 2.0		1.2 ± 0.8
PLA	4.2 ± 3.2		1.4 ± 1.3
POM	2.6 ± 2.0	1.9 ± 0.9	
silicone	1.9 ± 1.2	0.2 ± 0.1	
PSU	1.6 ± 2.6		
PTFE	2.4 ± 2.0		
PVOH	2.1 ± 2.6		
peek	2.8 ± 1.8		

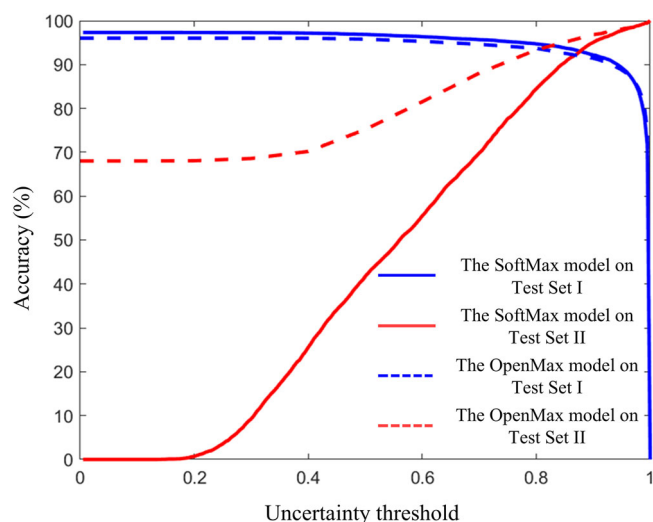
Results are presented as the mean Euclidean distance ± standard deviation between each spectrum and the centroid of its corresponding class.

For Zhu et al.<sup>35</sup> and Liu et al.<sup>23</sup>, only the classes overlapping with the 18 known classes in this study were included for comparison; blanks indicate that those classes were not covered in their datasets. PET polyethylene terephthalate, PU polyurethane, PC polycarbonate, PP polypropylene, PVC polyvinyl chloride, PMMA poly(methyl methacrylate), PEVA ethylene-vinyl acetate, PLA polylactic acid, POM polyoxymethylene, PSU polysulfone, PTFE polytetrafluoroethylene, PVOH polyvinyl alcohol, peek polyether ether ketone.

**Table 3 | Test set classification accuracy under different valid spectral ranges, with and without Type I data augmentation**

Valid spectral range	Accuracy (without Type I data augmentation)	Accuracy (with Type I data augmentation)
400 cm <sup>-1</sup> to 4000 cm <sup>-1</sup>	98.1%	98.3%
720 cm <sup>-1</sup> to 4000 cm <sup>-1</sup>	96.3%	97.9%
672 cm <sup>-1</sup> to 3168 cm <sup>-1</sup>	93.9%	97.5%
1240 cm <sup>-1</sup> to 3600 cm <sup>-1</sup>	96.5%	98.0%
704 cm <sup>-1</sup> to 1912 cm <sup>-1</sup>	93.0%	95.0%
776 cm <sup>-1</sup> to 1800 cm <sup>-1</sup>	90.0%	94.6%

Only a subset of representative spectral ranges is shown here. Several of the 16 spectral ranges listed in Supplementary Table 1 are highly similar (e.g., 400 to 4000 cm<sup>-1</sup> vs. 400 to 3992 cm<sup>-1</sup>) and yielded nearly identical classification accuracy. To avoid redundancy, only distinct and illustrative ranges are included in this table.



**Fig. 2 | Classification accuracy of the SoftMax model and the OpenMax model on Test Set I and Test Set II under varying uncertainty thresholds.** Test Set I refers to the subset of the test set containing spectra from known classes, whereas Test Set II refers to the subset of the test set containing spectra from unknown classes.

is a weighted combination of cosine distance and Euclidean distance, see Eq. 4 for the definition) and tail size values ranging from 15 to 40 (with an interval of 5) were evaluated. The results showed that cosine distance consistently outperformed Euclidean distance and Euclidean–cosine distance across all tail size values, as cosine distance substantially enhanced the model’s ability to identify the unknown classes (Supplementary Fig. 2), whereas the other two distance measures did not (Supplementary Figs. 3 and 4). Interestingly, Bendale and Boul<sup>33</sup> reported that the optimal distance measure was Euclidean–cosine distance when they developed OpenMax on image datasets (e.g., ImageNet 2012). This might suggest that the optimal distance metric may be influenced by data dimensionality and modality. An investigation of this difference is beyond the scope of the present study. When the tail size value was set to 35, the best performance was achieved (Supplementary Fig. 2e). Similar results were observed on the test set after running the same evaluation task (Supplementary Fig. 5). Therefore, all OpenMax-related results reported in the main text are based on the cosine distance with a tail size value of 35.

The OpenMax model achieved a classification accuracy of 96.0% on Test Set I, which was 1.3% lower than that of the SoftMax model. However, unlike the SoftMax model, the OpenMax-enhanced model was capable of open-set recognition. When applied to Test Set II, the OpenMax model successfully classified 68.0% of the samples as “unknown”.

When an uncertainty threshold was further applied to the OpenMax model, the success rate for identifying unknown classes could be increased. However, it is important to note that this additional improvement was fundamentally different from the 68.0% achieved by the model itself. The 68.0% reflects the model’s autonomous ability to assign inputs to the unknown class based on how distant they were from the known classes in feature space. In contrast, the extra improvement resulting from thresholding was a functional, human-defined decision. The results obtained after applying the uncertainty threshold to the OpenMax model are also shown in Fig. 2. As shown, when the threshold increased from 0 to 0.95, the classification accuracy of the OpenMax model on Test Set I gradually decreased to 90%. Once the threshold exceeded 0.95 and approached 1, the accuracy dropped sharply to 0%. Conversely, for Test Set II, the unknown detection rate remained at 68.0% between thresholds of 0 and 0.2, and then increased steadily to 100% as the threshold approached 1. Notably, the two performance curves intersected at a threshold of 0.80, at which the model achieved a balanced classification accuracy of 93.1% for both known and unknown classes. This threshold falls within the range  $0.81 \pm 0.03$  obtained from a 10-

fold cross-validation on Validation Set II, where the accuracies of the known and unknown classes were also approximately equal.

It is evident that as the uncertainty threshold increased, the behavior of the OpenMax model and the SoftMax model on Test Set I became nearly identical. Across low to mid-range thresholds, the accuracy of the OpenMax model remained consistently slightly lower, with a difference of no more than 1.3% compared to the SoftMax model. However, this difference gradually diminished as the threshold increased and eventually became negligible. In contrast, a clear distinction is observed on Test Set II, where the OpenMax model significantly outperformed the SoftMax model in detecting unknown classes, especially at lower uncertainty thresholds, albeit at the cost of slightly lower accuracy for the known classes.

The SoftMax model was evaluated on the combined dataset of Test Set I and Test Set II. After applying an uncertainty threshold of 0.87 to the classification outputs, an accuracy of 93.1%, a precision of 94.1%, and an F1 score of 93.4% were achieved, suggesting overall satisfactory performance. To examine these results in detail, a confusion matrix (Fig. 3) and a classification report (Table 4) were generated to allow further inspection of individual class-level results. This inspection revealed that most classes achieved relatively high F1 scores. However, PEVA and POM exhibited notably low F1 scores of 0.25 and 0.27, respectively, due to both low recall and low precision.

Among the 10 test spectra labeled as PEVA, three were not predicted as PEVA by the model. Although these three spectra were labeled as PEVA, they essentially contain only baseline signals without any characteristic PEVA absorption peaks (Supplementary Fig. 6). Likewise, among the six test spectra labeled as POM, two were not predicted as POM by the model. These two spectra similarly exhibit only baseline patterns without identifiable POM features (Supplementary Fig. 7). All of these spectra originated from OpenSpecy, suggesting that the version of OpenSpecy used in this study includes some low-quality entries.

Precision for PEVA and POM was low at 0.18 and 0.17, which indicated a tendency for the model to misclassify other classes as PEVA or POM. In total, 27 spectra were misclassified as PEVA, and 20 spectra were misclassified as POM.

Four test spectra labeled as PE, which is a class known to the model, were predicted as PEVA. Visual inspection showed that these four spectra closely match PEVA, as they contain the full set of characteristic PEVA peaks and are less consistent with PE (Supplementary Fig. 8). This strongly suggests mislabeling. All these four spectra are from OpenSpecy.

Two spectra labeled as ethylene vinyl acetate were predicted as PEVA. In fact, ethylene vinyl acetate is the full name of PEVA. During the initial data harmonization across sources, we might have focused on the abbreviations “PEVA” and “EVA” and overlooked the full name. As a result, the ethylene vinyl acetate spectra were incorrectly placed in Test Set II. This outcome reflects a human error rather than a model misclassification.

Three spectra labeled as cellulose acetate (CA), which is a class unknown to the model, were predicted as PEVA. These three CA spectra were plotted together with a reference PEVA spectrum (Fig. 4). Several peaks of the reference PEVA spectrum are numbered in the figure and represent characteristic PEVA peaks<sup>38</sup>. As seen, the CA spectra are highly similar to PEVA, having seven characteristic peaks of PEVA numbered 1–7, with peaks 3, 5, and 6 showing comparable absorption intensities. This similarity likely explains why the model classified the CA spectra as PEVA. The model did not overturn the PEVA decision despite intensity differences at peaks 1, 2, 4, and 7, the presence of the peak indicated by the dashed line, the absence of peak 8, or other differences; these cues likely carried low decision weight and therefore had little influence on the final prediction.

Deep learning models are often treated as black boxes whose decisions do not follow human rules, such as “if characteristic peaks appear, then classify as PEVA”, but instead arise from high-dimensional, nonlinear representations. To illuminate how the model leverages information across wavenumbers when making predictions, we employed Shapley Additive Explanations (SHAP), a game-theoretic interpretability framework that quantifies each input feature’s marginal contribution to the model output.

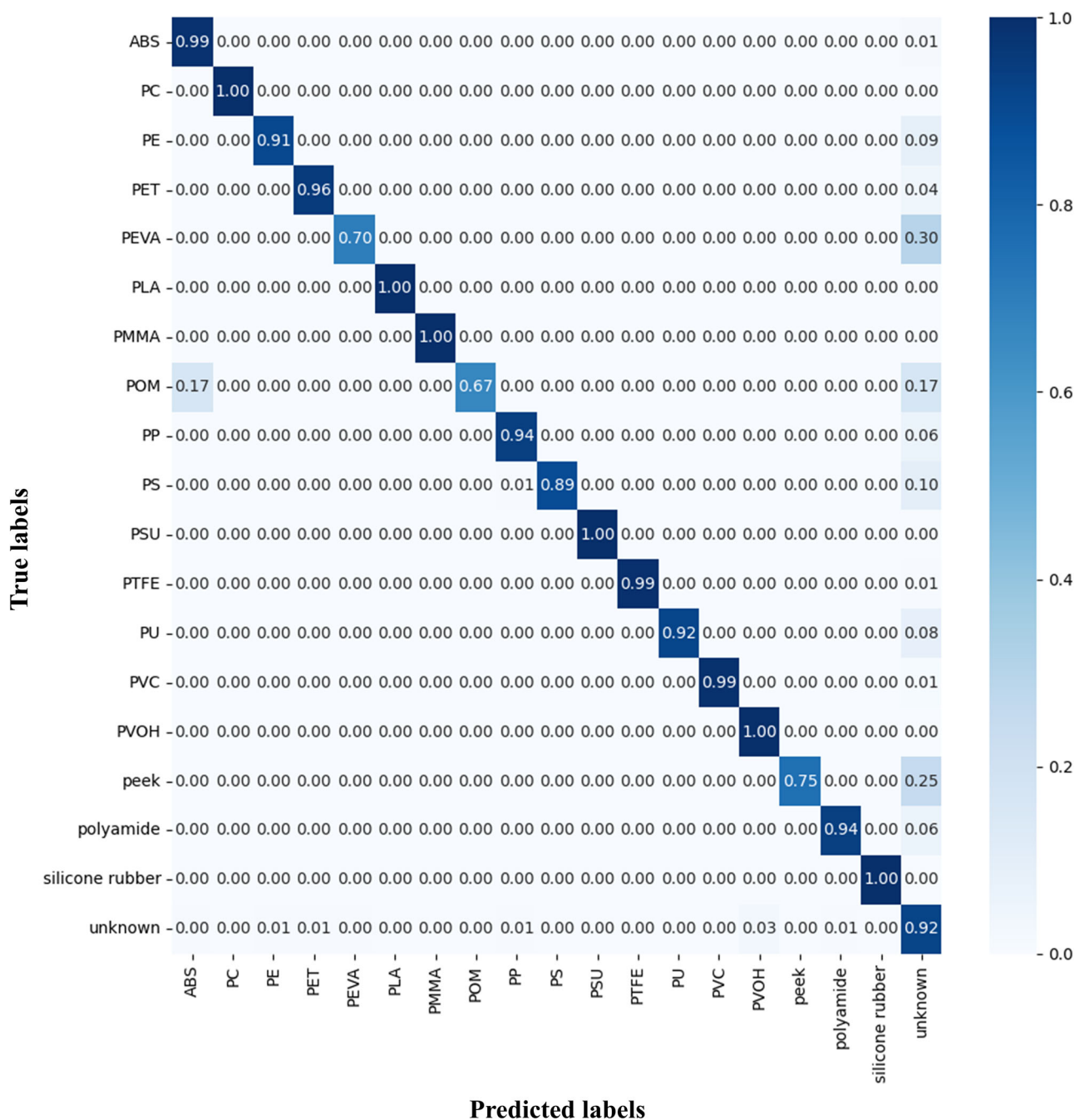


Fig. 3 | Normalized confusion matrix for the test set. Results were obtained with the SoftMax model, and an uncertainty threshold of 0.87 was applied.

We first computed SHAP values for all spectra that were misclassified as PEVA, in order to understand why these non-PEVA spectra were pushed into the PEVA class. As a representative example, we focused on a single CA spectrum (the top spectrum in Fig. 5), which, from the model’s point of view, belongs to an unknown class but was predicted as PEVA. Passing this spectrum through the SHAP explainer yielded per-class SHAP values. Among all 18 classes, the signed SHAP raw sum for PEVA is the largest at +6.23, whereas the corresponding sums for other classes are smaller in magnitude and often negative. This pattern indicates that the spectrum is pulled strongly towards PEVA and pushed away from many alternative classes.

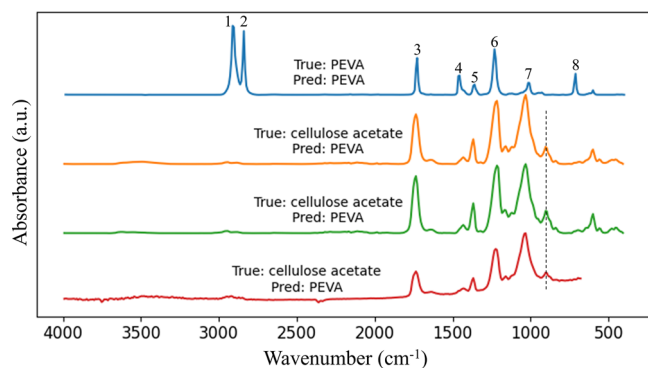
A closer inspection of the spectral-band contributions shows that this strongly PEVA-oriented decision is not produced by roughly uniform positive contributions from all wavenumbers. Instead, it is dominated by a small number of short intervals that exert a very strong positive pull. For the

PEVA class, the net signed SHAP contribution of the full 400–4000  $\text{cm}^{-1}$  spectrum is about +6.23, but within the 550–2000  $\text{cm}^{-1}$  region alone, the net contribution already reaches +5.52. Using a sliding-window analysis on this mid-wavenumber region and enforcing a non-overlap constraint, we identified four key sub-intervals (indicated by the semi-transparent yellow shading on the top spectrum in Fig. 5). The sum of their signed SHAP contributions corresponds to 138.44% of the +5.52 contribution from the entire 550–2000  $\text{cm}^{-1}$  band, that is, +7.64, which itself is about 122.66% of the net contribution from the entire 400–4000  $\text{cm}^{-1}$  spectrum. In other words, these four sub-intervals alone provide more positive evidence than the net effect of the full spectrum. If we conceptually remove these four sub-intervals from the spectrum, the remaining wavenumbers contribute a combined SHAP value of about –1.41, which is only around 18% of the magnitude of the four key intervals. The rest of the spectrum, therefore, does exert a negative counteracting effect, but this compensation is far too small

**Table 4 | Classification report for the test set**

Class	Precision	Recall	F1-score	Number of spectra
ABS	0.88	0.99	0.93	247
PC	0.81	1.00	0.90	69
PE	0.99	0.91	0.95	4447
PET	0.95	0.96	0.95	768
PEVA	0.18	0.70	0.25	10
PLA	0.94	1.00	0.97	138
PMMA	0.88	1.00	0.93	56
POM	0.17	0.67	0.27	6
PP	0.99	0.94	0.96	3902
PS	0.99	0.89	0.94	1151
PSU	0.97	1.00	0.99	142
PTFE	0.95	0.99	0.97	146
PU	0.96	0.92	0.94	170
PVC	0.99	0.99	0.99	469
PVOH	0.34	1.00	0.51	94
peek	1.00	0.75	0.86	4
polyamide	0.99	0.94	0.96	3790
silicone rubber	1.00	1.00	1.00	132
unknown	0.85	0.92	0.89	6279

Results were obtained with the SoftMax model, and an uncertainty threshold of 0.87 was applied.

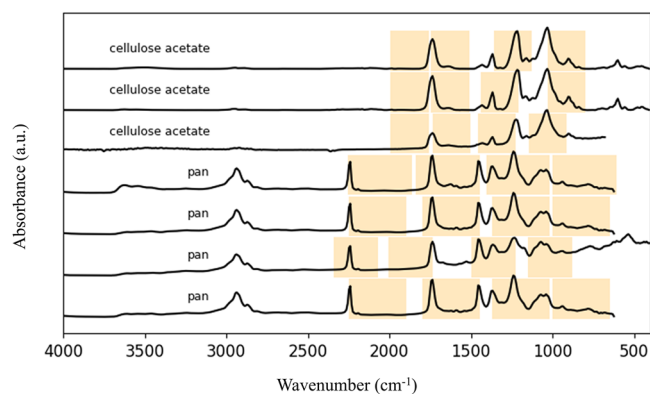


**Fig. 4 | Four spectra predicted as PEVA by the model.** The top spectrum is a PEVA reference spectrum with eight numbered characteristic peaks<sup>38</sup>. The dashed line indicates a peak present in cellulose acetate but absent in PEVA. “True” indicates the true label, “Pred” indicates the predicted label. Absorbance is shown in arbitrary units (a.u.).

to overturn the overall strongly PEVA-oriented pattern, and the net SHAP for PEVA remains close to +6.23.

These results indicate that the CA spectrum is classified as PEVA primarily because its local spectral structures in these four sub-intervals closely resemble the PEVA-specific patterns learned during training. In network terms, the relevant convolutional channels are strongly activated in these regions, and the ensuing activations are passed through relatively large positive weights connected to the PEVA output unit in the fully connected layer, producing a large positive contribution to the PEVA logit. For most other classes, SHAP contributions in the same regions are negative or weakly positive, so their logits are comparatively suppressed.

Figure 5 and Supplementary Figs. 9 and 10 show all spectra that were misclassified as PEVA, together with the key sub-intervals identified by the SHAP analysis as driving these misclassifications. For each misclassified spectrum, the dominant sub-intervals are overlaid as semi-transparent bands along the curve. These highlighted regions mark, for each spectrum,



**Fig. 5 | Unknown-class spectra misclassified as PEVA and their dominant evidence-bearing intervals.** Each curve represents a spectrum that the closed-set model incorrectly classified as PEVA; its true material label is indicated next to the corresponding trace. Semi-transparent bands overlaid on each spectrum mark the SHAP analysis-identified intervals that provide the strongest positive evidence for the PEVA output, whose cumulative directional contribution exceeds the net PEVA-directed effect of the full spectrum. pan denotes polyacrylonitrile. Absorbance is shown in arbitrary units (a.u.).

the parts of the wavenumber range that contribute most strongly to the PEVA logit and can be interpreted in the same way as the four key sub-intervals discussed above for the CA example.

20 spectra were misclassified as POM, all of which belong to classes unknown to the model. We subjected these spectra to the same SHAP-based analysis. The results are shown in Supplementary Figs. 11 and 12, where each spectrum is plotted together with its dominant SHAP-derived sub-intervals, overlaid as semi-transparent bands along the curve. For each misclassified spectrum, these highlighted regions indicate the parts of the wavenumber range that contribute most strongly to the POM logit and can be interpreted in the same way as for the CA example described above.

The OpenMax model was evaluated on the same test set, with an uncertainty threshold of 0.81 applied to the outputs. The resulting accuracy, precision, and F1 score were 93.6%, 94.5%, and 93.9%. These values are close to those of the SoftMax model with an uncertainty threshold of 0.87. Class level results are likewise similar, as shown in Supplementary Fig. 13 and Supplementary Table 3.

Small differences were observed. For PP, PSU, PVC, and silicone rubber, the per-class accuracy under OpenMax was lower by 0.01 compared with SoftMax. This decline is consistent with the mechanism by which OpenMax treats borderline samples more conservatively. OpenMax recalibrates scores using a distance-based scheme. When the activation pattern of a spectrum lies far from the mean activation of a class, the model reallocates part of its score mass away from that class toward the unknown category. This reduces the number of high-confidence assignments within those classes, slightly decreases the true positives, and appears as a modest drop in per-class accuracy.

In contrast, PE, PS, PU, and the unknown category showed higher per-class accuracy under OpenMax compared with SoftMax, by 0.01, 0.01, 0.03, and 0.02, respectively. These gains indicate that OpenMax suppresses false positives more effectively in cases of spectral similarity across neighboring classes, and assigns out-of-distribution or weak-evidence spectra more reliably to the unknown category. The resulting decision boundaries are therefore sharper, and residual errors are concentrated in genuinely ambiguous regions.

Overall, OpenMax matched the aggregate performance of SoftMax while offering a small improvement in robustness for classes that are prone to confusion and for the unknown class. The observed trade off at the chosen operating point is expected, with a slight reduction in coverage for several known classes balanced by improved control of spurious assignments.

**Table 5 | Comparison of classification accuracy among our models and two benchmark CNN models**

Model	Accuracy	Notes
Our SoftMax model	97.3%	Evaluated on the full Test Set I
Our OpenMax model	96.0%	Evaluated on the full Test Set I
Closed-set 1D CNN (pretrained)	47.6%	Evaluated only on Test Set I samples belonging to classes known to the model
Closed-set 2D CNN (reproduced)	56.4%	Evaluated only on Test Set I samples belonging to classes known to the model

Test Set I refers to the subset of the test set containing spectra from the 18 known classes included in this study.

Closed-set 1D CNN refers to the model of Liu et al.<sup>23</sup>; we used their released pretrained weights. Closed-set 2D CNN follows Zhu et al.<sup>35</sup>; since pretrained weights were not provided, we reproduced the model from the authors' released code and data. Our implementation was faithful to the originals, using the same software environment (e.g., Python/MATLAB), preprocessing, training/testing settings.

Benchmark studies in the field of CNN-based MP IR spectral classification include those by Zhu et al.<sup>35</sup> and Liu et al.<sup>23</sup>. The CNN model developed by Zhu et al.<sup>35</sup> was designed for the classification of 11 common plastic types. The CNN model proposed by Liu et al.<sup>23</sup> supports the classification of 15 typical MPs as well as over 100 types of plastic mixtures. For ground truth generation, they created mixtures by melting and combining known types of plastics and then acquired IR spectra from the resulting blends. Both models achieved test accuracies exceeding 90.0%. Notably, Liu et al.<sup>23</sup> further developed a software tool based on their trained models and made it freely available to the research community. Both studies made valuable contributions to the field; however, neither provided a solution for the handling of classes unknown to their models.

As a study closely aligned with their work, it is meaningful to compare our work with theirs in terms of model performance.

Table 5 summarizes the comparison results. Our models were evaluated on the entire Test Set I, which contains all 18 known classes. The models of the benchmark studies were only evaluated on the subset of Test Set I corresponding to the classes known to their models. This evaluation setup ensures a fair comparison. In addition, uncertainty thresholds were not applied in this comparison, as the two studies did not define or report any thresholding strategies; applying thresholds post hoc would therefore be inappropriate.

## Discussion

To achieve strong generalization, the dataset for model development must exhibit sufficient within-class diversity. For the classes shared across studies, our dataset shows higher within-class diversity than the datasets used in current state-of-the-art studies on CNN-based MP spectral classification (Table 2), which is expected to translate into better model generalization.

From an environmental MP perspective, we reasonably expect that intra-class diversity is an open and dynamic concept, i.e., it does not converge or stabilize with increasing data but rather continues to diverge as more real-world spectra are collected. This suggests that a fundamental strategy for ensuring diversity coverage is to link the model to a continuously updated spectral database, such as OpenSpecy, thereby maintaining strong generalization performance as new and variable inputs emerge in real-world applications.

The Type I data augmentation (i.e., reusing long data) proposed in this study originated from an intuitive idea. To illustrate this idea, consider an image classification model trained to recognize several animal categories, including cats. During testing, when presented with a stack of photos showing entire cats, the model performs remarkably well, almost always making correct classifications. However, when those same photos are cropped to show only half of each cat, the model's error rate increases noticeably. A reasonable assumption would be that the training set contains few examples of "half-cat" images, causing the model to rely more heavily on the global, holistic features present in full-cat images during learning. This observation naturally leads us to consider whether artificially increasing the number of "half-cat" images in the training set, by cropping and reusing existing full-cat photos, could help the model better learn local features. Our Type I data augmentation method builds on this idea, substituting the concept of partial images for that of partial spectra. The classification results

presented in Table 3 confirm its effectiveness: the model not only maintains high classification accuracy under full-spectrum inputs but also demonstrates significantly improved robustness when handling spectrally limited, narrower-range inputs. The accuracy improvement was particularly significant when fewer variables were involved, with an increase of up to 4.6%, as shown in the last row of Table 3.

CNN models thrive on data diversity: the richer and more heterogeneous the training set, the more completely a network can approximate the underlying data structures and generalize to unseen samples. Yet, aiming for this very diversity inevitably introduces inputs of unequal length, which raises a natural question: does input-length (number of variables) disparity have an influence on classification accuracy? Our analysis not only provides an answer (yes, shorter spectra are more prone to misclassification) but also demonstrates that our proposed Type I data augmentation strategy can effectively mitigate this issue. To the best of our knowledge, we are the first to systematically investigate this problem within a CNN framework and to propose a practical solution. The relevant results have important implications for the future application of CNN models in MP spectral classification, as they offer a viable strategy for addressing a key challenge in integrating data from diverse instruments and sources.

As noted in the introduction, whenever an input enters a closed-set CNN model, it can only be placed into one of the model's known classes, regardless of whether it truly belongs to a known class. The "Model development" section explains this mechanism. In the context of environmental MP detection, any particle may end up on the filter, be measured by IR spectroscopy, and then passed to the model. If no safeguards are implemented, for example, open-set recognition methods or uncertainty-based rejection thresholds, all non-plastic particles and polymers that are absent from the training set will inevitably be forced into one of the known MP labels.

Löder and Gerdt<sup>39</sup> collected a sediment sample and, after purification, found that only about 1.4% of particles and fragments were MP, while the remainder were minerals, cellulose, chitin, and other inorganic or biological materials. If a closed-set CNN were applied to this sample, a large number of false positives would be unavoidable for the remaining 98.6% of non-target material (under the reasonable assumption that none of the non-plastic substances were included in the training set). Similar issues could arise for freshwater plankton samples, wastewater samples, and biological tissue samples, which are likewise rich in non-plastic materials<sup>40</sup>. Separation can reduce, but not completely remove, such contaminants. Moreover, even if separation made the filter contain "only MPs," they might still not be among the model's known, common types; rare or novel polymers would still be pushed by a closed-set model into the "closest" known class. Introducing open-set recognition (e.g., OpenMax), which gives the model flexibility to assign some inputs to an "unknown" category, therefore has clear practical value and can suppress systematic misclassification.

On the trade-off: a 1.3% drop in known-class accuracy in exchange for 68.0% accuracy on unknowns. When is this worthwhile? Let  $p_u$  be the fraction of unknowns on the filter and  $p_k = 1 - p_u$  the fraction of knowns. Relative to the closed-set baseline, the net gain in overall accuracy from

adopting open-set recognition can be approximated by the following.

$$\Delta \approx 0.68p_u - 0.013p_k = 0.68p_u - 0.013(1 - p_u) = 0.693p_u - 0.013 \quad (2)$$

Open-set recognition is preferable when  $\Delta > 0$ , i.e.,

$$p_u > \frac{0.013}{0.693} \approx \frac{1}{53} \iff \text{unknown} : \text{known} \gtrsim 1 : 52 \quad (3)$$

Thus, as long as the unknown: known ratio is roughly  $\geq 1:52$  (unknowns  $\geq 1.9\%$ ), open-set recognition yields a positive overall benefit. Conversely, making the “loss of 1.3% in known-class accuracy” not worthwhile would require the known fraction on the filter to exceed 98.1%. In practice, current purification workflows (which target reductions in organic and inorganic contaminants) are not designed to reduce the fraction of types unknown to the model. Even if an exceptionally effective protocol raised the MP:non-MP ratio above 98.1%, that would not imply the known:unknown ratio also exceeds 98.1%. Therefore, in most environmental samples, adopting open-set recognition is not only reasonable but also acceptable for overall accuracy.

Based on the results summarized in Fig. 2, effective approaches to enabling CNN models to recognize unknown classes include applying an uncertainty threshold, applying OpenMax, and combining the uncertainty threshold with OpenMax.

The uncertainty threshold not only ensures the quality of predictions for known classes but also enables the model to reject or identify certain unknown inputs. However, to the best of our knowledge, no previous studies on CNN-based MP spectral classification have adopted or explicitly discussed thresholding SoftMax probabilities, including those that claim their methods are applicable to environmental MP spectral identification. The reason for this absence remains unclear, but we think a plausible explanation could be that there is currently no widely accepted standard or reference threshold for CNN models intended for MP spectral classification. This situation is in stark contrast to library search-based spectral identification methods, where the Hit Quality Index (HQI) threshold, commonly suggested to be around 0.7, has been widely adopted and has become a de facto standard<sup>41</sup>.

Based on the results of 10-fold validation on Validation Set II, the SoftMax model consistently showed that a threshold around  $0.87 \pm 0.01$  yielded balanced and high accuracy for both known and unknown classes, approximately 93.0% in each fold. However, we wish to emphasize that this threshold range is not intended as a general reference standard, either for the SoftMax model described here or for other CNN models. This is because the threshold was identified using the specific known and unknown samples present in Validation Set II. Once new data are introduced, the “optimal threshold range” may (significantly) shift. Fundamentally, although we made efforts to collect spectra from multiple sources for enhanced intra-class diversity, we are not yet confident enough to claim that the dataset is fully representative. Therefore, the “optimal threshold range” determined in this study should not be assumed to generalize across different datasets or application scenarios.

The key message is that incorporating an uncertainty threshold into CNN-based MP spectral classification is both useful and recommended, while the value of that threshold remains a dynamic, data-dependent quantity. In practice, when deploying the model on a new dataset, it is recommended to randomly sample a small portion of the data (e.g., around 10%), for expert labeling. This labeled calibration subset can then be used to re-estimate the uncertainty threshold to achieve the desired operating point. In this study, the goal is to maximize retention of known classes while effectively rejecting unknowns. If calibration fails to identify a threshold that provides strong and balanced performance for both known and unknown classes, this usually indicates a substantial distribution shift. In such cases, the model should be updated or retrained using additional data from the new domain.

To the best of our knowledge, just like the absence of the uncertainty threshold, no open-set recognition methods, such as OpenMax, have been applied in any published studies on MP spectral classification. According to the results shown in Fig. 2, OpenMax performed better in the lower range of uncertainty thresholds, with its advantage becoming more pronounced as the threshold decreased. However, in the higher threshold range, its added benefit diminished significantly, and at certain threshold values, its performance became nearly identical to that achieved by using the uncertainty threshold alone. In our case, when the OpenMax model was applied with a threshold of 0.81, it yielded the same result as the SoftMax model with a threshold of 0.87: both achieved a classification accuracy of 93.1% on the test set. These findings suggest that if a relatively low uncertainty threshold is chosen, incorporating OpenMax is recommended. Otherwise, its application may be unnecessary.

Many open-set recognition techniques exist, and this study examined only one. Future studies might consider evaluating a wider range of open-set methods to determine which performs the best for CNN based MP spectral classification.

Overall, the SoftMax model performed very well on the test set. At the class level, however, we observed that some spectra from classes other than PEVA or POM were predicted as PEVA or POM. Closer inspection showed that a small portion of these errors stemmed from human mistakes (e.g., mislabeling) and could have been avoided; the remainder were examined with SHAP.

Based on the SHAP analysis results, the mechanism behind misclassification was essentially the same for spectra predicted as PEVA and for those predicted as POM. In both cases, a small set of local spectral intervals exerted a very strong positive influence on the logit of a single known class, while the remaining wavenumbers contributed weaker, often negative, evidence that only partially offset this effect. As a result, the final logit for the incorrect known class remained high, and the corresponding SoftMax probability easily exceeded the threshold that was intended to filter out unknown spectra. This led to misclassifications with high apparent confidence.

This pattern suggests that the CNN model has learned to treat a few localized intensity patterns as defining features of each known polymer. When an unknown spectrum happens to exhibit local structures within the same short wavenumber intervals that resemble these learned patterns, while still differing in other regions, the model assigns disproportionate weight to these familiar local intervals and largely discounts contradictory information elsewhere in the spectrum. In other words, the prediction is driven by a handful of localized features, and the network does not respond adequately to global inconsistencies across the full spectrum.

To address this behavior, a feasible strategy is to include recurrently misclassified materials such as CA in subsequent training, either as explicit known classes or as labeled hard negative examples. By exposing the network to these locally similar but globally different spectra with correct supervision, the model can learn to assign lower PEVA or POM logits and reduced maximum SoftMax probabilities in such cases, which should in turn, reduce the frequency of high confidence misclassifications.

In summary, the low precision observed for PEVA and POM is attributed partly to human errors ( $\sim 13\%$ ) and primarily to the model’s over-reliance on a few PEVA- or POM-like local spectral intervals in unknown spectra ( $\sim 87\%$ ).

At the theoretical level, the setting of the tail size hyperparameter in OpenMax still leaves room for improvement. In the current implementation, as per Bendale and Boulton<sup>33</sup>, all classes used the same tail size value when fitting the Weibull distribution. This implicitly assumes that the tail behaviour of intra-class distance distributions is similar across classes. In practice, however, different classes often exhibit markedly different intra-class variability. Some classes have compact distributions with short tails, whereas others are more dispersed and exhibit longer tails. Using a single tail size value across all classes may therefore introduce systematic bias. Compact classes may suffer from overfitting of the tail, while dispersed classes may experience underfitting.

Future research could address this limitation by tuning the tail size separately for each class, which would allow the Weibull fitting to capture class-specific variability more accurately and improve calibration. This strategy has already been suggested in the image classification domain<sup>42</sup>.

As shown in Table 5, our CNN models achieved higher classification accuracies on Test set I compared to the models reported in two benchmark studies. This suggests better generalization performance, which is likely due to the higher intra-class diversity of our dataset, as detailed in Table 2.

## Methods

### Data collection

The datasets used in this study include both data collected in our lab and data obtained from freely available online sources.

In our lab, the mIRage IR microscope (also known as the O-PTIR microscope, Photothermal Spectroscopy Corp, Santa Barbara, California, USA) was used to acquire IR spectral data from reference polyamide (PA), PC, PE, PET, PP, PS, PSU, PTFE, PVC. In addition, PA microparticles released from teabags, as per Xu et al.<sup>43</sup>, and PP microparticles released from baby bottles, as per Su et al.<sup>44</sup>, were utilized. Additionally, Nicolet iN10 (Thermo Fisher Scientific) and Nicolet iS50 (Thermo Fisher Scientific) were used to acquire IR spectra from commercial polymeric films, including PA, PE, PET, PP, PS, PSU, PTFE, PVC, PLA, PU, PVOH, silicone rubber. These Nicolet instrument acquired spectra had been used in two previous studies<sup>45,46</sup>. All the spectra collected in our laboratory were manually labeled based on expert knowledge and known source materials.

As for the publicly available datasets, we sourced datasets from Zhu et al.<sup>35</sup> and OpenSpecy<sup>32</sup>. IR spectra of virgin plastic particles as well as plastic particles containing additives and dyes were acquired by Zhu et al.<sup>35</sup> using both FPA-FTIR and ATR-FTIR techniques. Based on their collected data, the authors developed a CNN-based model named PlasticNet, which achieved a classification accuracy of over 95%. Another growing and open database is OpenSpecy, which includes a huge number of IR spectra of various (micro)plastics and non-plastics collected from a wide range of sources. It continuously expands as users and developers add new data to it. We used the February 2024 updated version of OpenSpecy, which includes 8769 IR spectra of (micro)plastics from 15 sources and 8693 IR spectra of non-plastics from 12 sources (i.e., organizations and individuals).

Data cleaning was conducted following data collection, during which some spectra with ambiguous labels were removed. For example, spectra labeled as “plastic” without specifying the plastic type were excluded (removed 1101 spectra in this way). Subsequently, 12,772 spectra remained from our own sources, 7253 spectra remained from the dataset of Zhu et al.<sup>35</sup>, and 16,636 spectra remained from OpenSpecy.

Next, a preliminary examination of the remaining data was conducted, where we identified a total of 2666 classes, including 152 plastic classes and 2514 non-plastic classes. We ultimately selected 18 plastic classes for model development (Supplementary Table 4), as they represent plastic types that are commonly found in daily life and reported most frequently in recent MP studies. These 18 classes were used for model development, and we refer to them herein as the known classes. The remaining 2648 classes, comprising a total of 9419 spectra, were not used in model development, and we refer to them as the unknown classes here (see Supplementary Table 5 for their labels).

### Data augmentation

Data augmentation was used herein to expand the effective number of labeled samples by simulating various realistic perturbations and expected variations. Two types of data augmentation were used here, entitled Type I & II, as described below.

Type I data augmentation reuses spectra with broad spectral ranges by selecting subsets of narrower, specific ranges (i.e., reusing long data). Specifically, an analysis of all collected spectra showed that there were 16 distinct spectral ranges (Supplementary Table 1), with the broadest range being 400–4000  $\text{cm}^{-1}$  and the narrowest range being 769–1801  $\text{cm}^{-1}$ . Reusing long data means that for a spectrum with a broad spectral range, not only the full

spectrum itself, but also the sections within narrower, specific spectral ranges are used for model development (Supplementary Table 1). Type I data augmentation allowed the CNN model to effectively handle inputs across varying spectral ranges. Relevant results are presented in the “Effects of Type I data augmentation” section.

Type II data augmentation includes operations designed to simulate additive and multiplicative effects, such as: adding Gaussian noise (scaled by 10% of the spectrum’s standard deviation) to the spectrum; applying a random proportional rescaling of the entire spectrum (i.e., a  $\pm 10\%$  change in overall signal intensity); modifying the baseline of the spectrum by introducing a random linear tilt. The tilt was applied by adding a slope to the quadratic baseline, with tilt angles ranging approximately from  $-26.57$  to  $+26.57$  degrees. These operations simulate realistic sources of variability commonly observed in IR spectroscopy and were implemented as per Bjerrum et al.<sup>19</sup>, with slight modifications. The addition of Gaussian noise mimics detector and electronic noise, small fluctuations in signal acquisition, and environmental interference. The multiplicative scaling reflects variations in overall signal intensity caused by factors such as differences in sample thickness, optical path length, or instrument gain. The random baseline tilt reproduces the effects of scattering, baseline drift, and imperfect background correction, which can occur during sample preparation or measurement. To the best of our knowledge, Type I data augmentation is proposed here for the first time, whereas Type II data augmentation is a widely used approach in spectroscopic ML.

The data from the 18 known classes were randomly partitioned into a training set, Validation Set I, Validation Set II, and a test set. All 9419 spectra from the unknown classes were also divided, with 1/3 assigned to Validation Set II and 2/3 to the test set (see Supplementary Table 6 for the number of spectra from each class assigned to each subset). The two validation sets served different purposes: Validation Set I was used during model training for parameter tuning and early stopping, while Validation Set II was specifically used to determine the optimal uncertainty threshold range and to identify the distance measure and tail size required for Weibull fitting when applying OpenMax. Note: As the test set included data from both known and unknown classes, for convenience in results reporting and discussion, we refer to the subset of known-class spectra as Test Set I, and the subset of unknown-class spectra as Test Set II.

Next, data augmentation was applied to the training and Validation Set I to ensure that each class contained approximately 1000 spectra in the training set and approximately 200 spectra in Validation Set I. Specifically, Type I augmentation was first applied to each class in the training set. If a class exceeded 1000 spectra after Type I augmentation, random samples were removed to reduce the total to approximately 1000, and no further augmentation (i.e., Type II) was applied. If the resulting number was already close to 1000, Type II was also skipped. For classes with fewer than 1000 spectra after Type I, we applied additional Type II augmentation to reach approximately 1000. The same procedure was applied to Validation Set I, with the only difference being that the target number per class was set to approximately 200 instead of 1000. The final number of spectra in each subset is summarized in Supplementary Table 6.

### Other data treatments

After data augmentation and prior to model training, a series of pre-processing steps was sequentially applied to all spectra from both known and unknown classes. First, all spectra were carefully reviewed by experts to verify that they were recorded in absorbance units. Spectra in transmittance units were converted to absorbance. Second, all spectra were normalized to a range between 0 and 1 using min-max normalization. Third, spectra with a spectral range narrower than 400–4000  $\text{cm}^{-1}$  were zero-padded to extend their coverage to 400–4000  $\text{cm}^{-1}$ <sup>23</sup>. Fourth, all spectra were standardized to have a uniform spectral resolution of 8  $\text{cm}^{-1}$  using interpolation and/or extrapolation, resulting in a 451-dimensional vector for each spectrum. The use of interpolation and/or extrapolation was necessary to ensure that all inputs shared the same dimensionality, as this is a fundamental requirement of CNN architectures. The resolution of 8  $\text{cm}^{-1}$  was selected because it is the

predominant spacing in our dataset. It is important to note, however, that resampling may introduce artifacts, such as the smoothing of sharp absorption features or subtle alterations in peak shapes.

It is also important to note that, according to previous studies<sup>18,19</sup>, CNNs are capable of learning from raw spectral data without explicit pre-processing such as baseline correction or smoothing. Qualitative investigations have demonstrated that CNN kernel activations can perform functions analogous to traditional signal processing operations, including smoothing, derivative extraction, thresholding, and spectral region selection<sup>19</sup>. Hence, no baseline correction or additional denoising was applied to the data in the present study.

### Model development

CNN models developed for spectral classification tasks related to MP research include one-dimensional (1D) and two-dimensional (2D) CNN models<sup>17,23</sup>. 1D CNN models are constructed directly from spectral vectors and are used on spectral vectors, whereas 2D CNN models are trained and used based on 2D images that can be generated from spectral vectors. Both 1D and 2D CNN models have demonstrated excellent and comparable performance in MP spectral classification<sup>23</sup>. However, 2D CNN models are less efficient in that they require an additional step of converting spectral vectors into images, which can be time-consuming, particularly when large datasets are involved. Consequently, here, we chose to focus on 1D CNN models.

Common CNN network architectures include LeNet, AlexNet, VGG16, and ResNet, each with increasing complexity (e.g., more convolutional layers and parameters), resulting in a stronger ability to handle more sophisticated classification tasks but also demanding greater computational resources and longer training times. Previous studies have demonstrated that CNN models based on an architecture as simple as LeNet already perform well for spectral classification tasks<sup>18</sup>. Here, we adopted the network architecture proposed by Liu et al.<sup>18</sup>, which is similar to LeNet, and made slight modifications to it for CNN model training.

The CNN network architecture used in this study is illustrated in Fig. 6a. As seen, this architecture has three convolutional layers, each followed by a max-pooling layer. The combination of convolutional and pooling layers hierarchically extracts features or patterns from the input spectrum<sup>18</sup>. The extracted features are subsequently passed through three fully connected (FC) layers and finalized by a SoftMax layer, which produces a probability vector which sums to 1. We refer to the final FC layer (immediately before the SoftMax layer) as the Activation Vector (AV). The AV stores the logit for each class (denoted as  $S_i$ ).

The dimensionality of the AV and the SoftMax layer of this network architecture is the same as the number of known classes included in model training. For example, in Fig. 6a, the AV and the SoftMax layer are 18-dimensional because the illustration is based on the current work, where 18 known classes are included.

To mitigate against overfitting, batch normalization and dropout were applied in certain layers, as indicated in Fig. 6a.

In model training, the Adam optimizer was used to optimize the model parameters. The cross-entropy loss of Validation Set I was monitored by EarlyStopping (patience = 5). Model training was terminated when the loss of Validation Set I stabilized, and this marked the production of the model. Google Colab was the platform used for model training and validation, and the T4 GPU provided by Google Colab was used to speed up the process.

Note: Based on the network architecture and the training process described in this section, two models were trained: one was developed on all 18 known classes, while the other was developed on only six classes from the 18 classes.

The 6-class model was specifically developed for investigating the effects of Type I data augmentation.

As for the 18-class model, we refer to it as “the SoftMax model”. In the next paragraphs, we will further describe how OpenMax was applied to “the SoftMax model” to obtain a new model, which we refer to as “the OpenMax model.”

The OpenMax model was developed by applying the OpenMax technique directly to the SoftMax model according to Bendale and Boulton<sup>33</sup>, with slight modification. Briefly, step one: the SoftMax model was used to classify the training data. Step two: the AVs of the correctly classified samples were extracted. Step three: for each class, the mean activation vector (MAV) was computed. Step four: for each class, the distance between each AV and the MAV was calculated, resulting in a set of distances. Step five: for each class, a Weibull cumulative distribution function (CDF) was fitted based on the distance set. Weibull fitting is the core of OpenMax, and is motivated by Extreme Value Theory<sup>33</sup>, which justifies modeling the largest AV distances within each known class to estimate the likelihood that a new input is an outlier and thus potentially belongs to an unknown class.

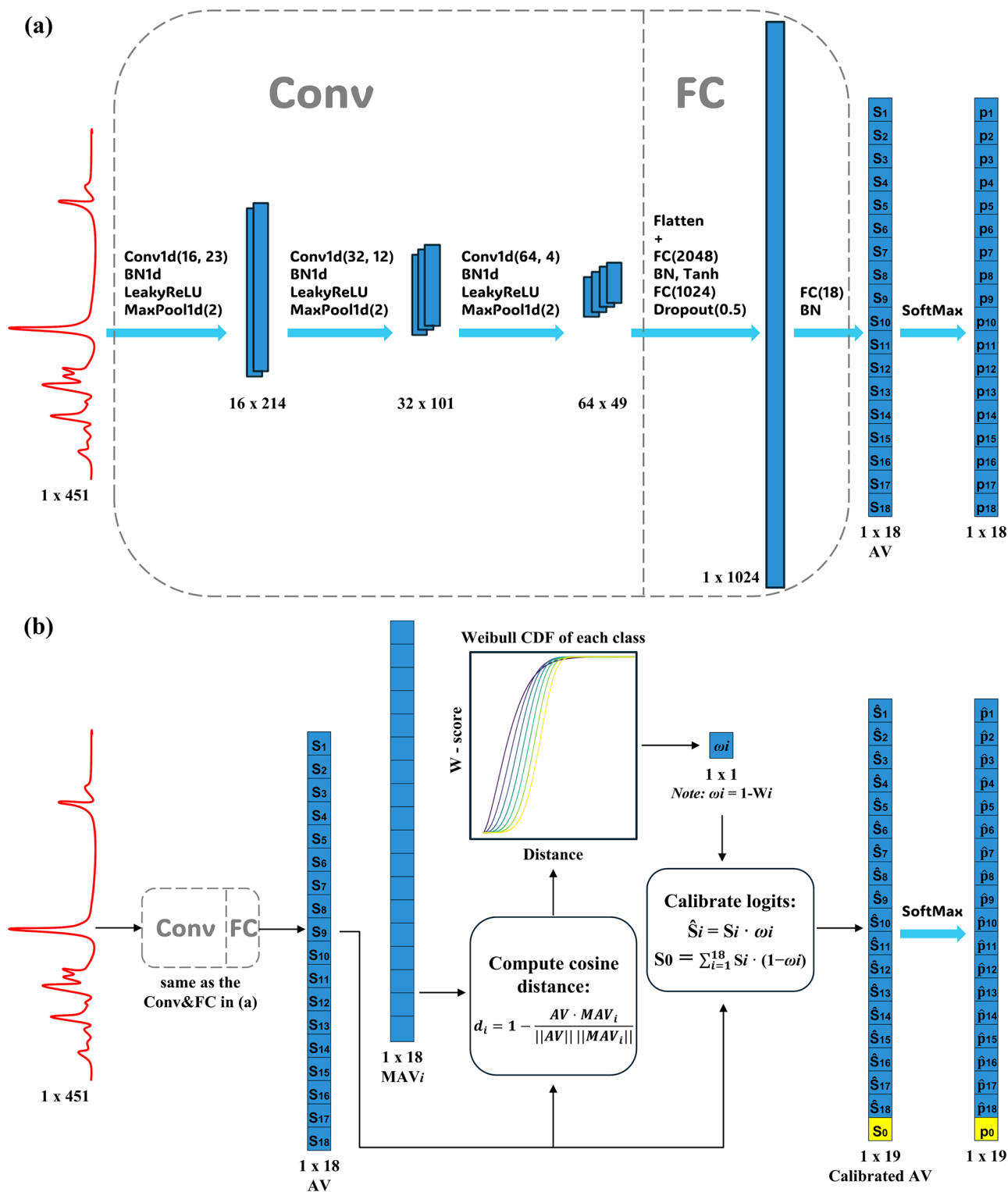
The implementation of the five steps described above resulted in the production of the OpenMax model. Among these five steps, two hyperparameters are particularly critical. The first is the distance measure described in step four. According to the OpenMax package, the available options are cosine distance, Euclidean distance, and Euclidean–cosine distance. The Euclidean–cosine distance is defined as

$$d_{\text{Euclidean-Cosine}} = \frac{1}{n} \times d_{\text{Euclidean}}(AV, MAV) + d_{\text{Cosine}}(AV, MAV) \quad (4)$$

where  $1/n$  is believed to scale the Euclidean distance to the same range as the cosine distance. However, this was not explained in either the original paper or the provided code. The authors, Bendale and Boulton<sup>33</sup>, used a factor of  $1/200$ , but did not specify how this value was determined. We reasonably assume that this coefficient was intended to normalize the Euclidean distance so that both distance components would be of comparable magnitude. In our dataset, the maximum Euclidean distance between AV and MAV was 19; therefore, we used  $1/9.5$  as the scaling factor to ensure that the Euclidean term was normalized to approximately the same range as the cosine distance.

The second hyperparameter is the tail size used in step five for estimating the parameters of the Weibull distribution. Following Bendale and Boulton<sup>33</sup>, these two hyperparameters should be selected jointly, and the optimal combination should be determined through validation using grid search. In this study, we conducted grid search over the three distance measures and tail size values from 15 to 40 in increments of 5 (Supplementary Table 7) and found that the best-performing configuration was cosine distance with a tail size value of 35 (All grid search results are presented Supplementary Figs. 2, 3 and 4).

Figure 6b illustrates the prediction phase of the OpenMax model. As seen, when an input is given to the OpenMax model, it is processed through the convolutional and FC layers, generating an AV. Subsequently, the cosine distance between the AV and the MAV of each known class is computed. These distances are then passed through the corresponding Weibull CDFs, finally yielding class-wise probability adjustment factors (Let  $\omega_i$  denote the adjustment factor for the  $i$ th class). These adjustment factors are then used to calibrate the original AV. Specifically, the cosine distance between the AV of the given input and the MAV of the  $i$ th class is fed into the Weibull CDF of the  $i$ th class, producing  $\omega_i$ . This  $\omega_i$  is then applied to calibrate the logit of the  $i$ th class (denoted as  $S_i$  here, which becomes  $\hat{S}_i$  after calibration). After calibrating all known-class logits, the model computes the unknown-class logit ( $S_0$ ). This value is derived based on the deviations of the AV of the given input from the known-class MAVs. The calibrated AV, now consisting of the calibrated logits and the unknown-class logit, is fed into a SoftMax layer, resulting in a probability vector where the sum of probabilities for the known classes plus the unknown class equals 1. This is in stark contrast to the probability vector produced by the SoftMax model, where only known classes are considered (probabilities for the known classes summing to 1). In essence, OpenMax “compresses” the probabilities of known classes, allocating the remaining probability mass to the unknown class. This mechanism enables OpenMax to detect inputs from unknown classes, or more intuitively, inputs “far from” the known training data.



**Fig. 6 | Model development illustration based on an example where 18 known classes are included.** **a** is the network architecture of the SoftMax model; **b** is the prediction phase of the OpenMax model. Conv1d ( $m, n$ ) refers to a one-dimensional convolutional layer, where  $m$  is the number of convolutional filters and  $n$  is the filter size. The stride of the convolution operation is 1. BN1d refers to one-dimensional batch normalization. MaxPool1d ( $k$ ) refers to a maximum pooling layer with a kernel size of  $k$  and a stride of  $k$ . FC ( $p$ ) stands for fully connected layers, where  $p$

denotes the number of neurons in the layer. Dropout (0.5) refers to a dropout layer with a rate of 50%.  $\omega_i$  represents the probability that the input does not belong to the  $i$ th class, and is defined as  $\omega_i = 1 - W_i$ , where  $W_i$  is the probability that the input belongs to the  $i$ th class. AV stands for Activation Vector. An AV stores the logit for each class, denoted as  $S_i$ , which becomes  $\hat{S}_i$  after calibration.  $MAV_i$  stands for the mean activation vector of the  $i$ th class.  $S_0$  is the unknown-class logit.

In more detail, OpenMax “compresses” the probabilities of the known classes by dynamically adjusting their logits ( $S_i$ ) using  $\omega_i$ , which represents the likelihood that the given input belongs to the typical distribution of the  $i$ th class; specifically, when an input’s AV is close to the MAV of the  $i$ th class (i.e., when  $\omega_i$  approaches 1), the original logit remains largely unchanged. Conversely, when the input’s AV deviates significantly from the MAV (i.e., when  $\omega_i$  approaches 0), the corresponding logit is suppressed, reducing the model’s confidence that the sample is in that class.

As mentioned earlier and illustrated in Fig. 6, regardless of whether the model is the SoftMax model or the OpenMax model, an input is ultimately transformed into a probability vector. Each element in this vector represents the predicted probability that the input belongs to a specific class. According to the standard approach, the class corresponding to the highest value (i.e., the highest probability) in the vector is taken as the model’s final prediction.

The uncertainty threshold is a manually defined threshold applied to this probability vector. If the probability of the model’s predicted outcome (i.e., the highest probability in the vector) does not exceed this threshold, the model’s prediction is considered (by who set the threshold to be) unreliable. Typically, such low-probability predictions are often rejected or flagged as “uncertain”. To the best of our knowledge, no existing studies that employ CNN models for spectral classification in MP research have used the uncertainty threshold.

In this study, we define that any input whose maximum predicted probability does not exceed the uncertainty threshold assigned a label of “unknown”, or placed into a class named “unknown”, rather than simply rejecting such uncertain predictions.

### Data availability

The datasets covered in this study are accessible at the following link: <https://drive.google.com/drive/folders/11MofhjEchgZelWPcCHUvIMRPNEQPALfUO?usp=sharing>. In case the link becomes unavailable, please contact the corresponding author.

### Code availability

The code used in this study has been annotated and is accessible at the following link: <https://drive.google.com/drive/folders/11MofhjEchgZelWPcCHUvIMRPNEQPALfUO?usp=sharing>. In case the link becomes unavailable, please contact the corresponding author.

Received: 7 July 2025; Accepted: 20 November 2025;

Published online: 02 February 2026

### References

- Manzo, S. & Schiavo, S. Physical and chemical threats posed by micro(nano)plastic to sea urchins. *Sci. Total Environ.* **808**, 152105 (2022).
- Ullah, R. et al. Micro (nano) plastic pollution in terrestrial ecosystem: emphasis on impacts of polystyrene on soil biota, plants, animals, and humans. *Environ. Monit. Assess.* **195**, 252 (2023).
- Gasperi, J. et al. Microplastics in air: are we breathing it in? *Curr. Opin. Environ. Sci. Health* **1**, 1–5 (2018).
- Mortensen, N. P., Fennell, T. R. & Johnson, L. M. Unintended human ingestion of nanoplastics and small microplastics through drinking water, beverages, and food sources. *NanoImpact* **21**, 100302 (2021).
- Du, F., Cai, H., Zhang, Q., Chen, Q. & Shi, H. Microplastics in take-out food containers. *J. Hazard. Mater.* **399**, 122969 (2020).
- Li, B. et al. Fish ingest microplastics unintentionally. *Environ. Sci. Technol.* **55**, 10471–10479 (2021).
- Uurasjärvi, E., Sainio, E., Setälä, O., Lehtiniemi, M. & Koistinen, A. Validation of an imaging FTIR spectroscopic method for analyzing microplastics ingestion by Finnish lake fish (*Perca fluviatilis* and *Coregonus albula*). *Environ. Pollut.* **288**, 117780 (2021).
- Azeem, I. et al. Uptake and accumulation of nano/microplastics in plants: a critical review. *Nanomaterials* **11**, 2935 (2021).
- Amato-Lourenço, L. F. et al. Microplastics in the olfactory bulb of the human brain. *JAMA Netw. Open* **7**, e2440018–e2440018 (2024).
- Leslie, H. A. et al. Discovery and quantification of plastic particle pollution in human blood. *Environ. Int.* **163**, 107199 (2022).
- Liu, S. et al. Detection of various microplastics in placentas, meconium, infant feces, breastmilk and infant formula: a pilot prospective study. *Sci. Total Environ.* **854**, <https://doi.org/10.1016/j.scitotenv.2022.158699> (2023).
- Tarafdar, A., Xie, J., Gowen, A., O’Higgins, A. C. & Xu, J.-L. Advanced optical photothermal infrared spectroscopy for comprehensive characterization of microplastics from intravenous fluid delivery systems. *Sci. Total Environ.* **929**, 172648 (2024).
- Veerasingam, S. et al. Contributions of Fourier transform infrared spectroscopy in microplastic pollution research: a review. *Crit. Rev. Environ. Sci. Technol.* **51**, 2681–2743 (2021).
- Primpke, S. et al. Critical assessment of analytical methods for the harmonized and cost-efficient analysis of microplastics. *Appl. Spectrosc.* **74**, 1012–1047 (2020).
- Moses, S. R. et al. Comparison of two rapid automated analysis tools for large FTIR microplastic datasets. *Anal. Bioanal. Chem.* **415**, 2975–2987 (2023).
- Hufnagl, B. et al. A methodology for the fast identification and monitoring of microplastics in environmental samples using random decision forest classifiers. *Anal. Methods* **11**, 2277–2285 (2019).
- Ishmukhametov, I., Batasheva, S. & Fakhrullin, R. Identification of micro- and nanoplastics released from medical masks using hyperspectral imaging and deep learning. *Analyst* **147**, 4616–4628 (2022).
- Liu, J. et al. Deep convolutional neural networks for Raman spectrum recognition: a unified solution. *Analyst* **142**, 4067–4074 (2017).
- Bjerrum, E. J., Glahder, M. & Skov, T. Data augmentation of spectral data for convolutional neural network (CNN) based deep chemometrics. *arXiv preprint* <https://doi.org/10.48550/arXiv.1710.01927> (2017).
- Ai, W. et al. Application of hyperspectral imaging technology in the rapid identification of microplastics in farmland soil. *Sci. Total Environ.* **807**, 151030 (2022).
- Luo, Y. et al. Component identification for the SERS spectra of microplastics mixture with convolutional neural network. *Sci. Total Environ.* **895**, 165138 (2023).
- Xu, L. et al. Study on detection method of microplastics in farmland soil based on hyperspectral imaging technology. *Environ. Res.* **232**, 116389 (2023).
- Liu, Y., Yao, W., Qin, F., Zhou, L. & Zheng, Y. Spectral classification of large-scale blended (Micro) plastics using FT-IR raw spectra and image-based machine learning. *Environ. Sci. Technol.* **57**, 6656–6663 (2023).
- Ren, L. et al. Identification of microplastics using a convolutional neural network based on micro-Raman spectroscopy. *Talanta* **260**, 124611 (2023).
- Ai, W., Chen, G., Yue, X. & Wang, J. Application of hyperspectral and deep learning in farmland soil microplastic detection. *J. Hazard. Mater.* **445**, 130568 (2023).
- Zhang, W. et al. A deep one-dimensional convolutional neural network for microplastics classification using Raman spectroscopy. *Vib. Spectrosc.* **124**, 103487 (2023).
- Qin, Y., Qiu, J., Tang, N., He, Y. & Fan, L. Deep learning analysis for rapid detection and classification of household plastics based on Raman spectroscopy. *Spectrochim. Acta Part A: Mol. Biomol. Spectrosc.* **309**, 123854 (2024).
- Zeng, G. et al. Deep convolutional neural networks for aged microplastics identification by Fourier transform infrared spectra classification. *Sci. Total Environ.* **913**, 169623 (2024).

29. Wei, D., Zhou, B., Torrabla, A. & Freeman, W. Understanding intra-class knowledge inside CNN. Preprint at <https://arxiv.org/abs/1507.02379#> (2015).
30. Shi, H. et al. Embedding deep metric for person re-identification: a study against large variations. In *Proc. Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings 732–748* (Springer, 2016).
31. Thompson, R. C. et al. Twenty years of microplastic pollution research — what have we learned? *Science* **386**, ead12746 (2024).
32. Cowger, W. et al. Microplastic spectral classification needs an open source community: open specy to the rescue!. *Anal. Chem.* **93**, 7543–7548 (2021).
33. Bendale, A. & Boulton, T. E. Towards open set deep networks. Preprint at <https://arxiv.org/abs/1511.06233> (2016).
34. Miller, D., Sunderhauf, N., Milford, M. & Dayoub, F. Class anchor clustering: a loss for distance-based open set recognition. In *Proc. IEEE/CVF Winter Conference on Applications of Computer Vision 3570–3578* (IEEE, 2021).
35. Zhu, Z., Parker, W. & Wong, A. Leveraging deep learning for automatic recognition of microplastics (MPs) via focal plane array (FPA) micro-FT-IR imaging. *Environ. Pollut.* **337**, 122548 (2023).
36. Xie, J., Gowen, A., Xu, W. & Xu, J. Analysing micro- and nanoplastics with cutting-edge infrared spectroscopy techniques: a critical review. *Anal. Methods* **16**, 2177–2197 (2024).
37. Choi, E., Choi, Y., Lee, H., Kim, J. W. & Oh, H. B. Development of a machine-learning model for microplastic analysis in an FT-IR microscopy image. *Bull. Korean Chem. Soc.* **45**, 472–481 (2024).
38. Adelnia, H., Bidsorkhi, H. C., Ismail, A. & Matsuura, T. Gas permeability and permselectivity properties of ethylene vinyl acetate/sepiolite mixed matrix membranes. *Sep. Purif. Technol.* **146**, 351–357 (2015).
39. Löder, M. G. & Gerdtts, G. Methodology used for the detection and identification of microplastics—a critical appraisal. In *Proc. Marine Anthropogenic Litter* (eds. Bergmann, M., Gutow, L., & Klages, M.) 201–227 (Springer, 2015).
40. Löder, M. G. et al. Enzymatic purification of microplastics in environmental samples. *Environ. Sci. Technol.* **51**, 14283–14292 (2017).
41. Clough, M. E. et al. Enhancing confidence in microplastic spectral identification via conformal prediction. *Environ. Sci. Technol.* **58**, 21740–21749 (2024).
42. Giusti, E., Ghio, S., Oveis, A. H. & Martorella, M. Proportional similarity-based Openmax classifier for open set recognition in SAR images. *Remote Sens.* **14**, 4665 (2022).
43. Xu, J.-L., Lin, X., Hugelier, S., Herrero-Langreo, A. & Gowen, A. A. Spectral imaging for characterization and detection of plastic substances in branded teabags. *J. Hazard. Mater.* **418**, 126328 (2021).
44. Su, Y. et al. Steam disinfection releases micro (nano) plastics from silicone-rubber baby teats as examined by optical photothermal infrared microspectroscopy. *Nat. Nanotechnol.* **17**, 76–85 (2022).
45. Xu, J.-L., Lesniak, A. & Gowen, A. A. Predictive modeling of the in vitro responses of preosteoblastic MC3T3-E1 cells on polymeric surfaces using Fourier transform infrared spectroscopy. *ACS Appl. Mater. Interfaces* **12**, 24466–24478 (2020).
46. Xu, J.-L. & Gowen, A. A. Time series Fourier transform infrared spectroscopy for characterization of water vapor sorption in hydrophilic and hydrophobic polymeric films. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **250**, 119371 (2021).
47. de Carvalho, A. R. et al. Urbanization and hydrological conditions drive the spatial and temporal variability of microplastic pollution in the Garonne River. *Sci. Total Environ.* **769**, 144479 (2021).
48. Davidson, J. et al. Polymer characterization of submerged plastic litter from Lake Tahoe, United States. *Appl. Spectrosc.* **77**, 1240–1252 (2023).
49. De Frond, H., Rubinovitz, R. & Rochman, C. M. muATR-FTIR spectral libraries of plastic particles (FLOPP and FLOPP-e) for the analysis of microplastics. *Anal. Chem.* **93**, 15878–15885 (2021).

## Acknowledgements

Funding for this research was provided by the Science Foundation Ireland (SFI)-Irish Research Council Pathway Programme Proposal ID 21/PATH-S/9290. We thank Professor Sedat Gündoğdu for kindly providing 12 IR spectra, which were used as an external test set after the model development. Of these, 10 spectra were correctly classified. These results are not included in the main text but are noted here for completeness.

## Author contributions

J.X. and J.-L.X. conceived the study and developed the central ideas. J.X. was responsible for data acquisition, code development, analysis, visualization, and manuscript writing. J.-L.X. and A.G. provided supervision and revision of the manuscript. J.-L.X. secured funding for the project.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s44454-025-00022-0>.

**Correspondence** and requests for materials should be addressed to Junhao Xie.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025