ACCURACY VERSUS COMPUTATIONAL TIME: A BENCHMARK OF CLASSIFICATION ALGORITHMS FOR BLACK PLASTIC IDENTIFICATION USING MWIR HYPERSPECTRAL IMAGING

Marco Papetti, Marco Diani, Marcello Colledani

Politecnico di Milano, Mechanical Department, via La Masa 1, Milan – 20156, Italy

ABSTRACT

Black-coloured polymers are undetectable by near-infrared sorting systems because of carbon-black absorption, and mid-wavelength infrared (MWIR) hyperspectral imaging has emerged as a viable alternative for their in-line identification. However, the weak MWIR spectra and the high data throughput make the choice of the classification algorithm a key implementation choice. This work presents a benchmark of sixteen chemometric, machine- and deep-learning classifiers for black-plastic identification, trained and evaluated under identical conditions on two independently acquired MWIR hyperspectral datasets, using both pixel-wise and object-wise metrics. In this work tree-based ensembles achieve the best performance (Random Forest, object-wise F1 = 0.81), consistently outperforming deeper architectures; notably, the shallow SCNN1D surpasses the deep DCNN1D in both accuracy and speed. Lightweight models therefore provide the most favourable accuracy-efficiency trade-off for real-time sorting, while residual errors remain confined to intra-family sub-type confusion.

Index Terms - Black plastic identification, Mid-wavelength infrared, Hyperspectral imaging, Classification algorithms.

1. INTRODUCTION AND OBJECTIVES

Accurate identification is a crucial step in improving plastic recycling rates. This process can be seen as the combination of two main factors: the sensing technology used to capture the material signal and the classification algorithm adopted for its identification.

From the sensing perspective, NIR spectroscopy is currently the industrial standard for automated polymer characterization due to its cost-effectiveness and rapid analysis capabilities. However, the carbon black absorption across the visible and NIR spectrum causes its inability to detect black-coloured polymers [1]. A spectral shift towards mid-wavelength infrared (MWIR) proved to overcome this limitation. Hence, the implementation of sensing technologies relying on this spectral range has taken place, with hyperspectral imaging (HSI) being one of the most promising. MWIR HSI has indeed emerged as a suitable technology for in-line black plastic identification [2], [3], and it has already been adopted at industrial level by Steinert and Tomra through the systems UniSort BlackEye and AUTOSORT Black. However, despite its advantages, this technology captures lower reflected signal and reduced signal-to-noise ratio compared to NIR/SWIR systems and presents a high computational burden associated with hyperspectral data processing. These characteristics make classification particularly challenging, since algorithms must extract discriminative information while remaining computationally efficient for in-line sorting.

On the algorithmic side, early studies regarding plastic classification mainly relied on linear multivariate methods such as PCA and PLS-DA [4]. Hierarchical PLS-DA (H-PLSDA) architectures were later introduced to improve scalability across multiple polymer classes [5]. More recently, machine learning and deep learning approaches - including ANNs, CNNs, SCNNs, and more complex DCNNs - have demonstrated excellent classification performance on MIR spectral data [2], [3], [6], [7]. Despite these promising results, direct comparisons among classification algorithms remain difficult because different studies employ different datasets, acquisition systems, preprocessing pipelines, and evaluation protocols. Moreover, in industrial practice, classifiers are frequently designed to discriminate only a limited number of polymer classes (e.g., PP vs PE or PP vs ABS, etc.), suggesting potential scalability issues in terms of both accuracy and computational efficiency when the number of target polymers increases. For these reasons, it remains unclear which algorithms provide the best trade-off between accuracy, robustness and computational efficiency for black plastics identification, and a main question emerges: does the potential accuracy gain offered by increasingly complex architectures actually justify the higher computational cost?

To address this gap, this work presents a systematic benchmark of representative state-of-the-art classification algorithms for the identification of black plastics. Each of them is adapted and optimized for the considered application before being trained and evaluated under identical experimental conditions, enabling a fair comparison in terms of precision, recall, F1-score, and computational time. Unlike studies based on laboratory-grade FTIR spectra [7], the proposed benchmark relies on data acquired using an MWIR HSI system. Consequently, the evaluated algorithms are

exposed to the instrumental and environmental variability that characterizes industrial acquisition, including illumination non-uniformity, specular reflections, detector noise, and the already mentioned low signal-to-noise ratio of MWIR measurements. This enables the assessment of not only the classification performance of each method but also its robustness under realistic in-line sorting conditions.

2. EXPERIMENTAL SETUP AND METHODOLOGY

2.1. Acquisition system

The hyperspectral images analysed in this study were acquired using a Specim FX50 hyperspectral camera. The camera operates in the 2720 – 5260 nm spectral range, capturing the signal across 255 spectral bands. In the experimental setup adopted for this work, the camera was positioned 62 cm above the sample, and image acquisition was performed with an integration time (IT) of 0.43 ms, a pixel length of 0.82 mm and a conveyor speed of 117 mm/s.

2.2. Computing environment

All experiments were run on a Lenovo ThinkPad P14s equipped with an Intel Core Ultra 7 155H CPU (16 cores / 22 threads, 31.5 GB RAM) and an NVIDIA RTX 500 Ada Laptop GPU (4 GB), under Windows 11 Pro. Classical classifiers were executed on the CPU (scikit-learn 1.9.0, LightGBM 4.6.0), whereas the neural-network models were trained and evaluated on the GPU (PyTorch 2.6.0, CUDA 12.4).

2.3. Plastic samples, training set and test set

A set of black plastic samples was collected from landfills, scrapyards, and household waste. Each sample was assigned to its polymer class based on the material marking moulded on the parts: ABS, HDPE, LDPE, PA, PET, PP, PP with EPM/EPDM, PP with talc (PP TD) and PS. The samples come from different sectors: packaging, building and construction, end of life vehicles, agriculture, waste from electrical and electronic equipment and household. No surface treatment was applied to preserve the original condition of the pieces, which were only cut into manageable tiles. The evaluation is performed on two independently acquired datasets: the training set comprises 129 sample cutouts from 16 image acquisitions, while the test set comprises 95 cutouts from 10 image acquisitions. The number of sample cutouts for each polymer class in the training and test sets is reported in Table 1. Whenever possible, different cutouts were used to populate the two datasets. For the underrepresented classes, each physical sample was divided into multiple tiles, each yielding a distinct cutout, to enlarge the dataset without reusing the same tile. When this was not feasible, the same physical sample was re-acquired.

2.4. Data processing and classification

2.4.1. Spectra calibration

The acquired spectra were calibrated. To perform this calculation, hyperspectral images of both a high reflectance standard reference, Ref_{100}, and a dark reference, d, are required. The acquisition of a frosted aluminium alloy tile is conducted for the white reference, while for the dark reference, the camera automatically covers the lens during the acquisition. Each of the two images is adopted for conversion of the sample measurement intensity (C) from counts to spectroscopic units according to Equation 1.

$$R = \frac{C - d}{Ref_{100} - d} \qquad \text{Equation 1}$$

2.4.2. Masking procedure

For each hyperspectral image, both in the training and in the test set, a binary mask isolates the plastic samples from the background. The regions of interest (ROIs) were segmented using the Segment Anything Model (SAM) [8] and manually validated by the operator. Since accurate automatic segmentation is not the focus of this work, this manual validation ensures reliable, high-quality masks and prevents segmentation errors from confounding the algorithm evaluation.

2.4.3. Spectra preprocessing

The adopted spectra preprocessing procedure consists in the application of Standard Normal Variate (SNV) followed by a Savitzky-Golay (SG) smoothing filter. The latter has been applied with a first order derivative, a third order polynomial degree and a window size of seven channels. This procedure is widely validated in the literature and has not been changed during the tests.

Table 1. Number of sample cutouts per polymer class in the training and test sets.

	ABS	HDPE	LDPE	PA	PET	PP	PP + EPM/EPDM	PP TD	PS
TRAIN	15	18	12	6	9	31	9	8	21
TEST	11	19	10	6	9	10	10	10	10

2.4.4. Classification algorithms

The algorithms compared in this study are drawn from two groups. The first comprises the main chemometric classifiers established in the plastic-analysis literature (LDA, PLS-DA, SIMCA, hierarchical PLS-DA, KNN, SVM, RF, Bagged Trees), for which [1] serves as the main reference. The second comprises the most recent machine- and deep-learning developments proposed for black plastic discrimination: the BPNN and the deep and shallow one-dimensional CNNs of Yang et al. [7], the MLP and CNN of Roming et al. [2] – building on the architecture of Xia et al. [6] – and the N-BEATS and LightGBM models [9]. For all trained algorithms, an automatic hyperparameter optimization through random search following the approach proposed by [11] was performed. For each algorithm, 10 hyperparameter configurations were evaluated through cross-validation, with balanced accuracy used as the selection criterion to mitigate the effect of class imbalance. The optimization was performed on a stratified subset of spectra representative of each class, and the selected configuration was then used to train the final models on the complete training dataset.

Table 2. Adopted algorithms and reference papers.

Reference Paper	Algorithm
[1]	RF
[1]	KNN
[1]	LDA
[1]	SVM
[1]	PLSDA
[1]	Bagged Trees
[10]	H-PLSDA
[6]	SIMCA
[7]	DCNN1D
[7]	SCNN1D
[7]	BPNN
[2]	MLP
[6]	CNN_1
[2]	CNN_2
[9]	N-BEATS
[9]	LightGBM

3. RESULTS AND DISCUSSION

The algorithms were initially trained and optimized on the presented nine classes (i.e., ABS, HDPE, LDPE, PA, PET, PP, PP with EPM/EPDM, PP with TALC and PS) in order to test the finest classification level achievable with the inspected samples. As a secondary benchmark, the same algorithms were also evaluated with a coarser configuration in which HDPE and LDPE were merged into a single PE class, and the three PP-based compounds were grouped into a single PP class. This resulted in a six-class problem (i.e., ABS, PE, PA, PET, PP and PS). The results obtained with the six classes serve as a reference and are discussed in section 3.4.

3.1. Overall classification performance

The classification performance was evaluated on 95 cutouts and a total of 499,294 spectra, while adopting precision, recall and F1 score as performance metrics. Each metric was computed in a pixel-wise (PW) and an object-wise (OW) fashion; OW predictions were obtained by assigning each cutout the modal class among its constituent pixels. Given the industrial application of the paper, the OW metrics are considered the most representative for assessing algorithm performance, with particular focus on F1 score as it enables fair comparison in imbalanced datasets. Table 3 shows the classification results. Overall, tree-based ensembles (RF, Bagged Trees and LightGBM) achieve the highest F1 scores both on the pixel-wise and on the object-wise classification, with deep architectures failing to outperform them despite their higher capacity. RF was the only algorithm achieving an F1 score higher than 0.80, with Bagged Trees positioning right below this threshold. The negative influence of low signal-to-noise ratio on the classification performance is clearly visible in the PW metrics. Here, the pixels affected by glares or located on the edge of the sample are often misclassified (see Figure 1). However, the OW classification overcomes this limitation for most of the samples.

Table 3. Classification performance metrics for pixel-wise (PW) and object-wise (OW) classification.

Algorithm	PW F1	PW Precision	PW Recall	OW F1	OW Precision	OW Recall
RF	0.66	0.71	0.69	0.81	0.85	0.84
BaggedTrees	0.63	0.66	0.65	0.79	0.84	0.82
LightGBM	0.67	0.74	0.70	0.76	0.85	0.81
CNN_2	0.64	0.72	0.68	0.76	0.82	0.80
SCNN1D	0.60	0.69	0.64	0.74	0.85	0.77
CNN_1	0.63	0.71	0.68	0.73	0.82	0.77
SIMCA	0.61	0.62	0.65	0.72	0.77	0.75
MLP	0.62	0.67	0.66	0.71	0.78	0.73
N-BEATS	0.62	0.66	0.65	0.70	0.75	0.70
BPNN	0.60	0.69	0.65	0.69	0.81	0.73
DCNN1D	0.57	0.68	0.63	0.67	0.80	0.73
LDA	0.53	0.70	0.59	0.67	0.80	0.73
PLSDA	0.52	0.65	0.56	0.63	0.73	0.69
H-PLSDA	0.47	0.61	0.52	0.57	0.66	0.63
SVM	0.51	0.62	0.58	0.54	0.53	0.65
KNN	0.46	0.62	0.49	0.54	0.75	0.59

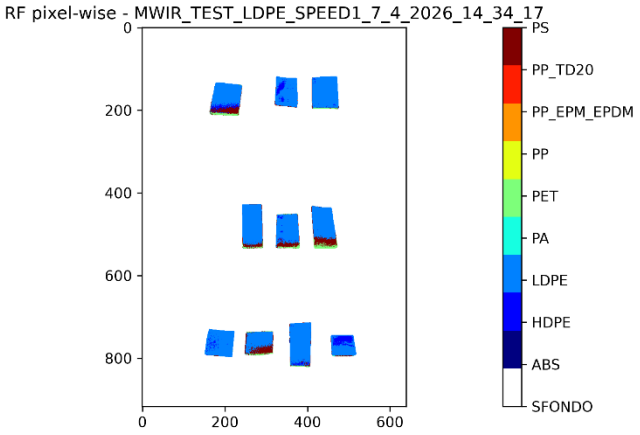


Figure 1. RF pixel-wise classification of LDPE test set.

3.2. Accuracy – Efficiency trade-off

Figure 2 reports the object-wise F1 of each algorithm against its inference time (log scale) on the nine-class problem. The Pareto front - the models not dominated in both accuracy and speed - is populated exclusively by tree-based ensembles and shallow networks, whereas the deep CNNs, N-BEATS and the SVM are strictly dominated. As previously outlined Random Forest achieves the highest accuracy (object-wise F1 score = 0.81) but at a comparatively high cost (15.7 s to classify the whole test set), while Bagged Trees reaches almost the same accuracy (0.79) in 2.1 s. Deep convolutional architectures provide no additional benefit: the best of them (CNN of Roming et al., F1 = 0.76) is matched by LightGBM (0.76) at one third of the inference time (29 s vs 9.6 s), and by the shallow SCNN1D (0.74) at two orders of magnitude lower inference time (29 s vs 0.29 s).

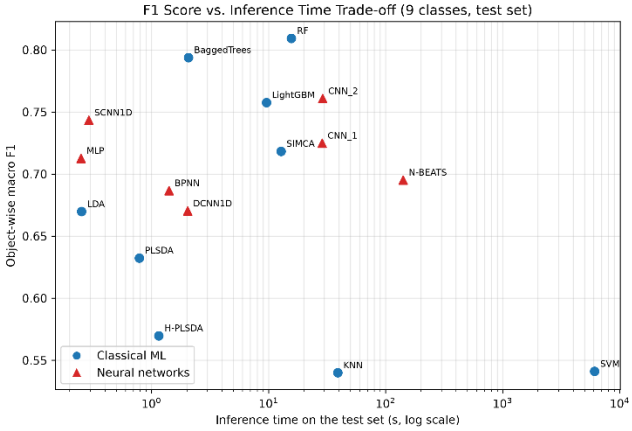


Figure 2. Comparison of object-wise F1 score versus inference time (log scale) on the nine-class classification. Inference time refers to the classification of the whole test set.

This directly answers the central question of this work: based on the obtained results, the additional complexity of the deep networks does not translate into better discrimination and speed; the shallow SCNN1D outperforms the deep DCNN1D and CNN_1 in both accuracy and speed. At the opposite extreme, the SVM combines low accuracy (0.54) with a prohibitive inference time (> 6000 s, ~ 80 spectra/s), making it unsuitable for in-line deployment. Overall, the best accuracy-efficiency trade-off is offered by lightweight models (i.e., tree ensembles and SCNN1D), with near-top accuracy and throughputs compatible with real-time industrial sorting.

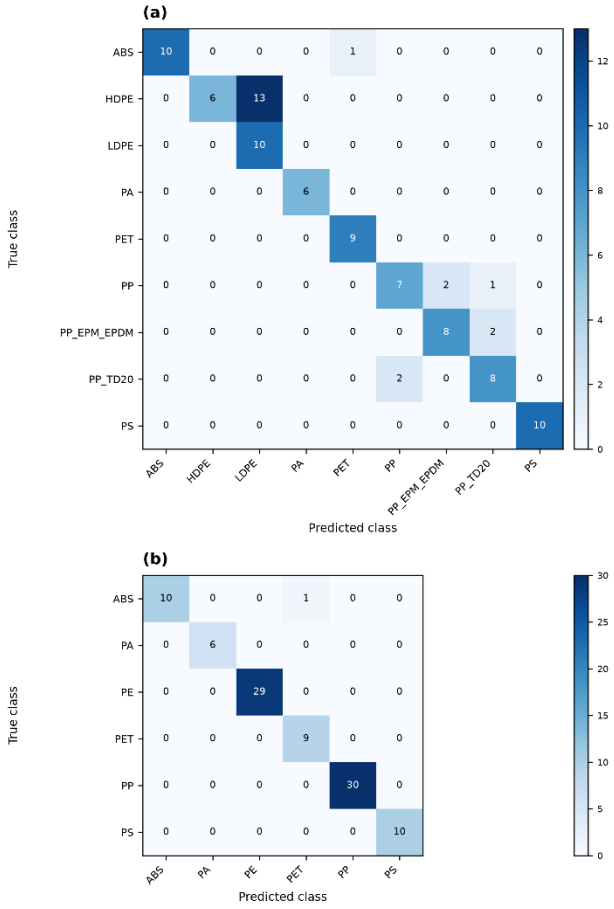


Figure 3. Object-wise confusion matrices obtained with Random Forest for (a) the nine-class and (b) the six-class classification configurations.

3.3. Error analysis

The obtained classification results are strongly influenced by the selected number of classes and their level of detail. The

OW confusion matrix of the RF on the nine-class classification (Figure 3a) shows that the misclassifications are almost exclusively concentrated within the same polymer family. The largest source of error is the confusion between HDPE and LDPE, with 13 of the 19 HDPE samples assigned to LDPE: both are polyethylenes and their mid-wavelength infrared response is dominated by the same CH_2 stretching and bending bands, so that the differences in density and chain branching produce only subtle spectral variations. A similar behaviour is observed among the three polypropylene-based classes (i.e., PP, PP with EPM/EPDM and PP with talc). Conversely, inter-family confusion is almost absent: chemically distinct polymers (ABS, PA, PET, PS) are recognised with no or negligible errors.

These observations motivate a coarser, family-level classification. Indeed, when HDPE and LDPE are merged into a single PE class and the three PP-based classes into a single PP class, the intra-family errors disappear, and the classification becomes nearly perfect: as shown in the six-class confusion matrix of Figure 3b, RF correctly classifies 94 of the 95 samples (OW F1 = 0.98). These results indicate that the remaining classification errors are primarily driven by the intrinsic spectral similarity between polymer sub-types rather than by an inability to distinguish chemically different polymer families.

4. CONCLUSIONS AND FUTURE DEVELOPMENTS

This work presents a benchmark of sixteen chemometric, machine learning, and deep learning algorithms for the identification of black plastics through MWIR hyperspectral imaging, trained and evaluated on independently acquired datasets under realistic acquisition conditions, and assessed at both pixel and object levels.

The results showed that tree-based ensemble methods consistently achieved the best trade-off between classification performance and computational efficiency, with Random Forest providing the highest object-wise F1-score among the evaluated algorithms. Conversely, more complex deep learning architectures did not provide a significant performance advantage despite their higher architectural complexity. These findings indicate that increasing model complexity does not necessarily translate into improved performance under realistic MWIR hyperspectral acquisition conditions, where noisy reflected spectra and industrial constraints play a crucial role.

Future work will focus on assessing the robustness and generalization capability of the evaluated algorithms under different acquisition configurations, including variations in the imaging setup, illumination conditions, and conveyor belt speeds. Extending the benchmark to additional polymer classes and more challenging industrial scenarios will further support the selection of reliable classification methods for real-world in-line sorting applications.

5. REFERENCES

[1] E. R. K. Neo et al., "A review on chemometric techniques with infrared, Raman and laser-induced breakdown spectroscopy for sorting plastic waste in the recycling industry," May 01, 2022, *Elsevier B.V.* doi: 10.1016/j.resconrec.2022.106217.

[2] L. Roming et al., "Black plastic identification by hyperspectral imaging in mid-wave infrared," *Waste Management*, vol. 209, p. 115175, Jan. 2026, doi: 10.1016/j.wasman.2025.115175.

[3] O. Tamin et al., "A review of hyperspectral imaging-based plastic waste detection state-of-the-arts," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 3, pp. 3407–3419, Jun. 2023, doi: 10.11591/ijece.v13i3.pp3407-3419.

[4] S. Serranti et al., "Characterization of post-consumer polyolefin wastes by hyperspectral imaging for quality control in recycling processes," *Waste Management*, vol. 31, no. 11, pp. 2217–2227, Nov. 2011, doi: 10.1016/j.wasman.2011.06.007.

[5] G. Bonifazi et al., "A hierarchical classification approach for recognition of low-density (LDPE) and high-density polyethylene (HDPE) in mixed plastic waste based on short-wave infrared (SWIR) hyperspectral imaging," *Spectrochim. Acta A Mol. Biomol. Spectrosc.*, vol. 198, pp. 115–122, Jun. 2018, doi: 10.1016/j.saa.2018.03.006.

[6] J. Xia et al., "Convolutional neural network with near-infrared spectroscopy for plastic discrimination," *Environ. Chem. Lett.*, vol. 19, no. 5, pp. 3547–3555, Oct. 2021, doi: 10.1007/s10311-021-01240-9.

[7] J. Yang et al., "Application of mid-infrared spectroscopy in plastic waste discrimination: chemometric methods and exploration of their potential in hyperspectral applications," *Measurement*, vol. 257, p. 118657, Jan. 2026, doi: 10.1016/j.measurement.2025.118657.

[8] A. Kirillov *et al.*, "Segment Anything," Apr. 2023.

[9] A. Sbrana et al., "Plastic Classification With NIR Hyperspectral Images and Deep Learning," *IEEE Sens. Lett.*, vol. 7, no. 1, pp. 1–4, Jan. 2023, doi: 10.1109/LSENS.2023.3234401.

[10] G. Bonifazi et al., "Black Plastic Waste Classification by Laser-Induced Fluorescence Technique Combined with Machine Learning Approaches," *Waste Biomass Valorization*, 2023, doi: 10.1007/s12649-023-02146-z.

[11] F. Gruber et al., "Classification of black plastics waste using fluorescence imaging and machine learning," *Recycling*, vol. 4, no. 4, 2019, doi: 10.3390/recycling4040040.