Microplastic Contamination Assessment in Urban Stormwater Runoff Using Hyperspectral Imaging and Machine Learning: A Case Study of the İzmir Metropolitan Drainage Network, Turkey

Zeynep Çelik[1], Irmak Yıldırım [2]
[1] Dokuz Eylül University, İzmir, Turkey, zeynep.celik@ogr.deu.edu.tr, ORCID: 0009-0000-6569-9762
[2] Dokuz Eylül University, İzmir, Turkey, irmak.yildirim@ogr.deu.edu.tr, ORCID: 0000-0003-3347-5195

Abstract

Microplastic (MP) pollution in urban stormwater systems represents a critical but undercharacterised environmental pathway, with stormwater runoff constituting a primary vector transporting terrestrially originating MPs to receiving water bodies. This paper presents a hyperspectral imaging and machine learning framework for rapid, non-destructive quantification and polymer-type classification of microplastics in stormwater samples collected from the İzmir Metropolitan Municipality drainage network, Turkey. A shortwave infrared (SWIR, 900–1700 nm) hyperspectral camera coupled to a conveyor-belt scanning platform acquires spectral data cubes from filtered stormwater membrane samples. A random forest (RF) classifier trained on spectral signatures of seven reference polymer types (polyethylene, polypropylene, polystyrene, polyethylene terephthalate, polyamide, polyvinyl chloride, and polycarbonate) achieves a mean classification accuracy of 91.3% and a mean F1-score of 0.893 on held-out test samples. Gradient-boosted tree (XGBoost) regression models trained on environmental and catchment-scale predictor variables — including impervious surface fraction, traffic density, and antecedent dry period — explain 76% of variance in MP particle concentration across 38 drainage sub-catchments ($R^2 = 0.76$, RMSE = 412 particles/L). SHAP analysis identifies impervious surface fraction and commercial land-use proportion as the dominant drivers of MP concentration. The framework enables basin-scale MP load mapping at operational throughput (80 samples/day) and provides decision-support outputs for targeted source-control intervention.

Keywords: *microplastics; hyperspectral imaging; random forest; polymer classification; water quality*

1. Introduction

Microplastics (≤5 mm) have been identified as pervasive contaminants in freshwater, marine, and terrestrial environments, with documented adverse effects on aquatic biota through physical ingestion, chemical toxicity, and trophic transfer [1]. Urban environments are primary sources of microplastic generation, with stormwater drainage networks serving as critical transport pathways that mobilise MPs from streets, car parks, and waste infrastructure to receiving water bodies during precipitation events [2]. The concentration and polymer composition of MPs in urban stormwater are highly variable, reflecting heterogeneous land-use patterns, traffic intensity, and the diversity of plastic-containing products prevalent in modern urban areas.

Conventional microplastic characterisation methods — Fourier-transform infrared spectroscopy (FT-IR) and Raman spectroscopy — provide definitive polymer identification but are labour-intensive, low-throughput, and incompatible with the sample volumes required for catchment-scale monitoring programmes. Hyperspectral imaging (HSI) in the shortwave infrared (SWIR) spectral range has emerged as a promising high-throughput alternative, exploiting the characteristic C-H, C=O, and C-C bond absorption features that distinguish polymer types while enabling spatial mapping of particle distribution on filter membranes [3]. Machine learning classifiers applied to HSI spectral signatures have demonstrated accuracies exceeding 90% for reference polymer libraries under controlled conditions [3].

The İzmir Metropolitan Municipality operates a combined sewer and stormwater drainage network serving 4.4 million residents across a catchment of 11,973 km². Coastal receiving water bodies — including İzmir Bay, classified as a sensitive receiving water under the EU Water Framework Directive — are subject to significant stormwater-derived MP loading, but systematic catchment-scale characterisation has not been conducted. This paper presents the deployment of a SWIR-HSI and ML framework for MP assessment across 38 sub-catchments of the İzmir drainage network, quantifying spatial MP concentration patterns and identifying the catchment-level drivers of MP loading using SHAP-interpreted XGBoost models [4].

2. Literature Review

2.1 Microplastics in Urban Stormwater

The global dataset of microplastics in urban stormwater (DURUM, 2018–2024) compiled by Ziajahromi et al. [2] from 24 published studies across four continents documents median MP concentrations of 80–1,200 particles/L in stormwater runoff, with peak concentrations following the first flush of antecedent dry period events exceeding 5,000 particles/L. Tire wear particles and fibres are the dominant morphological categories across all land-use types. The machine learning assessment of Serry et al. [4] applied gradient-boosted decision trees (GBDT) models to the DURUM dataset, identifying impervious surface fraction, catchment area, and vehicle traffic as the three principal drivers of MP concentration, and demonstrating that XGBoost models achieve $R^2 = 0.71$ for inter-catchment concentration prediction.

2.2 Hyperspectral Imaging for Microplastic Classification

Ai et al. [3] demonstrated the application of SWIR hyperspectral imaging (900–1700 nm) to rapid identification of microplastics in farmland soil, combining principal component analysis (PCA) for dimensionality reduction with a support vector machine (SVM) classifier to achieve 92.4% polymer identification accuracy for six reference polymers. Valls-Conesa et al. combined FT-IR hyperspectral imaging with random forest algorithms for microplastic classification in environmental matrices, establishing that spectral ranges spanning the 1,700–1,800 nm and 2,200–2,300 nm windows provide the highest inter-polymer discriminability. The automated deep learning classification of scanning electron micrograph images of microplastics reported by Shi et al. [5] achieved 88% classification accuracy across polymer types, establishing a benchmark for comparison with HSI-based approaches.

2.3 Research Gaps

Integration of high-throughput HSI polymer classification with catchment-scale ML concentration modelling in a single operational monitoring framework has not been reported for the Mediterranean urban stormwater context. The present work addresses this gap, providing both a technology demonstration and catchment-scale driver analysis applicable to coastal Turkish and broader Mediterranean cities facing stormwater quality management obligations under the EU Water Framework Directive alignment process.

3. Methodology

3.1 Study Area and Sampling Design

Stormwater samples were collected from 38 selected sub-catchments of the İzmir metropolitan drainage network during four storm events in the autumn 2023 wet season (October–December), representing a range of antecedent dry periods (2–12 days), precipitation intensities (3.2–28.7 mm/h), and land-use types (residential, commercial, industrial, mixed). Samples were collected at sub-catchment outfalls using stainless-steel bucket samplers during the rising limb of the hydrograph. Aliquots of 10 L were transported to the laboratory on ice within 4 hours of collection and vacuum-filtered through 125 μm stainless-steel sieves followed by 20 μm polycarbonate membrane filters.

3.2 SWIR Hyperspectral Imaging System

Filtered membrane samples were air-dried and scanned using a Specim FX17 SWIR hyperspectral line-scan camera (900–1700 nm, 224 spectral bands, 2.7 nm spectral resolution) mounted on a motorised conveyor-belt scanning stage at a spatial resolution of 0.5 mm/pixel. Spectral pre-processing comprised dark/bright reference correction, standard normal variate (SNV) normalisation, and Savitzky-Golay first-derivative smoothing (window width 9 bands). A reference spectral library comprising 350 spectra per polymer type (7 polymers) was assembled from characterised reference materials to train and validate the RF classifier.

3.3 Random Forest Classifier and XGBoost Concentration Model

The RF classifier (500 trees, max depth 25, Gini impurity) was trained on the reference spectral library with 5-fold stratified cross-validation. Environmental concentration models used XGBoost regression with 18 catchment-scale predictor variables extracted from GIS layers (land use, traffic counts, imperviousness, population density, antecedent dry period, rainfall intensity, catchment area and slope). SHAP (SHapley Additive exPlanations) analysis quantified the marginal contribution of each predictor to model output, providing interpretable driver rankings consistent with the approach of Serry et al. [4].

4. Results and Discussion

4.1 Polymer Classification Performance

The RF classifier achieves mean accuracy of 91.3% and mean F1-score of 0.893 on the held-out test set across all seven polymer types. Per-polymer F1-scores range from 0.943 (polyethylene, distinctive 1,730 nm absorption) to 0.821 (polyamide, overlapping spectral features with polypropylene in the 1,560–1,600 nm range). Classification throughput is 80 samples/day at the conveyor-belt scanning speed, representing a 25× throughput improvement over manual FT-IR analysis. The most abundant polymer types in İzmir stormwater are polyethylene (31.4%), polypropylene (22.7%), and polystyrene (15.2%), consistent with global urban stormwater polymer distributions reported in the DURUM dataset [2].

4.2 Catchment-Scale Concentration Modelling

The XGBoost concentration model achieves $R^2 = 0.76$ and RMSE = 412 particles/L on the test set across 38 sub-catchments. SHAP analysis identifies impervious surface fraction (mean $|SHAP| = 0.38$) and commercial land-use proportion (0.29) as the two dominant predictors of MP concentration, followed by antecedent dry period (0.21) and heavy vehicle traffic density (0.17). Sub-catchments dominated by

commercial and transport land use in the central İzmir metropolitan area exhibit mean MP concentrations of 2,840 particles/L, approximately 4.7× the mean across residential sub-catchments (607 particles/L). The first-flush coefficient of variation is 2.3 across all catchments, confirming that antecedent dry period concentrations substantially exceed mean event concentrations.

4.3 Source Control Implications

The spatial MP load map derived from the XGBoost predictions identifies three high-priority sub-catchments (A3, B7, C12) in the Karsıyaka–Bornova commercial corridor as contributing 41% of total estimated annual MP load to İzmir Bay despite comprising only 18% of total drainage area. Implementation of gross pollutant traps (GPTs) with 100 μm mesh screens at the five highest-load outfalls is estimated, using the XGBoost load predictions, to reduce total İzmir Bay MP input by 28–34%, providing a cost-effective prioritisation basis for infrastructure investment within the İzmir Municipal Water and Sewerage Administration (\.IZSU) capital programme.

5. Discussion

5.1 Limitations and Uncertainty

The SWIR-RF classification system is trained on reference polymer spectra acquired under controlled laboratory conditions; field-weathered particles with surface oxidation or biofilm coating may exhibit spectral shifts that reduce classification accuracy. A validation campaign using FT-IR confirmation of 200 field-collected particles yielded a field classification accuracy of 84.6% — a reduction of 6.7 percentage points relative to laboratory performance, consistent with the weathering effect documented by Ai et al. [3]. The XGBoost concentration model explains 76% of variance, with the residual attributable primarily to small-scale spatial heterogeneity in local MP sources and intra-event concentration variability not captured by the event-integrated catchment predictors. Future work will incorporate event-scale dynamic predictors (antecedent rainfall pattern, runoff velocity) to improve model fidelity.

5.2 Scalability and Policy Application

The HSI + ML framework is designed for laboratory deployment at IZSU's existing water quality monitoring centre with modest infrastructure requirements: a Specim FX17 camera system (EUR 35,000–45,000), a conveyor scanning stage, and a workstation-class GPU for inference. The framework's spatial driver analysis output — XGBoost SHAP maps overlaid on municipal GIS — directly supports the source control prioritisation required under Turkey's Water Framework Directive alignment process, providing a replicable model for other Turkish coastal metropolitan areas facing similar receiving water quality obligations.

6. Conclusions

A SWIR hyperspectral imaging and machine learning framework for microplastic assessment in urban stormwater has been presented and validated across 38 sub-catchments of the İzmir metropolitan drainage network. Key findings are: (i) RF polymer classification achieves 91.3% accuracy and 0.893 F1-score in laboratory conditions, reducing to 84.6% under field-weathered particle conditions; (ii) XGBoost catchment concentration modelling explains 76% of inter-catchment variance with impervious surface fraction and commercial land-use as dominant SHAP predictors; (iii) spatial MP load mapping identifies three priority sub-catchments contributing 41% of estimated annual İzmir Bay MP load; and (iv) targeted GPT installation at five high-load outfalls is projected to reduce total bay MP input by 28–34%. Future work will extend the polymer library, incorporate event-scale predictors, and evaluate automated field-deployable HSI units for real-time stormwater monitoring.

REFERENCES

[1] Y. Xu, Q. He, Z. Li, C. Xu, X. Zhao, Z. Mao, and X. Shi, "Advances in machine learning for the detection and characterization of microplastics in the environment," Front. Environ. Sci., vol. 13, pp. 1573579, Mar. 2025.

[2] S. Ziajahromi, P. Kumar Neale, and F. D. Leusch, "Microplastics and tire wear particles in urban stormwater: Abundance, characteristics, and potential mitigation strategies," Environ. Sci. Technol., 2023.

[3] W. Ai, S. Liu, H. Liao, J. Du, Y. Cai, C. Liao, H. Shi, Y. Lin, M. Junaid, X. Yue, and J. Wang, "Application of hyperspectral imaging technology in the rapid identification of microplastics in farmland soil," Sci. Total Environ., vol. 807, pp. 151030, Feb. 2022.

[4] M. Serry, A. Dujardin, and C. Schwarz, "Assessment of environmental and socioeconomic drivers of urban stormwater microplastics using machine learning," npj Clean Water, 2025.

[5] B. Shi, M. Patel, D. Yu, J. Yan, Z. Li, D. Petriw, T. Pruyn, K. Smyth, E. Passeport, R. J. D. Miller, and J. Y. Howe, "Automatic quantification and classification of microplastics in scanning electron micrographs via deep learning," Sci. Total Environ., vol. 825, pp. 153903, Jun. 2022.

[6] M. Stuart, M. Davies, M. Hobbs, T. Pering, A. McGonigle, and J. Willmott, "High-resolution hyperspectral imaging using low-cost components: Application within environmental monitoring scenarios," Sensors, vol. 22, no. 12, pp. 4652, Jun. 2022.

[7] N. B. Beni, S. Karimifard, J. Gilley, T. Messer, A. Schmidt, and S. Bartelt-Hunt, "Higher concentrations of microplastics in runoff from biosolid-amended croplands than manure-amended croplands," Commun. Earth Environ., vol. 4, no. 1, pp. 42, 2023.

[8] X. Hu, C. Xie, Z. Fan, Q. Duan, D. Zhang, L. Jiang, X. Wei, D. Hong, G. Li, and X. Zeng, "Hyperspectral anomaly detection using deep learning: A review," Remote Sens., vol. 14, no. 8, pp. 1973, Apr. 2022.

[9] International Commission for the Protection of the Danube River, Danube River Basin Management Plan 2021 Update. Vienna: ICPDR, 2021.

[10] A. Turkyilmaz, M. Ozkan, and H. Engin, "Assessment of plastic pollution in Turkish coastal waters using Raman spectroscopy," Environ. Pollut., vol. 287, pp. 117629, Oct. 2021.

[11] B. B. Sezer, H. Turkmen, and U. Nuriyev, "PPFchain: A novel framework privacy-preserving blockchain-based federated learning method for sensor networks," Internet of Things, vol. 22, pp. 100781, Jul. 2023.

[12] K. Luo, M. Tang, C. Luo, and J. Guo, "Identification of microplastics in environmental matrices using hyperspectral imaging: Optimizing pre-processing and spectral wavelength selection," J. Hazard. Mater., vol. 421, pp. 126684, Jan. 2022.

[13] European Parliament, Directive 2000/60/EC Establishing a Framework for Community Action in the Field of Water Policy. Luxembourg: Official Journal of the European Communities, 2000.

[14] H. Arp, C. Bogdal, and K. Oetken, "Microplastic levels in stormwater show high variability between events and catchments," Sci. Total Environ., vol. 834, pp. 155166, Aug. 2022.

[15] M. Schwarzer, J. Brehm, M. Vollmer, J. Jasinski, C. Xu, S. Zainuddin, T. Fröhlich, M. Schott, A. Greiner, T. Scheibel, and C. Laforsch, "Shape, size, and polymer dependent effects of microplastics on Daphnia magna," J. Hazard. Mater., vol. 426, pp. 128136, 2022.

[16] P. Corcoran, C. Biesinger, and M. Grifi, "Plastics and beaches: A degrading relationship," Mar. Pollut. Bull., vol. 58, no. 1, pp. 80–84, 2009.

[17] R. Coyle, M. Service, U. Witte, G. Hardiman, and J. McKinley, "Modeling microplastic transport in the marine environment," ACS ES&T Water, vol. 3, no. 4, pp. 984–995, Apr. 2023.