

# High-throughput microplastic screening system using multispectral fluorescence imaging under UV excitation

Asuka Sawamoto<sup>a</sup> and Shigeki Nakauchi<sup>b</sup>

<sup>a,b</sup>Department of Computer Science and Engineering, Toyohashi University of Technology, 1-1 Hibarigaoka, Tempaku-cho, Toyohashi, Aichi, Japan 441-8580

## ABSTRACT

The widespread contamination of microplastics (MPs) necessitates comprehensive monitoring across various environments. However, standard analytical methods, such as FTIR and Raman spectroscopy, are time-consuming and resource-intensive, limiting their utility for high-throughput screening. Here, we propose a rapid multispectral imaging method using polymer auto-fluorescence under UV excitation. Leveraging the distinct fingerprints observed in Excitation-Emission Matrices (EEMs), our approach optimizes spectral acquisition by selecting a minimal subset of effective spectral bands rather than measuring the full sets of spectra in the EEM. Using a Random Forest-based feature selection, we demonstrated that only 8 spectral bands are sufficient to classify 9 representative virgin polymers with an average accuracy of 89%. This significant reduction in spectral dimensionality facilitates the design of simplified hardware using commercially available UV-LEDs, offering a practical and accessible solution for high-throughput MP monitoring.

**Keywords:** Microplastics, Multispectral fluorescence imaging, UV excitation, High-throughput screening, Spectral optimization, Excitation-emission matrix

## 1. INTRODUCTION

The widespread accumulation of microplastics (MPs) in the natural environment is an increasingly serious international issue [1, 2]. Understanding MP pollution sources and dynamics requires comprehensively quantifying their distribution and flux across various media [3]. Specifically, tracking pollution sources and environmental behavior, identifying the specific material type, such as polyethylene (PE) from packaging vs. polyethylene terephthalate (PET) from bottles, of every single sample collected in field surveys is essential [4].

However, while conventional analytical methods such as Fourier Transform Infrared Spectroscopy (FTIR) and Raman spectroscopy excel at identifying the chemical structure of MPs, they rely on point-scanning, which requires an enormous amount of time for analysis [5, 6]. Consequently, comprehensive chemical profiling of the vast amounts of MPs in the environment is virtually impossible within a realistic timeframe. Therefore, for in-process applications, there is a strong demand for high-throughput solutions that achieve both practical classification accuracy and rapid analysis speeds [7].

Excitation-Emission Matrix (EEM) spectroscopy enables the non-destructive characterization of chemical compositions. When coupled with supervised machine learning algorithms, it facilitates the robust classification of complex environmental samples [8]. Furthermore, our research group has extended this approach to fluorescence imaging, demonstrating that EEM analysis can be effectively utilized to visualize the spatial distribution of chemical properties in biological samples [9]. In this study, we propose a multispectral imaging method specialized for polymer identification based on the Excitation-Emission Matrix (EEM). By utilizing inherent polymer auto-fluorescence under UV excitation, we capture essential spectral features as images. To begin with, toward the future construction of a high-throughput system, we conducted a feasibility study using a benchtop system equipped with a broadband light source. Furthermore, we identified a combination of spectral bands that minimizes measurement time while maintaining classification accuracy.

---

Further author information: (Send correspondence to A.S.)

A.S.: E-mail: sawamoto.asuka.us@tut.jp, Telephone: +81 90 7647 0330

## 2. PRINCIPLE

### 2.1 Fluorescence Characteristics

Most petroleum-based synthetic polymers emit specific auto-fluorescence under ultraviolet (UV) excitation due to their chemical structures, such as electron conjugated systems, and additives [10]. This fluorescence characteristic depends on the Excitation-Emission Matrix (EEM) and serves as a unique fingerprint for each polymer type. In this study, we utilize these EEM characteristics as the physical basis for MP classification. Generally, the fluorescence intensity  $F$  at an excitation wavelength  $\lambda_{ex}$  and an emission wavelength  $\lambda_{em}$  is described by the following simplified model:

$$F(\lambda_{ex}, \lambda_{em}) = K \cdot I_0(\lambda_{ex}) \cdot (1 - 10^{-\varepsilon(\lambda_{ex})cL}) \cdot \Phi(\lambda_{ex}, \lambda_{em}), \quad (1)$$

where  $K$  is the instrument constant (geometric factors and detection efficiency),  $I_0$  is the excitation light intensity,  $\varepsilon$  is the molar absorption coefficient,  $c$  is the concentration,  $L$  is the optical path length, and  $\Phi$  is the fluorescence quantum yield. Under the dilute solution or thin film approximation, Equation (1) becomes linear with respect to the material-intrinsic parameters  $\varepsilon$  and  $\Phi$ . Since the spectral profiles of  $\varepsilon$  (absorption characteristics) and  $\Phi$  (emission efficiency) differ depending on the polymer type, measuring these at multiple wavelengths enables material identification.

As a preliminary experiment, we measured the EEMs of the nine target polymers used in this study (HDPE, LDPE, PP, PS, PVC, PC, PMMA, PET, and ABS) using a fluorescence spectrophotometer (F-7000, Hitachi High-Tech). The obtained EEM spectra are shown in Fig 1. These results confirmed that each polymer possesses a specific fluorescence spectrum under UV excitation.

### 2.2 Data-Driven Band Selection

While high-dimensional data such as EEMs are essential for accurate polymer classification, acquiring and processing the full hyperspectral datacube is inefficient for real-time applications due to the massive data volume and spectral redundancy. Therefore, an approach is needed to selectively acquire only the minimal spectral bands that have a high contribution to discrimination. However, since spectral data possess strong correlations between bands and contain non-linear relationships, it is difficult to determine the optimal subset using simple linear models. In this study, we employed Random Forest (RF), which can account for non-linear interactions between features. In RF, the importance (Gini Importance) of a feature (spectral bands)  $X_j$  is defined as the sum of the decrease in impurity at the nodes where that feature was used for splitting, as shown in the following equation:

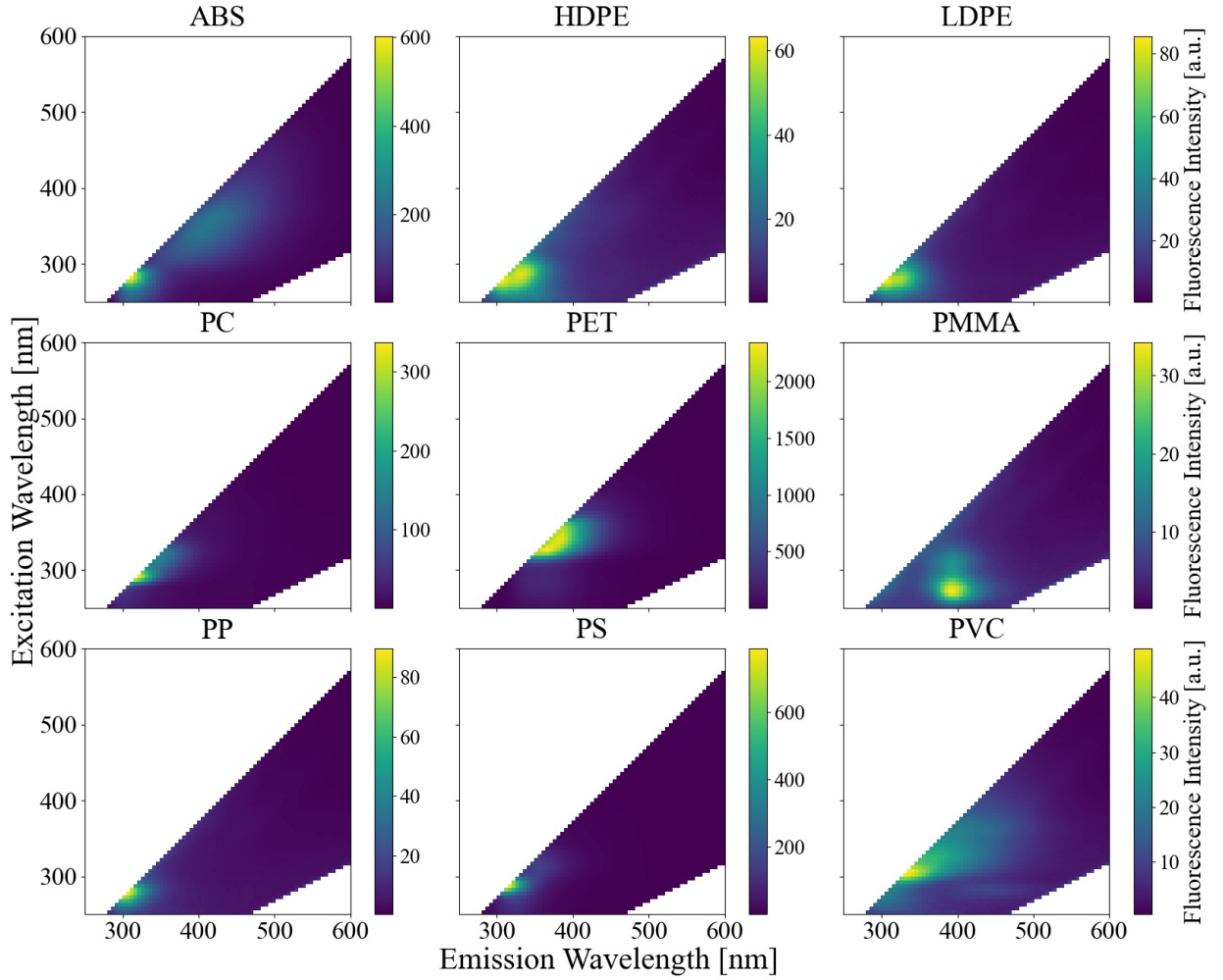
$$I_G(X_j) = \frac{1}{N_T} \sum_T \sum_{t \in T: v(s_t)=X_j} p(t) \Delta i(s_t, t), \quad (2)$$

where  $N_T$  is the total number of trees constituting the forest,  $v(s_t)$  is the variable used for splitting at node  $t$ ,  $p(t)$  is the proportion of samples reaching node  $t$ , and  $\Delta i(s_t, t)$  indicates the Gini Impurity Decrease due to the split. Based on this  $I_G(X_j)$ , we applied a step-wise method in which bands with high importance are sequentially added and evaluated. This allowed us to determine the optimal band configuration that minimizes imaging time while maintaining classification accuracy.

## 3. EXPERIMENTAL SETUP

### 3.1 Optical Configuration

To acquire comprehensive data for spectral optimization, we used a benchtop evaluation system equipped with a broadband light source [9]. The schematic diagram of the experimental apparatus is shown in Fig 2. We used a tunable light source (MAX-303, Asahi Spectra Co., Ltd.) equipped with a broadband xenon lamp. We selected the excitation wavelength via a rotating filter wheel and guided it to the sample through an optical fiber. The fluorescence emitted from the sample passed through an emission-side bandpass filter to remove scattered excitation light, and a UV-sensitive camera (BU-56DUV, BITRAN CORPORATION) captured the signal.

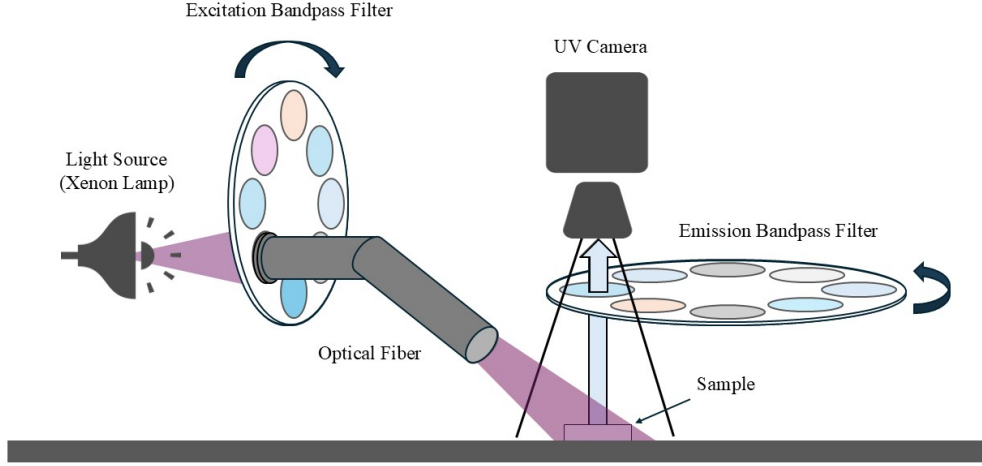


**Figure 1:** Measured Excitation-Emission Matrices (EEMs) of the nine types of virgin polymer pellets used in this study. The contour plots represent the fluorescence intensity [a.u.] at excitation (y-axis) and emission (x-axis) wavelengths [nm] for each polymer sample. The distinct spectral patterns (fingerprints) observed for each polymer type demonstrate the feasibility of the proposed classification method based on UV-excited fluorescence. Note: The diagonal regions corresponding to Rayleigh and Raman scattering have been computationally removed to enhance the visibility of fluorescence features.

To construct a comprehensive hyperspectral dataset, we combined excitation bandpass filters (center wavelengths: 260–380 nm) and emission bandpass filters (280–600 nm) at 20 nm intervals. By selecting valid combinations satisfying the Stokes shift (fluorescence wavelength > excitation wavelength), we acquired a total of 98 spectral images.

### 3.2 Samples and Data Acquisition

We used nine types of virgin polymer pellets (HDPE, LDPE, PP, PS, PVC, PC, PMMA, PET, and ABS), which account for the majority of global plastic production, as experimental samples. The size of each pellet is approximately 3–5 mm (Fig 3(a)). While these are transparent or white under visible light and difficult to distinguish its type, they exhibit characteristic fluorescence intensity differences under UV excitation (Fig 3(b), (c)). To improve data acquisition efficiency, we arranged the nine types of pellets within the same field of view during imaging. We acquired the datasets used for model calibration and evaluation in independent imaging sessions with different geometric arrangements:

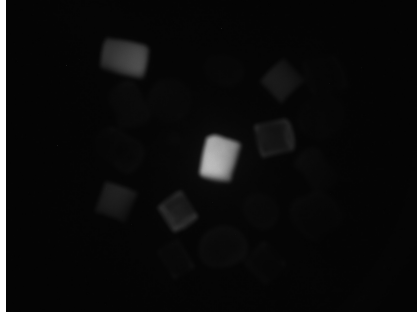


**Figure 2:** Schematic diagram of the benchtop evaluation system used for hyperspectral fluorescent image acquisition. A broadband xenon lamp and interchangeable bandpass filters were used to excite the samples, and the emitted fluorescence was captured by a UV-sensitive camera to construct a comprehensive dataset for spectral optimization.

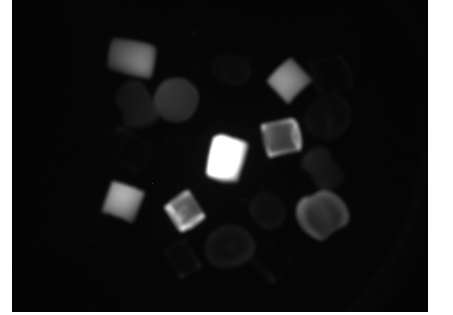
- **Calibration Set:** To cover spectral variations depending on the arrangement state and construct a robust model, we integrated images acquired in the following two configurations:
  1. *Adjacent Arrangement:* Pellets of the same type are arranged to touch and cluster together.
  2. *Dispersed Arrangement:* Pellets are randomly dispersed so that they do not touch each other.
- **External Validation Set:** A dataset used solely for the final performance evaluation of the model, independent of the calibration process. It consists only of images with the *Dispersed Arrangement*. This verified the pure identification capability for unknown data.



(a) Visible Light



(b) Ex 320 / Em 400



(c) Ex 360 / Em 480

**Figure 3:** Images of the experimental samples consisting of nine types of virgin polymer pellets. (a) Visible light image; most samples appear transparent or white. (b, c) Representative fluorescence images acquired at (b) Ex 320 nm / Em 400 nm and (c) Ex 360 nm / Em 480 nm. Note: The sample arrangement differs between visible and fluorescence images. Despite the random arrangement, characteristic fluorescence intensity differences are clearly observed.

## 4. METHOD

In this study, we propose a framework for high-throughput and accurate identification by selecting optimal wavelengths from acquired hyperspectral data. The process consists of three main stages: (1) Problem Formulation, (2) Pre-processing, and (3) Feature Selection.

## 4.1 Problem Formulation

Let the spectral intensity vector at each pixel be  $\mathbf{x} \in \mathbb{R}^D$  (where  $D = 98$  corresponds to the number of acquired spectral images described in Section 3.1). Let the class label of the target polymer be  $y \in \{C_1, \dots, C_9\}$ . The objective of this study is to learn a mapping function  $f: \mathbf{x} \rightarrow y$  that predicts the ground-truth label  $y$  from the input  $\mathbf{x}$ .

To verify the effectiveness of the proposed method, we define two experimental settings:

- **Baseline (Full-Spectrum):** A model that uses the entire wavelength band ( $D = 98$ ) as input features. This serves as the theoretical upper limit of accuracy.
- **Ours (Optimized-Band):** A model that uses only  $k$  important bands selected by the algorithm described below ( $k \ll 98$ ).

## 4.2 Pre-processing

The pixel intensity of the acquired fluorescence images strongly depends on the geometric shape of the sample, the distance from the light source, and the non-uniformity of illumination. These intensity variations hinder accurate polymer identification. Therefore, in this study, to achieve robust identification based solely on the polymer-specific spectral profile rather than absolute intensity, we applied Minimum Value Subtraction and L2 Normalization to the spectral data vector  $\mathbf{x}$  of each pixel:

$$\mathbf{x}' = \frac{\mathbf{x} - \min(\mathbf{x})}{\|\mathbf{x} - \min(\mathbf{x})\|_2}. \quad (3)$$

This process generates a normalized feature vector  $\mathbf{x}'$  that is independent of lighting conditions and sample arrangement.

## 4.3 Spectral Feature Selection Strategy

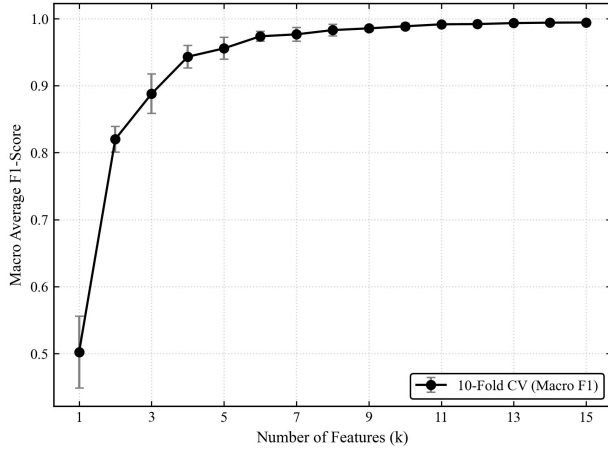
To determine the optimal subset of spectral bands, we implemented a step-wise approach based on the Random Forest Gini Importance described in Section 2.2. The specific procedure is as follows:

1. **Ranking and Stability Analysis:** Instead of a single calculation, we conducted a stability analysis using a 10-fold cross-validation scheme on the calibration set. In each fold, we calculated the feature importance  $I_G(X_j)$  and recorded the top-ranked bands. We defined the selection frequency as the number of times a band appeared in the top 10 across the 10 folds. The final ranking was determined based on this frequency to ensure robustness against data variations.
2. **Iterative Evaluation:** We sequentially added spectral bands to the model in descending order of the determined robust ranking. At each step, with the number of features  $k$  ranging from 1 to 10, we evaluated the discrimination performance using the Weighted Average F1-Score.
3. **Determination of Optimal Subset:** We adopted the Weighted Average F1-Score of all 9 classes as the evaluation metric. The optimal number of bands  $k$  was determined by identifying the point where the accuracy saturated in the internal validation and reached its peak in the external validation. This approach minimizes the number of required images while maximizing the generalization capability.

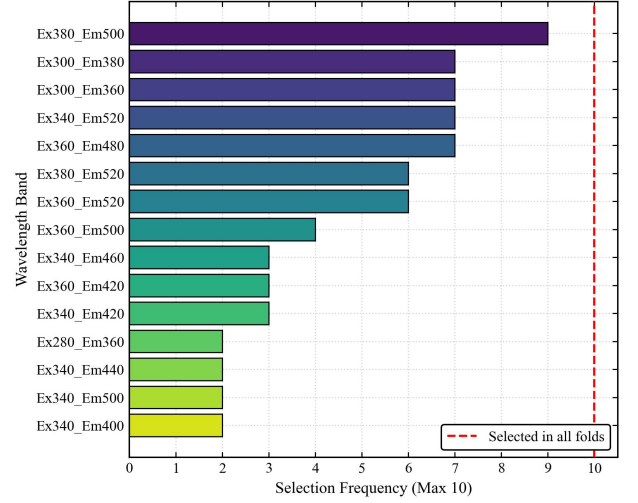
# 5. RESULTS

## 5.1 Optimization of Spectral Bands

The results of feature selection by the step-wise method using the calibration dataset are shown in Fig 4(a). As the number of spectral bands  $k$  increased, the Weighted Average F1-Score for all 9 classes rose rapidly, exceeding 0.95 at  $k = 5$ . The score improvement subsequently plateaued, and, with the accuracy reached a saturation level (approx. 0.99) around  $k = 11$ . Fig 4(b) shows the selection frequency of each spectral band obtained through 10-fold cross-validation. We observed that the top five bands were selected in more than 7 out of 10 folds.



(a) F1-Score vs. Number of bands  $k$ .



(b) Band selection frequency.

**Figure 4:** Results of the feature selection. (a) Dependence of the Weighted Average F1-Score (y-axis) on the number of spectral bands  $k$  (x-axis). Error bars represent standard deviation. (b) Selection frequency (x-axis) of each spectral band (y-axis) during 10-fold cross-validation.

## 5.2 Generalization Performance on External Validation Set

We validated the generalization capability of the selected subsets using the external validation dataset acquired independently from the training data. Fig 5 illustrates the Weighted Average F1-Score on the external set. Contrasting with the internal validation, the performance peaked at  $k = 5$  with an F1-score of 0.9171. Notably, this peak score slightly outperformed the baseline score of 0.916 obtained using the full spectrum of 98 bands. Furthermore, adding features beyond  $k = 5$  caused a decrease in accuracy, with the F1-score dropping to 0.880 at  $k = 10$ . Therefore, we concluded that  $k = 5$  is the optimal cutoff for this system.

To further investigate how the model distinguishes between polymers, we analyzed the class-wise performance. Fig 6 presents the F1-score transition for each of the 9 classes. We observed distinct behaviors depending on the material type:

- **Early Saturation:** Polymers such as ABS and PVC reached high accuracy exceeding 0.9 with only 1 band. Additionally, PC, PP, and PS demonstrated rapid convergence, reaching their baseline performance levels at  $k = 5$ .
- **Peak and Drop:** HDPE and PET showed a clear peak around  $k = 3$  to 5. Notably, for these materials, the F1-scores of the selected subset were higher than the full-spectrum baseline.
- **Slow Saturation:** PMMA and LDPE exhibited a slower convergence to the baseline accuracy compared to other materials.

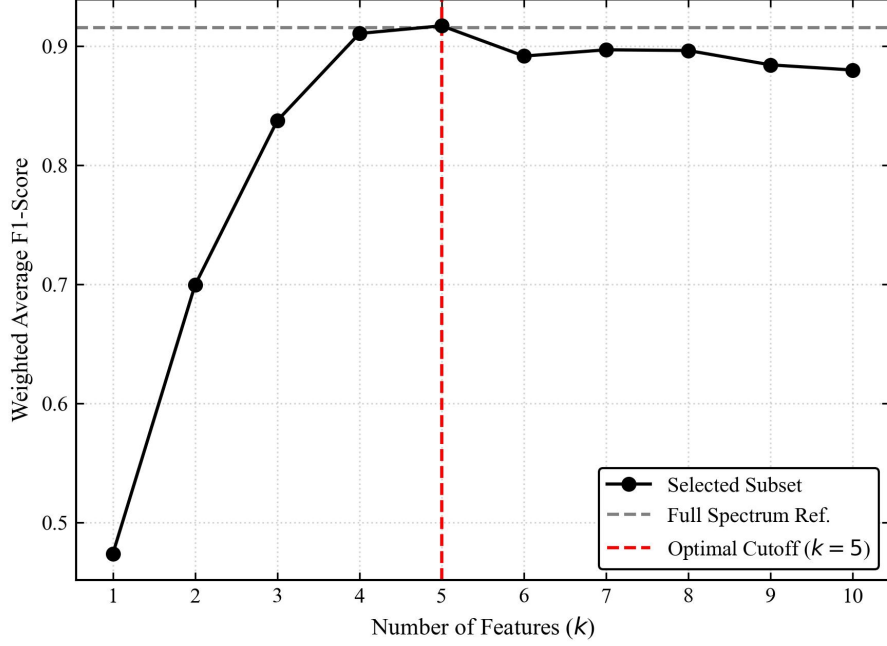
Fig 7 presents the confusion matrix obtained with the optimal subset of  $k = 5$ . The model achieved perfect classification (a recall of 1.00) for ABS, PS, and PVC. In contrast, mutual confusion was observed between PET and HDPE. Specifically, the model misclassified 35% of PET samples as HDPE, and 10% of HDPE samples as PET.

## 6. DISCUSSION

### 6.1 Interpretation of Optimization Behavior

In this study, we prioritized optimizing the overall average performance for high-throughput screening. The feature selection analysis, as shown in Fig ??(b), revealed that the top 5 spectral bands were selected in more





**Figure 5:** Generalization performance on the external validation dataset. Dependence of the Weighted Average F1-Score (y-axis) on the number of features  $k$  (x-axis). The black solid line shows the proposed method, the gray dashed line indicates the full-spectrum baseline (98 bands), and the red dashed line marks the optimal cutoff at  $k = 5$ .

than 7 out of 10 cross-validation folds. This high selection frequency demonstrates the robustness and stability of these features. While including additional bands beyond this optimal subset resulted in a performance decline for specific polymers such as HDPE and PET, the selected 5 bands successfully maximized the system-wide weighted average F1-score. This confirms that identifying a compact and stable feature subset is a rational strategy to ensure generalization capability.

## 6.2 High-Throughput Capability

The greatest advantage of this method is the overwhelming throughput achieved by area-scan imaging. While conventional point-scanning methods require time proportional to the number of measurement points, our method acquires spectral information for the entire field of view simultaneously. The experimental results demonstrate a speedup of several orders of magnitude compared to conventional methods.

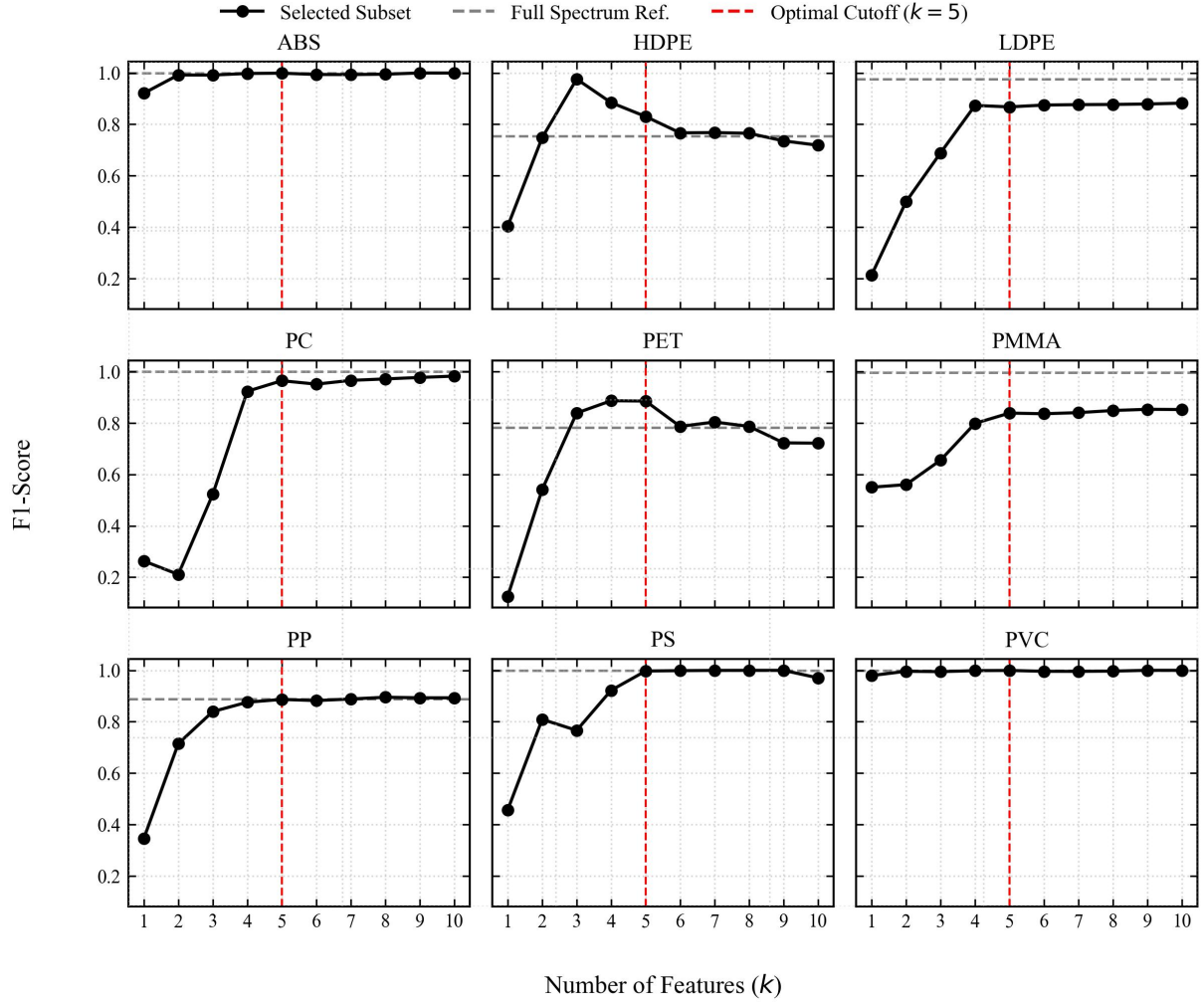
## 6.3 Limitations and Potential Improvements

There are several limitations to this study.

First, we observed some confusion between PET and HDPE (PET misclassification rate: approx. 35%). Analysis of the feature space using t-SNE revealed that while the clusters for PET and HDPE are spatially separated (indicating sufficient spectral distinguishability), a distinct distribution shift occurred between the calibration and validation datasets. Despite the inherent separability of the features, the Random Forest model misclassified a significant portion of PET validation samples as HDPE. This suggests that the model’s static decision boundaries could not robustly handle the observed distributional discrepancy between the datasets. Future work will focus on methods to compensate for such dataset shifts to ensure consistent classification.

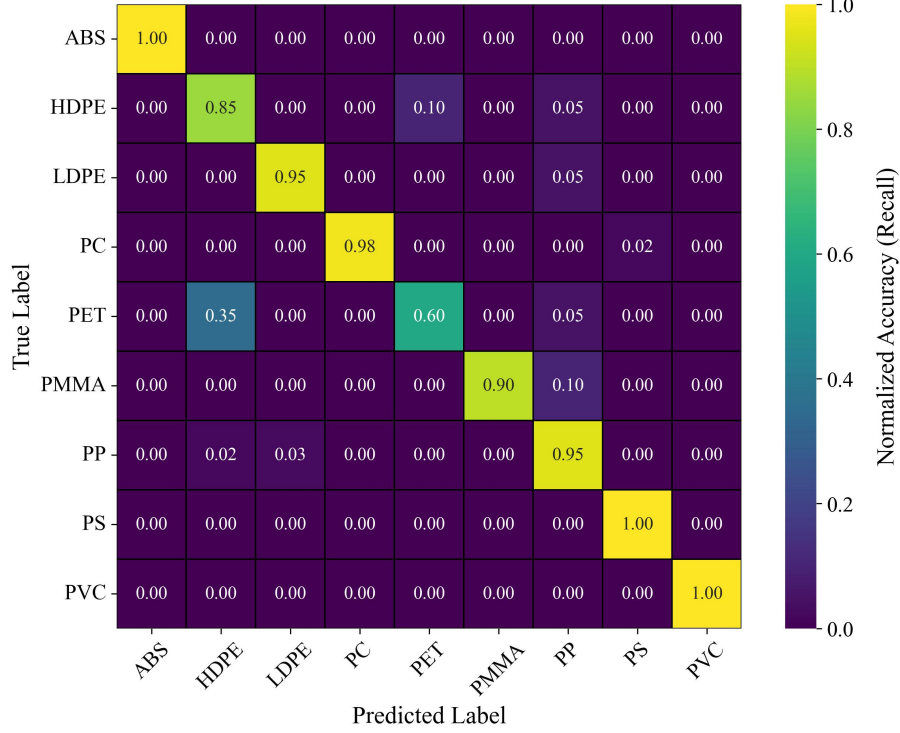
Second is the scope of direct accuracy comparison. Since this study focuses on throughput improvement, we did not perform strict particle-to-particle comparison with conventional methods.

Third, we conducted this experiment using virgin pellets. Since microplastics in actual marine environments may have altered fluorescence characteristics due to biofilm adhesion or weathering, further verification using real environmental samples is necessary.



**Figure 6:** Class-wise classification performance. F1-Score (y-axis) versus the number of features  $k$  (x-axis) for each material. The vertical red dashed line indicates the system-wide optimal cutoff ( $k = 5$ ), and the gray dashed line represents the full-spectrum baseline.





**Figure 7:** Normalized confusion matrix evaluated on the external validation dataset ( $k = 5$ ). The diagonal elements represent the classification accuracy (Recall) for each class. A distinct confusion is observed between HDPE and PET (True PET is misclassified as HDPE in 35% of cases), highlighting the spectral similarity between these polymers.

#### 6.4 System Design for Field Deployment

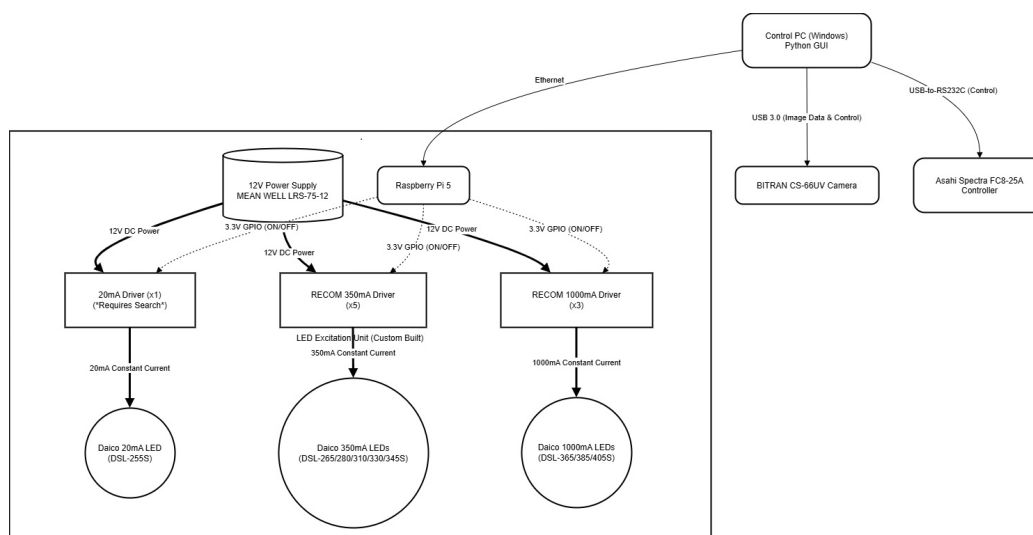
We conducted this study as a proof-of-principle for the future construction of a high-throughput system. Based on the verification results, we designed a compact dedicated system for on-site implementation. Fig 8 shows the block diagram of the proposed system. To replace the large benchtop equipment, we adopted high-power UV-LEDs as the light source. However, commercially available LEDs have a wider Full Width at Half Maximum (FWHM) compared to bandpass filters, and the selectable center wavelengths are discrete. Therefore, in this design, we selected a group of commercially available LEDs that approximately cover the spectral regions suggested to be effective in the benchtop experiment. In the future, we plan to perform re-optimization of features based on actual hardware characteristics using this compact system. We will adopt a single-board computer (SBC, e.g., Raspberry Pi 5) for the control system, aiming to realize a low-cost, standalone monitoring solution.

### 7. CONCLUSION

In this study, we demonstrated the feasibility of a high-throughput microplastic identification method using UV excitation fluorescence imaging. Through band selection using the step-wise method, we revealed that an average classification accuracy of 89% can be obtained across nine representative plastic types with only 8 spectral bands. The proposed method achieves significantly higher throughput compared to conventional methods and demonstrates high industrial practicality. Based on these findings, we also proposed a design for a compact standalone system using UV-LEDs. In the future, we will focus on the construction of this dedicated hardware and demonstration experiments using real marine samples.

### REFERENCES

- [1] Hale, R. C., Seeley, M. E., La Guardia, M. J., Mai, L., and Zeng, E. Y., “A global perspective on microplastics,” *J. Geophys. Res. Oceans* **125** (Jan. 2020).



**Figure 8:** Block diagram of the proposed compact dedicated system. High-power UV-LEDs and an SBC are used to achieve a standalone, low-cost solution.

- [2] Jambeck, J. R., Geyer, R., Wilcox, C., Siegler, T. R., Perryman, M., Andrady, A., Narayan, R., and Law, K. L., "Marine pollution. plastic waste inputs from land into the ocean," *Science* **347**, 768–771 (Feb. 2015).
- [3] Rocha-Santos, T. and Duarte, A. C., "A critical overview of the analytical approaches to the occurrence, the fate and the behavior of microplastics in the environment," *Trends Analyt. Chem.* **65**, 47–53 (Feb. 2015).
- [4] Hidalgo-Ruz, V., Gutow, L., Thompson, R. C., and Thiel, M., "Microplastics in the marine environment: a review of the methods used for identification and quantification," *Environ. Sci. Technol.* **46**, 3060–3075 (Mar. 2012).
- [5] Prata, J. C., da Costa, J. P., Duarte, A. C., and Rocha-Santos, T., "Methods for sampling and detection of microplastics in water and sediment: A critical review," *Trends Analyt. Chem.* **110**, 150–159 (Jan. 2019).
- [6] Shim, W. J., Hong, S. H., and Eo, S. E., "Identification methods in microplastic analysis: a review," *Anal. Methods* **9**(9), 1384–1391 (2017).
- [7] Serranti, S., Capobianco, G., Cucuzza, P., and Bonifazi, G., "Efficient microplastic identification by hyperspectral imaging: A comparative study of spatial resolutions, spectral ranges and classification models to define an optimal analytical protocol," *Sci. Total Environ.* **954**, 176630 (Dec. 2024).
- [8] Hall, G. J. and Kenny, J. E., "Estuarine water classification using EEM spectroscopy and PARAFAC-SIMCA," *Anal. Chim. Acta* **581**, 118–124 (Jan. 2007).
- [9] Bui, M. V., Rahman, M. M., Nakazawa, N., Okazaki, E., and Nakauchi, S., "Visualize the quality of frozen fish using fluorescence imaging aided with excitation-emission matrix," *Optics Express* **26**(18), 22954–22964 (2018).
- [10] Lakowicz, J. R., "Instrumentation for fluorescence spectroscopy," in [*Principles of Fluorescence Spectroscopy*], 27–61, Springer US, Boston, MA (2006).